

ABSTRACT

Title of dissertation: IMPROVING U.S. EXTREME
PRECIPITATION PREDICTION
AND PROCESS UNDERSTANDING
USING A MESOSCALE CLIMATE MODEL
MULTI-PHYSICS ENSEMBLE APPROACH

Chao Sun
Doctor of Philosophy, 2019

Dissertation directed by: Professor Xin-Zhong Liang
Department of Atmospheric and Oceanic Science

Despite many recent improvements, climate models continue to poorly simulate extreme precipitation. I attempted to improve prediction of extreme precipitation, focusing on daily 95th percentile (P95) events, and to better understand the source of model biases in three ways: 1) determine which physics processes P95 is most sensitive to and which parameterization schemes best represent these processes; 2) understand the underlying mechanisms through which these processes impact P95; and 3) maximize advantages from the ensemble of the best performing models.

First, to determine the sensitive processes affecting P95, I tested a 25-member ensemble of different physics configurations in the regional Climate-Weather Research and Forecasting model (CWRF) for 36-yr historical U.S. simulations. Of these, P95 simulation was most sensitive to cumulus parameterization. Overall, the ensemble cumulus parameterization best represented P95 seasonal mean spatial

patterns and interannual variations, while one traditional cumulus scheme generally overestimated P95 and the other three severely underestimated P95, especially over the Gulf States (GS) and the Central-Midwest States (CM) in convection-dominated seasons.

Second, I built structural equation models (SEMs) to identify the underlying processes through which cumulus parameterization affects precipitation. I discovered five distinct physical mechanisms, each involving unique interplays among water and energy supplies and surface and cloud forcings. The relative importance of these factors varied significantly by season and region. For example, water supply is the dominant factor for P95 in CM, but its effect reversed from positive in summer to negative in winter due to changes in the prevailing precipitation system. In contrast, the predominant factors affecting P95 in GS were cloud forcing in summer, but surface forcing in winter. Since the choice of cumulus parameterization affected how water and energy supplies acted through surface and cloud forcings, it determined CWRP's ability to simulate extreme precipitation.

Third, I improved P95 prediction by developing an optimized multi-model ensemble based on the Bayesian Model Averaging (BMA) approach. BMA is a model-selection method that weights ensemble members to create an optimal composite. However, many BMA methods rely on maximum likelihood estimation and thus may be flawed when the true solution is not among the ensemble, as is the case in extreme precipitation. To resolve this issue, I adapted three BMA variations to fit the needs of extreme precipitation problems. These methods significantly improved performance compared to both the ensemble mean and the single best model, and

provided a more reliable confidence interval.

My work shows that to improve extreme precipitation simulation, a better understanding of physics processes, especially cumulus processes, is critical. For this, I applied the SEM framework, for the first time in the climate community, to uncover the underlying physical mechanisms essential to regional extreme precipitation predictions. Furthermore, I adapted new BMA methods into extreme precipitation ensembles to maximize the benefits from the most physically advanced models. These advances may help improve the prediction of extreme precipitation occurrences and future changes, one of the most difficult modeling challenges and one with huge socioeconomic significance.

IMPROVING U.S. EXTREME
PRECIPITATION PREDICTION
AND PROCESS UNDERSTANDING
USING A MESOSCALE CLIMATE MODEL
MULTI-PHYSICS ENSEMBLE APPROACH

by

Chao Sun

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Xin-Zhong Liang, Chair/Advisor
Professor Da-Lin Zhang
Professor Michael N. Evans
Professor Russell R. Dickerson
Professor Raghu Murtugudde
Professor Wei-Kuo Tao

© Copyright by
Chao Sun
2019

Acknowledgments

First and foremost, I would like to thank my beloved advisor, Professor Xin-Zhong Liang for giving me an invaluable opportunity to work on challenging and exciting projects over the past years. He has always made himself available for help and advice. He even helped my project at 3 a.m. and ready to pick up my call whenever he is awake. Once I heard his wife urged him several times to eat, but he still worked with me on my project. He is a terrible workaholic, sometimes I really wonder whether he has time to sleep. His attitude towards research is ultra-strick, which showed me the right way to become a scholar. He edited and commented almost every word in my papers, and provided enormous detailed suggestions on how to write it. More important, whenever I was stuck in my project, he always can come up with a genius solution.

Furthermore, he also took great care of my life, supported me not only by scholarship but also by his wisdom and experience. Because of the protection from his big umbrella, I can focus on my studies peacefully for past years. Usually, he hosted several parties a year at his home. He and his family members prepared the most delicious food for us, which is really unforgettable. Sometimes, he even brought some fresh vegetables from his garden and let us enjoy them together. It has been a great pleasure to work with him and learn from such an extraordinary scientist.

I sincerely thank Jennifer Kennedy for careful editing. She is the most thoughtful and patient person I have ever seen on the planet of earth. She spent tremendous

time and effort in helping my writing and provided me many practical suggestions.

I owe my deepest thanks to my family. Words cannot express the gratitude I owe them.

The CWRF simulations and analyses were conducted on the supercomputers, including the University of Illinois' Blue Water, the Maryland Advanced Research Computing Center's Bluecrab, the Computational and Information Systems Lab of the National Center for Atmospheric Research, and the National Energy Research Scientific Computing Center of the U.S. Department of Energy. I thank Kenneth Kunkel for providing the COOP station data. The radiation data was obtained from the SRB/GEWEX product, courtesy of the Langley Research Center Atmospheric Science Data Center, USGS Earth Resources Observation and Science Center. The research was supported by U.S. National Science Foundation Innovations at the Nexus of Food, Energy and Water Systems under Grant EAR-1639327, 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.

Table of Contents

Acknowledgements	ii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
List of Abbreviations	xi
1 Introduction	1
1.1 Background and Motivation	1
1.2 Improving extreme precipitation simulation through a multiphysics approach	3
1.3 Improving extreme precipitation simulation by better physical understanding	5
1.4 Improving extreme precipitation simulation by optimal weightings	7
1.5 Organization	9
2 Sensitivity to physics parameterizations	10
2.1 Introduction	10
2.2 Model description, experiment design, observations, and extreme indices	14
2.3 General performance of seasonal mean and extreme precipitation	18
2.4 Physics sensitivity of regional extreme precipitation simulation	23
2.5 Summary and discussion	39
3 Dependence on cumulus parameterization and underlying mechanism	46
3.1 Introduction	46
3.2 Model description, experiment design, causal ingredients, and observations	50
3.3 Extreme precipitation dependence on cumulus parameterization	56
3.4 Correlation analysis of regional extreme precipitation biases	62
3.5 Process understanding of P95 biases by structural equation modeling	74
3.6 Summary and conclusions	88

4	Improvement by Markov Chain Monte Carlo based Bayesian model average	93
4.1	Introduction	93
4.2	Model description, observations	96
4.3	Definitions of extreme precipitation, OptiRankDSCV, MODE tool	99
4.4	Performance analysis of individual ensemble members	104
4.5	Bayesian model average methods	118
4.6	Ensemble performance analysis	120
4.7	Summary and conclusions	132
5	Future work: projections of future extreme precipitation changes and impacts	137
	Bibliography	140

List of Tables

2.1	Physical processes, parameterizations, and their references.	13
2.2	Ensemble experiment physics configurations.	15
3.1	Cumulus schemes closures, triggers, entrainments, and their references.	51
3.2	Observations available years and their references.	55
4.1	CPN experiment model configurations, parameterizations, and their references.	98
4.2	Extreme indicators, their definations and units.	100

List of Figures

2.1	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).	18
2.2	Same as Fig. 2.1 except for the number of rainy days (NRD).	20
2.3	Same as Fig. 2.1 except for the daily 95 th percentile precipitation (P95) [mm day^{-1}].	21
2.4	Boundary specification of the two key regions where ERI severely underestimated extreme precipitation: the Gulf States (GS) in spring and the Central to Midwest States (CM) in summer.	24
2.5	Comparison among ERI and all CWRf physics configurations in simulating 1980-2015 mean seasonal P95 [mm day^{-1}] biases (from observations) averaged over the GS (<i>left</i>) and CM (<i>right</i>) for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). They are separated by color into type I (<i>blue</i>) and type II (<i>red</i>) members depending on their cumulus schemes.	26
2.6	Geographic distributions of 1980-2015 mean daily 95 th 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).	29
2.7	Same as Fig. 2.6 except for the number of rainy days (NRD) biases.	31
2.8	Same as Fig. 2.6 except for the daily rainfall intensity (DRI) biases [mm day^{-1}].	32
2.9	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 GS (<i>left</i>) and CM (<i>right</i>) regions for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). Shown are the pattern correlation (<i>azimuthal</i>) and normalized standard deviation (<i>radius</i>) 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.	34

2.10	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 GS (<i>left</i>) and CM (<i>right</i>) 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.	36
2.11	Summer mean vertical potential temperature (θ , <i>star</i>) and water vapor (q_v , <i>curve</i>) tendency profiles among the five cumulus schemes (<i>color</i>) for 2004 as averaged over all the grids having rainfall greater than 50 [mm day ⁻¹] within the CM (<i>left</i>) and GS (<i>right</i>) 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.	42
3.1	The teleconnection patterns of the CM and GS regional mean precipitation interannual variations correlated with 850 hPa meridional wind distributions for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). Outlined separately for CM and GS is the core correlation area common to all seasons, where the V850 index is defined.	53
3.2	Seasonal P95 interannual variations during 1980-2015 averaged over the CM (<i>left</i>) and GS (<i>right</i>) regions for spring (MAM), summer (JJA), autumn (SON), and winter (DJF) [<i>from top down</i>] as observed (OBS), assimilated by NR2 and ERI, and simulated by CWRF using five cumulus schemes (ECP, NKF, TDK, NSAS, BMJ). Also shown (<i>bottom</i>) are the corresponding temporal correlations (COR, scaled upward on the <i>left</i>) and root mean square errors (RMSE, scaled downward on the <i>right</i>) with respect to observations during the whole period.	57
3.3	Seasonal mean RCT distributions averaged in the CM (<i>left</i>) and GS (<i>right</i>) regions according to total precipitation intensity bins at an interval of 5 [mm day ⁻¹] for spring (MAM), summer (JJA), autumn (SON), and winter (DJF) [<i>from top down</i>] as assimilated by NR2 and ERI, and simulated by five CWRF cumulus schemes (ECP, NKF, TDK, NSAS, BMJ). Marked with circles are the corresponding regional average 1980-2015 mean seasonal P95 values, while the respective observed values are depicted by the vertical lines.	61
3.4	Composite P95 bias (<i>blue</i>) and departure (<i>red</i>) correlations with those fields that had observational data (DRI, NRD, SWD, RET, OLR, CRE, CWP, T2m) as well as P95 departure correlations with rainfall components (PL, PC) in the CM (<i>left</i>) and GS (<i>right</i>) regions for spring (MAM), summer (JJA), autumn (SON), and winter (DJF) [<i>from top down</i>]. A star mark indicates the correlation is statistically significant. They are labeled with a number equal to the correlation coefficient times 100 if statistically significant.	65

3.5	Spring composite departure correlations across P95 and its all ingredient fields in the CM (<i>upper triangle</i>) and GS (<i>lower triangle</i>). They are coded each with a color and, if statistically significant, also labeled with a number equal to the correlation coefficient times 100. The diagonal represents the P95 correlation with each field named.	66
3.6	Same as Fig. 3.5 except for summer.	67
3.7	Same as Fig. 3.5 except for autumn.	68
3.8	Same as Fig. 3.5 except for winter.	69
3.9	The conceptual design of the experimental SEM for extreme precipitation (EP). The center oval represents the predictand (EP), while each outer oval defines one latent variable (LV) with a list in a brace of the designated manifest variables (MV) as the predictor candidates. There are four LVs, hypothetically representing energy supply (ES), water supply (WS), surface forcing (SF), and cloud forcing (CF). Each effect from one LV to another or to EP is expressed by an arrow for its direction and a coefficient along the line for its strength. The SEM consists of regression equations from these MVs through LVs to EP, as depicted at the lower right corner.	76
3.10	Spring finalist SEMs for CM (<i>upper</i>) and GS (<i>lower</i>). Each SEM panel includes its structure (<i>left</i>) with the active manifest and latent variables and their directional effect coefficients, its performance scores (<i>upper right corner</i>), and the relative importance of each latent variable's direct, indirect and total effects (<i>bottom</i>).	80
3.11	Same as Fig. 10 except for summer.	81
3.12	Same as Fig. 10 except for autumn.	82
3.13	Same as Fig. 10 except for winter.	83
4.1	Six subregions in the continental United States	101
4.2	Winter overall performance of 1989-2009 mean extreme indicators measured by multi-score in all six subregions. Color represents scores scaled by their range in each row. Horizontally, the locations of models represent the optimal rank aggregated from all 144 indicators, with the most <i>left</i> is the <i>best</i> performed.	105
4.3	Same as Fig4.2 except for spring.	107
4.4	Same as Fig4.2 except for summer.	109
4.5	Same as Fig4.2 except for autumn.	111
4.6	Geographic distributions of 1980-2009 mean winter P95 amount [mm day ⁻¹] (color) and results of MODE (grey) for observed (OBS), assimilated (ERI), and simulated by all CPN members. A grey area represents the 'interest' area identified by the MODE tool, with total interest score showing in the lower-left corner. The black lines represent the convex hull identified by MODE, which was used to calculate shape features.	113
4.7	Same as Fig4.6 except for spring.	114
4.8	Same as Fig4.6 except for summer.	115

4.9	Same as Fig4.6 except for autumn.	117
4.10	Geographic distributions of 2004-2009 (cross-validation) mean seasonal P95 amount [mm day ⁻¹] observed (OBS), assimilated (ERI) simulated by CWRF, composite ensemble mean (EnsMean), post-processed by bootstrapping Akaike method (B-Akaike), post-processed by Akaike method (Akaike), and post-processed by stacking method (stacking) for winter (DJF), spring (MAM), summer(JJA),and autumn(SON).	121
4.11	Geographic distributions of 2004-2009 (cross-validation) winter P95 confidence interval (CI) amount [mm day ⁻¹]. Estimated by bootstrapping observational data (OBS), estimated by bootstrapping ensembles (boot), estimated by applying bootstrapping Akaike method on ensembles (B-Akaike), estimated by applying Akaike method on ensembles (Akaike), and estimated by applying stacking method on ensembles (stacking). <i>left</i> lower (25%) CI; <i>left</i> upper (75%).	125
4.12	Same as Fig4.11 except for spring.	127
4.13	Same as Fig4.11 except for summer.	128
4.14	Same as Fig4.11 except for autumn.	130
4.15	Geographic distributions of 2004-2009 (cross-validation) IGN of BMA methods minus IGN of raw composite ensembles: bootstrapping Akaike method (B-Akaike), Akaike method (Akaike), and stacking method (stacking).	131
5.1	Geographic distributions of projected changes in extreme precipitation for different return levels.	138
5.2	Geographic distributions of projected changes in the number of rainy-days, with confidence interval estimated by the bootstrapping method.138	

List of Abbreviations

ACM	Asymmetric Convective Model
AIC	Akaike Information Criteria
AMS	American Meteorology Society
AOGCM	Atmosphere-Ocean General Circulation Models
AVHRR	Advanced Very High-Resolution Radiometer
BMA	Bayesian Model Averaging
BMJ	Betts-Miller-Janjic cumulus scheme
CAM	Community Atmosphere Model
CAPE	Convective Available Potential Energy
CCCMA	Canadian Centre for Climate Modelling and Analysis
CDD	Consecutive Dry Days
CDF	Cumulative Distribution Function
CF	Cloud Forcing
CFI	Comparative Fit Index
CI	Confidence Interval
CIN	Convective inhibition
CLASS	Canadian Land Surface Scheme
CM	Central to Midwest
CONUS	Contiguous United States
CORDEX	Coordinated Regional Climate Downscaling Experiment
CPN	CWRF Plus NA-CORDEX
CRCM	Canadian Regional Climate Model
CRE	cloud radiative effect
CSSP	Conjunctive Surface Subsurface Process Model
CWP	Cloud Water Path
CWRF	Climate Weather Research and Forecasting Model
DGM	Data Generating Model
DREAM	DiffeRential Evolution Adaptive Metropolis
DRI	Daily Rain Intensity
DSCV	Distance (rmse), Similarity (correlation), Consistency(LEPS) and Variability(MVI)
ECP	Ensemble Cumulus Parameterization
EM	Expectation-Maximization
ERI	ERA-Interim
ES	Energy Supply
ET	Evaporation Transpiration
ETS	Equitable Threat Score
FCH	Fraction of Cloud (high)
FCL	Fraction of Cloud (low)
FLG	Fu-Liou-Gu radiation transfer scheme

GEWEX	Global Energy and Water Exchanges
GS	Gulf States
GSFC	NASA Goddard Space Flight Center
HIRMAM	High Resolution Atmospheric Model
HPD	Highest Posterior Density
IGN	Ignorance skill score
LCL	Lifting Condensation Level
LEPS	Linear Error In Probability Space
LFC	Level of Free Convection
LOO	Leave-One-Out
LV	Latent Variables
MC	Moisture Convergency
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation
MODE	Method for Object-Based Diagnostic Evaluation
MOR	Morrison et al. two-moment microphysics scheme
MV	Manifest Variables
MVI	Model Variability Index
MYNN	Mellor–Yamada PBL scheme modified by Nakanishi–Niino
NARR	North American Regional Reanalysis
NASA	National Aeronautics and Space Administration
NCA	National Climate Assessment
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NKF	New Kain–Fritsch cumulus parameterization
NOAH	NCAR–NCEP unified land surface model
NRD/RD	Number of Rainy Days
NSAS	New Simplified Arakawa-Schubert scheme
NSE	Net Surface Energy
NUTS	No U-turn Sampler algorithm
OBS	Observation
OLR	Outgoing Longwave Radiation
P95	95th percentile precipitation
PBLH	Planetary Boundary Layer Height
PC	convective precipitation
PL	explicit precipitation
RB	Relative Biases
RCA	Rosby Centre regional atmospheric climate model
RCD	Regional Climate Downscaling
RCT	convective to total precipitation ratio

RET	reflecting more solar radiation
RI	Relative Importance
RIV	Ratios of Interannual Variability
RMSE/rmse	Root Mean Square Error
RRTMG	Rapid Radiative Transfer Model GCM Applications
SEM	Structural Equation Modeling
SF	Surface Forcing
SH	Sensible Heat
SRB	Surface Radiation Budget
SWD	Short Wave Downwelling
TAO	Tao et al. microphysics scheme
TDK	Tiedtke cumulus scheme
TPW	Total Precipitable Water
UW	University of Washington
WS	Water Supply

Chapter 1: Introduction

1.1 Background and Motivation

Extreme precipitation has profoundly harmed U.S. social and economic life. From 1980 to 2019, \$1.3 trillion has been lost due to extreme precipitation-related events (?), which have imposed a substantial burden on our society. ? showed that compared to other weather disasters, extreme precipitation dominated economic losses (accounting for more than 70%).

Extreme precipitation events affect every aspect of life. They pose a direct threat in urban areas with flood damage (e.g., \$22 million loss in the July 2016 Ellicott flood, \$1 billion loss in the 2010 Tennessee flood) (?). They increase the risks of landslides and mudslides (?). ? demonstrated that excessive rain can also damage buildings, increasing infrastructure risks. ? show that extreme precipitation disrupts transportation and increases the risk of motor vehicle collisions. Extreme precipitation events can also lead to health risks through waterborne diseases (?). Not only that, from 1980 to 2016, crop production lost \$10 billion U.S. dollars due to excessive rains (?). Overall, extreme precipitation exerts significant pressure on the government and insurance system, decreasing the stability of the society.

Records show that losses from many types of extreme precipitation are rising

consistently (?). For example, over three decades (1940–1990), losses from hurricanes increased eight times from \$5 billion to \$40 billion (?). During the same period, damage from floods rose from \$1 billion to \$6 billion (?). Meanwhile, hailstorms occurred more frequently, with losses of more than \$300 million (?).

Future projections show that the frequency and intensity of extreme precipitation events are increasing across the United States, which will lead to higher risks of property loss and more significant impacts on citizens' daily lives (????). Furthermore, within the warming climate, threats from extreme precipitation on food security are becoming more and more severe (???)

The goal of this study is to improve the prediction and process understanding of extreme precipitation in the United States. To do this, I adopted a mesoscale climate model multi-physics ensemble approach. To improve prediction, I first identified major problems in current extreme precipitation simulations, and key regions in which those problems occur. This analysis also provided a performance baseline for the following studies. Focusing on the most problematic key regions in extreme simulation, I performed multi-physics sensitivity experiments and found that mesoscale climate modeling of extreme precipitation is highly sensitive to your choice of cumulus schemes. Second, I focused on explaining the underlying physical mechanisms of why some Climate-Weather Research and Forecasting model (CWRF) schemes failed to capture extreme precipitation. I built a causality model to find the best fitting configuration from all possible configurations of structural equation models. I used this analysis model to explore the causes of problematic extreme precipitation simulation from a systematic point of view, including iden-

tifying the interaction between multiple crucial processes. Both the multi-physics sensitivity experiments and causal analysis showed that there are substantial differences in simulation performance and physics representation of extreme precipitation simulations among schemes. Therefore, weighting all schemes equally, regardless of performance does not produce the optimal ensemble outcome. To further boost the performance of the ensemble system, Chapter Four is dedicated to applying the Bayesian model averaging (BMA) method to derive optimal weights for each member. Using the selected physics configurations (particularly the ECP scheme) with the BMA method can improve the future projections of extreme precipitation. This improved understanding of the underlying mechanisms of extreme precipitation may help to guide future model development.

1.2 Improving extreme precipitation simulation through a multiphysics approach

Many studies have demonstrated the difficulties of extreme precipitation simulation, and the fact that the majority of numerical models still underperform and have sizable uncertainty (???). Meanwhile which processes extreme precipitation is most sensitive to is still a subject of debate. ? concluded that microphysics is the most critical process, specifically the ice phase process. However, even though the ice phase process is non-negligible, given the complexity of numerical models (and especially convection-cloud-radiation interactions) (?), they are not sufficient to resolve the underestimation in other situations. ? demonstrated the importance

of cumulus schemes in extreme precipitation simulations, while other researchers have shown the impact of resolution (????).

Consequently, Chapter Two aims to identify the processes to which extreme precipitation is most sensitive and to provide a reference point of the current status of extreme precipitation simulation, specifically for U.S mesoscale simulations. To achieve this goal, I conducted experiments using 25 different physics configurations covering eight major processes, and thoroughly analyzed their performance. The analysis identified not only problems in extreme simulation in current climate models, but also which processes were most sensitive to extreme precipitation simulation. The results showed that for mesoscale extreme precipitation simulations on a 30km grid, extreme precipitation was most sensitive to cumulus parameterization schemes. Overall, CWRf with ensemble cumulus parameterization (ECP) significantly outperformed other models and even outperformed the driving reanalysis.

Given the significant improvements in the best-performing configuration, this study provides a useful optimized configuration of physics schemes for future extreme precipitation projections. I also found the most sensitive process in mesoscale extreme precipitation simulation was the cumulus process. This analysis is highly valuable from a practical perspective, since the complexity of the climate modeling system, efficient model development requires identifying and focusing on key processes

1.3 Improving extreme precipitation simulation by better physical understanding

While capturing the observed characteristics of extreme precipitation is ambitious, the second goal, understanding the underlying physical mechanisms explaining model failure or success, is even more challenging. The previous analysis highlighted the importance of cumulus schemes. Given the importance of cumulus activity to precipitation, multiple potential hypotheses on why some cumulus schemes are insufficient to produce intense extreme precipitation have been put forward. ? emphasized the importance of convective to total precipitation ratios. In our experiments, however, we did not find a strong correlation between the convective ratio and extreme precipitation. Some studies have emphasized the importance of better-simulating intensities of cumulus activity: for example, ? proposed the use of an explicit convection solution, and ? suggested developing better convective closures and triggers. However, from the vertical tendency profiles of cumulus schemes, I found that schemes that simulated strong cumulus activity did not always produce sufficient P95 intensity. Furthermore, my experiments found no clear connection between triggers and performance. Meanwhile, ? highlighted that convective available potential energy (CAPE) is crucial in extreme precipitation simulations, while ? found that the highly uncertain term "time scale for CAPE consumption rate" plays a central role in extreme precipitation simulations. These studies indicate that the connection between cumulus activity and energy supply is important. Therefore, to

uncover why some schemes underestimate extreme precipitation, the analysis needs to include surface forcing, cloud forcing, energy supply, water supply, and their interactions.

To comprehensively analyze the interaction between different forcing and supply, I built an analysis model that find all potential configurations of the structural equation model (SEM) (?). The SEM can detect potential causal relationships within a web of knowledge, while also considering collinearity between factors. More importantly, the SEM can quantitatively measure the effects of each component. This ability of SEM is beneficial in the highly complex and interconnected relationships between each process in the earth system. For this analysis of biases, I included the 22 primary dynamic and thermodynamic fields most relevant to extreme precipitation. Through the SEM, I discovered five distinct physical mechanisms underlying the key biases in regional extreme precipitation, each involving interplays among water and energy supplies and surface and cloud forcing, with varying degrees of relative importance. For extreme precipitation, complex interactions were underlying the four major processes (two forcing and two supplies). The dominant processes changed with the transition of prevailing precipitation systems (convective precipitation to stratiform precipitation). The choice of cumulus parameterization affected how water and energy supplies acted through surface and cloud forcing and thus determined CWRF's ability to simulate extreme U.S. precipitation.

This analysis confirmed that the improvement of ECP was due to a better representation of the physics mechanism rather than a result of overfitting parameters. Thus, we can have more confidence in the ECP schemes in future projection.

Specifically, when the ECP scheme results disagree with other models in future projections, my findings indicate that the ECP outputs may be more trustworthy.

1.4 Improving extreme precipitation simulation by optimal weightings

Chapters Two and Three focus on the performance of individual schemes. However, given the regime dependence of parameters, no single scheme can represent the real climate, let alone extreme precipitation universally (??). In particular, the success of ECP also validated the benefits of combining ensemble members. Hence, Chapter Four combines ensemble results from different schemes to further boost extreme precipitation skill. Two methods were used to combine outcomes from the different ensemble members. The first method was the arithmetic mean of ensemble members (i.e. "composite ensemble"). This composite ensemble has the benefit of computational efficiency and stability and is theoretically straightforward to explain. However, ? demonstrated that equal weights in ensemble calculation cannot provide optimal ensemble mean outcomes for schemes with substantial differences in performance. Both of the previous sensitivity experiments and causal analyses demonstrated individual schemes perform differently in terms of extreme precipitation simulations. Chapter Four thus applied non-equal weighting methods. Among these, Bayesian Model Averaging (BMA) is generally considered to be the gold standard for making out of sample predictions (?). ? proposed a well-adopted expectation-maximization (EM) algorithm-based BMA method, which has success-

fully improved the performances of precipitation simulations (?). Therefore, this study tested the EM-based BMA method in extreme precipitation related ensemble. The EM-based method also served as a baseline to compare with other methods, including an Akaike information criteria-based BMA method (AIC,?), which is computationally more efficient. Moreover, this study also compared a bootstrapping version of Akaike weights, which considered the uncertainty associated with training data. Besides the above three BMA methods, Chapter Four also tested a stacking method. This final method was tested because BMA methods are based on a strong hypothesis: that the true data-generating model is among the potential candidate models. However, given the highly nonlinear complexity of extreme precipitation, the true model, earth system, is out of the ensemble member list (?). ? proposed the stacking method specifically to resolve this theoretical flaw.

? implemented the EM-based BMA method with linear bias correction. The linear bias correction ignored model uncertainty and led to "over-confident" prediction, in contrast to the probabilistic type of bias correction, which provides uncertainty information. Such uncertainty information is highly valuable in climate studies and projections. Hence, I also implemented the Markov Chain Monte Carlo (MCMC) based probabilistic bias correction in this study. The MCMC methods also enabled a more flexible model design (?), which allowed the use of extreme value distributions as the prior distribution.

Therefore, Chapter Four explores the potential benefits of using different methods to combine members (three variations of BMA, and stacking) as well as bias correction methods (linear bias correction and MCMC method) in extreme precipita-

tion ensemble simulations. The results demonstrated that both BMA and stacking methods improved the general performance of extreme precipitation not only in spatial pattern distribution but also in interannual variability. This improvement highlights the advantage of using more sophisticated methods for ensemble combinations. Meanwhile, there was no significant difference between different combinations or bias correction methods. Consequently, the more computationally efficient EM algorithm with linear bias correction is preferred. In addition to the above benefits, the EM also provided probabilistic information associated with each field, which can be greatly valuable in future projections and beneficial to decision makings.

1.5 Organization

Chapter Two (submitted to Journal of Climate) examines CWRF’s improvements in simulating U.S. extreme precipitation and investigates the responsive processes by comparing a large ensemble of long integrations using multiple physics configurations. Chapter Three (submitted to Journal of Climate) explores the physical mechanisms that can explain how cumulus parameterization determines CWRF’s ability to simulate U.S. extreme precipitation by comparing deep integrations with the five representative cumulus schemes (ECP, NKF, TDK, NSAS, and BMJ). In Chapter Four, to further boost the performance of extreme precipitation simulation, I explored different combination methods in extreme precipitation simulation to maximize extreme precipitation simulation skills.

Chapter 2: Sensitivity to physics parameterizations

2.1 Introduction

Since 1980, 230 billion-dollar weather disasters have occurred in the United States, causing more than \$1.5 trillion in total economic losses (?). Of the damage caused by these weather disasters, more than 70% was due to extreme precipitation (?). Greater risks are anticipated in the future, since the frequency and intensity of extreme precipitation events are increasing over the United States (?????). Despite the profound impact on society, prediction of extreme precipitation remains highly uncertain (???). In addition to observational data scarcity and discrepancy (??), data-to-model spatial scale mismatch (??), and unpredictable natural variability (?), the prediction uncertain arises from model deficiencies in representing organized convection systems and other physical processes (?????). A major problem is that most climate models tend to underestimate extreme precipitation (????). 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 (??). Others reported simultaneous overestimation of light rain and underestimation of extreme precipitation (???). However, more light rain could consume only slightly more energy, which would not noticeably weaken extreme precipitation

(?). ? 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, ? showed that a modified trigger function suppressed not only light rain but also heavy rain. ? 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 (?), 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 (??????). ? 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, ? 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 (?). 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 (?). 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 (????????????). CWRF is superior to the driving reanalysis and a popular regional climate model in capturing extreme precipitation characteristics over China (?). 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 U.S. 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 concludes the results. The companion paper (?) will demonstrate how cumulus parameterization dominates extreme precipitation simulation and explore the potential physical contributors to extreme event modeling ability.

Physical processes		Parameterizations	References
Cumulus (CU)		ECP	?; ? ;???
		NKF	? ;?
		TDK	? ;? ;? ;? ;?
		NSAS	?
		BMJ	?; ??
Microphysics (MP)		TAO	??
		THO	??; ?
		MOR	???
		WD6	?
Aerosol (AE)		A3D	?;?; ?. A3D uses aerosol mass loadings and optical properties or observed (MISR, MODIS)
		A2D	A2D uses aerosol optical depth distributions (?)
Cloud (CL)		XRL	Diagnostic cloud cover based on ? with modifications by ?
		CPL	Prognostic cloud cover based on ?
Radiation (RA)		GSFC	? ;?
		CCCMA	?; ? ;?
		FLG	??
		RRTMG	?
Boundary layer (BL)		CAM	? with updates to include gravity wave drag effect and orographic turbulence stress (?)
		MYNN	??
		ACM	? with updates on MOL calculation following WRF3.7.1.
		UW	?
Surface (SF)	Land	CSSP	????? ; ? ;? ;? ;? ;?
		NOAH	? ;? ;?
	Ocean	SST	?; ?
		XOML	??

Table 2.1: Physical processes, parameterizations, and their references.

2.2 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 (?), with several important updates including terrestrial hydrology (?), cloud-aerosol-radiation interaction (??), land surface characteristics (?), and upper oceans (?). Of particular relevance to this study, CWRF incorporates an ensemble cumulus parameterization (ECP) based on ?, which has outstanding performance in precipitation simulation, including extreme events and flooding (?), over oceans (?) and land (?), and in different climate regimes (?). 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 (?).

This study selected multiple schemes from each of the nine key CWRF parameterization processes (Table 2.1, ?) to form 25 CWRF physics configurations, as listed in Table 2.2. All CWRF simulations were conducted on a well-tested North American domain including the contiguous United States (??????). Horizontally, the domain was centered at (37.5 ° N, 95.5 °W), containing 138x195 points at grid spacing of 30 km using the Lambert conformal map projection. There were 36

Parameterization		CU			MP			THO			MOR	WD6	AE		CL		RA		BL			Land		Ocean	
		ECP	NKF	TDK	NSAS	BMJ	GSFC	THO	MOR	WD6	A2D	A3D	XRL	CLP	GSFC	CCCMA	FLG	RRTMG	CAM	MYNN	ACM	UW	CSSP	NOAH	SST
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	FMITDK			✓			✓					✓								✓				✓	
V	FMNSAS				✓		✓					✓								✓				✓	
W	FMBMJ					✓						✓										✓		✓	
X	FMTHO	✓					✓					✓										✓		✓	
Y	FMMOR	✓								✓		✓										✓		✓	

Table 2.2: Ensemble experiment physics configurations.

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 (?). The simulation began in October 1, 1979 and ran continuously until the end of 2015. Considering the first two 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 was originally described by ?, and has since been updated, including modifications in entrainment formulation (?) and parameterization closure (?). Given the non-Gaussian temporal and inhomogeneous spatial distributions of extreme precipitation (??), an extended observational period in a dense monitoring network is required to properly capture the statistical characteristics of extreme events (?). 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 (??). 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 ?, they were adjusted for topographic dependence using monthly mean data from the Parameter elevation Regression on Independent Slopes Model (?). 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 ?. 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 (?).

Numerous extreme indices have been recommended by the Expert Team on Climate Change Detection, Monitoring and Indices (??), and some are often used for observational analysis (e.g. ?) and model evaluation (e.g. ?). Here we used daily 95th percentile precipitation (P95) to analyze climatic extreme simulation skills in different seasons and distinct regions. 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 (??). 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 (???). However, improved P95 performance could be due to fake drizzling (??) or to an artificial shift in the precipitation intensity distribution (?), rather than improvement in the underlying physics processes. To improve the reliability of the P95 analysis, we also evaluated total number of rainy days (NRD) and average daily rainfall intensity ($\text{DRI} = \text{total accumulated precipitation amount} / \text{NRD}$). Together they provide additional information on whether P95 biases are associated with deficiencies in clear-day frequencies or in rainfall magnitudes.

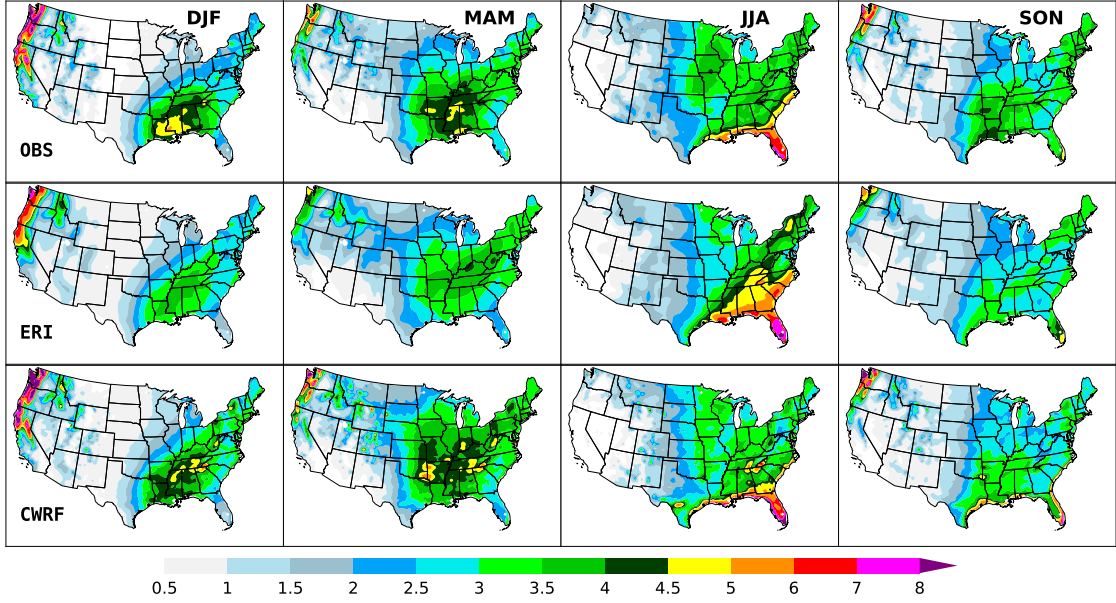


Figure 2.1: Geographic distributions of 1980-2015 mean seasonal precipitation amount [mm day^{-1}] observed (OBS), assimilated (ERI), and simulated by CWRf control ECP for winter (DJF), spring (MAM), summer (JJA), and autumn (SON).

2.3 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 2.1 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 $4.5 [\text{mm day}^{-1}]$ 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 $4.5 \text{ [mm day}^{-1}\text{]}$, which was more realistic than ERI. In summer, observations showed strong centers over the Midwest. ERI significantly underestimated precipitation over those centers, while CWRF produced both better intensity and a more reasonable distribution. ERI overestimated precipitation over the Gulf States, averaging over $4.5 \text{ [mm day}^{-1}\text{]}$ compared to the less than $4 \text{ [mm day}^{-1}\text{]}$ 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, 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 observations and surface measurements (?), 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 2.2 compares 36-year mean seasonal NRD distributions. CWRF captured finer structural details than ERI over the western U.S. 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 (?). In winter, both ERI and CWRF sim-

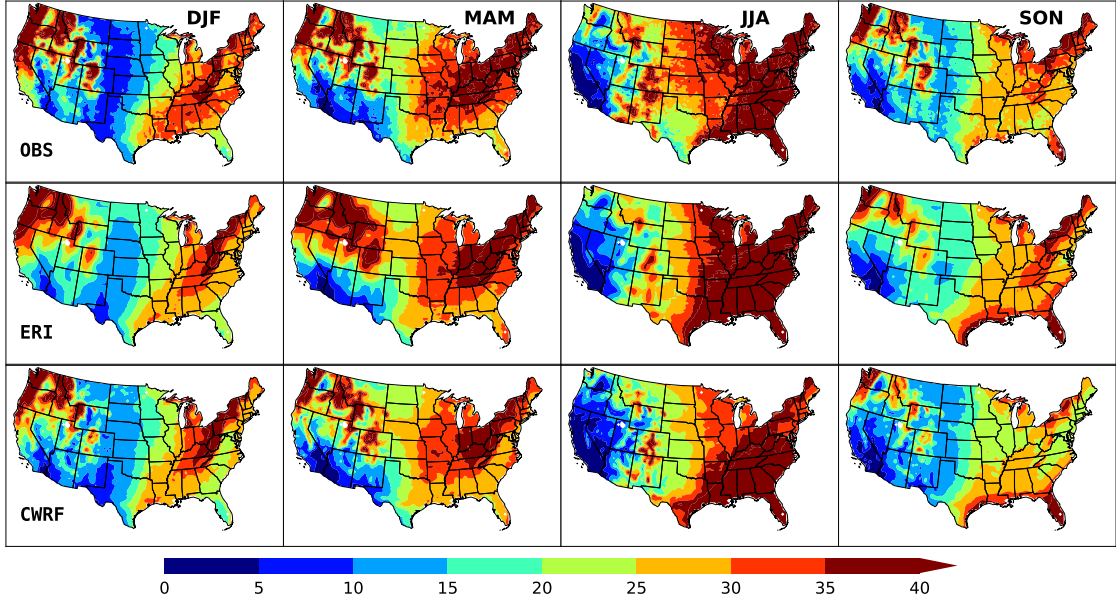


Figure 2.2: Same as Fig. 2.1 except for the number of rainy days (NRD).

ulated 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 (?) 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 2.3 compares 36-year mean seasonal P95 distributions. In all seasons,

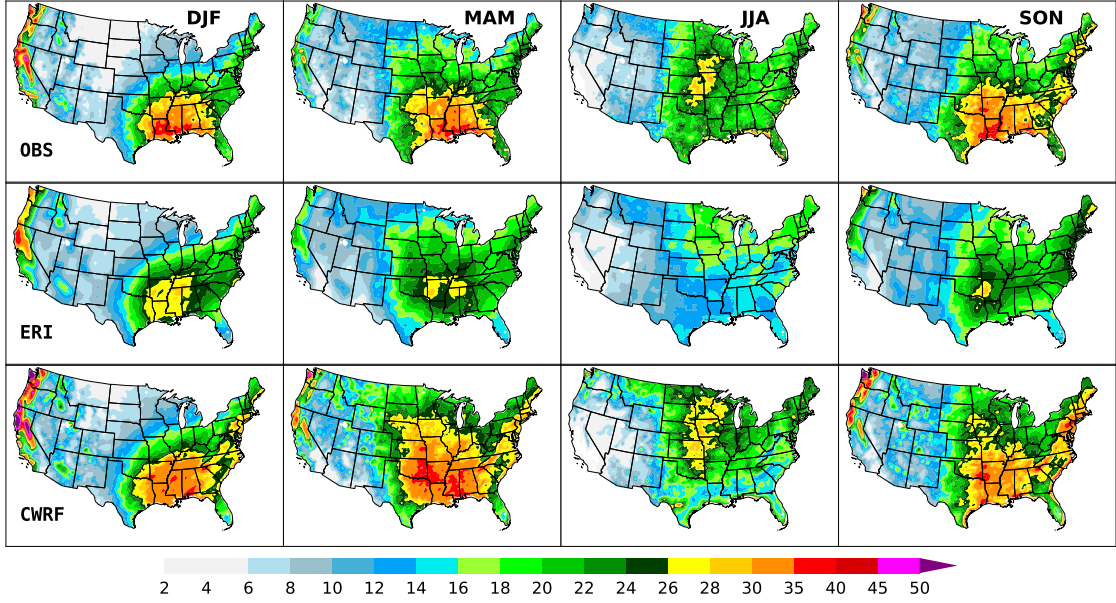


Figure 2.3: Same as Fig. 2.1 except for the daily 95th percentile precipitation (P95) [mm day⁻¹].

CWRF outperformed ERI over the western U.S. 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 (?), while the eastern sides are typically dry transitional zones partly controlled by the precipitation-shadowing effect (?). 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 (??). 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 spatiotemporal variations (?). CWRF also substantially out-

performed 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 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. 2.3). 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 reduced its skill.

2.4 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 CWRP: the Gulf States (GS) and the Central to Midwest States (CM). Since the driving large-scale synoptic conditions were same and the regional topographic forcing effects are expected to be small, the CWRP’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 CWRP’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 CWRP’s performance among its ensemble of 25 physics configurations to simulate extreme precipitation features. The comparison identifies which physical processes CWRP extreme precipitation simulation is most sensitive to, and which schemes or combinations best capture those processes.

CWRP’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. 2.4. 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 with radii smaller than

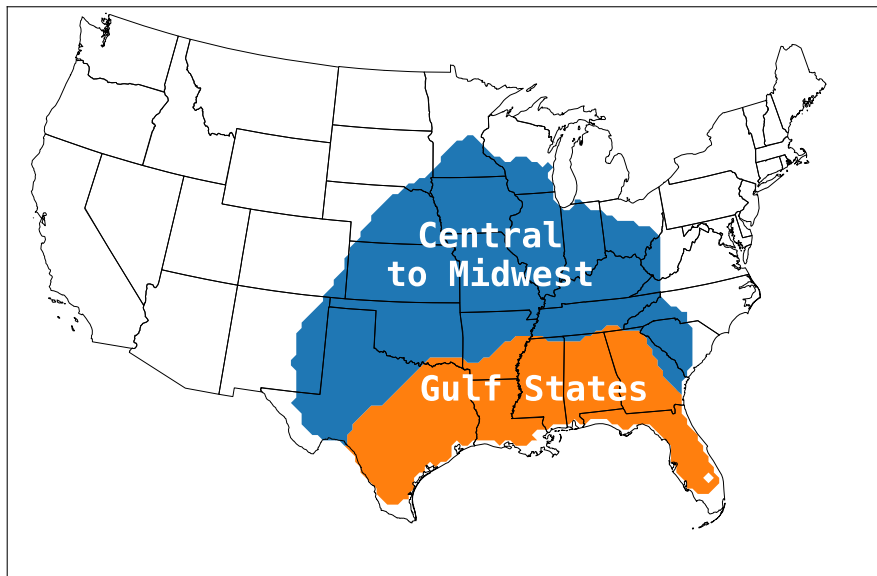


Figure 2.4: Boundary specification of the two key regions where ERI severely underestimated extreme precipitation: the Gulf States (GS) in spring and the Central to Midwest States (CM) in summer.

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

Figure 2.5 compares 36-year mean seasonal P95 biases (from observations) between ERI and the 25 CWRP physics configurations averaged over the two regions. The control CWRP corresponds to the ECP run, called in short the control ECP. For brevity, we refer these five CWRP runs directly by the name of the respective 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 CWRP run replacing the microphysics scheme of Tao et al. with that of Morrison et al. while the rest is identical to the control configuration. Other runs replacing two or more processes’ parameterization schemes differed from the control CWRP 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 CWRP configurations that include ECP cumulus scheme.

As apparent in Fig. 2.5, CWRP 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 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, CWRP 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, type I members produced reasonable extreme precipitation and relatively small biases, especially in winter and spring. 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 overestimated in autumn and underestimated in summer, while NKF members generally overestimated, 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 underestimated summer P95, whereas type I members of CWRf produced more realistic simulations. ERI also underestimated autumn P95, which CWRf type I members overestimated slightly. 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). Among all seasons, ERI was most realistic in spring and overestimated slightly in winter. Type I members significantly 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 under-

estimated (especially in spring) and TDK members overestimated, whereas NSAS members had small spring biases and moderate winter overestimates.

Figure 2.6 compares geographic distributions of seasonal P95 biases from observations over the contiguous United States for ERI and the five CWRf members differed 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. 2.1). 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 ?). 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. 2.6). 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 over 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

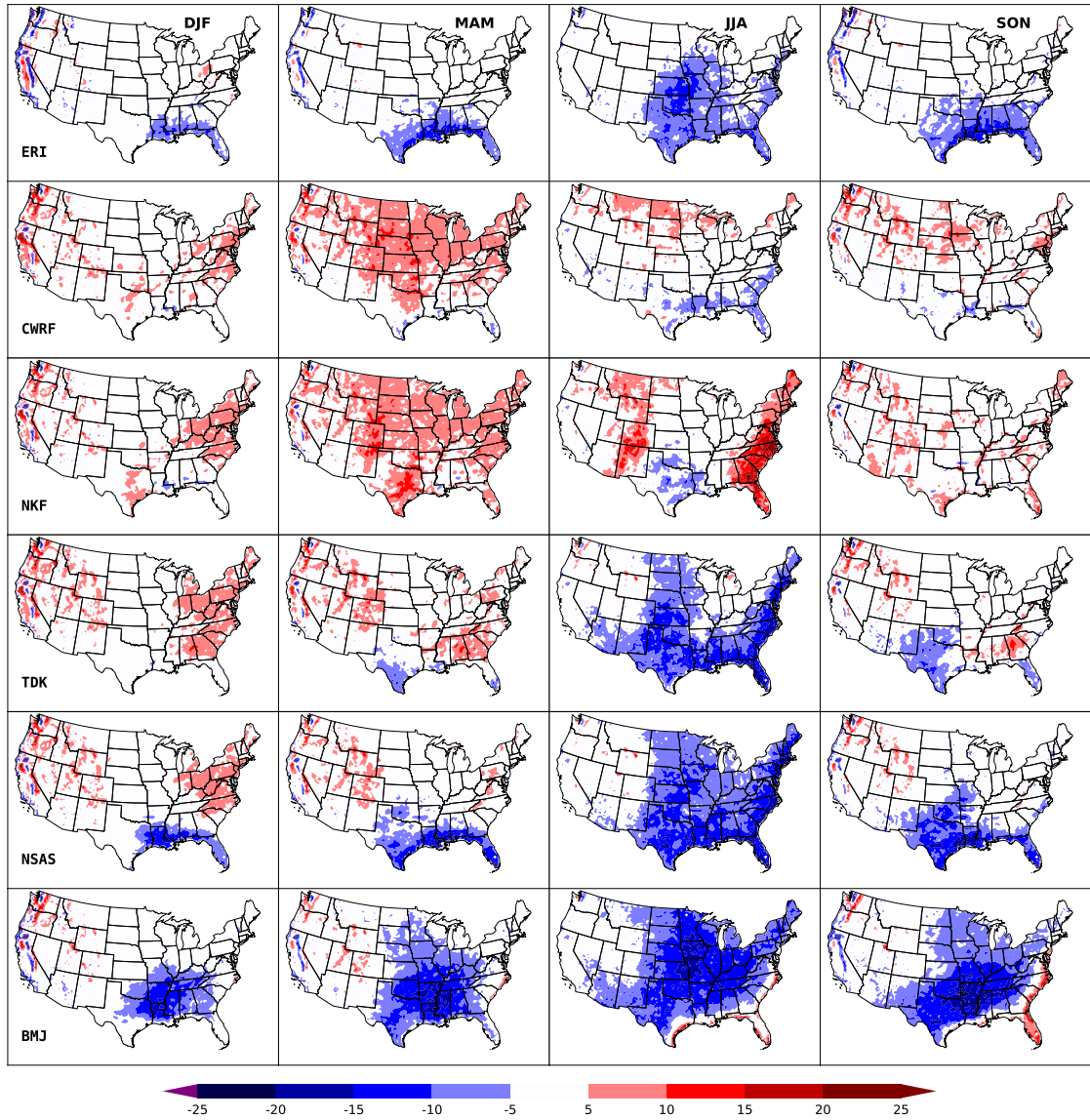


Figure 2.6: Geographic distributions of 1980-2015 mean daily 95th percentile precipitation (P95) biases from observations [mm day^{-1}] for winter (DJF), spring (MAM), summer (JJA), and autumn (SON) as assimilated (ERI) and simulated by five CWRf members varying only the cumulus scheme (ECP, NKF, TDK, NSAS, BMJ).

of cumulus parameterization closure assumptions to distinguish land versus oceans (??), 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 (?). Therefore, the ECP scheme, with a more comprehensive treatment of the land-ocean contrast helped produce sufficient convective activity and better extreme precipitation simulation.

Figure 2.7 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 United 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 2.8 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 were 0.88~0.94 for ERI, 0.89~0.91 for ECP, 0.88~0.90 for NKF, 0.74~0.90 for TDK, 0.92~0.94 for NSAS, and 0.96~0.98 for BMJ. Strong correlations indicated that un-

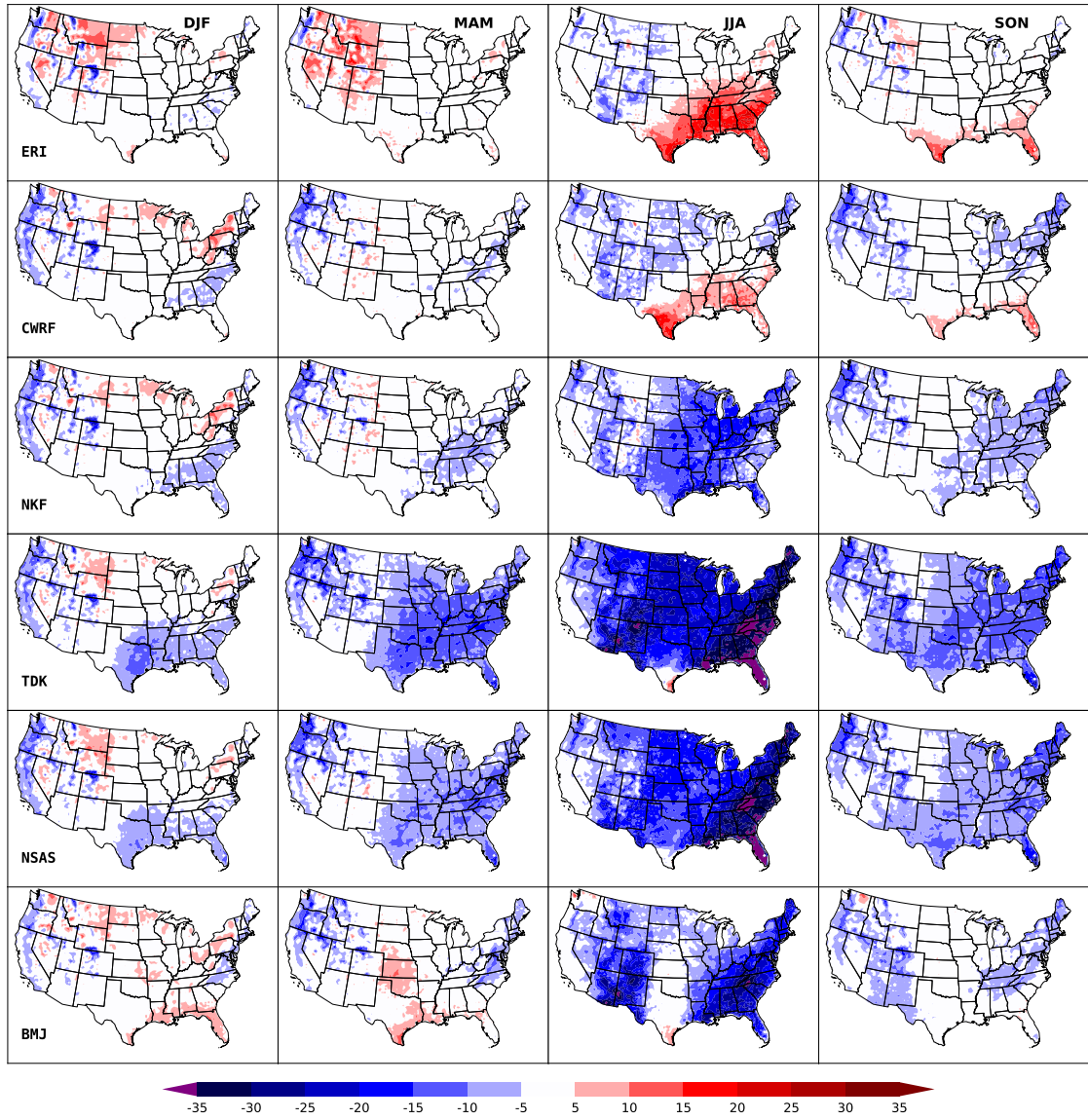


Figure 2.7: Same as Fig. 2.6 except for the number of rainy days (NRD) biases.

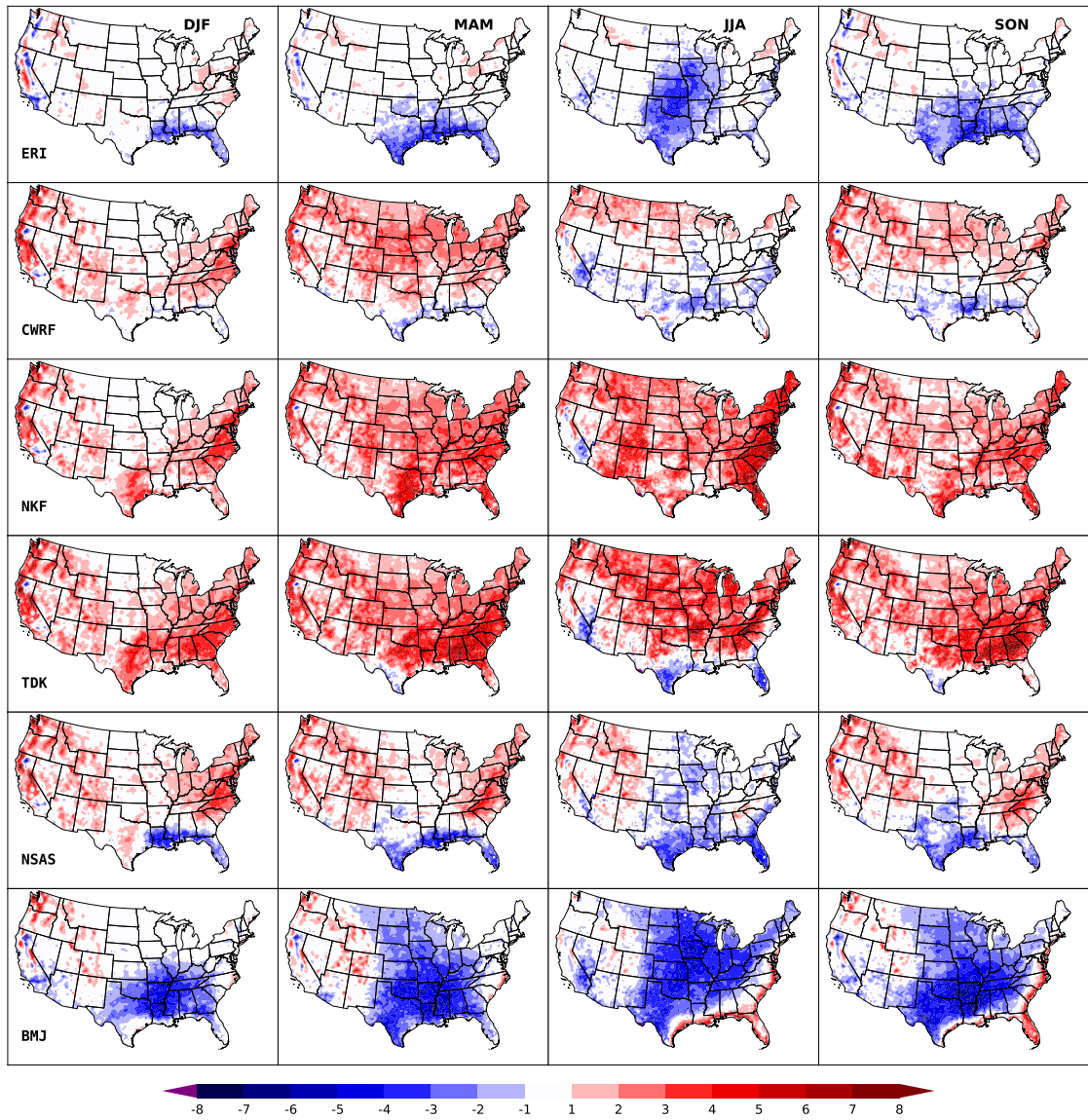
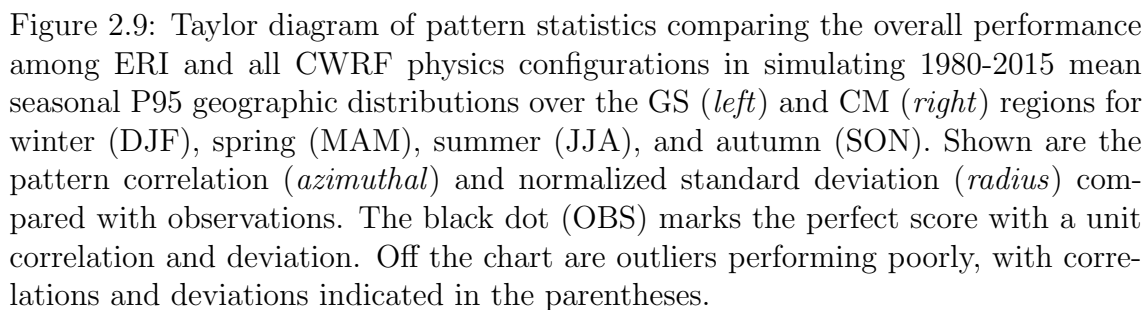


Figure 2.8: Same as Fig. 2.6 except for the daily rainfall intensity (DRI) biases [mm day^{-1}].

derestimations (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. 2.7), 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 2.9, using a radar diagram, compares seasonal P95 spatial pattern correlations and standard deviations of ERI and the 25 CWRP 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). CWRP 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. 2.6), which canceled each other to yield small regional mean P95 biases (Fig. 2.5) with significantly overestimated spatial deviations (Fig. 2.9). 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); 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. 2.9), with large positive and negative local errors (Fig. 2.6) canceled each other to yield small regional mean biases in spring and autumn (Fig. 2.5)

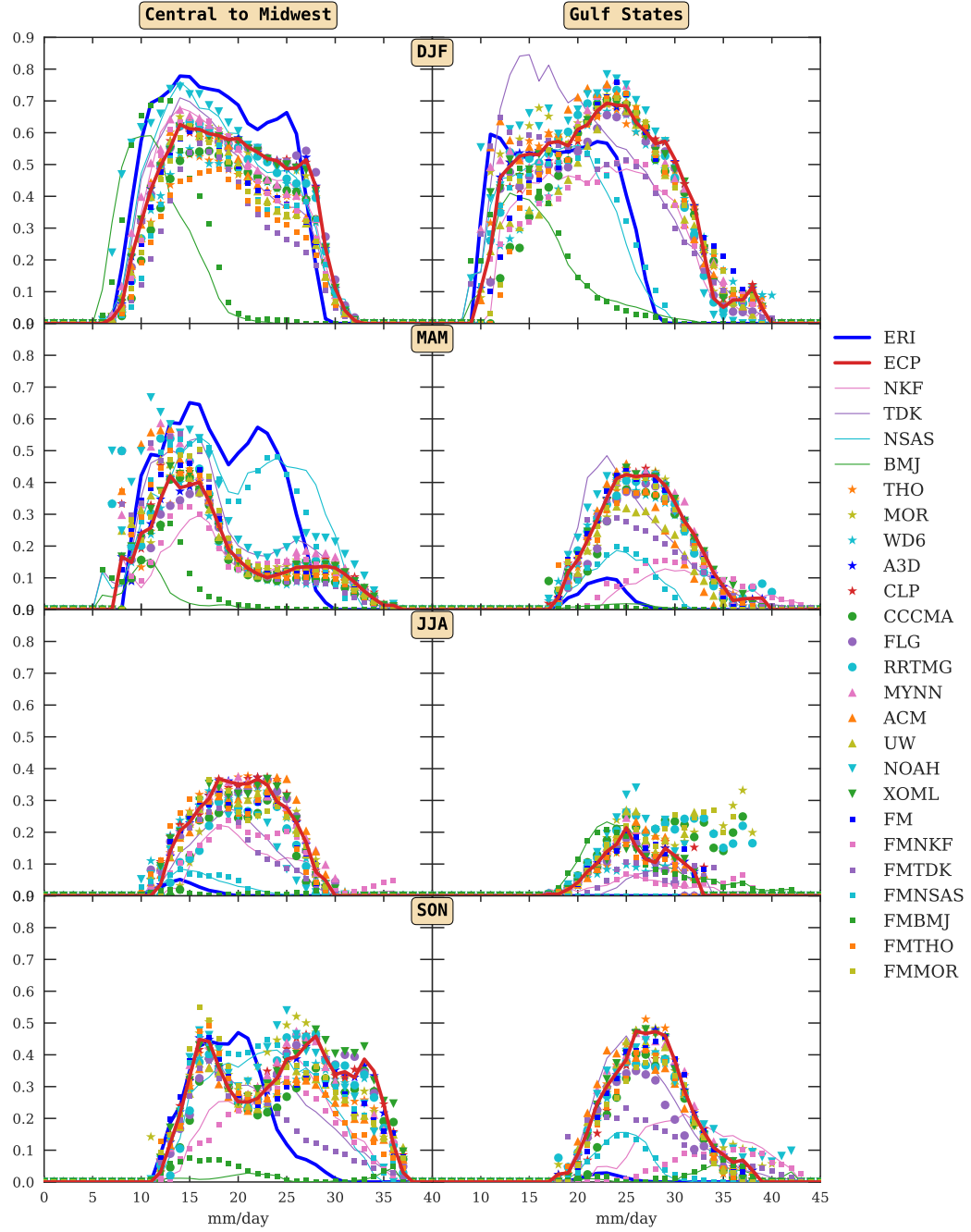


Figure 2.10: The equitable threat score (ETS) that measures the overall skill dependence on rainfall intensity for ERI and all CWRW physics configurations in simulating 1980-2015 mean seasonal P95 geographic distributions over the GS (*left*) and CM (*right*) regions for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). The x-axis depicts the P95 thresholds at a $1.0 \text{ [mm day}^{-1}]$ bin interval, while the y-axis scores the ETS values.

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 minus hits expected by chance (?). Figure 2.10 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 CWRP physics configurations. Overall, ETS skills were highest in winter and lowest in summer. This highlights the difficulty of simulating extreme events in summer, when convective precipitation dominates. 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. CWRP members using cumulus schemes other than ECP generally had lower ETS. NKF members scored systemat-

ically 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. 2.6). 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. 2.6) 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. 2.5). 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 2.6-2.8 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.

2.5 Summary and discussion

We analyzed CWRF’s improvements over ERI extreme precipitation simulation 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. The two groups 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 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. In summary, CWRF using the ECP cumulus scheme performed the best of all physics configurations and generally outperformed ERI, especially over the GS and CM regions in seasons dominated by convective precipitation. This success may reflect ECP's use of an optimized ensemble of parameterization closures based on the framework of ? to represent convection variations between land and oceans (???). Other cumulus schemes severely underestimated extreme events. In particular, CWRF members

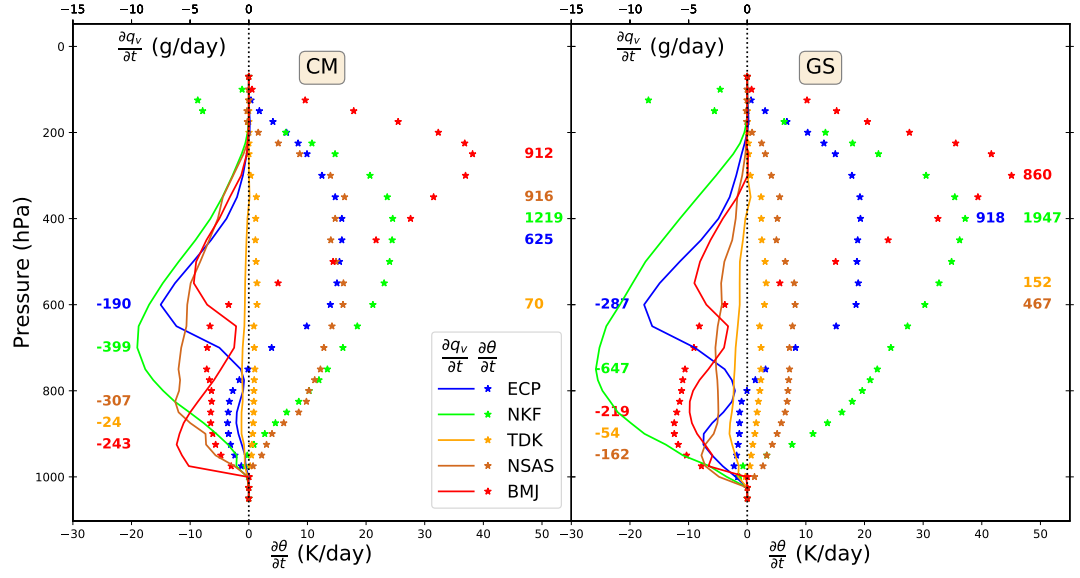


Figure 2.11: Summer mean vertical potential temperature ($\theta, star$) and water vapor ($q_v, curve$) tendency profiles among the five cumulus schemes (*color*) for 2004 as averaged over all the grids having rainfall greater than 50 [mm day⁻¹] within the CM (*left*) and GS (*right*) regions. Also labeled 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.

using TDK, NSAS, and BMJ underestimated P95 substantially in both GS and CM regions for summer, and also largely in autumn and spring. ERI, which used a variant of the TDK cumulus scheme (?), similarly underestimated P95, even though it assimilated pseudo-precipitation data. ? compared the convective heating profiles of these three cumulus schemes and showed that TDK favors boundary layer clouds, BMJ favors shallow and midlevel clouds, while NSAS exclusively favors deep clouds. They further found that the scheme generating more deep convection tended to produce less intense mean precipitation in the tropics due to reduced net cloud radiative cooling.

Figure 2.11 compares summer mean vertical temperature and humidity tendency profiles among the five cumulus schemes for 2004, in which the P95 geographic

distribution was most strongly correlated to the climatological mean. These profiles were calculated 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 (?). 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, likely due to heat loss from large detrainment and associated evaporation. On the other hand, NSAS generated a very weak warming in the entire cumulus tower, while TDK yielded an even weaker warming. All cumulus schemes resulted in drying throughout the cloud column as water vapor was depleted by precipitation. However, their drying's vertical distributions differed substantially. In the CM region, BMJ had distinct double peaks at 925 (primary)

and 550 (secondary) hPa, corresponding to the shallow and midlevel convection, respectively. ECP also simulated double peaks at 875 and 600 hPa, but the midlevel drying was predominant. In contrast, NKF produced a much stronger and deeper drying layer than ECP, with the predominant peak at 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 50% (ECP) and 62% (NKF) but decreased by 10% (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 at 825 hPa, whereas in the GS region, it produced a much weaker drying than both ECP and NKF, with the peak further down at 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. Our result agrees with ? in regarding convective heating profile contrasts among TDK, NSAS, and BMJ, but disagrees on the tendency of deeper convection for less precipitation. To the opposite, among all schemes, NKF produced the strongest warming and drying rates and correspondingly largest P95, whereas TDK and NSAS 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. Given our sensitivity analysis above, it is imperative to understand how these cumulus parameterizations and their interactions with other physics representations result in

improved 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 (?) and other subsequent papers in this series.

Chapter 3: Dependence on cumulus parameterization and underlying mechanism

3.1 Introduction

Extreme events are by definition rare, but are high-impact, hard-to-predict phenomena that are beyond our normal expectation (??). Frequency and intensity of extreme precipitation events have been observed increasing over the United States in the past decades and are projected to continue rising as global warming (???). However, realistic simulation of these events remains challenging, since their development depends on initial conditions, large-scale drivers, regional feedbacks, and stochastic processes (?). In particular, most climate models significantly underestimate extreme precipitation (????). This has often been identified as the drizzling problem, where models simulate too frequent light rain but inadequate heavy events. The problem exists even in the modern reanalyses that have already assimilated daily pseudo-precipitation observations (????). Increasing model resolution (such as to grid spacing of 10~50 km) may help, but cannot solve the problem (????????). Even cloud-permitting models with grid spacing down to 3-4 km may still underestimate extreme precipitation (?) or substantially overestimate it (?). The problem

could also be responsible for existing models' general underestimation of observed extreme precipitation trends, and have further ramification for their reliability in projecting future changes of these events (?), since biases can persist or propagate into projections (??). Therefore, it is imperative to improve model ability in simulating extreme precipitation. As the first of a pair, our earlier paper (?) demonstrated that the regional Climate-Weather Research and Forecasting model (CWRF, ?) significantly improved its driving European Center for Medium-Range Weather Forecasts' Interim Reanalysis (ERI, ?) in simulating extreme precipitation over the United States. This offered an opportunity to investigate how the improvement was made. To facilitate the study, we defined two distinct regions: the Gulf States (GS) and the Central to Midwest States (CM), where ERI substantially underestimated precipitation extremes while CWRF realistically captured them. We then compared the performance among an ensemble of 25 CWRF physics configurations in the two regions and found that the extreme precipitation simulation was most sensitive to cumulus parameterization. Of five tested schemes, the ensemble cumulus parameterization (ECP, ???) used in the CWRF control was the most skillful at reproducing seasonal mean spatial patterns of daily 95th percentile precipitation (P95), which is also a good indicator of rainfall intensity. In contrast, the new Kain-Fritsch scheme (NKF,?) produced a comparable P95 skill to ECP but at the expense of underestimating the number of rainy days (NRD). On the other hand, the modified ? scheme (TDK, ?) severely underestimated summer P95 because its stronger NRD underestimate could not balance its larger reduction in average daily rainfall intensity (DRI = total accumulated precipitation amount / NRD). Such TDK biases occurred sim-

ilarly to the new simplified Arakawa–Schubert (NSAS) scheme used in the National Centers for Environmental Prediction’s Global Forecast System (??), except that the underestimates covered broader areas in summer and extended into all seasons over the GS region. The Betts–Miller–Janjić scheme (BMJ, ?) underestimated P95 most severely in both regions for all seasons because its strong DRI underestimate accompanied with its NRD overestimate – the typical drizzling problem. The ERI, using another variant of the ? scheme (??), simulated realistically DRI through its observational assimilation, but still underestimated P95 largely in summer CM and all seasons’ GS mainly because of its overestimating NRD or again the drizzling problem. While capturing the observed characteristics of extreme precipitation is challenging, understanding the underlying physical mechanisms for model failure or success is even more difficult. After reviewing various issues on detection, simulation, and attribution of extreme events, ? and ? highlighted the pressing need to better understand and ultimately overcome model deficiencies. While numerous studies have focused on why future increases in regional precipitation extremes can be significantly greater than that expected from the Clausius–Clapeyron relationship (???), very few have investigated how and why current climate models fail to reproduce extreme events observed in the past. ? suggested that realistically simulating convective to total precipitation ratios is important. ? demonstrated that improving cumulus parameterization formulations, such as convective closures and triggers, is a key. ? speculated that replacing cumulus parameterization with an explicit convection solution is more desirable. ? showed that no single physics configuration performs best in all cases and the selection of parameterization schemes

has larger impact on model performance during more intensive rainfall events. ? found a large sensitivity of extreme precipitation simulation to physics representations of land surface, cumulus, and radiation processes, and argued that how much the convective available potential energy (CAPE) is generated and consumed is essential. ? identified extreme precipitation is sensitive to uncertain parameters in deep convection scheme, especially its time scale for CAPE consumption rate. ? attributed the underestimation of extreme precipitation to lacking ice phase processes in cloud microphysics schemes. Given the complexity of climate models and especially convection-cloud-radiation interactions (??), these studies did not fully address the underlying mechanisms and their relative contributions to systematic model biases at regional scales. As the second of the pair, this paper builds on our first study (?) to explore the physical mechanisms that can explain how cumulus parameterization determines CWRP’s ability in simulating U.S. extreme precipitation by comparing long integrations with the five representative cumulus schemes (ECP, NKF, TDK, NSAS, BMJ). Section 2 describes these cumulus schemes, the causal ingredients and observational data for evaluation, and the experiment design for sensitivity understanding. Section 3 examines extreme precipitation’s dependence on cumulus schemes, distinguishing seasonal-interannual variations over the GS and CM regions. Section 4 presents the complicated correlations of regional extreme precipitation biases with the causal ingredient fields. Section 5 uses the structural equation modeling approach to explore the potential causes and physical processes responsible for the regional extreme precipitation biases and their differences among ERI and CWRP configurations. Section 6 concludes the results with a summary for

the key mechanisms underlying these model biases and regional contrasts.

3.2 Model description, experiment design, causal ingredients, and observations

The CWRF model formulation, computational domain, and integration procedure for this study were identical to ?. Here we used the five ERI-driven CWRF 1980-2015 continuous integrations, whose model configurations differed only in cumulus parameterization, swapping the control ECP with NKF, TDK, NSAS, or BMJ schemes. We also compared these CWRF downscaling simulations with the driving ERI, which used a variant of the TDK scheme. Table 3.1 summarizes the major differences among these cumulus schemes in closure, trigger, and entrainment formulations as well as respective references. We further adopt ? as the common parameterization for shallow convection, so to eliminate the complication from those built in some cumulus schemes. ? presented a comprehensive sensitivity analysis of how these and other 9 cumulus schemes affect CWRF’s prediction of the 1993 and 2008 summer floods over the Central United States. ?? then focused on ECP to study the effects of its closures on summer precipitation simulations over the continental United States and coastal oceans.

To gain physical insight, we conducted a correlation analysis among model biases in P95, DRI, NRD and other dynamic and thermodynamic quantities. For this, we adopt the ingredients-based approach proposed by ? and used in ?. As lifting moist air to condensation produces precipitation, extreme events must re-

	Closure	Trigger	Entrainment	Reference
ECP	Ensemble of moisture convergence and vertical velocity closure, with different weights and ensemble algorithms over ocean and land	Maximum CAPE strength	Linear combinations of multiple entrainment rates, each member has its own updraft radius	?; ?; ???
NKF	Total instability adjustment	CAPE >0; Parcel temperature perturbation	Entrainments were proportional to mass-flux at the cloud base divided by updraft radius	?;?
TDK	Total instability adjustment	Moisture convergence	Entrainments equals to summation of turbulent part and organized part. Turbulent part calculated from vertical velocity. Organized part is a linear function of height in pressure	?;? ; ?; ? ; ?
NSAS	Quasi-equilibrium assumption	Lifting depth trigger	Entrainments inversely proportional to updraft radius	?
BMJ	Quasi-equilibrium assumption	Positive cloud work function threshold	No entrainment schemes	?; ??

Table 3.1: Cumulus schemes closures, triggers, entrainments, and their references.

sult from sustained high rainfall rates, which require rapid ascents of air masses containing ample water vapor contents and efficient rainouts. So the basic ingredients for extreme precipitation events include maximizing moisture supply, ascent strength, and precipitation efficiency. These conditions can be met in the weather systems with dominant deep moist convection, which prevails in warm seasons with abundant moisture availability and sufficient buoyant instability promoting strong updrafts. Convective rainfall rates tend to be higher than those from stratiform (i.e. stably stratified) precipitation systems, in which updrafts are relatively weak but widespread as maintained by topographic or synoptic forcings rather than buoyancy. Thus, an alternative for precipitation efficiency is the convective to total precipitation ratio (RCT). Moisture supply can be measured by column total precipitable water (TPW), which is associated with atmospheric moisture convergence (MC) and surface evapotranspiration (ET). Ascent strength may be quantified by CAPE and convective inhibition (CIN) for buoyant instability caused updrafts, and 700 hPa vertical velocity (W_{700}) for large-scale forced upward motions. Note that CAPE depicts the theoretical maximum updraft speed $w_{max} = \sqrt{CAPE}$ (?), and is separated from CIN at the level of free convection (LFC). They all links to the lifting condensation level (LCL), where the air parcel lifted dry adiabatically to become saturated.

For a deeper process understanding, we examined other relevant quantities. In particular, surface shortwave downwelling radiation (SWD) provides the source energy for ET and sensible heat flux (SH), which drive the development of CAPE and planetary boundary layer height (PBLH). The latter is expected to link with

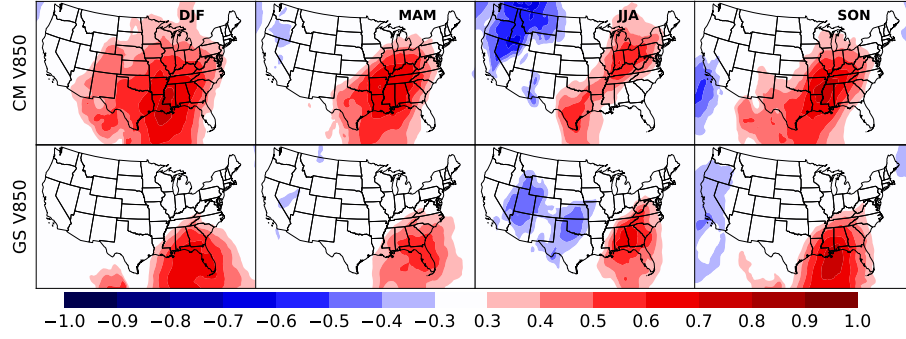


Figure 3.1: The teleconnection patterns of the CM and GS regional mean precipitation interannual variations correlated with 850 hPa meridional wind distributions for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). Outlined separately for CM and GS is the core correlation area common to all seasons, where the V850 index is defined.

LFC and LCL as well as the fraction of cloud in low layers (FCL), which in turn affects SWD and hence 2-m air temperature (T_{2m}) and specific humidity (Q_{2m}). Deep cumulus towers have higher and colder cloud tops, giving larger fraction of cloud in high layers (FCH), emitting less outgoing longwave radiation (OLR), and also thicker optical depths, reflecting more solar radiation (RET). As the water vapor provider, TPW limits total cloud water (liquid plus ice) water path (CWP) available to the warm and cold cloud microphysics processes. Furthermore, we calculated net surface energy (NSE) as net downward radiation minus (latent plus sensible) heat fluxes at the surface, which include respectively SWD, ET, and SH effects, while define cloud radiative effect (CRE) as the (OLR plus RET) difference between clear and total sky at the top of the atmosphere.

Dynamic quantities (except MC and W700) are more difficult to define because their relationships with precipitation are generally nonlocal (e.g., ???). Thus, we first examined their observed teleconnection patterns. This requires observed atmo-

spheric circulation data, which is lack. The best proxy is the NASA Modern-Era Retrospective Analysis for Research and Applications version 2 (NR2), with data available at grid spacing of 50 km from 1980 onward (?). This offers an independent reanalysis from the driving ERI (~ 80 km) so that our diagnosis based on the CWRF downscaling simulations is not biased toward ERI inherited circulation errors. Figure 3.1 shows the teleconnection patterns of the CM and GS regional mean precipitation interannual variations correlated with 850 hPa meridional wind distributions. The correlations were done for each season using monthly mean anomalies from daily maximum wind values to remove the influence of the annual and diurnal cycles. Strong positive correlations exist with 850 hPa southerly flows over the Southeast, with specific patterns that vary both in seasons and between CM and GS. For simplicity, we defined a low-level flow index (V850) as daily 850 hPa meridional wind averaged over the core correlation area that is common to all seasons but separated for CM and GS.

This study used daily precipitation and T2m data based on the National Weather Service Cooperative Observer network (COOP) with quality-controlled records from 8516 stations (??). They were analyzed with topographic adjustments onto CWRF grids following ?. Table 3.2 summarizes other daily observational data sources and available periods used in this study. These include: 1) CWP retrieved from the Advanced Very High-Resolution Radiometer (AVHRR)-based Thematic Climate Data Records (?) 2) SWD, RET, OLR and CRE from the Surface Radiation Budget release 3.0 Global Energy and Water Cycle Exchanges project (SRB/GEWEX) Daily Universal Time Data (?) ; 3) Furthermore, our di-

Meteorology variables	Available years	Sources
Precipitation (PR)	1980-2015	COOP data (?)
Temperature at 2-meter (T2m)		
Total precipitable water (TPW)	2001-2013	MODIS Daily Global data product (?)
Cloud (liquid plus ice) water path (CWP)	1984-2008	Advanced Very High-Resolution Radiometer (AVHRR) (??)
Shortwave downwelling at surface (SWD)	1984-2010	SRB3.0 (?)
Reflecting solar radiation at top of atmosphere (RET)		
Outgoing long-wave radiation clear/all (OLR)		
Cloud radiative effect (CRE)		

Table 3.2: Observations available years and their references.

agnostic comparison used all respective variables from ERI and NR2. These observational and reanalysis data were all interpolated onto CWRf 30-km grids using a higher-order patch recovery algorithm from the Earth System Modeling Framework regridding package (?) to reduce data mapping error. One important exception was for precipitation, whose daily values from COOP, ERI, and NR2 were all mapped onto the same CWRf grid by mass-conservative algorithms (?). This remapping procedure was to alleviate the impact of data scale mismatch on extreme precipitation analysis (?).

3.3 Extreme precipitation dependence on cumulus parameterization

? demonstrated that the simulation of climatological average seasonal P95 distributions is most sensitive to the choice of the cumulus parameterization among all 25 CWRf physics configurations. This section elaborates performance differences in interannual variability averaged over the GS and CM regions among ERI and CWRf five cumulus schemes (ECP, NKF, TDK, NSAS, BMJ). Here we focused on these CWRf configurations, where only the cumulus parameterization was changed and other physics schemes were kept identical.

Figure 3.2 compares CM and GS regional average P95 interannual variations during 1980-2015 for each season between NR2, ERI and five CWRf cumulus schemes against observations. Also shown are the corresponding temporal correlations and rmse with respect to observations during the whole period. These two statistics differ from those shown in Fig. 2.9, depicting performance in the regional-

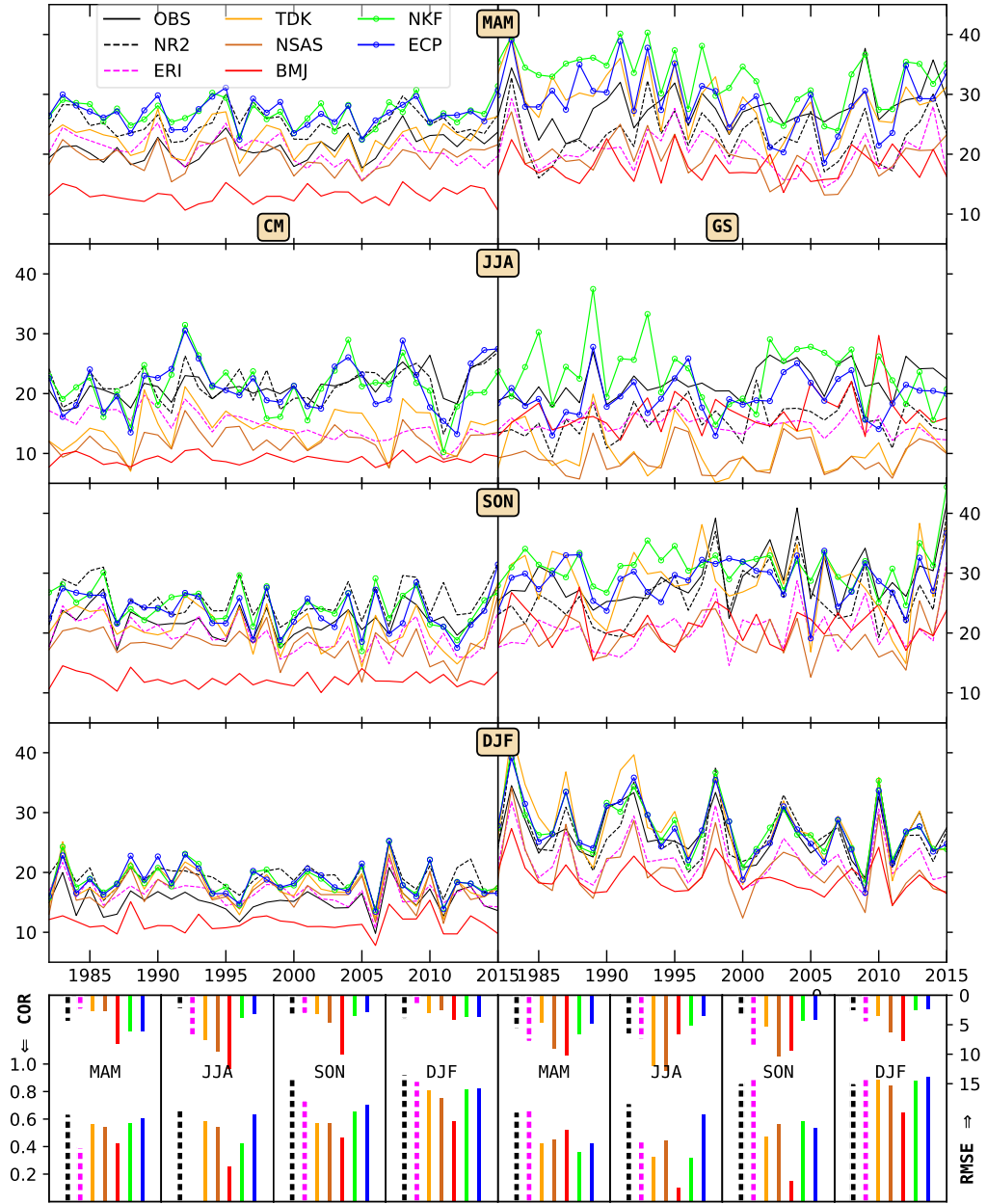


Figure 3.2: Seasonal P95 interannual variations during 1980-2015 averaged over the CM (*left*) and GS (*right*) regions for spring (MAM), summer (JJA), autumn (SON), and winter (DJF) [*from top down*] as observed (OBS), assimilated by NR2 and ERI, and simulated by CWRf using five cumulus schemes (ECP, NKF, TDK, NSAS, BMJ). Also shown (*bottom*) are the corresponding temporal correlations (COR, scaled upward on the *left*) and root mean square errors (RMSE, scaled downward on the *right*) with respect to observations during the whole period.

mean temporal characteristics for the former rather than the time-mean spatial distributions for the latter. In the CM region, ERI was the most realistic in winter, as compared to ECP’s slight overestimation of P95 magnitude and variability. ERI and ECP performed comparably well in autumn. Interestingly, in spring ERI had a notably lower correlation but smaller rmse than ECP. This resulted from ERI’s unrealistic decreasing trend and ECP’s slight overestimation. On the other hand, in summer ERI had a zero correlation and large rmse, whereas ECP was much more realistic. This occurred because ERI not only substantially underestimated P95 but also produced a large unrealistic decreasing trend. In the GS region, ERI’s correlation was notably higher than that of ECP in autumn and spring, comparable in winter but lower in summer. However, in all seasons ERI produced notably larger rmse than ECP due to its substantial systematic underestimation. In the CM region, NR2 performed systematically better than ERI in simulating P95 variations, with higher correlations and smaller rmse in all seasons. The skill difference was especially large in summer when ERI produced significant underestimation and fake decreasing trend. NR2 was also more realistic than ECP in autumn and winter, while they were comparable in spring and summer. In the GS region, NR2 and ECP were comparable in winter and both slightly better than ERI, which underestimated P95. In autumn, NR2 were more skillful than ERI (larger rmse by systematic underestimation) and ECP (smaller correlation by less interannual correspondence). In summer, NR2 and ERI both underestimated P95, having similar rmse larger than ECP, while ERI also incurred a notably smaller interannual correlation than NR2 and ECP. In spring, NR2 had a correlation comparable to ERI and higher than ECP, and an rmse

similar to ECP because of systematic underestimation, which ERI was more severe. Overall, NR2 more faithfully reproduced the P95 annual cycle and interannual variations in both regions than ERI, and hence could serve as a reasonable proxy for the reference when lacking observational data. Given that ERI assimilated pseudo-precipitation data, its interannual correspondence to observations is expected to be high, with a relatively high correlation and small rmse. [The same expectation applies to NR2.] The above comparison indicated otherwise. In cases where deep convection dominates (GS throughout the year and CM in summer and spring), ERI failed to capture the key processes responsible for precipitation extremes (?). Since CWRf did not directly assimilate surface observations but modeled precipitation as driven by ERI general circulation though lateral boundary conditions (??), ECP's skill matching or even exceeding that of ERI is an outstanding achievement. Of all the cumulus schemes examined, ECP performed best overall, with highest correlations and lowest rmse in both CM and GS regions for most seasons. One exception was autumn GS, where NKF produced slightly higher correlation but similar rmse. Another was in spring when TDK's skill was comparable for GS and better for CM (smaller rmse or less overestimation). Overall, NKF scored the second, TDK the third, NSAS the fourth, whereas BMJ was absolutely the worst with substantial systematic underestimation and little coherence to observations. NSAS generally underestimated, except for reasonable spring and underestimated winter in the CM region. TDK's performance was more mixed for both regions: similar to ECP in winter, slightly better than ECP in spring, close to NSAS in summer, and slightly worse than NKF in autumn. These results confirmed that cumulus parameteriza-

tion played a dominant role in extreme precipitation simulations, in terms of not only climatological mean spatial distributions but also regional mean interannual characteristics.

Figure 3.3 compares daily RCT averaged in the CM and GS regions for each season. Daily ratios at individual grids were first grouped according to total precipitation intensity bins at an interval of 5 [mm day⁻¹] and then averaged into those bins over all grids within the CM or GS region. Marked also are the corresponding regional average 1980-2015 mean seasonal P95 values. In a striking contrast, for all seasons NKF produced substantially high ratios while TDK yielded systematically lowest ratios across all precipitation intensities. The RCT ratios from light to heavy rain changed gradually in NKF, but much faster in TDK. Correspondingly, for both regions NKF notably overestimated spring and autumn P95 while TDK severely underestimated summer P95, otherwise they simulated P95 reasonably well. On the other hand, ECP produced intermediate ratios, which remained relatively stable and high for heavy precipitation events; its simulated P95 was most realistic among all schemes except for overestimation in CM spring and winter. To some extent, these results seem to agree with ? in that a balanced convective contribution to total precipitation may be associated with extreme events. Models simulating larger (smaller) convective ratios tend to overestimate (underestimate) P95, especially when convection activities prevail. However, the situation is more complex. ERI, NSAS, and BMJ also produced intermediate ratios, but all substantially underestimated P95 with exceptions only in CM spring and winter for ERI and NSAS. While NR2 closely matched TDK for both RCT and P95 in autumn and winter, they

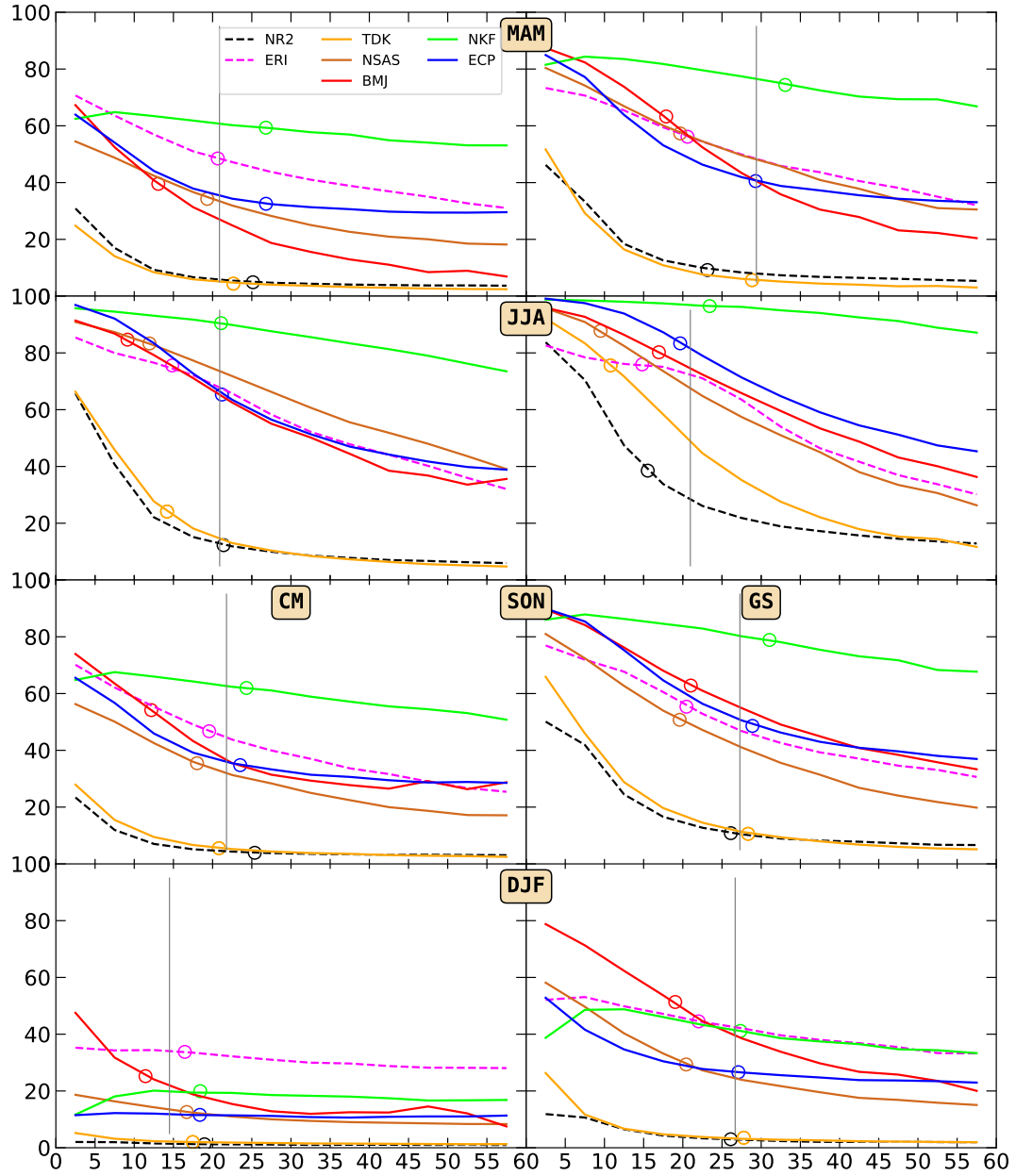


Figure 3.3: Seasonal mean RCT distributions averaged in the CM (*left*) and GS (*right*) regions according to total precipitation intensity bins at an interval of 5 [mm day⁻¹] for spring (MAM), summer (JJA), autumn (SON), and winter (DJF) [from top down] as assimilated by NR2 and ERI, and simulated by five CWRf cumulus schemes (ECP, NKF, TDK, NSAS, BMJ). Marked with circles are the corresponding regional average 1980-2015 mean seasonal P95 values, while the respective observed values are depicted by the vertical lines.

differed significantly in other seasons. In spring, although RCT was very similar, NR2 underestimated (overestimated) P95 in the GS (CM) region while TDK was realistic. In summer, for the CM region with almost same RCT, P95 was correctly simulated by NR2 but largely underestimated by TDK; for the GS region, NR2 produced much smaller RCT but less P95 underestimation than TDK. Therefore, the relative contribution of the parameterized versus resolved rainfall to extreme events may not be as important as originally thought, some other processes must be at play.

3.4 Correlation analysis of regional extreme precipitation biases

Tracing back specific causes for model biases is challenging, but necessary to guide future model improvement. This is particular so for extreme events as they are rare but of high impacts (?). Here we explored possible physical processes that lead to CWRF performance differences in extreme precipitation simulation. To do so, we first analyzed possible relationships between biases in P95 and other fields on extreme event days. For a given simulation of ERI or CWRF, we first identified the date of the P95 event in each season of a year at a specific grid. Then at that grid and respective to the dates of all P95 events as a function of season and year, we calculated the biases of simulated precipitation and their relevant variables from the corresponding observations. Although the specific date when the P95 event occurred in a season of a year was expected to differ from grid to grid, the composite biases retained the coherence among the variables in concern. These biases were averaged

over the CM and GS regions to yield interannual variations in each season and for every simulation of ERI and CWRf’s five cumulus members. The subsequent statistics (correlation coefficient and significance p-value) were based on these regional mean interannual variations of all six simulations between precipitation and each of the other variables for the entire period with observational data records as listed in Table 2. Thus, the total number of samples used in the statistics was 6 simulations times (25, 27, 36) years or (150, 162, 216) for (CWP, SWD/RET/OLR/CRE, P95/T2m) variables. We also included bias correlations with DRI and NRD, which were calculated from the same COOP precipitation data as for P95, and by definition seasonal values rather than on specific P95 event dates. All other fields, while essential for process understanding, lacked daily observational data and hence their departures from the respective NR2 result were used, having 6 simulations times 36 years or 216 samples. For any pair of variables, the correlation could be based on their biases if both have observations or otherwise their departures. This mixture, however, would cause two issues. First, the P95 event dates simulated by NR2 likely differ from those observed, and so the biases and departures would be based on different events between the fields with and without observational data. Second, the cross-field correlations and subsequent regression models would have to build upon the shortest common period (25 years or 150 samples), and so the result would suffer from large uncertainties. To alleviate these issues, we first established the consistency between the bias and departure correlations of P95 with those fields that had observational data, then examined the functional relationships among the departures of all fields, and finally built the regression models on these departures in

Section 5 to best capture P95 structural dependences on the key contributing fields. Correlations are considered as statistically significant if they pass the student's t-test at the one-tail 95% confidence level.

Figure 3.4 compares the above P95 composite bias and departure correlations with those fields that had observational data in all seasons, while Figs. 3.5-3.8 compare the composite departure correlations across all ingredient fields in respectively spring, summer, autumn, and winter. In both (CM, GS) regions, P95 and DRI biases were strongly correlated in phase, with coefficients increasing from summer (0.62, 0.71), autumn (0.80, 0.79), spring (0.89, 0.87) to winter (0.89, 0.89). A model simulating stronger (weaker) mean rainfall intensity likely produced higher (lower) extreme precipitation. The correlation with NRD was positive only for GS summer (0.35), reversed in winter (-0.41) and spring (-0.21), and insignificant in autumn; it was systematically much weaker for CM, significant only in summer (0.26) and winter (-0.23). Thus a simulation biased with more rainy days tended to overestimate summer but underestimate winter (and weakly spring) extreme precipitation; this tendency was more evident in the GS than CM region, and not obvious in autumn. Most of these bias correlations were faithfully reproduced by the departures from NR2 in both sign and magnitude.

The correlation between P95 and parameterized or convective precipitation (PC) departures was strong for both (CM, GS) regions in summer (0.54, 0.68), significant in spring (0.47, 0.48) and autumn (0.36, 0.43), and reversed or insignificant in winter (-0.21, -0.13). On the other hand, the correlation with resolved or explicit precipitation (PL) departures was significant for both (CM, GS) regions only in

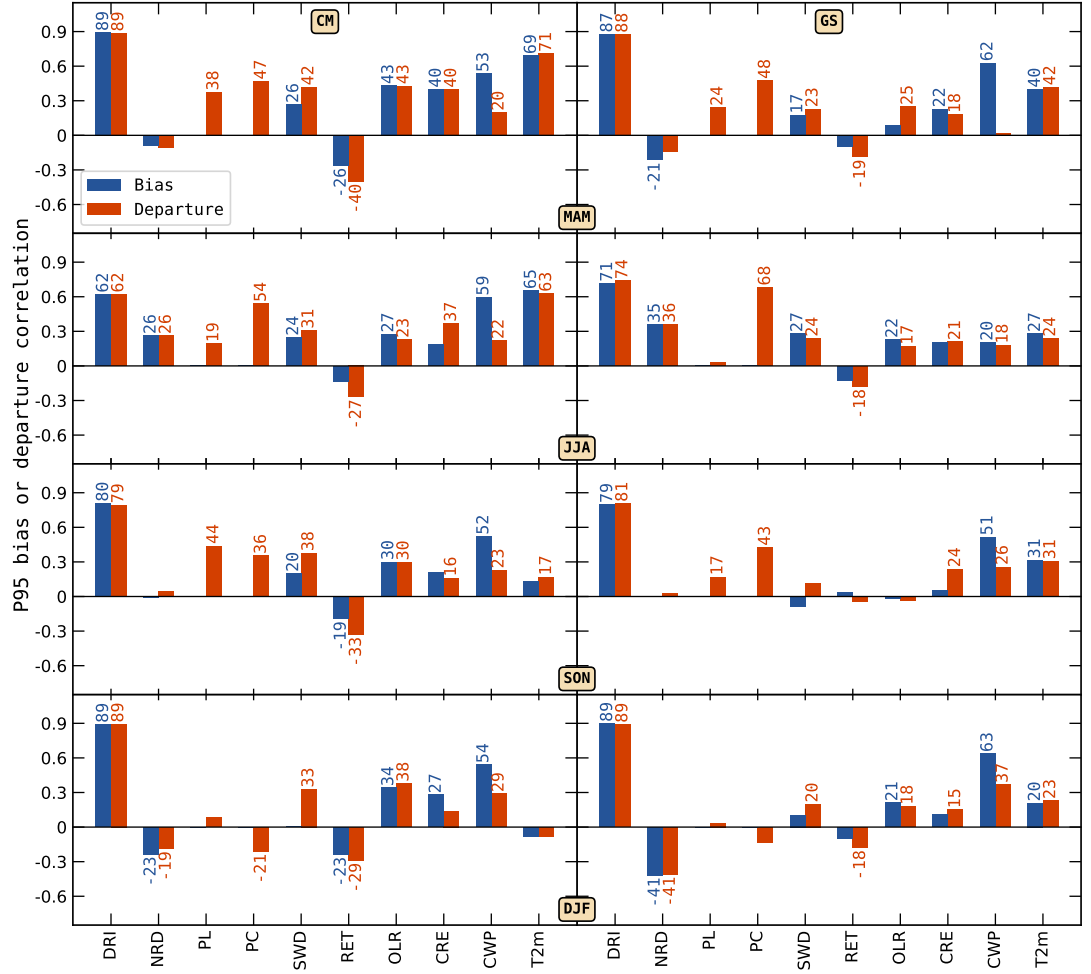


Figure 3.4: Composite P95 bias (*blue*) and departure (*red*) correlations with those fields that had observational data (DRI, NRD, SWD, RET, OLR, CRE, CWP, T2m) as well as P95 departure correlations with rainfall components (PL, PC) in the CM (*left*) and GS (*right*) regions for spring (MAM), summer (JJA), autumn (SON), and winter (DJF) [from top down]. A star mark indicates the correlation is statistically significant. They are labeled with a number equal to the correlation coefficient times 100 if statistically significant.

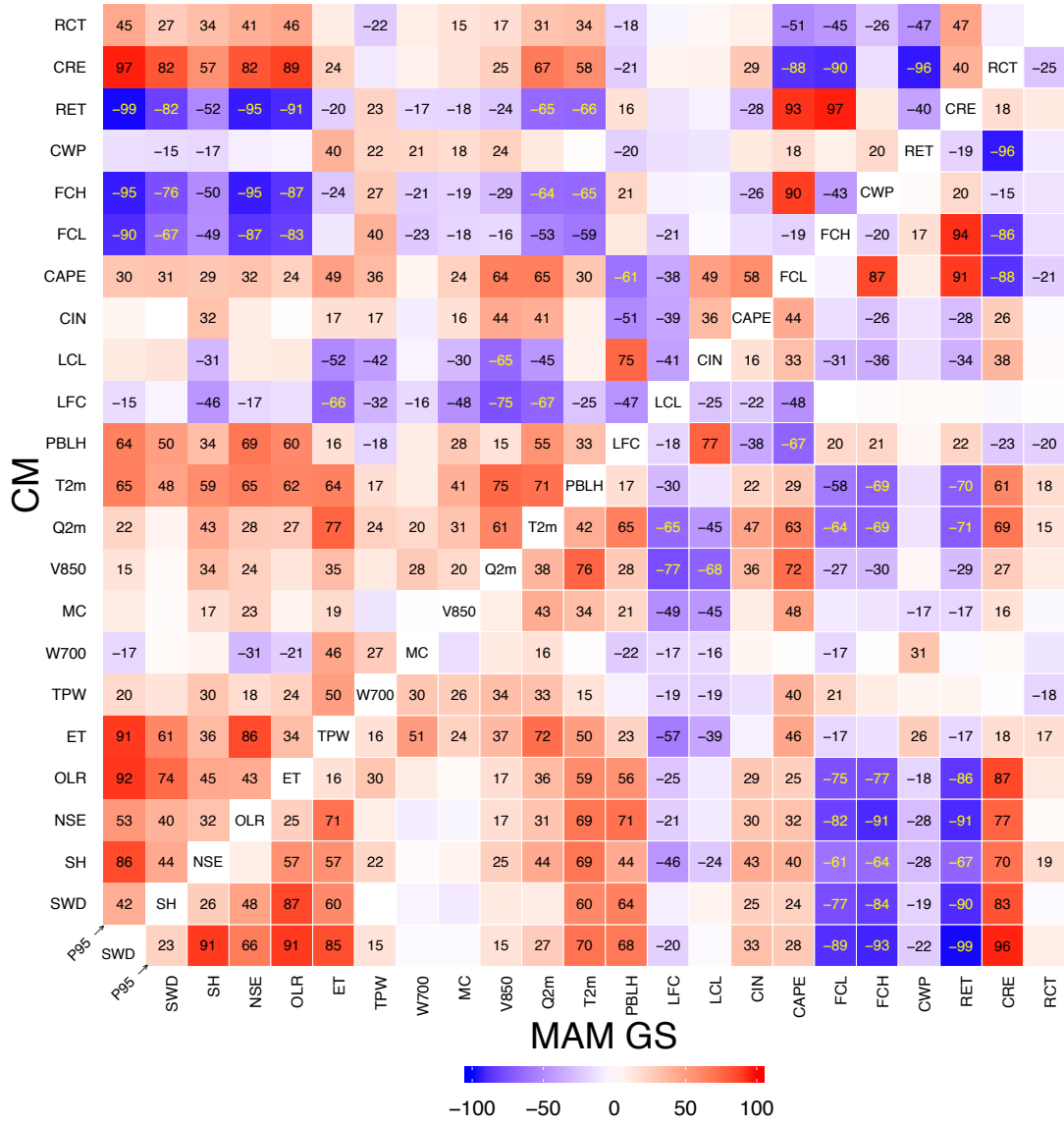


Figure 3.5: Spring composite departure correlations across P95 and its all ingredient fields in the CM (*upper triangle*) and GS (*lower triangle*). They are coded each with a color and, if statistically significant, also labeled with a number equal to the correlation coefficient times 100. The diagonal represents the P95 correlation with each field named.

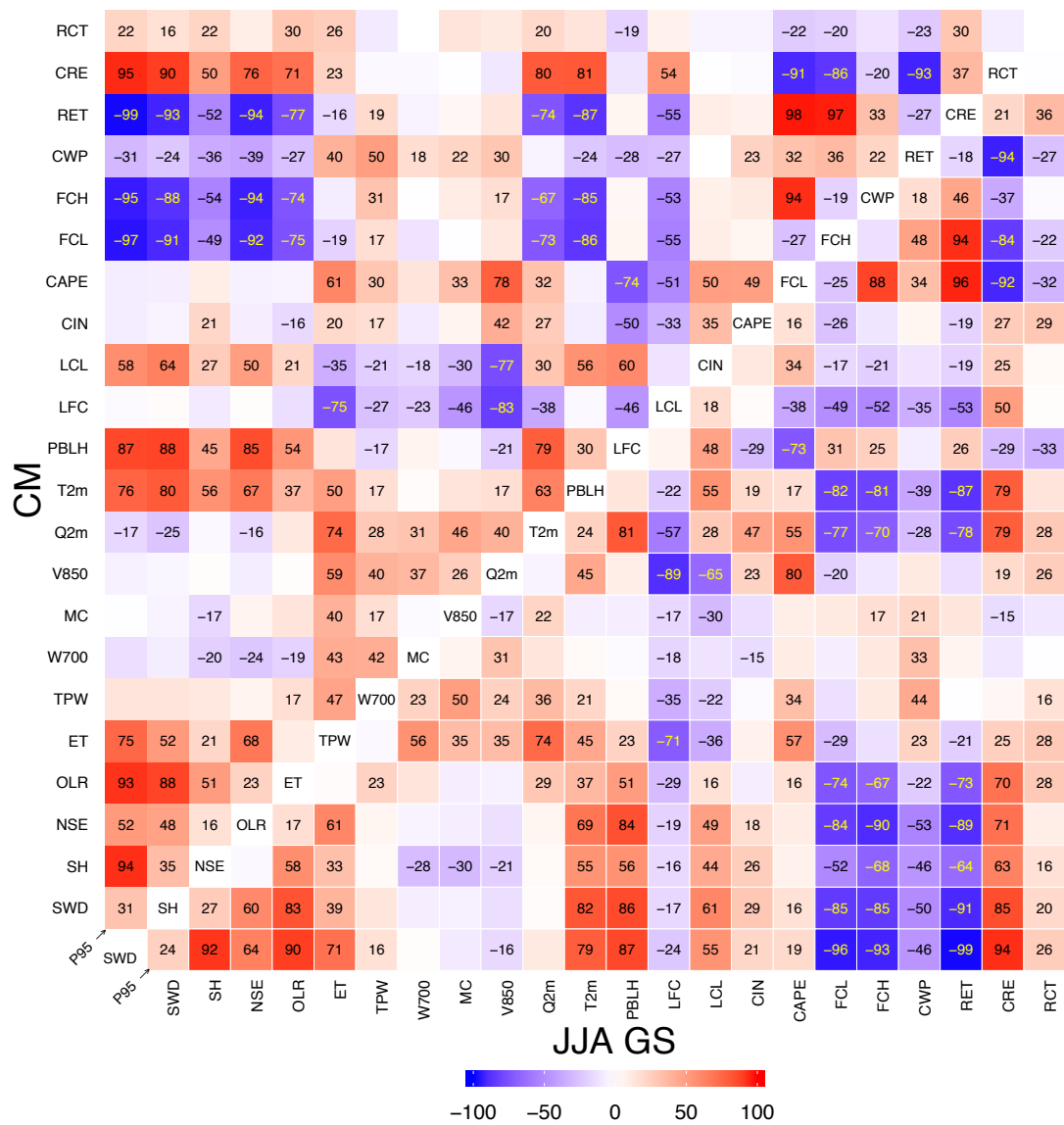
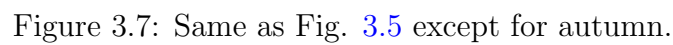
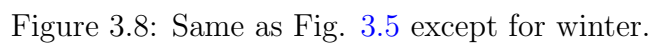


Figure 3.6: Same as Fig. 3.5 except for summer.





spring (0.38, 0.24) and autumn (0.44, 0.17), and CM summer (0.19). Therefore, for extreme precipitation simulation, cumulus parameterization is more sensitive than cloud microphysics, especially in summer. However, P95 and RCT departure correlations were negative and significant for both (CM, GS) regions in autumn (-0.46, -0.20) and winter (-0.51, -0.36), and GS spring (-0.25). As compared to the P95-PC correlations, they were drastically weakened in both summer and spring, totally reversed in autumn, and largely strengthened in winter. As discussed earlier, the RCT role on P95 was very much mixed.

Positive correlations between P95 and SWD biases were significant only for CM summer, spring and autumn (0.24, 0.26, 0.20) and GS summer and spring (0.27, 0.17). These bias correlations were generally captured by the corresponding departures from NR2, although systematically strengthened for CM spring to winter (0.42, 0.31, 0.38, 0.33) and GS spring, summer and winter (0.23, 0.24, 0.20). Positive correlations between P95 and ET departures were significant for both (CM, GS) regions only in spring (0.34, 0.16) and autumn (0.25, 0.20). In these transition seasons, the regional water recycling process plays a certain role in extreme precipitation formation, and a simulation with insufficient (excessive) water supply through ET could underestimate (overestimate) P95. Positive correlations between P95 and SH departures were significant for the CM region from spring to winter (0.44, 0.35, 0.43, 0.33), while generally reduced for the GS region (0.26, 0.27, 0.18, 0.32). Hence, surface energy supply through SH could affect P95 significantly. These results were coherent since SWD is the dominant source for ET and SH, although their effects on extreme precipitation strongly depended on regions and seasons due to other

feedback processes. P95 and NSE departures were correlated significantly only in CM from spring to autumn (0.32, 0.16, -0.36), again indicating feedback effects.

Positive correlations between P95 and OLR biases were significant in the CM region from spring to winter (0.43, 0.27, 0.30, 0.34). They were largely reduced in the GS region, significant only in summer (0.22). Meanwhile, negative correlations between P95 and RET biases were significant only in CM spring and winter (-0.26, -0.23). Consequently, positive correlations between P95 and CRE biases were significant only in CM spring winter (0.40, 0.27) and GS spring (0.22). Deeper convection produced smaller OLR that overcame larger RET to reduce CRE, a net warming effect to the earth. Underestimation of deep convection could reduce extreme precipitation, more so in the CM than GS region. As compared to these bias correlations, P95 correlations with OLR departures were very close in the CM region from spring to winter (0.43, 0.23, 0.30, 0.38), while in the GS region strengthened in spring (0.25), slightly weakened in summer and winter (0.17, 0.18), and still insignificant in autumn. For RET departures, the correlations were generally strengthened in the CM region from spring to winter (-0.40, -0.27, -0.33, -0.29), and also in the GS region in spring, summer and winter (-0.19, -0.18, -0.18) but still insignificant in autumn. For CRE departures, the correlations in both (CM, GS) regions were close in spring (0.40, 0.18), strengthened in summer (0.37, 0.21) and autumn (0.16, 0.24), and in winter only significant in GS (0.15). Considering large uncertainties in the satellite estimates, these departure and bias correlations agreed reasonably well.

Positive correlations between P95 and CWP biases were strong in the CM region from spring to winter (0.53, 0.59, 0.52, 0.54), and also in the GS region except

for weaker summer (0.62, 0.20, 0.51, 0.63). This agrees with ? in that the cloud microphysics process is important to extreme precipitation simulation. Positive correlations were also shown by the corresponding departures from NR2, although systematically weakened for CM spring to winter (0.20, 0.22, 0.23, 0.29) and GS summer to winter (0.18, 0.26, 0.37). Such departure and bias correlation differences could be partly due to large uncertainties in the satellite estimates. Positive correlations between P95 and TPW departures were strong in spring and summer for the CM region (0.50, 0.47) but systematically reduced for the GS region (0.16, insignificant); they were weaker in autumn and winter for the CM region (0.25, 0.19) and stronger for the GS region (0.18, 0.29). This agrees with Kunkel et al. (2013a) in that TPW is generally a good indicator for the upper limit of extreme precipitation. These results were consistent since atmospheric water vapor abundance (TPW) is necessary for cloud liquid and ice water (CWP) formation, and their impacts on P95 were systematically enhanced from the GS to CM region mainly because water supply is typically more limited (so stronger dependence) over inland than coastal areas.

Correlations between P95 and T2m biases were positive for the CM region, strong in spring (0.69) and summer (0.65) but insignificant in autumn and winter. They were reduced for the GS region in spring (0.40) and summer (0.27) but increased in autumn (0.31) and winter (0.20). These bias correlations were faithfully reproduced by the departures from NR2 in both sign and magnitude. In both regions and for all seasons, strong T2m departure correlations were found with NSE, Q2m, TPW, PBLH, and CAPE (positive) as well as LFC (negative), in the mag-

nitude range of 0.40~0.81 with very few exceptions. Consistently, greater NSE produced warmer T2m, wetter Q2m, more TPW, higher PBLH, lower LFC, and larger CAPE. In spring and summer, strong T2m departure correlations were also found with SWD, SH, ET, and OLR (positive) as well as FCL, FCH, and RET (negative), in the magnitude range of 0.48~0.82. Greater SWD led to warmer T2m, larger SH and ET fluxes, less FCL and FCH clouds, and so reduced RET and OLR. On the other hand, these correlations in autumn and winter were totally reversed in sign, and most often weakened in magnitude and some became insignificant. Therefore, surface-atmospheric and cloud-radiative interactions change substantially from spring and summer to autumn and winter, when the regional precipitation processes change from the dominance of deep convections to stratiform systems.

Positive correlations between P95 and PBLH departures were significant in CM spring, summer and winter (0.33, 0.30, 0.23), and GS spring and autumn (0.17, 0.21). They were identified with strong PBLH links to T2m (discussed above) and other fields. In spring and summer, strong PBLH departure correlations were found in both regions with SH, ET, SWD, OLR and CRE (positive) as well as FCL, FCH and RET (negative), in the magnitude range of 0.50~0.88. Greater SWD, SH and ET caused higher PBLH, elevated cumulus base (smaller FCL) while reduced cloud depth (smaller FCH), and thus resulted in less RET and more OLR and CRE. In autumn and winter, with a few exceptions, they were substantially reduced in magnitude, and some became insignificant (especially in CM) or changed in sign. This again could be identified with the seasonal change from convective to stratiform precipitation dominance.

The above analysis showed that P95 correlations with key fields' biases from observations were well captured by those with the corresponding departures from NR2. The bias and departure correlations resembled closely in both interannual variations and seasonal contrasts. Thus, we can analyze the relationships across the simulated departures from NR2 to determine the physical processes that may cause P95 biases from observations. Such a causal analysis would be impossible if based on observational data that were available to only a few fields. In addition, Figs. 3.4-3.8 contain much more complicated cross-field relationships than we discussed, and require advanced machine learning techniques to disentangle them in order to identify the most plausible mechanisms underlying the P95 departures or biases.

3.5 Process understanding of P95 biases by structural equation modeling

The previous two sections illustrated that the regional precipitation extremes resulted from variations in a complex coupled climate system, whose component processes interacted each other through direct and indirect effects. Among the select fields (Figs. 3.5-3.8) representing these processes existed strong correlations. They could act and/or counteract on extreme precipitation formation. Our objective was to build robust regression models of these fields to quantify their relative contributions and interpret the underlying processes affecting P95 simulation. Direct inclusion of all these highly correlated fields as predictors would cause multicollinearity, leading to unstable regression models that may suffer from double-counting partic-

ular physical factors and overfitting certain numerical parameters. These models can violate the parsimony principle and miss the big picture that the reality may present. Here we used the structural equation model (SEM) framework to solve the problem of multicollinearity. The framework is an extension of the confirmatory factor analysis and has the ability to test hypotheses on causal relationships in the presence of physically based experimental design (?).

An SEM consists of manifest (measurable) and latent (hypothetical) variables. All the fields listed in Figs. 3.5-3.8 were considered as the candidates for manifest variables. Based on the physical understanding from the previous two sections, we constructed four latent variables: energy supply, water supply, surface forcing, and cloud forcing. As discussed earlier, sustained energy and water supplies directly power the climate system processes for extreme precipitation. While surface forcing acts to couple surface energy and water sources with atmospheric precipitation processes, cloud forcing works to regulate such surface-atmospheric coupling in order to balance the energy and mass budgets of the earth system. Given P95 as the predictand and 22 manifest variables as the predictor candidates to choose from, huge flexibility exists in constructing these latent variables, including their member predictors and directional effects. Figure 3.9 illustrates our conceptual design of the experimental SEM for extreme precipitation, where each latent variable was designated with 4 or 6 exclusive predictor candidates that are strongly correlated. Energy and water supplies may each affect both surface and cloud forcings, and surface forcing may also affect cloud forcing, while all the four latent variables may finally affect P95. Therefore, the SEM reduces the dimensionality by using designated grouping

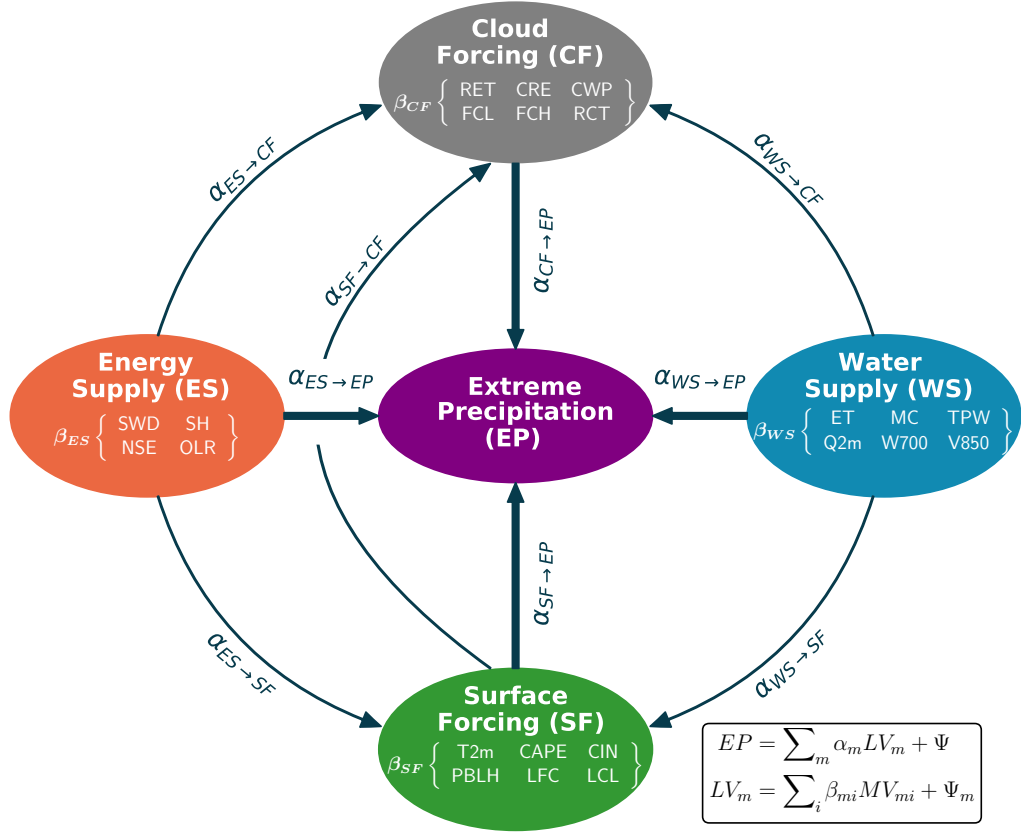


Figure 3.9: The conceptual design of the experimental SEM for extreme precipitation (EP). The center oval represents the predictand (EP), while each outer oval defines one latent variable (LV) with a list in a brace of the designated manifest variables (MV) as the predictor candidates. There are four LVs, hypothetically representing energy supply (ES), water supply (WS), surface forcing (SF), and cloud forcing (CF). Each effect from one LV to another or to EP is expressed by an arrow for its direction and a coefficient along the line for its strength. The SEM consists of regression equations from these MVs through LVs to EP, as depicted at the lower right corner.

to substitute the 22 manifest variables into the 4 latent variables. The dimensionality can be further reduced through system optimization that maximizes the SEM performance while minimizing the model complexity. The SEM, as a network of knowledge, can be made more robust by designating into each latent variable more tightly connected manifest variables to reduce the uncertainty of the group and so the whole network. We did this designation based on the physical understanding and correlation analysis discussed earlier with three additional considerations: the manifest variables selected for each latent variable were highly correlated so to enhance the unidimensionality of the group; they were close to each other on the hierarchical clustering dendrogram and so in the multidimensional space; they were consistent as much as possible among different seasons and regions. The resulted group of the manifest variables (exclusive) was listed for each latent variable in Fig. 3.9. They were linked to P95 through the respective latent variables and the latter were coupled to form the SEM. These links and couplings were made by a set of regression equations, each containing an error term. It is important to note that these error terms are not orthogonal to other regression fields, and may contain a significant portion of their covariance unexplained. The SEM so constructed represents the integrated impact on P95 through the active latent variables, each of which consists of the variations manifested coherently in the select measured fields as they are significantly related to P95. The manifest variables listed were still just the candidates, not the finalists. Assume each latent variable consist of at least two manifest variables. The total number of possible combinations for a single latent variable is the sum of choosing 2 to 4 (6) out of 4 (6) manifest variables, that is, 11

(57). At the latent variable level, one or both of energy and water supplies act on surface forcing, any combination of these three in turn affects cloud forcing, and any of these four finally impacts P95. Thus, there are 3, 7 and 15 combinations respectively for surface forcing, cloud forcing and P95. These counts, however, consider only direct effects between two paired variables, while their mixtures can exert an exponentially increased number of indirect effects. Together, there are 230,938,920 direct plus indirect effects on P95 or ~ 231 millions of alternative SEMs in each region (CM or GS) and each season. This study used the open-source ‘lavaan’ software version 0.6-3 (?) on the ‘R’ platform to construct the SEMs through unconstrained optimization, adopting the default algorithm configuration for all parameters and settings (including the maximum likelihood estimator). The goal of the optimization is to minimize the difference between measured and implied cross-covariance matrix and, as stated by ?, to “discover a model with three attributes: it makes theoretical sense, it is reasonably parsimonious, and it has acceptably close correspondence to the data”. Given the huge number of alternatives, the whole process of the optimization (brute-force search) requires machine learning via supercomputing to construct potential SEMs. Many of these alternatives failed to reach a solution with stable regression coefficients, and so were filtered. The number of successfully constructed SEMs was still tremendous. We therefore searched for the final SEM under the following conditions. First, the sign of total (direct plus indirect) effect of each included manifest variable on P95 must be preserved same as their original correlation if significant (Figs. 3.5-3.8). Second, the total explained variance (R^2) of P95 departures must be greater than 0.8. [All departures were

normalized with a zero mean and unit deviation.] Third, the comparative fit index (CFI, χ^2/df), a most popular indicator for the goodness of fit (χ^2/df), must be larger than 0.9. Fourth, a bootstrap resampling procedure must be succeeded to estimate the uncertainty range of every regression coefficient across all acting latent variables and P95; otherwise the SEM would be unstable in its structure, containing substantial uncertainties. Finally, the Akaike information criterion (AIC), an integrated measure of both model fitness and complexity, should be the lowest for the most preferred SEM that fits best but keeps parsimony (to avoid overfitting). ΔAIC indicated the need to consider other competitors only if their AIC differences from the minimum are smaller than 10.

The above selection rules led to a unique choice of a single best SEM for each region (CM or GS) and each season. Figures 3.10-3.13 illustrate these finalist SEMs as paired by the CM and GS regions for individual seasons. They include the active manifest variables to each latent variable, the strength (coefficient) and direction (arrow) of each effect, the four performance scores (R^2 , CFI, AIC and its increment to the next competitor ΔAIC). The net effect from a latent variable onto P95 is its direct effect (coefficient on immediate arrow line) plus indirect effects (product of all coefficients along each directional path) (χ^2/df). Below the result is interpreted in terms of the relative importance to P95, and is expressed as the percentage of any effect's coefficient over the sum of the net effect's absolute coefficient from all active latent variables. It is stated as [RI = %] in the text and also shown in Figs. 3.10-3.13. In CM spring (Fig. 3.10), the P95 departure was dominated by a positive direct effect of the surface forcing departure [RI = 85%], where greater CAPE and higher

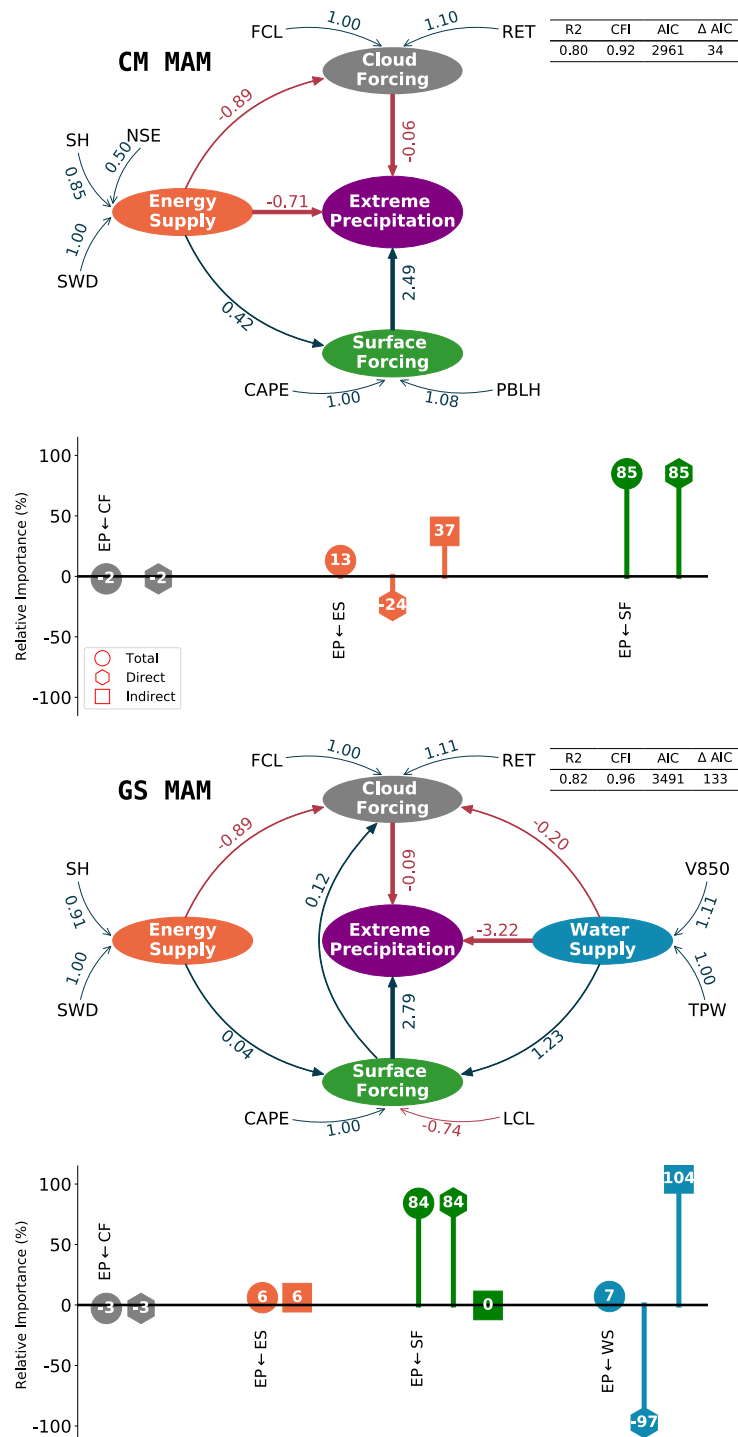


Figure 3.10: Spring finalist SEMs for CM (*upper*) and GS (*lower*). Each SEM panel includes its structure (*left*) with the active manifest and latent variables and their directional effect coefficients, its performance scores (*upper right corner*), and the relative importance of each latent variable's direct, indirect and total effects (*bottom*).

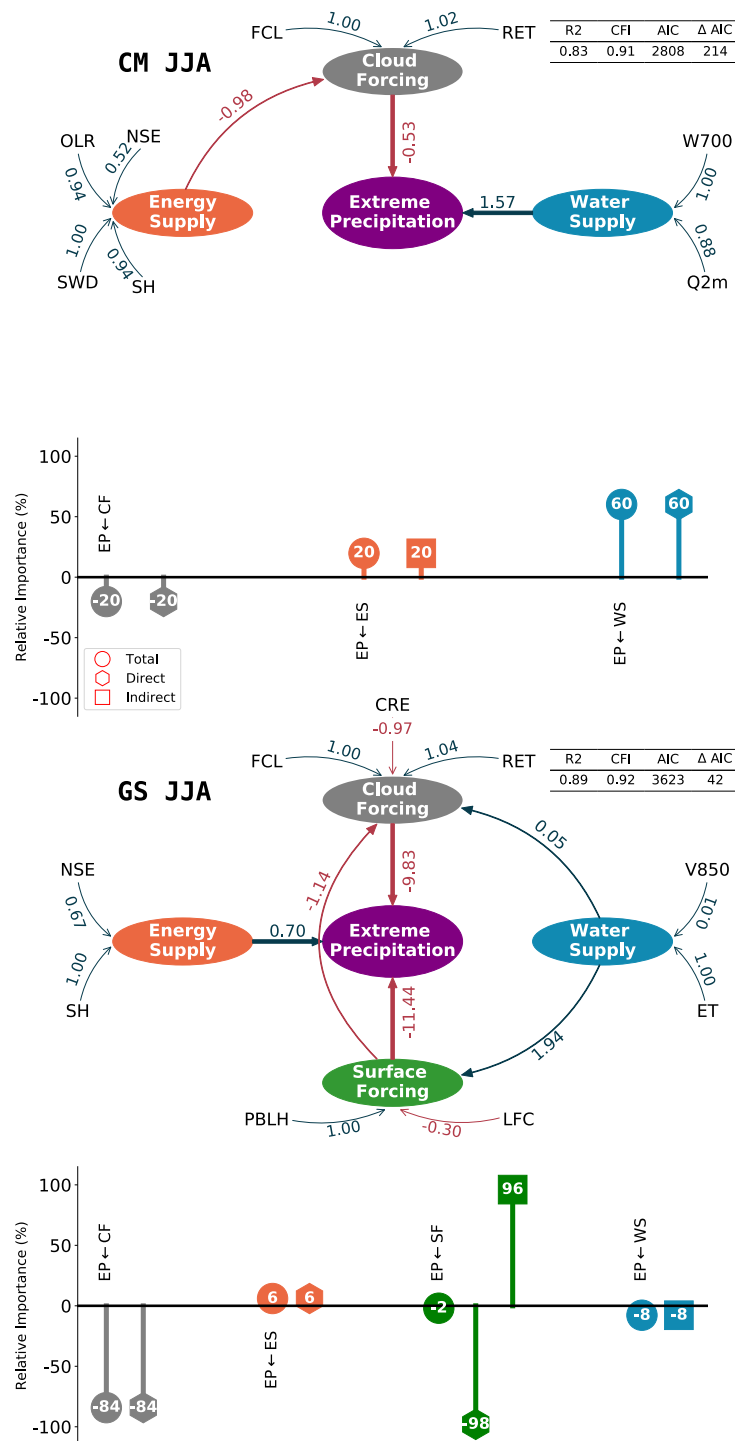


Figure 3.11: Same as Fig. 10 except for summer.

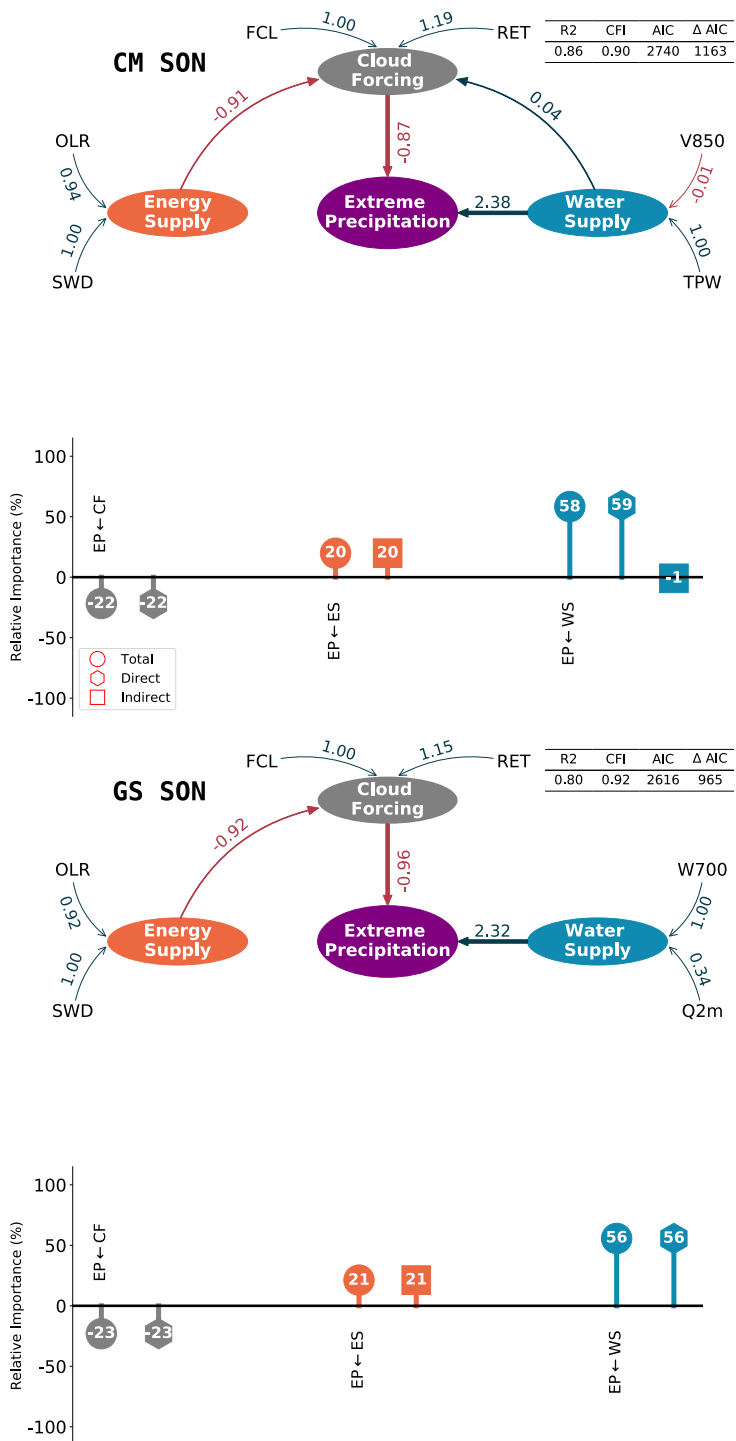


Figure 3.12: Same as Fig. 10 except for autumn.

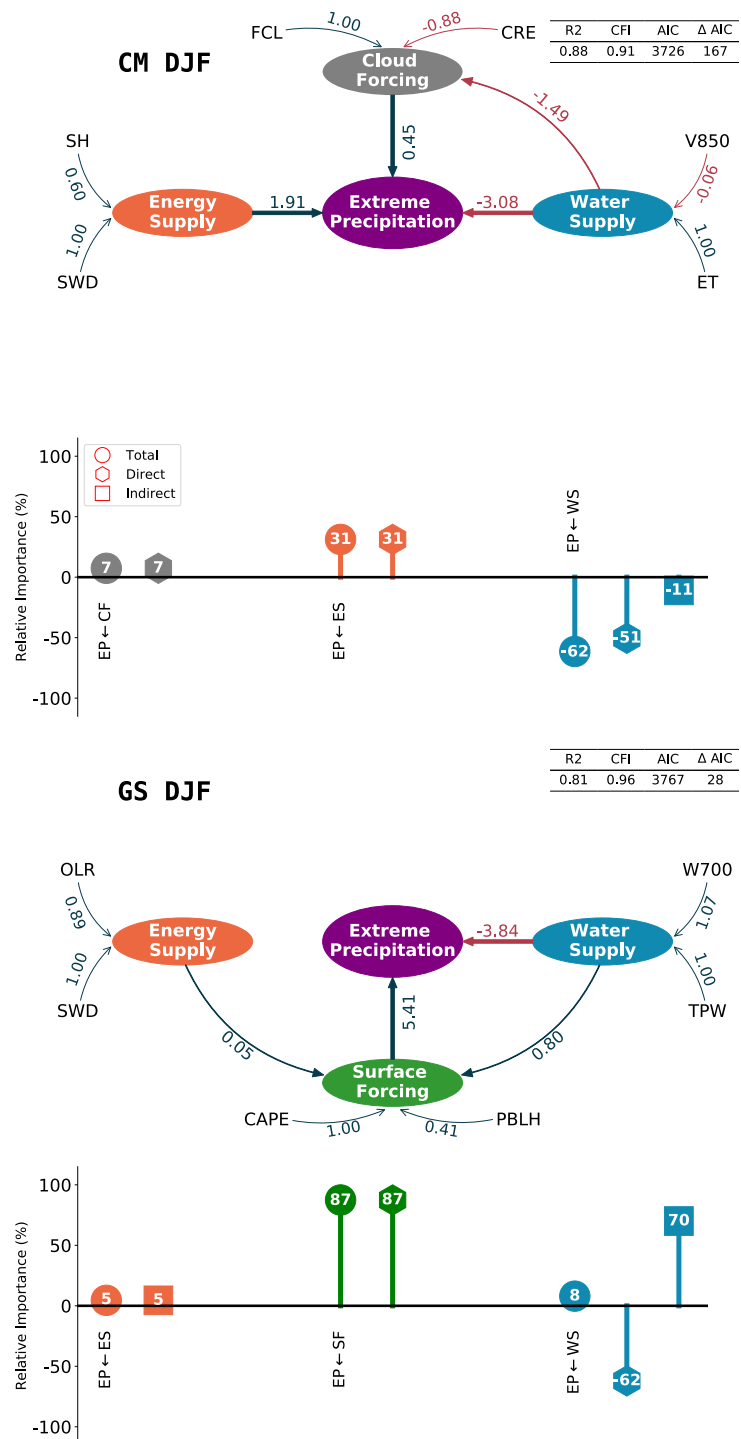


Figure 3.13: Same as Fig. 10 except for winter.

PBLH led to larger extreme precipitation. Such surface forcing was supported by larger energy supply [RI = 13%], which consisted of more solar radiation incoming to the surface (SWD), larger sensible heat release to the atmospheric boundary layer (SH), and $[1/2]$ net surface energy surplus (NSE). Meanwhile, precipitation directly consumed energy supply [RI = -24%]. Larger precipitation also depleted more low-level cloud (FCL), reflected less solar radiation (RET), and so reduced cloud forcing [RI = -2%], which in turn increased energy supply, causing a small positive feedback [RI = 2%]. Combining its direct and indirect (through mainly surface and trivially cloud forcings) effects, energy supply impacted P95 with total importance of 13%, which was smaller than surface forcing even without its support [RI = 85-22 or 63%]. Therefore, surface forcing and energy supply were respectively the first and second most important factors determining P95 in CM spring, while cloud forcing played a very minor role and water supply had negligible influence. In GS spring (Fig. 3.10), the P95 departure was dominated by a positive direct effect of the surface forcing departure [RI = 84%], where greater CAPE and lower $[3/4]$ LCL drove larger extreme precipitation. The latter depleted more low-level cloud (FCL), reflected less solar radiation (RET), and so reduced cloud forcing [RI = -3%], which in turn reduced surface forcing, causing a tiny negative feedback. Meanwhile, larger precipitation removed more water from the atmospheric column (TPW) and the moisture transport (V850), and thus directly consumed most of water supply [RI = -97%]. The latter, however, was replenished by surface forcing of a greater amount. Since its direct and indirect (through mainly surface and trivially cloud forcings) effects almost canceled each other, water supply impacted P95 with net

importance of only 7%. In contrast, energy supply (consisting of SWD and SH) had only indirect effects through surface and cloud forcings (both positive) on P95 with total importance of just 6%. Therefore, surface forcing was the predominant factor determining P95 in GS spring [84%], cloud forcing had a very minor effect [-3%], while both energy and water supplies shared evenly the remaining portion. In CM summer (Fig. 3.11), the P95 departure was dominated by a positive direct effect of the water supply departure [RI = 60%], where stronger upward motion (W700) and higher surface humidity (Q2m) led to larger extreme precipitation. The latter depleted more low-level cloud (FCL), reflected less solar radiation (RET), and so directly reduced cloud forcing [RI = -20%], which in turn indirectly increased energy supply, causing a strong positive feedback [RI = 21%]. The small surplus of energy supply over cloud forcing implied that the combined effect of SWD, SH, OLR, and [1/2] NSE outweighed that of RET and FCL by a little. Therefore, water supply and energy supply or cloud forcing were respectively the first and second most important factors determining P95 in CM summer, while surface forcing had negligible influence. In GS summer (Fig. 3.11), the P95 departure was dominated by a negative direct effect of the cloud forcing departure [RI = -84%], where larger precipitation depleted more FCL, reflected less RET, and yielded smaller CRE. Meanwhile, cloud forcing had a strong positive indirect feedback from surface forcing [RI = 97%], where mainly lower PBLH and secondarily [1/3] higher LFC lead to larger extreme precipitation. This indirect effect of surface forcing canceled most of its direct effect to produce a tiny negative net impact on P95 [RI = -2%]. In addition, energy supply, consisting of mainly SH and [2/3] NSE, exerted a weak positive direct effect

[RI = 6%]. On the other hand, water supply, mainly from the regional recycling ET, had only negative indirect effects through primarily surface and trivially cloud forcings [RI = -8%]. Therefore, cloud forcing was the predominant factor determining P95 in GS summer, while energy and water supplies exerted much weaker effects, and surface forcing had a trivial impact. The physical mechanism for the P95 departure in both CM and GS autumn (Fig. 3.12) was essentially identical to that in CM summer (Fig. 3.12). In all cases, water supply and energy supply or cloud forcing were identified as respectively the first and second most important factors determining P95, while surface forcing had negligible influence. Their corresponding RI values were [60%, 20%, -20%] for CM summer, [58%, 20%, -22%] for CM autumn, and [56%, 21%, -23%] for GS autumn. The manifest variables were identical (FCL, RET) for cloud forcing, common (SWD, OLR) for energy supply except with additional (SH, NSE) in CM summer, and same (W700, Q2m) for water supply except with a replacement (TPW, 1/100 V850) in CM autumn. The last subtle change was the additional tiny indirect effect from water supply through cloud forcing in CM autumn [RI = -1%]. In CM winter (Fig. 3.13), the P95 departure was determined by two opposite direct effects: negative water supply (ET, 3/50 V850) [RI = -51%] and positive energy supply (SWD, 3/5 SH) [RI = 31%]. The former also had a negative indirect effect through cloud forcing [RI = -11%], and hence caused the total effect [RI = -62%] even stronger than energy supply. Larger extreme precipitation was maintained by more surface energy supply, which consisted of primarily more incoming solar radiation (SWD) and secondary larger sensible heat release (SH). Meanwhile, precipitation directly consumed water supply

that predominantly recycled from surface evapotranspiration (ET). Strikingly, the direct effect of cloud forcing in winter was positive [RI = 7%], totally opposite from other seasons. Hence, larger low cloud amount (FCL), which reduced CRE, actually resulted in greater extreme precipitation. This is reasonable since winter precipitation is dominated by stratiform systems, where sustained water supply maintains low clouds while steadily precipitating. In contrast, convective precipitation prevails in other seasons, depleting clouds much faster. Therefore, water and energy supplies were the two critical counteractive factors determining P95 in CM winter, while cloud forcing played a secondary but positive role and surface forcing had negligible influence. In GS winter (Fig. 3.13), the P95 departure was dominated by a positive direct effect of surface forcing departure [RI = 87%], where greater CAPE and higher [2/5] PBLH led to larger extreme precipitation. Such surface forcing was supported by larger water supply [RI = 70%], which consisted of higher atmospheric water content (TPW) and stronger upward motion (W700). Meanwhile, larger precipitation directly consumed more water supply [RI = -62%]. Due to the near cancelation between its direct (negative) and indirect (positive) effects, water supply had only a small net impact on P95 [RI = 8%]. Energy supply had an even smaller effect [RI = 5%], which consisted of more incoming solar radiation (SWD) and outgoing longwave radiation (OLR). Therefore, surface forcing was the predominant factor determining P95 in GS winter, while water and energy supplies exerted much weaker effects, and cloud forcing had a trivial impact.

3.6 Summary and conclusions

In this study, we took the challenge to uncover the physical mechanisms that can explain how cumulus parameterization determines CWRP’s ability in simulating U.S. extreme precipitation as identified in our companion paper (?). The challenge arose from the rareness of extreme events, the lack of observational data, and the complexity of physical processes. To disentangle the problem, we made three key analyses. First, we analyzed interannual variations of spatial averages over the two distinct regions, CM and GS, where ERI substantially underestimated climatological mean P95 while CWRP realistically captured that. We found that, of all the cumulus schemes tested in CWRP, ECP best simulated P95 interannual variability (with highest correlations and lowest rmse) in both CM and GS regions for most seasons, and also performed generally better than ERI. Hence, cumulus parameterization affected significantly extreme precipitation simulation in terms of not only climatological mean spatial distributions but also regional mean interannual characteristics. However, we found that the relative contribution of the parameterized versus resolved rainfall (RCT) to extreme events were not as important as originally thought in the literature. In addition, we showed that NR2 more faithfully reproduced the P95 annual cycle and interannual variations in both regions than ERI, and hence we adopt it as the proxy reference when lacking observational data. Second, we analyzed interannual correlations of P95 biases (from observations) and/or departures (from NR2) with those of seasonal statistics (DRI, NRD) and of the P95 event-based composites (22 key fields). The composite was made for each field in

a same simulation by first identifying the date when the P95 event occurred in a season of a year at a specific grid and then averaging the field’s bias or departure on that date over all grids within the CM or GS region. We showed that the P95 bias correlations with all the fields that had good observational data (DRI, NRD, T2m, SWD, OLR, RET, CRE, CWP) were well captured by the corresponding departure correlations in both interannual variations and seasonal contrasts. Thus, it is reasonable to assume that the relationships underlying the simulated departures from NR2 could represent the mechanisms responsible for P95 biases from observations. We found that the departures of all six simulations, that is, ERI and CWRf’s five cumulus members (ECP, NKF, TDK, NSAS, BMJ), contained significant correlations across P95 and many of the 22 fields. These correlations, however, varied greatly in sign and magnitude, from -0.99 to +0.97, depending on season and region, and also were interdependent across multiple fields. They could act or counteract on extreme precipitation formation. They were so complex that a coherent picture on the plausible mechanisms for P95 departures could not be readily discerned.

Third, we sought machine learning based on the SEM framework to build robust regression models of these correlated fields to quantify their relative contributions and interpret the underlying processes affecting P95 simulation. The SEM is an extension of the confirmatory factor analysis and has the ability to test hypotheses on causal relationships in the presence of multicollinearity. Based on our physical understanding and clustering analysis, we constructed four latent variables: energy supply (SWD, SH, NSE, OLR), water supply (ET, MC, TPW, Q2m, W700, V850), surface forcing (T2m, CAPE, CIN, PBLH, LFC, LCL), and cloud forcing

(RET, CRE, CWP, FCL, FCH, RCT). We then defined the objective selection rules using four performance scores (R2, CFI, AIC, Δ AIC) and searched through ~ 231 millions of alternative SEMs in each region (CM or GS) and each season. We finally discovered a unique finalist SEM for each region and season that is physically reasonable, structurally parsimonious, and optimally fitting the data of the simulated P95 departure correlations with the responsive fields. They could be grouped to represent five distinct physical mechanisms as discussed below.

The finalist SEMs for CM summer as well as CM and GS autumn resembled closely, suggesting an essentially identical mechanism in which water supply [RI = 56~60%], energy supply [RI = 20~21%], and cloud forcing [RI = -20~-23%] jointly determined P95. Here stronger W700 and higher Q2m in CM summer and GS autumn or greater TPW in CM autumn caused larger P95, which in turn depleted more FCL and reflected less RET, and consequently increased both SWD and OLR. The SEMs for GS spring and winter were basically the same, suggesting a similar mechanism in which surface forcing was the predominant factor [RI = 84~87%] while energy and water supplies shared evenly the remaining portion [RI = 5~8%]. Here greater CAPE combining with lower [3/4] LCL in spring or higher [2/5] PBLH in winter caused larger P95, which directly consumed most of TPW plus moisture transported by southerly wind (V850) in spring or by ascending motion (W700) in winter, while water supply residual and energy supply from more SWD plus larger SH in spring or OLR in winter both supported surface forcing. The SEM for CM spring also revealed a mechanism in which surface forcing was predominant [RI = 85%], but it was supported by energy supply alone [RI = 13%]. In contrast, the

SEM for GS summer portrayed a mechanism in which cloud forcing was predominant [RI = -84%], because larger precipitation directly depleted more FCL, reflected less RET, and yielded smaller CRE. Meanwhile, water supply [-8%] was opposite to energy supply [6%]. On the other hand, the SEM for CM winter showed a different mechanism, in which water and energy supplies were the two critical counteractive factors [RI = -62%, 31%], while cloud forcing played a secondary but positive role [RI = 7%]. The effects of water supply and cloud forcing were both opposite to those in summer and autumn, since the prevailing precipitation system changed from convective to stratiform processes. Among the 22 fields listed as the possible manifest variables, 15 were selected at least once constituting the latent variables with notable importance [$|RI| > 3\%$] in the finalist SEMs. Of the eight SEMs, the select fields (occurrences) were SWD (7), FCL (5), SH (5), OLR (4), RET (4), CAPE (3), TPW (3), W700 (3), NSE (3), CRE (2), ET (2), PBLH (2), Q2m (2), V850 (1), and LCL (1). This order of sequence indicated the decreasing degree of their relevance to P95, while the relative importance of their actual effect on P95 was determined by the product sum of the strength coefficients along all their directional paths to P95. Both the relevance and importance depended strongly on season and region. Notably, MC, RCT, T2m, CWP, FCH, CIN, and LFC were not on the list. Missing MC is no surprise, since its correlation with P95 was always insignificant. Missing RCT confirms our initial finding that the relative contribution of the parameterized versus resolved rainfall was not important, although it had significant negative correlations with P95 in autumn and winter. However, missing T2m is not expected, especially in CM spring and summer when it had substantially

high correlations with P95 and other fields. The more representative fields such as SWD, FCL and SH must have incorporated the T2m role. Missing the other fields can be similarly understood. Unfortunately, on the final list of the select manifest variables, only SWD, RET, OLR and CRE had long-records of good quality observational data, making it difficult to objectively rank the actual model performance. It is even more difficult to directly compare CWRF against ERI since only the latter had constrained these radiation fields through satellite data assimilation. Nonetheless, our SEM analysis discovered the five distinct physical mechanisms that clearly explained how P95 was simulated differently by ERI and CWRF. In particular, CWRF using type II cumulus parameterization schemes (ECP, NKF) simulated both P95 and the four radiation fields more realistically than using type II schemes (TDK, NSAS, BMJ). The choice of cumulus parameterization affected how water and energy supplies acted through surface and cloud forcings, and thus determined CWRF's ability to simulate U.S. extreme precipitation. In our subsequent papers, we will conduct perturbation experiments, in which the model representation of surface-atmospheric and cloud-radiative interactions is altered, to further test and confirm these mechanisms responsible for extreme precipitation formation.

Chapter 4: Improvement by Markov Chain Monte Carlo based Bayesian model average

4.1 Introduction

Nowadays, many numerical models still perform poorly in extreme precipitation simulation (????). Many studies have investigated how to improve the performance of extreme precipitation by increasing the model resolution (????????), including complex crucial cloud-related processes (?), implementing a more comprehensive cumulus parameterization (?), developing more advanced dynamic models (?), or using machine learning algorithms to identify underlying physics mechanisms (?). However, given the regime dependence of parameters, no single scheme can represent the real climate, let alone extreme precipitation, universally (???). Hence, more advanced ensemble simulation of extreme precipitation is urgently needed (?). Ensemble methods have been proven to be one effective method to improve extreme precipitation simulation (????). Furthermore, they can reduce the uncertainty in projection (????) and provide valuable information on the reliability of future projections of extreme events (????).

There are two main types of ensemble-combination methods. The most basic

method involves taking the arithmetic mean of ensemble members (i.e., the “composite ensemble”). This method has the benefit of computational efficiency, stability, and is theoretically straight forward to explain. However, ? demonstrated that equal weights in ensemble calculation cannot provide optimal ensemble mean outcomes for schemes with substantial differences in performance. In Chapter Two and Chapter Three, both the sensitivity experiments and causal analysis demonstrated that individual schemes perform differently in terms of extreme precipitation simulations. Chapter Four thus applied non-equal weighting methods.

Among non-equal weighting methods, Bayesian Model Averaging (BMA) is the gold-standard for making out-of-sample predictions (?). ? proposed a well-adopted expectation-maximization (EM) algorithm-based BMA method, which has successfully improved the performance of precipitation simulation (?). The EM-based BMA method is both relatively easy to implement and computationally efficient. Furthermore ? provided the open source code, which greatly extended the accessibility of this method. Therefore, I implemented the EM-based BMA method in this extreme precipitation simulation study. The EM-based method also served as a baseline to compare with other methods evaluated in this study. These included the Akaike information criteria-based BMA method (AIC, Akaike 1998), which is computationally more efficient, the bootstrapping version of Akaike weights, which considers the uncertainty associated with training data, and the stacking method. because the first three of these are BMA methods, which are based on the hypothesis that the true data-generating model is among the potential candidate models. However, given the highly nonlinear complexity of extreme precipitation, the true

model, earth system, is out of the ensemble member list (?). Therefore, this study also tested the stacking algorithm proposed by ?, which resolves this theoretical flaw.

? implemented the EM-based BMA method using linear bias correction. Linear bias correction ignores model uncertainty, which can lead to an “over-confident” prediction, in contrast to the probabilistic type of bias correction, which provides more information on prediction uncertainty. That uncertainty information is highly valuable in climate studies and projections. Hence, this study also implemented Markov Chain Monte Carlo (MCMC) based probabilistic bias correction. The MCMC methods also enabled a more flexible model design (?), which allowed me to implement extreme value distributions as the prior distribution.

This chapter first examined the performance of individual ensemble members from the regional Climate-Weather Research and Forecasting model (CWRf, ?), as well as results from the North American CORDEX (NA-CORDEX project(?)) (together we call it CPN in this chapter), using a newly proposed Optimal Rank distance, similarity, consistency and variability (DSCV) framework and the Method for Object-Based Diagnostic Evaluation (MODE) tool. Next, it proposes and tests different ensemble combination methods to improve the performance of extreme precipitation simulations. Section 2 describes models in CPN, focusing on the physics configuration as well as the observational data used for evaluation. Section 3 explains the extreme metrics, performance scores, the OptiRankDSCV framework, and MODE. Next, section 4 analyzes the individual member performance in extreme precipitation simulation using the OptiRankDSCV framework and the MODE tool.

Section 5 describes the three BMA methods. Section 6 demonstrates and compares the extreme precipitation outcomes from BMA methods. Finally, section 7 summarizes the results.

4.2 Model description, observations

	Dynamic	SF	PBL	MP	CU	Ref
CWRF	Non-hydrostatic (No Nudging) 30km	CSSP (???????????)	CAM (?), with-modifications- from ?	TAO (??)	ECP (?????)	?
WRF	Non-hydrostatic (Nudging) 50km 25km	NOAH (???)	MYJ (?)	WSM3 (?)	??	?

RegCM4	Hydrostatic (No Nudging) 50km 25km	BATS (?)	?	SUBEX (?)	?	?
RCA	Semi- Lagrangian (No Nudging) 0.44°	RCALSS (?)	???	Prognostic equation for to- tal cloud water (?)	??	?
HIRHAM5	Semi- Lagrangian (No Nudging) 0.44°	?	ECHAM3 (????)	?	?	?
CRCM5	Semi- Lagrangian (No Nudging) 0.44°,0.22°,0.11°	CLASS3.5 (??)	?	?	??	?

CanRCM4	Semi-Lagrangian (Nudging) 0.44°, 0.22°	CLASS2.7 (??)	?	?	??	?
----------------	---	---------------	---	---	----	---

Table 4.1: CPN experiment model configurations, parameterizations, and their references.

The performance of individual models is crucial to an ensemble simulation study. The well-developed regional Climate-Weather Research and Forecasting model (CWRF, ?) is the perfect platform for this study, because ?? demonstrated the extraordinary skills of CWRF in precipitation related simulations over the continental United States as well as coastal oceans. ? demonstrated that the superior performance from CWRF is universal to another climate domain. Furthermore, ? showed that CWRF simulates extreme precipitation exceedingly well. Given the highly reliable performance from CWRF, this study included the control run simulation results from ?. CWRF’s superior performance is due to its systematically developed multiple physics processes that cover the land surface, planetary boundary layer, cumulus, and microphysics as well as the aerosol-cloud-radiation system (?????). Therefore, the physically advanced CWRF is a perfect match for the following adopted statistic ensemble method. his study focused on 1989-2009, which is the period covered by both CWRF and NA-CORDEX. These 21 years

are the common period between CWRf and the NA-CORDEX experiments. Table 4.1 summarizes the major physics parameterization schemes, model configurations, and references for all CPN members. To achieve consistent results and to minimize the error introduced by interpolation (?), I followed procedures in ? and adopted the conservative algorithm from the Earth System Modeling Framework regridding package to interpolate NA-CORDEX outcomes onto the CWRf grid. This study employed quality controlled observation data from the National Weather Service Cooperative Observer Network (COOP) (??). Given the topographic dependence, I preprocessed precipitation the data following the method described in ?, which used the slope adjustment algorithm: Parameter elevation Regression on Independent Slopes Model (?). Finally, the station data were gridded onto CWRf grid using the Cressman objective analysis method from ?.

4.3 Definitions of extreme precipitation, OptiRankDSCV, MODE tool

There is no single indicator that can comprehensively capture all aspects of extreme precipitation (?). Therefore, this study followed ? and applied six extreme precipitation related indicators. Table 4.2 summarizes the definitions, abbreviations, and units of those indicators. Similarly, there is no universally perfect skill score that can represent all differences in performance (?). Hence, this study proposed a systematic performance analysis framework: Optimal Rank distance (rmse), similarity (correlation), consistency ("linear error in probability space") and variability. The

Abbreviation	Definitions	units
RD	Number of rainy days, with precipitation greater than 1 mm day ⁻¹	days
DRI	Daily precipitation intensity. Total amount of precipitation divided by RD in the period.	mm day ⁻¹
R5D	Sum of five maximum daily precipitations.	mm day ⁻¹
R10	Number of rainy days with precipitation greater than 10 mm day ⁻¹	days
P95	95 percentile precipitation for precipitation greater than 1 mm day ⁻¹	mm day ⁻¹
CDD	Maximum number of consecutive dry day (precipitation smaller than 1 mm day ⁻¹)	days

Table 4.2: Extreme indicators, their definations and units.

rmse and correlation are common indicators used in performance measurements, and provide information enabling the validation by comparison with previous studies with previous studies. However, both rmse and correlation suffer from a damping effect in which simulations with less variability yield better scores. The linear error in probability space (LEPS) indicator resolves this problem, so this study adopted the LEPS indicator following the definition of ?

$$\begin{aligned}
LEPS = & 3(1 - |CDF_o(F_i) - CDF_o(O_i)| + \\
& CDF_o^2(F_i) - CDF_o(F_i) + \\
& CDF_o^2(O_i) - CDF_o(O_i)) - 1
\end{aligned} \tag{4.1}$$

where CDFo stands for observed cumulative distribution function; Oi stand for

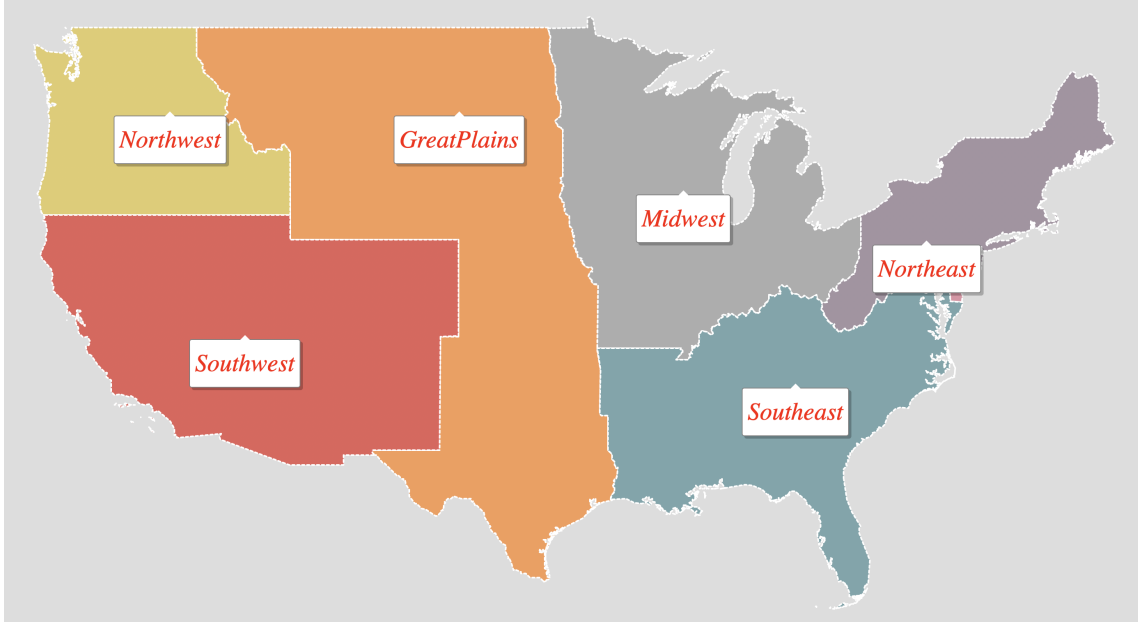


Figure 4.1: Six subregions in the continental United States

the observed value; and F_i stands for the forecast value. The LEPS value ranges from 0 to 1, with 0 representing a perfect score. This score measures the distance between simulated and observed values in cumulative probability space. The updated LEPS algorithm prevents the score from “bending back”, overestimating scores near the extremes (?), which is preferable in this extreme study.

The MVI is defined as in ?, and measures how the model simulates the variability of the observational field. This score provides valuable information in terms of how dispersed or spread out extreme precipitation cases are. Instead of summarizing all variables, I calculated MVI for each extreme precipitation indicator:

$$MVI = (\beta - \frac{1}{\beta})^2 \quad (4.2)$$

where β is the ratio of simulated to observed variance for a specific index in each region. The MVI ranges from 0 to 1 ,with 0 indicating a perfect score.

It is impossible to provide a single measurement containing all performance information. Therefore, this study adopted the methodology from multiple criteria decision making (MCDM). Each model is a potential candidate, and each score is one vote. In the same measurement, voters (measurement) prefer to the model (candidate) have higher score. Then, since the same model might perform differently in different climate regimes (e.g., it might perform well near coastal areas but poorly in mountain areas), I further divided the continental United States into six subregions according to the Third National Climate Assessment (NCA; ?), as shown in Figure 4.1. By using these subregions, the results can be fairer (since it prevents the gerrymandered districts effect seen in social science) and quickly compared to other studies in the future. Each model uses 144 indicators ("votes," four measurements multiplied by six regions multiplied by six extreme precipitation indicators). Using more indicators produced better coverage of overall ensemble performance.

Furthermore, the use of multiple scores rather than a single indicator can help to prevent overfitting in model development. The cross-entropy optimal rank aggregation algorithm was used to calculate the "optimal rank" results from these indicators (?). The weight of each subregion was their ratio of area relative to the contiguous United States. This optimal rank took into account all significant features of extreme precipitation, the four principal skill measurement scores, and the performance difference in many climate regions. Hence OptiRankDSCV represented the most comprehensive performance of each member.

OptiRankDSCV provided highly condensed information on performance from each model. However, this is not enough to understand the simulation spatial pat-

tern distribution. Hence, in addition to the optimal rank analysis, this study further investigated each model’s spatial pattern performance of P95 by using the Method for Object-Based Diagnostic Evaluation (MODE) tool. The Research Applications Laboratory in National Center for Atmospheric Research (NCAR), USA developed the MODE tool (?). The MODE provides a meaningful spatial score that considers several features factors (e.g., orientation, location, shape, and intensity percentile). This score mimics humans understanding of “regions of interest” in a objective, nonjudgmental way. The MODE first resolved the objects in the raw data using a convolution process, which replaces a field’s value with its surrounding grid points within a preset distance. After the convolution, the MODE tool applied a filter to remove precipitation values below 1 mmhour^{-1} (the AMS glossary definition of light rain). This filter helps to focus on the only impactful grid point in extreme precipitation simulation analysis. Then MODE tool then calculated related attributions for each object. When those attributes were ready, the MODE conducted the ”matching and merging” step. In the merging process, MODE tries to mimic a human meteorologist by clustering related grid points into a single physically meaningful object. In the matching process, MODE relates the most reasonable matching objects in forecast fields to similar observations by a sophisticated fuzzy algorithm (?). Finally, ”total interest” measured relative performance from models (?). The ”total interest” considers all-important pattern distribution features in space (e.g., centroid distance, orientation, and area).

4.4 Performance analysis of individual ensemble members

Figure 4.2 uses the OptiRankDSCV framework to compare the spatial pattern skill of winter extreme precipitation simulation of CPN in OptiRankDSCV framework. All statistics are based the 21-year period from 1989-2009. The scores were scaled to the 0-1 range by dividing the value range of each score by each indicator in every region. In winter the WRF-22 ranked and CWRF ranked second. In other models, higher resolutions simulations always almost performed better than their lower counterparts, except for CRCM5-UQAM. No single model outperformed other models in terms of all indicators in every region, but the optimal rank revealed a reasonable order of the skills of the different members. The best-performing member, WRF-22, had the best rmse for all regions except the Midwest. However, its improvement in rmse was accompanied by degradation in MVI, where it had the weakest performance in rainydays (RD), consecutive dry days (CDD) over Southeast and maximum five rainy days (R5D) over the Southwest. A better rmse accompanied by a weaker MVI indicates that the improvement from WRF-22 rmse may have been inflated by the damping effect, which causes underestimation of the variability of corresponding fields. The WRF-22 simulation of P95 over Southeast scored the lowest in almost all measurements. WRF-22's poor performance in MVI is accompanied by its low score in RD. As observed by ?, RD strongly correlates with P95 performance in winter. Hence, the WRF-22's weak performance in winter over the Southeast was partly due to its deficiency in RD simulation. CWRF had a stable performance and relatively high scores in most regions, whereas it performed rela-

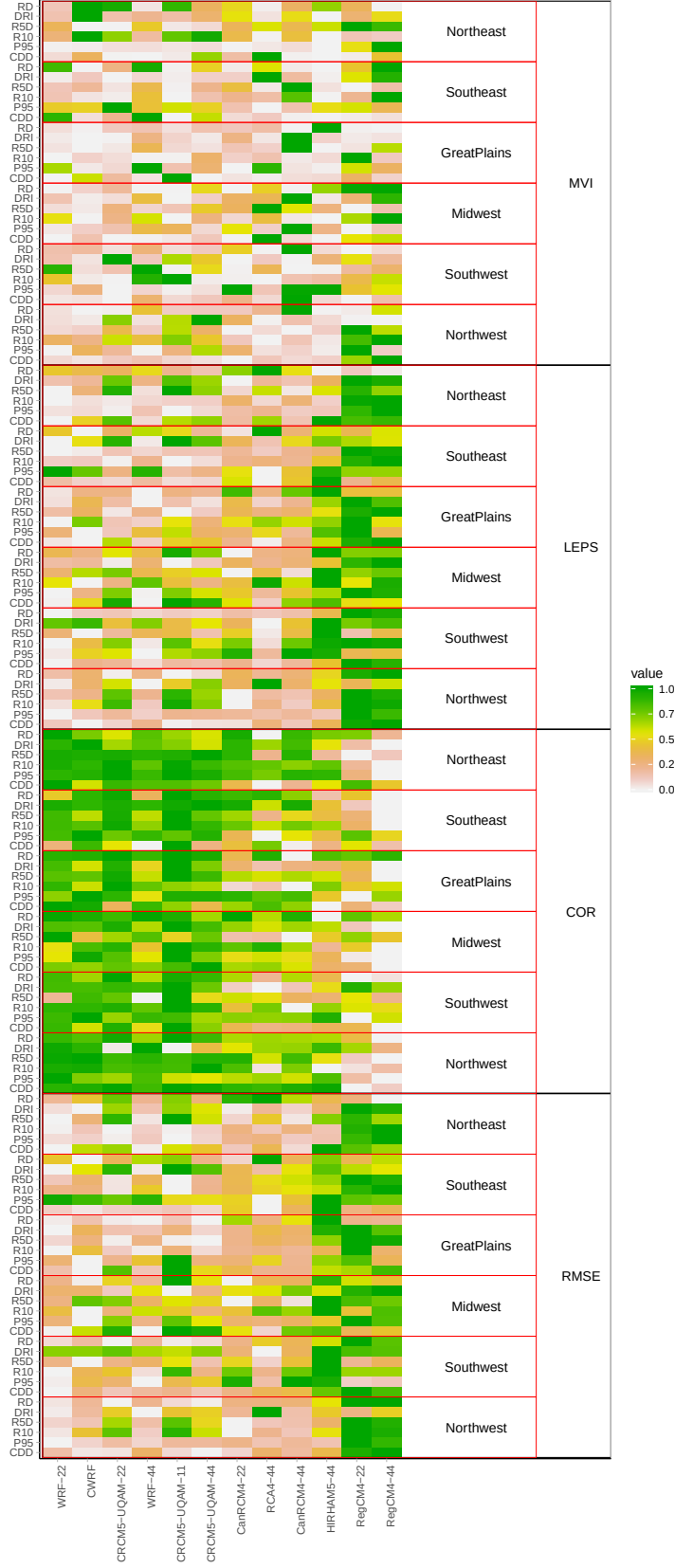


Figure 4.2: Winter overall performance of 1989-2009 mean extreme indicators measured by multi-score in all six subregions. Color represents scores scaled by their range in each row. Horizontally, the locations of models represent the optimal rank aggregated from all 144 indicators, with the most *left* is the *best* performed.

tively weak in the Northeast in terms of RD, DRI, and R10 variability. RegCM-44 ranked lowest in terms of overall performance, but its RD simulations over Midwest and Southwest were reasonably good.

Figure 4.3 compares the spatial pattern skill of extreme precipitation spring simulation using the OptiRankDSCV framework (1980-2009). Generally, the ranking of models in spring resembled that of in winter. CWRF and WRF ranked highest, followed by models from CRCM5. RCA4 and HIRLAM ranked in the middle, and RegCM4 had the weakest performance. Meanwhile, almost all models' high-resolution simulations performed better than their low-resolution simulations, except CRCM5, whose 44km simulation was the best. CWRF performed best in terms of overall extreme simulations. In terms of MVI, CWRF performed the best over almost every subregion, which highlights its ability to simulate pattern variability in extreme precipitation. Its performance in LEPS was mixed over Great Plains, where it earned the best score in Great Plains R10, but had a relatively lower score in RD and R5D. Similarly, in the Southwest, CWRF performed well in RD, R5D, and R10, but relatively weaker in DRI and P95. Interestingly, the second best member, which was CRCM5-UQAM-44, showed the same performance pattern in Southwest LEPS; it also had performed better in RD, R5D, and R10, but relatively weaker in DRI and P95. CWRF generally performed well in terms of pattern correlation, with the only outliers in the Southwest, where CWRF had some difficulty in RD simulation. Interestingly, CWRF scored poorly on this same indicator for both COR and RMSE, but received excellent MVI and acceptable LEPS scores. Hence, CWRF captured the RD in the Southwest by its distribution in probability

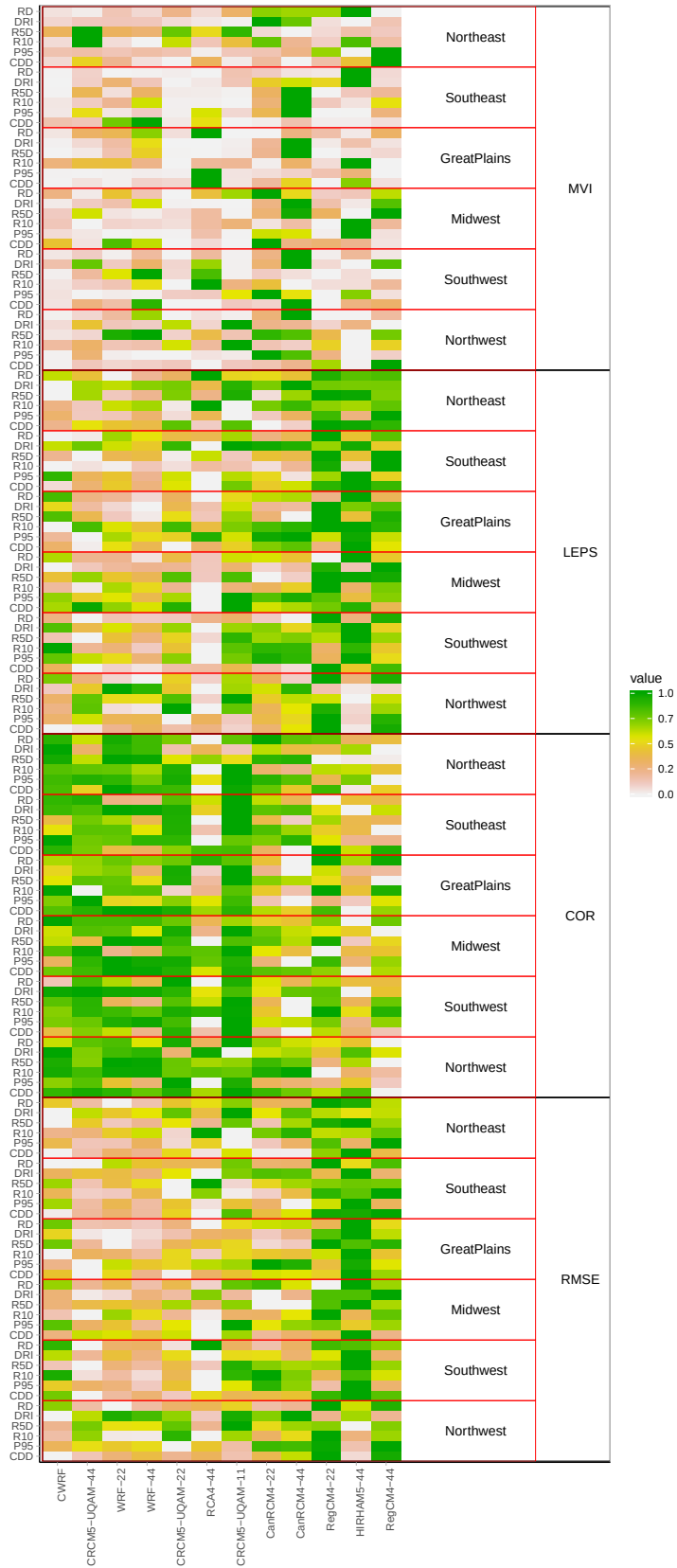


Figure 4.3: Same as Fig4.2 except for spring.

space as well as its variance, but did not correctly represented the shape of spatial distribution.

Figure 4.4 compares the spatial pattern skill of extreme precipitation summer simulation using the OptiRankDSCV framework (1980-2009). CWRF again was the best model this season. Overall, CWRF showed the greatest skill in terms of all measurements and overall subregions. Its outstanding performance was consistent with a previous study ?, which explained the physical mechanisms behind its success. Its only relatively weak scores were RD and CDD over Midwest, and DRI over the Southeast in terms of its MVI. Interestingly, the RegCM-44 had a relatively better score for these three indicators, which was rare since it performed poorly overall. Simulations from CRCM5 showed consistent skill, as in other seasons. CRCM5-44 ranked second, outperforming its high-resolution counterpart, as it did in autumn as well. This indicates that for CRCM5 a higher resolution simulation was not only computationally costly but also detrimental to its simulation skill in its current model framework. One potential explanation is that the assumptions in CRCM5 cumulus and stratiform precipitations were not compatible with a higher resolution simulation. WRF summer performance became relatively weaker compared to its performance in other seasons. Interestingly, like CRCM5, WRF's low-resolution simulation outperformed its higher resolution simulation. This only occurred in summer for WRF, and may be due to the fact that the most dominant summer precipitation type is convective, and higher resolution simulation does not result in a better understanding of convective activity. Without a better physical understanding and representation of convective activity, use of a higher resolution

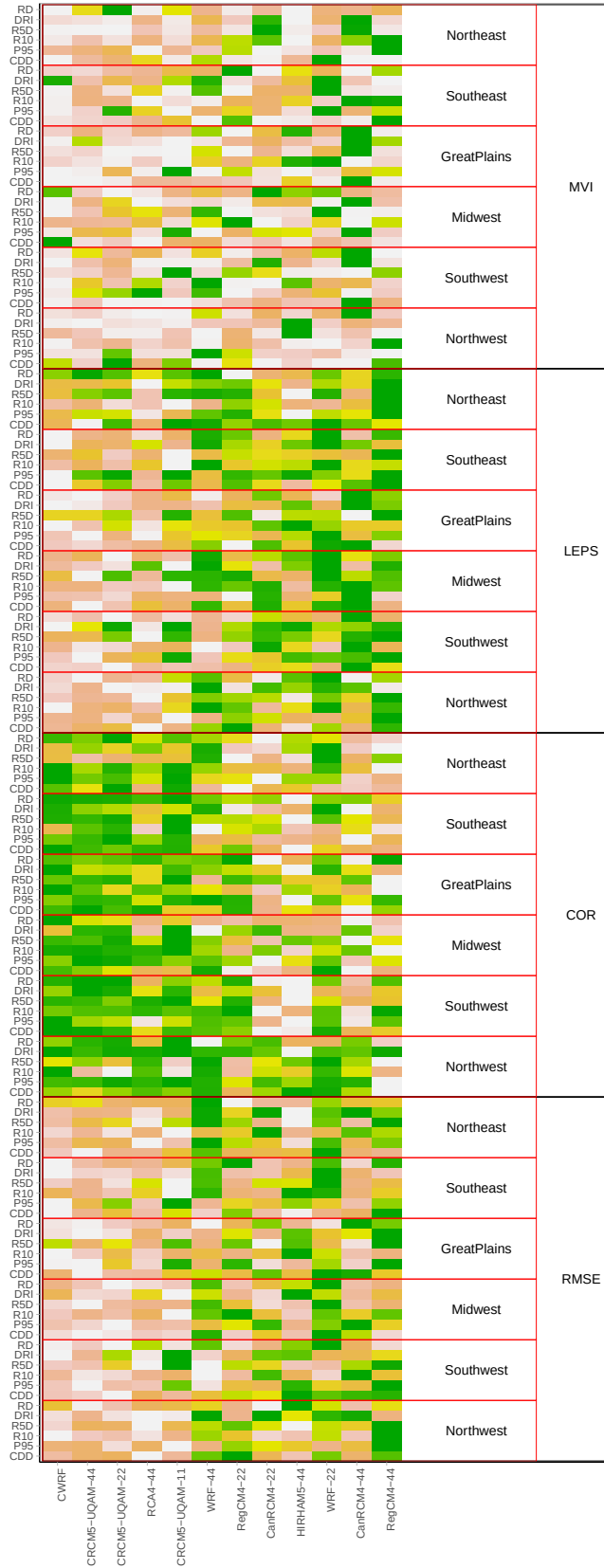


Figure 4.4: Same as Fig4.2 except for summer.

might be not only computationally more costly but also inaccurate. Summer RCA-44 had its highest rank in terms of its performance in all four seasons. Given the importance of convective related precipitation in summer, the improvement in RCA-44 summer performance might be due to its use of a modified Kain-Fritsch scheme compared other models, as ? pointed out that the principle difference in cumulus scheme used in RCA-44 was that the shallow convection was not precipitable.

Figure 4.5 compares the spatial pattern skill of extreme precipitation autumn simulation using the OptiRankDSCV framework (1980-2009). CWRF ranked highest of all members for the overall model performance, . and it's MVI scores were generally the best. However, there were two outliers: R5D in Northeast and CDD in the Midwest. CWRF's relatively weak MVI of CDD in the Midwest in both autumn and summer indicate that the causes of the underlying bias are connected, as ? showed that the physics causes in both summer and autumn over Midwest are the same. The WRF-22 simulation ranked second behind CWRF. WRF-22 generally had a better outcome in terms of RMSE, whereas its MVI was not as good as CWRF. WRF-22's relatively weaker performance in both LEPS and COR indicated this improvement in rmse from WRF-22 might have occurred due to the "damping effect" (underestimation in variability). The CRCM5-44 simulation skill ranked behind WRF-22. As mentioned before, CRCM5-44 outperformed its high-resolution counterpart in both summer and autumn. CRCM5-44, WRF-22, and WRF-44 all had low scores in pattern correlation of R5D over Great Plains. Over the Great Plains in summer, R5D may have a more significant social-economic impact than any other indicator in the same region. Thus, CWRF's stable performance from

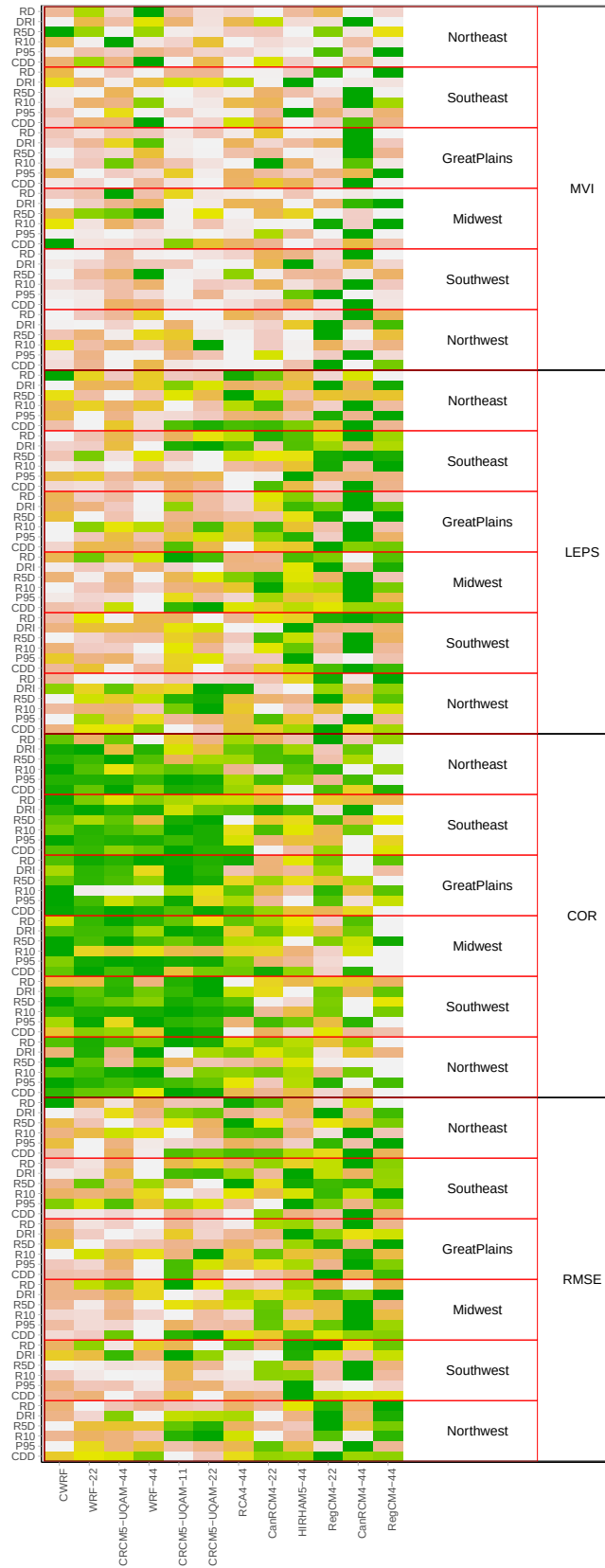


Figure 4.5: Same as Fig4.2 except for autumn.

CWRF exhibits its potential importance for future projection. For all CPN members in autumn, higher resolutions were linked to improved performance, except for CRCM5, whose low-resolution simulations always performed better.

The above comparisons used the newly proposed framework to aggregate information into a concise ranking order, which can then be used to guide the model development. The ranking can help to answer questions such as whether the newly implemented schemes can boost performance, or whether higher or lower resolution runs should be used. However, the ranking system cannot reveal the distribution of interesting fields, which is centrally useful to gaining a better physics understanding. Thus, the study employed the MODE tool to investigate the spatial pattern distributions. It would be impractical to analyze the spatial patterns of all fields (which is the reason for the development of a concise ranking system in the first place), so therefore I used P95 as the example field in the following analysis.

Figure 4.6 compares 21-year winter P95 distributions, including the total interest score (values range from 0 to 1, with 1 indicating the best performance.) ERI has a small area of the identified object (for objects differing greatly in size, the contribution of the weight of the centroid separation to the denominator is small). Given that larger area coverage indicates a more substantial economic impact, this comparison used a convolution radius of 5. Objects in which the simulations and observations matched are colored gray, while objects that did not match are colored dark blue. MODE identified the three largest objects, located in the north-west, Southwest, and Southeast coastal areas (largest). The interest values for the largest object were quite large (0.88-0.979), indicating that all models were capable of cap-

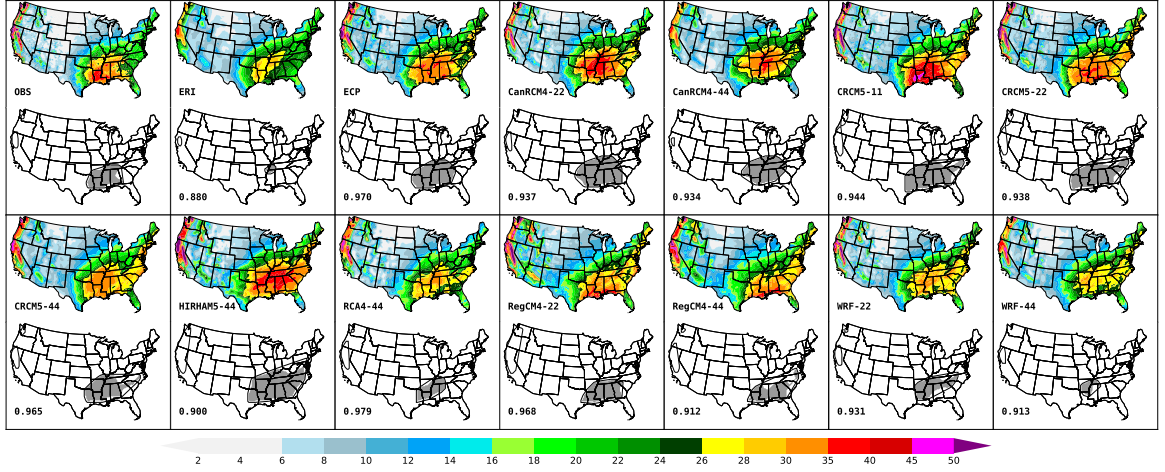


Figure 4.6: Geographic distributions of 1980-2009 mean winter P95 amount [mm day^{-1}] (color) and results of MODE (grey) for observed (OBS), assimilated (ERI), and simulated by all CPN members. A grey area represents the 'interest' area identified by the MODE tool, with total interest score showing in the lower-left corner. The black lines represent the convex hull identified by MODE, which was used to calculate shape features.

turing the essential extreme precipitation features in winter. However, the models captured somewhat different details. ERI had the lowest total interest values compared to other regional models, due to its inability to produce sufficiently intense precipitation. Meanwhile, its center of its maximum shifted inland. ERI also was unable to produce sufficiently intense precipitation near coastal lines. CWRF ranked second (0.97), with improved intensity but relatively oversized coverage. CanRCM4-22 scored slightly higher (0.973) than its lower resolution counterpart CanRCM4-44, because it could capture a more realistic P95 distribution near the coastal line, whereas CanRCM4-44 underestimated P95 near north Florida as well as in Louisiana and Texas. CRCM5 was the only model whose lower resolution simulation had a higher total interest than its high-resolution counterpart. CRCM5-11 significantly overestimated P95 over Louisiana, although the location of its maxi-

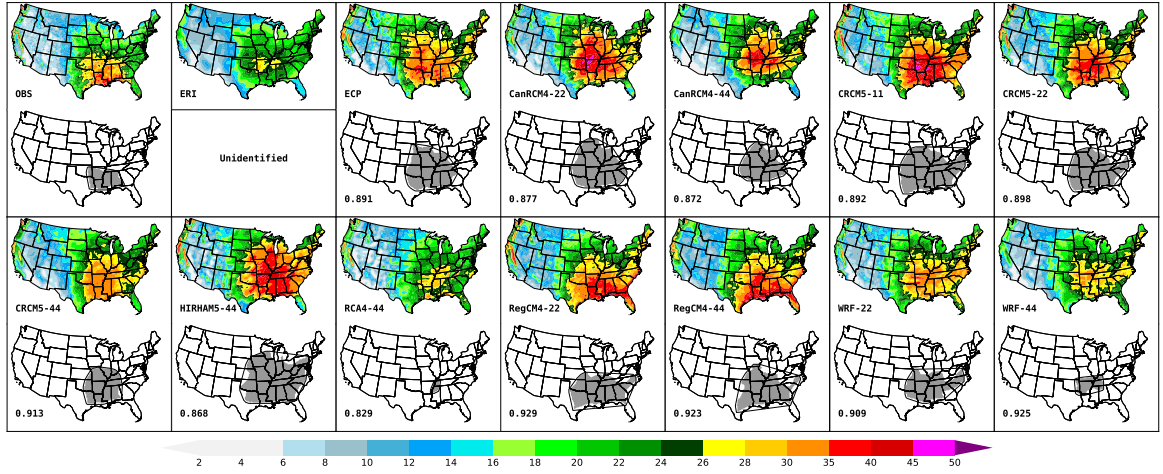


Figure 4.7: Same as Fig4.6 except for spring.

num center was reasonable compared to observations. HIRAM5-44 had the smallest total interest values of all the regional models; it not only overestimated the intensity but also misidentified the shape of P95. RCA4-44 ranked first by total interest score (0.979), but understated precipitation intensity. RegCM4-22 performed better than RegCM4-44, but both overestimated P95 near the Eastern coastal line. WRF-22 and WRF-44 had similar problems as RegCM simulations.

Figure 4.7 compares 21-year spring P95 distributions. The interest values in this season range from 0.868 to 0.929. ERI has a small area of the identified object. Due to ERI's significant underestimation of P95 in this season, MODE was unable to identify the region of interest. CWRF had a reasonable score of 0.981. The score was not highest was due to an oversized heavy precipitation region. The majority of models in this season shared the common problem of overestimating the extent of heavy precipitation to the far north of the U.S. CanRCM-22 performed slightly better than CanRCM-44, but both shifted the maximum center northward

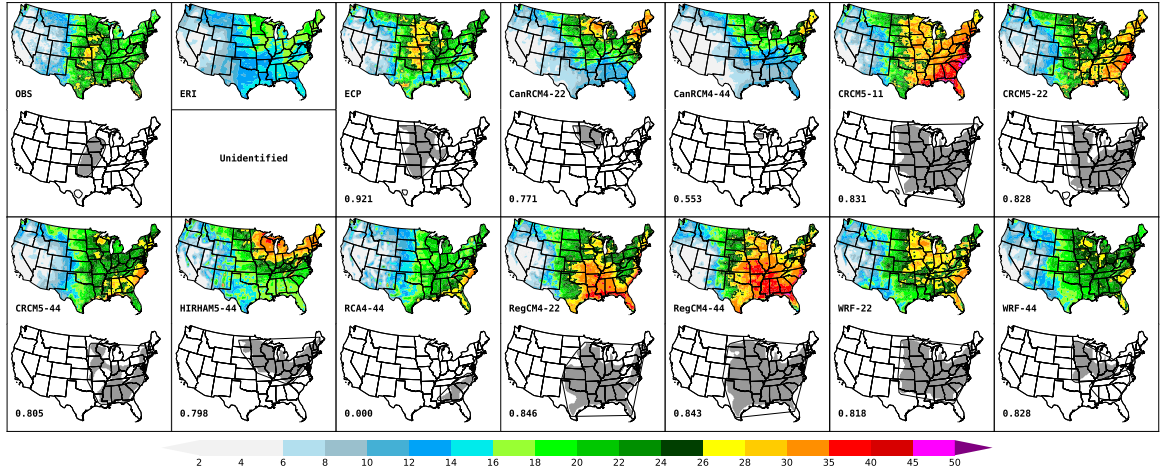


Figure 4.8: Same as Fig4.6 except for summer.

and overestimated its intensity [35mm day^{-1}]. Similarly, in winter, CRCM’s low-resolution simulation performed better than its high-resolution simulation, which overestimated the intensity as well as coverage. HIRAM5-44 still performed relatively poorly compared to other models with a total interest of 0.868, and it again overestimated both the intensity (mostly more than 35mm day^{-1}) and coverage. RCA4-44 had the smallest total interest in spring, with a systematic underestimation in P95. RegCM-22 scored a higher total interest (0.929) than its low-resolution counterpart RegCM-44. However, the difference was not significant. WRF-44 performed better than WRF22 in terms of total interest. Its biggest improvement was primarily due to the weaker intensity from the coarse resolution simulation.

Figure 4.8 compares 21-year summer P95 distributions. The interest values in this season range from 0.553 to 0.921. The relatively lower total interest shows the difficulty of simulating summer extreme precipitation. Due to ERI’s significant underestimation of P95, MODE could not identify an “interest region” from ERI

in summer. CWRF had the highest total interest score [0.921], which shows its outstanding performance in summer extreme precipitation simulation. It potentially overestimated P95 near big lakes, which might occur due to its lake models. CanRCM4-22 scored much higher than CanRCM-44, which barely produced any P95 greater than 26 [mm day⁻¹] Both had their maximum center shifted Northeast compared to observations. In summer, CRCM5 high-resolution simulation had higher total interest than its low-resolution counterpart. However, all CRCM5 members produced too much rain near coastal lines in summer, when observations showed no intense coastal precipitation. As in other seasons, HIRHAM5-44 produced too much rain northward, and the coverage of rain greater than 26 [mm day⁻¹] from HIRHAM5-44 was much higher than the one from CWRF. RCA-44 again underestimated P95, whose P95 were less than 26 [mm day⁻¹] over the majority of the land. As a result, MODE could not find a matching object in RCA4-44 to the “interest-region” over the Central Plains in observations. RegCM again overestimated the P95 and produced too much precipitation near the east and south coastal lines. WRF-44 had a higher total interest [0.828] than WRF higher resolution simulations [0.818], but the improvement was mostly due to shrinking overestimation bias. Neither WRF simulation could realistically capture the shape and center of P95, and both overestimated P95 near coastal regions.

Figure 4.9 compares the 21-year autumn P95 distributions. The total interest scores range from 0.846 to 0.980. Due to underestimation, ERI again could not identify an interest region. Again, CWRF earned the highest score (0.980). Compared to other models, CWRF was able to simulate both size and location in a reason-

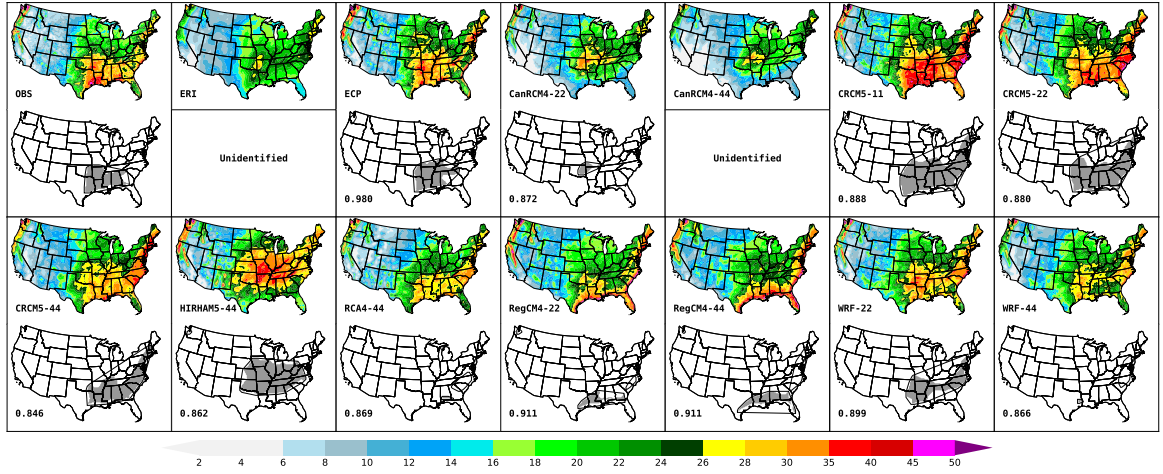


Figure 4.9: Same as Fig4.6 except for autumn.

able match to observations. The low-resolution CanRCM4 model could not identify an interest objects, while its high- resolution counterpart produced insufficient precipitation (0.872), and shifted the location inland. Higher resolution simulations from CRCM5 scored higher (0.888) and produced stronger precipitation, whereas low-resolution simulation produced less rain and scored lower (0.846) The moderate resolution produced a score in between (0.880). HIRHAM5 rained too heavily in the middle of the U.S. over a vast area of coverage, and its distribution was incorrectly shaped, and the center shifted to the north, which led to a relatively low score (0.862). The maximum center from RCA4-44 simulation shifted to the east coastal line, and did not produce intense enough precipitation near the south coast. Both simulations from RegCM4 had the same total interest (0.911): they had the common problem that produced too much rain near the coastal area but not enough rain inland. On the contrary, simulations from WRF underestimated precipitation near the coastal area, with stronger inland precipitation. The high-resolution sim-

ulation from WRF produced more precipitation in a reasonable location compared to the results from WRF-44 (0.866), which was too weak in general.

4.5 Bayesian model average methods

BMA is an ensemble method to weight each member by the marginal posterior probability that the M_k is the data generating model. Following ?, the forecast PDF $p(y)$ can be written as:

$$p(y) = \sum_{k=1}^K p(y|M_k)p(M_k|y^T) \quad (4.3)$$

Here M_k represents the k th ensemble member, y represents the extreme seasonal precipitation indicator in each grid point (e.g. P95), and y^T represents training data. This study used the period from 1989 to 2003 as training data y^T , and from 2004 to 2009 as cross-validation data. In the BMA assumption, $p(M_k|y^T)$ represents the posterior probability that M_k is the data generating model. ? calculated $p(M_k|y^T)$ using an expectation-maximization (EM) algorithm. $p(y|M_k)$ represents the probability forecast outcome from M_k . ? approximated $p(y|M_k)$ by a normal distribution:

$$y|f_k \sim N(a_k + b_k, \sigma^2) \quad (4.4)$$

The probability forecast model for $y|f_k$ is:

$$\begin{aligned}
\alpha &\sim Normal(0, 1) \\
\beta &\sim Normal(1, 1) \\
\mu_{\star} &\sim \alpha + \beta * f_k \\
\mu &= \begin{cases} \mu_{\star}, \mu_{\star} > 0 \\ 0, \mu_{\star} \leq 0 \end{cases} \\
y|f_k &\sim \Gamma(\mu, \sigma_{clim})
\end{aligned} \tag{4.5}$$

The above equations are the Tobit model (censored regression model) to describe the relationship between simulated and observed extreme precipitation. This probability model was calculated using the MCMC method (?) with the No U-turn Sampler (NUTS) algorithm (?). $p(y|M_k)$ is the weighting in the Bayesian model average, while previous studies calculated this variable using maximum likelihood, (??). Here I examined three new variations: Akaike weighting, Akaike weighting with bootstrapping, and stacking. Akaike weighting employs the information from the previous probability model to calculate the Akaike Information Criterion (AIC). The AIC was calculated with the LOO (Pareto-smoothed importance sampling Leave-One-Out cross-validation) algorithm (?). With the calculated AIC, the weight was calculated as (?)

$$w_i = \frac{e^{\frac{1}{2}AIC_i}}{\sum_1^K e^{\frac{1}{2}AIC_i}} \tag{4.6}$$

The above calculation relies on the accuracy of AIC, and since the AIC it-

self can have uncertainty, I also compared the Bootstrapping AIC outcome (?). The above calculations all have the strong assumption the data generating model (DGM) is among the ensemble members. However, this assumption is theoretically unrealistic: given the complexity of the earth system, none of the numerical models will be the true reality but only an approximation. Hence, the third option “stacking” algorithm is more theoretically favorable in extreme precipitation ensemble analysis. The stacking algorithm uses the predicted result $\hat{y}_k$ from the previous Bayesian model and calculates the weight by minimizing the logarithmic score in training periods (?).

$$\max \frac{1}{n} \sum_{i=1}^n \log \sum_{k=1}^K w_k p(y_i | y_{-i}, M_k), \text{ s.t. } 0 \leq w_k \leq 1, \sum_{k=1}^K w_k = 1 \quad (4.7)$$

4.6 Ensemble performance analysis

Figure 4.10 compares observed and simulated mean seasonal precipitation distributions for six cross-validation years (2004- 2009). Given the previously mentioned outstanding performance from CWRF, it represented the best performance a single model can achieve in terms of extreme precipitation simulation. In all seasons, CWRF produced a much realistic outcome than the driving ERI data, whose P95 was mostly weaker than actual observations. The mean of CPN members produced a smooth distribution of P95, which lost the detailed information of pattern features, and was also generally weaker than the outcome from CWRF and observations. On the other hand, results from BMA method further improved the outcome

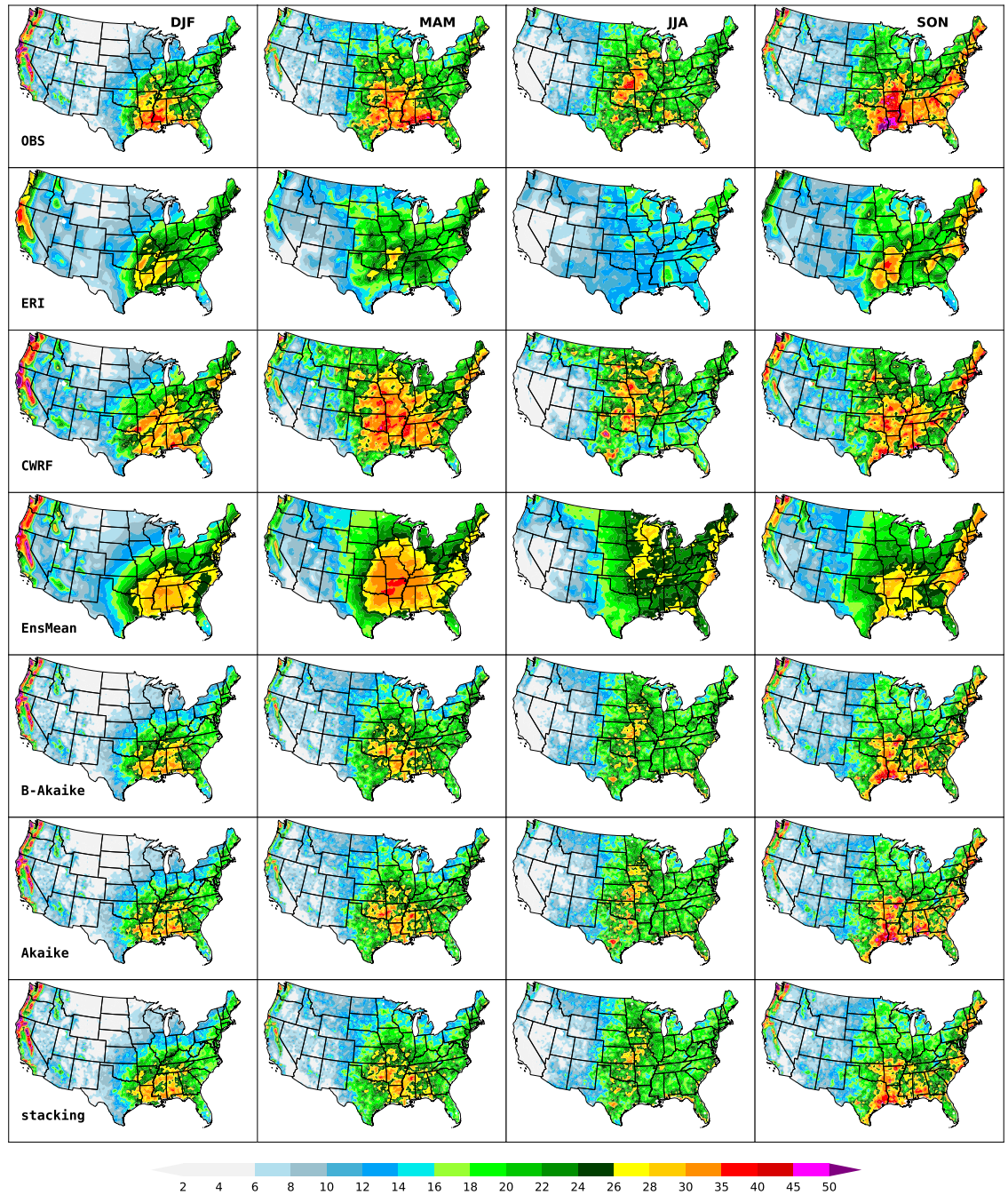


Figure 4.10: Geographic distributions of 2004-2009 (cross-validation) mean seasonal P95 amount [mm day⁻¹] observed (OBS), assimilated (ERI) simulated by CWRf, composite ensemble mean (EnsMean), post-processed by bootstrapping Akaike method (B-Akaike), post-processed by Akaike method (Akaike), and post-processed by stacking method (stacking) for winter (DJF), spring (MAM), summer(JJA),and autumn(SON).

from the CPN members, and the difference among different BMA methods was not significant. In winter, observed P95 was greater than 30 [mm day⁻¹] over the Gulf States. ERI could not produce sufficiently intense precipitation, and its precipitation area was much too small. CWRF, representing the best outcome from CPN members, produced sufficiently strong P95 and better coverage. The ensemble mean generated a broader P95 coverage with an intensity of more than 28 [mm day⁻¹], but was not strong enough to reach the observed value of 30 [mm day⁻¹]. All three BMA methods produced similar P95 distribution patterns, with Akaike and stacking producing relatively stronger P95 (greater than 30 [mm day⁻¹]). B-Akaike was relatively weaker compared to the other two. The BMA method also produced a better distribution pattern than other method; it preserved the details of distribution, and reduced CWRF's overestimation over the Northeast coast without compromising the intensity over the Gulf States. In spring, ERI missed most P95 greater than 30 [mm day⁻¹] over the Gulf States, whereas CWRF produced strong enough P95 but with the maximum center shifted inland. Again, the ensemble mean lost valuable details of pattern distribution, and it did not improve the overestimation problem near the east coast.

The difference between BMA methods was insignificant. All BMA methods reduced the overestimation problem over the Northeast and produced a more precise spatial distribution pattern with details. However, no BMA methods could produce sufficiently strong P95 over Texas (all produced less than 30 [mm day⁻¹]), and they also underestimated the P95 over the coast of Alabama. In summer, P95 from ERI were all less than 20 [mm day⁻¹]. ERI also failed to capture the shape and

location of the maximum P95. CWRf improved substantially, producing sufficiently strong P95 (greater than 35 [mm day⁻¹]) and better capturing the location of the maximum center. The ensemble mean captured the maximum location, but its P95 was not intense enough, and the problematic overestimation near the Northeast coast remained. The three results from BMA were quite similar, with the Akaike method slightly stronger than the other two. The BMA method produced a much more realistic pattern distribution, not only reducing problematic overestimation in the Northeast coast but also reducing CWRf's overestimation over the Northwest central region (North and South Dakotas) from greater than 35 [mm day⁻¹] to the observed 20 [mm day⁻¹]. In Autumn, ERI simulated P95 better than in other seasons. ERI captured the location of maximum P95 over the Gulf States, but still underestimated the intensity and did not produce sufficiently strong P95 near the south coast. CWRf produced stronger P95, but overestimated its area with more P95 greater than 15 [mm day⁻¹]. The CWRf-simulated central maximum did not pass 30 [mm day⁻¹], while observed P95 was higher than 45 [mm day⁻¹] over Louisiana. As in other seasons, the ensemble mean was less intense than the observed results and much detailed information was lost. The outcome from the BMA methods was consistent to that of other seasons. BMA results were more precise with regards to pattern distribution, and also were much stronger near the Gulf States coast (greater than 35 [mm day⁻¹]). The BMA method also improved P95 over Georgia, where both the ensembles mean and CWRf failed to produce sufficiently intense precipitation. In all seasons, the ensemble mean lost detailed distribution information over the Rocky Mountains, while the BMA appeared able

to capture those details.

Figure 4.11 compares the winter observed upper (75%) and lower (25%) confidence interval (CI) with bootstrapping, as well as BMA-estimated CI for P95 during the six cross-validation years (2004- 2009). For observations, the CI was calculated by bootstrapping the seasonal mean P95 1000 times during the six cross-validation years (6000 samples for each grid). The bootstrapping CI was calculated by bootstrapping all 12 members together 1000 times during the six cross-validation years (72000 samples for each grid). For BMA methods, CI was calculated from high probability density (HPD) of all posterior samples during training years (144000 samples for each grid). The bootstrapping CI from all CPN members was relatively small. For most areas, the difference between members was smaller than 1 [mm day⁻¹]. The under dispersive was because bootstrapping did not learn the information on uncertainty from the training data like the BMA did, but just estimated uncertainty from the cross-validation data itself. Hence, it was overconfident and underestimated uncertainty. Meanwhile, CI estimated by bootstrapping shows that CPN members produced too much rain over the Northeast coast and insufficiently strong precipitation near the coast of the Gulf States. The CI from three methods of BMA were very similar (Akaike method again produced slightly stronger precipitation), and the upper and lower CI were more dispersive than the simple composite ensemble results by bootstrapping. The relatively large CI provided more conservative judgments by learning from the training data. Furthermore, observed CI were mostly contained within the CI from BMA methods; specifically, for the Northeast coast, there was no unrealistic overestimation in low-end or high-end CI from BMA

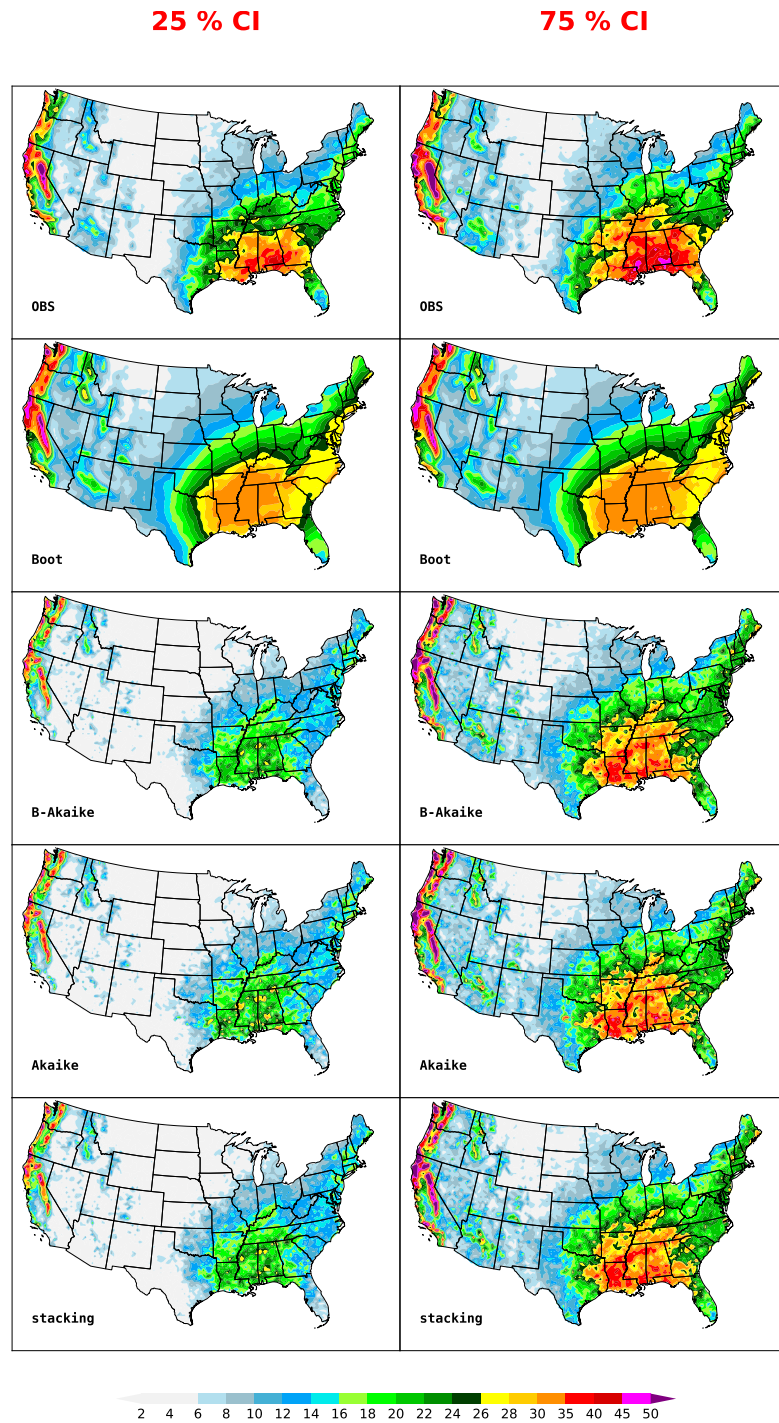


Figure 4.11: Geographic distributions of 2004-2009 (cross-validation) winter P95 confidence interval (CI) amount $[\text{mm day}^{-1}]$. Estimated by bootstrapping observational data (OBS), estimated by bootstrapping ensembles (boot), estimated by applying bootstrapping Akaike method on ensembles (B-Akaike), estimated by applying Akaike method on ensembles (Akaike), and estimated by applying stacking method on ensembles (stacking). *left lower* (25%) CI; *left upper* (75%).

methods.

Figure compares spring observed upper (75%), and lower (25%) CI with boot and BMA estimated CI for P95 during the six cross-validation years (2004-2009). The results from bootstrapping were again too optimistic since they did not account for model uncertainty. Meanwhile, the overestimation over the Northeast coast was very significant from the boot method, which makes sense because bootstrapping method cannot reduce bias in simulations. Furthermore, like the ensemble mean, boot-strapping ensemble members lost the details of pattern distribution. The results over the mountainous area were too smooth. More importantly, the scattered massive precipitation centers over the Gulf States became a continuous precipitation field. The BMA method results largely reduced the overestimation near the Northeast coast area. Meanwhile, they intensified P95 near the Gulf States coast, where bootstrapping underestimated. They also produced more detailed information on pattern distribution (e.g., over the Rocky Mountains and in the Great Plains).

Figure 4.13 compares the summer observed upper (75%) and lower (25%) CI with bootstrapping and BMA estimated CI for P95 during the six cross-validation years (2004- 2009). The CI from the boot method was less than 2 [mm day⁻¹] for most areas, which was even smaller than the observed natural variability alone. The same problems of overestimated P95 near the northeast coast and lost details of distribution patterns persisted in summer. The BMA methods largely improved performance over the Northeast coast. The difference between different BMA methods remained relatively small. The BMA methods estimated a relatively high upper level of P95 compared to observations over Texas. This estimation was quite rea-

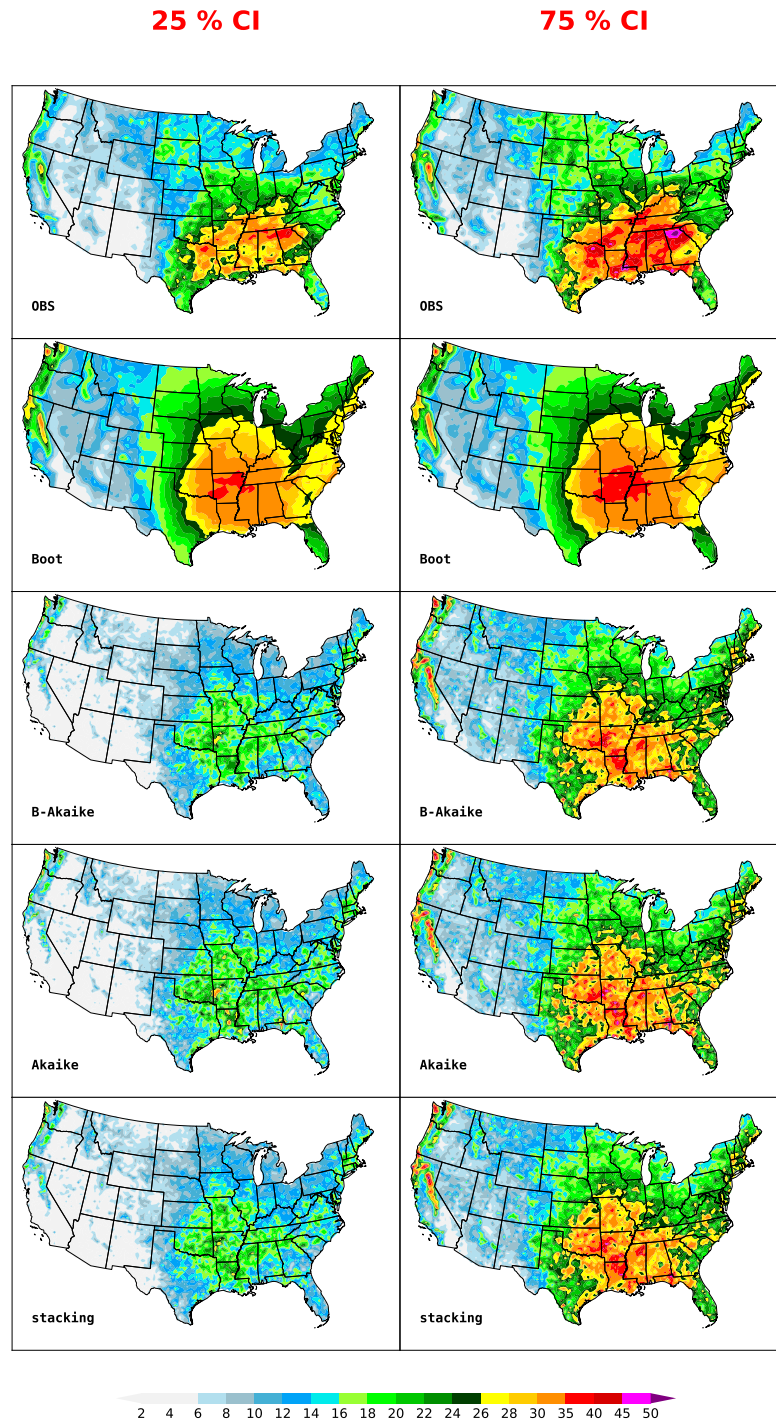


Figure 4.12: Same as Fig4.11 except for spring.

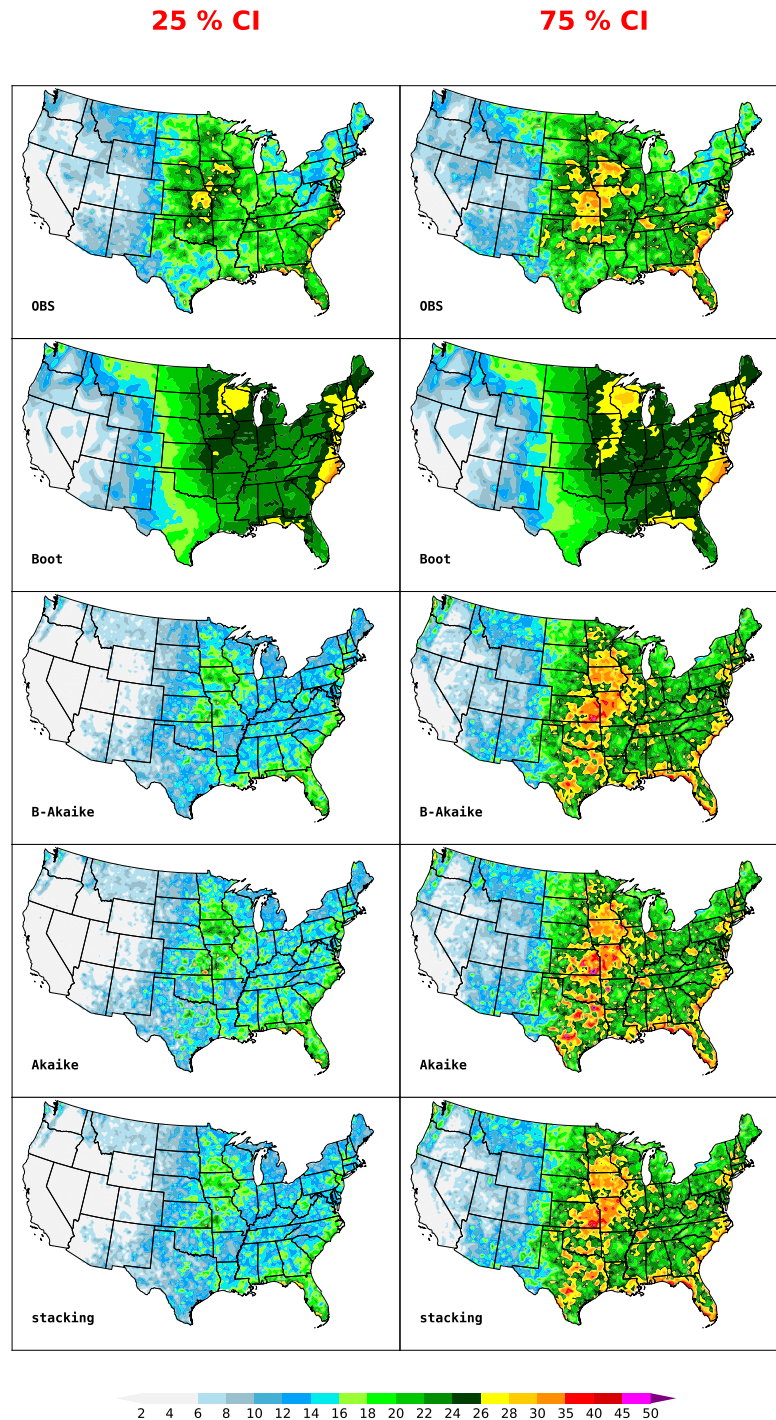


Figure 4.13: Same as Fig4.11 except for summer.

sonable, given that Texas owned the highest number of tornadoes (155) per state during 1981- 2010 (NOAA, 2019). Hence, the BMA methods not only considered the uncertainty from model error but also the uncertainty from natural variability. In this study, the CI of observation from the bootstrapping method underestimated nature variability due to limitations in the available test years.

Figure 4.14 compares autumn observed upper (75%), and lower (25%) CI with boot and BMA estimated CI for P95 during the six cross-validation years (2004-2009). Bootstrapping results exhibited the same problem as other seasons, except over Northeast, where the high observed P95 reduced the issues from overestimation. The BMA methods correctly produced higher CI over the northeast coast as well, which indicated that reducing P95 over the Northeast coast in other seasons was not merely shifting the magnitude. More interestingly, BMA methods produced substantial, strong P95 near eastern Texas, which did not appear in the relatively short period of observation data, which indicates, as in summer, that the BMA methods learned from the training data.

Figure 4.15 compares the change in the unbiased ignorance score (IGN) (Siegert, 2014) between the composite ensemble and the three BMA methods results during the six cross-validation years (2004-2009). IGN has many advantages. In particular, it does not assume the underlying shape of the probability density function (PDF). Hence it is a very desirable indicator, given the non-normality of extreme precipitation. Furthermore, for continuous simulation, ? found that IGN was the only proper local and smooth score. In this study, the IGN was calculated following the unbiased IGN definition from ?, which is suitable since the BMA meth-

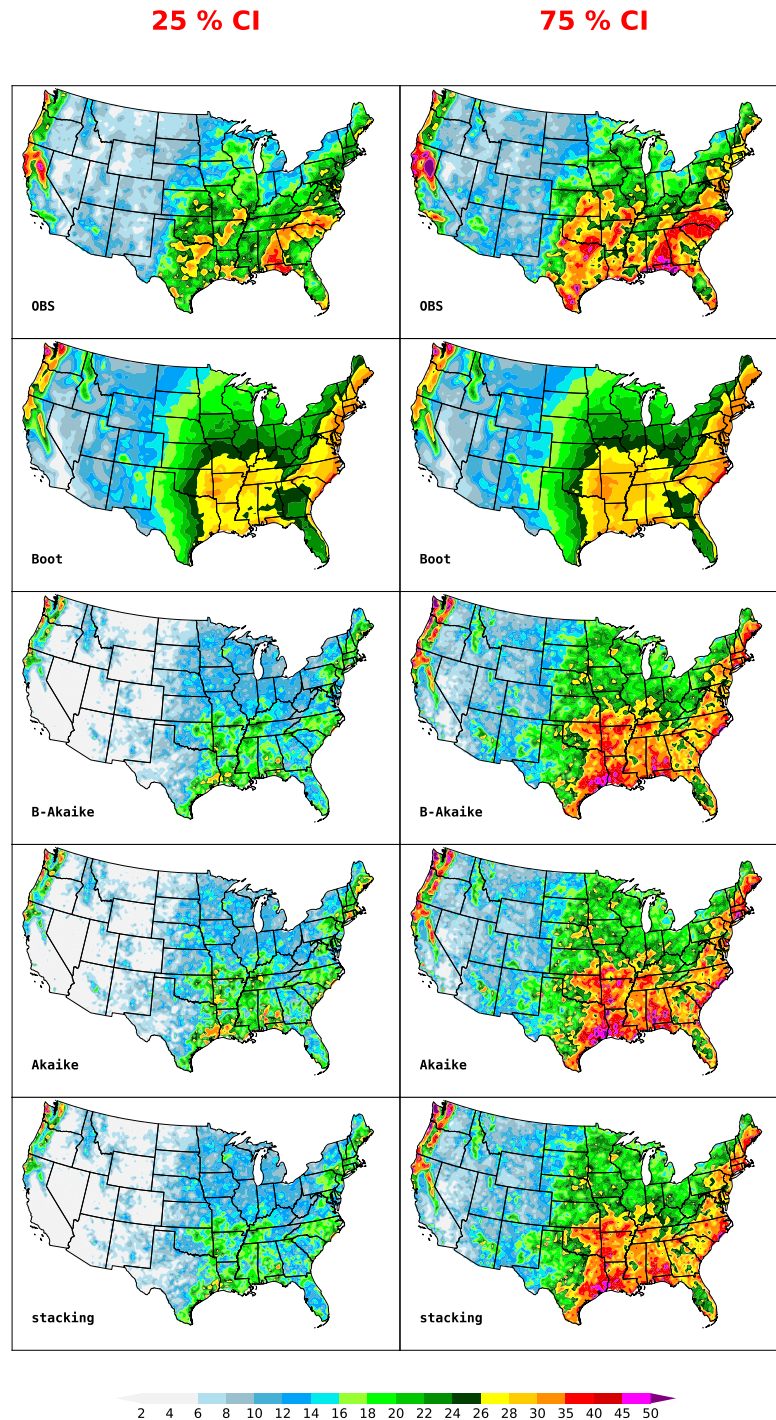


Figure 4.14: Same as Fig4.11 except for autumn.

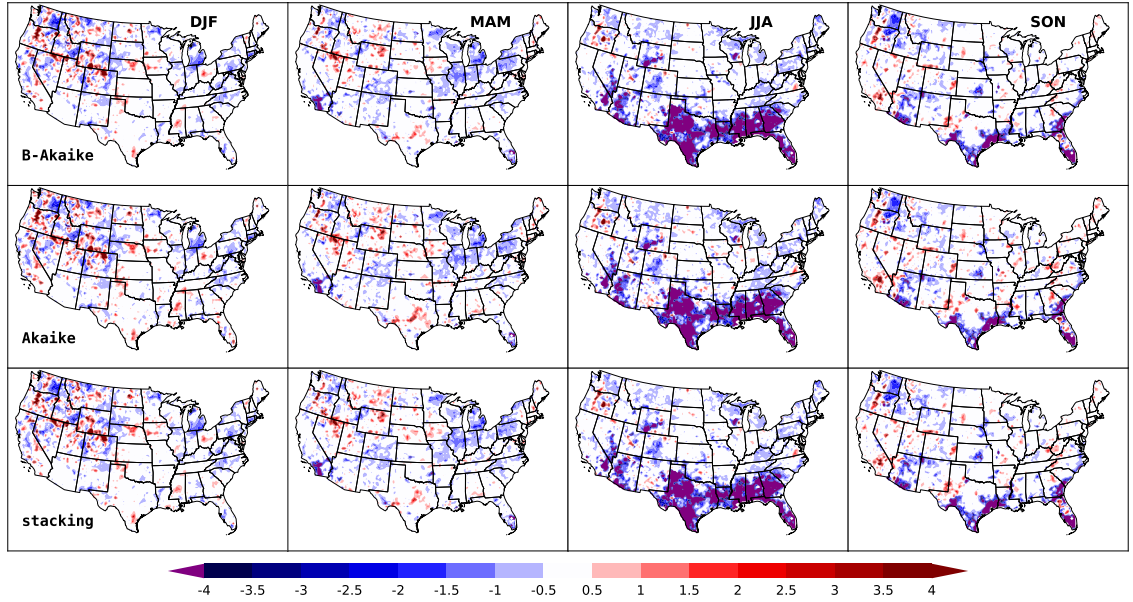


Figure 4.15: Geographic distributions of 2004-2009 (cross-validation) IGN of BMA methods minus IGN of raw composite ensembles: bootstrapping Akaike method (B-Akaike), Akaike method (Akaike), and stacking method (stacking).

ods had larger sample spaces than the raw data. IGN rewards both accuracy and sharpness. A lower IGN means better performance. In all seasons, over the most areas, the members based on BMA methods had smaller unbiased IGN compared to raw ensemble data, which indicates a consistent overall improvement from BMA processing. The magnitude of potential improvement in IGN changed with seasons, with the maximum in summer and the minimum in winter. The biggest improvements occurred in the summer Gulf States, which was the area with the highest risk of extreme precipitation, e.g., Texas. The regions with the highest skill moved away from land in autumn. Meanwhile, over the Rocky Mountains, improvements became significant in autumn. In winter, the changes were more scattered, and performance in some mountain areas even deteriorated. This might be because the P95 was too weak in the winter mountain area, and consequently the signal to noise ratio was too

weak in the training data over this area. In spring, there was a slight improvement over the Midwest region.

4.7 Summary and conclusions

This study was a continuation of a larger effort to improve extreme precipitation simulation over the U.S., as demonstrated in my previous studies (??). One central scientific question I aim to address is whether the ensemble method can effectively improve extreme precipitation simulation, and if so, what actual algorithm should be applied to improve the performance. To answer this question, this study defined three consecutive goals.

The first goal is to identify the best performing model in extreme precipitation. To achieve this goal, this study proposed a new MCDM-based OptiRankDSCV framework, with multiple extreme indicators and performance scores. The OptiRankDSCV framework provided a consistent, fair platform that demonstrates the overall performance from each model and can also act as a testbed for new model development. This study used the OptiRankDSCV framework to analyze the performance of the CPN members in 1989-2009 over the contiguous United States. The results showed that CWRF outperformed all other CPN members by multiple skill scores in terms of multiple extreme precipitation metrics over the whole United States.

Overall, this systematic analysis demonstrated that many models could improve performance in extreme precipitation simulation by using a higher resolution.

However, there are still outliers like CRCM5, whose low-resolution simulation always outperformed its high-resolution simulation. In summer, the reversal of this phenomenon is more prominent. On the other hand, the results showed that resolution was not the most critical factor in determining model performance. Many low-resolution simulations outperformed high-resolution simulations in all four seasons. Both of the above phenomena highlighted the importance of physics understanding to model development. A higher resolution may or may not be useful. In particular, more studies should focus on improving model performance through improved physics understanding.

Then this study used the MODE tool to demonstrate differences in models' ability to capture P95 pattern distribution. The MODE generally produced consistent results as in the OptiRankDSCV framework. Furthermore, the MODE diagnostic tool provided a human mimic analysis, which is more physically meaningful than pure single digits. CWRF showed outstanding performance by MODE measurements. Particularly in summer, it earned significant higher total interest values (0.921) than any other members. This result confirmed previous studies, and further provided a solid, nonjudgmental score to evaluate P95 distribution performance.

The second goal was to provide an ensemble method that is either more theoretically reasonable or more computationally efficient for climate study. After comparing different BMA methods and considering the particular requirements from a climate modeling point of view, I proposed MCMC-based BMA methods to do ensemble analysis of CPN members. The stacking method is the most theoretically sound choice for an open-M situation such as the earth system. The AIC-based

method was more computationally efficient. The B-AIC method included posterior uncertainty. There is no significant difference in their performance, therefore, the computational more efficient algorithm is a better choice.

The third goal was to test the newly proposed methods against both the ensemble mean and the best-performing simulation from CWRP. Results showed that BMA methods significantly improved the simulation's ability to capture the general spatial pattern. Compared with the mean of ensemble members, BMA methods reduced overestimation of P95 greater 25 [mm day⁻¹] in both spring and winter. Also, the BMA methods corrected the dislocation of P95 in summer and autumn. The BMA methods further significantly reduced P95 overestimation in the Northeast coastal area of the U.S. in both winter and spring. BMA method not only reduced overestimation, but BMA methods also enhanced the simulated intensity in the Gulf States coastal area in autumn. One more benefit of the BMA method was that the BMA did not compromise the accuracy of the simulation. The high-resolution details were able to be retained in the BMA methods output.

Compared to the optimal model output from CWRP, BMA methods reduced overestimation in both winter and spring. In summer, even though the intensity from BMA methods was weaker than observations, BMA reduced overestimation over the Midwest area (where CWRP simulated P95 greater than 25 [mm day⁻¹]) and produced a more reasonable shape of the distribution. In autumn, the BMA methods again provided a more precise distribution location, whose P95 maximum center (greater than 35 [mm day⁻¹]) was more similar to the observational data. Overall, the newly proposed BMA methods significantly improved the performance

of extreme precipitation, not only compared to the ensemble mean, but also compared to the best performing single model. Meanwhile, this study demonstrated that there was no significant difference between the three methods. This result has substantial practical value. Since the AIC method is the most computationally efficient but performs similarly to the other two theoretically more sound BMA methods (stacking and B-AIC), it is safe to adopt the AIC without worrying about the potential theoretical deficiency.

Probabilistic BMA methods can not only improve the performance of simulation distribution, but also can provide valuable information on model uncertainty. Hence, this study further compared the CI from BMA methods and bootstrapping outcomes from the raw ensemble model. In all seasons, high CI from BMA methods showed more reasonable coverage than the simple bootstrapping on the raw ensemble models. Compared to bootstrapping ensemble members, there was less area where bootstring observations exceeded the upper CI values estimated by BMA methods. On the other hand, bootstrapping ensemble members severely underestimated the CI range, which means that without BMA the simple bootstrapping method can be under-dispersive and overlook the actual uncertainty in the model simulation. Hence the uncertainty estimation by BMA methods were more conservative and reliable. This study also adopted the IGN score to further measure the performance of probabilistic extreme precipitation forecasts. This study considered the effect of sample size in this measurement; hence, I adopted the unbiased IGN score (?). The results show that overall the newly proposed methods successfully improved probability prediction compared to the composite ensemble results. In particular,

over the summer Gulf States, BMA methods significantly added more value than the composite method. This improvement was particularly useful in Texas, which suffered the most tornados of all U.S. states. As in the case of the ensemble mean output, there was no significant difference in terms of model uncertainty estimation from different BMA methods. This suggests that it is safe to adopt the computationally efficient AIC method accompanied by the MCMC algorithm to perform BMA.

Finally, the above analyses show that the BMA method using the MCMC algorithm not only performed better than the arithmetic ensemble mean, but also provided more reasonable uncertainty estimations. Furthermore, the AIC weights were practically equivalent to a more theoretically sounding stacking method or the bootstrapping version of AIC.

In our subsequent papers, I will implement this system of the BMA method into future projections, and specifically extreme precipitation projections. I hope that this newly proposed method will help reduce errors in extreme precipitation simulation. Furthermore, it may help to provide more reasonable and reliable uncertainty estimation information for extreme precipitation projection.

Chapter 5: Future work: projections of future extreme precipitation changes and impacts

The ultimate purpose of this study is to provide a better estimation of how extreme precipitation will change in the future. I have already conducted a preliminary analysis of the future extreme precipitation changes based on the NA-CORDEX data. This study calculated the change of extreme precipitation by using RCP85 (2006-2100) minus the historical (1950-2005) simulated extreme precipitation. This study adopted the extreme value theory and compared the changes in extreme values for different return levels.

As demonstrated in figure 5.1, generally, all over the U.S. extreme values systematically increased, with summer Texa and west coastal getting much stronger precipitation.

Meanwhile, the preliminary results on changes in rainy days figure 5.2 shows there will be less number of rainydays in the Gulf States and Arizona. Given the actual number of rainydays in Arizona is less than ten, this clearly shows over these areas, the rain will become less frequent. Meanwhile, as figure 5.2 shows, the future projections might also have under dispersive problem by a simple bootstrapping method.

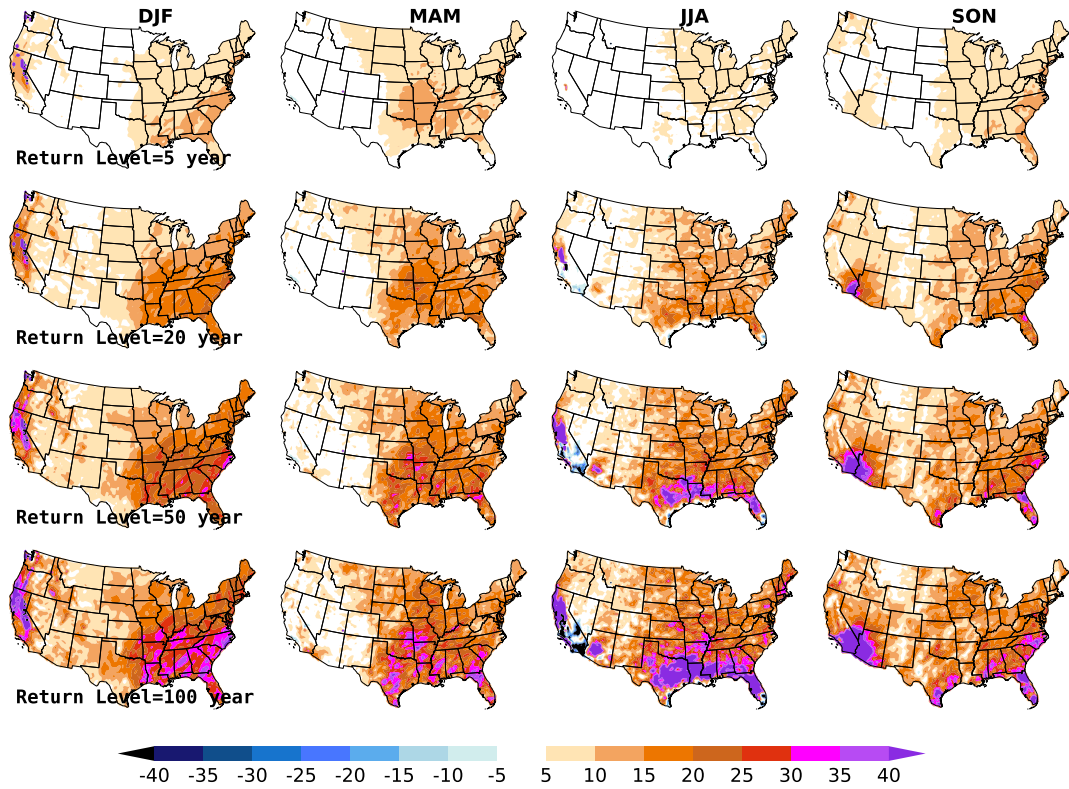


Figure 5.1: Geographic distributions of projected changes in extreme precipitation for different return levels.

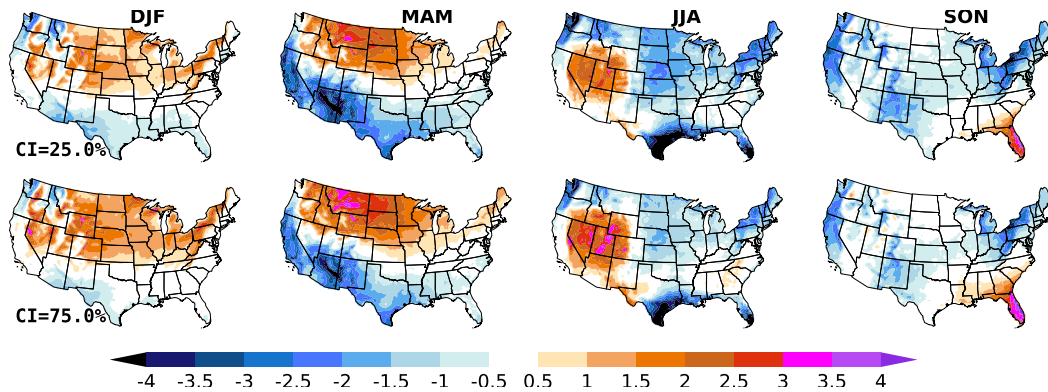


Figure 5.2: Geographic distributions of projected changes in the number of rainy-days, with confidence interval estimated by the bootstrapping method.

We hope to apply the newly proposed BMA methods into the future projections combined with the outcome from CWRP to provide a more reliable outcome. Ultimately make it a better world.

Bibliography

- Akaike, H. (1974), A new look at the statistical model identification, in *Selected Papers of Hirotugu Akaike*, pp. 215–222, Springer.
- Akaike, H. (1998), Information theory and an extension of the maximum likelihood principle, in *Selected papers of hirotugu akaike*, pp. 199–213, Springer.
- Alexander, L., X. Zhang, T. Peterson, J. Caesar, B. Gleason, A. Klein Tank, M. Haylock, D. Collins, B. Trewin, and F. Rahimzadeh (2006), Global observed changes in daily climate extremes of temperature and precipitation, *Journal of Geophysical Research: Atmospheres*, 111(D5).
- Allan, R. P., and B. J. Soden (2008), Atmospheric warming and the amplification of precipitation extremes, *Science*, 321(5895), 1481–1484.
- Allen, M., and P. Stott (2003), Estimating signal amplitudes in optimal fingerprinting, Part I: Theory, *Climate Dynamics*, 21(5-6), 477–491.
- Allen, M. R., and W. J. Ingram (2002), Constraints on future changes in climate and the hydrologic cycle, *Nature*, 419(6903), 224–232.
- Anderson, B. T., D. J. Gianotti, and G. D. Salvucci (2015), Detectability of historical trends in station-based precipitation characteristics over the continental United States, *Journal of Geophysical Research: Atmospheres*, 120(10), 4842–4859.
- Asadieh, B., and N. Krakauer (2015), Global trends in extreme precipitation: climate models versus observations, *Hydrology and Earth System Sciences*, 19(2), 877–891.
- Ashouri, H., S. Sorooshian, K.-L. Hsu, M. G. Bosilovich, J. Lee, M. F. Wehner, and A. Collow (2016), Evaluation of NASA’s MERRA precipitation product in reproducing the observed trend and distribution of extreme precipitation events in the United States, *Journal of Hydrometeorology*, 17(2), 693–711.
- Auld, H., and D. Maclver (2006), Changing weather patterns, uncertainty and infrastructure risks: emerging adaptation requirements, in *2006 IEEE EIC Climate Change Conference*, pp. 1–10, IEEE.
- Barnston, A. G., M. K. Tippett, M. L. L’Heureux, S. Li, and D. G. DeWitt (2012), Skill of real-time seasonal ENSO model predictions during 2002–11: is our capability increasing?, *Bulletin of the American Meteorological Society*, 93(5), 631–651.

- Bechtold, P., J.-P. Chaboureau, A. Beljaars, A. Betts, M. Köhler, M. Miller, and J.-L. Redelsperger (2004), The simulation of the diurnal cycle of convective precipitation over land in a global model, *Quarterly Journal of the Royal Meteorological Society*, 130(604), 3119–3137.
- Bechtold, P., M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. J. Rodwell, F. Vitart, and G. Balsamo (2008), Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales, *Quarterly Journal of the Royal Meteorological Society*, 134(634), 1337–1351.
- Bechtold, P., N. Semane, P. Lopez, J.-P. Chaboureau, A. Beljaars, and N. Bormann (2014), Representing equilibrium and nonequilibrium convection in large-scale models, *Journal of the Atmospheric Sciences*, 71(2), 734–753.
- Belmecheri, S., F. Babst, A. R. Hudson, J. Betancourt, and V. Trouet (2017), Northern Hemisphere jet stream position indices as diagnostic tools for climate and ecosystem dynamics, *Earth Interactions*, 21(8), 1–23.
- Bentler, P. M. (1990), Comparative fit indexes in structural models., *Psychological bulletin*, 107(2), 238.
- Bernard, J. M., and A. F. Smith (1994), *Bayesian theory*, John Wiley & Sons.
- Betts, A., and M. Miller (1986), A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets, *Quarterly Journal of the Royal Meteorological Society*, 112(473), 693–709.
- Bhattacharya, R., S. Bordoni, and J. Teixeira (2017), Tropical precipitation extremes: Response to SST-induced warming in aquaplanet simulations, *Geophysical Research Letters*, 44(7), 3374–3383.
- Blackadar, A. K. (1962), The vertical distribution of wind and turbulent exchange in a neutral atmosphere, *Journal of Geophysical Research*, 67(8), 3095–3102.
- Boyle, J., and S. A. Klein (2010), Impact of horizontal resolution on climate model forecasts of tropical precipitation and diabatic heating for the TWP-ICE period, *Journal of Geophysical Research: Atmospheres*, 115(D23).
- Boyles, R., A. Marshall, and F. Proschan (1985), Inconsistency of the maximum likelihood estimator of a distribution having increasing failure rate average, *The Annals of Statistics*, 13(1), 413–417.
- Bretherton, C. S., and S. Park (2009), A new moist turbulence parameterization in the Community Atmosphere Model, *Journal of Climate*, 22(12), 3422–3448.
- Brinkop, S., and E. Roeckner (1995), Sensitivity of a general circulation model to parameterizations of cloud–turbulence interactions in the atmospheric boundary layer, *Tellus A*, 47(2), 197–220.

- Brown, J. R., C. Jakob, and J. M. Haynes (2010), An evaluation of rainfall frequency and intensity over the Australian region in a global climate model, *Journal of Climate*, 23(24), 6504–6525.
- Burnham, K. P., and D. R. Anderson (2004), Multimodel inference: understanding AIC and BIC in model selection, *Sociological methods & research*, 33(2), 261–304.
- Cairo, A. (2016), *The truthful art: Data, charts, and maps for communication*, New Riders.
- Caldwell, P., H.-N. S. Chin, D. C. Bader, and G. Bala (2009), Evaluation of a WRF dynamical downscaling simulation over California, *Climatic Change*, 95(3-4), 499–521.
- Cattiaux, J., and A. Ribes (2018), Defining single extreme weather events in a climate perspective, *Bulletin of the American Meteorological Society*, (2018).
- Catto, J. L., and S. Pfahl (2013), The importance of fronts for extreme precipitation, *Journal of Geophysical Research: Atmospheres*, 118(19).
- Cha, D.-H., and D.-K. Lee (2009), Reduction of systematic errors in regional climate simulations of the summer monsoon over East Asia and the western North Pacific by applying the spectral nudging technique, *Journal of Geophysical Research: Atmospheres*, 114(D14).
- Chakraborty, A., and T. Krishnamurti (2006), Improved seasonal climate forecasts of the South Asian summer monsoon using a suite of 13 coupled ocean–atmosphere models, *Monthly weather review*, 134(6), 1697–1721.
- Changnon, S. (1997), Trends in hail in the United States, in *Proceedings of the Workshop on the Social and Economic Impacts of Weather*, National Center for Atmospheric Research, Boulder, CO, pp. 19–34.
- Chen, C.-T., and T. Knutson (2008), On the verification and comparison of extreme rainfall indices from climate models, *Journal of Climate*, 21(7), 1605–1621.
- Choi, H. I., and X.-Z. Liang (2010), Improved terrestrial hydrologic representation in mesoscale land surface models, *Journal of Hydrometeorology*, 11(3), 797–809.
- Choi, H. I., P. Kumar, and X.-Z. Liang (2007), Three-dimensional volume-averaged soil moisture transport model with a scalable parameterization of subgrid topographic variability, *Water resources research*, 43(4).
- Choi, H. I., X.-Z. Liang, and P. Kumar (2013), A conjunctive surface–Subsurface flow representation for mesoscale land surface models, *Journal of Hydrometeorology*, 14(5), 1421–1442.

- Choi, I.-J., E. K. Jin, J.-Y. Han, S.-Y. Kim, and Y. Kwon (2015), Sensitivity of diurnal variation in simulated precipitation during East Asian summer monsoon to cumulus parameterization schemes, *Journal of Geophysical Research: Atmospheres*, 120(23), 11–971.
- Chou, C., J. C. Chiang, C.-W. Lan, C.-H. Chung, Y.-C. Liao, and C.-J. Lee (2013), Increase in the range between wet and dry season precipitation, *Nature Geoscience*, 6(4), 263.
- Chou, M.-D., and M. J. Suarez (1999), *A Solar Radiation Parameterization for Atmospheric Studies*. Volume 15.
- Chou, M.-D., M. J. Suarez, X.-Z. Liang, M. M.-H. Yan, and C. Cote (2001), A thermal infrared radiation parameterization for atmospheric studies.
- Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher (2008), On the need for bias correction of regional climate change projections of temperature and precipitation, *Geophysical Research Letters*, 35(20).
- Christensen, O. B., M. Drews, J. H. Christensen, K. Dethloff, K. Ketelsen, I. Hebestadt, and A. Rinke (2007), The HIRHAM regional climate model. Version 5 (beta).
- Coppola, E., F. Giorgi, S. Rauscher, and C. Piani (2010), Model weighting based on mesoscale structures in precipitation and temperature in an ensemble of regional climate models, *Climate Research*, 44(2-3), 121–134.
- Coumou, D., and S. Rahmstorf (2012), A decade of weather extremes, *Nature Climate Change*, 2(7), 491–496.
- Cuijpers, J., and P. Duynkerke (1993), Large eddy simulation of trade wind cumulus clouds, *Journal of the Atmospheric Sciences*, 50(23), 3894–3908.
- Curriero, F. C., J. A. Patz, J. B. Rose, and S. Lele (2001), The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994, *American journal of public health*, 91(8), 1194–1199.
- Cuxart, J., P. Bougeault, and J.-L. Redelsperger (2000), A turbulence scheme allowing for mesoscale and large-eddy simulations, *Quarterly Journal of the Royal Meteorological Society*, 126(562), 1–30.
- Dai, A. (2006), Precipitation characteristics in eighteen coupled climate models, *Journal of Climate*, 19(18), 4605–4630.
- Dai, A., G. A. Meehl, W. M. Washington, T. M. Wigley, and J. M. Arblaster (2001), Ensemble simulation of twenty-first century climate changes: Business-as-usual versus CO₂ stabilization, *Bulletin of the American Meteorological Society*, 82(11), 2377–2388.

- Dai, Y., X. Zeng, R. E. Dickinson, I. Baker, G. B. Bonan, M. G. Bosilovich, A. S. Denning, P. A. Dirmeyer, P. R. Houser, and G. Niu (2003), The common land model, *Bulletin of the American Meteorological Society*, 84(8), 1013–1024.
- Daly, C., G. Taylor, and W. Gibson (1997), The PRISM approach to mapping precipitation and temperature, in *Proc., 10th AMS Conf. on Applied Climatology*, pp. 20–23, Citeseer.
- Daniels, A. E., J. F. Morrison, L. A. Joyce, N. L. Crookston, S.-C. Chen, and S. G. McNulty (2012), *Climate projections FAQ*.
- Davis, C. A., B. G. Brown, R. Bullock, and J. Halley-Gotway (2009), The method for object-based diagnostic evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program, *Weather and Forecasting*, 24(5), 1252–1267.
- Dee, D., S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balsameda, G. Balsamo, and P. Bauer (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the royal meteorological society*, 137(656), 553–597.
- Delage, Y. (1997), Parameterising sub-grid scale vertical transport in atmospheric models under statically stable conditions, *Boundary-Layer Meteorology*, 82(1), 23–48.
- Dickinson, E., A. Henderson-Sellers, and J. Kennedy (1993), Biosphere-atmosphere transfer scheme (BATS) version 1e as coupled to the NCAR community climate model.
- Dittus, A. J., D. J. Karoly, S. C. Lewis, L. V. Alexander, and M. G. Donat (2016), A multiregion model evaluation and attribution study of historical changes in the area affected by temperature and precipitation extremes, *Journal of Climate*, 29(23), 8285–8299.
- Doelling, D. R., R. Bhatt, B. R. Scarino, A. Gopalan, C. O. Haney, P. Minnis, and K. M. Bedka (2016), A consistent AVHRR visible calibration record based on multiple methods applicable for the NOAA degrading orbits. Part II: Validation, *Journal of Atmospheric and Oceanic Technology*, 33(11), 2517–2534.
- Donat, M. G., L. V. Alexander, N. Herold, and A. J. Dittus (2016), Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations, *Journal of Geophysical Research: Atmospheres*, 121(19), 11–174.
- Doswell III, C. A., H. E. Brooks, and R. A. Maddox (1996), Flash flood forecasting: An ingredients-based methodology, *Weather and Forecasting*, 11(4), 560–581.
- Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose (2010), Comprehensive automated quality assurance of daily surface observations, *Journal of Applied Meteorology and Climatology*, 49(8), 1615–1633.

- Easterling, D. R., G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns (2000a), Climate extremes: observations, modeling, and impacts, *science*, 289(5487), 2068–2074.
- Easterling, D. R., J. Evans, P. Y. Groisman, and T. Karl (2000b), Observed variability and trends in extreme climate events: a brief review, *Bulletin of the American Meteorological Society*, 81(3), 417.
- Ek, M., K. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. Tarpley (2003), Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *Journal of Geophysical Research: Atmospheres*, 108(D22).
- Evans, J. P., M. Ekström, and F. Ji (2012), Evaluating the performance of a WRF physics ensemble over South-East Australia, *Climate Dynamics*, 39(6), 1241–1258.
- Fan, J., T.-C. Hu, and Y. K. Truong (1994), Robust non-parametric function estimation, *Scandinavian journal of statistics*, pp. 433–446.
- Ferguson, T. S. (1982), An inconsistent maximum likelihood estimate, *Journal of the American Statistical Association*, 77(380), 831–834.
- Fischer, E. M., U. Beyerle, and R. Knutti (2013), Robust spatially aggregated projections of climate extremes, *Nature Climate Change*, 3(12), 1033.
- Fraley, C., A. E. Raftery, and T. Gneiting (2010), Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging, *Monthly Weather Review*, 138(1), 190–202.
- Francq, C., and J.-M. Zakoïan (2010), Inconsistency of the MLE and inference based on weighted LS for LARCH models, *Journal of Econometrics*, 159(1), 151–165.
- Frei, C., J. H. Christensen, M. Déqué, D. Jacob, R. G. Jones, and P. L. Vidale (2003), Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps, *Journal of Geophysical Research: Atmospheres*, 108(D3).
- Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. K. Tank, and T. Peterson (2002), Observed coherent changes in climatic extremes during the second half of the twentieth century, *Climate research*, 19(3), 193–212.
- Fu, Q., and K. Liou (1992), On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres, *Journal of the Atmospheric Sciences*, 49(22), 2139–2156.
- Fu, Q., and K. N. Liou (1993), Parameterization of the radiative properties of cirrus clouds, *Journal of the Atmospheric Sciences*, 50(13), 2008–2025.

- Gan, Y., X.-Z. Liang, Q. Duan, H. I. Choi, Y. Dai, and H. Wu (2015), Stepwise sensitivity analysis from qualitative to quantitative: Application to the terrestrial hydrological modeling of a Conjunctive Surface-Subsurface Process (CSSP) land surface model, *Journal of Advances in Modeling Earth Systems*, 7(2), 648–669.
- Gandin, L. S., and A. H. Murphy (1992), Equitable skill scores for categorical forecasts, *Monthly Weather Review*, 120(2), 361–370.
- Ghil, M., P. Yiou, S. Hallegatte, B. Malamud, P. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, and V. Kossobokov (2011), Extreme events: dynamics, statistics and prediction, *Nonlinear Processes in Geophysics*, 18(3), 295–350.
- Giorgi, F., and L. O. Mearns (2002), Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “reliability ensemble averaging” (REA) method, *Journal of Climate*, 15(10), 1141–1158.
- Giorgi, F., and C. Shields (1999), Tests of precipitation parameterizations available in latest version of NCAR regional climate model (RegCM) over continental United States, *Journal of Geophysical Research: Atmospheres*, 104(D6), 6353–6375.
- Giorgi, F., C. Jones, and G. R. Asrar (2009), Addressing climate information needs at the regional level: the CORDEX framework, *World Meteorological Organization (WMO) Bulletin*, 58(3), 175.
- Giorgi, F., E. Coppola, F. Solmon, L. Mariotti, M. Sylla, X. Bi, N. Elguindi, G. Diro, V. Nair, and G. Giuliani (2012), RegCM4: model description and preliminary tests over multiple CORDEX domains, *Climate Research*, 52, 7–29.
- Gleckler, P. J., K. E. Taylor, and C. Doutriaux (2008), Performance metrics for climate models, *Journal of Geophysical Research: Atmospheres*, 113(D6).
- Gneiting, T., and A. E. Raftery (2007), Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association*, 102(477), 359–378.
- Gregory, D., J.-J. Morcrette, C. Jakob, A. Beljaars, and T. Stockdale (2000), Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecasting System, *Quarterly Journal of the Royal Meteorological Society*, 126(566), 1685–1710.
- Grell, G. A., and D. Dévényi (2002), A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, *Geophysical Research Letters*, 29(14), 38–1.
- Groisman, P. Y., T. R. Karl, D. R. Easterling, R. W. Knight, P. F. Jamason, K. J. Hennessy, R. Suppiah, C. M. Page, J. Wibig, and K. Fortuniak (1999), Changes in the probability of heavy precipitation: important indicators of climatic change, *Climatic Change*, 42(1), 243–283.

- Groisman, P. Y., R. W. Knight, D. R. Easterling, T. R. Karl, G. C. Hegerl, and V. N. Razuvaev (2005), Trends in intense precipitation in the climate record, *Journal of climate*, 18(9), 1326–1350.
- Han, J., and H.-L. Pan (2011), Revision of convection and vertical diffusion schemes in the NCEP global forecast system, *Weather and Forecasting*, 26(4), 520–533.
- Harding, K. J., and P. K. Snyder (2015), The relationship between the Pacific–North American teleconnection pattern, the Great Plains low-level jet, and North Central US heavy rainfall events, *Journal of Climate*, 28(17), 6729–6742.
- Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications.
- Haylock, M., and N. Nicholls (2000), Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910–1998, *International Journal of Climatology*, 20(13), 1533–1541.
- Hegerl, G. C., E. Black, R. P. Allan, W. J. Ingram, D. Polson, K. E. Trenberth, R. S. Chadwick, P. A. Arkin, B. B. Sarojini, and A. Becker (2018), Challenges in quantifying changes in the global water cycle, *Bulletin of the American Meteorological Society*, 99(1).
- Herman, G. R., and R. S. Schumacher (2016), Extreme precipitation in models: An evaluation, *Weather and Forecasting*, 31(6), 1853–1879.
- Herold, N., L. Alexander, M. Donat, S. Contractor, and A. Becker (2016), How much does it rain over land?, *Geophysical Research Letters*, 43(1), 341–348.
- Herold, N., A. Behrangi, and L. V. Alexander (2017), Large uncertainties in observed daily precipitation extremes over land, *Journal of Geophysical Research: Atmospheres*, 122(2), 668–681.
- Herrera, S., L. Fita, J. Fernández, and J. M. Gutiérrez (2010), Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel simulations over Spain, *Journal of Geophysical Research: Atmospheres*, 115(D21).
- Higgins, R., Y. Yao, E. Yarosh, J. E. Janowiak, and K. Mo (1997), Influence of the Great Plains low-level jet on summertime precipitation and moisture transport over the central United States, *Journal of Climate*, 10(3), 481–507.
- Hoffman, M. D., and A. Gelman (2014), The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo., *Journal of Machine Learning Research*, 15(1), 1593–1623.
- Holton, J. R. (2012), An introduction to dynamic meteorology, *American Journal of Physics*.

- Holtslag, A., and B. Boville (1993), Local versus nonlocal boundary-layer diffusion in a global climate model, *Journal of Climate*, 6(10), 1825–1842.
- Holtslag, A., E. De Bruijn, and H. Pan (1990), A high resolution air mass transformation model for short-range weather forecasting, *Monthly Weather Review*, 118(8), 1561–1575.
- Hong, S.-Y., J. Dudhia, and S.-H. Chen (2004), A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation, *Monthly Weather Review*, 132(1), 103–120.
- Hooper, D., J. Coughlan, and M. Mullen (2008), Structural equation modelling: Guidelines for determining model fit, *Articles*, p. 2.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins (2008), Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, *Journal of Geophysical Research: Atmospheres*, 113(D13).
- Iguchi, T., W.-K. Tao, D. Wu, C. Peters-Lidard, J. A. Santanello, E. Kemp, Y. Tian, J. Case, W. Wang, and R. Ferraro (2017), Sensitivity of CONUS summer rainfall to the selection of cumulus parameterization schemes in NU-WRF seasonal simulations, *Journal of Hydrometeorology*, 18(6), 1689–1706.
- Iorio, J., P. Duffy, B. Govindasamy, S. Thompson, M. Khairoutdinov, and D. Randall (2004), Effects of model resolution and subgrid-scale physics on the simulation of precipitation in the continental United States, *Climate Dynamics*, 23(3-4), 243–258.
- Jamili, A. (2016), Robust job shop scheduling problem: Mathematical models, exact and heuristic algorithms, *Expert systems with Applications*, 55, 341–350.
- Janjić, Z. I. (1994), The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes, *Monthly Weather Review*, 122(5), 927–945.
- Janjić, Z. I. (2000), Comments on “Development and evaluation of a convection scheme for use in climate models”, *Journal of the Atmospheric Sciences*, 57(21), 3686–3686.
- Kain, J. S. (2004), The Kain–Fritsch convective parameterization: an update, *Journal of applied meteorology*, 43(1), 170–181.
- Kain, J. S., and J. M. Fritsch (1993), Convective parameterization for mesoscale models: The Kain-Fritsch scheme, in *The representation of cumulus convection in numerical models*, pp. 165–170, Springer.

- Kang, I.-S., Y.-M. Yang, and W.-K. Tao (2015), GCMs with implicit and explicit representation of cloud microphysics for simulation of extreme precipitation frequency, *Climate Dynamics*, 45(1-2), 325–335.
- Karl, T. R., and R. W. Knight (1998), Secular trends of precipitation amount, frequency, and intensity in the United States, *Bulletin of the American Meteorological society*, 79(2), 231–241.
- Key, J. T., L. R. Pericchi, and A. F. Smith (1999), Bayesian model choice: what and why, *Bayesian statistics*, 6, 343–370.
- Kharin, V. V., and F. W. Zwiers (2002), Climate predictions with multimodel ensembles, *Journal of Climate*, 15(7), 793–799.
- Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl (2007), Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, *Journal of Climate*, 20(8), 1419–1444.
- Khoei, A., and S. Gharehbaghi (2007), The superconvergence patch recovery technique and data transfer operators in 3D plasticity problems, *Finite Elements in Analysis and Design*, 43(8), 630–648.
- Kirschbaum, D., R. Adler, D. Adler, C. Peters-Lidard, and G. Huffman (2012), Global distribution of extreme precipitation and high-impact landslides in 2010 relative to previous years, *Journal of hydrometeorology*, 13(5), 1536–1551.
- Kirtman, B. P., and D. Min (2009), Multimodel ensemble ENSO prediction with CCSM and CFS, *Monthly Weather Review*, 137(9), 2908–2930.
- Kline, R. B. (2016), *Principles and practice of structural equation modeling*, Guilford publications.
- Krishnamurti, T., C. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran (1999), Improved weather and seasonal climate forecasts from multimodel superensemble, *Science*, 285(5433), 1548–1550.
- Kunkel, K. E., and X.-Z. Liang (2005), GCM simulations of the climate in the central United States, *Journal of Climate*, 18(7), 1016–1031.
- Kunkel, K. E., K. Andsager, and D. R. Easterling (1999), Long-term trends in extreme precipitation events over the conterminous United States and Canada, *Journal of Climate*, 12(8), 2515–2527.
- Kunkel, K. E., K. Andsager, X.-Z. Liang, R. W. Arritt, E. S. Takle, W. J. Gutowski Jr, and Z. Pan (2002), Observations and regional climate model simulations of heavy precipitation events and seasonal anomalies: A comparison, *Journal of Hydrometeorology*, 3(3), 322–334.

- Kunkel, K. E., D. R. Easterling, D. A. Kristovich, B. Gleason, L. Stoecker, and R. Smith (2012), Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous United States, *Journal of Hydrometeorology*, 13(3), 1131–1141.
- Kunkel, K. E., T. R. Karl, H. Brooks, J. Kossin, J. H. Lawrimore, D. Arndt, L. Bosart, D. Changnon, S. L. Cutter, and N. Doesken (2013), Monitoring and understanding trends in extreme storms: State of knowledge, *Bulletin of the American Meteorological Society*, 94(4), 499–514.
- Lesk, C., P. Rowhani, and N. Ramankutty (2016), Influence of extreme weather disasters on global crop production, *Nature*, 529(7584), 84–87.
- Leung, L. R., Y. Qian, X. Bian, W. M. Washington, J. Han, and J. O. Roads (2004), Mid-century ensemble regional climate change scenarios for the western United States, *Climatic Change*, 62(1-3), 75–113.
- Li, F., W. D. Collins, M. F. Wehner, D. L. Williamson, J. G. Olson, and C. Algieri (2011), Impact of horizontal resolution on simulation of precipitation extremes in an aqua-planet version of Community Atmospheric Model (CAM3), *Tellus A: Dynamic Meteorology and Oceanography*, 63(5), 884–892.
- Li, J., and H. Barker (2005), A radiation algorithm with correlated-k distribution. Part I: Local thermal equilibrium, *Journal of the atmospheric sciences*, 62(2), 286–309.
- Li, J., and K. Shibata (2006), On the effective solar pathlength, *Journal of the atmospheric sciences*, 63(4), 1365–1373.
- Li, Y., K. Guan, G. D. Schnitkey, E. DeLucia, and B. Peng (2019), Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the United States, *Global change biology*.
- Liang, X.-Z., and F. Zhang (2013), The cloud–aerosol–radiation (CAR) ensemble modeling system, *Atmospheric Chemistry and Physics*, 13(16), 8335–8364.
- Liang, X.-Z., K. E. Kunkel, and A. N. Samel (2001), Development of a regional climate model for US Midwest applications. Part I: Sensitivity to buffer zone treatment, *Journal of Climate*, 14(23), 4363–4378.
- Liang, X.-Z., L. Li, K. E. Kunkel, M. Ting, and J. X. Wang (2004a), Regional climate model simulation of US precipitation during 1982–2002. Part I: Annual cycle, *Journal of Climate*, 17(18), 3510–3529.
- Liang, X.-Z., L. Li, A. Dai, and K. E. Kunkel (2004b), Regional climate model simulation of summer precipitation diurnal cycle over the United States, *Geophysical Research Letters*, 31(24).

- Liang, X.-Z., M. Xu, W. Gao, K. Kunkel, J. Slusser, Y. Dai, Q. Min, P. R. Houser, M. Rodell, and C. B. Schaaf (2005a), Development of land surface albedo parameterization based on Moderate Resolution Imaging Spectroradiometer (MODIS) data, *Journal of Geophysical Research: Atmospheres*, 110(D11).
- Liang, X.-Z., H. I. Choi, K. E. Kunkel, Y. Dai, E. Joseph, J. X. Wang, and P. Kumar (2005b), Surface boundary conditions for mesoscale regional climate models, *Earth Interactions*, 9(18), 1–28.
- Liang, X.-Z., M. Xu, H. Choi, K. Kunkel, L. Rontu, J.-F. Geleyn, M. D. Müller, E. Joseph, and J. X. Wang (2006), Development of the regional climate-weather research and forecasting model (CWRf): Treatment of subgrid topography effects, in *Proceedings of the 7th Annual WRF User’s Workshop*, Boulder, CO, pp. 19–22.
- Liang, X.-Z., M. Xu, K. E. Kunkel, G. A. Grell, and J. S. Kain (2007), Regional climate model simulation of US–Mexico summer precipitation using the optimal ensemble of two cumulus parameterizations, *Journal of Climate*, 20(20), 5201–5207.
- Liang, X.-Z., K. E. Kunkel, G. A. Meehl, R. G. Jones, and J. X. Wang (2008), Regional climate models downscaling analysis of general circulation models present climate biases propagation into future change projections, *Geophysical research letters*, 35(8).
- Liang, X.-Z., M. Xu, X. Yuan, T. Ling, H. I. Choi, F. Zhang, L. Chen, S. Liu, S. Su, and F. Qiao (2012a), Regional climate-weather research and forecasting model, *Bulletin of the American Meteorological Society*, 93(9), 1363–1387.
- Liang, X.-Z., M. Xu, W. Gao, K. R. Reddy, K. Kunkel, D. L. Schmoldt, and A. N. Samel (2012b), A distributed cotton growth model developed from GOSSYM and its parameter determination, *Agronomy journal*, 104(3), 661–674.
- Liang, X.-Z., C. Sun, X. Zheng, Y. Dai, M. Xu, H. I. Choi, T. Ling, F. Qiao, X. Kong, and X. Bi (2018), CWRf performance at downscaling China climate characteristics, *Climate Dynamics*, pp. 1–26.
- Lim, K.-S. S., and S.-Y. Hong (2010), Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models, *Monthly weather review*, 138(5), 1587–1612.
- Ling, T., X.-Z. Liang, M. Xu, Z. Wang, and B. Wang (2011), A multilevel ocean mixed-layer model for 2-dimension applications, *Acta Oceanol. Sin.*, 33(3), 1–10.
- Ling, T., M. Xu, X.-Z. Liang, J. X. Wang, and Y. Noh (2015), A multilevel ocean mixed layer model resolving the diurnal cycle: Development and validation, *Journal of Advances in Modeling Earth Systems*, 7(4), 1680–1692.
- Louis, J.-F. (1979), A parametric model of vertical eddy fluxes in the atmosphere, *Boundary-Layer Meteorology*, 17(2), 187–202.

- Mahoney, K., M. Alexander, J. D. Scott, and J. Barsugli (2013), High-resolution downscaled simulations of warm-season extreme precipitation events in the Colorado Front Range under past and future climates, *Journal of Climate*, 26(21), 8671–8689.
- Martynov, A., R. Laprise, L. Sushama, K. Winger, L. Šeparović, and B. Dugas (2013), Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation, *Climate dynamics*, 41(11-12), 2973–3005.
- 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, *Climate Dynamics*, 22(2-3), 183–204.
- McInnes, K., K. Walsh, G. Hubbert, and T. Beer (2003), Impact of sea-level rise and storm surges on a coastal community, *Natural Hazards*, 30(2), 187–207.
- Meehl, G. A., L. Goddard, J. Murphy, R. J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M. A. Giorgetta, A. M. Greene, and E. Hawkins (2009), Decadal prediction: can it be skillful?, *Bulletin of the American Meteorological Society*, 90(10), 1467.
- Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston (2012), An overview of the global historical climatology network-daily database, *Journal of atmospheric and oceanic technology*, 29(7), 897–910.
- Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl (2011), Human contribution to more-intense precipitation extremes, *Nature*, 470(7334), 378–381.
- Moradkhani, H., C. M. DeChant, and S. Sorooshian (2012), Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method, *Water Resources Research*, 48(12).
- Morrison, H., and J. Milbrandt (2010), Comparison of two-moment bulk microphysics schemes in idealized supercell thunderstorm simulations, *Monthly Weather Review*.
- Morrison, H., and J. A. Milbrandt (2015), Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests, *Journal of the Atmospheric Sciences*, 72(1), 287–311.
- Morrison, H., G. Thompson, and V. Tatarskii (2009), Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes, *Monthly Weather Review*, 137(3), 991–1007.
- Nakanishi, M., and H. Niino (2006), An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog, *Boundary-Layer Meteorology*, 119(2), 397–407.

- Nakanishi, M., and H. Niino (2009), Development of an improved turbulence closure model for the atmospheric boundary layer, *Journal of the Meteorological Society of Japan. Ser. II*, 87(5), 895–912.
- Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, and E. Rosero (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, *Journal of Geophysical Research: Atmospheres*, 116(D12).
- NOAA (2018), NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters.
- NOAA (2019), NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters.
- Nordeng, T. E. (1994), Extended versions of the convective parametrization scheme at ECMWF and their impact on the mean and transient activity of the model in the tropics., *Research Department Technical Memorandum*, 206, 1–41.
- Oleson, K., G.-Y. Niu, Z.-L. Yang, D. Lawrence, P. Thornton, P. Lawrence, R. Stöckli, R. Dickinson, G. Bonan, and S. Levis (2008), Improvements to the Community Land Model and their impact on the hydrological cycle, *Journal of Geophysical Research: Biogeosciences*, 113(G1).
- Pal, J. S., E. E. Small, and E. A. Eltahir (2000), Simulation of regional-scale water and energy budgets: Representation of subgrid cloud and precipitation processes within RegCM, *Journal of Geophysical Research: Atmospheres*, 105(D24), 29,579–29,594.
- Palmer, T. N. (2000), Predicting uncertainty in forecasts of weather and climate, *Reports on Progress in Physics*, 63(2), 71.
- Park, S., and C. S. Bretherton (2009), The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the Community Atmosphere Model, *Journal of Climate*, 22(12), 3449–3469.
- Pathirana, A., H. B. Deneke, W. Veerbeek, C. Zevenbergen, and A. T. Banda (2014), Impact of urban growth-driven landuse change on microclimate and extreme precipitation—A sensitivity study, *Atmospheric Research*, 138, 59–72.
- Pendergrass, A. G., and D. L. Hartmann (2014), The atmospheric energy constraint on global-mean precipitation change, *Journal of Climate*, 27(2), 757–768.
- Pfahl, S., P. A. O’Gorman, and E. M. Fischer (2017), Understanding the regional pattern of projected future changes in extreme precipitation, *Nature Climate Change*, 7(6), 423.

- Pielke, R., and C. Landsea (1998), WEATHER FORECASTS, *Weather Forecasts*, 13, 621.
- Pielke, R. A. (1999), Nine fallacies of floods, *Climatic Change*, 42(2), 413–438.
- Pihur, V., S. Datta, and S. Datta (2009), RankAggreg, an R package for weighted rank aggregation, *BMC bioinformatics*, 10(1), 62.
- Platnick, S., M. D. King, K. G. Meyer, G. Wind, N. Amarasinghe, B. Marchant, G. T. Arnold, Z. Zhang, P. A. Hubanks, and B. Ridgway (2015), MODIS cloud optical properties: User guide for the Collection 6 Level-2 MOD06/MYD06 product and associated Level-3 Datasets, Version, 1, 145.
- Pleim, J. E. (2007), A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing, *Journal of Applied Meteorology and Climatology*, 46(9), 1383–1395.
- Potts, J., C. Folland, I. Jolliffe, and D. Sexton (1996), Revised “LEPS” scores for assessing climate model simulations and long-range forecasts, *Journal of Climate*, 9(1), 34–53.
- Preacher, K. J., and A. F. Hayes (2008), Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models, *Behavior research methods*, 40(3), 879–891.
- Prein, A. F., R. M. Rasmussen, K. Ikeda, C. Liu, M. P. Clark, and G. J. Holland (2017), The future intensification of hourly precipitation extremes, *Nature Climate Change*, 7(1), 48.
- Qian, Y., H. Yan, Z. Hou, G. Johannesson, S. Klein, D. Lucas, R. Neale, P. Rasch, L. Swiler, and J. Tannahill (2015), Parametric sensitivity analysis of precipitation at global and local scales in the Community Atmosphere Model CAM5, *Journal of Advances in Modeling Earth Systems*, 7(2), 382–411.
- Qiao, F., and X.-Z. Liang (2015), Effects of cumulus parameterizations on predictions of summer flood in the Central United States, *Climate dynamics*, 45(3-4), 727–744.
- Qiao, F., and X.-Z. Liang (2016), Effects of cumulus parameterization closures on simulations of summer precipitation over the United States coastal oceans, *Journal of Advances in Modeling Earth Systems*, 8(2), 764–785.
- Qiao, F., and X.-Z. Liang (2017), Effects of cumulus parameterization closures on simulations of summer precipitation over the continental United States, *Climate Dynamics*, 49(1-2), 225–247.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005), Using Bayesian model averaging to calibrate forecast ensembles, *Monthly weather review*, 133(5), 1155–1174.

- Randles, C., A. Da Silva, V. Buchard, P. Colarco, A. Darmenov, R. Govindaraju, A. Smirnov, B. Holben, R. Ferrare, and J. Hair (2017), The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation, *Journal of climate*, 30(17), 6823–6850.
- Rasch, P., and J. Kristjánsson (1998), A comparison of the CCM3 model climate using diagnosed and predicted condensate parameterizations, *Journal of Climate*, 11(7), 1587–1614.
- Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax (2007), Daily high-resolution-blended analyses for sea surface temperature, *Journal of Climate*, 20(22), 5473–5496.
- Roeckner, E., K. Arpe, L. Bengtsson, M. Christoph, M. Claussen, L. Dümenil, M. Esch, M. A. Giorgetta, U. Schlese, and U. Schulzweida (1996), The atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate.
- Roeckner, E., G. Bäuml, L. Bonaventura, R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, I. Kirchner, L. Kornblueh, and E. Manzini (2003), The atmospheric general circulation model ECHAM 5. PART I: Model description.
- Rosseel, Y. (2012), Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA), *Journal of statistical software*, 48(2), 1–36.
- Rubin, D. B. (1981), The bayesian bootstrap, *The annals of statistics*, pp. 130–134.
- Samuelsson, P., S. Gollvik, and A. Ullerstig (2006), The land-surface scheme of the Rossby Centre regional atmospheric climate model (RCA3), SMHI.
- Samuelsson, P., C. G. Jones, U. Willen, A. Ullerstig, S. Gollvik, U. Hansson, E. Jansson, C. Kjellstrom, G. Nikulin, and K. Wyser (2011), The Rossby Centre Regional Climate model RCA3: model description and performance, *Tellus A: Dynamic Meteorology and Oceanography*, 63(1), 4–23.
- Schär, C., N. Ban, E. M. Fischer, J. Rajczak, J. Schmidli, C. Frei, F. Giorgi, T. R. Karl, E. J. Kendon, and A. M. K. Tank (2016), Percentile indices for assessing changes in heavy precipitation events, *Climatic Change*, pp. 1–16.
- Schulz, J., P. Albert, H.-D. Behr, D. Caprion, H. Deneke, S. Dewitte, B. Durr, P. Fuchs, A. Gratzki, and P. Hechler (2009), Operational climate monitoring from space: the EUMETSAT Satellite Application Facility on Climate Monitoring (CM-SAF), *Atmospheric Chemistry and Physics*, 9(5), 1687–1709.
- Scinocca, J., V. Kharin, Y. Jiao, M. Qian, M. Lazare, L. Solheim, G. Flato, S. Biner, M. Desgagne, and B. Dugas (2016), Coordinated global and regional climate modeling, *Journal of Climate*, 29(1), 17–35.

- Semmler, T., and D. Jacob (2004), Modeling extreme precipitation events—a climate change simulation for Europe, *Global and Planetary Change*, 44(1), 119–127.
- Siegert, S., C. A. Ferro, and D. B. Stephenson (2014), Evaluating ensemble forecasts by the Ignorance score—Correcting the finite-ensemble bias, *arXiv preprint arXiv:1410.8249*.
- Siler, N., and G. Roe (2014), How will orographic precipitation respond to surface warming? An idealized thermodynamic perspective, *Geophysical Research Letters*, 41(7), 2606–2613.
- Sillmann, J., V. Kharin, X. Zhang, F. Zwiers, and D. Bronaugh (2013), Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate, *Journal of Geophysical Research: Atmospheres*, 118(4), 1716–1733.
- Sillmann, J., T. Thorarinsdottir, N. Keenlyside, N. Schaller, L. V. Alexander, G. Hegerl, S. I. Seneviratne, R. Vautard, X. Zhang, and F. W. Zwiers (2017), Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities, *Weather and climate extremes*, 18, 65–74.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers (2005), A description of the advanced research WRF version 2, Tech. rep., National Center For Atmospheric Research Boulder Co Mesoscale and Microscale
- Sloughter, J. M. L., A. E. Raftery, T. Gneiting, and C. Fraley (2007), Probabilistic quantitative precipitation forecasting using Bayesian model averaging, *Monthly Weather Review*, 135(9), 3209–3220.
- Smith, A. B., and J. L. Matthews (2015), Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates, *Natural Hazards*, 77(3), 1829–1851.
- Smith, L. A. (2002), What might we learn from climate forecasts?, *Proceedings of the National Academy of Sciences*, 99, 2487–2492.
- Stephens, G. L., T. L’Ecuyer, R. Forbes, A. Gettelmen, J.-C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes (2010), Dreary state of precipitation in global models, *Journal of Geophysical Research: Atmospheres*, 115(D24).
- Subin, Z. M., W. J. Riley, and D. Mironov (2012), An improved lake model for climate simulations: Model structure, evaluation, and sensitivity analyses in CESM1, *Journal of Advances in Modeling Earth Systems*, 4(1).
- Sun, C., and X.-Z. Liang (“submitted”,a), Improving U.S. extreme precipitation simulation: Sensitivity to physics parameterizations, *Journal of Climate*.

- Sun, C., and X.-Z. Liang ("submitted",b), Improving U.S. extreme precipitation simulation: Dependence on cumulus parameterization and underlying mechanism, *Journal of Climate*.
- Sun, D.-Z., J. Fasullo, T. Zhang, and A. Roubicek (2003), On the radiative and dynamical feedbacks over the equatorial Pacific cold tongue, *Journal of Climate*, 16(14), 2425–2432.
- Sun, Y., S. Solomon, A. Dai, and R. W. Portmann (2006), How often does it rain?, *Journal of Climate*, 19(6), 916–934.
- Sundqvist, H. (1978), A parameterization scheme for non-convective condensation including prediction of cloud water content, *Quarterly Journal of the Royal Meteorological Society*, 104(441), 677–690.
- Tao, W.-K., J. Simpson, and M. McCumber (1989), An ice-water saturation adjustment, *Monthly Weather Review*, 117(1), 231–235.
- Tao, W.-K., J. Simpson, D. Baker, S. Braun, M.-D. Chou, B. Ferrier, D. Johnson, A. Khain, S. Lang, and B. Lynn (2003), Microphysics, radiation and surface processes in the Goddard Cumulus Ensemble (GCE) model, *Meteorology and Atmospheric Physics*, 82(1), 97–137.
- Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183–7192.
- Team, S. S. (2008), SRB Data, Hampton, VA, USA: NASA Atmospheric Science Data Center (ASDC).
- Thompson, G., and T. Eidhammer (2014), A study of aerosol impacts on clouds and precipitation development in a large winter cyclone, *Journal of the atmospheric sciences*, 71(10), 3636–3658.
- Thompson, G., R. M. Rasmussen, and K. Manning (2004), Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis, *Monthly Weather Review*, 132(2), 519–542.
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall (2008), Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization, *Monthly Weather Review*, 136(12), 5095–5115.
- Tiedtke, M. (1989), A comprehensive mass flux scheme for cumulus parameterization in large-scale models, *Monthly Weather Review*, 117(8), 1779–1800.
- Tran, M.-N. (2011), A criterion for optimal predictive model selection, *Communications in Statistics—Theory and Methods*, 40(5), 893–906.

- Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons (2003), The changing character of precipitation, *Bulletin of the American Meteorological Society*, 84(9), 1205–1217.
- Tripathi, O. P., and F. Dominguez (2013), Effects of spatial resolution in the simulation of daily and subdaily precipitation in the southwestern US, *Journal of Geophysical Research: Atmospheres*, 118(14), 7591–7605.
- Turner, B. M., and P. B. Sederberg (2012), Approximate Bayesian computation with differential evolution, *Journal of Mathematical Psychology*, 56(5), 375–385.
- USGCRP, D. (2017), *Climate Science Special Report: Fourth National Climate Assessment, Volume I*, Page, US Global Change Research Program, Washington, DC, USA.
- Vehtari, A., A. Gelman, and J. Gabry (2017), Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC, *Statistics and Computing*, 27(5), 1413–1432.
- Verseghy, D., N. McFarlane, and M. Lazare (1993), CLASS—A Canadian land surface scheme for GCMs, II. Vegetation model and coupled runs, *International Journal of Climatology*, 13(4), 347–370.
- Verseghy, D. L. (1991), CLASS—A Canadian land surface scheme for GCMs. I. Soil model, *International Journal of Climatology*, 11(2), 111–133.
- Vrugt, J. A., C. Ter Braak, C. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3), 273–290.
- Wang, J., and V. R. Kotamarthi (2015), High-resolution dynamically downscaled projections of precipitation in the mid and late 21st century over North America, *Earth’s Future*, 3(7), 268–288.
- Wang, W., and N. L. Seaman (1997), A comparison study of convective parameterization schemes in a mesoscale model, *Monthly Weather Review*, 125(2), 252–278.
- Watanabe, M., H. Shiogama, T. Yokohata, Y. Kamae, M. Yoshimori, T. Ogura, J. D. Annan, J. C. Hargreaves, S. Emori, and M. Kimoto (2012), Using a multiphysics ensemble for exploring diversity in cloud–shortwave feedback in GCMs, *Journal of Climate*, 25(15), 5416–5431.
- Wehner, M. F., R. L. Smith, G. Bala, and P. Duffy (2010), The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model, *Climate dynamics*, 34(2-3), 241–247.

- Wilcox, E. M., and L. J. Donner (2007), The frequency of extreme rain events in satellite rain-rate estimates and an atmospheric general circulation model, *Journal of Climate*, 20(1), 53–69.
- Wilson, D. R., A. Bushell, A. M. Kerr-Munslow, J. D. Price, C. J. Morcrette, and A. Bodas-Salcedo (2008), PC2: A prognostic cloud fraction and condensation scheme. II: Climate model simulations, *Quarterly Journal of the Royal Meteorological Society*, 134(637), 2109–2125.
- Wuebbles, D. J., K. Kunkel, M. Wehner, and Z. Zobel (2014), Severe weather in United States under a changing climate, *Eos, Transactions American Geophysical Union*, 95(18), 149–150.
- Wuebbles, D. J., D. W. Fahey, and K. A. Hibbard (2017), Climate science special report: fourth national climate assessment, volume I.
- Xie, J., and M. Zhang (2017), Role of internal atmospheric variability in the 2015 extreme winter climate over the North American continent, *Geophysical Research Letters*, 44(5), 2464–2471.
- Xie, S., M. Zhang, J. S. Boyle, R. T. Cederwall, G. L. Potter, and W. Lin (2004), Impact of a revised convective triggering mechanism on Community Atmosphere Model, Version 2, simulations: Results from short-range weather forecasts, *Journal of Geophysical Research: Atmospheres*, 109(D14).
- Xie, S.-P., C. Deser, G. A. Vecchi, M. Collins, T. L. Delworth, A. Hall, E. Hawkins, N. C. Johnson, C. Cassou, and A. Giannini (2015), Towards predictive understanding of regional climate change, *Nature Climate Change*.
- Xu, K.-M., and D. A. Randall (1996), A semiempirical cloudiness parameterization for use in climate models, *Journal of the atmospheric sciences*, 53(21), 3084–3102.
- Xu, M., X.-Z. Liang, A. Samel, and W. Gao (2014), MODIS consistent vegetation parameter specifications and their impacts on regional climate simulations, *Journal of Climate*, 27(22), 8578–8596.
- Yang, Z.-L., G.-Y. Niu, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, L. Longuevergne, K. Manning, D. Niyogi, and M. Tewari (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins, *Journal of Geophysical Research: Atmospheres*, 116(D12).
- Yao, Y., A. Vehtari, D. Simpson, and A. Gelman (2018), Using stacking to average Bayesian predictive distributions (with discussion), *Bayesian Analysis*, 13(3), 917–1003.
- Yuan, X., and X.-Z. Liang (2011), Evaluation of a Conjunctive Surface-Subsurface Process Model (CSSP) over the contiguous United States at regional-local scales, *Journal of Hydrometeorology*, 12(4), 579–599.

- Yuan, X., X.-Z. Liang, and E. F. Wood (2012), WRF ensemble downscaling seasonal forecasts of China winter precipitation during 1982–2008, *Climate dynamics*, 39(7–8), 2041–2058.
- Zhang, C., Y. Wang, and K. Hamilton (2011a), Improved representation of boundary layer clouds over the southeast Pacific in ARW-WRF using a modified Tiedtke cumulus parameterization scheme, *Monthly Weather Review*, 139(11), 3489–3513.
- Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa (2007), Detection of human influence on twentieth-century precipitation trends, *Nature*, 448(7152), 461–465.
- Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers (2011b), Indices for monitoring changes in extremes based on daily temperature and precipitation data, *Wiley Interdisciplinary Reviews: Climate Change*, 2(6), 851–870.
- Zhang, X., H. Wan, F. W. Zwiers, G. C. Hegerl, and S.-K. Min (2013), Attributing intensification of precipitation extremes to human influence, *Geophysical Research Letters*, 40(19), 5252–5257.
- Zhu, J., and X.-Z. Liang (2013), Impacts of the Bermuda high on regional climate and ozone over the United States, *Journal of Climate*, 26(3), 1018–1032.
- Zobel, Z., J. Wang, D. J. Wuebbles, and V. R. Kotamarthi (2018), Evaluations of high-resolution dynamically downscaled ensembles over the contiguous United States, *Climate dynamics*, 50(3–4), 863–884.