Evaluation of the WRF physics ensemble using a multivariable integrated evaluation approach over the Haihe river basin in northern China

Danqiong Dai[1,2] · Liang Chen[1,2] · Zhuguo Ma[1,2] · Zhongfeng Xu[1]

Abstract

A crucial step in the application of the Weather Research and Forecasting (WRF) model to regional climate research is the selection of the proper combinations of physical parameterizations. In this study, we examined the performance of various parametrization schemes in the WRF model in terms of precipitation and temperature over the Haihe river basin in northern China. The WRF experiments were integrated with 13-km horizontal resolution and driven by ERA-INTERIM reanalysis data over the period from 1st June to 31st August, 2016. Fifty-eight members of physics combinations derived from five types of physics options were assessed against the available observational temperature and precipitation data by utilizing the multivariable integrated evaluation (MVIE) method. Our results indicated that the best combination of physical schemes consisted of CAM5.1 microphysics, MRF PBL, BMJ cumulus, CAM Longwave/Shortwave radiation, and Noah Land Surface schemes. The optimal setup's differences with the observational data, temporally and spatially, were much smaller than other setups in terms of surface air temperature and precipitation, which proves that the optimal setup showed better performance than the other setups. Further analysis of the sensitivities of model outputs to different types of physics options suggests that the microphysics, planetary boundary layer (PBL), and cumulus schemes have a more significant impact on the model performances than the radiation scheme and Land Surface schemes.

Keywords WRF · Haihe River Basin · MVIE method · Surface temperature · Precipitation

1 Introduction

With advantages such as low computational cost and a range of output variables, applications of numerical weather prediction (NWP) models have become increasingly common in climate research and operational weather forecasting in recent decades. Among these models, general circulation models (GCMs) provide comprehensive predictions of large-scale climate events (Gillett and Thompson 2003; Osborn 2004), and regional climate models (RCMs) can accomplish high-resolution runs over restricted areas and explore local circulation dynamics at lower computational cost (Dickinson et al. 1989; Wang et al. 2004; Giorgi 2006). Typically, the Advanced Research WRF model (ARW), designed to serve both regional operational forecasting and atmospheric research needs, is increasingly in use throughout the world as a regional climate model (Skamarock et al. 2008; Bukovsky and Karoly 2009; Argüeso et al. 2011).

Prior to utilizing the WRF model for climate simulations, it is crucial that model outputs be analyzed against observational data to assess their ability to capture spatial and temporal distributions. To our knowledge, variations in output from the WRF model depend on many factors, including the model itself (Giorgi and Bi 2000; Christensen et al. 2001), boundary conditions (Von Storch et al. 2000; Denis et al. 2002), geographic region (Seth and Giorgi 1998; Landman et al. 2005), and parameterization schemes (Ratnam and Kumar 2005; Tegoulias et al. 2017). Given these factors, the choice of parameterization schemes in WRF is one substantial source of model uncertainty (Jerez et al. 2013; Mooney et al. 2013). A wide

✉ Liang Chen
chenliang@tea.ac.cn

1 Key Laboratory of Regional Climate and Environment for Temperate East Asia, RCE-TEA, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

2 University of Chinese Academy of Sciences, Beijing, China

range of parameterizations are available in WRF, which differ in their level of complexity and their representation of physical processes. However, there is no single configuration optimal for all locations, variables, and at every possible timescale (Fernández et al. 2007; Borge et al. 2008). It is therefore necessary to identify the optimal set of schemes applicable to the domain of interest (Ferreira et al. 2014; Chen et al. 2014; Pieri et al. 2015).

Over the years, significant developments have been achieved in the methods for evaluating the sensitivities to WRF parameterizations or the simulations of models (Fernández et al. 2007; Hu et al. 2010; Evans et al. 2012; Crétat et al. 2012; Li et al. 2016; Giannaros et al. 2019). Furthermore, a variety of measures, such as metrics (Willmott et al. 1985; Taylor 2001; Gleckler et al. 2008) and statistical approaches (Perkins et al. 2007; Quan et al. 2016; Budakoti et al. 2019), have been applied to calculate some key traceable components across different scales, such as precipitation, temperature, and surface flux (Tian et al. 2017; Yáñez-Morroni et al. 2018). Gallus and Bresch (2006) utilized WRF to model multiple events to compare the sensitivity of warm season rainfall forecasts to changes in model physics, dynamics, and initial conditions. They adopted statistical evaluation indexes, such as the equitable threat score (ETS; Schaefer 1990) and bias, to measure forecast accuracy and explore the key impacting factors of different types of rainfall events. Bastidas et al. (2006) carried out a comparison of land surface model sensitivity within a multicriteria framework. Bukovsky and Karoly (2009) discussed the impacts of changes in convective and land surface parameterizations, nest feedbacks, sea surface temperature, and WRF version on mean precipitation in four-month-long simulations by utilizing a subjective evaluation. Argüeso et al. (2011) evaluated the WRF sensitivity to eight different combinations of cumulus, microphysics, and planetary boundary layer (PBL) parameterization schemes over southern Spain for the period 1990–99. Cohen et al. (2015) summarized the key characteristics of the various PBL parameterization schemes employed to simulate the southeastern USA cold season severe thunderstorm environment. Their method of evaluating the performance of models focused on using a framework for error analysis often applied in economic forecast analysis. Hasan et al. (2018) made a comparison between observed and simulated rainfall over Bangladesh using 19 different combinations of microphysics and cumulus schemes available in WRF. The study found a combination of the Stony Brook University microphysics schemes with the Tiedtke cumulus scheme was the most suitable scheme for reproducing these events. However, an evaluation of the performance of high-resolution models and a quantitative method for evaluating the overall performance in simulating multiple fields are still lacking. Additionally, the number of parameterizations has increased, and no robust evaluation of

the five types of physics options is available over the Haihe River Basin.

The goal of this study was to evaluate the skill of different physics scheme combinations in reproducing precipitation and temperature over the Haihe River Basin. The results will provide useful information about the optimal parameterization sets over the river basin and suggestions for which physics options may be sensitive to simulating the key component. The study region is characterized by being the political, economic, cultural, and transport center of China. Also, severe shortage of water resources and related environmental problems in this region, have become major critical challenges for regional social and economic development. Another important highlight of the present study that distinguishes it from others focuses on utilizing the multivariable integrated evaluation (MVIE) method (Xu et al. 2016, 2017) to measure the overall performance of WRF in simulating multiple fields, The MVIE method consists of multiple statistics that measure model performance from different aspects. It provides a framework that can evaluate model ability to simulate individual variables and the overall model ability to simulating multiple fields. The MVIE method is particularly suitable for this study that compares and ranks various WRF simulations in terms of their ability to simulate multiple fields.

Section 2 provides the model setup and descriptions of the study area and the evaluation method. Section 3 describes the evaluation process of each scheme option, and a brief discussion of the robustness of the optimal schemes is presented in Sect. 3.6. Our main conclusions are given in Sect. 4.

2 Methodology

2.1 Study site

The study area chosen was the Haihe River Basin (Fig. 1), located between 35°–43°N and 112°–120°E in northern China. The basin is surrounded by the Taihang Mountains to the west, the Bohai Sea to the east, the Mongolia Plateau to the north, and the lower reaches of the Yellow River to the south. The Beijing-Tianjin-Hebei megacity is located in the Haihe River basin where three major rivers, Haihe river, Luanhe river, and Tuhaimajia river flow through. The mountain areas are concentrated in the western parts of the basin, and the central, eastern, and southern areas are plain areas. This entire area belongs to the Temperate East Asia monsoon climate zone, which is mainly hot and wet in summer and cold and dry in winter. Additionally, this area is distinguished by an average annual temperature of 10.4 °C and average annual precipitation and evaporation of approximately 541.6 and
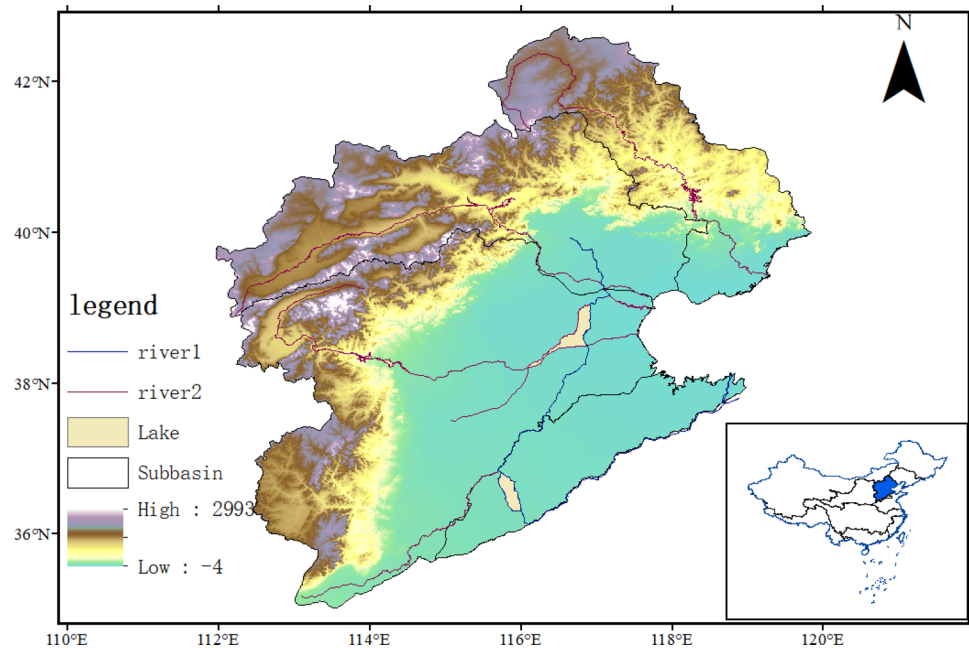
470 mm, respectively (He et al. 2015). Rainfall shows strong spatial and temporal variability in this area (Wang et al. 2019). Over 70% of the annual precipitation occurs during the summer season (June–August) in the form of rainstorms. The strongest evaporation was observed in the winter wheat-growing season (March–June), with about 55% of the annual total. Drought often occurs in spring as a result of low precipitation, high evaporation, and higher temperatures. The Haihe river basin is the economic, political, and cultural center of China, with a high density of population and huge demands for resources. And this area has limited water resources and often suffers from drought disaster in recent decades associated the weakening monsoon circulation (Wang et al. 2001; Ding et al. 2008). Drought is one of these major weather-related disasters and has serious environmental, social and economic impacts (Ma et al. 2006,2007). Many previous studies focused on the spatiotemporal characteristics of drought and the causes of drought information (Ding et al. 2010, Li et al. 2015, Duan et al. 2017). And a mesoscale climate model (i.e., WRF) is a powerful tool to simulate meteorological and land surface conditions relating to drought. However, one of the most important uncertainties of WRF simulation is related to the choice of optimal physical schemes. Only a few studies have been conducted in this area using the WRF model to investigate the changes in regional climate. Therefore, it is of importance to evaluate the overall performance of WRF in this area in order to determine an optimal set of parameterization combinations applicable to the Haihe River Basin.

2.2 Model configuration and datasets

In this study, the numerical experiments were conducted using the Advanced Research WRF model version 3.9.1 (Table 1), which is a non-hydrostatic, primitive-equation model providing many different physical and running options for a wide spectrum of applications at very different scales, from large-eddy simulations to climate simulations (see https://www2.mmm.ucar.edu/wrf/users). Physical options are composed of various parameterization schemes to describe sub-grid-scale processes, such as microphysics, cumulus convection, near-surface physics, land surface physics, planetary boundary-layer physics, and atmospheric longwave and shortwave radiative transfer. For each physical option, there are many parameterization schemes available

Table 1 Overview of WRF model configurations

Model version	WRF/ARW (3.9.1)
Vertical layers	51 layers (50 hPa)
Horizontal resolution	13 km
Horizontal grid	87 (E–W) × 80 (S–N)
Time step	180 s
Initial and boundary conditions	ERA-interim
Simulation period	June 1st–Aug 31st, 2016
Terrain and land-use data	MODIS
Default parameterization schemes	Cumulus–Kain Fritsch (new eta)
	Microphysics–WSM6
	PBL/surface layer scheme–YSU/ MM5
	Radiation–CAM3
	LSM–Noah

in the WRF model (Table 2). For the study area, we chose the default combination of parameterization schemes that is often used for regional climates in WRF reference physical suits. They are WSM6 microphysics scheme, the YSU planetary boundary-layer scheme, the KF cumulus scheme, the CAM shortwave and longwave radiation, and the Noah Land Surface scheme. The horizontal resolution of the simulation area encompassing the Haihe River Basin was set at 13 km, with a total of 87 points in the east–west direction and 80 points in the north–south direction. In the vertical direction, there were 51 stretched vertical levels topped at 50 hPa. The initial and lateral boundary conditions were obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) six-hourly ERA-Interim reanalysis. Land use categorization in WRF 3.9.1 was determined from Moderate Resolution Imaging Spectroradiometer (MODIS) data. The time step of the WRF model was set to 180 s, and the output frequency of the WRF model was 3 h. Simulations were carried out for the four-month period of summer (May 1st–August 31st) in 2016. The first month was discarded for spin-up purpose. The other three months, i.e., June, July, and August 2016, were used to evaluate performance of different combinations of physical parameterizations. These simulations were assessed in terms of temperature and precipitation against hourly rainfall and daily surface air temperature observation from 261 stations in the Haihe River Basin.

2.3 Experimental design and evaluation methodology

Use of the rule of statistical permutation and combination of five types of physics options to generate different scheme sets would have resulted in millions of experiments. Therefore, in order to reduce the computational cost effectively, we analyzed the interactions between WRF parameterization schemes and further utilized the step-wise refinement method to determine the optimal combination of physics parameterizations from five types of scheme options (Stergiou et al. 2017). As shown in Fig. 2, the first set of WRF simulations were kept the same as the default model configuration, except that the microphysics schemes were chosen from 19 kinds of schemes. The second set of WRF simulations was consistent with the default schemes, except that the microphysics scheme was adjusted for the optimal scheme, which further modified the choices for the planetary boundary-layer physics. A third set of experiments were performed, which were consistent with the second set of simulations, except that the cumulus schemes were assessed on the basis of the optimal PBL scheme. There was a concern that free combinations of all existing longwave and shortwave radiation schemes would not be able to run the WRF model successfully. Thus, some combinations of both longwave and shortwave radiation schemes available to begin the WRF

model were applied to the fourth part of this study when the model configurations were consistent with the third set of simulations and the cumulus scheme was found to be the best option. Finally, the fifth set of experiments focused on sensitivity analysis of the land surface schemes in relation to selection of the other four best-performing schemes. Therefore, in order to carry out all the experiments for this study, we chose to select the best setup for runs supporting climate simulation enhancement over the Haihe River Basin.

To comprehensively evaluate and rank the WRF simulations, we employed a multivariable integrated evaluation (hereafter MVIE) method developed by Xu et al.(2017). The MVIE framework consists of three levels of statistical metrics (Table 3). The first level of metrics, consists of the commonly used mean error (ME), standard deviation (hereafter SD) value, correlation coefficient (CORR), and root-mean-square deviation (hereafter RMSD). These metrics assess model performance in terms of individual variables (i.e. temperature and precipitation). For example, the standard deviation (SD) measures the variance of a scalar field. Correlation coefficient describes the pattern similarity between the two scalar fields (here referring to model outputs and observations). RMSD measures the overall difference of two scalar fields. The RMSD is a function of mean error, standard deviation, and correlation coefficient (Xu and Han 2020). The second level of metrics include the root-mean-square length (hereafter RMSL) of a vector field, the vector field similarity coefficient (hereafter VSC), and the root-mean-square vector deviation (hereafter RMSVD). These metrics can be defined in centered or uncentered forms. In this study, we employed the centered statistical metrics. The centered RMSL (cRMSL) is analogous to standard deviation but it describes the variance of a vector field. The centered VSC (cVSC) is analogous to Pearson's correlation coefficient but it measures the pattern similarity between two vector fields. Similarly, the centered RMSVD (cRMSVD) is analogous to RMSD except that it measures the overall difference between two vector fields. Similar to the Taylor diagram (Taylor 2001), one can show these vector statistics on a vector field evaluation (VFE) diagram. The VFE diagram is a generalized Taylor diagram, which can summarize model performance in simulating vector fields (Xu et al. 2016). The VFE diagram can intuitively reveal to what extent the overall RMSVD of various fields are separately attributable to the bias in variance (represented by the RMSL) and the pattern similarities (represented by VSC). The VFE methods were successfully applied to evaluate model performance in simulating vector fields (e.g. Wang et al. 2019; Huang et al. 2018, 2020).

Based on the vector statistics, Xu et al. (2017) devised a MVIE method to evaluate model performance in simulating multiple fields. The general idea of MVIE is to normalized various scalar fields and group them into a multiple

Table 2 Physics parameterization scheme assessed in this study

Physics parameterization scheme		Acronym
Microphysics	Kessler Scheme	Kessler
	Lin et al. Scheme	Lin
	WSM3 Single-moment 3-class Scheme	WSM3_ice
	WSM5 Single-moment 5-class Scheme	WSM5
	Ferrier	Ferrier
	WSM6 Single-moment 6-class Scheme	WSM6_Grau
	Goddard Scheme	Goddard_GCE
	Thompson Scheme	Thompson
	Milbrandt-Yau Double Moment Scheme	Milbrandt-Yau
	Morrison 2-moment Scheme	Morrison
	CAM V5.1 2-moment 5-class Scheme	CAM5.1
	SBU Stony-Brook University Scheme	SBU_YLIN
	WDM5 Double Moment 5-class Scheme	WDM5
	WDM6 Double Moment 6-class Scheme	WDM6
	NSSL 2-moment Scheme	NSSL17
	NSSL 2-moment Scheme with CCN prediction	NSSL18
	NSSL 1-moment 7-class Scheme	NSSL19
	NSSL 1-moment 6-class Scheme	NSSL21
	Aerosol-aware Thompson Scheme	Thompson_aer
Planetary boundary layer (PBL)/Suface Layer	Yonsei University Scheme / MM5	YSU
	Mellor–Yamada–Janjic Scheme / Eta	MYJ
	Quasi–normal Scale Elimination Scheme	QSNE
	Mellor–Yamada Nakanishi Niino Level 2.5/MM5	MYNN2-1
	MYNN2/Eta	MYNN2-2
	MYNN2/MYNN	MYNN2-5
	Mellor–Yamada Nakanishi Niino Level 3/MM5	MYNN3-1
	MYNN3/Eta	MYNN3-2
	MYNN3/MYNN	MYNN3-5
	ACM2/MM5 Asymmetric Convection Model 2 Scheme	ACM2-1
	ACM2/Pleim-Xiu	ACM2-7
	Bougeault–Lacarrere Scheme/MM5	BouLac-1
	BouLac/Eta	BouLac-2
	University of Washington Scheme/MM5	UW-1
	UW/Eta	UW-2
	Shin-Hong/MM5 Scale–aware Scheme	Shin-Hong
	MRF/MM5	MRF
Cumulus convection	Kain–Fritsch Scheme	KF
	BMJ Betts–Miller–Janjic Scheme	BMJ
	GF Grell–Freitas Ensemble Scheme	GFE
	OSAS Old Simplified Arakawa–Schubert	G3
	G3 Grell 3D Ensemble Scheme	Tiedtke
	Tiedtke Scheme	KF_PDF
	NSAS New Simplified Arakawa–Schubert	NSAS
	New Tiedtke Scheme	NewTiedtke
	GD Grell–Devenyi Ensemble Scheme	GD
	old KF Old Kain–Fritsch Scheme	oldKF
Longwave radiation	CAM Longwave Scheme	CAM
	RRTMG Longwave Scheme	RRTMG
	RRTMG Fast Version	Fast_RRTMG

Table 2 (continued)	Physics parameterization scheme		Acronym
	Shortwave radiation	CAM Shortwave Scheme	CAM
		RRTMG Shortwave Scheme	RRTMG
		RRTMG Fast Version	Fast_RRTMG
	Land surface	5–layer Thermal Diffusion	TD5
		Unified Noah Land Surface Model	Noah
		Noah–MP Land Surface Model	NoahMP
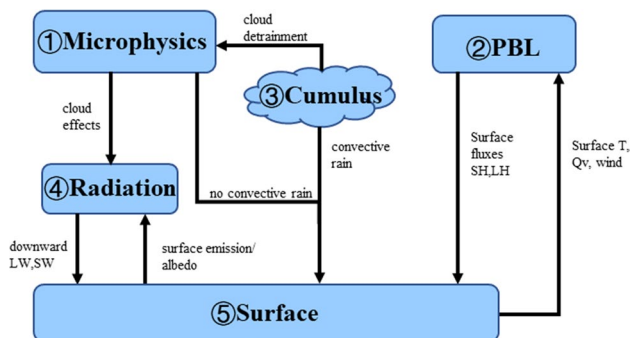
Fig. 2 Interactions between the WRF parameterization schemes

dimensional vector field. Then one can evaluate the modelled constructed vector fields to the observed one using the vector statistical metrics. To summarize and rank the overall model performance, Xu et al. (2017) defined a multivariable integrated evaluation index (MIEI), which takes multiple statistics, e.g., mean error, pattern similarity, and variance, of multiple variables into account. In this study, we normalize the temperature and precipitation and group them into a two-dimensional vector field. Thus, the grouped vector field derived from model can be assessed against that derived from observation by using the VFE diagram and MIEI. The computation of three levels of metrics consists of the following steps:

Step 1: Prepare the two groups of datasets (the model data and observations) for evaluation. The original model field for uncentered mode is produced by interpolating WRF model output to the closest observational station location. In this study, the anomaly field for centered mode is produced by the original field minus the mean value.

Step 2: Calculate the first level of metrics for individual variables in Table 3. In centered mode, the SD, cRMSD, CORR, ME of the anomaly field are calculated and displayed in Result and discussion part. In uncentered mode, the rms, uCORR, RMSD of the original field are calculated and not showed here.

Step 3: Calculate the second level of metrics for multivariable integrated field in Table 3. For the centered mode in this study, centered RMSL (cRMSL), centered VSC (cVSC), centered RMSVD (cRMSVD) and VME are calculated in terms of anomaly fields. For the uncentered mode (not shown here), the RMSL, VSC, and RMSL are calculated in terms of the original field.

Step 4: Calculate the third level of metrics for summarizing overall performance in Table 3, MIEI is calculated in terms of the original field to rank the performance of the different parameterizations.

3 Results and discussion

3.1 Microphysics

The other WRF parameterizations, except for the microphysics option, were generally held as default options in the present analysis, which represents a sensitivity analysis of the simulation results with different microphysics schemes. Microphysics schemes contain explicitly resolved water vapor, cloud, and precipitation processes. In the current version of the ARW, microphysics is able to accommodate any number of mass mixing-ratio variables, and other quantities such as number concentrations. In this study, we chose 19 commonly used microphysics parameterizations, and we calculated three levels of metrics for each WRF simulation with certain microphysics parameterizations (Table 4). It is immediately obvious that the option 11, CAM5.1, shows the minimum MIEI (0.71) which is selected as the best microphysics parameterization scheme.

The metrics table (Table 4) of the centered statistics decomposes the original field into mean and anomaly fields for evaluation. The anomaly fields are further evaluated from the perspective of variance (SD, cRMSL), pattern similarity (CORR, cVSC) and overall difference between the model and observation (cRMSD, cRMSVD). Note that the mean error (ME) is additionally computed for the centered statistics, as the centered statistics exclude mean error. For temperature and precipitation, the mean error (ME) values of CAM5.1 have the minimum values (− 0.96 and 1.21). Undoubtedly, the vector mean error (VME) of CAM5.1, measuring the difference between two mean vector fields, shows the minimum value (1.54) among 19 kinds of microphysics schemes. For the centered mode, the cRMSLs, reflecting the total standard

Table 3 Formulas of the statistical metrics in the multivariable integrated evaluation (MVIE) method

Acronyms	Denotation	Formula
Statistics for individual variables		
ME	Mean error	$ME_i = \overline{a}_i - \overline{b}_i, \ \overline{a}_i = \sum_{j=1}^{N} a_{ij}, \overline{b}_i = \sum_{j=1}^{N} b_{ij}$
SD	Standard coefficient	$L_{ai} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} a_{ij}^2}$ or $L_{bi} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} b_{ij}^2}$
rms	Root-mean-square	
CORR	Correlation coefficient	$R_{ui} = \dfrac{\frac{1}{N} \sum_{j=1}^{N} a_{ij} b_{ij}}{\sigma_{ai} \sigma_{bi}}$
uCORR	Uncentered correlation coefficient	$\sigma_{ai} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} a_{ij}^2}$ or $\sigma_{bi} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} b_{ij}^2}$
RMSD	Root-mean-square-difference	$RMSD_i = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (a_{ij} - b_{ij})^2}$
cRMSD	Centered root-mean-square difference	
Statistics for multivariable integrated field		
RMSL	Root mean square length	$L_A = \sqrt{\sum_{i=1}^{M} L_{ai}^2}$ or $L_B = \sqrt{\sum_{i=1}^{M} L_{bi}^2}$
cRMSL	Centered root-mean-square length	
VSC	Vector similarity coefficient	$R_v = \frac{1}{L_A L_B} \sum_{i=1}^{M} \sigma_{ai} \sigma_{bi} R_{ui}$
cVSC	Centered vector similarity coefficient	
RMSVD	Root-mean-square vector difference	$RMSVD = \sqrt{\sum_{i=1}^{M} (\frac{1}{N} \sum_{j=1}^{N} (a_{ij} - b_{ij})^2)}$
cRMSVD	Centered root-mean-square vector difference	
VME	Vector mean error	$VME = \sqrt{\sum_{1}^{M} (\overline{a}_i - \overline{b}_i)^2}$
Multivariable integrated evaluation index		
MIEI	Multivariable integrated evaluation index (MIEI)	$MIEI = \frac{1}{M} \sum_{i=1}^{M} (L_{ai}^* - L_{bi}^*)^2 + 2(1 - R_v)$, $\quad L_{ai}^* = \frac{L_{ai}}{L_{bi}}$ or $L_{bi}^* = \frac{L_{bi}}{L_{bi}}$

Here assume the vector fields (***A*** and ***B***) are derived from a climate model simulation and observation, respectively. Each vector field (A and B) consists of N dimensions (in time or/and space, here in time, j = 1, 2,…,N), and has M scalar fields (here refer to precipitation and temperature, i = 1,2,…,M). "a" refers to a certain scalar field of model results. "b" refers to observational field. "*" refers to normalization process. The centered statistics are the same as the uncentered statistics, except that the original field is replaced by the anomaly field

deviation (SD) values across all components (including precipitation and temperature here) of the anomaly field varied from 0.98 to 1.07 among the 19 different simulations. And it clearly indicated that which schemes overestimated or underestimated the amplitude error of anomaly field, as characterized by greater or smaller normalized cRMSLs, respectively. For example, some schemes (e.g., Kessler, Goddard_GCE) underestimated the overall cRMSL over the Haihe River Basin in summer. In contrast, other schemes (e.g., WSM6_Grau) overestimated the overall cRMSL, reaching 1.07. As for the second statistic, the cVSC (in terms of anomaly fields) reflects the pattern similarity between two vector fields. The CAM V5.1 2–moment 5–class Scheme (hereafter CAM5.1, Eaton 2011), capture the simulation of precipitation and temperature, in close agreement with the observations, and the cVSC both reach a highest value of 0.78.

All schemes show a high correlation with observations of surface temperature, but precipitation shows lower correlations with the corresponding in situ observations, with a range between 0.58 and 0.64. The cRMSVD measures the overall cRMSDs of temperature and precipitation and indicates the overall difference of anomaly field in terms of the simulation of multiple variables. The optimal microphysics scheme CAM5.1 has the minimum cRMSVD of 0.94, which reflects that the optimal one has the minimum overall error in terms of anomaly fields. There was a significant difference between the cRMSDs in terms of the two variables, surface temperature, and precipitation. The cRMSD of surface temperature ranges from 1.08 to 1.14, and the one as regards precipitation ranges from 0.77 to 0.91. Under these conditions, the overall differences of anomaly field were primarily from temperature and precipitation second. The MIEI values in Table 3 take both

Table 4 Three levels of statistics in MVIE in terms of 19 kinds of microphysics schemes; the lighter colors show which schemes performed better

METRICS	ME		VME	SD		cRMSL	CORR		cVSC	cRMSD		cRMSVD	MIEI
	T2	PR		T2	PR		T2	PR		T2	PR		
Kessler	-1.34	1.50	2.02	1.12	0.83	0.98	0.82	0.66	0.75	1.12	0.77	1.00	0.75
Lin	-1.19	2.43	2.70	1.09	1.03	1.06	0.86	0.60	0.73	1.09	0.91	1.06	0.78
WSM3_ice	-1.04	2.63	2.83	1.08	1.05	1.06	0.87	0.61	0.74	1.08	0.91	1.05	0.78
WSM5	-1.10	2.62	2.84	1.09	1.03	1.06	0.88	0.60	0.74	1.09	0.91	1.05	0.78
Ferrier	-1.28	2.16	2.51	1.10	0.98	1.04	0.88	0.60	0.75	1.10	0.88	1.02	0.76
WSM6_Grau	-1.09	2.61	2.83	1.10	1.03	1.07	0.88	0.60	0.75	1.10	0.90	1.05	0.78
Goddard_GCE	-2.40	0.57	2.46	1.11	0.84	0.99	0.63	0.63	0.63	1.11	0.80	1.21	0.75
Thompson	-1.05	2.71	2.90	1.09	1.01	1.06	0.88	0.61	0.75	1.09	0.89	1.04	0.77
Milbrandt-Yau	-1.15	2.61	2.86	1.11	1.01	1.06	0.87	0.62	0.75	1.11	0.87	1.03	0.76
Morrison	-1.17	2.27	2.55	1.08	0.98	1.03	0.87	0.62	0.75	1.08	0.86	1.02	0.76
CAM5.1	-0.96	1.21	1.54	1.14	0.87	1.01	0.89	0.66	0.78	1.14	0.78	0.94	0.71
SBU_YLIN	-1.03	2.80	2.98	1.12	0.98	1.06	0.89	0.60	0.75	1.12	0.88	1.02	0.77
WDM5	-1.53	2.20	2.68	1.11	0.99	1.05	0.78	0.62	0.70	1.11	0.87	1.12	0.76
WDM6	-1.53	2.13	2.62	1.12	0.98	1.05	0.76	0.62	0.69	1.12	0.86	1.14	0.77
NSSL17	-1.01	2.61	2.80	1.12	0.98	1.06	0.90	0.58	0.75	1.12	0.91	1.04	0.77
NSSL18	-0.99	2.68	2.86	1.13	0.99	1.06	0.90	0.61	0.76	1.13	0.87	1.00	0.76
NSSL19	-1.94	1.24	2.30	1.10	0.90	1.01	0.69	0.64	0.67	1.10	0.81	1.16	0.76
NSSL21	-1.91	1.31	2.32	1.11	0.90	1.01	0.71	0.64	0.68	1.11	0.81	1.15	0.77
Thompson_aer	-1.04	2.73	2.92	1.10	1.01	1.06	0.88	0.61	0.75	1.10	0.88	1.03	0.77

The MIEI column ranks the different models according to how they performed. The RMS (RMSL) values here are the ratio of modeled to observed RMS (RMSL) values. The manipulation of the metrics refers to Table 2.

the amplitudes and pattern similarities of various variables into account, and can therefore provide a comprehensive evaluation of model performance in terms of original field. In general, the MIEI values of all of the schemes varied from 0.71 to 0.78, which suggests that the performances of all model simulations were sensitive to the choice of microphysics scheme options. For example, the CAM5.1 scheme (option 11) show smaller MIEI values (smaller than 0.71) than those of the other schemes. Note that the default scheme (WSM6_Grau scheme, option 6) had a larger MIEI value (0.78) than that of the optimal scheme.

In comparison with the MIEI, the VFE diagram can provide statistics on model performance that are more comprehensive, and also clearly shows the differences between models and observations as well as the differences between the various models. A VFE diagram in terms of original field showing the overall RMSL, VSC, and RMSVD over the Haihe River Basin is shown in Fig. 3. It can readily be seen that the optimal scheme (CAM5.1 scheme, option 11) is a better performer than the default scheme (option 6, WSM6_Grau). The CAM5.1 scheme shows the maximum VSC (0.75), indicating that the scheme applied in the climate model can generally better reproduce the spatial pattern of the two variables relative to other models. Similarly, in comparison with the default scheme (option 6), the optimal scheme (option 11) shows a smaller RMSVD
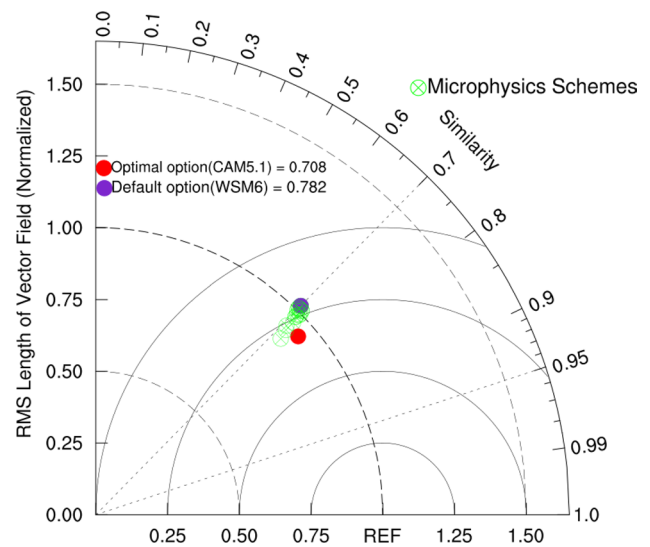

Fig. 3 Vector field evaluation (VFE) diagram for the 19 kinds of microphysics parameterizations over the Haihe River Basin. The purple symbol denotes the default scheme set and the red symbol denotes the optimal scheme set

(0.96) describing the overall difference between two vector fields (here referring to model field and observation field). However, in terms of the property of the RMSL

measuring the overall magnitude of multiple variables, the RMSL of the CAM 5.1 schemes is 0.94 smaller than 1.0, which indicates that the model simulation underestimate the magnitude of the vector field.

The CAM 5.1 denotes the version of 5.0 of the Community Atmosphere Model (CAM), which contains significant enhancements to the representation of atmospheric process. Especially, the revised cloud macrophysics scheme provides a more transparent treatment of cloud processes and imposes full consistency between cloud fraction and cloud condensate. A prognostic, two-moment formulation for cloud droplet and cloud ice, and liquid mass and number concentrations are included in Stratiform microphysical processes. And the scheme describes ice supersaturation and features activation of aerosols to form cloud drops and ice crystals (Eaton 2011, http://www.cgd.ucar.edu/amp/research/modeling.html).

3.2 pbl/surface layer scheme

The main objectives for developing PBL schemes are concluded as follows: (1) calculating the fluxes of momentum, heat, and water vapor within the atmospheric boundary layer. (2) predicting the atmospheric boundary layer depth, the amount of cloud, and consequently the vertical redistribution of heat and mass. It can be divided into two classes of PBL scheme (Xie et al 2012), the first one is turbulent kinetic energy prediction (MYJ, MYNN, Bougeault-Lacarrere, TEMF, QNSE, CAM UW), the second one is diagnostic non-local (YSU, GFS, MRF, ACM2) [available from http://www2.mmm.ucar.edu/wrf/users/

tutorial/201601/physics.pdf]. To determinate which PBL scheme performed best, Table 5 gives all of the statistical metrics values of various PBL schemes on the basis of the evaluation method. According to the monotonic property of MIEI values, it can readily be seen that the MRF scheme (option 99), followed by the ACM2-1 and ACM2-7 schemes, has a smaller MIEI value (0.70) than the other models. BouLac-1 and BouLac-2 show the worst performance with respect to the greatest MIEI values (above 0.74). Undoubtedly, the cVSC and cRMSVD values of the MRF schemes have been optimized to a greater or less extent than all of the other PBL schemes and the best microphysical schemes (CAM5.1). For example, the cVSC values of all PBL schemes range from 0.76 to 0.79, and the MRF scheme has a greatest value (0.79) among the others. The centered correlation coefficients (CORR) across the two individual variables (surface temperature and precipitation) show a better result. Specifically, the CORR values of surface temperature ranges from 0.89 to 0.91, which implies that the models reproduce the spatial pattern of temperature quite well. Compared with surface temperature, precipitation shows weaker centered correlation coefficients (CORR) ranging from 0.62 to 0.68. However, the MRF scheme selected for the optimal PBL scheme has a higher CORR value for precipitation (0.68). Similarly, all schemes show smaller cRMSL values, ranging from 0.99 to 1.04, than the microphysics schemes. The cRMSL values of MRF scheme approaching a value of 1.0, reveals the small amplitude error of the anomaly field. For mean field, the mean error (ME) of each individual

Table 5 Three levels of statistics in MVIE in terms of 17 kinds of PBL/Surface Layer schemes; the lighter colors show which schemes performed better

METRICS	ME		VME	SD		cRMSL	CORR		cVSC	cRMSD		cRMSVD	MIEI
	T2	PR		T2	PR		T2	PR		T2	PR		
YSU	-0.96	1.21	1.54	1.14	0.87	1.01	0.89	0.66	0.78	0.52	0.78	0.94	0.71
MYJ	-0.93	1.83	2.05	1.06	0.94	1.00	0.91	0.64	0.78	0.44	0.83	0.94	0.72
QNSE	-1.49	1.73	2.28	1.09	0.92	1.01	0.91	0.63	0.78	0.46	0.82	0.94	0.74
MYNN2-1	-0.91	1.43	1.70	1.11	0.95	1.03	0.90	0.65	0.78	0.49	0.82	0.95	0.73
MYNN2-2	-0.63	1.46	1.59	1.09	0.95	1.02	0.91	0.65	0.79	0.45	0.81	0.93	0.72
MYNN2-5	-0.92	1.50	1.76	1.12	0.95	1.04	0.89	0.65	0.78	0.51	0.81	0.96	0.73
MYNN3-1	-1.26	1.37	1.86	1.09	0.96	1.03	0.90	0.65	0.78	0.47	0.81	0.94	0.73
MYNN3-2	-1.07	1.32	1.70	1.07	0.95	1.01	0.92	0.66	0.79	0.43	0.81	0.92	0.72
MYNN3-5	-1.39	1.35	1.94	1.10	0.96	1.03	0.90	0.66	0.79	0.48	0.81	0.94	0.73
ACM2-1	-0.95	0.74	1.20	1.14	0.83	1.00	0.89	0.66	0.79	0.52	0.77	0.93	0.70
ACM2-7	-1.14	0.77	1.37	1.15	0.82	1.00	0.89	0.68	0.79	0.52	0.75	0.91	0.71
BouLac-1	-0.70	2.41	2.51	1.07	0.97	1.02	0.88	0.62	0.76	0.52	0.86	1.00	0.74
BouLac-2	-0.44	2.42	2.46	1.05	0.97	1.01	0.90	0.62	0.76	0.47	0.86	0.98	0.74
UW-1	-1.08	1.18	1.60	1.11	0.89	1.01	0.90	0.64	0.78	0.48	0.81	0.94	0.72
UW-2	-0.85	1.14	1.42	1.08	0.89	0.99	0.91	0.64	0.79	0.44	0.80	0.92	0.71
Shin-Hong	-1.03	1.40	1.74	1.14	0.89	1.02	0.89	0.66	0.78	0.51	0.79	0.94	0.71
MRF	-0.35	0.69	0.77	1.16	0.79	1.00	0.89	0.68	0.79	0.53	0.74	0.91	0.70

variable in all PBL schemes shows a decreasing tendency relative to all microphysics schemes above. Regarding to the vector mean error (VME), the optimal MRF scheme shows a smaller mean error (0.77) of simulating multiple variables. And the optimal one has minimum mean error of temperature (-0.35) and precipitation (0.69).

The application of the VFE diagram (Fig. 4) to the Haihe River Basin illustrates the overall correspondence between model results and observed behavior. For example, the optimal PBL scheme (MRF/MM5) shows a very high value of VSC (0.76), which indicates that the spatial pattern of all scalar fields (surface temperature and precipitation) is very close to that of the observational field. It should be noted that the optimal microphysics scheme (CAM5.1) and the default combination of parameterizations have weaker values of VSC of 0.75 and 0.69, respectively. The RMSLs of all schemes, composed of the RMSs of surface temperature and precipitation, are significantly different, and are also not well suited to performance evaluation here. The RMSL of the MRF scheme (0.93) is farther from the value of 1.0 than that of the default scheme (1.02), which indicates that the optimal one indeed underestimates the amplitude error of multiple variables. Similarly, the RMSVDs of the MRF scheme (0.95) are smaller than those of all other PBL schemes and the default scheme (1.11), which indicates that the MRF scheme performs better. The MIEI serves as a more concise evaluation index for generating the ranking of all model performances in simulating surface temperature and precipitation, and a smaller MIEI value shows a better model performance. Thus, the MRF scheme, which has the advantages of the aspects of three statistical quantities (i.e.,
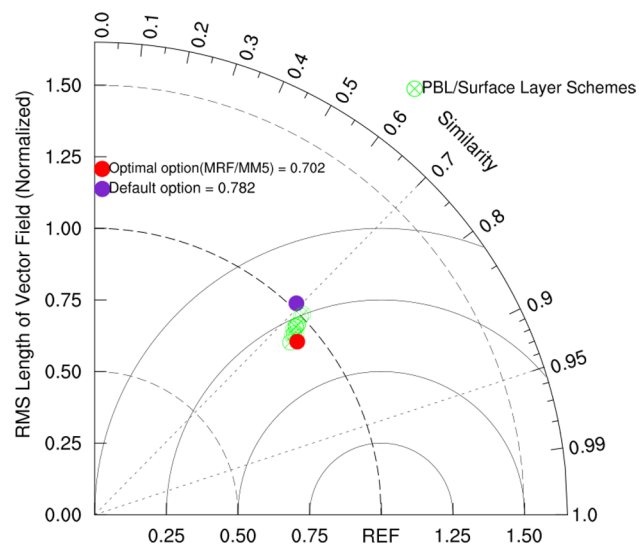
RMSL, VSC, RMSVD), has a smaller value of MIEI (0.70) than the default scheme (0.78) and the optimal microphysics scheme (0.71). This fact strongly suggests that the combination of the optimal microphysics and PBL schemes can enhance the predictive performance for precipitation and surface temperature.

The MRF scheme as the optimal scheme has been widely used for atmospheric numerical models because of its smaller computational demands and because it produces reasonable results. Also, the MRF scheme, on the basis of a nonlocal-K approach proposed by Troen and Mahrt (1986), substantially improves the precipitation forecast by enhancing the convective processes at the right place and by suppressing abnormal rainfall events (Hong and Pan 1996; Balzarini et al. 2014).

3.3 Cumulus scheme

After completing two sets of simulations, we conducted a third simulation group focusing on the cumulus scheme on the basis of the microphysics and PBL options set to CAM5.1 and MRF, respectively. Cumulus schemes are intended to represent vertical fluxes under circumstances of unsolved updrafts and downdrafts and compensating motion outside clouds. Additionally, cumulus parameterizations are theoretically valid only for coarser grid sizes (greater than 10 km), where the release of latent heat is necessary in the convective columns. According to the results given in Table 6, the BMJ scheme shows a smaller MIEI (0.68) relative to all other schemes, and the MIEI of all cumulus schemes ranged from 0.68 to 0.75. In contrast, the GFE scheme had a greater MIEI (0.75) characterized by a worse performance of the climate model. For mean field, the VME (vector mean error), showing mean errors in terms of multivariable integrated field, ranges from 0.94 to 1.93. Regarding to the mean error (ME) of the optimal scheme (BMJ), both precipitation and temperature are in a smaller value (-0.63 and -0.75). In terms of the mean error (ME) values of other options, either temperature or precipitation is in high values. For anomaly field, the centered root-mean-square vector difference (cRMSVD), providing a comprehensive evaluation of anomaly field, varied from 0.83 to 0.98. It can readily be seen that there are significant differences between the amplitudes of various variables, relative to the RMSD statistical quantities, among the various cumulus schemes. For example, the RMSD of temperature, ranging from 0.44 to 0.56, is much smaller than that of precipitation, which ranges from 0.68 to 0.85. The next statistical quantity to discuss is the centered vector similarity coefficient (cVSC). The spatial patterns of all simulation results are very close to the observational anomaly field, which corresponds to very high values of cVSC (0.77–0.81). The centered root-mean-square length (cRMSL) of all schemes ranges from
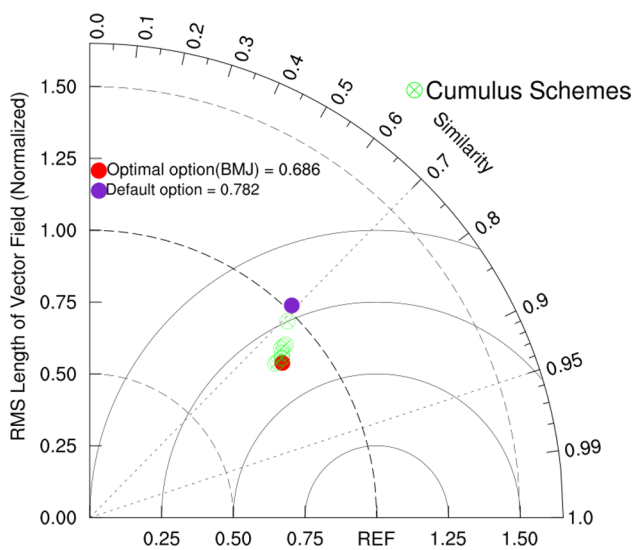

Fig. 4 VFE diagram for the 17 kinds of PBL/Surface Layer schemes over the Haihe River Basin. The purple symbol denotes the default scheme set and the red symbol denotes the optimal scheme set

Table 6 Three levels of statistics in MVIE in terms of 10 kinds of cumulus schemes; the lighter colors show which schemes performed better

METRICS	ME		VME	SD		cRMSL	CORR		cVSC	cRMSD		cRMSVD	MIEI
	T2	PR		T2	PR		T2	PR		T2	PR		
KF	-0.94	0.75	1.21	1.15	0.82	1.00	0.89	0.67	0.79	0.52	0.76	0.92	0.71
BMJ	-0.75	-0.63	0.98	1.19	0.74	0.99	0.89	0.73	0.81	0.54	0.68	0.87	0.68
GFE	-1.28	1.45	1.93	1.09	0.96	1.03	0.90	0.63	0.77	0.49	0.85	0.98	0.75
G3	-1.03	-0.33	1.08	1.13	0.81	0.98	0.89	0.70	0.80	0.51	0.72	0.88	0.71
Tiedtke	-0.74	-1.00	1.24	1.20	0.77	1.00	0.88	0.73	0.81	0.56	0.68	0.88	0.70
KF_PDF	-1.05	0.50	1.17	1.12	0.82	0.98	0.89	0.68	0.80	0.50	0.74	0.90	0.70
NSAS	-0.98	0.13	0.99	1.13	0.73	0.95	0.89	0.70	0.80	0.51	0.72	0.88	0.70
NewTiedtke	-0.90	-0.25	0.94	1.18	0.77	1.00	0.90	0.72	0.81	0.51	0.69	0.86	0.69
GD	-1.08	-0.26	1.11	1.09	0.78	0.95	0.89	0.72	0.81	0.49	0.70	0.85	0.69
oldKF	-1.48	-0.24	1.50	1.02	0.73	0.89	0.91	0.71	0.81	0.44	0.71	0.83	0.70

Fig. 5 VFE diagram for the 10 kinds of cumulus schemes over the Haihe River Basin. The purple symbol denotes the default scheme set and the red symbol denotes the optimal scheme set

0.89 to 01.03. And the cRMSL of the optimal one (0.99) is approaching 1.0, which reveals less amplitude error of multivariable integrated field. In general, the optimal BMJ scheme has smaller values of MIEI, cRMSVD, VME and greater values of cVSC, which suggests that the optimal one has improvements in model results.

The VFE diagram shown in Fig. 5 also provides some guidance for evaluating the original field's performance of various cumulus schemes as well as the default scheme. As shown in Fig. 5, all cumulus schemes have improved model performance compared to the default combination of schemes. For example, the optimal BMJ scheme has a distinct advantage in promoting model simulation accuracy over the default scheme. The MIEI value of the BMJ scheme (0.68) is much smaller than that of the WSM6 scheme

(0.78). Specifically, the BMJ scheme shows a smaller root-mean-square vector difference (RMSVD,0.88) and a higher vector similarity coefficient (VSC, 0.78) relative to the RMSVD (1.11) and VSC (0.69) of the default one. However, it should be noted that, as regards the RMSL, the default scheme (1.02) approaches 1.0 more than the BMJ scheme (0.86). The VFE diagram can clearly illustrate that the overall RMSVD of the BMJ scheme is more associated with the improvement in pattern similarity than with the systematic difference in vector length.

With respect to the optimal cumulus scheme, the Betts–Miller–Janjic (BMJ) scheme derived from the Betts–Miller convective adjustment scheme has been optimized in concept to the parameter values recommended by Janjic (1994, 2000). Furthermore, cumulus schemes fall into two main classes: (i) the adjustment type containing only one BMJ scheme and (ii) the mass-flux type containing all other schemes in WRF.

3.4 Longwave/shortwave radiation scheme

The radiation scheme is responsible for atmospheric heating resulting from radiative flux divergence and surface downward longwave and shortwave radiation for the ground heat budget. Longwave radiation schemes compute mainly clear-sky and cloud upward and downward radiation fluxes. Shortwave radiation schemes compute clear-sky and cloudy solar fluxes, and most schemes consider downward and upward fluxes in particular (Dudhia scheme only has downward flux). The longwave/shortwave radiation schemes simulation group is assessed next (see Table 7). Because some of the radiation schemes were unable to start WRF successfully, we selected nine kinds of radiation schemes capable of outputting model results. There is a subtle difference among the MIEI values of all radiation schemes, which range from 0.67 to 0.68. Thus, we select the optimal option by the performance of mean fields and anomaly fields. For mean fields,

Table 7 Three levels of statistics in MVIE in terms of 9 kinds of Longwave/Shortwave schemes; the lighter colors show which schemes performed better

METRICS	ME		VME	SD		cRMSL	CORR		cVSC	cRMSD		cRMSVD	MIEI
	T2	PR		T2	PR		T2	PR		T2	PR		
CAM_SW03	-0.03	-0.84	0.84	1.22	0.72	1.00	0.90	0.75	0.82	0.55	0.66	0.86	0.68
CAM_SW04	-0.26	-0.92	0.95	1.23	0.72	1.01	0.90	0.76	0.82	0.55	0.65	0.85	0.68
CAM_SW24	-0.27	-0.90	0.94	1.23	0.73	1.01	0.90	0.76	0.82	0.55	0.65	0.85	0.68
RRTMG_SW03	0.94	-0.34	1.00	1.31	0.75	1.07	0.91	0.74	0.82	0.58	0.67	0.88	0.67
RRTMG_SW04	0.75	-0.45	0.88	1.33	0.75	1.08	0.91	0.75	0.82	0.59	0.66	0.88	0.68
RRTMG_SW24	0.73	-0.51	0.90	1.34	0.74	1.08	0.91	0.76	0.82	0.59	0.65	0.88	0.67
Fast_RRTMG_SW03	0.94	-0.40	1.02	1.31	0.74	1.06	0.91	0.74	0.82	0.58	0.67	0.88	0.67
Fast_RRTMG_SW04	0.76	-0.43	0.87	1.33	0.75	1.08	0.91	0.74	0.82	0.59	0.67	0.89	0.68
Fast_RRTMG_SW24	0.80	-0.30	0.86	1.33	0.76	1.08	0.90	0.74	0.81	0.62	0.68	0.91	0.68

the CAM scheme, as with the default radiation scheme, has a minimum vector mean error (VME) value of 0.84, indicating its less mean difference. For anomaly fields, most of the simulations in this group have greater centered vector similarity coefficient (cVSC) and smaller centered root-mean-square difference (cRMSVD) than the earlier simulation groups, which demonstrates that the process of selecting optimal physics schemes enhances the quality of the simulations. The centered root-mean-square length (cRMSL) of all radiation schemes shows a steady trend exceeding 1.0. Specifically, the cRMSL of the optimal scheme (CAM) is 1.00, indicating its good performance in amplitude error of anomaly field. And the optimal one has a larger cVSC (0.82) and smaller cRMSVD (0.86) among other option. The CAM longwave radiation schemes documented fully by Collins et al. (2006) are linked to resolved clouds and cloud fractions and are able to handle several trace gases. Additionally, the CAM shortwave radiation scheme using cloud fractions and overlapping assumptions for unsaturated regions has a monthly zonal ozone climatology and can handle the optical properties of several aerosol types and trace gases. All in all, the CAM radiation scheme is especially suited to regional climate simulations by having an ozone distribution that varies during the simulation according to monthly zonal-mean climatological data.

3.5 Land surface parameterization scheme

The representation of LSM schemes results from the increased interest in land-use activities and the need to simulate regional climates more precisely. The presence of vegetation, orographic features, and surface heterogeneity, all suitably represented in LSM, all influence surface albedo (radiative transfer), surface roughness (momentum transfer), and surface hydrology (sensible and latent heat transfer; runoff). The last simulation group focuses on the evaluation of the land surface schemes given in Table 8. In most runs, there is little difference in the MIEIs of various LSM (Land Surface Model) schemes, and the optimal Noah scheme has a relatively small MIEI value (0.68). Particularly for mean field, the mean error of individual variables in this group is superior to all other groups. For example, the mean error (ME) of temperature in Noah simulation is much less than other option and approximately approaching 0. The VFE diagram in Fig. 6 shows that, although not all radiation and LSM schemes show significant differences in modeling results from the optimal choice, the selection of the best combination of schemes optimizes model performance more than the default scheme.

The optimal LSM scheme selected above was the Noah LSM scheme, which is the successor to the OSU LSM scheme described by Chen and Dudhia (2001), and has a four-layer soil temperature and moisture model with canopy moisture and snow cover prediction. Additionally, the Noah LSM not only predicts soil ice and fractional snow cover effects, but also has an improved urban treatment, and considers surface emissivity properties (Koren et al. 1999; Ek et al. 2003).

Table 8 Three levels of statistics in MVIE in terms of 3 kinds of Land Surface schemes; the lighter colors show which schemes performed better

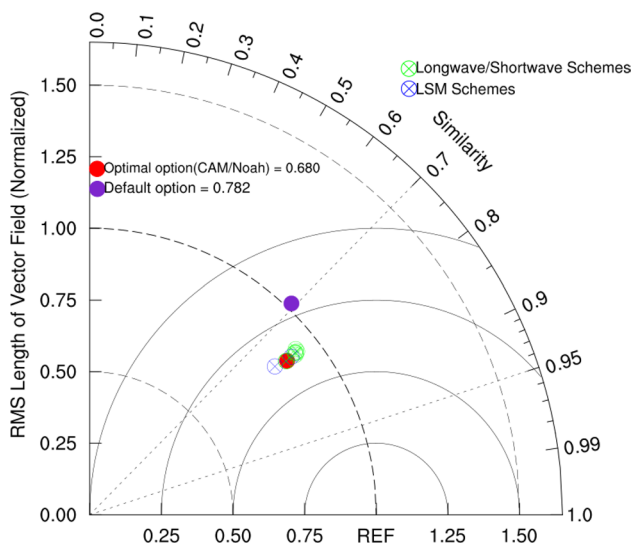| METRICS	ME		VME	SD		cRMSL	CORR		cVSC	cRMSD		cRMSVD	MIEI
	T2	PR		T2	PR		T2	PR		T2	PR		
TD5	-1.89	-0.96	2.12	1.03	0.65	0.88	0.92	0.77	0.84	0.40	0.65	0.76	0.70
Noah	-0.03	-0.84	0.84	1.22	0.66	1.00	0.90	0.75	0.82	0.55	0.66	0.86	0.68
NoahMP	0.28	-0.51	0.59	1.13	0.65	0.95	0.86	0.76	0.81	0.57	0.65	0.86	0.68

Fig. 6 VFE diagram for the 9 kinds of radiation schemes and 3 types of LSM schemes over the Haihe River Basin. The purple symbol denotes the default scheme set and the red symbol denotes the optimal scheme set. Because the optimal LSM scheme (Noah) is consistent with the default scheme, we discuss the two types of physics options here together in the VFE diagram

3.6 Identifying the optimal parameterization scheme combination

The results of the test of the sensitivity of precipitation and surface temperature simulated by the WRF model to different parameterization scheme combinations have been determined, and the best performing scheme set consists of the CAM5.1 microphysics scheme, the MRF PBL scheme, the BMJ cumulus scheme, the CAM for shortwave/longwave radiation scheme, and the Noah land surface scheme. In order to identify the best parameterization setup with the overall best performance in terms of simulating temperature and precipitation, several approaches were utilized in this study. One crucial issue is selection of the VFE diagram providing various statistics from all groups of WRF simulations (Fig. 7). Some manipulation of the statistical metrics is necessary to rank the model performance of all simulation members in the study; thus, the RMSL, VSC, and RMSVD statistical quantities have been integrated to access all 58 sets of simulations shown in Fig. 7. As shown in the figure, different simulation groups are clearly separated from each other. The performance of members in different groups in the VFE diagram has been discussed before, so here we focus on analyzing whether the evaluation method enhances the performance of each option on basis of spatial mean values and temporal mean values in all scalar fields (i.e. surface temperature and precipitation). The earlier work involved five stages of evaluation work, and it can be seen from Fig. 7 that the RMSVD and VSC statistical quantities
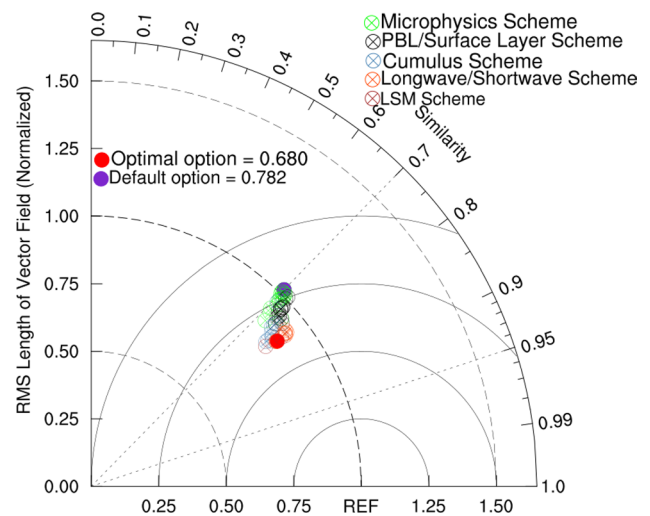

Fig. 7 VFE diagram for all 58 kinds of simulation above the Haihe River Basin. Different colors represent the different physics options, and the red (purple) symbol is the best proper (default) scheme set over the basin

have been much enhanced in the process of selecting five kinds of physics parameterizations. The difference between the default model setup and the optimal set of schemes can also be seen from Fig. 7. Specifically, compared with the default scheme, the optimal scheme has a higher VSC (0.79; that of the default is 0.69) and a smaller RMSVD (0.87; that of the default is 1.11). Some analysis of the features of the RMS values of various scalar fields is necessary prior to ranking the RMSL values of the vector field here. Thus, relative to surface temperature, the normalized RMS of the optimal scheme (1.00) approaches 1.0 more closely than that of the default scheme (0.95). In contrast, the optimal scheme has underestimated the RMS of precipitation (0.72), which is different than the overestimate of the default scheme (1.07). All in all, the optimal combination of physics parameterizations is beneficial for improving the reproduction of surface temperature and precipitation in the WRF model over the Haihe River Basin.

Next, in order to identify the impact of the relative differences between different parameterization schemes on the bias between model outputs and observations, we calculated the spatial mean bias of the best schemes of all five options examined above to present the spatial bias characteristics of precipitation and surface temperature (Fig. 8) over the Haihe River Basin. The statistic for precipitation and temperature is calculated as the spatial average of the differences in the three months' mean fields. In addition, in order to identify whether the optimal physics parameterizations set is optimized, we also calculated the statistics of the default scheme to demonstrate the improvement in the other model simulations. As shown in Fig. 8, we can conclude that the evaluation approach described above greatly

Fig. 8 Spatial characteristics of the mean model deviations from the observational data for precipitation (left) and surface temperature (right)
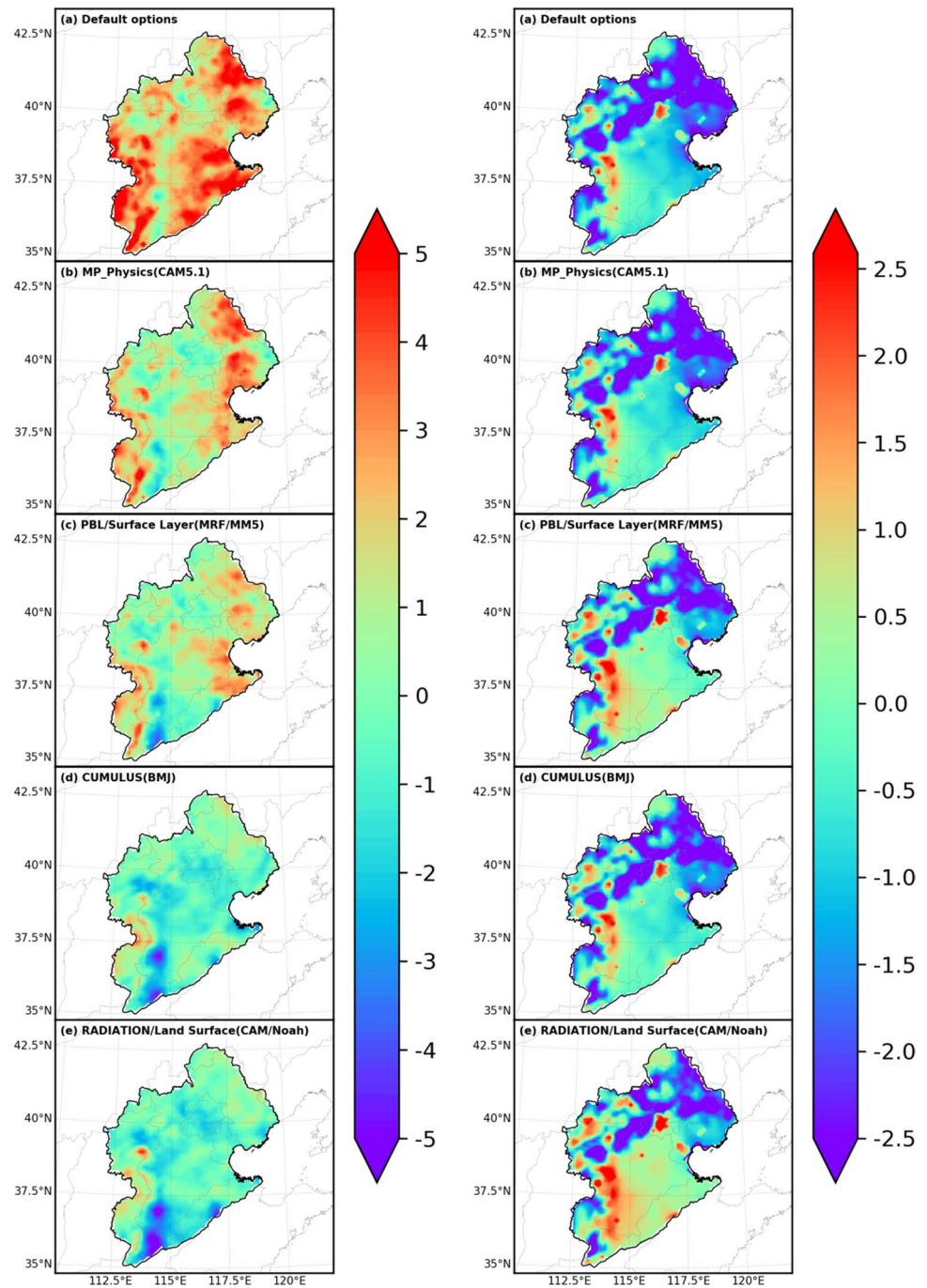
enhances the ability of the model to simulate precipitation. The bias values for precipitation using the default option shown in the WRF model greatly overestimate rainfall. This overestimation of rainfall in summer is located over most regions in the Haihe River Basin, and overestimations of up to 4–5 mm/day are located in the northeastern part, the southeast plain area, and the western hilly area of the basin. Looking at the optimal microphysics option, the simulation in some areas also produces overestimations of more than 3 mm/day. In the optimal PBL schemes,

there is no significant difference between the model and the observational data, and the positive and negative bias values are all distributed between − 3 and + 3 mm/day. The optimal cumulus scheme option has a minimum magnitude (between − 1 and + 1 mm/day) of bias values spread over the Haihe River Basin. With respect to the last simulation (referring to the optimal radiation/LSM scheme, the optimal LSM scheme remains as the default scheme, and so the radiation scheme coincides with the LSM scheme), it can readily be seen that the remarkable bias values of

almost all models have mostly disappeared when using the optimal combination of schemes. The summer precipitation predictions remain poor over a very small area located in the southern hilly area, with a relatively small underestimation, while the remaining parts maintain a high correspondence with the optimal cumulus schemes characterized by no severe failures.

For the surface temperature variable (Fig. 8), the majority of the simulations show a magnitude of temperature biases with a range between − 2.5 °C and + 2.5 °C. The underestimate with a magnitude of − 2.5 °C is located in most parts of the hilly areas (the northern and western parts), and the model deviation in the central plateau area show a small overestimate (above 0.5 °C) or a small underestimate (above − 1 °C). There is a similar pattern to the temperature simulated by each physics combination (shown in Fig. 8). However, the trend of particular model bias in some parts of the Haihe River Basin shows the improvement in the temperature forecast in the final simulation. For example, the central area shifts from a small underestimate (above − 1 °C) to a small overestimate (above 0.5 °C). Furthermore, the peak values of the overestimates in several parts of the hilly areas have disappeared. Although the overestimates in some parts have intensified in the final simulation, we can see that the simulation of surface temperature has improved to some extent as a result of using the optimal physics schemes combination.

We analyzed the relative differences in the impacts of all combinations on the temporal characteristics of surface temperature (Fig. 9) and precipitation (Fig. 10) by calculating the spatial averages over the Haihe River Basin at the daily timescale. As shown in Fig. 9, the first four plots show the daily spatial averages of all simulations during the study period. Specifically, all of the simulated temperatures show underestimates compared with the observed temperatures. Note that there is a steady trend in simulating the temperature by selecting the optimal options above, and these simulated temperatures gradually approach the observed temperature data, which indicates that the evaluation method we have chosen is capable of selecting the best scheme set to further enhance the model's prediction accuracy.

Figure 10 shows plots of regional averages for the amount of simulated rainfall falling on each day in summer. By examining each physics parameterization in Fig. 10, our study focuses on discussing the optimal scheme for each option examined above. Although there is no particular amplitude error among the various scheme sets, it could be seen that the optimal setup (red line) is in more close proximity to the observational one (yellow line) than other setups in Fig. 10. The amount of rainfall distributed on the daily timescale, in Haihe river basin located in semi-arid region, is small and may account for the unremarkable amplitude errors between the model and the observational data. All in all, the model can simulate the overall temporal distribution of observational precipitation to some degree.

Fig. 9 Temporal characteristics of the spatial averages of surface temperature for all simulations over the Haihe River Basin during the summer of 2016. The gray lines denote the improper schemes, and the other colored lines denote the optimal schemes for each physics option
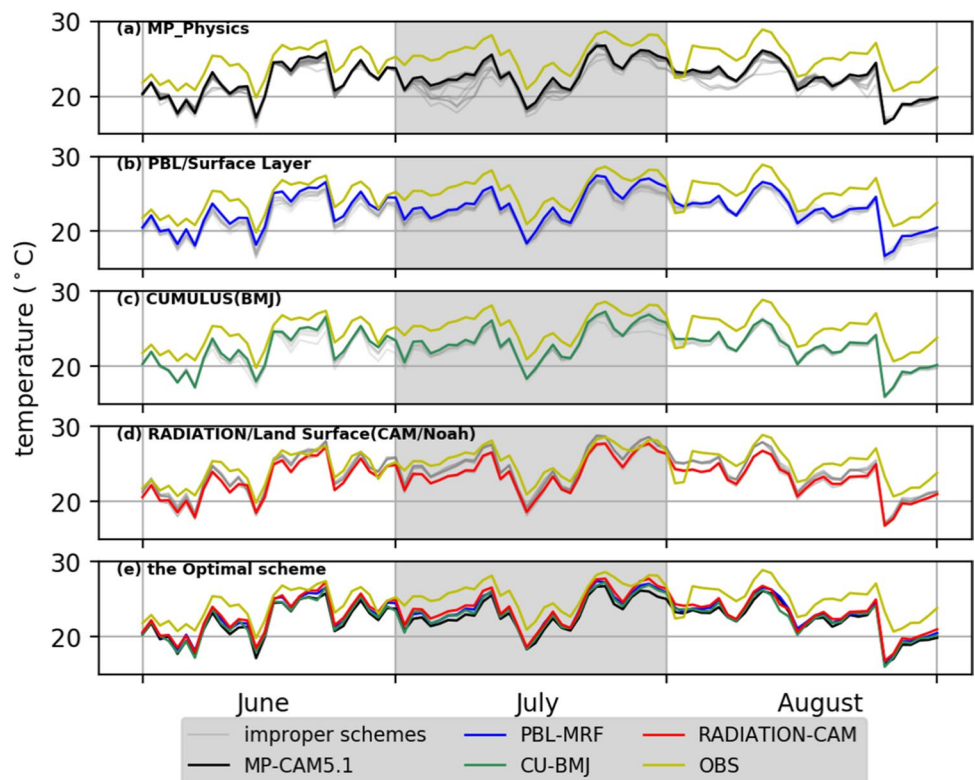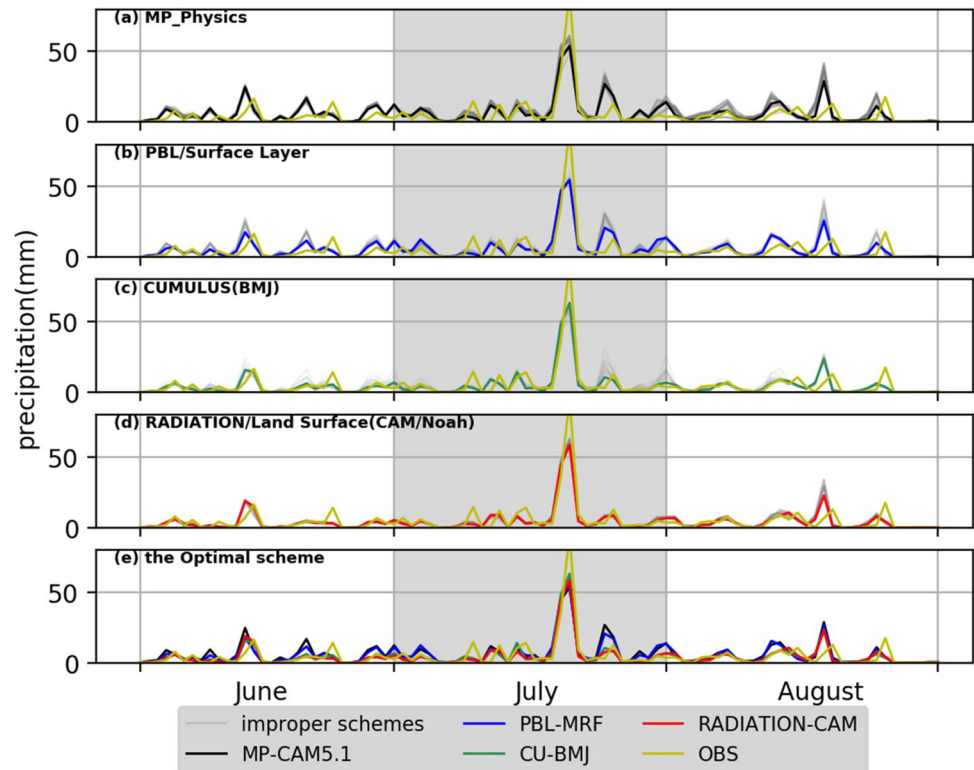
Fig. 10 Temporal characteristics of the spatial averages of precipitation for all simulations over the Haihe River Basin during the summer of 2016. The gray lines denote the improper schemes, and the other colored lines denote the optimal schemes for each physics option

4 Conclusions

According to previous studies, the simulation of rainfall and temperature is fundamental to climate research using the Advanced Research Weather Research and Forecasting (WRF) model, and particularly because the WRF model performs poorly when simulating rainfall. The selection of the optimal combination of schemes from a wide range of physical parameterization sets is beneficial for improving the performance of the model, even at the risk of high computational cost. Given the two reasons above, in this study, various physics combinations have been used to simulate the surface temperature and precipitation during a summer season in the Haihe River Basin for the purpose of optimizing the performance of the WRF model. In contrast with previous studies, we chose a multivariable integrated evaluation (MVIE) method, which groups various scalar fields (here referring to temperature and precipitation) into a vector field in order to thoroughly identify the most appropriate combination of schemes for the WRF model for the Haihe River Basin. Furthermore, the 58 members of the physics combinations that were chosen, which were not determined by permutation and combination of all parameterizations from the five physics options because of the computational cost, but by utilizing a stepwise refinement method based on an analysis of the interactions between the WRF parameterization schemes. In brief, this study has applied a more comprehensive evaluation methodology to assess more efficiently

the existing options that the WRF model offers (here referring to the five kinds of options, MP, PBL, CU, RA, and LSM) with as small a computational cost as possible.

In the evaluation process, various metrics were calculated and integrated to produce a preferred combination of physics parameterizations, which contains CAM5.1 MP, MRF PBL, BMJ cumulus, CAM radiation, and Noah LSM schemes, in comparison with the default scheme set (referring to WSM6 MP, YSU PBL, KF cumulus, CAM radiation, and Noah LSM schemes). Analysis of the multivariable integrated field revealed that the optimal scheme for each option examined above has a smaller MIEI value, which is much attributed to the improvement of centered vector similarity coefficient (cVSC) and smaller mean error. Obviously, the optimal setup has a higher cVSC value than other setups, which shows a good performance in pattern similarity of anomaly field. With respect to mean field, the optimal one for each option, in terms of both precipitation and temperature, showed a decreasing trend in mean error. Especially for surface temperature, the optimal scheme for each option reduced mean errors, gradually approaching a value of 0.

For this domain, we recommend the appropriate scheme set that contains the CAM5.1 MP, MRF PBL, BMJ cumulus, CAM radiation, and Noah LSM. To identify the advantages of the set of schemes selected by the MVIE approach, the VFE diagram with all sets of schemes clearly illustrates that the best set of schemes improved the pattern similarity and the RMS vector difference compared with the default

scheme set. The spatial mean bias plots of temperature and precipitation also demonstrate that the best scheme shows much smaller deviations from the observed values than the default scheme, and, in particular, the spatial improvement in the prediction of precipitation is obvious. The temporal plots of the spatial averages of temperature and precipitation show that the optimal scheme for each option has gradually improved the simulation of the spatial averages with the selection of successive schemes, and surface temperature matches the observational data better than precipitation does. Perhaps a more robust conclusion drawn from the overall analysis results is that the model output is more sensitive to the choice of microphysics schemes and PBL schemes than the cumulus scheme, but the radiation and LSM schemes have no significant impact on the model results.

It should be acknowledged that there will be uncertainties in aspects of observation data. The in-situ observational data will not only contains errors from instruments and measuring practices, but also should take a longer time series into consideration. In further work, greater advantage should be taken of more sets of observational data, such as Climatic Research Unit (CRU) gridded data, Global Historical Climatology Network (GHCN) temperature data, and Global Precipitation Climatology Center (GPCC) precipitation data. Also, the results for the optimal scheme set in this study can increase confidence in regional modeling over the Haihe River Basin for subsequent studies. It is worth pointing out that MVIE needs to be further improved to continue to be used in higher resolution convection-permitting models (Liu et al. 2017; Li et al. 2019). Unsurprisingly, the model evaluators can apply the assessment method used in this study to the domains of interest involved with other variables or extreme events.

References

Argüeso D, Hidalgo Muñoz JM, Gámiz Fortis SR et al (2011) Evaluation of WRF parameterizations for climate studies over southern Spain using a multistep regionalization. J Clim 24(21):5633–5651

Balzarini A, Angelini F, Ferrero L et al (2014) Sensitivity analysis of PBL schemes by comparing WRF model and experimental data. Geosci Model Develop Discuss 7:6133–6171

Bastidas LA, Hogue TS, Sorooshian S et al (2006) Parameter sensitivity analysis for different complexity land surface models using multicriteria methods. J Geophys Res 111(D20):1–19

Borge R, Alexandrov V, del Vas JJ, Lumbreras J, Encarnacioń R (2008) A comprehensive sensitivity analysis of the WRF model for air quality applications over the IberianPeninsula. Atmos. Environ. 42:8560–8574

Budakoti S, Singh C, Pal PK (2019) Assessment of various cumulus parameterization schemes for the simulation of very heavy rainfall event based on optimal ensemble approach. Atmospheric Res 218:195–206

Bukovsky MS, Karoly DJ (2009) Precipitation simulations using WRF as a nested regional climate model. J Appl Meteorol Clim 48:2152–2159

Chen F, Dudhia J (2001) Coupling an advanced land surface hydrology model with the Penn State-NCAR MM5 modeling system. Part I. Model implementation and sensitivity. Mon Weather Rev 129:569–585

Chen F, Liu C, Dudhia J et al (2014) A sensitivity study of high-resolution regional climate simulations to three land surface models over the western United States. J Geophys Res Atmospheres 119:7271–7291

Christensen O, Gaertner M, Prego J et al (2001) Internal variability of regional climate models. Clim Dyn 17:875–887

Cohen AE, Cavallo SM, Coniglio MC et al (2015) A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather environments. Weather Forecast 30:150224120634008

Collins WD, Rasch PJ, Boville BA et al (2006) The formulation and atmospheric simulation of the Community Atmosphere Model version 3 (CAM3). J Clim 19:2144–2161

Crétat J, Pohl B, Richard Y et al (2012) Uncertainties in simulating regional climate of Southern Africa: sensitivity to physical parameterizations using WRF. Clim Dynam 38:613–634

Denis B, Laprise R, Caya D et al (2002) Downscaling ability of one-way nested regional climate models: the big-brother experiment. Clim Dynam 18(627):646

Dickinson RE, Errico RM, Giorgi F et al (1989) A regional climate model for the western United States. Clim Change 15:383–422

Ding Y, Wang Z, Sun Y (2008) Inter-decadal variation of the summer precipitation in East China and its association with decreasing Asian summer monsoon. Part I: observed evidences. Int J Climatol 228:1139–1161

Ding Y, Wang Z, Sun Y (2010) Inter-decadal variation of the summer precipitation in East China and its association with decreasing Asian summer monsoon. Part I: observed evidences. Int J Climtol 28(9):1139–1161

Duan Y, Ma Z, Yang Q (2017) Characteristics of consecutive dry days variations in China. Theoret Appl Climatol 130(1–2):711–711

Eaton B (2011) User's guide to the community atmosphere model CAM-5.1. NCAR. http://www.cesm.ucar.edu/models/cesm1.0/cam.

Ek MB, Mitchell KE, Lin Y, Rogers E, Grunmann P, Koren V, Gayno G, Tarpley JD (2003) Implementation of Noah land surface model advancements in the National Centers for Environmental Prediction operational mesoscale Eta model. J Geophys Res 108:8851

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

Fernández J, Montávez JP, Sáenz J et al (2007) Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: annual cycle. J Geophys Res Atmos 112(D4):D04101

Ferreira JA, Carvalho AC, Carvalheiro L et al (2014) On the influence of physical parameterisations and domains configuration in the simulation of an extreme precipitation event. Dyn Atmos Oceans 68:35–55

Gallus JS, Bresch J (2006) Comparison of impacts of WRF dynamic core, physics package, and initial conditions on warm season rainfall forecasts. Mon Weather Rev 134:2632

Giannaros C, Melas D, Giannaros TM (2019) On the short-term simulation of heat waves in the Southeast Mediterranean: sensitivity of the WRF model to various physics schemes. Atmos Res 218:99–116

Gillett NP, Thompson DWJ (2003) Simulation of recent Southern Hemisphere climate change. Science 302:273–275

Giorgi F (2006) Regional climate modeling: status and perspectives. J Physique IV 139:101–118

Giorgi F, Bi X (2000) A study of internal variability of a regional climate model. J Geophys Res 105:29503–29516

Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res 113:D06104

Hasan MA, Islam AKMS (2018) Evaluation of microphysics and cumulus schemes of WRF for forecasting of heavy monsoon rainfall over the southeastern hilly region of Bangladesh. Pure Applied Geophys 175:4537–4566

He J, Yang XH, Li JQ et al (2015) Spatiotemporal variation of meteorological droughts based on the daily comprehensive drought index in the Haihe River basin, China. Nat Hazards 75:199–217

Hong SY, Pan HL (1996) Nonlocal boundary layer vertical diffusion in a medium-range forecast model. Mon Weather Rev 124:2322–2339

Hu XM, Nielsen-Gammon JW, Zhang F (2010) Evaluation of three planetary boundary layer schemes in the WRF model. J Appl Meteorol Climatol 49:1831–1844

Huang F, Xu Z, Guo W (2018) Evaluating vector winds in the Asian-Australian monsoon region simulated by 37 CMIP5 models. Clim Dyn 53(1–2):491–507

Huang F, Xu Z, Guo W (2020) The linkage between CMIP5 climate models' abilities to simulate precipitation and vector winds. Clim Dyn 54(11–12):4953–4970

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

Janjic ZI (2000) Comments on "Development and Evaluation of a Convection Scheme for Use in Climate Models." J Atmos Sci 57:3686

Jerez S, Montavez JP, Jimenez-Guerrero P et al (2013) A multi-physics ensemble of present-day climate regional simulations over the Iberian Peninsula. Clim Dyn 40:3023–3046

Koren V, Schaake JC, Mitchell KE, Duan QY, Chen F, Baker JM (1999) A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. J Geophys Res 104:19569–19585

Landman WA, Seth A, Camargo SJ (2005) The effect of regional climate model domain choice on the simulation of tropical cyclone-like vortices in the southwestern Indian Ocean. J Climate 18:1263–1274

Li W, Guo W, Xue Y et al (2016) Sensitivity of a regional climate model to land surface parameterization schemes for East Asian summer monsoon simulation. Clim Dyn 47:2293–2308

Li Y, Li Z, Zhang Z et al (2019) High-resolution regional climate modeling and projection over western canada using a weather research forecasting model with a pseudo-global warming approach. Hydrol Earth Syst Sci 23(11):4635–4659

Li MX, Ma Z (2015) Soil moisture drought detection and multi-temporal variability across China. Sci China-Earth Sci 58(10):1798–1813

Liu CH, Ikeda K, Rasmussen R et al (2017) Continental-scale convection-permitting modeling of the current and future climate of North America. Clim Dyn 49(1–2):71–95

Ma ZG, Fu C (2006) Some evidences of drying trend over North China from 1951 to 2004. Chin Sci Bull 51(23):2913–2925

Ma ZG, Fu C (2007) Evidences of drying trend in the global during the later half of 20th century and their relationship with large-scale climate background. Sci China Ser D-Earth Sci 50(5):776–788

Mooney PA, Mulligan FJ, Fealy R (2013) Evaluation of the sensitivity of the Weather Research and Forecasting Model to parameterization schemes for regional climates of Europe over the period 1990–95. J Clim 26:1002–1017

Osborn TJ (2004) Simulating the winter North Atlantic Oscillation: the roles of internal variability and greenhouse gas forcing. Clim Dyn 22:605–623

Perkins SE, Pitman AJ, Holbrook NJ et al (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20:4356

Pieri AB, Von Hardenberg J, Parodi A et al (2015) Sensitivity of precipitation statistics to resolution, microphysics, and convective parameterization: a case study with the high-resolution WRF climate model over Europe. J Hydrometeorol 16:1857–1872

Quan J, Di Z, Duan Q et al (2016) An evaluation of parametric sensitivities of different meteorological variables simulated by the WRF model. Quart J Royal Meteorolog Soc 142:2925–2934

Ratnam JV, Kumar KK (2005) Sensitivity of the simulated monsoons of 1987 and 1988 to convective parameterization schemes in MM5. J Clim 18:2724–2743

Schaefer JT (1990) The critical success index as an indicator of warning skill. Wea Forecast 5:570–575

Seth A, Giorgi F (1998) The effects of domain choice on summer precipitation simulation and sensitivity in a regional climate model. J Clim 11:2698–2712

Stergiou I, Tagaris E, Sotiropoulou RE (2017) Sensitivity Assessment of WRF Parameterizations over Europe. Proceedings 1(5):119.

Skamarock WC, et al. (2008) A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN–4751STR. pp 113.

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res: Atmos 106:7183–7192

Tegoulias I, Kartsios S, Pytharoulis I et al (2017) The influence of WRF parameterisation schemes on high resolution simulations over Greece Perspectives on Atmospheric Sciences. Springer, Cham, pp 3–8

Tian J, Liu J, Wang J et al (2017) A spatio-temporal evaluation of the WRF physical parameterisations for numerical rainfall simulation in semi-humid and semi-arid catchments of Northern China. Atmos Res 191:141–155

Troen IB, Mahrt L (1986) A simple model of the atmospheric boundary layer; sensitivity to surface evaporation. Boundary-Layer Meteorol 37:129–148

Von Storch H, Langenberg H, Feser F (2000) A spectral nudging technique for dynamical downscaling purposes. Mon Wea Rev 128:3664–3673

Wang H (2001) The weakening of the Asian monsoon circulation after the end of 1970's. Adv Atmos Sci 18(3):376–386

Wang D, Yu X, Jia G et al (2019) Sensitivity analysis of runoff to climate variability and land-use changes in the Haihe Basin mountainous area of north China. Agric Ecosys Environ 269:193–203

Wang Y, Leung LR, McGregor JL et al (2004) Regional climate modeling: progress, challenges, and prospects. J Meteor Soc Japan Ser II 82:1599–1628

Willmott CJ, Ackleson SG, Davis RE et al (1985) Statistics for the evaluation and comparison of models. J Geophys Res Oceans 90:8995–9005

Xie B, Fung JCH, Chan A et al (2012) Evaluation of nonlocal and local planetary boundary layer schemes in the WRF model. J Geophys Res Atmos. https://doi.org/10.1029/2011JD017080

Xu Z, Hou Z, Han Y et al (2016) A diagram for evaluating multiple aspects of model performance in simulating vector fields. Geosci Model Devel 9:4365–4380

Xu Z, Han Y, Fu C (2017) Multivariable integrated evaluation of model performance with the vector field evaluation diagram. Geosci Model Devel Discuss 10:1–31

Xu Z, Han Y (2020) Short communication comments on 'DISO: A rethink of Taylor diagram.' Int J Climatol 40(4):2506–2510

Yáñez-Morroni G, Gironás J, Caneo M et al (2018) Using the Weather Research and Forecasting (WRF) model for precipitation forecasting in an Andean region with complex topography. Atmosphere 9:304