https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

An objective dynamic multivariable weighting method for reducing uncertainty in WRF parameterization scheme selection

Tianyu Gou[1,2,3], Yaoyang Deng[1,2,3], Jun Niu[1,2,3], Shaozhong Kang[1,2,3]

[1]State Key Laboratory of Efficient Utilization of Agricultural Water Resources, Beijing 100083, China
[2]National Field Scientific Observation and Research Station on Efficient Water Use of Oasis Agriculture in Wuwei of Gansu Province, Wuwei 733009, China
[3]Center for Agricultural Water Research in China, China Agricultural University, Beijing 100083, China

Correspondence to: Shaozhong Kang (kangsz@cau.edu.cn)

Abstract. The selection of optimal physical parameterization schemes is a major source of uncertainty in WRF model simulations. A comprehensive evaluation of model performance requires simultaneous consideration of multiple output variables. However, existing multivariate approaches often rely on subjective or overly simplistic equal-weighting strategies and lack an objective mechanism to quantify variable importance. Such limitations can obscure significant biases in poorly simulated variables. To overcome this issue, this study proposes an objective dynamic weighting method for multivariate evaluation. The method employs a two-layer weighting framework based on two statistical metrics: the mean relative error, which measures the simulation accuracy of a variable, and the coefficient of variation of the absolute error, which reflects the sensitivity of a variable to the physical process under evaluation. The approach is applied and validated in assessing WRF parameterization schemes across two climatically distinct environments: the arid Northwest and the humid coastal Southeast of China. The results show that the method assigns greater weights to poorly simulated variables, such as precipitation and wind speed, thereby enabling the identification of more physically plausible and robust parameterization scheme combinations. Compared with the equal-weighting method, the scheme combinations obtained using this approach produce a lower Multivariate Integrated Evaluation Index (MIEI), a higher correlation coefficient (R), a lower Root Mean Square Error (RMSE), and exhibit superior performance in independent extreme-year validations. By dynamically incorporating both simulation performance and sensitivity specific to each variable, the method offers a more rigorous and objective framework for model evaluation and uncertainty reduction.

Keywords: Model evaluation; WRF model; Parameterization scheme; Multivariable; Dynamic weights

1 Introduction

In the domains of climate change detection and attribution, future climate projection, and adaptation research, Earth system models (ESMs) serve as indispensable tools (Reichler and Kim, 2008; Knutti et al., 2010). However, the reliability of these model-based studies is closely tied to the simulation performance of the models themselves (Knutti et al., 2010). With advancements in computational power and a growing understanding of Earth system dynamics, the range of physical process

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

representations within models has expanded, and the interactions among these processes have become increasingly complex (Reichstein et al., 2019). As a result, there is a pressing need for more efficient and systematic evaluation approaches to accurately assess the consistency between model outputs and observational data, thereby offering a robust scientific foundation for future research endeavors (Eyring et al., 2019; Hausfather et al., 2020).

Given the extensive spatial and temporal variability in climate characteristics, the evaluation of Earth system models (ESMs) is a complex and challenging task. Current studies typically utilize performance metrics to quantify a model's ability to reproduce climate characteristics or processes (Baker and Taylor, 2016). For single variables, researchers assess the model's simulation capability by comparing the correlation, root mean square error, and variance ratio between the simulated and observed data (Lambert and Boer, 2001). For example, the Taylor diagram (Taylor, 2001) integrates these three metrics into a single graphical representation via geometric relationships. However, such methods are limited in their ability to comprehensively characterize the overall model performance (Pincus et al., 2008), particularly when addressing research questions involving multiple variables. Owing to the complex coupling and interactions between climate variables, achieving balanced simulation performance across different variables is often challenging. To evaluate the overall performance of the model (Gleckler et al., 2008), proposed the Model Climate Performance Index (MCPI), which quantifies the model's overall bias by calculating the mean relative error across multiple variables. Xu et al. (2017) introduced a multivariate integrated evaluation (MVIE) method that combines multiple scalar fields into a vector field to improve the overall assessment of climate models.

The Weather Research and Forecasting (WRF) model, a regional climate model utilized within the framework of the Coordinated Regional Downscaling Experiment (CORDEX) under the World Climate Research Programme (WCRP), is distinguished by its inclusion of multiple parameterization schemes for various physical processes, including microphysics, land surface, planetary boundary layer, cumulus, and radiation (Skamarock et al., 2019). These schemes differ in complexity and in their representations of physical processes, and varying combinations can result in substantial disparities in the simulation of key meteorological variables (Yu et al., 2022b).

However, due to the complex interactions among physical processes and the varying adaptability of individual parameterization schemes, no universally optimal configuration exists that can consistently perform well across different spatial, temporal, and output variable contexts (Fernández et al., 2007; Borge et al., 2008). The extensive range of freely selectable parameterization scheme combinations introduces significant uncertainty into the simulation performance of the WRF model (Mooney et al., 2013). Consequently, identifying and evaluating the optimal combination of parameterization schemes tailored to a specific study region is a critical prerequisite for WRF-based research (Chen et al., 2014; Ferreira et al., 2014).

Another major challenge stemming from the diversity of parameterization schemes is the substantial computational cost associated with model evaluation (Wang et al., 2021). For instance, in WRF version 4.3, the model provides 28 microphysics schemes, 8 land surface schemes, 9 longwave radiation schemes, 8 shortwave radiation schemes, 13 planetary boundary layer schemes, 8 surface schemes, and 13 cumulus schemes—yielding a theoretical total of over 14.5 million possible combinations.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

Conducting an exhaustive evaluation of all these configurations would impose an immense computational burden, rendering such an approach practically unfeasible for most researchers. To address this, and to balance improvements in model performance with the efficient use of computational resources, researchers commonly employ sensitivity analyses of physical process parameterization schemes. By systematically varying the parameterization schemes for individual physical processes, they assess the corresponding effects on simulation outputs, thereby identifying the most suitable scheme for each physical process (Li et al., 2014; Yu et al., 2022a).

Although the WRF model generates multiple output variables, researchers often prioritize the evaluation of key or difficult-to-simulate variables, such as precipitation. This univariate evaluation strategy may introduce biases in certain study regions, as it disproportionately emphasizes the accuracy of a single variable while neglecting the significant roles of other essential meteorological variables in regional climate dynamics (Lv et al., 2020). In response, the advancement of multivariate evaluation methodologies has prompted a shift toward more comprehensive assessments of the model's performance across multiple variables. This multidimensional approach facilitates a more objective and holistic evaluation of simulation capabilities, while also enhancing the understanding and quantification of uncertainties associated with parameterization schemes (Righi et al., 2020; Dai et al., 2021; Lu et al., 2021).

Although both univariate and multivariate integrated evaluation methods have contributed valuable insights into model comparison and simulation performance assessment, current multivariate approaches still exhibit notable limitations in objectively attributing sources of uncertainty. This is especially true for models like WRF that incorporate multiple physical process parameterization schemes, which exacerbates the uncertainty. First, this limitation is manifested in the averaging of simulation performance across different variables. Existing multivariate evaluation methods commonly allocate equal weights to all variables or rely on subjective expert judgment for weighting. However, the model's simulation capabilities for various variables differ across regions. The overall uncertainty in model performance is not contributed equally by all meteorological variables but is instead dominated by large simulation biases in a few key limiting variables (e.g., precipitation, wind speed). This is typically because these variables involve more complex nonlinear physical processes. Traditional equal-weighting methods overlook these differences in simulation accuracy, allowing variables with large biases to be masked by an averaging effect within the multivariate framework.

Second, the limitation is also evident in the neglect of the differential sensitivity of variables to various physical processes. In the sequential evaluation of individual physical processes, the optimal parameterization scheme selected for one process influences the evaluation of subsequent processes. This dependency can lead to the propagation and accumulation of errors, making the method highly reliant on the evaluation accuracy at each step. The influence of various physical processes on different output variables in WRF is inherently uneven. For example, microphysics and cumulus parameterizations exert a greater influence on precipitation simulations. Hence, it is essential to apply differentiated weights to variables when evaluating the parameterization schemes of distinct physical processes. Existing studies tend to apply a uniform weight setting across all physical processes, ignoring the coupled relationships and sensitivity differences of variables to different physical processes. Consequently, equal-weighting or subjective weighting methods cannot objectively reflect the simulation uncertainties

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

associated with each variable and physical process, challenging the scientific validity and reliability of the evaluation results. How to objectively attribute the sources of model uncertainty and effectively integrate this attribution into the evaluation system has become a pressing challenge.

To address the limitations outlined above, this study introduces an innovative dynamic weighting evaluation method, grounded in existing multivariate evaluation frameworks. This method aims to construct an evaluation framework that can objectively identify and quantify the uncertainty of each variable, dynamically amplify the influence of these variables during the evaluation and select parameterization scheme combinations that are more physically sound and robust in independent validations. The core of this method lies in a two-layer dynamic weighting mechanism: first, it assigns weights based on the model's varying simulation capabilities for different output variables to amplify the impact of key limiting variables; second, it dynamically adjusts these weights during the evaluation of different physical processes according to the sensitivity of each variable to those processes. As an objective weighting strategy based on the statistical characteristics of simulated variables, the dynamic weighting method provides substantial flexibility and adaptability. It enables real-time adjustment of weight distributions according to regional characteristics and the influence of various physical processes, thereby minimizing simulation errors arising from inappropriate or static weight assignments. This approach is particularly well-suited for evaluating complex models with multiple interacting physical processes, such as the WRF model. To validate the universality and robustness of the proposed method, we will conduct a comparative application of this method and the equal-weighting method in two distinct regions: the arid region of Northwest China and the coastal region of Southeast China. Furthermore, we will use an independent extreme climate year for validation. This design aims to comprehensively assess the method's effectiveness under different dominant physical process backgrounds.

The remainder of this paper is organized as follows. Sect. 2 describes the study area, model configuration, and data sources, as well as the proposed dynamic weighting evaluation method and its implementation. In Sect. 3, we systematically evaluate the performance of a large number of parameterization combinations and determine the regionally optimal physical parameterization scheme combination and compare the performance of the dynamic weighting method with the equal weighting method. An analysis of the robustness of the dynamic weighting method is further presented in Sect. 3.5. Sect. 4 discusses the weight allocation mechanism of the dynamic weighting method, along with its applicability and generalizability. Finally, Sect. 5 summarizes the main conclusions of this study.

2 Method

2.1 Study area

This study selects the arid region of Northwest China and the coastal region of Southeast China as its research areas. This choice is primarily based on the significant differences between the two regions in terms of climate background, underlying surface characteristics, and dominant physical processes. The two regions exhibit distinct hydro-thermal patterns but share similarities and complementarities in the frequency of extreme climate events and the complexity of climate dynamics. The

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

arid region of Northwest China is characterized by extreme temperature fluctuations and dust weather, whereas the coastal region of Southeast China is marked by intense precipitation events and frequent typhoons. These extreme climatic conditions present substantial challenges to the accuracy and reliability of regional climate models, making them ideal testbeds for evaluating model simulation capabilities under extreme climate scenarios. By independently applying the same evaluation methodology to these two regions, this study aims not only to verify the method's general capability to stably identify optimal parameterization schemes across different climate backgrounds but also to provide a critical scientific basis for future research on water resource management in arid regions and the development of rainstorm and flood early warning systems in coastal urban clusters.

2.1.1 Arid region of Northwest China

The arid region of Northwest of China (73°–106°E, 35°–50°N) lies at the core of the Eurasian continent, encompassing the entire Xinjiang region, the Hexi Corridor in Gansu Province, and parts of Inner Mongolia, as illustrated in Fig. 1(b). The region features complex topography, including prominent mountain ranges such as the Kunlun Mountains, Qilian Mountains, and Tianshan Mountains, as well as extensive inland basins like the Tarim and Junggar Basins. Its deep inland location, remote from maritime moisture sources, coupled with its position within a subsiding air divergence zone north of the Tibetan Plateau, severely limits moisture influx from the ocean, resulting in an extremely arid climate. Annual precipitation typically falls below 200 mm, with hyper-arid zones such as the Tarim Basin and Turpan receiving less than 50 mm per year, making this one of the driest regions globally. While a few southeastern subregions exhibit characteristics of a temperate monsoon climate, the vast majority of the area experiences a temperate continental or alpine climate, marked by cold, dry winters, hot summers, low precipitation, high evaporation rates, and significant diurnal temperature variation. Owing to its geographical position and distinctive climatic conditions, the region is highly sensitive to global climate change and regional anthropogenic activities—particularly regarding water cycle dynamics, temperature variability, and extreme climate phenomena such as droughts and heatwaves. This sensitivity underscores the region’s importance for understanding climate impacts in inland arid zones and across the Tibetan Plateau (Fu et al., 2022). Furthermore, due to its extreme aridity and climatic complexity, existing climate models often face challenges in accurately simulating regional climate change patterns. A comprehensive assessment and study of climate simulation performance in this region holds immense value for predicting climate change in arid areas globally.

2.1.2 Coastal region of Southeast China

In contrast to the arid region of Northwest, the coastal region of Southeast China is among the most precipitation-abundant areas in the country. Situated at the convergence of tropical and subtropical climatic zones, the region features a diverse landscape of coastal plains, mountains, and hills. It is also one of China’s most densely populated and highly urbanized regions, as illustrated in Fig. 1(c), and serves as a major hub of economic activity. Heavily influenced by the southeast monsoon and frequent typhoons, the region benefits from a substantial influx of warm, moist air, resulting in annual precipitation exceeding 1,500 mm across most areas. The coastal region of Southeast is characterized by relatively high temperatures, minimal diurnal

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

temperature variation, and a general equilibrium between precipitation and evaporation. This stands in sharp contrast to the arid, rainfall-deficient Northwest Arid Region, where large temperature fluctuations dominate. However, under the influence of global climate change and accelerated urbanization, the frequency and intensity of extreme climate events, such as typhoons, heavy rainfall, and droughts, have increased markedly in this region (Yang et al., 2015; Li et al., 2021), posing serious risks to both human populations and ecological systems. Accordingly, a comprehensive investigation into the regional climate system, particularly the coupled interactions among oceanic, atmospheric, and terrestrial components, is essential for enhancing climate prediction accuracy, mitigating the impacts of extreme events, and fostering sustainable regional development.

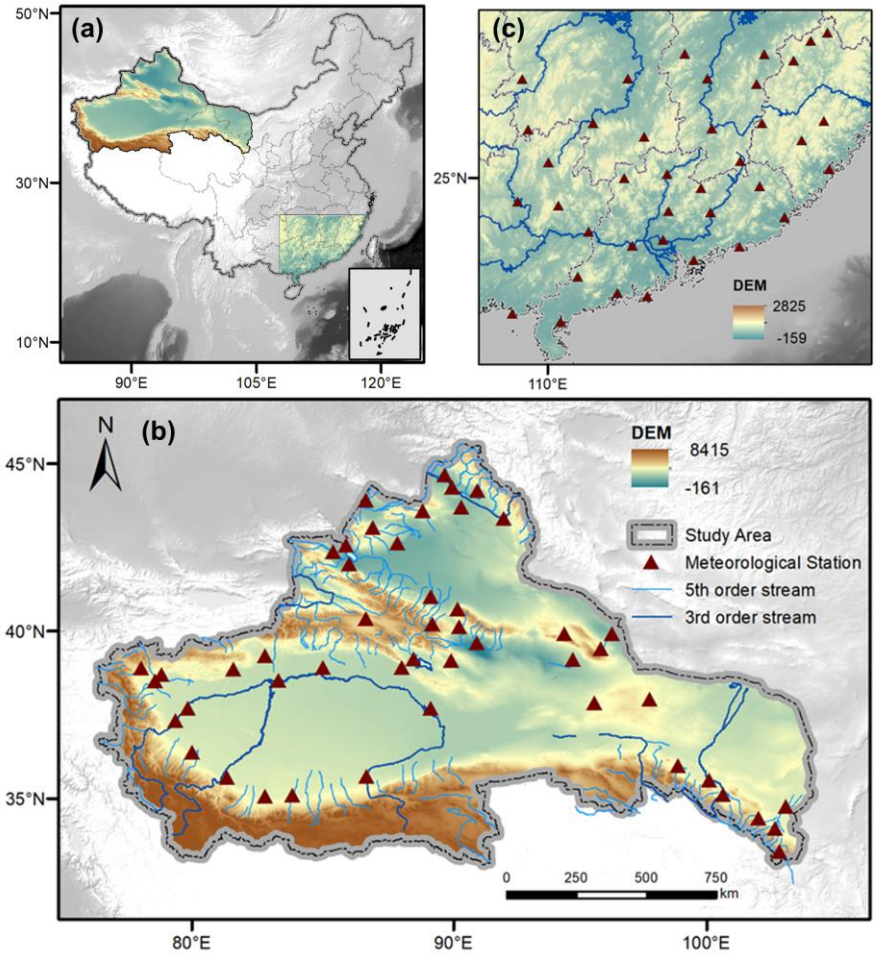


Figure 1. (a) Geographical location of the study area. (b) Overview of the arid region of Northwest China. (c) Overview of the coastal region of Southeast China.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

2.2 Model configuration and datasets

In this study, the Advanced Research WRF model version 4.3.3 was used for numerical simulations. Different model configurations were set for the two study regions (Table 1). For the arid region of Northwest China, the simulation domain covers 30°N to 52°N and 66°E to 114°E, with a horizontal grid resolution of 9 km and a total of 380 × 240 grid points. The model had 38 vertical levels, with the top of the model at 50 hPa. For the coastal region of Southeast China, the simulation domain covers 20°N to 29°N and 108°E to 120°E, with the same horizontal grid resolution of 9 km and a total of 120 × 105 grid points. The vertical configuration was identical to that of the arid region of Northwest.

Table 1. Overview of WRF model configurations.

	Northwest China	Southeast China
Model version	WRF/ARW (4.3.3)	WRF/ARW (4.3.3)
Horizontal resolution	9km	9km
Horizontal grid	380 (E–W) × 240 (S–N)	120 (E–W) × 105 (S–N)
Initial and boundary conditions	ERA5	ERA5
Simulation period	June 1st–Aug 31st, 2007	June 1st–Aug 31st, 2009
Land-use data	GLC_FCS30D	GLC_FCS30D
Default parameterization schemes	Microphysics-Thompson	Microphysics-Thompson
	PBL/surface layer scheme–MYJ	PBL/surface layer scheme–MYJ
	Cumulus–Tiedtke	Cumulus–Tiedtke
	Radiation–RRTMG	Radiation–RRTMG
	LSM–Noah LSM	LSM–Noah LSM

For this study, the period from May 1 to August 31 was selected for 2007 in the arid region of Northwest China and 2009 in the coastal region of Southeast China to screen for optimal parameterization scheme combinations using both dynamic and equal weighting methods. Consequently, simulations conducted from June to August were utilized to evaluate the performance of various physical parameterization scheme combinations. The selection of simulation years for the two study regions was based on meteorological station data within each region. This involved comparing daily meteorological observations, including temperature, surface temperature, precipitation, pressure, evaporation, relative humidity, sunshine duration, and wind speed and direction, across eight meteorological variables for each year from 1991 to 2020 with the corresponding 30-year climatological averages. The analysis identified 2007 and 2009 as the year most representative of the climatological state over the past three decades for the arid region of Northwest and the coastal region of Southeast, respectively (see Tables S1 and S2 in Supporting Information S1).

To evaluate the robustness of the proposed method, the year 2010, which was characterized by extreme climate events common to both study regions, was selected as an independent validation period. The optimal scheme obtained from the

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

195 screening period was applied to the validation period for simulation, and its performance was subsequently evaluated. The year 2010 is a representative year of extreme climate in China, during which the arid region of Northwest experiencing drought and the coastal region of Southeast suffering from severe flooding. From June to July in the Northwest, persistent high temperatures and markedly reduced rainfall, 30% to 80% below average, were observed. Concurrently, southern China experienced heavy rainfall in June and August, leading to significant flooding in the coastal region of Southeast, with 200 substantial loss of life and property. Designating 2010 as the validation period not only allows for an assessment of the method's robustness but also verifies its simulation capabilities under extreme climate conditions.

This study follows the evaluation order suggested by existing research, namely, microphysics, planetary boundary layer, cumulus, radiation, and land surface processes, and sequentially selects the optimal parameterization scheme for each physical process to ultimately form the overall optimal combination (Dai et al., 2021; Stergiou et al., 2017). For this study, a near-205 complete ensemble of the parameterization schemes available within WRF model was utilized, as presented in Table 2. The final number of scheme combinations was constrained by incompatibilities identified between certain physical process parameterizations. This resulted in the configuration of 42 (the dynamic weighting method) /43 (the equal weighting method) combinations for the arid domain in Northwest China and 54 (the dynamic weighting method) /53 (the equal weighting method) combinations for the coastal domain in Southeast China. (see Tables S3-S4 and S5-S6 in Supporting Information S1).

210

Table 2. Physics parameterization scheme assessed in this study.

	Microphysics	Planetary Boundary Layer/Surface Layer	Cumulus	Radiation	Land surface
1	8 Thompson	2 MYJ+2 MYJ	6 Tiedtke	1 RRTM+1 Dudhia	2 Noah LSM
2	1 Kessler	1 YSU+1 MM5	1 KF	3 CAM+3 CAM	1 SLAB
3	2 Lin	3 GFS+3 GFS	2 BMJ	5 New Goddard	4 NoahMP
4	3 WSM3	4 QNSE+4 QNSE	3 GF	99 GFDL+99 GFDL	5 CLM4
5	4 WSM5	5 MYNN2.5+1 MM5	5 Grell3		
6	5 Eta/Ferrier	5 MYNN2.5+2 MYJ	7 ZM		
7	6 WSM6	5 MYNN2.5+5 MYNN	11 MsKF		
8	9 Milbrandt	6 MYNN3+ 1 MM5	14 NSAS		
9	10 Morrison	6 MYNN3 + 2 MYJ	16 New Tiedtke		
10	11 CAM	6 MYNN3 + 5 MYNN	93 GD		
11	13 Lin-SBU	7 ACM2+1 MM5	99 Old KF		
12	14 WDM5	7 ACM2+7 Pleim-Xiu			
13	16 WDM6	8 BouLac + 1 MM5			
14	17 NSSL 2-17	8 BouLac + 2 MYJ			
15	18 NSSL 2-18	9 UW + 1 MM5			
16	19 NSSL 1-19	9 UW + 2 MYJ			

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

17	21 NSSL 1-21	11 Shin-Hong+1 MM5
18	24 WSM7	12 GBM + 1 MM5
19	26 WDM7	99 MRF+1 MM5
20	28 Thompson aerosol-aware	
21	50 P3	

The model's initial and boundary conditions were provided by the ECMWF's fifth-generation atmospheric reanalysis dataset (ERA5), with a spatial resolution of 0.25°×0.25° and a temporal resolution of 6 h. The reanalysis dataset is available on the European Center for Medium-Range Weather Forecasting (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels ; https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels).

Owing to the lack of ground validation and limitations in spatial and temporal resolution, the default MODIS land use dataset in the WRF model exhibits discrepancies with actual land use results in the study areas. This study uses the 1985–2022 global 30-m land use product (GLC_FCS30D) developed by the Aerospace Information Research Institute, Chinese Academy of Sciences, as the surface input data (Zhang et al., 2024), available for download from https://zenodo.org/records/8239305. This dataset meets the requirements for continuous and consistent land use data in climate modeling, with a spatial resolution of 30 m and a time span from 1985–2022, making it suitable for more detailed regional simulation studies. The land use types in the dataset were reclassified according to the IGBP20 classification standard.

To comprehensively assess the performance of the WRF model, observational data from meteorological stations were used to validate the simulated variables. Daily meteorological observation data were obtained from the China Meteorological Data Service Centre (https://data.cma.cn/en). A total of 49 meteorological stations from the arid region of Northwest China and 38 from the coastal region of Southeast China, which provided long-term daily meteorological data from 1991–2020, were collected. The spatial distribution of the meteorological stations in the study area is shown in Fig. 1. These station-based observations underwent rigorous calibration and quality control, and thus provide a more accurate depiction of surface meteorological conditions than gridded datasets. Based on the availability of meteorological variables at these stations, the output variables selected for evaluation in this study included 2-m air temperature (T2), ground surface temperature (GST), air pressure (PRS), precipitation (PRE), relative humidity (RH), meridional wind speed (U10), and zonal wind speed (V10).

2.3 Model evaluation and variable weighting method

2.3.1 Taylor diagram and related statistical metrics

This study utilizes several multivariate evaluation methods to describe the performance of the WRF model. The normalized Taylor diagram (Taylor, 2001) is used to assess the performance of different physical schemes in the sensitivity experiments to find the optimal combination of WRF schemes. The Taylor diagram simultaneously displays three complementary statistical metrics related to model performance: the center root mean square error (cRMSE), standard deviation (σ), and correlation

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

coefficient (R). In addition, the vector statistical method (Xu et al., 2016) is employed, allowing the statistical metrics to reflect the combined results of multiple variables.

$$\sigma_{fn} = \sqrt{\sum_{m=1}^{M} \omega_m \left(\frac{1}{T} \sum_{t=1}^{T} \left(f_{mnt} - \overline{f_{mn}} \right)^2 \right)} \tag{1}$$

$$\sigma_r = \sqrt{\sum_{m=1}^{M} \omega_m \left(\frac{1}{T} \sum_{t=1}^{T} \left(r_{mt} - \overline{r_m} \right)^2 \right)} \tag{2}$$

$$R_n = \frac{\sum_{m=1}^{M} \omega_m \left(\frac{1}{T} \sum_{t=1}^{T} \left(f_{mnt} - \overline{f_{mn}} \right) \left(r_{mt} - \overline{r_t} \right) \right)}{\sigma_{fn} \sigma_r} \tag{3}$$

$$cRMSE_n = \sqrt{\sum_{m=1}^{M} \omega_m \left(\frac{1}{T} \sum_{t=1}^{T} \left(\left(f_{mnt} - \overline{f_{mn}} \right) - \left(r_{mt} - \overline{r_t} \right) \right)^2 \right)} \tag{4}$$

In the equation, m represents the output variable, n represents the model's parameterization scheme, and t refers to T discrete points in time and/or space. f_{mnt} represents the simulated values of the m variable at the t discrete points for the model's n parameterization scheme, whereas r_{mt} represents the observed values of the m variable at the t discrete points. ω_m represents the weight of the m variable during the statistical calculation.

2.3.2 Model performance indicators

Although Taylor diagrams provide a multidimensional perspective on model performance through multiple statistical indicators, they are limited in their ability to fully quantify model superiority in complex evaluation scenarios. While it is widely recognized that a single performance metric cannot capture all aspects of model evaluation across multiple dimensions, the establishment of a unified evaluation standard is essential for this study to more intuitively illustrate the process of parameterization scheme selection. To this end, the multivariable integrated evaluation index (MIEI) (Zhang et al., 2021) is adopted as the primary metric for assessing model simulation performance.

MIEI integrates both the magnitude and pattern similarity of model-simulated variables and maintains a strictly monotonic relationship with overall model performance: the lower the MIEI value, the better the simulation accuracy. This makes MIEI particularly well-suited for ranking the comprehensive performance of models in multivariable simulations (Lv et al., 2020). A detailed description of the formulation and application of this method is provided in (Xu et al., 2016; Xu et al., 2017; Zhang et al., 2021).

$$MIEI_n^2 = \sum_m^M \omega_m \left(L_{mn}^* - 1 \right)^2 + 2\left(1 - R_n \right) \tag{5}$$

$$L_{mn}^* = \begin{cases} \sigma'_{mn}, & \sigma'_{mn} \le 1 \\ \dfrac{1}{\sigma'_{mn}}, & \sigma'_{mn} > 1 \end{cases} \tag{6}$$

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

$$\sigma'_{mn} = \frac{\sigma_{fmn}}{\sigma_{rm}} \tag{7}$$

In the equation, m represents the output variable, n represents the model's parameterization scheme. ω_m represents the weight of the m variable during the statistical calculation. σ'_{mn} represents the ratio of the σ value of the m variable in the n parameterization scheme of the model to the observed σ value. R_n represents the correlation coefficient between the simulated values and the observed values of the n parameterization scheme of the model.

2.3.3 Dynamic weight evaluation approach

The dynamic weighting method proposed in this study is mainly used to calculate the variable weights ω_m in the evaluation metrics mentioned above. This is an objective weighting method that is independent of the statistical indicators used, so that it can be applied to all multi-objective evaluation methods. The proposed dynamic weighting method is built upon a simple yet powerful principle: a variable should be given more importance in the evaluation if (1) the model struggles to simulate it accurately, and (2) its simulation is highly sensitive to the choice of parameterization schemes for a given physical process. The foundation of this method is based on two statistical indicators: the mean relative error ($\overline{MRE_m}$) of the variable and the coefficient of variation ($CV_{|ME|m}$) of the variable's absolute error.

The mean relative error ($\overline{MRE_m}$) represents the difference between the simulated value and the observed value of a specific variable for a given physical parameterization process. This indicates the simulation ability of that physical parameterization process for the variable.

$$\overline{MRE_m} = \frac{1}{N}\sum_{n=1}^{N}\left(MRE_{mn}\right) \tag{8}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\frac{|ME|_{mn}}{\frac{1}{T}\sum_{t=1}^{T}|r_{mt}|}\right) \tag{9}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\frac{1}{T}\sum_{t=1}^{T}|f_{mnt} - r_{mt}|}{\frac{1}{T}\sum_{t=1}^{T}|r_{mt}|}\right) \tag{10}$$

In the equation, f_{mnt} represents the simulated values of the m variable at the t discrete points for the model's n parameterization scheme, while r_{mt} represents the observed values of the m variable at the t discrete points. $|ME|_{mn}$ represents the absolute error of the m variable in the n parameterization scheme of the model.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

The coefficient of variation of the absolute error ($CV_{|ME|m}$) represents the degree of dispersion when a physical parameterization process simulates a specific variable. This indicates the sensitivity of the variable to the physical parameterization process.

$$CV_{|ME|m} = \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(|ME|_{mn} - \overline{|ME|_m}\right)^2}}{\overline{|ME|_m}} \tag{11}$$

$$= \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}|f_{mnt} - r_{mt}| - \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}|f_{mnt} - r_{mt}|\right)\right)^2}}{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}|f_{mnt} - r_{mt}|\right)} \tag{12}$$

In the equation, f_{mnt} represents the simulated values of the m variable at the t discrete points for the model's n parameterization scheme, whereas r_{mt} represents the observed values of the m variable at the t discrete points. $|ME|_{mn}$ represents the absolute error of the m variable in the n parameterization scheme of the model. $\overline{|ME|_m}$ indicates the absolute error of the m variable for all parameterization schemes of the model.

The base of the method is to assign higher weights to variables with larger relative errors (poor simulation performance) or higher coefficients of variation (more sensitive to the physical process). To achieve this, $Softmax = \frac{e^x}{\sum_{n=1}^{N} e^x}$ is employed to convert the two statistical indicators (relative error and coefficient of variation) into a weight value between 0 and 1 for each variable. The Softmax function is a widely used activation function in classification tasks, transforming raw numerical outputs into a probability distribution in which each output lies between 0 and 1, and the sum of all outputs equals 1. The exponential property of the Softmax function non-linearly amplifies the influence of variables with greater error or higher sensitivity. This ensures that the most problematic or responsive variables are assigned substantially higher weights, thereby establishing a clear hierarchy of importance. This characteristic is ideally suited to the objective of this study, which is to prevent the dilution of critical simulation biases. It is important to note, however, that the exponential form used in the Softmax function is not inherently fixed. In practice, alternative functional forms can be employed to scale statistical outputs, depending on the structure and characteristics of the data.

$$\omega_{m1} = \frac{e^{\overline{MRE_m}}}{\sum_{n=1}^{N} e^{\overline{MRE_m}}} \tag{13}$$

$$\omega_{m2} = \frac{e^{CV_{|ME|m}}}{\sum_{n=1}^{N} e^{CV_{|ME|m}}} \tag{14}$$

$$\omega_m = \omega_{m1} \times \omega_{m2} \tag{15}$$

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

In the equation, ω_{m1} and ω_{m2} represent the weight values of variable m derived from the mean relative error and the coefficient of variation of the absolute error, respectively. ω_m represents the weight of the m variable.

The specific application process of this method in WRF model evaluation includes the following steps:

Step 1: Run all the parameterization schemes in the first physical parameterization process (e.g., microphysics schemes) and calculate the absolute and relative errors of the simulation results for different variables compared with the observed values.

Step 2: Calculate the mean relative error and the coefficient of variation of the absolute error for the simulation results of each parameterization scheme for different variables to quantify the degree of error and sensitivity of different variables.

Step 3: Use the Softmax function to convert the two statistical indicators (mean relative error and coefficient of variation) into two sets of weights, which serve as the importance measure for each variable in the simulation process.

Step 4: Based on the obtained weights, further calculate the statistical indicators for each scheme, such as the RMSE, R, σ, and MIEI, to determine the best scheme for the first physical parameterization process.

Step 5: Use the best scheme identified in Step 4, continue simulating the next physical parameterization process and repeat the above steps until the best scheme for all physical parameterization processes is determined.

3 Result

3.1 Optimal parameterization scheme combination for the arid region of Northwest

Figure 2 shows the weights assigned to variables in the WRF simulations of different physical processes for the arid region of Northwest using the dynamic weighting method. Figure 2(a) shows the weights ω_{m1} assigned on the basis of the mean relative error ($\overline{MRE_m}$), where wind speed and precipitation received average weights of 0.45 and 0.24, respectively, while all other variables had weights below 0.01 The higher weights for wind speed and precipitation indicate that all physical processes in the WRF model have significant errors in simulating these variables, reflecting the model's weaker simulation capability for these variables. Therefore, more attention needs to be given to these variables in the model.

Figure 2(b) shows the weights ω_{m2} assigned on the basis of the coefficient of variation ($CV_{|ME|m}$) of the absolute error, where the weights for different variables showed considerable variation during the simulation process, reflecting significant differences in the sensitivity of the variables to different physical processes. Precipitation and temperature were the variables exhibiting the most pronounced variability. Precipitation carried weight values of 0.17 and 0.18 in the microphysics and cumulus processes, respectively, indicating its high sensitivity to these two physical processes. In contrast, temperature demonstrated greater sensitivity to radiation and land surface processes, with corresponding weight values of 0.184 and 0.201. Relative humidity showed a more pronounced response during the planetary boundary layer process, with a weight value of 0.176.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

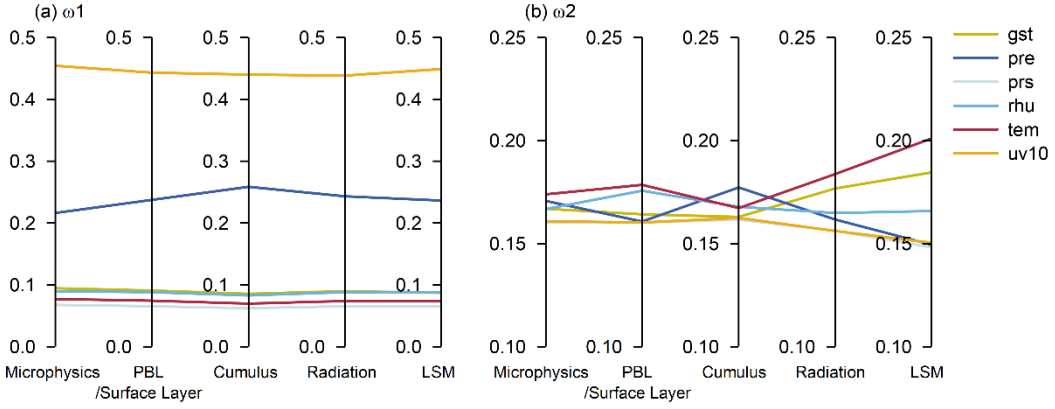
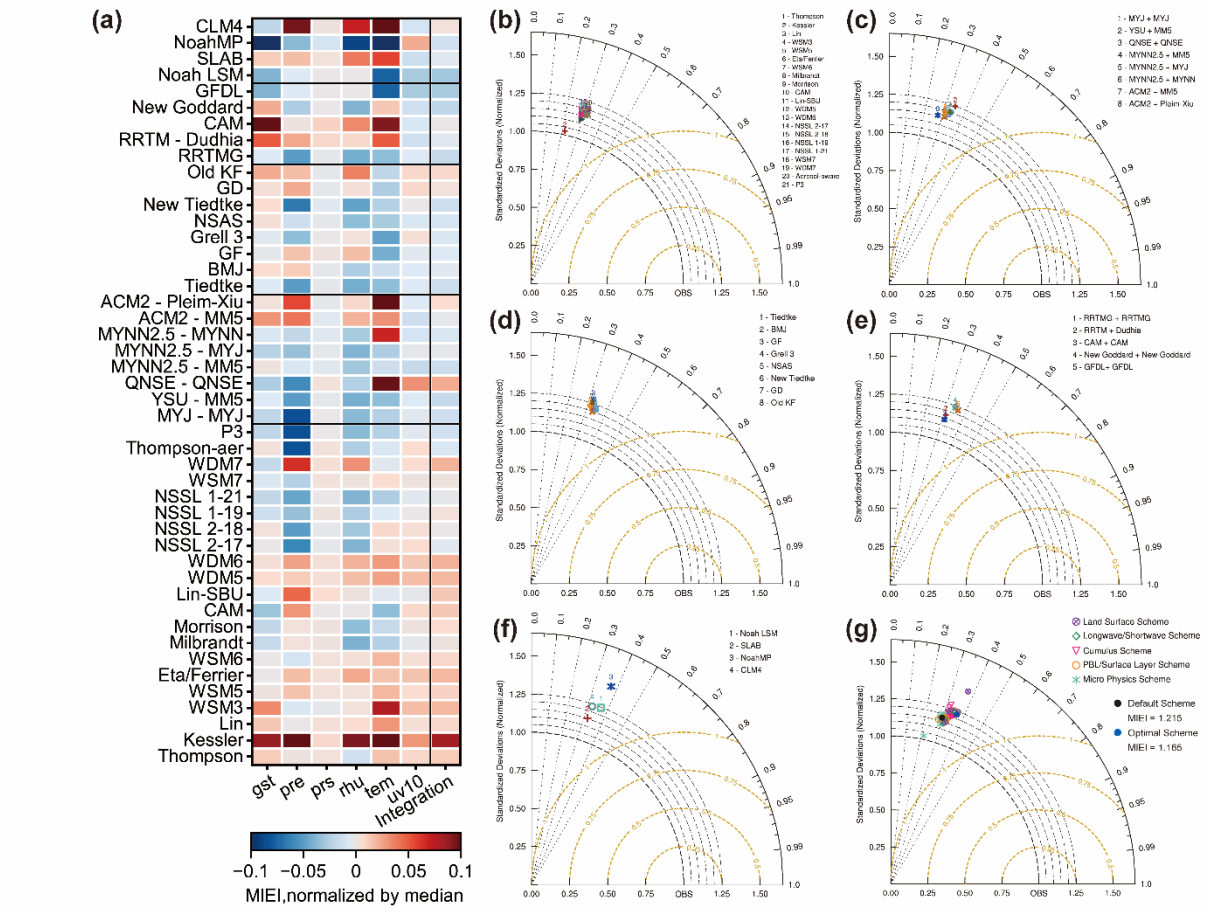


Figure 2. Weight distributions of the variables associated with the different physical processes of the WRF simulation in the arid region of Northwest. (a) Based on the mean relative error, (b) Based on the coefficient of variation of the absolute error.

Figure 3(a) displays the standardized MIEI results for evaluating WRF parameterization schemes in the arid region of Northwest, providing an assessment of simulation performance across various output variables and in the overall integrated evaluation. Analysis of the simulation results for individual variables revealed substantial variability in precipitation simulation performance among different parameterization schemes. Notably, the P3 and MYJ schemes exhibited superior performance in simulating precipitation, as indicated by their lower MIEI values, which reflected greater accuracy. In contrast, the Kessler and CLM schemes yielded higher MIEI values, suggesting larger simulation errors. Although CLM, as a land surface model, does not directly govern precipitation-related processes, its influence on the hydrological cycle and land-atmosphere interactions significantly affects precipitation outcomes. In the simulations of surface and air temperature, the impact of specific physical processes on these variables is more direct. For example, the Noah-MP and CLM4 schemes achieved the lowest and highest MIEI values, respectively, for land surface processes, underscoring the critical role of land surface parameterizations in air temperature simulation. Additionally, the YSU scheme demonstrated strong performance in simulating wind speed within the planetary boundary layer process and was subsequently selected as the optimal scheme for further assessment. Consequently, the subsequent parameterization schemes for cumulus, radiation, and land surface processes also demonstrated good performance for wind speed.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025



355 **Figure 3.** Evaluation results of parameterization schemes in the arid region of Northwest. (a) Normalized MIEI and (b-f) Taylor diagrams of different physical processes. (b) Microphysics, (c) Planetary boundary layer, (d) Cumulus, (e) Radiation, (f) Land surface, and (g) Taylor diagrams of a total of 44 combinations of parametric schemes for all physical processes

In the multivariate integrated evaluation (rightmost column), the MIEI value serves as the primary criterion for selecting parameterization schemes in this study. As the evaluation of physical processes advances, a progressive decrease in MIEI
360 values was observed, reflecting a corresponding improvement in model simulation accuracy. The integrated evaluation identified the P3 (microphysics), YSU (planetary boundary layer), Tiedtke (cumulus), GFDL (radiation), and Noah LSM (land surface) schemes as the optimal choices indicating that their combination yielded the most suitable configuration for the arid region of Northwest. However, it is important to note that the optimal parameterization scheme does not necessarily deliver the best performance across all individual variables. For instance, within the radiation process, the GFDL scheme exhibited
365 slightly inferior performance compared to the RRTMG scheme in simulating relative humidity and precipitation. Nonetheless, its superior simulation results for wind speed and temperature led to a lower overall MIEI value, establishing GFDL as the relatively best-performing scheme in the integrated evaluation.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

Compared to MIEI, the Taylor diagram offers a more comprehensive statistical evaluation of model performance, highlighting the discrepancies between different parameterization schemes and observational data. Figure 3(b-f) present the centered root mean square error (cRMSE), standard deviation (σ), and correlation coefficient (R) for each scheme across the various physical processes. In the microphysics process, the optimal parameterization scheme, P3, exhibited a relatively low RMSE of 1.232 and the highest correlation coefficient of 0.321. Most microphysics schemes demonstrated elevated standard deviation ratios (ranging from 1.156 to 1.205), indicating a tendency to overestimate the amplitude of anomalies in the simulated fields. For the planetary boundary layer process, the YSU scheme emerged as optimal, with the highest R value (0.346). Although YSU performed well in terms of correlation, it still overestimated the amplitude of anomalies, as reflected in the standard deviation. In the cumulus process, the Tiedtke scheme demonstrated superior performance, with the lowest RMSE (1.277). For the radiation process, the GFDL scheme yielded the highest correlation coefficient (0.363). Although the CAM scheme showed better performance in terms of RMSE and standard deviation, its lower correlation coefficient (0.312) led to a lower integrated evaluation score compared to GFDL. In the land surface process, although the Noah LSM scheme did not achieve the best results in RMSE (1.282), correlation coefficient (0.365), its overall balanced performance across all indicators rendered it the most suitable parameterization scheme in the integrated evaluation.

Figure 3(g) presents the Taylor plot for the combined simulations of all 44 parameterization scheme combinations. Each color denotes a different physical process option, with the blue symbol representing the optimal scheme and the black symbol indicating the default scheme for the study area. The plot revealed a clear separation in the statistical outcomes associated with different physical processes. Following the systematic selection of parameterization schemes across five physical processes, both the RMSE and R showed marked improvement. Notably, the optimal combination of parameterization schemes significantly outperformed the default configuration. The optimal scheme achieved a higher correlation coefficient (R = 0.364) compared to the default (R = 0.295), and a lower RMSE (1.277 versus 1.300), demonstrating its superior capacity to reproduce meteorological variables in the WRF model for the arid region of Northwest. A detailed analysis of individual physical processes indicated that microphysical, cumulus, and land surface schemes primarily contributed to reducing RMSE and enhancing overall simulation accuracy, though they exerted limited influence on the correlation coefficient. The planetary boundary layer process and the radiation process mainly affected the correlation coefficient and had a relatively weak influence on the root mean square error.

3.2 Comparison of dynamic weighting and equal weighting methods in the arid region of Northwest

In addition to employing the dynamic weighting method to determine the optimal parameterization scheme for the arid region of Northwest, this study also applies an equal-weighting approach, maintaining a consistent evaluation framework for both methods. Figure 4 presents the simulation results for various output variables obtained using the two weighting strategies.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

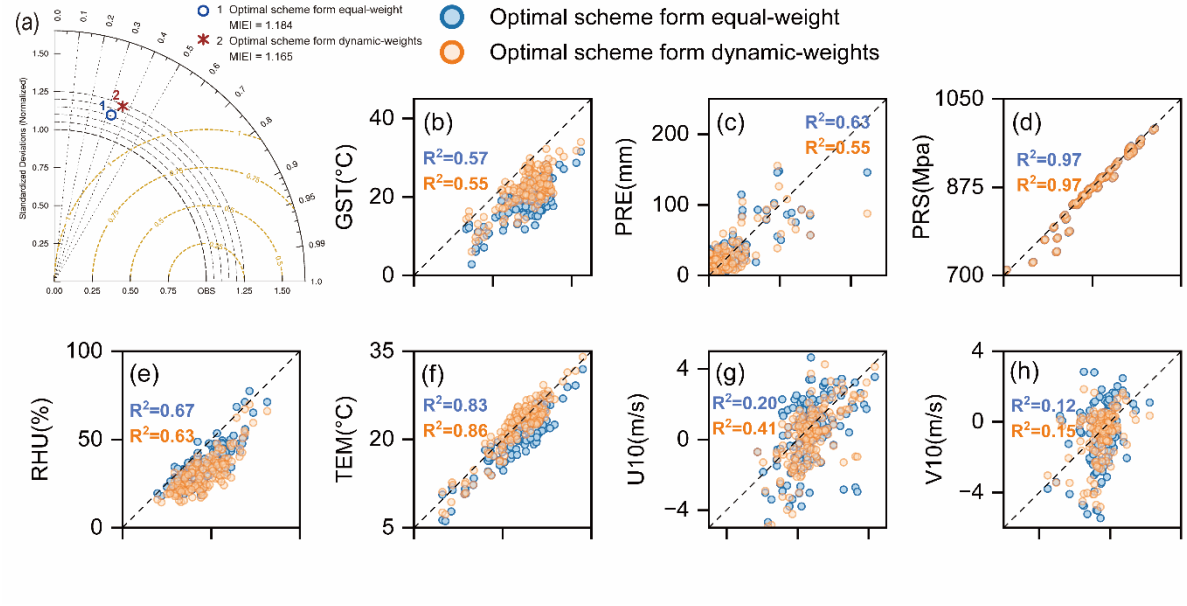


Figure 4. Comparison of simulation results obtained from the optimal parameterization schemes using the dynamic weight and equal weight methods. (a) Taylor plots of the optimal parameterization schemes obtained by different methods. (b-h) Scatter plots of the optimal parameterization schemes obtained via different methods for different meteorological elements. (b) Surface temperature, (c) precipitation, (d) Air pressure, (e) Relative humidity, (f) Air temperature, (g) Zonal wind, (h) Meridional wind.

Figure 4(a) displays the Taylor plots for the optimal parameterization schemes derived from the dynamic and equal-weighting methods, providing a visual comparison of their respective performances. The optimal scheme identified through dynamic weighting achieved a lower MIEI value (1.165) compared to that derived from the equal-weighting method (1.184). More specifically, the dynamic weighting approach yielded a higher correlation coefficient (R = 0.364) versus the equal-weighting method (R = 0.323).

Figure 4(b-h) shows scatter plots of the optimal parameterization schemes obtained via dynamic weighting and equal weighting methods, which present a comparison of the monthly mean values (where precipitation is the monthly cumulative value) and the observed values of the two optimal parameterization schemes at different stations. Among all the variables, air pressure performed the best. Both schemes showed extremely high fitting degrees and simulated the changing trend of air pressure relatively accurately. For the simulation results of surface temperature and relative humidity, the dynamic weighting scheme yielded a slightly lower R^2 compared to the equal weighting scheme. During the temperature simulation process, the dynamic weighting scheme demonstrated superior performance under high-temperature scenarios, achieving an R^2 of 0.86. In contrast to the air temperature results, the dynamic weighting scheme performed relatively poorly in the precipitation simulation, and R^2 decreased to 0.55. For the simulation of wind speed, neither scheme performed well in representing the zonal and meridional components. However, the R^2 values of the dynamic weight scheme reached 0.41 and 0.15, which were significantly higher than those of the equal weight scheme (0.20 and 0.12), improving the simulation of the wind speed. Although the simulation performances of some variables have slightly decreased, on the whole, the combination of the optimal

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

parameterization schemes with dynamic weights simulated the changes in meteorological elements in the arid region of Northwest China better than the optimal parameterization schemes obtained via the equal-weight method.

3.3 Optimal parameterization scheme combination for the coastal region of Southeast

Figure 5 shows the weight distributions of different variables across physical processes in the WRF simulations for the coastal region of Southeast, as determined by the dynamic weighting method. Figure 5(a) represents the weights ω_{m1} calculated on the basis of the mean relative error ($\overline{MRE_m}$). The simulation weights for wind speed and precipitation remained the highest, with average weights of 0.36 and 0.31, respectively. Compared with those of the arid region of Northwest, the weights of the other variables were generally greater. The weight for the wind speed has decreased noticeably, indicating that the WRF model has reduced the disparity between the wind speed simulation and other variables in the coastal region of Southeast, suggesting a more balanced simulation capability across different variables in this region.

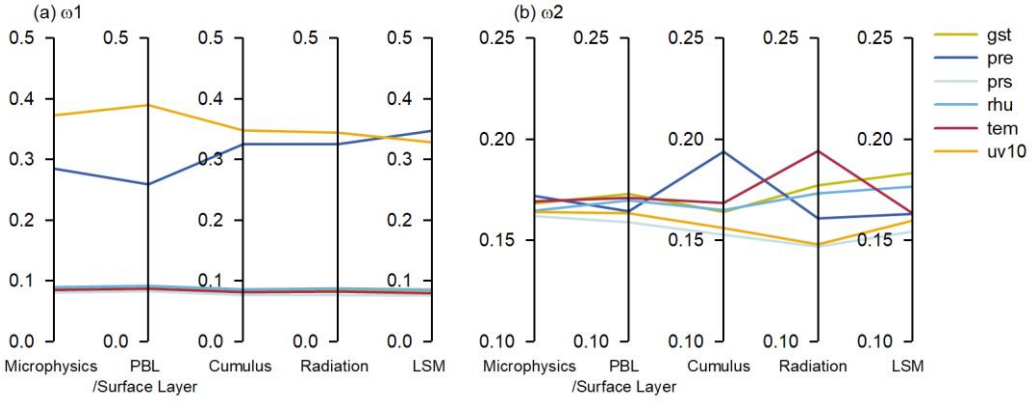


Figure 5. Weight distributions of the variables in the WRF simulations of the different physical processes in the coastal region of Southeast. (a) Mean relative error and (b) coefficient of variation of the absolute error

Figure 5(b) represents the weights ω_{m2} calculated on the basis of the coefficient of variation ($CV_{|ME|m}$) of the absolute error. There was substantial variability in the weights assigned to different variables throughout the simulation process, reflecting marked differences in their sensitivity to various physical processes. Precipitation exhibited the greatest variation in weight across physical processes, with a maximum value of 0.194 in the cumulus process, highlighting the direct influence of cumulus on precipitation simulation. In contrast, air temperature and surface temperature demonstrated higher sensitivity to the radiation process, with corresponding weight values of 0.194 and 0.177, respectively.

Figure 6(a) presents the standardized MIEI results for the evaluation of WRF parameterization schemes in the coastal region of Southeast. Analysis of the simulation performance across different variables revealed substantial variability in precipitation simulation across physical processes, with the microphysics process generally exhibiting the lowest performance. This outcome reflected the high sensitivity of precipitation to microphysical processes. Within this category, the WDM6 scheme was identified as the optimal parameterization scheme and was subsequently incorporated into the following stages of

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

evaluation. As a result, the parameterization schemes for the remaining physical processes demonstrated improved performance in precipitation simulation. For wind speed, the planetary boundary layer and land surface processes exerted a more direct influence, yielding superior simulation results relative to other physical processes. Surface temperature, air temperature, and relative humidity exhibited significant fluctuations and variability in MIEI values throughout the evaluation. These variations were attributed to the complex influence of multiple physical processes and the dominant role played by previously selected optimal parameterization schemes.

The MIEI values derived from the multivariate integrated evaluation showed a consistent decline as the assessment progresses, indicating incremental improvements in overall model performance with each stage of physical process evaluation. Based on the multivariate evaluation, the optimal combination of parameterization schemes for the coastal region of Southeast included WDM6 (microphysics), Shin-Hong-MM5 (planetary boundary layer), Grell 3 (cumulus), GFDL-GFDL (radiation), and Noah LSM (land surface). As with the arid region of Northwest, the optimal combination does not guarantee superior simulation performance for all variables. For instance, in the cumulus process, the Grell 3 scheme performed exceptionally well in simulating precipitation and relative humidity but was slightly less effective for air temperature and wind speed. Nevertheless, due to the higher weight assigned to precipitation in the cumulus evaluation, Grell 3 remained the preferred scheme in the integrated analysis.

Figure 6(b-f) present Taylor diagrams for different physical process parameterization schemes. In the microphysical process, the optimal parameterization scheme, WDM6, had a relatively small RMSE (1.042), demonstrating good consistency between the simulated and observed values of anomalous fields. The correlation coefficients for all parameterization schemes within the microphysics process ranged from 0.42 to 0.47, exhibiting only minor variation, which suggested that the microphysics process exerted limited influence on the model's correlation performance. In contrast, within the planetary boundary layer process, the optimal scheme demonstrated a well-balanced simulation capability, with an RMSE of 1.066 and a correlation coefficient (R) of 0.488, both ranking among the highest across evaluated schemes. In the cumulus category, the Grell 3 scheme emerged as the optimal choice, yielding the lowest RMSE (1.032) and a relatively high correlation coefficient (0.478), while also effectively capturing the amplitude of anomalous field variations. For the radiation and land surface processes, the GFDL and Noah LSM schemes, respectively, achieved superior correlation coefficients compared to other schemes. Although their RMSE were relatively elevated, these schemes were nonetheless regarded as optimal due to their overall strong performance in the integrated evaluation.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

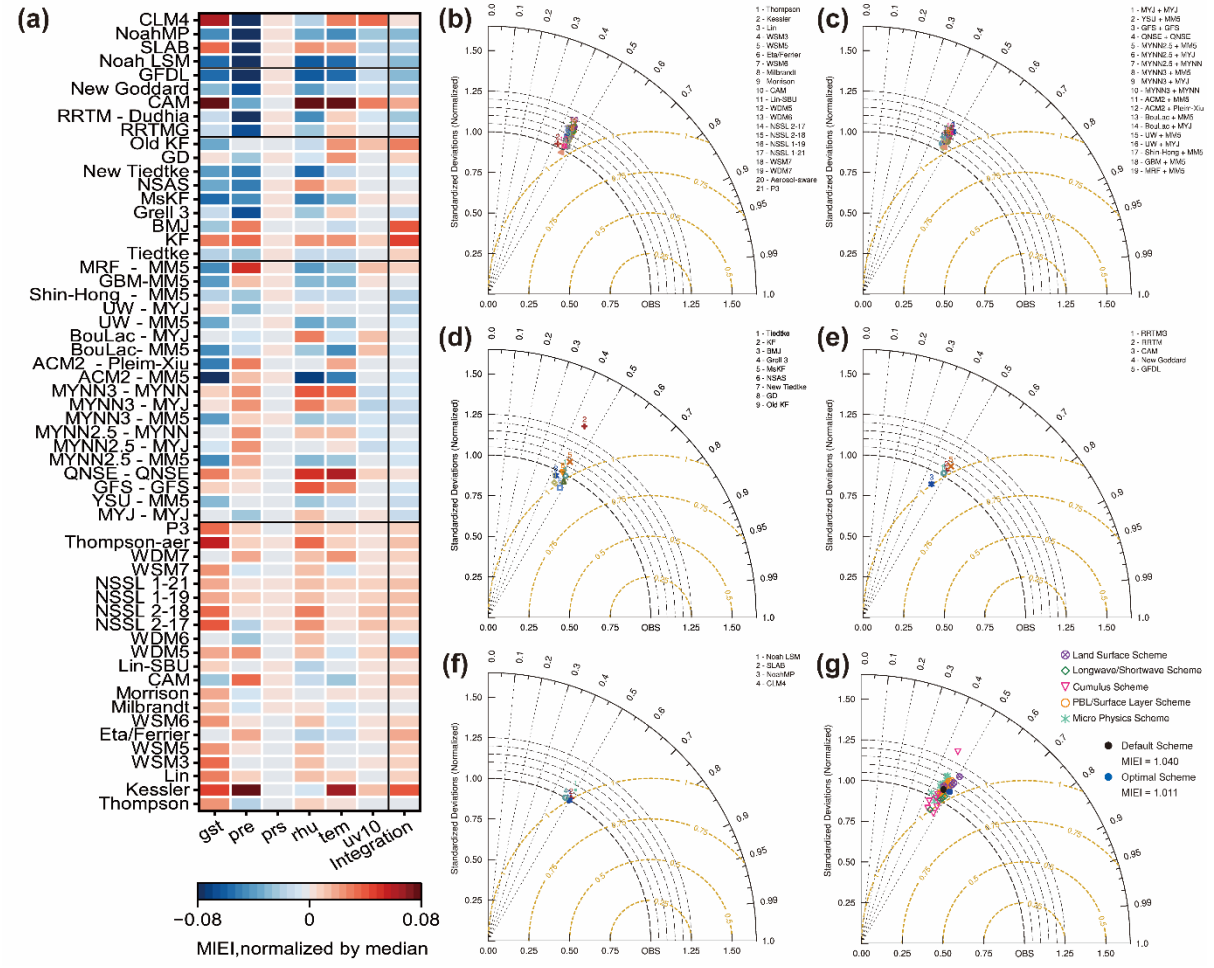


Figure 6. Evaluation results of parameterization schemes in the coastal region of Southeast. (a) Normalized MIEI and (b-f) Taylor diagrams of different physical processes. (b) Microphysics, (c) Planetary boundary layer process, (d) Cumulus, (e) Radiation, (f) Land surface, and (g) Taylor diagrams of a total of 54 combinations of parametric schemes for all physical processes.

Figure 6(g) presents the Taylor diagram summarizing the combined simulation results of all 54 parameterization scheme combinations. Following the sequential selection of optimal schemes for the five physical processes, both RMSE and R exhibited notable improvements. Compared with the default scheme, the optimal configuration achieved a higher R value (0.502 versus 0.470) and a lower RMSE (1.039 versus 1.067), underscoring its enhanced ability to reproduce various meteorological variables in the WRF model for the coastal region of Southeast. Independent analysis of each physical process revealed that microphysical, cumulus, and planetary boundary layer processes primarily contributed to reducing RMSE and enhancing overall simulation accuracy. In contrast, the radiation and land surface processes mainly influenced the correlation coefficient, with comparatively limited effects on the root mean square error.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

3.4 Comparison of dynamic weight and equal weight methods in the coastal region of Southeast

Figure 7 compares the performance of the optimal parameterization schemes derived using the dynamic weighting method and the equal-weighting method in simulating various output variables. Figure 7(a) displays the Taylor plots corresponding to each method. The optimal scheme obtained through dynamic weighting exhibited a lower MIEI value (1.011) than that derived from the equal-weighting method (1.020). Specifically, the dynamic weighting scheme achieved a higher correlation coefficient (R = 0.502) compared to the equal-weighting scheme (R = 0.499), and a smaller RMSE (1.039 vs. 1.106), indicating superior overall simulation accuracy.

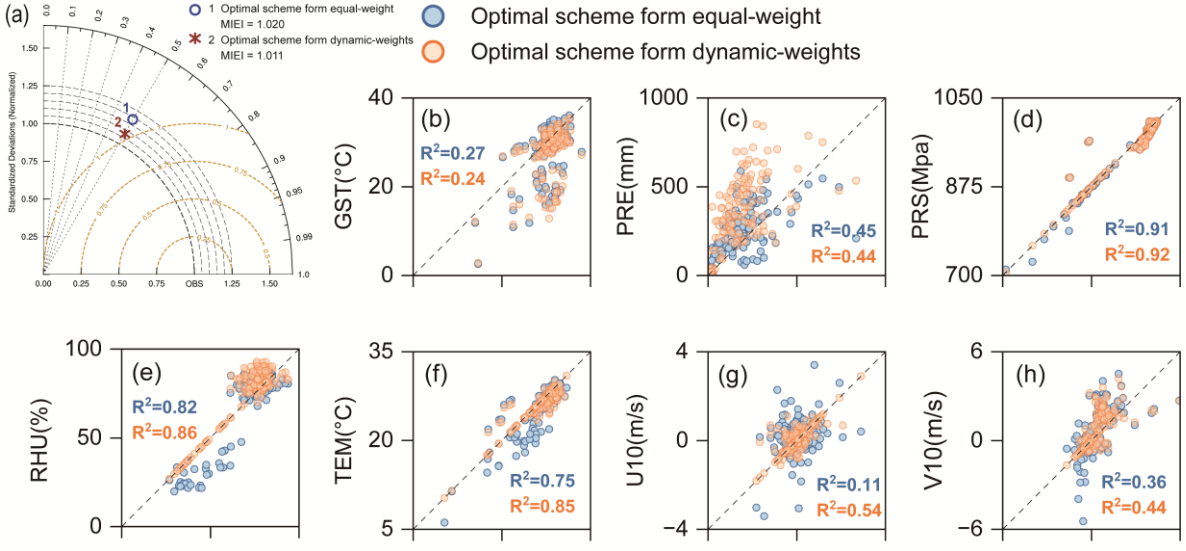


Figure 7. Comparison of simulation results obtained from the optimal parameterization schemes using the dynamic weight and equal weight methods. (a) Taylor plots of the optimal parameterization schemes obtained by different methods. (b-h) Scatter plots of the optimal parameterization schemes obtained via different methods for different meteorological elements. (b) Surface temperature, (c) precipitation, (d) Air pressure, (e) Relative humidity, (f) Air temperature, (g) Zonal wind, (h) Meridional wind.

Figure 7(b-h) present scatter plots comparing the observed and simulated results for each output variable using the respective optimal parameterization schemes. In the simulation of surface temperature, the dynamic weighting scheme yielded an R^2 value of 0.24, slightly lower than that of the equal-weighting method. However, it performed better in capturing high-temperature scenarios. For precipitation, both schemes showed comparable correlation, with no significant difference in performance. Air pressure emerged as the best-simulated variable under both schemes, with extremely high levels of agreement. The dynamic weighting scheme achieved an R^2 value of 0.92, effectively capturing the variability in air pressure. In the case of relative humidity, the dynamic weighting scheme outperformed the equal-weighting approach, with R^2 values of 0.86 and 0.82, respectively. The simulation of air temperature further highlighted the advantage of the dynamic weighting approach, which yielded a substantially higher correlation and an R^2 of 0.85. For wind variables, the dynamic weighting scheme also

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

demonstrated superior performance. The R^2 values for zonal and meridional wind were 0.54 and 0.44, respectively, compared to 0.11 and 0.36 under the equal-weighting scheme. Notably, the equal-weighting method exhibited a relatively large deviation in simulating zonal wind. Overall, the optimal parameterization scheme derived from the dynamic weighting method more accurately simulated the variability of meteorological elements in the coastal region of Southeast than the scheme obtained through the equal-weighting approach.

3.5 Comparison of dynamic weight and equal weight methods for parameterization schemes in extreme climate years

To evaluate the robustness of the dynamic weighting method, this study applied the optimal parameterization schemes, derived from the year most representative of the climatological mean for the arid region of Northwest China and the southeastern coastal region, to a 2010 validation period. The aim is to assess the robustness and adaptability of each scheme under differing climatic scenarios.

Figure 8 compares the simulation performance of different meteorological variables using both weighting methods for the year 2010. Figure 8(a) illustrates the comparison for the arid region of Northwest. The results indicated that the dynamic weighting scheme yielded superior performance relative to the equal-weighting approach. Specifically, the dynamic weighting method achieved a lower RMSE (1.372) compared to the equal-weighting method (1.437), and a lower MIEI value (1.310 vs. 1.313), reflecting greater simulation accuracy under extreme climatic conditions.

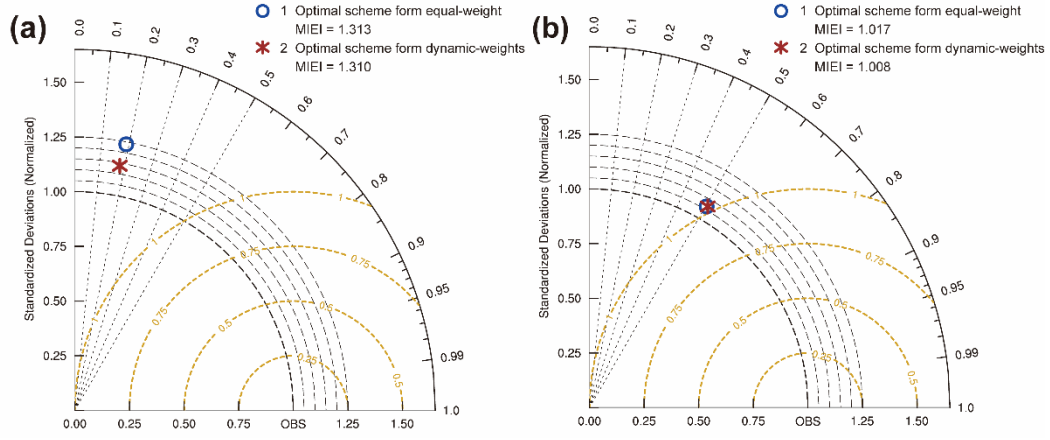


Figure 8. Comparison of the optimal parameterization schemes obtained via the dynamic weight and equal weight methods in the 2010 simulation results. (a) the arid region of Northwest; (b) the coastal region of Southeast.

Figure 8(b) shows the performance of the two schemes in the coastal region of Southeast. In this area, the simulation effect of the dynamic weight scheme had a higher correlation coefficient R (0.508; that of the equal weight method was 0.504) and a lower RMSE (1.028; that of the equal weight method was 1.030), with an MIEI value of 1.008, which was lower than the MIEI of 1.017 for the equal-weight scheme. This result indicated that the dynamic weight method had greater simulation accuracy when simulating meteorological variables in years with extreme climates and could better reflect meteorological changes under extreme climate conditions.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

4 Discussion

The dynamic weighting method proposed in this study—based on the average relative error and the coefficient of variation of the absolute error across different output variables—introduces a more refined mechanism for assigning weights in the evaluation of multiple physical processes. This approach substantially enhances the accuracy of the WRF model in multivariable meteorological simulations. As demonstrated in Figs. 2 and 5, the dynamic weighting method identifies precipitation and wind speed as key constraining variables in two climatically distinct regions. This finding aligns with previous evaluation studies that have highlighted precipitation as a particularly challenging target variable (Yu et al., 2022b; Wang and Tan, 2023). This suggests that despite the different prevailing climate backgrounds, the WRF model may face common physical process challenges in accurately simulating precipitation associated with local strong convection and the near-surface wind field related to surface friction. In multivariable simulation studies, as the number of variables increases, those that are more difficult to simulate are often underrepresented in the evaluation process. The dynamic weighting method addresses this issue by ensuring that critical and sensitive variables consistently receive appropriate weighting, thereby maintaining their influence and visibility throughout the model evaluation.

Through temporal cross-validation, this study confirms that the optimal combination of schemes selected by the dynamic weighting evaluation method exhibits robustness across different years. Furthermore, the successful application of this method in two climatically distinct areas—the arid region of Northwest China and the coastal region of Southeast China—verifies its potential for broad applicability across different climatic regions. This demonstrates that the method serves as a general paradigm broadly applicable to model evaluation in different regions. Under any specific regional climate background, it can effectively capture the key constraining variables of a simulation, thereby enabling a more precise assessment of parameterization schemes.

In multi-physics process simulations, the model's capacity to simulate various variables differs significantly, with the average relative error more indicative of its performance on specific variables than of discrepancies among the physical processes themselves. This implies that, while interactions between physical processes markedly influence the simulation outcomes, the disparities in the simulation capabilities of the current physical mechanisms remain relatively minor. Consequently, relative error serves as the primary criterion for determining weight distribution. As the model's physical mechanisms continue to advance, the simulation accuracy across different variables is expected to become more balanced (Pincus et al., 2008; Reichler and Kim, 2008). Therefore, distributing weights on the basis of the sensitivity of variables to physical processes will become increasingly important in future simulations.

The weights derived from the coefficient of variation of absolute errors did not exhibit significant regional differences, indicating that the sensitivity weighting of variables is primarily influenced by the physical mechanisms underlying the parameterization schemes. For instance, the pronounced sensitivity of precipitation to microphysics and cumulus processes is consistent with findings from previous studies (Tian et al., 2017; Liu et al., 2018). Similarly, the strong correlation between

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

surface temperature and land surface processes further supports the validity of this method in effectively capturing the sensitivity relationships between physical processes and meteorological variables (Liu et al., 2023).

The WRF model offers many parameterization schemes that can be freely combined. However, this high degree of flexibility necessitates substantial computational resources to support the model's operation (Li et al., 2014). The piecewise simulation method employed in this study, which is based on the sensitivity of individual physical processes, effectively simplifies the complex multiplicative combinations of physical process parameterization schemes into additive combinations, thereby reducing the computational burden. However, in multivariate evaluations, the intricate interactions between physical processes and variables introduce challenges related to uncertainty (Zeng et al., 2016; Jia et al., 2023). To address this, the dynamic weighting method dynamically adjusts variable weights during each stage of physical process simulation, ensuring that each process emphasizes the simulation of key elements most influential to specific variables. This approach helps to mitigate the uncertainty stemming from inter-process interactions caused by changes in parameterization schemes.

In existing studies, the common method of sequentially selecting the optimal parameterization scheme involves bringing the optimal parameterization scheme from each physical process into the simulation of the next physical process. For instance, as illustrated in Fig. 3(a), the selection of the YSU planetary boundary layer scheme leads to improved wind speed simulation in subsequent processes, suggesting that the sequence in which physical processes are simulated can also introduce uncertainty. Throughout the evaluation process, such uncertainty is subject to continuous propagation and accumulation. Consequently, the resulting set of parameterization schemes is not one that is globally optimal, but rather one that is preferentially optimized for the initial scheme selection. However, the current literature lacks a robust theoretical or experimental basis for the prescribed simulation sequence. Therefore, a necessary and significant direction for future research is to quantify and mitigate the uncertainties arising from both the simulation sequence and the initial choice of parameterization schemes.

Notably, the dynamic weighting method proposed in this study is not only applicable to parameterization scheme evaluation in the WRF model but also has broad applicability, especially in complex climate simulations and multi-model assessments. In comprehensive evaluations of complex climate models, such as multi-model ensembles (e.g., CMIP6), the dynamic adjustment of each variable's weight via two indicators on the basis of all model data structures—the average relative error and the coefficient of variation of absolute error—effectively reduces errors caused by differences in model structure and parameters (Knutti et al., 2010). This characteristic enables the dynamic weighting method to provide a more precise and objective framework for model performance evaluation in multi-model assessments.

5 Conclusion

This study proposes an optimization method for WRF model parameterization schemes based on dynamic weight evaluation, which improves the accuracy of multivariable climate simulations by dynamically adjusting the weights of each variable. Compared to traditional equal-weighting methods, the dynamic weighting approach offers a more accurate representation of overall model performance by accounting for simulation errors across different output variables and the sensitivity of these

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

variables to various physical processes. Validation in both the arid region of Northwest and the coastal region of Southeast confirms the effectiveness of the proposed method.

In the arid region of Northwest, the optimal parameterization scheme combination derived using the dynamic weighting method includes P3, YSU, Tiedtke, GFDL, and Noah LSM. This configuration significantly enhances the simulation accuracy of air temperature and wind speed compared to that achieved with the equal-weighting method. Similarly, in the coastal region of Southeast, the dynamic weighting method demonstrates its superiority, with the optimal combination comprising WDM6, Shin-Hong-MM5, Grell 3, GFDL-GFDL, and Noah LSM. This configuration yields more accurate temperature and wind speed, the dynamic weighting scheme provides more accurate results, improving the model's simulations, reflected in improved correlation coefficients and reduced RMSE values. Compared to the equal-weighting method, it more effectively captures the trends and magnitudes of meteorological variability. Nonetheless, despite notable improvements in the simulation of near-surface meteorological variables, certain biases persist—particularly in the simulation of precipitation and wind speed. This suggests that, while the dynamic weighting method substantially enhances simulation performance, further refinement of the model's physical mechanisms is necessary, especially for accurately representing complex weather systems and extreme events.

As the parameterization schemes for different physical processes in climate models continue to evolve and with the increasing frequency of extreme climate events, the need for rapid and accurate simulations and predictions of climate change becomes even more critical. The dynamic weight method offers a flexible and precise tool that can optimize the evaluation process on the basis of the performance of different variables and models, thereby better addressing the uncertainties and complexities of climate change. The widespread application of this method not only optimizes existing climate models but also provides more reliable scientific foundations for climate change response strategies and policy-making.

Code and data availability. The Weather Research and Forecasting (WRF) model version 4.3.3 used in this study is freely available online and can be downloaded from https://www2.mmm.ucar.edu/wrf/users/download/get_source.html (Skamarock et al., 2019). The code used in this study to calculate the dynamic weights is available at https://doi.org/10.5281/zenodo.17414002 (Gou et al., 2025). The Zenodo repository also provides the WRF model simulation results and the evaluation process. Any future updates or additional scripts will be hosted there.

Author contributions. TG and SK conceptualized the study. TG and YD designed the methodology. TG developed the model code, performed the simulations and drafted the manuscript. JN and SK supervised the work and contributed to the writing and editing of the paper.

Competing interests. The authors have no relevant financial or non-financial interests to disclose.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

Acknowledgments. The authors acknowledge NCAR for the WRF model and ECMWF for the ERA5 reanalysis datasets. The computational work for this article was partially performed on resources of the National Supercomputing Centre in Tianjin. We thank Dr. Zhongfeng Xu from Institute of Atmospheric Physics, Chinese Academy of Sciences who provided guidance for MIEI methods.

Financial support. This research was supported by the National Key Research and Development Program of China (No. 2022YFD1900503).

References

Baker, N. C. and Taylor, P. C.: A framework for evaluating climate model performance metrics, J. Clim., 29, 1773-1782, doi:10.1175/JCLI-D-15-0114.1, 2016.

Borge, R., Alexandrov, V., Del Vas, J. J., Lumbreras, J., and Rodríguez, E.: A comprehensive sensitivity analysis of the WRF model for air quality applications over the Iberian Peninsula, Atmos. Environ., 42, 8560-8574, doi:10.1016/j.atmosenv.2008.08.032, 2008.

Chen, F., Liu, C., Dudhia, J., and Chen, M.: A sensitivity study of high-resolution regional climate simulations to three land surface models over the western United States, J. Geophys. Res. Atmos., 119, 7271-7291, doi:10.1002/2014JD021827, 2014.

Dai, D., Chen, L., Ma, Z., and Xu, Z.: Evaluation of the WRF physics ensemble using a multivariable integrated evaluation approach over the Haihe river basin in northern China, Clim. Dyn., 57, 557-575, doi:10.1007/s00382-021-05723-x, 2021.

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., and Hoffman, F. M.: Taking climate model evaluation to the next level, Nat. Clim. Change, 9, 102-110, doi:10.1038/s41558-018-0355-y, 2019.

Fernández, J., Montávez, J. P., Sáenz, J., González-Rouco, J. F., and Zorita, E.: Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: Annual cycle, J. Geophys. Res. Atmos., 112, D04101, doi:10.1029/2005JD006649, 2007.

Ferreira, J. A., Carvalho, A. C., Carvalheiro, L., Rocha, A., and Castanheira, J. M.: On the influence of physical parameterisations and domains configuration in the simulation of an extreme precipitation event, Dyn. Atmos. Oceans, 68, 35-55, doi:10.1016/j.dynatmoce.2014.08.001, 2014.

Fu, J., Kang, S., Zhang, L., Li, X., Gentine, P., and Niu, J.: Amplified warming induced by large-scale application of water-saving techniques, Environ. Res. Lett., 17, 034018, doi:10.1088/1748-9326/ac4b52, 2022.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, J. Geophys. Res. Atmos., 113, D06104, doi:10.1029/2007JD008972, 2008.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

Gou, T., Deng, Y., Niu, J., and Kang, S.: An objective dynamic multivariable weighting method for reducing uncertainty in WRF parameterization scheme selection, Zenodo [code], https://doi.org/10.5281/zenodo.17414002, 2025.

Hausfather, Z., Drake, H. F., Abbott, T., and Schmidt, G. A.: Evaluating the performance of past climate model projections, Geophys. Res. Lett., 47, e2019GL085378, doi:10.1029/2019GL085378, 2020.

Jia, W., Zhang, X., Wang, H., Wang, Y., Wang, D., Zhong, J., Zhang, W., Zhang, L., Guo, L., and Lei, Y.: Comprehensive evaluation of typical planetary boundary layer (PBL) parameterization schemes in China–Part 1: Understanding expressiveness of schemes for different regions from the mechanism perspective, Geosci. Model Dev., 16, 6635-6670, doi:10.5194/gmd-16-6635-2023, 2023.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in combining projections from multiple climate models, J. Clim., 23, 2739-2758, doi:10.1175/2009JCLI3361.1, 2010.

Lambert, S. J. and Boer, G. J.: CMIP1 evaluation and intercomparison of coupled climate models, Clim. Dyn., 17, 83-106, doi:10.1007/PL00013736, 2001.

Li, L., Li, W., and Jin, J.: Improvements in WRF simulation skills of southeastern United States summer rainfall: physical parameterization and horizontal resolution, Clim. Dyn., 43, 2077-2091, doi:10.1007/s00382-013-2031-2, 2014.

Li, X., Wang, C., and Lan, J.: Role of the South China Sea in Southern China rainfall: Meridional moisture flux transport, Clim. Dyn., 56, 2551-2568, doi:10.1007/s00382-020-05603-w, 2021.

Liu, D., Yang, B., Zhang, Y., Qian, Y., Huang, A., Zhou, Y., and Zhang, L.: Combined impacts of convection and microphysics parameterizations on the simulations of precipitation and cloud properties over Asia, Atmos. Res., 212, 172-185, doi:10.1016/j.atmosres.2018.05.017, 2018.

Liu, L., Zhang, X., Han, C., and Ma, Y.: A performance evaluation of various physics schemes on the predictions of precipitation and temperature over the Tibet Autonomous Region of China, Atmos. Res., 292, 106878, doi:10.1016/j.atmosres.2023.106878, 2023.

Lu, S., Guo, W., Xue, Y., Huang, F., and Ge, J.: Simulation of summer climate over Central Asia shows high sensitivity to different land surface schemes in WRF, Clim. Dyn., 57, 2249-2268, doi:10.1007/s00382-021-05876-9, 2021.

Lv, M., Xu, Z., and Yang, Z.: Cloud resolving WRF simulations of precipitation and soil moisture over the central Tibetan Plateau: An assessment of various physics options, Earth Space Sci., 7, e2019EA000865, doi:10.1029/2019EA000865, 2020.

Mooney, P. A., Mulligan, F. J., and Fealy, R.: 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, doi:10.1175/JCLI-D-11-00676.1, 2013.

Pincus, R., Batstone, C. P., Hofmann, R. J. P., Taylor, K. E., and Glecker, P. J.: Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models, J. Geophys. Res. Atmos., 113, D14209, doi:10.1029/2007JD009334, 2008.

Reichler, T. and Kim, J.: How well do coupled models simulate today's climate?, Bull. Am. Meteorol. Soc., 89, 303-312, doi:10.1175/BAMS-89-3-303, 2008.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat, F.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, doi:10.1038/s41586-019-0912-1, 2019.

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., Vegas-Regidor, J., Bock, L., Brötz, B., and de Mora, L.: Earth System model evaluation tool (ESMValTool) v2. 0–technical overview, Geosci. Model Dev., 13, 1179-1199, doi:10.5194/gmd-13-1179-2020, 2020.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D. O., Liu, Z., Berner, A. J., Wang, A. W., Powers, A. J. G., Duda, A. M. G., Barker, A. D., and Huang, A. X.-Y.: A Description of the Advanced Research WRF Model Version 4.3, NCAR [code], https://doi.org/10.5065/1dfh-6p97, 2019.

Stergiou, I., Tagaris, E., and Sotiropoulou, R.-E. P.: Sensitivity assessment of WRF parameterizations over Europe, Proceedings, 1, 119, doi:10.3390/ecas2017-04138, 2017.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res. Atmos., 106, 7183-7192, doi:10.1029/2000JD900719, 2001.

Tian, J., Liu, J., Yan, D., Li, C., and Yu, F.: Numerical rainfall simulation with different spatial and temporal evenness by using a WRF multiphysics ensemble, Nat. Hazards Earth Syst. Sci., 17, 563-579, doi:10.5194/nhess-17-563-2017, 2017.

Wang, C., Qian, Y., Duan, Q., Huang, M., Yang, Z., Berg, L. K., Gustafson Jr, W. I., Feng, Z., Liu, J., and Quan, J.: Quantifying physical parameterization uncertainties associated with land-atmosphere interactions in the WRF model over Amazon, Atmos. Res., 262, 105761, doi:10.1016/j.atmosres.2021.105761, 2021.

Wang, X. and Tan, Z.: On the combination of physical parameterization schemes for tropical cyclone track and intensity forecasts in the context of uncertainty, J. Adv. Model. Earth Syst., 15, e2022MS003381, doi:10.1029/2022MS003381, 2023.

Xu, Z., Han, Y., and Fu, C.: Multivariable integrated evaluation of model performance with the vector field evaluation diagram, Geosci. Model Dev., 10, 3805-3820, doi:10.5194/gmd-10-3805-2017, 2017.

Xu, Z., Hou, Z., Han, Y., and Guo, W.: A diagram for evaluating multiple aspects of model performance in simulating vector fields, Geosci. Model Dev., 9, 4365-4380, doi:10.5194/gmd-9-4365-2016, 2016.

Yang, B., Zhang, Y., Qian, Y., Huang, A., and Yan, H.: Calibration of a convective parameterization scheme in the WRF model and its impact on the simulation of East Asian summer monsoon precipitation, Clim. Dyn., 44, 1661-1684, doi:10.1007/s00382-014-2118-4, 2015.

Yu, E., Ma, J., and Sun, J.: Developing a climate prediction system over Southwest China using the 8-km weather research and forecasting (WRF) model: system design, model calibration, and performance evaluation, Weather Forecast., 37, 1703-1719, doi:10.1175/WAF-D-21-0188.1, 2022a.

Yu, E., Bai, R., Chen, X., and Shao, L.: Impact of physical parameterizations on wind simulation with WRF V3.9.1.1 under stable conditions at planetary boundary layer gray-zone resolution: a case study over the coastal regions of North China, Geosci. Model Dev., 15, 8111–8134, doi:10.5194/gmd-15-8111-2022, 2022b.

https://doi.org/10.5194/egusphere-2025-5362
Preprint. Discussion started: 19 December 2025

Zeng, X., Wang, B., Zhang, Y., Zheng, Y., Wang, N., Wang, M., Yi, X., Chen, C., Zhou, Z., and Liu, H.: Effects of land surface schemes on WRF-simulated geopotential heights over China in summer 2003, J. Hydrometeorol., 17, 829-851, doi:10.1175/JHM-D-14-0239.1, 2016.

Zhang, M., Xu, Z., Han, Y., and Guo, W.: An improved multivariable integrated evaluation method and tool (MVIETool) v1.
725 0 for multimodel intercomparison, Geosci. Model Dev., 14, 3079-3094, doi:10.5194/gmd-14-3079-2021, 2021.

Zhang, X., Zhao, T., Xu, H., Liu, W., Wang, J., Chen, X., and Liu, L.: GLC_FCS30D: The first global 30-m land-cover dynamic monitoring product with a fine classification system from 1985 to 2022 using dense time-series Landsat imagery and continuous change-detection method, Earth Syst. Sci. Data Discuss., 16, 1353–1381, doi:10.5194/essd-16-1353-2024, 2024.

730