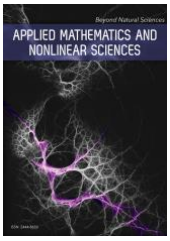
sciendo Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-17



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Ensemble numerical weather prediction model to improve the efficiency of Henan parameterization scheme

Xiaojiu Ma[1,†], Jinggang Wang[1], Zhe Liu[1], Bo Liu[1], Liang Yuan[1], Pei Pei[1], Xin Sun[2]

1. State Grid Henan Company, Zhengzhou, Henan, 450000, China.
2. State Grid Henan Electric Power Research Institute, Zhengzhou, Henan, 450000, China.

Submission Info

Communicated by Z. Sabir
Received April 28, 2024
Accepted August 21, 2024
Available online October 9, 2024

Abstract

China has highly emphasized the research and operational application of numerical weather prediction. This paper determines the objective function parameters, such as CAPE and SRH, to apply an ensemble numerical prediction model in weather forecasting. Preprocessing and evaluating rainfall data is necessary to construct the WRF-ARW numerical weather prediction model. The WRF-ARW model is applied to simulate the weather forecasts in Henan Province, and the difficulties and challenges faced in the efficient implementation of the parameterized scheme are outlined. The WRF-ARW model's prediction errors for the maximum rainfall and total rainfall in Henan Province range from 1.78%-13.51% and 0.16%-3.78%, respectively, which are significantly less than 15%, and the model is more predictive than the others. The raw data test set's credibility ranges from 0.957 to 0.997, which is close to 1, indicating that the raw data collected in this paper are highly credible. The WRF-ARW model's qualification rates for forecasting maximum rainfall and total rainfall are 86.7% and 93.3%, respectively, and its overall accuracy is grade B and grade A, respectively. The pass rates for the peak occurrence time of maximum rainfall and total rainfall were 93.3% and 86.7%, respectively, and the overall prediction accuracy was Grade A and Grade B, respectively. The WRF-ARW model is effective in weather forecasting throughout Henan Province. In summary, the WRF-ARW model is very effective in improving the efficiency of ensemble numerical weather prediction and parameterization schemes in Henan Province.

Keywords: WRF-ARW model; CAPE; SRH; Ensemble numerical; Weather prediction; Parameterized scheme.
AMS 2010 codes: 97P10

†Corresponding author.
Email address: hnzcqyc@163.com

ISSN 2444-8656
https://doi.org/ 10.2478/amns-2024-2813

1 Introduction

Numerical weather prediction refers to the method of predicting the weather phenomena in a certain period in the future by solving the numerical calculations depicting the weather evolution process through a large-scale computer based on the actual atmospheric conditions and under certain conditions of initial and marginal values. The foresight period of numerical weather prediction is generally able to reach 1~15 days or even longer [1-2]. Henan Province is located in the Central Plains. The terrain is extremely complex: the northwestern part of the Taihang Mountains, the western part of the Zhongtiao Mountains, the Banshan Mountains, the southwestern part of the Xiong'er Mountains, the outer Fangshan Mountains, the FuNiu Mountains, the central part of the Songshan Mountains, MangShan Mountains, the southern part of the TongBaShan Mountains, the Dabie Mountains, and so on, and in the east of the region for the plains, the central part of the north of the Yellow River running across the east and west. These complex topographies have a great influence on the boundary layer meteorological elements, especially on the distribution of wind fields [3]. Every year in summer, many strong convective weather occurs in Henan, such as thunderstorms and hailstorms over a wide area of Henan, gales and hailstorms in Zhengzhou, and shipwrecks caused by gales in Xiaolangdi, Jiyuan, all of which caused great losses to people's lives and properties [4]. Although every strong convective weather has its large-scale weather background, the influence of small- and medium-scale systems and topography is not negligible, which increases the difficulty of strong convective forecasting in our practical work [5]. Therefore, in order to improve the timeliness and accuracy of the ensemble numerical weather prediction model in Henan and to provide some reference ideas for the forecasters, it is of great significance to study the effect of the efficiency of parameterization scheme on the strong convective weather in Henan [6-8].

In order to improve the accuracy of weather forecasting, people have never stopped on the road of exploring weather forecasting techniques. Murray, S.A stated that as the space weather community has been exploring and researching more and more deeply on the application of emerging technologies and combining them with the traditional body of knowledge established based on the physical principles, which has achieved remarkable results in operational forecasting work, the application of ensemble techniques in space weather forecasting operations will continue to push The application of ensemble technology in space weather forecasting operations will continue to promote the development of operational forecasting work to a higher level [9]. Ren, X et al. pointed out that the new idea of "deep learning-based" brings infinite possibilities for future development, which can help people better understand the subtle connections behind complex systems, can also provide more accurate and reliable weather regression analysis results, and most importantly, will promote the rapid development of the entire industry in the direction of intelligence and autonomy, and further meet people's growing demand for high-quality services [10]. Wang, T et al.'s method of multi-model ensemble prediction based on a convolutional neural network and feed-forward neural network is very effective for improving the temperature prediction in Henan Province, and this study provides an important reference for the future in-depth understanding of the SAT weather in Henan Province in the next few tens of hours and provides reliable support for the application in related fields [11]. Zheng et al. emphasized that in order to improve the prediction ability of extreme precipitation events, it is necessary to strengthen the research and improvement of the interaction mechanism between the atmospheric circulation systems within each model, and only by more accurately grasping the complex relationship between various meteorological elements and incorporating them into a more complete and reliable numerical weather prediction system can we effectively deal with the problem of extreme weather events such as heavy rainfall that may have catastrophic impacts [12].

Nowadays, ensemble forecasting is a widely used numerical forecasting technique, particularly in Chinese numerical weather prediction research. In this paper, a brief overview of the application of ensemble numerical modeling in weather forecasting is first presented, and then the algorithms for

determining specific parameters, such as CAPE and SRH, are introduced to provide support for the construction of the objective function of the WRF-ARW numerical weather prediction model. The pre-processing and evaluation methods for the collected raw rainfall datasets are briefly discussed. The WRF-ARW model is utilized to simulate the weather prediction in Henan Province, and the difficulties and challenges that come with implementing the parameterized scheme are discussed. It is expected to guide future development trends of ensemble numerical computation in the efficient implementation of parameterized schemes for weather forecasting in Henan Province.

2 Method

2.1 Application of Ensemble Numerical Prediction Models to Weather Forecasting

2.1.1 Application in Data Assimilation

Numerical weather prediction systems provide important information on the atmosphere, oceans, and land surface by extensively assimilating a large number of observations from the ground, satellites, and ships with the help of data assimilation techniques. With the improvement of the quality of observation data, especially the increasing number of observation data from satellites, as well as the continuous upgrading of Earth system models and data assimilation techniques, the level of modern medium-term weather prediction has been significantly enhanced. Data-driven methods, represented by deep learning, have become a hot topic in weather forecast data assimilation.

2.1.2 Applications in the development of numerical refinement

The evolution of numerical weather prediction models continues, with significant improvements in model resolution. For subgrid processes that cannot be resolved by models, such as small-scale radiation, convection, and diffusion processes within the atmospheric system and with the outside world, researchers have developed various parameterization schemes to describe the heat and momentum gain and loss of subgrid processes to improve the quality of numerical forecasts. The deep neural network parameterization schemes driven by big data can significantly improve model forecasts and computational efficiency.

2.1.3 Applications in target observation

The ensemble numerical prediction method is able to mine the hidden high-dimensional information from the massive simulation data of numerical weather prediction so as to construct the numerical weather prediction product bias characteristics, which opens up a new direction for the numerical weather prediction bias revision research and business practice. Supported by enough training data, the results of the ensemble numerical prediction model are generally better than traditional statistical post-processing methods. It is foreseeable that the quality of forecast products, and thus the early warning capability for catastrophic weather and climate, will be significantly improved by continuously improving the neural network model as well as the optimization algorithm, driven by the meteorological, oceanic big data.

2.1.4 Applications in Probabilistic Forecasting

By reanalyzing the data or observations for training, data-driven models can avoid the limitations present in numerical forecasting models. Large ensembles help improve the prediction level of extreme weather events in both short-term and long-term forecasts. The Earth system, as a complex

system, is governed by physical laws at different spatial and temporal scales. The numerical solution of partial differential equations (PDEs) using finite difference, finite element, and spectral methods has been a great success in interfacing multi-scale physical processes. Through the in-depth study of the artificial intelligence framework with embedded physical constraints, a new generation of data-driven and physically-driven numerical weather prediction systems are developed, which can ensure the accuracy, real-time, and reliability requirements of numerical prediction products.

2.2 WRF-ARW numerical weather prediction model construction

The WRF model [13-14] was developed as a mesoscale numerical weather prediction system for atmospheric research and operational forecasting applications. Two main dynamic cores are present: the data assimilation system and the software architecture. The physical range of the model is also extremely wide, from tens of meters in the low-altitude phase to thousands of kilometers in the high-altitude phase for effective meteorological applications. The WRF model can be simulated under real natural atmospheric conditions, such as observation data and analysis data, or under idealized conditions for meteorological simulations. In this paper, we mainly use the WRF-ARW of the WRF model [15-16].

In accordance with Laprise's method, the following derivation process can be carried out firstly for the following equations: take the coordinates of the terrain perpendicular to the static air pressure, take η as the perpendicular mass coordinates, and take σ as the corrected surface coordinates corresponding to it, which have the following form:

$$\eta = (p_h - p_{ht}) / \mu \tag{1}$$

$\mu = p_{hs} - p_{ht}$ here, where Ph is the component of the hydrostatic equilibrium of the air pressure, p_{ht} is the air pressure at the top of the boundary, and p_{hs} is the air pressure at the surface of the terrain. At (x, y) grid points within the model area, $u(x, y)$ is the mass of the air column per unit horizontal area at the corresponding grid point, so for the conservative quantities above, we can express them in the form of flux approximations:

$$\bar{V} = \mu_{\bar{V}} = (U, V, W), \Omega = \mu\eta, \Theta = \mu\theta \tag{2}$$

Next, we define the deviation of the perturbation quantity with respect to the reference state to obtain:

$$p = \bar{p}(z) + p' \tag{3}$$

$$\varphi = \bar{\varphi}(z) + \varphi' \tag{4}$$

$$\alpha = \bar{\alpha}(z) + \alpha' \tag{5}$$

$$\mu = \bar{\mu}(z) + \mu' \tag{6}$$

In general, the coordinate plane in η-coordinates has a non-horizontal character, so the reference state quantity $\bar{p}, \bar{\varphi}, \bar{\alpha}$ therein is usually expressed as a function of (x, y, η). We can directly refer to the above perturbation quantities to eliminate the static equilibrium part within the equation, and then applying the above-conserved quantities associated with the reference state deviations, we obtain the Laprise system of equations of the following form, i.e.:

$$\frac{\partial U}{\partial t}+(\nabla\circ\bar{W})_{\eta}+\mu_{\alpha}\frac{\partial p'}{\partial x}+(\eta\mu\frac{\partial\bar{\mu}}{\partial x})\alpha'+\mu\frac{\partial\varphi'}{\partial x}+\frac{\partial\varphi}{\partial x}(\frac{\partial p'}{\partial\eta}-\mu')=F_{U} \tag{7}$$

$$\frac{\partial V}{\partial t}+(\nabla\circ\bar{v}V)_{\eta}+\mu_{\alpha}\frac{\partial p'}{\partial y}+(\eta\mu\frac{\partial\bar{\mu}}{\partial y})\alpha'+\mu\frac{\partial\varphi'}{\partial y}+\frac{\partial\varphi}{\partial y}(\frac{\partial p'}{\partial\eta}-\mu')=F_{V} \tag{8}$$

$$\frac{\partial W}{\partial t}+(\nabla\circ\bar{v}W)_{\eta}-g(\frac{\partial p'}{\partial\eta}-\mu')=F_{W} \tag{9}$$

$$\frac{\partial\Theta}{\partial t}+(\nabla^{\circ}\bar{v}\Theta)_{\eta}=F_{\Theta} \tag{10}$$

$$\frac{\partial\mu}{\partial t}+(\nabla^{\circ}\bar{W})_{\eta}=0 \tag{11}$$

$$\frac{\partial\varphi}{\partial t}+(\nabla^{\circ}\bar{v}\Phi)_{\eta}=gw \tag{12}$$

After obtaining the above system of Laprise equations in the form of flux forecasts, the following static pressure diagnostic relations need to be satisfied:

$$\frac{\partial\varphi'}{\partial\eta}=-\bar{\mu}\alpha'-\alpha\eta' \tag{13}$$

and the gas equation of the state

$$p=\left(\frac{R\Theta}{p_{0}\mu_{\alpha}}\right)^{r} \tag{14}$$

Here g is the gravitational acceleration, the Possion index r takes the value of $r=c_{p}/c_{v}=1.4$, and the other parameters are meteorologically generalized quantities.

The five key parameters in the new KFCPS are related to the rate of descending air flux and onset height, the ambient entrained flux rate, the turbulent kinetic energy (TKE) in the subcloud layer, and the time of depletion of the convective available potential energy (CAPE). In WRF, the CPS of KF is considered important, but the range of possible values is quite wide.

The strengths of both the downward and entrained fluxes are proportional to the mass flux of the upward flux at the top of the USL in the KF. In this study, two parameters, P_{d} and P_{e}, are defined as additional scaling factors that regulate the rates of the downwelling and entrainment fluxes, respectively, to 1/2 to 2 times their original values.

$$\frac{M_{d}^{USL}}{M_{u}^{USL}}=2\times(1-RH)\times2^{P_{d}},P_{d}\in(-1,1) \tag{15}$$

$$\frac{\delta M_{e}}{M_{u}^{USL}}=\frac{-0.03\times\delta p}{R}\times2^{P_{e}},P_{e}\in(-1,1) \tag{16}$$

In Eqs. M_u^{USL} and M_u^{USL} are the rising and falling mass fluxes at the top of USL, respectively. RH is the average relative humidity of the ambient air from the starting layer of the descending airflow to the bottom of the cloud. R is the cloud radius, δp is the pressure thickness of the modeled layer, and δM_e is the maximum possible entrainment rate of the layer.

In criterion KF, it is assumed that the downdraft starts from $150hPa$ above USL. The starting height of the downdraft P_h controls the structure of the downdraft and also influences the atmospheric properties in the subclouds. We set the range of P% to $50-350hPa$ to allow for a greater degree of freedom in the downdraft structure from larger and narrower to shorter and wider.

The TKE and average CAPE consumption times are referred to as P_t and P_c, and take values of $5m^2s^{-2}$ and $2700s$ in the standard KF. The range of intervals for values taken for P_t is $[3m^2s^{-2},12m^2s^{-2}]$, and the range of intervals for values taken for P_c is $[900s,7200s]$.

2.3 Environmental instability parameters of the WRF-ARW model

2.3.1 Convective Effective Potential Energy CAPE

The area enclosed by the laminar curve and the state curve between the upper free convection height (LFC) and the equilibrium height (EL) is called the positive area (PA), which is proportional to the magnitude of the kinetic energy generated by the positive buoyancy force between the LFC and the EL. The positive area on the thermodynamic diagram is currently referred to as the daily convective effective potential energy CAPE [17-18], defined in the following equation:

$$CAPE = g\int_{Z_{LFC}}^{Z_{EL}} (\frac{T_{vp}-T_{ve}}{T_{ve}})dz \tag{17}$$

Where T_v - imaginary temperature, e, p denotes physical quantities related to the environment and to the gas block, respectively: Z_{LFC} - free convection height, which is the height at which $T_{vp}-T_{ve}$ turns from a negative to a positive value; Z_{EL} - equilibrium height, which is the height at which $T_{vp}-T_{ve}$ the height from positive to negative values. Geometrically, CAPE is proportional to the positive area on a thermodynamic diagram (e.g., T-LnP diagram), characterizing the unstable energy of strong convection in the atmosphere.

1) Steps for calculating convective effective potential energy CAPE

   (1) Select the horizontal coordinate as temperature and the vertical coordinate as logarithmic air pressure.

   (2) Temperature, dew point laminar curve production: the vertical layer of airborne data temperature, air pressure, and dew point temperature data, coordinates are connected to each point in turn, the temperature, dew point laminar curve can be obtained.

   (3) State curve production: after selecting the starting lifting point, according to the dry adiabatic process to pick up the condensation height (T_L, P_L).

$$T_L = \frac{2840}{3.5\ln T_k - \ln e - 4.805} + 55 \tag{18}$$

$$P_L = P_0 \left(\frac{273 + T_k}{273 + T_L}\right)^{C_{\mu d}/R_d} \tag{19}$$

where T_k, P_0 are the starting point temperature and air pressure. According to the principle of conservation of potential temperature θ in a dry adiabatic process, the corresponding temperature values are calculated by iteration. The formula for calculating the potential temperature θ is:

$$\theta = T_k \left(\frac{1000}{P_0}\right)^{0.2854(1-0.28\times10^{-3} r)} \tag{20}$$

Where T_k is the absolute temperature at the starting point, T is the temperature at the pick-up condensation height, P_0 is the air pressure at the starting point, r is the mixing ratio of the gas block, and e is the water vapor pressure at the pick-up condensation height. After the gas block reaches the lifting condensation height, according to the principle of pseudo-equivalent potential temperature conservation, use the iterative or bifurcation method to calculate the state curve, pseudo-equivalent potential temperature θ_{se} calculation formula for:

$$\theta_{se} = \theta \exp\left[\left(\frac{3.376}{T_L} - 0.00254\right) \times r(1 + 0.81\times10^{-3} r)\right] \tag{21}$$

(4) Compare the lamination curve with the state curve, and when the temperature of the gas block is greater than the ambient temperature, the unstable energy integration starts and ends when the temperature of the gas block is less than the ambient temperature. The value obtained from the integration is the size of CAPE.

2) Convective effective potential energy calculated by density temperature

The density temperature T_ρ is a widely used quantity in recent years in the theory of wet-atmosphere convection. Similar to the process of introducing imaginary temperature T_v, the density temperature T_ρ is defined as:

$$T_\rho = T \frac{1 + \gamma/\varepsilon}{1 + \gamma_\tau} \tag{22}$$

Where T is the temperature, γ_τ is the mixing ratio of the water substance, $\gamma_\tau = \gamma + \gamma_i + \gamma_i$, $\gamma, \gamma_i, \gamma_i$ are the mixing ratios of water vapor, liquid water, and ice respectively. ε is the ratio of dry air gas constant Rd to water vapor gas constant Rv. When there is no water condensate, $\gamma_\tau = \gamma$, then:

$$T_\rho = T \frac{1 + \gamma/\varepsilon}{1 + \gamma} \cong T \cdot (1 + 0.608\gamma) = T_v \tag{23}$$

T_v is a special case of T_ρ. Density temperature T_ρ can be regarded as the temperature at which air should be when the density of dry air equals the density of wet air and air containing solid water at the same pressure. The introduction of density temperature T_ρ then reflects the temperature of a multiphase system, i.e. $\alpha = R_d T_\rho / p$.

In the atmospheric wet process, we consider T_ρ instead of T_v to calculate the convective effective potential energy, which gives CAPE as:

$$CAPE = g\int_{z_\mu}^{z_\mu} \frac{1}{T_\mu}(T_\mu - T_\mu)dz \tag{24}$$

2.3.2 Parameters related to wind shear

Helicity is a parameter used to measure the strength of the incoming flow of a t-storm as well as the horizontal vorticity component along the direction of the incoming flow. Helicity is the volume fraction of the dot product of the wind speed vector and the vorticity vector, expressed as:

$$H_t = \iiint_r \vec{V} \cdot \nabla \times \vec{V} d\tau \tag{25}$$

The dot product of the wind speed vector and the vorticity vector is called the local helicity, denoted as:

$$H_D = \bar{V} \cdot \nabla \times \bar{V} \tag{26}$$

$\bar{V}$ in the above two equations is the three-dimensional wind speed. Defined as:

$$H_{s-r-T} = \int_0^h (\vec{V} - \vec{C}) \cdot \vec{\omega} dz \tag{27}$$

The storm-relative helicity size unit is m^2s^{-2}. The average storm-relative helicity) is simply a high average of the total helicity. The above equations can be rewritten as formulas based on the properties of analytic geometry:

$$H_{s-r-T} = \sum_{n=0}^{N-1}[(u_{n+1} - c_x)(v_n - c_y) - (u_n - c_x)(v_{n+1} - c_y)] \tag{28}$$

In this case, the storm speed was determined by selecting the mass-weighted average wind in the 850 hPa to 400 hPa air layer, with the wind deflected 30° to the right and 75% of the magnitude of the wind speed as the storm speed at that point.

2.4 Precipitation data assessment and fusion methods

2.4.1 Precipitation Data Assessment Methods

Before fusing the numerical forecast data, the precipitation data needs to be evaluated as a way to judge the reliability of that precipitation data. In this paper, relative deviation, mean absolute error, correlation coefficient, and root mean square error are selected, which are defined as follows:

$$RelativeBias = 1 - \frac{\sum_{i=1}^{n} y_{pred}^{i}}{\sum_{i=1}^{n} y_{obs}^{i}} \tag{29}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_{pred}^{i} - \bar{y}| \tag{30}$$

$$\mathrm{CC} = 1 - \frac{\Sigma_{i=1}^{n}\left(y_{pred}^{i} - y_{obs}^{i}\right)^{2}}{\Sigma_{i=1}^{n}\left(y_{obs}^{i} - \bar{y}\right)} \tag{31}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{pred}^{i} - y_{obs}^{i}\right)^{2}}{n}} \tag{32}$$

Where n is the number of observations at each grid point, y_{pred}^{i} represents the predicted precipitation for that grid, y_{obs}^{i} represents the measured observed precipitation for that grid, y_{pred}^{-} represents the mean value of the predicted precipitation for that grid, and y_{obs}^{-} represents the mean value of the measured observed precipitation for that grid. Y_{pred}, Y_{obs} represents the forecast precipitation sequence and observed precipitation sequence, respectively.

The principle of the empirical orthogonal function analysis method is to decompose the variable field that changes with time into a spatial function that does not change with time and a temporal function that only depends on the change of time, and its expression is as follows:

$$X = EOF_{m\times m} \times PC_{m\times n} \tag{33}$$

Where: m is the number of spatial lattice points; n is the length of the time series; EOF is the spatial modality, which mainly expresses the spatial distribution characteristics of the element; PC is the time-related part, reflecting the change of each spatial modality over time. Whether the spatial modality obtained in the actual data is random, the North test is needed to determine how many modes the data set needs to be divided into before conducting the EOF analysis. North et al. pointed out in their study that the $EOF_{m\times m}$-matrix eigenroot error at the 95% confidence level is:

$$\Delta\gamma = \gamma\sqrt{\frac{2}{N}} \tag{34}$$

γ is the characteristic root and N is the effective degree of freedom of the data. Check γ in sequential order by labeling the margin of error. If there is an overlap of error ranges between the two γ before and after, the test of significance has not been passed. Determine the modality of the data set based on the significance test.

2.4.2 Precipitation Data Fusion Methods

For the fused data, the data were extracted with a resolution of $0.1°\times0.1°$ and a latitude/longitude range of $28°\sim29°30'N$, $120°30'\sim122°$. The data were converted to decimals in the interval (0, 1) using the min-max normalization method, and finally, the daily 6-hour cumulative data were sorted by date.

$$x_{normalization}=\left(\frac{x-Min}{Max-Min}\right) \tag{35}$$

Where, $x_{normalization}$ is the value after the data normalization process, x is the raw precipitation data, Min is the minimum value of the whole data set and Max is the maximum value of the whole data set. The downscaling method used is Inverse Distance Weight Interpolation (IDW). The principle of IDW is shown below:

$$D_i=\sqrt{(x_i-A)^2+(y_i-B)^2} \tag{36}$$

$$Z_{(A,B)}=\frac{\Sigma_{i=1}^{N}(\frac{Z_i}{D_i})^2}{\Sigma_{i=1}^{N}(\frac{1}{D_i})^2} \tag{37}$$

Where D_i is the distance from (x_i,y_i) to grid point (A,B) and $Z_{(A,B)}$ is the estimate on grid point (A,B), i.e., the estimate obtained by accumulating the nearest N grid points to point (A,B) according to the weight of the distance. In this paper, take $N=4$ for calculation.

3 Results and discussion

3.1 Analysis of simulation results of the WRF-ARW model

3.1.1 Parameter Optimization and Accuracy Grading

In the process of parameter rate determination, the selection of the objective function is involved. The objective function is used to reflect the degree of agreement between the measured weather process and the simulated process, and the purpose of the automatic parameter rate determination is to find the parameter value that optimizes the value of the objective function, so the selection of the objective function is extremely important. The deterministic coefficient is used as the objective function for parameter optimization, and the WRF-ARW algorithm is used as the algorithm for parameter optimization to rate the model to obtain the optimal parameter combinations for the study basin.

According to the evaluation standard of weather forecasting specification, three evaluation items of maximum precipitation, total precipitation and precipitation process are selected to evaluate the simulation effect of rainy days. Three error indexes of relative error of maximum precipitation, relative error of total precipitation and certainty coefficient are selected to analyze comprehensively. The permissible error of total precipitation is proposed to be 10%, and that of maximum precipitation is proposed to be 15%. The qualified evaluation standard is that the relative error is less than the

permitted error, and the overall accuracy level of the simulation can be graded according to the size of the qualified rate or the certainty coefficient, as shown in Table 1 below.

Table 1. The prediction accuracy scale

Grade Level	A level	B level	C level	D level
Conformity rate (%)	90-100	85-90	70-85	≤70
Certainty coefficient	[1-0.9]	[0.9-0.85]	[0.85-0.7]	[≤0.7]

3.1.2 WRF-ARW Model Predictions for the Detection Set

Before assessing the credibility of the test set data, the WRF-ARW model needs to predict the test set, and the predicted value of the model is used as the truth estimate.

The prediction results of the WRF-ARW model for the maximum rainfall test set in Henan Province are shown in Figure 1. From Fig. 1, it can be seen that the predicted and actual observed values of the WRF-ARW model for the maximum rainfall in Henan Province are very consistent in the numerical magnitude and data trend, and the error between its predicted and true values ranges from 1.78% to 13.51%, which are all less than 15%. It shows that the model has a better prediction effect, which illustrates that the WRF-ARW model fully learns the characteristics of the data and has better prediction ability.

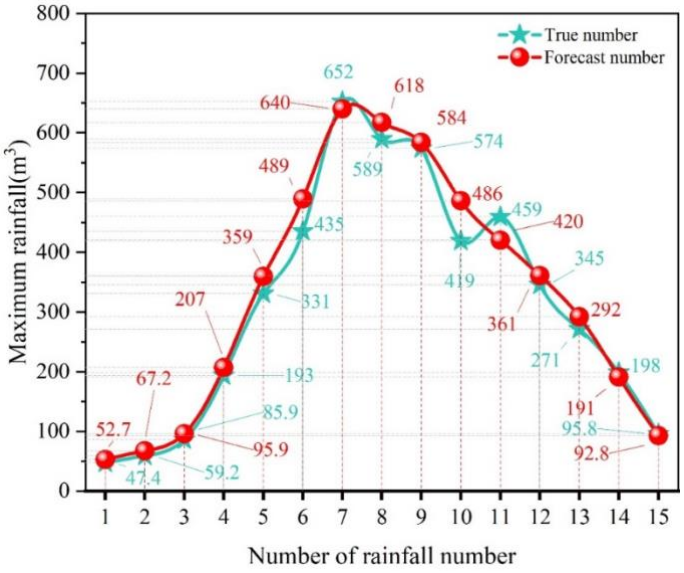


Figure 1. The WRF-ARW model is the prediction curve for maximum rainfall

The prediction results of the WRF-ARW model for the total rainfall detection set in Henan Province are shown in Fig. 2. The experimental results show that the predicted values of the WRF-ARW model and the actual observed values are very consistent in the numerical magnitude and data trends, and the errors between its predicted values and the real values are less than 10% in the range of 0.16%-3.78%. The model's prediction effect is extremely good, which indicates that it fully learns the data's characteristics and has a good prediction ability.

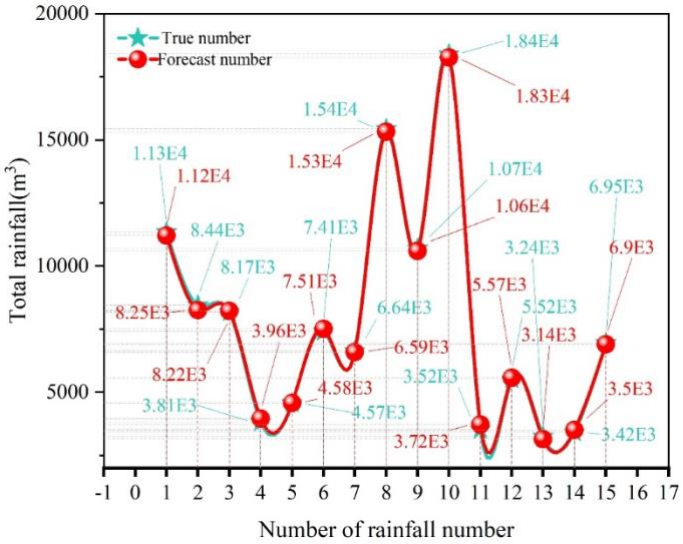


Figure 2. The prediction curve of total rainfall is based on the WRF-ARW model

3.1.3 Data credibility assessment for information fusion

Based on the prediction results of the model, the data credibility of the test set is evaluated according to the credibility measure algorithm, and the results of the data credibility evaluation of the test set of raw data are shown in Figure 3. For the undisturbed test set data, the data credibility given by this assessment method is very high. The confidence values of all the test samples are in the range of 0.957-0.997, and the confidence level is greater than 0.9 and close to 1. Thus, the raw data collected in this paper has a high level of confidence. The WRF-ARW model-based credibility assessment algorithm has a high degree of accuracy in assessing raw data, as demonstrated by this.

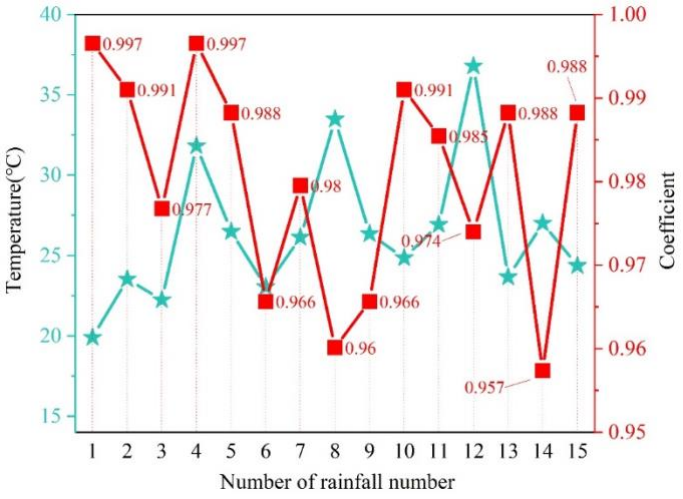


Figure 3. The credibility assessment of the original data set

3.2 Analysis of weather forecasting results of Henan by WRF-ARW model

3.2.1 Prediction of maximum rainfall by the WRF-ARW model

In order to analyze the precipitation simulation effect of the WRF-ARW model in Henan Province, the daily rainfall data from 10 national standard meteorological stations and 5 rainfall stations in

Henan Province, totaling 15 stations, were selected as measured data, and the data period was from 2022 to 2023. The daily precipitation data of the 10 national standard weather stations were provided by the National Meteorological Information Center of the China Meteorological Administration (CMA), and a group company in Henan Province provided the daily precipitation data of the 5 rainfall stations.

The maximum rainfall and relative error of the forecast for 2022-2023 in Henan Province are displayed in Figure 4 below. All the relative errors of the predicted maximum rainfall at 15 points are within 10%, as shown by the experimental results. The WRF-ARW model's prediction of maximum rainfall at number 12th has the largest relative error of 9.21% among them. This was followed by the numbered 5th, 14th, and 6th determinations with relative errors of 8.19%, 7.92%, and 7.35%, respectively. Number 7 was the measurement number with the lowest relative error, with a relative error of 1.23%. Overall, the relative errors of the 15 times of maximum rainfall in Henan Province predicted by the WRF-ARW model during 2022-2023 are all within 10%, indicating that the experimental data predicted by the WRF-ARW model have high credibility. The accuracy of the maximum rainfall predicted by the WRF-ARW model is >90%, and the model accuracy parameter reaches A level.

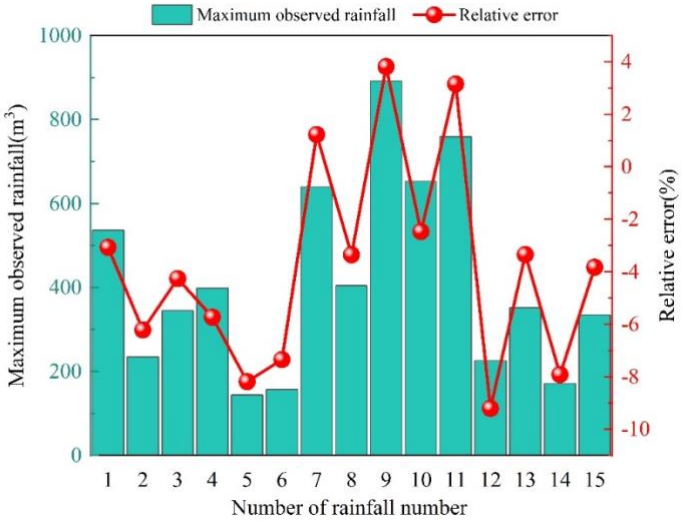


Figure 4. The WRF-ARW model predicts maximum rainfall and error in 2022 to 2023

3.2.2 Prediction of maximum rainfall by the WRF-ARW model

The results of total rainfall and relative errors predicted based on the WRF-ARW model for 2022-2023 in Henan Province are shown in Figure 5 below. According to the experimental results, the predicted total rainfall errors at all 15 points are within 10%. The total rainfall predicted at number 10 has the largest relative error of 6.81%, followed by numbers 9 and 4, which have relative errors of 6.2% and 5.47%, respectively. The remaining numbering has a lower relative error rate of less than 5%. Especially numbered 3rd, 14th, and 15th have relative errors of 0.14%, 0.26%, and 0.37% for the total precipitation predicted by using the WRF-ARW model, and the relative error of the total precipitation predicted by the WRF-ARW model is <1%. It can be seen that the similarity between the total rainfall values predicted by the WRF-ARW model and the measured values 15 times in Henan Province from 2022 to 2023 is infinitely close to that of the measured values, which indicates that the experimental data are extremely credible. The accuracy of the prediction results of the total rainfall by the WRF-ARW model is >90%, and the parameters of the model accuracy are still up to the level of A. The total rainfall predicted by the WRF-ARW model is >1%.

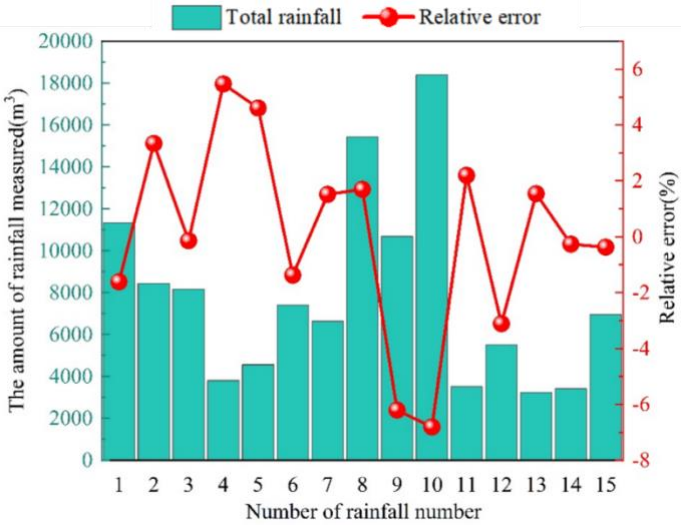


Figure 5. The WRF-ARW model predicts total rainfall and error in 2022 to 2023

3.2.3 Efficiency statistics of parameterized schemes for the WRF-ARW model

The qualification rate of the WRF-ARW model weather forecasting scheme is depicted in Table 2 below. It is found that after the post-processing of the WRF-ARW model forecast precipitation data, the number of forecast maximum rainfall events in Henan Province totaled 15 maximum rainfall events, and the number of qualified events of maximum rainfall was 13, with a qualification rate of 86.7%, and the overall forecast accuracy was grade B. The maximum rainfall time of peak occurrence was 14, with a qualification rate of 93.3%. The total number of rainfall events in Henan Province after the post-processing of precipitation data forecasted by the WRF-ARW model is 15, and the total number of qualified rainfall events is 14, with a qualification rate of 93.3%, and the overall forecast accuracy is A. The total number of qualified rainfall events in Henan Province after the post-processing of precipitation data forecasted by the WRF-ARW model is 15, and the total number of qualified rainfall events is 14, with a qualification rate of 93.3%, and the overall forecast accuracy is A. The number of qualified rainfall peak times was 13, with a qualification rate of 86.7%, and the overall accuracy of the forecast was Grade B. It was predicted that the maximum and total rainfall would be 15, with a qualification rate of 14 and a qualification rate of 93.3%. Both the maximum rainfall and total rainfall have been certified with a B grade. Overall, the WRF-ARW model is very effective in improving the efficiency of parameterized scenarios in Henan Province.

Table 2. The results of the quality rate of the WRF-ARW model weather forecast scheme

Weather forecast project	Total number	Qualified number	Conformity rate (%)	Grade Level
Maximum rainfall	15	13	86.7%	B level
Emergence time	15	14	93.3%	A level
Total rainfall	15	14	93.3%	A level
Emergence time	15	13	86.7%	B level

3.3 Difficulties and Challenges in Implementing Parametric Program Efficiency

3.3.1 Deficiencies in the Use of Parametric Schemes for Low Precipitation

The post-processing model of the WRF-ARW parameterization scheme used in this paper is not good at extracting information of low values, which will lead to the difficulty of extracting the smaller precipitation events, which will often cause a certain degree of error on the accuracy of weather prediction, thus affecting the accuracy of the weather forecast, so it is necessary to process the raw data to make the spatial information of the small amount of precipitation can be obtained by the model to obtain a higher-quality Spatial precipitation information. We can compare additional deep-learning models and examine the ones that are better suited for precipitation post-processing.

3.3.2 Quantitative post-processing of WRF-ARW models

The model involved in this study is a precipitation prediction model for Henan Province, which utilizes the division of subunits and the Muskingum method of convergence for precipitation forecasting. In subsequent studies, the rationality of the conclusions can be further verified using numerical weather prediction coupled with distributed weather models. Meanwhile, because of the lack of station data, low selection rate, and not very abundant precipitation fields for validation, more continuous and detailed hour-by-hour rainfall and runoff data from real stations should be collected for the validation of the model stability rate so as to improve the accuracy of the precipitation forecasts. The watershed involved in this study is in the central-southern region, while in the northern region, because of the low precipitation, the amount of data may not be enough to support the post-processing of precipitation with the deep learning model as in this paper, but the method can be used to do the rainfall grading. Models such as TOPMODEL can be used to perform precipitation forecasting work in the northern region.

3.3.3 Trends in future modeling of collective values

In order to meet the future business needs for refined and seamless weather forecasting and climate prediction, as well as the challenges of the future development of heterogeneous multicore high-performance computers, we will develop a new framework for high-precision scalable and conservative atmospheric modeling and scale-adaptive physical processes, and gradually carry out the study of coupled numerical forecasting technology to meet the challenges of the future development of local 100-meter and global 1,000-meter level resolution and coupled numerical forecasting.

4 Conclusion

This paper first introduces the application of the ensemble numerical forecasting model in weather forecasting and then constructs the WRF-ARW numerical weather forecasting model by selecting the deterministic coefficient as the objective function of parameter optimization for the specific climatic characteristics of the Henan region. After that, the WRF-ARW model's parameters are optimized and graded for accuracy, and the reliability and validity of the original dataset are assessed using this model. Finally, the WRF-ARW model is used to simulate weather prediction in Henan Province, and the impact of the difficulties encountered in implementing the parameterized scheme is analyzed.

1) The errors between the predicted and real values of maximum rainfall and total rainfall in Henan Province by the WRF-ARW model are 1.78%-13.51% and 0.16%-3.78%, which are less than 15% and 5%, respectively. The model's prediction effect is better, which

demonstrates that the WRF-ARW model fully understands the data's characteristics and has a stronger prediction ability.

2) The WRF-ARW model forecasts the maximum rainfall and the time of peak occurrence of maximum rainfall with 13 and 14 qualified fields, respectively, with a qualification rate of 86.7% and 93.3%, and the overall forecasting accuracy is grade B and grade A, respectively. The total rainfall and peak rainfall occurrence time were qualified in 14 and 13 cases, with pass rates of 93.3% and 86.7%, respectively, and the overall prediction accuracy was Grade A and Grade B, respectively. The WRF-ARW model is highly effective in weather forecasting in Henan Province overall.

References

[1] Krishnamurti, T. N., & Bounoua, L. (2018). An introduction to numerical weather prediction techniques. CRC press.

[2] Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L. V., Hegerl, G., ... & Zwiers, F. W. (2017). Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. Weather and climate extremes, 18, 65-74.

[3] Tang, J., Xu, X., Zhang, S., Xu, H., & Cai, W. (2023). Response of remote water vapor transport to large topographic effects and the multi-scale system during the "7.20" rainstorm event in Henan Province, China. Frontiers in Earth Science, 11, 1106990.

[4] Guo, X., Cheng, J., Yin, C., Li, Q., Chen, R., & Fang, J. (2023). The extraordinary Zhengzhou flood of 7/20, 2021: How extreme weather and human response compounding to the disaster. Cities, 134, 104168.

[5] Huang, Y., Guo, B., Sun, H., Liu, H., & Chen, S. X. (2021). Relative importance of meteorological variables on air quality and role of boundary layer height. Atmospheric Environment, 267, 118737.

[6] Scher, S., & Messori, G. (2019). Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. Geoscientific Model Development, 12(7), 2797-2809.

[7] Cho, D., Yoo, C., Im, J., & Cha, D. H. (2020). Comparative assessment of various machine learning-based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas. Earth and Space Science, 7(4), e2019EA000740.

[8] Xu, M., Zhao, C., Gu, J., Feng, J., Li, G., & Guo, J. (2023). Appropriately representing convective heating is critical for predicting catastrophic heavy rainfall in 2021 in Henan Province of China. Environmental Research Communications, 5(5), 051002.

[9] Murray, S. A. (2018). The importance of ensemble techniques for operational space weather forecasting. Space Weather, 16(7), 777-783.

[10] Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K., & Wang, X. (2021). Deep learning-based weather prediction: a survey. Big Data Research, 23, 100178.

[11] Wang, T., Zhang, Y., Zhi, X., & Ji, Y. (2023). Multi-model ensemble forecasts of surface air temperatures in Henan Province based on machine learning. Atmosphere, 14(3), 520.

[12] Zheng, L., Li, T., & Liu, D. (2023). Evaluation of sub-seasonal prediction skill for an extreme precipitation event in Henan province, China. Frontiers in Earth Science, 11, 1241202.

[13] Anna Jiang, Wanshun Zhang, Xin Liu, Feng Zhou, Ao Li, Hong Peng & Hao Wang. (2024). Improving hydrological process simulation in mountain watersheds: Integrating WRF model gridded precipitation data into the SWAT model. Journal of Hydrology131687-131687.

[14] Zhang Weihang, Tian Meng, Hai Shangfei, Wang Fei, An Xiadong, Li Wanju... & Sheng Lifang. (2024). Improving the Forecasts of Coastal Wind Speeds in Tianjin, China Based on the WRF Model with Machine Learning Algorithms. Journal of Meteorological Research(3),570-585.

[15] Ahmed Raghad H., Roomi Thaer O. & Hussain Hazim H.. (2023). High Resolution Dynamical Downscaling of Air Temperature and Precipitation using WRF-ARW Model Over Iraq. IOP Conference Series: Earth and Environmental Science(2),

[16] Sivaramakrishna, Surireddi Satya Venkata, Rao, Bodda Ravi Srinivasa, Satyanarayana, Gubbala Chinna, Rao, Nellipudi Nanaji, Panda, Roshmitha, Sai, Singuru Madhu... & Rao, Dodla Venkata Bhaskar. (2022). Simulation of Regional Climate over the Indian subcontinent through dynamical downscaling using WRF-ARW model. Theoretical and Applied Climatology(1-2),1-23.

[17] Tamamadin Mamad, Lee Changkye, Kee Seong-Hoon & Yee Jurng-Jae. (2023). Prediction of Convective Available Potential Energy and Equivalent Potential Temperature using a Coupled WRF and Deep Learning for Typhoon Identification. IOP Conference Series: Earth and Environmental Science(1),

[18] Wahiduzzaman Md, Ali Md. Arfan, Luo Jing Jia, Wang Yu, Uddin Md. Jalal, Shahid Shamsuddin... & Haque Md. Emdadul. (2021). Effects of convective available potential energy, temperature and humidity on the variability of thunderstorm frequency over Bangladesh. Theoretical and Applied Climatology(1-2),325-346.