A Novel Integrated Learning Model for Rainfall Prediction CEEMD- FCMSE -Stacking

Xianqi Zhang
North China University of Water Resources and Electric Power
Kai Wang (2416932728@qq.com)
North China University of Water Resources and Electric Power
Tao Wang
North China University of Water Resources and Electric Power

Research Article

Keywords: Rainfall Forecasting, CEEMD, RCMSE, Stacking Integrated Learning, Xixia

Posted Date: January 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1006290/v1

A novel integrated learning model for rainfall prediction CEEMD- FCMSE -Stacking

Xianqi Zhang[a,b,c] , Kai Wang[a,*], Tao Wang[a]

[a]*Water Conservancy College, North china University of Water Resources and Electric Power, Zhengzhou 450046, China*

[b]*Collaborative Innovation Center of Water Resources Efficient Utilization and Protection Engineering, Zhengzhou 450046, China*

[c]*Technology Research Center of Water Conservancy and Marine Traffic Engineering, Henan Province, Zhengzhou 450046, China*

***Corresponding authors**: Kai Wang. Email: 2416932728@qq.com;phone: +86-152-9042-1442

Authors:

Xianqi Zhang. E-mail: zxqi@163.com; phone: +86-158-3719-7937

Kai Wang. Email: 2416932728@qq.com; phone:+86-152-9042-1442

Tao Wang. E-mail: 1124149584@qq.com; phone: +86-157-3834-7223

Abstract: Scientific prediction of precipitation changes has important guiding value and significance for revealing regional spatial and temporal patterns of precipitation changes, flood climate prediction, etc. Based on the fact that CEEMD can effectively overcome the interference of modal aliasing and white noise, fine composite multi-scale entropy can reorganize the same FCMSE value to reduce the modal component and improve the computational efficiency, and Stacking ensemble learning can effectively and conveniently improve the fitting effect of machine learning, a rainfall prediction method based on CEEMD-fine composite multi-scale entropy and Stacking ensemble learning is constructed, and it is applied to the prediction of monthly precipitation in the Xixia. The results show that, under the same conditions, the CEEMD-RCMSE-Stacking model reduces the root mean square error by 83.48% and 62.08%, and the mean absolute error by 83.25% and 61.84%, respectively, compared with the single Stacking model and CEEMD-LSTM, while the goodness-of-fit coefficients improve by 15.94% and 2.34%, respectively, which means that the CEEMD-RCMSE-Stacking model has higher prediction performance. The CEEMD-RCMSE-Stacking model has higher prediction performance.

Keywords: Rainfall Forecasting, CEEMD, RCMSE, Stacking Integrated Learning, Xixia

Introduction

The contradiction between water supply and demand is a major problem facing China's social development at present, while precipitation is an important way of recharging regional water resources, and long-term precipitation forecasts play an important role in the rational allocation and use of water resources (Qing et al. 2019). Precipitation is influenced by a variety of uncertainties and its process is very complex. Most of the precipitation series are non-stationary series, which not only have the characteristics of trend and periodicity, but also have randomness, sudden changes and "multi-time scale" structure, which makes the precipitation prediction accuracy relatively low. This makes the accuracy of precipitation forecasts relatively low. In recent years, scholars at home and abroad have conducted a lot of research on precipitation forecasting and have achieved fruitful results. Bidroha et al. (2020) A downscaling method based on support vector regression (SVR) to downscale rainfall at several locations in a study area to obtain future rainfall forecasts, and the performance of the SVR method is compared with downscaling methods based on correlation vector machines and deep learning to demonstrate the effectiveness of the method. Yu et al.(2018) explored short-long time variation information in the original rainfall time series, where Support Vector Regression (SVR) was used for short-period component prediction and Artificial Neural Networks (ANN) for long-period component prediction, with results showing superior performance to traditional methods. Fethi et al. (2018) Prediction of rainfall using data retrieved from the Meteosat Second Generation (MSG) Rotationally Enhanced Visible and Infrared Imager (SEVIRI) based on the Random Forest (RF) algorithm, showing that the rainfall estimated by the scheme correlates well with that observed by rain gauges. Khan et al. (2020) The analysis was carried out using 30 years of rainfall data from 1986 - 2016 in the Langat River Basin, Malaysia. Discrete wavelet transform decomposition of the calculated drought time series and prediction of high frequency subseries using artificial neural networks. Kavya et al. (2020) An adaptive empirical model decomposition-artificial neural network model is proposed for predicting summer rainfall in India. Liu et al. (2021) Integration of improved K-Nearest Neighbour, Remote Sensing and Geographic Information Systems to analyse extreme rainfall data from tourist sites. Sun et al. (2021) The possibility of using the RF-SVR statistical downscaling model for extreme rainfall simulations during the flood season was explored, and the downscaling effects of the RF-SVR statistical downscaling model were compared with those of the SVR model. The results show that the deviations of daily rainfall in the Luan River basin simulated by the RF-SVR model are significantly reduced, and the prediction of extreme rainfall in the basin can be improved. Guo et al. (2010) The ANN-based statistical downscaling method is studied and explored to establish the statistical relationship between large-scale climate observations and measured precipitation through ANN and apply it to study the precipitation changes in the Han River basin under future climate scenarios. Lin et al.

(2021) The integrated models such as bagging integrated model and stacking integrated model were established respectively to improve the effect of short time runoff forecasting in the Andun water basin, and the results show that its stacking integrated model has a better effect on the prediction of small flow incoming water. Scholars at home and abroad have investigated the use of combinatorial methods to optimise model parameters or modify parts of the code to improve the accuracy of rainfall prediction models, however there is less research on integrated learning model methods using each single prediction model. The Stacking algorithm brings together the strengths of different models and allows the raw data to be analysed from multiple perspectives, resulting in better predictive performance. In this paper, an adaptive noise-complete ensemble empirical modal decomposition (CEEMD) method is first introduced to decompose the raw precipitation series and calculate the fine composite multiscale entropy (RCMSE) of each decomposition component. The sequences of components with similar entropy values are then reorganised into new sequences to reduce model complexity and improve computational efficiency. In the prediction stage, the reconstructed series are based on the popular KNN, RF, SVR and ANN algorithms mentioned above, and the Stacking algorithm is used to combine the advantages of the four models to analyse the raw data from multiple perspectives, thus improving the model prediction performance and applying it to the monthly precipitation prediction at the Xixia station.

1 Research methods

1.1 CEEMD algorithm

The CEEMD algorithm adaptively adds white noise to the decomposition process to solve the modal mixing problem that occurs with EMD, and to overcome the low efficiency and noise residuals of EEMD decomposition(Sang et al. 2019).The specific steps of the CEEMD algorithm implementation are as follows:

(1) The i-th signal $y^i(t)$ obtained by adding adaptive white noise to the original rainfall signal can be expressed as

$$y^i(t)=y(t)+\varepsilon_0\omega^i(t),\ t=1,2,\ldots,n \tag{1}$$

where n is the sample point of the rainfall data; t is the rainfall data period (1 mouth). $y(t)$ is the original wind power signal. ε_0 is the standard deviation of the noise. $\omega^i(t)$ is the i-th white noise signal.

(2) The EMD decomposition of $y^i(t)$ A is performed until the first modal component $IMF_1(t)$ is obtained

$$IMF_1(t)=\frac{1}{n}\sum_{i=1}^{n} IMF_1^{\mathrm{i}}(t) \tag{2}$$

Calculate the first residual signal $r_1(t)$

$$r_1(t)=y(t)-IMF_1(t) \tag{3}$$

(3) The EMD algorithm is used to decompose the signal $r_1(t)+\varepsilon_1 E_1(\omega^i(t))$ to obtain the second modal component $IMF_2(t)$

$$IMF_2(t)=\frac{1}{n}\sum_{i=1}^{n} E_1\left(r_1(t)+\varepsilon_1 E_1\left(\omega^i(t)\right)\right) \tag{4}$$

For each of the remaining stages, i.e. k=2, ... , k, the kth residual signal is calculated as follows

$$r_k(t)=r_{k-1}(t)-IMF_k(t) \tag{5}$$

Repeat the procedure in step (3) to obtain the k+1st modal component as follows

$$IMF_{k+1}(t)=\frac{1}{n}\sum_{i=1}^{n} E_1\left(r_k(t)+\varepsilon_k E_k\left(\omega^i(t)\right)\right) \tag{6}$$

Steps (4) and (5) are repeated until the residual signal can no longer be decomposed, resulting in K modal components. The decomposed final residual signal is

$$R(t) = y(t) - \sum_{i=1}^{K} IMF_k(t) \tag{7}$$

The original rainfall signal is therefore decomposed into

$$y(t) = \sum_{i=1}^{K} IMF_k(t) + R(t) \tag{8}$$

1.2 Fine-composite multiscale entropy

The FCMSE algorithm is an algorithm that measures the complexity of a time series as the sample entropy of the time series at different scales, which compensates for the shortcomings of the MSE and CMSE algorithms (Kang et al. 2021). The main steps are as follows:

(1)For the rainfall time series $x = \{x_1, x_2, \cdots, x_n\}$, its τ - th scale coarse-grained process is

$$y_{k,j}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i, 1 \le j \le \frac{n}{\tau}, 1 \le k \le \tau \tag{9}$$

Coarse-grained time series at different time scales were obtained using the coarse-graining described above.

(2) Under the scale factor τ , the number of all τ coarse-grained sequence matching vector pairs $n_{k,\tau}^{m+1}$ and pairs $n_{k,\tau}^{m}$ is calculated.

(3) Calculate the mean $\overline{n_{k,\tau}^{m+1}}$ and the value $\overline{n_{k,\tau}^{m}}$ of τ the $n_{k,\tau}^{m+1}$ and $n_{k,\tau}^{m}$ at $1 \le k \le \tau$. Define the FCMSE under the scale factor B as the logarithm of the ratio of N and the value M, i.e.

$$RCMSE\left(x,\tau,m,r\right) = -In\frac{\overline{n_{k,\tau}^{m+1}}}{\overline{n_{k,\tau}^{m}}} \tag{10}$$

According to equation (10), the FCMSE appears as undefined entropy only when all $n_{k,\tau}^{m+1}$ and $n_{k,\tau}^{m}$ are zero.

1.3 Stacking Algorithms

In the Stacking Integrated Learning Model, the predictive power of each base learner is analysed individually, while the combined effect of each base learner is compared, so that the Stacking Integrated Learning Model achieves the best prediction results (Shi et al. 2019). The base learner has to be selected for good performance but with different model principles. The K-Nearest Neighbor algorithm is a well-theorized classical machine learning method that is simple to implement and efficient to train (Lu et al. 2021). Random Forest is one of the first machine learning algorithms that improves the prediction accuracy of various models by constructing different training sets to improve the variance of each classification model (Dong et al. 2020，Xiao et al 2020). Support vector regression shows good fitting performance for non-linear regression prediction of high-dimensional, complex data (Xiong et al. 2006).Better generalization and fitting performance of artificial neural network models (Gorai et al.2021). In summary, this paper selects KNN, RF, SVR and ANN as the base learners, and the ANN algorithm with strong generalization ability is used as the meta-learner.

2 CEEMDAN-RCMSE-Stacking Model construction

Based on the above-mentioned CEEMD algorithm, fine composite multiscale entropy and Stacking algorithm, a short-term precipitation prediction model is built in this paper, as shown in Figure 1. The specific ideas are as follows.

(1) The raw precipitation data were decomposed into s modal components and 1 residual component using the CEEMD algorithm.

(2) The RCMSE values of each component are calculated and sequences with similar entropy values are superimposed to form a new sequence of components.

(3) The reconstructed component data were divided into a training set D and a test set T.

(4) Using 5-fold cross-validation, the training set D is randomly divided into 5 equal parts. Let each base learner train 1 copy of the training set and use the remaining 4 copies as the test set for prediction. Combine the predictions of the 4 base learners into the training set D' of the meta-learner.

(5) Let each base learner make predictions on the test set T. The average of the predictions is used as the test set T' for the meta-learner.

(6) A meta-learner is used to train the new training set D',and this meta-learner is used to predict the new test set T'. The final prediction is output.

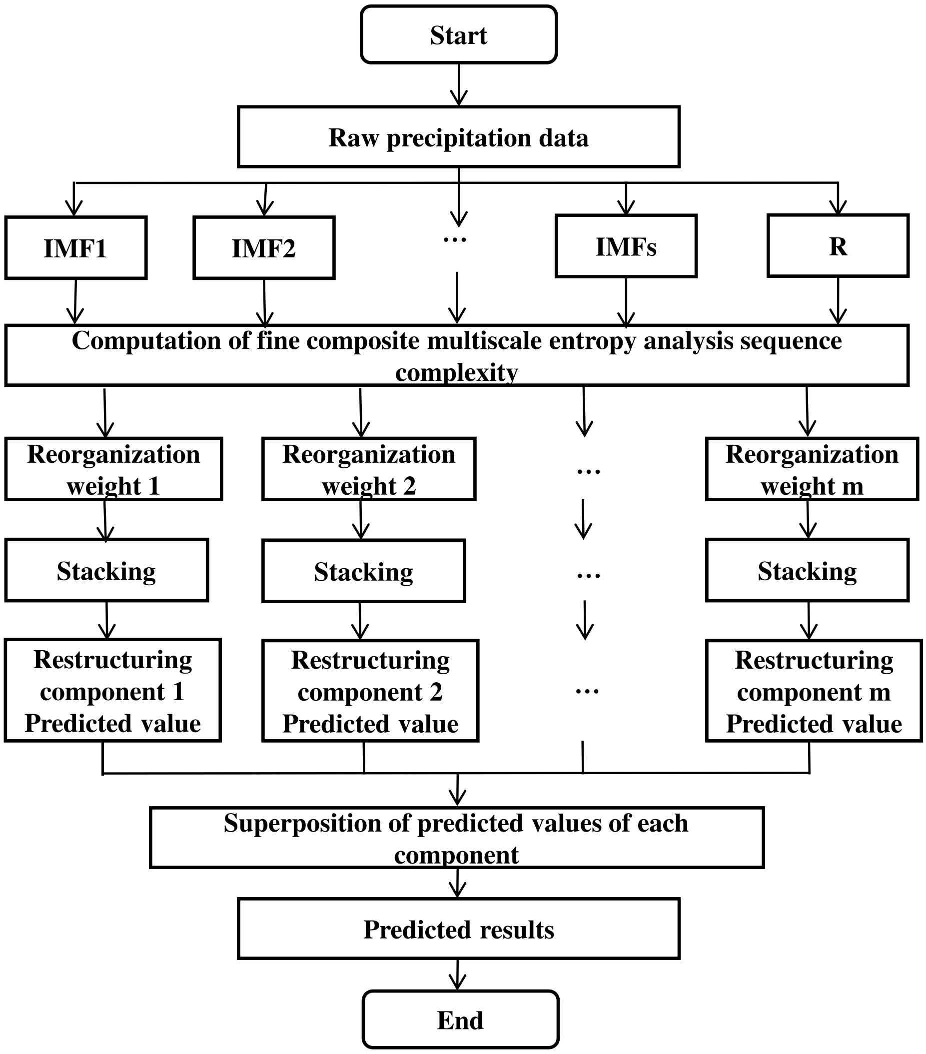

Fig1 Flow of CEEMDAN-RCMSE-Stacking based prediction model

3 Example applications

The data used in this paper are from actual precipitation measurements in the Western Gorge, as shown in Figure 2, sampled from 1960 to 2019. A total of 720 data points were used, with the first 660 data points used for the training set and the last 60 data points for the test set. In this paper, root mean square error (RMSE), mean absolute error (MAE) and goodness-of-fit coefficient R^2 are used to evaluate the model error.

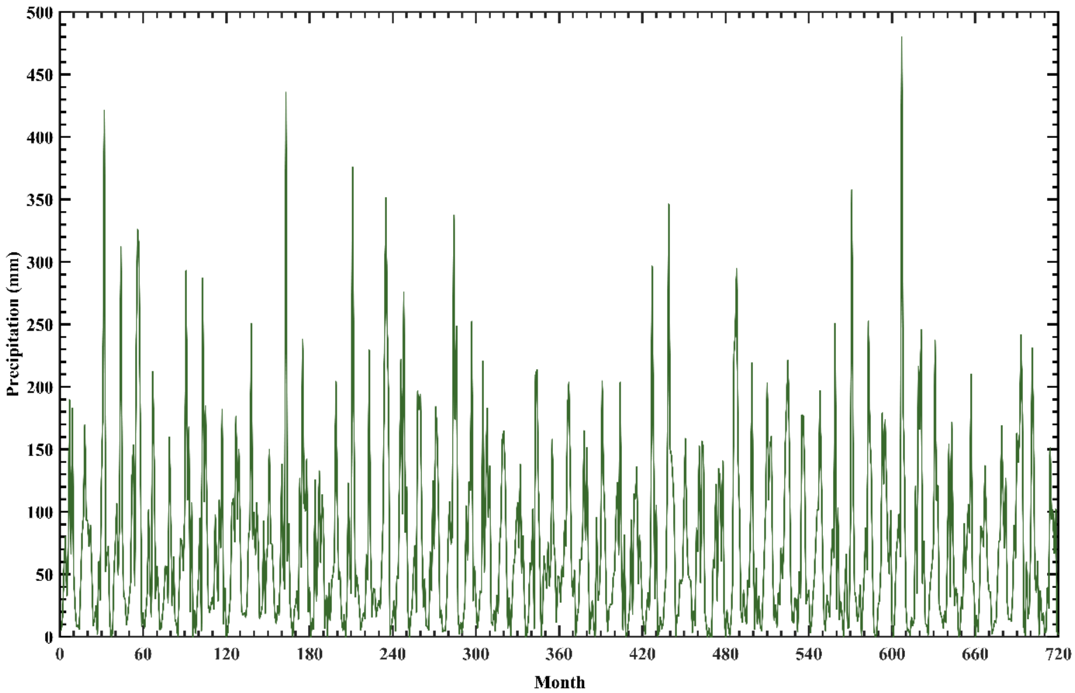

Fig2 Monthly precipitation sequence in Xixia

3.1 Precipitation sequence CEEMD-RCMSE decomposition recombination test

In this paper, the CEEMD algorithm is used to decompose the raw wind power data series by adding the number of white noise groups $NR = 80$, the noise standard deviation $N_{std} = 0.05$, and the maximum number of iterations $MaxIter = 100$. The decomposition results are shown in Figure3.

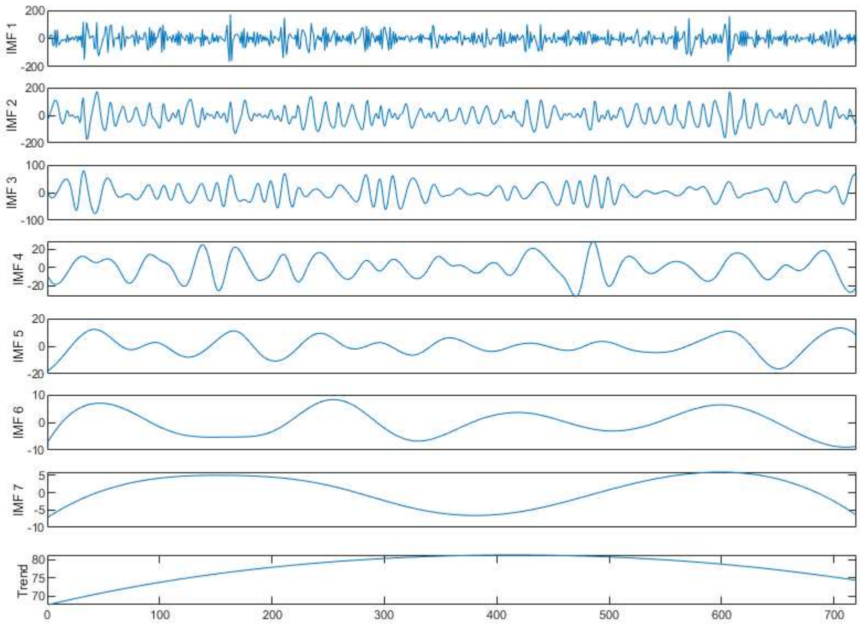

Fig3 CEEMD decomposition of monthly precipitation series results

As can be seen from Figure 3, the precipitation series is decomposed by CEEMD into 7 modal components and 1 residual component at different frequency scales. Due to the non-smoothness of the series makes the decomposition result in a large number of component series, if the 5-fold method is used to model the prediction of each component directly it will lead to a sharp increase in computational effort. This paper therefore uses the RCMSE method to calculate the fine composite multiscale entropy values of each component in order to assess the complexity of each component, and based on this, the components are combined and reorganised. RCMSE test parameters: embedding dimension $m = 2$, conditional threshold r takes 0.2 times standard deviation of component sequence time lag $tau = 1$. FCMSE values for each IMF component and residual are listed in Table 1.

Table 1 FCMSE values for IMF components and residuals

Component sequences	FCMSE values	Component sequences	FCMSE values
IMF1	0.1605	IMF5	0.0443
IMF2	0.1044	IMF6	0.0243
IMF3	0.0639	IMF7	0.0089

IMF4	0.0560	R	0.0023

As can be seen from Table 1, the RCMSE values for each component series are mainly distributed in the four numerical neighbourhoods of 0.15, 0.1, 0.05 and 0.005. Accordingly, in this paper, the component sequences are recombined and the results are as follows: F1=IMF1, F2=IMF2, F3=IMF3+IMF4+IMF5+IMF6 and F4=IMF7 +R are the recombined components. Figure 4 shows the rainfall recombination component sequences.

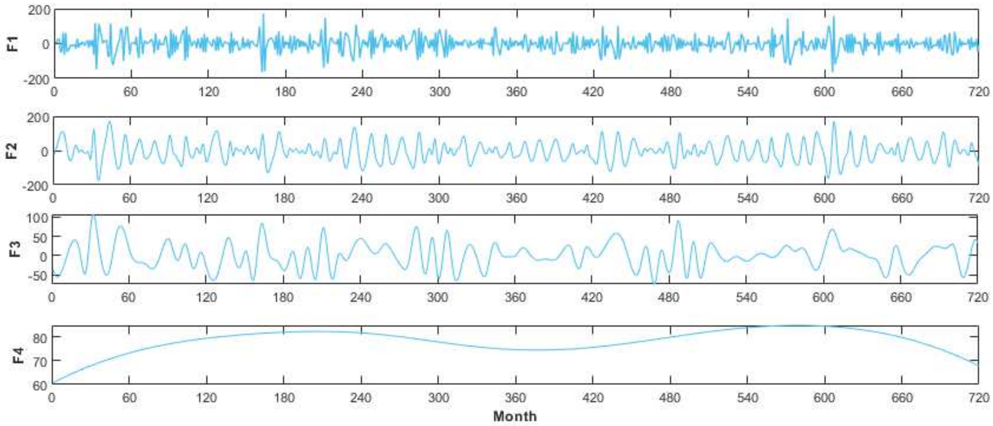

Fig4 FCMSE recombination component sequence results

The reorganised four components were imported into the Stacking model for prediction and the results are shown in Figure 5.

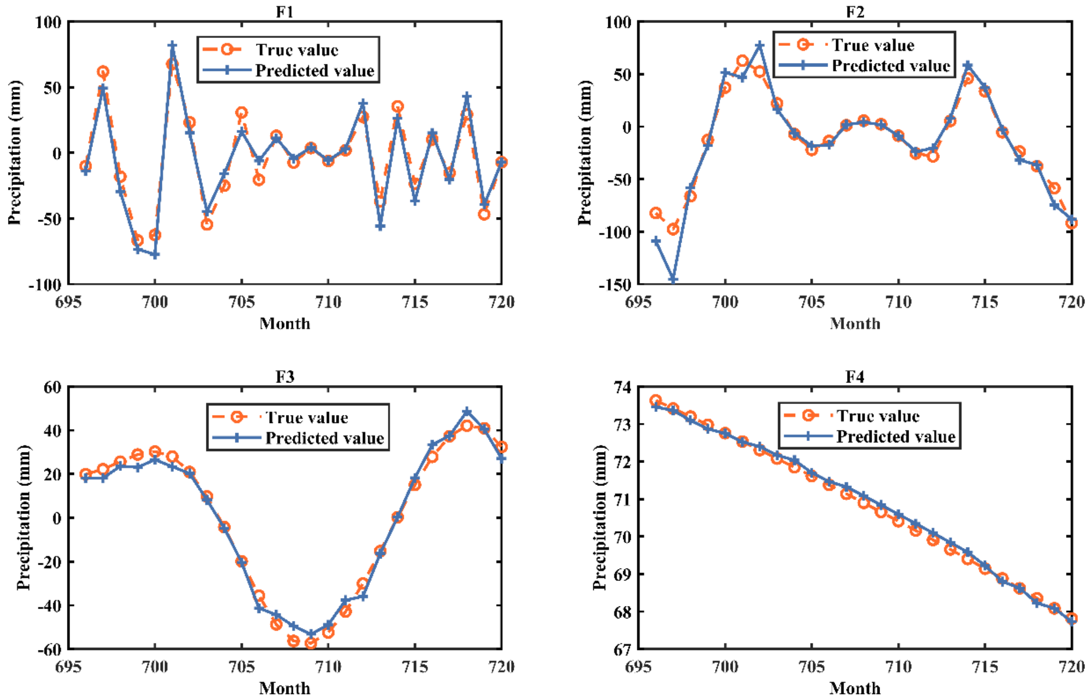

Fig5 F1-F4 prediction results

As can be seen from Figure 5, the smoothness of the Xixia precipitation time series is enhanced and the volatility is significantly reduced after the CEEMD decomposition. The prediction errors from F1 to F4 are getting lower and lower, which indicates that the training is getting better and better.

3.2 Cross-sectional comparison test of Stacking model precipitation prediction

In order to compare the prediction accuracy of the Stacking prediction model, this paper will use four base learners to do a cross-sectional comparison test of the prediction of modal F3, taking into account the moderate frequency fluctuation and smoothness of modal F3, which is more likely to reflect the fitting effect differently. The values of each model fit error evaluation metric and precipitation prediction plots for the F3 modal are shown in Figure 6 and listed in Table 2.

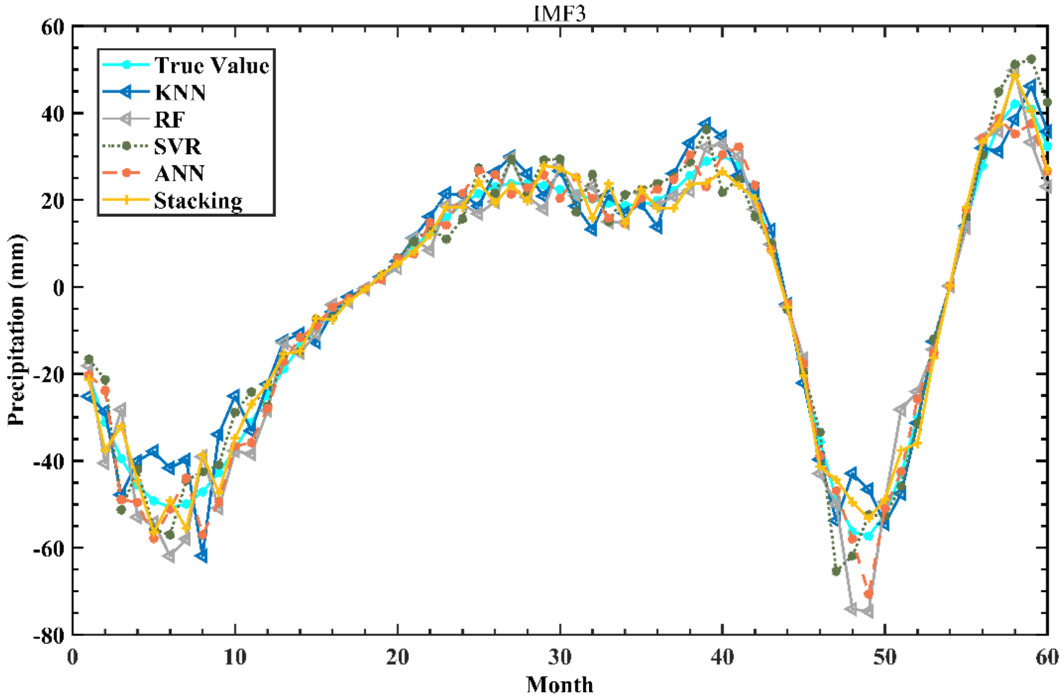

Fig6 Fitting of each model for the F3 mode

Table 2 Relative errors in the prediction results for the F3 trend term

Predictive models	RMSE	MAE	R^2
KNN	26.78	20.39	0.8158
RF	26.04	19.17	0.8227
SVR	26.42	19.24	0.8213
ANN	25.14	18.16	0.8432
Stacking	22.52	15.10	0.8591

It can be seen from Table 2 that in the base learner, the RMSE and MAE of the ANN algorithm are the smallest, which are 25.14 and 18.16, respectively. From the fitting effect, ANN model fitting degree is the highest. This indicates that the model performs best and generalises best, and empirically demonstrates the rationality and validity of setting the meta-learner of the Stacking model as an ANN model in this paper. The Stacking model, which integrates four models, achieves better prediction performance than the single model, with prediction errors RMSE and MAE 10.42% and 16.85% lower respectively than the ANN model, which is the best performing of the single models. This is 15.91% and 25.94% lower than the KNN model, which is the worst performing of the single models, and the fit to the actual values of the modal F3 is 5.82% better.

3.3 Five-year rainfall prediction test

To further verify the generalisability and accuracy of the CEEMDAN-RCMSE-Stacking model proposed in this paper, a total of 720 data points were selected from the five-year 12-month precipitation data of the hydrological station, with the first 55 years (660 data points) as the test set and the remaining 60 data points as the test set. The comparison model is the LSTM model, a deep learning algorithm widely used in natural language processing (Liu et al. 2020). The results of 60 months of wind power predictions under each prediction model are shown in Table 4.

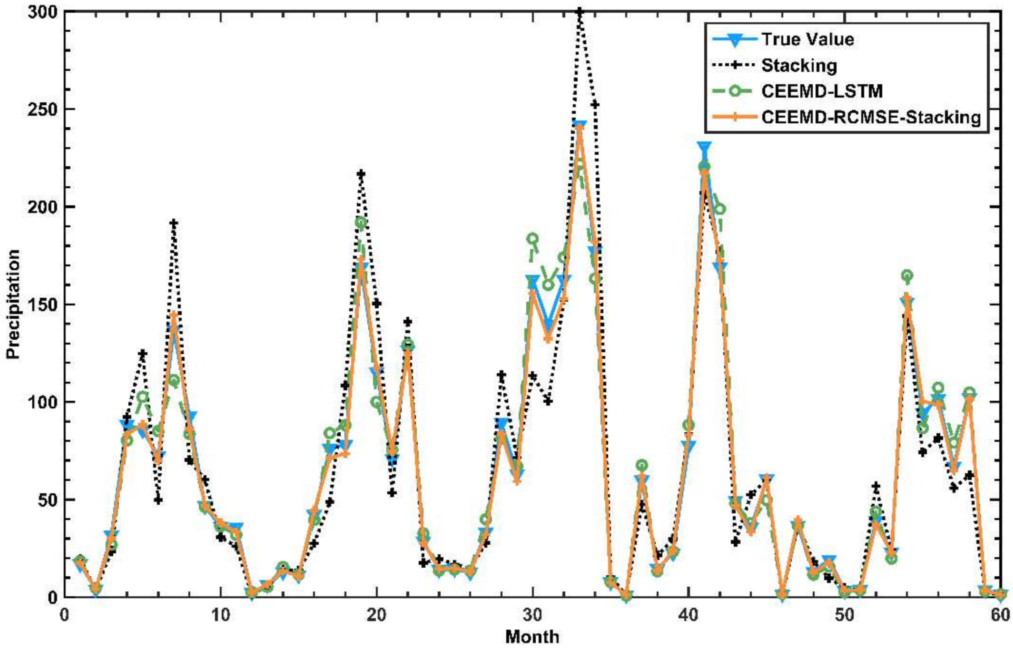

Fig7 Prediction results of the CEEMD-RCMSE-Stacking model compared with other models

As can be seen from Figure 7, the single Stacking model has a large error between the predicted and measured values, and the fit is the worst, with the predictions obtained using the CEEMD-LSTM decomposition method deviating significantly from the peak of the measured runoff series. The combined CEEMD-RCMSE-Stacking model is the best fit and the prediction results are closer to the actual rainfall, verifying the accuracy and superiority of the model proposed in this paper.

Table 3 Comparison of prediction results from different models

Month	True value	Stacking		CEEMD-LSTM		CEEMD-RCMSE-Stacking	
		Predicted value	Relative error %	Predicted value	Relative error %	Predicted value	Relative error %
1	17	19.77	16.28	18.53	9.00	17.53	3.14
2	4	3.74	6.62	4.74	18.49	4.13	3.36
3	31.7	23.38	26.25	26.91	15.11	30.32	4.36
4	88.4	92.39	4.52	80.42	9.03	83.21	5.87
5	85.6	124.82	45.82	102.59	19.85	88.51	3.40
6	72.2	49.93	30.84	85.07	17.82	69.71	3.45
7	136.9	191.60	39.96	111.25	18.74	144.43	5.50
8	92.6	70.28	24.10	83.53	9.80	86.34	6.76
9	46.7	60.23	28.97	46.27	0.93	46.87	0.37
10	36.2	30.82	14.85	36.25	0.14	38.03	5.04
11	35.9	26.06	27.40	32.01	10.84	33.46	6.80
12	2.4	1.61	33.02	2.59	7.93	2.50	3.99
13	6.2	5.86	5.43	5.36	13.55	6.53	5.25
14	13	14.17	8.97	15.47	19.03	13.52	3.97
15	11.3	13.65	20.82	11.60	2.69	10.86	3.88
16	42.3	27.61	34.74	39.64	6.29	44.36	4.87
17	76	48.78	35.82	84.05	10.59	71.39	6.06
18	78.2	108.46	38.69	88.10	12.66	73.53	5.98
19	168.8	216.88	28.48	192.01	13.75	173.04	2.51
20	115.3	150.39	30.44	100.02	13.25	117.44	1.85
21	70.9	53.51	24.53	75.49	6.47	74.81	5.51
22	126.6	141.01	11.39	129.47	2.27	125.12	1.17
23	28.6	17.69	38.15	32.66	14.21	27.97	2.21
24	14.1	19.69	39.61	13.68	2.98	14.55	3.16
25	14.5	16.92	16.66	14.20	2.05	14.92	2.87
26	12.7	14.32	12.76	13.62	7.25	13.16	3.64
27	33.3	27.87	16.30	39.93	19.91	31.94	4.09
28	89.3	113.91	27.56	81.31	8.95	83.44	6.56
29	63.1	71.58	13.44	67.13	6.38	59.36	5.92
30	162.8	113.53	30.26	183.63	12.79	155.61	4.42
31	139.7	100.42	28.12	159.97	14.51	132.41	5.22
32	162.7	152.36	6.36	174.01	6.95	152.82	6.07
33	241.7	299.75	24.02	221.76	8.25	240.16	0.64
34	177.3	252.15	42.22	163.15	7.98	181.99	2.64
35	7.8	10.42	33.59	7.88	0.97	7.61	2.47
36	1.2	1.30	8.23	0.97	19.45	1.18	1.84

37	59.9	47.65	20.45	67.63	12.91	62.06	3.61
38	14.5	21.50	48.28	13.33	8.06	13.65	5.83
39	22.6	29.79	31.81	23.36	3.35	24.03	6.35
40	77.7	83.76	7.80	88.18	13.49	83.22	7.10
41	231.1	207.00	10.43	220.46	4.60	217.30	5.97
42	168.9	176.71	4.62	198.80	17.70	172.25	1.99
43	49.5	28.33	42.77	48.08	2.88	47.03	5.00
44	35.6	52.45	47.33	37.95	6.59	33.52	5.84
45	60.4	60.52	0.20	49.71	17.70	61.32	1.52
46	1.6	0.97	39.28	1.84	15.19	1.57	2.11
47	36.8	36.94	0.38	36.66	0.38	39.47	7.26
48	13	18.42	41.71	11.64	10.49	12.69	2.41
49	19.3	9.80	49.25	15.83	17.99	17.92	7.16
50	3.3	4.82	45.94	2.67	19.08	3.47	5.03
51	3.9	3.79	2.84	3.68	5.75	3.80	2.58
52	39.1	57.01	45.80	43.80	12.03	37.17	4.93
53	23	24.04	4.54	19.85	13.69	23.06	0.24
54	151.1	144.01	4.69	164.85	9.10	153.33	1.48
55	94.5	74.20	21.48	86.60	8.36	100.26	6.10
56	101.6	81.47	19.81	107.33	5.64	98.63	2.93
57	67	55.83	16.66	79.18	18.17	65.06	2.90
58	102	62.49	38.74	104.78	2.73	101.48	0.51
59	3.7	2.38	35.60	3.17	14.40	3.45	6.67
60	1.4	1.43	2.07	1.49	6.58	1.35	3.33
Average relative error %		24.29		10.30		4.06	
RMSE		22.52		9.81		3.72	
MAE		15.10		6.63		2.53	
R2		0.8591		0.9732		0.9960	

The statistical results of the evaluation criteria in Table 3 show that the combined model CEEMD-FCMSE-Stacking is much better than the single Stacking forecasting model, with a 15.94% improvement in the coefficient of goodness of fit R^2 and 83.48% and 83.25% reductions in root mean square error RMSE and mean absolute error MAPE, respectively. The CEEMD-LSTM also achieves a certain level of fit, but not as good as the CEEMD-FCMSE-Stacking model at the peak points. Compared with the CEEMD-LSTM model, the combined CEEMD-RCMSE-Stacking model improved the goodness-of-fit coefficient R2 by 2.34%, and reduced the root mean square error RMSE and mean absolute error MAPE by 62.08% and 61.84%.

To improve the accuracy of medium- and long-term precipitation forecasts, this paper investigates the practicality and feasibility of the Stacking integrated learning model for runoff time series, constructs the CEEMD-FCMSE-Stacking prediction model, and verifies and compares the prediction effects of different models with the monthly precipitation data of the Xixia. The model performs well in precipitation forecasting tasks and is suitable for short-term precipitation forecasting with large seasonal fluctuations and for developing flood and drought plans, and can be used effectively for time series analysis in hydrology and related fields to mitigate the risk of climate extremes.

4 Conclusion

(1) To address the characteristics of intermittent and fluctuating precipitation, this paper introduces the CEEMD algorithm to decompose the precipitation series, which effectively reduces the non-smooth characteristics of the original series. At the same time, the CEEMD algorithm overcomes the possible modal confounding problem in EMD and provides a good basis for the Stacking model to make predictions. At the same time, the complexity of the CEEMD decomposition components is calculated using fine composite multiscale entropy, and they are restructured by FCMSE values, reducing the model complexity and computational scale.

(2) The CEEMD-FCMSE-Stacking model is a better fit than the CEEMD-LSTM coupled neural network model for the location of abrupt changes in precipitation data, and is more reasonably detailed in terms of reflecting the true variability of the series.

The mean relative error of 4.06%, RMSE and MAE of 3.72 and 2.53 respectively, are both low, and the coefficient of goodness of fit R^2 of 0.9960 is very close to 1. The prediction accuracy is better than that of the CEEMD-LSTM model. The empirical results show that the model overcomes the limitations of the coupled model and substantially improves the generalization ability and accuracy of the precipitation prediction.

(3) Although the overall prediction fitting degree of the established CEEMD-RCMSE-Stacking model is relatively high, RMSE and MAE are relatively large. When selecting the base learner and meta-learning, conservative methods are used to select the simple learning methods that are common in recent years. There is no bold choice that has not been applied to the precipitation prediction method. In the subsequent work, further research will be carried out to improve the accuracy of the Stacking integrated learning model.

Availability of data and materials Data and materials are available from the corresponding author upon request.

Author contribution All authors contributed to the study conception and design. writing and editing: Xianqi Zhang and Kai Wang; chart editing: Tao Wang; preliminary data collection: Kai Wang. All authors read and approved the final manuscript.

Funding This work was supported by the Key Scientific Research Project of Colleges and Universities in Henan Province (CN) [grant numbers 17A570004].

Declarations

Ethical approval and consent to participate Not applicable.

Consent to publish Written informed consent for publication was obtained from all participants.

Competing interests The authors declare no competing interests

References

Bidroha B, Maria N, Alan O C.(2020) New Approach to Multisite Downscaling of Precipitation by Identifying Different Set of Atmospheric Predictor Variables[J]. Journal of Hydrologic Engineering,2020,25(5):

Dong N, Ke K W, Li J S. (2020) Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: A case study[J]. Applied Soft Computing Journal, 93:

Fethi O, Mourad L, Soltane A. (2018) Improvement of rainfall estimation from MSG data using Random Forests classification and regression[J]. Atmospheric Research, 211:

Guo J, Guo S L, Chen H. (2010). ANN statistical downscaling method for predicting precipitation changes in the Han River basin[J]. Journal of Wuhan University (Engineering Edition), 43(02):148-152.

Gorai A K, Himanshu V K, Santi C. (2021)Development of ANN-Based Universal Predictor for Prediction of Blast-Induced Vibration Indicators and its Performance Comparison with Existing Empirical Models[J]. Mining, Metallurgy & Exploration:

Kang W H, Xu T Q, Wang Y G. (2021) Short-term wind power forecasting based on CEEMDAN-fine composite multiscale entropy and Stacking integrated learning[J/OL]. Water Resources and Hydropower Technology (in Chinese and English):1-13.

Kavya J, Maya L. Pai S.(2020) Adarsh. Adaptive EEMD-ANN hybrid model for

Indian summer monsoon rainfall forecasting[J]. Theoretical and Applied Climatology:

Khan Md. Munir H, Muhammad N S, (2020). Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting[J]. Journal of Hydrology, 590:

Liu S, Liu R, Tan N Z. (2021) A Spatial Improved-kNN-Based Flood Inundation Risk Framework for Urban Tourism under Two Rainfall Scenarios[J]. Sustainability, 13(5):

Lu K, Wu W, Lin G R. (2021) Combined prediction method of passenger hub aggregation based on KNN regression[J]. Journal of Jilin University (Engineering Edition),51(04):1241-1250.

Liu X, Zhao N, Guo JY. (2020) Monthly precipitation prediction on the Tibetan Plateau based on LSTM neural network[J]. Journal of Geoinformation Science, 22(08):1617-1629.

Lin, Y G, Du Y, Meng Y. (2021) Study on the influence of different integrated models on short time runoff forecasting in small watersheds[J/OL]. China Rural Water Conservancy and Hydropower:1-15.

Li Y W. (2021) Research and application of runoff time series prediction based on integrated learning[D]. University of Electronic Science and Technology.

Qing H, Miao B X, Pan, H W. (2019) Improving Monsoon Precipitation Prediction Using Combined Convolutional and Long Short Term Memory Neural Network[J]. Water, 11(5):

Shi J Q, Zhang J H. (2019) Load forecasting method based on multi-model fusion Stacking integrated learning approach[J]. Chinese Journal of Electrical Engineering, 39(14):4032-4042.

Song Y T, Zhao X H, Zhu X P, Xi D J. (2019) Monthly runoff prediction of the upper Fen River based on CEEMD-BP model[J]. People's Yellow River, 41(08):1-5.

Sun A, Li J Z, Feng P. (2021). Applicability of RF-SVR downscaling model in the Luan River basin[J]. Journal of Water Resources and Water Engineering, 32(02):31-37.

Xiao S G, Mo L, Zhang X. (2020) Short-term load forecasting based on CEEMDAN+RF+AdaBoost[J]. Hydropower Energy Science, ,38(04):181-184+175.

Xiong W L, Xu B G. (2006) Study on the optimal selection method of SVR parameters based on PSO[J]. Journal of System Simulation, (09):2442-2445.

Yu X, Ling G, Li H H. (2018) A SVR－ANN combined model based on ensemble EMD for rainfall prediction[J]. Applied Soft Computing Journal,2018: