IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 19 August 2023, accepted 27 August 2023, date of publication 6 September 2023, date of current version 12 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3312185

RESEARCH ARTICLE

Comparative Study of Multi-Combination Models for Medium- and Long-Term Runoff Prediction in Weihe River

HAO LIU[1,2], WEI LIU[1,2], JUNGANG LUO[3], AND JING LI[4]
[1]Changjiang Schinta Software Technology Company Ltd., Wuhan 430010, China
[2]Changjiang Survey, Planning, Design and Research Company Ltd., Wuhan 430010, China
[3]Department of Hydrology and Water Resources, School of Water Conservancy and Hydropower, Xi'an University of Technology, Wuhan 430010, China
[4]Hydrology Bureau of Changjiang Water Resources Commission, Wuhan 430010, China

Corresponding author: Hao Liu (liuhao_sya@foxmail.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFC3005504.

ABSTRACT The characteristics of hydrological data include nonconsistency and nonlinearity. The prediction accuracy can be improved through the combination of both the decomposition algorithm and the runoff model. Previous studies have typically focused on the combination of a single decomposition algorithm and model. These studies have compared the prediction accuracy before and after decomposition, ignoring the role of multiple decomposition algorithms and models. Considering the limitations of previous single combinations of decomposition algorithms and models, this study will explore the unique features of hydrological data by using a combination of five algorithms, including Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), TIME series decomposition (TIME), Variational Mode Decomposition (VMD), and Singular Spectrum Analysis (SSA). The study constructed models for Prophet, Long Short-Term Memory (LSTM), Multiple Regression (MLR), Random Forest Regression (RFR), Gradient Boosting Regression (GBR), and Support Vector Regression (SVR). Thirty combined prediction models were then developed and used to forecast medium and long-term runoff at Xianyang Station. To comprehensively evaluate the forecasted runoff results, multiple evaluation metrics were used. The prediction accuracy improved after using EMD and TIME decomposition, but the difference was insignificant, and TIME decomposition was the least effective. VMD, EEMD, and SSA, on the other hand, yielded higher data quality. The combined model achieved an NSE above 0.70, demonstrating good prediction results. Of the thirty combined models, the SSA-SVR and SSA-LSTM models were most accurate, with a verification NSE of 0.90. This study developed a comprehensive, reliable, and accurate combination prediction model by employing multiple decomposition algorithms and models. These findings provide a framework for characteristics-driven watershed runoff prediction and water resources scheduling.

INDEX TERMS Decomposition algorithm, singular spectrum analysis, LSTM, combination model, medium and long term runoff forecast.

I. INTRODUCTION

Runoff prediction is an important component of hydrological forecasting that is founded on the objective hydrological law. It involves the qualitative or quantitative prediction of hydrological variables at a particular time along the hydrological cross-section. This is achieved by utilizing present or past hydrometeorological data in conjunction with relevant theories and methods.

In recent decades, the development of computer information technology and hydrological prediction theory has led to a growing number of new prediction model methods proposed by scholars to enhance the theoretical framework of runoff prediction. Process-driven and data-driven models are the two primary categories of runoff prediction models. The data-driven model is commonly used

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar.

in medium and long-term runoff prediction as it does not require an in-depth exploration of hydrophysical mechanism and hydrological movement process. This technique mainly focuses on simulating the optimal relationship between independent and dependent variables. In this model, the predictive factors serve as input data, while runoff factors are used as output, with the optimal algorithm being used to simulate the correlation between the two. The most frequently used models in medium and long-term runoff prediction are the ARIMA model, ANN model, BP neural network model, LSTM model, SVR model, and MLR mode [1], [2], [3], [4].

Hydrological data is characterized by inconsistency and nonlinearity, resulting in low precision in predicting runoff by traditional methods. Numerous studies have demonstrated that data decomposition algorithms can address hydrological data's nonlinearity, enhance data quality and augment the model's prediction accuracy. Yan et al. [5] improved runoff prediction accuracy by combining EEMD and ARIMA, resulting in better results than the traditional ARIMA model. Kişi [6] improved the modeling by coupling wavelet decomposition with the GRNN neural network model, achieving significantly higher prediction accuracy. Weiyao [7] decomposed monthly runoff into EEMD and predicted separately by ANN and LSTM models. The results exhibit an excellent performance with the Nash efficiency coefficient above 0.75. Guoyong et al. [8] developed the EEMD-LSSVM model by reconstructing EEMD and LSSVM and verified its reliability by comparing with other models. Furthermore, numerous studies suggest that EMD, Wavelet Decomposition [6] and VMD can effectively enhance the accuracy of model prediction. Xie et al. [9] em-ployed the EMD method and weighted Markov chain to establish EMDMK and EMD-WDD-MK models, which yield the highest prediction accuracy according to the results. Huang et al. [10] proposed the VMD-DBN-IPSO, a hybrid model-based "feature decomposition-learning reconstruction," to leverage hydrometeorological information, develop high-precision prediction models, and swiftly and accurately predict runoff.

At present, much of the research in the field focuses on combinations of either a single decomposition algorithm and a single model, or a single decomposition algorithm and multiple models. However, the lack of prediction of combinations of multiple decomposition algorithms and models results in the inability to accurately select the combination model with the best prediction effect. Without comprehensive consideration of the combination prediction results, accuracy and comprehensiveness suffer. Therefore, this paper utilizes a five-pronged approach, including empirical mode decomposition, ensemble empirical mode decomposition, time series decomposition, variational mode decomposition and singular spectrum analysis, in order to explore the inherent characteristics of hydrological data. Following this, we build thirty models, including Prophet, short and long memory neural networks, multiple regression, random forest regression, gradient lifting regression, and support vector regression models. We then apply thirty combined prediction models to the medium and long term runoff prediction of Xianyang Station. We use a variety of evaluation indices to provide comprehensive evaluation of the predictions in order to gain a better understanding of the hydrological characteristics of the region and forecast runoff in advance, providing necessary support for future water control, flood control and drought relief work.

The main innovation of this paper lies in the combination calculation of multiple models and multiple decomposition algorithms. Compared with the comparative analysis of a single model and a single algorithm, the combination model calculation can more comprehensively determine the optimal model for runoff prediction in the current study area and the decomposition algorithm more applicable to the model. EMD, EEMD, TIME, VMD and SSA have all been verified on the basis of predecessors, and these decomposition algorithms can improve the prediction accuracy of a single model. Therefore, this paper selects the decomposition algorithm with proven conclusions to calculate the combined model.

II. DATA AND METHODS

A. DATA COLLECTION

The Shaanxi Hydrology and Water Resources Exploration Center provided the runoff data for the hydrological station in this study, while the meteorological data was obtained from China's surface climatological data daily value dataset V3.0. Specifically, it included the daily runoff data from Weijiapu Station and Xianyang Station for the period between January 1, 1970, and December 31, 2019, covering the Weihe River Basin. Additionally, precipitation, evaporation, air pressure and temperature data from Changwu and Wugong meteorological stations were included in the period ranging from January 1, 1960, to December 31, 2019.

B. DATA PROCESSING

1) DATA DECOMPOSITION ANALYSIS

a: EMPIRICAL MODE DECOMPOSITION

The empirical mode decomposition (EMD) proposed by Vautard et al. [11] is a powerful approach used to analyze unstable and nonlinear data. This method can effectively decompose complex input signals into a small number of intrinsic mode functions (IMFs), making it superior to other decomposition methods such as wavelet. In particular, the EMD technique provides great advantages in processing non-stationary time series.

The daily flow data of Xianyang Station is decomposed by EMD, and the decomposition results are shown in Figure 1.

b: SINGULAR SPECTRUM ANALYSIS

SSA [12] uses the singular value decomposition (SVD) in linear algebra to construct the corresponding singular value sequence and decompose the input signal into independent signal components.

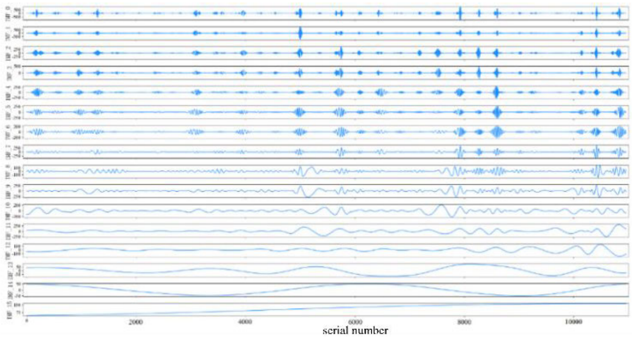

FIGURE 1. Empirical mode decomposition of rainfall flow data.

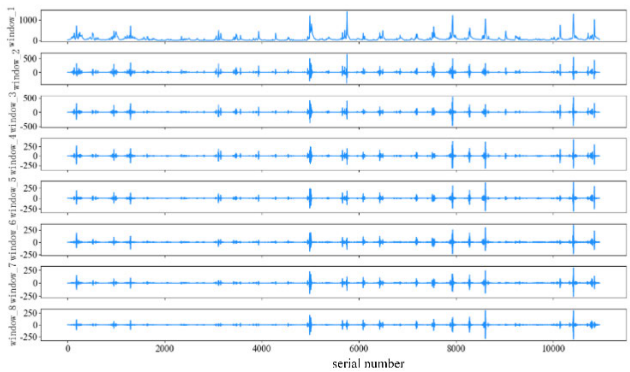

FIGURE 2. Singular spectrum analysis of rainfall flow data.

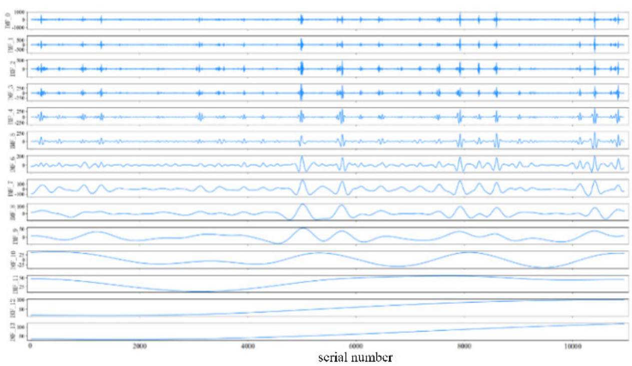

FIGURE 3. Empirical mode decomposition of rainfall flow data.

The daily flow data of Xianyang Station is decomposed by SSA, and the number of Windows ''L'' is set as 8. The decomposition results are shown in Figure 2.

c: SINGULAR SPECTRUM ANALYSIS

On the basis of EMD decomposition, EEMD [13] introduced ''Gaussian white noise a(t)''. a(t) is uniformly distributed in the whole sequence with a mean of zero, and its influence on the original sequence is eliminated after multiple calculations, so as to solve the mode aliasing problem of EMD decomposition.

The daily flow data of Xianyang Station is decomposed by EEMD, the decomposition results are shown in Figure 3.

d: VARIATIONAL MODE DECOMPOSITION

VMD [14] is an improvement on the basis of EMD algorithm. Through the adaptive solution of the structurally constrained variational equation, the signal components are effectively decomposed into a series of single-oscillation modal components.

The daily flow data of Xianyang Station is decomposed by VMD, the decomposition results are shown in Figure 4.

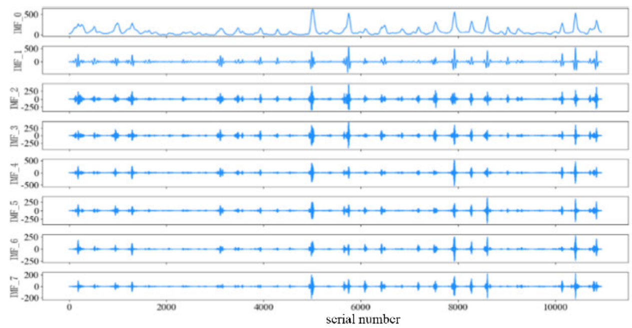

FIGURE 4. Variational modal decomposition of rainfall flow data.

2) PREDICTOR LAG

The correlation of predictive factor series and its lag series is an important content in feature mining of time series prediction data. The higher the correlation of data, the better the prediction accuracy. The correlation of series with different lags varies greatly, which has a significant impact on the accuracy of subsequent prediction. The predictive factor sequences of different lags were constructed, and the relationship between sequence correlation and lag was observed to select the lag period with the strongest correlation. The hysteresis correlation diagram of predictive factor sequence and the autocorrelation analysis diagram of predictive target are shown in Figure 5.

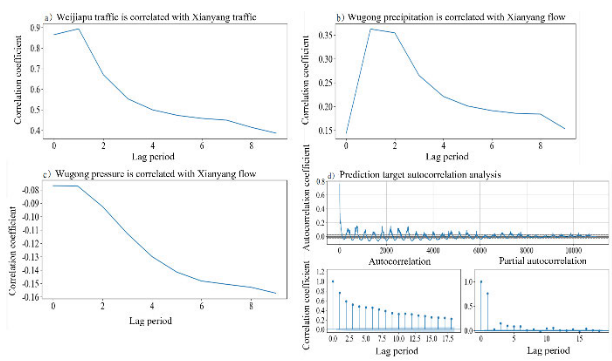

FIGURE 5. Prediction target autocorrelation and partial correlation diagram.

As can be seen from Figure 5, different data have different lag periods and different correlation coefficients. In subfigure a), the correlation coefficient between Xianyang flow and Weijiapu flow decreases with the increase of lag time, and the highest correlation coefficient is close to 0.9. In subfigure b), the highest correlation coefficient between Xianyang discharge and Wugong precipitation is 0.35, and the lag time is 1. In subfigure c), the correlation between Xianyang flow rate and martial arts pressure is poor, and the correlation is negative. Subfigure d) is the autocorrelation analysis of Xianyang flow sequence.

C. RUNOFF PREDICTION MODEL

The multi-factor model considers other key factors other than the forecast target, and comprehensively analyzes the statistical law between each factor and the forecast object, so as to build the model for prediction. The multi-factor prediction models mainly studied in this paper include Prophet model, long and short term memory neural network model, multiple regression model, random forest regression model, gradient lifting regression model and support vector regression.

1) PROPHET MODEL

Prophet [15] is a time series prediction model proposed by FaceBook in 2018. Prophet model has a wide range of flexible applications. After users set parameters, the model will automatically complete manual model construction, prediction evaluation, problem description and visual inspection prediction, and constantly cycle correction.

2) LONG SHORT-TERM MEMORY

LSTM [16] is a cyclic neural network, which is modified on the basis of RNN so that it can learn long-term dependent information.

3) MULTIPLE LINEAR REGRESSION

When multi-factor time series data are available, the linear regression relationship between multiple variables and prediction target variables can be established, and the target variables can be predicted from the combination of multiple variables.

4) RANDOM FOREST REGRESSION

RFR [17] is a regression model based on RF algorithm of random forest. RFR model adopts decision binary tree. The RFR model is trained to make the binary tree continuously slice variables and points, measure the quality of the chopped variables and points, and use the exhaustive method to traverse all the features and their values, and finally find the optimal slice variables and points.

5) GRADIENT BOOSTING REGRESSION

GBR [18] algorithm is derived from gradient descent algorithm. In the process of machine learning, continuous training needs to minimize the loss function $L(\theta)$, θ is the parameter of the solution required by the model. Gradient descent method is an iterative algorithm, which needs to initialize θ_0 and iterate continuously until the loss is minimized. The gradient lift method is similar to the gradient descent method in that the gradient value is iteratively raised to minimize losses.

6) SUPPORT VECTOR REGRESSION

SVR [19] is the application of support vector machine (SVM) in the field of regression estimation of nonlinear systems. The main processes of SVR include variable selection and variable processing. In this paper, RBF is selected as the kernel function of SVR.

D. EVALUATION INDEX AND MODEL PARAMETER OPTIMIZATION

1) EVALUATION INDEX

In the sampling of model parameters and the evaluation of prediction results, the evaluation indexes used were Nash efficiency coefficient (NSE), mean absolute error (MAE), root mean square error (RMSE) and mean percentage error (MAPE). The mean square error (MSE) was selected as the evaluation index for model training and pruning. The closer the value of NSE is to 1, the higher the prediction accuracy is, and the smaller the value of other indexes is, the higher the prediction accuracy is.

2) MODEL PARAMETER OPTIMIZATION

The model parameters were sampled by grid search GS, random search RS, TPE algorithm and CMA-ES algorithm. Taking the SSA-GBR model as an example, a total of 30 years of Xianyang average flow sequence from 1990 to 2019 was used, the rate period was from 1990 to 2013, and the validation period was from 2014 to 2019. Table 1 shows the prediction results of the four sampling algorithms under different iterations. The objective function of the sampling algorithm is the maximum NSE.

The analysis of Table 1 shows that: 1) When the number of iterations is small, TPE can better find the optimal parameter set; 2) When the number of iterations is sufficient, RS and GS can find a better set of parameters.

TABLE 1. Evaluation table of parameters optimization calculation results of SSA-GBR model.

Sampling algorithm	50 iterations					200 iterations				
	Evaluation index of SSA-GBR validation period				Time consuming	Evaluation index of SSA-GBR validation period				Time consuming
	NSE	MAE	RMSE	MAPE		NSE	MAE	RMSE	MAPE	
RS	0.54	41.45	69.49	0.58	7s	0.64	41.9	61.42	0.7	33s
GS	0.61	40.02	64.3	0.59	7s	0.62	40.3	62.88	0.58	33s
TPE	0.63	39.02	62.34	0.57	10s	0.64	38.5	59.77	0.56	40s
CMA-ES	0.52	40.74	70.89	0.61	9s	0.57	40.2	67.44	0.59	35s

TABLE 2. Parameter optimization calculation loss table.

Pruning algorithm	Test 1		Test 2		Test 3		Test 4		Test 5		Mean value	
	Time	MSE	Time	MSE	Time	MSE	Time	MSE	Time	MSE	Time	MSE
Primitive	807	2910	863	2122	853	2467	819	3152	846	2610	838	2652
Median pruning	350	3815	347	2541	362	3003	328	3401	278	2657	333	3083
Threshold pruning	331	3816	344	3126	351	2891	298	2797	291	3726	323	3271
HyperBand	175	3855	190	4160	159	5639	164	3578	207	4687	179	4384
ASHA	84	5352	41	5633	57	5633	59	5426	52	4278	59	5264

TABLE 3. Combination model rate periodic and validation period NSE evaluation index results.

Model / Decomposition	Rate period						Verification period					
	SVR	LSTM	GBR	RFR	Prophet	MLR	SVR	LSTM	GBR	RFR	Prophet	MLR
Raw data	0.44	0.40	0.46	0.46	0.37	0.34	0.33	0.36	0.40	0.40	0.26	0.32
SSA	**0.92**	0.90	**0.99**	**0.98**	**0.85**	**0.84**	**0.92**	**0.90**	**0.87**	**0.81**	**0.84**	**0.83**
VMD	0.89	**0.92**	0.98	0.97	0.80	0.80	0.87	0.86	0.81	0.77	0.77	0.76
EEMD	0.85	0.86	0.93	0.97	0.74	0.74	0.80	0.71	0.76	0.77	0.70	0.69
EMD	0.83	0.71	0.92	0.95	0.63	0.62	0.65	0.70	0.64	0.58	0.60	0.60
TIME	0.40	0.41	0.61	0.62	0.39	0.37	0.33	0.34	0.38	0.35	0.33	0.33

TABLE 4. Combination model rate periodic and validation period MAE evaluation index results.

Model / Decomposition	Rate period						Verification period					
	SVR	LSTM	GBR	RFR	Prophet	MLR	SVR	LSTM	GBR	RFR	Prophet	MLR
Raw data	27.4	38.5	34.7	34.1	41.0	42.4	32.1	37.3	34.0	32.8	73.1	38.3
SSA	**13.6**	24.1	**6.3**	**6.6**	**24.5**	**24.8**	**16.8**	**24.3**	**18.6**	**19.4**	**28.8**	**29.3**
VMD	15.1	**18.5**	12.4	8.0	28.3	28.4	20.2	23.1	23.8	23.1	33.5	33.7
EEMD	16.8	21.8	20.1	7.5	30.7	30.5	23.6	27.7	26.9	24.5	36.2	36.2
EMD	17.0	29.3	20.8	9.3	36.3	36.4	28.0	32.7	34.7	34.0	38.6	38.1
TIME	28.3	36.5	32.8	25.7	42.2	41.6	31.6	42.7	35.4	35.0	44.3	41.9

TABLE 5. Combination model rate periodic and validation period MAPE evaluation index results.

Model / Decomposition	Rate period						Verification period					
	SVR	LSTM	GBR	RFR	Prophet	MLR	SVR	LSTM	GBR	RFR	Prophet	MLR
Raw data	0.29	0.36	0.90	0.82	1.23	1.82	3.25	4.42	4.05	3.64	10.17	4.52
SSA	**0.22**	1.46	**0.19**	**0.08**	**0.68**	**0.62**	**1.64**	**3.96**	**2.25**	**2.34**	**3.69**	**3.37**
VMD	0.25	**0.60**	0.36	0.11	0.79	0.78	2.43	3.32	3.25	2.95	4.28	3.89
EEMD	0.28	0.66	0.67	0.10	0.92	0.89	3.10	3.06	4.04	4.03	3.85	3.66
EMD	0.26	0.89	0.65	0.12	1.04	1.08	3.12	2.46	3.48	3.80	3.89	3.75
TIME	0.28	0.89	0.83	0.36	1.27	1.22	3.47	6.76	4.92	4.05	6.64	6.80

The key to the performance of machine learning algorithm depends on a set of determined super-parameters. The above mentioned parameter configuration can be selected adaptively by a variety of sampling algorithms, but there is still the problem of resource waste. Therefore, it is necessary to accelerate random search through adaptive resource allocation and stop in advance. In order to avoid too many iterations, it is very necessary to automatically stop the hopeless test at the early stage of training. When the training loss is always at the same level, it is considered that continuing training will not bring better results, so parameter optimization, namely pruning, should be ended. In this study,

TABLE 6. Combination model rate periodic and validation period RMSE evaluation index results.

Model / Decomposition	Rate period						Verification period					
	SVR	LSTM	GBR	RFR	Prophet	MLR	SVR	LSTM	GBR	RFR	Prophet	MLR
Raw data	115	107	113	113	122	125	137	134	130	129	144	138
SSA	**43.9**	43.5	**9.7**	**23.8**	**60.4**	**60.9**	**48.0**	**52.4**	**60.9**	**76.7**	**65.4**	**65.6**
VMD	50.2	**39.1**	21.1	26.5	68.3	68.6	61.2	62.9	73.0	79.6	79.9	80.0
EEMD	59.7	52.3	39.6	26.4	78.0	78.2	74.6	90.5	81.3	79.4	92.0	91.9
EMD	62.8	74.3	43.5	32.9	94.2	94.8	98.3	92.3	101	109	106	105
TIME	119	106	96.2	95.1	120	122	137	136	132	135	136	136

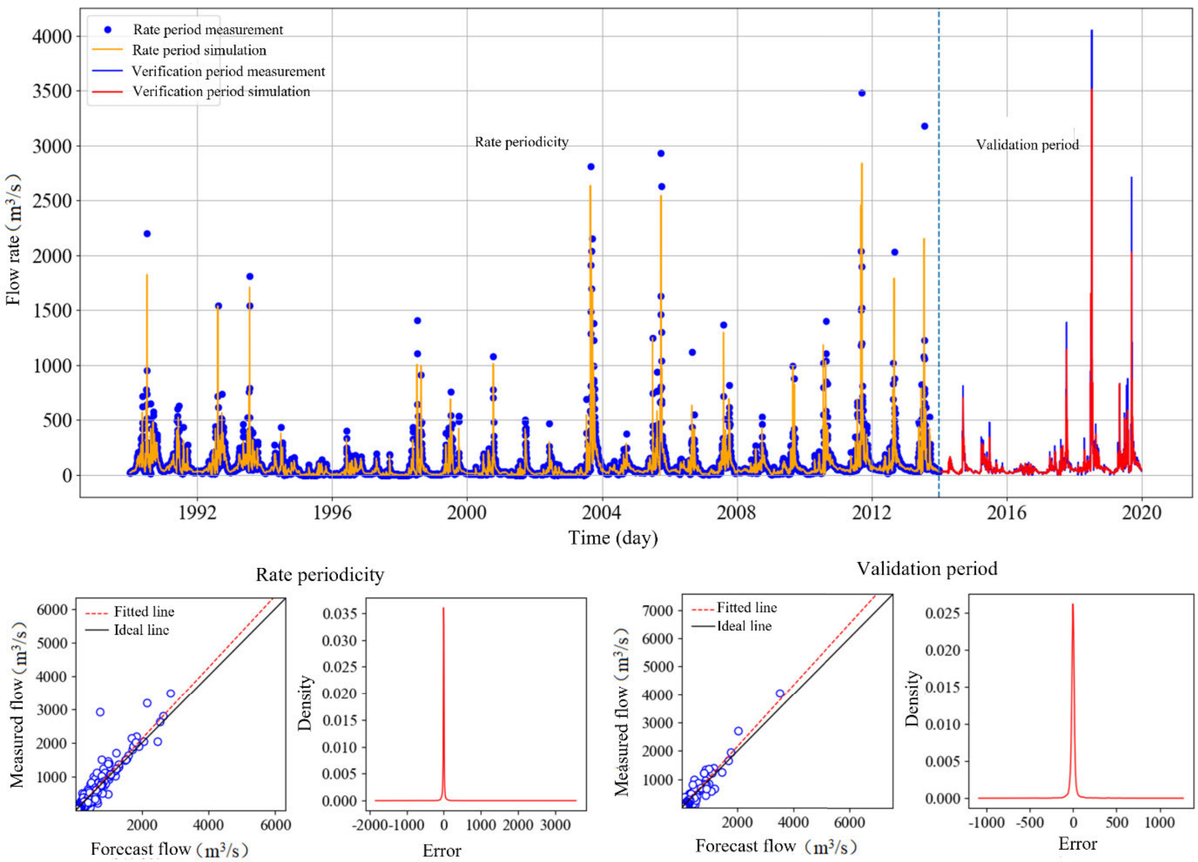

FIGURE 6. SSA-SVR flow measurement and simulation comparison chart.

four commonly used pruning algorithms were adopted: median pruning algorithm, threshold pruning algorithm, Hyperband algorithm and ASHA algorithm.

Training pruning is the main way to reduce the calculation loss of parameter optimization. Taking LSTM model as an example, the average flow data of Xianyang from 1990 to 2019 were taken to optimize the model parameters. The parameter sampling algorithm was TPE algorithm, and the optimization objective was the minimum MSE, batch size of batch training sample was 32. The number of training epochs is 100. In order to reduce the randomness of the sampling path of the parametric sampling algorithm, five tests were conducted on each pruning algorithm, and the average value was taken to compare the advantages and disadvantages of each algorithm. The preheating training times of median and threshold pruning algorithms were 20, and the comparison pruning was performed every 5 training sessions. Table 2 shows the calculation loss table of parameter optimization.

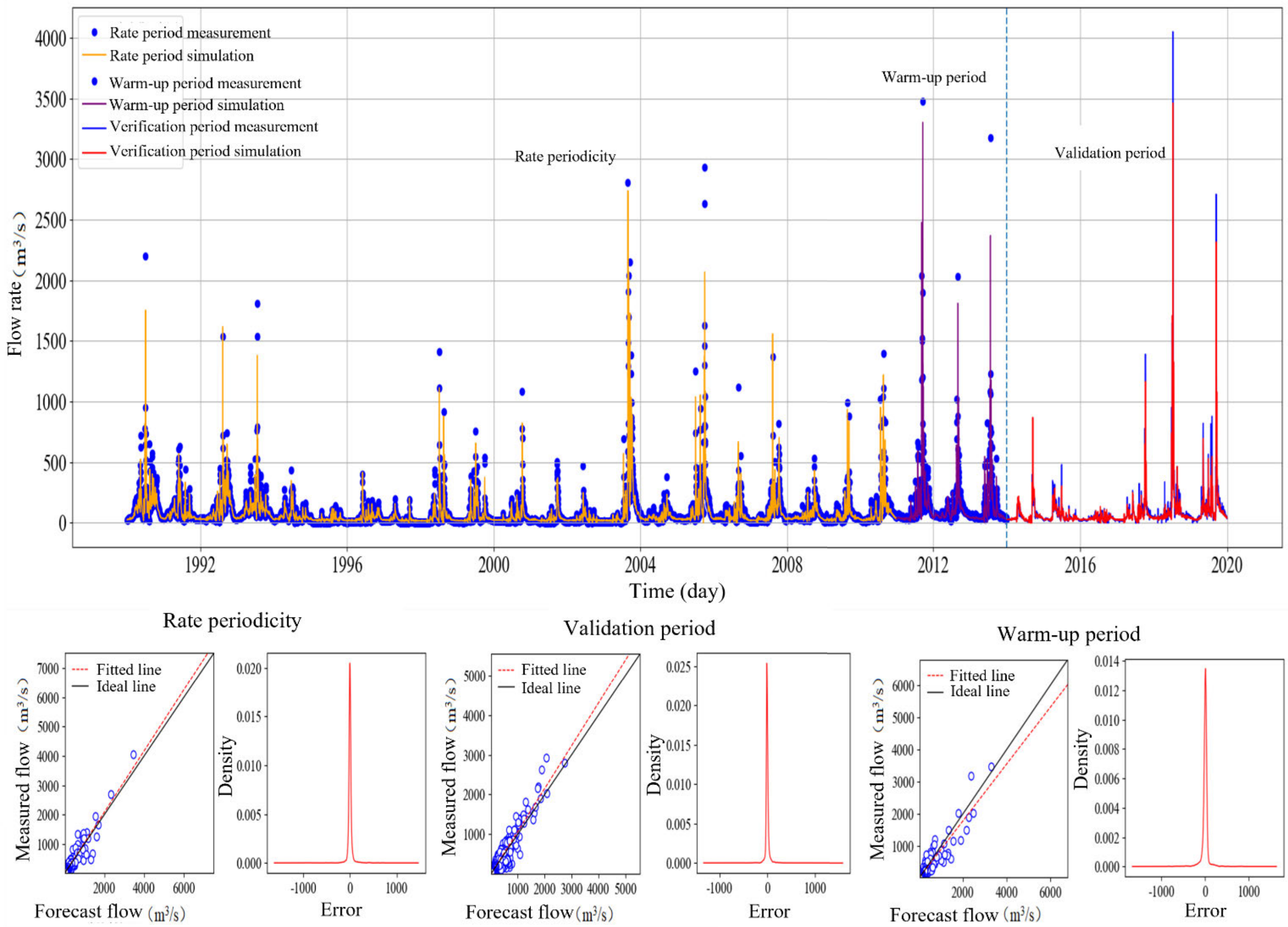

FIGURE 7. SSA-LSTM model flow measurement and simulation comparison chart.

It can be seen from the Table 2 that: 1) the use of pruning algorithm can greatly reduce the time loss of parameter optimization and avoid spending time on hopeless parameter configuration; 2) Median and threshold pruning need to be preheated for training, and the calculation time of these two algorithms is relatively long; 3) For the MSE error, if the threshold of the threshold pruning algorithm is set too large, most of the parameter configurations will be pruned away, while if the threshold is set too small, the hopeless parameter configurations cannot be pruned away. 4) ASHA pruning algorithm can reduce the calculation time of parameter optimization to the shortest, but the accuracy of calculation results is lower than the other three algorithms.

III. RESULTS AND DISCUSSION

The runoff meteorological data is used to forecast the medium and long term runoff, with the periodic rate accounting for 80% of the whole time series length and the verification period accounting for 20%. The runoff prediction results are evaluated through the comprehensive evaluation index system, and the model with the best prediction effect is selected.

The multifactor prediction was calculated by using the data feature mining of Xianyang discharge, Weijiapu discharge and Wugong meteorological (precipitation, pressure) daily data from 1990 to 2019. The prediction period was 3 days, in which the periodic data were from 1990 to 2013 (24 years), and the verification period was from 2014 to 2019 (6 years). TPE algorithm was used to optimize the model parameters, and the number of iterations was 50. Table 3 to Table 6 are the evaluation results of all combination models with evaluation indexes of NSE, MAE, MAPE and RMSE, respectively. Figure 6 and Figure 7 show the measured simulation comparison of predicted flow of SSA-SVR and SSA-LSTM models, the two combination models with the highest prediction accuracy.

Conclusions can be drawn from the table and figure:

(1) Among the five data decomposition algorithms, SSA, VMD and EEMD greatly helped to improve the accuracy of the simulation prediction results. Compared with the undecomposed raw data, NSE value increased to at least 0.69, and SSA greatly helped to improve the prediction accuracy. EMD and TIME have little effect on the accuracy of prediction results, and TIME has the worst effect.

(2) Among the six prediction models, SVR, LSTM and GBR models have higher prediction accuracy. Compared with other models, they have higher adaptability to the six data sets, and the accuracy of prediction indicators is relatively higher.

(3) Among the 30 combination models, the NSE index of the prediction results of the combination models using SSA and VMD algorithms for data decomposition is above 0.75 in the verification period. Notably, SSA-SVR and SSA-LSTM models delivered higher prediction accuracy. Specifically, the validation period NSE, MAE, MAPE, and RMSE indexes of the SSA-SVR model were 0.92, 16.8, 1.64 and 47.95, respectively, while the corresponding indexes of SSA-LSTM were 0.90, 24.3, 3.96, and 52.4. The refined writing follows standard academic style guidelines, enhancing concision, clarity, and readability.

This study considered a comprehensive range of decomposition algorithms and model combinations for the prediction of runoff. Additionally, multiple evaluation indexes were utilized to assess the accuracy of the predictions, leading to the selection of the combination model with the highest precision. The default parameters for the decomposition algorithm were utilized in the data analysis and processing, thus limiting adaptability and adjustability. Although efforts were made to enhance the suppleness of model integration and parameter adjustment, there is still ample room for further improvement.

REFERENCES

[1] R. C. Deo, O. Kisi, and V. P. Singh, "Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model," *Atmos. Res.*, vol. 184, pp. 149–175, Feb. 2017.

[2] A. S. Tokar and P. A. Johnson, "Rainfall-runoff modeling using artificial neural networks," *J. Hydrologic Eng.*, vol. 4, no. 3, pp. 232–239, Jul. 1999.

[3] L. Lingjie, W. Yintang, H. Qingfang, L. Dingzhong, and Z. Anfu, "Long term runoff forecast of reservoir based on random forest and support vector machine," *J. Water Resour. Transp. Eng.*, vol. 4, pp. 33–40, Apr. 2020.

[4] G. Junbao, Y. Zhongbo, Y. Chuanguo, and X. Shiqin, "A study on medium and long term runoff prediction model based on physical causes," *Hydroelectr. Energy Sci.*, vol. 38, no. 5, pp. 35–37, 2020.

[5] L. Yan, Y. Yun, N. Lei, and S. Qiuyu, "EEMD-ARIMA prediction of runoff at the Manas River outlet," *Res. Soil Water Conservation*, vol. 24, no. 6, pp. 273–280, 2017.

[6] Ö. Kişi, "A combined generalized regression neural network wavelet model for monthly streamflow prediction," *KSCE J. Civil Eng.*, vol. 15, no. 8, pp. 1469–1479, Nov. 2011.

[7] S. Weiyao, "Research on the prediction of Weihe River main stream runoff based on EEMD and deep learning," Water Conservancy Project, Xi'an Univ. Technol., Xi'an, China, Tech. Rep., 2020, doi: 10.27398/d.cnki.gxalu.2020.000845.

[8] Z. Guoyong, W. Yonggang, Y. Linming, and W. Pengfei, "EEMD-LSSVM annual runoff combination prediction model based on parameter optimization," *J. Water Resour. Water Eng.*, vol. 24, no. 6, pp. 1–5, 2013.

[9] Y. Guohui, Q. Changlu, and C. Fulong, "Annual runoff prediction of Manas River based on EMD-WDD-MK model," *China Rural Water Resour. Hydropower*, vol. 11, pp. 83–89, Mar. 2021.

[10] T. Xie, G. Zhang, J. Hou, J. Xie, M. Lv, and F. Liu, "Hybrid forecasting model for non-stationary daily runoff series: A case study in the Han River Basin, China," *J. Hydrol.*, vol. 577, Oct. 2019, Art. no. 123915.

[11] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc., Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[12] R. Vautard, P. Yiou, and M. Ghil, "Singular-spectrum analysis: A toolkit for short, noisy chaotic signals," *Phys. D, Nonlinear Phenomena*, vol. 58, nos. 1–4, pp. 95–126, Sep. 1992.

[13] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[14] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[15] G. A. Papacharalampous and H. Tyralis, "Evaluation of random forests and prophet for daily streamflow forecasting," *Adv. Geosci.*, vol. 45, pp. 201–208, Aug. 2018.

[16] W. Yuanhao, "Global temperature prediction analysis based on ARIMA model and LSTM neural network," *Sci. Technol. Innov.*, vol. 35, pp. 166–170, Jun. 2021.

[17] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006.

[18] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[19] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

HAO LIU received the M.S. degree in hydrology and water resources from the Xi'an University of Technology, China, in 2022. He is currently an Engineer with Changjiang Water Resources Commission, Wuhan, China. He is also an Engineer at Changjiang Schinta Software Technology Company Ltd., Changjiang Schinta Software Technology Company Ltd., is a subsidiary of Design Group. His research interests include machine learning and runoff prediction.

WEI LIU received the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China. He is currently an Engineer with Changjiang Water Resources Commission, Wuhan. He is also an Engineer at Changjiang Schinta Software Technology Company Ltd., Changjiang Schinta Software Technology Company Ltd., is a subsidiary of Design Group. His research interests include deep learning and data mining.

JUNGANG LUO received the Ph.D. degree from the Xi'an University of Technology, China, in 2009. He is currently a Professor of hydrology with the School of Water Resources and Hydropower, Xi'an University of Technology. His research interests include water conservancy informatization and water resource management and allocation decision-making.

JING LI received the M.S. degree in hydrology and water resources from the Xi'an University of Technology, China, in 2022. She is currently an engineer with Hydrology Bureau of Changjiang Water Resources Commission, Wuhan, China. Her research interests include machine learning and urban hydrology.