Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2022.030101
ISSN 2790-0983 Vol. 3, Issue 1: 1-9

Medium and Long Term Runoff Forecast of Huaihe River Basin Based on Machine Learning

Yani Mai and Ku Yang*

Guangzhou College of SCUT, Guangzhou, China

**corresponding author*

Keywords: Machine Learning, Huaihe River Basin, Medium and Long-Term Runoff Forecasting, BOA-EEMD-LSTM Model

Abstract: Water resources are the source of life and the basis of ecology, and medium- and long-term runoff forecasting plays an important role in the overall planning of water resources. However, complex factors such as climate change affect the formation of runoff, making the runoff process more complex and posing a great challenge to hydrological forecasting. Therefore, this paper investigates the medium- and long-term runoff forecasting in the Huaihe River basin based on machine learning. The paper takes the improvement of medium- and long-term runoff prediction accuracy as the background, and combines the specific requirements of practical projects and research ideas to characterise the runoff series in the Huaihe River basin. The results of runoff simulation and forecasting under future climate change are also analyzed using the BOA-EEMD-LSTM model.

1. Introduction

Runoff prediction is an important element in hydrological forecasting research. Its prediction results can provide a basis for flood and drought prevention, reservoir scheduling and hydropower generation, and the optimal allocation of water resources in the basin and regional development planning require the spatial distribution and dynamic changes of water resources [1-2]. Medium and long-term hydrological forecasts are scientific predictions of future hydrological conditions over a longer period of time based on preliminary hydro-meteorological data [3]. As an important part of disaster mitigation and prevention, medium- and long-term hydrological forecasting has been of great interest to the national economy [4]. Runoff forecasting involves multiple disciplines and fields, and the analysis of the spatial and temporal characteristics of runoff in the target basin is a prerequisite for medium- and long-term forecasting [5-6].

With the development of industry in China, water resources problems are becoming more and more prominent, and in this context many scholars have conducted research on medium- and long-term runoff forecasting with good results [7]. For example, Sara Parvinizadeh et al. proposed a high-precision runoff prediction model based on machine learning that would improve early warning capability for floods and droughts, in which the GRNN model was selected as the optimal runoff prediction model, and the GRNN model used flood propagation time to predict flow and water level, and the results showed that the GRNN model performed well in runoff prediction [8]. A hybrid model for monthly runoff time series prediction was developed by Umut Keskin et al. A grey wolf optimiser was used to optimise the input hidden weights of the ELM method and a generalised inverse method was used to determine the hidden output weights, comparing the performance of various prediction methods, and simulation results showed that the method outperformed traditional prediction methods in several quantitative metrics [9]. The use of machine learning for prediction of runoff is a direction worthy of further investigation.

The increasing prominence of water environment and water resources problems in the basin hinders coordinated regional development, so this paper investigates medium- and long-term runoff forecasting in the Huaihe River basin based on machine learning [10]. The main content of this paper is divided into three parts: the first part is the construction of the runoff prediction model, which mainly includes the setting of the simulation scenario and the model workflow; the second part is the single model analysis, which is divided into two parts of the research analysis, including the analysis of the influencing factors of runoff change and the single model prediction analysis; the third part is the analysis of the runoff prediction results, which mainly investigates the model validity and the future climate change under The third part is the analysis of runoff forecasting results, which focuses on the model validity and future runoff forecasting under climate change, and verifies the validity of the BOA-EEMD-LSTM model through the analysis.

2. Construction of the Runoff Prediction Model

2.1. Setting of the Simulation Scenario

The core of the hydrological modelling approach is the reliability of the model and the strict control of variables [11]. In order to optimise the applicability of the model, a base period with a more moderate climate change over a longer time span and closer to the natural properties of the basin is used to derive the corresponding model parameters, which are then applied in subsequent simulations to ensure that the runoff values obtained from the model are closer to the natural runoff generated by the corresponding scenario [12-13]. In the comparative analysis, the most reasonable set of parameters was used to carry out the simulations and ensure that the input conditions were identical for each scenario except for the input variables to be studied, to ensure that the model simulation errors were kept to a minimum [14-15]. Scenarios K1 and K3 were set to differ only in meteorological data, and only in land use between scenarios K2 and K3.

The model simulation results in unaffected natural runoff. The difference between the measured flow L2 in the change period (2007-2022) and the measured flow L1 in the base period (1995-2006) is the value of runoff change due to the combined effect of climate change, land use change and human activities. The specific calculations are as follows.

$$\Delta L_A = P_1 - P_2 \tag{1}$$

$$\Delta L_B = P_{K3} - P_{K1} \tag{2}$$

$$\Delta L_C = P_{K2} - P_{K3} \tag{3}$$

$$\Delta L_D = \Delta L_A - \left(\Delta L_B + \Delta Q_C\right) \tag{4}$$

L_{K1}, L_{K2} and L_{K3} represent the annual runoff volumes (m3/s) simulated by the SWAT model under the K1, K2 and K3 hydrological simulation scenarios, respectively. The unit of runoff data used in the calculation is the multi-year average runoff (m3/s).

2.2. Model Workflow

This section proposes a prediction model approach based on a combination of Bayesian optimization, an integrated empirical modal decomposition algorithm and a long and short term memory neural network. The Bayesian optimization algorithm is used to find the hyperparameters in the algorithm, the EEMD algorithm is used to process the data, and finally the LSTM is used to predict each sub-series and synthesize the components to obtain the final runoff prediction results. The BOA-EEMD-LSTM model is based on the "optimisation-decomposition-prediction-synthesis" process, and the model workflow is shown in Figure 1.

Step 1: Pre-process the data, read the runoff series, and normalise the data set [16].

Step 2: The BOA algorithm is used to optimise the hyperparameters of the LSTM network model to obtain the optimised hyperparameter result values [17].

Step 3: Decomposition of the runoff data into residual series Res and K intrinsic modal functions IMF using signal processing EEMD.

Step 4: The runoff sequence components are input into the LSTM for training, and the trained model is validated using a test set.

Step 5: The data are subjected to an inverse normalisation operation and the accuracy evaluation metric is used for model evaluation [18].

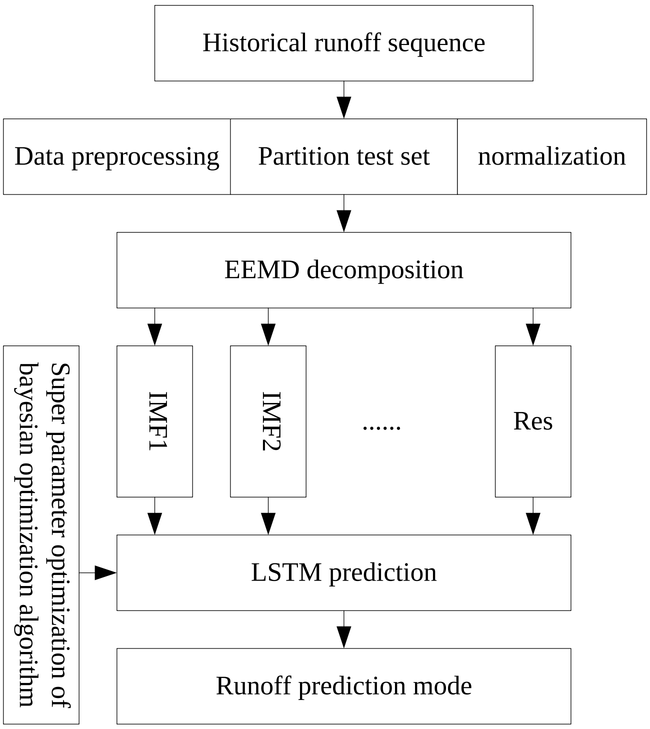

Figure 1. Flow chart of runoff prediction

3. Single Model Analysis

3.1. Analysis of Factors Influencing Runoff Changes

The causes of runoff changes in the Huaihe River basin are briefly analysed through the changes in rainfall characteristic values and flood characteristic values. Indicators such as multi-year average rainfall, rainfall ephemeris, average rainfall intensity, flood ephemeris, receding ephemeris, flood peak flow, minimum daily flow during flood season and basin storage capacity were selected to analyse the causes of changes in runoff mechanisms. The values of the variation of each indicator are shown in Table 1.

Table 1. Variation in flow production at different stages

	Base period	Change period	Rate of change(%)
Average annual rainfall (mm)	876.4	867.9	-0.97
Rainfall duration (h)	64.7	69.1	6.8
Flood duration (h)	275.3	297.5	8.06
Water recession duration (h)	20.8	24.2	16.35
Peak discharge (m^3/s)	865.6	824.8	-4.71
Minimum daily flow in flood season (m^3/s)	1.55	1.87	17.11
Basin storage capacity (mm)	79.68	88.63	11.23

As can be seen from Table 1, the rainfall factors do not change much in both the base and change periods, with rainfall only decreasing by 0.97% in the change period relative to the base period, and factors such as rainfall ephemeris not changing much in the two periods, but flood ephemeris and receding ephemeris increase by 8.06% and 16.35%, and the trends in rainfall and flood characteristic values such as peak flow and flood ephemeris are not consistent, indicating that rainfall is not the This indicates that rainfall is not the main influence on the change of these flood characteristics. In addition, according to the data in Table 1, the minimum daily flow during the flood season increased by 17.11% over the two phases of change, and the basin storage capacity increased from 79.68mm to 88.63mm, and the changes in these characteristic values are closely related to the increase in the area of vegetation such as forest and grass. The increase in vegetation leads to an increase in soil water retention, resulting in an increase in flood ephemeris, a decrease in peak flow and an increase in basin storage capacity, thus inferring that changes in substrate conditions in the Huaihe River basin are the main cause of the changes in flood characteristics and the main factor in the changes in runoff regime in the Huaihe River basin.

3.2. Single Model Prediction Analysis

In recent years, with the continuous development of computer technology, artificial intelligence algorithms have been extended to various fields, and have been applied in the field of runoff prediction. SVR, SARIMA and LSSVM are representative machine learning algorithms, and single model prediction is carried out for these three methods and described for each model separately. The results of the single model prediction comparison are shown in Table 2.

As observed in Table 2, the SARIMA model has the highest NSE and r values among the three models. The Nash efficiency coefficient NSE of the SARIMA model increased by 92.59% and 51.27% compared to the SVR and LSSVM models respectively, and the correlation coefficient R of the SARIMA model increased by 45.48% and 33.10% compared to the ELM and LSSVM models respectively. This represents the higher accuracy of the SARIMA model runoff prediction and the stronger correlation between the predicted and original series. Of the three models, the SARIMA model achieved the lowest values of RMSE, MAE and MAPE. SARIMA model is relatively accurate because it takes into account the seasonality factor, but the traditional time series model SARIMA also has disadvantages such as: high data requirements (requiring the time series to be stable, or stable after differencing), prediction accuracy decreases with time, focusing on univariate data with linear relationships and fixed manual diagnostic time dependence, etc.; low SVR generalisation error rate, low computational overhead, easy to interpret the results, but the parameter selection has a large impact; LSTM is simple to implement and has better convergence, and is less prone to gradient disappearance or explosion. The LSTM is simple to implement, has better convergence, and is less prone to problems such as gradient disappearance or explosion, and does not require operations such as data differencing compared to SARIMA, but requires certain requirements for model tuning, and the model computation time is relatively long, so some subsequent optimisation is required to obtain better prediction results. Therefore, a single prediction model is not suitable for medium and long-term runoff forecasting.

Table 2. Single model prediction error comparison

	SVR	LSTM	SARIMA
R	0.398	0.435	0.579
NSE	0.216	0.275	0.416
RMSE(m^3)	5346.72	4187.44	3862.75
MAPE(m^3)	0.379	0.275	0.125
MAE(m^3)	3316.78	2978.25	2153.89

4. Analysis of Runoff Forecasting Results

4.1. Analysis of Model Validity

The prediction accuracy of all prediction models was compared and the results are shown in Table 3.

Table 3. Comprehensive comparison results

	RMSE(m^3)	MAE(m^3)	MAPE(m^3)	DC	Accuracy judgment
BOA-EEMD-LSTM	2057.68	1657.34	0.114	0.97	A
EMD-LSTM	2925.75	2146.77	0.138	0.88	B
EEMD-BP	2675.36	1846.53	0.119	0.91	B
SVR	5346.72	3316.78	0.379	0.72	C
LSTM	4187.44	2978.25	0.275	0.78	C
SARIMA	3862.75	2135.89	0.125	0.85	B

As can be seen from Table 3, the BOA-EEMD-ISTM model can fit the data better, and in runoff prediction, the forecast accuracy is divided into four grades: ABC and unqualified forecast, which represent the forecast accuracy grade in actual forecasting work. the BOA-EEMD-ISTM model achieves a forecast accuracy of 0.97, with an accuracy grade of A, which is higher than the single model SVR, LSTM, and The forecast accuracy of the BOA-EEMD-ISTM model is the highest among all the models, which indicates that it can reach the level of qualified forecasts in practical applications. Overall the BOA-EEMD-LSTM model has the best prediction accuracy, proving the effectiveness of the proposed model.

4.2. Analysis of Runoff Forecasting under Future Climate Change

Precipitation, maximum and minimum temperatures for the future period (2022-2030) were used as input data for the meteorological variables in the BOA-EEMD-LSTM model, and the BOA-EEMD-LSTM model was used to simulate future runoff. The analysis of future runoff changes used the observed values from 1995-2006 as the base period runoff, and the simulation scenarios were as described in 2.1 of this paper. The results of future runoff changes are shown in Figure 2, and the future monthly runoff changes are shown in Figure 3.

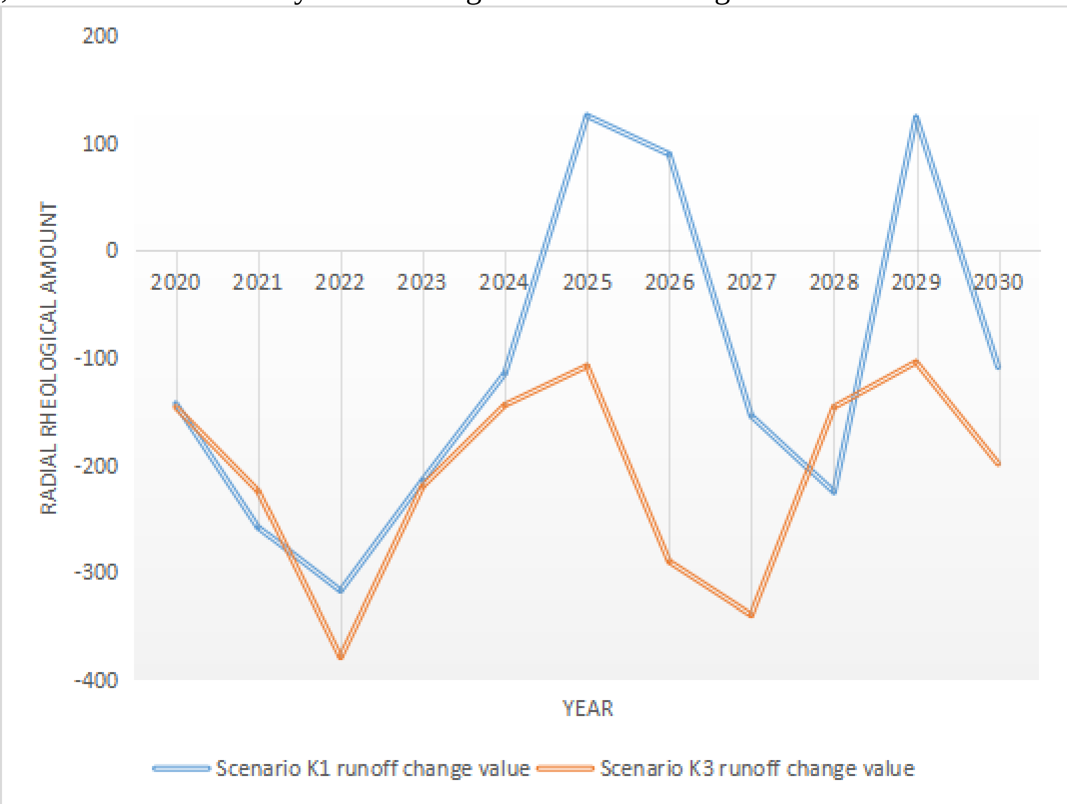

Figure 2. Changes in runoff under different scenarios

It can be seen from Figure 2 that the future runoff for scenario K1 has large interannual fluctuations, with the years 2022, 2023 and 2024 decreasing by 316.78 m3/s, 213.79 m3/s and 114.24 m3/s, respectively, and the years 2027 and 2028 decreasing by 153.75 m3/s and 224.68 m3/s, respectively, while the years 2025, 2026 and 2029 runoff shows an increasing trend, and its changes are generally consistent with the precipitation trends in the corresponding years. Scenario K3 shows small inter-annual fluctuations in future runoff, all showing a decreasing trend. The lower future runoff in 2022, 2026 and 2027 indicates a higher probability of drought and a significant reduction in river runoff in those years. Overall, scenarios K1 and K3 show an overall decreasing trend in future runoff.

As can be seen in Figure 3, compared to the base period, the greatest reduction in runoff is seen in June and July for scenarios K1 and K3, with June runoff at 977 m3/s for scenario K1 and 876 m3/s for scenario K3, and July runoff at 875 m3/s for scenario K1 and 819 m3/s for scenario K3, with less change in the remaining months. The runoff under both scenarios shows a decreasing trend, which is consistent with the monthly precipitation trend. In addition the runoff in

August-September under the future scenario is larger, increasing the probability of causing autumn floods, suggesting that abnormal climate changes may intensify due to the combined effects of El Niño and La Niña events. Overall, the future scenario shows a general trend of decreasing monthly runoff and an increased probability of autumn flooding.

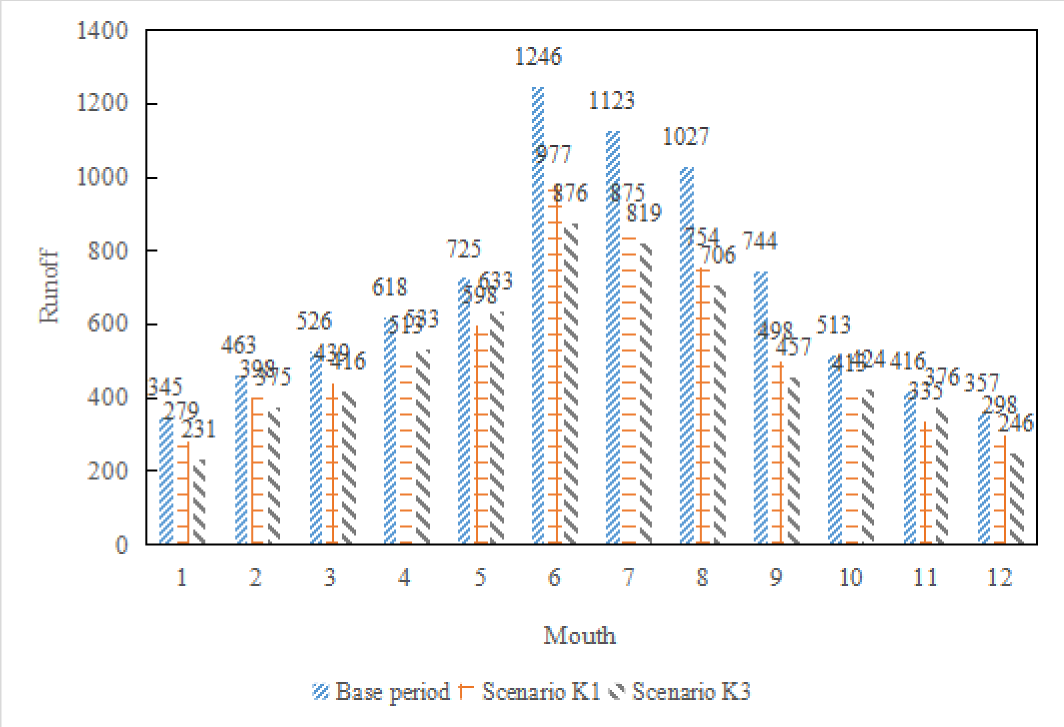

Figure 3. Projected monthly runoff changes

5. Conclusion

As the level of social science and technology continues to improve, the level of China's ability to unify water resources management and allocation and basin water regulation continues to rise, so the reliability of the methods and results of medium- and long-term runoff forecasting face higher requirements, highlighting the importance of medium- and long-term runoff forecasting. In this paper, a BOA-EEMD-LSTM medium- and long-term runoff forecasting model is constructed based on machine learning in the Huaihe River basin. By comparing the model with a single model and a composite model, it is found that the BOA-EEMD-LSTM model has the best prediction accuracy of 0.97, which proves the effectiveness of the model. The overall trend of runoff decreases and the probability of autumn flooding increases. There are many areas for improvement in this paper.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Elisabete A. De Nadai Fernandes, Gabriel A. Sarries , Yuniel T. Mazola, Robson C. de Lima, Gustavo N. Furlan, Marcio A. Bacchi: Machine learning to support geographical origin traceability of Coffea Arabica. Adv. Artif. Intell. Mach. Learn.2(1): 273-287 (2022). https://doi.org/10.54364/AAIML.2022.1118

[2] Cameron Cooper: Using Machine Learning to Identify At-risk Students in an Introductory Programming Course at a Two-year Public College. Adv. Artif. Intell. Mach. Learn. 2(3): 407-421 (2022). https://doi.org/10.54364/AAIML.2022.1127

[3] Dmitry V. Vinogradov: Algebraic Machine Learning: Emphasis on Efficiency. Autom. Remote. Control. 83(6): 831-846 (2022). https://doi.org/10.1134/S0005117922060029

[4] Suresh Dara O, Swetha Dhamercherla, Surender Singh Jadav, Ch Madhu Babu, Mohamed Jawed Ahsan. Machine Learning in Drug Discovery: A Review. Artif. Intell. Rev. 55(3): 1947-1999 (2022). https://doi.org/10.1007/s10462-021-10058-4

[5] Leandro Miranda, Jose Viterbo, Flavia Bernardini: A survey on the use of machine learning methods in context-aware middlewares for human activity recognition. Artif. Intell. Rev.55(4): 3369-3400 (2022). https://doi.org/10.1007/s10462-021-10094-0

[6] Muhammad Waqas, Shanshan Tu, Zahid Halim, Sadaqat ur Rehman, Ghulam Abbas, Ziaul Haq Abbas: The role of artificial intelligence and machine learning in wireless networks security: principle, practice and challenges. Artif. Intell. Rev.55(7): 5215-5261 (2022). https://doi.org/10.1007/s10462-022-10143-2

[7] Athar Hussain, Jatin Kumar Singh, A. R. Senthil Kumar, HarneK. R: Rainfall-Runoff Modeling of Sutlej River Basin (India) Using Soft Computing Techniques. Int. J. Agric. Environ. Inf. Syst. 10(2): 1-20 (2019). https://doi.org/10.4018/IJAEIS.2019040101

[8] Sara Parvinizadeh, Mohammad Zakermoshfegh, Maryam Shakiba 0: A simple and efficient rainfall-runoff model based on supervised brain emotional learning. Neural Comput. Appl.34(2): 1509-1526 (2022). https://doi.org/10.1007/s00521-021-06475-9

[9] Umut Keskin, M. Remzi Sanver, H. Berkay Tosunlu: Monotonicity violations under plurality with a runoff: the case of French presidential elections. Soc. Choice Welf. 59(2): 305-333 (2022). https://doi.org/10.1007/s00355-022-01397-4

[10] Amir Alizadeh, Ahmad Rajabi, Saeid Shabanlou , Behrouz Yaghoubi, Fariborz Yosefvand: Modeling long-term rainfall-runoff time series through wavelet-weighted regularization extreme learning machine. Earth Sci. Informatics 14(2): 1047-1063 (2021). https://doi.org/10.1007/s12145-021-00603-8

[11] Vahid Nourani, Huseyin Gokgekus, Tagesse Gichamo: Ensemble data-driven rainfall-runoff modeling using multi-source satellite and gauge rainfall data input fusion. Earth Sci. Informatics 14(4): 1787-1808 (2021). https://doi.org/10.1007/s12145-021-00615-4

[12] Vikas Kumar Rana, Tallavajhala Maruthi Venkata Suryanarayana: GIS-based multi criteria decision making method to identify potential runoff storage zones within watershed. Ann. GIS 26(2): 149-168 (2020). https://doi.org/10.1080/19475683.2020.1733083

[13] Sandeep Samantaray, Abinash Sahoo: Prediction of runoff using BPNN, FFBPNN, CFBPNN algorithm in arid watershed: A case study. Int. J Knowl. Based Intell. Eng. Syst. 24(3): 243-251 (2020). https://doi.org/10.3233/KES-200046

[14] Adnan Ahmad Tahir , Samreen Abdul Hakeem, Tiesong Hu, Huma Hayat, Muhammad Yasir : Simulation of snowmelt-runoff under climate change scenarios in a data-scarce mountain environment. Int. J. Digit. Earth 12(8): 910-930 (2019). https://doi.org/10.1080/17538947.2017.1371254

[15] Athar Hussain, Jatin Kumar Singh, A. R. Senthil Kumar, Harne K. R: Rainfall-Runoff Modeling of Sutlej River Basin (India) Using Soft Computing Techniques. Int. J. Agric. Environ. Inf. Syst. 10(2): 1-20 (2019). https://doi.org/10.4018/IJAEIS.2019040101

[16] Owen Dafydd Jones: Runoff on rooted trees. J. Appl. Probab. 56(4): 1065-1085 (2019). https://doi.org/10.1017/jpr.2019.61

[17] Amir Salimi , Ali Rafiee: A grid interpolation technique for anomaly separation of stream sediments geochemical data based on catchment basin modelling, U-statistics and fractal. Earth Sci. Informatics 15(1): 151-161 (2022). https://doi.org/10.1007/s12145-021-00712-4

[18] Alison Garza-Alonso, Michael V. Basin, Pablo Cesar Rodriguez-Ramirez: Predefined-Time Backstepping Stabilization of Autonomous Nonlinear Systems. IEEE CAAJ. Autom. Sinica 9(11): 2020-2022 (2022). https://doi.org/10.1109/JAS.2022.105953