Amirkabir Journal of Civil Engineering

Amirkabir J. Civil Eng., 53(7) (2021) 621-624
DOI: 10.22060/ceej.2020.17520.6589

Capability Evaluation of Hybrid Wavelet-Principal Component Analysis-Random Forest Approach in Simulating the River Flow

F. Azarpira, S. Shahabi*

Department of Water Engineering, Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, Kerman, Iran.

ABSTRACT Simulating the flow for managing the water allocation in drought and wet periods is of great importance. According to the researches conducted during several decades in this regard, computational intelligence methods combined with wavelets are known to be effective. In this paper, Wavelet-Principal Component Analysis-Random Forest (WPCARF) hybrid approach is proposed to model the daily flow of the Polroud river. In the proposed model, first, hydrometric data is preprocessed by wavelet transform and applied to the PCA along with meteorological data. Afterward, their output vectors were entered into the random forest network. The results have shown that the PCA algorithm can improve the performance accuracy and speed of the model, despite reducing the input vectors and simplifying them. Also, it can integrate a model with increased simulation time and input vectors uncertainty having a lower impact on model capability leading to a more uniform decreasing trend. Furthermore, preprocessing the data accompanied by PCA could enhance the agreement index by 5 and 8 percent during one and three days of the simulation and increase the model ability for a more accurate simulation of river flow. On the other hand, results for the best-proposed hybrid model during the one-day-ahead simulation time were R=0.911 and RMSE=7.095 m^3/s, while these values were R=0.817 and RMSE=8.681 m^3/s in the best hybrid model for three-day-ahead simulation time. This indicates the adequate capacity of the proposed hybrid model for long-term simulation times.

Review History:

Received: Dec. 11, 2019
Revised: Aug. 28, 2020
Accepted: Aug. 29, 2020
Available Online: Sep. 10, 2020

Keywords:

Daily streamflow
Polroud river
Preprocessing
Simulation time
Time series

1. INTRODUCTION

Streamflow simulation is one of the most vital issues in surface water resources and it is taken into account as the basis for many hydraulic and hydrological planning goals. The streamflow simulation is dependent on many factors, which is not possible to model and formulate the whole of these factors. In this regard, computational intelligence methods have been developed to Figure out the problem [1-4]. Different researchers have found that the use of data preprocessing has regularly improved the performance of these models [5-8]. Additionally, several researchers have developed some methods (e.g., PCA) to select the best inputs for the streamflow simulation model, leading to the model simplification and improvement of its performance [9-10]. In addition to studies that have only employed the hydrometric data for the streamflow simulation model, a combination of hydrometric and metrological data has been utilized in several studies, and they have reported promising results [11-12].

2. MATERIAL AND METHODS

In this paper, the Wavelet-Principal Component Analysis-Random Forest (WPCARF) approach was proposed to simulate the daily streamflow for Polroud river, which is located in the northern part of Iran. To achieve this purpose, hydrometric and hydrological data were utilized for analyzing the process. This approach contains several novel techniques, and thus it has a significant advantage in comparison with similar studies.

In this paper, the streamflow simulation is accomplished by integrating several approaches. These approaches are comprised of the wavelet transform, principal component analysis (PCA), and random forest (RF). Fig. 1 depicts a step-by-step flowchart for the proposed hybrid approach (WPCARF).

The Polroud river basin was investigated in this study. It is situated in the south of the Caspian Sea and it is originated from the Chakroud Mountains, after passing through the Roudsar plain, it eventually flows into the Caspian Sea (Fig. 2).

The used dataset is related to the Toullat hydrometric station and the data have been recorded during the water year of 1977 to 2017. Moreover, the information of the Ramsar synoptic station was employed to gather the metrological parameters. Furthermore, 70% and 30% of data were utilized for training and testing phases, respectively. The proposed models have been evaluated considering four different indicators, including *R*, *RMSE*, *NSE*, and I_a.

*Corresponding author's email: s.shahabi@kgut.ac.ir

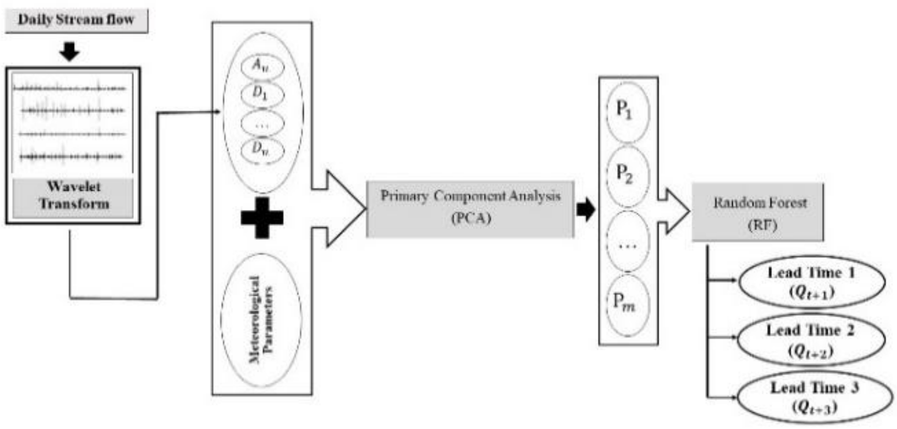

Fig. 1. Schematic diagram of proposed hybrid model

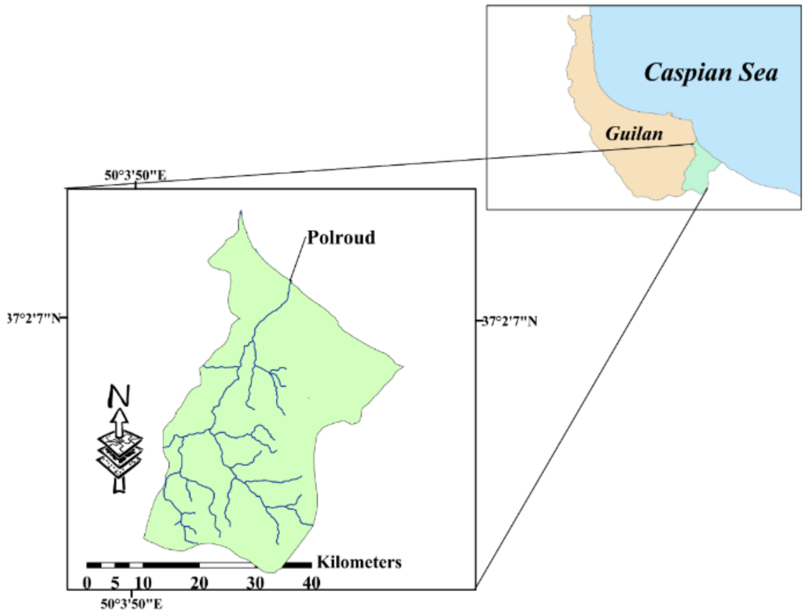

Fig. 2. The geographical location of Polroud river and Roudsar city

3. RESULTS AND DISCUSSION

To perform the modeling process, the hydrometric and metrological data were utilized at five different times (*t*, *t*-1, *t*-2, *t*-3, *t*-4) and a single time (*t*), respectively. In hybrid models, the Haar and Dmey functions were employed to decompose the time series up to the decomposition level of 7. PCA was used to select the most effective input vectors for entering the RF. Table 1 summarizes the best results.

According to Table 1, an increase in the simulation time declines their performance accuracy. However, this process has a much higher effect on the RF model than other hybrid models. The proposed hybrid approach provided a very smaller decrease in the model efficiency and accuracy at higher simulation times than the single models or even other hybrid models proposed by other hydrologists [7]. Although the PCA has significantly simplified the model by decreasing the number of input vectors, it has not reduced the model accuracy.

The comparison of results revealed that the WPCARF model operates better than PCARF in the simulation of the maximum values. It is observed that the proposed model could properly simulate the daily discharge values of the Polroud river. This model has a deficiency, leading to the reduction of model performance. This issue is concerned with simulating the maximum values at higher simulation times. Nevertheless, it can be solved by considering more components of the PCA.

4. CONCLUSION

This study developed a novel approach for the streamflow simulation by integrating wavelet transform, Principal Component Analysis, and random forest during the time intervals of the next one, two, and three-day-ahead. The input data of the models were comprised of hydrometric and metrological data. In each group of data, the influences of the preprocessing and the levels of principal component analysis

Table 1. Simulation of Polroud daily streamflow using different hybrid and standalone approaches

LT (day)	Model	Testing Data	
		R	*RMSE* (m³/s)
1	RF	0.83	8.30
	PCA- RF	0.67	12.49
	PCA- haar- RF	0.83	9.46
	PCA- dmey- RF	0.91	7.10
2	RF	0.73	10.35
	PCA- RF	0.67	11.44
	PCA- haar- RF	0.81	8.99
	PCA- dmey- RF	0.86	7.84
3	RF	0.71	10.38
	PCA- RF	0.62	11.82
	PCA- haar- RF	0.79	9.30
	PCA- dmey- RF	0.82	8.68

were investigated on the model. The findings of this study demonstrated that the proposed hybrid model has a desirable performance in simulating the river flow (especially at higher forecast times) and provides simpler models for the analysis process.

REFERENCES

[1] Solomatine, D. P., and Ostfeld, A., 2008. "Data-driven modeling: some past experiences and new approaches". Journal of hydroinformatics, 10(1), Jan, pp. 3-22.

[2] Khairuddin, N., Aris, A. Z., Elshafie, A., Sheikhy Narany, T., Ishak, M. Y., and Isa, N. M., 2019. "Efficient forecasting model technique for river streamflow in tropical environment". Urban Water Journal, 16(3), Mar, pp. 183-192.

[3] Tongal, H., and Booij, M. J., 2018. "Simulation and forecasting of streamflows using machine learning models coupled with base flow separation". Journal of Hydrology, 564, Sep, pp. 266-282.

[4] Hussain, D., and Khan, A. A., 2020. "Machine learning techniques for Monthly River flow forecasting of Hunza River, Pakistan". Earth Science Informatics, Feb, pp. 1-11.

[5] Nourani, V., Baghanam, A. H., Adamowski, J., and Kisi, O., 2014. "Applications of hybrid wavelet–artificial intelligence models in hydrology: a review". Journal of Hydrology, 514, Jun, pp. 358-377.

[6] Nourani, V., Davanlou Tajbakhsh, A., Molajou, A., and Gokcekus, H., 2019. "Hybrid wavelet-M5 model tree for rainfall-runoff modeling". Journal of Hydrologic Engineering, 24(5), May, p. 04019012.

[7] Sun, Y., Niu, J., and Sivakumar, B., 2019. "A comparative study of models for short-term streamflow forecasting with emphasis on wavelet-based approach". Stochastic Environmental Research and Risk Assessment, 33(10), Oct, pp. 1875-1891.

[8] Roushangara, K., and Ghasempourb, R., 2019. "Monthly precipitation prediction improving using the integrated model based on kernel-wavelet and complementary ensemble empirical mode decomposition". CEEJ (XML). In Persian.

[9] Hassanzadeh, Y., Abdi, K. A., Shafiei, N. M., and Khoshtinat, S., 2015. "Daily streamflow forecasting of Nooranchay river using the hybrid model of Artificial Neural Networks-Principal Component Analysis". Journal of Soil and Water Science, 25 (3), pp. 53- 63. In Persian.

[10] Ehteram, M., Afan, H. A., Dianatikhah, M., Ahmed, A. N., Ming Fai, C., Hossain, M. S., and Elshafie, A., 2019. "Assessing the predictability of an improved ANFIS model for monthly streamflow using lagged climate indices as predictors". Water, 11(6), Jun, p. 1130.

[11] Diop, L., Bodian, A., Djaman, K., Yaseen, Z. M., Deo, R. C., El-Shafie, A., and Brown, L. C., 2018. "The influence of climatic inputs on stream-flow pattern forecasting: case study of Upper Senegal River". Environmental earth sciences, 77(5), Mar, p. 182.

[12] Hadi, S. J., and Tombul, M., 2018. "Monthly streamflow forecasting using continuous wavelet and multi-gene genetic programming combination". Journal of Hydrology, 561, Jun, pp. 674-687.

HOW TO CITE THIS ARTICLE

F. Azarpira, S. Shahabi, Capability Evaluation of Hybrid Wavelet-Principal Component Analysis-Random Forest Approach in Simulating the River Flow, Amirkabir J. Civil Eng., 53(7) (2021) 621-624.

DOI: 10.22060/ceej.2020.17520.6589