



## 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<sup>3</sup>/s, while these values were  $R=0.817$  and  $RMSE=8.681$  m<sup>3</sup>/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



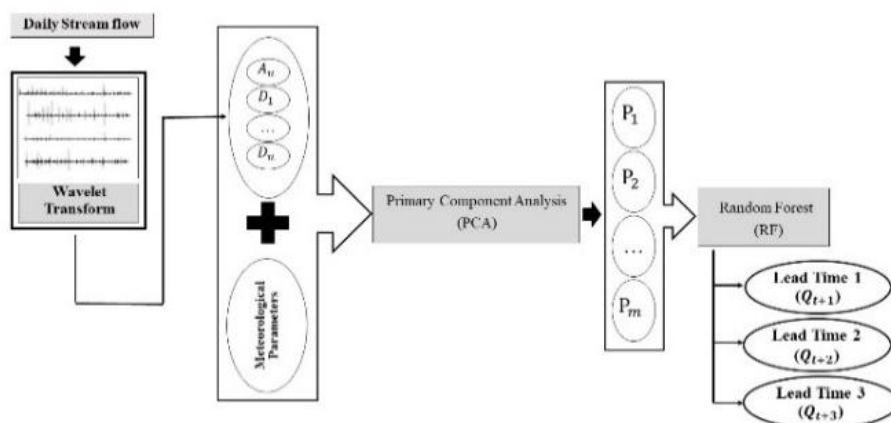


Fig. 1. Schematic diagram of proposed hybrid model

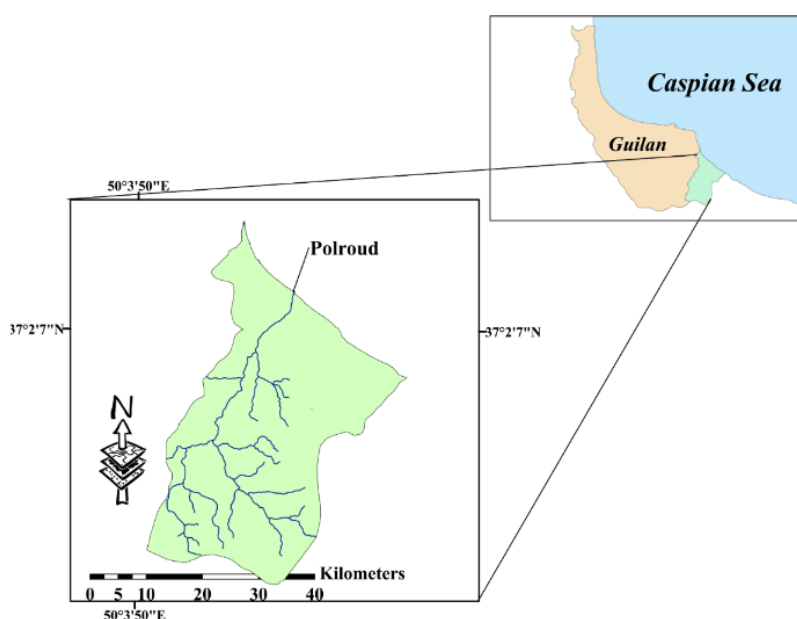


Fig. 2. The geographical location of Polroud river and Roudsar city

### 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 (m <sup>3</sup> /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.

<p><b>HOW TO CITE THIS ARTICLE</b></p> <p><i>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.</i></p> <p><b>DOI: 10.22060/ceej.2020.17520.6589</b></p>	
---	---





## ارزیابی قابلیت رویکرد ترکیبی موجک-تحلیل مؤلفه‌ی اصلی-جنگل تصادفی در شبیه‌سازی جریان رودخانه

فریبا آذرپیرا، سجاد شهابی\*

دانشکده‌ی مهندسی عمران و نقشه‌برداری، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران.

### تاریخچه داوری:

دریافت: ۱۳۹۸/۰۹/۲۰

بازنگری: ۱۳۹۹/۰۶/۰۷

پذیرش: ۱۳۹۹/۰۶/۰۸

ارائه آنلاین: ۱۳۹۹/۰۶/۲۰

### کلمات کلیدی:

رودخانه پلرود

پیش پردازش

جریان روزانه

زمان شبیه‌سازی

سری زمانی

**خلاصه:** شبیه‌سازی جریان جهت مدیریت تخصیص آب در دوره‌های ترسالی و خشکسالی بسیار حائز اهمیت است. با توجه به پژوهش‌هایی که در طول چندین دهه در این خصوص صورت گرفته، روش‌های هوش محاسباتی در ترکیب با موجک، به عنوان روش‌هایی کارآمد شناخته شده‌اند. در این مقاله، رویکرد ترکیبی موجک (W)-تحلیل مؤلفه اصلی (PCA)-جنگل تصادفی (RF)، جهت شبیه‌سازی جریان روزانه‌ی رودخانه‌ی پلرود، پیشنهاد شده است. در مدل پیشنهادی، ابتدا داده‌های هیدرومتری توسط تبدیل موجک پردازش شده و به همراه داده‌های هواشناسی به الگوریتم PCA اعمال شدند. سپس بردارهای خروجی آن به شبکه‌ی جنگل تصادفی وارد گردید. نتایج نشان داد که الگوریتم PCA علی‌رغم کاهش بعد بردارهای ورودی و ساده‌سازی آن، می‌تواند دقت و سرعت عملکرد مدل را ارتقا بخشد. همچنین، مدل را به گونه‌ای منسجم کند که افزایش زمان شبیه‌سازی و عدم قطعیت بردارهای ورودی، بر قابلیت مدل کمتر اثر بگذارد و روند نزولی آن، از شیب هموارتری برخوردار باشد. علاوه بر این، پیش‌پردازش داده به همراه PCA، توانسته در زمان شبیه‌سازی یک و سه روز، شاخص توافق را، به ترتیب، ۵ و ۸ درصد بهبود ببخشد و قابلیت مدل در شبیه‌سازی درست‌تر جریان رودخانه را افزایش دهد. از طرفی، نتایج بهترین مدل ترکیبی در زمان شبیه‌سازی یک روز دارای ضریب همبستگی ۰/۹۱۱ و جذر متوسط مربعات خطای ۷/۰۹۵ مترمکعب در ثانیه است که این مقادیر در بهترین مدل ترکیبی برای زمان شبیه‌سازی سه روز به ۰/۸۱۷ و ۸/۶۸۱ مترمکعب در ثانیه می‌رسد، که این موضوع به خوبی قابلیت بسیار مناسب مدل ترکیبی پیشنهادی برای زمان‌های شبیه‌سازی بلندمدت‌تر را نشان می‌دهد.

### ۱- مقدمه

بر قابلیت آن‌ها بیفزاید. به ویژه استفاده از تبدیل موجک<sup>۱</sup> (WT) که نه تنها خطاهای موجود در سری را کاهش می‌دهد، بلکه بی‌نظمی سری‌های زمانی را نیز اصلاح می‌کند [۵-۸]. همچنین، جهت انتخاب پارامترهای مناسب می‌توان از روابط خطی خودهمبستگی<sup>۲</sup> (ACF) و خودهمبستگی جزئی<sup>۳</sup> (PACF) [۹-۱۰]، آزمون گاما<sup>۴</sup> [۱۱-۱۲]، تحلیل مؤلفه اصلی<sup>۵</sup> (PCA) [۱۳-۱۵] و برخی روش‌های دیگر استفاده کرد [۱۶]. به‌طور مثال نوری و همکاران جریان ماهانه‌ی سد علویان رودخانه‌ی صوفی‌چای را با دو روش رگرسیون بردار پشتیبان<sup>۶</sup>

شبیه‌سازی جریان از جمله مباحث مهم آب‌های سطحی و مبنای بسیاری از برنامه‌های هیدرولیکی و هیدرولوژیکی است. این فرآیند متأثر از عوامل مختلفی است که مشارکت همه‌ی آن‌ها در مدل‌سازی مقدور نیست. به‌همین دلیل، روش‌های داده محور، به ویژه روش‌های مبتنی بر هوش محاسباتی که از دقت و سهولت اجرای بیشتری برخوردار هستند، رواج یافته‌اند [۱-۴]. این روش‌ها می‌توانند با پیچیدگی‌های (ناایستایی، غیرنرمال بودن) موجود در سری‌های هیدرولوژیکی مقابله کنند. اما، به‌دلیل تأثیرپذیری خروجی مدل‌ها از مقادیر ورودی، آماده‌سازی و پیش‌پردازش داده‌ها می‌تواند

1 Wavelet Transform (WT)

2 Auto- Correlation Function (ACF)

3 Partial- Auto Correlation Function (PACF)

4 Gamma Test (GT)

5 Principle Component Analysis (PCA)

6 Support Vector Regression (SVR)

\* نویسنده عهده‌دار مکاتبات: s.shahabi@kgut.ac.ir



و پایین‌دست، تبخیر و تعرق، دما و بارش استفاده کردند. آن‌ها نشان دادند که هر یک از پارامترها در افزایش قابلیت مدل‌های ANN، برنامه‌ریزی ژنتیک چندژنی<sup>۶</sup> (MGGP) و تبدیل موجک پیوسته<sup>۷</sup> - برنامه‌ریزی ژنتیک چندژنی - شبکه‌عصبی (WMGGPNN) مؤثرند، فقط ممکن است زمان اثرگذاری آن‌ها متفاوت باشد (الزام به ایجاد تأخیر) [۲۱]. ادنان و همکاران جهت ارزیابی جریان ماهانه‌ی دو ایستگاه مختلف از رودخانه سوات<sup>۸</sup> پاکستان، از داده‌های بارش، دما و جریان استفاده کردند. آن‌ها نشان دادند نه تنها، داده‌های هواشناسی دقت مدل‌ها را افزایش می‌دهند، بلکه برای یک ایستگاه، استفاده از دما به تنهایی نیز، نتایج مطلوبی را ارائه می‌دهد [۲۲]. در پژوهشی دیگر، پارامترهای جریان، دما، تبخیر/تعرق، بارش و رطوبت نسبی رودخانه‌ی امامه تهران با استفاده از روش همبستگی متقابل مورد ارزیابی قرار گرفت، که نتیجه‌ی آن شصت و دو سناریو مختلف شد. جهت مدل‌سازی این سناریوها از دو روش برنامه‌ریزی بیان ژن<sup>۹</sup> و درخت تصمیم M۵ استفاده شد. نتایج نشان داد که هر دو مدل، برای سناریو متشکل از همه‌ی پارامترهای ذکر شده، توانستند بهترین عملکرد را داشته باشند [۲۳].

هرچند تعدد مطالعات انجام شده در شبیه‌سازی جریان قابل ملاحظه است، لیکن در این مقاله، از رویکرد ترکیبی موجک-تحلیل مؤلفه‌ی اصلی-جنگل تصادفی به منظور شبیه‌سازی جریان روزانه‌ی رودخانه‌ی پلرود واقع در شرق استان گیلان استفاده شده که دربرگیرنده‌ی چندین رویکرد نوین در مقایسه با مطالعات مشابه است، که هرچند هر یک از آن‌ها به تنهایی واجد شرایط نوینی می‌باشد، ولی ترکیب آن‌ها و با ساختار مورد اشاره، برای نخستین بار بکار گرفته می‌شود. بر این اساس، نخست برخلاف غالب مطالعات صورت پذیرفته، سری‌های زمانی مورد استفاده با تبدیل موجک‌ها تجزیه می‌شوند، سپس با در نظرگرفتن پارامترهای هواشناسی و هیدرومتری، از میان تمامی سری‌های حاصل (که خود با دو موجک مادر متفاوت و در سطوح تجزیه‌ی مختلف حاصل شده‌اند) بهترین مؤثرترین ترکیبات خطی با استفاده از تحلیل مؤلفه‌ی اصلی انتخاب می‌شوند. در ادامه ترکیبات مؤلفه‌ی اصلی منتخب به الگوریتم جنگل تصادفی وارد می‌شود.

(SVR) و شبکه‌عصبی<sup>۱</sup> (ANN) تخمین زدند. برای انتخاب بهینه‌ی ورودی‌ها، از میان پارامترهای جریان، بارش، دما (حداقل، حداکثر و میانگین) و تابش خورشید، در سه زمان ( $t$ ,  $t-1$ ,  $t-2$ )، از سه روش PCA، آزمون گاما و روش‌های انتخاب رو به جلو<sup>۲</sup> (FS) استفاده کردند. هر یک از این سه روش، هجده پارامتر ورودی را، به ترتیب به ۵، ۵ و ۷ بردار کاهش دادند. در نهایت نیز نتایج مدل PCA-SVR به سایر مدل‌ها برتری یافتند [۱۷]. در سال ۲۰۱۵ نیز برای شبیه‌سازی جریان روزانه‌ی رودخانه نوران‌چای، داده‌های جریان، بارش، ساعات آفتابی و دمای هوا (حداقل و حداکثر)، جهت بررسی، انتخاب شدند. به طوری که از روش PCA برای انتخاب مؤثرترین آن‌ها استفاده شد. نتایج نشان داد که پیش‌پردازش داده توسط PCA سبب افزایش قابلیت مدل ANN می‌شود [۱۸]. در مطالعه‌ی دیگر احترام و همکاران ضمن بررسی اثر شاخص‌های هواشناسی بر جریان ماهانه‌ی حوضه‌ی در شمال غربی ایران، با استفاده از سیستم استنتاج فازی-عصبی تطبیقی<sup>۳</sup> (ANFIS)، از PCA به منظور کاهش بردارهای ورودی استفاده کردند، که نتایج افزایش قابلیت مدل ANFIS به همراه PCA را نشان می‌دهد [۱۹].

از جمله عوامل مؤثر بر جریان، فاکتورهای هواشناسی هستند که با توجه به چرخه‌ی هیدرولوژی بهم مرتبط می‌شوند. اما سؤال اساسی این است که آیا پارامترهای هواشناسی در یافتن الگوی بهینه‌ی جریان آتی مؤثر هستند؟ جهت پاسخ‌گویی به این سوال، پژوهش‌هایی صورت گرفته است که در ادامه به اختصار به برخی از مهم‌ترین آن‌ها اشاره می‌شود.

دایاپ و همکاران به منظور شبیه‌سازی جریان روزانه‌ی رودخانه سنگال<sup>۴</sup>، با توجه به شرایط اقلیمی استوایی، پارامترهای تبخیر/تعرق و بارش را به همراه جریان روزانه به کار بردند. آن‌ها با تشکیل دو ترکیب از داده‌های هواشناسی و هیدرومتری و داده‌های هیدرومتری به این نتیجه رسیدند که داده‌های هواشناسی، هر چند کم، اما می‌توانند در بهبود عملکرد SVR و شبکه‌ی عصبی رگرسیون تعمیم یافته<sup>۵</sup> (GRNN) نقش داشته باشد [۲۰]. هادی و تامبول جهت شبیه‌سازی جریان ماهانه‌ی رودخانه‌ی در ترکیه، از پارامترهای جریان بالادست

1 Artificial Neural Network (ANN)

2 Forward Selection Techniques (FS)

3 Adaptive Neuro- Fuzzy Inference System (ANFIS)

4 Senegal

5 Generalized Regression Neural Network (GRNN)

6 Multi- Gene- Genetic Programming (MGGP)

7 Continuous Wavelet Transform (CWT)

8 Swat

9 Gene Expression Programming (GEP)



## ۲- مواد و روش‌ها

### ۲-۲- جنگل تصادفی (RF)

جنگل تصادفی مجموعه‌ای از درخت‌های مدل CART<sup>۱</sup> است که جهت بهبود عملکرد آن ارائه شده است. این درخت‌ها با استفاده از روش نمونه‌گیری Bootstrap، از مجموعه‌ی بردارهای ورودی، ساخته می‌شوند. به طوری که، ورودی انتخابی همه‌ی آن‌ها از یک توزیع تبعیت می‌کند. مجموعه‌ی درخت‌ها علی‌رغم ساختار منحصربه‌فرد و مستقل‌شان، از همبستگی زیادی نسبت به هم برخوردار هستند که به همراه مقاومت و تعداد زیاد آن‌ها، سبب همگرایی خطای مدل می‌شود. هم‌چنین، ویژگی‌های ذکر شده به همراه تخمین خطای داخلی مدل، سبب افزایش قابلیت آن در بررسی تعداد بیشتر ویژگی‌ها و انتخاب مؤثرترین پارامترها می‌شود. در این شبکه‌ی درختی، فرآیند هرس انجام نمی‌شود، بنابراین برای پیشگیری از بیش‌پدازش مدل، میزان خطای مدل محدود می‌شود [۲۶].

### ۲-۳- تحلیل مؤلفه‌ی اصلی (PCA)

تحلیل مؤلفه‌ی اصلی از جمله روش‌های آماری چندپارامتری است که با برقراری یک ارتباط خطی بین ویژگی‌های (متغیرهای) ورودی اولیه‌ی مدل، از حجم اطلاعات ورودی کاسته و مؤثرترین بردارهای با ضریب همبستگی صفر را انتخاب می‌کند. خصوصیت اخیر موجب می‌شود تا مؤلفه‌های منتخب جنبه‌های مختلف از مدل را بررسی کنند. یک مؤلفه‌ی اصلی را می‌توان به صورت زیر نوشت:

$$\begin{aligned} PC_1 &= a_1^1 A_1 + a_1^2 D_1 + \dots + a_1^n X_n \\ PC_2 &= a_2^1 A_1 + a_2^2 D_1 + \dots + a_2^n X_n \\ &\vdots \\ PC_n &= a_n^1 A_1 + a_n^2 D_1 + \dots + a_n^n X_n \end{aligned} \quad (1)$$

که در آن  $PC_1$  تا  $PC_n$  تعداد مؤلفه‌های اصلی است که در این مقاله از ۱۳ تا ۵۲ مورد متغیر است. همچنین  $a_i^j$  ضریب اُامین مؤلفه‌ی اصلی و اُامین متغیر می‌باشد، که در این تحقیق ورودی تحلیل مؤلفه‌ی اصلی خود براساس فرآیند پیش پردازش موجک بوده است. بدین ترتیب متغیرهای تجزیه شده توسط موجک به صورت تقریب و جزئیات ( $A_i$  و  $D_i$ ) بوده و سایر پارامترها با  $X_i$  نمایش داده

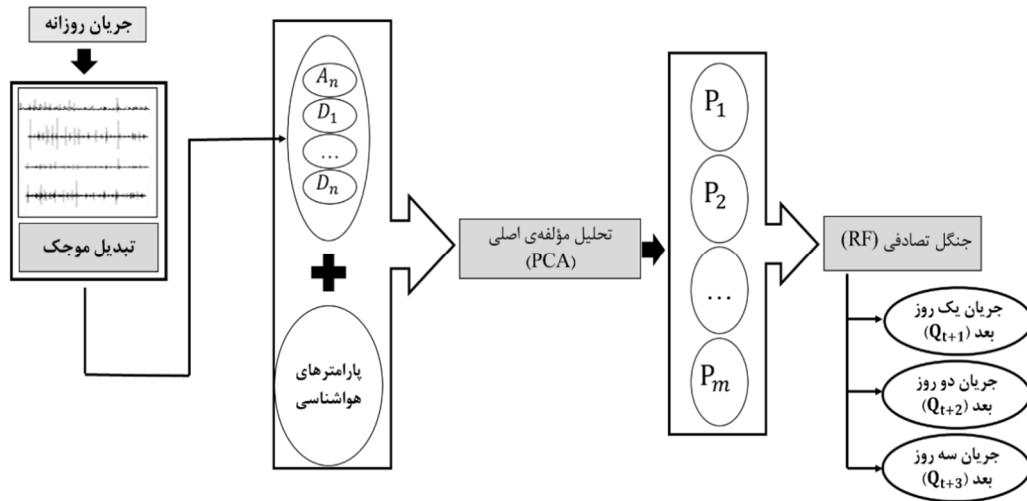
در این مقاله به منظور شبیه‌سازی جریان از ترکیب چندین رویکرد مختلف استفاده شده است. بر این اساس، در ادامه نخست ضمن معرفی مختصری از تبدیل موجک‌ها، تحلیل مؤلفه‌ی اصلی و رویکرد جنگل تصادفی، منطقه‌ی مورد مطالعه معرفی شده و در نهایت شاخص‌های مورد استفاده جهت ارزیابی مدل‌ها ارائه خواهند شد.

### ۲-۱- تبدیل موجک گسسته (W)

تبدیل موجک فرآیندی است که سبب شناسایی الگوهای پنهان در سری‌زمانی، از طریق تجزیه‌ی سری می‌شود. در واقع، سری را به مجموعه‌ی فرکانسی بالا (زیرسری جزئیات) و پایین (زیرسری تقریب) تقسیم (فیلتر) می‌کند. بررسی و تحلیل این دو زیرسری، به ترتیب توسط توابع واحدی به نام‌های تابع مقیاس و تابع موجک صورت می‌گیرد که به مجموعه‌ی آن‌ها موجک مادر گفته می‌شود. موجک‌های مادر توابع شناخته‌شده‌ای هستند که در دو دسته‌ی Orthogonal و Biorthogonal طبقه‌بندی می‌شوند. توابع Orthogonal موجک‌هایی هستند که حالت خطی ندارند و توابع مقیاس و موجک آن‌ها در مجموعه‌ی فضا دوبه‌دو متعامد هستند. در حالی که توابع Biorthogonal حالت خطی دارند و دو زیرتابع آن‌ها فقط در مقیاس خود باهم متعامدند. علاوه بر موجک مادر، پارامتر سطح تجزیه و یا انتقال در تجزیه سری نقش دارد. به طوری که، تعداد مراحل تجزیه و مقیاس بررسی زیرسری‌ها را مشخص می‌کند. مقیاس تابعی نمایی عکس انتقال است که نوع مبنای آن منجر به دسته‌بندی تبدیل موجک، به دو دسته‌ی تبدیل موجک پیوسته و گسسته می‌شود. در نوع پیوسته مبنای به صورت  $2^j$  ( $j=1, 2, 3, \dots$ ) و در گسسته مبنای  $2^j$  می‌باشد.

در تمامی مراحل تجزیه، زیرسری جزئیات توسط تابع مقیاس، در میانگین فرکانسی صفر، نمونه‌برداری و خصوصیات بارز آن در قالب ضرایب موجک ارائه می‌شود. زیرسری تقریب (کلیات) نیز توسط تابع موجک برای ادامه‌ی روند فرآیند فیلترکردن آماده می‌شود. در نهایت نیز به تعداد پارامتر انتقال زیرسری جزئیات و یک زیرسری تقریب ایجاد می‌شود که قابلیت بازسازی به سری اولیه را دارد [۲۴-۲۵].

1 Classification and Regression Tree



شکل ۱. فلوچارت رویکرد ترکیبی پیشنهادی  
 Fig. 1. Flowchart of proposed hybrid approach

می‌دهد، آن‌ها را در ابعاد کمتری به شبکه‌ی داده‌کاو اعمال کند. در مدل پیشنهادی، دبی جریان روزانه‌ی رودخانه‌ی پلرود، ابتدا توسط تبدیل موجک پیش‌پردازش می‌شود و سپس، در پنج گام زمانی  $t$ ،  $t-1$ ،  $t-2$ ،  $t-3$  و  $t-4$  به همراه پارامترهای هواشناسی (بارش، دمای هوای حداقل، حداکثر و میانگین، نقطه‌ی شبنم، فشار هوای ایستگاه و فشار هوا) به الگوریتم PCA اعمال می‌شوند. به کمک این الگوریتم، مؤثرترین ترکیبات خطی از پارامترهای ورودی شناسایی شده و به عنوان بردارهای ورودی مورد استفاده قرار می‌گیرند. در نهایت نیز، بردارهای استخراجی از PCA به جنگل تصادفی اعمال می‌شوند. جهت سنجش قابلیت مدل نیز، روش PCA-RF پیشنهاد شده است که در آن ابتدا دبی جریان در پنج گام زمانی مذکور و پارامترهای هواشناسی به PCA اعمال و نتایج آن به RF وارد می‌شوند. شکل ۱ فلوچارت رویکرد ترکیبی پیشنهادی را نشان می‌دهد.

#### ۵-۲- منطقه‌ی مطالعاتی

حوضه‌ی مطالعاتی در این پژوهش، رودخانه‌ی پلرود، واقع در شرق استان گیلان است که از کوه‌های چاکرود سرچشمه گرفته و پس از گذر از دشت و جلگه‌ی رودسر، در نهایت به دریای کاسپین می‌ریزد. حوضه‌ی این رودخانه دارای مساحت ۱۷۴۰ کیلومتر مربع بوده و طول آبراهه‌ی اصلی حوضه (رودخانه) ۶۷ کیلومتر، میانگین شیب حوضه ۵ درصد، حداقل ارتفاع (از سطح دریا) ۵۲ متر و حداکثر آن ۳۹۳۱ متر می‌باشد. موقعیت قرارگیری رودخانه‌ی پلرود و حوضه‌ی آن در

شده‌اند.

در معادله‌ی (۱) ضرایب  $a_i^j$  از حل معادله‌ی زیر به دست می‌آیند:

$$|R - \lambda I| = 0 \quad (2)$$

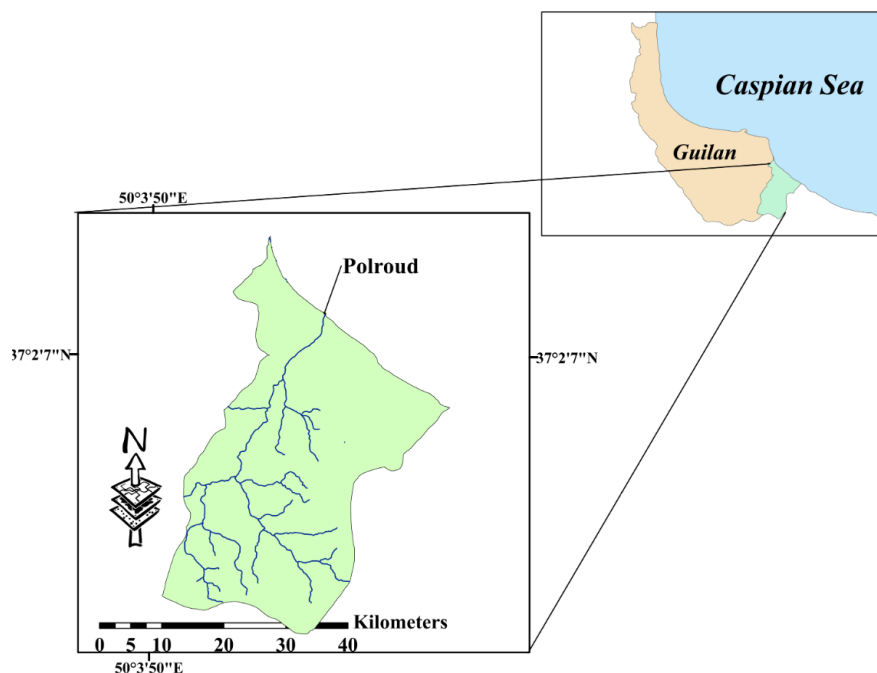
که در آن  $I$  ماتریس واحد و  $R$  ماتریس همبستگی بین متغیرهای اولیه و  $\lambda$  مقادیر ویژه می‌باشد. همچنین تناسب جامعه‌ی آماری نیز توسط آزمون KMO<sup>۱</sup> سنجیده شده که مقادیر بهینه‌ی آن در محدوده‌ی ۰/۵ تا ۱ واقع بوده و هر چه به یک نزدیکتر باشد، تناسب بهتری را نشان می‌دهد [۲۷].

#### ۴-۲- مدل ترکیبی موجک-تحلیل مؤلفه‌ی اصلی-جنگل تصادفی (WPCARF)

عوامل پنهان و آشکار متعددی در جریان رودخانه نقش دارند که تحلیل و بررسی همه‌ی آن‌ها، نیازمند صرف زمان و انرژی زیادی است. از طرفی با توجه به محدودیت اندازه‌گیری پارامترها، در هیدرولوژی، استفاده از داده‌ها با تقریب و خطا همراه است. بر همین اساس، در این پژوهش، روشی پیشنهاد شده است که بتواند از خطاها و عدم قطعیت‌های موجود در سری جریان بکاهد و همچنین، علی‌رغم اینکه عوامل مختلف هواشناسی را در الگوییابی جریان آینده شرکت

1 Kaser Meyer Olkin





شکل ۲. موقعیت جغرافیایی رودخانه‌ی پلرود و شهرستان رودسر  
**Fig. 2. Geographical location of Polroud River and Roudsar city**

جدول ۱. پارامترهای آماری داده‌های استفاده شده در مقیاس روزانه

**Table 1. Statistical parameters of used data (Daily)**

پارامترها	حداقل	حداکثر	متوسط	انحراف معیار	ضریب چولگی
دبی ( $m^3/s$ )	۰/۰۵۵	۳۶۲	۱۴/۳۳۱	۱۵/۳۱۰	۵/۵۶۰
بارش (mm)	۰	۳۴۰/۲۰۰	۳/۳۴۵	۱۲/۹۷۰	۹/۱۹۶
دمای حداقل هوا ( $^{\circ}C$ )	-۵/۶۰۰	۲۸/۲۰۰	۱۳/۲۳۳	۷/۱۷۸	-۰/۰۴۸
دمای حداکثر هوا ( $^{\circ}C$ )	۰/۴	۳۸	۱۹/۶۰۹	۷/۴۰۶	-۰/۰۳۶
دمای متوسط هوا ( $^{\circ}C$ )	-۱/۵	۳۰/۵	۱۶/۴۲۰	۷/۱۳۹	-۰/۰۰۱
نقطه‌ی شبنم ( $^{\circ}C$ )	-۱۷/۰۲۴	۲۶/۱۸۲	۱۲/۹۹۰	۶/۶۸۱	-۰/۱۰۹
فشار هوا (mbar)	۹۹۳/۶۰۳	۱۰۴۳/۹۵۰	۱۰۱۵/۲۰۶	۶/۹۱۱	۰/۳۶۸
فشار هوای منطقه (mbar)	۹۹۵/۸۵۰	۱۰۴۳/۳۰۰	۱۰۱۷/۵۸۹	۶/۹۶۶	۰/۳۶۵

### ۲-۶- شاخص‌های ارزیابی مدل

با توجه به عدم قطعیت مدل‌های هیدرولوژیکی از آمار و احتمال به‌عنوان معیار ارزیابی عملکرد این مدل‌ها استفاده می‌شود. بر همین اساس، در این مطالعه، از چهار شاخص مختلف استفاده شده که عبارتند از:

۱. ضریب همبستگی<sup>۱</sup> ( $R$ )، بیانگر ارتباط خطی بین مقادیر

شهرستان رودسر و استان گیلان در شکل ۲ نشان داده شده است. جهت استخراج دبی‌های روزانه‌ی پلرود، از داده‌های سال آبی ۵۷-۱۳۵۶ تا ۹۶-۱۳۹۵ ایستگاه هیدرومتری طولات واقع بر روی این رودخانه، استفاده شده است. علاوه بر این، جهت گردآوری پارامترهای هواشناسی، از اطلاعات ایستگاه سینوپتیک رامسر استفاده شده است. مشخصات آماری داده‌های مورد استفاده در جدول ۱ آورده شده است.

1 Correlation Coefficient

جدول ۲. معیارهای ارزیابی مدل  
Table 2. Model Assessment Criteria

شاخص	معادله مربوط به شاخص	تعریف متغیرها	محدوده نتایج	تفسیر
$R$	$R = \frac{\sum_1^n (D_i^o - D_m^o)(D_i^f - D_m^f)}{\sqrt{\sum_1^n (D_i^o - D_m^o)^2 \sum_1^n (D_i^f - D_m^f)^2}}$	مقادیر مشاهداتی: $D_i^o$	$[-1, +1]$	$(0, -1]$ : ارتباط غیرمستقیم ۰: بدون ارتباط $[0, 1)$ : ارتباط مستقیم بیشترین انسجام در مقادیر حدی ۱ و -۱
$RMSE$	$RMSE = \sqrt{\frac{\sum_1^n (D_i^o - D_i^f)^2}{n}}$	مقادیر مدل: $D_i^f$	$[0, +\infty]$	مقادیر کوچک‌تر و نزدیک به صفر نشان‌دهنده ی خطای کمتر مدل است. این شاخص هم‌بعد پارمتر مورد بررسی است $(0, -\infty]$ : ناکارآمدی مدل
$NSE$	$NSE = 1 - \frac{\sum_1^n (D_i^f - D_i^o)^2}{\sum_1^n (D_i^o - D_m^o)^2}$	مشاهداتی: $D_m^o$	$(-\infty, 1]$	$[0, 1)$ : مقادیر بیشتر این بازه کارایی بیشتر مدل را نشان می‌دهد.
$I_a$	$I_a = \frac{1}{\sum_1^n ( D_i^p - D_m^o  +  D_i^o - D_m^o )^2}$	مقادیر مدل: $D_m^f$	$[0, 1]$	مقدار یک نماینده دقیق‌ترین مدل است.

تجزیه شده‌اند. هار موجکی گسسته و اولین موجک شناخته شده است. دیمی موجکی با پهنای باند وسیع و حساس به مقادیر حدی است.

جهت انتخاب مناسب‌ترین بردارهای ورودی، برای هر دو مدل منفرد و ترکیبی (با ده نوع ویژگی)، داده‌های هواشناسی ( $t$ ) و هیدرومتری ( $t-4$  تا  $t$ )، به الگوریتم PCA و در نهایت، نتایج آن به RF اعمال شدند. همچنین، سناریو مطرح شده، بدون استفاده از PCA و پیش‌پردازش نیز منظور شده است. تا تأثیر این دو عامل نیز بر روند مدل سازی مورد بررسی قرار بگیرد. علاوه بر این، تمامی سناریوها برای سه زمان شبیه‌سازی یک، دو و سه روز انجام شده‌اند تا میزان قابلیت مدل پیشنهادی برای زمان‌های مختلف نیز ارزیابی گردد.

به منظور اختصار صرفاً نتایج یک مورد از تحلیل مؤلفه‌ی اصلی ارائه می‌شود. جهت کسب اطلاعات بیشتر در مورد این تحلیل می‌توان به کتب مرجع مراجعه نمود [۳۰]. به عنوان مثال برای تابع موجک مادر هار و در سطح تجزیه‌ی ۴، جدول ۳ واریانس متغیرهای ورودی را نشان می‌دهد.

بر اساس جدول ۳، ۳۲ متغیر ورودی به ۱۱ متغیر کاهش داده شده‌اند. مقدار KMO برای این مجموعه متغیرهای ورودی ۰/۶۴ می‌باشد، که تناسب رضایت بخش جامعه‌ی آماری را نشان می‌دهد.

محاسباتی و مشاهداتی است [۲۸].

۲. جذر میانگین مربعات خطا<sup>۱</sup> ( $RMSE$ )، که نشان‌دهنده‌ی انحراف معیار مقادیر محاسباتی نسبت به مقادیر مشاهداتی است.

۳. ضریب کارایی ناش ساتکلیف<sup>۲</sup> ( $NSE$ )، نشان‌دهنده درجه‌ی آزادی مدل نسبت به مقادیر مشاهداتی است.

۴. شاخص توافق<sup>۳</sup> ( $I_a$ )، که میزان درستی و دقت مدل را نشان می‌دهد [۲۹].

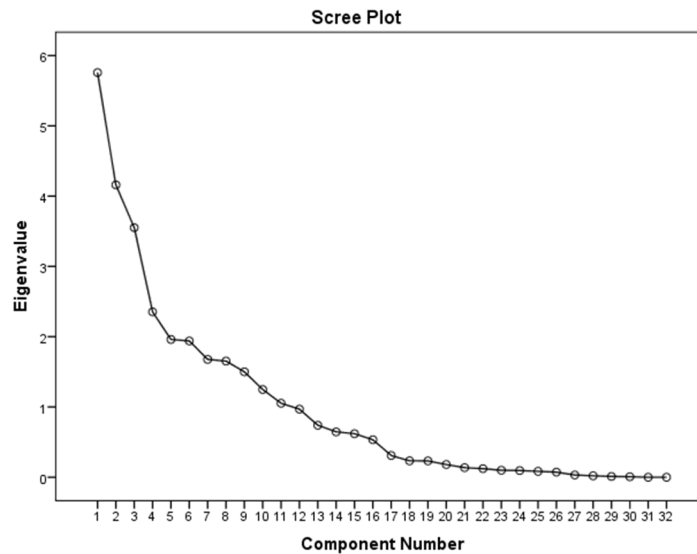
### ۳- نتایج و بحث

شبیه‌سازی جریان دارای عدم قطعیت‌های فراوانی است که ارائه راهکاری جهت کاهش آن بسیار مورد توجه است. از این‌رو در این پژوهش، از مدل هوش مصنوعی در ترکیب با موجک، داده‌های هواشناسی و الگوریتم تحلیل مؤلفه‌ی اصلی جهت انتخاب ورودی‌ها استفاده شده است. جهت مدل سازی، داده‌های هیدرومتری در پنج زمان ( $t-4$ ،  $t-3$ ،  $t-2$ ،  $t-1$ ،  $t$ ) و داده‌های هواشناسی در زمان  $t$  مورد استفاده قرار گرفتند که در مدل ترکیبی WRF داده‌های هیدرومتری در پنج سطح تجزیه ( $3-7$ )، توسط دو موجک مادر هار<sup>۴</sup> و دیمی<sup>۵</sup>

1 Root Mean Square Error  
2 Nash-Sutcliffe Model Efficiency Coefficient  
3 Index of Agreement  
4 haar  
5 dmeq

جدول ۳. واریانس متغیرهای ورودی  
Table 3. Total variance explained

شماره مؤلفه	ارزش کلی	درصد پراکندگی	درصد پراکندگی تجمعی	ارزش کلی	درصد پراکندگی	درصد پراکندگی تجمعی
۱	۵/۷۵۸	۱۷/۹۹۳	۱۷/۹۹۳	۵/۷۵۸	۱۷/۹۹۳	۱۷/۹۹۳
۲	۴/۱۶۰	۱۳	۳۰/۹۹۳	۴/۱۶۰	۱۳	۳۰/۹۹۳
۳	۳/۵۵۲	۱۱/۱	۴۲/۰۹۲	۳/۵۵۲	۱۱/۱	۴۲/۰۹۲
۴	۲/۳۵۳	۷/۳۵۴	۴۹/۴۴۷	۲/۳۵۳	۷/۳۵۴	۴۹/۴۴۷
۵	۱/۹۶۰	۶/۱۲۶	۵۵/۵۷۳	۱/۹۶۰	۶/۱۲۶	۵۵/۵۷۳
۶	۱/۹۴۰	۶/۰۶۱	۶۱/۶۳۴	۱/۹۴۰	۶/۰۶۱	۶۱/۶۳۴
۷	۱/۶۷۷	۵/۲۴۰	۶۶/۸۷۵	۱/۶۷۷	۵/۲۴۰	۶۶/۸۷۵
۸	۱/۶۵۳	۵/۱۶۶	۷۲/۰۴۱	۱/۶۵۳	۵/۱۶۶	۷۲/۰۴۱
۹	۱/۵۰۱	۴/۶۸۹	۷۶/۷۳۰	۱/۵۰۱	۴/۶۸۹	۷۶/۷۳۰
۱۰	۱/۲۴۹	۳/۹۰۲	۸۰/۶۳۲	۱/۲۴۹	۳/۹۰۲	۸۰/۶۳۲
۱۱	۱/۰۵۳	۳/۲۹۰	۸۳/۹۲۲	۱/۰۵۳	۳/۲۹۰	۸۳/۹۲۲
۱۲	۰/۹۶۹	۳/۰۲۸	۸۶/۹۵۰	۰/۹۶۹	۳/۰۲۸	۸۶/۹۵۰
۱۳	۰/۷۴۰	۲/۳۱۲	۸۹/۲۶۲	۰/۷۴۰	۲/۳۱۲	۸۹/۲۶۲
۱۴	۰/۶۴۶	۲/۰۱۸	۹۱/۲۸۰	۰/۶۴۶	۲/۰۱۸	۹۱/۲۸۰
۱۵	۰/۶۲۰	۱/۹۳۶	۹۳/۲۱۶	۰/۶۲۰	۱/۹۳۶	۹۳/۲۱۶
۱۶	۰/۵۳۴	۱/۶۶۸	۹۴/۸۸۴	۰/۵۳۴	۱/۶۶۸	۹۴/۸۸۴
۱۷	۰/۳۰۹	۰/۹۶۷	۹۵/۸۵۱	۰/۳۰۹	۰/۹۶۷	۹۵/۸۵۱
۱۸	۰/۲۳۴	۰/۷۳۰	۹۶/۵۸۱	۰/۲۳۴	۰/۷۳۰	۹۶/۵۸۱
۱۹	۰/۲۳۲	۰/۷۲۶	۹۷/۳۰۷	۰/۲۳۲	۰/۷۲۶	۹۷/۳۰۷
۲۰	۰/۱۸۱	۰/۵۶۵	۹۷/۸۷۲	۰/۱۸۱	۰/۵۶۵	۹۷/۸۷۲
۲۱	۰/۱۳۶	۰/۴۲۶	۹۸/۲۹۸	۰/۱۳۶	۰/۴۲۶	۹۸/۲۹۸
۲۲	۰/۱۲۲	۰/۳۸۱	۹۸/۶۸۰	۰/۱۲۲	۰/۳۸۱	۹۸/۶۸۰
۲۳	۰/۰۹۹	۰/۳۱۱	۹۸/۹۹۱	۰/۰۹۹	۰/۳۱۱	۹۸/۹۹۱
۲۴	۰/۰۹۵	۰/۲۹۷	۹۹/۲۸۷	۰/۰۹۵	۰/۲۹۷	۹۹/۲۸۷
۲۵	۰/۰۸۴	۰/۲۶۲	۹۹/۵۴۹	۰/۰۸۴	۰/۲۶۲	۹۹/۵۴۹
۲۶	۰/۰۷۳	۰/۲۲۸	۹۹/۷۷۷	۰/۰۷۳	۰/۲۲۸	۹۹/۷۷۷
۲۷	۰/۰۳۲	۰/۱	۹۹/۸۷۷	۰/۰۳۲	۰/۱	۹۹/۸۷۷
۲۸	۰/۰۲	۰/۰۶۴	۹۹/۹۴۱	۰/۰۲	۰/۰۶۴	۹۹/۹۴۱
۲۹	۰/۱۱	۰/۰۳۴	۹۹/۹۷۵	۰/۱۱	۰/۰۳۴	۹۹/۹۷۵
۳۰	۰/۰۰۸	۰/۰۲۴	۹۹/۹۹۹	۰/۰۰۸	۰/۰۲۴	۹۹/۹۹۹
۳۱	.	۰/۰۰۱	۱۰۰	.	۰/۰۰۱	۱۰۰
۳۲	.	.	۱۰۰	.	.	۱۰۰



شکل ۳. نمایش شماره مؤلفه در برابر مقادیر ویژه  
**Fig. 3. The plot of component number and eigenvalues**

جدول ۴. شبیه‌سازی جریان روزانه‌ی پلرود با استفاده از رویکردهای مختلف ترکیبی و منفرد

**Table 4. Simulation of Polrود daily stream flow using different hybrid and standalone approaches**

داده‌های آزمون				سطح تجزیه	مدل	زمان شبیه‌سازی (روز)
I <sub>a</sub>	NSE	RMSE (m <sup>3</sup> /s)	R			
۰/۸۹۹	۰/۶۸۶	۸/۲۹۸	۰/۸۲۹	-	RF	۱
۰/۷۷۹	۰/۴۳۸	۱۲/۴۹۲	۰/۶۶۵	-	PCA- RF	
۰/۸۸۸	۰/۶۷۸	۹/۴۵۶	۰/۸۲۶	۳	PCA- haar- RF	
۰/۹۴۲	۰/۸۱۹	۷/۰۹۵	۰/۹۱۱	۵	PCA- dmey- RF	
۰/۸۲۹	۰/۵۲۷	۱۰/۳۵۳	۰/۷۲۷	-	RF	۲
۰/۷۸۲	۰/۴۴۷	۱۱/۴۳۷	۰/۶۷۰	-	PCA- RF	
۰/۸۸۵	۰/۶۵۹	۸/۹۸۶	۰/۸۱۲	۵	PCA- haar- RF	
۰/۹۱۷	۰/۷۴۰	۷/۸۳۸	۰/۸۶۱	۳	PCA- dmey- RF	
۰/۸۱۶	۰/۵۰۲	۱۰/۳۸۳	۰/۷۱۰	-	RF	۳
۰/۷۴۸	۰/۳۸۰	۱۱/۸۲۴	۰/۶۲۲	-	PCA- RF	
۰/۸۶۷	۰/۶۱۶	۹/۳۰۵	۰/۷۸۵	۶	PCA- haar- RF	
۰/۸۸۵	۰/۶۶۶	۸/۶۸۱	۰/۸۱۷	۶	PCA- dmey- RF	

در تمامی مدل‌ها، افزایش زمان شبیه‌سازی سبب کاهش دقت عملکرد آن‌ها می‌شود. جز در مدل PCA- RF که در زمان ۲ روز بیشترین دقت را دارد. در روند مدل‌سازی از انتخاب تصادفی داده‌ها، به نسبت ۷۰ به ۳۰، برای آموزش و ارزیابی رویکرد RF استفاده شده است. در نتیجه، این استثناء می‌تواند ناشی از انتخاب تصادفی جامع

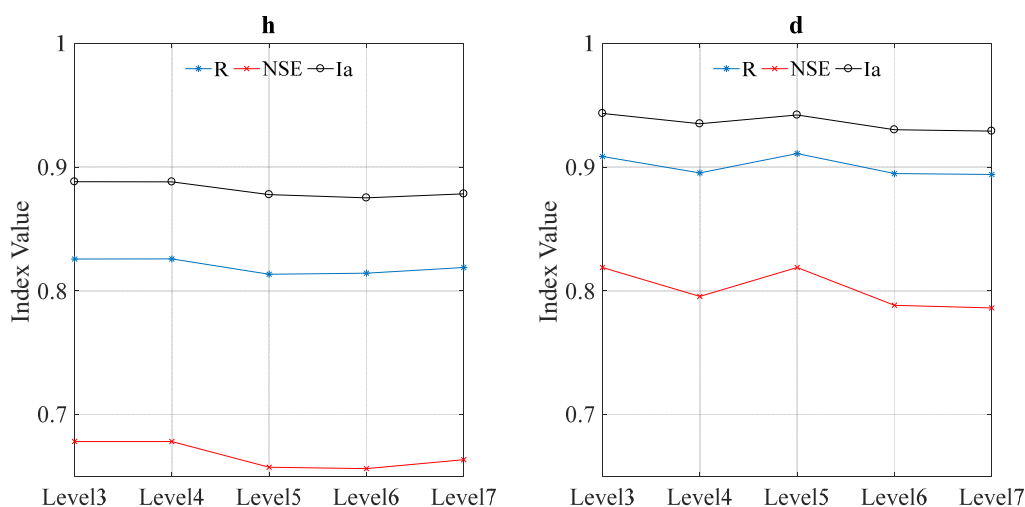
همچنین شکل ۳ مقادیر ویژه‌ی متغیرهای ۳۲گانه را نشان می‌دهد که با جدول ۳ در ارتباط است. این شکل به خوبی میزان مشارکت مؤلفه‌های انتخابی را در مدل منتخب نشان می‌دهد. پس از مدل‌سازی، بهترین نتایج آن‌ها در جدول ۴ آورده شده است. بر این اساس مشاهده می‌شود که:

مدل‌های PCA-RF از سه بردار ابتدایی PCA برای مدل‌سازی استفاده شده است که در واقع، ورودی مدل RF به یک چهارم تقلیل یافته است و همزمان، نتایج  $R$  و  $I_a$  زمان یک روز، به ترتیب، ۲۰ و ۱۱ درصد کاهش یافته‌اند. به دلیل اینکه، PCA بردارهای مستقلی از هم، که هر یک شامل خصوصیات مستقلی از ورودی‌ها هستند، ارائه می‌دهد. ممکن است بردارهای ناکافی از PCA استفاده شده و یا اینکه همبستگی بین داده‌ها، خود در یافتن الگوی جریان بسیار مؤثر بوده باشد (کاهش ۲۰ درصدی  $R$ ). در مدل‌های ترکیبی نیز، PCA با تقلیل ورودی‌ها به یک چهارم تا یک سوم تعداد اولیه، توانسته عملکرد مدل را افزایش دهد که کمترین اثر را بر روی زمان یک روز، که از پیچیدگی کمتری برخوردار است، داشته است.

در مدل‌های ترکیبی، مشاهده می‌شود که نوع موجک مادر و سطح تجزیه بسیار حائز اهمیت است، به گونه‌ای که در مدل‌های موجک دیمی که پهنای نمونه‌برداری وسیع‌تر و مشابه‌تری با سری‌زمانی مبدأ دارد، نتایج مطلوب‌تری ارائه شده است. اختلاف عملکرد این دو نوع موجک در زمان‌های ۲ و ۳ روز کمتر است. زیرا، افزایش زمان برای مدل‌های دقیق‌تر مؤثرتر بوده و کاهش بیشتری را در آنها به وجود می‌آورد. علاوه بر موجک مادر، سطح تجزیه نیز عامل مهمی است. با بررسی کلیه نتایج، ملاحظه می‌شود که حساسیت دو نوع مدل، به سطح تجزیه، در زمان شبیه‌سازی ۱ و ۲ روز به صورت نوسانی و در زمان شبیه‌سازی ۳ روز، صعودی است. همان‌طور که در شکل ۴ مشاهده می‌شود، حساسیت موجک دیمی به این مورد بیشتر است.

و مناسب داده‌ها، در این زمان نسبت به زمان یک روز، برای آموزش شبکه باشد. پس از بررسی دقیق نتایج ملاحظه می‌شود که روند کاهش آن‌ها، با افزایش زمان، برای مدل RF شدیدتر است. به طوری که، در مدل RF میزان  $R$  بین ۱۲-۱۴ درصد و  $I_a$  ۸-۹ درصد کاهش یافته است. در حالی که، در مدل‌های همراه با PCA، میزان  $R$  بین ۵-۱۰ درصد و  $I_a$  ۲-۶ درصد کاهش داشته است. شاخص توافق، نشان می‌دهد که دقت عملکرد مدل‌های PCARF و WPCARF، بسیار کمتر تحت تأثیر زمان قرار می‌گیرد و بیشترین کاهش (۶٪) مربوط به موجک دیمی در زمان شبیه‌سازی ۳ روز است که نشان از حساسیت و دقت بالای این مدل ترکیبی است. همان‌طور که در جدول ۴ ملاحظه می‌شود رویکرد ترکیبی پیشنهادی موجب شده است تا آفت کارایی و دقت مدل در زمان‌های شبیه‌سازی بیشتر بسیار کمتر از مدل‌های منفرد و یا حتی سایر مدل‌های ترکیبی پیشنهاد شده توسط سایر هیدرولوژیست‌ها باشد. این مورد یکی از مهمترین چالش‌هایی است که همواره محققین مختلف در شبیه‌سازی جریان رودخانه و در زمان‌های دورتر با آن مواجه هستند. این موضوع از مقایسه‌ی نتایج تحقیق حاضر با نتایج سایر پژوهشگران این شاخه [۷] آشکارتر نیز خواهد شد.

علاوه بر این، مشاهده می‌شود که در هر زمان، نتایج مدل‌های RF بهتر از PCA-RF شده و در زمان یک روز، با اختلاف کمی به PCA-RF haar- RF برتری یافته است. اما در سایر مدل‌های ترکیبی، حضور همزمان پیش‌پردازش و PCA سبب بهبود عملکرد مدل شده‌اند. در

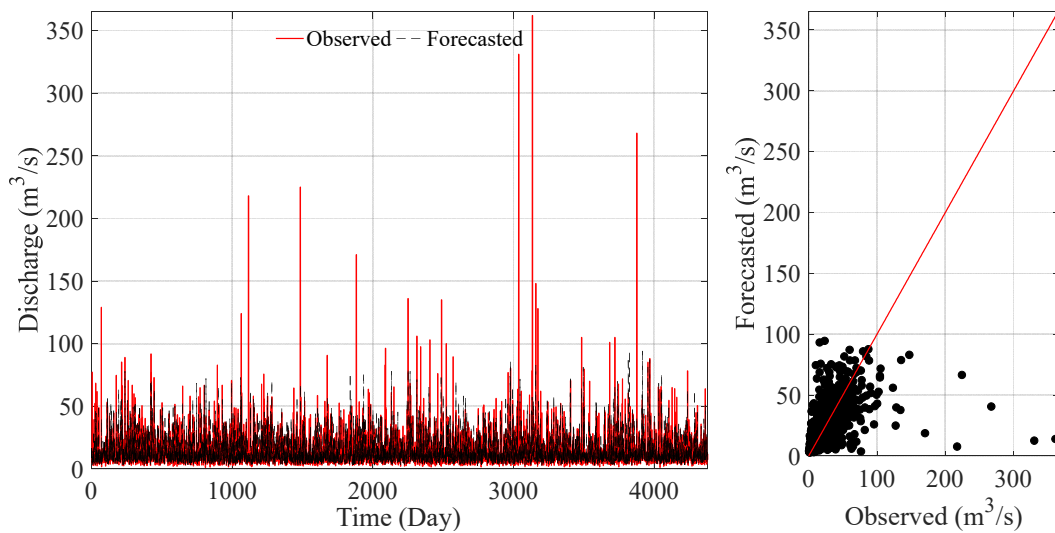


شکل ۴. نمودار تغییرات شاخص‌های ارزیابی نسبت به سطح تجزیه دو نوع موجک مادر هار (h) و دیمی (d) در زمان شبیه‌سازی یک روز

Fig. 4. Variation of validation indices and decomposition level of two mother wavelets (Haar and Dmey) at one-day-ahead river flow simulation

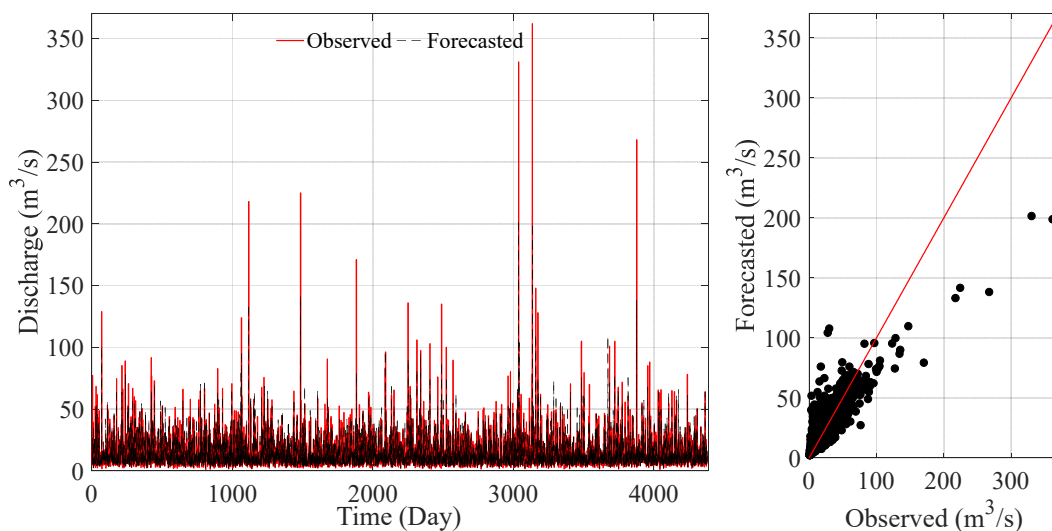
مقادیر مشاهداتی برآورد شده است. در ادامه و در شکل های ۷-۸ بهترین نتایج حاصل از مدل ترکیبی پیشنهادی برای زمان شبیه سازی دو و سه روز نمایش داده شده است. همان گونه که در این شکل ها مشاهده می شود، مدل پیشنهادی توانسته به خوبی مقادیر دبی روزانه ی رودخانه ی پلرود را شبیه سازی کند. تنها موردی که کارایی مدل را تا حدودی کاهش داده است، ضعف مدل در شبیه سازی مقادیر حداکثر در زمان های شبیه سازی بالاتر است که این نقیصه با مشارکت مؤلفه های بیشتری

از مقایسه شکل های ۳ و ۴ که برای زمان شبیه سازی یک روز هستند، مشاهده می شود که مدل WPCARF نسبت به مدل PCARF، در شبیه سازی مقادیر بیشینه عملکرد بهتری دارد. همچنین براساس نمودار پراکندگی آن ها، انسجام مدل ترکیبی با موجک در راستای خط ۴۵ درجه بیشتر است و این می تواند به دلیل کاهش خطا و عدم قطعیت های سری زمانی توسط تبدیل موجک باشد. شکل های ۵ و ۶ نیز کاهش عملکرد مدل ترکیبی در مقادیر بیشینه را نشان می دهد، بدین صورت که مقادیر بیشینه کمتر از



شکل ۵. نمودارهای دبی- زمان و پراکندگی مدل RF، برای زمان شبیه سازی یک روز.

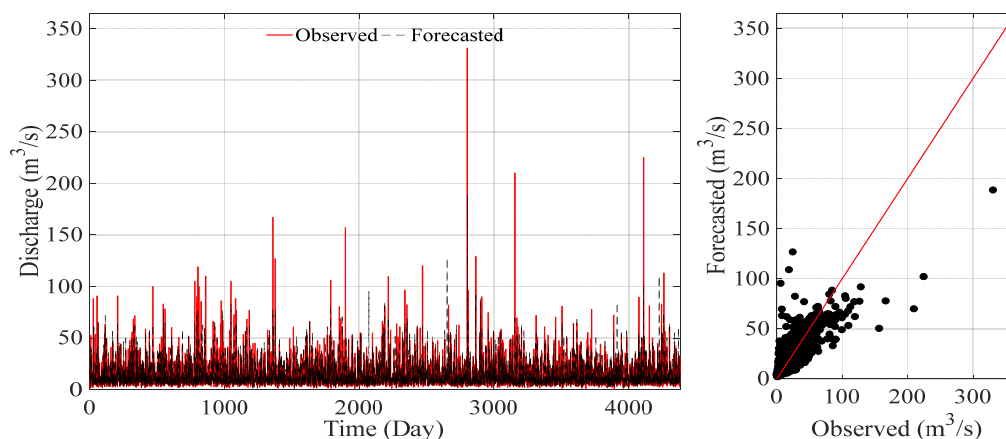
**Fig. 5. Discharge-time and scatter plots of standalone model at one-day-ahead river flow simulation**



شکل ۶. نمودارهای دبی- زمان و پراکندگی مدل Dmey- RF در سطح تجزیه ی پنجم، برای زمان شبیه سازی یک روز

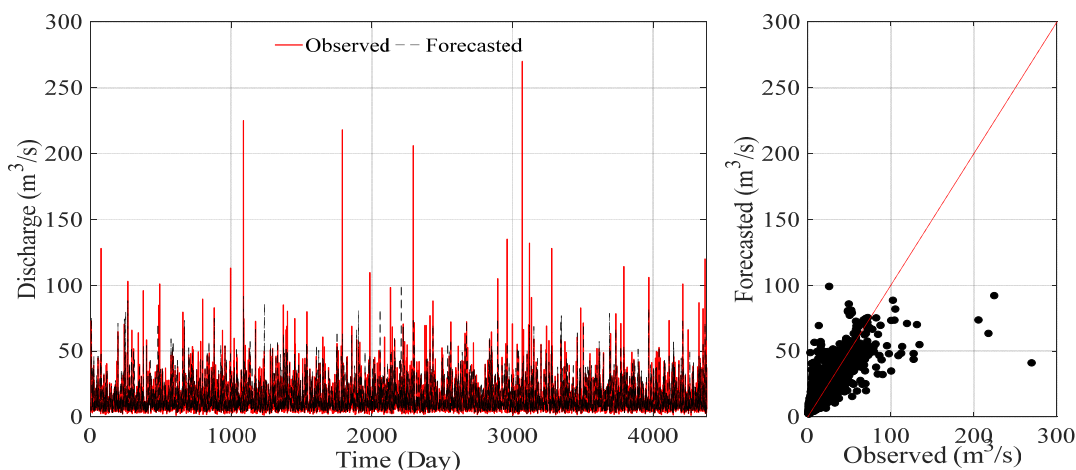
**Fig. 6. Discharge-time and scatter plots of Dmey-RF hybrid model at one-day-ahead river flow simulation (decomposition level=5)**





شکل ۷. نمودارهای دبی- زمان و پراکندگی مدل Dmey- RF در سطح تجزیه‌ی سوم، برای زمان شبیه‌سازی دو روز

**Fig. 7. Discharge-time and scatter plots of Dmey-RF hybrid model at two-day-ahead river flow simulation (decomposition level=3)**



شکل ۸. نمودارهای زمان- دبی و پراکندگی مدل Dmey- RF در سطح تجزیه‌ی ششم، برای زمان شبیه‌سازی سه روز

**Fig. 8. Discharge-time and scatter plots of Dmey-RF hybrid model at three-day-ahead river flow simulation (decomposition level=6)**

اطلاع از مقادیر آینده‌ی جریان رودخانه همواره برای هیدرولوژیست‌ها و برنامه‌ریزان و مدیران منابع آب واجد اهمیتی حیاتی است. این مقاله شبیه‌سازی جریان روزانه‌ی رودخانه‌ی پلرود را برای دوره‌های یک، دو و سه روز آینده مورد توجه قرار داد. این پژوهش یک رویکرد ترکیبی نوین را به منظور شبیه‌سازی جریان رودخانه ارائه داده است که در آن از ترکیب موجک (W)، تحلیل مؤلفه‌ی اصلی (PCA) و جنگل تصادفی (RF) استفاده شده است. داده‌های ورودی مدل‌ها شامل داده‌های هیدرومتری و هواشناسی می‌باشد. در این مقاله تأثیر فرآیند پیش‌پردازش، سطوح تجزیه و تحلیل مؤلفه‌ی اصلی بر کارایی مدل ارزیابی شده است. نتایج این تحقیق کارایی بالای مدل پیشنهادی را نشان داد، به‌گونه‌ای که مدل

از PCA تا حدود زیادی قابل جبران است؛ که البته دارای این اشکال می‌باشد که از سادگی مدل خواهد کاست، که آن هم برای زمان‌های دورتر غیرقابل اجتناب است.

در جمع بندی نتایج حاصل از اجرای مدل‌های منفرد و ترکیبی پیشنهادی می‌توان گفت که به‌رغم آنکه استفاده از تحلیل مؤلفه‌ی اصلی موجب شده تا بردارهای ورودی به شدت کاهش یافته و ورودی‌های بسیار کمتر و مدل ساده‌تری حاصل شود، لیکن نه تنها از دقت مدل نکاسته که حتی توانسته دقت آن را خصوصاً در زمان‌های شبیه‌سازی بالاتر ارتقا داده و نتایج قابل‌قبول‌تری را در عین سادگی مدل ارائه کند.

#### ۴- نتیجه‌گیری

- prediction improving using the integrated model based on kernel-wavelet and complementary ensemble empirical mode decomposition, CEEJ (XML). In Persian.
- [9] A. D. Mehr, An improved gene expression programming model for streamflow forecasting in intermittent streams, *Journal of hydrology*, 563 (2018) 669-678.
- [10] M. Rezaie-Balf, S. Fani Nowbandegani, S. Z. Samadi, H. Fallah, S. Alaghmand, An ensemble decomposition-based artificial intelligence approach for daily streamflow prediction. *Water*, 11(4) (2019) 709.
- [11] F. J. Chang, P. A. Chen, Y. R. Lu, E. Huang, K. Y. Chang, Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control, *Journal of Hydrology*, 517 (2014) 836-846.
- [12] S. Kabir, S. Patidar, G. Pender, Investigating capabilities of machine learning techniques in forecasting stream flow, In *Proceedings of the Institution of Civil Engineers-Water Management*, 173 (2) (2020) 69- 86.
- [13] R. Noori, A. Farokhnia, S. Morid, H. Riahi Madvar, (2009). Effect of input variables preprocessing in artificial neural network on monthly flow prediction by PCA and wavelet transformation, *Journal of Water and Wastewater*, (1) (2009) 13-22. In Persian.
- [14] M.R. Najafi, H. Moradkhani, T. C. Piechota, Ensemble streamflow prediction: climate signal weighting methods vs. climate forecast system reanalysis, *Journal of Hydrology*, 442 (2012) 105-116.
- [15] C. Prieto, N. Le Vine, D. Kavetski, E. García, R. Medina, Flow prediction in ungauged catchments using probabilistic Random Forests regionalization and new statistical adequacy tests, *Water Resources Research*, 55(5) (2019) 4364-4392.
- [16] H.A. Afan, M.F. Allawi, A. El-Shafie, Z.M. Yaseen, A.N. Ahmed, M.A. Malek, A. Sefelnasr, Input attributes optimization using the feasibility of genetic nature inspired algorithm: Application of river flow forecasting, *Scientific Reports*, 10(1) (2020) 1-15.
- [17] R. Noori, A.R. Karbassi, A. Moghaddamnia, D. Han, M.H. Zokaei-Ashtiani, A. Farokhnia, M.G. Gousheh, Assessment of input variables determination on the

پیشنهادی ضمن ساده‌سازی مدل‌های شبیه‌سازی و استفاده از ۳ تا ۱۰ بردار ورودی حاصل از PCA به جای ۱۰ تا ۵۲ بردار ورودی، همزمان توانست کارایی مدل را ارتقا دهد، این موضوع خصوصاً در ارتقای زمان‌های شبیه‌سازی دو و سه روز کاملاً قابل ملاحظه است. ضمن اینکه در زمان‌های شبیه‌سازی دورتر سطح تجزیه‌ی بالاتر برای رسیدن به بهترین کارایی مدل مورد نیاز است. در پایان، این تحقیق کارایی رضایت‌بخش رویکرد ترکیبی پیشنهادی در شبیه‌سازی جریان رودخانه، همراه با ارائه‌ی مدل‌های ساده‌تر، را نشان داد.

## مراجع

- [1] D. P. Solomatine, A. Ostfeld, Data-driven modelling: some past experiences and new approaches, *Journal of hydro informatics*, 10(1) (2008) 3-22.
- [2] N. Khairuddin, A.Z. Aris, A. Elshafie, T. Sheikhy Narany, M.Y. Ishak, N.M. Isa, Efficient forecasting model technique for river stream flow in tropical environment. *Urban Water Journal*, 16(3) (2019) 183-192.
- [3] H. Tongal, M. J. Booij, M, Simulation and forecasting of stream flows using machine learning models coupled with base flow separation, *Journal of hydrology*, 564 (2018) 266-282.
- [4] D. Hussain, A.A. Khan, A, Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan, *Earth Science Informatics*, (2020) 1-11.
- [5] V. Nourani, A. H. Baghanam, J. Adamowski, O. Kisi, Applications of hybrid wavelet-artificial intelligence models in hydrology: a review, *Journal of Hydrology*, 514 (2014) 358-377.
- [6] N. Nourani, A. Davanlou Tajbakhsh, A. Molajou, H. Gokcekus, Hybrid wavelet-M5 model tree for rainfall-runoff modeling, *Journal of Hydrologic Engineering*, 24(5) (2019) 04019012.
- [7] Y. Sun, J. Niu, B. Sivakumar, A comparative study of models for short-term streamflow forecasting with emphasis on wavelet-based approach. *Stochastic Environmental Research and Risk Assessment*, 33(10) (2019) 1875-1891.
- [8] K. Roushangara, R. Ghasempourb, Monthly precipitation

- using hydro-meteorological data as inputs, *Journal of Hydrology*, 586 (2020) 124371.
- [23] M. Motamednia, A. Nohegar, A. Malekian, M. Saberi Anari, K. Karimi Zarchi, Impacts of combining meteorological and hydrometric data on the accuracy of streamflow modeling, *Environmental Resources Research*, 7(2) (2019) 147-164.
- [24] Y-F. Sang, A review on the applications of wavelet transform in hydrology time series analysis, *Atmospheric Research*, 122 (2012) 8-15.
- [25] S. Mallat, *A wavelet tour of signal processing: the sparse way 3rd edn*, New York: Academic, (2008).
- [26] L. Breiman, Random forests. *Machine learning*, 45(1) (2001) 5-32.
- [27] H. Abdi, L. J. Williams, *Principal component analysis*, Wiley interdisciplinary reviews: computational statistics, 2(4) (2010) 433-459.
- [28] F. Anctil, M.H. Ramos, Verification Metrics for Hydrological Ensemble Forecasts, *Handbook of Hydro Meteorological Ensemble Forecasting*, (2019) 893-922
- [29] C. J. Willmott, on the validation of models, *Physical geography*, 2(2) (1981) 184-194.
- [30] B. G. Tabachnick, L. S. Fidell, *Experimental designs using ANOVA*, Belmont, CA: Thomson/Brooks/Cole, (2007).
- SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction, *Journal of Hydrology*, 401(3-4) (2011) 177-189.
- [18] Y. Hassanzadeh, K. A. Abdi, N. M. Shafiei, S. Khoshtinat, Daily streamflow forecasting of Nooranchay river using the yybrid model of Artificial Neural Networks-Principal Component Analysis, *Journal of Soil and Water Science*, 25 (3) (2015) 53- 63. In Persian.
- [19] M. Ehteram, H.A Afan, M. Dianatikhah, A. N. Ahmed, C. Ming Fai, M. S. Hossain, A. Elshafie, Assessing the predictability of an improved ANFIS model for monthly streamflow using lagged climate indices as predictors, *Water*, 11 (6) (2019) 1130.
- [20] L. Diop, A. Bodian, K. Djaman, Z. M. Yaseen, R.C. Deo, A. El-Shafie, L. C. Brown, The influence of climatic inputs on stream-flow pattern forecasting: case study of Upper Senegal River, *Environmental Earth Sciences*, 77(5) (2018) 182.
- [21] S.J. Hadi, M. Tombul, (2018). Monthly streamflow forecasting using continuous wavelet and multi-gene genetic programming combination, *Journal of Hydrology*, 561 (2018) 674-687.
- [22] R. M. Adnan, Z. Liang, S. Heddam, M. Zounemat-Kermani, O. Kisi, B. Li, Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin

چگونه به این مقاله ارجاع دهیم

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) 2837-2850.

DOI: [10.22060/ceej.2020.17520.6589](https://doi.org/10.22060/ceej.2020.17520.6589)



