BioOne COMPLETE

Forecasting Daily River Flow Using an Artificial Flora–Support Vector Machine Hybrid Modeling Approach (Case Study: Karkheh Catchment, Iran)

Authors: Dehghani, Reza, Torabi Poudeh, Hassan, Younesi, Hojatolah, and Shahinejad, Babak

Source: Air, Soil and Water Research, 13(1)

Published By: SAGE Publishing

URL: https://doi.org/10.1177/11786221209696 59

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Forecasting Daily River Flow Using an Artificial Flora–Support Vector Machine Hybrid Modeling Approach (Case Study: Karkheh Catchment, Iran)

Reza Dehghani, Hassan Torabi Poudeh, Hojatolah Younesi and Babak Shahinejad
Faculty of Agriculture, Lorestan University, Khorramabad, Iran.

Air, Soil and Water Research
Volume 13: 1–10
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1178622120969659

ABSTRACT: In this study, the hybrid support vector machine–artificial flora algorithm method was developed and the obtained results were compared with those of the support vector–wave vector machine model. Karkheh catchment area was considered as a case study to estimate the flow rate of rivers using the daily discharge statistics taken from hydrometric stations located upstream of the dam in the statistical period of 2008 to 2018. Necessary criteria including coefficient of determination, root mean square error (RMSE), mean absolute error (MAE), and Nash–Sutcliffe coefficient were used to evaluate and compare the models. The results illustrated that the combined structures provided acceptable results in terms of river flow modeling. Also, a comparison of the models based on the evaluation criteria and Taylor's diagram demonstrated that the proposed hybrid method with the correlation coefficient of $R^2 = 0.924$ to 0.974, RMSE = 0.022 to 0.066 m^3/s, MAE = 0.011 to 0.034 m^3/s, and Nash-Sutcliffe (NS) coefficient = 0.947 to 0.986 outperformed other methods in terms of estimating the daily flow rates of rivers.

KEYWORDS: Artificial flora, support vector machine, wavelet, Karkheh catchment

RECEIVED: August 23, 2020. **ACCEPTED:** October 9, 2020.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Hassan Torabi Poudeh, Faculty of Agriculture, Lorestan University, Khorramabad, Iran. Email: torabi.ha@lu.ac.ir

Introduction

The most important concern in managing flood and preventing the ensuing economic and life-threatening damages is accurate estimation of river flow. Accordingly, application of reliable methods to the prediction of river flow to plan for timely use of water resources is gaining growing significance.[1] In other words, accurate river flow forecasting can play a vital role in water resources planning and management. However, various factors affect this phenomenon, making its analysis difficult. Hence, it is necessary to incorporate influential factors in a model for estimating river flow at an acceptable level.[2,3] Today, intelligent systems are widely used for estimating nonlinear phenomena. One of the methods that has been considered in the field of hydrology is the support vector machine (SVM) model. This model has good performance and optimization algorithms have been applied to it in recent years to increase its accuracy and reduce its error rate. In metaheuristic algorithms, due to the addition of velocities with random values to the problem variables, they may be inadvertently transferred out of their defined ranges. On the contrary, based on the values of discrete variables in other algorithms, the answers obtained in all iterations are in the domain of the problem. As a result, finding a global optimum solution to some particular cases takes a longer amount of time, causing the problem to be trapped in local optima.[4] Therefore, the algorithm of artificial flora (AF), which is a combination of continuous and discrete optimizations, has been developed for large-scale problems to shorten the time required to achieve a global optimal solution and prevent trapping in local optima. This algorithm has an acceptable ability to solve nonlinear problems with large dimensions at an appropriate convergence speed. With these in mind, this study combines the AF algorithm with SVM. In recent years, a number of studies have attempted to present smart hybrid models for forecasting river flow rate. In the following, some cases are presented.

Huang et al[5] predicted monthly flow in the Huaxian Station in China using a SVM and their results proved the high accuracy of the proposed model. Sedighi et al[6] predicted the rainfall runoff process in Rudak catchment area located in northeastern Tehran by artificial neural networks (ANNs) and SVM using 92 Modis sensors within the statistical period of 2003-2005. They demonstrated the acceptable performance of the SVM model in estimating runoff. In another study, Ghorbani et al[7] used supportive modeling machines and ANNs to predict the daily flow of the Cypress River in Texas. They employed correlation coefficient and root mean square error (RMSE) to evaluate the models and demonstrated the proper performance of the SVM in predicting river flow and its better accuracy than ANNs. Samadianfard et al[8] proposed a hybrid model comprising SVM regression and fly algorithm and compared its performance with the decision tree model in estimating Dubai River and Venar located in Iran. Superior performance of the proposed hybrid model was proven in this research. Having employed support models and decision trees to predict the monthly flow of the Swat River in Pakistan, Adnan et al[9] showed effectiveness of the SVM model. Rajaee et al[10] used a combination of wavelet conversion and SVM models, nephropathy, ANNs, and genetic planning to predict the daily flow of the Dunbe River in Serbia. The results showed that the hybrid model comprising vector machine model and wavelet model experienced less serious errors than other hybrid models did. Alizadeh et al[11] examined the hybrid model of support vector-wavelet machine to predict the daily flow of the Souris River in the northern United States and observed the efficiency and accuracy of the proposed model. Hussain and Ahmed Khan[12]

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

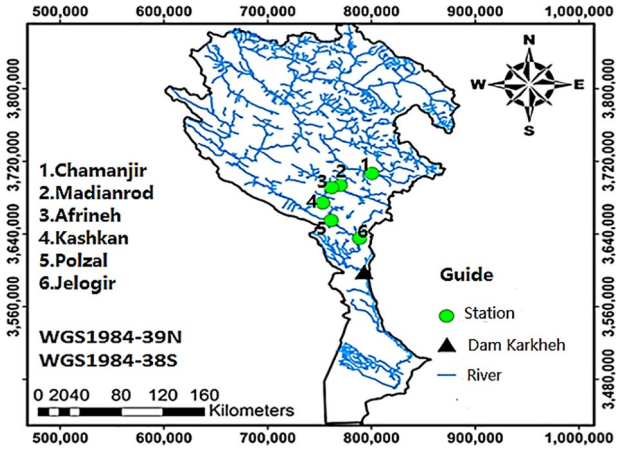


Figure 1. The studied region.

conducted a study on predicting the flow of the Hanza River in Pakistan employing the supportive vector machine models, ANNs, and random forest. The results showed better performance of the random forest model.

The rivers of Karkheh catchment area are generally considered as the most important watersheds in Iran. They constitute the major source of water supply to the adjacent areas for agriculture and drinking purposes. However, the drastic reduction in their flow indicates the necessity of simulating river flow in this basin and presenting measures to manage water more than ever. Therefore, the aim of this study was to predict the daily flow of Karkheh catchment rivers using a hybrid model comprising SVM and AF as well as to compare the results with those of the hybrid SVM-wavelet model.

Materials and Methods

The studied region

Karkheh basin with an area of 51 640 square kilometers in southwestern Iran is located in the range of 30° to 35° N and 46° to 49° E. The Karkheh catchment is part of the Persian Gulf catchment area, which is bounded in the north by the Sirvan, Sefidrood, and Qarachai river basins; in the west by the Iran-Iraq border area; in the south by a part of the western borders of the country; and in the east by the Dez River. The Karkheh river is 900 km long and it is the third largest river in the country in terms of the average annual discharge (8.5 billion cubic meters). Figure 1 shows the selected stations of the Karkheh catchment area, which did not have any missing or homogeneous data. The data were obtained from the Lorestan Regional Water Company and the Khuzestan Water and Electricity Organization (Table 1).

Table 1. Station characteristics.

NUMBER	STATION	LATITUDE	LONGITUDE	AREA (KM^2)
1	Chamanjir	33° 26′ 37″	48° 14′ 38″	1140
2	Madianrod	33° 18′ 20″	47° 48′ 59″	780
3	Afrineh	33° 18′ 51″	47° 53′ 22″	800
4	Kashkan	39° 19′ 52″	47° 53′ 39″	820
5	Polzal	32° 25′	48° 9′	90
6	Jelogir	32° 9′	47° 43′	120

Support vector machine

SVM was developed in the early 1990s by Vapnik[13] and Misra et al.[14] Support vector machine embodies the structural risk minimization (SRM) principle, which minimizes the expected error of a learning model, reduces the overfitting problem, and enables better generalization.[13] It is an efficient learning system based on optimization theory that uses minimization of structure error and leads to an optimum response [46]. In the regression model, SVM is a function related to the Y-dependent variable and an X-independent one. Similar to regression issues, the relation between independent and dependent variables is assumed to be clear, as given below[15]

$$f(x) = W^T.\varnothing(x) \quad b \tag{1}$$

$$y \quad f(x) \quad \text{noise} \tag{2}$$

If W^T is the coefficient vector, b is fixed for the regression function properties and $\varnothing$ is Kernel function, whose form is given below. These properties are further corrected by training the support vector model using data collection.[16] To calculate W and b, error function (equation (3)) in SVM- ε must be minimized (equations (3) and (4))[13]

$$\frac{1}{2}W^T.W + C\sum_{i=}^{N}\varepsilon_i + C\sum_{i=}^{N}\varepsilon_i^* \tag{3}$$

$$W^T.\varnothing(X_i) + b - y_i \le \varepsilon + \varepsilon_i^* \tag{4}$$

$$y_i - W^T.\varnothing(X_i) - b \le \varepsilon + \varepsilon_i,\ \varepsilon_i,\ \varepsilon_i^* \ge \quad ,\ i = \ ,\ ,\&,\ N \tag{5}$$

where C is a true and positive value that determines any deviation in the model training error. $\varnothing$ is kernel, N is the number of samples, and ε_i and ε_i^* are deficient variables. Support vector machine function is re-written as follows

$$f(x) = \sum_{i=}^{N}\bar{\alpha}_i\varnothing(x_i)^T.\varnothing(x) + b \tag{6}$$

where $\bar{\alpha}_i$ is the Lagrange coefficient. $\varnothing(x)$ is calculated in a special space.[13] To solve the problem, a common pattern in the support vector model is the kernel function

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

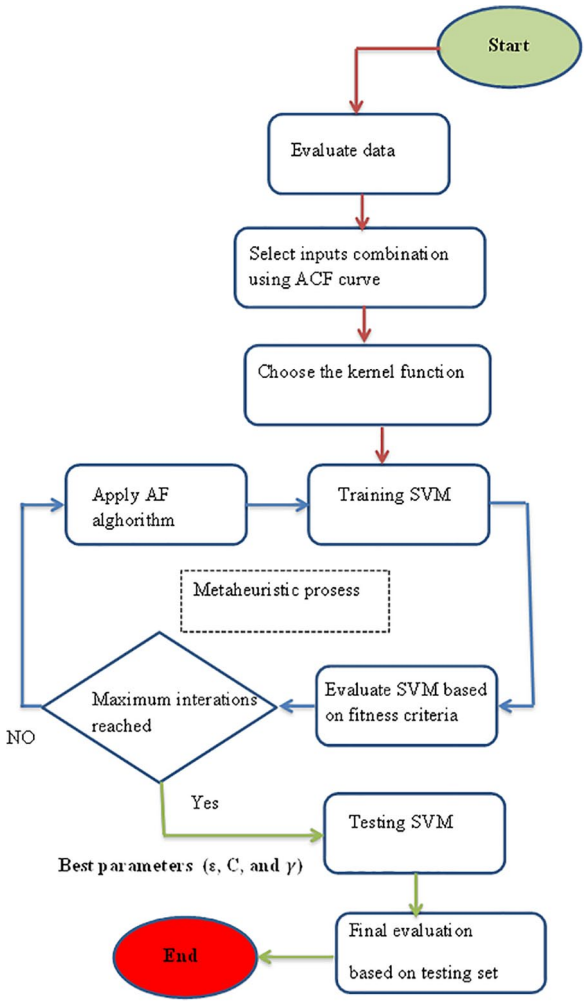


Figure 2. Flowchart of AF algorithm.
AF indicates artificial flora; SVM, support vector machine.

$$K\left(X_J, X\right) = \varnothing\left(X_i\right)^T \sqrt{b - ac} \tag{7}$$

Different Kernel functions have been used for ε-SVM fabrication. Different kernel functions in the support vector model include polynomial kernel, radial basis functions (RBFs), and linear kernel, which, due to their popularity and widespread use,[17,18] have been employed in this study. Of note, vector machine calculations were conducted based on coding in MATLAB software and the parameters were optimized

$$k\left(x, x_j\right) = \left(t + x_i \cdot x_j\right)^d \tag{8}$$

$$K\left(x, x_i\right) = \exp^{f()}\left(-\frac{\left\| x - x_i \right\|}{\sigma}\right) \tag{9}$$

$$k\left(x, x_j\right) = x_i \cdot x_j \tag{10}$$

Algorithm of AF

Biological bases. Flora disperses its grains in different ways, which are divided into autochory and allochory. Autochory involves self-dispersal of grains, while allochory is the process of distributing grains through external forces. Autochory provides a condition for independent migration of flora to an appropriate environment. On the contrary, allochory provides conditions for migration to far regions. The different methods for grain distribution reduce the probability of plant extinction. Natural environment under a harsh condition and competition may reduce flora distribution. Following the migration of flora to a new environment, flora species develop.[19] Flora migration can change the distribution region and cause development, extinction, and emergence of flora. Flora cannot move and is not smart. However, it can find the best place for life. Flora randomly distributes grains during migration and reproduction. A grain can survive for a while. Flora survives and distributes grains in its surrounding environment. It develops and adapts to the environment under a harsh condition. Before the extinction of flora in a region, it may grow in a new environment. Grains may grow in a new region and replicate by multi-replication. Flora finds an optimum region for growth, development, extinction, and growth.[19]

AF algorithm theory

AF algorithm is composed of 4 main elements: main flora, child flora, flora position, and distribution distance. Child flora acts as the grain for the main florae and it cannot distribute grain. Distribution distance means grain distribution distance. There are 3 behavior patterns: development behavior, distribution behavior, and selection behavior.[20-22] Development behavior means flora development for adaptation to environmental behavior.[23-25] Distribution behavior stands for the movement of grains. Grains can move using allochory and autochory. Selection behavior suggests survival and extinction for environmental reasons. Figure 2 shows a flowchart for the establishment of AF algorithm.

Wavelet transform

A wavelet transform is presented as a replacement method for Fourier transformation and its purpose is to dominate the degradation of frequency within a short amount of time. For the transform wavelets such as short-time transformation, the signal is divided into windows.[26] The most important difference between the above-mentioned 2 methods is the changes of frequency type in wavelet transform, in which scale is found rather than frequency. Based on wavelet transform, high scales are expanded and thus the details can be analyzed.[27] A wavelet means a small wave and it is a small part of the main signal whose energy is concentrated in time. The mother signal can be degraded to wavelets and different scales. Wavelets include

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Table 2. Selected combinations of input parameters.

NUMBER	INPUT STRUCTURE	OUTPUT
1	$Q(t-1)$	$Q(t)$
2	$Q(t-1)$, $Q(t-2)$	$Q(t)$
3	$Q(t-1)$, $Q(t-2)$, $Q(t-3)$	$Q(t)$
4	$Q(t-1)$, $Q(t-2)$, $Q(t-3)$, $Q(t-4)$	$Q(t)$

the transformed and dilated samples with fluctuations. Based on the properties of wavelets, time series of continuous wavelet transform (CWT) can be analyzed.[28] Wavelet transformation is defined in the continuous and discrete forms.

Continuous wavelet transform is defined based on equations (11) and (12) as follows[27]

$$\mathrm{CWT}_f^{\psi}(s,\tau) = \Psi_f^{\psi}(s,\tau) = \frac{1}{\sqrt{|s|}}\int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{t-\tau}{s}\right)dt = \langle f(t), \psi_{s,\tau}(t)\rangle \quad (11)$$

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right) \quad (12)$$

Equation (12) is the relationship between 2 variables of s and τ, where s is the scaling parameter and τ is the translation parameter. In addition, * shows the mixed paired, ψ is the window function for the mother wavelet, and $/(\sqrt{s})\psi((t-\tau)/s)$ is the wavelet of transformation and scale change for the mother wavelet.[28] The term "mother" is used because all the transformed and dilated (daughter wavelet) versions are obtained from the function. The mother wavelet is a pattern for other windows, showing the vector cross of 2 functions in the signal space.

Evaluation criteria

In this study, to evaluate the accuracy and efficiency of the models, the indicators of coefficient of determination (R^2), RMSE, mean absolute error (MAE), and Nash-Satcliffe (NS) coefficient are used according to the given relationships.[29] The best values for these 4 criteria are one, zero, zero, and one, respectively

$$R \quad \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})\ \sum_{i=1}^{N}(y_i-\bar{y})}} \quad -1 \le R \le 1 \quad (13)$$

$$\mathrm{RMSE} \quad \sqrt{\frac{}{N}\sum_{i=1}^{N}(x_i-y_i)} \quad \le \mathrm{RMSE} \le 1 \quad (14)$$

$$\mathrm{MAE} = \frac{}{n}\sum|x_i-y_i| \quad \le \mathrm{MAE} \le \quad (15)$$

$$\mathrm{NS} = -\frac{\sum_{i=}^{N}(x_i-y_i)}{\sum_{i=}^{N}(x_i-\bar{y})} \quad -\infty \le \mathrm{NS} \le \quad (16)$$

In the above relations, x_i and y_i are the observational and computational values in the ith temporal step, respectively; N is the number of temporal steps; and $\bar{x}$ and $\bar{y}$ are the means of the observational and computational values, respectively.

Results and Discussion

Combinational selection of input variables is an important step for modeling. Hence, the cross-correlation between input and output variables was calculated and input parameters were selected for obtaining an optimum model for predicting the flow rate of the river of Karkheh catchment. The results are shown in Table 2. In Table 3, $Q(t-1)$, $Q(t-2)$, $Q(t-3)$, and $Q(t-4)$ columns show river flow at times $t-1$, $t-2$, $t-3$, and $t-4$ and $Q(t)$ shows river flow at time t. To facilitate a better understanding of the nature of the mechanism, pattern complexity and memory are increased, while the model precision decreases. To model the river flow, most of the efficient data were used as the training data. This study investigated the effects of streamflow using return flow. The cross-correlation

Table 3. Cross-correlation between input and output variables.

STATION	OUTPUT	INPUT			
		$Q(T-1)$	$Q(T-2)$	$Q(T-3)$	$Q(T-4)$
Chamanjir	$Q(t)$	0.940	0.921	0.893	0.854
Madianrod		0.890	0.864	0.814	0.754
Afrineh		0.914	0.892	0.865	0.813
Kashkan		0.925	0.904	0.876	0.834
Polzal		0.920	0.896	0.885	0.822
Jelogir		0.924	0.897	0.880	0.826

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Table 4. Analysis of AF-support vector machine for selected stations.

STATION	TRAINING				TESTING			
	R^2	RMSE (M^3/S)	MAE (M^3/S)	NS	R^2	RMSE (M^3/S)	MAE (M^3/S)	NS
Chamanjir	0.956	0.033	0.012	0.978	0.974	0.022	0.011	0.986
Madianrod	0.905	0.075	0.037	0.923	0.924	0.066	0.034	0.947
Afrineh	0.944	0.056	0.028	0.965	0.955	0.036	0.018	0.974
Kashkan	0.951	0.030	0.014	0.972	0.968	0.025	0.014	0.977
Polzal	0.923	0.064	0.031	0.937	0.941	0.054	0.028	0.962
Jelogir	0.936	0.061	0.030	0.951	0.948	0.048	0.025	0.967

Abbreviations: AF, artificial flora; MAE, mean absolute error; NS, Nash-Satcliffe; RMSE, root mean square error.

between input and output data was higher than 0.750 and different combinations of input parameters were used for estimating the optimum model for Karkheh catchment. The data were obtained from hydrometric stations of Chamanjir, Madianrod, Afrineh, Kashkan, Polzal, and Jologir over the years 2008-2018. The total number of 2920 records for training and other 730 records for assessing accuracy were selected. It should be mentioned that 80% of the data were selected for training and 20% for testing randomly.[30,31] Cross-correlation between input and output variables is shown in Table 3.

The results for support vector model-AF algorithm

In this study, a hybrid method comprising the SVM and AF algorithm is proposed. The optimal values of the characteristics of the SVM model including ε and C were determined. Also, different kernels were examined and based on their performances as well as the used kernel functions, the RBF function was adopted due to its higher accuracy in estimating the daily flow rate of rivers.[32,33] In this function, the characteristic of γ must be determined. Therefore, in general, to predict the daily flow rate of rivers by using the SVM model, it is necessary to calculate the optimal values of the 3 mentioned characteristics, namely ε, C, and γ, for which the best values are determined by AF algorithm. By using the developed models, the model with the least error could be determined and its characteristics be selected as the optimal values of ε, C, and γ. The AF algorithm was inspired by the migration and reproduction behavior of flora, comprising 3 main behaviors including evolution, distribution, and selection. This algorithm is able to prevent reaching a local optimal solution. It incorporates both self-pollination and cross-pollination behaviors. While the former searches around itself for the optimum solution, the latter explores a broader space, which improves the capability of the algorithm to find the optimum solution and increases the convergence speed to the optimal solution. The results of the hybrid SVM-AF algorithm are given in Table 4. According to the table, the proposed hybrid model for the basin station of the catchment area is more accurate and less erroneous due to the lack of intervention of the base flow along the river. The correlation coefficient of $R^2 = 0.924$ to 0.974, RMSE = 0.022 to 0.066 m^3/s, MAE = 0.011 to 0.034 m^3/s, and Nash-Sutcliffe (NS) coefficient = 0.947 to 0.986 were achieved at the validation step of the model. Figure 3 shows the distribution diagram of the proposed hybrid model at the validation step, indicating the best fit line of computational values y = x. In this figure, the estimated and observational values, except for a few points, are on the semiconductor line, indicating their equality on (y = x). Also, as can be seen in the figure, the hybrid model has an acceptable performance in predicting the maximum and minimum with high proximity to the actual values.

The results of *SVM*-wavelet

To evaluate the results of the hybrid model, first, the input parameters were broken down into sub-signals using wavelet conversion and then the mentioned sub-signals were added to the model of the backup vector machine as input, constituting the combined model. One of the most important and fundamental points in this study was the study of different wave functions and it was observed that the Mexican cap wave had better performance than other functions. Table 5 shows the results of the hybrid model for the selected stations of the Karkheh catchment area. The table indicates that the proposed hybrid model for Chamanjir station had higher accuracy and lower error with the correlation coefficient of $R^2 = 0.915$ to 0.964, RMSE = 0.031 to 0.084 m^3/s, MAE = 0.015 to 0.068 m^3/s, and NS coefficient = 0.930 to 0.978. Figure 4 demonstrates the best fit line (y = x) for the distribution diagram of the computational values of the support wave vector machine in the validation stage. In this figure, the estimated and observational values except for a few points are on the semiconductor line (y = x), indicating their equality. Also, as observed in the figure, the hybrid model has an acceptable

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

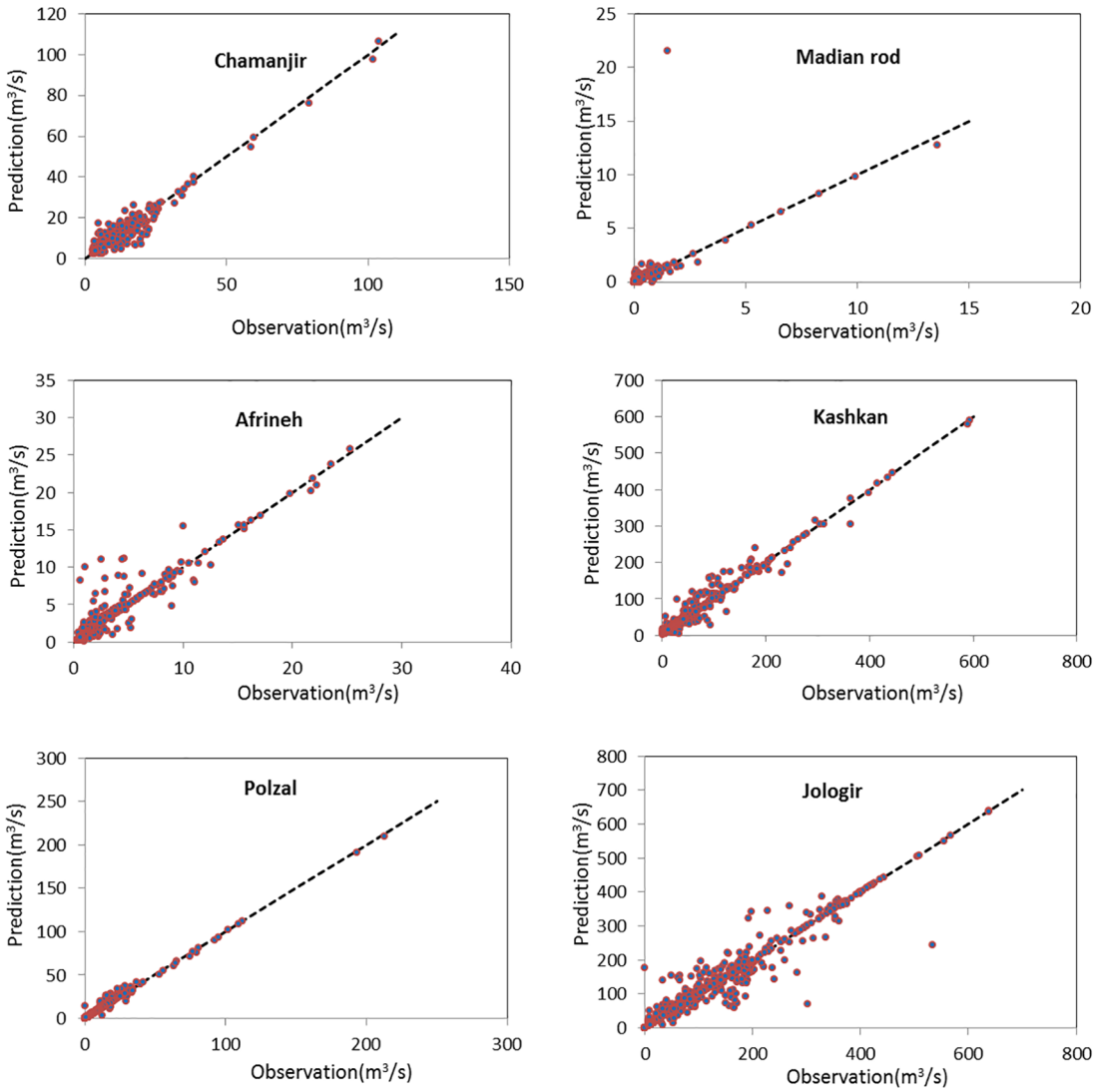


Figure 3. Scatter figure of observational and calculated amounts by AF-support vector machine in assessing accuracy phase.
AF indicates artificial flora.

Table 5. Analysis of the wavelet-support vector machine for selected stations.

STATION	TRAINING				TESTING			
	R^2	RMSE (M³/S)	MAE (M³/S)	NS	R^2	RMSE (M³/S)	MAE (M³/S)	NS
Chamanjir	0.942	0.048	0.023	0.968	0.964	0.031	0.015	0.978
Madianrod	0.887	0.095	0.075	0.910	0.915	0.084	0.068	0.930
Afrineh	0.921	0.074	0.058	0.934	0.936	0.079	0.058	0.945
Kashkan	0.932	0.063	0.037	0.948	0.956	0.044	0.021	0.968
Polzal	0.894	0.088	0.074	0.918	0.927	0.081	0.063	0.941
Jelogir	0.914	0.084	0.063	0.925	0.942	0.077	0.051	0.957

Abbreviations: MAE, mean absolute error; NS, Nash-Satcliffe; RMSE, root mean square error.

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

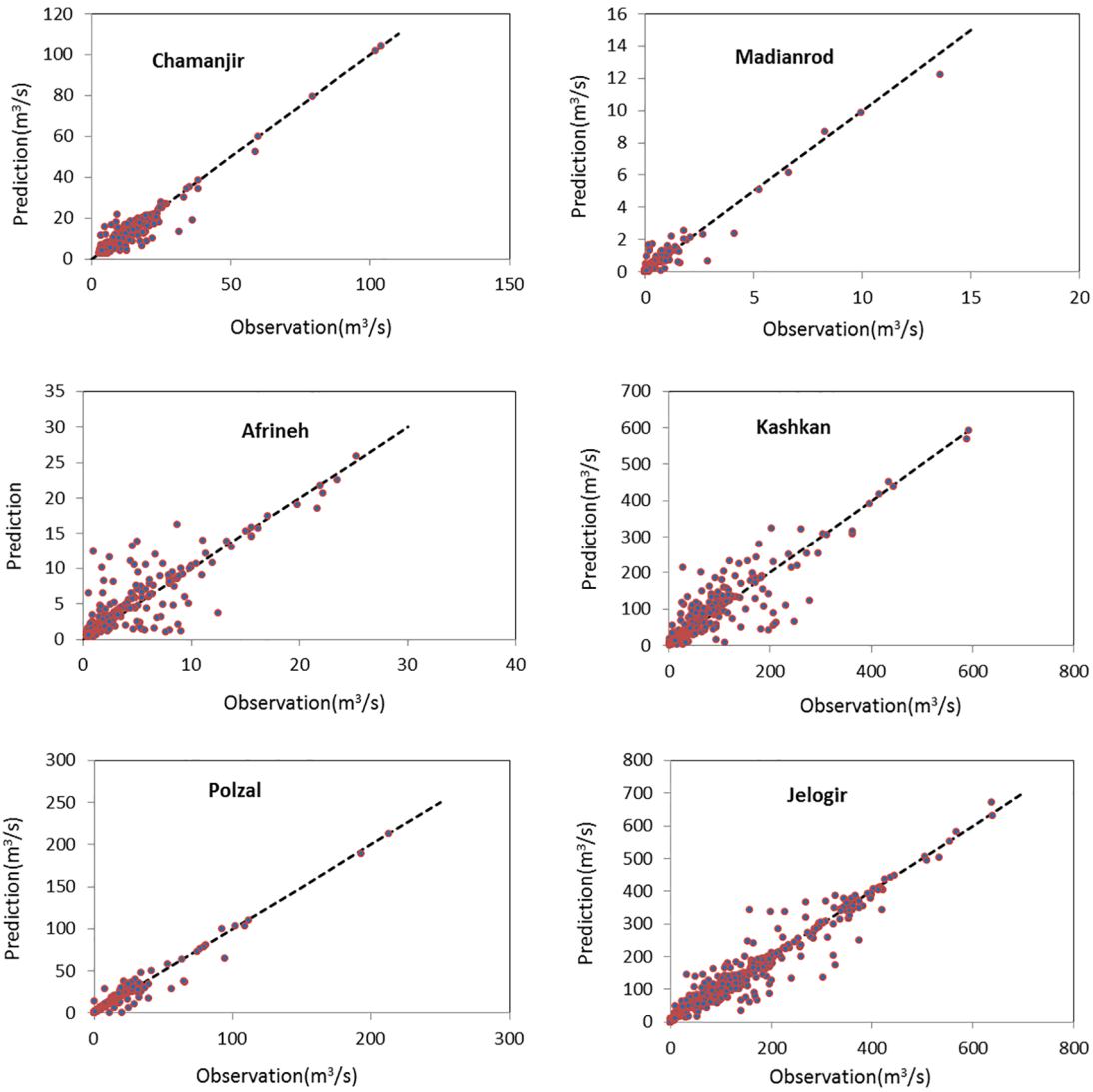


Figure 4. Scatter figure of observational and calculated amounts for the wavelet-support vector machine.

performance in predicting intermediate values with high proximity to the actual values.

Comparison of the performances of the models

By considering the optimal results of each hybrid artificial intelligence model and comparing the findings, the capability of both models to simulate the flow in the Karkheh catchment area was proved (Figure 5). Figure 5 illustrates the diagrams of the observed and calculated values for the studied models with respect to time in all of the studied stations. As observed earlier, the SVM-AF optimization algorithm model has shown an acceptable ability to estimate the minimum and maximum values. Moreover, the SVM-wavelet model exhibits appropriate performance in estimating the intermediate values such that they will be close to the observed values. Figure 6 displays the diagrams of relative error of the studied models with respect to the observed values. In this figure, the SVM-AF optimization algorithm has lower error than the SVM-wavelet such that the relative error values of the latter model are higher for all of the studied stations.[34,35]

Taylor diagrams were used to analyze and evaluate the models used in the study, as shown in Figure 7. A clear advantage of Taylor's diagram is that it uses 2 common correlation statistics: the correlation coefficient and the standard deviation.[36] The closer the predicted value to the observational value is in terms of correlation coefficient and standard deviation, the higher the predictability will be. Taylor's performance chart shows that the AF-SVM model has the highest efficiency and performance, because the predicted standard deviation value has the closest distance to the standard deviation of observational data and the correlation coefficient shows the highest value. According to all the evaluation criteria, the models with the highest predictive power of AF-SVM and WSVM have the lowest predictability.

Conclusions

In this study, an attempt was made to evaluate the performance of the models in simulating the daily flow of rivers in the Karkheh

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

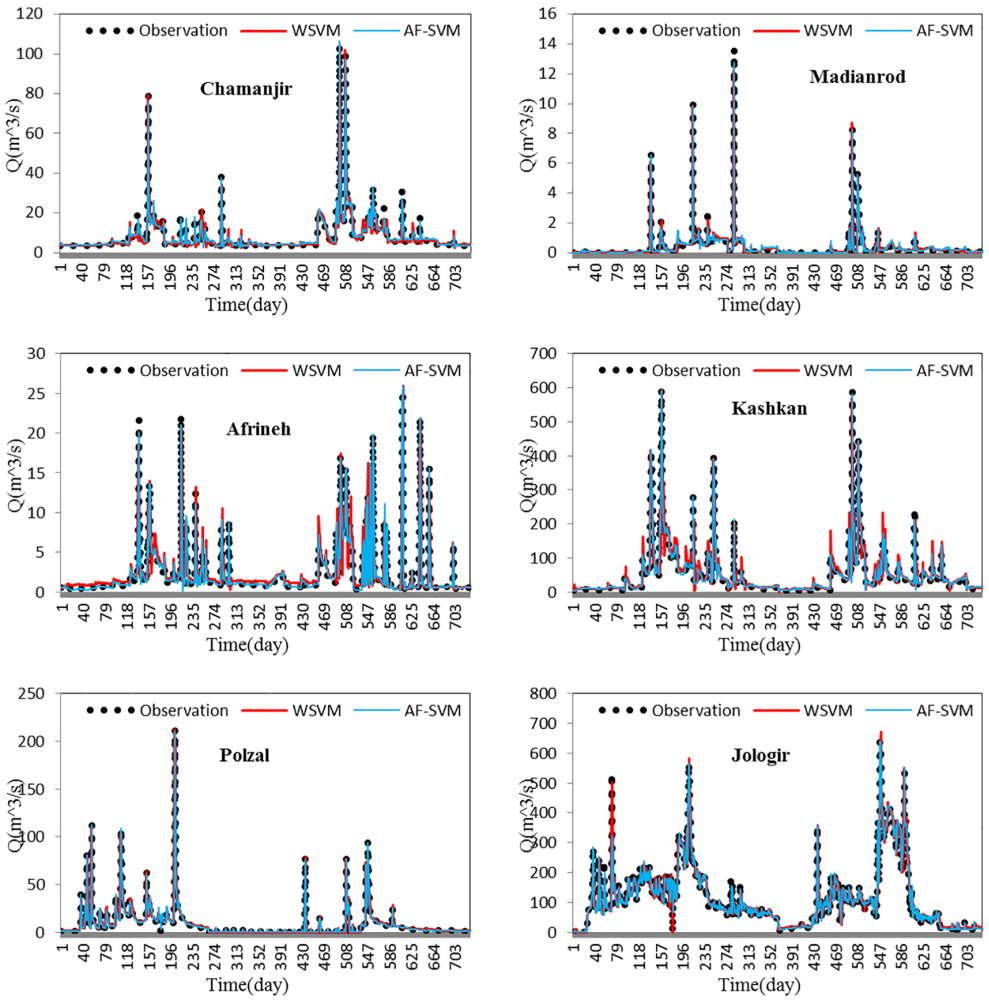


Figure 5. Computational and observational values of the studied models in the validation section.
AF indicates artificial flora; SVM, support vector machine; WSVM, wavelet-support vector machine.

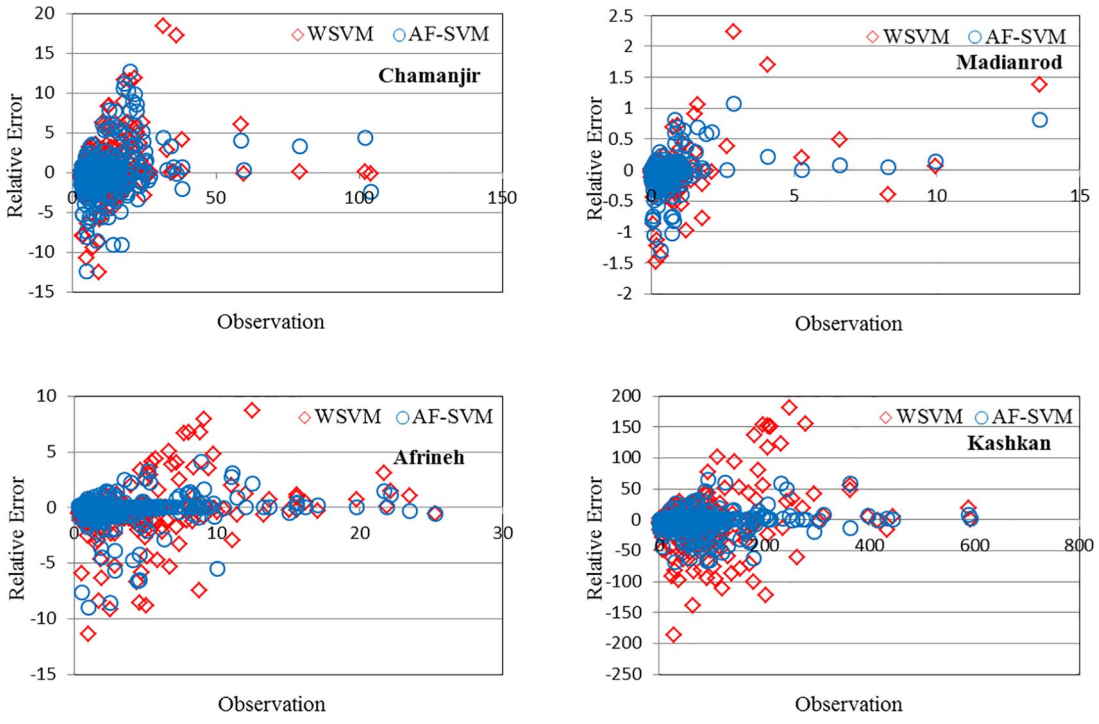


Figure 6. (continued)

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

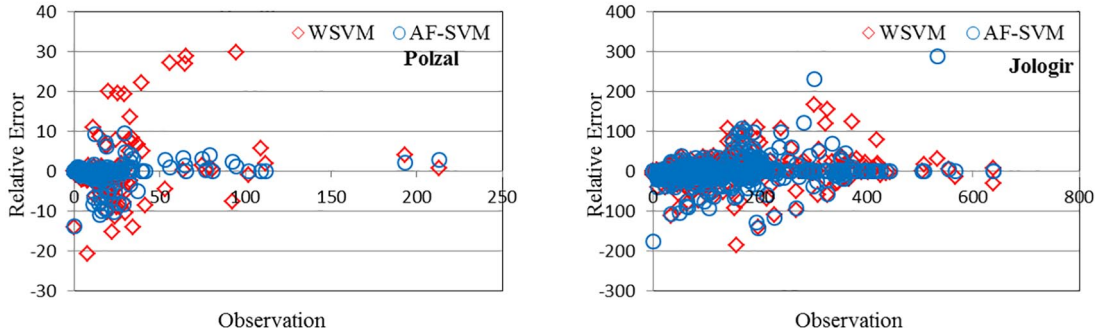


Figure 6. Diagram of relative error for the models studied in the validation section.
AF indicates artificial flora; SVM, support vector machine; WSVM, wavelet-support vector machine.

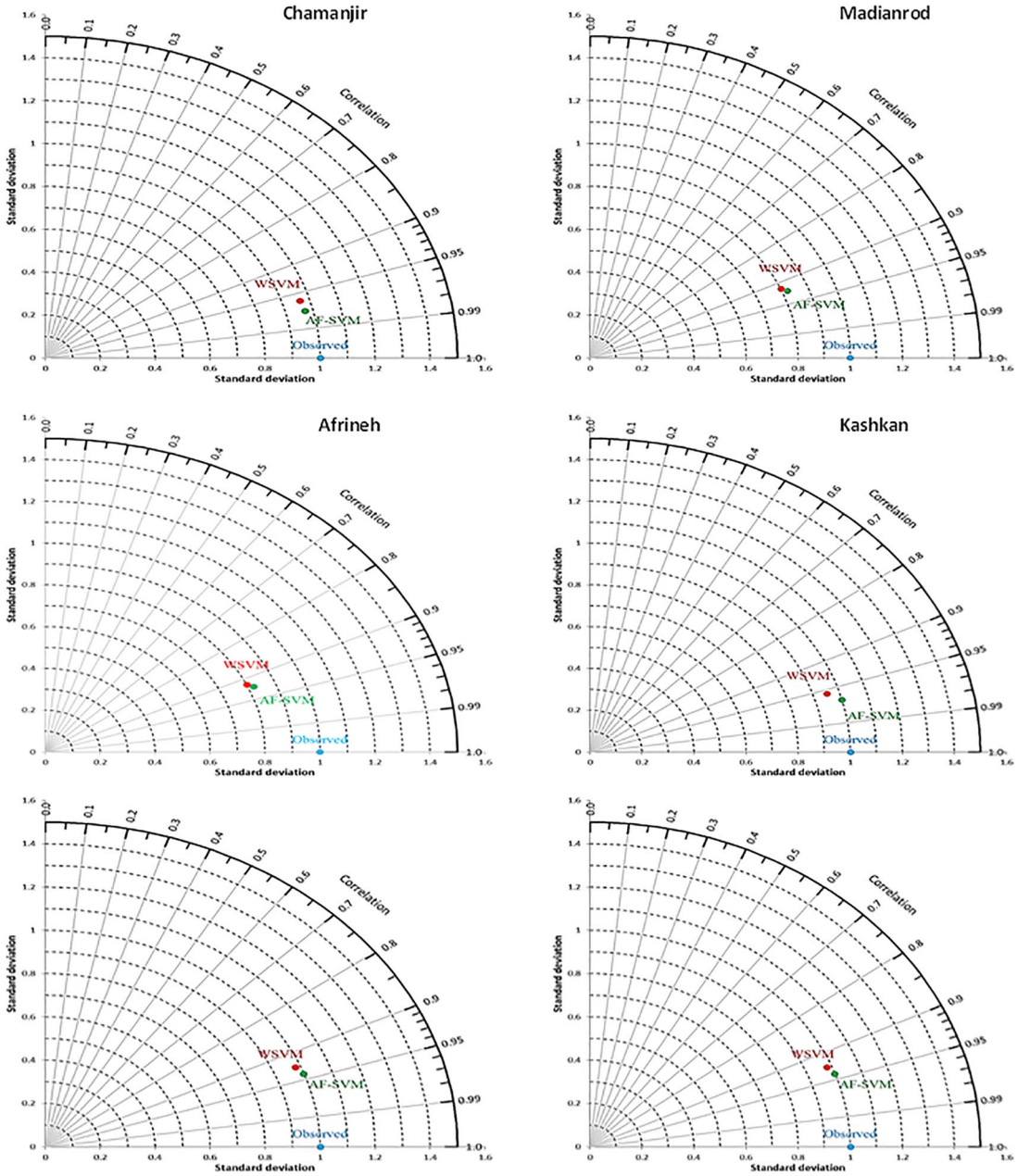


Figure 7. Taylor diagram of the studied stations.
AF indicates artificial flora; SVM, support vector machine; WSVM, wavelet-support vector machine.

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

catchment area using data from stations. The employed models were the SVM-AF hybrid model and the support vector wave machine. The observational values for the flow were compared with the predicted values using evaluation criteria. The research results can be summarized as follows: both models, namely the hybrid model of SVM-AF and support vector wave machine, achieved better results with structures consisting of 1 to 4 times delays than those with other structures. Also, according to the evaluation criteria, it was concluded that both models could predict the daily flow rates of the rivers with relatively high accuracy. Meanwhile, the proposed hybrid model of SVM-AF showed higher accuracy and lower error. Taylor's diagrams showed that the hybrid model was more accurate. In general, it can be stated that high accuracy of the hybrid model was due to the optimization of the parameters of the backing machine model by the AF algorithm with the best possible values, which could be due to the capability of the algorithm to find the optimal location and its increased convergence speed. Overall, this study supported the effectiveness of the combined model of SVM-AF in predicting the daily flow of rivers. Given that the decision to exploit water resources and implement management strategies for many uses (especially agriculture and industry) depends on the accurate estimation of river flow, the proposed hybrid model can be an appropriate tool for managerial decision making. It is recommended that one use hybrid models of SVM with new optimization algorithms such as creative gunner, ski, chicken crowding, and cat crowding and compare the results. Moreover, the proposed model in this study can be applied to other hydrological phenomena.

Author Contributions

All authors contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

REFERENCES

1. Zhang G, Patuwo BE, Hu YM. Forecasting with artificial neural networks: the state of the art. *Int J Forecasting*. 1998;14:35-62.
2. Kalteh AM. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Comput Geosci*. 2013;54:1-8.
3. Edossa DC, Babel MS. Forecasting hydrological droughts using artificial neural network modeling technique. Paper presented at: Proceedings of 16th SANCIAHS National Hydrology Symposium; October 1-3, 2012:1-10; Pretoria, South Africa.
4. Chen H, Zhu Y. Optimization based on symbiotic multi-species coevolution. *J Appl Math Comput*. 2008;22:179-194.
5. Huang S, Chang J, Huang Q, Chen Y. Monthly streamflow prediction using modified EMD-based support vector machine. *J Hydrol*. 2014;511:764-775.
6. Sedighi F, Vafakhah M, Javadi MR. Rainfall–runoff modeling using support vector machine in snow-affected watershed. *Arab J Sci Eng*. 2016;41:4065-4076.
7. Ghorbani MA, Khatibi R, Geol A, Fazelifard MH, Azani A. Modeling river discharge time series using support vector machine and artificial neural networks. *Environ Earth Sci*. 2016;75:675-685.
8. Samadianfard S, Jarhan S, Salwana E, Mosavi A, Shamshirband S, Akib S. Support vector regression integrated with fruit fly optimization algorithm for river flow forecasting in Lake Urmia Basin. *Water*. 2019;11:1934-1945.
9. Adnan R, Liang Z, Heddam S, Kermani M, Kisi O, Li B. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *J Hydrol*. 2019;19:432-448.
10. Rajaee T, Khani S, Ravansalar M. Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: a review. *Chemometr Intell Lab*. 2020;8:1324-1336.
11. Alizadeh F, Gharamaleki A, Jalilzadeh M, Akhoundzadeh A. Prediction of river stage-discharge process based on a conceptual model using EEMD-WT-LSSVM approach. *Water Resour.* 2020;47:41-53.
12. Hussain D, Ahmed Khan A. Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. *Earth Sci Inform*. 2020;13:939-949.
13. Vapnik VN. *Statistical Learning Theory*. New York, NY: Wiley; 1998.
14. Misra D, Oommen T, Agarwal A, Mishra SK, Thompson AM. Application and analysis of support vector machine based simulation for runoff and sediment yield. *Biosyst Eng*. 2009;103:527-535.
15. Hamel L. *Knowledge Discovery With Support Vector Machines*. Hoboken, NJ: *John* Wiley; 2009.
16. Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J Hydrol*. 2011;396:128-138.
17. Basak D, Pal S, Patranabis DC. Support vector regression. *Neural Inf Process*. 2007;11:203-225.
18. Vapnik V, Chervonenkis A. The necessary and sufficient conditions for consistency in the empirical risk minimization method. *Pattern Recognit Image Anal*. 1991;1:283-305.
19. Cheng L, Wu X, Wang Y. Artificial flora (AF) optimization algorithm. *Appl Sci*. 2018;8:329.
20. Rosin CD, Belew RK. Methods for competitive co-evolution: finding opponents worth beating. Paper presented at: Proceedings of the International Conference on Genetic Algorithms; July 15-19, 1995:373-381; Pittsburgh, PA.
21. Pagie L, Mitchell MA. Comparison of evolutionary and coevolutionary search. *Int J Comput Intell Appl*. 2002;2:53-69.
22. Wiegand RP, Sarma J. Spatial embedding and loss of gradient in cooperative coevolutionary algorithms. Paper presented at: Proceedings of the International Conference on Parallel Problem Solving from Nature; September 18-22, 2004:912-921; Birmingham, UK.
23. Hillis WD. Co-evolving parasites improve simulated evolution as an optimization procedure. *Physica D*. 1990;42:228-234.
24. Cartlidge J, Bullock S. Combating coevolutionary disengagement by reducing parasite virulence. *Evol Comput*. 2004;12:193-222.
25. Williams N, Mitchell M. Investigating the success of spatial coevolution. Paper presented at: *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*; June 2005:523-530; *Washington, DC*.
26. Vapnik VN. *The Nature of Statistical Learning Theory*. New York, NY: Springer; 1995.
27. Wang D, Safavi AA, Romagnoli JA. Wavelet-based adaptive robust M-estimator for non-linear system identification. *AIChE J*. 2000;46:1607-1615.
28. Shin S, Kyung D, Lee S, Taik Kim J, Hyun J. An application of support vector machines in bankruptcy prediction model. *Expert Syst Appl*. 2005;28:127-135.
29. Zhu YM, Lu XX, Zhou Y. Suspended sediment flux modeling with artificial neural network: an example of the Longchuanjiang River in the Upper Yangtze Catchment. *Geomorphology*. 2007;84:111-125.
30. Nagy H, Watanabe K, Hirano M. Prediction of sediment load concentration in rivers using artificial neural network model. *J Hydraul Eng*. 2002;128:558-559.
31. Kisi O, Karahan M, Sen Z. River suspended sediment modeling using fuzzy logic approach. *Hydrol Process*. 2006;20:4351-4362.
32. Liong SY, Sivapragasam C. Flood stage forecasting with support vector machines. *J Am Water Resour As*. 2002;38:173-186.
33. Lin JY, Cheng CT, Chau KW. Using support vector machines for long-term discharge prediction. *Hydrol Sci J*. 2006;51:599-612.
34. Khatibi R, Ghorbani MA, Naghshara S, Aydin H, Karimi V. A framework for 'Inclusive Multiple Modelling' with critical views on modelling practices – Applications to modelling water levels of Caspian Sea and Lakes Urmia and *Van*. *J Hydrol*. 2020;578:923-944
35. Ghorbani MA, Khatibi R, Singh P, et al. Continuous monitoring of suspended sediment concentrations using image analytics and deriving inherent correlations by machine learning. *Sci Rep*. 2020;537(2):859–870
36. Taylor E. Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res*. 2001;106:7183-7192.

Downloaded From: https://complete.bioone.org/journals/Air,-Soil-and-Water-Research on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use