Journal of Hydrosciences and Environment

Runoff simulation and prediction using Support Vector Regression (SVR) and SWAT hydrological model

M. Azhdari Moghaddam[1,*], M. Rashki[2], S. Sanayee[3]

1,. Professor of Civil Engineering Department, University of Sistan and Baluchestan. Zahedan, Iran*
2. Assistant Professor, University of Sistan and Baluchestan. Zahedan, Iran
3. MSc Student, Department of Civil Engineering, University of Sistan and Baluchestan

A R T I C L E I N F O

Article history:

Received: 2 Mar 2020
Accepted: 26 Oct 2020

Keywords:

SWAT hydrological model, Support Vector Regression (SVR), Time-series forecast, Kahir River.

A B S T R A C T

The application of dynamic and continuous time series has drawn attention due to the complexities of the rainfall-runoff process and the simplification of multiple regression and static methods. On the other hand, forecasting river flow is one of the main topics in flood control. This paper reports the results of applying the SWAT hydrological model to analyze the rainfall-runoff relationship for a 6-year period in the Kahir catchment basin, Sistan and Baluchistan province. The model output was calibrated by the SUF12 optimizer algorithm, and the data entered the model again. Then, the model output was used to forecast future periods using Support Vector Regression (SVR) and essential codes in MATLAB. The acceptable results of the SVR model regarding data prediction can be used as another method to estimate parameters and inputs.

1. Introduction

Regression methods are ordinary ways to assess the runoff-rainfall relationship. It is even used for complex rainfall-runoff processes. The first hypothesis in these methods is steady rainfall across the catchment. This hypothesis is almost true in small catchments (Fang et al,. 2008). Also, depending on the surface and soil conditions of the catchment basin, a part of rainfall known as direct runoff is different even if the rainfall volume is similar. This shows that rainfall transfer to runoff does not have a static state, and a dynamic process according to environmental conditions can give us a more realistic response from the catchment. Another point is that rainfall obviously outweighs runoff. Accordingly, these methods estimate the runoff coefficient, which is between 0 and 1, as a constant value, but this coefficient is not constant all the yearlong. Dynamic and continuous hydrological models can conquer multiple modeling weaknesses of the above-mentioned methods in terms of time like the Soil and Water Assessment Tool (SWAT), which can be connected to GIS software packages and can take a large number of data (Zekai et al. 2007).

On the other hand, flow forecast in rivers is considered one of the most important elements in water resource management. Forecasting is basically performed by regression models. These models fit some relationships between input and output data according to accessible information and such relationships can estimate proper output for these inputs. Single variable and multivariate regression equations and Auto-Regressive (AR) models are some of the simple models. With the advancement of science in the field of artificial intelligence and machine learning in recent years, other methods such as Artificial Neural Network (ANN) and Support Vector Machine (SVM) are widely used in linear and nonlinear regression questions (Nemati et al. 2014).

Burlando et al. (2012) used Autoregressive Moving Average Model (ARMA) to forecast hourly rainfall data. They compared the rainfall point and network data and achieved acceptable results. Runoff-rainfall forecast has been extensively discussed in different fields (Demirel et al. 2009). Maier and Dandy (2012) studied the literature concerning ANN applications and designing with ANN. Comoro Nick et al. (2013) compared and predicted the efficiency of time-series hydrologic models in the Czech Republic. They reported the high efficiency of the above-mentioned models to forecast hydrologic processes. Daml and Yalsyn (2005) studied the flood forecast with time-series applications in the Mississippi River. Their results indicated that time series can construct daily flow rates and verify predictions.

In another study in this regard, using the SWAT model, they concluded that winter months and inaccurate flow rate results were two effective factors in the studied catchment basin in South-Eastern Pennsylvania. However, the ANN model did not have this problem (Srivastava et al. 2006).

* Corresponding author's email: mazhdary@eng.usb.ac.ir

DOR: 20.1001.1.23455608.2020.4.8.2.3

Morid et al. (2010) used the ANN and SWAT models simultaneously. They found that ANN outperformed SWAT for basic flow rates in summer, fall, and winter. On the contrary, the SWAT model could perform better modeling in spring when rainfall and sudden peak flow rates were frequent.

Demirel et al. [4] compared the ANN and SWAT models to predict river flow. They concluded that the ANN model enjoyed better performance compared to the SWAT model in forecasting maximum flow rate [4]. Bashari and Saha Torouno (2007) used the least square in SVM to predict river flow in Malaysia. They concluded that this was an appropriate model for river flow forecast.

Bhagwat and Maity (2015) studied the potential to use the least square SVM to forecast river flow rate in a catchment basin in India. In Iran, Sahraee et al. [10] used SVM to forecast data measurement in the Kashkan catchment basin. The results showed the high accuracy of the model. Aghili et al. (2004) used the Artificial Neural Fuzzy Inference System (ANFIS) method according to C_means fuzzy clustering.

Bashari et al. (2007) used the Holt and Winters method and Autoregressive Moving Average Model (ARMA) advised in the Box-Vinjer method for 21 years in the Karkheh catchment basin. They concluded that higher rank methods enjoy better accuracy. Nemati et al. (2014) used SVR and ANN methods to estimate the Saeed Abad River flow rate in East Azerbaijan. They concluded that SVR enjoys better performance rather than ANN.

Seyed Ghasemi et al. (2014) used SDSM to simulate climatic parameters and then the SWAT model to simulate the Zayande Roud River for future periods under various climatic scenarios. According to frequent applications of the SWAT hydrological model and the time-series tool, this paper intends to estimate the flow rate in the Kahir catchment basin in the Kahir dam under construction using SVM.

2. STUDY SITE

Kahir desert is located in the southern part of Sistan and Baluchistan province, 80 km west of Chabahar in the north of Pozam Gulf. The geographical coordinates are 60°45' eastern longitude and 25°25' north latitude. Kahir village is the largest oasis in the region. It has a population of 900. It is located near the Kahir River. It is linked to Konarak with a 60-km asphalt road. Rainfall is low in the region, but short-term rainstorms are frequent and severe. The average rainfall is annually reported at 137 mm. The only and most important river, which enters Kahir, is the Nik Shahr River known as the Kahir River in the southern section. The area of the catchment basin is almost 4700 km^2. Fig. 1 shows the location of this catchment (Regional Water Company in Sistan and Baluchistan, 2010).

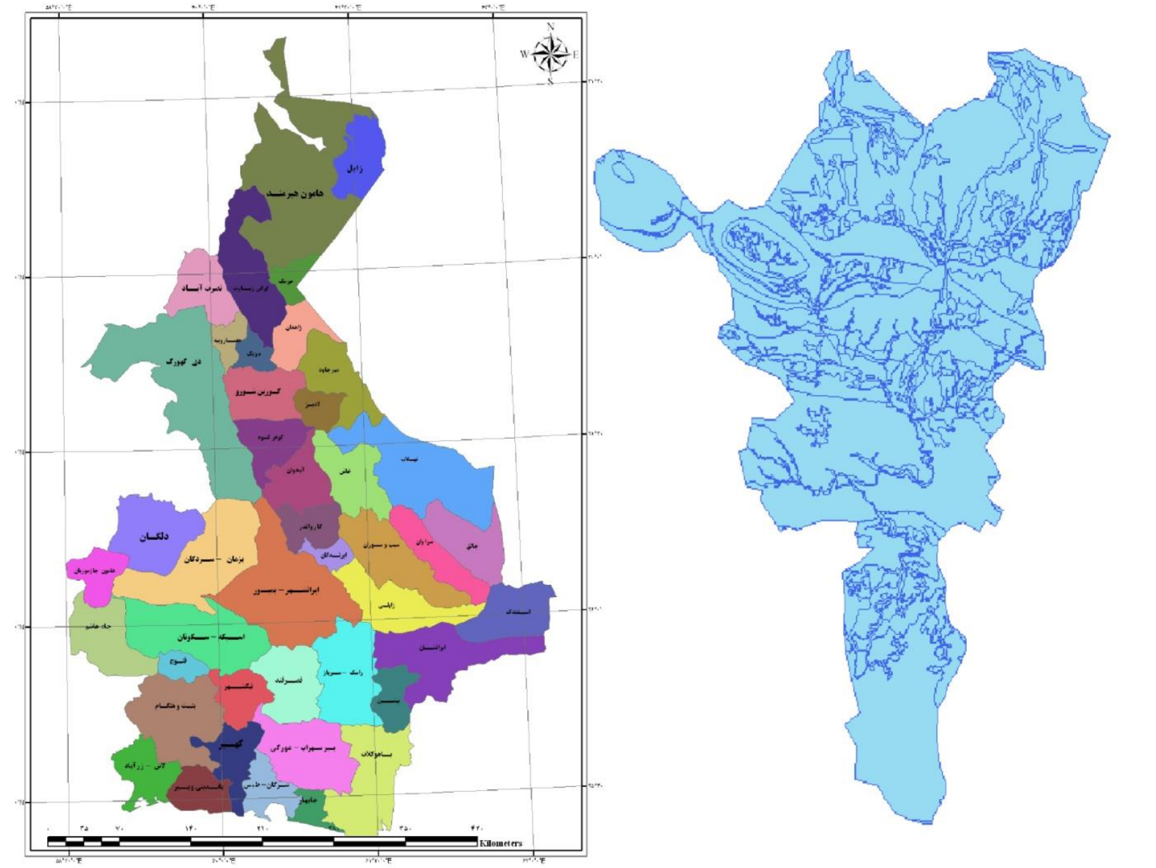


Fig 1. The location of the Kahir catchment basin

3. Materials and Methods

3.1. SWAT HYDROLOGICAL METHOD

The smallest working unit is the Hydrologic Response Unit (HRU) obtained through a mixture of slope maps, land use, and soil. The HRU formation process is shown in Fig. 2. The water balance equation used in the model is as follows:

$$S_{wt} = SW_{\circ} + \sum (R_{day} - Q_{surf} - E_a - W_{seep} - Q_{gw}) \qquad (1)$$

where S_{wt} is the final water content, t is time (day), S_{w0} is initial water in the soil, is rainfall per day, R_{day} is surface runoff per day, E_a is daily evaporation level, Q_{surf} is the amount of water that has penetrated to the cortical area in the soil profile, and Q_{gw} is the penetration level to underground aquifers. The basic maps required are the digital elevation model, land use maps, and soil maps. All three must be provided in the raster model. Other related information including meteorological data, water quality, effective factors on canal and surface water, underground water, water withdrawal, land management, information about tank water quality, and other fields need to be taken into account in the model. In this model, the border map of the sub-catchment basins and the stream network can be extractable from DEM maps in a software environment. However, it is optional. The mentioned maps are available in advance considering the fact that the objective of this paper is river flow simulation (Saneyee Moghadam et al. 2015).

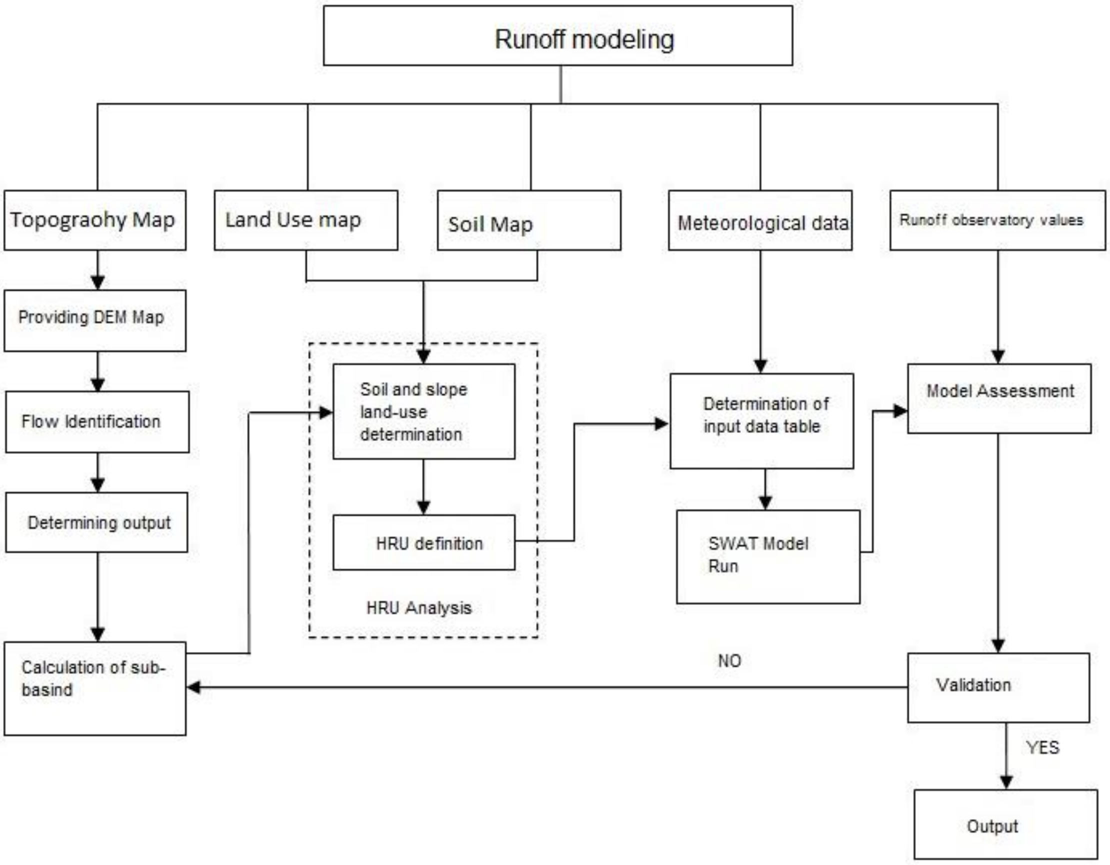


Fig 2. The runoff calculation flow chart with the help of hydrologic SWAT

3.2 SUPPORT VECTOR REGRESSION (SVR)

Support Vector Regression (SVR) was introduced by Vapnik in 1995. SVR minimizes operational risk. It means that an equation is selected for estimation among all estimator equations with equal error so that the error is minimized in case of entering new data. In SVR, regression is performed by estimation of an interval in which we need to find the support vector in order to find this interval. These support vectors are obtained by solving quadratic programming as follows. In SVR, the objective is to minimize the following function by the mentioned constraints.

$$Minimize\ \frac{1}{2}\|w\|^2 \quad (2)$$

$$subject\ \ to \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon \\ -y_i + \langle w, x_i \rangle + b \le \varepsilon \end{cases} \quad (3)$$

where x_i is the training data, w is the weight vector, y_(i) is the objective data, and b is the bias. The inner product of ⟨w,x_i ⟩ +b is an estimation for predictor data and ε is an acceptable error in the forecast model, which plays the threshold role. The kernel function used in SVR can be various functions; however, the most common one is the radial bias function kernel function. This function is used

because of its appropriate efficiency (Nemati et al. 2014). Figure 3 shows that a non-linear function with accepted ε error is transferred by a kernel function into a linear function with an acceptable ε range.

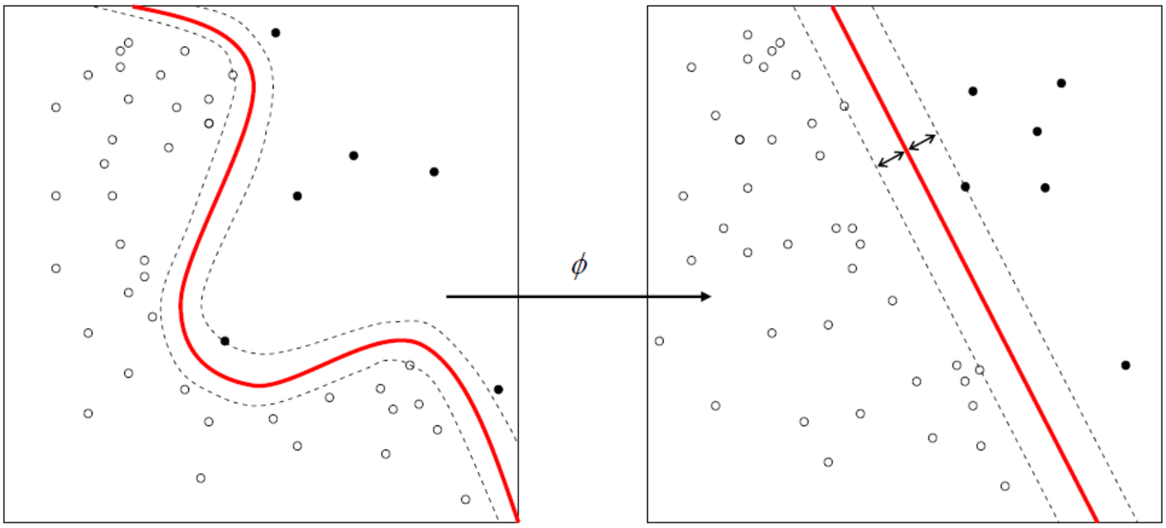


Fig 3. The kernel function and error interval

3.3. PREDICTION MODEL

After entering the 6-year period information, we reached the result analysis model. After the model results are obtained, the model needs to be calibrated. Figure 4 shows SWAT modeling of the upstream catchment basin of the Kahir dam. Figure 2 shows the sequence by which the information is entered into the model given that the final model is obtained by the superposition of basins, sub-basin, rivers, soil layers, and land use.

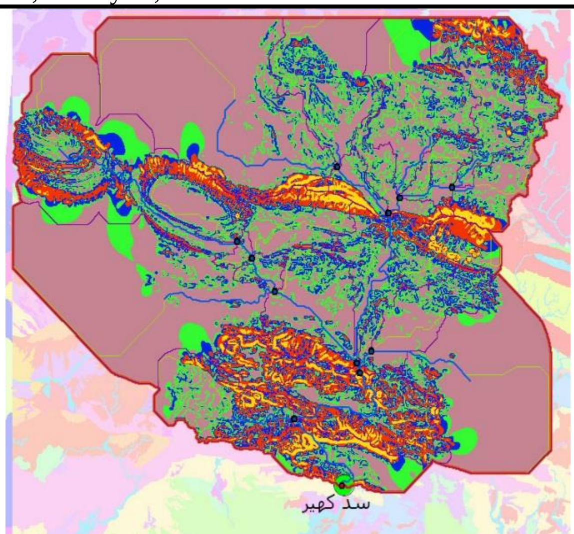
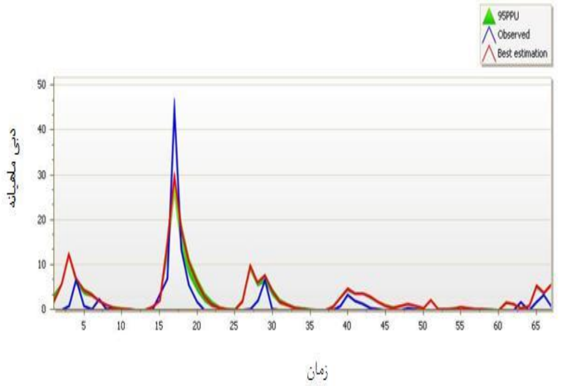


Fig 4. SWAT modeling of the Kahir dam upstream

The SWAT model can be calibrated in two ways: manual calibration and auto-calibration. This paper has adopted auto-calibration and the SUF12 algorithm. Calibration was initially performed on a monthly basis and then it was performed on a daily basis. Here, sensitivity analysis was performed and important parameters in the basin were separated from the less important ones. Multiple parameters, which are mainly related to soil properties, are known as important factors in a sensitivity analysis. To calibrate a model, the optimum values of algorithms are used frequently. In each implementation, in the case of the acceptability of the optimization results, the optimum values of the parameters are used in the calibration stage. In case of the unacceptability of the results, optimization is re-implemented. In this paper, SUF12 was run more than 10 times, and the model was calibrated.



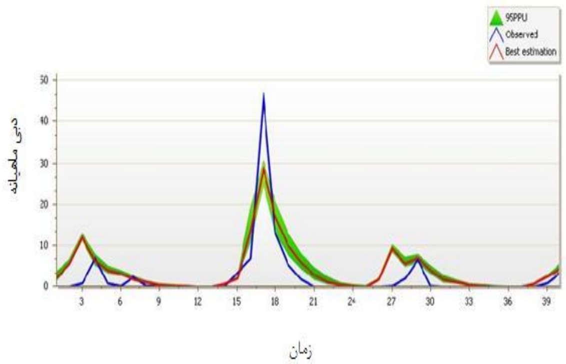


Fig 5. Calibration and verification of hydrograph observations and simulations

The efficiency assessment process is not only important during the model development and calibration process but also when reporting the results to other researchers. Various indicators are introduced and used to this end. In this paper, the results of the simulations were evaluated in the model calibration stage and verified with some indicators including R2 (correlation coefficient), BR2 (weight correlation coefficient), and NS (Nash Sutcliffe). The results were estimated by SWAT-Cup software package and SUF12 algorithm. The errors are listed after 10 times repetitions.

Table 1. The results of the model efficiency evaluation

	R^2	NS	BR^2	MSE	SSQR
Simulated flow	0.72	0.68	0.4651	18.6703	14.2997

According to Zu et al. (2009), if NS is greater than 0.75, the model is perfect, if it is between 0.36 and 0.75, it is considered satisfactory, and if it is less than 0.36, it is unacceptable. Here, the NS of the model is reported satisfactory with a value of 0.68, showing the high capability of the model to forecast runoff-rainfall.

After the SWAT model calibration and data re-entry into the software, the model was re-implemented. We finally reached an optimized runoff-rainfall result. Data in this study include daily 6-year data from 2006 to 2013. Also, Kahir flow rate station data were measured in the Kahir dam in the mentioned interval, and they were considered the output values. In this paper, model performance was evaluated by studying the difference between observed and forecasted values of dependent variables or studying the errors. The error validation method used in this paper is R2. The correlation between forecasted and real values ranges from 1 for complete correlation (complete equality of observed and forecasted values) to 0 showing the lack of correlation. The closer the values to 1, the better the performance of the model. Another indicator is Root Mean Square Error (RMSE). Lower values of this indicator show less model error and proper performance. Mean Average Error (MAE) is the absolute mean of observed and predicted error. A lower value of this indicator shows the better performance of the model. These indicators are listed as follows. y_t is the real values of dependent values, $\hat{y}_t$ is the forecasted values, and n is the number of variables.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}} \quad (4)$$

$$MAE = \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)}{n} \quad (5)$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y}_t)^2} \quad (6)$$

Various kernel functions have been used. Radial Basis Functions (RBF) give the best result compared to other kernel functions. After selecting RBF, calibration was performed on the parameter (8) and SVR minimization was considered as criteria to select calibration parameters. The parameter calibration needs to be performed by test data because training data give better prediction of targeted training data performed by increased and reduced (8). On the other hand, fitness conditions happen and the model cannot forecast flow rate values well for test data (Kian Pisheh et al. 2010)

$$K(x,z) = \exp(-\frac{\|x - z\|^2}{2\delta^2} \quad (7)$$

$$\frac{1}{2\delta} = \gamma \quad (8)$$

A total number of 6 years and 2190 data were studied. As much as 75 percent were used for training and the rest were used for model calibration. While we are working with monthly data in regression studies, the support vector does not act successfully because the interval of the month is relatively large for rivers so that the SVM is not able to learn information well in such intervals; however, correlation is well maintained in SVM with the daily interval. That is why output daily flow rate data of the SWAT model were used. The results are presented in Table 2 for 1-6 forecast intervals.

Forecasted and calculated results by SVR are shown for the period of one day in Figure 6. The data observed in the Kahir flow rate station and the data calculated by SVR diagrams are shown. An important factor in the diagram is that a great volume of zero is associated with drought and the seasonal river. On the other hand, noticeable flow rate peaks were observed as well. The simulation of such behaviors seems extremely difficult for most prediction algorithms because it is difficult to discover the logical relationship among data.

Table 2. SVR performance to forecast the Kahir River flow rate

Day	RMSE	MAE	R^2
1	0.075152	0.0135	0.76049
2	0.099859	0.028449	0.57822
3	0.10983	0.030913	0.49122
4	0.11665	0.033883	.46406
5	0.13423	0.03632	0.44399
6	0.14135	0.041918	0.427640

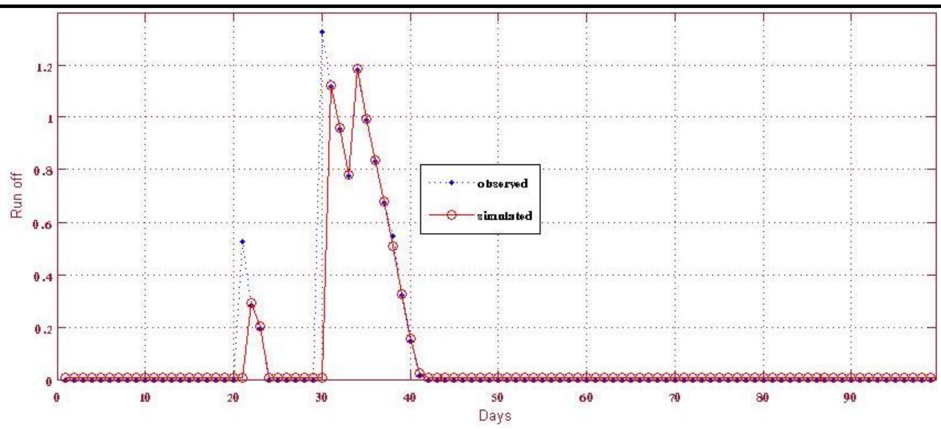


Fig 6. The one-day time-series prediction by SVR

4. Conclusion

We made the hydrologic model for the Kahir catchment basin by SWAT software. Then, the results were calibrated by the SUF12 optimizer algorithm and SWAT-CUP software. Reaching to NS coefficient of 0.68 proves that SWAT provides satisfactory modeling. The results were entered into the model again, and the optimum results were extracted. We forecasted the flow rate for a period of 1 to 6 coming days with SVR. The results were compared with real ones, indicating a good estimation of SVR. The results are listed in Table 2. The Kahir River is a seasonal one with a flow rate of zero for most of the year and it has sudden flow rates and severe rainfalls since it is located near Chabahar and Mansoon flows. However, the algorithm is well able to forecast. A considerable number of zeros for most algorithms such as neural networks make this issue difficult to find an appropriate pattern among data. It is noteworthy that it is better for data to be entered on a daily basis in order to make SVR capable of providing an appropriate estimation of data. Data obtained by the model can be studied for rainfall as well. These data were entered into the SWAT model again in order to use for future predictions.

References

1. Fang, X., Thompson, D., Cleveland, T., Pradhan, P., Time of concentration estimated using watershed parameters determined by automated and manual methods, Journal of Irrigation and drainage Engineering 134 (2), 202-211., 2008.

2.Zekai ¸Sen., and Essam Wagdani. Aquifer heterogeneity determination through the slope method., Article first published online: 1 OCT 2007.DOI: 10.1002/hyp.6737., 2007.

3. Nemati, Amir Reza, Mahmoud Zakeri Nayeri, and Saber Moazami Goudarzi, 2014, Runoff simulation and prediction using SVR and ANN, 1st national conference of development in civil, architecture, electric and mechanical engineering, Omran Bana

Tadbir engineering Company and Golestan University, Golestan University, Goelstan.

4. Demirel, M. C., Venancio, A., and Kahya, E. (2009). Flow forecast by SWAT model and ANN in Pracana basin, Portugal. Advances in Engineering Software, 40(7), 467–473.
5. Maier, H. R., and Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. Environmental Modelling and Software, 15(1), 101–124.
6. Srivastava P, McNair JN, Johnson TE. Comparison of process-based and artificial neural network approaches for stream flow modelling in an agricultural watershed. J Am Water Resource Association 2006;42(3):545–63.
7. Morid S, Gosain AK, Keshari AK. Comparison of the SWAT model and ANN for daily simulation of runoff in snowbound ungauged catchments. In: Fifth International conference on hydro informatics, Cardiff, UK; 2002.
8. Kalteh, Aman Mohammad, 2014, Using Support Vector Regression to predict river flow rate in Shafa roud River in Gilan province, 8th national conference of Civil engineering, Noushiravani industrial university. Babol.
9. Bhagwat, P.P. and Maity, R. (2012). "Multistep-Ahead River Flow Prediction Using LS-SVR at Daily Scale" Journal of Water Resources and Protection, 4, pp528-539.
10. Sahraee, Hamideh, Behnaz, Ali Ghanbari, 2013, Hec-hms model evaluation to estimate flood in Khoshk River catchment basin, 5th Water management resources in Iran, Water resources engineering and science association, Shahid Beheshti University.
11. Kaheh , Mahdi; Aghili, Roya; and Moazed, Hadi, 2009, inflow forecast to Shahid Abbaspour reservoir using adaptive neuro-fuzzy inference system (ANFIS) and based on c-mean clustering, 8th international conference of river engineering, Shahid Chamran University, Ahvaz.
12. Mahdi, B., Vafakhah, Mahdi, 2010, Comparison of Time-series analysis methods to predict monthly flow rate in Karkheh catchment basin, Seasonal research and science Journal of water and irrigation engineering.
13. Seyed Ghasemi, Samaneh, Ahmad Abrisham Chi, and Masoud Tajrishi, 2006, Evaluation of flow changes in Zayande Roud River as a result of climate change, 2nd conference of water resource management, Industrial University of Isfahan, Iran Water resource engineering and science Association.
14. Regional Water Company in Sistan and Baluchistan, 2nd studies of Kahir dam, Hydraulic Structure report, Pajouhab Consultant Committee, December, 2010.
15. Saneyee Moghadam, Soroosh, Risk analysis in overflows using fuzzy reliability, M.S. thesis, Hydraulic structures, Sistan and Baluchistan University, Zahedan, January 2013.
16. Support vector machine. https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Kernel_Machine.png.
17. Kian Pisheh, Ghasem; Ahmadi, Azadeh; and Moridi, Ali; 2010; Rainfall prediction model using Support Vector Machine. 1st conference of applied researches of water resources in Iran, industrial university of Kermanshah.