International Journal of Geography and Geology
2016 Vol. 5, No. 6, pp. 113-118
ISSN(e): 2305-7041
ISSN(p): 2306-9872
DOI: 10.18488/journal.10/2016.5.6/10.6.113.118

RUNOFF PREDICTION BY SUPPORT VECTOR MACHINE FOR CHALOUS RIVER BASIN OF IRAN

P. Mirzaee[1]† --- **R. Fazloula**[2]
[1,2]*Water Engineering Department, Sari Agricultural Sciences and Natural Resources University, Sari, Iran*

ABSTRACT

Runoff is the result from the comprehensive action of climate conditions and drainage area underlying surface. Rainfall, evaporation, temperature, wind speed, solar radiation and relative humidity are the most important factor which effect on runoff. Prediction of runoff amounts is performed using Support Vector Machine (SVM). In this paper, the prediction of runoff for Chalous River basin along the Caspian Sea is investigated. A model based on SVM approach is proposed to runoff, predicated on a total of 8 years daily data sets, including field investigation records for the Chalous River Basin along the southern shoreline of Caspian Sea. This study addresses the question of whether Support Vector Machine (SVM) approach could be used to predict runoff. Results revealed that SVM provides an effective means of efficiently recognizing accurately predicting the runoff and the prediction of the future runoff evolution trend with this model will provide the basis for water regulation and water resources reasonable configuration.

Keywords: Runoff, Chalous River, Support vector machine, Validation.

Received: 7 September 2015/ **Revised:** 2 March 2016/ **Accepted:** 27 March 2016/ **Published:** 4 April 2016

Contribution/ Originality

This study uses new estimation methodology as well as support vector machine (SVM) to predict runoff amount based on hydrological condition in Chalous river basin from north of Iran. This study addresses the capability of SVM in runoff prediction.

1. INTRODUCTION

Runoff simulation and prediction in watersheds is a requirement for many practical applications involving conservation, environmental disposal and water resources management. Yet, the rainfall-runoff process is a complex, non-linear and dynamic hydrological phenomenon to simulate due to the spatial–temporal variability and interrelationships of underlying climatic and physiographic variables (Zhang and Govindaraju, 2003). The development of rainfall-runoff models has undergone substantial changes since Sherman pioneered the unit hydrograph theory in 1932 (Liong *et al.*, 2002). Based on the description of the governing processes, rainfall-runoff models can be classified as either physically based (knowledge-driven) or system theoretic (data-driven). Physically based models involve a detailed interaction of various physical processes controlling the hydrologic behavior of a system. However, system theoretic models are instead based primarily on observations (measured data) and seek to characterize the system response from those data using transfer functions (Wu and Chau, 2011). In recent years, data-driven modeling approaches are being widely used as surrogate for physically based models, as they overcome some limitations associated with physically based approaches. As an example of data-driven models, Support Vector

† Corresponding author

Machine (SVM) techniques are advocated as an appropriate and sensible method for the combination of simulated river flows of a suite of rainfall-runoff models.

Support vector machines (SVMs), which are based on the statistical learning theory and were introduced by Vapnik (1995) are a relatively new class of models in the data-driven prediction field. Although SVMs have remarkable successes in various fields, there are few studies on their applications in water resources and hydrology (Behzad *et al.*, 2009). Liong and Sivapragasam (2002) demonstrated that SVM models show good generalization performance in their applications on flood forecasting and rainfall-runoff modeling. Bray and Han (2004) described an exploration in using SVM models in flood forecasting. A SVM model was used to forecast flows at different time scales: seasonal flow volumes, hourly stream flows and long-term discharges (Asefa *et al.*, 2006; Lin *et al.*, 2006).

This treatment aims to develop a SVM for the prediction of runoff for Chaloos River basin in north of Iran. Following the aims of the study, first reviews of SVM methodology, then a brief explanation of the case histories under consideration, and the phenomena of modeling with SVM are presented. Finally the developed SVM model is described and compared with data measurements.

2. PRINCIPLES OF MODELING USING SVM

The SVM has recently emerged as an elegant pattern recognition tool and a better alternative to Artificial Neural Network (ANN) methods. The method has been developed by Vapnik (1998) and is gaining popularity due to many attractive features. The formulation is based on Structural Risk Minimization (SRM) which has been shown to be superior to the Empirical Risk Minimization (ERM) used in conventional neural networks (Vapnik, 1998). This section of the paper serves an introduction to relatively new technique.

SVM is primarily a classier method that performs classification tasks by constructing hyper-planes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables. To construct an optimal hyper-plane, SVM employees an iterative training algorithm, this is used to minimize an error function. According to the form of the error function, SVM models can be classified into classification and regression groups.

For Classification and the regression type of SVM, training involves the minimization of the error function:

$$\frac{1}{2}\omega^T\omega + C\sum_{i=1}^{N}\xi_i \qquad (1)$$

Subject to the constraints:

$$y_i(\omega^T\varphi(x_i) + b) \geq 1 - \xi_i \; and \; \xi_i \geq 0, i = 1, \ldots, N \qquad (2)$$

Where C is the capacity constant, ω is the vector of coefficients, b a constant and $\xi_i$are parameters for handling non-separable data (inputs). The index i labels the N training cases. Note that $y\epsilon \pm 1$ is the class labels and x_i is the independent variables.

The kernel φ is used to transform data from the input (independent) to the feature space. The larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting.

SVM supports a number of kernels for use in Support Vector Machines models. These include polynomial and radial basis function (RBF) as popular:

Linear: $K(x_i, x_j) = {x_i}^T x_j,$ (3)

Polynomial: $K(x_i, x_j) = (\gamma {x_i}^T x_j + r)^d, \; \gamma > 0,$ (4)

Radial basis function (RBF): $K(x_i, x_j) = e^{-\gamma(x_i - x_j)^2}, \; \gamma > 0,$ (5)

Where γ, *r* and *d* are parameters of the kernel functions and are entered manually. In the following we will only focus on the RBF kernel. Thus, two parameters C and γ must be determined. Unfortunately, It is not known beforehand which C and γ are best for a given problem.

3. OVERVIEW OF CHALOUS RIVER

The Chalous is the longest river in north of Iran. The length of the Chalous River is 5,464 km with a drainage area of 752,440 km², it passes through Mazandaran province and autonomous regions and reaches the Caspian Sea in north of Iran. Chalous basin which is located between 36° 08′-36° 36′N latitude and 50° 58′-51° 40′E longitude was selected for this study. Chalous river is part of great chalous catchment and there is one of the seven sub-basins of the Caspian Sea in Iran

The studied basin area is approximately 1631.429 km² and perimeter is 210.66 km. Geographical location of Chalous River basin is shown in figure 1.

Fig-1. Location map of the study area.

Topography of the areacan generally be divided in to three areas, hills (50 m-500m), mountains (500 m-1000 m), plateaus (less than 50 m and often less than 25 m). Generally Chalous river basin due to high levels of relative height difference of land units classified in mountain region.

The amount of elevation difference between the highest and lowest point is more than 4000m in whole of the catchment. In view of the landuse types of the entire catchment, it can be generally divided in to agricultural areas, urbanized areas, natural forest and pastureland. Annual average rainfall in the catchment is 836mm and has an average discharge 13.8 m3/s. in this paper the 8 daily years runoff database collected to model with SVM as seen in figure 2.

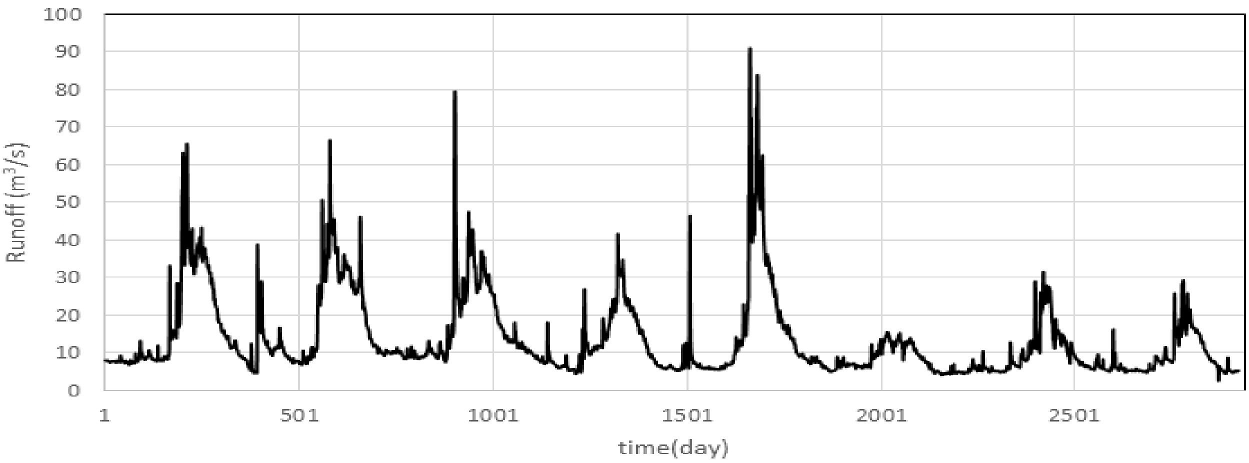

Fig-2. Database used in this study

4. MODELING RUNOFF PREDICTION USING SVM

By means of a SVM fitting, a model can be represented as time series. In order to develop the evolved SVM, the database is divided into two different sets, namely, training and testing. The training set consists of 75 percent of data pairs. The testing set, which consists of 25 percent of data unforeseen during the training process, is merely used for testing the trained SVM models. It should be noted that the training and testing sets are randomly selected from the data sets with approximately the same statistical properties. In order to illustrate the model's predictive performance in comparison with observed data, training set is shown in Fig. 3, predicted and measured values are properly close.

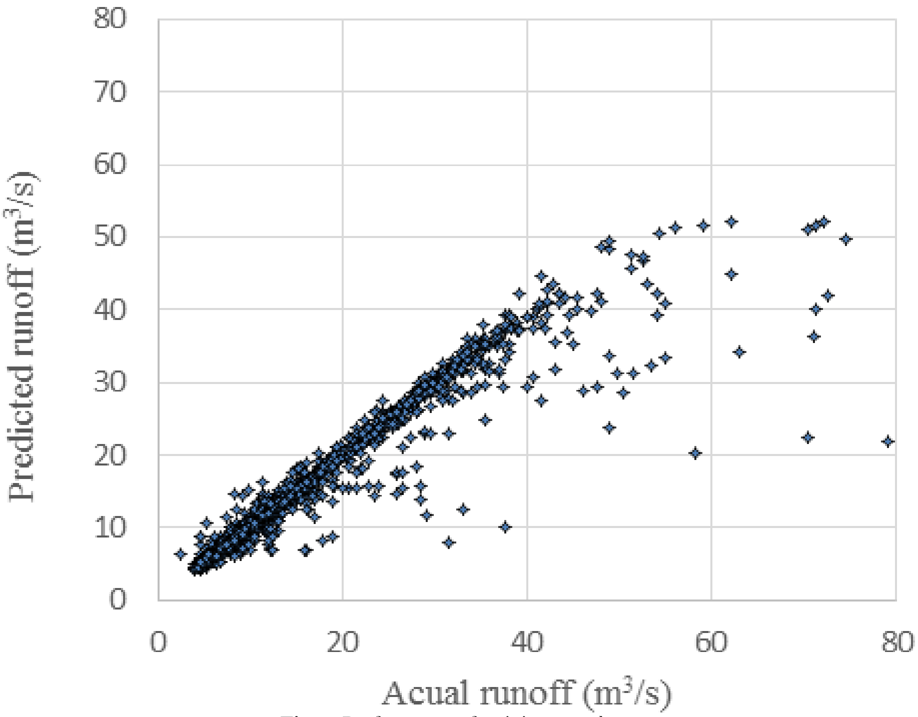

Fig-3. Performance of training test data

As presented in Table 1, the statistically assessed accuracy of the model is determined by R^2 (absolute fraction of variance), RMSE (root mean squared error), MSE (mean squared error), and MAD (mean absolute deviation) which are defined as follow:

$$R^2 = 1 - \left[\frac{\sum_{i=0}^{M}(Y_{i(Model)} - Y_{i(Actual)})^2}{\sum_{i=1}^{M}(Y_{i(Actual)})^2}\right] \quad (6)$$

$$RMSE = \left[\frac{\sum_{i=0}^{M}(Y_{i(Model)} - Y_{i(Actual)})^2}{M}\right]^{1/2} \quad (7)$$

$$MSE = \frac{\sum_{i=0}^{M}(Y_{i(Model)} - Y_{i(Actual)})^2}{M} \quad (8)$$

$$MAD = \frac{\sum_{i=1}^{M}|Y_{i(Model)} - Y_{i(Actual)}|}{M} \quad (9)$$

Table-1. Statistical information for the SVM model for predicting runoff

Statistic	R^2	MSE	MAD	RMSE
Neural training	0.95	11.79	1.12	3.41
Neural testing	0.96	11.98	1.21	3.43

The ability of the SVM model in predicting unforeseen data is tested for the testing dataset. As it is illustrated in figure 4results from the model agree well with measured values.

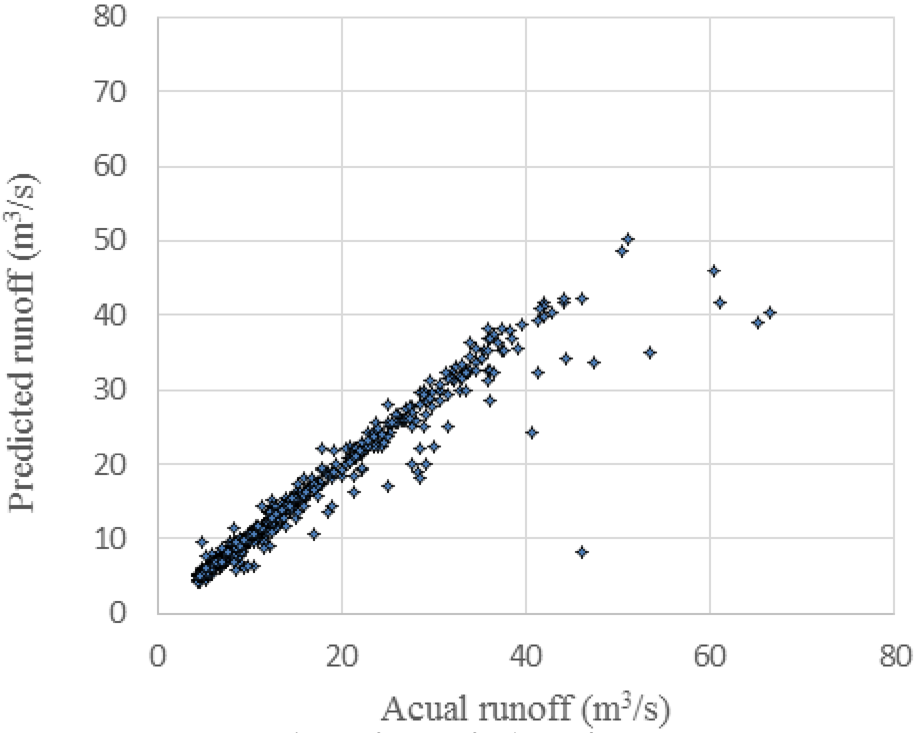

Fig-4. Performance of testing test data

5. VALIDATION

In this part all of the 8 years daily data with the predictions has been shown in figure 5. According to this paper methods, it can be supposed the best fit for mention basin.

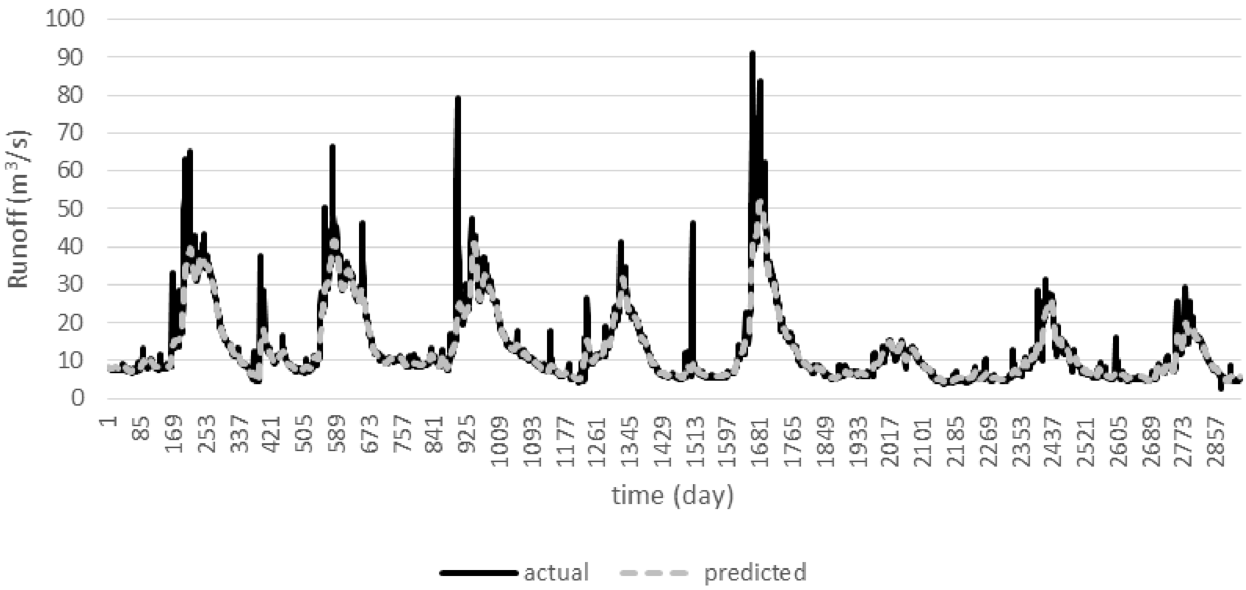

Fig-5. Performance of this study approach for all of the database

6. CONCLUSIONS

In this study, it has been attempted to deploy a system identification technique to develop the *runoff*. The evolved Support Vector Machine (SVM) have been used to obtain a model for the prediction of runoff. The results shows the best correlation.

Results obtained from this study and previous researches reveal that SVM model derived from a local dataset should not be implemented for different sites with significantly varying features. Therefore, these proposed relationships should be used with caution in hydrology engineering and should be checked against measured.

Funding: This study received no specific financial support.

Competing Interests: The authors declare that they have no competing interests.

Contributors/Acknowledgement: All authors contributed equally to the conception and design of the study.

REFERENCES

Asefa, T., M. Kemblowski, M. McKee and A. Khalil, 2006. Multitime scale stream flow predictions: The support vector machines approach. Journal of Hydrology, 318(1): 7–16.

Behzad, M., K. Asghari, M. Eazi and M. Palhang, 2009. Generalization performance of support vector machines and neural networks in runoff modeling. Expert Systems with Applications, 36(4): 7624–7629.

Bray, M. and D. Han, 2004. Identification of support vector machines for runoff modeling. Journal of Hydroinformatics, 6(4): 265–280.

Lin, J.Y., C.T. Cheng and K.W. Chau, 2006. Using support vector machines for long-term discharge prediction. Hydrologica Sciences Journal, 51(4): 599–612.

Liong, S.Y., R.R. Gautam and S.T. Khu, 2002. Genetic programming:A new paradigm in rainfall runoff modeling. Journal of the American Water Resources Association, 38(3): 705–718.

Liong, S.Y. and C. Sivapragasam, 2002. Flood stage forecasting with support vector machines. Journal of the American Water Resources Association, 38(1): 173–186.

Vapnik, V., 1995. The nature of statistical learning theory. New York: Springer.

Vapnik, V.N., 1998. Statistical learning theory. Adaptive and learning systems for signal processing, communications, and control. New York: Wiley.

Wu, C.L. and K.W. Chau, 2011. Rainfall-runoff modeling using artificial neural network coupled with singular spectrum analysis. Journal of Hydrology, 399(3-4): 394–409.

Zhang, B. and R.S. Govindaraju, 2003. Geomorphology-based artificial neural networks (GANNs) for estimation of direct runoff over watersheds. Journal of Hydrology, 273(1): 18–34.