Scientia Iranica A (2016) 23(3), 856–863

Sharif University of Technology
**Scientia Iranica**
*Transactions A: Civil Engineering*
www.scientiairanica.com

# Improving the accuracy of K-nearest neighbour method in long-lead hydrological forecasting

**M. Azmi[a,*] and F. Sarmadi[b]**

a. *Department of Civil Engineering, Faculty of Engineering, Clayton Campus, Monash University, VIC, 3008, Australia.*
b. *School of Earth, Atmosphere, and Environment, Faculty of Science, Clayton Campus, Monash University, VIC, 3008, Australia.*

Received 5 December 2014; received in revised form 1 August 2015; accepted 15 September 2015

**KEYWORDS**
K-NN;
Hydrological forecasting;
HBMO;
Zayandeh-rud;
Iran.

**Abstract.** The nonparametric regression method of K-Nearest Neighbour (K-NN) has been used in a variety of eco-hydrological issues. In this study, some techniques were presented to improve the accuracy of the K-NN method in forecasting accumulated 9-month inflow, from 1971 to 2001, of Zayandeh-rud dam in Iran, from winter to the end of the following summer. The considered improving techniques consisted of: 1) selecting the best data preprocessing functions, 2) selecting the best number of neighbours, 3) selecting the best distance functions, 4) specifying the best weights of predictors at distance functions, and 5) adding the ability of extrapolation to K-NN using a proposed method. Final results showed that the use of the mentioned techniques had promoted the accuracy of K-NN's forecast, meaningfully. The results of goodness-of-fit criteria for the optimized K-NN in comparison with a regular K-NN presented an increase by 31% in correlation coefficient (from 65% to 96%), a decrease from 31% to 8% in volume error, and finally a drop from 54% to 25% in the root mean square error.



## 1. Introduction

Data-driven methods are powerful means to be used in different issues such as pattern recognition, categorizing, estimating, and forecasting. These methods can be divided into two main groups of parametric and nonparametric models, in which K-Nearest Neighbour (K-NN) method is one of the most acknowledged nonparametric ones [1]. In the past decades, K-NN has been applied in a variety of areas such as density estimation [1], rainfall-runoff forecasting [2], resampling hydrologic time series [3], generating regional climate scenarios [4], short-term traffic flow prediction [5], wind power forecasting [6], short-term rainfall predictions [7], probabilistic streamflow forecasts [8], modelling hydrological time series [9], short-term foreign exchange forecast [10], and long-term rainfall probabilistic predictions [11].

Regarding long-lead hydrological forecast by K-NN, Araghinejad and Burn [12] showed that the use of a combination of K-NN and geostatistical methods in hydrological forecasting can lead to more reliable results in long-term management of water resources. Further, Asadiani Yekta [13] presented that a combination of K-NN algorithm and ANFIS model could be successfully applied in estimating inflow suspended load to dams. Later, Azmi et al. [14] proposed a K-NN based data fusion method for short-term and long-term hydrological forecasting.

Generally, some advantages of K-NN method in forecasting and estimating issues are simplicity, no calibration stage, modelling the nonlinear processes, and ability to cope with numerous predictors [9,15]. However, there are limitations such as requiring enough historical time series to recognize similar events and

*. *Corresponding author. Tel.: +613 9905 4957; Fax: +613 990 54944*
*E-mail addresses: mohammad.azmi@monash.edu (M. Azmi); fahimeh.sarmadi@monash.edu (F. Sarmadi)*

inability at extrapolating output, which need to be addressed appropriately [8,9].

## 2. General algorithm of K-NN in hydrological forecasting

The concept of K-NN method is based on observing values of the predictor variables in the current time and then searching for similar conditions amongst historical events. These similar conditions can be considered as possible solutions depending on the degree of similarity. In this method, Kernel function is used as a nonparametric distribution function. The general algorithm of K-NN is as follows:

1. Preprocessing values of all predictor variables;
2. Setting a matrix with $m$ columns (the number of predictors) and $n+1$ rows (the length of time series);
3. Assuming the last row of the above matrix as a vector of predictors at the current time ($x_{j,t}$, $j = 1:m$);
4. Assuming the remaining rows as a matrix of predictors at historical time series ($x_{j,t-i}$, $j=1:m$, $i=1:n$);
5. Defining the vector $Q$ with $n$ rows of independent variable values from $t-n$ to $t-1$;
6. Using distance function, distances between $x_{j,t}$ and $x_{j,(t-i)}$ are calculated as:
$$\text{Dist}(t-i) = f\left(w_j, x_{j,(t-i)}, x_{jt}\right), \tag{1}$$
where, $w_j$ is the weight of predictor variable $j$;
7. Sorting distances vector (Dist) from minimum to maximum ($S$Dist) and correspondingly sorting vector $Q$ based on $S$Dist;
8. Selecting the best number of neighbours ($k$) based on some methods which will be explained in Section 3.3;
9. Applying a discrete Kernel function [1] to give weights to $k$ neighbours as follows:
$$S(e) = \frac{1/S\text{Dist}(t-e)}{\sum_{e=1}^{k} 1/S\text{Dist}(t-e)} \quad e = 1...k \tag{2}$$
10. Forecasting the values as follows:
$$\text{Forecast} = S \times Q^T, \tag{3}$$
where, $T$ is the transpose operation.

## 3. The best selection of effective components at K-NN

The main components of K-NN include data preprocessing functions, distance functions, weights of predictors at distance functions, the number of neighbours, and extrapolation stage.

### 3.1. Data preprocessing functions

It is necessary to make the range of all variables consistent. The following functions are commonly used in data preprocessing [16]:

1. Auto scaling:
$$X_{ij} = \frac{x_{ij-m_j}}{\sigma_j}, \tag{4}$$
where, $x_{ij}$ is the value of predictor $j$ at time $i$; $m_j$ and $\sigma_j$ are the average and standard deviation of predictor $j$ at historical time series, respectively;
2. Range scaling:
$$X_{ij} = \frac{x_{ij} - L_j}{U_j - L_j}, \tag{5}$$
where, $L_j$ and $U_j$ are the minimum and maximum values of variable $j$ at historical time series, respectively;
3. Maximum scaling:
$$X_{ij} = \frac{x_{ij}}{U_j}. \tag{6}$$
4. Profiles:
$$X_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2 \sum_{j=1}^{m} x_{ij}^2}}. \tag{7}$$
5. Principal Component Analysis (PCA). This method has been used in a variety of applications to reduce the volume of information and number of variables contributing to a model [17]. In fact, after performing PCA on predictor variables, the aggregated variables are considered as new predictor variables.

### 3.2. Distance functions

The following distance functions are commonly used in K-NN method. In the following distance functions, $x_{ij}$ is the value of predictor $j$ at time $i$, and $w_j$ is the weight of predictor $j$:

1. Euclidean distance function:
$$\text{Dist}(t-i) = \sqrt{\sum_{j=1}^{m} w_j \left(x_{j,(t-i)} - x_{j,t}\right)^2}. \tag{8}$$
2. Manhattan distance function [18]:
$$\text{Dist}(t-i) = \sum_{j=1}^{m} w_j \left|x_{j,t} - x_{j,(t-i)}\right|. \tag{9}$$
3. Mahalanobis distance function [19,20]:
$$\text{Dist}(t-i) = \sqrt{\left(x_t - x_{t-i}\right)^T C^{-1} \left(x_t - x_{t-i}\right)}, \tag{10}$$
where, $C^{-1}$ is the inverse of covariance matrix.

4. Camberra distance function [18]:

$$\mathrm{Dist}(t-i)=\sum_{j=1}^{m} w_j\left(\frac{x_{j,t}-x_{j,(t-i)}}{x_{j,t}+x_{j,(t-i)}}\right)^2. \tag{11}$$

5. Lance-Villiams distance function [18]:

$$\mathrm{Dist}(t-i)=\frac{\sum_{j=1}^{m} w_j\left|x_{j,t}-x_{j,(t-i)}\right|}{\sum_{j=1}^{m} w_j\left|x_{j,t}+x_{j,(t-i)}\right|}. \tag{12}$$

6. Cosine Coefficient function [21]:

$$\mathrm{Dist}(t-i)=\frac{1}{m}\left(\sum_{j=1}^{m} w_j\frac{x_{j,t}\times x_{j,(t-i)}}{{x_{j,(t-i)}}^2}\right), \tag{13}$$

where, $m$ is the total number of predictor variables.

### *3.3. Selecting the number of neighbours and specifying weights at distance functions*

Three common methods to determine the number of neighbours are:

1. Personal experiences (Empirical methods): Different researchers have tried to find explicit equations to estimate the best number of neighbours and weights of predictor variables at distance functions. Here, the equation $K=\sqrt{n}$ is used as an approximation of the best number of neighbours [22]. In this equation, $n$ is the length of historical time series and $K$ is the approximated number of neighbours in K-NN method. The performance of this equation improves with increase in the length of time series. Further, weights of predictor variables at distance functions can be specified by user subject to professional experiences;
2. Trial and error methods: This method tries to determine the best number of neighbours and weights of predictor variables at distance functions by following a trial and error process. Generalized Cross-Validation (GCV) defined by Tarboton et al. [22] is presented as follows:

$$\mathrm{GCV}=\frac{\sum_{i=1}^{n}\mathrm{err}_i^2/n}{\left(1-1/\sum_{l=1}^{k}1/l\right)^2}, \tag{14}$$

where, $n$ is the total number of samples (here, time spans), $\mathrm{err}_i$ is the forecasts errors at time $i$, and $k$ is the best number of neighbours. Here, number of neighbours and weights of predictor variables, which lead to minimum value of GCV, are considered as the optimum solution. The main problem of this method is to determine the space of solutions as well as being highly time consuming, which can be persuasive to consider evolutionary optimization methods;
3. Evolutionary optimization algorithms: Evolutionary optimization methods are indeed the advanced form of the previous technique (trial and error process). In the current research, Honey-Bee Mating Optimization (HBMO) algorithm was applied to derive the optimum values of number of neighbours and weights of predictors at distance functions via minimizing the values of GCV (Eq. (14)). Over the past decade, HBMO has been used in different environmental subjects [23,24]. More information about the procedure and state-of-art of HBMO algorithm can be found at Sabbaghpour et al. [25].

### *3.4. Extrapolation in K-NN method*

In order to overcome the limitation of K-NN method in extrapolating the forecasts, the following method is proposed:

1. Deriving errors of forecasts of the training stage:

$$E=[Z_i]-\lfloor\hat{Z}_i\rfloor, \tag{15}$$

where, $Zi$ and $\hat{Z}_i$ are the vectors of observed and forecasted values of the dependent variable at training stage, respectively;
2. Running test stage;
3. Combining errors of the training stage with predicted values of the test stage:

$$Z_p=\text{K-NN}_{Z_i}(x_i)+\text{K-NN}_E(x_i), \tag{16}$$

where, $\text{K-NN}_{Z_i}(x_i)$ is the predicted value of the dependent variable at the test stage, $\text{K-NN}_{E_i}(x_i)$ is the error of the dependent variable at the training stage, and $Z_P$ is the predicted value of the dependent variable at the test stage. The sizes of training and test datasets were considered 75% and 25%, respectively [14,26,27]; further, It has been tried to have both extreme and normal events in both training and test datasets to reflect the abilities of this proposed method, more realistically.

## 4. Case study

Zayandeh-rud River is the main water resource for terrestrial ecosystem, and agricultural and municipal consumptions of Esfahan City located in the centre of Iran; and Zayandeh-rud dam, with a volume of 1470 MCM, controls the streamflow of Zayandeh-rud River (Figure 1). Monthly data of inflow to Zayandeh-rud dam during a 30-year period from 1971 to 2001 is used at this study. Previous research in this area of study has identified that the accumulated 3-month streamflow of autumn and the average of Southern Oscillation climate Index (SOI) from summer to the end of autumn can be considered as the most appropriate

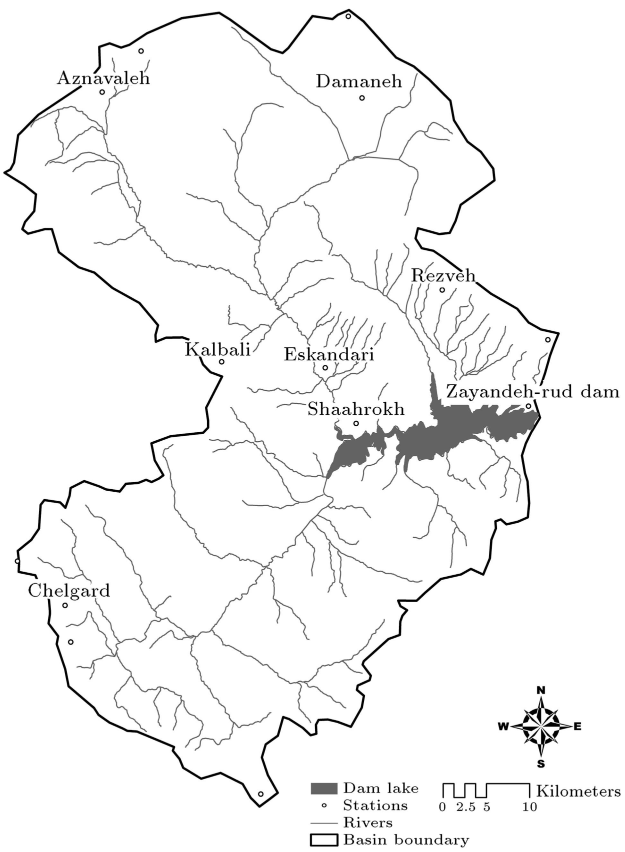


**Figure 1.** The location of Zayandeh-rud dam basin.

predictor variables for forecasting the accumulated 9-month inflow from winter to the end of the following summer [14,26].

Here, three goodness-of-fit criteria are employed in comparing the results. In the following equations, $obs_i$ and $for_i$ are the observed and forecasted values of the dependent variable, respectively, and $n$ is the total number of samples (time spans):

1. Root of Mean Square Error (RMSE):

$$\text{RMSE} = \frac{\sqrt{\sum_i^n (obs_i - for_i)^2}}{n}. \tag{17}$$

2. Volume Error (VE):

$$\%VE = \frac{\sum_i^n \left| \frac{obs_i - for_i}{obs_i} \right|}{n}. \tag{18}$$

3. Correlation coefficient:

$$\text{Corr}\% = \frac{Cov(obs, for)}{\sigma_{obs} \times \sigma_{for}}, \tag{19}$$

where, $Cov(obs, for)$ is the covariance between the observed and predicted values and $\sigma_{obs}$ and $\sigma_{for}$ are standard deviations of the observed and predicted values, respectively.

## 5. Results and discussion

In the current research, five data preprocessing functions, six distance functions, and three approaches (with 90 combinations) were applied for specifying the best number of neighbours and weights of predictors. Table 1 presents the best derived results amongst all the mentioned combinations with/without taking benefit from the proposed extrapolation method.

As for the Empirical method, due to the length of time series (30 time spans), the best number of neighbours was calculated 5 ($k = 5$). Further, considering the correlation between the predictors and dependent

**Table 1.** Final results of the best effective components of K-NN method for long-lead forecasting of inflow to Zayandeh-rud dam.

| | Methods | Preprocessing functions | Distance functions | Weights of predictors | | The best no. of neighbours | Extrapolation method | Whole data | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 3-month streamflow of autumn | 6-month SOI | | | CORR | %V | RMSE |
| For the best combinations of preprocessing and distance functions | Empirical & trial-error processes | Maximum scaling | Euclidean | 0.8 | 0.2 | 5 | No | 80 | 17 | 42 |
| | | | | | | | Yes | 84 | 14 | 35 |
| | Evolutionary optimization algorithm | Range scaling | Mahanalobis | 0.63 | 0.37 | 6 | No | 92 | 10 | 34 |
| | | | | | | | Yes | 96 | 8 | 25 |
| For the worst combinations of preprocessing and distance functions | Empirical & trial-error processes | Profile | Cosine | 0.8 | 0.2 | 5 | No | 65 | 31 | 54 |
| | Evolutionary optimization algorithm | Profile | Lance-William | 0.63 | 0.37 | 6 | No | 75 | 28 | 50 |

variable, weights of predictors were considered as 0.8 for accumulated autumn streamflow ($w_{\text{streamflow}} = 0.8$) and 0.2 for average of SOI from summer to the end of autumn ($w_{\text{SOI}} = 0.2$). In the trial-error method, weights were considered similar to those of the empirical method and then GCV was calculated for a range of $k$ from 1 to 10.

According to Table 1, the best data preprocessing and distance functions for the empirical and trial-error processes were Maximum Scaling and Euclidean, respectively. RMSE, VE, and Corr for the above-mentioned combination were derived 42, 17, and 80, respectively. With the same $k$ and weights, the worst results of RMSE, VE, and Corr were evoked as 54, 31, and 65 for the combination of profile function as the data preprocessing function and cosine function as the distance function.

When evolutionary optimization algorithm of HBMO was used, the best effective component was presented as $k = 6$, $w_{\text{streamflow}} = 0.63$, $w_{\text{SOI}} = 0.37$, range scaling as the data preprocessing function, and Mahanalobis as the distance function. RMSE, VE, and Corr for this combination were calculated 34, 10, and 92, respectively. With the same $k$ and weights, the worst results had RMSE, VE, and Corr equal to 50, 28, and 75 for the combination of profile function as the data preprocessing function and Lance-William function as the distance function.

The positive influence of employing the proposed extrapolation method at K-NN is shown in Table 1. The results present the capability of this method to present more accurate forecasts, especially in case of the existing extreme events. The mentioned extrapolation method has improved the values of RMSE, VE, and CORR to 35, 14, and 84 for the Empirical and the Trial-Error process methods, as well as to 25, 8, and 96 for the HBMO method.

The results of goodness-of-fit criteria for the optimized K-NN in comparison with a regular K-NN presents an increase by 31% in CORR (from 65% to 96%), a decrease from 31% to 8% in VE, and finally a drop from 54% to 25% in RMSE. Figures 2 to 5

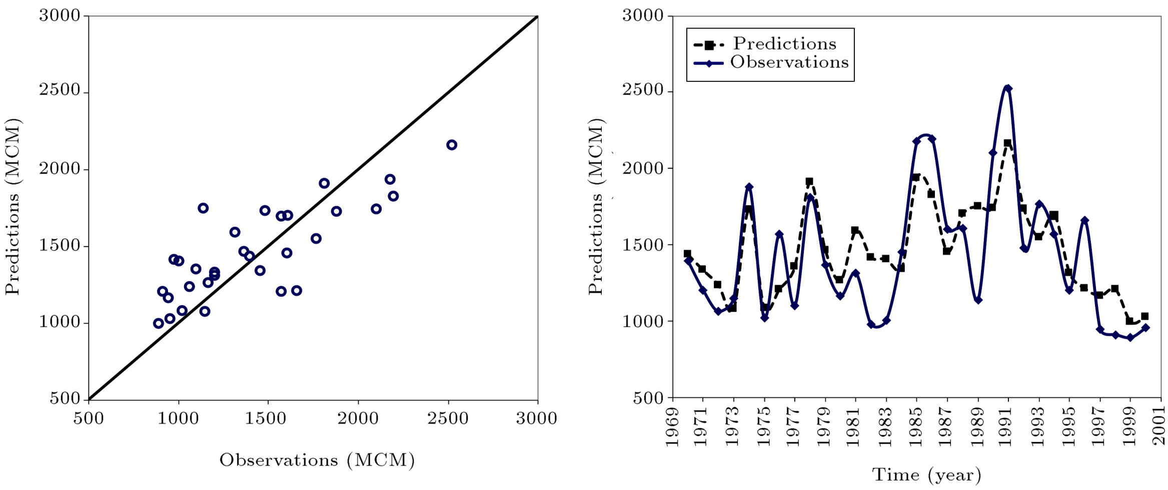


**Figure 2.** Final results based on the empirical and trial-error methods without applying the extrapolation method.

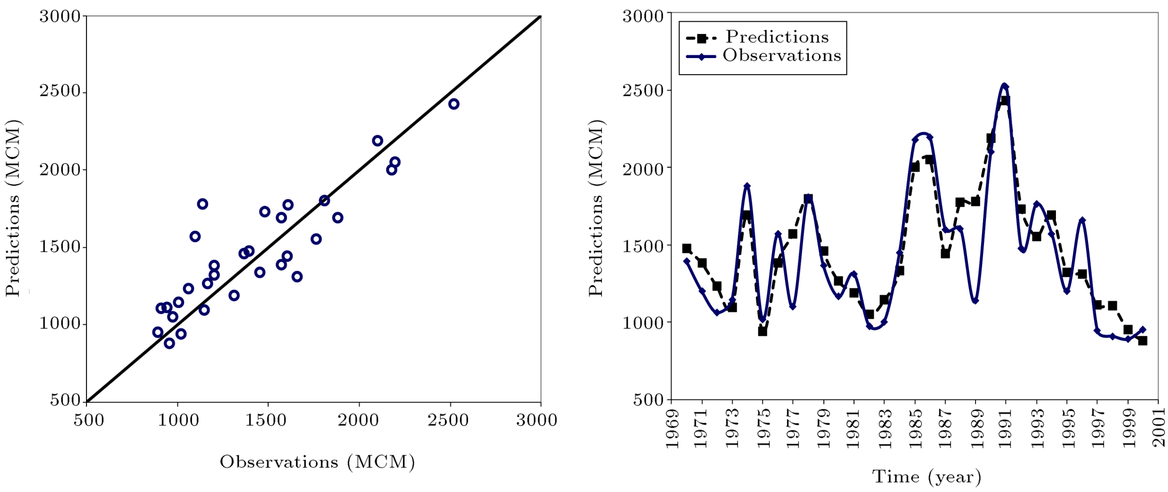


**Figure 3.** Final results based on the empirical and trial-error methods with applying the extrapolation method.

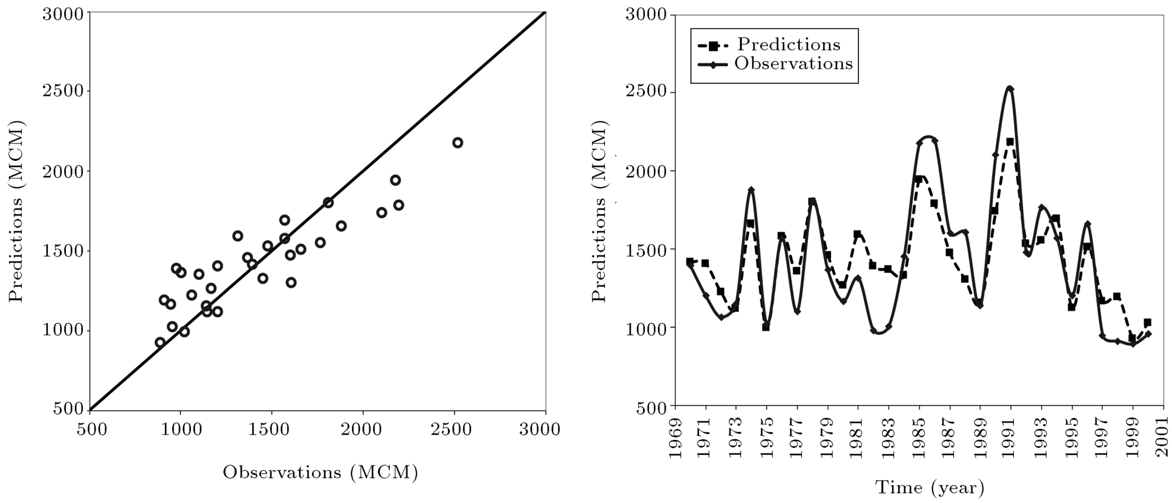


**Figure 4.** Final results based on HBMO method without applying the extrapolation method.

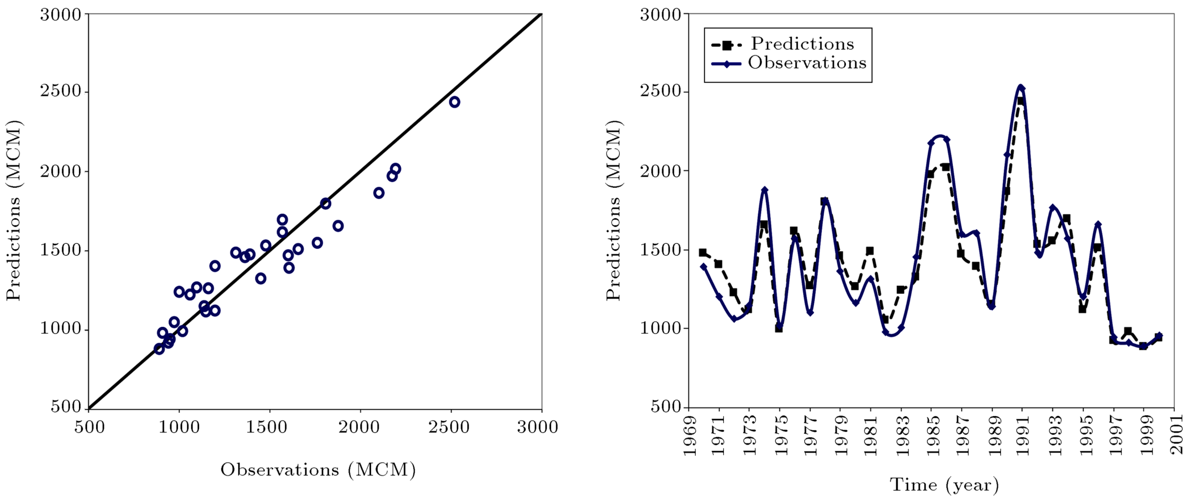


**Figure 5.** Final results based on HBMO method with applying the extrapolation method.

indicate the forecasts based on different approaches with/without considering the proposed extrapolation method. Final results present the increase in the accuracy of forecasts by using the proposed extrapolation method in comparison with common K-NN, especially for extreme events such as 1982, 1986-7, 1991, and 1997 to 2001. It is worth noting that even by using extrapolation method, the forecasts of maximum extreme events (e.g., 1992) are still underestimated, while the forecasts of minimum extreme events (e.g., 1997 to 2001) are overestimated, which shows the necessity of further research in this issue.

## 6. Summary, conclusion, and future research direction

This research aimed at showing the potentials, abilities, and disadvantages of K-NN method in long-lead streamflow forecasting of Zayandeh-rud River inflow to Zayandeh-rud storage reservoir from 1971 to 2001. A variety of combinations between preprocessing and distance functions along with different methods to estimate the best number of neighbours and weights of predictors at distance functions were considered to increase the accuracy of forecasts. Moreover, a proposed method was introduced and applied to decrease the errors of forecasts, especially for extreme events.

Overall, the results presented remarkable improvements: an increase by 31% in CORR (from 65% to 96%), a decrease from 31% to 8% in VE, and finally a drop from 54% to 25% in RMSE; however, it seems that the proposed extrapolation method has not been completely successful (underestimations of maximum extreme events and overestimations of minimum extreme events). However, finally, using more complicated methods such as HBMO in optimizing the best number of neighbours and weights of predictors at distance functions leads to growing the accuracy in terms of complexities and time consuming. It seems that Empirical methods can be considered as a serious competitive alternative. The latter point, besides introducing more effective extrapolation methods, is a topic which needs to be considered as a direction for future research.

## References


1. Yakowitz, S.J. "Nonparametric density estimation, prediction, and regression for Markov sequences", *J. Am. Stat. Assoc.*, **80**, pp. 215-221 (1985). DOI: 10.1016/0047-259X(89)90091-2.

2. Karlsson, M. and Yakowitz, S. "Nearest-neighbour methods for nonparametric rainfall-runoff forecasting", *Water Resources Research*, **23**(7), pp. 1300-1308 (1987). DOI: 10.1029/WR023i007p0 1300

3. Lall, U. and Sharma, A. "A nearest neighbour bootstrap for resampling hydrologic time series", *Water Resources Research*, **32**(3), pp. 679-694 (1996). DOI: 10.1029/95WR02966

4. Yates, D., Gangopadhyay, S., Rajagopalan, B. and Strzepek, K. "A technique for generating regional climate scenarios using a nearest-neighbour algorithm", *Water Resoures Research*, **39**(7), pp. 1114-1121 (2003). DOI: 10.1029/2002WR001769

5. Zhang, L., Liu, Q., Yang, W., Wei, N. and Dong, D. "An Improved K-nearest neighbor model for short-term traffic flow prediction", *Procedia: Social and Behavioral Sciences* (2013). DOI: 10.1016/j.sbspro.2013.08.076

6. Mangalova, E. and Agafonov, E. "Wind power forecasting using the k-nearest neighbours algorithm", *International Journal of Forecasting*, **30**(2), pp. 402-406 (2014). DOI: 10.1016/j.ijforecast.2013.07.008

7. Toth, E., Brath, A. and Montanari, A. "Comparison of short-term rainfall prediction models for real-time flood forecasting", *Journal of Hydrology*, **239**(4), pp. 132-147 (2000). DOI: 10.1016/S0022-1694(00)00344-9

8. Piechota, T.C., Chiew, F.H.S., Dracup, J.A. and McMahon T.A. "Development of an exceedance probability streamflow forecast using El Nino-southern oscillation and sea surface temperatures", *ASCE Journal of Hydrologic Engineering*, **6**, pp. 20-28, http://dx.doi.org/10.1061/(ASCE)1084-0699(2001) 6:1(20) (2001).

9. Jayawardena, A.W., Li, W.K. and Xu, P. "Neighbour selection for local modelling and prediction of hydrological time series", *Journal of Hydrology*, **258**, pp. 40-57 (2002). DOI: 10.1016/S0022-1694(1)00557-1

10. Meade, N. "A comparison of the accuracy of short-term foreign exchange forecasting methods", *International Journal of Forecasting*, **18**(1), pp. 67-83 (2002). DOI: 10.1016/S0169-2070(01)00111-X

11. Sharma, A., Luck, K.C., Cordery, I. and Lall, U. "Seasonal to interannual rainfall probabilistic forecast for improved water supply management: Part 2 - predictor identification of quarterly rainfall using ocean-atmosphere information", *Journal of Hydrology*, **239**, pp. 240-248 (2000). DOI: 10.1016/S0022-1694(00)00347-4

12. Araghinejad, S. and Burn, H.D. "Probabilistic forecasting of hydrological events using geostatistical analysis", *Hydrological Sciences Journal- Des Sciences Hydrologiques*, **50**(5), pp. 57-66 (2005). DOI:10.1623/hysj.2005.50.5.837

13. Asadiani Yekta, A.H. "Comparison between inflow suspended load of Ekbatan Storage using ANFIS and K-NN method", *7th Conference on Hydraulics*, Shiraz, Iran, April, pp. 246-253 (2007).

14. Azmi, M., Araghinejad, S. and Kholghi, M. "Multi-model data fusion for hydrological forecasting using K-nearest neighbour method", *Iranian Journal of Science & Technology, Transaction B*, **34**(B1), pp. 81-92 (2010).

15. Sorjamaa, A., Reyhani, N. and Lendasse, A. "Input and structure selection for K-NN approximator", *8th International Conference on Artificial Neural Networks*, IWANN 2005, *Lecture Notes in Computer Science*, Springer, pp. 958-992 (2005).

16. Todeschini, R. "K-nearest neighbour method: Influence of data preprocessings and metrics", *Chemometrics and Intelligent Laboratory Systems*, **6**, pp. 213-220 (1989).

17. Jolliffe, I.T., *Principal Component Analysis*, Springer Publication, Second Ed. (2002).

18. Lance, G.N. and Williams, W.T. "Computer programs for hierarchical polythetic classification", *Computer Journal*, **9**(1), pp. 60-64 (1966). DOI:10.1093/comjnl/9.1.60

19. Sharif, M. and Burn, H.D. "Simulating climate change scenarios using an improved K-nearest neighbour model", *Journal of Hydrology*, **325**, pp. 179-196 (2006). DOI: 10.1016/j.jhydrol.2005.10. 015

20. Sharif, M. and Burn, H.D. "Improved K-nearest neighbor weather generating model", *Journal of Hydrologic Engineering*, **12**(1), pp. 42-51 (2007). http://dx.doi.org/10.1061/(ASCE)1084-0699(2007)12:1(42).

21. Jones, W.P. and Furnas, G.W. "Pictures of relevance: A geometric analysis of similarity measures", *Journal of the American Society for Information Science*, **36**(6), pp. 420-442 (1987).

22. Tarboton, D.G., Sharma, A. and Lall, U. "The use of non-parametric probability distribution in streamflow modeling", in *Proceedings of the Sixth South African National Hydrological Symposium*, Ed. S.A. Lorentz et al., University of Natal, Pietermaritzburg, South Africa, September 8-10, pp. 315-327 (1993).

23. Afshar, A., Bozorg Haddad, O., Marino, M.A. and Adams B.J. "Honey-bee mating optimization algorithm for optimal reservoir operation", *J. Franklin Institute*, **344**(5), pp. 452-462 (2007). DOI: 10.1016/j.jfranklin.2006.06.001

24. Bozorg Haddad, O., Adams, B.J. and Marino, M.A. "Optimum rehabilitation strategy of water distribution systems using the HBMO algorithm", *Journal of Water Supply: Research and Technology-AQUA*, **57**(5), pp. 1-16 (2008). DOI: 10.2166/aqua.2008.151

25. Sabbaghpour, S., Naghashzadehgan, M., Javaherdeh, K. and Bozorg Haddad, O. "HBMO algorithm for calibrating water distribution network of Langarud city", *Water Science & Technology*, **65**(9), pp. 1564-9 (2012). DOI: 10.2166/wst.2012.045

26. Araghinejad, S., Azmi, M. and Kholghi, M. "Application of artificial neural network ensembles in probabilistic hydrological forecasting", *Journal of Hydrology*, **407**(1-4), pp. 94-104 (2011). DOI: 10.1016/j.jhydrol.2011.07.011

27. Gharun, M., Azmi, M. and Adams, M.A. "Short-term forecasting of water yield from forested catchments after bushfire: A case study from southeast Australia", *Journal of Water*, **7**(2), pp. 599-614 (2015). DOI: 10.3390/w7020599


## Biographies

**Mohammad Azmi** received his BEng degree in Water Engineering from University of Tehran, Iran, in 2006, and then received his MEng degree in Water Resources Engineering from the same university in 2008. In 2009, he joined some consulting engineering corporations as an expert, head expert, and project manager assistant. Meantime, he cooperated with University of Tehran as a Research/Teacher Assistant as well as Invited Lecturer. In 2013, he commenced his PhD degree in the Department of Civil Engineering, Monash University, which is expected to be completed in 2016. His main research interests consist of hydrometeorlogical forecasting, drought/flood monitoring, and water resources modelling.

**Fahimeh Sarmadi** received her BEng degree in Water Engineering from University of Tehran, Iran, in 2010, and then received her MEng degree in water engineering from IKIU, Iran, in 2012. During 2013-14, she was involved in some research projects, of which the results are published in forms of journal and conference papers. In 2015, she commenced her PhD degree at the School of Earth, Atmosphere, and Environment (EAE), Monash University, which is expected to be completed in 2018. Her main research interests consist of meteor-climatology, cloud simulations, hydrometeorlogical forecasting, and drought/flood monitoring.