Application of artificial neural network ensembles in probabilistic hydrological forecasting

Shahab Araghinejad *, Mohammad Azmi, Majid Kholghi
College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran

Abstract

Ensemble techniques are used in regression/classification tasks with considerable success. Due to the flexible geometry of artificial neural networks (ANNs), they have been recognized as suitable models for ensemble techniques. The application of an ensemble technique is divided into two steps. The first step is to create individual ensemble members, and the second step is the appropriate combination of outputs of the ensemble members to produce the most appropriate output. This paper deals with the techniques of both generation and combination of ANN ensembles. A new performance function is proposed for generating neural network ensembles. Also a probabilistic method based on the K-nearest neighbor regression is proposed to combine individual networks and to improve the accuracy and precision of hydrological forecasts. The proposed method is applied on the peak discharge forecasting of the floods of Red River in Canada as well as the seasonal streamflow forecasting of Zayandeh-rud River in Iran. The study analyses the advantages of the proposed methods in comparison with the conventional empirical methods such as conventional artificial neural networks, and K-nearest neighbor regression. The utility of the proposed method for forecasting hydrological variables with a conditional probability distribution is demonstrated. The results of this study show that the application of the ensemble ANNs through the proposed method can improve the probabilistic forecast skill for hydrological events.

Keywords:
Artificial neural network ensembles
Nearest neighborhood
Hydrological forecasting
Canada
Iran

1. Introduction

Forecasting future hydrological conditions is one of the most important issues in water resources planning and management. Forecasting of hydrological variables is used for warning of floods and droughts, reservoir operation, contract negotiation, and irrigation scheduling. Hydrological forecasting methods seek to develop a mapping function between independent and dependent variables. Even though, efforts have been dedicated by many researchers to improve conceptual and empirical methods of forecasting, there is still a need to improve the accuracy and reliability of operational forecasts. Ensemble techniques have been recently applied in hydrology as an approach to enhance the skill of forecasts.

Ensemble techniques have been used in regression/classification tasks with considerable successes (Krogh and Vedelsby, 1995). Normally, the output of an ensemble is a weighted sum, where the weights are determined from the training or validation of data. The motivation for this procedure is based on the idea that by combining the outputs of several individual predictors one might improve on the performance of a single generic one (Krogh and Vedelsby, 1995). Operational hydrological forecasting may benefit from the ability of combining information derived from multiple sources, such as the individual outputs of different forecasting models. Through the application of the so called data fusion methods, both raw and processed information can be fused into the useful outputs including higher level decision. The approach of using ensemble models in hydrological forecasting has been applied by researchers such as See and Abrahart (2001, Abrahart and See (2002), Anctil and Laouzon (2004), Georgakakos et al. (2004), McIntyre et al. (2005), and Kim et al. (2006). See and Abrahart (2001) used data fusion approach for continuous river level forecasting where data fusion was the amalgamation of information from multiple sensors and/or different data sources. Abrahart and See (2002) evaluated six data fusion strategies and found that the data fusion using an ANN model provides the best solution. Anctil and Laouzon (2004) used stop training, and Bayesian regularization, as well as data fusion approaches such as stacking, bagging and boosting to generalize ANN results in the field of streamflow predictions. They concluded that the results provided by those approaches improved performance of the ANNs compared to the standard ANNs without generalization. Georgakakos et al. (2004) used the ensemble approach to simulate uncertainty in streamflow simulation. McIntyre et al. (2005) used the basis of ensemble mod-
eling and model averaging for regionalization of rainfall–runoff models in ungauged basins. Kim et al. (2006) demonstrated that some statistical methods, such as the simple average, constant coefficient regression, switching regression, and ANNs can be exploited as combining methods. They combined ensemble streamflow prediction scenarios of several rainfall–runoff models for forecasting the monthly inflow to the Daechegung multipurpose dam in Korea. Besides, they took advantages of correction methods, such as optimal linear and ANNs correction methods to improve the probabilistic forecasting accuracy.

Due to the flexible geometry of ANNs, they have been recognized as suitable models to be used in the ensemble techniques. ANN ensembles offer a number of advantages over a single ANN since they have the potential for improving generalization. Cannon and Whitfield (2002) and Shu and Burn (2004) used ANN ensembles in the hydrology context. Cannon and Whitfield (2002) used ANN ensembles to predict changes in streamflow conditions in British Columbia, Canada through a downscaling model. Shu and Burn (2004) applied artificial neural network ensembles in pooled flood frequency analysis for estimating the index flood and the 10-year flood quintiles. They used data fusion method to combine individual ANN models in order to enhance the final estimation.

The forecast variable may be estimated in a deterministic or probabilistic sense. In a probabilistic approach, the output of a forecasting model consists of probabilities of occurrence of different values or categories of rainfall or streamflow during a specific condition. A probabilistic forecast procedure predicts a full range of values that are expected rather than relies solely on a single value forecast (Anderson et al., 2001; Krzysztofowicz, 2001; Araghinejad and Burn, 2005). Developing probabilistic forecasts is of significance in risk-based decision making and it is encouraging in case of facing with the changing climate. Categorical forecasts of streamflow are important for effective water resources management. Typically, they are obtained by generating ensemble forecasts of streamflow and counting the proportion of ensembles in the desired category (Regonda et al., 2006). Regonda et al. (2006) developed a method of predicting the probability of the leading mode (or principal component) of the basin streamflows above a given threshold and subsequently translating the predicted probabilities to all the sites in the basin.

This paper uses ensemble techniques to benefit from the generalization ability of ANNs and probabilistic modeling of K-nearest neighbor method in probabilistic hydrological forecasting. This study deals with the both techniques of creating ensemble members and combining ANN ensembles. This research contributes previous works by proposing a new performance function for creating ANN ensembles as well as applying a probabilistic method to combine various ensembles. The proposed methods are tested on the Red River in Canada and Zayande–rud River in Iran.

The remainder of the paper is structured as follows. The next section presents the procedure of forecasting containing the proposed methods for creating and combining ANN ensembles. It is followed by description of two case studies and the data that are used in the study. The paper continues with the presentation and discussion of the results and ends with a conclusion.

2. Artificial neural network ensembles

The general equation of a hydrological event forecasting model is

\[ y_i = f(X_i) + \varepsilon_i \quad i = 1, \ldots, n \]  

(1)

where \( X \) is the vector of predictors, \( y \) the forecast variable, \( \varepsilon \) the model error and \( n \) is the number of observation data. In the case of using multiple artificial neural networks with similar predictors, Eq. (1) is changed to the following form:

\[ \hat{Y}_i = \left[ \begin{array}{c} \hat{y}_{i1} \\ \hat{y}_{i2} \\ \vdots \\ \hat{y}_{im} \end{array} \right] = \left[ \begin{array}{c} f_1(X_i) \\ f_2(X_i) \\ \vdots \\ f_m(X_i) \end{array} \right] + \left[ \begin{array}{c} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{im} \end{array} \right] \quad i = 1, \ldots, n \]  

(2)

where \( m \) is the number of ANN members used to forecast \( y \). \( \hat{Y}_i \) is the matrix of estimations of \( y \) provided by different individual ANNs. \( f_j(X) \) the output of \( j \)th artificial neural network (\( j = 1, \ldots, m \)), and \( \varepsilon_i \) is the forecast error associated with the \( j \)th ANN (\( j = 1, \ldots, m \)).

The ensemble techniques are divided into two steps. The first step is to create individual ensemble members (\( f_p,s \)), and the second step is the appropriate combination of outputs from the ensemble members to provide a final output. Using the combining approach \( \hat{Y} \) is sum up to a unique estimation of \( y \). The following sections present the proposed methods of this paper in both creating and combining steps.

2.1. The proposed method for creating ensemble members

The goal of creating ensemble ANNs is to design networks with diverse generalization ability. The following approaches are the most widely used ones (Sharkey, 1999): (1) manipulating the set of initial random weights, (2) varying the topologies, (3) manipulating the training set, (4) varying the training algorithm. Among the mentioned approaches, using different training algorithms by diverse performance functions makes it possible to train specific networks with special ability in estimation of certain values of the dependent variable more accurate than the other values of the same data. For instance, Coulibaly et al. (2001a,b) used a specific performance function known as PLC (peak and low flow criterion) to improve the ability of the ANNs in extreme value forecasting.

The PLC for an input set \( k \) is specified as

\[ PLC = P_k \times L_k \]  

(3)

where \( P_k \) is the peak flow criterion given by

\[ P_k = \left( \frac{\sum_{i=1}^{n_p}(y_{pi} - \bar{y}_k)^2}{\sum_{i=1}^{n_p} y_{pi}^2} \right)^{0.25} \]  

(4)

where \( n_p \) is the number of peak flows greater than one-third of the mean flow observed, \( y_{p} \) and \( \bar{y}_k \) are the observed and the computed peak flows, respectively.

\( L_k \) is the low flow criterion, which is given by

\[ L_k = \left( \frac{\sum_{i=1}^{n_l}(y_{li} - \bar{y}_k)^2}{\sum_{i=1}^{n_l} y_{li}^2} \right)^{0.25} \]  

(5)

where \( n_l \) is the number of low flows lower than one-third of the mean flow observed, and \( y_{l} \) and \( \bar{y}_k \) are observed and the computed low flows, respectively. \( P_k \) provides a more accurate measure of the model performance for the peak flow periods, whereas the \( L_k \) is a better performance indicator for the low flow periods.

However, the PLC is a good measure for indicating goodness of fit at high and low flows, we need a performance function, which is flexible to fit at different quantiles of a hydrological variable as desired. Here we introduce another performance function to deal with this flexibility as follows:

\[ PF = \sum_{i=1}^{n} \left[ \frac{10\log_{10} \frac{\bar{y}_i}{\hat{y}_i}}{1 + \log_{10} \bar{y}_i} \right] \]  

(6)
where \( y_i \) and \( \hat{y}_i \) are the observed and the estimated dependent variables respectively, \( z_{im} \) is the values between \(-10\) and \(10\), which are obtained by rescaling of dependent variables to \([-10, 10]\) at the \(m\)th try for creating individual network. To calculate \( z \) numbers, a specific value of dependent variable, \( y_{im} \) is considered as \(10\), the value with maximum distance to \( y_{im} \) is considered as \(-10\), and the other dependent variables are re-scaled between these two numbers. Choosing different values of \( y_{im} \), the performance function of Eq. (6) makes an individual ANN biased to certain values of dependent variables close to \( y_{im} \). Actually, it considers an exponential weighting function as shown in Fig. 1 to indicate the preference of some \( m\)th forecasts of \( y_{im} \). Three different weighting functions are associated with the three different quantiles of the cumulative distribution function shown in Fig. 1. As it is shown in Fig. 1, \( y_{im} \) could be defined as:

\[
F(y_m) = P[y \leq y_m] = b
\]

where \(0 < b < 1\), and takes different values during selecting different \( y_{im} \). \( y_{im} \) varies during the creation of different networks. It is suggested that \( y_{im} \)'s are chosen such that they cover the whole range of dependent variable. In the other hand it is suggested that \( b \) would take symmetrical values between 0 and 0.5, and 0.5 and 1.

### 2.2. The proposed method for combining ANN members

A spread range of experimental and statistical methods have been proposed for combining ANN members through an ensemble technique. Relying on the user’s experience; simple and weighted averaging; using a combining artificial neural network model; and error analysis (See and Abrahart, 2001; Shu and Burn, 2003) are examples of techniques used in combining of ANN ensembles. This study presents a statistical technique based on the well known nonparametric \(K\)-nearest neighbor technique.

The recognition of the nonlinearity of the underlying dynamics of hydrological processes, have spurred the growth of nonparametric methods. Nonparametric estimation of probability densities and regression functions are pursued through weighted local averages of the dependent variable. This is the foundation for nearest neighbor methods. \(K\)-nearest neighbor (K-NN) methods use the similarity (neighborhood) between observations of predictors and similar sets of historical observations (successors) to obtain the best estimate for a dependent variable (Karlsson and Yakowitz, 1987; Lall and Sharma, 1996). Prairie et al. (2006) developed a lag-1 modified K-NN approach applied to stochastic streamflow simulation. This approach was similar to what is used in autoregressive methods. The process contained two main steps of obtaining the conditional mean and bootstrapping. Results were found to show better performance in comparison with the conventional parametric periodic autoregressive and nonparametric index sequential models. Gangopadhyay et al. (2005) used this method to generate ensembles and derive precipitation and temperature from large scale weather model.

In this paper, using K-NN technique, the outputs of individual networks are combined by the method of nearest neighborhood in a way that in the averaging process, the networks with better performance during the hydroclimatological conditions similar to the current condition, are assigned by greater values of weights. Furthermore, the forecasting errors of each specific network are contributed into the real time forecasting to improve the forecast resolution. The summary of the explained procedure is shown in Fig. 2.

The explained concept is used through the algorithmic procedure of the following paragraphs to estimate the probable realizations of the dependent variable.

Let \([X_i] (i = 1, 2, 3, \ldots , \text{number of observed data})\) be the past records of a \(P\)-dimensional vector of predictors and \(y_i \) is the past records of associated dependent variable. At each instant \(t\), considering the observation of \([X_i]\), the following procedure is proposed to forecast \(y_t\).

1. Determine \(m\) values from the cumulative distribution function of dependent variable considering different values of \(b\) as described in Eq (7).
2. Train \(m\) individual networks using the performance function of Eq. (6). Evaluate individual networks in forecasting all \(n\) values of historical \(y\)'s. Compute matrix of \([A] = [a_{pq}]_{m \times n}\) where \(a_{pq} = 1\) if \(m\)th network results in the best forecast during \(i\)th experience; otherwise \(a_{pq} = 0\). Also compute \([\bar{E}] = [e_{ip}]_{m \times 1}\) where \(e_{ip}\) is the minimum forecast error in estimation of \(i\)th \(y\).
3. For each instant \(t\), compute \(m\) forecasts of \(y_t\), using \(m\) individual networks and develop \([\{\tilde{Y}_t\} = [y_{1t}, y_{2t}, \ldots , y_{mt}]^T\) where \(y{i}_{it} (i = 1, 2, \ldots , m)\) is the forecast provided by \(m\)th individual network at real time \(t\).
4. Compute \([F] = [A] \times [\tilde{Y}_t] = [f_1 \ f_2 \ \ldots \ f_m]\)  
5. Compute Mahalanobis distance (Yates et al., 2003) between the current vector of predictors, \([X_i]\), and the past vectors predictors, \([X_j]\) \(i = 1, 2, 3, \ldots, \text{number of observed data}\). By using this approach the effect of the scale of predictors through the mapping process is minimized. Moreover, significance of a predictor over the others is considered automatically by the correlation analysis. The distance metric between \([X_i]\) and \([X_j]\) is defined as

\[
d_i = \sqrt{(X_i - X_j)C_i^{-1}(X_i - X_j)^T}
\]

where \(T\) represents the transpose operation and \(C_i^{-1}\) is the inverse of the covariance matrix.

6. Sort the Mahalanobis distances in ascending order and retain the first \(K\) nearest neighbors. \(K\) could be chosen by the heuristic method recommended by Yates et al. (2003) as \(K = \sqrt{n}\), where \(n\) is the number of observed data.

7. A discrete probability distribution that gives higher weights to the closer neighbors is used to determine the contribution of each individual network in real time forecasting. Weights are assigned to each of \(q \ (q = 1, \ldots, K)\) neighbors according to the metric defined by

\[
w_q = \frac{1/q}{\sum_{i=1}^{q} 1/q}
\]

8. Set up the weight matrix of \([W] = [w_i]\) where \(w_i = w_q\) for the \(q\) neighbors determined by the previous step; otherwise \(w_q = 0\).

9. The weighted output is then calculated as

\[
\hat{y}_t = [W]^T ([F] - [\bar{E}]) = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}^T \begin{bmatrix} f_1 & f_2 & \cdots & f_n \\ e_1 & e_2 & \cdots & e_n \end{bmatrix}
\]

10. \(\hat{y}_t\) is considered as the estimated expected value of the dependent variable at time \(t\). Different realizations of dependent variable are known as \(f_i - e_i\), in condition that their associated \(w_i\) is not equal to zero. Obviously, dealing with the different values of \(f_i - e_i\) and their associated \(w_i\) will result in the discrete form of conditional probability function of the dependent variable at time \(t\).
3. Case study I: forecasting flood peak discharge

Flooding of the Red River in Canada typically results from snowmelt, often in combination with spring precipitation events. The nature of the Red River, with its headwaters in the USA, results in considerable warning of pending flood events within the Canadian portion of the watershed (Burn, 1999). The low slope of the river channel, and the consequent low water velocities, facilitates advanced warning of flooding conditions. The major causal parameters of Red River floods, based on previous flood studies, are (Warkentin, 1999): Index of soil moisture at freeze-up the previous autumn, based on weighted basin precipitation from May to October; Average degree-days per day at Grand Forks during the active melt period; Total basin precipitation from 1 November of previous year to the start of active melt during the flood year; Total basin precipitation from the start of active spring melt to the date of the spring crest at Emerson; and Index of the south–north time phasing of the runoff based on the percentage of tributary peaks experienced on the date of the mainstream peak at specific points from Halstad to Winnipeg (percent of worst possible). The above five parameters are used as predictors of the flood peak discharge of Red river.

3.1. Application of the models

Three different models including, nonparametric $K$-nearest neighbor regression ($K$-NN), conventional multilayer perceptron network (MLP), an MLP trained by the PLC performance function (MLP–PLC), and artificial neural network ensembles (ANNE) developed by the methods presented in this paper are applied as the flood forecasting models of this case study. $K$-nearest neighbor regression uses the procedure explained in the Lall and Sharma (1996), and Araghinejad and Burn (2005) with a difference that it uses the Mahanoble distance as presented in Eq. (8). MLP is trained by the well known performance function of $e_j^n$, where $e$ is the difference between actual and estimated values of dependent variable, and $S$ is number of neurons in the hidden layer of the network. MLP–PLC is trained by the performance function as explained in Eqs. (3)–(5). ANNE uses the procedure explained in previous sections. Five individual ANNs are trained to be considered as ensemble members. The five individual members are associated to $b$ values of 0.05, 0.25, 0.5, 0.75, and 0.9 as discussed in Eq. (6).

In all models, the 60 years of annual peak discharge data from 1940 to 1999 are used for calibration and validation. The data is divided to three sets of 1940–1959; 1960–1979; and 1980–1999. At each experience, two sets out of three sets are used as calibration data and the remaining set is used as the validation data. The point forecasts are compared by three statistics of root mean square error (RMSE), percent of volume error (%VE), and the linear correlation between actual and forecast data (CORR) in a deterministic manner. Each statistic is reported as the average of what is obtained for three validation data set.

Environmental and water resources datasets are frequently noisy and need analyzing by some noise processes (Cawley et al.,...
Analysis issues can be used in assessing the impacts of climate change and to improve decision making processes (Cawley et al., 2007). The best forecasting models are those which are able to quantify the estimation errors in a probabilistic manner. Carney and Cunningham (2006) discussed that the probability density forecasting for regression can be applied to two main goals: sharpness and calibration. These play a main role in selecting the best forecasting model. One of the most reliable approaches to evaluation of predictive performance is based on the paradigm of maximizing the sharpness of the predictive distribution subject to calibration (Gneiting et al., 2007).

A good probabilistic forecast method should produce an output distribution that is both accurate and precise. A forecast distribution is accurate if some fixed probability interval contains the true actual value. The precision of a forecast distribution is measured by its narrowness. The linear error in probability space (LEPS) score is one measure of skill for an exceedance probability forecast, which evaluates both accuracy and precision. The theory of the LEPS score for a single forecast, and its advantages relative to other statistics, are discussed by Potts et al. (1996). Piechota et al. (2001) modified the LEPS score for exceedance probability forecast. In this paper, the LEPS score developed by Piechota et al. (2001) is applied for the evaluation of the probabilistic forecast skill of ANNE and KNN models, which produce probabilistic forecasts, explicitly. Instead, standard deviation of residuals (SDR) is compared as a statistic representing the skill of a probabilistic forecast in MLP and MLP–PLC models.

The distance between forecast and observed values is defined as

\[ S = 3(1 - |p_f - p_v| + p_f^2 - p_f^2 + p_v^2 - p_v^2) - 1 \]  

where \( p_f \) is the cumulative probabilities of the observations found from the unconditional exceedance probability curve and \( p_f \) is the cumulative probability of the forecast found from the unconditional exceedance probability curve. The unconditional exceedance probability curve of the predicted value is based on the historical observations and is generated by dividing the rank of each historical value by the total number of years in the record. For each forecast, \( p_f \) remains constant while \( p_v \) varies depending on the cumulative probability of forecast. Different \( p_f \) values with equal increments are chosen from the cumulative probability of forecast. The LEPS score for the ensemble forecasts is then calculated by

\[ S' = \sum_k S_k \]  

where \( k \) represents different values of the forecast variable in a forecast probability function. \( S' \) is calculated for all pairs of predictors and corresponding dependent values in the validation data set. The average LEPS skill score \( (SK) \) is defined as

\[ SK = \frac{\sum_{j=1}^{n} 100S'_j}{\sum_{j=1}^{n} S_{mj}} \]  

where \( n \) is the number of pairs of predictors and dependent values in the data set. For each pair of observed and forecast values, if \( S' \) is positive, \( S_{mj} \) is the sum of the best possible forecast. The best

---

**Table 1**

Performance statistics of the forecasting results in case study I.

<table>
<thead>
<tr>
<th>Models</th>
<th>Calibration LEPS</th>
<th>SDR</th>
<th>Validation LEPS</th>
<th>SDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-NN</td>
<td>75.2 38.7 0.43</td>
<td>41.2</td>
<td>543.2 137.8 42.0</td>
<td>61.2</td>
</tr>
<tr>
<td>MLP</td>
<td>29.4 11.9 0.95</td>
<td>-</td>
<td>261.7 106.2 37.3</td>
<td>0.82</td>
</tr>
<tr>
<td>MLP–PLC</td>
<td>38.5 15.4 0.96</td>
<td>-</td>
<td>198.4 89.1 27.6</td>
<td>0.90</td>
</tr>
<tr>
<td>ANNE</td>
<td>25.7 11.3 0.98</td>
<td>56.3</td>
<td>154.3 55.8 16.5</td>
<td>0.96</td>
</tr>
</tbody>
</table>

---

**Fig. 3.** The results of point forecasts of Red-river peak flood discharge for validation data set using K-NN (a), MLP (b), MLP–PLC (c), and ANNE (d).
Table 2  
Performance statistics of low value forecasting in case study I.

<table>
<thead>
<tr>
<th>Models</th>
<th>Validation RMSE</th>
<th>%VE</th>
<th>CORR</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-NN</td>
<td>110.1</td>
<td>98.0</td>
<td>0.49</td>
</tr>
<tr>
<td>MLP</td>
<td>143.9</td>
<td>102.1</td>
<td>0.21</td>
</tr>
<tr>
<td>MLP-PLC</td>
<td>71.3</td>
<td>29.1</td>
<td>0.82</td>
</tr>
<tr>
<td>ANNE</td>
<td>37.1</td>
<td>21.7</td>
<td>0.96</td>
</tr>
</tbody>
</table>

possible forecast assumes $p_f = p_v$ for all pairs of observed and forecast values. If $S$ is negative, $S_{00}$ is the sum of the worst possible forecast. The worst possible forecast is calculated considering $p_f = 1$ and $p_v = 0$ (which gives the minimum $S$ equal to $-1$) when actual $p_v$ is less than 0.5 and is calculated considering $p_f = 0$ and $p_v = 1$ (which also gives the minimum $S$ equal to $-1$) when the actual $p_v$ is greater than or equal to 0.5. Considering these conditions, the formulation for determining $S_{00}$ is the same as determining $S$. The value of the LEPS score is a continuous number between 100 and $-100$. The LEPS score will take the value of 100 for a perfect forecast and is equal to $-100$ when all forecast values obtained from the probability function of the forecast have the maximum possible difference with the observed value (i.e., $p_f = 0$ and the $p_v$ values = 1, or $p_f = 1$ and the $p_v$ values = 0). According to Piechota et al. (2001), forecasts with a LEPS score greater than 10 represent a good forecast.

Table 3  
Performance statistics of high value forecasting in case study I.

<table>
<thead>
<tr>
<th>Models</th>
<th>Validation RMSE</th>
<th>%VE</th>
<th>CORR</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-NN</td>
<td>967.2</td>
<td>40.1</td>
<td>0.67</td>
</tr>
<tr>
<td>MLP</td>
<td>421.7</td>
<td>22.0</td>
<td>0.89</td>
</tr>
<tr>
<td>MLP-PLC</td>
<td>29.5</td>
<td>3.2</td>
<td>0.98</td>
</tr>
<tr>
<td>ANNE</td>
<td>192.8</td>
<td>11.7</td>
<td>0.97</td>
</tr>
</tbody>
</table>

As shown in Table 1, point estimation forecast by ANNE results in the minimum RMSE and %VE and maximum CORR in comparison with the other models. It is demonstrated in Table 1, ANNE which benefits from the conjunctive use of MLP and K-NN, outperforms both of them with a significant difference according to the obtained performance statistics. Performance statistics obtained by MLP–PLC are the most similar to what are obtained by ANNE. However, the overall performance of ANNE shows its supremacy over MLP–PLC, the detailed performance of these models should be compared by investigating their ability in extreme value forecasting as shown in Tables 2 and 3.

Table 2 presents the results of forecasting low flood discharge values obtained by different models. As it is demonstrated in Table 2, ANNE still outperforms other models. The results of forecasting high flood discharge values are shown in Table 3. MLP–PLC results in better forecasts in this category of observed values.

Fig. 4 shows the upper and lower bounds of forecast obtained by ANNE during the validation period. The average bandwidth of forecasts is +21% and $-11\%$. Five out of sixty forecasts are out of range of forecast values obtained by individual ANNs.

A predictive probability distribution function is presented to express the degree of certainty about the realization of a peak discharge. An exceedence probability forecast can be used depending on an assumed level of risk. For example, taking a 10% risk corresponds to a peak discharge that has a 90% probability of nonexceedence, which means that the probability of occurrence of a flood greater than the considered value is 10%.

As far as the probabilistic forecast is concern, ANNE outperforms other models as it produces forecasts with greater LEPS score and lower standard deviation for errors (Table 1). It is concluded by the results shown in Table 1, that MLP–PLC outperforms MLP in potential probabilistic forecasting as it results in residuals with less standard deviation.

As shown in Fig. 5, probabilistic forecasts are produced with different spreads. More reliable forecasts are produced with less estimation variance (spread). Hence the most reliable forecast from the three years is produced in 1995 and the most uncertain forecast is produced in 1997. According to the CDFs shown in Fig. 5, the peak discharge associated with 10% risk in years 1987, 1995, and 1997 are equal to 2520, 4100, and 2025 cms, respectively. The best estimate (associated with 50% probability) for these years are equal to 2255, 3987, and 1954 cms, respectively. In comparison with the observed flood peak discharge in these years, which are equal to 2329, 4597, and 1866 cms, it is concluded that better and more reliable forecasts are produced in years 1995 and 1987 with LEPS scores of 87.2 and 64.7.
respectively. The forecast values in these years are produced with less bias error and less forecast spread.

4. Case study II: seasonal streamflow forecasting

Zayandeh-rud River is the main surface resource for irrigation demands in the central part of Iran, especially the Isfahan metropolitan area. As water and energy demands increase in Isfahan, water withdrawals from the river increase and it is critical that climate variability is incorporated into water resources related decision-making. Zayandeh-rud reservoir with 1470 million cubic meters volume, controls streamflow upstream of Isfahan city (Fig. 6). The total annual average inflow to Zayandeh-rud reservoir is about 1600 million cubic meters of which an average annual flow of 600 million cubic meters is transferred from the adjacent Karoon River basin. Monthly inflow data to the Zayandeh-rud reservoir for the 30 year period from 1972 to 2002 are used in this

Fig. 5. Examples of CDF provided by ANNE in case study I.
Fig. 6. The map of Zayandeh-rud river basin and the Zayandeh-rud reservoir.

Fig. 7. The results of point forecasts of Zayandeh-rud streamflow for validation data set using K-NN (a), MLP (b), MLP–PLC (c), and ANNE (d).
study. Total streamflow from April to June (spring streamflow), which mainly results from the winter snow pack, are used as predicted value in this study. NAO index, which seem to be effective for predicting climate variations of the central and southern parts of Iran, is considered as a predictors of Zayandeh-rud River spring streamflow (Araghinejad et al., 2006). The North Atlantic Oscillation (NAO) involves a negative correlation in winter months between sea-level pressures in the subtropical Atlantic high and the Icelandic low. The North Atlantic Oscillation (NAO) index is the difference between normalized sea level pressure over the Azores and Iceland. The usual index is given by the December–March average of this measure (Jones et al., 1997). NAO indexes, winter snow budget as well as the winter streamflow are considered as predictors of total spring streamflow. Like the previous case study, K-NN, MLP, MLP–PLC, and the proposed method of ANN ensembles are used to forecast the predicted variable of this case study. These models are trained and calibrated using the same methods as discussed in case study I.

4.1. Results and discussion of case study II

Point forecasts of seasonal streamflow by K-NN, MLP, MLP–PLC, and ANNE are shown in Fig. 7 for 30 years cross validation data set of 1972–2002. Evaluation statistics of those forecasts for calibration and validation are shown in Table 4. As it is shown in Table 4, ANNE outperforms MLP–PLC and K-NN as it results in less RMSE and %VE and more CORR. In this case, the second best model is obtained as MLP. ANNE is a better model in probabilistic forecasting than K-NN as it results in greater LEPS score. Meanwhile, MLP outperforms MLP–PLC in over all probabilistic forecasting as it results in residuals with less standard deviation.

Table 5 presents the results of forecasting low streamflows obtained by different models. As it is demonstrated in Table 5, MLP–PLC outperforms other models. The results of forecasting high streamflow volumes are shown in Table 6. ANNE results in better forecasts in this category of observed values.

Fig. 8 shows the upper and lower bounds of forecast obtained by ANNE during the validation period. The average bandwidth of forecasts is $+14.2\%$ and $-11.3\%$. Six out of thirty forecasts are out of range of forecast values obtained by individual ANNs.

A predictive probability distribution function of years 1993, 2000, and 2002 are presented to express the degree of certainty about the realization of a seasonal streamflow of Zayandeh-rud in these years. The most reliable forecast among those three examples is produced in 1993 and the most uncertain forecast is produced in 2000. According to the CDFs shown in Fig. 9, the best estimates for these years are equal to 1596, 801, and 674 MCM, with LEPS scores of 75.6, 9.4, and 42.3; respectively. In comparison with the observed streamflow volume in these years, which are equal to 1719, 651, and 536 MCM, it is concluded that better and more reliable forecasts are produced in year 1993 according to the LEPS score.

![Figure 7](image.png)

**Fig. 8.** The bandwidth of forecasts provided by ANNE in case study II.
5. Summary and conclusion

Using the concept of artificial neural network ensembles, a procedure for forecasting hydrological variables in a probabilistic manner was proposed. The proposed method benefits from the training flexibility of artificial neural networks as well as the statistical technique of the nearest neighbor method. Individual networks are trained in which each one of them is biased to a specific quantile from the distribution function of the dependent variable. During real time forecasting, the outputs of individual networks are combined by the method of nearest neighborhood in a way that in the averaging process the networks with better performances during the hydroclimatological conditions similar to the current condition, are assigned by greater values of weights.

The method was applied into two contrasting case studies of flood peak discharge forecasting in Red river, Canada, and seasonal streamflow forecasting in Zayandeh-rud River, Iran. The results were compared with the three models of conventional ANN, $K$-NN regression models, and the networks trained by a specific performance function developed for extreme value forecasting. The results were compared in terms of performance statistics during calibration and validation periods. The results demonstrated that the proposed method outperforms other models in

![CDF examples](image-url)
point estimation forecasts. The overall comparison of the results of applying four models in terms of forecasting extreme and normal values showed the supremacy of the proposed method of ANN ensembles, however, the network trained by the PLC performance function was better in some case of extreme value forecasting. Furthermore, the ability of the proposed method in probabilistic forecasting was demonstrated.

An encouraging aspect of the proposed model is that it could be used to produce probability distribution function of dependent variables explicitly. One primary aim in this study was to present the conceptual procedure to show the applicability of the ensemble techniques and how it can be implemented for real situations of the probabilistic hydrological forecasting. Thus the stage is cast for more detailed studies to follow, including emphasizing on the extrapolation of data, and developing the application of single ANNs into the probabilistic forecasting, which is suggested for further study.

References


