Comparison of radial basic function (RBF) and STATISTICA in daily flow forecasting

Mahdi Moharrampour¹ and Mohammad Kherad Ranjbar²

¹ Department of civil Engineering, Islamic Azad University, buinzahra branch, Qazvin, Iran
² Same technical and vocational training college, Islamic Azad University, Karaj branch, Karaj, Iran
E-mail: kherad_m62@yahoo.com

*Corresponding Author: Mahdi Moharrampour
E-mail: m62.mahdi@yahoo.com

Abstract. Two decades with the advent of methods based on artificial intelligence and genetics. Algorithm directly based upon different parameters to predict water engineering is highly developed. Accurate prediction of flow in rivers, always as one of the most important factors in safe and economic design of facilities and structures related to river water has been considered by researchers. In this study, the method of radial basic function RBF and STATISTICA model, are used to forecast daily river flow in north of Iran and the results of these models are compared with Observed daily values. In this paper using information from the four Station hydrometer from the basin for 18 years, statistics from 1368 to 1386 and after normalization to 75% data for training and 25% data for testing were chosen. The results showed that the ability of the radial basic function RBF model output better than the STATISTICA software.

Introduction:
In recent years, the broader application of Computational Intelligence Computational Intelligence (CI) in solving problems that are very large impact parameters and some physical processes are highly nonlinear relationship between them is governed. Methods based on intelligent use of knowledge lies in the data, trying to extract the intrinsic relationship between them and replicate it in other situations. The most sophisticated methods include artificial neural networks (computing neurons), fuzzy logic (approximate conclusions) and genetic algorithms [1-4]. One of the newest and most dynamic areas of research that many researchers are currently in various scientific fields has attracted. On the other hand, one of the important issues in water resources management to the instantaneous flow rate of correct diagnosis, short and long term future and how Dubai is expected to be relying on the accurate prediction than the use of management planning to be done [5,6]. In this study, STATISTICA software was used too for the first time in order to predict the daily discharge [7,8]. This paper compares two expert models in daily flow forecasting, the RBF neural network and STATISTICA model, are used to forecast daily river flow in north of Iran and the results of these models are compared with Observed daily values Ghara-soo River is the case study and Ghara-soo river data is used for this article.

Case study area and data
Ghara-soo River basin is in Golestan province, northeast of Iran. This basin is located (54) to (54-45) east latitude and (36-36) to (36-59) north longitude. Basin area is 1678.1 km2. Maximum height of this basin is 3200 meters from sea level and the length of Main River is 108.005 km. Fig. 1 shows the natural plan and location of Ghara-soo River.
More than 4 rain measurement stations are existed over this river, but because of lack of statistics for all stations, in this research 4 stations are used. Gharasoo station as exit discharge of this basin and Ziarat, Shastkalateh and Kordkooy as input of this basin in three different locations. (Table 1)

**Table 1:** specification of Gharasoo basin stations

<table>
<thead>
<tr>
<th>Province</th>
<th>Code</th>
<th>Location</th>
<th>River</th>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golestan</td>
<td>12-050</td>
<td>Gharasoo</td>
<td>Gharasoo</td>
<td>54-03-00</td>
<td>36-50-00</td>
</tr>
<tr>
<td>Golestan</td>
<td>12-043</td>
<td>Naharkhoran</td>
<td>Ziarat</td>
<td>54-28-00</td>
<td>36-46-00</td>
</tr>
<tr>
<td>Golestan</td>
<td>12-045</td>
<td>Shastkalate</td>
<td>Shastkalate</td>
<td>54-20-00</td>
<td>36-45-00</td>
</tr>
<tr>
<td>Golestan</td>
<td>12-049</td>
<td>Ghaz mahalle(pole jadde)</td>
<td>Kordkooy</td>
<td>54-05-00</td>
<td>36-47-00</td>
</tr>
</tbody>
</table>

**Preprocessing data**

Preprocessing of data includes selection of effective variable, selection of training and test patterns and normalizing the patterns [9, 10]. The goal of normalizing is that all values in one pattern are in a range [11]. Pattern normalizing exchanges all values to a specified interval such as [0-1] or [-1-1].

After normalizing all patterns, period of case study was selected between 1989 to 2007 (18 years). For this period, there are 6550 daily patterns for every station. 75% of these data are used for support vector machine training and 25% of these data are used for test. Fig. 2 shows daily flow hydrograph of Gharasoo River for training period and Fig. 3 shows daily flow hydrograph of Gharasoo River for test period.
Designed and developed simulation models using radial basic function neural network software MATLAB

Or Radial Basis Function RBF neural networks are networks with a hidden layer. Gaussian functions is to stimulate them. A neuron has a Gaussian starts with education and training in each iteration a neuron is added to the network. In each iteration a Gaussian added to the network structure and the data is spread. In default on any of the data is Gaussian. I like the fuzzy input space with the Gaussian software, we are. But to achieve the proper answer to this problem and prevent punctures broad Gaussian over fitness level on data from $10^{-n}$ to increase the amount of output produced increases in the neighboring. The wide experience of over fitness to prevent punctures in the education system for highly trained) Because each data set shows a Gaussian (It does not respond well to test data. The RBF network and MLP has only one layer parameters can be changed in the maximum number of neurons is 5 that we start to reach a suitable solution, we increase. In all simulations, 75% of completely random data is used for training And the remaining 25% is set aside for network testing.

In the methods of teaching - oriented, the number and type of model input parameters is important. Since the structure of the neural network input, there is a constant and uniform, the results can be presented in the articles will help. Accordingly, the following five input pattern has been studied:
\[Q(t) = f\{P_g(t), P_n(t), P_{sh}(t), P_p(t), P_g(t-1), P_n(t-1), P_{sh}(t-1), P_p(t-1)\}\]

\[Q(t) = f\{Q_n(t), Q_n(t-1), Q_p(t), Q(t-1), Q(t-2)\}\]

\[Q(t) = f\{P_g(t), P_n(t), P_{sh}(t), P_p(t), P_g(t-1), P_n(t-1), P_{sh}(t-1), P_p(t-1), Q(t-1), Q(t-2)\}\]

\[Q(t) = f\{Q(t-1), Q(t-2)\}\]

\[Q(t) = f\{P_g(t), P_n(t), P_{sh}(t), P_p(t), P_g(t-1), P_n(t-1), P_{sh}(t-1), P_p(t-1), P_g(t-1)\}\]

In the above Relations:

- \(Q\): The average daily discharge stations Ghara Soo
- \(Q_n\): Daily average discharge station dining
- \(Q_p\): Daily average discharge of Station Road Bridge
- \(P_n\): The average daily rainfall station dining
- \(P_{sh}\): The average daily rainfall stations sixty Kalate
- \(P_p\): The average daily rainfall stations, bridges, roads
- \(P_g\): The average daily rainfall stations Ghara Soo

The other patterns were built near each of their results with one of the top models have not gained, and here the expression patterns.

Here is the question raised is which of the five models in the best performance will determine the daily discharge? To answer this question, the input patterns are evaluated. In the present study to compare the performance of models of the correlation coefficient and RMSE R2 is used.

### Table 2: Summary of results of the RBF neural network performance

<table>
<thead>
<tr>
<th>Input pattern</th>
<th>A pattern</th>
<th>B pattern</th>
<th>C pattern</th>
<th>D pattern</th>
<th>E pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of neurons in the input layer</td>
<td>14</td>
<td>6</td>
<td>10</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>The number of neurons in the hidden layer</td>
<td>15</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>R2</td>
<td>0.97503</td>
<td>0.97141</td>
<td>0.97521</td>
<td>0.97538</td>
<td>0.85502</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.025595</td>
<td>0.027442</td>
<td>0.025142</td>
<td>0.025495</td>
<td>0.061629</td>
</tr>
</tbody>
</table>

What the tables (2) can be concluded that the model is trained to be able to predict the proper discharge of the river that day. input pattern with the second quarter as the input parameters (1-10-2) and (5) Output network training and testing data shows.
The answer would be better compared with the zoom on the graph (Figure 6). Data is high because the number of graphs can be seen a little sloppy and messy. The diagram in Figure 7 below you can see how to reduce the network training error:

**Figure 4:** Comparison of model outputs in the test phase (first model).

**Figure 5:** Comparison of model outputs in the training phase (first model).

**Figure 6:** Comparison of model outputs to focus on the test (first model).
Predicting the course of the river by using software STATISTICA

By using the data related to the structure of discharge proposed exit point with the software and STATISTICA had predicted the results with the results of the structure proposed by comparison, SVM and just as in the table (3) is determined by the structure proposed by STATISTICA software performance as a result.

Table 3: Results of STATISTICA function

<table>
<thead>
<tr>
<th>ways</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATISTICA</td>
<td>0/095025</td>
</tr>
</tbody>
</table>

Figure 8 and 9 shows the observed discharge and predicted discharge with STATISTICA software. Figure 10 shows the results of both observed and predicted discharge. It shows that predicted maximum discharge is lower than observed discharge [2].

**Table 3: Results of STATISTICA function**

<table>
<thead>
<tr>
<th>ways</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATISTICA</td>
<td>0/095025</td>
</tr>
</tbody>
</table>

**Observed**

Figure 8 Daily discharge hydrograph of Gharasoo Station (Observed)
Fig. 9 Daily discharge hydrograph of Gharasoo Station (Predicted)

Conclusion
1. 5 input patterns presented, patterns involving Nagorno Station, discharge before and two days before hand, the results were acceptable offer.
2. What the tables (2) can be concluded that the model is trained for the fourth input pattern 2 as input parameters (1-10-2) has More accurate forecasts than other models for daily river discharge do. The result is that a lot of inputs do not necessarily mean it is not anticipated.
3. The prediction of river flow, rainfall and discharge days before and two days before the day of the Station, a major role in the Nagorno-hand plays, so the results can be seen as. According to what mentioned above, we can say that the radial basic function (RBF) is successful than software STATISTICA.

6. References


11/24/2012