EFFECT OF ENSO ON BLACK SEA STREAMFLOWS

Ali Ihsan Marti¹, Humeyra Bilge²

¹Selcuk University, Engineering Faculty, Department of Civil Engineering, Konya, Turkey
²Civil Engineer, M.Sc. Student, Selcuk University, Konya, Turkey

Abstract

The climates of the countries on the world were sometimes considerably affected from the oceanic-atmospheric events that ENSO (El Nino Southern Oscillation) is one of the most significant one changing the hydrologic parameters and triggering the climate changes on the world. Hence, the objective of this study was to determine whether the streamflow data of Turkey was affected from the ENSO events or not by considering 20 streamflow gauging stations existing at the Black Sea coasts of Turkey. For this purpose, synthetic monthly streamflow data corresponding to the El Nino years of the time series were generated using Feed Forward Back Propagation Artificial Neural Network model and replaced with the original data of the considered time series. Then, two streamflow data sets for each station (synthetic and original) were obtained in order to make comparison between their statistical parameters (variance, mean, population, autocorrelation) to find out whether the synthetic and the original data differ from each other or not. As a result, significant El Nino effects were determined for the streamflow data of Turkey's Black Sea Region and the surroundings especially in terms of variance, autocorrelation and population parameters which should be considered for long term drinking water, irrigation, energy and environmental planning purposes. The results of this study can be extended for all the countries surrounding the Black Sea to obtain the regions affected from ENSO more precisely in the subsequent studies.

Key words

Southern Oscillation; Turkey; Black Sea; streamflow; ENSO


*Corresponding author: Tel.: +90-530-4580981, Fax: +90-332-2410635
E-mail address: alihsan@selcuk.edu.tr
1 INTRODUCTION

El Nino Southern Oscillation (ENSO), the irregular, intensive and extensive ocean-atmosphere phenomenon occurring due to the changes in the usual atmospheric pressure patterns and in the sea surface temperature at some parts of the Pacific Ocean, affects the climatic characteristics of the world in terms of temperature, rainfall, streamflow, evaporation, etc. El Nino (warm phase) and La Nina (cold phase) are two contrary phases of ENSO [1]. Despite the irregular repetition, El Nino occurs approximately every four years on average. Streamflow is the most important hydrologic variable used for reliable forecasting and analysis of the hydrometeorological variables on the world in order to execute accurate water resources planning and operating processes of the countries. Redmond and Koch [2] and Kahya and Dracup [3,4] studied the regional and global hydrologic and climatologic influences of ENSO events [5].

In recent years, widespread uses of estimation processes for hydrometeorological parameters by using artificial neural network (ANN) methods were encountered frequently. ANN is a black box model that is successfully applied for the behavior estimation of the non-linear systems and the variables whose interrelations are not exactly known. Therefore, since ANN models have easy practicability and not requiring more data features, they present alternative solutions against the numerical estimation models [5].

ENSO influences on hydrologic and climatologic variables in global and regional scales have been extensively studied in the literature. The most comprehensive global-scale studies were carried out by Ropelewski and Halpert [6] using the data from over 2000 rainfall stations worldwide. Redmond and Koch [2], Kahya and Dracup [3], Dracup and Kahya [4], Maurer, Lettenmaier et al. [7], Twine et al. [8], and Gobena and Gan [9] illustrated the streamflow variability in the North America and its relationship to ENSO occurrences. For streamflow variable, Dettinger et al. [10] studied multi-scale streamflow variability in relation to ENSO events using over 700 stations worldwide. Similarly Chiew and McMahon [11] investigated the global ENSO–runoff teleconnections using data from 581 catchments. Additionally, the precipitation data of Sao Paolo were estimated by Ramirez et al. [12] using ANN techniques where ANN was stated as the best method in comparison to the multi-linear regression method. An ANN model was used by Hall [13] for determining the daily rainfall probability in a two-year time series in America, Kulligowski and Baros [14] estimated a 6-hour rainfall data in two basins using ANN model, Shrestha and Kostaschuk [15] examined the impacts of ENSO on mean-monthly streamflow variability in Nepal and found that ENSO-related below normal streamflows in two core regions, Sen et al. [16] proposed ENSO templates that can be used for streamflow prediction, Freiwan and Cigizoglu [17] made estimations for the monthly precipitation data of Jordan by applying Feed Forward Back Propagation ANN model. It is probable that the ENSO-streamflow relationship is more noticeable than the ENSO-rainfall relationship for the reason that precipitation variability is higher than that in streamflow due to the fact that streamflow integrates information spatially. [1] [18]

The objective of this study was to determine whether the warm phase of ENSO, i.e. El Nino, affects the streamflow records of the Black Sea Region of Turkey or not. For this aim, new streamflow records for the past El Nino years were generated using the model of Feed Forward Back Propagation Artificial Neural Network (FFBPANN) in order to analyze the possible differences between the generated and the historical streamflow data by using statistical tests [5]. As a result, two data sets were obtained that one was the original historical data set, and the other was the historical data set whose monthly streamflow values corresponding to the El Nino years were replaced with the synthetic streamflow data. Then,
the comparison of the historical and synthetic data was performed in terms of mean, population, variance and autocorrelation coefficient parameters to determine the probable existence of any ENSO effect on the Black Sea Region streamflows. [18]

2 MATERIAL

This study’s data network was formed from 20 streamflow gauging stations uniformly distributed in the Black Sea Region of Turkey (Figure 1). The streamflow data set spans from 1962 to 2000. The selection of the stations of that study has been made considering the data distributing around Turkey homogeneously, not having any missing records. No interference should have occurred on streamflow, and there should not be constructed any structures inside the basin of the station which will damage the homogeneity of the data. Thinking of other possible homogeneity-damaging factors that could not be detected, the Standard Normal Homogeneity Test (SNHT) and Pettitt Test were used to determine the homogeneity of the selected stations and the non-homogeneous ones were not used in the subsequent calculations. The homogeneous stations used in Karabork and Kahya [19] were also used in this study without making any additional homogeneity tests [5]. In the simulation phase of our analysis, we considered the El Nino years as stated in Table 1.

![The Black Sea Region](image)

Fig. 1: The Black Sea Region (the region inside the bold line) studied in this study [20]

Tab. 1: El Nino Years [21] [22]

<table>
<thead>
<tr>
<th>El Niño Years</th>
</tr>
</thead>
</table>

3 METHODS

At first, considering the non-occurrence of ENSO events in the historical records, the original monthly streamflow values corresponding to El Nino years were removed and replaced by the estimated monthly streamflow values generated by Feed Forward Back Propagation Artificial Neural Network model (FFBPANN) [5].
The first two years prior to the ENSO event year which were supposed to have high correlations with the values of the estimated year were taken as the input variables of the model. Furthermore, the preceding third and fourth data values were also considered as the extra input variables of the FFBPANN model. In order to increase the performance of the model, the mean of each month was calculated without taking the data corresponding to the El Nino years in the data set. The data generation with the FFBPANN model was performed using the MATLAB computer program [23].

After the completion of the synthetic data generation, the comparison of the generated data with the original data should be done. Therefore, Scenario A and Scenario B defined in the following were formed:

Scenario A - Includes two different 39 years long time series; a) the original annual streamflow data, ii) the synthetic annual streamflow data (the values of the original data remained same except the data of the ENSO event years replaced with the generated ones using FFBPANN ).

Scenario B - Includes two different streamflow time series for ENSO event years a) the original streamflow data for 10 ENSO event years, b) the synthetic streamflow data generated with FFBPANN model for 10 ENSO event years [23].

Feed Forward Back Propagation Artificial Neural Network (FFBPANN): Artificial neural networks can be considered as the black box generating outputs corresponding to the input variables introduced into the models. The ability of human brain in solving difficult operations and understanding complex examples, especially sometimes learning some information without knowing any physical relationships but only trying, has inspired the scientists to develop artificial neural network methods [24]. Artificial neural network (ANN) methods produce unique solution systems by training themselves using the input variables entered to the models. Feed Forward Back Propagation Artificial Neural Network method is the mostly used ANN method in generating synthetic data. It has three different layers involving input, hidden and output layers connected to each other and each formed from many neurons.

FFBPANN method has two steps. First one is the feed forward step transferring the external input information of the input cells to the forward in order to calculate the output information signal of the output layer. Second one is the back propagation step making alterations on the connection forces depending on the differences between the calculated information signals of the output layer and the observed information signals [24]. At the beginning of a training period, the connection forces are assigned randomly, and then the learning algorithm changes the force until the training is completed successfully in the iterations. The input layer has many neurons while the output layer has only one neuron. \( X_1, \ldots, X_k \) are the input values of the input neurons. The input values are multiplied with the first intermediate connection weights \( w_{ij} \) in hidden neurons where \( j = 1, \ldots, h \), and the results are collected through the \( i \) index and become the inputs of the hidden layers [23].

\[
H_j = \sum_{i=1}^{k} w_{ij}x_i 
\]

Here, \( H_j \) is the input of the hidden neuron, \( w_{ij} \) is the connection weight from neuron \( i \) to neuron \( j \). Each hidden neuron produces a hidden neuron output \( H_O_j \) by the aid of a sigmoid function. \( H_O_j \) is defined as in the following:
\[ H_{o_j} = f(H_j) = \frac{1}{1 + \exp[-(H_j + \theta_j)]} \]  

(2)

\( H_{o_j} \) output becomes the input of the subsequent layer’s input. The input reaching the output neurons is calculated using the following equation [23].

\[ I_{O_n} = \sum_{j=1}^{h} w_{jn} H_{O_j} \quad n = 1, \ldots, m \]  

(3)

The neural network output values are obtained by treating these input values using the previously defined sigmoid function. The subsequent weight arrangement or the learning period will be provided using the back propagation algorithm. The output values will not be the same of the destination values. Average square error is calculated between the output and destination values. The goal of back propagation algorithm is to minimize the average square error by iteration. Therefore, the input weights are updated distributing the error gradients calculated by the average square error, and the process starts again and repeated until the average square error is minimized to an optimum level or a pre-specified iteration number is reached [24] [23].

Now a test procedure including the comparison of variances, means, autocorrelation coefficients and populations of the aforementioned data sets will be applied to compare the generated synthetic data with the original historical data.

Variances: The \( F \)-test was used to test the variances of the original and hypothetical series for all testing scenarios. \( F \) statistics can be determined using Eq. 4;

\[ F = \frac{s_1^2}{s_2^2} \quad (s_1^2 > s_2^2) \]  

(4)

Here, \( s \) is the standard deviation and \( s^2 \) is the variance. After calculating the F value, the value from \( F \)-distribution and the calculated F value were compared with each other, and if the calculated value exceeds the critical value obtained from the \( F \)-distribution table for the required significance level, then the difference between the variances will be accepted as significant [25].

Means: \( t \)-test was used for the comparison of the mean values of the two data sets [25]. The data set length was 39 for Scenario A and 10 for Scenario B as they would be same for both historical and synthetic data sets \((n_1 = n_2)\). If a positive correlation was not determined between the samples, the following \( t \)-model would be used [26]

\[ t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sum x_1^2}{n_1} - \frac{\sum x_2^2}{n_2} - 2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \]  

(5)

Here \( n_1, n_2 \) are the sample lengths, \( \bar{X} \) is the sample mean. However, the following formula should be used in case of positive correlation.
Populations: After the data samples were tested in terms of variances and means, the non-parametric Mann-Whitney U test was applied to the couple of original and synthetic data sets for both scenarios to understand whether they are from the same populations or not. The U-statistics of the Mann Whitney Test is defined as in the following [27];

\[ U = \min(U_1, U_2) \]  
\[ U_1 = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - R_1 \]  
\[ U_2 = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - R_2 \]

Here, \( n_1, n_2 \) values are the sample lengths. Both data series were taken together and arranged from the lowest to the highest value by assigning ranks to each member of the data sets. \( R_1 \) and \( R_2 \) values were calculated as a result of the summation of the ranks of the first and second data sets separately. Since U-statistics asymptotically follows the normal distribution, the expected value and the standard deviation of the calculated U-statistics are expressed in Eqs. 10 and 11;

\[ \mu_U = \frac{n_1 n_2}{2} \]  
\[ \sigma_u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \]

and the standard normal variable will be computed as in the following;

\[ z = \frac{U - \mu_U}{\sigma_U} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \]

The calculated standard normal variable will be evaluated using the two-tailed test to understand whether any significant difference between two data samples exists or not [5].

Autocorrelations: The original and synthetic streamflow series were examined to see whether ENSO events influence the autocorrelation structure of the observed series or not. The existence of statistically significant positive lag-1 autocorrelation coefficient \( (r_1) \) is the important indication of the persistence characteristics of a streamflow time series. Therefore, we assume that the \( r_1 \) is a representative of the autocorrelation structure of the original and synthetic streamflow series and can be expressed as in Eq.13 [25].
Here, k=1, \( \bar{x}_t \) is the mean of the first N-k terms and \( \bar{x}_{t+k} \) is the mean of the last N-k terms. The confidence intervals at the 95% significance level can be determined by Eq.14 [25].

\[
r_k(\%95) = \frac{-1 \pm \sqrt{N-k-1}}{N-k}
\]

### 4 RESULTS AND DISCUSSION

The estimation of the parameters related to hydroclimatology has great importance in planning and operating water resources and structures. In this context, the streamflow data generation was performed in this study using Feed Forward Back Propagation Artificial Neural Network model and the generated synthetic streamflow data was compared with the original historical data in terms of variances, means, autocorrelation coefficients and populations of both data sets for each streamflow gauging station of the Black Sea Region of Turkey. The comparison of the aforementioned parameters was carried out considering 90% and 95% statistically significant levels [18].

The comparison of the original and synthetic data variances was performed by the F-test, and the results given in Table 2 for the stations in the Black Sea Region of Turkey presented only one significant variance difference (in station 2218) between the two series according to Scenario A. However, for Scenario B, the variance differences between two series considerably occurred with 95% significance level for nearly half of the stations.

Using the t-test for the comparison of the means of the original and synthetic series, considerably less significant values for both scenarios were obtained (Table 2) that only in two stations (101, 1314) the mean values presented differences. However, this result can not be generalized for the whole Black Sea Region.

Mann Whitney U-test was applied for the comparison of the original and synthetic data set populations to understand whether both of the time series are the members of the same population or not? And the results are tabulated in Table 2. However, both scenarios A and B did not present any statistically significant difference between the original and generated streamflow data sets for all the stations of the Black Sea Region.

The possible changes that can occur in the autocorrelation structure of the streamflow data of the Black Sea Region due the tropical El Niño events were analyzed using the equations 13 and 14 to compare the autocorrelation structures based on the lag-1 correlation of the synthetic and original series, and the results are tabulated in Table 2. It is known that the first autocorrelation coefficient (\( r_1 \)) is the most important component of the autocorrelation structure of the hydrologic time series [1].

**Case I:** If the calculated \( r_1 \) values of both synthetic and historical series do not fall within the 95% confidence intervals, both series are said to have no ENSO effect, i.e. no serial dependency property at lag-1 [18]. For Scenario A there are eleven stations (1302, 1307, 1314, 1335, 1401, 1402, 1418, 1524, 2218, 2232, 2304) presenting this behavior. And for
Scenario B there are seven stations (1237, 1307, 1314, 1335, 1402, 2218, 2232) conforming this conclusion.

Case 2: If the calculated r\textsubscript{1} value is significant for the generated synthetic series, but not for the historical series, then it will be understood that ENSO events can change the autocorrelation structure of the streamflow series in an eliminating form. This case is observed in three stations (2213, 2233, 2323) for Scenario A and in 2 stations (1524, 2213) for Scenario B. For these stations, it can be concluded that El Nino events damage the autocorrelation structure of the streamflow data that is expected to be important. Actually, this outcome is a normal outcome for the El Nino events that are seasonal anomalies [18].

Table 2: Brief tabulation of statistical comparisons between historical and generated streamflow data using Feed Forward Back Propagation Artificial Neural Network (FFBPANN)

<table>
<thead>
<tr>
<th>Station Number</th>
<th>Variance Scenario A</th>
<th>Variance Scenario B</th>
<th>Mean Scenario A</th>
<th>Mean Scenario B</th>
<th>Mann Whitney Scenario A</th>
<th>Mann Whitney Scenario B</th>
<th>Autocorrelation Scenario A</th>
<th>Autocorrelation Scenario B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Obs. ANN</td>
<td>Obs. ANN</td>
</tr>
<tr>
<td>101</td>
<td>■</td>
<td>■</td>
<td></td>
<td></td>
<td>+■</td>
<td>-■</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1221</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td>-■</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1237</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1243</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td>-■</td>
<td>+■</td>
<td>+■</td>
</tr>
<tr>
<td>1302</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td>-■</td>
</tr>
<tr>
<td>1307</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1314</td>
<td>■</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1335</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1401</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1402</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1413</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td>-■</td>
<td></td>
<td>+■</td>
</tr>
<tr>
<td>1418</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td>-■</td>
</tr>
<tr>
<td>1524</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td>-■</td>
<td></td>
<td>+■</td>
</tr>
<tr>
<td>1528</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td>-■</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2213</td>
<td>■</td>
<td>□</td>
<td></td>
<td></td>
<td>+■</td>
<td>□</td>
<td>-■</td>
<td></td>
</tr>
<tr>
<td>2218</td>
<td>■</td>
<td>□</td>
<td></td>
<td></td>
<td>+■</td>
<td>□</td>
<td>+■</td>
<td>□</td>
</tr>
<tr>
<td>2232</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td>-■</td>
<td></td>
<td>+■</td>
</tr>
<tr>
<td>2233</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td>+■</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2304</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-■</td>
<td></td>
</tr>
<tr>
<td>2323</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+■</td>
<td>-■</td>
</tr>
</tbody>
</table>

□ significant at 90 % level  ■ significant at 95 % level

Case 3: If the calculated r\textsubscript{1} value is significant for the original series but not for the generated synthetic series, our perception will be similar to Case 2 that ENSO events can change the autocorrelation structure of the streamflow series in an eliminating form [1]. This case is observed in one station (1237) for Scenario A and 7 stations (1221, 1302, 1401, 1413, 1418, 2233, 2304) for Scenario B. For these stations, it can be said that ENSO events change some characteristics of the autocorrelation structure of the streamflow series by increasing the internal dependency.

Case 4: If the r\textsubscript{1} values computed for both series exceeds the 95% confidence intervals, our perception will be “ENSO events have no influences on the autocorrelation structure of the streamflow series unless the opposite sign of the computed significant r\textsubscript{1} values appears”. For Scenario A, five stations (101, 1221, 1243, 1413, 1528), and for scenario B two stations (101,
presented this behavior. However, we had one station (2323) having opposite signs for $r_1$ values which proves the ENSO effect on the autocorrelation structure of the streamflow series of this station. Additionally, if both series have no difference, it can be concluded that El Nino events have no influence on the original series [18].

5 CONCLUSION

Consequently, considerable number of streamflow stations presented significant differences between the synthetic and historical time series when the variances and the autocorrelation coefficients of the series are considered. It is an important outcome that the identified scenarios A and B were distinctive in reflecting the ENSO modulation on streamflow variance among all stations in the entire study. Then, Scenario B presented more significant variance differences than that of Scenario A.

The masked El Nino effects among the 39-year time series can be explained by the non-occurrence of population differences for Scenario A in all stations. Although, there were also differences in terms of population, however the differentiation could not reach to the lowest significant level of 90%. Moreover, it can be stated that the ENSO events caused modification in the persistency characteristic of Turkish streamflow patterns, and artificial neural network models are good alternatives for the processes related to synthetic data generation where physical forecasting models can not be applied.

Since the data generated in this study using FFBPANN model and the data generated in Kahya and Marti [1] using Radial Based Artificial Neural Network (RBANN) presented similar results that both ANN methods can be used for streamflow data generation where other forecasting models do not work.

As a conclusion, all results from the above statistical tests performed in this study showed that El Niño events have influences on the streamflow values of Turkey’s Black Sea Region. However, since only the Black Sea Coasts of Turkey was investigated in this study, more evident influences of ENSO events can be observed and determined when the same analysis is applied to the streamflow time series of all Black Sea countries (Georgia, Russia, Romania, Bulgaria, Ukraine, etc.). Furthermore, this kind of studies applied on streamflow, precipitation or evaporation data provides a solid incentive for the water resources planners to consider the large-scale atmospheric oscillation patterns like the North Atlantic and/or Southern Oscillations for planning and managing purposes.

ACKNOWLEDGEMENT

This paper was supported by The Coordinating Office of Scientific Research Projects of Selcuk University (BAP – Selcuk University).

REFERENCES


