HYDROLOGICAL VARIABILITY OF THE
EUROPEAN PART OF TURKEY*

H. AKSOY **
Dept. of Civil Engineering, Istanbul Technical University, Hydraulics Division 34469 Maslak, Istanbul, Turkey

Abstract– The european part of Turkey, the Thrace region, is of agricultural and industrial importance for the country and hence irrigational and industrial water demand drastically increases, whereas most of the rivers in the region have no flow in summer. Two gauged hydrological basins are studied in order to analyze the following streamflow characteristics of the region: Randomness, jump, trend, and best-fit probability distribution. Selected gauging stations are found with random annual streamflow data for a 40 year observation period (1961-2000). No jump is observed in the annual mean streamflow time series of the rivers; however, the presence of insignificant negative trends is indicated by the trend test. Results for the probability distribution goodness-of-fit test show gamma distribution to be the best fit distribution for the rivers.

Keywords– Thrace region, Turkey, streamflow, randomness, jump, trend, best-fit probability distribution

1. INTRODUCTION

The design, planning and management of water-related structures are based on streamflow data that should be analyzed to determine the statistical characteristics of the streamflow time series as well as physiographical characteristics of the hydrological basin. Therefore preliminary streamflow data analysis plays an important role in hydrology and water resources studies. Insufficient information of the observed data increases uncertainties [1] in the design of water resources structures such as the spillway on a dam [2]. In hydrological practice, it is unusual to have long enough streamflow data. Therefore, the data set in hand is simulated in order to extract information such as a 100-year return period event [3, 4].

Structural characteristics consist of consistency (homogeneity), trend, randomness, and jump-type discontinuities as well as the probability law that the data set obeys. The time series is called consistent if the elements of the series come from the same population, which also means that no jump is observed throughout the observation period. A trend-free time series has neither negative nor positive tendencies. If there is no intervention by humans on the streamflow data, then it is expected that the time series is random. Persistency, quantified in the serial correlation, shows the effect of the subsequent elements in the time series and hence makes the series non-random.

In literature, a number of studies are devoted to the analysis of the hydrological time series. For instance, Hubert [5] developed a software that divides a given time series into as many subseries as possible; all differences between the mean values of two consecutive subseries are significant. Fanta et al. [6] investigated the variability of annual streamflow data of southern Africa by using 502 gauging stations in the region. The methodology consists of the Fisher test to determine the randomness of the series; a simple linear trend test to detect the presence of a trend; and the segmentation procedure of Hubert [5] to find the points, if any, where the time series has a jump. In a more recent study [7], preliminary

*Received by the editors December 3, 2005; final revised form September 17, 2006.
streamflow analysis prior to water resources planning was done by means of the double mass curve to check whether the streamflow is consistent; by means of the Spearman Rank Order Correlation nonparametric test to check whether the time series has a trend; by means of a nonparametric run test to check whether the time series is random; and by means of the probability plot correlation coefficient (PPCC) test to determine what probability distribution is the best fit.

In this study, a combination of tests mentioned above was used in order to determine the streamflow characteristics of two hydrological gauging stations selected from river basins in the European part of Turkey. In the following sections, structural characteristics of hydrological data (streamflow in this study) are given together with tests employed for the analysis of the data. The study area and data are then defined. Results obtained are discussed and conclusions are finally given.

2. TESTING STREAMFLOW CHARACTERISTICS

a) Randomness

Randomness in a hydrological time series means that the data result from natural causes. If there is no randomness, then there is persistency which is typically quantified by a serial correlation coefficient. An adopted version of a simple nonparametric run test [7] was used in this study. The test consists of the following steps:

a. The median of the observations is determined.
b. Each of the data items are examined whether or not it exceeds the median. If a data item exceeds the median, this is defined as a case of success, \( S \), otherwise a case of failure, \( F \).
c. The number of successes and failures are counted and denoted by \( n_1 \) and \( n_2 \), respectively.
d. The total number of runs (\( R \)) in the data set is determined. A run is a continuous sequence of successes until it is interrupted by a failure or vice versa.
e. The test statistics is computed by

\[
z = \frac{R - \left( \frac{2n_1n_2}{n_1 + n_2} - 1 \right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}}
\]

where \( z \) has a standard normal distribution under the null hypothesis, \( H_0 \), that the sequence of successes and failures is random.
f. Critical values of the standard normal distribution are obtained for the chosen significance level, \( \alpha \), and denoted by \( \pm z_{\alpha/2} \).
g. The computed statistics \( z \) is compared to the critical values \( \pm z_{\alpha/2} \). \( H_0 \) is rejected if \( z < -z_{\alpha/2} \) or \( z > z_{\alpha/2} \).

b) Jump

The automatic segmentation procedure [5] breaks the time series into as many segments with significantly different means as possible. The procedure is based on the minimization of the segmentation cost defined as the sum of squares of deviations from the average of each segment. If two or more segments are identified, the starting year of the last segment was chosen as the first year for splitting the time series. A comparison was then made between the segments before and after the chosen year. Once segmentation is completed the jump analysis was performed, for which details are given below.

a. The time series is divided into several segments by the automatic segmentation procedure [5].
b. The mean of two consecutive segments ($\bar{y}_1$ and $\bar{y}_2$) are calculated and the length of the segments ($N_1$ and $N_2$) are determined.

c. The t-statistics is calculated by

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

with $N_1 + N_2 - 2$ degrees of freedom. $s$ in Eq. (2) is the pooled variance given by

$$s = \sqrt{\frac{\sum_{i=1}^{N_1} (y_{i1} - \bar{y}_1)^2 + \sum_{j=1}^{N_2} (y_{j2} - \bar{y}_2)^2}{N_1 + N_2 - 2}}$$

\[ (3) \]

d. The null hypothesis, that shift in the mean is insignificant, is rejected if the sample t statistics in Eq. (2) is greater than the critical value of Student’s t-distribution ($t_{1-\alpha/2, N_1 + N_2 - 2}$) with $N_1 + N_2 - 2$ degrees of freedom.

c) Trend

The presence of a trend in a hydrological time series can be considered an indication of possible long-term changes such as climate change. The environment from which the data are taken out may be affected by human impact as well as natural phenomena. Climate change (global temperature increase) can be considered the most important among others. It can be expected that the significant increase in droughts and floods will become more common and even more severe as a result of climate change.

The Spearman Rank Order Correlation nonparametric test was used to investigate the existence of a trend that might be found in the time series. A step by step explanation of the test is as follows for a time series $y_i$, $i = 1, ..., n$, observed in time $i$:

a. Ranks, $R_{yi}$, are assigned to $y_i$, such that the rank 1 is assigned to the largest $y_i$ and the rank $n$ to the least $y_i$. Ties in the $y_i$ are handled by assigning each of the tied observations the mean rank of those rank positions that they occupy.

b. The difference

$$d_i = R_{yi} - i$$

is computed.

c. The coefficient of the trend, $r_s$, is computed by

$$r_s = \frac{1 - \frac{6}{n} \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

Under the null hypothesis that the time series has no trend, the variable

$$t = r_s \sqrt{\frac{n - 2}{1 - r_s^2}}$$

has a Student’s t-distribution with $n - 2$ degrees of freedom.

d. The critical values of the t-distribution for the chosen significance level, $\alpha$, and $n - 2$ degrees of freedom are obtained. For a two-tailed test, the critical values are denoted by $\pm t_{\alpha/2,n-2}$.

e. The values of $t$ are compared to the critical values. $H_0$ is rejected if $t > t_{\alpha/2,n-2}$ or $t < -t_{\alpha/2,n-2}$. 
**d) Best-fit distribution**

Several commonly used methods such as the Chi-squared and Kolmogorov-Smirnov tests are available for testing the goodness-of-fit of theoretical probability distribution functions to the hydrological time series data. In this study, the Probability Plot Correlation Coefficient (PPCC) test was used. This test is based on the correlation coefficient between the ordered sample $y_i$ ($y_1 \leq y_2 \leq \ldots \leq y_n$) and their corresponding fitted quantiles, $w_i = G^{-1}(1 - p_i)$, where $p_i$, $i = 1, 2, \ldots, n$, is the exceedance probability of $y_i$ and $G^{-1}(\bullet)$ is the inverse of the cumulative distribution function of the theoretical probability distribution considered for the sample. The test has the following steps:

a. The sample is ordered from the smallest to the largest.

b. The exceedance probability, $p_i$, $i = 1, 2, \ldots, n$, of each data item in the ordered sample is calculated by

$$p_i = \frac{R_{pi} - 0.4}{n + 0.2}$$  \hspace{1cm} (7)

c. The standard normal variate at $p_i$, $z_{pi}$ is calculated by

$$z_{pi} = \frac{(1 - p_i)^{0.135} - p_i^{0.135}}{0.1975}$$  \hspace{1cm} (8)

d. For normal distribution, for instance, quantiles corresponding to the data items in the sample are calculated by

$$w_i = \bar{y} + S_y z_{pi}$$  \hspace{1cm} (9)

Equation (9) changes depending on what probability distribution is selected.

e. The correlation between $y_i$ and $w_i$ is calculated by

$$r = \frac{1}{n - 1} \sum_{i=1}^{n} \frac{y_i - \bar{y}}{S_y} \frac{w_i - \bar{w}}{S_w}$$  \hspace{1cm} (10)

where $\bar{y}$ and $\bar{w}$ are the means, and $S_y$ and $S_w$ the standard deviations of $y$ and $w$, respectively.

### 3. STUDY AREA

The study area is the European part of Turkey, the Thrace region (Fig. 1) located between 40°-42° N and 26°-29° E. The region is bordered by the Black Sea to the North, the Marmara Sea and Aegean Sea to the South, Greece to the West and Bulgaria to the Northwest. This region is of agricultural and industrial importance for the country and hence irrigational and industrial water demand drastically increases during summer, in particular when most of the rivers in the region have no flow.

The geological structure in northwestern Turkey is a triangular-shaped tertiary sedimentary basin formed by extension in the late Middle Eocene to the latest Oligocene times [9]. The topography of the region is mainly composed of the Strandja Mountains lying parallel to the Black sea with the highest peak (1018 m above mean sea level) close to the Bulgarian border. In the south (the coast of the Marmara Sea), there are single peaks such as the Tekir Mountain and the Isiklar Mountain. Between these mountainous areas is the Ergene river basin with tributaries flowing towards the south from the north and to the north from the south.

Land in this region is used for farming, mainly wheat, sunflower, and rice. Additionally, sugar cane, sesame, corn, onion, garlic, bean, watermelon, melon, and zucchini are produced. A dry farming system without fallowing is practiced with wheat-sunflower rotation [10]. Not only dry farming, but also irrigational farming is popular in the region.
The region is characterized by a modified Mediterranean type of climate with influences from the Black Sea maritime and Balkan continental effects. Consequently, winters are cold and summer seasons are rather warm with long sunshine duration and high relative humidity. It can be considered a snow area when compared to the rest of the country, especially as storms under the Balkan continental effect leave snow over the region. Long term annual mean in the region is about 13.5 °C. The average rainfall is about 600 mm with 35% in winter, 25% in spring, 27% in autumn, and 13% in summer. Annual relative humidity in the region is 73%.

The region was studied widely. Intermittency is an important feature of the rivers in the region [11, 12]. Statistical and hydrological analysis of the low flow regime of the region was performed [13].
recession curve is a tool used in low flow analysis. Therefore seasonal variation in the recession curve of
daily streamflow records in the region was investigated [14, 15]. Baseflow index is another low flow
indicator [16]. A probabilistic approach [17] was developed for the investigation of the recession part of
the daily streamflow hydrograph and was applied to the data set given in Table 1. A trend analysis study
[18] covering all hydrological basins in Turkey has been completed with the results comparable to those
obtained in this study. Statistical type analysis is a common practice in hydrological studies not only for
dry conditions, but also for high water levels in the stream [19].

Table 1. Characteristics of the river basins studied

<table>
<thead>
<tr>
<th>River (Station number)</th>
<th>Ergene (105)</th>
<th>Hayrabolu (106)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drainage area (km²)</td>
<td>10194.8</td>
<td>1381.2</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>10</td>
<td>45</td>
</tr>
<tr>
<td>Record period (years)</td>
<td>1961-2000 (40)</td>
<td>1961-2000 (40)</td>
</tr>
<tr>
<td>Average flow (m³/s)</td>
<td>27.49</td>
<td>4.93</td>
</tr>
<tr>
<td>Standard deviation (m³/s)</td>
<td>16.17</td>
<td>3.65</td>
</tr>
<tr>
<td>Coefficient of variation</td>
<td>0.588</td>
<td>0.740</td>
</tr>
<tr>
<td>Coefficient of skewness</td>
<td>0.722</td>
<td>0.967</td>
</tr>
<tr>
<td>Serial correlation coefficient</td>
<td>0.251</td>
<td>0.298</td>
</tr>
<tr>
<td>Annual flow depth (mm)</td>
<td>71.9</td>
<td>92.0</td>
</tr>
<tr>
<td>Zero flow (%)</td>
<td>5.00 *</td>
<td>5.48 *</td>
</tr>
</tbody>
</table>

* Based on the observed data (not extended data)

4. DATA

This study is based on results obtained through an international project [20] in which, in order to obtain
comparable results for all participating countries, a constraint in the size of a hydrological basin was set.
According to this, homogeneous hydrological basins at a size of not less than 1000 km² and not bigger
than 10000 km² were selected. Two stations (105 and 106 operated by Electrical Power Resources, Survey
and Development Administration, EIEI, of Turkey) were found eligible to select (Fig. 1). The analysis,
here, is only based on these two stations. The rivers on which the stations are installed are in the Ergene
river basin, the most important river in the region. The Hayrabolu River (106) is a tributary of Ergene. The
gauging station 105 is on the Ergene River itself. The Ergene River joins Meric (Evros, Maritsa), a
boundary river between Turkey and Greece with an upstream area in Bulgaria (Fig. 1). The region has
perennial main streams with intermittent tributaries. Intermittency [11, 12], important characteristics of the
rivers in the region, is due to the low ground water table level. However, intermittency is not considered
completely natural due to recent water abstractions from the streams and groundwater for farming and
industry. Based on two maps (one dated 1967 and the other showing the current situation) obtained from
the State Hydraulics Works (DSI) of Turkey, it was seen that groundwater was densely extracted in the
region after the mid 1980s in particular. This resulted in substantial drops in the groundwater level. For
instance, the groundwater level which measured (from the aquifer base) 125 m in 1967 was dropped to
about 60 m at the current situation.

The minimum daily streamflow discharges of the rivers consist of zero flows (because of their
intermittent structure) as well as those higher than zero. The zero flow percentages calculated by using the
daily discharges are 5% and 5.48% for stations 105 and 106, respectively. Floods have discharges as high
as 75 to 150 times of annual mean streamflow discharges for the Ergene and Hayrabolu rivers,
respectively. Seasonality is another particular feature of the streams in the region with a high amount of
water discharges throughout the wet seasons and low (even zero) flows during dry weather conditions
(e.g. summer season).
As mentioned above, this study is extracted from a project on the effect of climate change on streamflow variability in rivers in the Balkan Peninsula [20, 21]. The duration of the observed data is 26 years (from 1969 to 1994) for 105, and 34 years (from 1969 to 2002) for 106. As the data set does not cover the basic period, 1961-1990 the popular climatological baseline period of the World Meteorological Organization [22], the data set was extended beyond the record available by simple regression analysis. The regression is based on the correlation between the stations 105 and 101 (Seytan river a third hydrometric station in the region with a longer record period, but not eligible to be included in the study because of its smaller basin size), 106 and 101, and 105 and 106. High linear dependence was found between the stations as expected, because they are all very close to each other and under the same hydrological regime (Fig. 2). Backward data extension (1961-1968) for stations 105 and 106 was performed using the correlation between each of these stations (105 and 106) with station 101. Forward data extension (1995-2000), however, was done by using the dependence structure between the stations 105 and 106 as we already have up to date data for 106 (2002).

\[ y = 3.9825x + 7.3771 \]
\[ R^2 = 0.8095 \]

\[ y = 2.84x - 2.0522 \]
\[ R^2 = 0.6686 \]

\[ y = 13.882x - 6.1872 \]
\[ R^2 = 0.8558 \]

Fig. 2. Linear regressions between stations used for data extension,
(a) 105 and 106, (b) 106 and 101, (c) 105 and 101

The time series of the extended annual mean streamflow data are given in Fig. 3, and the histograms in Fig. 4. Based on the histograms, it is obvious that the streamflow data set is not normally distributed because of the tail (positive skew). Table 1 illustrates the long term characteristics of the rivers based on the extended data.

5. RESULTS

The above detailed tests were employed for the two hydrological basins defined previously. A summary of the results obtained is given below, together with the physical interpretations. Test results are based on the extended time series in Fig. 3. In a previous study [23], the same analysis was performed for the observed data of 26 years (1969-1994) only. Having these results in hand made a comparison to the results in this study available.
Fig. 3. The extended time series of hydrometric stations selected for the study

Fig. 4. Histograms of the extended time series

**a) Randomness test**

Results of the randomness test are presented in Table 2 in which the number of successes and failures ($n_1$ and $n_2$, respectively), total number of runs ($R$), and of the standard normal statistics and their critical value for the selected significance level, 5%, are given. In Table 2, the decision based on the results of the test is also given. It is then concluded that the time series of annual mean streamflow data of both river basins are random.

Table 2. Results of the randomness test

<table>
<thead>
<tr>
<th>River</th>
<th>$n_1$</th>
<th>$n_2$</th>
<th>$R$</th>
<th>$z$</th>
<th>$z_{cr}$ ($\alpha = 0.05$)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ergene</td>
<td>20</td>
<td>20</td>
<td>19</td>
<td>-0.641</td>
<td>-1.960</td>
<td>RANDOM</td>
</tr>
<tr>
<td>Hayrabolu</td>
<td>20</td>
<td>20</td>
<td>19</td>
<td>-0.641</td>
<td>-1.960</td>
<td>RANDOM</td>
</tr>
</tbody>
</table>

An increase in the time interval reduces the correlation coefficient between the subsequent flow values. For instance, the correlation coefficients of 0.838 and 0.592 calculated for the daily average flows of the Ergene and Hayrabolu Rivers, respectively, were reduced, as given in Table 1, to 0.251 and 0.298 for the annual mean flow. Low values of the annual lag-one correlation coefficients give support to the conclusion on the randomness of the annual mean flow data of both river basins. Extending data sets made no difference in their character related to the randomness as the same results were obtained for both stations in the previous study with historical data only [23].

**b) Jump test**

The automatic segmentation procedure [5] was run first. It was found that both river basins could be divided into two segments (Fig. 5). Table 3 summarizes the results of the jump analysis in which the first and last years of each segment were given. Additionally, the average and length of each segment ($\bar{Y}_1$, $\bar{Y}_2$, and $N_1$, $N_2$) are presented. In Table 3 the variance, $s$, of the data set, defined in Eq. (3), the test statistics and then critical value and the result based on the comparison of the test statistics and the critical value are also given. It is seen that the Ergene river has a negative jump in 1983 and Hayrabolu river has a positive one in 1998, both insignificant. The negative jump in the Ergene River was, however, found significant
while only the historical data was used [23]. The insignificant positive jump in the Hayrabolu River was shifted back to 1987 at the downward direction for the historical data and found insignificant [23]. The reason for this change can be explained by the use of longer streamflow data with a maximum in 1999 that shifted the jump to 1998 and turned it upward for the longer data. It should also be mentioned that the backward data extension has an effect on the location of the jumps as the segmentation procedure is based on the minimization of the differences with the average of each segment.

When both time series are eye-inspected, a drop is seen in the mid 1980s. Indeed, as mentioned before, analysis of the historical data alone [23] has already pointed out negative jumps (though the one in Hayrabolu was insignificant) during this period. This is the beginning of a dry period when precipitation in the region decreases and falls down below the long-term average and lasts for more than ten years. At the same time, it is the period when the extensive use of groundwater for industrial and agricultural purposes in the region starts. Intermittent and perennial streams in the region become drier with lower amounts of precipitation and a lessening groundwater level. The region however is under the effect of heavy rains that naturally result in high level flows in the stream course, even floods. This causes damage to properties, agricultural lands and even loss of lives. Therefore, the negative jumps eye-inspected and determined in the time series can easily be interrupted. The maximum flow observed in the Hayrabolu River for the year 1999 is such an example.

c) Trend test

Table 4 summarizes the results of the trend test. The Spearman Rank Order Correlation coefficient, $r_s$, and the test statistics and their critical value for 5% of the significance level are given in Table 4. Based on the comparison of these statistics, it is concluded that both data sets have insignificant negative trends. These results are shown in Fig. 6. Trend analysis with the historical data [23] resulted in significant negative trends in both stations. The effect of the length of the data sets is obvious in these contradictory results.
d) Goodness-of-fit test for probability distribution

Histograms in Fig. 4 and the statistics in Table 1 indicate that the streamflow records in the selected river basins have non-normal distributions with positive tails. The PPCC test was employed for six probability distribution functions listed in Table 5. The probability distribution function with the highest correlation coefficient is selected as the best-fit probability distribution. As indicated in Table 5, both data sets were found to best fit by gamma distribution, while it was the three-parameter log-normal distribution for the historical data [23]. However, it should be noted that both the gamma and the three-parameter log-normal distributions are positively skewed and have a better fit than the normal distribution to the data sets.

<table>
<thead>
<tr>
<th>Table 5. Results of the PPCC test</th>
</tr>
</thead>
<tbody>
<tr>
<td>River</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Ergene</td>
</tr>
<tr>
<td>Hayrabolu</td>
</tr>
</tbody>
</table>

6. CONCLUSION

This study reports the analysis of streamflow data taken from the European part of Turkey, the Thrace region. Two hydrometric stations (Ergene and Hayrabolu) from the Ergene river basin were used for the hydrological analysis. Data available for the hydrometric stations starts from 1969 to recent years (1994 for Ergene and 2002 for Hayrabolu). Data sets were extended backward to 1961 (for both stations) and forward to 2000 (for Ergene only). Although the number of stations used is limited to only two; results may highlight the characteristics of streamflow records and possible changes in the region.

Analysis of streamflow data consists of checking whether or not the data set is random, homogeneous, and trend free. The best-fit probability distribution function is also searched. Results show that both hydrometric stations are with a random streamflow regime that implies no persistency between the streamflow data of the subsequent years. However, at the same time, the regime is under the influence of both natural interventions such as dry weather conditions and man-made changes such as increasing water abstraction from the rivers and the groundwater system for irrigation and industry, and dam construction for different kinds of purposes. The results point out insignificant trends in the time series. How these trends might be linked to climate change or variability factors is a continuing effort through an international cooperation [20] in which negative trends, for the Balkan Peninsula, were described in precipitation and streamflow with expected higher variability than before, which might be a signal for more severe extremes in the future. Jumps determined by the segmentation procedure are found...
insignificant. The annual mean streamflow in both stations are found far from the normal distribution, but to obey gamma distribution. Results obtained by using a shorter streamflow data consisting of only observations highlight the importance of the length of the time series used in such analysis.

This study is station-based. A regional analysis can be considered once a homogenous hydrological region is defined.

Acknowledgement- This study is partly based on the results of the project “Assessment of the climate change impact on the elements of the hydrological cycle” led by Dr. Snejana Dakova, of the Bulgarian Academy of Sciences, of whom I am thankful for her collaboration. My sincere thanks go to Dr. Pierre Hubert of Ecole de Mines de Paris, France for sharing his software to use in the jump analysis, and to the referees and Dr. Mahmood Yaghoubi, the editor for their comments.

REFERENCES


