Spatial grouping of annual streamflow patterns in Turkey

Ercan Kahya[1]
Istanbul Technical University, Civil Engineering Department, Hydraulic Division, 34469 Maslak Istanbul, Turkey

M. Cüneyd Demirel
Istanbul Technical University, Institute of Science and Technology, 34469 Maslak Istanbul, Turkey

Thomas C. Piechota
University of Nevada, Las Vegas, Department of Civil and Environmental Engineering, 4505 Maryland Parkway, Box 454015, Las Vegas, NV 89054-4105, USA

Abstract. This paper describes a procedure grouping streamflow patterns across Turkey that exhibit similar annual flow behaviors. Streamflow information is the integrated response of the river basin, such as topography, soil and vegetation to external impacts such as climate. A regionalization of annual data is carried out using k-means analysis. Streamflow series from 80 gauging stations were chosen to characterize geographic differences in year to year streamflow variability. The eight-cluster solutions of k-means analysis are stable and interpretable with reference to spatial and temporal variations. It was shown that western Turkey is dominated by a relatively large cluster, indicating strong influences of the Mediterranean dynamics. The resultant map was given at the 8-cluster level, showing the spatial variability of homogeneous streamflow regions.

1. Introduction

Streamflow data is used in comparative hydrology and design of structures built along stream channels, in evaluating flood hazards, and in defining the available water supply for the region. The main climatic factors of precipitation, temperature, sunshine, humidity, soil moisture, topography and wind all affect discharge that make the deterministic approaches for quantifying it more complicated (Riggs, 1985; Haines et al., 1988).

Turkey is located in a semiarid zone where water is limited and a scarce resource since precipitation is limited to rainfall and snow. The hydrological characteristics of the country represent high spatial and temporal variability. Comparative hydrology is used to explore broad-scale characteristics of hydrological processes leading the interaction between atmosphere and land surface. Stahl (2001) studied drought across Europe by correlating the monthly

[1] Professor. Hydraulic Division
Civil Engineering Department
Istanbul Technical University
34469 Maslak Istanbul, Turkey
Tel: (212) 285-3002
e-mail: kahyae@itu.edu.tr

averages of the Regional Streamflow Deficiency Index (RDI) series of the 19 large clusters to the NAO indexes and found weak correlations. The Europe domain has a strong seasonal regime; therefore, seasonal variability was important to assess the impact of climate changes on this complex hydrological system (Stahl, 2001). Dettinger and Diaz (2000) focused on the monthly streamflow series in global scale and indicated that the timing and amplitude of seasonality in streamflow depends on the local month of maximum precipitation and the extent to which precipitation is trapped in snow and ice at most gauges. Demirel (2004) tested three different standardization techniques to determine the finest data preprocessing solution for a robust clustering. The standardization by range performed better, and Wards clustering method was recommended for other studies (Demirel, 2004).

This work advances the pluviometric characterization of Turkey streamflow, and better knowledge of its regional streamflow patterns. The structure of the paper starts with a description of the study area, dataset used and the details of the methodology employed (Section 2, 3 and 4 respectively). In Section 5, the emergent annual flow patterns are presented by using a new approach of k-means analysis which was not previously applied by the authors (i.e., Demirel, 2004; Kahya and Demirel, 2007).

2. Description of study area

The study area covers the entire country and extends from 26-45° of longitude east and 36-42° of latitude north (Figure 1). The spatial distributions of 80 streamflow gauging stations with continuous recording are not quite uniform. Monthly streamflow records compiled by EIE (General Directorate of Electrical Power Resources Survey and Development Administration) were shown to satisfy the homogeneity condition at a desirable confidence by Kahya and Karabörk (2001).

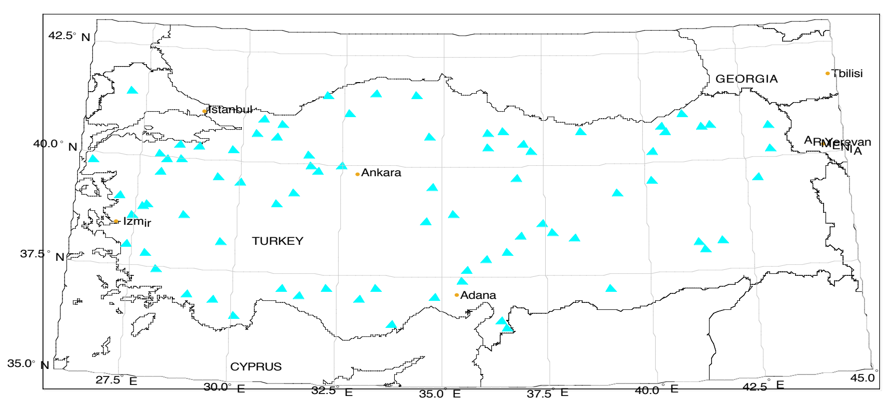

Figure 1. The streamflow stations used in the analysis.

3. Data

Annual mean discharge values from 80 stations with unimpaired station characteristics were used to develop k-means analysis. The stations located on streams where flows are regulated or where diversions significantly affect flows were deleted from a network of 257 gauging stations by using high resolution (100000:1 or 25000:1) maps of Turkey. The records extracted for this study span from 1964 to 1994.

The original data was not well separated in the Eigen space hence standardization was applied prior to analysis (Figure 2). Standardization of the data was not only necessary to achieve physically meaningful classifications but it was also a useful precursor to the application of k-means analysis (Demirel, 2004).

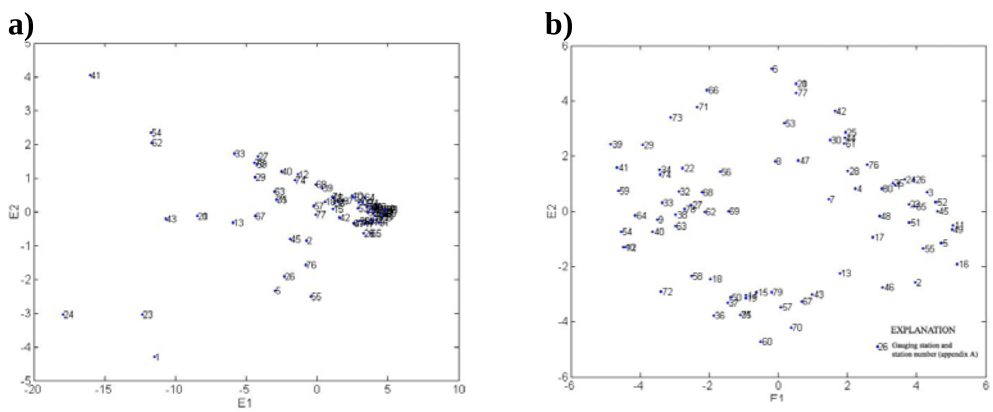

Figure 2. (a) Annual streamflow data dispersion by multidimensional scaling, **(b)** the same but for standardized annual streamflow data, adapted from Demirel (2004).

4. Methods

For streamflow S_{it} for station i in year t, the streamflow index, Z_{it} was computed by the following standardization by the range equation

$$Z_{it} = \frac{S_{it}}{\max(S_{it}) - \min(S_{it})} \qquad \text{Eq. (1)}$$

where $\bar{S}_i$ is the average, σ_i the standard deviation of all measurements at station: for i = 1,2, …, 80 and t =1,2, …, 31.

The reduction of the data by another multivariate analysis method potentially loses some relevant information; therefore, principal component analysis (PCA) or factor analysis was not applied in this study (Arabie et al., 1996). However it is observed that the first two principal components are explaining more than 80 % of the variance (Figure 3).

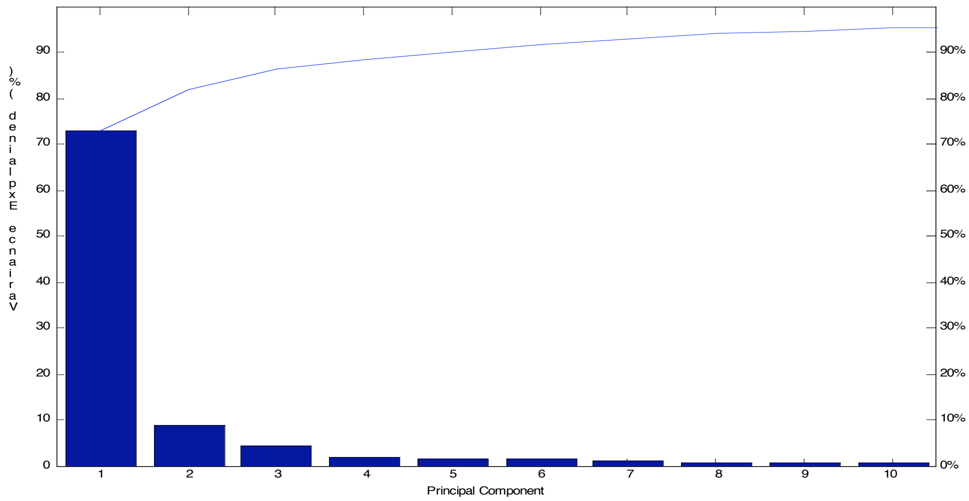

Figure 3. Principal components of the data.

For the spatial grouping, the nonhierarchical k-means algorithm using re-allocation criteria (called as *exact assignment test* in the literature), as implemented in the Matlab, was used (Url-2; Bacher, 2002). The squared Euclidean distance was taken as the similarity index. A hierarchical tree plot generated by the Average Linkage method was also considered as reference for deciding how many clusters to create. However, the number of clusters is an issue and prone to subjectivity. Therefore it was, in most cases, dictated by the researcher's experience. It was decided among the others that the solution of 8 clusters was satisfactory for annual streamflow. These clusters shown on a map are presented and discussed in next section.

The k-means analysis is used to identify hydrologically homogeneous zones showing similar patterns or behavior so that the hydrological effects can be compared within these regions. Hydrologic predictions and transferring information from one area to another with analogous characteristics will become possible if a robust scheme of regionalization is established (Andrade, 1997). Standardization by range is used due to its superior efficiency in streamflow studies (Demirel, 2004).

The steps in the k-means algorithm (Url-1) were as follows:

1. Define the K points into the data space represented that will be clustered. These points are the initial group centres.

2. Assign each point (station) to the cluster that has the closest centre.

3. When all points have been assigned to one cluster, reorder the positions of the K centres.

4. Repeat Steps 2 and 3 until the centres no longer move. This produces a separation of the stations into clusters from which the metric to be minimized can be calculated.

Finally, this algorithm aims to minimize the *objective function*, in this case a squared error function (MacQueen, 1967). The objective function can be stated as (Url-1):

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad \text{Eq. (2)}$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure (herein squared Euclidean) between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centres.

5. K-means Analysis at Annual Domain

We applied the k-means method which defines the centers at 80 cluster level and stops iteration at a minimum total sum of difference. Apart from hierarchical methods, k-means does not provide a dendrogram at the end which allow the user to define the clustering levels. In Figure 4, the blue region is the largest cluster covering most of the entire western costal zone indicating Mediterranean dynamics in Turkey. Mediterranean rainfalls very often have a torrential character and the region is a transition zone between the mid-latitude low pressures and the subtropical highs as a result of being between the latitudes 36 and 44 degree North (Romero *et al.* 1999). Since Turkey is located at latitude 36–42°N, there a significant pluviometric difference between the north and south of the basin, and more importantly, this imposes strong seasonal differences (e.g., dry summers and wet winters) (Romero *et al.* 1999).

The Black Sea coast is considered the most complicated zone having wet conditions and varying streamflow characteristics (State Institute of Statistics, 1995). The northwestern Turkish coast is steep and rocky, and rivers cascade through gorges of the coastal ranges. Access to inland is limited because of few narrow valleys in the Black Sea region. As a result, the coast has always been isolated. The differences in streamflow patterns across the three regions in the far eastern parts of the Black Sea coastline might be explained through the topographical perspective although rainfall shows almost uniform distribution in the small red cluster in the east Black Sea region (Figure 4).

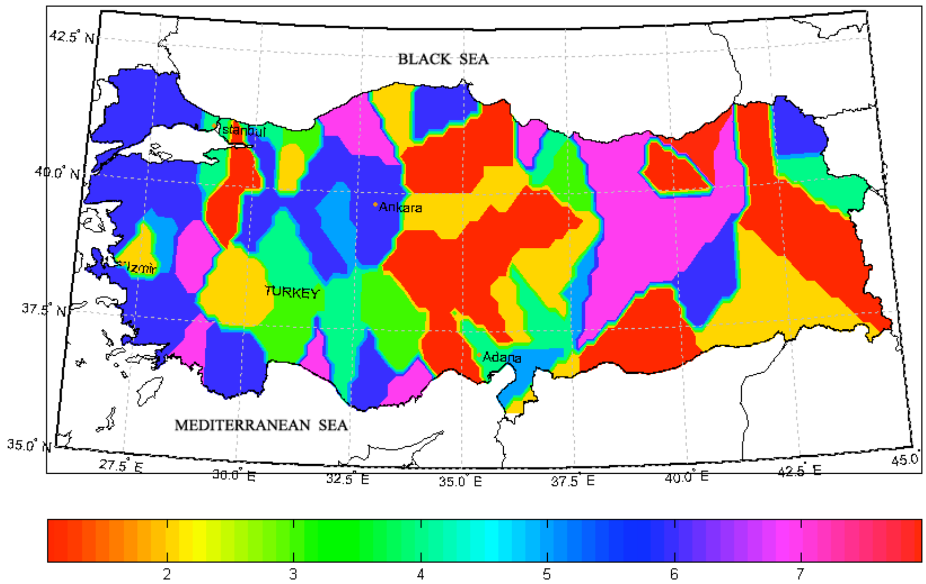

Figure 4. K-means analysis of annual mean streamflow.

The recent studies by Kahya and Karabörk (2001) and Kahya and Kalaycı (2004) documented a mapped streamflow patterns in Turkey, showing the Southern Oscillation extreme phases related regions and the detectable significant trend related regions, respectively. Both studies resulted in fewer regions than that in Figure 4. There are, however, some similarities with the climate regions redefined by Ünal et al. (2003).

The silhouette diagram (Figure 5) using the cluster indices output from k-means analysis was plotted to get an idea of how well-separated the resulting clusters are (Url-2). From this diagram, some of the points appearing in the second cluster have a silhouette value greater than 0.9, indicating that cluster is well separated from the neighboring clusters. However, 2 points is too low to define an acceptable streamflow grouping scheme in Turkey. The third and fifth clusters contain many points with low silhouette values and also some negative values, indicating that those two clusters are not well separated. Eventually we have come to a conclusion that the solution of 8-cluster has more or less uniform silhouette values and adequate magnitude to cover the spatial differences in streamflow characteristics.

Southern Turkey is mainly dominated by karsts and rises sharply from the sea level to the elevations up to 2800 meters. Mountains in this region prevent the Mediterranean influences from extending inland, leaving the interior of Turkey to experience continental climate conditions with distinct seasons. A grouping of eight regions is quite useful for our study, because it helps recognizing smaller regions in this country having highly variable climatic and topographic conditions. With this technique, we can go further in regrouping in high order clusters to obtain even smaller streamflow regions that satisfy local and micro level studies.

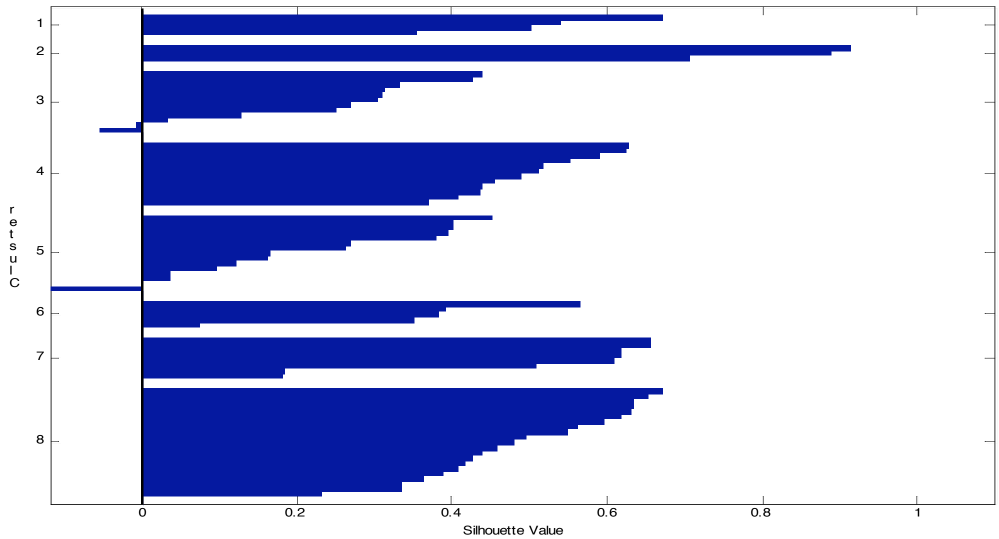

Figure 5. Silhouette diagram.

6. Summary

The results of this study provided insights into the regionalization of annual streamflow patterns in Turkey. We applied the analysis of k-means which forces all individual stations to be allocated in identifiable clusters or pattern groups. As a result, 8 pattern groups have been determined for annual streamflow. Although only k-means method was used here, a hybrid approach, such as the PCA and k-means coupling, is also recommended for future similar grouping studies to see whether any improvement upon the indications of the present study would be obtained using a different approach, or not.

Acknowledgements. This research is supported by Istanbul Technical University Research Activities Secretariat (PN# 30695).

References

Andrade E. M. de., 1997: Regionalization of average annual runoff models for ungaged watersheds in arid and semiarid regions. *Ph.D. Thesis*, University of Arizona, Arizona.

Arabie, D., L. J. Hubert, and G. De Soete, 1996: *Clustering and classification*. World Scientific Publ., River Edge, NJ.

Bacher, J., 2002: *Cluster analysis, lecture notes*, Nuremberg, 2002.

Demirel, M.C., 2004: Cluster Analysis of Streamflow Data over Turkey. *M.Sc. Thesis*, Istanbul Technical University, Istanbul.

Dettinger, M.D., and H.F. Diaz, 2000: Global characteristics of streamflow and variability. *Journal of Hydromoteorology*, **1**. 289-310.

Kahya, E., and M.Ç. Karabörk, 2001: The analysis of El Nino and La Nina signals in streamflows of Turkey. *International Journal of Climatology*, **21**, 1231-1250.

Kahya, E., and S. Kalaycı, 2004: Trend analysis of streamflow in Turkey. *Journal of Hydrology*, **289**, 128-144.

Kahya, E., and M.C. Demirel, 2007: Evaluation of multivariate statistical methods for characterizing annual streamflow regimes in Turkey. EGU General Assembly, April 15-20, 2007 Vienna, Austria (accepted).

Haines, A.T., B.L. Finlayson, and T.A. McMahon, 1988: A global classification of river regimes. *Applied Geography*, **8**, 255-272.

MacQueen, J. B., 1967: Some methods for classification and analysis of multivariate observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability"*, Berkeley, University of California Press, **1**, 281-297.

Riggs H.C., 1985: Streamflow characteristics. Elsevier, Amsterdam, 1985.

Romero, R., C. Ramis, and J.A. Guijarro, 1999: Daily rainfall patterns in the Spanish Mediterranean area: an objective classification. *International Journal of Climatology*,**19**, 95-112.

Stahl, K., 2001: Hydrological drought: a study across Europe. *Ph.D. Thesis*, Albert-Ludwigs-Universität, Freiburg.

State Institute of Statistics,1995: *Environmental Statistics & River Basin Statistics*, Publication No: 2152, pp.120, Ankara.

Url-1<*http://www.elet.polimi.it/upload/matteucc/Clustering/tutorial_html/ kmeans.html*>, accessed at 10.01.2007.

Url-2 <http://www.mathworks.com/>, accessed at 19.01.2007.

Ünal, Y., T. Kindap, M. Karaca, 2003: Redefining of climate zones of Turkey. *International Journal of Climatology,* **23**, 1045-1055.