Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi
Journal of Agricultural Faculty of Gaziosmanpasa University
http://ziraatdergi.gop.edu.tr/

Araştırma Makalesi/*Research Article*

JAFAG
ISSN: 1300-2910
E-ISSN: 2147-8848
(2020) 37 (3), 109-114
doi:10.13002/jafag4699

Impact of Change Point Presence in Extreme Rainfall Series on Hydrological Design

Kadri YÜREKLİ[1*]
[1]Tokat Gaziosmanpasa University, Agriculture Faculty, Department of Biosystem Engineering, Tokat
(orcid.org/0000-0003-4938-663X)
*e-mail: kadriyurekli@yahoo.com

Alındığı tarih (Received): 12.06.2020 Kabul tarihi (Accepted): 30.08.2020
Online Baskı tarihi (Printed Online): 09.09.2020 Yazılı baskı tarihi (Printed): 30.12.2020

Abstract: The study was conducted to on analyze the variations in maximum rainfall data sequences and the impact of change on these datasets through frequency analysis. This goal was achieved with Pettitt test considered in the detection of change-point and, with Generalized Extreme Value (GEV) distribution for frequency analysis. Based on the GEV distribution, quantile values at different return periods were estimated for two data sets before and after the change-point for each site. The couple of quantile at each return period had a significant differences. It was emphasized that it was imperative to eliminate the change in data in obtaining reliable data required in the design of hydraulic projects. When the double-mass curve method was applied to data with a change point, it was confirmed that a reliable data would be able to be reached.

Key Words: Maximum rainfall, change-point, Pettitt test, GEV distribution.

Hidrolojik Tasarımda Ekstrem Yağmur Serilerindeki Değişim Noktasının Etkisi

Öz: Çalışma, maksimum yağış veri dizilerinin değişiminin analizi ve bu veri setlerindeki değişimin frekans analizleri üzerindeki etkisi üzerine yürütülmüştür. Bu hedefe, değişim noktasının saptanmasında dikkate alınan Pettitt testi ve frekans analizi için Genelleştirilmiş Ekstrem Değer (GEV) dağılımı ile ulaşılmıştır. GEV dağılımına dayanarak, farklı tekrarlanma periyotlarındaki yağmur değerleri her istasyon için değişim noktasından önceki ve sonraki veriler için tahmin edildi. Her tekrarlanma periyodundaki yağmur çiftleri önemli farka sahip oldu. Hidrolik projelerin tasarımında gerekli olan güvenilir verinin temininde, verideki değişimin giderilmesinin zorunlu olduğu vurgulanmıştır. Değişim noktasına sahip bir veriye çift birikimli eğri yöntemi uygulandığında güvenilir bir veriye ulaşılabileceği teyit edildi.

Anahtar Kelimeler: Maksimum yağmur, değişim noktası, Pettitt test, GEV dağılımı.

1. Introduction

Heavy rains caused by climate change, occurring in most parts of the world, led to serious damages on environment, agriculture, economy and communication infrastructure, and most importantly result in loss of life. It is possible to eliminate or minimize the damage caused by heavy rains with hydrological and hydraulic engineering applications. In the context of engineering services, it is imperative to predict what the future amounts of the extreme rainfalls would be. Since hydrological events occur under the influence of many factors together, probability distributions are used to estimate the probable amount of a hydrological variable in the future. However, before the frequency analysis of any hydrological data, the existing data must be suitable for the analysis. In other words, the data to be analyzed should not change during the observation period. The hydraulic structure to be planned based on the design value to be estimated from the data that has variability over time will not provide the expected benefit.

Parametric or non-parametric statistical approaches are used to determine the presence of a change in hydrological data. However, since hydrological data often have a skewed distribution characteristic, that is, it does not satisfy the condition of normality in most cases, nonparametric approaches should be preferred in the analysis of homogeneity. Yurekli (2015) and Yuksel (2019) in their study applied the non-parametric Mann-Whitney U test to rainfall data sequences. Mahmood Agha et al. (2017) employed the Standard Normal Homogeneity (SNH), Buishand (BR), Von Neumann (VNR), and Pettitt tests for annual and seasonal rainfall data. Wijngaard et al. (2003) also used the same tests preferred in Agha et al. (2017) for the daily rain and temperature data. Kazemzadeh and Malekian (2018) analyzed the homogeneity of streamflow series by using SNH, BR, VNR, Pettitt, and Cumulative Deviation (CD) tests. Sahin and Cigizoglu (2010) carried out the homogeneity analysis of Turkish meteorological data based on a relative test and four absolute homogeneity tests. Chu et al. (2012) detected a change-point in extreme rainfall data by applying the CUSUM method. Yurekli (2019) revealed change-point of seasonal reference evapotranspiration (ETo) dataset based on the CUSUM test. The sequential version of rank statistic was used to detect the approximate starting point of change in rainfall data sets (Yurekli, 2015). Duhan and Pandey (2013) achieved the detection of abrupt change in the studied data with Mann-Whitney Pettitt (MWP) and CD test.

Knowledge about the frequency distribution behavior of hydrological variables has crucial importance in the hydraulic structure design. In this regard, determining the most appropriate distribution of probability for the available data is a matter to be considered as much as the data is suitable for statistical analysis. Not having reasonable data available and not choosing a sufficiently successful probability distribution would lead to a quantile estimation below or above the design value of the hydraulic structure. This improper analysis would affect everything regarding hydraulic structure, such as design, operation, and maintenance (Yurekli, 2005). Frequency analyses of hydrological data are carried out on a point and regional scale (Anlı, 2009). Since this study is not based on the performance of point or regional frequency analysis, the main goal of the study is related to how the presence of the change in the data affects the frequency analysis. In this regard, this study is targeted (a) to reveal the existence of change-point in the studied data, (b) to analyze the data groups before and after the change-point, (c) to evaluate the ability of considered probability distributions to represent the data in groups.

2. Material and Methods

2.1. Data and Homogeneity Analysis

In order to assess the effect of the presence of the change point in a hydrological data on the shape of its frequency distribution, the maximum rainfall depths of different durations (1-h- and 2-h-duration for Bitlis and Tunceli rainfall stations, 12-h-duration for Ağrı, Elazıg, and Hakkari rainfall stations) were used as a material in the study. Some features associated with these stations are in Table 1. The locations of the stations in the Euphrates-Tigris basin are given in Figure 1. Erdogan (2020) tested with Mann-Whitney U (MWU) method whether these data sequences had homogeneity. The results of the MWU test applied to these data sets pointed out that there was inhomogeneities in the whole data sets (Table 1). The test statistic values (Z_U) of these data sequences were greater than the critical value (±1.96) from the standard normal distribution table at a 5% confidence level for the two-tailed. These results emphasize the presence of a change point in the data taken into consideration.

2.2. Detection of Change-point in time series

The Pettitt test, which was first described by Pettit (1979) to determine the change point in time series, is widely used in determining the change point in climate data. This non-parametric approach is accomplished as follows (Smadi and Zghoul, 2006).

First, the ranks are determined by arranging a given data in ascending order, the "$\mathbf{U_k}$" statistic is calculated according to Equation 1.

$$\mathbf{U_k} = \mathbf{2}\sum_{\mathbf{i=1}}^{\mathbf{k}} \mathbf{M_i} - \mathbf{k}(\mathbf{n}+\mathbf{1}) \qquad k = 1,2,\ldots,n \qquad (1)$$

Where "$\mathbf{M_i}$ and n" are the rank of ith rainfall depth in the series and observation length, respectively. The absolute maximum value of the "$\mathbf{U_k}$" value gives the change point (Equation 2).

$$K = \max_{1 \le k \le n} |U_k| \quad (2)$$

The K-value is compared to the critical "K_α" value obtained from the equation below at a 5% confidence level (α).

$$K_\alpha = 2.exp\left(\frac{-6K^2}{n^2+n^3}\right) \quad (3)$$

Table 1. The features associated with the rain gauge stations, and the MWU results
Çizelge1. *Yağmur istasyonlarıyla ilgili özellikler ve MWU sonuçları*

Meteorological Stations	Data Period	Latitude (E)	Longitude (N)	Elevation (Meter)	MWU Test (Z_U)		
					1h	2h	12h
Bitlis	1966-2015	38° 24'	42° 60’	1545	-2.31	-2.17	---
Tunceli	1968-2010	37° 44’	38° 13’	1019	-2.67	-2.74	---
Agrı	1967-2015	39° 43'	43° 30’	1672	---	---	-2.21
Elazıg	1957-2015	38° 40’	39° 14’	1027	---	---	-3.01
Hakkari	1968-2010	37° 34’	43° 46’	1758	---	---	-2.55

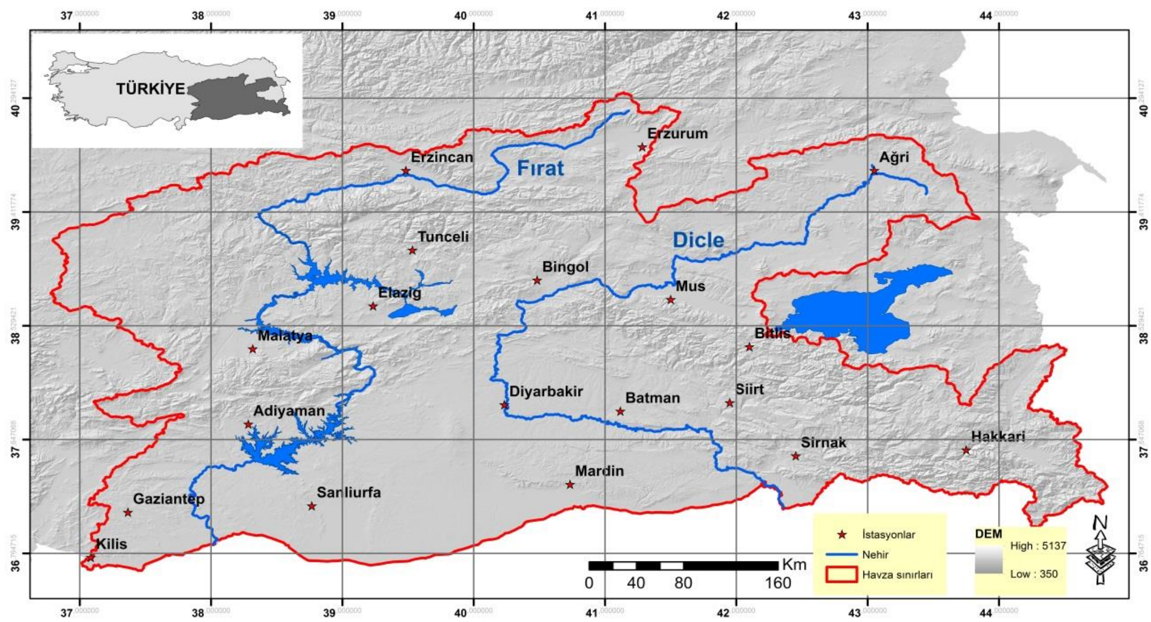

Figure 1. The locations of the stations on the Euphrates-Tigris Basin
Şekil 1. *İstasyonların Fırat-Dicle Havza’sı üzerindeki konumları*

2.3. Quantile prediction based on Frequency Analysis

As can be seen from the literature, there are many probability distribution relationships used in frequency analysis of hydrological data. However, among these, the probability distribution that provides the best compliance with the data is preferred. As a matter of fact, there is no search for the most appropriate distribution to data among the objectives of this study. The aim of the study is to determine the effect of change in data over time on frequency analysis. Nevertheless, the study of Erdoğan (2020) was taken into consideration. The study is associated with the regional frequency analysis applied to the maximum rainfall depths belonging to meteorology stations in the Euphrates-Tigris basin. Generalized Extreme Values (GEV) Distribution was suggested as a regional probability distribution in the study. Based on this reference, the GEV distribution was selected for the study.

In Anlı (2009), it is possible to find detailed information about some distributions and how to obtain the parameters of the distributions based on the l-moments method. The general relationship of GEV distribution is given below.

$$F = exp\{-[1 - k(x - \xi)/\alpha]^{1/k}\} \quad (4)$$

$$x = \xi + \alpha\{1 - (-logF)^k\}/k \quad (5)$$

parameters; ζ (location), α (scale) and k (shape)

Estimates of parameters based on l-moments are also determined from the following equations (Hosking and Wallis, 1997).

$$k \approx 7.8590c + 2.9554c^2,$$

$$c = \frac{2}{3+\tau_3} - \frac{log2}{log3} \quad (6)$$

$$\alpha = \frac{\lambda_2 k}{(1-2^{-k})\Gamma(1+k)} \quad (7)$$

$$\xi = \lambda_1 - \alpha\{1 - \Gamma(1+k)\}/k \quad (8)$$

3. Results and Discussion

The maximum rainfall sequences belonging to different durations were tested with the Pettitt method to reveal the change-point in the data set. The results of the test are given in Table 2. As can be seen in the table, the K-value of each data calculated was greater than the value of the critical K_α. Theise results indicate that the considered data sets had a significant change-point. This proves disruption of the technical conditions of the place where the measurement made, and/or the natural functioning of the hydrologic cycle.

The duration curves with their average values of both groups formed for each maximum rainfall data based on these results showed the variation for the maximum rainfall of 1h- and 12h-durations at Bitlis and Elazıg sites (Figure 2). The figures associated with the remaining stations were not given in order not to increase the volume of the study. However, their duration curves also had very similar behavior to that of these two figures

Table 2. Pettitt test results
Çizelge 2. *Pettitt testi sonuçları*

Site	1-h			2-h			12-h		
	YCP*	K	K_α	YCP	K	K_α	YCP	K	K_α
Bitlis	1992	300	252.31	1990	254	252.31	---	---	---
Tunceli	1982	284	201.54	1981	324	201.54	---	---	---
Agri	---	---	---	---	---	---	1997	262	244.83
Elazig	---	---	---	---	---	---	1991	424	331.13
Hakkari	---	---	---	---	---	---	1996	284	201.54
YCP*, Year of Change-Point									

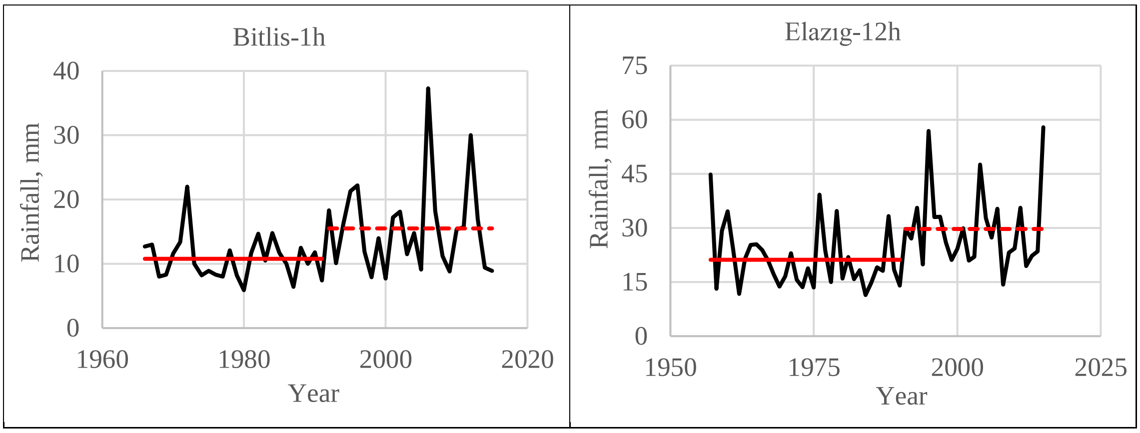

Figure 2. Duration curve of maximum rainfall
Şekil 2. *Maksimum yağmurların zaman grafiği*

The l-moments dealing with the annual maximum rainfall sequences before and after a change point are provided in Table 3. In particular, it seems that there were large differences between the groups of each rainfall data across the coefficients of variation (L-CV), skewness (L-CS) and kurtosis (L-CK). In fact, it is partially understood from these results that the groups of each rain data had different frequency distribution. However, in order to confirm the accuracy of this conclusion, it is necessary to rivet with the estimation of quantiles belonging to each group from the probability distribution relationship considered. For this purpose, Generalized Extreme Values (GEV) Distribution was applied to the rainfall data of each group and, quantiles were estimated for different return periods. Based on GEV distribution, the quantile estimations of both groups for different return periods dealing with the probability of non-exceedance highlighted that there was a meaningful difference at the same return period between each other. In Figure 3, Quantile estimations were given for Bitlis and Elazıg stations. It is evident in Figure 3, where the deviation between quantiles is greater, especially in small return period values.

Table 3. L-moments of data groups (G1and G2) before and after the change-point
Çizelge 3. *Değişim noktasından önceki ve sonraki veri gruplarının (G1ve G2) L-momentleri*

Site	Data	L-Moments						
		L1	L2	L3	L4	L-CV	L-CS	L-CK
Bitlis	1h-G1	10.773	1.795	0.311	0.313	0.167	0.173	0.174
	1h-G2	15.504	3.749	0.995	0.717	0.242	0.265	0.191
	2h-G1	14.717	2.140	0.392	0.284	0.145	0.183	0.133
	2h-G2	19.115	3.977	0.892	0.889	0.208	0.224	0.224
Tunceli	1h-G1	8.236	0.738	0.180	0.016	0.090	0.243	0.021
	1h-G2	15.293	4.397	1.453	0.599	0.288	0.330	0.136
	2h-G1	10.254	0.947	-0.006	-0.058	0.092	-0.006	-0.061
	2h-G2	18.920	4.740	1.548	0.585	0.251	0.326	0.123
Agri	12h-G1	23.953	4.588	0.870	0.405	0.192	0.190	0.088
	12h-G2	17.626	4.097	0.989	0.533	0.232	0.241	0.130
Elazig	12h-G1	21.174	4.443	1.253	0.647	0.210	0.282	0.146
	12h-G2	29.712	5.780	1.661	1.189	0.195	0.287	0.206
Hakkari	12h-G1	30.757	8.022	1.356	1.233	0.261	0.169	0.154
	12h-G2	15.927	5.235	1.463	1.580	0.329	0.280	0.302

L-CV, Coefficient of Variation
L-CS, Coefficient of Skewness, is referred to as "τ3" in the Equation 3
L-CK, Coefficient of Kurtosis

After determining the presence of a change in maximum rainfall data and negative effects on frequency analysis, the effect of the change in the existing data should be eliminated in order to reliably estimate the design value required in the planning phase of hydraulic structures. For this purpose, the double-mass curve method (DMCM) was applied to the maximum rainfall data in the study, new rainfall data for the years after the change-point was calculated. In order to check the results of the application based on the DMCM, the results of quantiles of different return periods dealing with 1h- and 12h-durations of Bitlis and Elazıg stations were shown in Figure 4. While obtaining this figure, two data sets for the same station were considered in predicting the quantiles based on the GEV. The first is the data (1h-G1 and 12h-G1) set of the period before the change point. The second one (1h-ND and 12h-ND) consists of the data set before the change and the regulated data after the change point. The GEV distribution was applied to these four data sets and quantiles were predicted. The figure shows that there was an insignificant deviation between the quantile values.

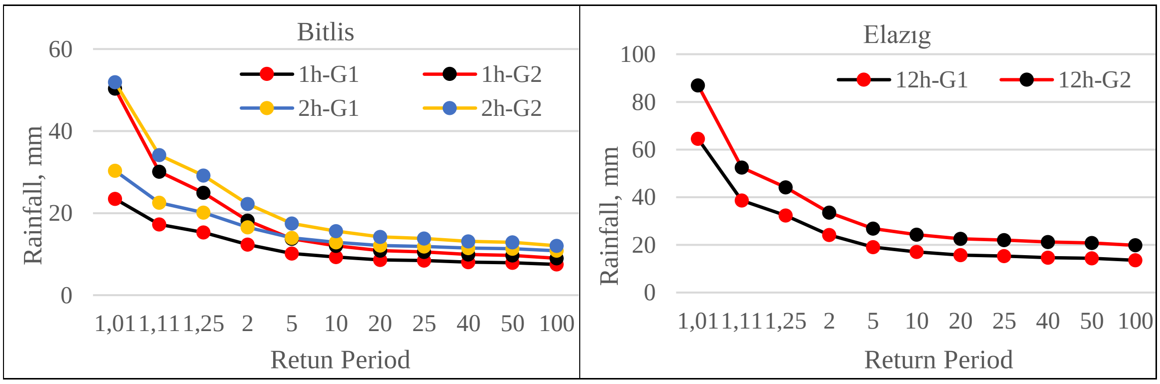

Figure 3. Maximum rainfall amounts at different return periods
Şekil 3. *Farklı tekrarlanma periyotlarında maksimum yağmur miktarları*

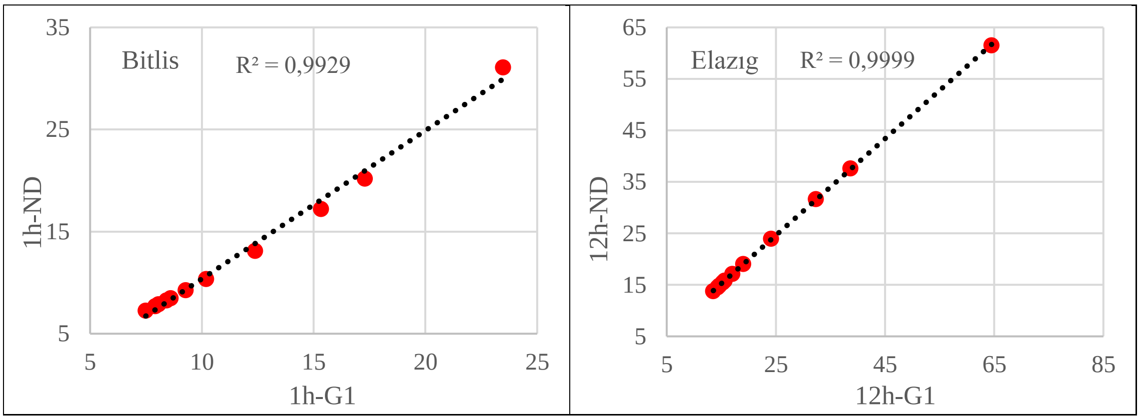

Figure 4. Application of double-mass curve method
Şekil 4. *Çift birikimli eğri yönteminin uygulanması*

4. Conclusion

In this study, it was aimed to reveal the effect of change in hydro-meteorological time series on frequency analysis. For this purpose, annual maximum rainfall amounts belonging to different durations (1h-, 2h- and 12h-) were used. Firstly, the year in which there was a change in rainfall data was determined and Generalized Extreme Value (GEV) distribution was applied to the rainfall data (for two groups of each site) before and after this change year. Based on the GEV distribution, quantile values were predicted for rainfall depths with the different durations and quantiles values of the groups were compared with each other. The results highlighted a meaningful deviation between the values. The conclusion made from the study was that the data studied was essential for the reliable estimation of design values, which are very important for the cost and functions of hydraulic structures.

References

Anlı AS. (2009). Ankara'da Meydana Gelen Yağmurların L Moment Yöntemleri ile Bölgesel Frekans Analizi. Ankara Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi.

Chu HJ, Pan TY and Liou JJ. (2012).Change-point detection of long-duration extreme precipitation and the effect on hydrologic design: a case study of south Taiwan. Stochastic Environmental Research and Risk Assessment, 26(8): 1123-1130.

Duhan D and Pandey A. (2013). Statistical analysis of long term spatial and temporal trends of precipitation during 1901–2002 at Madhya Pradesh, India. Atmospheric Research, 122: 136-149.

Erdogan M. (2020). Fırat-Dicle Havzasında Farklı Şiddetteki Maksimum Yağmurların L-Momentler Yaklaşımıyla Bölgesel Frekans Analizi. Tokat Gaziosmanpasa Universitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Tokat.

Hosking JRM and Wallis JR. (1997). Regional Frequency Analysis: An Approach Based on LMoments. Cambridge University Press: Cambridge, U.K.

Kazemzadeh M and Malekian A. (2018). Homogeneity analysis of streamflow records in arid and semi-arid regions of northwestern Iran. Journal of Arid Land, 10: 493-506.

Mahmood Agha OM, Bagcacı SC and Şarlak N. (2017). Homogeneity analysis of precipitation series in North Iraq. IOSR Journal of Applied Geology and Geophysics, 5(3): 57-63.

Pettitt AN. (1979). A non-parametric approach to the change point problem. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(2): 126-135.

Sahin S and Cigizoglu HK. (2010). Homogeneity analysis of Turkish meteorological data set. Hydrological Processes, 24(8): 981-992.

Smadi MM and Zghoul A. (2006). A sudden change in rainfall characteristics in Amman, jordan during the mid 1950s. American Journal of Environmental Sciences, 2(3): 84-99.

Wijngaard JB, Kleink Tank AMG and Konnen GP. (2003). Homogeneity of 20th century european daily temperature and precipitation series. International Journal of Climatology, 23(6): 679-692.

Yuksel S. (2019). Kızılırmak Havzasında Mevsimsel Ve Yıllık Yağmurlardaki Değişkenliğin Saptanması. Tokat Gaziosmanpasa Universitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Tokat.

Yurekli K. (2005). Regional frequency analysis of maximum daily rainfalls based on l-moment approach. Journal of Agricultural Faculty of Gaziosmanpasa University, 22 (1): 37-44.

Yurekli K. (2015). Impact of climate variability on precipitation in theUpper Euphrates–Tigris Rivers Basin of Southeast Turkey. Atmospheric Research, 154: 25-38.

Yurekli K. (2019).Detecting Change-point in Seasonal Reference Evapotranspiration (ETo) datasets. Journal of Agricultural Faculty of Gaziosmanpasa University, 36 (3): 206-210. doi:10.13002/jafag4595.