Quality Control and Homogeneity of Annual Precipitation Data in Büyük Menderes Basin, Turkey

Ercan Yeşilırmak[1 *], Selin Akçay[1], Necdet Dağdelen[1], Talih Gürbüz[1], Fuat Sezgin[1]

[1] Adnan Menderes University, Faculty of Agriculture, Department of Agricultural Structures and Irrigation, Aydın, Turkey

[*] Corresponding author: Phone: +90 256 772 70 23, Fax: +90 256 772 72 33, e-mail: eyesilirmak@adu.edu.tr

ABSTRACT

Precipitation is one of the most impotant climatic factors affecting agricultural production. Knowledge about spatial variability of precipitation amount over an agricultural area, its temporal change not only throughout a year but also over long-term span, start, end and length of rainy period, risk of wet and dry periods would be needed for appropriate agricultural planning and water management issues. However, analysis of long-term precipitation data for various purposes to be accurate, precipitation data must be homogeneous. It is defined that, as for other climatic time series, a homogeneous precipitation time series is to be affected by only natural weather and climatic conditions. Non-climatological factors such as changes in instrument, relocation of station, changes in observation practices make any climatic time series inhomogeneous. In this study, a quality control process involving outlier trimming and homogeneity checking were applied to 20 annual precipitation time series of various lenghts in Buyuk Menderes Basin, Turkey. Homogeneity analysis were performed using the Pettitt test and the Buishand range test. The results of the tests showed that 8 out of 20 stations can be considered to be inhomogeneous whose change points were found to be significant at 5% level by either one or both tests.

Keywords: Precipitation, Turkey, Büyük Menderes Basin, quality control, homogeneity, outlier

INTRODUCTION

Precipitation is one of the most impotant climatic factors affecting agricultural production. Knowledge about spatial variability of precipitation amount over an agricultural area, its temporal change not only throughout a year but also over long-term span, start, end and length of rainy period, risk of wet and dry periods would be needed for appropriate agricultural planning and water management issues. However, analysis of long-term precipitation data for various purposes to be accurate, high quality precipitation data must be used. Thus outliers and homogenization arise as important issues (Gonzalez-Rouca et al., 2001).

Detection of outliers has been considered an important part of quality control work. Outliers are data points that depart significantly from the trend of the remaining data (Naoum and Tsanis, 2003). They can be due to measurement errors or extreme meteorological events (Göktürk et al., 2008). When outliers are undoubtedly erroneous measurements those extreme data can be rejected and the problem is converted into one of missing data treatment (Gonzalez-Rouca et al., 2001). When outliers have a physical background the question arises whether they should be corrected or not (Barnett and Lewis, 1994), because, extreme data carry very valuable climatological information that should not be dismissed (Gonzalez-Rouca et al., 2001). On the other hand, outliers can affect the estimation of sample statistics during the use of nonresistant techniques (Göktürk et al., 2008). In order to retain the information of extreme events while not influencing nonresistant statistics too much, outliers can be replaced by a threshold value specific for each time series (Barnett and Lewis, 1994). Following Gonzalez-Rouca et al. (2001) and Göktürk et al. (2008) this approach was adopted as the quality control procedure in this work.

A homogeneous climate series is defined as one where variations are caused only by changes in weather and climate (Conrad and Pollak 1950). Most of the long-term climatic time series have been affected by a number of non-climatic factors that make these data unrepresentative of actual climate variations occurring over the time (Peterson et al., 1998). These non-climatic factors which makes data inhomogeneous are changes in location of the stations, instruments, formulae used to calculate means, observing practices and station environment (Göktürk et al., 2008). If a precipitation time series is homogeneous, all variability and changes of the series then can be considered due to the atmospheric processes (Karabörk et al., 2007).

There exists many methodologoies for detection of homogeneity of climatological time series. Firstly these methods can be grouped into two categories, direct or indirect methods, depending on availability or use of station history files known as metadata. Direct methods use metadata and indirect methods use a variety of statistical and graphical techniques to determine inhomogeneities (Peterson et al., 1998). The indirect homogeneity tests of a climatic time series could be classified into two groups: absolute tests and relative tests. The absolute tests depend on the use of a single station's records, whereas relative tests depend on the use of neighbouring stations' data that are supposedly homogeneous (Karabörk et al., 2007). Some relative homogeneity tests which do not require homogeneous reference series have become available (Szentimrey, 1999).

Numerous quality control, homogeneity testing and adjustment studies for many climatological time series were conducted at various temporal scales worldwide: for rainfall data in Kenya (Kipkorir, 2002), for daily air temperature and pressure series in Uppsala, Sweden (Bergström and Moberg, 2002), for precipitation and temperature series in Central America and northern South America (Aguilar et al., 2005), for precipitation in Taiwan (Yu et al., 2006), precipitation in Denmark (Frich et al., 1997), temperature and precipitation in Switzerland (Begert et al., 2005), precipitation in the Southwest of Europe (Gonzalez-Rouca et al., 2001), rainfall in Spain (Llasat and Quintas, 2004).

A number of studies checking data quality, testing and adjusting homogeneity for precipitation data in Turkey were conducted. Karabörk et al. (2007) checked the homogeneity of 212 precipitation records in Turkey for the period 1973-2002 by the Standard Normal Homogeneity Test (SNHT) and Pettitt Test. Authors found that 43 out of 212 stations were inhomogeneous based on the criteria that stations being considered inhomogeneous if at least one of the tests rejects the homogeneity. Göktürk et al. (2008) performed outlier trimming and homogeneity checking/correction on the monthly precipitation time series of various lengths from 267 stations in Turkey, by using the Standard Normal Homogeneity Test for homogeneity analysis. Sönmez and Kömüşçü (2007) tested the homonegeity status of monthly rainfall totals from 156 stations for 1977-2006 period by using Kruskal-Wallis Homogeneity test and found 16 stations being inhomogeneous. Em et al. (2007) assessed the homogeneity of annual precpitation totals recorded between 1970 – 2003 at 15 stations in GAP region of Turkey by using Swed-Eisenhart run test and graphical analysis method, and found only data of one station being inhomogeneous.

The purpose of this study is to check the quality and homogeneity of precipitation time series recorded at various stations within Büyük Menderes basin which could be used in later for water management, hydrology, climate change and variability studies. This study differs from other studies conducted for Turkish precipitation data in that it includes precipitation data recorded not only by State Meteorological Service of Turkey (DMI) but also by State Water Works of Turkey (DSI) which were not investigated before in terms of homogeneity.

MATERIALS and METHODS

Data

In this study, time series of annual precipitation totals from 20 stations within Büyük Menderes Basin were used. Locations of station are shown in Figure 1 and the list of stations is given in Table 1. Data were provided by the State Meteorological Service of Turkey (DMI) and by State Water Works of Turkey (DSI).

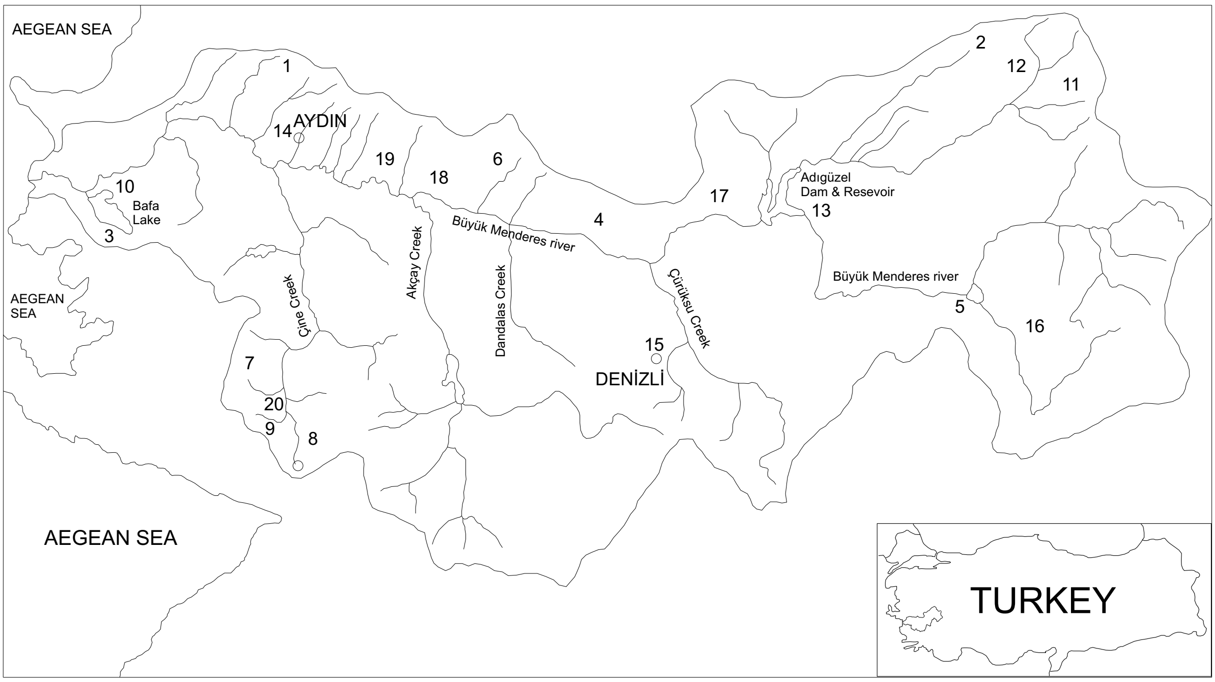

Figure 1. Distribution of 20 precipitation stations over Büyük Menderes Basin used in the study.

METHODS

Quality Control

The identification of outliers has been the primary emphasis of quality control work (Gonzalez-Rouca et al., 2001; Göktürk et al., 2008). Outliers are values greater than a threshold value specific for each time series, defined by

$$P_{out} = q_{0.75} + 3IQR$$

where $q_{0.75}$ is the third quartile and IQR is the interquartile range. In order to reduce the size of distribution tails and make a safer use of the nonresistant homogeneity testing methods used later, also to keep the information from extreme events, outlier values of each annual precipitation series were replaced by the unique P_{out} value (Gonzalez-Rouca et al., 2001; Göktürk et al., 2008).

Homogeneity Analysis

In this study, two methods to test the homogeneity in annual precpitation time series were used. These are the Buishand Range test (Buishand, 1982) and the Pettitt test (Pettitt, 1979). The mathematical formulation of the tests which were adopted from Wingaard et al. (2003) are given below. In the formulation given below, Y_i (i is the year from 1 to n) is the annual series to be tested, $\bar{Y}$ is the mean and s the standart deviation.

Buishand Range Test: In this test, the adjusted partial sums are defined as

$$S_0^* = 0 \quad \text{and} \quad S_k^* = \sum_{i=1}^{k}(Y_i - \bar{Y}) \quad k = 1,....,n$$

When a series is homogeneous the values of S_k^* will fluctuate around zero, because no systematic deviations of the Y_i values with respect to their mean will appear. If a break is present in year K, then S_k^* reaches a maximum (negative shift) or minimum (positive shift) near the year $k=K$. The $(S_k^*/s)/\sqrt{n}$ is depicted in the graphs representing the results of this test. The significance of the shift can be tested with the 'rescaled adjusted range' R, which is the difference between the maximum and the minimum of the S_k^* values scaled by the sample standard deviation:

$$R = (\max_{0 \le k \le n} S_k^* - \min_{0 \le k \le n} S_k^*)/s$$

Buishand (1982) gives critical values for $R/\sqrt{n}$.

Pettitt Test: This test is a non-parametric rank test. The ranks $r_1,......,r_n$ of the $Y_1,......,Y_n$ are used to calculate the statistics:

$$X_k = 2\sum_{i=1}^{k} r_i - k(n+1) \quad k = 1,....,n$$

The X_u is depicted in the graphs representing the results of this test. If a break occurs in year E, then the statistic is maximal or minimal near the year $k=E$:

$$X_E = \max_{1 \le k \le n} |X_k|$$

The significance level is given by Pettitt (1979).

RESULTS and DISCUSSION

The results of quality control process are given in Table 1 in which P_{out} values and extreme year(s) corrected for each station are tabulated. This table shows the variation of data that reaches maximum values along mountainous southern and northern border of the basin, whereas, lowest values occuring on central lowland part of the basin.

Table 1. The list of precipitation stations and the results of outlier trimming process.

S/N	Station Name	Data Period	Missing Data	P_{out} (mm)	Extreme year(s) replaced by P_{out}	State agency from which data were provided
1	Somak	1970 – 2005	1974	1757.5	–	DSI
2	A. Karacahisar	1964 – 2005	–	1215.9	–	DSI
3	Bafa-Çamiçi	1967 – 2005	1991, 1992	1571.8	–	DSI
4	Burhaniye	1963 – 1999	1993	1065.5	–	DSI
5	Işıklı Gölü	1963 – 2005	–	737.7	–	DSI
6	Kayran	1971 – 2005	–	1329.7	–	DSI
7	Kırıkköy	1968 – 2005	1974	1795.6	–	DSI
8	Kozağaç-Muğla	1962 – 2003	–	2542.5	–	DSI
9	Kozağaç-Yatağan	1962 – 2005	1963	1995.4	1981	DSI
10	Sarıkemer	1968 – 2001	1971, 1976	1467.0	–	DSI
11	Serban	1967 – 2000	–	957.7	–	DSI
12	Yavaşlar	1964 - 2001	1965, 1982	896.6	–	DSI
13	Yeşiloba	1968 - 2005	1974, 1975, 1976, 1977	1570.4	–	DSI
14	Aydın	1960 – 2007	–	1262.6	–	DMI
15	Denizli	1960 - 2007	–	1098.3	–	DMI
16	Dinar	1960 - 2006	–	1011.0	–	DMI
17	Güney	1960 - 2007	–	1018.8	1968	DMI
18	Nazilli	1960 - 2007	1969, 1970, 1971	1069.8	–	DMI
19	Sultanhisar	1961 - 2007	1967	1340.2	–	DMI
20	Yatağan	1961 - 2006	–	1378.8	–	DMI

The total number of corrected values is only two: one in Kozağaç-Yatağan and the other in Güney. In Kozağaç-Yatağan, total annual precipitation in year 1981 of 2033.3 mm is higher than and repleced by corresponding P_{out} value of 1995.4 mm. In Güney station, total annual precipitation in year 1968 of 1215.1 mm is higher than P_{out} value of 1018.8 mm, and it was replaced. The neighboring stations of Kozağaç-Yatağan and Güney have total annual precipitations which were not higher than their corresponding P_{out} values in the same years, thus one can conclude that these two outliers could be considered as erroneous measurements rather than as natural variation.

After quality control (outlier trimming) process, annual total prcipitation time series were tested for homogeneity. In this study two homogeneity testing methods were used. The selected methods are Buishand Range Test and Pettitt Test. The results of the tests applied are given in Table 2 for each station.

Table 2. The results of homogeneity tests. Significant change points at 5% level shown in bold.

S/N	Station Name	Data Period	Missing Data	Pettitt	Buishand
1	Somak	1970 – 2005	1974	1981	1981
2	A. Karacahisar	1964 – 2005	–	1977	1977
3	Bafa-Çamiçi	1967 – 2005	1991, 1992	1993	1993
4	Burhaniye	1963 – 1999	1993	1969 1981	1969 1981
5	Işıklı Gölü	1963 – 2005	–	1970 1984	1970
6	Kayran	1971 – 2005	–	1981	**1981**
7	Kırıkköy	1968 – 2005	1974	**1984**	**1986**
8	Kozağaç-Muğla	1962 – 2003	–	1983	1983
9	Kozağaç-Yatağan	1962 – 2005	1963	**1984**	**1983**
10	Sarıkemer	1968 – 2001	1971, 1976	1984	1984
11	Serban	1967 – 2000	–	1972 1976	1971 **1976**
12	Yavaşlar	1964 - 2001	1965, 1982	**1971**	**1971**
13	Yeşiloba	1968 - 2005	1974, 1975, 1976, 1977	**1986**	**1985**
14	Aydın	1960 – 2007	–	1986	1986
15	Denizli	1960 - 2007	–	1969 1981	1969 1981
16	Dinar	1960 - 2006	–	1969	**1969**
17	Güney	1960 - 2007	–	1969 1983	**1969** **1983**
18	Nazilli	1960 - 2007	1969, 1970, 1971	1983 1993	1983
19	Sultanhisar	1961 - 2007	1967	1971 1981	1971 1981
20	Yatağan	1961 - 2006	–	1971 1984	1971 1984

The results of Buishand Range test showed that 8 out of 20 stations have an inhomogeneity. On the other hand, according to Pettitt test 4 out of 20 stations were found to be inhomogeneous. Pettitt test detected nonsignificant change points at three stations, namely in Kayran, Serban and Güney stations, which Buishand Range test found significant change points at 5% significance level. Totally, 8 out of 20 stations are considered to be inhomogeneous whose change points were found to be significant at 5% level by either one or both tests.

Another outcome of the tests is that both tests detected change points at around the same years at almost all stations. In other words, the test results confirmed outcomes of one another in terms of timing of change point. For example, both tests detected a change point around 1985 in Yeşiloba station, as shown in Figure 2.

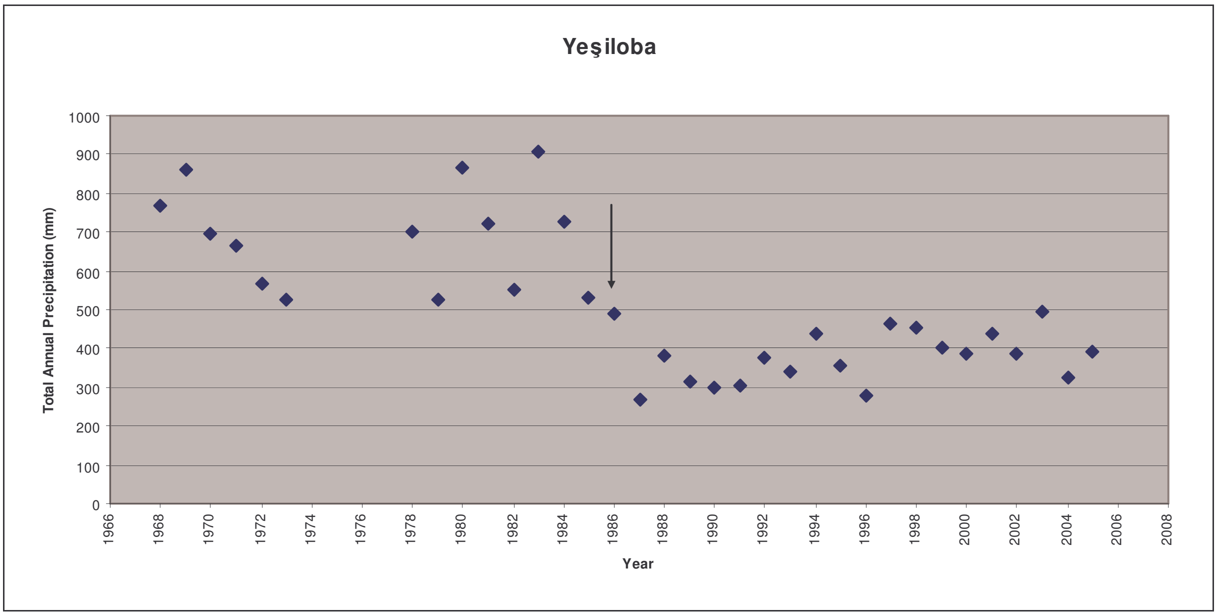

Figure 2. Annual precipitation time series of station Yeşiloba and detected inhomogeneity during 1985.

The test results depicts also that some neighboring stations exhibit simultaneous ocuurances of inhomogeneity. For instance, both tests detected a shift at the stations Kırıkköy and Kozağaç-Yatağan situated on southern border of the basin around the year 1984. On the other hand, two stations, namely Kozağaç-Muğla and Yatağan, neighboring to Kırıkköy and Kozağaç-Yatağan have change points around the same year which are not significant at 5% level. These simultaneous inhomogeneties may arise from simultaneous changes in observational routines (Karabörk et al., 2007).

Since thistorical metadata of the stations was not available in this study, no analysis could be made for the possible causes of the detected homogeneities. Some neighboring stations have simultaneous inhomogeneties or change points which are not significant at 5% level, therefore, it would be inappropriate to use the 'relative' homogeneity tests for Turkish precipitation data, as stated by Karabörk et al. (2007).

REFERENCES

Aguilar, E., T. C. Peterson, P. Ramirez Obando, R. Frutos, J. A. Retana, M. Solera, J. Soley, I. Gonzalez Garcia, R. M. Araujo, A. Rosa Santos, V. E. Valle, M. Brunet, L. Aguilar, L. Alvarez, M. Bautista, C. Castanon, L. Herrera, E. Ruano, J. J. Sinay, E. Sanchez, G. I. Hernandez Oviedo, F. Obed, J. E. Salgado, J. L. Vazquez, M. Baca, M. Gutierrez, C. Centella, J. Espinosa, D. Martinez, B. Olmedo, C. E. Ojeda Espinoza, R. Nunez, M. Haylock, H. Benavides, and R. Mayorga. 2005. Changes in precipitation and temperature extremes in Central America and northern South America, 1961–2003. J. Geophys. Res. 110, D23107, doi:10.1029/2005JD006119.

Alexandersson, H. 1986. A homogeneity test applied to precipitation data. Int. J. Climatol. 6: 661–675.

Barnett, V. and T. Lewis. 1994. Outliers in Statistical Data. John Wiley and Sons, New York.

Begert M., T. Schlegel and W. Kirchhofer. 2005. Homogeneous temperature and precipitation series of Switzerland From 1864 to 2000. Int. J. Climatol. 25: 65–80.

Bergström, H. and A. Moberg. 2002. Daily air temperature and pressure series for Uppsala (1722–1998). Climatic Change. 53: 213–252.

Buishand, TA. 1982. Some methods for testing the homogeneity of rainfall records. J. Hydrol. 58: 11–27.

Conrad, V. and C. Pollak. 1950. Methods in Climatology. Harvard University Press, Cambridge.

Em, A., N. Hamidi and Z.F. Toprak. 2007. GAP Bölgesinde Yıllık Toplam Yağışların Değişimi ve Homojenlik Analizi. In: Proceedings of I. Türkiye İklim Değişikliği Kongresi – TİKDEK 2007, (Eds., A. Öztopal and Z. Şen), İstanbul, pp. 379-393.

Frich P., S. Rosenørn, H. Madsen and J. J. Jensen. 1997. Observed Precipitation in Denmark, 1961-90. Danish Meteorological Institute Technical Report 97-8, Copenhagen.

Gonzalez-Rouca, J.F., J.L. Jimenez, V. Quesada and F. Valero. 2001. Quality control and homogeneity of precipitation data in the Southwest of Europe. J. Climate. 14: 964-978.

Jaruskova, D. 1994. Change-point detection in meteorological measurement. Mon. Weather Rev. 124: 1535–1543.

Göktürk, O.M., D. Bozkurt, Ö.L. Şen and M. Karaca. 2008. Quality control and homogeneity of Turkish precipitation data. Hydrol. Process. 22: 3210–3218.

Karabörk, Ç., E. Kahya and A. Ü. Kömüşçü. 2007. Analysis of Turkish precipitation data: homogeneity and the Southern Oscillation forcings on frequency distributions. Hydrol. Process. 21: 3203–3210.

Kipkorir, E.C. 2002. Analysis of rainfall climate on the Njemps Flats, Baringo District, Kenya. J. Arid Environ. 50: 445–458

Llasat, M.C. and L. Quintas. 2004. Stationarity of monthly rainfall series since the middle of the XIXth Century. Application to the case of Peninsular Spain. Natural Hazards. 31: 613–622.

Naoum, S. and I.K. Tsanis. 2003. Temporal and spatial variation of annual rainfall on the Island of Crete, Greece. Hydrol. Process. 17: 1899–1922.

Peterson, T.C., D.R. Easterling, T.R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E.J. Forland, I. Hanssen-Bauer, H. Alexandersson, P. Jones and D. Parker. 1998. Homogeneity adjustments of *in situ* atmospheric climate data: a review. Int. J. of Climatol. 18: 1493–1517.

Pettitt, A.N. 1979. A non-parametric approach to the change-point detection. Applied Statistics. 28: 126–135.

Sönmez, İ. and A.Ü. Kömüşçü. 2007. K-Ortalamaları Kümeleme Yöntemi ile Türkiye Yağış Bölgelerinin Yeniden Tanımlanması ve Alt-Periyodlardaki Değişimleri. In: Proceedings of I. Türkiye İklim Değişikliği Kongresi – TİKDEK 2007, (Eds., A. Öztopal and Z. Şen), İstanbul, pp. 360-378.

Szentimrey, T. 1999. Multiple analysis of series for homogenisation (MASH). In: Proceedings of 2nd Seminar for Homogenisation of Surface Climatological Data. Budapest, Hungary. WMO WCDMP, 41, pp. 27–46

Wijngaard, J.B., A. M. G. Klein Tank and G. P. Können. 2003. Homogeneity of 20th century European daily temperature and precipitation series. Int. J. Climatol. 23: 679–692

Yu, P.S., T.C. Yang and C.C. Kuo. 2006. Evaluating long-term trends in annual and seasonal precipitation in Taiwan. Water Resour. Manag. 20: 1007–1023.