

## Missing data analysis and homogeneity test for Turkish precipitation series

MAHMUT FIRAT<sup>1</sup>, FATİH DIKBAS<sup>2</sup>, A CEM KOÇ<sup>2</sup> and  
MAHMUD GUNGOR<sup>2</sup>

<sup>1</sup>Civil Engineering Department, Faculty of Engineering, İnönü University,  
Malatya, 44280 Turkey

<sup>2</sup>Civil Engineering Department, Faculty of Engineering, Pamukkale University,  
20017, Denizli, Turkey  
e-mail: mfirot@gmail.com

MS received 1 December 2009; revised 22 September 2010; accepted 15 October 2010

**Abstract.** In this study, missing value analysis and homogeneity tests were conducted for 267 precipitation stations throughout Turkey. For this purpose, the monthly and annual total precipitation records at stations operated by Turkish State Meteorological Service (DMI) from 1968 to 1998 were considered. In these stations, precipitation records for each month was investigated separately and the stations with missing values for too many years were eliminated. The missing values of the stations were completed by Expectation Maximization (EM) method by using the precipitation records of the nearest gauging station. In this analysis, 38 stations were eliminated because they had missing values for more than 5 years, 161 stations had no missing values and missing precipitation values were completed in the remaining 68 stations. By this analysis, annual total precipitation data were obtained by using the monthly values. These data should be hydrologically and statistically reliable for later hydrological, meteorological, climate change modelling and forecasting studies. For this reason, Standard Normal Homogeneity Test (SNHT), (Swed–Eisenhart) Runs Test and Pettitt homogeneity tests were applied for the annual total precipitation data at 229 gauging stations from 1968 to 1998. The results of each of the testing methods were evaluated separately at a significance level of 95% and the inhomogeneous years were determined. With the application of the aforementioned methods, inhomogeneity was detected at 50 stations of which the natural structure was deteriorated and 179 stations were found to be homogeneous.

**Keywords.** Precipitation; missing values completion; EM method; homogeneity tests.

## 1. Introduction

The accuracy and reliability of climate change, flood and drought modelling, water resources planning, determination of rainfall-runoff relationship, and river flow estimation models vary according to the quality of the data used. The factors such as method of gauging and data collection, the conditions around the station, station relocation, and the reliability of the measurement tool affect the homogeneous precipitation records. For this reason, the data recorded at gauging stations should be tested and checked for reliability and homogeneity prior to their use in the research studies. The data length and missing values at stations to be used in regional studies; the number of stations representing the area and the quality of the data are also very important in the development of an accurate model. For a better representation of an area, it is important to complete the series of the stations having missing values due to various reasons. A lot of methods such as time series models, Markov's models, multiple regression models, the nearest neighbour algorithm, neural networks and genetic algorithm, etc. are proposed for missing data analysis. In our study, Expectation Maximization (EM) algorithm is used for completing the series of stations with missing values. The EM algorithm combines statistical methodology with algorithmic applications and it receives interest in the solution of missing value problems (Dempster *et al* 1977). The EM algorithm is an iterative method for incomplete data and it increases the relationship between the missing value and the unknown parameters of a data model. In the application of this method, the missing values are initially calculated by using the estimated model parameters. Nelwamondo *et al* (2007) carried out a comparative study by using artificial neural networks and EM method for completing missing values. Schneider (2001) used the EM method for the completion and analysis of the missing values in climate series. Kim & Ahn (2009) applied the EM method to forecast the missing values in daily precipitation series. They recommended that the EM method can be successfully applied for missing data analysis.

The homogeneity tests of time series may be classified in two groups as 'absolute method' and 'relative method'. In the first method, the test is applied for each station separately. In the second method, the neighbouring (reference) stations are also used in the testing (Wijngaard *et al* 2003). In literature, a lot of methods were proposed for testing the homogeneity of meteorological variables like precipitation and temperature (Modarres 2008; Tomozeiu *et al* 2005; Klingbjør & Moberg 2003; Ducr'e-Rubitaillé *et al* 2003; Staudt *et al* 2007). Wijngaard *et al* (2003) used the SNHT method, Buishand test, Pettitt test and Von Neumann tests for testing the homogeneity of daily precipitation and temperature series. Mihajlovic (2006) used SNHT method to test the homogeneity of monthly total precipitation series used in monitoring of meteorological drought over Pannonian part of Croatia. Hanssen-Bauer & Førland (1994) applied homogeneity analysis on the 75 years long precipitation series of 165 stations in Norway by using SNHT method. Tayanç *et al* (1998) carried out a comparative evaluation by using Kruskal-Wallis and Wald-Wolfz methods to determine the inhomogeneous structure in the Turkish temperature series. Slonosky *et al* (1999) used various methods to test the homogeneity of surface pressure series of 51 stations with long years of observations in Europe. They stated that the SNHT method shows a good performance when a suitable reference series is obtained for comparison-evaluation and correction. Tuomenvirta *et al* (2000) used the SNHT method to test the reliability and homogeneity of the monthly maximum and minimum temperature series. Yıldırım *et al* (2004) tested the homogeneity of precipitation and riverflow series by using Kruskal-Wallis method in their study. Karabörk *et al* (2007) applied SNHT and Pettitt test to detect the inhomogeneity at 212 precipitation gauging stations for 1973-2002. They stated that 43 out of 212 stations had an inhomogeneity.

Kahya & Kalayci (2004) investigated the trend characteristics of Turkish stream flow data using four non-parametric trend tests (Sen's test, Spearman's Rho, Mann–Kendall and Seasonal Mann–Kendall tests). Freiwan & Kadioğlu (2008) used the annual, seasonal and monthly, maximum and minimum precipitation series to analyse the climate change in Jordan. RUN test was applied before the analysis to test the homogeneities of the annual, seasonal and monthly precipitation series. Özçelik (1996) used the Kruskal–Wallis test, the Swed–Eisenhart runs test, and graphical analysis to test the homogeneity of precipitation series. Partal & Kahya (2006) aimed to determine trends in the long-term monthly and annual precipitation series from 1929 to 1993 using non-parametric tests such as Mann–Kendall and Sen's T tests. They reported that January, February and especially September were determined to have strong decreasing trends, as opposed to other months showing either positive or negative trend in less stations. Moreover, it was stated that there were downward trends in the annual mean precipitation series, predominantly in western and southern Turkey, but a few upward trends were found in the central part of Turkey.

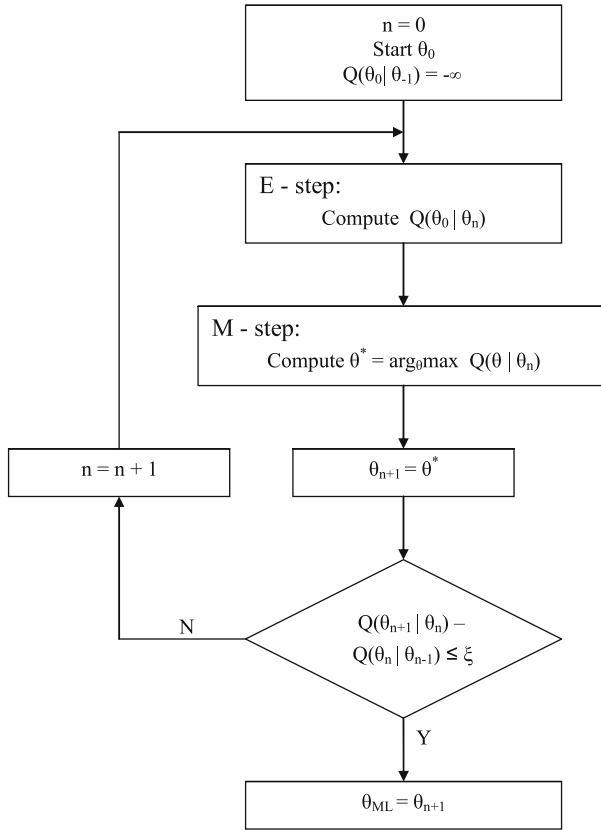
Basically this study has two purposes; (i) completion of missing values in the precipitation series recorded gauging stations throughout Turkey, (ii) testing the homogeneity of the annual total precipitation series. To complete the missing values, monthly total precipitation values are considered and EM method was applied at each station. For missing data analysis, daily precipitation series were not used because of the large number of zeros in daily records. The SNHT, Swed–Eisenhart Runs and Pettitt tests were used to test homogeneity and the inhomogeneity at stations was detected.

## 2. Expectation maximization (EM) method

The EM algorithm, which is proposed by Dempster *et al* (1977) to solve the problems faced in maximum likelihood methods, combines statistical methodology with algorithmic application and it receives attention for the solution of various missing value problems (Dempster *et al* 1977). The EM algorithm is a general method for incomplete data and it increases the relation between the missing data and the unknown parameters of the data model. Finding the model parameters is easy when the missing values are known. Similarly, when the parameters are known, it is possible to make estimations for the missing values. The EM algorithm, which is an iterative method, was proposed based on the reciprocal dependence between the model parameters and the missing values. If the data space is properly chosen, the EM algorithm can be estimated effectively the missing data values. The EM algorithm consists of two main steps; conditional expectation (called E-step) step and maximization (called M-step) step. The conditional expectations of missing data and estimates of model parameters are calculated by equation (1) in the E step. M step finds the estimates of the model parameter to maximize complete—data log likelihood function from E-step. These steps are iterated until the iterations converge (Schneider 2001). The details of EM methods can be obtained from Maclachlan & Krishnan (1997). The EM algorithm alternates the expectations (E) and maximization (M) steps for updating the estimate  $\theta_n$  of the unknown parameters  $\theta$  at iteration. The conditional expectations of missing data, given observed data and estimates of model parameters are calculated by equation (1) in the E step.

$$Q(\theta_0|\theta_n) = E_{Z|x,\theta_n}[\log L(\theta; x, z)], \quad (1)$$

where  $L(\theta; x, z)$  is the likelihood function,  $\theta$  is parameter vector,  $\theta_n$  is the estimate of the model parameters,  $x$  is observed data  $z$  is the missing data. In M step, the model parameters



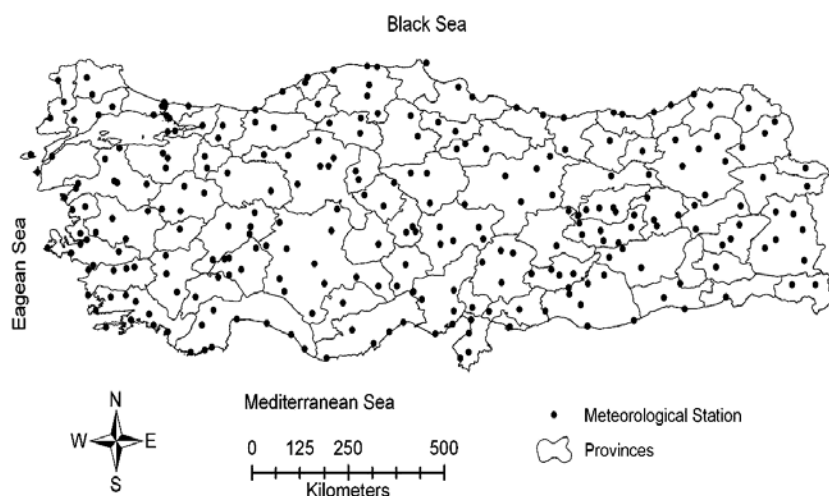
**Figure 1.** The flow chart for EM algorithm.

can be calculated using equation (2) to maximize complete—data log likelihood function from E-step. The flow chart for EM algorithm is demonstrated in figure 1.

$$\theta^* = \arg_{\theta} \max Q(\theta | \theta_n). \quad (2)$$

### 3. Homogeneity test

The quality and reliability of the data recorded at meteorological stations depends on many factors. Precipitation records at gauging stations are affected by the location of the station, the tool and method of data recording and collection and the observation quality and the time series might have inhomogeneity. For this reason, the reliability and quality of the data to be used in the modelling of hydrology and water resources processes should be tested statistically. It can be stated that the natural structure of the observation values is not deteriorated when the precipitation time series have a homogenous structure. There are many methods proposed and applied for testing homogeneity of meteorological series. The methods for testing the homogeneity of the series may be classified into two groups as ‘absolute method’ and ‘relative method’ (Karabörk *et al* 2007). In the first method, the test is applied for each station individually. Alternatively in the second method, neighbouring (reference) stations are also used for the testing process (Wijngaard *et al* 2003). However, it is difficult to



**Figure 2.** The locations of the investigated precipitation stations.

find reference stations with a high correlation and a homogeneous structure in wide regions (Tayanç *et al* 1998). For this reason, the absolute method was used for homogeneity test in our study owing to high spatial variation of precipitation stations. In this study, Standard Normal Homogeneity Test (SNHT), (Swed–Eisenhart) Run and Pettitt tests are used for detecting of the inhomogeneity of annual total precipitation records at gauging stations. The SNHT method was proposed by Alexanderson (1986) to detect the inhomogeneity in the time series. The SNHT detects the inhomogeneity at the beginning and/or towards of the series. The mathematical details of SNHT method can be seen in the studies of Alexanderson (1986), Alexandersson & Moberg (1997) and Gonzalez-Rouco *et al* (2001). The Pettitt test developed by Pettitt (1979), which is a non-parametric test that detects one change point in the observed time series, is more sensitive to detect the inhomogeneous structures in the middle of the time series (Costa & Soares 2009).

#### 4. Study area and available data

In this study, 267 precipitation stations operated by Turkish State Meteorological Service (DMI) throughout Turkey were considered for completing missing values and testing of the homogeneity. For this purpose, monthly and annual precipitation data covering the years between 1968 and 1998 was considered. This time range was determined by evaluating the records of the stations for using as much stations as possible. Monthly total precipitation values were obtained by summing up daily precipitation data observed by the stations. Similarly, the annual total precipitation data were calculated by adding monthly total precipitation data. The total annual precipitations at stations with missing data were calculated after completing the missing data. The geographical locations of stations are shown in figure 2.

In general, The Black Sea Region located in the north of Turkey receives the most precipitation. Turkey has a semi-arid climate characteristic and major spatial variations of precipitation are observed.

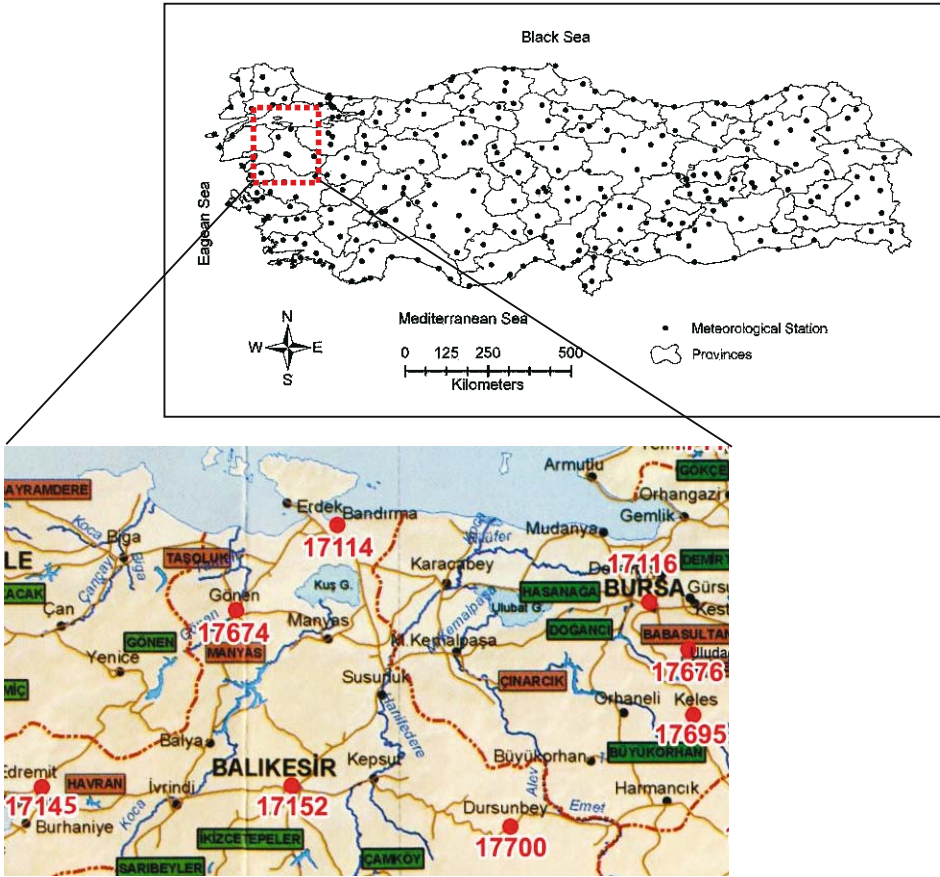


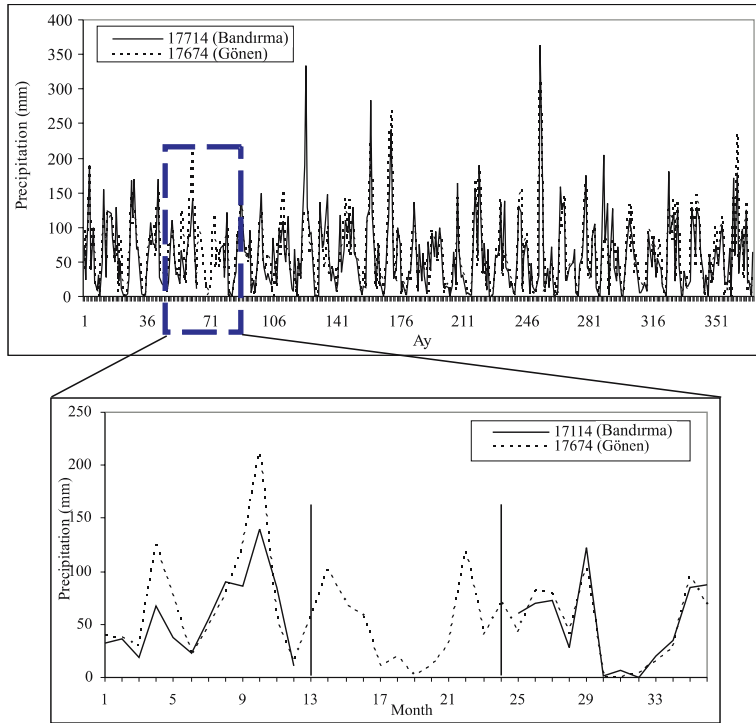
Figure 3. The locations of 17114 (Bandırma) and 17674 (Gönen) stations.

## 5. Results and discussion

As stated above, this study mainly consists of two steps. In the first step, the observation records of all stations were analysed one by one and the missing values were completed. Secondly, homogeneity test was applied for annual total precipitation data. For this, three different methods are used and the results were evaluated.

### 5.1 Missing value analysis

The missing values in the precipitation series were determined and they were completed by using the EM method. For this, monthly total precipitation data of 267 meteorology stations throughout Turkey for period from 01.10.1967 to 30.09.1998. For each station missing values were determined in monthly scale. As a result, the stations with more than five years of missing values were eliminated and the stations having missing data values for less than 5 years were completed by using EM method. To complete the missing values, the nearest neighbouring stations to the station with missing values were used. The results of the estimation were evaluated by correlation coefficient and the runs of the time series of the estimated and neighbouring stations. Bandırma station (station no. 17114) located in western Turkey was



**Figure 4.** Monthly total precipitation series of the stations 17114 and 17674.

selected to demonstrate the steps for the completion of the missing values (figure 3). It was determined that the monthly total precipitation values between 01.01.1973 and 31.12.1973 were missing in this station. For the completion of the missing values, the nearest neighbouring station, Gönen station (station no. 17674), was chosen as reference station. The comparison of the monthly total precipitation series of the stations 17114 (Bandırma) and 17674 (Gönen) is shown in figure 4.

The missing precipitation values for each month between 01.01.1973 and 31.12.1973 were estimated separately. As stated before, the estimation results were evaluated by calculating correlation coefficients and by comparing the precipitation series of each station. The comparison of the precipitation values of the completed station and neighbouring station can be seen in figure 5.

As can be seen from the figure, there is a good agreement between the estimated values of stations 17114 and 17674 for period from 01.01.1973 to 31.12.1973. Here, good results were obtained because the long year monthly total precipitations were used. In the determination of the missing values, 267 precipitation stations were similarly analysed. In this analysis, 38 stations were eliminated because they had missing values for more than 5 years, 161 stations had no missing values and missing precipitation values were completed in the remaining 68 stations. The location of these stations is shown in figure 6.

Figure 6 shows that a majority of the stations with missing values are in eastern Turkey. Similarly, the eliminated stations because of having too much missing values are also generally in the eastern part of Turkey. Table 1 shows the stations with missing values and the stations used in the completion of the missing values.

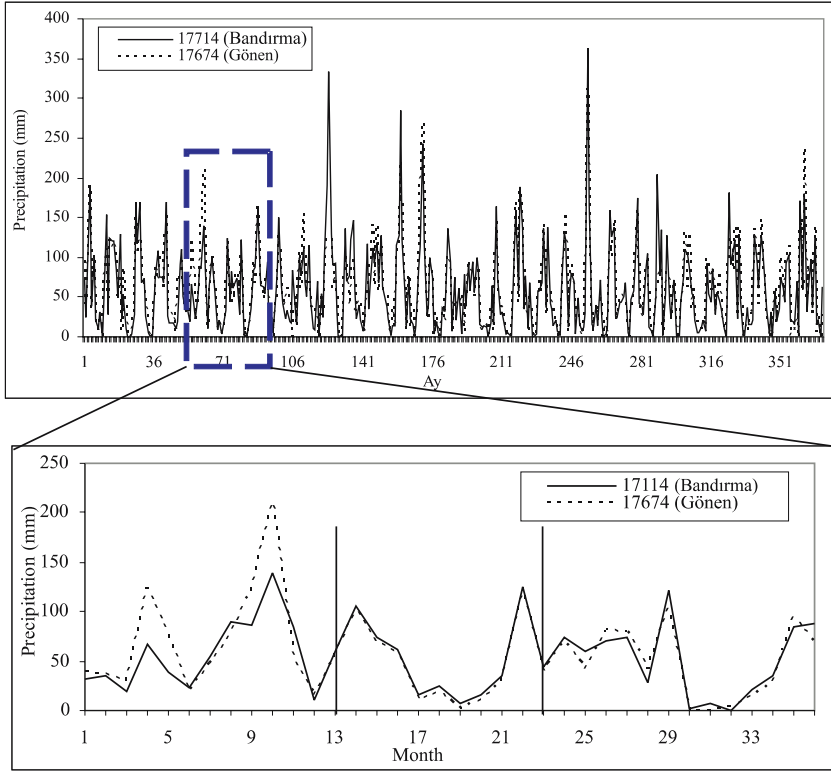


Figure 5. Comparison of the precipitation values of the completed station and neighbouring station.

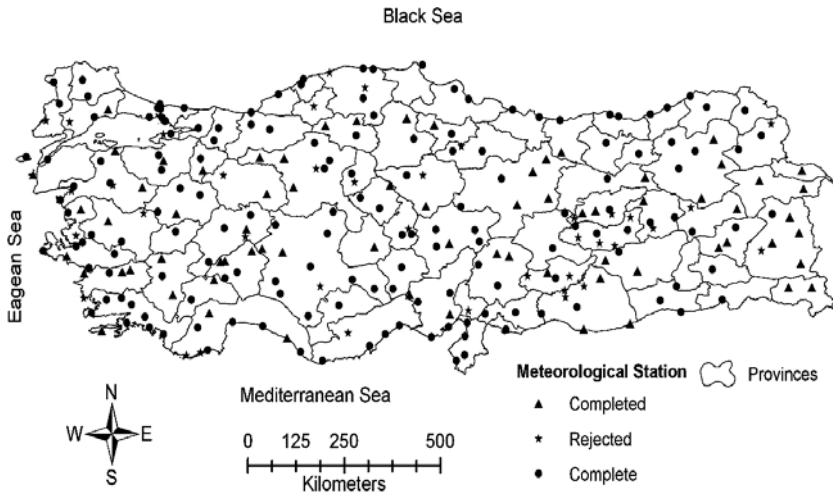


Figure 6. The distribution of the stations after the completion of the missing values.



**Table 1.** The stations with missing values and the stations used in the completion of the missing values.

The station with missing values			The station with missing values		
No	Name	Reference station	No	Name	Reference station
17054	CORLU	17056	17750	GEDIZ	17155
17083	MERZIFON	17085	17760	BOGAZLIYAN	17140
17099	AGRI	17780	17762	KANGAL	17090
17100	IGDIR	17099	17768	CEMISGEZEK	17165
17111	BOZCAADA	17110	17770	HOZAT	17165
17114	BANDIRMA	17674	17780	MALAZGIRT	17810
17172	VAN BOLGE	17812	17786	MURADIYE	17784
17184	AKHISAR	17186	17796	BULVADIN	17190
17192	AKSARAY	17250	17812	OZALP	17172
17205	TATVAN	17848	17820	SEFERIHISAR	17221
17232	KUSADASI	17854	17824	GUNEY	17237
17238	BURDUR	17240	17826	SENIRKENT	17828
17280	DIYARBAKIR	17282	17828	YALVAC	17862
17285	HAKKARI	17920	17830	AKSEHIR	17240
17310	ALANYA	17954	17832	ILGIN	17244
17646	CERKES	17080	17837	TOMARZA	17840
17648	ILGAZ	17080	17848	BITLIS	17204
17652	OSMANCIK	17083	17850	SULTANHISAR	17860
17668	OLTU	17688	17860	NAZILLI	17850
17676	ULUDAG-ZIRVE	17678	17864	ULUBORLU	17826
17679	NALLIHAN	17680	17868	AFSIN	17840
17680	BEYPAZARI	17130	17870	ELBISTAN	17866
17682	SEBINKARAHISAR	17684	17872	DOGANSEHIR	17199
17684	SUSEHRI	17682	17874	CERMIK	17270
17690	HORASAN	17096	17880	BASKALE	17172
17700	DURSUNBEY	17700	17890	ACIPAYAM	17237
17716	ZARA	17090	17912	SIVEREK	17914
17718	TERCAN	17096	17920	YUKSEKOVA	17285
17720	DOGUBEYAZIT	17152	17922	DATCA	17296
17726	SIVRIHISAR	17728	17926	KORKUTELI	17892
17728	POLATLI	17726	17934	POZANTI	17936
17732	CICEKDAGI	17160	17960	CEYHAN	17908
17734	DIVRIGI	17090	17968	CEYLANPINAR	17270
17738	KIGI	17165	17980	AKCAKALE	17270
17740	HINIS	17096			

In the completion of the missing values, the distance between the stations to be completed and to be used in the completion, geographical location, the climate properties of the location of the station are quite important for the model results. Especially, in the Central Anatolian region where arid climate is effective, it was very hard to find suitable reference stations for completing the missing values in summer months. If a general evaluation is made on the 68 completed stations, the correlations obtained for the completion process of summer months are at lower levels than the other months. The results of EM method for each station are evaluated by calculating correlations and by comparing the runs of the series of both stations. According to the results, it is thought and proposed that the EM method can successfully be used for completing the missing values in precipitation series.

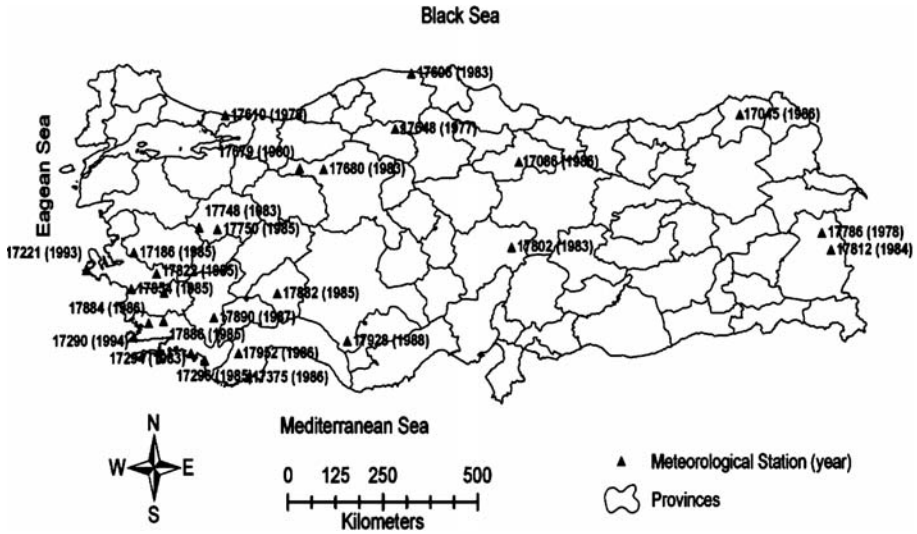
**Table 2.** Comparison of results of methods at 95% significant level.

Break year (test statistics) at 95% significant level							
Station No	SNHT	Pettitt	Run test	Station No	SNHT	Pettitt	Run test
17022	1995 (7-65)	–	–	17748	1983 (7-45)	1983 (154)	–
17026	1969 (7-61)	–	–	17750	1985 (8-42)	1985 (132)	–
17045	–	1986 (135)	–	17784	1969 (7-87)	–	–
17086	–	1986 (138)	–	17786	1976 (9-12)	1978 (140)	–
17110	1985 (9-66)	–	–	17802	–	1983 (164)	–
17186	–	1985 (142)	–	17806	1970 (8-12)	–	–
17221	–	1993 (42)	–	17812	1991 (12-86)	1984 (180)	–
17250	1975 (8-40)	–	–	17822	1985 (8-36)	1985 (154)	–
17261	1996 (8-65)	–	–	17824	1970 (24-96)	–	–
17265	1970 (8-48)	–	–	17826	1972 (11-26)	–	–
17282	1970 (11-86)	–	–	17836	–	–	2-2
17285	1970 (8-35)	–	–	17850	1985 (8-19)	1985 (134)	–
17290	–	1994 (31)	–	17854	1995 (7-77)	1985 (144)	–
17294	–	1983 (130)	–2.554	17870	1970 (10-59)	–	–
17296	–	1985 (128)	–	17880	1970 (10-32)	–	–
17300	–	–	–2.554	17882	–	1985 (86)	–
17375	1970 (8-31)	1986 (124)	–	17884	1969 (11-80)	1986 (138)	–
17602	1969 (8-71)	–	–	17886	–	1985 (130)	–2.554
17606	1984 (8-99)	1983 (150)	–	17890	1987 (9-5)	1987 (140)	–
17608	1995 (7-82)	–	–	17920	1970 (8-44)	–	–
17610	–	1978 (134)	–	17926	1970 (16-67)	–	–
17648	1976 (13-03)	1977 (136)	–	17928	–	1988 (132)	–
17679	1980 (18-11)	1980 (210)	–	17950	1970 (8-17)	–	–
17680	–	1983 (106)	–	17952	–	1986 (134)	–
17686	1995 (8-47)	–	–	17968	1970 (7-75)	–	–

### 5.2 Homogeneity test

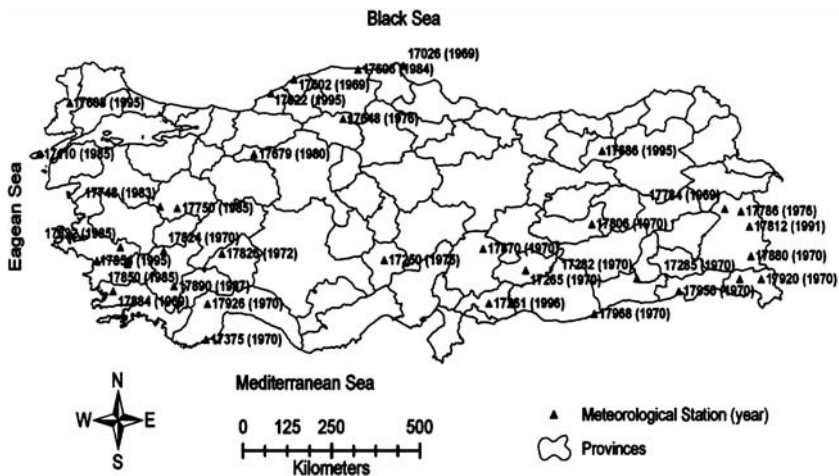
As stated above, the homogeneity of the annual total precipitation time series of the stations throughout Turkey were tested by using SNHT, Swed–Eisenhart Runs and Pettitt tests. As a result of the analysis explained in the previous section, annual total precipitation values for 229 stations were obtained. In the application of homogeneity methods, observation series of each station were considered separately. The results of each method were evaluated for a significance level of 95% and the inhomogeneities were detected. The evaluation of the SNHT results and the details of the criteria to be considered in the determination of the inhomogeneities or break years are given in the study made by Khaliq & Quarda (2007). Table 2 shows the list of stations having an inhomogeneity and the comparative test statistics calculated by the three methods. Figures 7 and 8 show the locations of the homogeneous and inhomogeneous stations determined by the application of the Pettitt and SNHT methods respectively.

In table 2, it is seen that the number of stations passing the critical test value by Pettitt test at a 95% significance level was 27. The results of the Pettitt test show that the inhomogeneity is generally detected between the years 1983 and 1988. As can be seen from the table, inhomogeneity was detected in 1983 in 5 stations, in 1985 in 8 stations and in 1986 in 5 stations by using the Pettitt test. Table 2 shows that only 4 stations were found to be inhomogeneous by using the RUN test.

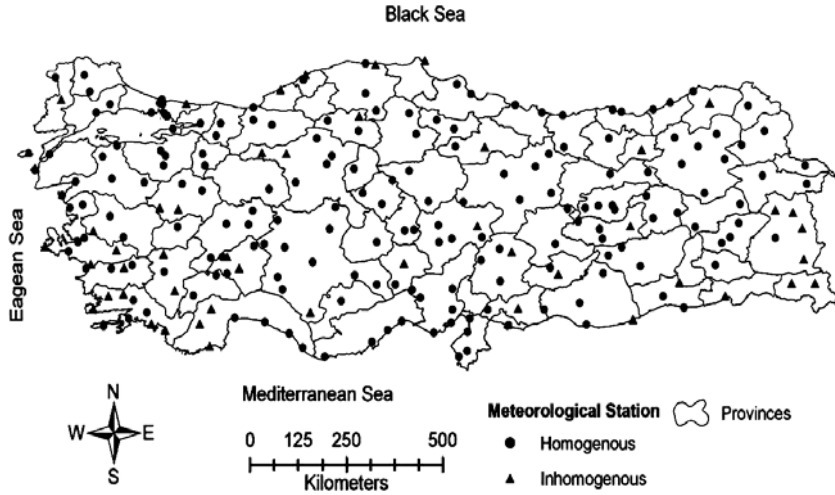


**Figure 7.** Homogeneity test results of Pettitt Test for precipitation series (the inhomogeneous stations and years).

Table 2 shows that 34 stations were found inhomogeneous by applying the SNHT method at a significance level of 95%. In the evaluation of the SNHT results, the stations with a test statistic higher than 7.54 are considered to be inhomogeneous for a 31 years long data. The results show that the inhomogeneity detected by using SNHT method is mostly at the beginning (1969–1970) or towards the end (1995–1996) of the series. The table also shows that the stations 17265, 17282, 17870, 17880, 17950, 17968, 17375, 17806, 17920, 17285, 17824 and 17926 have an inhomogeneity during 1970. Moreover, inhomogeneity was also detected during 1969 for the stations 17602, 17784, 17026 and 17884. The map shows that



**Figure 8.** Homogeneity test results of SNHT for precipitation series (the inhomogeneous stations and years).



**Figure 9.** The stations with and without homogeneous structure according to the homogeneity tests.

the stations with inhomogeneity in 1970 are located in Eastern Anatolia region. According to these results, it can be said that the inhomogeneity detected at gauging stations located in the same region is caused by the variations in the natural climate conditions.

It is also seen in the table that both the SNHT and Pettitt tests for some stations detected inhomogeneity in the same year. Inhomogeneity was detected during 1985 for the stations 17822, 17850 and 17750 by both of the tests. It is seen from the map that these stations are in the same region. Inhomogeneity was also detected during 1986 for the neighbouring 17296 (SNHT), 17882 (Pettitt), 17886 (Pettitt) stations. It can be deduced from these findings that the inhomogeneity might be related to the variations of natural meteorological conditions. With the application of three methods, at 50 gauging stations inhomogeneity were detected. The locations of these homogeneous and inhomogeneous stations are shown together in the map (figure 9).

Karabork *et al* (2007) applied the SNHT and Pettitt test to detect the inhomogeneity at 212 precipitation gauging stations for the period 1973–2002 throughout Turkey. The results of methods showed that 43 out of 212 stations had an inhomogeneity. The comparison of the results of this study and the paper presented by Karabörk *et al* (2007) shows that the stations for which the homogeneity is deteriorated are similar. In both the studies the inhomogeneity was detected at stations 17026, 17606 and 17648 located in northern Turkey. Also, it was determined that in both of the studies, inhomogeneity was detected at stations 17884, 17854 and 17186 located in southwest Turkey and at stations 17110, 17608, 17748 and 17750 located in western Turkey. According to these evaluations, it was seen that the obtained results and detected stations are similar in both of the studies.

## 6. Conclusions

In this study, missing value analysis and homogeneity tests were applied for the precipitation series of meteorology stations operated by DMI throughout Turkey. For this purpose, 267 stations having observations from 1968 to 1998 were analysed. Firstly, the missing value

analysis was carried out and the missing precipitation records at stations were completed using the EM method. For the completion of the missing values, the series of the nearest neighbouring stations were used. For this, each month of the year was evaluated separately and estimations were made for each month by using long year monthly precipitation series of the station with missing values and the neighbouring station used for completion. The results of estimation were evaluated by the correlation coefficients and the runs of the series of the estimated station and the neighbouring station. In the analysis, 38 stations were eliminated because they had missing values for more than 5 years, 161 stations had no missing values and missing precipitation values were completed in the remaining 68 stations.

In the second phase of the study, SNHT, Pettitt and RUN homogeneity tests for the annual total precipitations were applied for testing the reliability of the data. In the application of these methods, the observation records of each station were evaluated separately. The results of methods were evaluated separately at a significance level of 95% and the inhomogeneity and break years were detected. At a significance level of 95%, inhomogeneity was detected at 34 stations with the application of SNHT method on annual total precipitation series. The number of stations passing the critical test values by applying Pettitt and RUN test was 27 and 4 respectively. With the application of the three methods, inhomogeneity was detected in 50 stations of which the natural structure was deteriorated and 179 stations were found to be homogeneous. As a result of the study, it was seen that SNHT and Pettitt tests are more sensitive in the determination of inhomogeneity at gauging stations. The results show that these methods can be used successfully for testing the homogeneity of precipitation series. It is hoped that the results of this study would be a reference for future studies in the fields of hydrology and meteorology in Turkey.

This research was supported by TUBITAK (Turkish National Science Foundation) under the project number of 107Y318. The authors acknowledge the cooperation of DMI (Turkish State Meteorological Service) authorities for providing data and information. The authors are grateful to anonymous reviewers for their helpful and constructive comments on an earlier draft of this paper.

## References

- Alexanderson H A 1986 A homogeneity test applied to precipitation data. *J. Climatol.* 6: 661–675
- Alexanderson H, Moberg A 1997 Homogenization of Swedish temperature data Part I: Homogeneity test for linear trends. *Int. J. Climatol.* 17: 25–34
- Costa A C, Soares A 2009 Homogenization of climate data: Review and new perspectives using geostatistics. *Mathematical Geosci.* 41: 291–305
- Dempster A P, Laird N M, Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Soc., Series B* 39: 1–38
- Ducr'e-Rubitaillé J-F, Vincent L A, Boulet G 2003 Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol.* 23: 1087–1101
- Freiwan M, Kadioğlu M 2008 Climate variability in Jordan. *Inter. J. Climatol.* 28: 69–89
- Gonzalez-Rouco J F, Jimenez J L, Quesada V, Valero F 2001 Quality control and homogeneity of precipitation data in the Southwest of Europe. *J. Climate* 14: 964–978
- Hanssen-Bauer I, Førland E 1994 Homogenizing long Norwegian precipitation series. *J. Climate* 7: 1001–1013
- Kahya E, Kalaycı S 2004 Trend analysis of streamflow in Turkey. *J. Hydrol.* 289: 128–144

- Karabörk M Ç, Kahya E, Kömüşçü A Ü 2007 Analysis of Turkish precipitation data: Homogeneity and the Southern Oscillation forcings on frequency distributions. *Hydrological Processes* 21: 3203–3210
- Khaliq M N, Quarda T B M J 2007 On the critical values of the standard normal homogeneity test (SNHT). *Int. J. Climatol.* 27: 681–687
- Kim T W, Ahn H 2009 Spatial rainfall model using a pattern classifier for estimating missing daily rainfall data. *Stoch. Environ. Res. Risk Assess.* 23: 367–376
- Klingbjer P, Moberg A 2003 A composite monthly temperature record from Tornedalen in northern Sweden, 1802–2002. *Int. J. Climatol.* 23(12): 1465–1493
- Maclachlan G J, Krishnan T 1997 The EM algorithm and extensions (New York: John Wiley & Sons) 288 p
- Mihajlovic D 2006 Monitoring the 2003–2004 meteorological drought over pannonian part of Croatia. *Int. J. Climatol.* 26: 2213–2225
- Modarres R 2008 Regional frequency distribution type of low flow in north of Iran by L-moments. *Water Resources Management* 22: 823–841
- Nelwamondo F V, Mohamed S, Marwala T 2007 Missing data: A comparison of neural network and expectation maximization techniques. *Curr. Sci.* 93(11): 1514–1521
- Özçelik D 1996 Homogeneity analysis of the precipitation series located in Turkey. MSc thesis, Istanbul Technical University (in Turkish)
- Partal T, Kahya E 2006 Trend analysis in Turkish precipitation data. *Hydrological Processes* 20(9): 2011–2026
- Pettit A N 1979 A non-parametric approach to the change-point detection. *Appl. Statist.* 28(2): 126–135
- Schneider T 2001 Analysis of incomplete climate data: Estimating of mean values and covariance matrices and imputation of missing values. *J. Climate* 14: 853–871
- Staudt M, Esteban-Parra M J, Castri-Diez Y 2007 Homogenization of long-term monthly Spanish temperature data. *Int. J. Climatol.* 27: 1809–1823
- Slonosky V C, Jones P D, Davies T D 1999 Homogenization techniques for European monthly mean surface pressure series. *J. Climate* 12(8): 2658–2672
- Tayanc M, Dalfes H N, Karaca M, Yenigun O 1998 A comparative assessment of different methods for detecting inhomogeneities in Turkish temperature data set. *Int. J. Climatol.* 18: 561–578
- Tomozeiu R, Stefan S, Busuioc A 2005 Winter precipitation variability and large-scale circulation patterns in Romania. *Theor. Appl. Climatol.* 81: 193–201
- Tuomenvirta H, Alexandersson H, Drebs A, Frich P, Nordli P O 2000 Trends in Nordic and Arctic temperature extremes and ranges. *J. Climate* 13(5): 977–990
- Wijngaard J B, Klein Tank A M G, Können G P 2003 Homogeneity of 20th century European daily temperature and precipitation series. *Int. J. Climatol.* 23: 679–692
- Yıldırım Y E, Türkeş M, Tekiner M 2004 Time-series analysis of long-term variations in stream-flow data of some stream-flow stations over the Gediz basin and in precipitation of the Akhisar station. *Pakistan J. Biol. Sci.* 7(1): 17–24