HYDROLOGICAL PROCESSES
Hydrol. Process. **24**, 981–992 (2010)
Published online 26 January 2010 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.7534

Homogeneity analysis of Turkish meteorological data set

Sinan Sahin[1]* and H. Kerem Cigizoglu[2]

[1] *Civil Engineering Department, Engineering Faculty, Namik Kemal University, 59860 Tekirdag, Turkey*
[2] *Istanbul Technical University, Faculty of Civil Engineering, Hydraulics Division, 34469 Istanbul, Turkey*

Abstract:

The missing value interpolation and homogeneity analysis were performed on the meteorological data of Turkey. The data set has the observations of six variables: the maximum air temperature, the minimum air temperature, the mean air temperature, the total precipitation, the relative humidity and the local pressure of 232 stations for the period 1974–2002. The missing values on the monthly data set were estimated using two methods: the linear regression (LR) and the expectation maximization (EM) algorithm. Because of higher correlations between test and reference series, EM algorithm results were preferred. The homogeneity analysis was performed on the annual data using a relative test and four absolute homogeneity tests were used for the stations where non-testable series were found due to the low correlation coefficients between the test and the reference series. A comparison was accomplished by the graphics where relative and absolute tests provided different outcomes. Absolute tests failed to detect the inhomogeneities in the precipitation series at the significance level 1%. Interestingly, most of the inhomogeneities detected on the temperature variables existed in the Aegean region of Turkey. It is considered that theseinhomogeneities were mostly caused by non-natural effects such as relocation. Because of changes at topography at short distance in this region intensify non-random characteristics of the temperature series when relocation occurs even in small distances. The marine effect, which causes artifical cooling effect due to sea breezes has important impact on temperature series and the orograhpy allows this impact go through the inner parts in this region. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS Turkey; missing value estimation; homogeneity; meteorological data set; temperature

Received 13 March 2009; Accepted 12 October 2009

INTRODUCTION

The reliable measurements of the climate data are the essential foundation for the quantitative climate analyses. Unfortunately, there are several factors affecting the quality of the climate data and these factors must be understood and considered both for scientific and climatic analyses. Although there are universally accepted standards/recommendations for instrument installation and observations, the measurement practises and instruments may differ from station to station in a given country, and also there may be changes in an individual station from time to time. As a result, these factors cause variations in station time series.

Conrad and Pollack (1950) define a homogeneous time series as one in which variations are caused only by the weather and the climate. The factors causing variations on the long-term time series are, location of the stations, instruments, formulae used to calculate means, observing practices and station enviroment (Peterson *et al*., 1998). At most old observatories, instruments, location, observer, or other environmental factors have been altered (Jones *et al*., 1985; Karl and Williams, 1987; Peterson *et al*., 1998). Inhomogeneities can be detected in two ways; these are relative and absolute tests. In general, it is recommended (e.g. Peterson *et al*., 1998) to apply homogeneity tests relatively, i.e. testing with respect to neighbouring station that is supposedly homogeneous. If the two series are not sufficiently correlated, absolute tests, which use only the single station series are considered more powerful than relative tests (Wijngaard *et al*., 2003).

In this work, we selected one relative and four absolute test methods to test the departures from the homogeneity in the stations time series. We used the bivariate test developed by Maronna and Yohai (1978) for the relative homogeneity testing and when the correlations between reference series and testing series are low, we used four absolute tests proposed by Wijngaard *et al*. (2003). The four absolute test methods selected to detect inhomogeneities in the time series are: the standard normal homogeneity test (SNHT) for a single break (Alexandersson, 1986), the Buishand range test (Buishand, 1982), the Pettitt test (Pettitt, 1979) and the Von Neumann ratio test (Von Neumann, 1941). We labelled the time series 'suspect' if at least three of the four absolute tests reject the homogeneity, and these stations' time series were considered inhomogeneous. When two tests rejects the homogeneity, we labelled the stations time series 'doubtful' and when one or zero tests reject homogeneity we labelled these series 'useful'. The same procedure was used by Wijngaard *et al*. (2003).

In the international literature, only limited studies were found that the analyses the homogeneity of the

* Correspondence to: Sinan Sahin, Civil Engineering Department, Hydraulics Division, Engineering Faculty, Namık Kemal University, 59860 Corlu/Tekirdag, Turkey. E-mail: ssahin@corlu.edu.tr

Turkish meteorological data sets. Tayanç *et al.* (1998) analysed the homogeneity of the temperature variables belonging to 82 stations using the relative homogeneity tests. They applied Kruskal–Wallis (K–W) homogeneity test and Wald–Wolfowitz runs test and analysed the efficiencies of these tests generating the artificial time series. Karabork *et al.* (2007) performed two absolute homogeneity tests in 212 meteorological stations for the precipitation time series, which were SNHT and Pettitt test. Stations were considered inhomogeneous if at least one of the tests rejects the homogeneity at the significance level 0·5%. Türkeş *et al.* (2008) used for determining whether the time-series are homogeneous or not, the non-parametric K–W test. Sample size of sub-periods and the significance level of the test were taken as $n_j = 5$ and the $\alpha = 0{\cdot}05$, respectively. Türkeş (1999) were accomplished statistical evaluations of homogeneity for the annual and seasonal precipitation total series, and annual aridity index and temperature series by the non-parametric K–W test for the homogeneity of means of the sub-periods. Türkeş (1996) examined records of the monthly rainfall totals in order to obtain homogeneous records and to ensure a good geographical representation where possible. First, a total of 130 stations have been selected for the analysis. A preliminary examination of the monthly data showed that a number of the monthly records had missing data. Second, statistical evaluation of homogeneity of the rainfall data was checked by using the non-parametric K–W homogeneity test on the subperiods' means. When the station history files that include information for most of the remaining 91 stations were examined, it seemed that about 70% of the stations had been moved to new locations during the study period. Most of these relocations vary between several metres and 5 km. Türkeş (1996) checked the homogeneity of maximum and minimum temperature records by applying the non-parametric K–W test for homogeneity of means to annual and seasonal series (Sneyers, 1990). They have also verified the homogeneity of the variances using the same test. In this case, ranks of the absolute values of deviations from the general mean were used. Third, they have made a subjective assessment of each statistically significant inhomogeneity using additional information available from the plotted graphs and station history file. Then the climatological significance of each has been assessed. Finally, they have omitted six more stations from the analysis through this evaluation. Türkeş (2002b) performed a homogeneity analysis in order to detect homogeneity in mean annual and seasonal temperature series using the non-parametric K–W test (Sneyers, 1990; Türkeş, 1996; Türkeş *et al.*, 1996) of both 7- and 10-year sub-periods. The analysis was carried out for the means and variances of both 7 and 10-year sub-periods. This objective analysis was carried out not only to detect inhomogeneity (inconsistency) in the overall series, but also to examine whether the recent observations of about the last decade (in which increased spring and particularly summer minimum temperatures were dominant at many stations) affected the consistency of the temperature series.

Inhomogeneities can bias a time series and lead to the misinterpretations of the studied climate. It is important, therefore, to remove the inhomogeneities or at the least determine the possible error they may cause. The complete metadata are needed to ensure that the final data user has no doubt about the conditions in which the data are recorded, gathered and transmitted, in order to extract the accurate conclusions from their analysis. Unfortunately, in Turkey, the data sets analysed in this study have quite limited metadata preventing the researchers to make effective inferences from the homogeneity analyses.

The purpose of this work is to detect the inhomogeneities in the time series of all meteorological stations in order to determine the reliable climatic series for the future climate analyses. First, the missing values of the meteorological time series will be completed. To obtain more reliable results, we will apply two missing value methods [expectation maximization (EM) and linear regression (LR) method] and decide the best method according to criterions used in this study. One relative (bivariate test) and four absolute test methods will be used to test the departures from the homogeneity in the stations' time series. The relative test will be the first option in homogeneity analysis but when the requirements of the relative test have not met, four absolute tests will be used to detect inhomogeneities. The results of the homogeneity tests applied in this study will be demonstrated with graphichs to compare their results.

None of the available international studies cover the employment of the relative and the absolute homogeneity tests (four absolute tests in this study) for the various climatic time series (annual total precipitation, maximum, minimum and mean temperature, relative humidity, local pressure) on the national scale (232 stations in this study). The evaluation of the homogeneity results for the several climatic variables at the same station provides a quite extensive perspective about the quality of the data of this station.

THE DATA AND THE METHODOLOGY

Data

The climatic data (1974–2002) used in this study were provided by State Meteorological Service of Turkey. The data belong to 250 meteorological stations distributed all over the Turkey. These monthly data cover the observations of six variables: the maximum air temperature T_x, the minimum air temperature T_n, the mean air temperature T_m, the total precipitation P_t, the relative humidity R_h and the local pressure P_s. Because the availability of the climatic measurements varies spatially and temporally, the dataset contains several missing values. The total number of missing values decreases with respect to time, as seen in the Figure 1. Especially, a sharp decrease was detected between the years 1982 and 1987 for P_s

values. Suprisingly, in April 2002, a great increase has occured in the total number of the missing values. The variables of T_x, T_n, T_m and R_h have the same pattern of missing values. In Figure 1, T_x, T_n and R_h do not appear because of the overlapping lines.

During the missing data detection analyses, it was noticed that 18 meteorological stations had missing data with percentage higher than 5%. Distribution of these values was grouped in continuous months, which affect the missing value estimation and some homogeneity tests. Therefore, these stations were excluded from the study. The considered 232 stations had the climatic time series for five parameters for the time interval 1974–2002. The local pressure time series, however, were measured on 211 stations differing from the other five climatic parameters. The number of 229 stations decreased to 211 similarly after eliminating 18 stations. The data were collected on the monthly basis. The missing data analysis was then performed for five parameters for 232 stations and for P_s for 211 stations. The total numbers of missing monthly climatic measurements between 1974 and 2002 were plotted in Figure 1 for five climatic parameters for 232 stations and for P_s for 211 stations. The missing data were completed using two methods, the LR and the EM method.

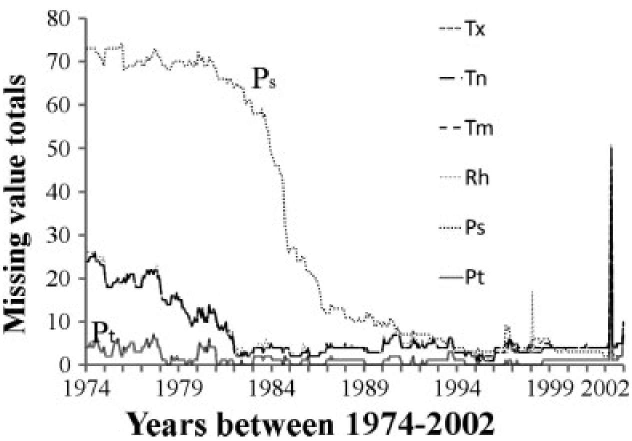

Figure 1. Missing value totals throughout the monthly recording period

The missing value interpolation

Because the set of Turkish meteorological data is incomplete, we estimated the monthly missing values for the further analysis. We applied two methods for the missing value interpolation. First one was the LR and the other one was the EM algorithm. The EM algorithm is based on the iterated LR analyses in order to compare how the iterative processing effects the missing value estimation (see Appendix for the details of the LR and EM algorithm).

We used the reference series to estimate the missing values of the time series in a given station. We used 16 stations that have missing values greater than 4% for an effective comparison. To compare the results, we computed the mean of variables, the mean of correlation coefficients between the reference and the test stations, standard deviation and the skewness of these 16 stations after filling in missing values with estimated values. As seen in Table I, both missing value estimation methods almost provided the same results except P_s. Although the mean of values and standard deviations were nearly the same, there was an important difference between mean of correlation coefficients and skewness for P_s variable. Generally, the skewness of the variables is very low.

When data of 18 stations were excluded from dataset, the number of missing values decreased significiantly as shown in Table II. Thus, EM algorithm and LR method became more applicable to our dataset and influences on homogeneity tests were minimized. For example, number of missing values for precipitation series decreased to 0·62% from 2·1% as seen in the Table II.

We checked temperature variables in order to determine the physically non-plausible values that mean maximum temperature is cooler than minimum temperature, after and before missing value analysis. During analyses, no physically non-plausible values were detected.

The geographical distribution of 250 stations is shown in Figure 2. The stations were numbered from 1 to 250 to point their locations. Complete information table for these stations was omitted here due to the scarcity of

Table I. Comparison of missing value estimation methods applied to monthly Turkish meteorological data set

Variable	Num. of stations (mis. val. >4%)	Num. of esti. mis.val.	Mean of variables	Mean of corr. coef.	SD	Skewness
P_t (EM)	16	275	43·057 mm	0·858	0·079	−1·362
P_t (LR)			43·142 mm	0·857	0·078	−1·375
T_x (EM)	16	480	16·547 °C	0·994	0·004	−1·613
T_x (LR)			16·546 °C	0·994	0·004	−1·51
T_n (EM)	16	479	3·803 °C	0·99	0·007	−1·774
T_n (LR)			3·807 °C	0·989	0·007	−1·671
T_m (EM)	16	492	10·090 °C	0·995	0·004	−2·368
T_m(LR)			10·088 °C	0·995	0·004	−2·203
R_h (EM)	16	490	63·479%	0·86	0·104	−1·947
R_h (LR)			63·600%	0·856	0·105	−1·948
P_s (EM)	25	1916	923·35 mbar	0·679	0·296	−0·884
P_s (LR)			925·74 mbar	0·508	0·298	0·293

Table II. Basic statistics of monthly data and total number of missing values after omitting 18 stations. 'Total 1' denote the total nunber of missing values for 250 stations (229 for P_s), 'Total 2' denote the total number of missing values for 232 stations (211 for P_s) and 'Percent 1' and 'Percent 2' denote the percentage of missing values

Variable	Max.	Min.	Mean	Total 1	Percent1 (%)	Total 2	Percent2 (%)
Precipitation P_t (mm)	907·2	0	51·169	1805	2·1	498	0·62
Maximum air temperature T_x (°C)	45·687	−12·245	18·621	2530	2·9	773	0·96
Mean air temperature T_n (°C)	37·374	−18·384	12·842	2580	3·0	784	0·97
Minimum air temperature T_m (°C)	28·794	−26·357	8·039	2539	2·9	779	0·96
Relative humudity R_h (%)	95·618	1·442	64·473	2612	3	850	1·05
Pressure P_s (mbar)	1029·2	761·5	938·14	10 241	13·9	8584	11·7

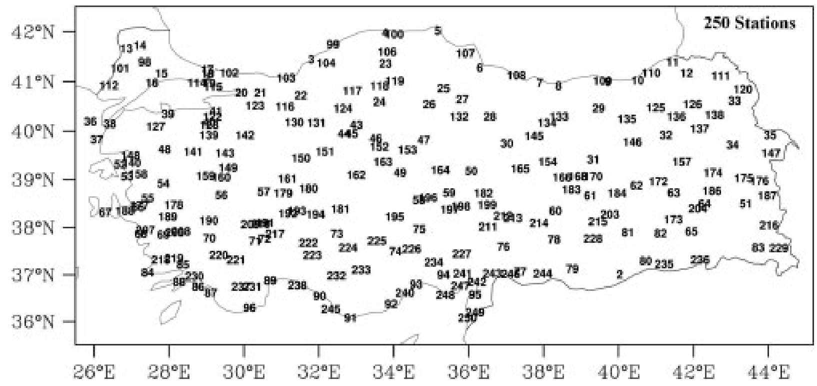

Figure 2. Geographical distribution of 250 stations of Turkey

space. Numbers which do not appear on the Figure 2 belong to the overlapping stations.

THE ABSOLUTE HOMOGENEITY TESTS

The absolute tests applied in this study were commonly used in the climatology to detect inhomogeneities in the meteorological time series. These absolute tests were selected on the basis of their different sensitivities where the break is likely to be expected. Machiwal and Jha (2008) evaluated the performance of three statistical homogeneity tests. Based on the known physical parameters affecting the homogeneity, the cumulative deviations and the Bayesian tests were found to be superior to the classical Von Neumann test. A similar finding was reported by Buishand (1982). Zanchettin *et al.* (2008) used the Craddock test of cumulative deviations (Craddock, 1979) for checking the homogeneity of the reconstructed normalized time series of the discharge and the precipitation measured in Po River, Italy. Rusticucci and Renom (2008) checked the homogeneity of the temperature variables measured in 17 stations in the different periods and at the different significance levels. The three homogeneity tests used in their work were, the SNHT, the Buishand range test and the homogeneity test proposed by Vincent (1998). Evaluating the breaks detected as well as for the future attempts to correct the series from the artificial steps was poor due to the lack of metadata support. The other most recent studies are: Wijngaard *et al.* (2003), Wulfmeyer *et al.* (2006) and Feng *et al.* (2004). The selected four absolute test methods for this ongoing study are recommended by Wijngaard *et al.* (2003): the SNHT for a single break (Alexandersson, 1986), the Buishand range test (Buishand, 1982), the Pettitt test (Pettitt, 1979), and the Von Neumann ratio test (Von Neumann, 1941). The details of these four absolute test methods are explained in detail in Appendix. A classification is made depending on the number of the absolute tests rejecting the null hypothesis as in Wijngaard *et al.* (2003). If one or zero tests reject the null hypothesis at the 1% level, we labelled it class 1 which means 'useful'. If two tests reject the null hypothesis at the 1% level, we labelled it class 2 which means 'doubtful' and if three or four tests reject the null hypothesis at the 1% level we labelled it class 3 which means 'suspect'.

Null hypothesis (H_0). All these tests suppose that the annual values Y_i of the testing variable Y are independent and identically distributed.

Alternative hypothesis (H_1). The SNHT, the Buishand range and the Pettitt test assume that a step-wise shift in the mean (a break) is present. Von Neumann ratio test assumes that the series is not randomly distributed.

Wijngaard *et al.* (2003) presented the detailed mathematical developments of the tests. We simply presented the main lines of these tests.

THE RELATIVE HOMOGENEITY TEST

The bivariate test

In climatology, the bivariate test has many applications to detect the inhomogeneities in stations' time series.

Potter (1981) tested the homogeneity of the annual precipitation series from the northeast United States using the bivariate test. Bücher and Dessens (1991) used the same test to check the inhomogeneities in the surface temperature time series from the Pyrenees in France. The bivariate test has also been used to demonstrate abrupt shifts in climatological variables that have a climatic origin. Lettenmaier *et al*. (1994) used the bivariate test to evaluate the relative changes in the streamflow relative to the precipitation, the streamflow relative to the temperature and the precipitation relative to the temperature across the United States. Gan (1995) computed similar analysis comparing the precipitation versus maximum temperature for Canada and North-eastern USA. Beaulieu *et al*. (2008) compared the performance of the bivariate test and the seven other relative homogeneity tests after an extensive literature review was performed. The bivariate test showed the second best performance for the absolute errors in the position and the magnitude in the case of series with a single shift. On series with two shifts, the bivariate test had the best performance for the absolute errors in position and magnitude. For the series with three shifts, the bivariate test was preferred because the bayesian test which showed better performance also detected a high number of nonexistent shifts in the homogeneous time series. Kirono and Jones (2007) described the use of the bivariate test for detecting and adjusting discontinuities in Class A pan evaporation time series for 28 stations across Australia. The results showed that 92% of the inhomogeneities detected by the bivariate test is consistent with the station metadata.

The bivariate test determines whether there is a shift in the mean of one station relative to a reference series, neighbouring stations or a regionally representative series. This test assumes that n two dimensional random vectors $\{x_i, y_i\}$ are serially independent and normally distributed. See Appendix for the details of the bivariate test.

Creating reference series

To meet the requirements of the bivariate test, reference series must be built with closely related stations. To provide this, weighted averages of the reference stations' time series must be calculated. This is supplied using the distance or correlation coefficients between the nearest stations. Taking weighted averages with the distance between the nearest stations to create reference series preserves geographical vicinity. However, using correlation coefficents as weight factors to create reference series provide mostly correlated but similar inhomogeneities with the tested series. Besides, it is not possible to create reference series for each test station due to the low correlations between the test and the reference stations in Turkey (Table III). For these reasons, we used the distance from the nearest stations as the weight factor to create the reference series. In this work, we generated the following formula to build the reference series:

Table III. Max/min/mean correlation coefficients, standard deviation and skewness between test end reference series

Variable	Num. of stations	Correlation coefficients Max.	Min.	Mean	Standard deviation	Skewness
P_t	232	0·977	0·288	0·853	0·102	−2·655
T_x	232	0·999	0·962	0·995	0·006	−2·793
T_n	232	0·998	0·94	0·991	0·007	−2·763
T_m	232	0·999	0·973	0·996	0·005	−2·787
R_h	232	0·974	−0·012	0·754	0·199	−1·413
P_s	211	0·997	−0·66	0·607	0·359	−0·787

$$G_i = \frac{\sum_{j=1}^{k_i} \frac{1}{d_j^2} Q_{ij}}{\sum_{j=1}^{k_i} \frac{1}{d_j^2}} \qquad i = 1, \ldots\ldots, n \qquad (1)$$

where n is the number of years with G_i being the weighted average of the nearest stations, k_i is the number of reference sites used in the time step i. In this work, we limited the maximum number of the reference sites with 5. d_j is a weight factor for the jth reference series that was defined as the square of the normalized distance between the test series and jth reference series. Q_{ij} the measured value at the jth reference station.

As given in Table III, the correlation coefficents between the test and the reference series were excellent for the temperature variables. However, we encountered low correlations between the test and the reference series, especially in P_s and R_h reference series. In order to achieve reliable homogeneity test results, we used a lower treshold of 0·7 for the correlation value between the test and the candidate reference stations.

COMPARISON OF THE RESULTS OBTAINED BY THE RELATIVE AND THE ABSOLUTE HOMOGENEITY TESTS

Applying both relative and absolute tests gave us the opportunity to compare their results and indicated why the relative tests were considered to be more powerful than the absolute tests. We selected station 193 as an example where relative and absolute tests found different results. Figure 3 shows the distribution of the annual T_x and its reference series values for the period 1974–2002. The correlation coefficient between the test and the reference series was 0·97. In 1993 (showed with square symbol), there is a sudden increase observed in both T_x and its reference series values and the rapid variation in the mean of T_x. It can be concluded that this variation was caused by climatic factors, because the same increase was observed in the reference series. As expected, absolute tests, which used only the single station series were failed

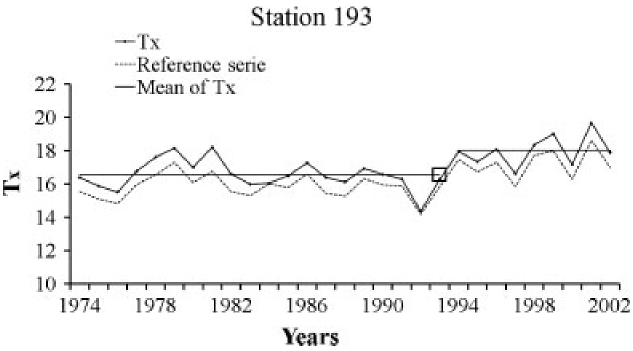

Figure 3. The test and the reference series of the station 193

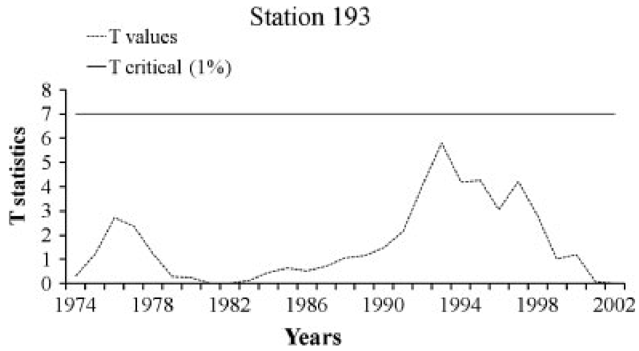

Figure 4. The bivariate test results

to distinguish this real climate variation under the null hypothesis at the 1% significance level.

The T statistics of the bivariate test reached to its maximum in 1993 and did not exceed the critical value at significance level 1% as shown in Figure 4. As a result, bivariate test found these stations T_x time series as homogeneous. Similarly, T values calculated with absolute tests reached to its maximum in 1993. Note that Von Neumann ratio test does not provide information on the year of break. Because three of the four absolute tests rejected the null hypotesis at the significance level 1%; this station was defined as inhomogeneous by the absolute tests. Only the Buishand range test did not reject the null hypothesis at the significance level 1%. All of the tests results applied to T_x variable of station 193 were given in Table IV.

Although absolute tests found clear sign of inhomogeneity, this stations T_x series was found to be homogenous by the bivariate test. Because test and reference series are sufficiently correlated (>0·7), we considered the source of inhomogeneity as a climate variability and accepted it as homogenous. We applied the same procedure during the homogeneity analysis in all stations.

Annual total precipitation of each station was tested by the relative and the absolute tests and the results are given in Table V. Von Neumann ratio test results were shown with X symbol because this test does not give information on the year of break. Absolute tests failed to detect the inhomogeneities in the annual precipitation time series at significance level 1%. Of 232 stations, 2 were found to be inhomogeneous and

Table IV. Homogeneity test results applied to T_x variable measured at station 193

Test	T statistic	T critical (1%)	Break year	Result
Bivariate	5·808	6·95	1993	Passed
Snht	10·652	10·45	1993	Rejected
Buishand	1·51	1·7	1993	Passed
Pettitt	140	133	1993	Rejected
V.Neumann	1·161	1·2	—	Rejected

labelled 'suspect' according to the absolute test results. These stations numbers were 186 and 187. Absolute test results indicated that such discontinuities are difficult to detect in annual total precipitation at the significance level 1% recorded in Turkey. Reducing significance level to 0·5% could be capable of detecting inhomogeneities (Karabork *et al.*, 2006); however, 93·96% of the test series were sufficiently correlated with reference series, so we used bivariate test as a first option in these stations. The geographical distribution of the inhomogeneous precipitation stations is shown in Figure 5.

According to the bivariate test, 30 of 232 stations were found to be inhomogeneous. Interestingly, most of the inhomogeneities were found close to end of the series. Vivès and Jones (2005) showed that the bivariate test is sensitive to the deviations in the mean near the beginning and the end of the time series. Therefore, the test should be applied with caution in these situations. An alternative is to use the metadata or another statistical technique to confirm the inhomogeneity. Unfortunately, these inhomogeneities could not be explained due to the lack of the metadata. Some researchers do not accept unexplained inhomogeneities 5 years or more from the end of the series (Gonzàlez-Rouco *et al.*, 2000; Göktürk M.O, *et al.*, 2008) because of an increased probability for high T values near the ends (Hawkins, 1977). The usage decision of these stations is left to the user.

In Table VI, the results of the SNHT, Buishand, Pettitt, Von Neumann and the bivariate test applied to all variables are given. It is noticeable that total numbers of the inhomogeneities detected by the bivariate test were always greater than the number of stations labelled 'suspect'. However, the SNHT test detected more inhomogeneities according to the other absolute tests for the temperature variables. Because the four absolute tests used in the homogeneity analysis have different sensitivities to different changes in a station's data series, the results from these methods sometimes have discrepancies. Similar problems were also reported with these methods (Wijngaard *et al.*, 2003; Feng *et al.*, 2004).

Figure 6 shows the test results for the temperature variables. The black dots denote that the two or more of the temperature variables have inhomogeneities, whereas the empty shapes of black dots denote that only a single temperature variable has inhomogeneities. Interestingly, as seen in Figure 6, most of the black dots were gathered in the left side of the map, which is in the Aegean region.

Table V. List of inhomogeneous precipitation time series

Station number	Latitude	Longitude	Break years detected by relative and absolute tests				
			Bivariate	SNHT	Buishand	Pettitt	V. Neumann
5	42·02	35·17	1982	—	—	—	—
9	41·00	39·72	1987	—	—	1987	—
15	41·17	27·80	2001	—	1981	—	—
33	40·62	43·10	2000	—	—	—	—
36	40·20	25·90	2001	—	—	—	X
40	40·18	29·07	2001	—	—	—	—
56	38·68	29·40	2001	—	—	—	—
57	38·75	30·53	1974	—	—	—	—
58	38·58	34·67	2000	—	—	—	—
98	41·40	27·35	2001	—	—	—	—
112	40·93	26·40	2001	—	—	—	X
118	40·92	33·63	1975	1975	—	—	X
125	40·48	41·00	2000	—	—	—	—
130	40·18	31·35	1998	—	1981	—	X
153	39·62	34·37	2001	—	—	—	—
158	39·12	27·18	2001	—	—	—	X
159	39·08	28·98	1981	1981	—	1981	—
182	38·72	36·40	1981	—	—	1981	—
183	38·80	38·75	1999	—	—	—	—
186	38·46	42·30	1996	1996	—	1994	X
187	38·67	43·98	1983	—	1983	1983	X
189	38·23	27·97	2001	—	—	—	—
190	38·15	29·07	1998	—	—	—	—
192	38·30	31·18	2001	—	—	—	—
196	38·63	34·92	2001	—	—	—	—
198	38·45	35·80	2000	—	—	—	—
220	37·42	29·33	2001	—	—	—	—
231	36·75	30·20	2001	—	—	—	—
232	36·98	32·47	1987	—	—	1987	—
233	37·11	33·13	1981	—	—	—	—

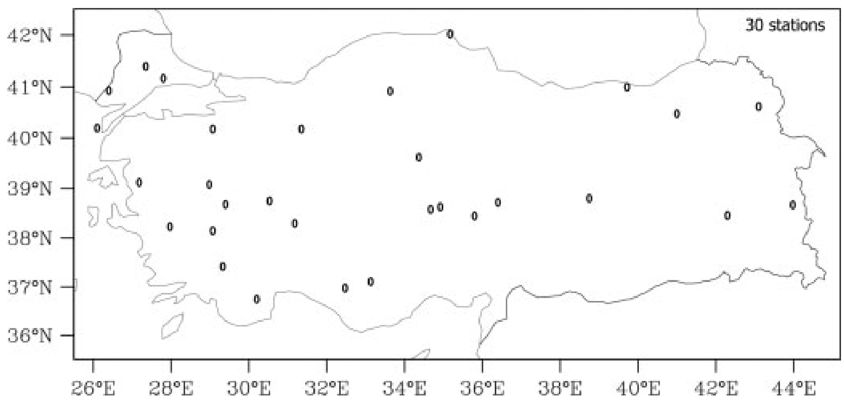

Figure 5. Geographical distribution of inhomogeneous precipitation stations

In particular, most of the black circles that symbolize the number of the inhomogeneities detected in all temperature variables were existed in the western and middle subregions of Turkey's Mediterranean and Aegean coastal belt. The causes of these inhomogeneities in temperature series were considered the relocation of stations, trends and in terms of the long-term variations (Türkeş *et al.*, 2002b). The sign and magnitude of the trends in those series do not depict any natural inter-seasonal differences after the relocation occurred because the artificial cooling (warming) effect of the station's relocation has suppressed the natural variability, and in particular, the trend characteristics of the temperature series at that station (Türkeş *et al.*, 2002b). These inhomogeneites must be carefully analysed for quantitative climate analyses.

Figure 7 shows the test results of P_s and R_h. The black square dots denote that P_s and R_h have inhomogeneities, the black circle dot denotes that R_h time series have inhomogeneities and the empty circle denotes that P_s have homogeneities in the corresponding stations time series. The results show that the inhomogeneities of these stations scatter randomly.

Table VI. Relative and absolute homogeneity test results applied to all variables. The values in paranthesis denote the percentage of total number of stations. Absolute tests results were set in the categories 'useful', 'doubtful', and 'suspect'. Total number of stations was 232 (for *Ps* 211)

Variable	SNHT	Buishand test	Pettitt test	V. Neumann Test	Class 1 'useful'	Class 2 'doubtful'	Class 3 'suspect'	Bivariate test
P_t	3	3	6	7	227(98%)	3(1%)	2(1%)	30(13%)
T_x	113	44	99	102	129(56%)	42(18%)	61(26%)	97(42%)
T_n	88	54	92	59	150(65%)	52(22%)	30(13%)	107(46%)
T_m	83	21	69	43	177(76%)	26(11%)	29(13%)	69(30%)
R_h	137	150	134	194	60(26%)	37(16%)	135(58%)	192(83%)
P_s	87	82	99	104	114(49%)	23(10%)	74(32%)	161(69%)

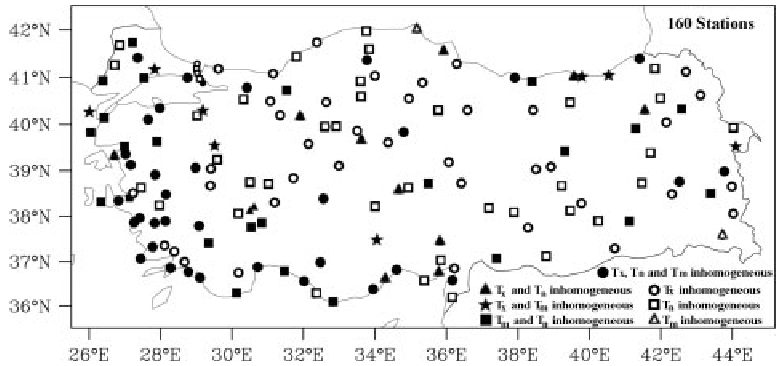

Figure 6. Geographical distribution of inhomogeneous temperature variables

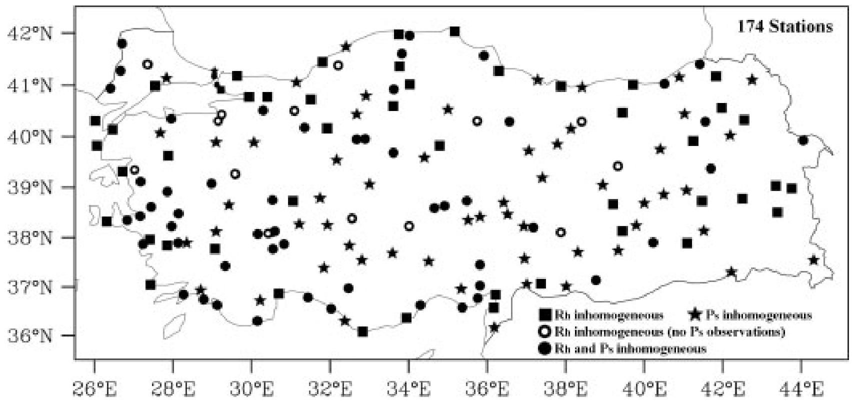

Figure 7. Geographical distribution of inhomogeneous Ps and Rh variables

DISCUSSION AND CONCLUSION

In this work, homogeneity tests and missing value interpolation of the meteorological time series of Turkey were carried out. This is the most comprehensive study in the international literature examining the homogeneity of six different meteorological variables using both relative and absolute homogeneity tests. However, it is the first countrywide detailed homogeneity analysis in Turkish meteorological data.

We used the bivariate test developed by Maronna and Yohai (1978) to apply the relative homogeneity test, and we used the SNHT for a single break (Alexandersson, 1986), the Buishand range test (Buishand, 1982), the Pettitt test (Pettitt, 1979) and the Von Neumann ratio test (Von Neumann, 1941) to apply the absolute homogeneity tests. The most common approach for building a reference time series is to calculate for each year a weighted average of the data from the neighbouring stations or sections of the neighbouring station time series using the correlation coefficients. In this study, we used the distance between the test and the reference stations as a weight factor while creating reference series. Using the nearest stations and the best-correlated series at the same time is an ideal and desired situation while taking weighted averages. Usually, the nearest station is not the best-correlated station with the test station in the Turkish meteorological data.

As a result, the employment of the distance between the test and the reference stations caused lower correlated series but it preserved geographical vicinity. Because of the excellent correlations for the temperature variables (Table III), the use of the distance of the nearest stations as a weight factor was considered to be the best option for creating reference series for the temperature time series. If the correlation coefficents between the test and reference series were under 0·7, absolute homogeneity test results were used.

According to the bivariate test, 30 of 232 stations precipitation times were found to be inhomogeneous which was an expected case. However, only a few stations have abrupt changes in the series; thus, most of the statistical inhomogeneities are very likely related to the long-period fluctuations and significant trends, both of which are accepted within other non-randomness characteristics of the series of climatological observations (WMO, 1966; Sneyers, 1990, 1992; Türkeş, 1996, 1999; Türkeş *et al*., 2002a,b).

Most of the inhomogeneities detected in temperature variables existed in the Mediterranean and the Aegean coastal belt. A non-climatological jump in the mean of these temperature series may result either from an abrupt change associated with relocation of a station or from a steep trend (a rapid increase or decrease) in temperature values because of different factors, such as urbanization (i.e. urban heat island effect or urban cooling effect, respectively) (Türkeş *et al*., 2002). It is considered that these inhomogeneities were mostly caused by non-natural effects such as relocation. Due to changes at topography at short distance in this region intensify non random characteristics of the temperature series when relocation occurs even in small distances. The marine effect, which causes artifical cooling effect due to sea breezes has important impact on temperature series (Türkeş, 1999; Türkeş *et al*., 2002b) and the orograhpy allows this impact go through the inner parts in this region.

As a missing value interpolation procedure, we applied two methods: LR and EM algorithm. Because missing values of 18 stations were grouped, we excluded data of these stations in order to obtain more accurate results. Using more then one method provided us the opportunity to compare their results. Generally, results of these two methods were nearly the same except the *Ps* variable. It can be concluded that EM algorithm results were more reliable than LR results, considering correlation coefficents between the test and the reference series.

APPENDIX

LR method

Interpolating the missing values with LR is a simple and convenient method and shows good performance. A simple linear model can be defined as below:

$$y = \beta_0 + \beta_1 x \qquad (2)$$

where β_0 is the intercept β_1 and is the slope. Let the difference between the observed value of y and the straight line $(\beta_0 + \beta_1 x)$ be the error ε. It is convenient to think of ε as a statistical error; that is, it is a random variable that accounts for the failure of the model to fit the data exactly.

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (3)$$

Equation (3) is called an LR model. Customarily x is called the independent (regressor) variable and y is called the dependent variable. Because Equation (3) involves only one regressor variable, it is called a simple LR model.

The EM algorithm

The EM algorithm, like all methods for incomplete data that ignore the mechanism causing the gaps in the data set, rests on the assumption that the missing values in the dataset are missing at random, in the sense that the probability that a value is missing does not depend on the missing value (Rubin, 1976). A short description of EM algorithm is outlined below, for more properties and details of EM algorithm, please see Schneider (2001). Let $X \in R^{n\times p}$ be a data matrix, n the number of the records and p is the variables with missing values. The objective of the EM algorithm is to estimate the mean $\mu \in R^{1\times p}$of the records and the covariance matrix $\Sigma \in R^{p\times p}$ from the incomplete data set. The relationship between the variables with missing values and the variables with available values is modelled with a LR model;

$$x_m = \mu_m + (x_a - \mu_a)B + e \qquad (4)$$

x being the record where $x = X_i$ $(i = 1, \ldots, n)$ with missing values and $x_m \in R^{1\times p_m}$ is a vector consisting of the remaining p_{m} values for which, in the given record, the values are missing, $x_a \in R^{1\times p_m}$ consist of the variables, in the given record, the values are available, $\mu_a \in R^{1\times p_m}$ is the partitioned part of the mean μ for the available values, and the part $\mu_m \in R^{1\times p_m}$ is the mean values of the variables in the given record, the residual $e \in R^{1\times p_m}$ is a vector with mean zero, the matrix $B \in R^{p_a\times p_m}$ is a matrix of regression coefficients. In each iteration of the EM algorithm, estimates of the mean μ and of the covariance matrix Σ are taken as given, and from these estimates, the conditional maximum likelihood estimates of the matrix of regression coefficients B and of the covariance matrix $C \in R^{p_m\times p_m}$ of the residual are computed for each record with missing values. With the estimated regression model for each record, the missing values are then filled in with the imputed values, and the new estimates of the mean μ and of the covariance matrix Σ are computed from the completed dataset and from the estimates of the residual covariance matrices C.

SNHT

$\overline{Y}$ is the mean and Y_i is the annual series to be tested (i is the year from 1 to n), s the standard deviation. The test statistic $T(k)$ is defined by Alexandersson (1986) as following:

$$T(k) = k\overline{z}_1^2 + (n-k)\overline{z}_2^2 \quad k = 1, \ldots .n \qquad (5)$$

where

$$\bar{z}_1 = \frac{1}{k}\sum_{i=1}^{k}(Y_i - \overline{Y})/s \text{ and}$$

$$\bar{z}_2 = \frac{1}{n-k}\sum_{i=k+1}^{n}(Y_i - \overline{Y})/s \quad (6)$$

The mean of the first k years and the last $n-k$ years of the record is compared. T(k) reaches its maximum value when a break is located at the year K. The distribution of T(k) according to years is depicted in the graphs to represent the results. The test statistic T_0 is defined as:

$$T_0 = \max_{1\le k\le n} T_{(k)} \quad (7)$$

If T_0 exceeds the critical value, the null hypothesis will be rejected. As seen in Table VII, the critical values are dependent on the sample size.

The SNHT is more sensitive to breaks near the beginning and the end of a series relatively easily.

Buishand range test

This test is calculated as follows:

$$S_0^* = 0 \text{ and } S_k^* = \sum_{i=1}^{k}(Y_i - \overline{Y}) \qquad k = 1, \ldots, n \quad (8)$$

The term of S_k^* is the partial sum of the given series. If there is no significant change in the mean, the difference between Y_i and $\overline{Y}$ will fluctuate around zero. The significance of the change in the mean is calculated with 'rescaled adjusted range' R, as the following:

$$R = (\max_{o\le k\le n} S_k^* - \min_{o\le k\le n} S_k^*)/s \quad (9)$$

Buishand (1982) were calculated the critical values for $R/\sqrt{n}$ (Table VIII).

The Buishand range test is more sensitive to the breaks in the middle of a time series (Hawkins, 1977).

Pettitt Test

Pettitt test is a non-parametric test based on the Wilcoxon test (Pettitt, 1979). It can also be derived from the Mann–Whitney U-test. The ranks $r_1, \ldots r_n$ of the $Y_1, \ldots Y_n$ are used to calculate the statistics:

$$X_k = 2\sum_{i=1}^{k} r_i - k(n+1) \qquad k = 1, \ldots, n \quad (10)$$

If a break occurs in year E, the absolute value of X_k reaches to its maximum.

$$X_E = \max_{1\le k\le n} |X_k| \quad (11)$$

Critical values were given by Pettitt (1979) as given in Table IX.

This test is more suitable for detecting the breaks near the middle of the series

Von Neumann ratio test

The Von Neumann ratio was defined as:

$$N = \sum_{i=1}^{n-1}(Y_i - Y_{i+1})^2 \Big/ \sum_{i=1}^{n}(Y_i - \overline{Y})^2 \quad (12)$$

If the sample contains a break, then the value of N tends to be lower than this expected value (Buishand, 1981). If the sample has rapid variations in the mean, then values of N may rise above 2 (Bingham and Nelson, 1981). Only this test does not give information on the year of break. Table X gives critical values for N.

Bivariate test

Null hypothesis (H_0). $\{x_i, y_i\}$ have the same bivariate normal distribution, $N(\mu_x, \mu_y, \sigma_x{}^2, \sigma_y{}^2, \rho)$, with all parameters unknown.

Alternative hypothesis (H_1). For some $0 < i_0 < n$ and $d \ne 0$, the distribution of $\{x_i, y_i\}$ is $N(\mu_x, \mu_y, \sigma_x{}^2, \sigma_y{}^2, \rho)$, for $i \le i_0$ and is $N(\mu_x, \mu_y + d, \sigma_x{}^2, \sigma_y{}^2, \rho)$ for $i > i_0$.

It is necessary to standardize the regional series $\{x'j\}$ and the test series $\{y'j\}$ of the length n by their mean and standard deviation to be able to use the critical values of statistic T_0.

Table VII. 1% Critical values for the statistic T_0 of the single shift SNHT as a function of n (calculated from the simulations carried out by Jaruskova (1994))

n	20	30	40	50	70	100
1%	9·56	10·45	11·01	11·38	11·89	12·32

Table VIII. 1% Critical values for of $R/\sqrt{n}$ the Buishand range test as a function of n (Buishand, 1982); the value of $n = 70$ is simulated

n	20	30	40	50	70	100
1%	1·60	1·70	1·74	1·78	1·81	1·86

Table IX. 1% Critical values for X_E of the Pettitt test as a function of n; values are based on simulation

n	20	30	40	50	70	100
1%	71	133	208	293	488	841

Table X. 1% Critical values for N of the Von Neumann ratio test as a function of n. For $n \le 50$ these values are taken from Owen (1962); for $n = 70$ and $n = 100$ the critical values are based on the asymptotic normal distribution of N (Buishand, 1981)

n	20	30	40	50	70	100
1%	1·04	1·20	1·29	1·36	1·45	1·54

Table XI. 1% Critical values for the statistic T_0 of the bivariate test as a function of n (Potter, 1981)

n	20	30	40	50	70	100
1%	1·04	1·20	1·29	1·36	1·45	1·54

Let

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} x'_j,\ \overline{Y} = \frac{1}{n}\sum_{j=1}^{n} y'_j \quad (13)$$

$$S_x = \left[\frac{1}{n}\sum_{j=1}^{n}(x'_j - \overline{X})^2\right]^{1/2} \text{ and}$$

$$S_y = \left[\frac{1}{n}\sum_{j=1}^{n}(y'_j - \overline{Y})^2\right]^{1/2} \quad (14)$$

$$x_j = \frac{(x'_i - \overline{X})}{S_x},\ \mathrm{y}_j = \frac{(y'_i - \overline{Y})}{S_y} \text{ for all } i < n \quad (15)$$

Computation of the test statistics:

$$\text{Let } X_i = \frac{1}{i}\sum_{j=1}^{i} x_j \text{and } Y_i = \frac{1}{i}\sum_{j=1}^{i} y_j \text{for all } i < n \quad (16)$$

$$S_{xy} = \sum_{j=1}^{n} x_j y_j \quad (17)$$

$$F_i = n - \frac{[X_i^2 ni]}{(n-i)} \text{ for all } i < n \quad (18)$$

$$D_i = \frac{(S_{xy}X_i - nY_i)n}{(n-i)F_i} \text{ for all } i < n \quad (19)$$

$$T_i = \frac{i(n-i)D_i^2 F_i}{(n^2 - S_{xy}^2)} \text{ for all } i < n \quad (20)$$

$T_0 = \max_{i<n}[T_i]$ and i_0 is the value of i for which T_i is a maximum.

Maronna and Yohai (1978) computed the T_i critical values by simulation for $n = 10$, 15, 20, 30 and 70 and Potter (1981) extended the results to $n = 100$, under the null hypothesis N (0,0,1,1,ρ). These are given in Table XI for the significance level 1%.

REFERENCES

Alexandersson H. 1986. A homogeneity test applied to precipitation data. *Journal of Climate* **6**: 661–675.

Beaulieu C, Seidou O, Ouarda J, Zhang X, Boulet G, Yagouti A. 2008. Intercomparison of homogenization techniques for precipitation data. *Water Resources Research* **44**: W02425.

Bingham C, Nelson LS. 1981. An approximation for the distribution of the Von Neumann ratio. *Technometrics* **23**: 285–288.

Bücher A, Dessens J. 1991. Secular trend of surface temperature at an elevated observatory in the Pyrenees. *Journal of Climate* **4**: 859–868.

Buishand TA. 1981. The analysis of homogeneity of long-term rainfall records in the Netherlands. KNMI Scientific Report, WR 81-7, De Bilt, The Netherlands.

Buishand TA. 1982. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology* **58**: 11–27.

Conrad V, Pollak C. 1950. *Methods in Climatology*. Harvard University Press: Cambridge; 459.

Craddock JM. 1979. Methods for comparing annual rainfall records for climatic purpose. *Weather* **34**: 332–346.

Feng S, Hu Q, Qian W. 2004. Quality control of daily meteorological data in China, 1951–2000: a new dataset. *International Journal of Climatology* **24**: 853–870.

Gan TY. 1995. Trends in air temperature and precipitation for Canada and north-eastern USA. *International Journal of Climatology* **15**: 1115–1134.

Göktürk OM, Bozkurt D, Şen OL, Karaca M. 2008. Quality control and homogeneity of Turkish precipitation data. *Hydrological Processes* **22**: 3210–3218.

Gonzàlez-Rouco JF, Jiménez JL, Quesada V, Valero F. 2000. Quality control and homogeneity of precipitation data in the Southwest of Europe. *Journal of Climate* **14**: 964–978.

Hawkins PM. 1977. Testing a sequence of observations for a shift in location. *Journal of American Statistical Association* **72**: 180–186.

Jones PD, Raper SCB, Santer BD, Cherry BSG, Goodess CM, Kelly PM, Wigley TML, Bradley RS, Diaz HF. 1985. A Grid Point Surface Air Temperature Data Set for the Northern Hemisphere. Technical Report. TRO22, U.S. Department of Energy, Carbon Dioxide Research Division: Washington, DC; 251 pp.

Karabork MC, Kahya E, Komuscu AU. 2007. Analysis of Turkish precipitation data: homogeneity and the Southern Oscillation forcings on frequency distributions. *Hydrological Processes* **21**: 3203–3210.

Karl TR, Williams CN Jr. 1987. An approach to adjusting climatological time series for discontinuous inhomogeneities. *Journal of Climate and Applied Meteorology* **26**: 1744–1763.

Kirono DGC, Jones RN. 2007. A bivariate test for detecting inhomogeneities in pan evaporation time series. *Australian Meteorological Magazine* **56**: 93–103.

Lettenmaier DP, Wood EF, Wallis JR. 1994. Hydro-climatological trends in the continental United States, 1948–1988. *Journal of Climate* **7**: 586–607.

Machiwal D, Jha MK. 2008. Comperative evaluation of statistical tests for time series analysis: application to hydrological time series. *Hydrological Sciences [Journal-des Sciences Hydrologiques]* **53**(3): 353–366.

Maronna R, Yohai VJ. 1978. A bivariate test for the detection of a systematic change in mean. *Journal of American Statistics Assocation* **73**: 640–645.

Owen DB. 1962. *Handbook of Statistical Tables*. Addison Wesley: Reading, UK.

Peterson TC, Vose R, Schmoyer R, Razuvaev V. 1998. Global Historical Climatology Network (GHCN) quality control of monthly temperature data. *International Journal of Climatology* **18**: 1169–1179.

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

Potter KW. 1981. Illustration of a new test for detecting a shift in mean precipitation series. *Monthly Weather Review* **109**: 2040–2045.

Rubin DB. 1976. Inference and missing data. *Biometrika* **63**: 581–592.

Rusticucci M, Renom M. 2008. Variability and trends in indices of quality-controlled daily temperature extremes in Uruguay. *International Journal of Climatology* **28**: 1083–1095.

Schneider T. 2001. Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *American Meteorological Society* **14**: 853–871.

Sneyers R. 1990. *On the Statistical Analysis of Series of Observations*. World Meteorological Organization (WMO): Geneva; 192. Technical Note No. 143.

Sneyers R. 1992. *Use and Misuse of Statistical Methods for the Detection of Climate Change*. In Climate Change Detection Project, Report on the Informal Planning Meeting on Statistical Procedures for Climate Change Detection, WCDMP, No. 20; J76–J81.

Tayanç M, Dalfes N, Karaca M, Yenigün O. 1998. A comparative assessment of different methods for detecting inhomogeneties in Turkish temperature data set. *International Journal of Climatology* **18**: 561–578.

Türkeş M. 1996. Spatial and temporal analysis of annual rainfall variations in Turkey. *International Journal of Climatology* **16**: 1057–1076.

Türkeş M, Sümer UM, Kılıç G. 1996. Observed changes in maximum and minimum temperatures in Turkey. *International Journal of Climatology* **16**: 463–477.

Türkeş M. 1999. Vulnerability of Turkey to desertification with respect to precipitation and aridity conditions. *Turkish Journal of the Engineering and Environmental Sciences* **23**: 363–380.

Türkeş M, Sümer UM, Kılıç G. 2002a. Persistence and periodicity in the precipitation series of Turkey and associations with 500 hPa geopotantial heights. *Climate Research* **21**: 59–81.

Türkeş M, Koç T, Sarı F. 2008. Spatiotemporal variability of precipitation total series over Turkey. *International Journal of Climatology* **29**: 1056–1074.

Türke§ M, Sümer M, Ysmail D. 2002b. Re-evaluation of trends and changes in mean, maximum and minimum temperatures of Turkey for the period 1929–1999. *International Journal of Climatology* **22**: 947–977.

Vincent LA. 1998. A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate* **11**: 1094–1104.

Vivès B, Jones RN. 2005. *Detection of Abrupt Changes In Australian Decadal Rainfall (1890–1989)*. CSIRO Atmospheric Research Technical Paper No: 73.

Von Neumann J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics* **13**: 367–395.

Wijngaard JB, Klein Tank AMG, Können GP. 2003. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology* **23**: 679–692.

WMO. 1966. *Climatic Change*. *WMO Technical Note*, No. 79, Geneva: Secretariat of the World Meteorological Organization (WMO), 79 pp.

Zanchettin D, Traverso P, Tomasino M. 2008. Po River discharges: a preliminary analysis of a 200-year time series. *Climatic Change* **89**: 411–433.