Journal of Agrometeorology 10(2) : 127- 135 (Dec. 2008)

Testing homogeneity, stationarity and trend in climatic series at Udaipur – a case study

DEEPESH MACHIWAL and ARUN SHARMA
SWE Department, College of Technology and Engineering, MPUAT, Udaipur – 313 001 (Raj.)
Email : dmachiwal@rediffmail.com

ABSTRACT

Three seasonal climatological time series, i.e., rainfall, minimum temperature, and maximum temperature of Udaipur in three seasons, i.e. summer, rainy and winter were tested for the presence of homogeneity, stationarity, and trend components. Box plot indicated normality in the rainy season rainfall. Minimum temperature series of rainy season was more uniform and normal than summer and winter season time series. Maximum temperatures during rainy season follow normal distribution. Homogeneity of seasonal time series was tested by using von-Neumann, Cumulative Deviations and Bayesian tests, which indicated that all seasonal rainfall and maximum temperature time series were homogenous. Homogeneity was present only in summer and rainy season minimum temperature time series. Stationarity and trends in the time series were detected by using Mann-Whitney test and Mann-Kendall test, respectively, which suggested that stationarity was present in all seasonal maximum and minimum temperature, and winter season rainfall time series and absent in summer and rainy season rainfall time series. Based on the results of Mann-Kendall tests, all the time series under study were free from any kind of trend.

Key words: Climatic time series, homogeneity, stationarity, trend.

The main intent of time series analysis is to detect and describe quantitatively each of the generating processes underlying a given sequence of observations (Shahin *et al.*, 1993). In hydrology, time series analysis is used for building mathematical models to generate synthetic hydrologic records, to forecast hydrologic events, to detect trends and shifts in hydrologic records, and to fill in missing data and extend records (Salas, 1993).

Most statistical analysis of hydrologic time series data at the usual time scale encountered in water resources planning studies are based on a set of fundamental assumptions, which are: the time series is homogenous, stationary, and free from trends. Homogeneity implies that the data in the series belong to one population, have a time invariant mean; non-homogeneity arises due to changes in the method of data collection and environment in which data collected (Fernando and Jayawardena, 1994). The feature of homogeneity tests is discussed by Buishand (1982) and Jayawardena and Lau (1990).

A time series is said to be strictly stationary, if its statistical properties do not vary with change of time origin and can not exhibit any trend. That is why, sometimes trend tests are utilized to check the stationarity of the hydrologic time series. Any trend in a time series can be expressed by suitable linear or non-linear model. However, the linear models are widely used in hydrological studies than the non-linear (Shahin *et al.*, 1993). Various parametric and non-parametric statistical tests have been reported in the literature for detecting the homogeneity (Buishand, 1982; Kanji, 2001), stationarity and trend (Jayawardena and Lai, 1989; Shahin *et al.*, 1993; Salas, 1993; Khan, 2001; Adeloye and Montaseri, 2002) in the hydrologic time series. However, the most powerful and widely-used tests for testing homogeneity are Cumulative Deviation, Bayesian, and von-Neumann tests (Machiwal and Jha, 2006). Similarly, Mann-Whitney test and Mann-Kendall test are reported to be superior over other tests for detecting stationarity and trend in a hydrologic time series (Machiwal and Jha, 2006). Therefore, these tests were used in present study.

MATERIALS AND METHODS

Data

Daily rainfall data of Udaipur (lat. 24°35' N, long.

73°42 E) for a period of 86 years (1921 to 2006) and daily temperature (maximum and minimum) data for a period of 30 years (1977 to 2006) were collected from the Agro-Meteorological Observatory of College of Technology and Engineering, Udaipur, Rajasthan and three seasonal climatological time series were developed. Three seasons in a year were considered as: rainy season (15 June to 15 October), winter season (16 October to 15 February), and summer season (16 February to 14 June). All the nine time series were tested for homogeneity, stationarity, and trend.

Graphical analysis of climatological time series

Time pattern of each individual climatological series was evaluated by plotting the time series over time scale. In addition box plots were drawn for comparing statistical characteristics, which indicate five important statistical properties of climatological time series such as median, largest value, smallest value, lower quartile upper quartile (USEPA, 1998).

Upper quartile: This mark of this line shows that 25% data above this mark and 75% data below this mark.

Lower quartile: This mark of this line shows that 75% data above this mark and 25% data below this mark.

Outliers: These marks, severe and mild outliers, show values, which respectively, are outside (far away or comparatively closer) the normal range or normal distribution range of data. Box plots were drawn by using STATISTICA (StatSoft, Inc., 2004) software.

Testing homogeneity of the time series

Homogeneity was tested by applying von-Neumann test, Cumulative Deviations test, and Bayesian test.

von Neumann test

The von Neumann ratio (N) was used to test the absence or presence of homogeneity, which is closely related to the first-order serial correlation coefficient (WMO, 1966) and can be defined as below:

$$N = \sum_{t=1}^{n-1}\left(x_t - x_{t+1}\right)^2 \Big/ \sum_{t=1}^{n}\left(x_t - \bar{x}\right)^2 \qquad ...(1)$$

Where, x_t = variable constituting the sequence in time; n = total number of records; and $\bar{x}$ = average of the x_t's.

For a homogenous time series, the expected value of the von Neumann ratio is 2 and it tends to be smaller than 2 for non-homogenous time series (Owen, 1962).

Cumulative deviations test

Tests for homogeneity are based on the adjusted partial sums or cumulative deviations from the mean, which are expressed as (Buishand, 1982):

, k = 1, 2, ..., n ...(2)

Rescaled adjusted partial sums (S_k^{**}) are obtained by dividing S_k^*'s by the sample standard deviation (D_x).

$$S_k^{**} = S_k^* \big/ D_x, \quad k = 1, 2, ..., n \qquad ...(3)$$

$$S_k^* = \sum_{t=1}^{k}\left(x_t - \bar{x}\right)$$

with $D_x^2 = \frac{1}{n}\sum_{t=1}^{n}\left(x_t - \bar{x}\right)^2$ ...(4)

The values of 's are not dependent on the unit of the variable, and hence homogeneity tests are based on the rescaled adjusted partial sums.

Sensitivity to the departures from homogeneity is defined by the following statistic:

$$Q = \max_{0 \le k \le n}\left|S_k^{**}\right| \qquad ...(5)$$

High values of Q indicate non-homogeneity in the time series. Another statistic, used for testing homogeneity is the range (R). It is defined as:

$$R = \max_{0 \le k \le n}\left|S_k^{**}\right| - \min_{0 \le k \le n}\left|S_k^{**}\right| \qquad ...(6)$$

Critical values of Q for specified values of 'n' are given by Buishand (1982), Doob (1949) while the critical values of the distribution of 'R' under the null hypothesis are given by Wallis and O'Connell (1973) in a graphical form.

Bayesian test

Bayesian test was developed by Chernoff and Zacks (1964) and later modified by Gardner (1969). The Gardener's test-statistic () for a two-sided test on a shift in the mean at an unknown point can be written as:

$$\tilde{G} = \sum_{k=1}^{n-1} p_k \left(S_k^* / \sigma_Y\right)^2 \quad ...(7)$$

Where, p_k = prior probability that the shift occurs just after k^{th} observation. For p_k independent of k, the test-statistic (U) can be expressed as:

$$U = \frac{1}{n(n+1)} \sum_{k=1}^{n-1} \left(S_k^{**}\right)^2 \quad ...(8)$$

However, for p_k proportional to $[k(n-k)]^{-1}$, the test-statistic can be written as follows:

$$A = \sum_{k=1}^{n-1} \left(Z_k^{**}\right)^2 \text{, k = 1, 2, ..., n} \quad ...(9)$$

Where = weighted rescaled partial sums and are computed as:

$$Z_k^{**} = \left[\left\{k(n-k)\right\}^{-1/2} S_k^*\right] / D_x \quad ...(10)$$

Large values of U and A test-statistics indicate departure from critical values (Buishand, 1982).

Testing stationarity by Mann-Whitney test

Stationarity of the time series was tested by applying Mann-Whitney test as defined by (Snedecor and Cochran, 1980):

$$u_c = \frac{\sum_{t=1}^{n1} R(x_t) - n1(n1+n2+1)/2}{\left[n1 n2 (n1+n2+1)/12\right]^{1/2}} \quad ...(11)$$

Where $R(x_t)$ is the rank of the observation x_t in ordered series z_t. n_1 and n_2 are sample sizes of two equal half subseries of the entire series.

Trend detection by using Mann-Kendall test

This is a nonparametric test Mann (1945), Kendall (1975) for exploring a trend in a time series without specifying the type of trend (i.e., linear or nonlinear). This test was reported as an excellent tool for trend detection (e.g., Hirsch et al., 1982; Gan, 1992). Considering the time series x_t (t = 1, 2, …, n), each value of the series (x_t) is compared with all subsequent values (x_{t+1}) and a new series z_k is generated as (Salas, 1993):

$$\begin{aligned} z_k &= 1 && \text{for } x_t > x_{t'} \\ z_k &= 0 && \text{for } x_t = x_{t'} \\ z_k &= -1 && \text{for } x_t < x_{t'} \end{aligned} \quad …(12)$$

Where, k is given by:

…(13)

The Mann-Kendall statistic (S) is defined as follows (Hirsch *et al.*, 1982):

$$S = \sum_{t'=1}^{n-1} \sum_{t=t'+1}^{n} z_k \quad …(14)$$

$$k = (t'-1)(2n-t')/2 + (t-t')$$

Moreover, the above test-statistic for n>40 may be written as (Hirsch *et al*., 1982):

$$u_c = \frac{S+m}{\sqrt{V(S)}} \quad …(15)$$

Where,

,

$$V(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{i=1}^{g} e_i (e_i - 1)(2e_i + 5)\right]$$

...(16)

In Eqns. (15) and (16), m = 1 for S<0 and m = -1 for S>0, g is the number of tied groups, and e_i is the number of data in the i^{th} tied group. The value of the test-statistic u_c is taken as zero for S = 0. Now, if the computed absolute value of u_c is greater than the critical value of the standard normal distribution, the hypothesis of an upward or downward trend cannot be rejected at the α significance level.

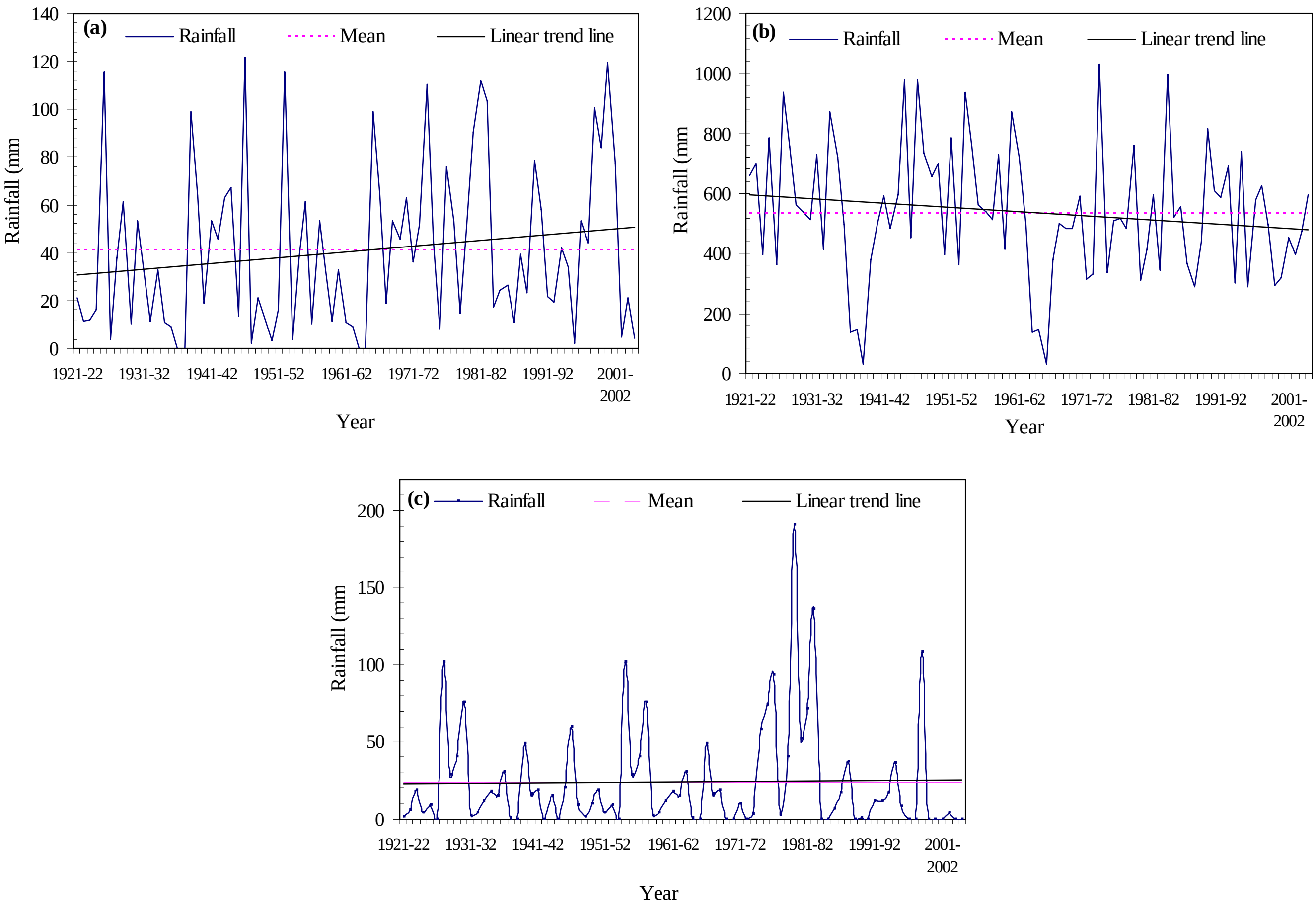

Fig. 1: Time patterns of rainfall time series in (a) summer, (b) rainy, and (c) winter seasons

RESULTS AND DISCUSSION

Time plots and box plots

A total of nine time plots and nine box plots were drawn for nine seasonal climatological time series of Udaipur. Variations in seasonal rainfall and temperature time series (i.e., summer, rainy, and winter) are observed though time plots and three of the time plots for rainy season rainfall, summer season maximum temperature, and winter season minimum temperature are shown in Figs. 1-3, respectively as an example. It can be seen from Figs. 1(a-c) that summer rainfall has an increasing linear trend whereas rainy season rainfall depicts a slight decreasing trend. However, winter rainfall does not have any trend. Maximum temperatures in summer and winter (Figs. 2(a,c)) show an increasing trend while no trend is seen in rainy season (Fig. 2(b)). Minimum temperatures are slightly increasing in summer and decreasing in winter (Figs. 3(a,c)). In rainy season, minimum temperature time series has no trend (Fig. 3(b)).

Box plots for rainfall and temperature time series of summer, rainy, and winter seasons are plotted in Figs. 5-7. Box plots (Fig. 4) indicate that range of rainfall occurrence is large during rainy season (i.e., about 10 to 1010 mm) compared to other two seasons (i.e., summer and winter). It is seen from rainy season box plot that median is at the centre of the box plot, which indicates similarity in the rainy season rainfall. Even upper and lower whiskers are also of similar length in case of rainy season rainfall. The above features confirm normality of the annual rainy season rainfall of Udaipur (with a value of about 500 mm annual rainy season rainfall), which is free from outliers. Rainfall box plots of summer and winter seasons are of no importance because of very low rainfall.

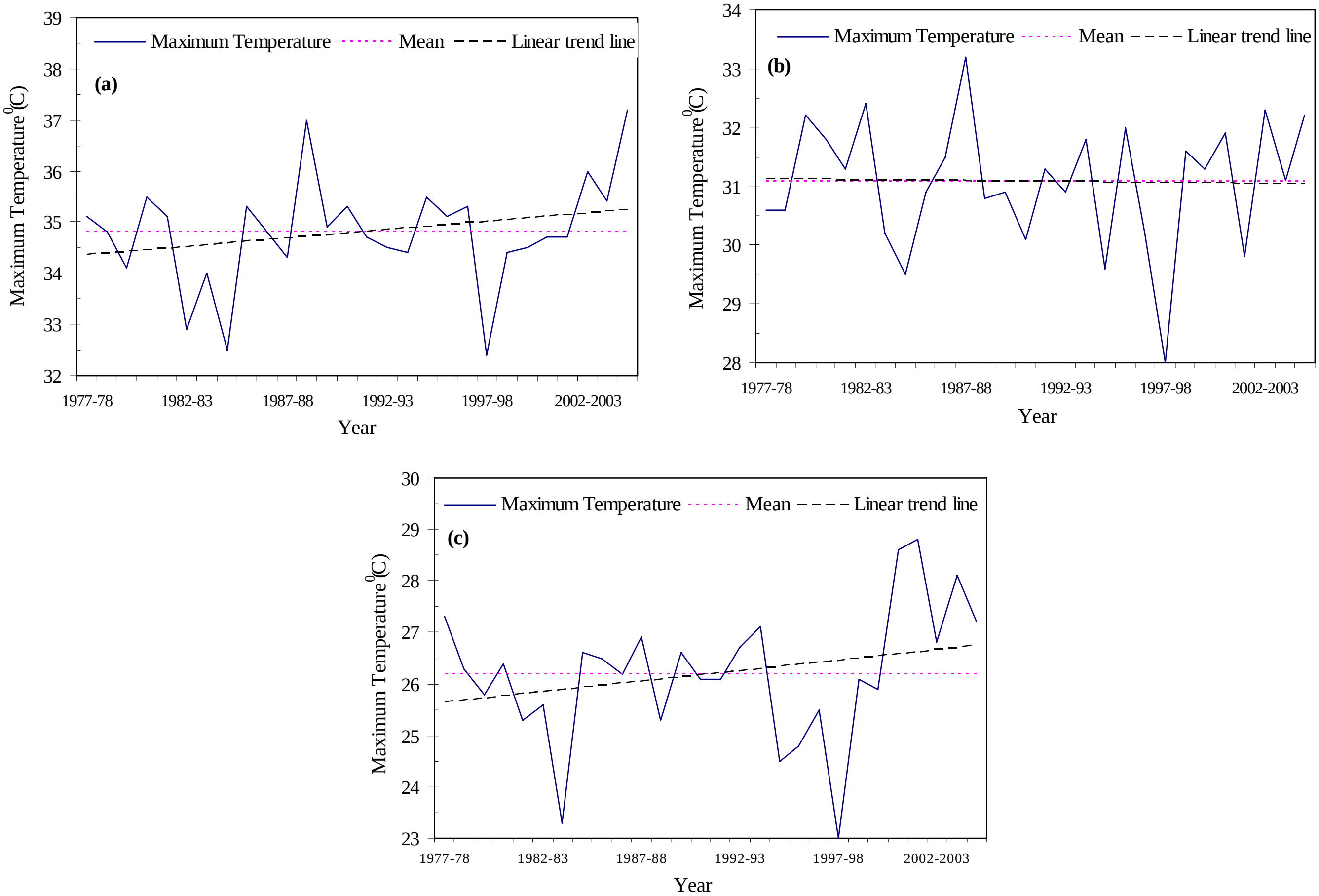

Fig. 2: Time patterns of maximum temperature in (a) summer, (b) rainy, and (c) winter seasons

Variations or range of maximum temperature (Fig. 5) is large for the rainy and winter seasons as compared to summer season. However, outliers are less in rainy season (only one) and winter season (two outliers) as compared to summer season (five outliers). Maximum temperature ranges in between 34 and 36^{0}C. It is indicated that maximum temperatures during rainy season may follow normal distribution.

Minimum temperatures during rainy season (Fig. 6) are higher than that in summer season. Minimum temperature time series of rainy season is found more uniform and normal than summer and winter season time series based on number of outliers present in the time series.

Testing homogeneity

Homogeneity of all the nine seasonal time series considered in this study was tested by using three tests; results are presented in Table 1.

Results of von-Neumann, Cumulative Deviations, and Bayesian tests applied to rainfall time series (Table 1) show that none of the seasonal rainfall time series is perfectly homogenous based on von-Neumann test (value of test-statistic, N is not very much closer to 2). However, value of N is more close to 2 in case of rainy and summer season rainfall series as compared to winter rainfall and these both series can be considered more close to homogenous than winter rainfall. Values of calculated test-statistics, 'Q' and 'R' of Cumulative Deviations test (Table 1) are less than their critical values (i.e. 1.284 and 1.598, respectively) for all seasonal rainfall time series of Udaipur, which indicates presence of homogeneity. Similarly, the values of calculated test-statistics, 'A' and 'U' of Bayesian test are also found less than their critical values (i.e. 2.48

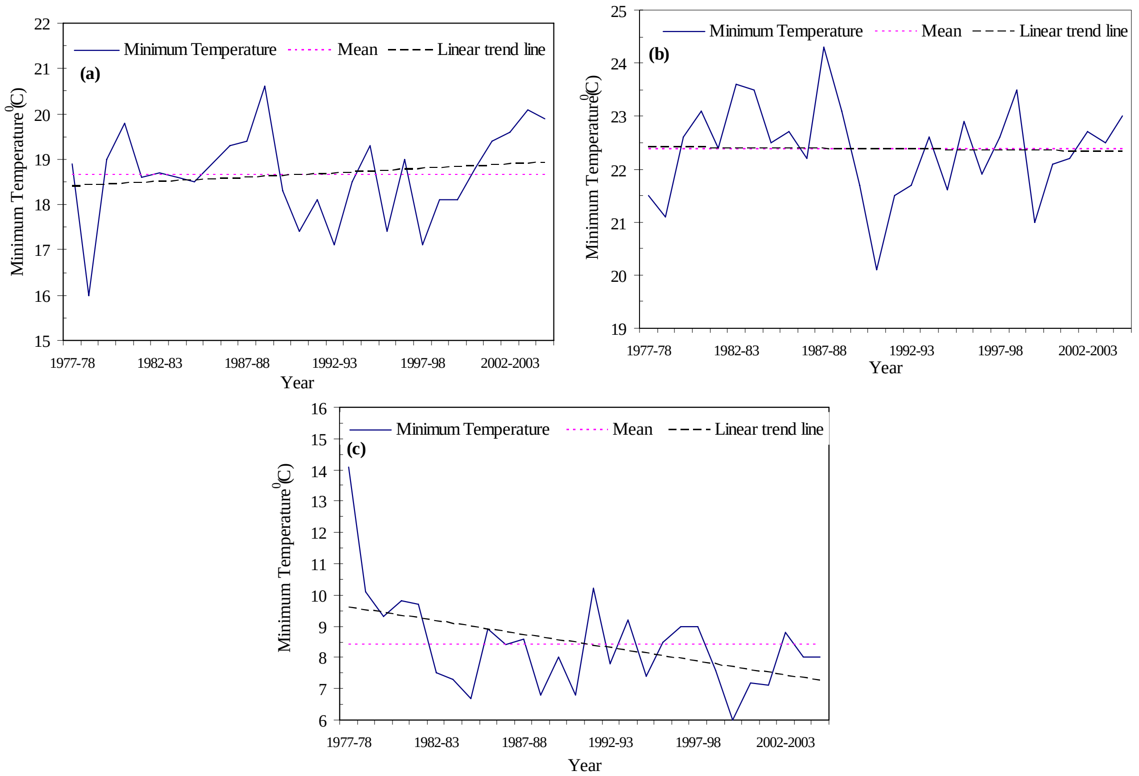

Fig. 3: Time patterns of minimum temperature in (a) summer, (b) rainy, and (c) winter seasons

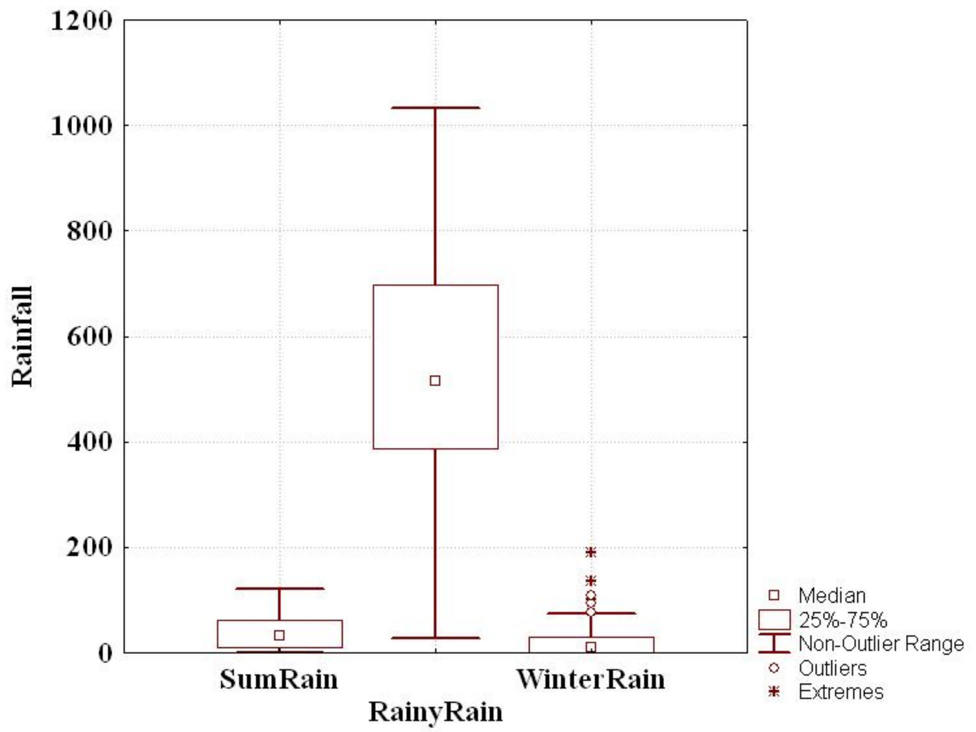

Fig. 4: Box plots for seasonal rainfall time series

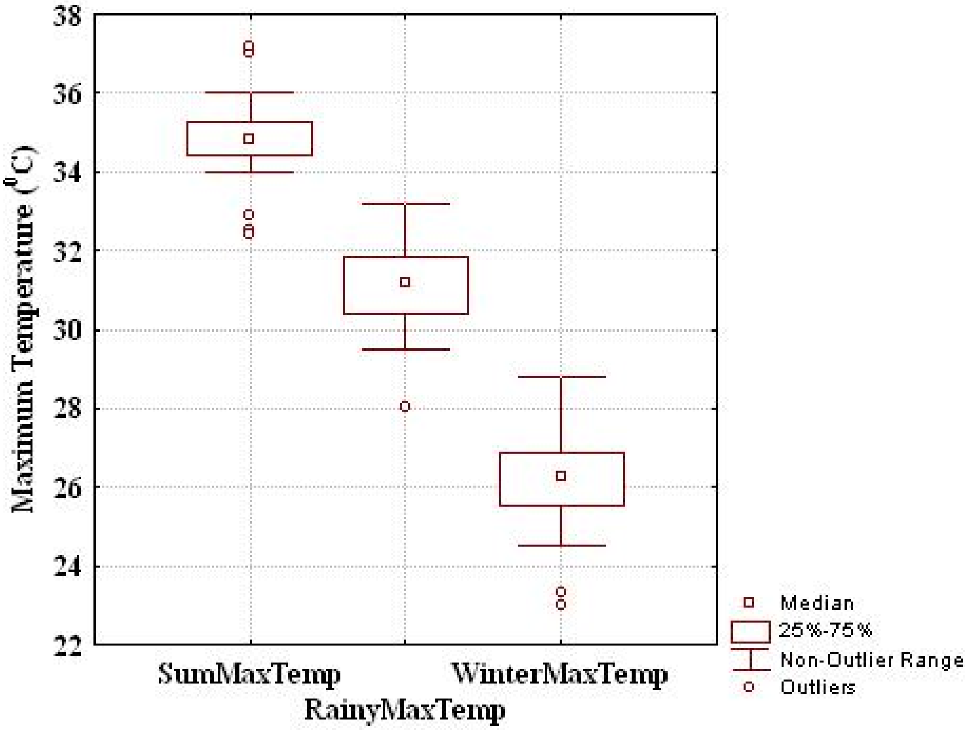

Fig. 5: Box plots for seasonal maximum temperature time series

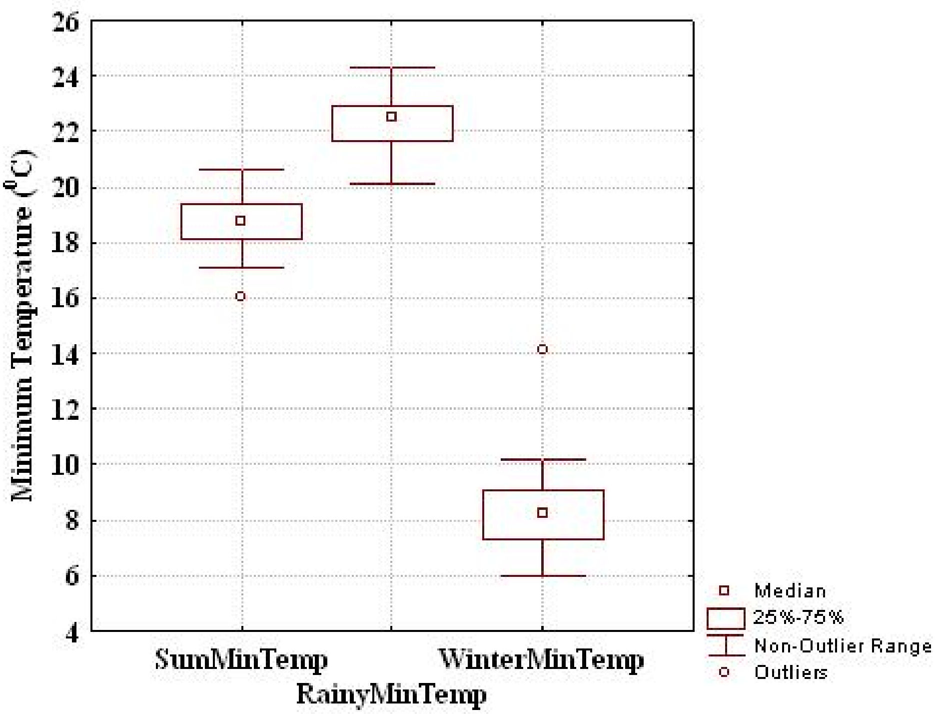

Fig. 6: Box plots for seasonal minimum temperature time series

Table 1: Results of three homogeneity test for three seasonal climatic time series

Season	Calculated test-statistics		
	Rainfall	Maximum Temperature	Minimum Temperature
von-Neumann test			
Summer	1.75	1.64	1.55
Rainy	1.73	2.13	1.46
Winter	1.64	1.31	1.009
Cumulative Deviation test (Q-statistic)			
Summer	1.189	0.7755	0.8331
Rainy	1.1945	0.4654	0.8864
Winter	0.8285	1.222	1.3138
Cumulative Deviation test (R-statistic)			
Summer	1.1551	0.7216	0.7981
Rainy	1.1943	0.4576	0.8849
Winter	0.8156	1.2030	1.2630
Bayesian test (A-statistic)			
Summer	1.57	1.50	1.24
Rainy	1.32	0.43	0.67
Winter	0.85	1.92	3.99
Bayesian test (U-statistic)			
Summer	0.297	0.210	0.151
Rainy	0.247	0.062	0.107
Winter	0.138	0.278	0.562

Note: value in the boldface indicates absence of homogeneity

Neumann test (Table 1). Calculated test-statistic values of Cumulative Deviations ('Q' and 'R') and Bayesian tests (A' and 'U') for summer and rainy seasons minimum temperature time series are found less than their critical values (1.236, 1.486, 2.42, and 0.445, respectively). Hence, rainy and summer season minimum temperature time series of Udaipur are considered homogenous. Test-statistic, 'R' of Cumulative Deviations test indicated homogeneity of the winter minimum temperature time series.

Testing stationarity

Stationarity of the climatological time series as tested by Mann-Whitney test (Table 2) shows that calculated test-statistics (in case of maximum and minimum temperature time series) are always less than their critical values (i.e. ±1.96) at 5% level of significance. Thus, it is considered that all seasonal temperature time series of Udaipur are stationary in nature. Similar results can be observed in case of winter season rainfall time series. However, summer and rainy seasons' rainfall time series of Udaipur are found non-stationary as values of calculated test-statistics of Mann-Whitney test are more than their critical values (i.e. ±1.96).

Trend detection

Results on trend component in the climatological time series as tested by Mann-Kendall test are presented in Table 2. Values of calculated test-statistics in case

Table 2: Results of Mann-Whitney test and Mann-Kendall test for rainfall and temperature time series

Season	Calculated test-statistics		
	Rainfall	Maximum temperature	Minimum temperature
(a) Mann Whitney test			
Summer	-2.03	-0.51	0.28
Rainy	2.59	-0.32	0.55
Winter	0.67	-1.33	0.51
(b) Mann-Kendall test			
Summer	-1.43	-0.25	-0.15
Rainy	1.91	-0.08	-1.99
Winter	1.27	-0.27	-0.01

Note: value in the bold face indicates absence of stationarity or presence of trend.

of all seasonal rainfall, maximum temperature, and minimum temperature time series of Udaipur are less than their critical values (i.e. ±1.96) at 5% level of significance; calculated test-statistic value is slightly more than critical value for rainy season minimum temperature. Overall, these time series are considered free from any kind of trend.

CONCLUSIONS

Box plots indicate normality of the rainy season rainfall time series. Minimum temperature time series of rainy season is found more uniform and normal than summer and winter season time series. Maximum temperatures during rainy season may follow normal distribution. All three seasonal rainfall series are found homogenous based on Cumulative Deviations and Bayesian tests. Homogeneity is present in summer and rainy season minimum temperature time series. Stationarity is present in all seasonal maximum and minimum temperature time series of Udaipur. Winter season rainfall time series is found stationary in nature; however, stationarity is not present in summer and rainy season rainfall time series. Based on the results of Mann-Kendall tests, all three seasonal long-term time series of rainfall, minimum temperature, and maximum temperature of Udaipur are found to be free from any kind of trend.

REFERENCES

Adeloye, A.J. and Montaseri, M. (2002). Preliminary streamflow data analyses prior to water resources planning study. *Hydrol. Sci. J,* 47(5): 679-692.

Buishand, T.A. (1982). Some methods for testing the homogeneity of rainfall records. *J. Hydrol.*, 58: 11-27.

Chernoff, H. and Zacks, S. (1964). Estimating the current means of a normal distribution which is subjected to change in time. *Annals Math. Statistics,* 35: 999-1018.

Doob, J.L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Annals Math. Statistics,* 20:393-403.

Fernando, D.A.K. and Jayawardena, A.W. (1994). Generation and forecasting of monsoon rainfall data. 20th WEDC Conference on Affordable Water Supply and Sanitation, Colombo, Sri Lanka, pp. 310-313.

Gan, T.Y. (1992). Finding trends in air temperature and precipitation for Canada and North-eastern United States. In: Kite, G.W. and Harvey, K.D. (editors), Using Hydrometric Data to Detect and Monitor Climatic Change. Proceedings of NHRI Workshop No. 8, National Hydrology Research Institute, Saskatoon, SK: 57-78.

Gardner Jr., L.A. (1969). On detecting changes in the mean of normal variates. *Annals Math. Statistics*, 40: 116-126.

Hirsch, R.M., Slack, J.R. and Smith, R.A. (1982). Techniques of trend analysis for monthly water quality data. *Water Resources Res.*, 18(1): 107–121.

Jayawardena, A. W. and Lai F. (1989). Time series analysis of water quality data in Pearl River, China. *J. Environ. Eng*., ASCE, 115(3): 590-607.

Jayawardena, A.W. and Lau, W.H. (1990). Homogeneity tests for rainfall data. Hong Kong Engineer, pp. 22-25.

Kanji, G. K. (2001). 100 Statistical Tests. Sage Publication, New Delhi, 111 pp.

Kendall, M.G. (1975). Rank Correlation Methods. Charles Griffin, London, U.K.

Khan, A. R. (2001). Analysis of hydro-meteorological time series in the upper Indus basin: Searching evidence for climatic change. International Water Management Institute, Working Paper 23, Pakistan Country Series Number 7.

Machiwal, D. and Jha, M.K. (2006). Time series analysis for water resources planning and management: a review. *J. Hydrol. Hydromechanics*, 54(3): 237-257.

Mann, H.B. (1945). Non-parametric tests against trend. *Econometrica*, 13: 245-259.

Owen, D.B. (1962). Handbook of Statistical Tables. Addison-Wesley, Reading, Mass.

Salas, J.D. (1993). Analysis and Modeling of Hydrologic Time Series. In: D. R. Maidment (editor-in-chief), Handbook of Hydrology, McGraw-Hill, Inc., USA, pp. 19.1-19.72.

Shahin, M., Van Oorschot, H.J.L., and De Lange, S.J. (1993). Statistical Analysis in Water Resources Engineering. A. A. Balkema, Rotterdam, the Netherlands, 394 pp.

Snedecor, G.W. and Cochran, W.G. (1980). Statistical Methods. The Iowa State University Press, Ames, Iowa.

StatSoft, Inc. (2004). STATISTICA (data analysis software system), version 6. www.statsoft.com.

USEPA (1998) Guidance for data quality assessment: Practical methods for data analysis. United States Environmental Protection Agency (USEPA), Quality Assurance Division, EPA QA/G-9, QA97 Version, pp. 2.3-3 to 2.3-5.

Wallis, J. R. and O'Connell, P. E. (1973). Firm reservoir yield —How reliable are historic hydrological records? *Hydrol. Sci. Bull.*, 18: 347-365.

WMO (1966). Climatic change. Technical Note 79, the World Meteorological Organization, Geneva.

Received: June 2007; Accepted: March 2008