Modelling the Monthly and Annual Temperature Series of Quetta, Pakistan

Farhat Iqbal
Department of Statistics
University of Balochistan, Quetta
farhatiqb@gmail.com

Sohail Chand
College of Statistical and Actuarial Sciences
University of the Punjab, Lahore
sohail.stat@gmail.com

Abstract

The monthly average temperature series of Quetta – Pakistan from 1950 – 2000 is examined. A straight line is fitted to the data and seasonal variation and trend in temperature for each month of the year were obtained. An overall model is constructed as large variations in the monthly slopes were observed. In order to describe the seasonal pattern and trend in temperature, corresponding to the different months, both sine/cosine waves and sine/cosine waves multiplied by the time were included in the model as independent variables. The lag-1 autocorrelation was found in the residual of the model and hence another model was fitted to the pre-whiten series that shows a good fit ($R^2 = 0.95$) and is free from correlated residuals. Both parametric and non-parametric tests applied to each month temperature show significant trend in all months except February and March.

Keywords: Temporal and seasonal variations, Pre-whiten series, Autocorrelation, Man-Kendall test.

1. Introduction

In the last century, a significant change in the global climate has been observed. During the last hundred years, the global average surface air temperature has increased by 0.74 ^{O}C (IPCC, 2007). In climatology, detecting trend in temperature has become one of the exciting research area. It is observed that temperature changes are not uniform globally. Yue and Hashino (2003) pointed that regional variations can be much larger, and there exist substantial spatial and temporal variations between climatically different regions. Many researchers throughout the world investigated trend detection in temperature and precipitation series (see Serra et al., 2001; Turkes and Sumer, 2004; ZerLin et al., 2005; Partal and Kahya 2006; Karpouzos et al., 2010; Croitoru et al., 2012, among others).

The non-parametric Man-Kendall (MK) tests (Mann, 1945; Kendall, 1975) is mostly used to identify hydroclimatologic trend and possible climate variations. Lettenmaier et al. (1994) adopted the Mann–Kendall test and investigated evidences of long-term trends in mean temperature, among other variables, over the USA. Long term trends in Japanese annual and monthly precipitation was studied by Yue and Hashino (2003) and significant negative trend was found. The MK test suffers in the presence of serial correlation in data. The presence of positive serial correlation in the series tends to increase the probability of type-I error (von Storch, 1995). Therefore, it is important to ascertain the presence of serial correlation, if any, before applying the MK test. In order to remove the effect of serial correlation on the MK test, Kulkarni and von Storch (1995) and von Storch (1995) proposed to "pre-whiten" the time series. This method id generally used in practice before applying the MK test for trend detection (Douglas et al. (2000); Zhang et al. (2000, 2001); Burn and Elnur (2001)).

In the present study, the monthly average temperature of Quetta – Pakistan is examined. Quetta is the provincial capital of Balochistan, the largest province of Pakistan in terms of area. Quetta is located in the West of Pakistan at 30.18^o N, 67.00^o E. The data were collected from January 1950 to December 2000. The average over each months of the daily average temperature, where daily average temperatures are the mean of maximum and minimum daily temperatures, are taken as the monthly mean temperature. The statistical analysis of the data starts by fitting a simple linear regression line to the average monthly temperature. Positive linear trends were observed for all months but different slopes were observed from month to month. Based on these findings, an overall model that had sine/cosine curves for both seasonal variations and trend in the temperature is constructed. The non-parametric MK test is also applied to detect the trend in average temperature of the station under study. The average annual temperature of the Quetta station is also analyzed.

This papers is organized as follows: Section 2 introduces the models for the modelling of monthly and annual temperature; Section 3 proposes the full model for the monthly temperature and finally Section 4 concludes the paper.

2. Modelling the Monthly and Annual Temperature

We start with analyzing the temperatures for each month separately by plotting the average temperature of Quetta against time. A positive trend in almost all months is evident in Figure 1 which corresponds to the global warming and urbanization. Based on this we consider the trends to be linear and fit a linear regression model

$$y_t^i = a^i + b^i t + e_t^i,$$

where the months *(t)* constitutes the independent variable and the average monthly temperature y_t^i $(i = 1, \dots, 12)$, the dependent variable. In the above model a^i is the regression constant (intercept) which represents the average temperature for month i in Quetta at the beginning of the series, b^i is the regression parameter (slope) which represents the rise of temperature for month *i*, e_t^i are the residuals for month *i* and *t* varies from 1 to *n*.

After fitting 12 simple linear regression equations to the monthly temperature series, we obtain the estimates of the intercepts and slopes for each month. Figure 2 below shows the plots of the estimates of intercepts, slopes and standard deviations for each month. Interestingly, we can see that the intercepts describe the large variation in the temperature of Quetta during the year and follow the temperature seasonal curve with the minimum value of 3.31^0C in January to the maximum of 26.96^0C in July. The slopes which measure the rise in the temperature in each month vary from 0.02^0C in March to 0.07^0C in November. The standard deviations of errors for few months of the year are found large.

Next, through statistical testing, we determine if the values of the slopes increase (or decrease) over some period of time. In this paper both the parametric and non-parametric statistical tests are used for detecting a significant trend. For this, the null hypothesis for no trend is tested against the alternative hypothesis of a trend in time series.

The parametric test is based on the assumption that the ratio of the slope and its standard errors is distributed as Student's t with $(n - 2)$ degrees of freedom. One of the assumption of the parametric methods are that the data need to be independent and normally distributed. A non-parametric methods are sometime better than the parametric methods (Hirsch et al., 1991) as no such assumptions about data are required and the power is almost same as the parametric tests. The MK test is the rank-based non-parametric test widely used in climatology for assessing the significance of the trend (see Onate and Pou, 1996 and Domonkos et al., 2003). The null hypothesis of the MK test is that a sample of data $\{Y_t : t = 1, \dots, n\}$ is independently and identically distributed against the alternative hypothesis that a monotonic trend exists in $\{Y_t\}$. The MK test is based on the statistic S. The standard normal statistic Z is estimated for the cases that $n > 10$ as:

$$Z = \begin{cases} (S-1)/\sqrt{Var(S)} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ (S+1)/\sqrt{Var(S)} & \text{if } S < 0 \end{cases}$$

where

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}(Y_j - Y_i),$$

$$\operatorname{sgn}(\theta) = \begin{cases} 1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases},$$

and

$$\operatorname{Var}(S) = \left[n(n-1)(2n+5) - \sum_k k(k-1)(2k+5) \right] / 18$$

In the $\operatorname{Var}(S)$, k is the extent of any given tie and $\sum_k$ denotes the summation over all ties. The null hypothesis of no trend is rejected if $|Z| > Z_{\alpha/2}$, (two-sided test) where α is the significance level.

2.1 Trend Analysis

We apply both the parametric and non-parametric tests for trend on monthly and annually average temperatures and the results are presented in Table I. It can be seen from the results in Table I that increasing trends are found in each month of the year except February and March. The null hypothesis of no trend is also rejected for mean annual temperature of Quetta at 5% level. Both parametric and non-parametric tests confirm the existence of trend in almost all months of the year.

Table 1: Parametric and Non-Parametric Trend analysis of average monthly temperature of Quetta

	Type of trend	
	Parametric Test	Non − parametric Test
January	b^*	b
February	c	c
March	c	c
April	b	b^{**}
May	b	b
June	b	b
July	b^*	b
August	b	b
September	b	b
October	b	b
November	b	b
December	b	b
Annual	b^*	b^*

b, b^*, b^{**} mean statistically significant increasing trend at 1, 5 and 10% levels of significance respectively and c means no trend.

3. A proposed Model

Instead of using different model for each month we construct a full model for the series of the monthly temperature and then test that the slopes are significantly different. The full model constructed using 12 different intercepts and 12 different slopes is

$$y_t = \sum_{i=1}^{12} a^i x_{t,i} + \sum_{i=1}^{12} b^i \, t x_{t,i} + e_t, \qquad (1)$$

where $x_{t,i}$ serve as dummy variables indicating if the observation y_t corresponds to the month i and $t = 1, \ldots, n$. We can also wirte the above model with the help of sine/cosine waves of period 12/*i*, *i*=1,…,6 as

$$y_t = \sum_{i=1}^{6} \left[\alpha_i \cos\left(\frac{2\pi it}{12}\right) + \beta_i \sin\left(\frac{2\pi it}{12}\right)\right] + \sum_{i=1}^{6} \left[\gamma_i t' \cos\left(\frac{2\pi it}{12}\right) + \delta_i t' \sin\left(\frac{2\pi it}{12}\right)\right] + e_t, \qquad (2)$$

where standardized time $t' = t/n$ is used.

The seasonal variation of temperature and variation of slopes with each month of year are described by the first and the second summation, respectively. The results of fitting the above model to the average monthly temperature of Quetta is presented in Table II. The values of 24 coefficients of the full model along with their standar errors and *p*-values are shown in the table. The *p*-values help us to identify the non-significant regression coefficients and allow us to fit a more parsimonious model to the data. The R^2 value of 0.97 indicates a very good fit for the model. The estimated standard deviation of residuals was 1.51. The Durbin-Watson test (Durbin and Watson, 1971) is also applied to check any dependence in the residuals of the model. The lag-1 autocorrelation is confirmed by the low *p*-value of the DW statistic.

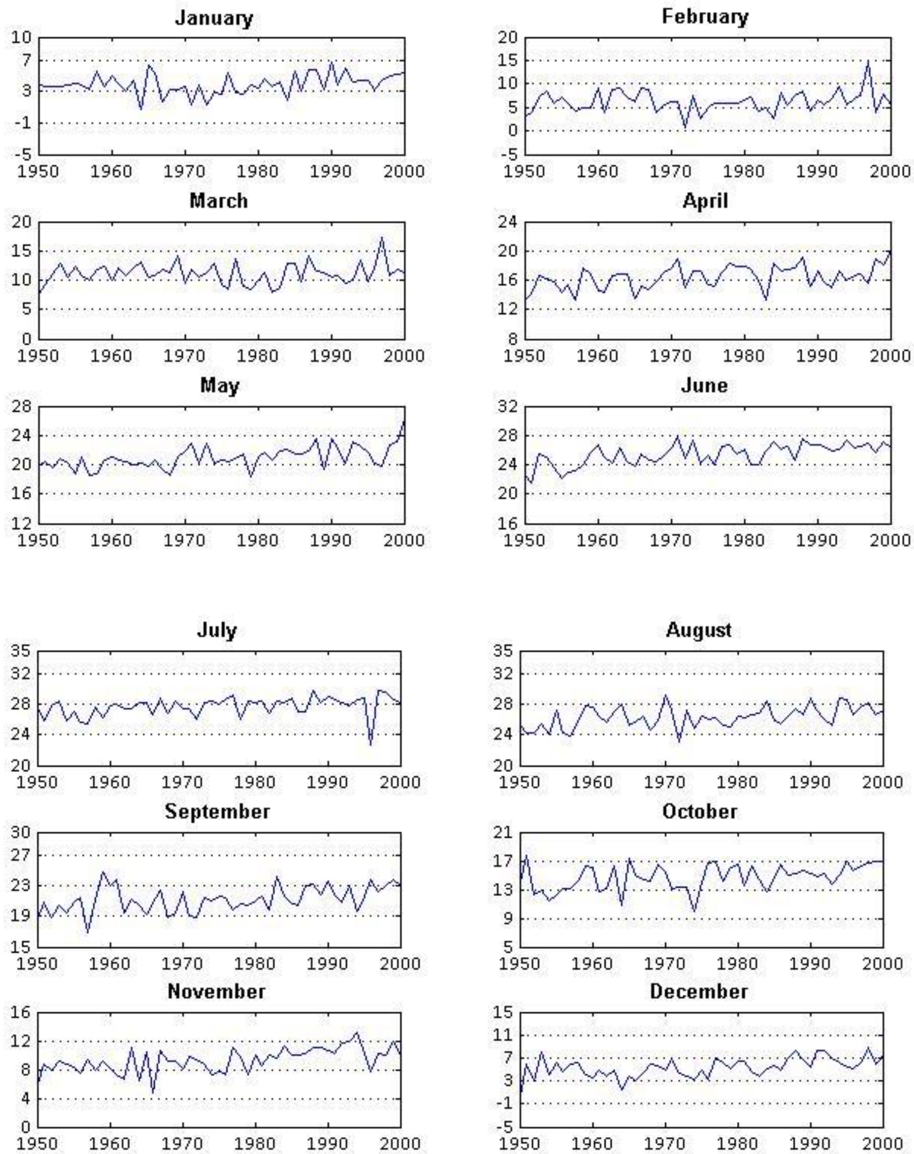

Figure 1: Average temperature of Quetta for each month from 1950 – 2000.

Based on the results of the full model [Eq. 2], we pre-whiten the series using $Y_t^* = Y_t - \hat{\rho}\, Y_{t-1}$, where $\hat{\rho} = 0.2649$ is the lag-1 autocorrelation coefficient of residuals from the full model. We fit a reduced model with only significant regression coefficients on the pre-whiten series. The significant variables found are the intercept, the sine wave (periods 12, 6 and 4), the cosine wave (periods of 12 and 6), the time (t') and the time multiplied by the sine wave of period 6. The result of fitting this reduced model is summarized in Table III. All the coefficients in the reduced model are found significant and the model has R^2 of 0.9531. The standard deviation of the reduced model is 1.46 and the lag-1 auto correlation coefficient was not found significant. Hence the residuals do not depend on their past values.

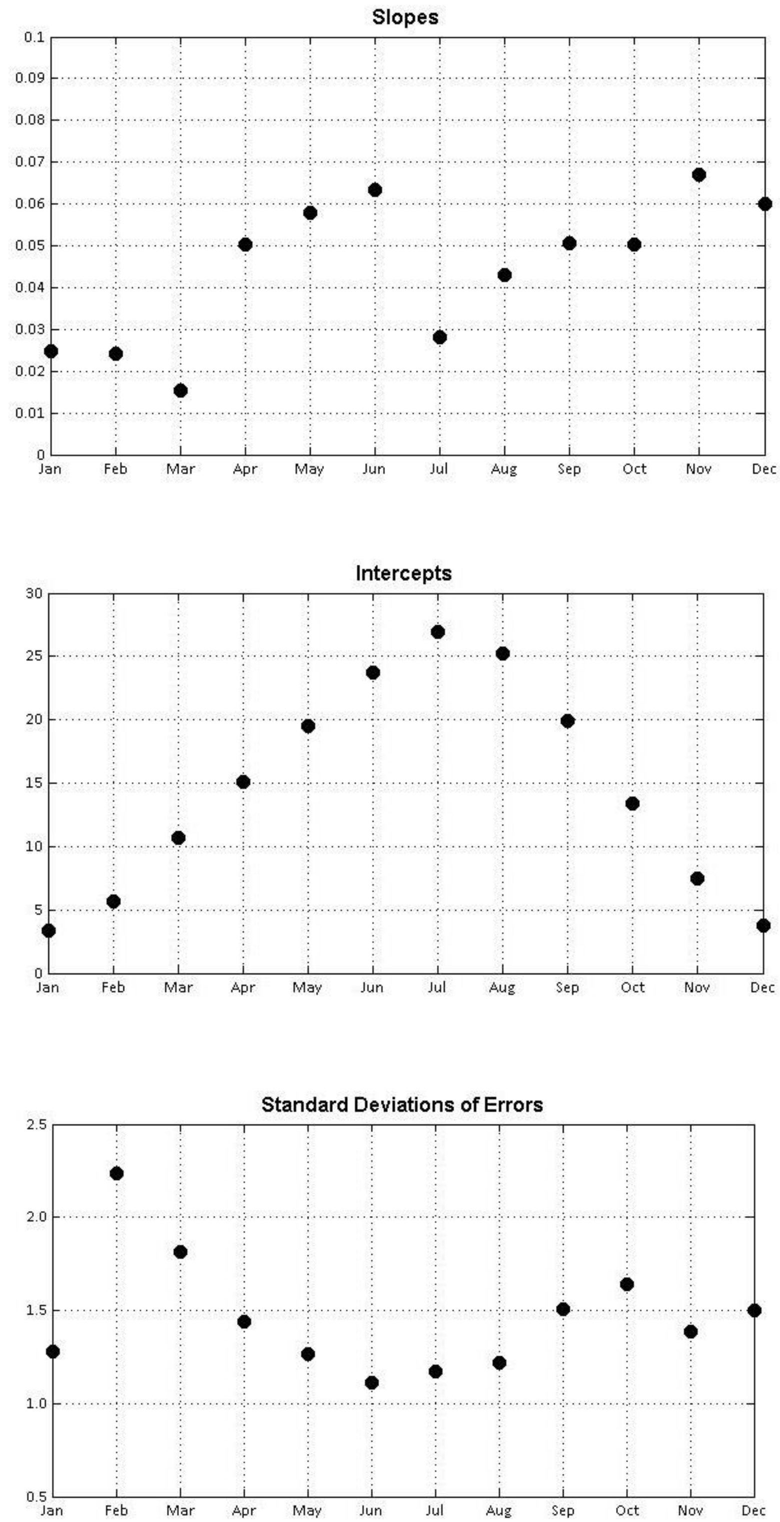

Figure 2: The slopes, intercepts and standard deviations of residuals plot for each month for the monthly temperature of Quetta from 1950 – 2000.

Table II: Result of the full model applied to average monthly temperature of Quetta

Variables	Coefficient	Standard Error	*p*-value
Intercept	14.5779	0.1223	0.0000
cos12	-10.2497	0.1731	0.0000
sin12	-5.0832	0.1727	0.0000
cos6	-0.6886	0.1731	0.0000
sin6	0.8152	0.1727	0.0000
cos4	0.2088	0.1731	0.2282
sin4	-0.4074	0.1727	0.0186
cos3	-0.0328	0.1731	0.8500
sin3	0.1285	0.1727	0.4573
cos2.4	0.0364	0.1731	0.8337
sin2.4	0.0821	0.1727	0.6345
cos2	-0.0854	0.1223	0.4852
time	2.2728	0.2115	0.0000
t' cos12	-0.0645	0.2991	0.8294
t' sin12	-0.4910	0.2991	0.1012
t' cos6	0.5315	0.2991	0.0761
t' sin6	-0.7753	0.2991	0.0097
t' cos4	0.1312	0.2991	0.6610
t' sin4	0.1923	0.2991	0.5204
t' cos3	0.1419	0.2991	0.6610
t' sin3	-0.2812	0.2991	0.3476
t' cos2.4	-0.1505	0.2991	0.6149
t' sin2.4	-0.2127	0.2991	0.4774
t' cos2	0.1988	0.2115	0.3474
R^2	0.9687	S	1.5100
$\hat{\rho} = 0.2649$	DW =1.4697	*p*-value	0.0000

Table III: Result of the reduced model applied to pre-whiten average monthly temperature of Quetta

Variables	Coefficient	Standard Error	*p*-value
Intercept	10.7096	0.1191	0.0000
cos12	-10.2811	0.1072	0.0000
sin12	-5.3282	0.1071	0.0000
cos6	-0.4231	0.0933	0.0000
sin6	0.7985	0.1877	0.0000
sin4	-0.3135	0.0811	0.0001
time	27.4780	3.3569	0.0000
t' sin6	-8.9213	3.8965	0.0224
R^2	0.9513	S	1.4650
$\hat{\rho} = 0.0056$	DW =1.9802	*p*-value	0.6760

We also analyze the average annual temperatures 1951 – 2000 (*n*=51). The following regression model is fitted to this data set:

$$y_t = a + b\,t + e_t,$$

where the (t) is the number for years and varies from 1 to n and y_t is the average annual temperature. As in the previous models, a is the intercept term and represents the average temperature of Quetta at the start of the series, b, the slope of the regression model represents the rise of temperature.

Figure 3 shows the time series plot of the average annual temperature in Celsius of Quetta from 1951 – 2000. The graph clearly shows an increasing trend in the annual temperature. The regression line is also plotted on the same graph. The estimated values of the intercept and slope are 14.56 and 0.05, respectively. Both these estimates are found significant (*p*-values < 0.001) and the value of $R^2 = 0.51$. The Durbin-Watson test was also applied on the residuals obtained from this regression and the lag-1 autocorrelation coefficient was not found significant.

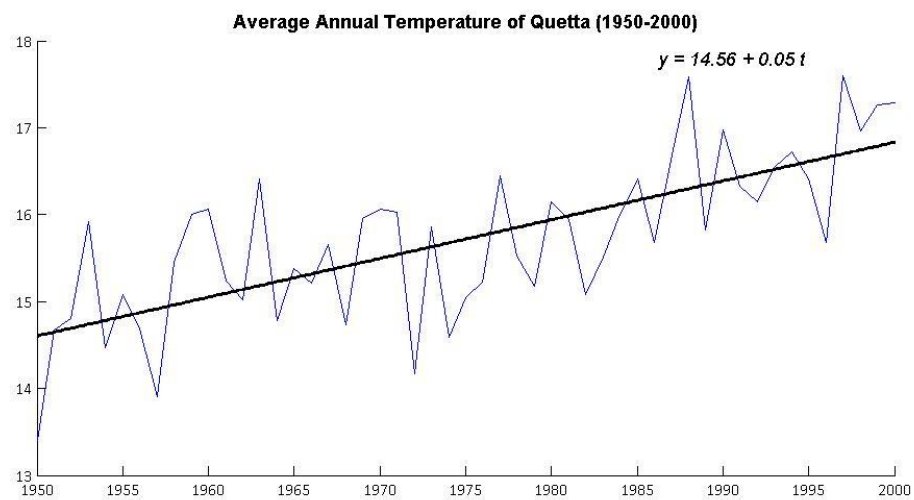

Figure 3: Average annual temperature along with least square fit line for Quetta from 1950 – 2000.

4. Conclusion

Annual and monthly average temperature of Quetta, Pakistan for the period of 1950 – 2000 were analyzed using both parametric and non-parametric statistical tests. A straight line to the monthly temperature data were fitted and an overall model is also constructed. The time series is pre-whitened as the lag-1 autocorrelation was found. Our analysis showed that the monthly temperature of Quetta has significant increased in all months except February and March.

Acknowledgments

We thank the Editor and the referees for their helpful comments and suggestions that greatly improved the previous version of this paper.

References

1. Burn, D.H., Hag, Elnur, M.A. 2001. Climate change impact using hydrologic variables, in CCAF Workshop Meeting.
2. Croitoru, A.E., HolobacaI, H., Catalin, L.C., Moldovan, F., and Imbroane, A. 2012. Air temperature trend and the impact on winter wheat phenology in Romania. *Climate Change.* 111(2): 393 – 41.
3. Douglas, E.M., Vogel, R.M., and Kroll, C.N. 2000. Trends in floods and low flows in the United States: impact of spatial correlation. *Journal of Hydrology*. 240: 90 – 105.
4. Domonkos, P., Kysel, J.Y., Piotrowicz, K., Petrovic P., and Likso, T. 2003. Variability of extreme temperature events in South-central Europe during the 20th century and its Relationship with the large scale circulation. *International Journal of Climatology* 23: 978 – 1010.
5. Durbin, J., and Watson, G.S. 1971. Testing for serial correlation in least squares regression. III. *Biometrika.* 58(1): 1 – 19.
6. Hirsch, R.M., Alexander, R.B., and Smith, R.A. 1991. Selection of methods for the detection and estimation of trends in water quality. *Water Resources Research* 27: 803 – 814.
7. IPCC. 2007. Summary for policymakers. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt K B, Tignor M and Miller HL (eds) Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change; Cambridge University Press.
8. Karpouzos, D.K., Kavalieratou, S., and Babajimopoulos, C. 2010. Trend analysis of precipitation data in Pieria Region (Greece). *European Water*. 30: 31 – 40.
9. Kendall, M.G., 1975. Rank Correlation Methods. Griffin, London.
10. Kulkarni, A., and Von Storch, H. 1995. Monte Carlo experiments on the effect of serial correlation on the Mann–Kendall test of trend. Meteorologische Zeitschrift. 4(2), 82 – 85. Germany.
11. Lettenmaier, D.P., Wood, E.F., and Wallis, J.R. 1994. Hydro-climatological trends in the continental United States, 1948–88. *Journal of Climate*. 7, 586 – 607.
12. Mann, H.B. 1945. Nonparametric tests against trend. *Econometrica*, 13: 245 – 259.
13. Onate, J.J., and Pou, A. 1996. Temperature variations in Spain since 1901: A preliminary analysis. *International Journal of Climatology* 16: 805 – 815.
14. Partal, T. and Kahya, E. 2006 Trend analysis in Turkish precipitation data *Hydrological Processes*. 20: 2011 – 2026.
15. Serra, C., Burgueno, A., and Lana, X. 2001. Analysis of maximum and minimum daily temperatures recorded at Fabra observatory (Barcelona, NE Spain) in the period 1917–1998. *International Journal of Climatology*. 21: 617 – 636.
16. Turkes, M., and Sumer, U.M. 2004. Spatial and temporal patterns of trends and variability in diurnal temperature ranges of Turkey. *Theoretical and Applied Climatology*. 77: 195 – 227.

17. von Storch, H. 1995. Misuses of statistical analysis in climate research. In Analysis of Climate Variability: Applications of Statistical Techniques, eds. by H.V. Storch and A. Navarra, Springer–Verlag Berlin, 11–26.

18. Yue, S., and Hashino, M. 2003. Long term trends of annual and monthly precipitation in Japan. *Journal of the American Water Resources.* 39(3): 587 – 596.

19. Zer Lin, W., Chung, T.H., Ho, W.C., and Hsien, T.W. 2005. Urbanization-induced regional climate change on the western plain of Taiwan for the period 1964–1999; Int. Conf. on Environment, Ecosystems and Development, Venice, Italy. November 2 – 4, 2005.

20. Zhang, X., Vincent, L.A., Hogg, W.D., and Niitsoo, A. 2000. Temperature and precipitation trends in Canada during the 20th century. *Atmosphere Ocean*. 38(3): 395 – 429.

21. Zhang, X., Harvey, K.D., Hogg, W.D., and Yuzyk, T.R. 2001. Trends in Canadian streamflow *Water Resources Research*. 37(4): 987 – 998.