West Indian Journal of Engineering Vol. 23, No. 2, (January 2001) Technical Paper (Stone) 9 - 16

Changing Seasonal Rainfall Patterns in Trinidad: Myth or Reality?

R.J. Stone*

Department of Food Production,
The University of the West Indies,
St. Augustine, Trinidad

Seasonal rainfall amounts at five stations in Trinidad were statistically analysed to detect the presence of deterministic components such as trends, jumps and cycles in the data series. The statistical procedures employed were the sample auto-correlation function, the runs test of randomness and the Wald-Wolfowitz test of independence. The results show that all the rainfall data series are both random and stationary at the 5% significance level indicating the absence of trends, jumps and cycles. Consequently, these results demonstrate that the recent claims of changing seasonal rainfall patterns in Trinidad due to global warming (drier dry seasons, wetter wet and transitional seasons) have no valid statistical basis, are inaccurate and therefore misleading. In addition, the study provides evidence to support the use of frequency analysis as the appropriate statistical method for predicting seasonal rainfall amounts in Trinidad.

1. Introduction

The issue of global warming due to increasing atmospheric concentration of greenhouse gases is now engaging the attention of engineers and scientists worldwide. It is of grave concern due to its anticipated adverse effects such as climate change and the concomitant sea level rise, which in turn, could have severe socio-economic and environmental consequences. Since it is not possible to carry out physical experiments to determine the effects of enhanced greenhouse gas-induced global warming at various geographical locations, climatologists have resorted to the use of General Circulation Models (GCM). GCMs are computer models based upon physical laws represented by mathematical equations that are solved using numerical methods as applied to a three-dimensional grid over the globe [1]. These models are used to predict the effects of a build-up of greenhouse gases on climate change.

Recently, Singh [2,3] utilised the Canadian Climate Centre's GCM to examine a climate change scenario for the Greater and Southern Caribbean which include the island of Trinidad. The Canadian Climate Centre's GCM climate change scenario projects more extreme rainfall conditions for the region, that is to say, drier dry seasons (January to April) and wetter wet seasons (May to August). Wetter conditions are also projected for what Singh [2] called the transitional season (September to December).

Singh [2,3] also smoothed seasonal rainfall data from several locations in Trinidad using 5-year and 10-year moving averages and concluded that the changing rainfall patterns detected are in agreement with the projections of the Canadian Climate Centre's GCM and demonstrate evidence of recent climate changes, which may be interpreted as early signals of global warming. Over two and a half decades ago, Gani [4] scanned a selection of the literature on climatic change and pointed out that several widely quoted claims of evidence for climate change are of doubtful statistical validity and raised the question of how well alleged climate changes are distinguished from random fluctuations. The author pointed out that there is very little statistical analysis in the climatologists' work, and some of what there is, is either superficial or wrong. The author further emphasised the need for more scientific rigour and advised that the problems tackled by climatologists called for the application of careful logic and the use of well-tested statistical methods.

Moreover, several other authors [5-8] have pointed out the inappropriateness of using moving-average procedures on random time series data. They caution that with these series, the application of moving-average procedures may introduce spurious correlation and cycles into the data that could lead to inaccurate results and misleading conclusions.

* Corresponding author: E-mail: rstone@fans.uwi.tt

Pertinent discussion will be published in July, 2002 West Indian Journal of Engineering if received by December, 2001.

In view of Gani's [4] comments and the inherent potential danger of using moving-average procedures, a careful examination of Singh's [2,3] work was undertaken to assess the validity of the findings. The examination revealed a lack of statistical rigour and a fundamental flaw in the approach, namely, the failure to examine the data sets for basic statistical properties such as randomness and stationarity, which is fundamental to the analysis of time series data [9].

The assessment of these two properties is essential for detecting the existence of trends, jumps, and cycles in the data and to inform the data analyst as to the appropriate candidate mathematical models that could be used to represent the data, for example, moving average, autoregressive, and autoregressive-moving average [10]. The practical engineering relevance of this assessment in hydrologic studies and water resources system design was highlighted by Matalas [11].

This study was therefore undertaken to assess the statistical properties of seasonal rainfall amounts in Trinidad with respect to randomness and stationarity in order to ascertain the presence or absence of trends, jumps, and cycles in the data series, and in so doing, to examine the validity of recent claims of changing seasonal rainfall patterns in Trinidad. In addition, the study seeks to identify an appropriate modelling approach to describe seasonal rainfall amounts for prediction purposes.

2. Statistical Properties

2.1 *Randomness*

A sequence of values collected over time on a particular variable is a time series. A time series can be composed of a quantity either observed at discrete times, averaged over a time interval, or recorded continuously with time [12]. Seasonal rainfall amounts therefore constitute a discrete time series.

Testing for randomness is a fundamental aspect of time series analysis. A random series is one in which the observations are serially uncorrelated. Randomness could also be defined in terms of the observations being independent. Independence is a stronger notion of randomness than serial uncorrelatedness. The essential difference is that it is sometimes possible to make non-trivial predictions (use current and past values to predict any feature of future values) from a series which has the uncorrelated property but impossible to do so in a series which possesses the property of independence. Independence means that no observation in the data series has any influence on any observations following it. Even if the events in a series are serially uncorrelated, they may not be independent [12].

2.3 *Stationarity*

The idea that past and future values of a time series will be similar statistically is an informal expression of what is called stationarity. A formal definition of a stationary time series is one in which its mean, variance and covariance do not vary with time. The observations therefore fluctuate around a constant level and there is no tendency for their spread to increase or decrease over time. Different time slices of a stationary time series can be regarded as having the same underlying mean, variance and covariance [13]. Types of non-stationarity include trends, jumps and cycles [14].

3. Statistical Tests

3.1 *Sample Autocorrelation Function*

The basic tool for the analysis of time series in the time domain is the sample autocorrelation function (ACF). The ACF is expressed as the variation of the autocorrelation coefficient as a function of lag. The kth-order autocorrelation coefficient, r_k, is given by [13]:

$$r_k = \frac{\sum_{i=1}^{N-k}[(x_i - \bar{x})(x_{i+k} - \bar{x})]}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \qquad (1)$$

where, N is the number of data points in the series, x_i the ith observation, $\bar{x}$ the mean of the N data points and k the lag.

An examination of the ACF may lead to the conclusion that the series is random (in the sense of uncorrelatedness), or that it exhibits a pattern of serial correlation that can perhaps be modelled by a particular stochastic process. An ACF that is close to zero for all nonzero lags suggests that an appropriate model for the data might be independent, and identically distributed (IID) noise. For an IID sequence, about 95% of the sample autocorrelations should fall between the bounds of $\pm 1.96 / N^{1/2}$ at the 5% significance level (95% confidence level). If we compute the sample autocorrelations up to lag 40 and find that more than two or three values fall outside the bounds, or that one value falls far outside the bounds, we reject the IID hypothesis [15]. If not, we accept the IID hypothesis at the 5% significance level.

The ACF can also be used to test for stationarity. The ACF of a stationary time series tends either to die down with increasing lag k or to cut off after a particular lag k = q. For the special case of q = 0, if 95% of the r_k values are within $\pm 1.96 / N^{1/2}$, then the times series is considered stationary at the 5% significance level. If the time series is non-stationary, then the ACF will neither cut off nor die down quickly, but rather will die down extremely slowly [16].

3.2 The Runs Test

The runs test is used to detect departures in randomness of a time series caused by trends or periodicities (cycles). Replacing each measurement in the order in which they are collected by a plus (+) symbol if it falls above the median, by a minus (-) symbol if it falls below the median, and omitting all measurements that are exactly equal to the median, we generate a sequence (or runs) of plus and minus symbols that are tested for randomness using the following procedure [17].

Let n_1, n_2 and L be the number of minus, plus and runs respectively. Then, for situations where both n_1 and n_2 are greater than 10, the sampling distribution of L approaches the normal distribution with mean, $\bar{L}$, and variance, Var(L), given by the following:

$$\bar{L} = \frac{2n_1 n_2}{n_1 + n_2} + 1 \tag{2}$$

$$Var(L) = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)} \tag{3}$$

The quantity $\iota = (L - \bar{L}) / (Var(L))^{1/2}$ follows a standardised normal distribution (mean 0 and variance 1) and can be used to test the hypothesis of randomness at the 5% significance level, by comparing $|\iota|$ with the critical value of standard normal variate, $z_{0.025} = 1.96$. If $|\iota| < 1.96$, we accept the null hypothesis of randomness at the 5% significance level. If not, we reject the null hypothesis of randomness at the 5% significance level.

3.3 The Wald-Wolfowitz Test

The Wald-Wolfowitz (W-W) test is used to check for independence in the data series [14]. For a sample of size $N(x_1, x_2, \ldots, x_N)$, the W-W test considers the statistic R such that:

$$R = \sum_{i=1}^{N-1} x_i x_{i+1} + x_1 x_N \tag{4}$$

In the case where the elements of the sample are independent, R follows a normal distribution with mean, $\bar{R}$, and variance, Var(R) given by:

$$\bar{R} = (s_1^2 - s_2) / (N - 1) \tag{5}$$

$$Var(R) = (s_2^2 - s_4) / (N - 1) - \bar{R}^2$$

$$+ (s_1^4 - 4s_1^2 s_2 + 4s_1 s_3 + s_2^2 - 2s_4)$$

$$/ (N - 1)(N - 2) \tag{6}$$

where, $s_k = N.m_k$ and m_k is the kth moment of the sample about the origin.

The quantity $r = (R - \bar{R}) / (Var(R))^{1/2}$ follows a standardised normal distribution (mean 0 and variance 1) and can be used to test the hypothesis of independence at the 5% significance level, by comparing $|r|$ with the critical value of standard normal variate, $z_{0.025} = 1.96$. If $|r| < 1.96$, we accept the null hypothesis of independence at the 5% significance level. If not, we reject the null hypothesis of independence at the 5% significance level.

4. Rainfall Data Series

The data series used in the analysis were obtained from the Piarco Meteorological Office and the Water Resources Agency. **Figure 1** shows the geographical locations of the five (5) stations which were chosen to be representative of Trinidad covering the northern, central, and southern regions of the island. Only stations with more than 50 years continuous monthly data were used in the analysis based on the recommendation of Box and Jenkins [10] for deriving a reliable ACF. The locations and the corresponding periods of the rainfall data used were: Piarco (Meteorological Office), 1946-1998; Port of Spain (Botanical Gardens), 1862-1996; Couva (Exchange), 1941-1998; Cedros (Perseverance Estate), 1925-1996; and Rio Claro (Poole Syndicate), 1932-1991. The dry, wet and transitional season groupings were identified following the procedure adopted by Singh [2]. The dry season rainfall amounts were obtained by summing up the monthly rainfall for January, February, March and April, the wet season rainfall amounts by summing up the monthly rainfall for May, June, July and August, and the transitional season rainfall amounts by

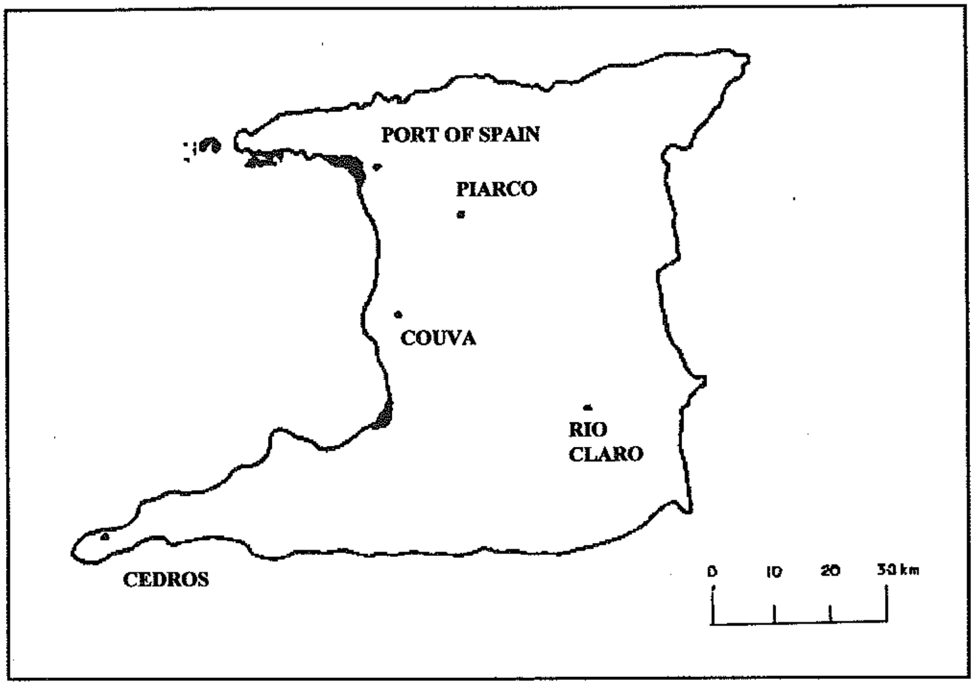

FIGURE 1: *Geographical Location of the Five (5) Stations*

summing up the monthly rainfall for September to December. The number of data points for each season at Piarco, Port of Spain, Couva, Cedros and Rio Claro was 53, 135, 58, 72 and 60 respectively.

5. Results

Table 1 presents a summary of the seasonal and annual rainfall characteristics at the five (5) stations. A salient feature of the rainfall data is their significant spatial and temporal variability. For example, the mean annual rainfall at Rio Claro is approximately 1.85 times the mean annual rainfall recorded at Cedros. Mean rainfall amounts in the wet seasons are greater than those of the transitional seasons, but generally, by only moderate amounts. However, both wet and transitional season mean rainfall amounts are significantly greater than their dry season counterparts.

Another interesting feature of the data is the inter-seasonal variability. The dry season rainfall exhibited the highest variability with the coefficient of variation (CV) varying from 45.5% to 64.2% while the annual rainfall was generally least variable with the CV varying from 14.8% to 23.9%. Wet and transitional season rainfall amounts showed similar variabilities which were generally slightly higher than the variability of annual rainfall amounts, ranging from 19.0% to 29.0%.

The critical test limits $\pm 1.96 / N^{1/2}$ of the ACFs at the 5% significance level for Piarco, Port of Spain, Couva, Cedros and Rio Claro are ±0.27, ± 0.17, ± 0.26, ± 0.23, and ±0.25 respectively. Examination of the ACFs provided the following results: for Piarco, only one point, $r_5 = 0.33$ for the transitional season, strayed out of the 0.27 bounds; for Port of Spain, only one point, $r_6 = 0.21$ for the dry season, strayed out of the 0.17 bounds while two points, $r_3 = 0.21$ and $r_4 = 0.21$, strayed out of the 0.17 bounds for the wet season; for Couva, only one point, $r_2 = 0.34$ for the wet season, and one point for the transitional season, $r_9 = 0.31$, strayed out of the 0.26 bounds; for Cedros, none of the points strayed outside the 0.23 bounds; and for Rio Claro none of the points strayed outside the 0.25 bounds. Since 95% or more points were within the critical test limits for each data series, the conditions of both randomness and stationarity are therefore satisfied for all seasons at all stations at the 5% significance level.

Table 2 shows the results of the runs test of randomness for the three (3) seasons at the five (5) locations. The values of the test statistic, $|t|$, varies from 0 to 1.64. All the values are therefore less than 1.96 and the hypothesis of randomness cannot be rejected at the 5% significance level, and therefore confirms the absence of trends or cycles in the data series.

Table 3 presents the results of the W-W test of Independence for the three (3) seasons at the five (5) locations. The values of the test statistic, $|r|$ varies from

TABLE 1: *Seasonal and Annual Rainfall Characteristics at the Five Stations*

Station	Mean (mm)	Median (mm)	Standard Deviation (mm)	Coeff. of Variation (%)	Minimum (mm)	Maximum (mm)
PIARCO						
Dry	198.9	176.3	99.1	49.8	47.3	573.9
Wet	899.8	921.2	175.1	19.0	569.6	1423.2
Transitional	740.8	715.6	176.5	23.8	430.1	1148.9
Annual	1839.5	1816.9	271.9	14.8	1265.3	2603.2
PORT OF SPAIN						
Dry	192.6	169.3	111.7	58.0	1.5	611.7
Wet	709.6	697.0	197.1	27.8	249.2	1335.2
Transitional	664.3	656.6	139.2	21.0	263.8	1122.4
Annual	1566.5	1560.0	313.2	20.0	861.5	2412.9
COUVA						
Dry	174.5	151.3	101.5	58.2	36.5	621.7
Wet	751.7	730.3	169.2	22.5	408.0	1214.2
Transitional	609.5	595.3	134.7	22.1	341.1	909.5
Annual	1535.7	1509.8	229.7	15.0	1086.2	2043.4
CEDROS						
Dry	288.5	239.5	185.3	64.2	35.3	1301.1
Wet	622.8	607.2	142.7	22.9	380.9	1241.8
Transitional	540.0	517.9	156.6	29.0	259.6	1014.5
Annual	1451.3	1412.8	347.1	23.9	1002.8	3498.2
RIO CLARO						
Dry	434.9	389.3	197.9	45.5	123.9	1129.7
Wet	1137.1	1117.3	248.5	21.9	560.0	1770.4
Transitional	1108.8	1064.3	287.6	25.9	678.2	1938.3
Annual	2680.8	2684.9	480.1	17.9	1687.6	3922.7

TABLE 2: *Test Statistics of the Runs Test of Randomness*

Station	Dry Season	Wet Season	Transitional Season
Piarco	0.42	0.69	0.97
Port of Spain	0.43	0.60	1.64
Couva	0.26	0.26	0.00
Cedros	0.71	0.71	0.24
Rio Claro	0.52	0.52	1.30

TABLE 3: *Test Statistics of the Wald-Wolfowitz Test of of Independence*

Station	Dry Season	Wet Season	Transitional Season
Piarco	1.47	1.08	1.42
Port of Spain	0.51	1.74	0.44
Couva	0.63	1.24	0.21
Cedros	0.77	0.47	1.29
Rio Claro	0.48	1.72	1.43

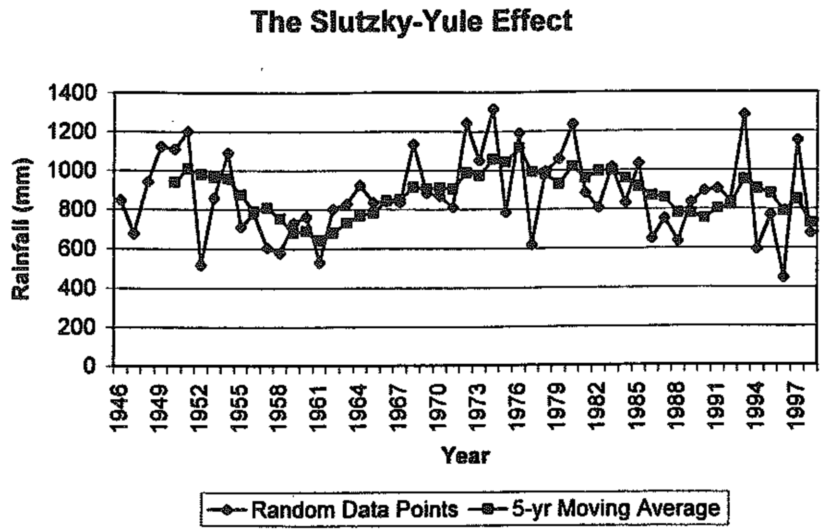

FIGURE 2: *Graphical Demonstration of the Slutzky-Yule Effect*

0.21 to 1.72. All the values are therefore less than 1.96 and the hypothesis of independence cannot be rejected at the 5% significance level and again indicates the absence of trend and supports the stronger notion of randomness.

6. Discussion

From the preceding results, it is clear that all the data series exhibit the statistical properties of randomness and stationarity and are therefore free from deterministic components of trends, jumps, or cycles. These results directly contradict the previously mentioned findings of Singh [2,3].

The explanation for the contradiction lies in the approach adopted, namely, the use of moving averages. The moving average analysis demonstrates the well-known statistical phenomenon referred to as the Slutzky effect [6] or the Slutzky-Yule effect [5,8] which was discovered independently by Slutzky, a Russian mathematician and Yule, an American mathematician, more than seven (7) decades ago in 1927 [6].

The phenomenon occurs when a random series is subjected to a moving average. In this situation, the moving average produces a series smoother than the original series but which has a quite high positive lag 1 autocorrelation and may present the appearance of systematic oscillations. Characteristically, the moving average series exhibits varying intervals between peaks and troughs, so that instead of a strict "period", there is a distribution of intervals. Likewise, the amplitudes of the movements show substantial variations from one peak to the next [5].

The moving average procedure therefore generates oscillations even though none exist in the original data. At the time of its discovery, this caused shock and horror to users of moving averages as they realised that they might have been inventing rather than discovering hidden patterns in their data [5,6].

An example of the Slutzky-Yule effect is demonstrated in **Figure 2**. Here a normally distributed random data series was generated using the popular Microsoft Excel spreadsheet. The statistical properties used (mean and standard deviation) for the data series are those of the wet season rainfall at Piarco (See **Table 1**). A five-year moving average of the random data points, also using the Microsoft Excel spreadsheet, clearly demonstrates the phenomenon described above. An ACF analysis was also carried out on the moving-average data to confirm quantitatively its non-random behaviour. The lag 1, lag 2, lag 3 and lag 4 autocorrelation coefficients were 0.84, 0.75, 0.55 and 0.42 respectively, which demonstrates the significant autocorrelation present in the moving-average data.

The oscillatory behaviour described and demonstrated above is precisely what is observed in the figures of moving averages of seasonal rainfall amounts presented by Singh [2]. However, the moving-average results were not viewed as oscillations but were interpreted as increasing and decreasing trends of various durations.

For example, for the station at Rio Claro, Singh [2] states: The station at Rio Claro on the foothills of the Central Range shows that dry season total rainfall has stayed more or less at around 400 mm, except for a slight increase in the 1960s and 1970s and a slight decrease in the 1980s (**Figure 17(a)**). Wet season total rainfall, however, increased steadily from about slightly in excess of 700 mm to about 1200 mm in the early 1980s before dropping off to about 1000mm in recent years (**Figure 17(b)**). The transition season shows a similar trend, increasing from about

1000 mm in the 1950s to about 1500 mm in the late 1980s and then declining slightly in recent years (**Figure 17(c)**).

In spite of obvious oscillations in the moving-average generated graphs, Singh [2] concludes that more extreme rainfall conditions in the dry and wet seasons and a tendency towards increasing rainfall in the transitional season therefore seems to be the trend.

Given the fact that the data series are random and stationary, the present study therefore demonstrates that Singh's [2,3] conclusions of changing seasonal rainfall patterns in Trinidad are without any valid statistical basis, inaccurate and therefore misleading. Moreover, the data series can be described as random noise superimposed on a constant mean. The use of frequency analysis is therefore appropriate and is consequently recommended for predicting seasonal rainfall amounts in Trinidad.

7. Conclusions

The major conclusions that can be drawn from the present study with respect to seasonal rainfall amounts in Trinidad are as follows:

1. Seasonal rainfall amounts are both random and stationary and therefore free from deterministic components such as trends, jumps and cycles. The data series can therefore best be described as random noise superimposed on a constant mean.

2. Recent claims of changing seasonal rainfall patterns (i.e., drier dry seasons, and wetter wet and transitional seasons) have no valid statistical basis, are inaccurate and therefore misleading.

3. The use of frequency analysis is the appropriate statistical method for modelling seasonal rainfall amounts for prediction purposes.

Acknowledgement

The author is grateful to the United States Information Agency (USIA) for funding his visit as a Research Scholar to the Department of Biological and Agricultural Engineering, University of Georgia, Athens, USA where the literature survey and data analysis for this study were carried out. In addition, he wishes to express his appreciation to the staff at the Piarco Meteorological Office and the Water Resources Agency for providing the data used in the study.

References

[1] Trenberth, K.E. (1997). *The Use and Abuse of Climate Models.* Nature, Vol. 386, No. 6621, pp. 131-133.

[2] Singh, B. (1997). *Climate Changes in the Greater and Southern Caribbean.* International Journal of Climatology, Vol. 17, No. 10, pp. 1093-1114.

[3] Singh, B. (1997). *Climate-related Global Changes in the Southern Caribbean: Trinidad and Tobago.* Global and Planetary Change, Vol. 15, pp. 93-111.

[4] Gani, J. (1975). *The Use of Statistics in Climatological Research.* Search, Vol. 6, No. 11-12, pp. 504-508.

[5] Kendall, M. (1976). *Time-Series.* 2nd edition. Charles Griffin & Company Ltd., London.

[6] Gottman, J.M. (1981) *Time-Series Analysis: A Comprehensive Introduction for Social Scientists.* Cambridge University Press, Cambridge.

[7] Barry, R.G. and Chorely, R.J. (1987). *Atmosphere, Weather and Climate.* 5th edition. Methuen, London.

[8] Jassby, A.D. and Powel, T.M. (1990). *Detecting Changes in Ecological Time Series.* Ecology, Vol. 71, No. 6, pp. 2044-2052.

[9] Harvey, A.C. (1993). *Time Series Models.* 2nd edition. Harvester Wheatsheaf, New York.

[10] Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

[11] Matalas, N.C. (1967). *Time Series Analysis: Water Resources Research.* Vol. 3, No. 3, pp. 817-829.

[12] Haan, C.T. (1977). *Statistical Methods in Hydrology.* The Iowa State University Press, Iowa.

[13] Wilks, D.S. (1995). *Statistical Methods in the Atmospheric Sciences*. Academic Press, California.

[14] Bobée, B. and Ashkar, F. (1991). *The Gamma Family and Derived Distributions Applied in Hydrology*. Water Resources Publications, Colorado.

[15] Brockwell, P.J. and Davis, R.A. (1996). *Introduction to Time Series and Forecasting*. Springer-Verlag, New York.

[16] Bowerman, B.L. and O'Connell, R.T. (1979). *Time Series and Forecasting: An Applied Approach*. Duxbury Press, Massachusetts.

[17] Walpole, R.E. and Myers, R.H. (1989). *Probability and Statistics for Engineers and Scientists*. 4th edition. Macmillan Publishing Company, New York.