MEMÓRIA DE ESTATÍSTICA

DA

UNIVERSIDADE FEDERAL DE PERNAMBUCO

RAINFALL AT FORTALEZA IN BRAZIL REVISITED

PEDRO A. MORETTIN

AFRÂNIO R. DE MESQUITA

JACIRA G. C. DA ROCHA

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

MEMÓRIAS DE ESTATÍSTICA - UFPE

As Memórias de Estatística se constituem numa rápida veiculação de trabalhos realizados por pesquisadores da UFPE e aceitos para publicação em revistas periódicas, atas de simpósio, colóquios, congressos, etc., e que sejam considerados de bom nível científico pelo Corpo Editorial.

São aceitos, excepcionalmente, para publicação textos de conferências realizadas no Departamento de Estatística por professores visitantes e conferências externas realizadas por pesquisadores do Departamento de Estatística, quando o conteúdo das mesmas for de divulgação recomendável por abordar temas relevantes e/ou resultados inéditos.

DEPARTAMENTO DE ESTATÍSTICA - UFPE
Endereço: Av. Professor Luiz Freire, s/n
Cidade Universitária
50 000 - Recife - PE
Tiragem: 50 exemplares

RAINFALL AT FORTALEZA IN BRAZIL REVISITED

Pedro A. Morettin*, Afranio R. de Mesquita** & Jacira G. C. da Rocha***

* Department of Statistics, University of São Paulo

** Department of Physical Oceanography, University of São Paulo

*** Department of Statistics, Federal University of Pernambuco.

RAINFALL AT FORTALEZA IN BRAZIL REVISITED

Pedro A. Morettin - Department of Statistics, University of São Paulo.

Afranio R. de Mesquita - Department of Physical Oceanography, University of São Paulo

Jacira G.C. da Rocha - Department of Statistics, Federal University of Pernambuco.

1. INTRODUCTION

The series of rainfalls at Fortaleza, Ceará, Brazil has been analysed by several authors recently, its importance depending on the fact that it is probably the longest series available for the study of the severe droughts that affect the Brazilian North-East.

The series consists of 131 years of annual data, from 1849 to 1979. It has been argued that this series is not appropriate for forecasting purposes, since the climate of Fortaleza is influenced by the sea and hence is not representative of the rest of the area. This question has been discussed by Girardi and Teixeira (1978) who have shown that there is a great similarity belween the behavior of the rainfall at Fortaleza and that at other sites of the region.

Recent studies as in Markham (1974), using "seasonalized" annual totals, 1849-1970, concluded that there were thirteen and twenty-six year periodicities in the data, and provided speculation as to the causes of these apparent periodicities. See also Markham (1967).

Jones and Kearns (1976) reanalysed the same data and concluded that the hypothesis that the series consists of statistically independent observations could not be rejected at the 10% Level. They based their conclusions on tests of serial correlation, estimated spectrum and cumulative periodogram.

Girardi and Teixeira (1978) used the same periodicities found by Markham to predict a severe drought in the area from 1978 to 1983. Further studies are those of Almeida et al.(1980) and Kantor (1982). This last author used the method of maximum entropy (Burg, 1975) to produce forecasts for the series.

In this paper we will provide a careful analysis of the rainfall series in order to find if the data supports the existence of periodicities. A mixed-spectrum analysis will be carried out and several tests will be used to detect the presence of harmonic terms.

In section 2 we present the data and some preliminary remarks. In section 3 we describe the statistical time series methodology we will use. The analysis of the series is performed in section 4 and a tentative model is discussed in section 5. We conclude the paper with some further comments.

2. THE DATA AND PRELIMINARY REMARKS

The observations are given in Appendix A and their plot is presented in Figure 1. From a visual inspection of both, it is not easy to detect noticeable trend and periodical patterns. The sample mean is 1425mm, the sample variance 230,52mm^2 while the minimum and maximum values are 468mm (1877) and 2512mm (1974), respectively.

The climate of the region may be classified as semi-arid, and the corresponding drought may be termed seasonal, occurring when the Inter-Tropical Convergence Zone (ITCZ) does not move up to the region in the period February-April.

During the period considered (1849-1979) major droughts occurred in 1877-1879, 1888-1889, 1898, 1900, 1903-1904, 1907-1908, 1915, 1919, 1932, 1936, 1951, 1953, 1958.

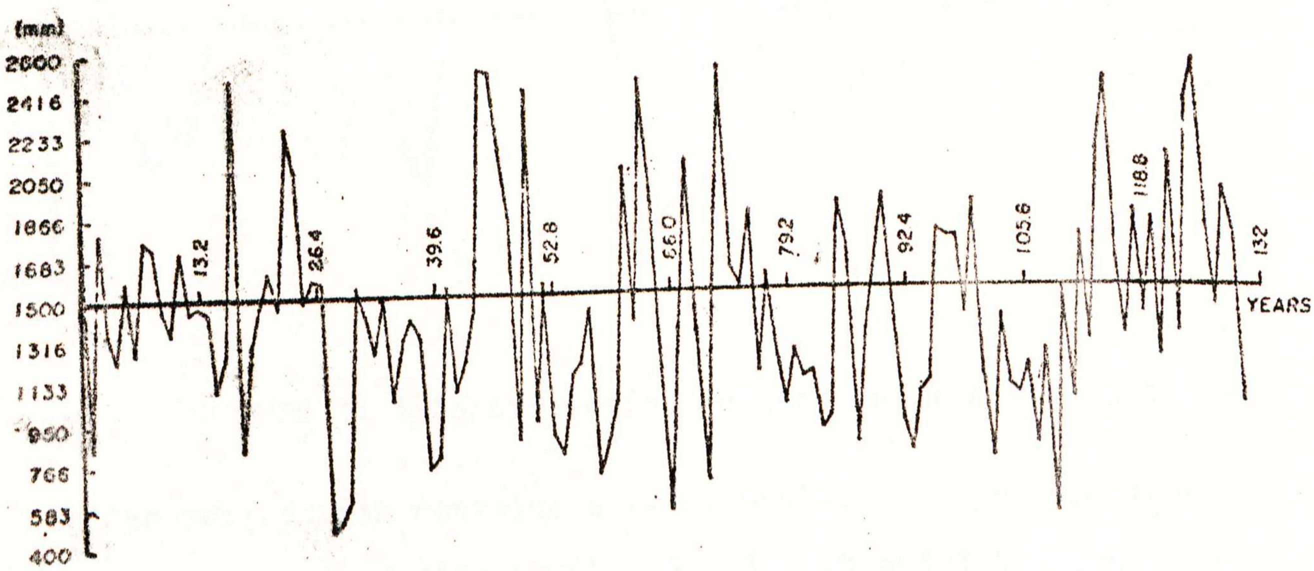

FIGURE 1: The Rainfall Series, 1849-1979

The least squares line for the raw data is

$$x_t = 1390.3 + 0.537 \quad (t-1849)$$

which would indicate a very small positive trend; but this is not statistically significant, as can also be seen using other trend tests (Cox and Stuart tes, for example; see Conover, 1971,p.130).

Figure 2 shows the plot of the autocorrelation function, and the periodogram of the data is given in Figure 3. The first sample autocorrelation is r_1 = 0.24 (significant at 5% level); this shows that there is a low year-to-year dependence. The remaining values tend to oscillate and do not show a definite periodical pattern.

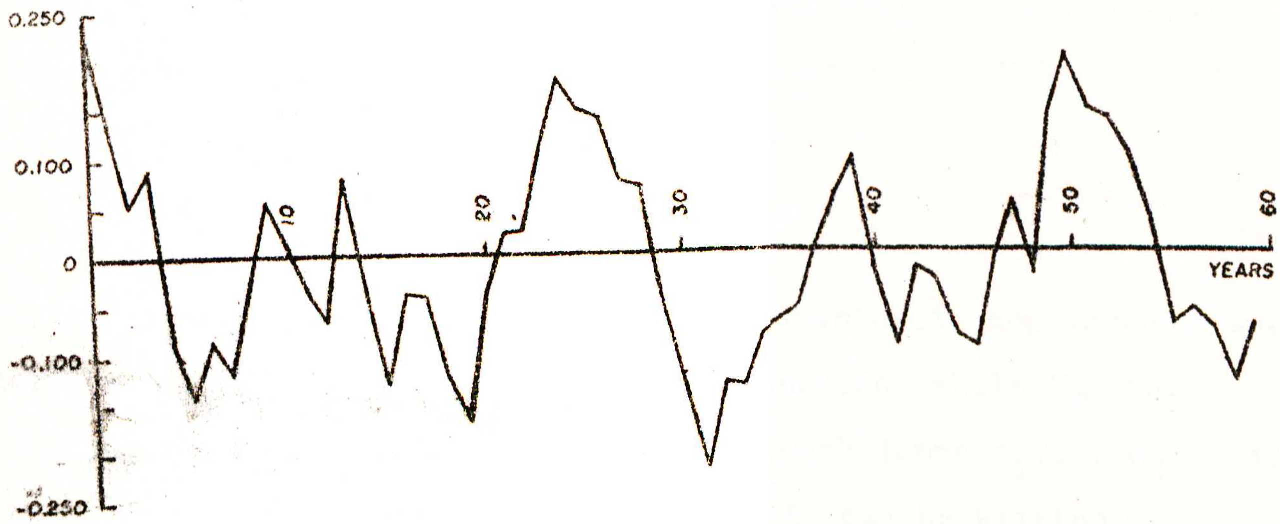

FIGURE 2: Autocorrelation Function of the Rainfall series

The periodogram contains several peaks, the more prominent being those corresponding to 65, 26, 13, 4.8 and 3.6 years. The significance of these peaks will be discussed in section 4.

3. A MIXED SPECTRUM ANALYSIS FOR THE RAINFALL SERIES

The conclusions drawn by Jones and Kearns (1976), based on a shorter series, suggest that the (mean-corrected) rainfall at Fortaleza behaves like a white noise series, with a constant spectrum. If the series contains periodicities, its spectrum would have a mixed form; that is, peaks (corresponding to the periodical components) will emerge from a continous spectrum.Then a model that seems adequate to describe the series may by developed as follows.

Denote by X_t, $t=1,2,\ldots,T$, the observations of a discrete parameter, zero mean, stationary process, with a mixed spectrum. This means that the spectral distribution function $F(\lambda)$ of the process X_t, $t=0, \pm1,\pm2,\ldots$ may by written

$$F(\lambda) = F_c(\lambda) + F_d(\lambda) \tag{3.1}$$

where F_c (the continuous component) is absolutely continuous and $F_c'(\lambda) = F(\lambda)$ is the spectral density function, while F_d (the discrete component) is a step function with jumps $p_1,\ldots,p_J$ at frequencies $\lambda_1,\ldots,\lambda_J$. It follows that X_t can be written

$$X_t = Y_t + Z_t \tag{3.2}$$

where Y_t and Z_t are uncorrelated processes, Y_t corresponding to the continuous part of the spectrum and Z_t to the discrete part. Further:

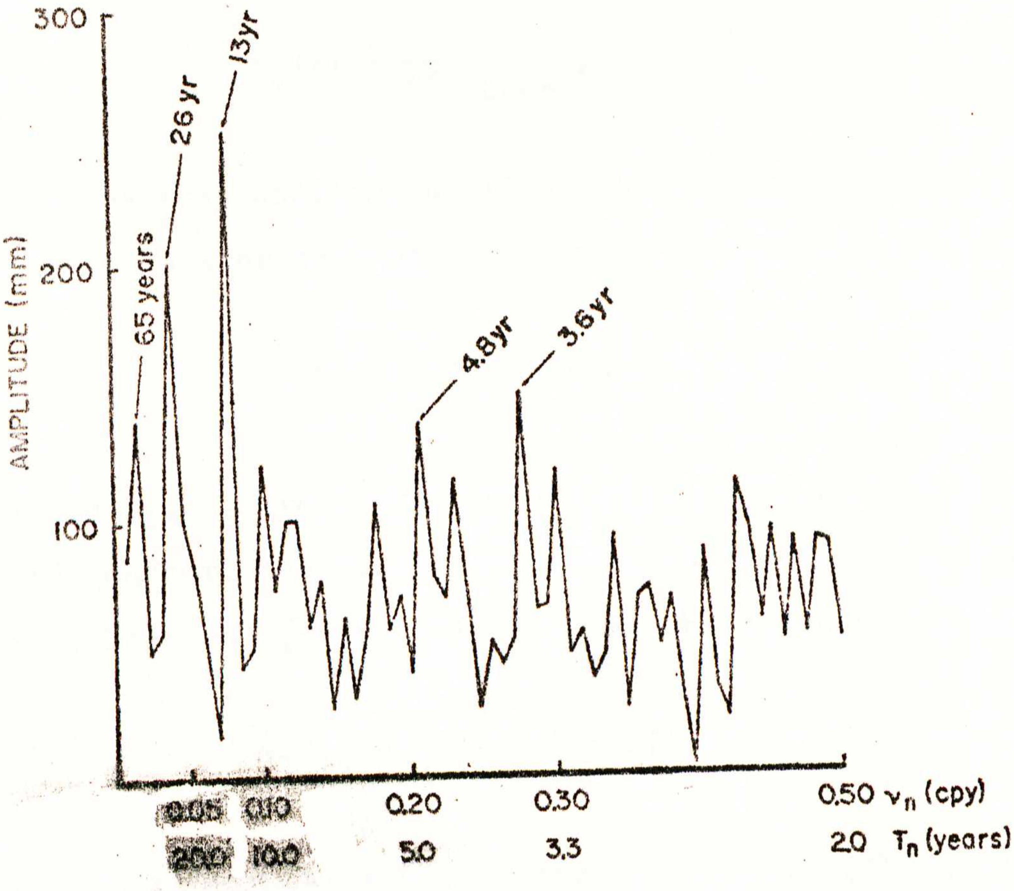

FIGURE 3: Periodogram of the Rainfall Series

i) Y_t is a linear process, that is,

$$Y_t = \sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k} \tag{3.3}$$

where ε_t is a white noise series, with mean zero, variance σ_ε^2 , $E(\varepsilon_t \varepsilon_s) = 0$, $s \neq t$, and β_k are constants satisfying $\sum_{k=0}^{\infty} \beta_k^2 < \infty$ and

$\sum_{k=0}^{\infty} k|\beta_k|<\infty$. Let $\gamma_Y(k)$ denote the autocovariance function of Y_t and

$$f_y(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_Y(k)e^{-i\lambda k} \tag{3.4}$$

its spectral density function, assuming that $\sum_{k=-\infty}^{\infty} |\gamma_Y(k)|<\infty$;
ii) Z_t is a process of the form

$$Z_t = \sum_{J=1}^{J} A_j \cos(\lambda_j t + \phi_j') \tag{3.5}$$

where A_j, λ_j are unknown constants, $j=1,\ldots,J$, and ϕ_j are independent, identically distributed, rectangular random variables on $[-\pi,\pi]$. If $\gamma_Z(k)$ is the autocovariance of Z_t, then

$$\gamma_Z(k) = \frac{1}{2} \sum_{j=1}^{J} A_j^2 \cos(\lambda_j k) \tag{3.6}$$

and $\gamma_X(k) = \gamma_Y(k) + \gamma_Z(k)$. The $\{A_j, j=1,\ldots,J\}$ forms the discrete (or line) spectrum of X_t and $f_Y(\lambda)$ the continuous spectrum.

A mixed spectrum analysis of X_t consists of:

i) estimating the amplitudes A_j and the frequencies λ_j, in order to obtain the discrete spectrum of X_t;

ii) estimating the continuous spectrum $f_Y(\lambda)$ of X_t.

It follows that the first step in the analysis is to test for the presence of periodical components, i.e., to test the null hypothesis that $A_j=0$, for all j. If we find that there are J periodical components, we estimate the λ_j, A_j and remove the contribution of these components from X_t; after this is done, the spectrum $f_Y(\lambda)$ is estimated from the residuals $X_t-\hat{Z}_t$, using standard techniques of spectrum estimation.

Several tests are available for testing the existence of periodic components in a set of data. These tests were not considered by the previous outhors who dealt with the rainfall series.

We decided to apply all the tests below since we are not aware of any study comparing their powers for finite samples.

a) An extension of Fisher's g-test (Fisher, 1929, Whittle,1952)

b) Whittle's test (Whittle, 1952);

c) Hannan's test (Hannan, 1961);

d) Bartlett's test (Bartlett, 1955);

e) Priestley's $P(\lambda)$ test (Priestley, 1962a,b).

A comparision of the asymptotic powers of the Whittle's test, the grouped periodogram test and the $P(\lambda)$ test was given by Priestley (1962b). We also used a white noise test based on the cumulative periodogram (Jenkins and Watts, 1968).

In Appendix B we briefly describe these tests.

4. ANALYSIS OF THE RAINFALL SERIES

The tests mentioned in section 3 (and described in Appendix B) were applied to the data. Unless otherwise stated we shall assume for all tests an overall significance level $\alpha = 0.05$.

(a) *Extension of Fisher's Test*

We used the suggestion of Whittle (1952) (see Appendix B). The value of $g^{(1)}$ is 0.136, and for m = 65 we obtain the critical value 0.0961. Therefore, the peak corresponding to 13.1 years is significant.

Applying the test to the second largest ordinate, correcting the denominator of (B.2), we observe $g^{(2)} = 0.1012$. Since, for m = 64, the critical value is 0.0984, we accept the presence of a periodical component with period 131/5 = 26.2 years.

For the third largest ordinate we obtaing $g^{(3)} = 0.0623$, which is not significant and we reject a periodicity of 131/36 = 3.64 years.

(b) *Whittle's Test*

The largest periodogram ordinate occurs at j = 10 (the frequency is $\lambda_{10} = 2\pi \times 10/131 \approx \pi/6$), with $g_W = 0.145$, which is significant (the critical value is 0.0964). Therefore, we accept a periodical component with period of 13.1 years.

For the second largest (j=5), g_W=0.092, which is greater than the critical value 0.0699 and we conclude that a 26.2 year periodicity is present.

For the third largest, $g_W = 0.056$, which is not significant (0.061 is the critical value), and we reject the existence of a harmonic term with period 3.64 years.

(c) *Bartlett's Test*

Let us choose k=4, 5, 7, 10, 20. The corresponding values of $g_{k,\ell}^{(B)}$ are given in Table 1, for some values of ℓ. For the 5% significance level, the critical values are given by $0.05k/|T/2| = k(1-g)^{k-1}$ and these are also shown in Table 1. We see that the only significant value is that for k=4 and ℓ=3, corresponding to the period of 13.1 years. We recall that taking k small is the only way to obtain a good approximation of $g_{k,\ell}$, using $g_{k,\ell}^{(B)}$. On the other hand, we do not expect to get segnificant results for large values of k, since this implies that we are assuming that the spectrum $f_Y(\lambda)$ is approximately constant over a broad band of frequencies.

TABLE 1: Values of $g_{k,\ell}^{(B)}$ and g

k \ ℓ	1	2	3	9	g
4	.5904	.6670	.9267*	.7417	.9084
5	.5478	.7625	.4276		.8335
7	.4421	.6748			.6973
10	.4115	.2528			.5492
20	.2983	.1724			.3143

(d) *Hannan's Test*

For the largest ordinate, we obtaing g_H=0.145 , which is significant against the critical value 0.13135, and the harmonic with period 13.1 is accepted. Testing for the second largest ordinate, we get g_H=0.092, which is greater than the tabled value

0.0699, and we accept as significant the peak corresponding to 26.2 years.

It is easy to see that we reject further periods. As expected, the test gives the same results as the Whittle test, due to the choice of the window $W(\theta)$ as that corresponding to the truncated periodogram.

(c) *The $P(\lambda)$ Test*

The calculations were done with m=25 and n=55; other values were tried, but we will only describe these results. The graph of $P(\lambda)$ (Figure 4) shows the presence of well defined peaks at $4\pi/131$, $10\pi/131$, $20\pi/131$, $30\pi/131$ and $40\pi/131$.

For $\lambda=4\pi/131$, we have $J_2=1.225$ and for $\alpha=1\%$, we have $\alpha_0=2.33$ and since $J_2<\alpha_0$, we see that the peak at this λ is not significant.

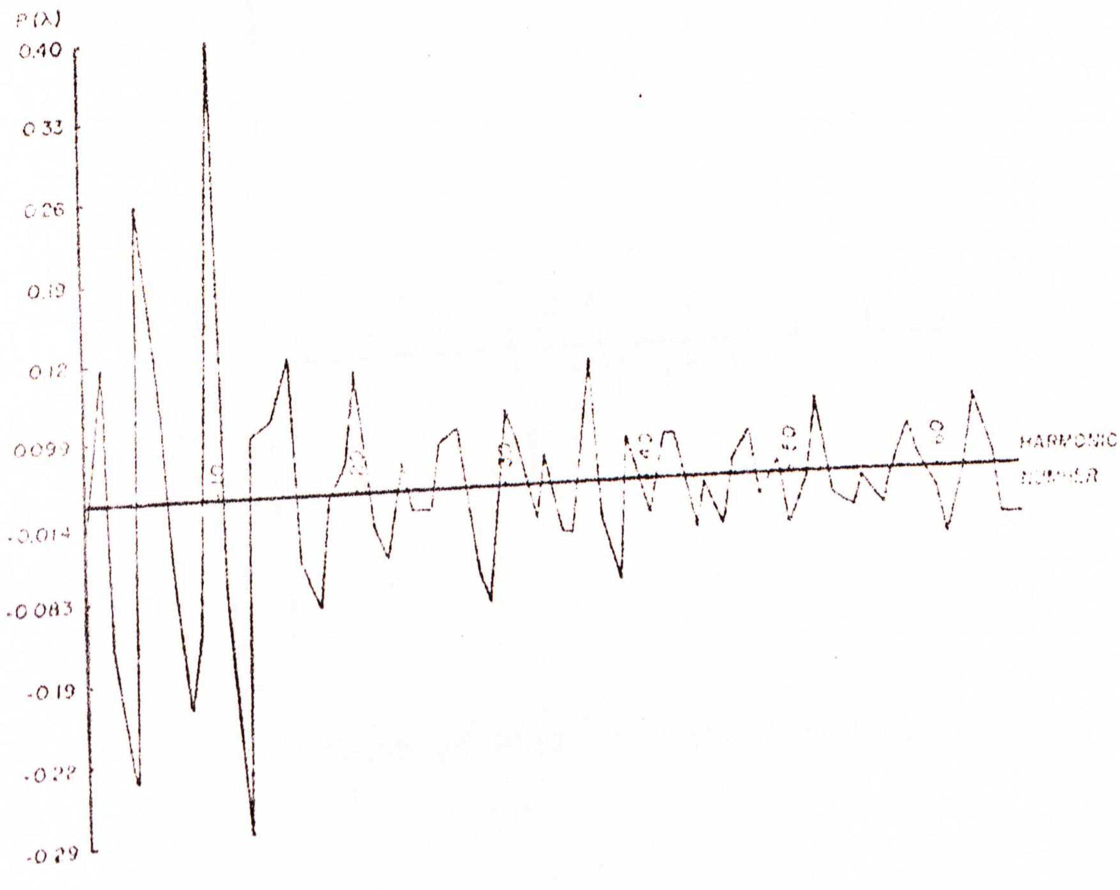

FIGURE 4: Graph of $P(\lambda)$ for the Rainfall Series

Next, in order of frequency, we test the peak at $\lambda=10\pi/131$ / and obtain $J_2=2.603$, which is significant ($\alpha_0=2.58$, corresponding to $\frac{\alpha}{2} = 0.005$) and we accept a harmonic component with period 26.2 years.

Removing the contribution of this significant harmonic term at the frequency $\hat{\lambda}_1 = 10\pi/131$, we obtain the new graph of $P(\lambda)$, which we call $P'(\lambda)$ (Figure 5). This has peaks at frequencies $20\pi/131$ and $72\pi/131$, and testing again in order of frequency we find that the peak at $\lambda=20\pi/131$ is significant, since $J_2=3.026$, is greater than the critical value $\alpha_0=2.71$. Hence we accept a periodicity of 13.1 years.

Removing again the contribution of this component we get $P''(\lambda)$ (Figure 6) with

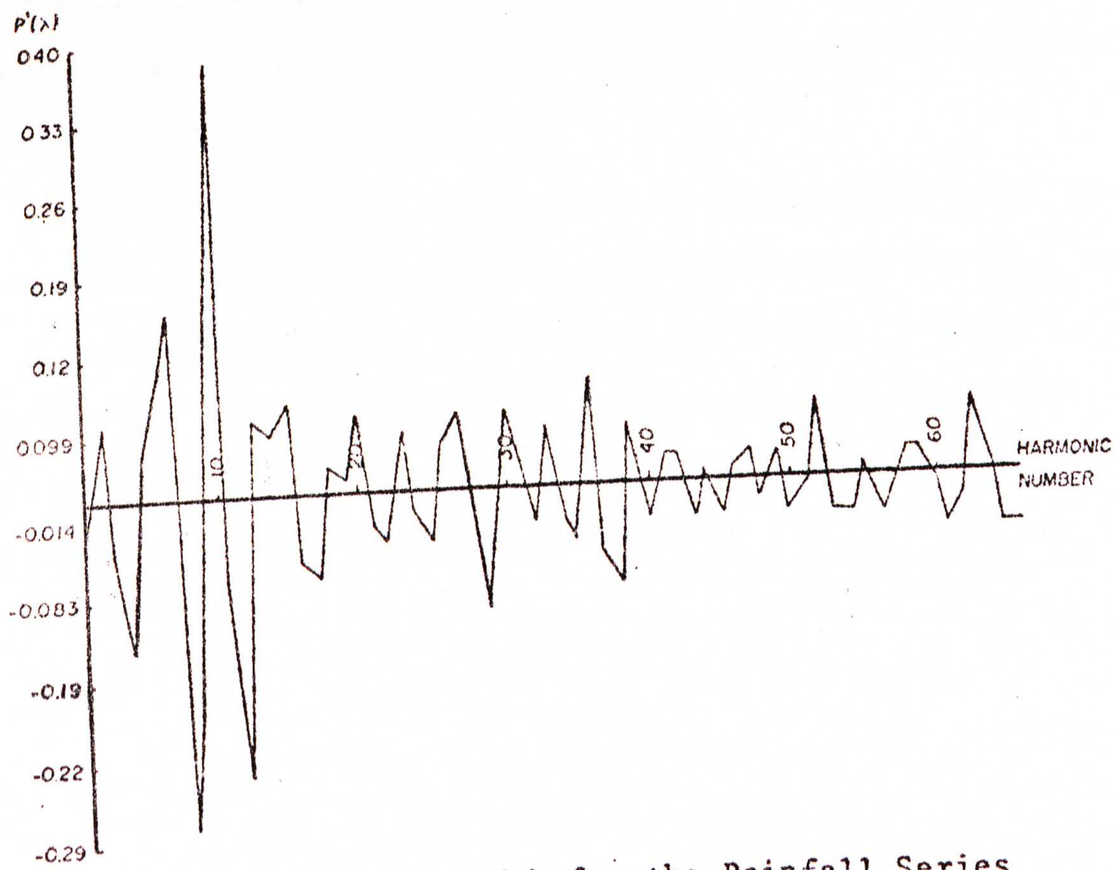

FIGURE 5: Graph of $P'(\lambda)$ for the Rainfall Series

peaks at the frequencies $46\pi/131$, $54\pi/131$, $66\pi/131$ and $72\pi/131$. For $\lambda=46\pi/131$ we find $J_2=0.810$, which is not significant and therefore we reject a periodicity of 5.7 years. The remaining peaks are not significant.

(f) *The White Noise Test*

The use of the test described in section B.2 of Appendix B on the rainfall data gave the result shown in Figure 7. It differs from the one obtained by Jones and Kearns (1976), since in the present case the series cannot be taken as white noise. The only difference in the calculation is that the present series has 10 years of data points more than the series used by them. This suggests that even other periods could eventually be added to the set of significant periodicities, as more years of observations allow them to come up from the overall background continuous spectrum.

The results of the application of the several tests are summarized in Table 2, and seem to indicate that the 13.1 and 26.2 years periodicities are present in the data, since four out of the five tests detected them both, while all the tests detected the 13.1 year periodicity. Therefore, we shall proceed to establish a model for the series.

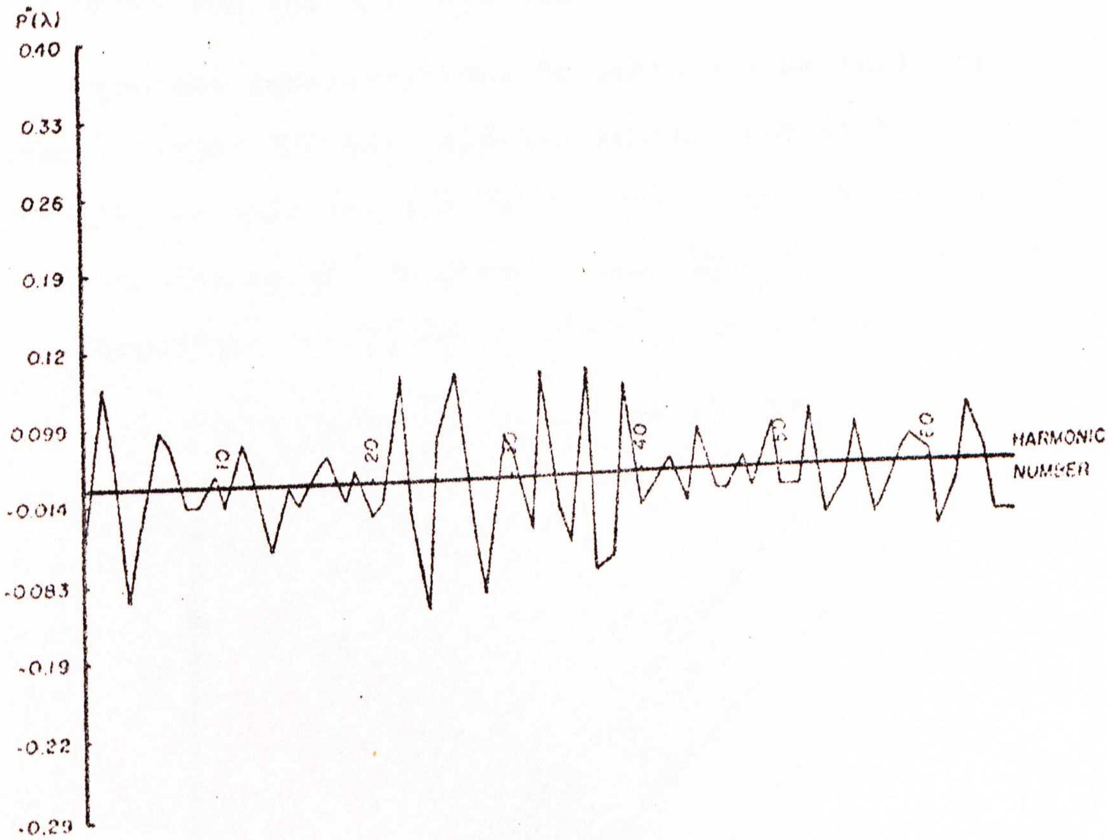

FIGURE 6: Graph of P''(λ) for the Rainfall Series

TABLE 2: Summary of the application of the tests

Test	Periods or remarks
Ext. Fisher	13.1, 26.2 years
Whittle	13.1, 26.2 years
Bartlett	13.1 years
Hannan	13.1, 26.2 years
Priestley	13.1, 26.2 years
White Noise	series is not white woise

5. A MODEL FOR THE RAINFALL SERIES

From the considerations in section 4 we shall try two harmonic terms for the rainfall series: one with frequency $\lambda_1=10\pi/131=0,2398$ and the other with frequency $\lambda_2=20\pi/131=0.4796$, corresponding to the periods of 26.2 and 13.1 years respectively.

Therefore, we can write (3.5), with J=2, as

$$Z_t = \mu + A_1\cos(\lambda_1 t+\phi_1) + A_2\cos(\lambda_2 t+\phi_2) \qquad (5.1)$$

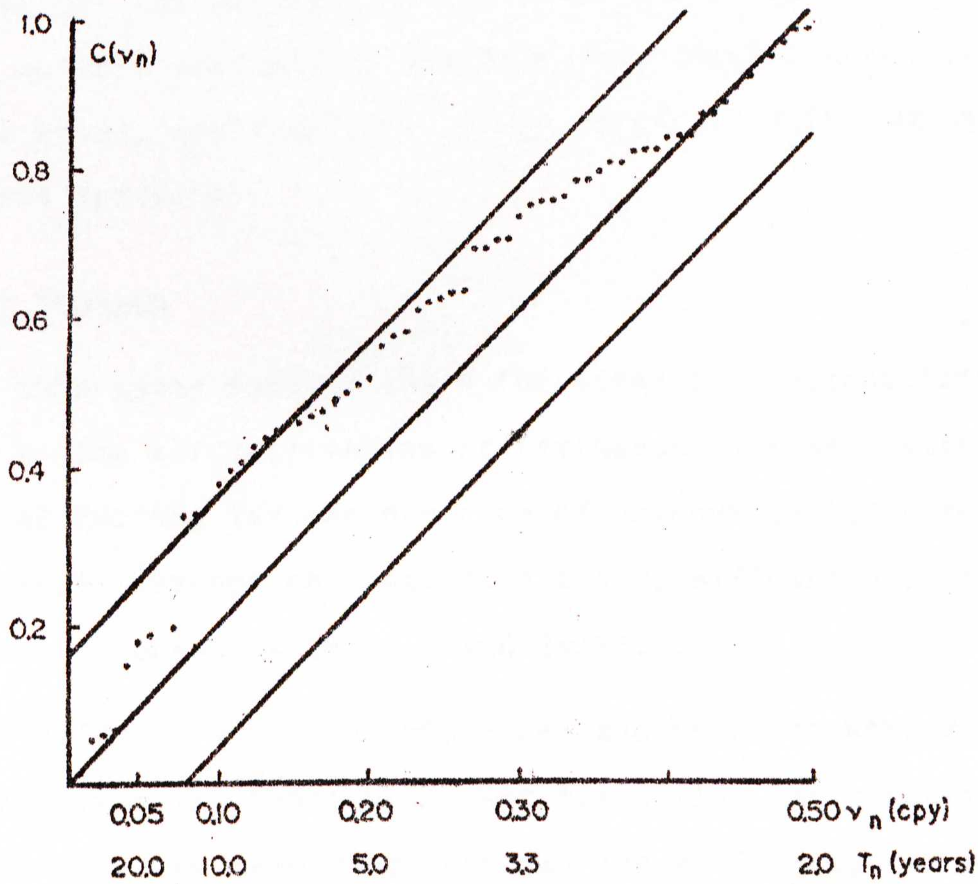

FIGURE 7: Normalized Accumulated Periodogram for the Rainfall Series, with Theoretical and Confidence Lines

and least squares estimates of A_i and $\phi_i (i=1,2)$ and μ are given / by

$\hat{A}_1$ = 202.5mm, $_1$ = 114°

$\hat{A}_2$ = 255.6mm $_2$ = 126°

$\hat{\mu} = \bar{X}$ = 1424.8mm,

respectively. Thus we can write the model for X_t as:

$$X_t = 1424.8 + 202.5 \cos(0.2398t + 1.99) + 255.6 \cos(0.4796t + 2.20) + Y_t \quad (5.2)$$

where Y_t has a continuous spectrum.

Let $\hat{Z}_t$ be the estimated harmonic polynomial and $\hat{Y}_t = X_t - \hat{Z}_t$ be the residual series. Figure 8 shows the plots of the original and $\hat{Z}_t$ series.

It is readily seen that the sample autocorrelation of the $\hat{Y}_t$ series are all nonsignificant. Also, a white noise test applied to this residual series gives the results shown in Figure 9. Hence we may conclude that the Y_t series is white noise, with a constant spectrum.

6. FINAL REMARKS

In this paper several tests for detecting periodicities were applied to the rainfall series at Fortaleza in Brazil with the purpose of testing for the presence of unknown periodic terms. All the tests, except one, detected two significant harmonic terms with frequencies $2\pi/26.2$ and $2\pi/13.1$.

The evidence given in this paper and by other workers seems to suggest that the hypothesis that the series has a continuous (uniform) spectrum should be rejected and a mixed spectrum model is more appropriate for it.

But, as we have emphasized before, care should be taken in drawing definite conclusions. As we gather more observations, further analyses could be carried out and eventually other suspected periodicities, such as the one corresponding to 65 years, might be confirmed.

A suggested model for the series is given by (5.2), where Y_t is a white noise process. We also fitted an autoregressive process to the data, using both the Yule-Walker and Burg's method for estimating the coefficients and the FPE criterion of Akaike to determine the order of the model. In both cases a fourth-order model was obtained. The corresponding (maximum entropy or auto-regressive) spectral estimates did not detect any harmonic term. To resolve the periodicities of 13.1 and 26.2 years, a much higher order (of about T/2) was necessary.

The statistical model presented predicts a long period (1980 -1985) of below average precipitation annual means. This, so far, is in fair agreement with the recent reports on the drought of the area.

The agreement draws attention to the periods of 13 and 26 years, pointed out in the analysis, and to their physical explanation. This has long been a matter of

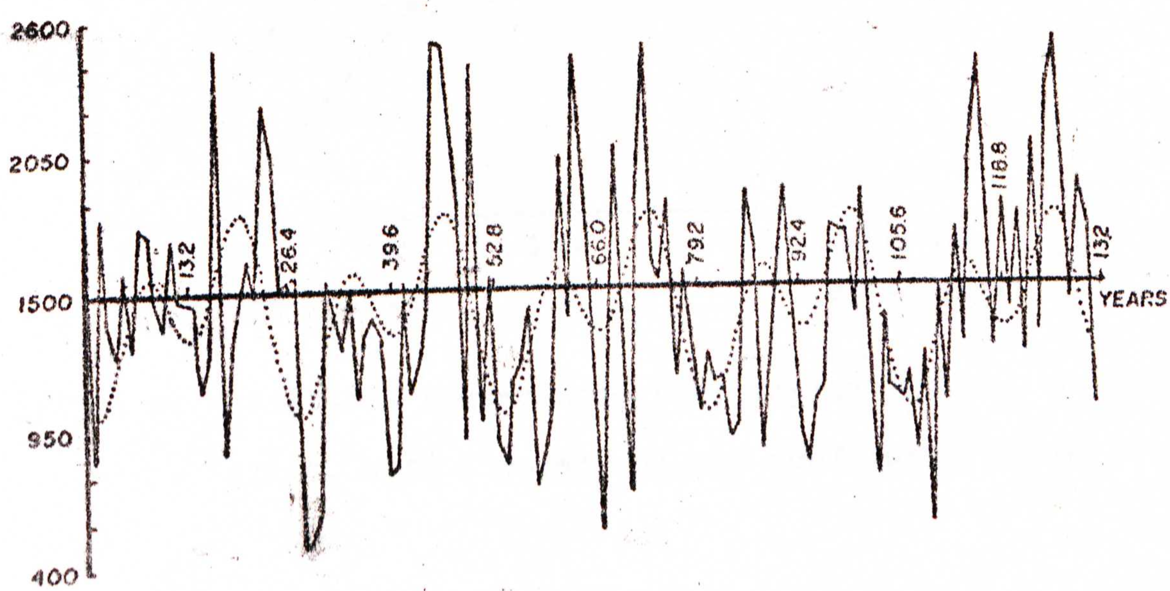

FIGURE 8: Original Series (Solid Curve) and Fitted Z_t (Dotted Curve)

concern, and many causes have been suggested to explain the occurrence of the drought in the norteast area of Brazil. Some try to relate it to solar activity, others try to find a possible connection with the occurrence of the "El Niño" on the coast of Peru (Caviedes, 1973). The 13 years cycle may be a result of a slightly out of phasing between the annual and the quasi-biannual period oscillations of the trade winds of the South Atlantic (Reiter, 1979).

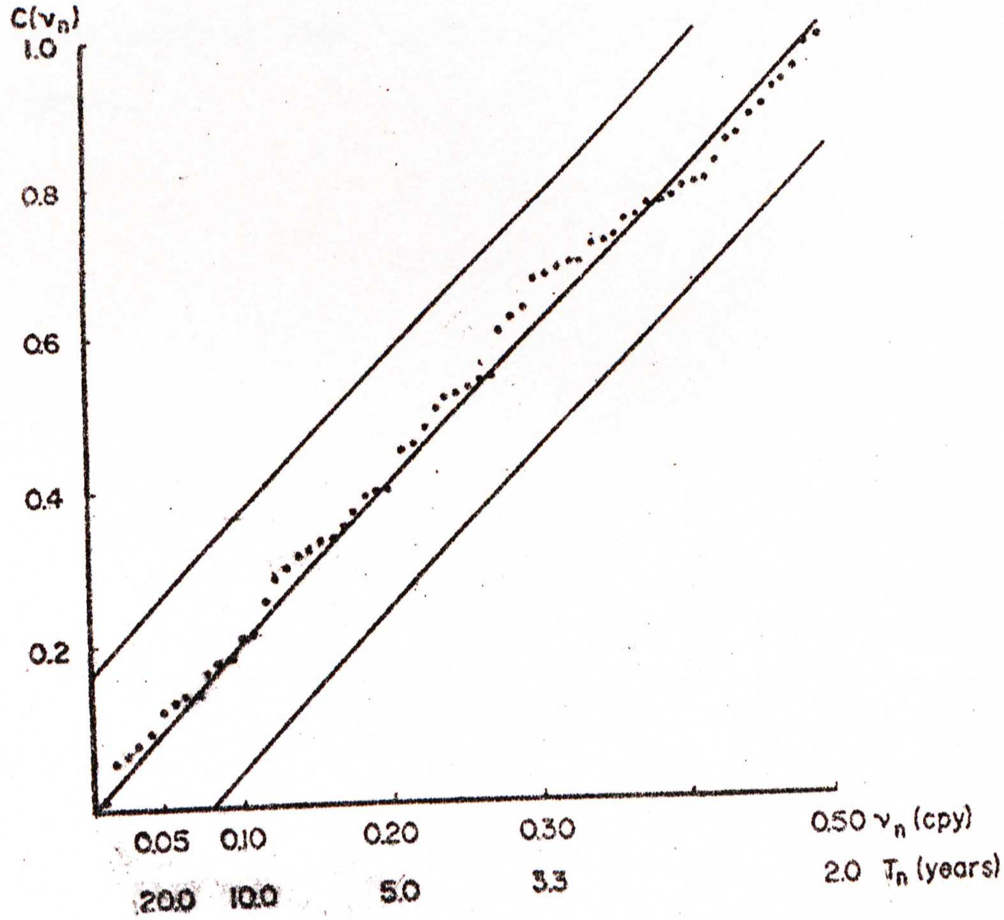

FIGURE 9: Normalized Accumulated Periodogram for the Residual Series, with Theoretical and Confidence Lines

Recent preliminary analyses of the mean sea level, made by the authors at the port of San Francisco, USA, show a periodicity of 12.4 years, among others. The evidence so far gathered seems to indicate that there is a great deal of oceanic causation in defining the 13 years periodicity of the Fortaleza rainfall. But certainly large amounts of data, not yet available, must be collected on several oceanographic, meteorological and hydrological time series, before definite conclusions can be drawn.

AKNOWLEDGENT: The authors would like to thank two referees and the Editor for comments which improved the presentation of the paper.

This work was partially supported by CNPq, FAPESP CAPES and FINEP grants.

APPENDIX A: The Rainfall Series, 1849 - 1979

Year	Rainfall	Year	Rainfall	Year	Rainfall	Year	Rainfall
1849	2001	1882	1246	1915	530	1948	1384
1850	852	1883	1508	1916	1328	1949	1881
1851	1806	1884	1047	1917	2077	1950	1114
1852	1356	1885	1307	1918	1319	1951	747
1853	1233	1886	1399	1919	656	1952	1378
1854	1950	1887	1320	1920	1847	1953	1068
1855	1273	1888	736	1921	2496	1954	1032
1856	1770	1889	784	1922	1595	1955	1152
1857	1734	1890	1534	1923	1513	1956	806
1858	1457	1891	1077	1924	1847	1957	1225
1859	1357	1892	1211	1925	1137	1958	504
1860	1716	1893	1430	1926	1571	1959	1493
1861	1445	1894	2505	1927	1195	1960	1011
1862	1468	1895	2491	1928	995	1961	1737
1863	1452	1896	2144	1929	1230	1962	1258
1864	1098	1897	1839	1930	1107	1963	2102
1865	1238	1898	863	1931	1133	1964	2428
1866	2478	1899	2414	1932	879	1965	1630
1867	832	1900	940	1933	937	1966	1288
1868	1289	1901	1545	1934	1888	1967	1839
1869	1470	1902	878	1935	1661	1968	1385
1870	1628	1903	789	1936	820	1969	1805
1871	1459	1904	1136	1937	1313	1970	1192
1872	2256	1905	1189	1938	1586	1971	2039
1873	2058	1906	1430	1939	1911	1972	1299
1874	1487	1907	697	1940	1447	1973	2331
1875	1581	1908	834	1941	916	1974	2512
1876	1569	1909	1015	1942	780	1975	1778
1877	468	1910	2051	1943	1042	1976	1417
1878	503	1911	1373	1944	1090	1977	1941
1879	597	1912	2446	1945	1750	1978	1752
1880	1539	1913	1905	1946	1724	1979	996
1881	1423	1914	1512	1947	1726		

SOURCE: Kantor (1982). The figures are in mm of rainfall

APPENDIX B

B.1. *Tests for Detecting Periodicities*

Given T observations $X_1,\dots,X_T$ of X_t, let

$$I_X^{(T)}(\lambda) = \frac{2}{T}\left|\sum_{t=1}^{T} X_t e^{-i\lambda t}\right|^2 \tag{B.1}$$

be the *periodogram* of these values. We shall evaluate (B.1) at the frequencies $\lambda_j = 2\pi j/T$, $j=0,1,\dots,[T/2]$, called Fourier frequencies, and call $I_j^{(T)} = I_X^{(T)}(\lambda_j)$.

(a) *Fisher's Test and an Extension*

Fisher's test is used for testing the value of the largest peak in the periodogram. The model (3.2) is assumed for X_t, but Z_t is now assumed to be white noise. We test $H_0:A_j=0$, for all j, under the condition that X_t is Gaussian.

Writting $I_j = I_j^{(T)}$ for brevity, Fisher's g-statistic is given by

$$g = \max(I_j) / \sum_{j=1}^{m} I_j \tag{B.2}$$

where $m = [T/2]$. Fisher (1929) derived (under H_0) the exact distribution of g, which is given by

$$P(g>z) = m(1-z)^{m-1} - \binom{m}{2}(1-2z)^{m-1} + \dots + (-1)^{k-1}(1-kz)^{m-1}, \tag{B.3}$$

where $k = [1/z]$. Tables of the distribution are given by Fisher (1929) and Shimshoni (1971).

If we reject H_0, we conclude that there is a periodicity in X_t at the frequency corresponding to $\max(I_j)$; if this occurs for $j=j'$, then this frequency is $\hat{\lambda}_0 = 2\pi j'/T$.

Whittle (1952) suggests that we may test for the next largest ordinate by omitting the term $I_{j'}$ from the denominator of (B.2) and adjust the value of m to m-1.

If we know that X_t has exactly r periodic components, when H_0 is false, then a test based on

$$g^{(r)} = I^{(r)} / \sum_{j=1}^{m} I_j \qquad (B.4)$$

can be used, where $I^{(r)}$ is the r-th largest ordinate. See Grenander and Rosenblatt (1957) for the distribution of $g^{(r)}$, of which (B.3) is a special case.

(b) *Whittle's Test*

We now return to the general model (3.2). To test the null hypothesis H_0, we use the statistic

$$g_W = \max_j \left[I_j / 2\pi \hat{f}_Y(\lambda_j) \right] / \sum_{j=1}^{m} \left[I_j / 2\pi \hat{f}_Y(\lambda_j) \right] \qquad (B.5)$$

and refer to Fisher's g-distribution with m degrees of freedom.

If $C_X(s) = T^{-1} \sum_{t=1}^{T-|s|} X_t X_{t+|s|}$ is the sample autocovariance, $\hat{f}_Y(\lambda)$ is the truncated periodogram estimate,

$$\hat{f}_Y(\lambda) = \frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} C_X(s) e^{-i\lambda s} \qquad (B.6)$$

$\ell < T$ being the truncation point.

(c) *Hannan's Test*

Hannan (1961) considers the problem of testing if a periodic component corresponds to a jump in the spectral distribution function $F(\lambda)$. The null hypothesis is that F is absolutely continuous, with derivative $f(\lambda)$, a smooth function, and the alternative hypothesis is that F has a jump.

The test statistic is essentially g_W, except that $f_Y(\lambda)$ is estimated through a windowed spectral estimate $f_Y^*(\lambda)$, of the form

$$f_Y^*(\lambda) = \int_{-\pi}^{\pi} I_X^{(T)}(\alpha)\, W_M(\lambda-\alpha)\, d\alpha. \tag{B.7}$$

Here $W_M(\lambda)$ is the spectral window and M is the truncation point. Under the null hypothesis, the statistic

$$g_H = \max_j \left[I_j/2\pi f_Y^*(\lambda_j)\right] / \sum_{j=1}^{m} \left[I_j/2\pi f_Y^*(\lambda_j)\right] \tag{B.8}$$

has approximately the Fisher distribution with m degrees of freedom.

(d) *Bartlett's (the grouped periodogram) Test*

Bartlett suggested a test for the purpose of separating spectral peaks with narrow bandwidths. Given a sample of T observations, it will be practically impossible to distinguish harmonic components from peaks in the continuous spectrum, with widths less than $2\pi/T$. See Bartlett (1957) and Priestley (1981).

Thus we will assume that $f_Y(\lambda)$ has bandwidth $B_f \geq 2\pi/T$. Let $k < B_f$ and divide the periodogram ordinates into $|T/2k|$ sets, each containing k ordinates. Let

$$g_{k,\ell} = I_{j'}/2\pi f_Y(\lambda_{j'}) / \sum_{j=(\ell-1)k+1}^{\ell k} I_j/2\pi f_Y(\lambda_j)$$

$$(\ell = 1,2,\ldots, \left[T/2\right] k^{-1}) \tag{B.9}$$

where $I_{j'}/2\pi f_Y(\lambda_{j'}) = \max\{I_j/2\pi f_Y(\lambda_j) : (\ell-1)k+1 \leq j \leq \ell k$

Under H_0, $g_{k,\ell}$ has asymptotically the same distribution as the Fisher's g-statistic with k degrees of freedom. Then $g_{k,\ell}$ may be approximated by

$$g_{k,\ell}^{(B)} = I_{j'}/\sum_j I_j \tag{B.10}$$

and even when $f_Y(\lambda)$ is unknown the test may be carried ou considering $g_{k,\ell}^{(B)}$ as having a Fisher distribution with k degrees of freedom. It can be shown that $g_{k,\ell}^{(B)}$ can differ considerably from $g_{k,\ell}$ and there is not any systematic way of choosing k. A compromise must be achieved between making k smaller than B_f, and sufficiently large to retain sufficient degrees of freedom.

Assuming that we may use $g_{k,\ell}^{(B)}$, it remains to adjust the significance level of the test. If α is the level for the original Fisher's test applied to k ordinates, then the approximate significance level for a test based on $g_{k,\ell}^{(B)}$ is $\alpha' = \alpha k/\left[T/2\right]$.

Then the procedure to follows is: i) choose somevalues of k; ii) for each k, test for the significance of the peaks, using

$g^{(B)}_{k,\ell}$; usually there will be no need to test all values of ℓ (look at the periodogram); iii) the critical value for $g^{(B)}_{k,\ell}$ based on a significance level α is approximately given by $\alpha k/[T/2]=k(1-g)^{k-1}$ using only the first term in (B.3).

(c) *Priestley's P(λ) Test*

The preceding tests have a number of disadvantages, which are summarized in Priestley (1981, p.625). The P(λ) test, developed by Priestley (1962a,b) is not based on the periodogram but on the autocorrelation function of the observed series.

We saw in section 3 that the autocovariance function of the series X_t is given by $\gamma_X(s)=\gamma_Y(s)+\gamma_Z(s)$, where $\gamma_Y(s)\to 0$, as $|s|\to\infty$, since Y_t has a purely continuous spectrum, and $\gamma_Z(s)$ is a combination of cosine waves with the same frequencies as Z_t, and therefore $\gamma_Z(s)\to 0$, as $|s|\to\infty$.

Therefore, if some of the A_j are nonnull, the autocovariance of X_t does not wear off as $|s|\to\infty$, and its tail will behave like a linear combination of cosine waves with the same frequencies as those of the periodic component.

The P(λ) test exploits this behavior of $\gamma_X(s)$, performing a Fourier analysis of the tail of $\gamma_X(s)$. Let m be such that $\gamma_Y(s) \simeq 0$, for $|s| > m$. To test $H_0: A_j=0$, all j, proceed as follows:

(i) compute $\hat{f}_m(\lambda)$ and $\hat{f}_n(\lambda)$, estimates of $f_Y(\lambda)$ using some spectral window, with truncation points m and n, $n > 2m$;

(ii) compute $P(\lambda) = \hat{f}_n(\lambda) - \hat{f}_m(\lambda)$, for $\lambda = 2\pi j/T$, $j = 0,1,\ldots,[T/2]$, and plot $P(\lambda)$ against λ. If $A_j \neq 0$, then $P(\lambda)$ will have well defined peaks;

(iii) test each peak in *order of frequency*; if the first peak appears at $\lambda_0 = 2\pi p/T$, subdivide the frequency range $(0,\pi)$ at intervals $2\pi/m$ on both sides of λ_0 and form

$$J_q = \left(\frac{T}{m}\Lambda_{n,m}^{-1}\right)^{\frac{1}{2}} \sum_{s=0}^{q} P^*\left(\frac{2\pi s}{m} + \delta\right) \left[\hat{g}(\pi)/2\pi\right]^{-\frac{1}{2}} \tag{B.11}$$

for $q=0,1,\ldots,[T/2]$ and test whether $\max(J_q)<\alpha_0$, where α_0 is the upper $100\alpha\%$ point of the standard normal, for a given significance level α. In (B.11), $P^*(\lambda) = P(\lambda)/C_Y(0)$

$$\hat{g}(\alpha) = \frac{1}{4\pi}\left\{ 2 \sum_{s=-m+1}^{m-1} r_Y^2(s) - \sum_{s=-2m+1}^{2m-1} r_Y^2(s) \right\}$$

J is chosen so that $\lambda_0 = 2\pi p/T = 2\pi s/m+\delta$, for some integer s, and

$$\Lambda_{n,m} = 2\pi \int_{\pi}^{\pi} \{W^{(1)}(\alpha) - W^{(2)}(\alpha)\}^2 d\alpha \tag{B.12}$$

In (B.12), $W^{(1)}(\lambda)$ and $W^{(2)}(\lambda)$ are the spectral windows corresponding to the truncation points n and m, respectively. If the Bartlett window is used for $W^{(1)}$ and $W^{(2)}$ we obtain

$$\Lambda_{n,m} = \frac{2}{3} n - \frac{4}{3} m + \frac{2}{3n}^2 .$$

In the formulae for $P^*(\lambda)$ and $\hat{g}(\pi)$, $C_Y(s)$ denote the sample autocovariance and sample autocorrelation functions of Y_t, respectively;

(v) if $\max(J_q)>\alpha_0$, then the peak at λ_0 is judged significant and the amplitude of the harmonic term at λ_0 is then estimated by

$$\hat{A}_0^2 = 8\pi P(\lambda_0)/(n-m) \quad (n>2m) \tag{B.13}$$

(vi) the effect of this harmonic component is then removed, computing

$$C_Y^{(1)}(s) = C_Y(s) - \frac{1}{2}\hat{A}_0^2 \cos(s\hat{\lambda}_0), \tag{B.14}$$

and testing whether there are other harmonic terms, recomputing $P(\lambda)$ using $C_Y^{(1)}(s)$ and examining its peaks in order of frequency

If α is the overall significance level, the significance level for testing the j-th peak in order of frequency, should be α/j. We continue until no further peaks of $P(\lambda)$ significant.

B.2. *A White Noise Test*

This is not a periodicity peak detector, but it tests if an observed time series can be regarded as a realization of a white noise process. Let us denote frequency in cycles per unit time by ν and let $\nu_j = j/T$. If $\{Z_t, t=1,...,T\}$ are observations of a stochastic process, denote by $F_Z(\nu)$ its spectrum and $I^{(T)}(\nu)$ the periodogram.

If Z_t is white noise, then $f_Z(\nu) = 2\sigma_Z^2$, $0 \leq \nu \leq \frac{1}{2}$, and

$$F_Z(\nu) = \int_0^\nu f_Z(\alpha)\,d\alpha = \begin{cases} 0, & \nu < 0 \\ 2\sigma_Z^2\nu, & 0 \leq \nu < \frac{1}{2} \\ \sigma_Z^2, & \nu \geq \frac{1}{2} \end{cases}$$

$F_Z(\nu)$ is the accumulated spectrum. Since $I^{(T)}(\nu)$ is an estimator of $f_Z(\nu)$, an estimator of $F_Z(\nu_j)$ is $T^{-1} \sum_{i=1}^{j} I_Z(\nu_i)$ and therefore

$$C(\nu_j) = \sum_{i=1}^{j} I_Z(\nu_i)/(T\hat{\sigma}_Z^2) \tag{B.15}$$

is an estimator of $F_Z(\nu_j)/\sigma_Z^2$, where $\hat{\sigma}_Z^2$ is an estimator of the variance of the process. $C(\nu_j)$ is the (normalized) accumulated periodogram. For a white noise process, the graph of $C(\nu_j)$ xν_j will be scattered around the line passing through (0,0) and (0.5,1).

To judge the deviations of $C(\nu_j)$ from this theoretical line a test of significance of the Kolmogorov-Smirnov type is used. See Jenkins and Watts (1968) and Priestley (1981) for details.

REFERENCES

ALMEIDA, F.C., CALHEIROS, R.V., DIAS, P.L.S., XAVIER, T.M.B.S., KANTOR, I.J., KOUSKY, V.E., MEIRA, G.M., MOLION, L.C.B., PARADA, N.J., SRIVATSANGAM, S. and GRAY, W.M. (1980): Contribuição ao estudo da previsão de seca e modificação artificial do tempo e do clima do Nordeste brasileiro. INPE-1812 - RPE/180, São José dos Campos, INPE.

BARTLETT, M.S. (1957): Discussion on "Symposium on spectral approach to time series". J. Roy. Statist. Soc., B, 19, 1-63

BURG, J.P. (1975): Maximum entropy spectral estimates. PhD Dissertation, Stanford University.

CAVIEDES, C.N. (1973): Secas and El Niño: Two simultaneous climatic hazards in South América. Proc. Assoc. Georgr., 5, 44-49.

CONOVER, W.J. (1971): Practical Nonparametric Statistics, New York, Wiley.

FISHER, R.A. (1929): Tests of significance in harmonic analysis, Proc. Roy. Soc., A, 125, 54-59.

GIRARDI, C. and TEIXEIRA, L. (1978): Prognóstico do tempo a longo prazo. Relatório Técnico ECA-06/80, S. José dos Campos, CTA/IAE.

GRENANDER, U. and ROSENBLATT, M. (1957): Statistical analysis of stationary time series. New York, Wiley.

HANNAN, E.J. (1961): Testing for a jump in the spectral function. J. Roy. Statist. Soc., B, 23, 394-404.

JENKINS, G.M. and WATTS, D.G. (1968): Spectral analysis and its applications, San Francisco, Holden-Day.

JONES, R.H. and KEARNS, J.P. (1976): Fortaleza, Cearã, Brazil rainfall. Journal of Applied Meteorology, 15, 307-308.

KANTOR, I.J. (1982): Previsibilidade da sẽrie de precipitação de chuvas de Fortaleza, pelo mẽtodo de mãxima entropia de Burg . INPE-2546-RPE/420, São Josẽ dos Campos, INPE.

MARKHAM, C.G. (1967): Climatological aspects of drought in northeastern Brazil. PhD Dissertation, University of California, Berkeley.

MARKHAM, C.G. (1974): Apparent periodicities in Rainfall at Fortaleza, Cearã, Brazil. Journal of Applied Meteorology, 13, 176-179.

PRIESTLEY, M.B. (1962a): Analysis of stationary processes with mixed spectra-I. J. Roy. Statist. Soc., B, 24, 215-233.

PRIESTLEY, M.B. (1962b): Analysis of stationary processes with mixed spectra-II. J. Roy. Statist. Soc., B, 24, 511-529.

PRIESTLEY, M.B. (1981): Spectral analysis and Time Series, Vol.1. London, Academic Press.

REITER, E. (1979): On the dynamic forcing of short term climate fluctuation by feedback mechanisms. Environmental Research Papers, Colorado State University, Fort Collins, USA.

SHIMSHONI, M. (1971): On Fisher's test of significance in harmonic analysis. Geophys. J. R. Astron. Soc., 23,373-377.

WHITTLE, P. (1952): The simultaneous estimation of a time series harmonic components and covariance structure. Trabajos Estad., 3, 43-57.

POSTSCRIPT

After the paper was completed, we traced A.F. Siegel (1980) Testing for periodicity in a time series, Journal Amer. Statist. Assoc., 75, 345-348 (thanks to a remark by J.K. Ord, during the ITSM in Toronto). We applied the test to our series and, at the 5% significance level, 3.6, 13 and 26 years periodicities were accepted; while, at the 1% level, 13 and 26 years periodicities were detected.

MEMÓRIA DE ESTATÍSTICA

DA

UNIVERSIDADE FEDERAL DE PERNAMBUCO

1 - *Improved Likelihood Ratio Statistics for Generalized Linear Models - Gauss Moutinho Cordeiro - 1984 - 19 páginas.*

2 - Corrélations Entre Paramètres Hydrophobes et Electroniques dans la Serie Benzylidene-2-Furannones-3 - S. L. Galdino, I. R. Pitta, M. A. Campos, A. J. Camarotti et M. N. Ramos - 1984 - 14 páginas.

3 - Estimação de Máxima Verossimilhança pelo Filtro de Kalman- Annibal Parracho Sant'Anna - IM - UFRJ - 1984 - 46 páginas.

4 - Comparison of Spectral Methods for the Construction of Confidence Intervals for Simulation Outputs - Carlos Alberto Barbosa Dantas & Enivaldo Carvalho da Rocha - 1984 - 13 páginas.

5 - Rainfall at Fortaleza in Brazil Revisited - Pedro A. Morettin, Afrânio R. de Mesquita & Jacira Guiro Carvalho da Rocha - 1984 - 31 páginas.