RT-MAE-8303

RAINFALL AT FORTALEZA, CEARÁ, BRASIL REVISITED

by

P.A. Morettin, A.R. de Mesquita
and J.G.C. da Rocha

Palavras Chaves: Rainfall, drought, periodicities,
(Key words) tests.

Classificação AMS: 62M10, 62M15
(AMS classification) 62P99

- Agosto de 1983 -

RAINFALL AT FORTALEZA, CEARÁ, BRAZIL REVISITED

P.A. Morettin, A.R. de Mesquita
and J.G.C. da Rocha

ABSTRACT

The 131 years of annual rainfall data of Fortaleza, Ceará, Brazil is reanalyzed in order to determine if they support the existence of periodicities. The use of several tests (Fisher, Whittle, Hannan, Bartlett and Priestley) seem to suggest that thirteen and twenty-six year periodicities are present in the series. In light of these results, a preliminary model is given for the series.

Key words: Rainfall series, periodical components, tests for periodicities, white noise test.

RAINFALL AT FORTALEZA, CEARÁ, BRAZIL REVISITED

P.A. Morettin and A.R. de Mesquita
University of São Paulo, São Paulo, Brazil
and
J.G.C. da Rocha
Federal University of Pernambuco, Brazil

1. INTRODUCTION

The series of rainfalls at Fortaleza, Ceará, Brazil, has been analysed by several authors recently, its importance lying on the fact that it is probably the longest series available for the study of the severe drought that affects 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 climate of Fortaleza is influenced by the sea and hence is not representative of the rest of the area. The question of representativeness of the Fortaleza series has been discussed by Girardi and Teixeira (1978) who have shown that there is a great similarity between the behavior of the rainfall at Fortaleza and other sites of the region.

Recent studies as in Markham (1974), using "seasonalized" annual totals, 1849-1970, concluded that there were a thirteen and a twenty-six year periodicities in the data, and provided speculations for 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 carefull 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 Exhibit 1. From a visual inspection of both it is not easy to detect noticeable trend and periodical patterns. The sample mean is 1425 mm, the sample variance is 230,526 mm^2, while the minimum and maximum values are 468 mm (1877) and 2512 mm (1974), respectively.

The weather 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 its latitudes 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.

- INSERT EXHIBIT 1 HERE -

The least squares line for the raw data is

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

which indicates a very small positive trend; for example, it gives an increase of 70 mm for the year of 1979, relative to the year of 1849.

Exhibit 2 shows the plot of the autocorrelation function, and the periodogram of the data is given in Exhibit 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.

- INSERT EXHIBITS 2 AND 3 HERE -

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, casts the doubt that the (mean-corrected) rainfall series at Fortaleza, behaves like a white noise series, with a constant spectrum. If the series contains periodicities, its spectrum will have a mixed form, that is, peaks (corresponding to the periodical components) will emerge from a continuous 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 be 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,

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 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 periodical components in a set of data. These tests were not considered by the previons authors who dealt with the rainfall series.

We decided to apply 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 comparison of the asymptotic powers of the Whittle's test, the grouped periodogram test and the $P(\lambda)$ test was given by Priestley (1962 b). 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 obtain $g^{(3)} = 0.0623$, 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 $\lambda_{10} = 2\pi\times 10/131 \simeq \pi/6$), 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 ordinate (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 ordinate, $g_W = 0.056$, 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=5,10,20. The corresponding values of g_k are given in Exhibit 4. For 0.05 significance level, the critical values are given by

$$\frac{0.05\ k}{[T/2]} = k(1-g)^{k-1}$$

and these are shown in Exhibit 4. We see that the calculated values of g_k are smaller than the corresponding critical values for all values of k, leading us to conclude that the mixed spectrum structure is not adequate for the series. This may be explained perhaps by the

the fact that g_k does not assume large values when the periodogram of the series has many peaks.

EXHIBIT 4: Values of g_k and g.

k	g_k	g
5	0.5657	0.8335
10	0.4286	0.5492
20	0.2893	0.3143

(d) Hannan's Test

For the largest ordinate, we obtain 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(θ) as that corresponding to the truncate periodogram.

(e) The P(λ) Test

The calculations were done with m=25 and n=55; other values tried, but we will only describe the results for m=25 and n=55. The graph of P(λ) (Exhibit 5) shows the presence of well defined peaks at 4π/131, 10π/131, 20π/131, 30π/131 and 40π/131.

For λ = 4π/131, we have J_2= 1.225 and for α= 1%, we have α_0= 2.33, and since $J_2<\alpha_0$, we see that the peak at this λ is not significant.

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 $P'(\lambda)$ (Exhibit 6). 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 greather than the critical value $\alpha_0= 2.71$. Hence we accept a periodicite of 13.1 years.

Removing again the contribution of this component we get $P''(\lambda)$ (Exhibit 7), with 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.

- INSERT EXHIBIT 5, 6, 7 HERE -

(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 Exhibit 8. It differs from the one obtained by Jones and Kearns (1976), since in the present case the series cannot be taken as a 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.

- INSERT EXHIBIT 8 HERE -

The results of the application of the several tests are summarized in Exhibit 9. Results seem to indicate that the 13.1 and 26.2 years periodicities are present in the data, since four out of five tests detected them. Therefore, we shall proceed to establish a model for the series.

EXHIBIT 9: Summary of the application of the testes.

Test	Periods or remarks
Ext. Fisher	13.1, 26.2 years
Whittle	13.1, 26,2 years
Bartlett	none
Hannan	13.1, 26.2 years
Priestley	13.1, 26.2 years
White Noise	passed

5. A MODEL FOR THE RAINFALL SERIES.

From the considerations in section 4 we shall entertain 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

$$Z_t = \mu + A_1 \cos(\lambda_1 t + \phi_1) + A_2 \cos(\lambda_2 t + \phi_2), \tag{5.1}$$

and least squares estimates of A_i and ϕ_i, i=1,2 and μ are given by

$$\hat{A}_1 = 202.56,\ \hat{\phi}_1 = 114^\circ$$
$$\hat{A}_2 = 255.64,\ \hat{\phi}_2 = 126^\circ$$
$$\hat{\mu} = \bar{X} = 1424.83,$$

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

$$X_t = 1424.83+202.56\ \cos(0.2398t+1.99)+255.64\ \cos(0.4796t+2.20)+Y_t, \qquad (5.2)$$

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. Exhibit 10 shows the plots of the original and $\hat{Z}_t$ series.

It is readily seen that the sample autocorrelations of the $\hat{Y}_t$ series are all nonsignificant. Also, a white noise test applied to this residual series gives the results shown in Exhibit 11. Hence we may conclude that the Y_t series is white noise, with a constant spectrum.

- INSERT EXHIBITS 10 AND 11 HERE -

6. FINAL REMARKS

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

The evidences given in this paper and by other workers seem 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, so as we gather more observations further analyses could be carried out and eventually other suspected periodicities, 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 for the periodicities of 13.1 and 26.2 years, a much higher order (of about T/2) was necessary.

Questions concerning the forecasting of the series were not dealt with in this paper and will be pursued elsewhere.

REFERENCES

Almeida,F.C. et al. (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. (1955): Stochastic processes. Methods and Applications, The Cambridge University Press.

Burg,J.P. (1975): Maximum entropy spectral estimates. Ph.D. Dissertation, Stanford University.

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, John Wiley & Sons.

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, Ceará, Brazil rainfall. Journal of Applied Meteorology, 15, 307-308.

Kantor,I.J. (1982): Previsibilidade da série de precipitação de chuvas de Fortaleza, pelo método da máxima entropia de Burg. INPE-2546-RPE/420, São José dos Campos, INPE.

Markham,C.G. (1967): Climatological aspects of drought in northeastern Brazil. Ph.D. Dissertation, University of California, Berkeley.

Markham,C.G. (1974): Apparent periodicities in Rainfall at Fortaleza, Ceará, 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.

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.

APPENDIX A: The Rainfall Series, 1849-1979.

Year	Rainfall	Year	Rainfall	Year	Rainfall
1849	2001	1871	1459	1893	1430
1850	852	1872	2256	1894	2505
1851	1806	1873	2058	1895	2491
1852	1356	1874	1487	1896	2144
1853	1233	1875	1581	1897	1839
1854	1590	1876	1569	1898	863
1855	1273	1877	468	1899	2414
1856	1770	1878	503	1900	940
1857	1734	1879	597	1901	1545
1858	1457	1880	1539	1902	878
1859	1357	1881	1423	1903	789
1860	1716	1882	1246	1904	1136
1861	1445	1883	1508	1905	1189
1862	1468	1884	1047	1906	1430
1863	1452	1885	1307	1907	697
1864	1098	1886	1399	1908	834
1865	1238	1887	1320	1909	1015
1866	2478	1888	736	1910	2051
1867	832	1889	784	1911	1373
1868	1289	1890	1534	1912	2446
1869	1470	1891	1077	1913	1905
1870	1628	1892	1211	1914	1512

The Rainfall Series, 1849-1979, Continued.

Year	Rainfall	Year	Rainfall	Year	Rainfall
1915	530	1937	1313	1959	1493
1916	1328	1938	1586	1960	1011
1917	2077	1939	1911	1961	1737
1918	1319	1940	1447	1962	1258
1919	656	1941	916	1963	2102
1920	1847	1942	780	1964	2428
1921	2496	1943	1042	1965	1630
1922	1595	1944	1090	1966	1288
1923	1513	1945	1750	1967	1839
1924	1847	1946	1724	1968	1385
1925	1137	1947	1726	1969	1805
1926	1571	1948	1384	1970	1192
1927	1195	1949	1881	1971	2093
1928	995	1950	1114	1972	1299
1929	1230	1951	747	1973	2331
1930	1107	1952	1378	1974	2512
1931	1133	1953	1068	1975	1778
1932	879	1954	1032	1976	1417
1933	937	1955	1152	1977	1941
1934	1888	1956	806	1978	1752
1935	1661	1957	1225	1979	996
1936	820	1958	504		

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

APPENDIX B

B.1. Tests for Detecting Hidden Periodicities

Given T observations $X_1,\ldots,X_T$ 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 $\lambda_j = 2\pi j/T$, $j=0,1,\ldots,[T/2]$, called Fourier frequencies, and call $I_j^{(T)} = I_X^{(T)}(\lambda_j)$.

(a) Fisher's Test and 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.

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

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

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

(b) Whittle's Test

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

$$g_W = \frac{\max_j [I_j/2\pi\hat{f}_Y(\lambda_j)]}{\sum_{j=1}^{m} [I_j/2\pi\hat{f}_Y(\lambda_j)]} \qquad (B.3)$$

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,

$$\hat{f}_Y(\lambda) = \frac{1}{2\pi} \sum_{s=-(\ell-1)}^{\ell-1} C_X(s) e^{-i\lambda s}, \qquad (B.4)$$

$\ell < T$ being the truncation point.

(c) Hannan's Test

Hannan (1961) considers the problem of testing if a periodical 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

$$f_Y^*(\lambda) = \int_{-\pi}^{\pi} I_X^{(T)}(\alpha)\ W_M\ (\lambda-\alpha)d\alpha; \qquad (B.5)$$

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

$$g_H = \frac{\max_j \left[I_j / 2\pi f_Y^*(\lambda_j) \right]}{\sum_{j=1}^{m} \left[I_j / 2\pi f_Y^*(\lambda_j) \right]} \tag{B.6}$$

has approximately the Fisher distribution with m degrees of freedom.

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

An alternative test, suggested by Bartlett (1955), is used 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 whose widths are less than $2\pi/T$.

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

$$g_k = \frac{I_{j'}/2\pi f_Y(\lambda_{j'})}{\sum_{j=(\ell-1)k+1}^{\ell k} I_j/2\pi f_Y(\lambda_j)}, \quad \ell = 1, \ldots, [T/2K] \frac{1}{k}, \tag{B.7}$$

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

Under H_0, g_k has, asymptotically, the same distribution as the Fisher's g-statistic with k degrees of freedom. If α is the significance level when using Fisher's test to k ordinates, then the approximate significance level for the test based on g_k becomes $\alpha' = \alpha k / [T/2]$.

(e) Priestley's $P(\lambda)$ Test

The preceding tests have a number of disadvantages, which are summarized in Priestley (1981, p.625). The $P(\lambda)$ 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 $\gamma_Z(s) \not\to 0$, 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 periodical component.

The $P(\lambda)$ 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

$$J_q = \left(\frac{T}{m} \Lambda_{n,m}^{-1}\right)^{\frac{1}{2}} \sum_{s=0}^{q} P^*\left(\frac{2\pi s}{m} + \delta\right) \left[\hat{g}(\pi)/2\pi\right]^{-\frac{1}{2}}, \qquad \text{(B.8)}$$

for $q = 0,1,\ldots,[T/2]$ and test whether $\max(J_q) \leq \alpha_0$, where α_0 is the upper 100α% point of the standard normal, for a given significance level α. In (B.8),

$$P^*(\lambda) = P(\lambda)/C_Y(0),$$

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

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

$$\Lambda_{n,m} = 2\pi\int_{-\pi}^{\pi}\{W^{(1)}(\alpha) - W^{(2)}(\alpha)\}^2\, d\alpha. \tag{B.9}$$

In (B. 9), $W^{(1)}(\lambda)$ 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

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

In the formulae for $P^*(\lambda)$ and $\hat{g}(\pi)$, $C_Y(s)$ and $r_Y(s)$ 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 signficant and the amplitude of the harmonic term at λ_0 is then estimated by

$$\hat{A}_0^2 = \frac{8\pi P(\lambda_0)}{n-m}, \quad n > 2m; \tag{B.10}$$

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

$$C_Y^{(1)}(s) = C_Y(s) - \frac{1}{2}\hat{A}_0^2 \cos(s\hat{\lambda}_0), \qquad (B.11)$$

and test 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 $P(\lambda)$ are 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 $\{Z_t, t=1,\ldots,T\}$ 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, them $f_Z(\nu) = 2\sigma_Z^2$, $0\leq\nu\leq\frac{1}{2}$, and

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

$F_Z(\nu)$ is the accumulated spectrum. Since $I^{(T)}(\nu)$ is an estimator of $f_Z(\nu)$, an estimator of $F_Z(\nu_j)$ is $\frac{1}{T}\sum_1^j I_Z(\nu_j)$, and therefore

$$C(\nu_j) = \frac{\sum_{i=1}^{j} I_Z(\nu_j)}{T\,\hat{\sigma}_Z^2} \qquad (B.12)$$

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 $C(\nu_j)\times\nu_j$ 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.

EXHIBIT 1 - THE RAINFALL SERIES, 1849-1979

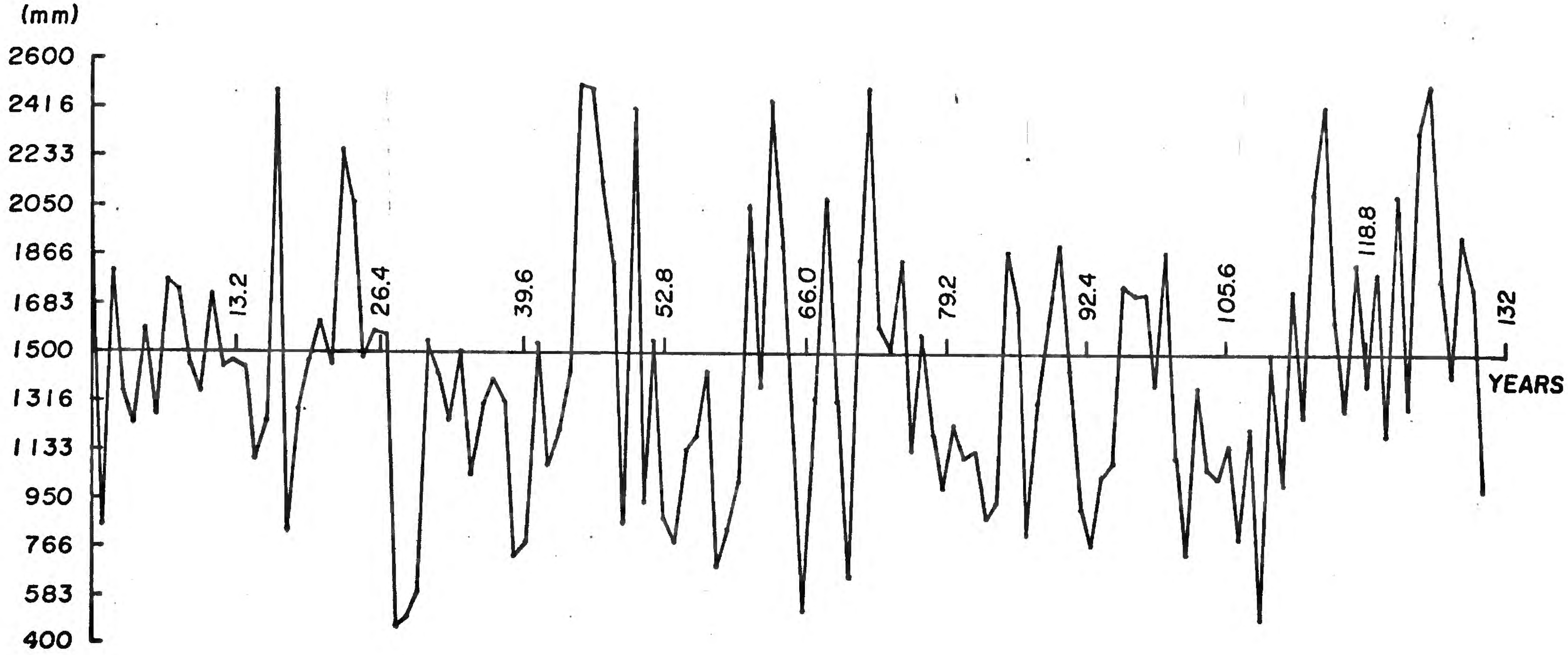


EXHIBIT 2 - AUTOCORRELATION FUNCTION OF THE RAINFALL SERIES

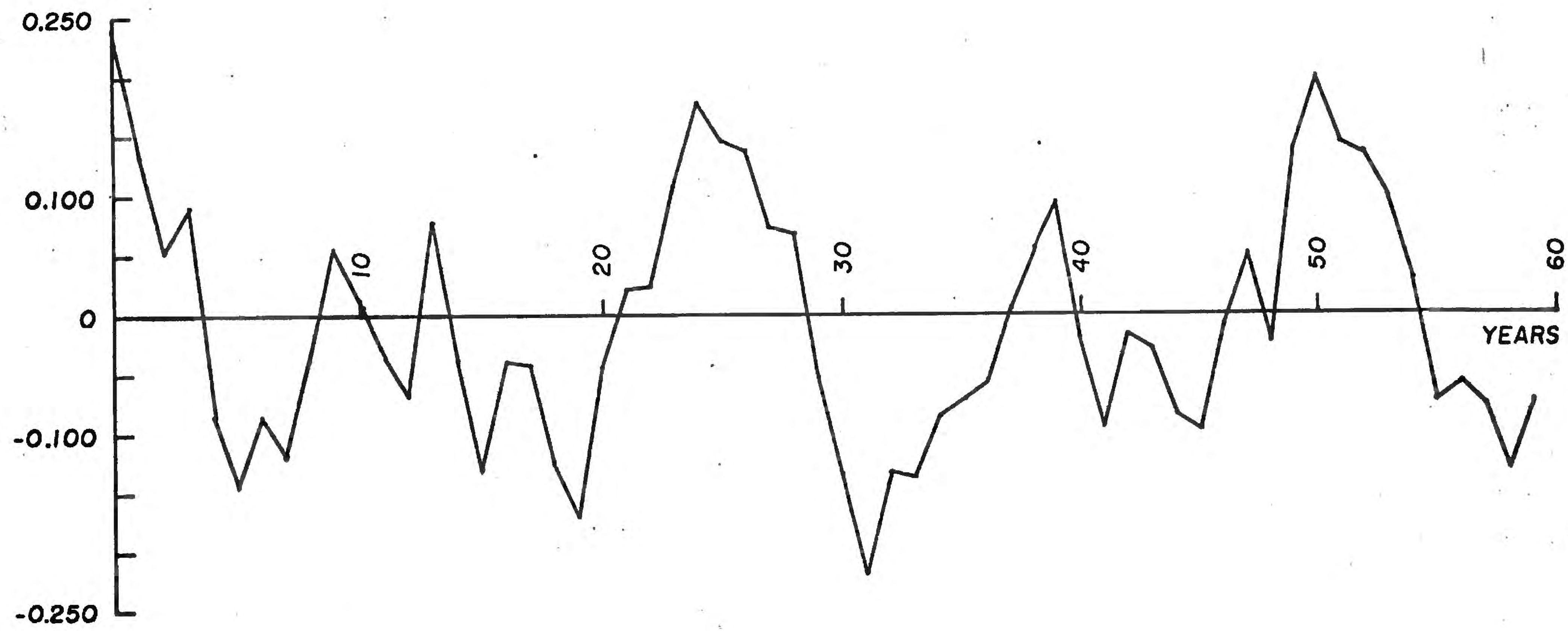


EXHIBIT 3 - PERIODOGRAM OF THE RAINFALL SERIES

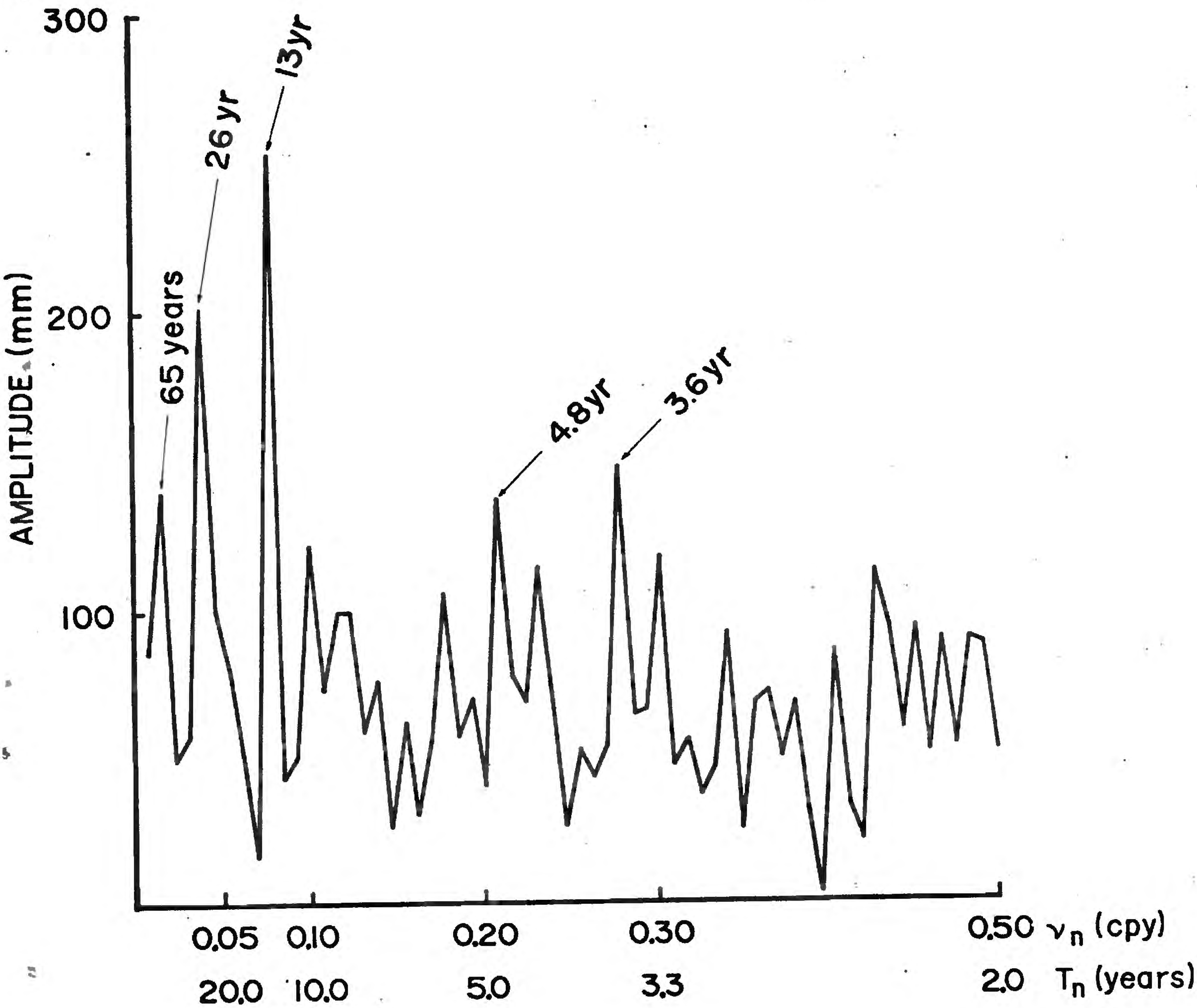


EXHIBIT 5 - GRAPH OF P(λ) FOR THE RAINFALL SERIES

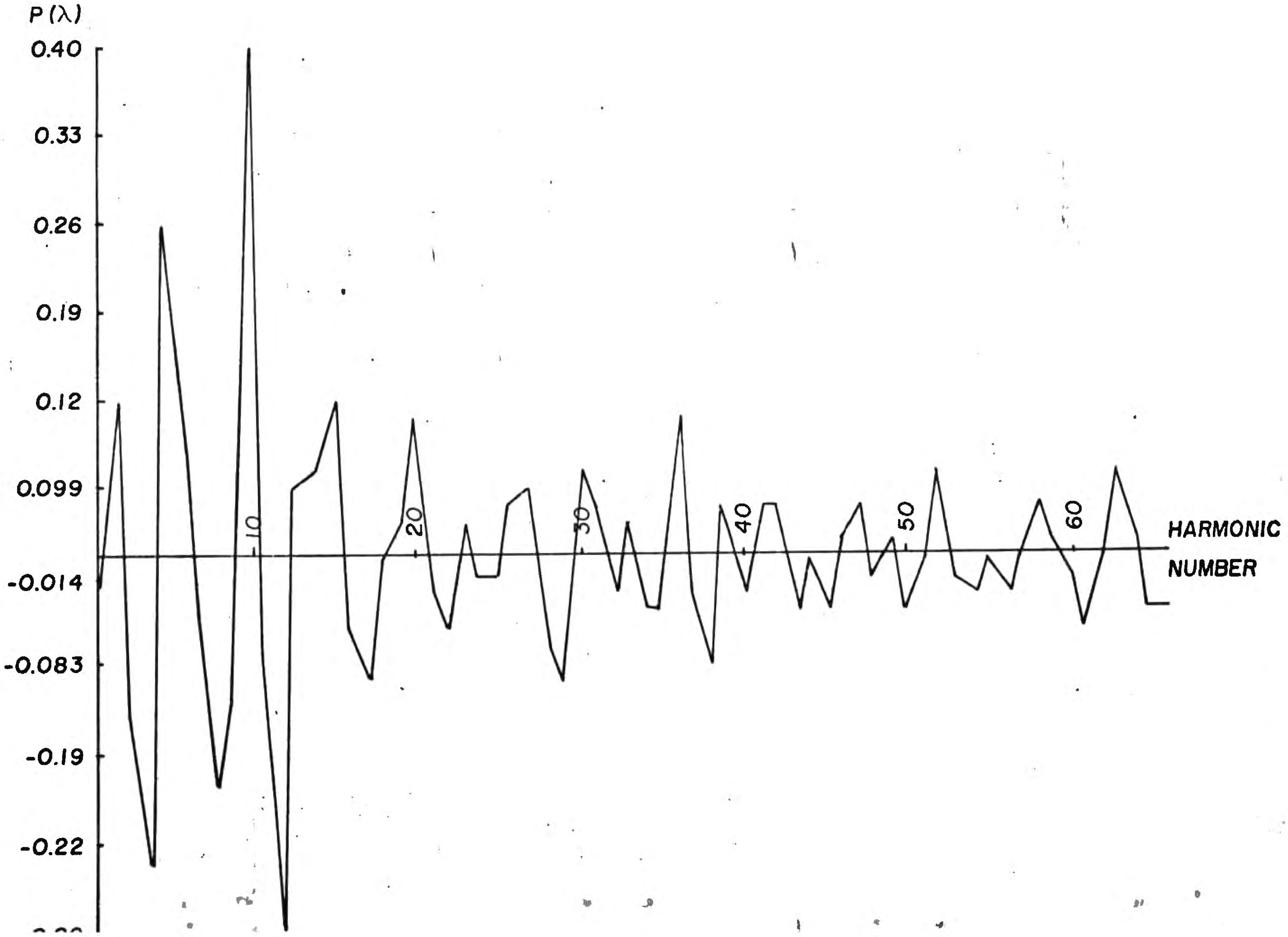


EXHIBIT 6 - GRAPH OF P'(λ) FOR THE RAINFALL SERIES

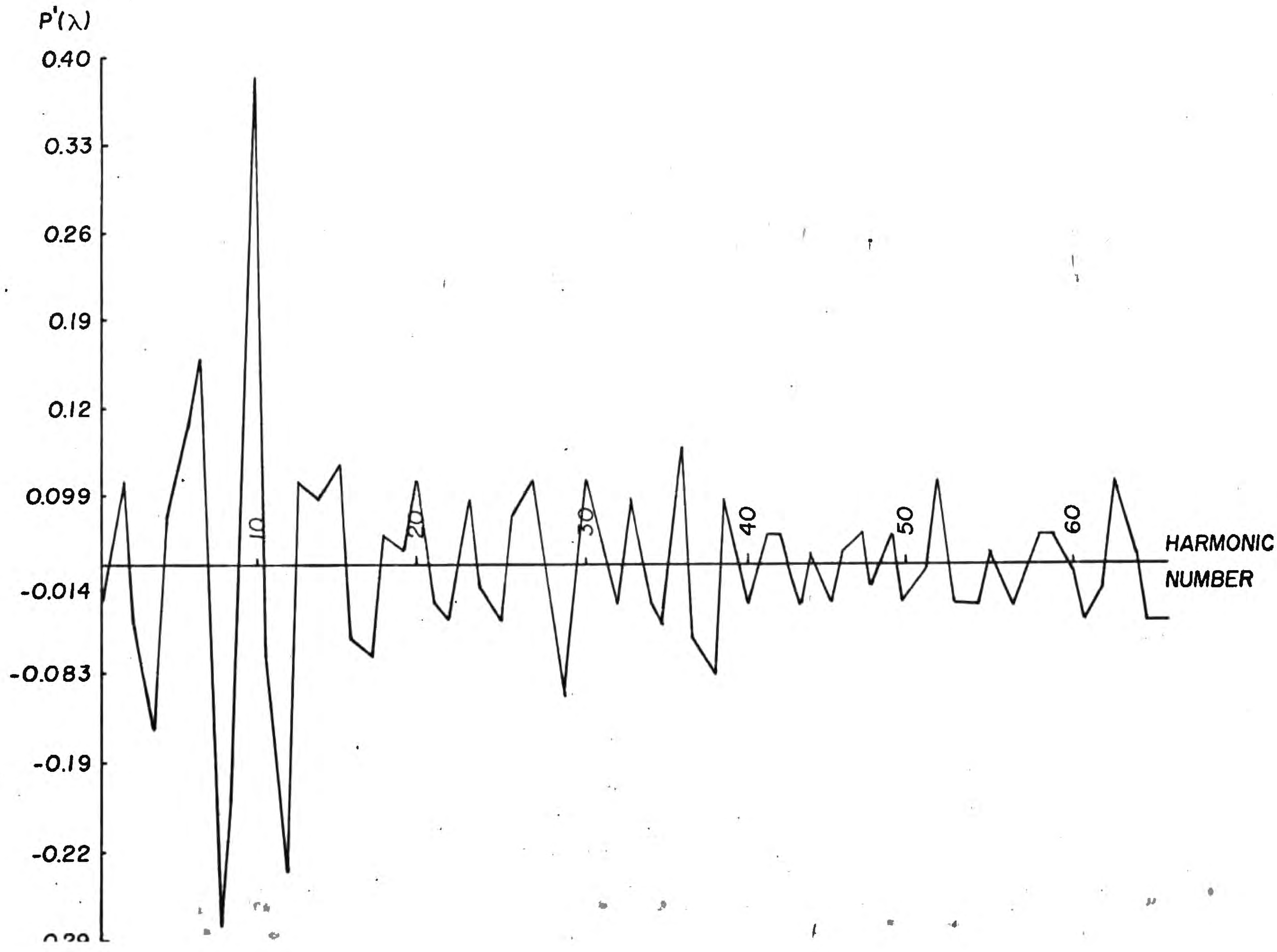


EXHIBIT 7 - GRAPH OF P''(λ) FOR THE RAINFALL SERIES

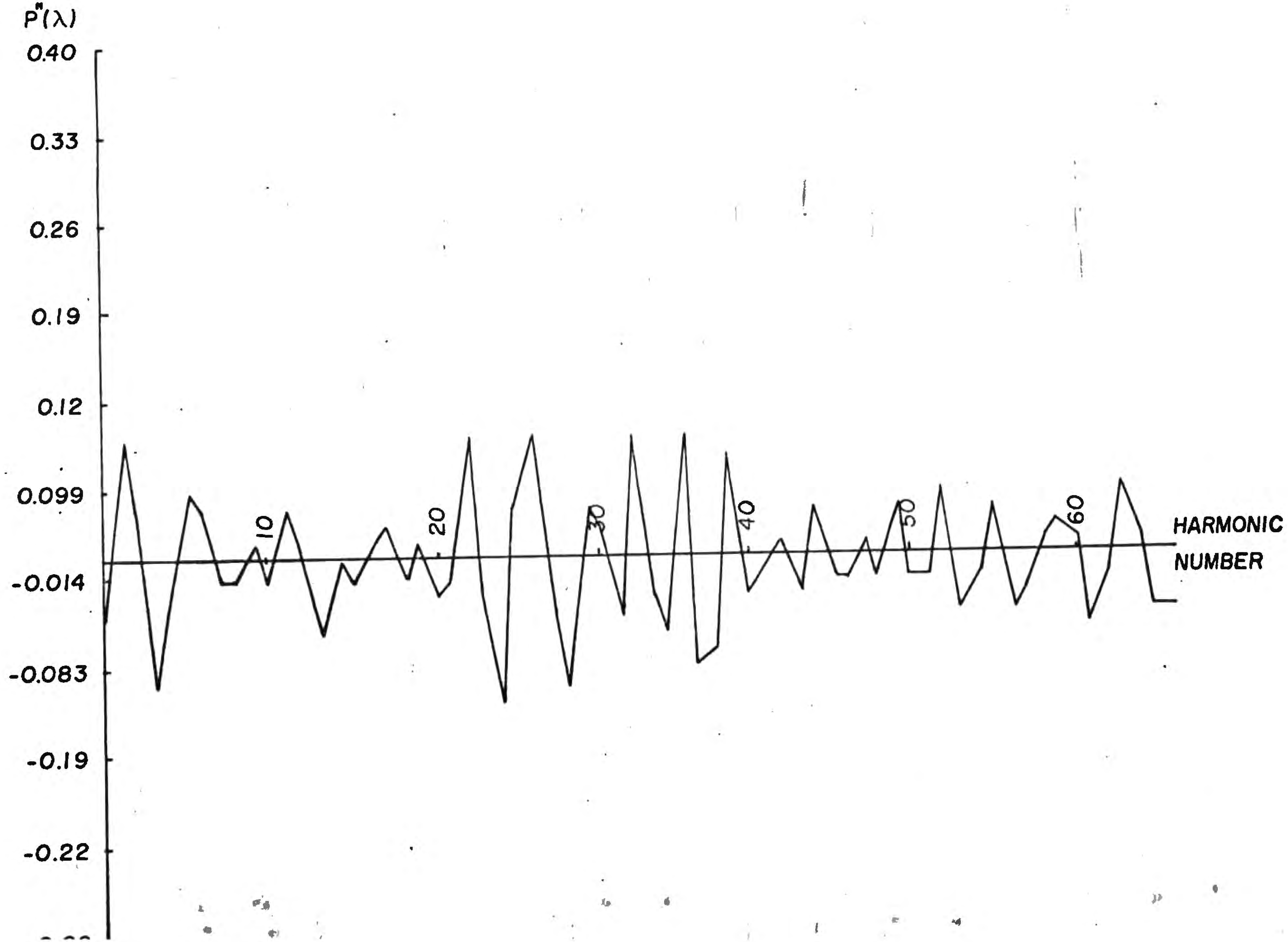


EXHIBIT 8 - NORMALIZED ACCUMULATED PERIODOGRAM FOR THE RAINFALL SERIES, WITH THEORETICAL AND CONFIDENCE LINES

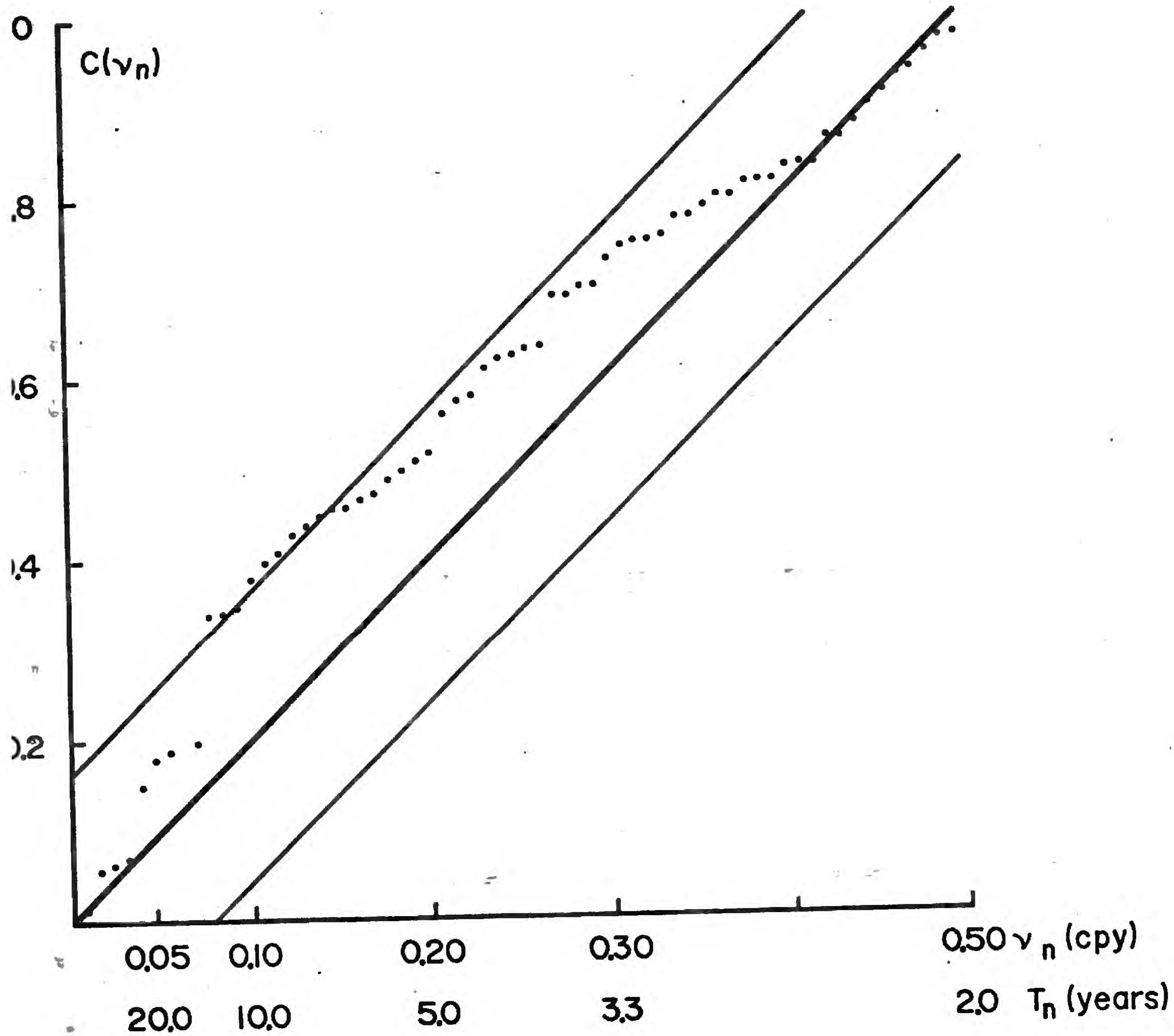


EXHIBIT 10 - ORIGINAL SERIES (SOLID CURVE) AND FITTED Z_t (DOTTED CURVE)

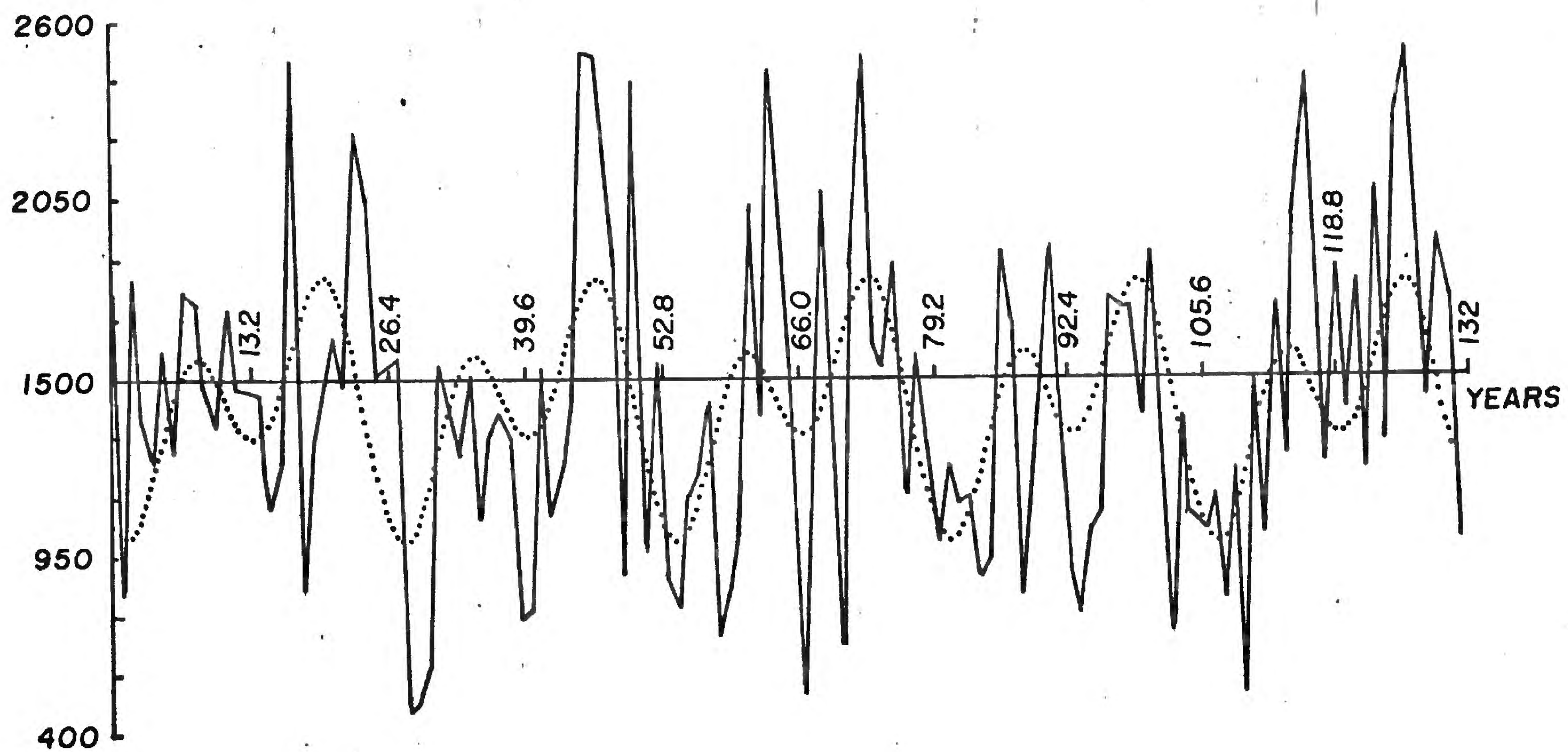
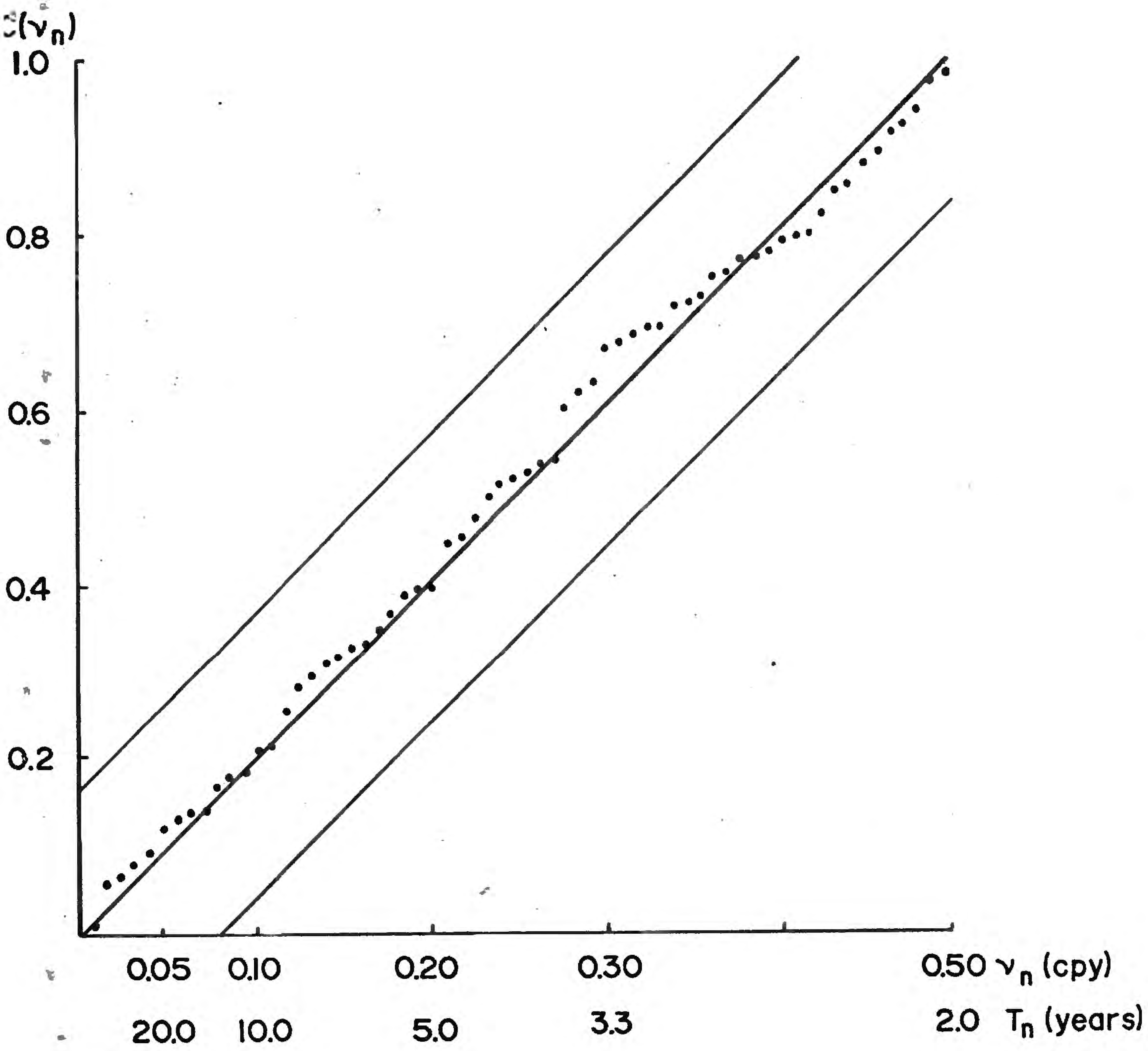


EXHIBIT 11 - NORMALIZED ACCUMULATED PERIODOGRAM FOR THE RESIDUAL SERIES, WITH THEORETICAL AND CONFIDENCE LINES



RELATÓRIO TÉCNICO

DO

DEPARTAMENTO DE ESTATÍSTICA

TÍTULOS PUBLICADOS

7901 - BORGES, W. de S. On the limiting distributios of the failure time of composite material. São Paulo, IME-USP, 1979, 22p.

7902 - GALVES, A.; LEITE, J.G.; ROUSSIGNOL, M. The invariance principle for the one-dimensional symmetric simple exclusion process. São Paulo, IME-USP, 1979. 9p.

8001 - MENTZ, R.P. et al. Exploratory fitting of autoregressive and moving average models to well-behaved time series data. São Paulo, IME-USP, 1980. 16p.

8002 - MORETTIN, P.A. Walsh spectral analysis. São Paulo, IME-USP, 1980. 27p.

8003 - RODRIGUES, J. Robust estimation and finite population. São Paulo, IME-USP, 1980. 13p.

8004 - BORGES, W. de S. & RODRIGUES, F.W. On the axiomatic theory of multistate coherent structures. São Paulo, IME-USP, 1980, 10p.

8005 - MORETTIN, P.A. A central limit theorem for stationary processes. São Paulo, IME-USP, 1980. 5p.

8101 - DANTAS, C.A.B. & COLUCCI, E. A Simulation program for emergency services-II, São Paulo, IME-USP, 1981, 14p.

8102 - ANDJEL, E.D. Invariant measures for the zero range process. São Paulo, IME-USP, 1981, 55p.

8103 - ANDJEL, E.D. The asymmetric simple exclusion process on Z^d. São Paulo, IME-USP, 1981, 13p.

8104 - MORETTIN, P.A. & TOLOI, C.M.C., Accuracy of forecasting with special reference to the Box-Jenkins and Bayesian Methodologies. São Paulo, IME-USP, 1981, 41p.

8105 - PINO, F.A. & MORETTIN, P.A., Intervention analysis applied to Brazilian coffee and milk times series. São Paulo IME-USP, 1981, 36p.

8106 - BORGES, W.S. & RODRIGUES, J., Testing for new better than used in expectation. São Paulo, IME-USP, 1981, 7p.

8107 - FAHMY, S.; PEREIRA, C.A.B.; PROSCHAN, F., The influence of the sample on the posterior distribution. São Paulo, IME-USP, 1981, 17p.

8108 - PERES, C.A., Asymptotic efficiency of the likelihood ratio conditional test for multinomial distributions. São Paulo IME-USP, 1981, 29p.

8109 - PERES, C.A., Testing the effect of blocking in a randomized complete block design (RCBD). São Paulo, IME-USP, 1981, 14p.

8110 - BASU, D. & PEREIRA, C.A.B., On the Bayesian analysis of categorical data: the problem of nonresponse. São Paulo, IME-USP, 1981, 13p.

8201 - BASU, D. & Pereira, C.A.B., Conditional independence in statistics. São Paulo, IME-USP, 1982, 37p.

8202 - BASU, D. & PEREIRA, C.A.B., A note on Blackwell sufficiency and a Skibinsky characterization of distributions. São Paulo, IME-USP, 1982, 12p.

8203 - PERES, C.A., On the interpretation of the parameters of the quadratic model for cell survival ofter irradiation. São Paulo, IME-USP, 1982., 22p.

8204 - GALVES, A., et al. Rescaling the stirring process. São Paulo IME-USP, 1982, 23p.

8205 - RODRIGUES, J., On the asymptotic theory for the fixed size confidence ellipsoids. São Paulo, IME-USP, 1982, 14p.

8206 - PEREIRA, C.A.B. & RODRIGUES, J., Robust linear prediction in finite populations. São Paulo, IME-USP, 1982, 14p.

8207 - MORETTIN, P.A., Walsh-Fourier transforms. São Paulo, IME-USP 1982, 15p.

8208 - PERES, C.A. & MORETTIN, P.A., Building bridges between the academic and real worlds - some observations from South America. São Paulo, IME-USP, 1982, 16p.

8209 - PEREIRA, C.A.B. & ROGATKO, A., The Hardy-Weinberg equilibrium under a Bayesian perspective. São Paulo, IME-USP, 1982, 16p.

8210 - MORETTIN, P.A., The Levinson algorithm and its applications in time series analysis. São Paulo, IME-USP, 1982, 16p.

8211 - RODRIGUES, J., A Note on Maximized and Conditional Likelihood Functions. São Paulo, IME-USP, 1982, 9p.

8301 - PEREIRA, C.A.B., Stopping rules and conditional inference in 2x2 contingence tables. São Paulo, IME-USP, 1983, 7p.

8302 - BOLFARINE, H., PEREIRA, C.A.B. & RODRIGUES, J., Robust Linear Prediction in Finite Populations: A Bayesian Perspective. São Paulo, IME-USP, 1983, 21p.