Two Reports on Trend Analysis:*

a) An Elementary Trend Analysis of Rio Negro Levels at Manaus, 1903-1985

b) Consistent Detection of a Monotonic Trend Superposed on a Stationary Time Series

David R. Brillinger

Statistics Department
University of California

Technical Report No. 141
January 1988

*Research supported by National
Science Foundation Grant DMS-8316634.

Department of Statistics
University of California
Berkeley, California

An Elementary Trend Analysis of Rio Negro Levels at Manaus, 1903-1985

David R. Brillinger

Statistics Department
University of California, Berkeley

ABSTRACT

There is concern that flooding of the Amazon River has been aggravated by deforestation in the Amazon Basin. In the paper this issue is naively formalized to one of obtaining an estimate $\hat{\beta}$ of β in the model

$$Y(t) = \alpha + \beta t + E(t)$$

(with E a zero mean stationary mixing noise) and then of examining the hypothesis that $\beta = 0$. A crucial part of the problem is that of obtaining a suitable uncertainty estimate of the sampling error of $\hat{\beta}$. The approach adopted here is semiparametric in character. The parameter β is estimated by least squares and the standard error is estimated nonparametrically. Some evidence is found for a linear trend in the daily stage values but not in the vector annual series of minimum, maximum, square root of exceedance area, proportion of exceedance days. Further cross-spectral analysis of the monthly mean series with monthly mean sunspots is carried out as is a bispectral analysis of seasonally adjusted daily values. No evidence is found of a linear time invariant relationship with sunspots, but there is some evidence of nonlinearity and nonGaussianity.

1. INTRODUCTION

Quite a number of scientific problems of contemporary interest come down to a question of whether or not some temporal trend is present in a time series of concern. This work presents an elementary addressing of the question in the case of Amazon River water height recorded at one location, Manaus. Daily river height values have been recorded there for the Rio Negro since 1903. It seems of some interest to examine this data for trend, given the many developments that have taken place in the Amazon Basin during this century.

2. DISPLAYING THE DATA

Daily stage, that is height, readings have been made since 1903 at Manaus, 18 km up the Rio Negro estuary, from the Amazon River. Because of the minimal slope of the water surface, the Manaus values are felt to provide a reasonable approximation to the Amazon's values in the region, see Sternberg (1987). The readings have been made available to Professor. O'R. Sternberg of my University by Manaus Harbour Ltd., Portobras. In all there are 30529 readings continuing through 31 July 1986.

These data may be graphed in several ways, having in mind that one is concerned with whether or not some trend is present. Figure 1 is a graph of the annual mean stage. Figure 2 is a perspective plot of monthly mean values by month and year. Figure 3 provides graphs of the annual "shape" and of the deviations from that shape for the 83 complete years available. These graphs each bring out distinct aspects of the data. Figure 1, being the annual mean, has smoothed out the annual variation. If there were a strong trend present, it should show itself in this plot. In 1926 there were some dramatically low values. They possibly relate to a great forest fire, see Sternberg (1987). In the early 70's the series was high, but since then it seems to have returned to a more usual level. Figures 2 and 3 illustrate the strong seasonal effect present in the river fluctuations. The details of the two graphs of Figure 3 are: the top graph is the 50% trimmed mean of the annual curves. Specifically, if $X_k(l)$ denotes the value on the l-th day of the k-th year, then the 50% trimmed mean is defined as

$$\hat{\theta}(l) = \underset{k}{mean}\{X_k(l) : \| X_k - \hat{\theta} \| \le d\} \tag{2.1}$$

where d is the median of the distances $\| X_k - \hat{\theta} \| = (\sum |X_k(l) - \hat{\theta}(l)|^2)^{1/2}$. It is computed iteratively. The properties of this statistic are developed in Folledo (1983). The very deviant curve in Figure 3 is for the year, 1926, of the forest fire.

3. FREQUENCY-SIDE ANALYSIS

3.1 Univariate analysis

Fourier techniques often lead to elementary analyses of time series measurements. The discrete Fourier transform of the time series stretch $Y(t)$, $t=0,...,T-1$ is the collection of values

$$y_j = \sum_{t=0}^{T-1} Y(t) \exp\{-2\pi i j t/T\} \tag{3.1}$$

for $j = 0,...,T-1$. The top graph of Figure 4 is a plot of the log of the periodogram, $|y_j|^2/2\pi T$, for $Y(t)$ the seasonally adjusted Manaus series, with seasonal adjustment consisting of subtracting out the 50% trimmed mean of Figure 3. The flatness of the graph at low frequencies is noteworthy . There is no substantial peak at the annual frequency, suggesting that the seasonal adjustment has been effective. (The Fourier transform here was computed via a fast Fourier transform algorithm for the last 30420 values of the series.)

Turning now to the central problem of the paper, consider the model

$$Y(t) = S(t,\theta) + E(t) \tag{3.2}$$

with S a trend term, with θ an unknown parameter and with E a zero mean noise series. Let $s_j(\theta)$ and ε_j denote the Fourier transform values for S and E respectively computed as in expression (3.1). Then equation (3.2) yields

$$y_j = s_j(\theta) + \varepsilon_j \tag{3.3}$$

for $j = 0,...,T-1$. This alternate form is often a great practical simplification. In the case that the noise process E is stationary and mixing (in various senses), the ε_j satisfy a central limit theorem. Specifically, as $T \to \infty$ and $2\pi j/T \to \lambda$, the variate ε_j may be approximated by a (complex) normal variate with mean 0 and variance $2\pi T f_{EE}(\lambda)$. (Here f_{EE} is the power spectrum of the series E. Its specific definition is given in the Appendix.) Further, J distinct ε_j values, with each $2\pi j/T$ near λ, are asymptotically independent. The dramatic practical implication of this result is that one may be able to act as if one is dealing with an additive (nonlinear) regression model with Gaussian errors. (Details of specific assumptions involved and results may be found in Brillinger (1983). This approach was pioneered, for linear systems, in Akaike (1965).)

Now consider specifically the case of a series with a linear trend, $S(t,\theta) = \alpha + \beta t$, where $\theta = (\alpha,\beta)$. The issue of whether a trend is present is the question: is $\beta = 0$? To develop this idea further write

$$\Delta_j = \sum_{t=0}^{T-1} \exp\{-2\pi i j t/T\} \quad \text{and} \quad \Gamma_j = \sum_{t=0}^{T-1} t \exp\{-2\pi i j t/T\} \tag{3.4}$$

These are the Fourier transforms of the constant and linear terms of the model respectively. These quantities may be evaluated by elementary algebra to find that $\Delta_0 = T$ and $\Gamma_0 = T(T-1)/2$, while for $j \ne 0$ one has $\Delta_j = 0$ and $\Gamma_j = T/(exp\{-2\pi i j/T\}-1)$. The bottom graph of Figure 4 is a plot of $|\Gamma_j|$ on the same scale as the periodogram in the upper graph. The bulk of the mass of $|\Gamma_j|$ is seen to lie at the lower frequencies. Elementary examination of the two graphs suggests that the likelihood of a strong linear trend in the Manaus series is small, there being insufficient fall-off at the low frequencies. A formal development follows.

For the linear model, one has the regression relationships

$$y_0 = \alpha\Delta_0 + \beta\Gamma_0 + \varepsilon_0 \tag{3.5}$$
$$y_j = \beta\Gamma_j + \varepsilon_j$$

for $j = 1,\ldots,T-1$ with the ε_j approximately independent normal variates with mean 0 and variance $2\pi T f_{EE}(2\pi j/T)$. It will be assumed henceforth that $f_{EE}(0) \neq 0$. The least squares estimate of β for the model (3.5) is

$$\hat{\beta} = \mathrm{Re}\{\sum_{j=1}^{J} y_j \overline{\Gamma}_j / \sum_{j=1}^{J} |\Gamma_j|^2\} \tag{3.6}$$

This variate is distributed asymptotically as a normal variate with mean β and variance $\pi T f_{EE}(0)/\sum|\Gamma_j|^2$ for $J < T/2$. Because of the rapid fall-off in magnitude of the Γ_j, the estimate $\hat{\beta}$ will generally be close to the ordinary least squares estimate of β, that is to

$$\sum_{j=1}^{T-1} y_j \overline{\Gamma}_j / \sum_{j=1}^{T-1} |\Gamma_j|^2 = \sum_{t=0}^{T-1} Y(t)(t-m) / \sum_{t=0}^{T-1} (t-m)^2$$

with $m = (T-1)/2$.

The variance, $\sigma^2 = 2\pi T f_{EE}(0)$ may be estimated via the residual sum of squares, specifically by

$$\hat{\sigma}^2 = \frac{2}{2J-1} \sum_{j=1}^{J} |y_j - \hat{\beta}\Gamma_j|^2 \tag{3.7}$$

The distribution of this variate will be approximately that of $\sigma^2\chi^2_{2J-1}/(2J-1)$, as is usual in the case of regression with Gaussian errors.

Now one can get approximate P-values for the hypothesis that $\beta = 0$ via the test statistic

$$\hat{\beta}\sqrt{2\sum|\Gamma_j|^2}/\hat{\sigma} \tag{3.8}$$

This statistic has approximately a t-distribution with $2J-1$ degrees of freedom.

As remarked above, the estimate computed is basically ordinary least squares. One has a concern that, because of the autocorrelated nature of the error series, it may not be efficient. In this connection it is to be noted that Grenander (1954) has shown that among linear unbiased estimates, the ordinary least squares estimate is asymptotically efficient. Asymptotically, there is no need for weights.

The analysis described refers to a linear trend. It is clear that polynomial trends may be handled in the same fashion. Further, in some cases it may be well worth employing some of the traditional time series "touch-up" techniques, such as prewhitening and tapering. The effects of these are to make the analysis more sensitive. Tapering was essential in the bispectral analysis reported below.

3.2 Vector-variate analysis

Taking a single characteristic value for each year, eg. yearly maximum is one way to "eliminate" the seasonal. Continuing to introduce a second approach, other characteristics than the daily stage values may be better measures of flooding, and perhaps such should be combined into a multivariate analysis. Vector extensions of the results of the previous section are therefore needed. In this connection, in the following, vector and matrix-valued quantities will be written in bold-faced type.

Let $\mathbf{Y}(t)$ denote a row-valued time series with r rows. The model of consideration is now

$$\mathbf{Y}(t) = \boldsymbol{\alpha} + \boldsymbol{\beta} t + \mathbf{E}(t) \tag{3.9}$$

The estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \mathrm{Re}\{ \sum_{j=1}^{J} \mathbf{y}_j \overline{\Gamma}_j / \sum_{j=1}^{J} |\Gamma_j|^2 \} \tag{3.10}$$

with $\mathbf{y}_j$ denoting the empirical Fourier transform of the vector series. This estimate will have large-sample covariance matrix $\frac{1}{2}\Sigma/\sum|\Gamma_j|^2$ where $\Sigma = 2\pi T \mathbf{f}_{EE}(0)$ with $\mathbf{f}_{EE}$ the spectral density matrix of the error series.

The error covariance matrix may be estimated by

$$\hat{\Sigma} = \mathrm{Re}\{ \frac{2}{2J-r} \sum_{j=1}^{J} \overline{(\mathbf{y}_j - \hat{\boldsymbol{\beta}}\Gamma_j)}^{\tau} (\mathbf{y}_j - \hat{\boldsymbol{\beta}}\Gamma_j) \} \tag{3.11}$$

with τ denoting matrix transpose. This statistic is distributed as Wishart, $W_r(2J-r, \Sigma)/(2J-r)$. As a test statistic of the hypothesis $\boldsymbol{\beta} = 0$ one can now consider the quantity

$$T^2 = (\hat{\boldsymbol{\beta}}\hat{\Sigma}^{-1}\hat{\boldsymbol{\beta}}^{\tau})(2 \sum|\Gamma_j|^2) \tag{3.12}$$

which is such that $(2J-2r+1)T^2/(r(2J-r))$ is distributed asymptotically as $F_{r,2J-2r+1}$. These last are standard multivariate results for normal variates. (See for example Rao (1973), Section 8b.)

4. RESULTS

The Fourier procedures described in Section 3 were implemented with the Rio Negro data presented in Section 2.

4.1 Univariate results

The seasonally adjusted series of daily water stages at Manaus is analyzed as in Section 3.1 . The number of Fourier values employed in the estimate (3.6) is $J = 25$. This number was chosen so that the t-distribution was reasonable stable and yet so that one stayed far below the seasonal frequency. The value obtained for the estimated slope, (3.6), is .0000251 . The value obtained for the t-statistic, (3.8), is 1.686 . with 49 degrees of freedom. The P-value, for examining the alternate hypothesis, $\beta > 0$, is .049 . There is some evidence for a positive linear trend.

4.2 Vector results

In an attempt to broaden the analysis, a 4-vector series was analyzed. Its components were the annual minimum, the annual maximum, the square root of the annual area of the series above the stage level of 26m and the annual proportion of days about the 26m level. These variates were chosen as possibly providing better measures of flooding and other aspects of river behaviour. For the analysis, the annual series were extended to 84 years. (To get estimates of the last 5 months of 1986, the first 7 months data were regressed on the corresponding 50% trimmed mean curve values.)

The value obtained for T^2 of (3.12) was 4.885 . The corresponding F-statistic was .960 with degrees of freedom 4 and 11, ($J = 9$). The P-value obtained was .467 . In this case there is no evidence that $\boldsymbol{\beta} \neq 0$. It is to be noted that the daily analysis was based on 30529 observations while this annual analysis was based on but 84. It may be expected to be considerably less sensitive.

5. DISCUSSION

The topic of the paper has been an "elementary" trend analysis. There are a multitude of other analyses that could have been carried out. A few will be mentioned. Other definitions and computations of a trend component are possible. More subtle seasonal adjustment procedures might have been employed, specifically procedures involving fitting a trend and seasonal simultaneously. Robust/resistant techniques could have been employed. Models with random effects might have been employed, so too might parametric models.

It may be remarked that the first analysis above was based on daily data. It was possible to have analyzed monthly or annual means instead. An advantage of employing the daily series is the large number of degrees of freedom at low frequencies available for estimating $f_{EE}(0)$.

6. TWO FURTHER ANALYSES

Having come this far with this data set it seems worth while to report the results of two further analyses. The first is a cross-spectral analysis of the monthly means with the series of mean monthly sunspot numbers. The second is a bispectral analysis of the daily stages.

6.1 Coherence with sunspots

There has long been a concern that sunspots affect the Earth's climate. (See for example Pittock (1978).) A cross-spectral analysis was carried out with the Rio Negro data to examine this possibility. Monthly mean values were employed. These values were graphed in Figure 2 for the Rio Negro. The values were seasonally adjusted by removing overall monthly means. Monthly mean relative sunspot numbers were found in Waldmeir (1961) for the years up to 1960 and thereafter in the International Astronomical Union Quarterly Bulletin on Solar Activity.

Figure 5 provides an estimate of the coherence between these two series. It is based on 996 monthly values and has 40 degrees of freedom. The horizontal line in the graph is the level that has probability 5% of being exceeded in the case of no relationship. (The level is computed as in Exercise 8.16.22 in Brillinger (1975).) The level is exceeded 5 times out of 101. This analysis has produced no evidence for a linear time invariant relationship between the Rio Negro values and the sunspot series.

6.2
Bispectral analysis

Bispectral analysis is of use in discerning nonGaussianity of a time series and in examining it for nonlinearity.

The bicoherence at bifrequency (λ,μ) for the stationary series Y is defined as

$$\frac{|f_{YYY}(\lambda,\mu)|^2}{f_{YY}(\lambda)f_{YY}(\mu)f_{YY}(\lambda+\mu)} \qquad (6.1)$$

where f_{YYY} is the series' bispectrum. (The bispectrum is defined in the Appendix.) Among the properties of the bicoherence are that it vanishes if Y is a Gaussian series and that it is constant in the case that Y is a linear process, that is in the case that $Y(t)=\sum a(t-u)\Psi(u)$ with Ψ white noise. (These properties are developed in Brillinger (1965).)

The estimation of the bispectrum and bicoherence is discussed in the Appendix. For the seasonally Manaus daily data, the series was split into $L = 119$ disjoint segments of length $V = 256$. A cosine taper was applied to each segment, (because of the fall-off of the spectrum and the short segment length).

Figure 6 provides contour and perspective plots of the bicoherence estimate. There is evidence of nongaussianity and nonlinearity, specifically there are elevated values near (0,0) and (33.,33.). (The elevated values in the latter case may be due in part to bias remaining in the estimate.) A formal test could have been applied here, see eg. Hinich (1982), but the work presented here is simply a pilot investigation.

ACKNOWLEDGEMENTS

The author thanks Carolina and Hilgard Stemberg for several enjoyable discussions. The work was partially supported by the National Science Foundation Grant MCS-8316634.

REFERENCES

AKAIKE, H. (1965). On the statistical estimation of the frequency response function of a system having multiple input. Ann. Inst. Statist. Math. 17, 185-210.

BRILLINGER, D. R. (1965). An introduction to polyspectra. Ann. Math. Statist. 36, 1351-1374.

BRILLINGER, D. R. (1975). Time Series: Data Analysis and Theory. Holt, Rinehart and Winston, New York.

BRILLINGER, D. R. (1983). The finite Fourier transform of a stationary process. pp. 21-37 in Handbook of Statistics, Vol. 3 (eds. D. R. Brillinger and P. R. Krishnaiah). Elsevier, Amsterdam.

FOLLEDO, M. (1983). Robust/resistant Methods in the Estimation of the Evoked Response Curve. Ph.D. Thesis, University of California, Berkeley.

GRENANDER, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. Ann. Math. Statist. 25, 252-272.

HINICH, M. (1982). Testing for Gaussianity and linearity of a stationary time series. J. Time Series Anal. 3, 169-176.

PITTOCK, A. B. (1978). A critical look at long-term sun-weather relationships. Rev. Geophysics Space Physics 16, 400-420.

RAO, C. R. (1973). Linear Statistical Inference and Its Applications, Second Edition. J. Wiley, New York.

STERNBERG, H. O'R. (1987). Aggravation of floods in the Amazon River as a consequence of deforestation? Geografiska Annaler 69A, 201-219.

WALDMEIR, M. (1961). The Sunspot Activity in the Years 1610-1960. Schulthess, Zurich.

APPENDIX

Specific expressions are set down here for pertinent parameters and their estimates.

A.1 Parameters

Let $Y(t)$, $t=0,\pm1,\pm2,\ldots$ denote a stationary time series. Let it have mean c_Y, covariance function $c_{YY}(u) = cov\{Y(t+u),Y(t)\}$ and third moment function $c_{YYY}(u,v) = E\{[Y(t+u) - c_Y][Y(t+v) - c_Y][Y(t) - c_Y]\}$. The power spectrum of Y at frequency λ is defined by

$$f_{YY}(\lambda) = \frac{1}{2\pi}\sum c_{YY}(u)e^{-iu\lambda}$$

The bispectrum at bifrequency (λ,μ) is defined by

$$f_{YYY}(\lambda,\mu) = \frac{1}{(2\pi)^2}\sum\sum c_{YYY}(u,v)e^{-i(u\lambda+v\mu)}$$

The fundamental domain of bifrequencies for the bispectrum is $0 \le \mu \le \lambda \le \pi$ and $0 \le 2\lambda+\mu \le 2\pi$. This is the region of Figure 6.

A.2 Estimates

There are a variety of fashions by which a bispectrum may be estimated. In this paper the technique of segmenting and averaging is employed because of the ease with which it may be programmed. Specifically let the data be broken into L stretches of length V, so $T = LV$. Next compute the tapered Fourier transform of the l-th stretch,

$$d_Y^V(\lambda;l) = \sum_{v=0}^{V-1} h\left(\frac{v+1}{V+1}\right)Y(lV+v)e^{-iv\lambda}$$

for $l = 0,...,L-1$. (The taper employed in the work of this paper is Tukey's namely $h(u) = (1 - \cos 2\pi u)/2$.) Next form the third order periodogram of the l-th stretch

$$I_{YYY}^V(\lambda,\mu;l) = \frac{1}{(2\pi)^2 h_3} d_Y^V(\lambda;l)d_Y^V(\mu;l)\overline{d_Y^V(\lambda+\mu;l)}$$

where $h_3 = \sum h((v+1)/(V+1))^3$. The estimate of the bispectrum is now

$$f_{YYY}^T(\lambda,\mu) = \frac{1}{L}\sum_{l=0}^{L-1} I_{YYY}^V(\lambda,\mu;l)$$

In forming the estimate of the bicoherence, as defined by expression (6.1), the power spectrum is estimated by similarly averaging the second-order periodograms of the L stretches.

Figure Captions

Figure 1. The annual mean of daily stage values. The horizontal line is at the level of the mean value of all the data.

Figure 2. Monthly mean stage values plotted by year and month.

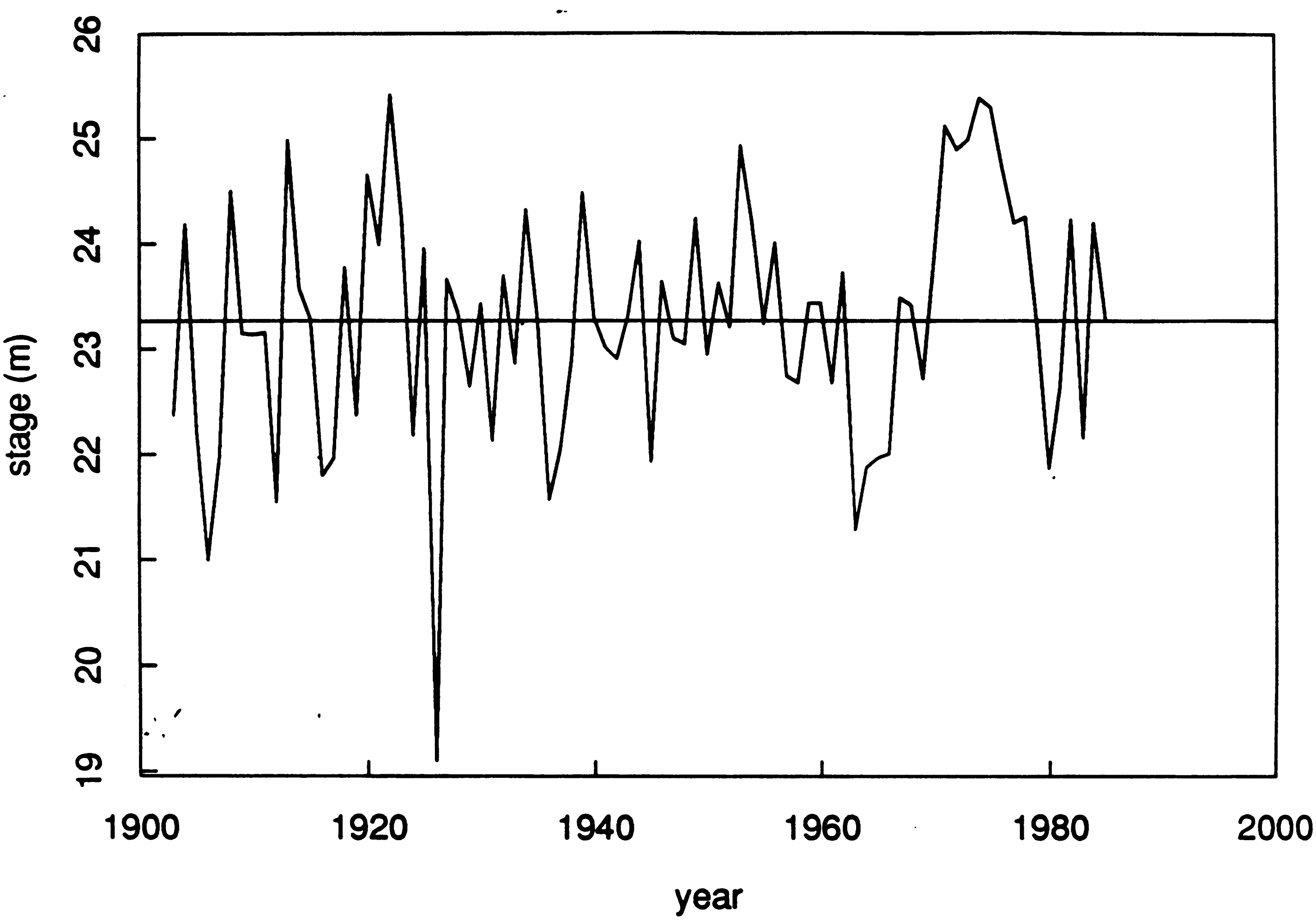
Figure 3. The top graph is the 50% trimmed mean of the annual curves, as defined by (2.1). The bottom graph provides the deviations of each year's curve from the 50% trimmed mean curve. The very low curve corresponds to a year of a serious forest fire.

Figure 4. The top graph is the logarithm of the periodogram of the daily values. The lower graph is the corresponding function for the case of an exact linear trend.

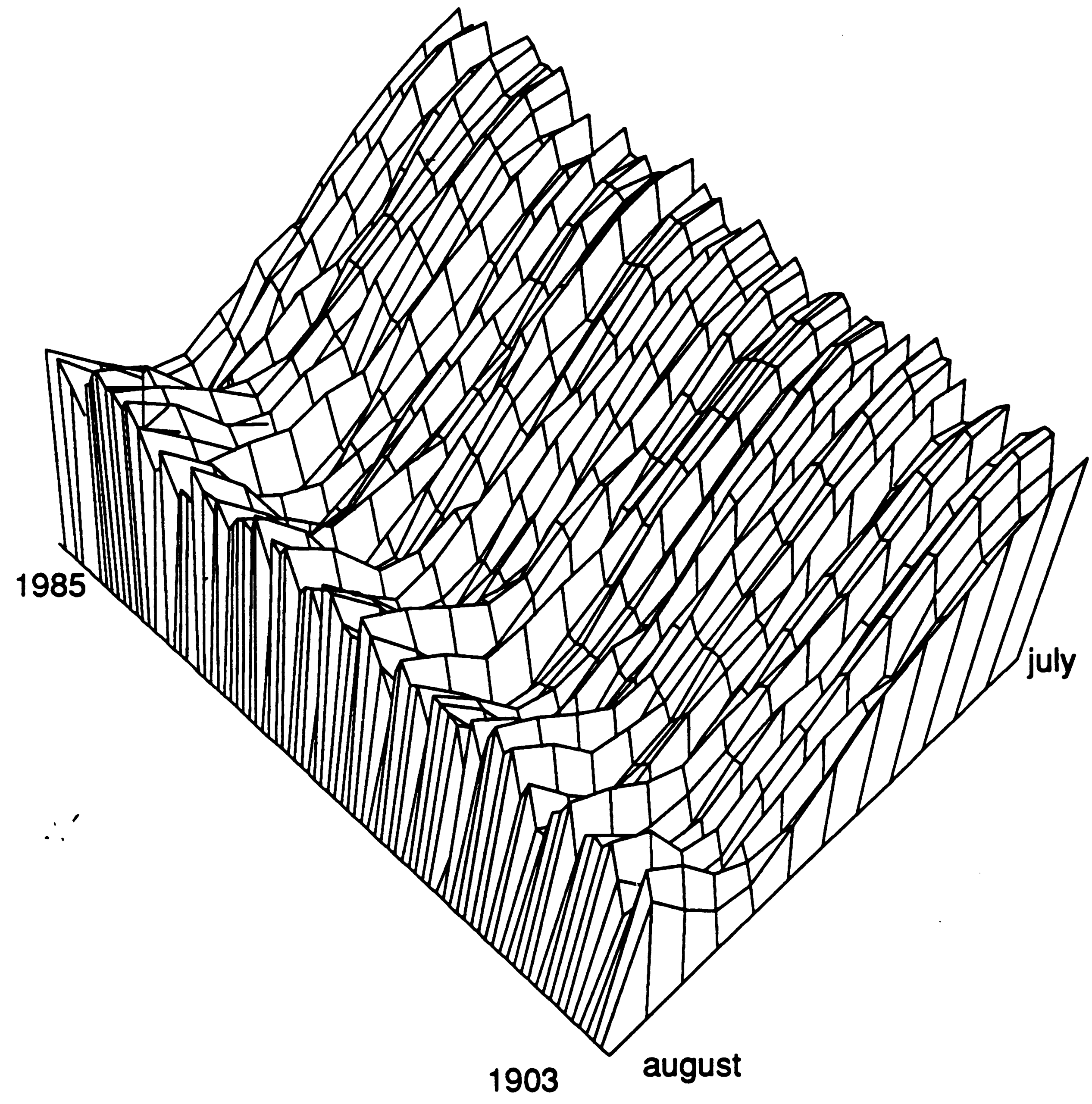
Figure 5. The estimated coherence of the seasonally adjusted monthly mean Manaus stages with monthly mean sunspot numbers. The horizontal line is exceeded with probability 5% in the case of no relationship.

Figure 6. The estimated bicoherence of the seasonally adjusted Manaus daily stage values.

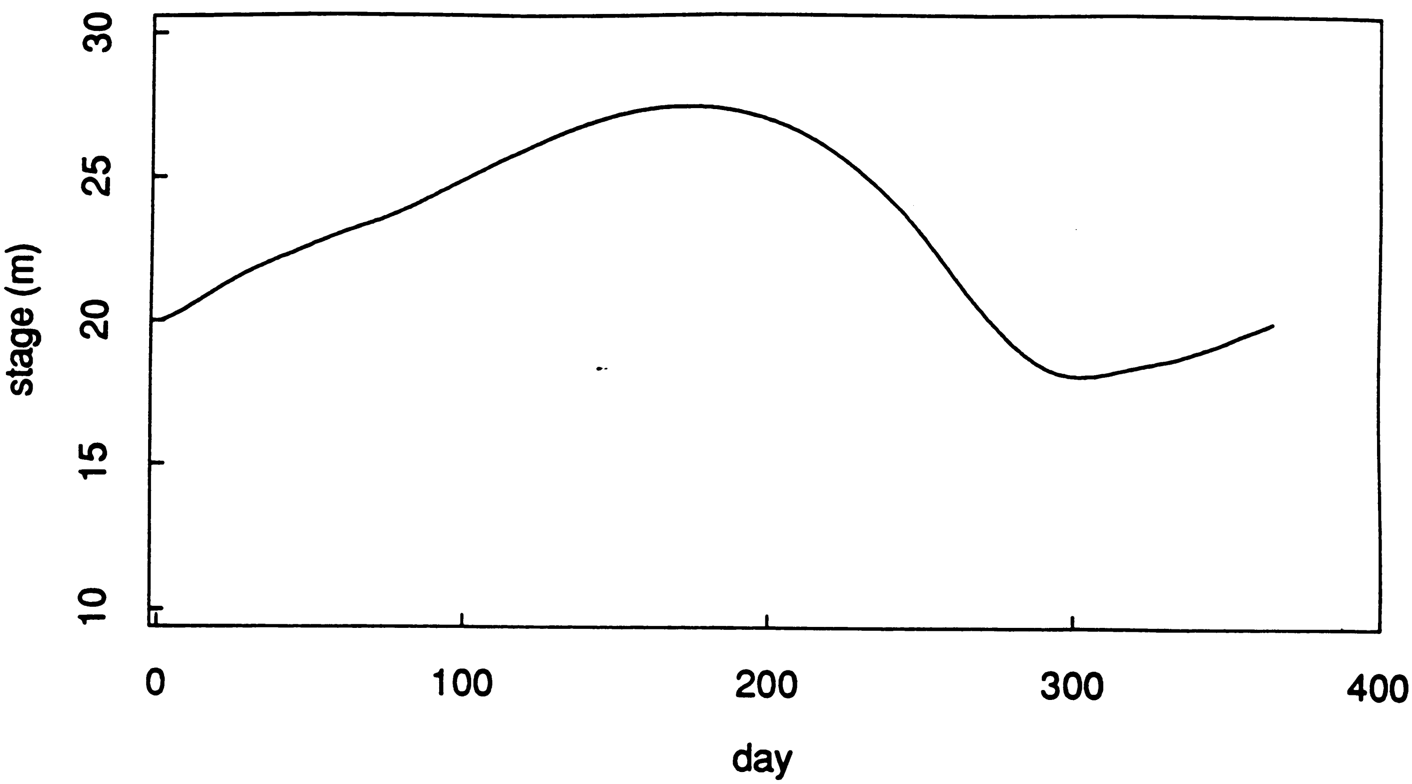
Annual Mean Stage Rio Negro at Manaus



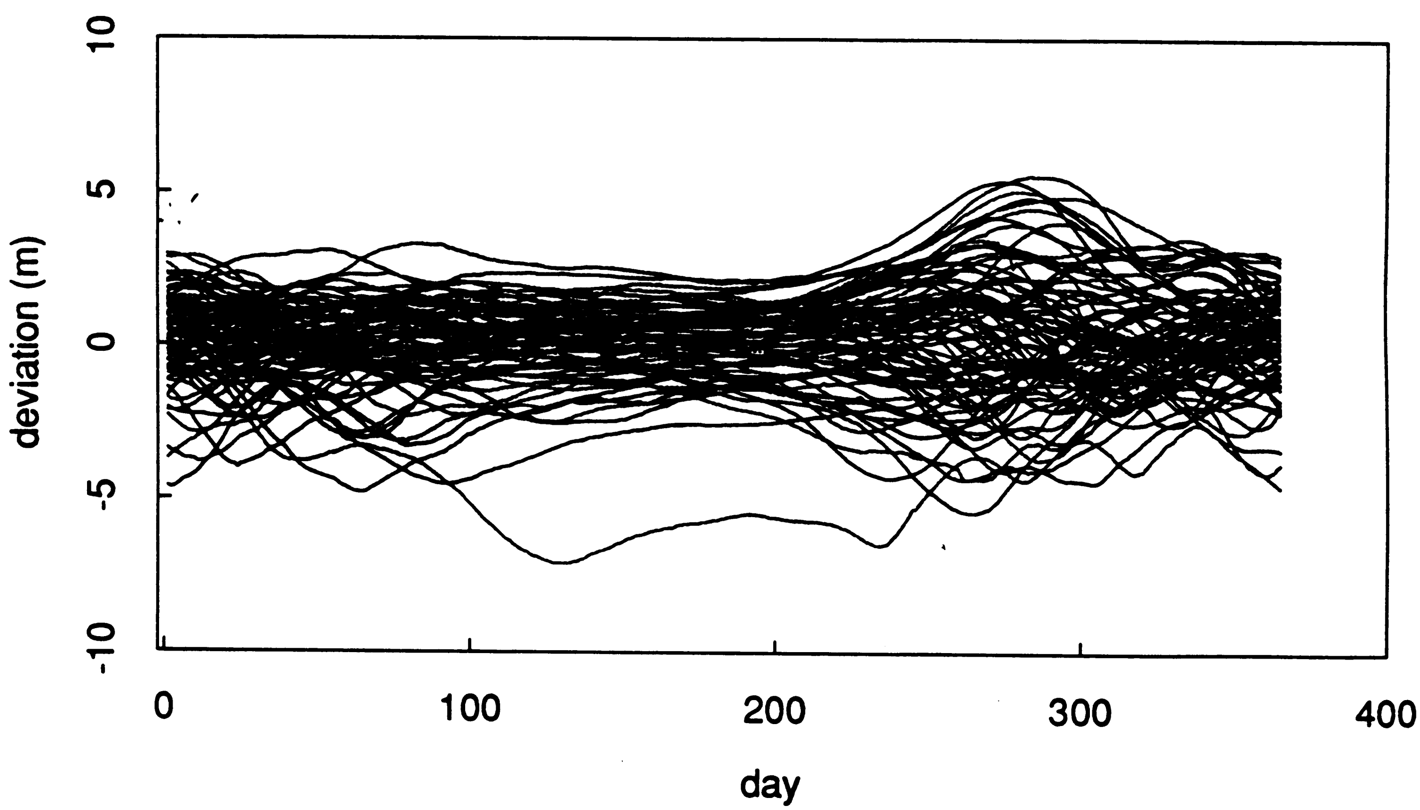
Monthly Mean River Height

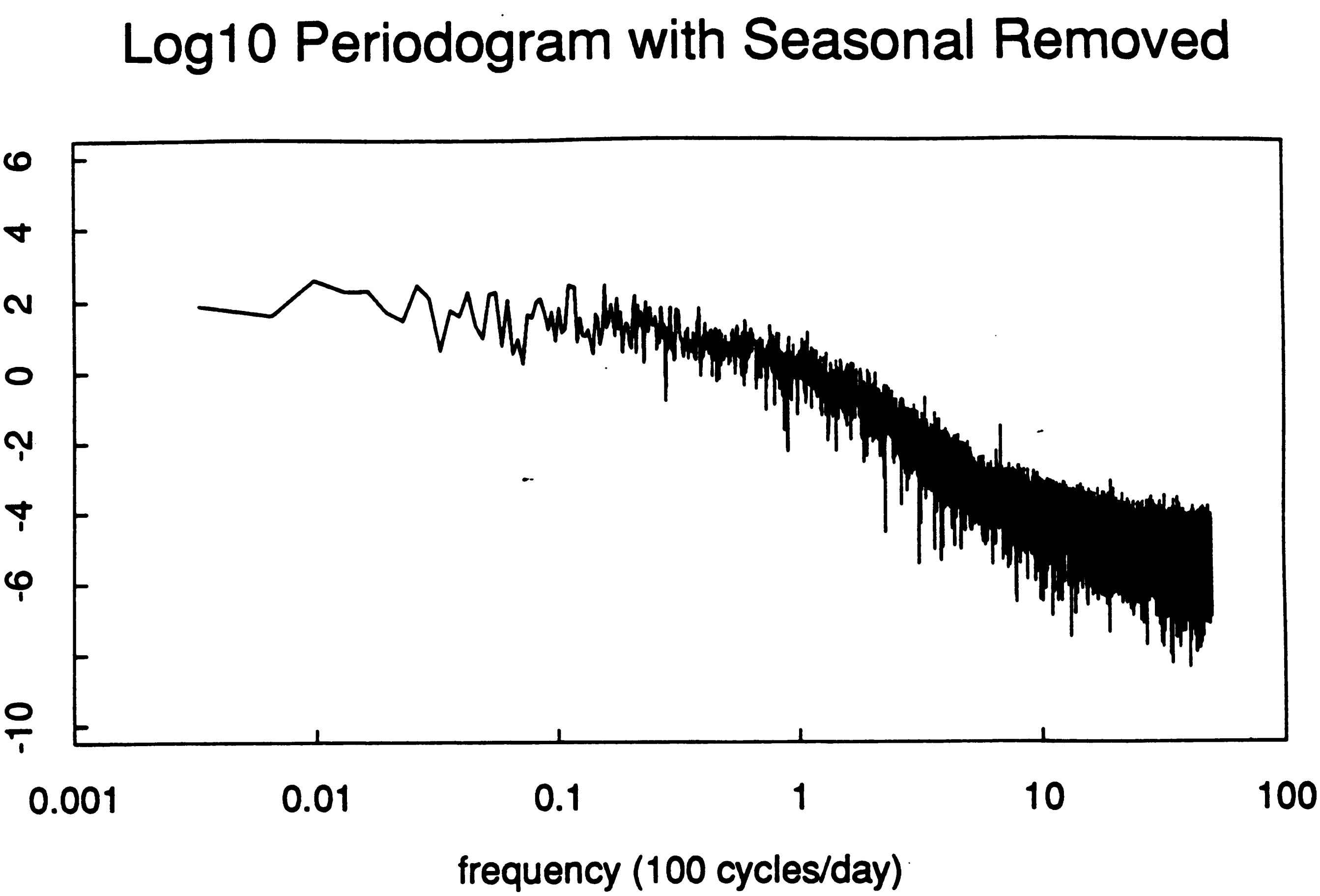


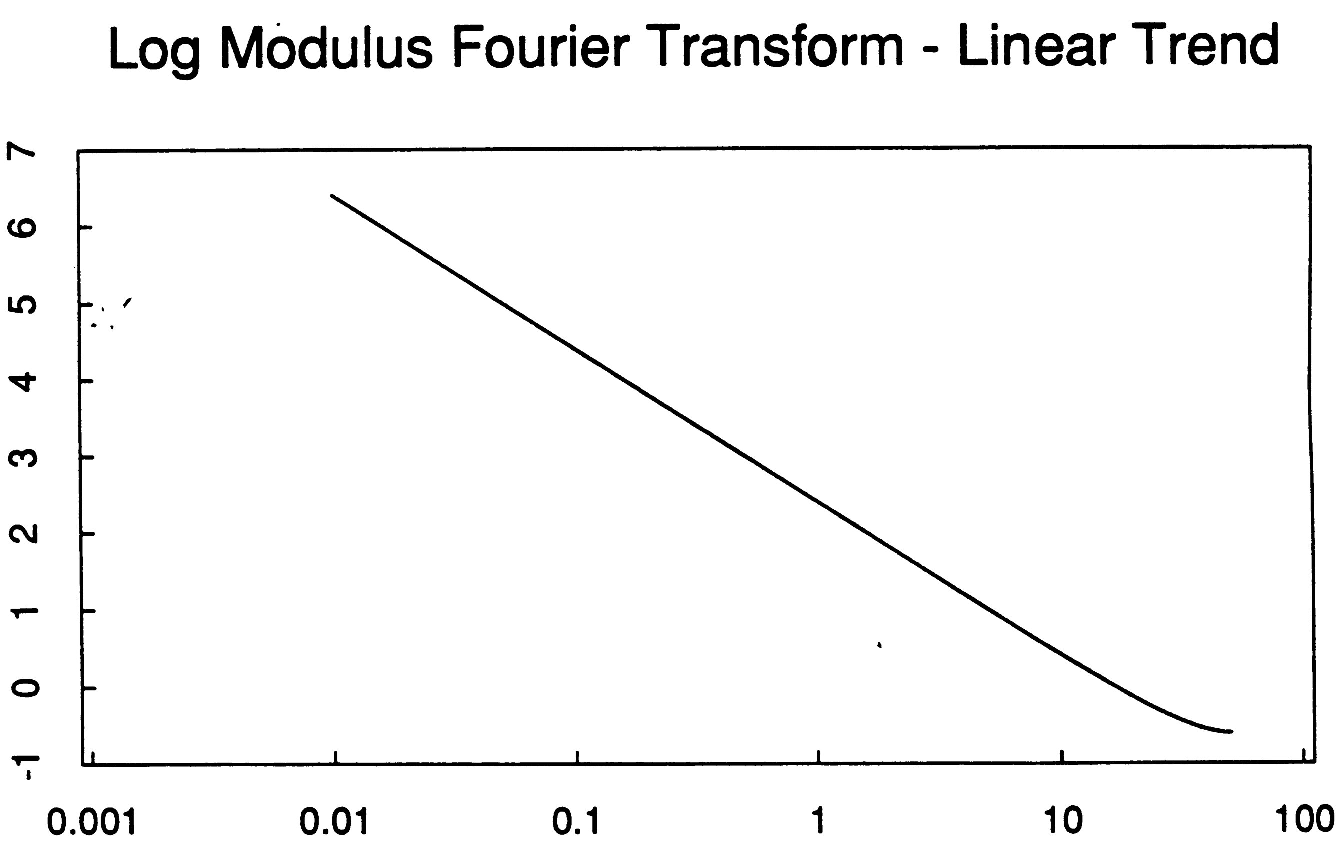
50% Trimmed Mean

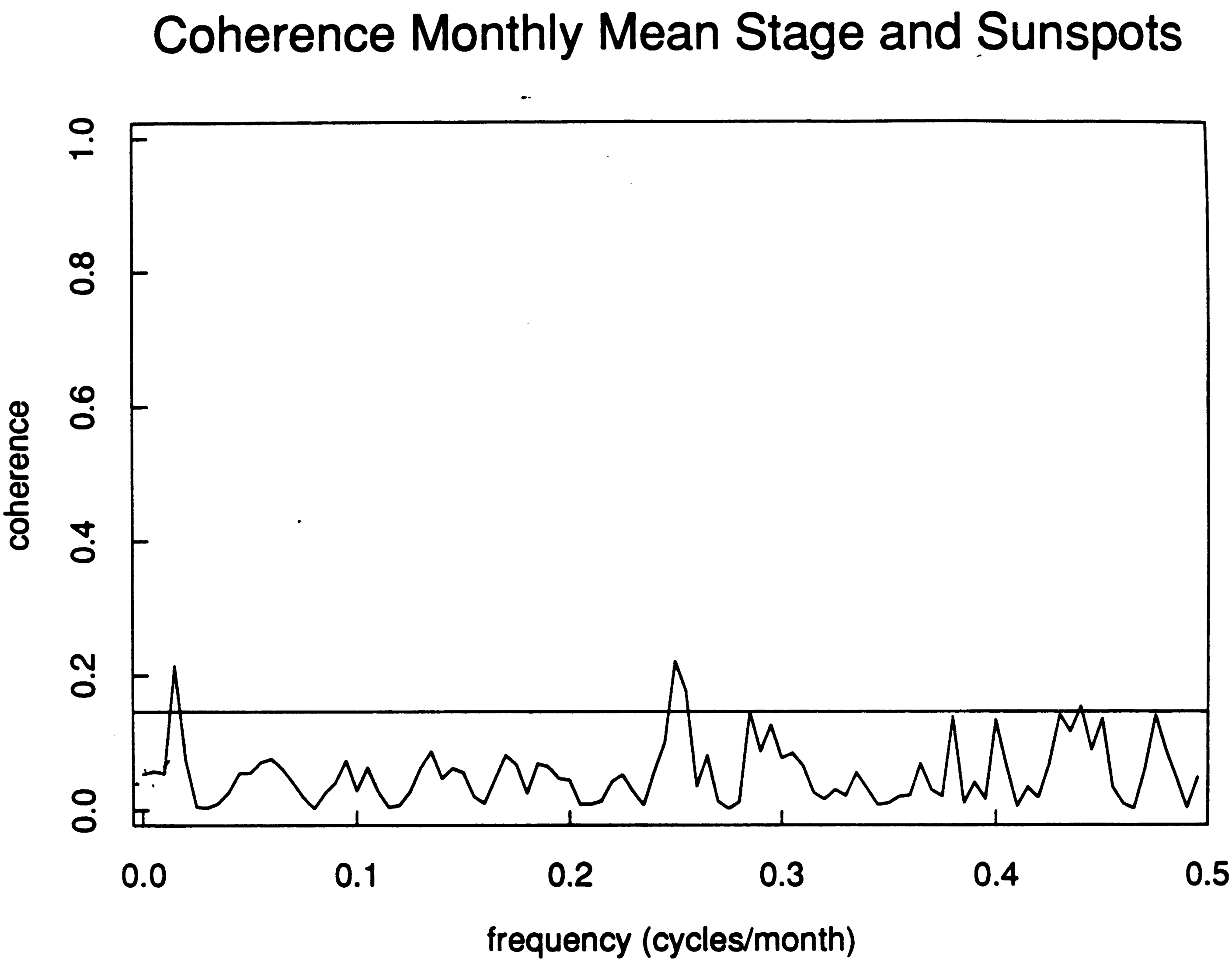


Deviations from Trimmed Mean

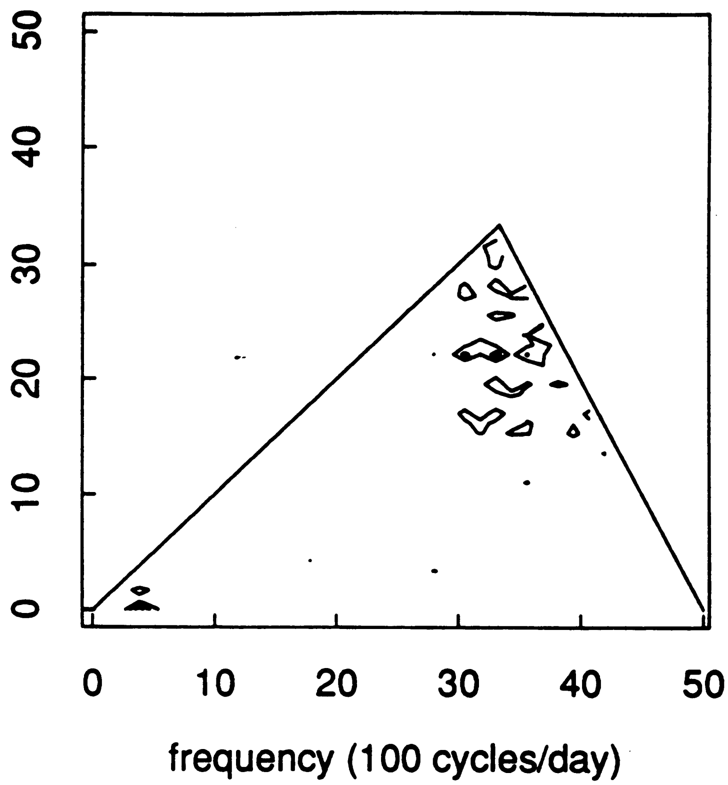


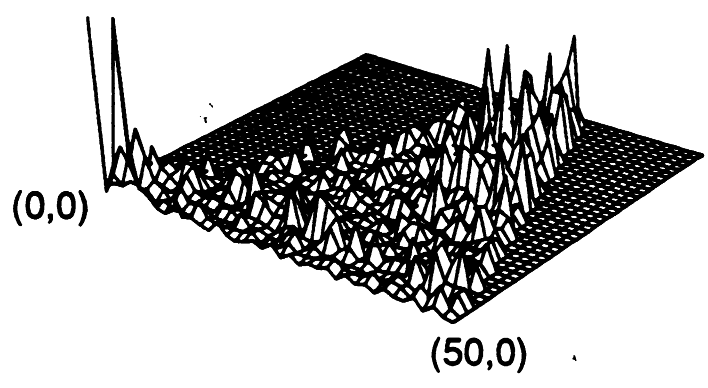

Log10 Periodogram with Seasonal Removed
6
4
2
0
-2
-4
-6
-10
0.001
0.01
0.1
1
10
100
frequency (100 cycles/day)


Log Modulus Fourier Transform - Linear Trend
7
6
5
4
3
2
1
0
-1
0.001
0.01
0.1
1
10
100


Coherence Monthly Mean Stage and Sunspots
coherence
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
frequency (cycles/month)

Bicoherence Estimate





Consistent detection of a monotonic trend superposed on a stationary time series

BY DAVID R. BRILLINGER

Statistics Department, University of California, Berkeley, California, USA

SUMMARY

Consider a time series made up of a signal and a stationary autocorrelated error series. A statistic is proposed for examining the hypothesis that the signal term is constant versus the hypothesis that it is monotonic in time. The statistic is the ratio of a linear combination of the time series values (with coefficients introduced by Abelson and Tukey (1963)), to an estimate of the standard error of the linear combination. The statistic has asymptotic power 1 for a broad class of monotonic alternatives. The procedure is illustrated for the series of river heights at a location on the Rio Negro in Brazil where there is concern that the height is rising due to deforestation of the Amazon Basin. The prob-value obtained is .025 .

Some key words: Amazon Basin, asymptotic methods, central limit theorem, change, consistent test, monotonic trend, power, Rio Negro, stationary time series

1. INTRODUCTION

1.1 Preamble

In the study of a scientific phenomenon via time series data, a fundamental question that sometimes arises is; is there a trend in the series? In particular one may mention the case of the thickness of the atmosphere's ozone layer possibly decreasing with increasing use of chlorofluorocarbons, Stolarski (1988), and the case of the long-term height of the Amazon river increasing with deforestation, Sternberg (1987). If there are changes in these two cases, it seems pertinent to view the changes as monotonic. These are particular examples of circumstances where a stimulus applied is increasing with time, and so there could well be a monotonic response effect.

To begin, consider the time series model

$$Y(t) = S(t) + E(t) \qquad (1.1)$$

$t = 0, \pm 1, \pm 2, \cdots$ where $S(.)$ is a deterministic signal and $E(.)$ is a zero mean stationary noise series. Of interest is the hypothesis: $S(t)$ equal constant, versus the alternative: $S(t) \le S(t+1)$ for all t, with strict inequality for some t.

Given data $Y(t)$, $t=0,...,T-1$ the procedure studied in this paper is based on a linear combination, $\sum c(t)Y(t)$, involving particular coefficients $c(.)$.

1.2 Earlier work in the autocorrelated time series case

Time series researchers have long approached the problem of trend analysis via the technique of fitting a parametric form for $S(.)$ and then examining the parameter estimates obtained. For example, one might fit $S(t) = \alpha+\beta t$ and consider the hypothesis $\beta = 0$. Grenander (1954) has worked out the asymptotics in this case. Specifically, given the data $Y(0)$, $Y(1)$, ..., $Y(T-1)$, his methods lead to the result that the statistic

$$\hat{\beta} = \sum_{t=0}^{T-1}(t - \bar{t})Y(t)/\sum_{t=0}^{T-1}(t - \bar{t})^2 \tag{1.2}$$

has mean β and asymptotic variance $2\pi f_{EE}(0)/\sum(t - \bar{t})^2$ where $f_{EE}(\lambda)$ denotes the power spectrum of the series E(.) at frequency λ and $\bar{t} = (T-1)/2$. Further, under regularity conditions, $\hat{\beta}$ is asymptotically normal, (see eg. Brillinger (1975), Theorem 5.11.1), and so for large T one may compute approximate P-values, confidence intervals and the like.

An alternate approach to trend analysis that is currently under development, see eg. Akaike (1980), Kitagawa (1981), involves what may be viewed as either Bayesian models for trend with a smoothness prior or models with a stochastic trend. These models do not have monotonic trends however, and so are not directly pertinent to the problem of this work.

1.3 Earlier work in the independent case

When the error series, E(.), is white noise, (that is a sequence of independent identically distributed random variables), quite a number of techniques, both parametric and nonparametric, have been proposed. Lombard (1987) is one recent reference and Shaban (1980) provides a bibliography of some results

The work for the case of independent errors that is pertinent to the time series problem of present concern, is that of Abelson and Tukey (1963). These authors address the problem of determining a statistic linear in the data and sensitive to monotonic mean function departure, as follows.

Determine coefficients, $\mathbf{c} = \{c(t), t=0,1,...,T-1\}$, with mean $\bar{c} = 0$ to

$$\max_{\mathbf{c}} \min_{\mathbf{S}} \frac{|\sum(c(t)-\bar{c})(S(t)-\bar{S})|^2}{\sum(c(t)-\bar{c})^2\sum(S(t)-\bar{S})^2} \tag{1.3}$$

where $\mathbf{S} = \{S(0) \le S(1) \le \cdots \le S(T-1)\}$. The coefficients Abelson and Tukey (1963) obtained are

$$c(t) = c^T(t) = \sqrt{t(1 - \frac{t}{T})} - \sqrt{(t+1)(1 - \frac{t+1}{T})} \tag{1.4}$$

The value at the extreme for expression (1.3) is $1/\sum c(t)^2 \approx 2/\log T$ for large T. This value is achieved for the step-function signals $S(t) = 0$ for $t \le t_0$ and $S(t) = 1$ for $t > t_0$. A plot of the coefficients $c(t)$ is provided by Figure 1 for the case of $T = 100$. (The shape will be the same for other values of T.) The linear combination, $\sum c(t)Y(t)$, is seen to strongly contrast the beginning and ending levels of the data, as seems intuitively reasonable given that one is looking for a monotonic trend across the time series values.

In the case that the E(t) are independent normals with known variance σ^2, a test statistic is provided by $\sum c(t)Y(t)/\sigma\sqrt{\sum c(t)^2}$. In practice if σ^2 is unknown but one has an independent chi-squared estimate of σ^2, then one can compute a t-statistic.

1.4 Notation and structure of the paper

Specific assumptions, theorems and proofs have been placed in Appendices. The results are of asymptotic character, but are hoped to provide useful finite sample approximations. The term "monotonic" will allow equality of some consecutive values in the sequence. The symbol $\sum$, unsubscripted, will refer to summation over $t = 0, 1,..., T-1$. It will be taken that $f_{EE}(0) \neq 0$.

2. THE PROPOSED PROCEDURE

2.1 The statistic

The problem of specific concern in this paper is that of the model (1.1) with an autocorrelated noise series, E(.).

Consider $\sum c(t)Y(t)$ with $Y(t)$ given by expression (1.1) and with the coefficients, $c(t)$ given by (1.4). One has

$$E\{\sum c(t)Y(t)\} = \sum c(t)S(t) \tag{2.1}$$

and

$$var\{\sum c(t)Y(t)\} = \sum_{s,t} c(s)c(t)c_{EE}(s-t) \tag{2.2}$$

where $c_{EE}(u) = cov\{E(t+u),E(t)\}$. It is shown in Appendix 3 that for large T expression (2.2) is approximately $2\pi f_{EE}(0)\sum c(t)^2$ where $f_{EE}(.)$ is the noise power spectrum. Further if the noise series is normal, then $\sum c(t)Y(t)$ is normal, while if the noise series is mixing, then the linear combination is asymptotically normal (see Appendix 3).

If one has $\hat{f}_{EE}(0)$, a consistent estimate of $f_{EE}(0)$, then in the case that $S(t)$ is constant one may approximate the distribution of the statistic

$$\sum c(t)Y(t)/\sqrt{2\pi \hat{f}_{EE}(0)\sum c(t)^2} \tag{2.3}$$

by a standard normal, and thereby compute an approximate P-value or carry out formal tests of significance. The problem of constructing such a consistent estimate of $f_{EE}(0)$ is now addressed.

2.2 Estimation of the variance

Suppose that the trend function $S(.)$ has the form $g(t/T)$ where $g(.)$ has a Lipshitz integral modulus of continuity of order α (see (A.4) below). The signal is taken to have this form in order that it may be present the whole of a time interval tending to infinity with T. Let the signal be estimated by the running mean

$$\hat{S}(t) = \sum_{s=-V}^{V} Y(t+s)/(2V+1) \tag{2.4}$$

$t = V+1,V+2,...,T-1-V$ for moderate sized V. The noise series may then be estimated by the residuals $\hat{E}(t) = Y(t)-\hat{S}(t)$. An estimate of $f_{EE}(0)$ may be based on these last.

Specifically denote the discrete Fourier transform of the residuals by

$$\varepsilon_j = \sum_{t=V+1}^{T-1-V} \hat{E}(t)\exp\{-\frac{2\pi itj}{T}\} \tag{2.5}$$

for $j = 0,1,...,T-1$. Define the transfer function values

$$a_j = sin(2\pi j[2V+1]/2T)/[2V+1]sin(2\pi j/2T) \tag{2.6}$$

For chosen L and M compute, $\hat{f}_{EE}(0)$, the smoothed periodogram spectral estimate

$$\sum_{j=M}^{L+M-1} \frac{1}{2\pi T}|\varepsilon_j|^2 / \sum_{j=M}^{L+M-1}(1-a_j)^2 \approx \sum_{j=M}^{L+M-1} \frac{1}{2\pi T}(1-a_j)^2|y_j|^2 / \sum_{j=M}^{L+M-1}(1-a_j)^2 \tag{2.7}$$

Here y_j denotes the discrete Fourier transform of the original series. Frequencies from M/T to $(L+M-1)/T$ cycles/unit time are involved, where both of these are to be near

0. The terms $(1-a_j)^2$ are introduced to compensate for the effect of the filtering operation on the noise series. In the data analysis below it seemed reasonable to take $M = 1$, but in the asymptotics developed M is set to $M = \delta T/V$ for some fixed δ.

Specific assumptions are given in the Appendix under which the estimate (2.7) is consistent as $T \to \infty$.

3. POWER

It has been indicated that the variate $\sum c(t)Y(t)$ may be approximated by a normal with mean $\sum c(t)S(t)$ and with variance $2\pi \hat{f}_{EE}(0) \sum c(t)^2$. This leads, for example, to approximating a probability such as

$$Prob\,\{\sum c(t)Y(t)/\sqrt{2\pi \hat{f}_{EE}(0)\sum c(t)^2} > d\,\} \tag{3.1}$$

for some given d by

$$1 - \Phi\left[d - \sum c(t)S(t)/\sqrt{2\pi f_{EE}(0)\sum c(t)^2}\right] \tag{3.2}$$

where $\Phi(.)$ is the normal cumulative. Now, from the results of Abelson and Tukey (1963) one has

$$\frac{|\sum c(t)S(t)|^2}{\sum c(t)^2} \geq \frac{2\sum(S(t) - \bar{S})^2}{\log T} + o(1) \tag{3.3}$$

for all monotonic $S(t)$. It follows that a test based on rejecting the hypothesis of constant mean when the statistic (2.3) exceeds d has asymptotic power 1 for $S(.)$ such that $\sum(S(t) - \bar{S})^2/\log T$ tends to infinity with T. (This is the case under Assumption A.2 below.)

4. THE EXAMPLE OF THE AMAZON RIVER

Daily stage, that is height, readings have been made since 1903 at Manaus, 18 km. up the Rio Negro estuary from the Amazon River in Brazil. In all 30529 readings were available for analysis. Many developments have taken place in the Amazon basin this century, particularly a steady deforestation, so it seems of interest to examine this river stage series for monotonic trend.

Figure 2 provides a graph of the data. A strong seasonal effect is present. Figure 3 is a graph of the 8 year running mean level as defined by expression (2.4) with $V = 4*365.25$. The seasonal effect is no longer apparent. (This running mean can be quickly computed via a fast Fourier transform algorithm.)

In order to obtain some idea of the low frequency character of the time series it is useful to look at the periodogram. Figure 4 provides the periodogram of the data at the lower frequencies. The high peak occurs at the seasonal frequency. (This periodogram was computed via a fast Fourier transform of the $N = 30720$ observations obtained by adding 191 zeros to the end of the given data).

Turning to a formal examination of the data for a monotonic trend, expression (2.3) is evaluated. The spectrum at 0 is estimated by expression (2.7) with $M = 1$ and $L = 25$. The value then obtained for the statistic (2.3) is 1.961. The corresponding P-value for the hypothesis of monotonic increase is .025. By way of comparison if a linear trend is fit to the series and the power spectrum at 0 estimated from the residuals of that fit, then the P-value obtained is .040 . In summary there is some evidence for increasing trend (or change) in this data.

5. DISCUSSION AND CONCLUDING REMARKS

The non-centrality parameter occurring in expression (3.2) is $\sum c(t)S(t)/\sqrt{2\pi f_{EE}(0)\sum c(t)^2}$. Under Assumption A.2 below, this is of order of magnitude $T/\sqrt{\log T}$ and tends to infinity with T. If one employed the statistic $\hat{\beta}$ of (1.2) instead, the corresponding non-centrality parameter would be of order $\sqrt{T}$ which tends to infinity more slowly. The corresponding test is therefore consistent, but has asymptotic efficiency 0 relative to the first.

It is clear that one could form other estimates of $S(t)$, such as the (robust) linear smoother of Cleveland (1979), or the monotonic smoother of Friedman and Tibshirani (1984); however the running mean is quickly computed and its statistical properties are simply derived. Further, it would appear that not a lot would be gained by using these other smoothers in the estimation of $f_{EE}(0)$.

ACKNOWLEDGEMENTS

The Rio Negro data was provided by Professor H. O'Reilly Sternberg of the Geography Department, University of Calfomia, Berkeley. The work was supported by National Science Foundation Grant DMS-8316634. The computations were carried out on a Sun 3/50 work station with double precision arithmetic.

APPENDIX 1

Assumptions

The cumulant functions of the stationary series E(.) are defined by

$$c_{E \cdots E}(u_1, \ldots, u_{k-1}) = cum\{E(t+u_1),\ldots,E(t+u_{k-1}),E(t)\} \tag{A.1}$$

for $k = 2,3,\ldots$ and the power spectrum at frequency λ by

$$f_{EE}(\lambda) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} c_{EE}(u)e^{-i\lambda u} \tag{A.2}$$

when this series converges.

Assumption A.1 (Mixing). For $k = 2,3,\ldots$

$$\sum_{u_1} \cdots \sum_{u_{k-1}} |c_{E \cdots E}(u_1,\ldots,u_{k-1})| < \infty \tag{A.3}$$

with $u_1, \ldots, u_{k-1}$ running from $-\infty$ to ∞.

Assumption A.2. *The signal has the form $S(t) = S^T(t) = g(t/T)$, with $g(.)$ square integrable, and where there is an α such that the integral modulus of continuity of $g(.)$ satisfies*

$$\sup_{|v| \le h} \int_0^1 |g(u+v)-g(u)|^2 du = O(h^\alpha) \tag{A.4}$$

as $h \to 0$.

Condition (A.4) allows steps in $g(.)$. In particular if $g(u) = 0$ for $u \le u_0$ and $= 1$ for $u > u_0$, then $\alpha = 1$. In the case that $g(.)$ has a bounded derivative, $\alpha = 2$.

APPENDIX 2

Lemmas

Lemma A.1. *For the coefficients* $c(.)$ *of (1.4), for fixed* $u, u_1, \ldots, u_{k-1}$ *and for* $k = 2, 3, \ldots$ *one has*

$$\sum c(t)^2 \approx \frac{1}{2}\log T \tag{A.5}$$

$$\lim_{T\to\infty} \sum c(t+u)c(t)/\sum c(t)^2 = 1 \tag{A.6}$$

$$\lim_{T\to\infty} \sum c(t+u_1)\cdots c(t+u_{k-1})c(t)/[\sum c(t)^2]^{k/2} = 0 \tag{A.7}$$

Proof. By elementary analysis.

Next define

$$\tilde{S}(t) = \sum_{s=-V}^{V} S(t+s)/(2V+1) \tag{A.8}$$

the running mean of the signal. One has

Lemma A.2. *Let* $S(t) = g(t/T)$ *with* $g(.)$ *satisfying Assumption A.2. Let* $V = V^T$ *be such that* $V/T \to 0$ *as* $T \to \infty$. *Then*

$$\sum_t |S(t) - \tilde{S}(t)|^2 = \frac{1}{T}\sum_j |s_j - \tilde{s}_j|^2 = O(V^{\alpha}T^{1-\alpha}) \tag{A.9}$$

Proof. The first equality follows from Parseval's Theorem. Next, neglecting end effects as one may,

$$\sum |S(t) - \tilde{S}(t)|^2 \le \frac{1}{2V+1}\sum_{|s|\le V}\sum |S(t) - S(t+s)|^2 \le \sup_{|s|\le V}\sum |S(t) - S(t+s)|^2$$

$$\le \sup_{|s|\le V}\sum |g(\frac{t}{T}) - g(\frac{t+s}{T})|^2 \approx T \sup_{|v|\le V/T}\int |g(u+v) - g(u)|^2 du$$

and one has the indicated result.

APPENDIX 3

Theorems

Theorem A.1. *Suppose that the series* E(.) *satisfies Assumption A.1 . Suppose that* $c(.)$ *is given by (1.4). Then*

$$\lim_{T\to\infty} var\{\sum c(t)Y(t)\}/\sum c(t)^2 = 2\pi f_{EE}(0) \tag{A.10}$$

Proof. From expression (2.2) the variance may be written

$$\sum_{u=-T+1}^{T-1} c_{EE}(u) \sum_{t=0}^{T-1-|u|} c(t+u)c(t)$$

The result now follows from (A.3) and the Dominated Convergence Theorem.

Theorem A.2. *Under the conditions of Theorem A.1, the variate* $\sum c(t)Y(t)$ *is asymptotically normal.*

Proof. The standardized cumulant of order k may be written

$$\sum_{u_1}\cdots\sum_{u_{k-1}} c_{E\cdots E}(u_1,\ldots,u_{k-1})\sum_t c(t+u_1)\ldots c(t+u_{k-1})c(t)/[\sum c(t)^2]^{k/2}$$

This tends to 0 for $k > 2$ by (A.3), (A.7) and the Dominated Convergence Theorem.

The asymptotic normality then follows from Lemma P4.5 in Brillinger (1975), page 403.

Theorem A.3. *Let the series* E(.) *satisfy Assumption A.1 and let the signal* $S(.)$ *satisfy Assumption A.2. Suppose that* $L = L^T$ *and* $V = V^T$ *tend to* ∞ *as* $T \to \infty$ *in such a way that* L/T, V/T, $V^\alpha T^{1-\alpha}/L \to 0$. *Suppose that* $M = \delta T/V$ *for fixed* δ. *Then* $\hat{f}_{EE}(0)$ *is a consistent estimate of* $f_{EE}(0)$.

Proof Since $Y(t) = S(t) + E(t)$ one has $y_j = s_j + \varepsilon_j$ and

$$|\hat{\varepsilon}_j|^2 = |s_j - \hat{s}_j|^2 + (\overline{s_j - \hat{s}_j})(\varepsilon_j - \hat{\varepsilon}_j) + (s_j - \hat{s}_j)(\overline{\varepsilon_j - \hat{\varepsilon}_j}) + |\varepsilon_j - \hat{\varepsilon}_j|^2$$

To begin, one shows that the first term here has negligeable effect on (2.6). Because $M = \delta T/V$, the $(1 - a_j)$ are bounded away from 0 and it is enough to consider the order of magnitude of

$$\frac{1}{L}\sum_j \frac{1}{T}|s_j - \hat{s}_j|^2$$

From (A.9) this is $O(L^{-1}T^{1-\alpha}V^\alpha)$ and so tends to 0 under the indicated limit processes.

Next one requires that

$$\sum\frac{1}{2\pi T}|\varepsilon_j - \hat{\varepsilon}_j|^2 \Big/ \sum(1-a_j)^2 \approx \sum\frac{1}{2\pi T}(1-a_j)^2|\varepsilon_j|^2 \Big/ \sum(1-a_j)^2$$

is a consistent estimate of $f_{EE}(0)$. This follows from classic arguments like those of Grenander and Rosenblatt (1957) and Parzen (1957).

Finally, from the previous two results, the cross product terms may be neglected.

REFERENCES

ABELSON, R.P. and TUKEY, J.W. (1963). Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order. *Ann. Math. Statist.* 34, 1347-1369.

AKAIKE, H. (1980). Seasonal adjustment by Bayesian modelling. *J. Time Series Analysis* 1, 1-13.

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory.* New York: Holt, Rinehart and Winston.

CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. American Statistical Assoc.* 74, 829-836.

FRIEDMAN, J. and TIBSHIRANI, R. (1984). The monotone smoothing of scatterplots *Technometrics* 26, 243-250.

GRENANDER, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math, Statist.* 25, 252-272.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series.* New York: J. Wiley.

KITAGAWA, G. (1981). A nonstationary time series model and its fitting by a recursive filter. *J. Time Series Analysis* 2, 103-116.

LOMBARD, F. (1987). Rank tests for changepoint problems. *Biometrika* 74, 615-24.

PARZEN, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* 28, 329-348.

SHABAN, S. A. (1980). Change point problem and two-phase regression: an annotated bibliography. *International Statistical Review* 48, 83-93.

STERNBERG, H. O'R. (1987). Aggravation of floods in the Amazon River as a consequence of deforestation? *Geografiska Annaler* 69A, 201-219.

STOLARSKI, R. S. (1988). The Antartic ozone hole. *Scientific American* 258, 30-36.

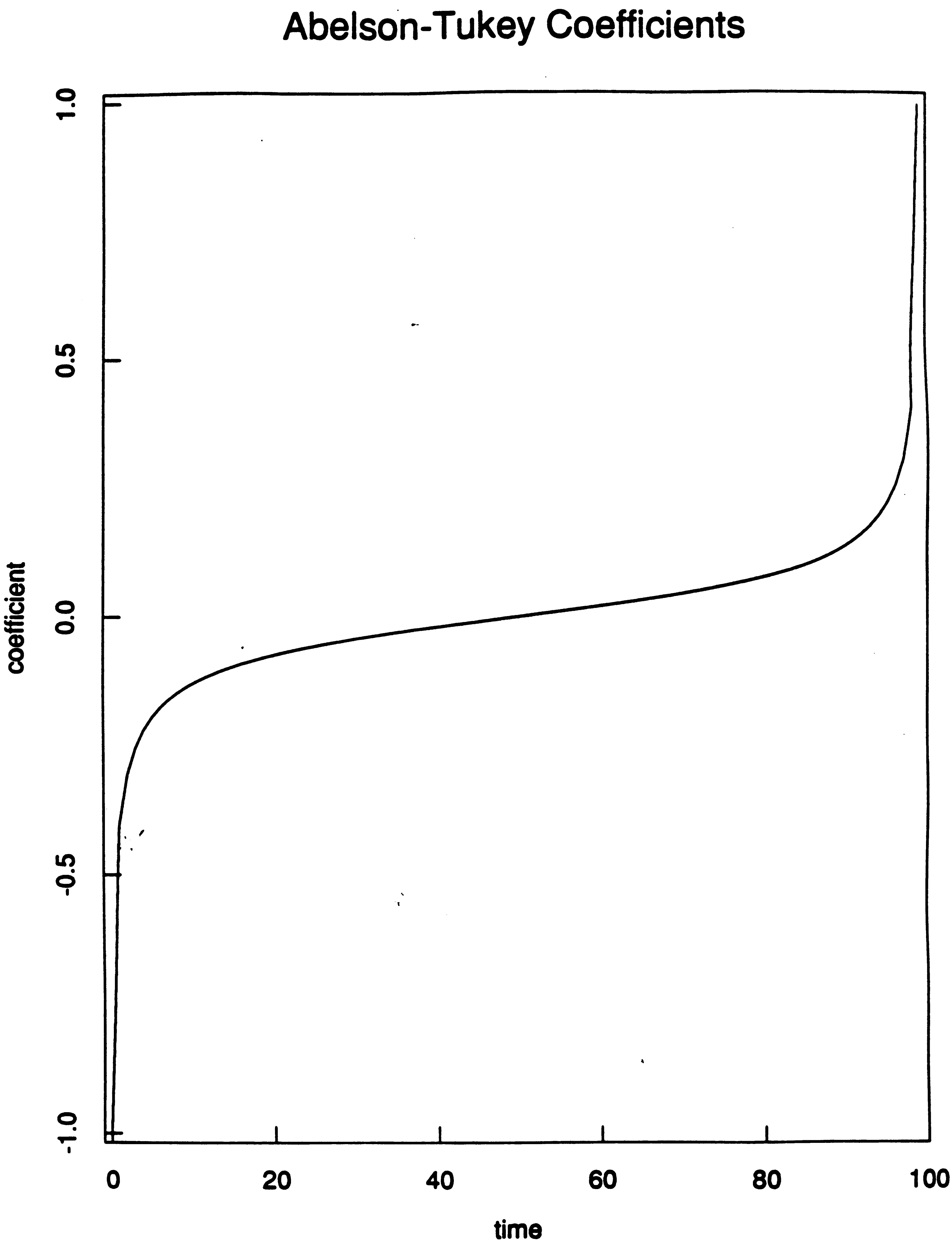
Figure Captions

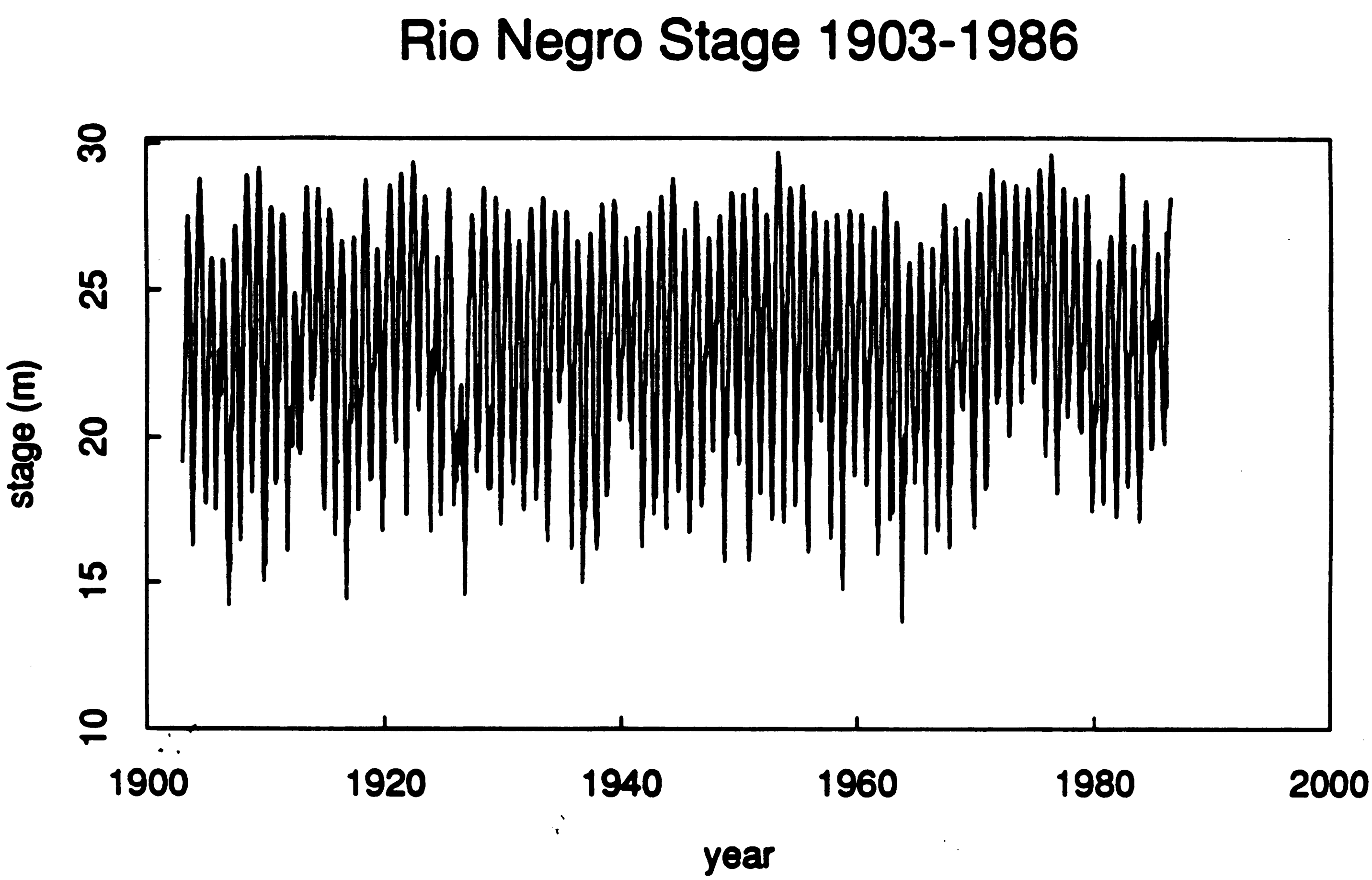
Figure 1. The coefficients of expression (1.4) in the case that $T = 100$.

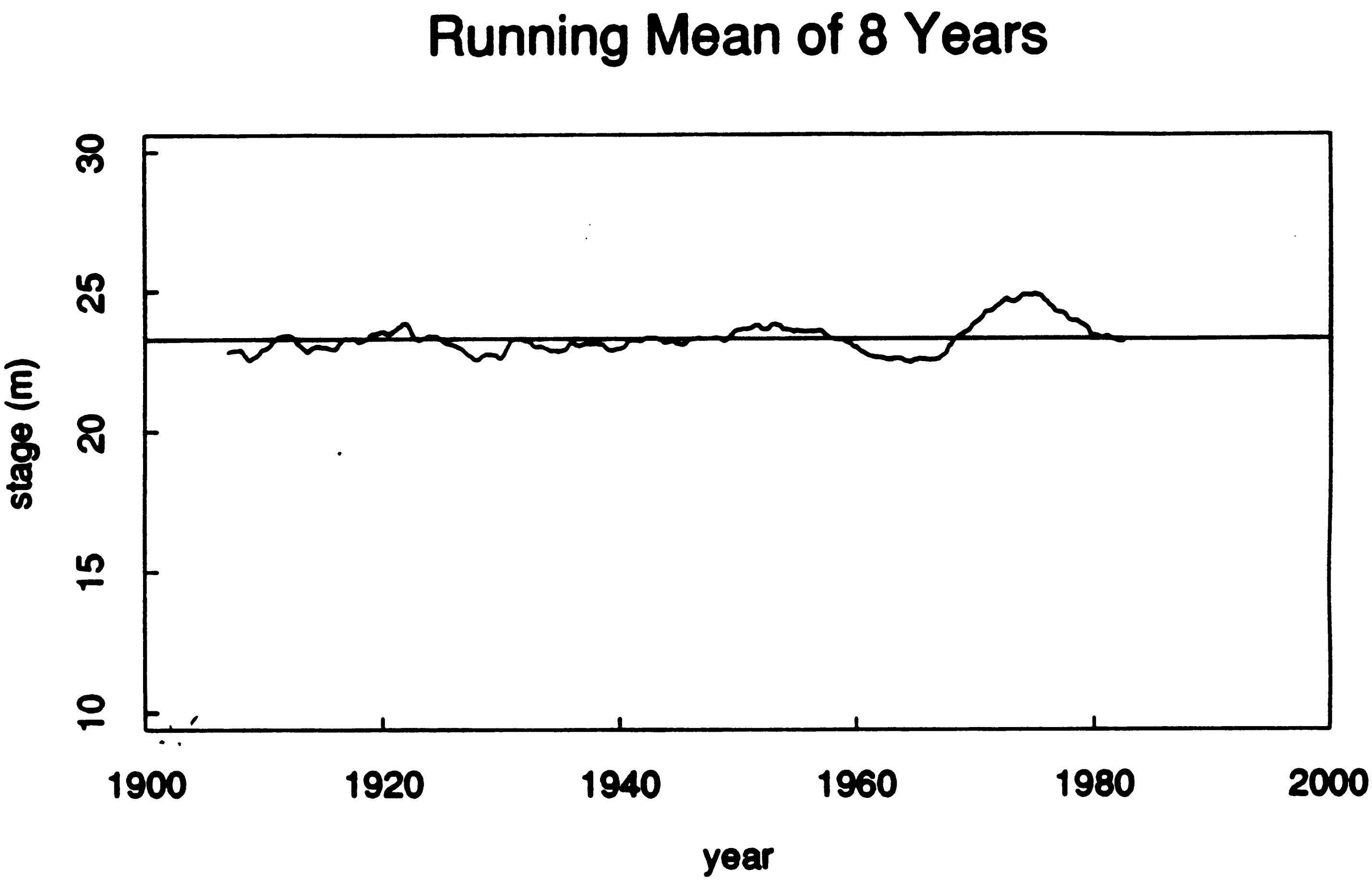
Figure 2. Graph of the daily river height level as measured in meters from a reference point at Manaus.

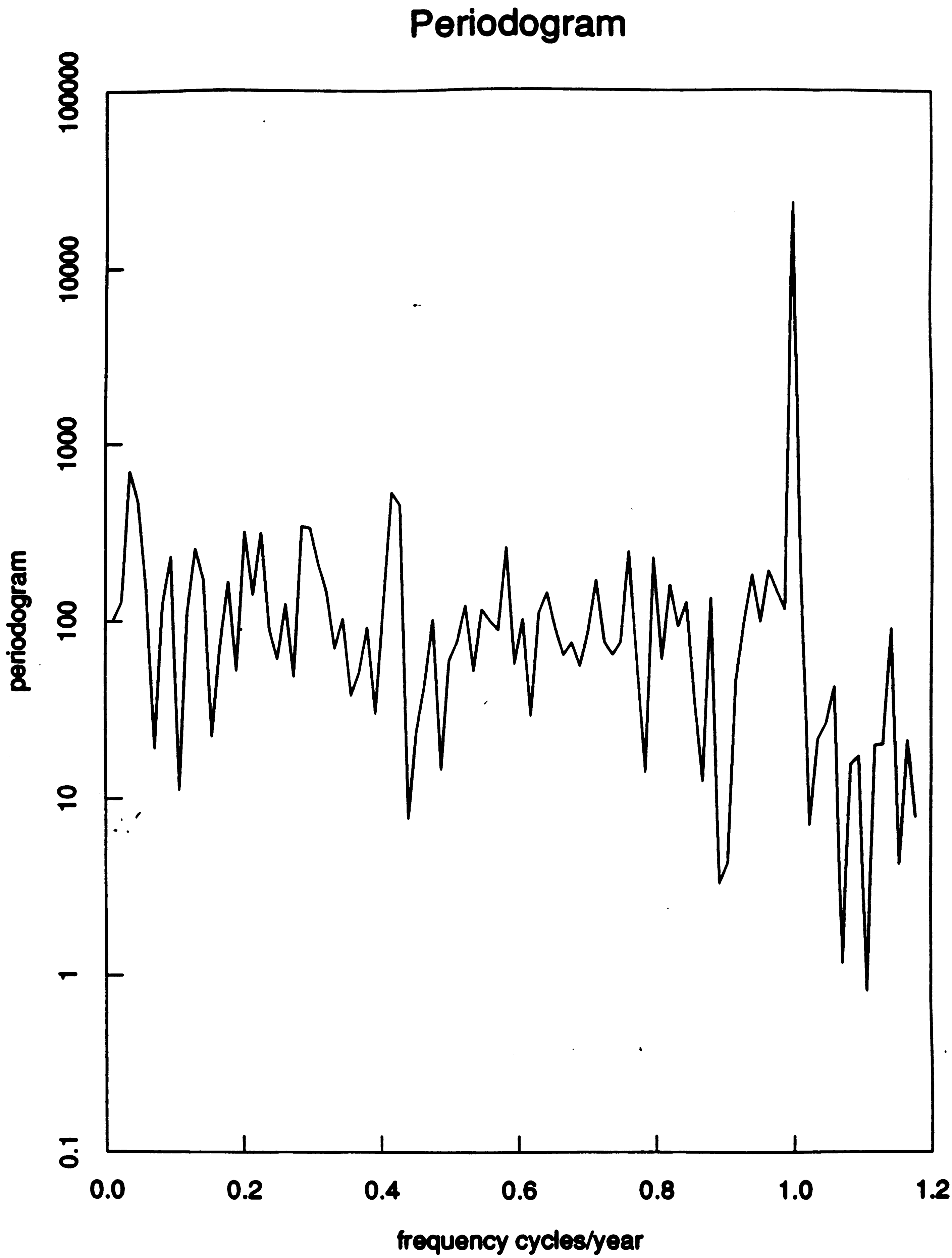
Figure 3. An eight year running mean of the series of Figure 2. The horizontal line is at the overall mean level.

Figure 4. The low frequency portion of the periodogram of the data of Figure 2.


Abelson-Tukey Coefficients
coefficient
1.0
0.5
0.0
-0.5
-1.0
0
20
40
60
80
100
time


Rio Negro Stage 1903-1986
stage (m)
10
15
20
25
30
1900
1920
1940
1960
1980
2000
year


Running Mean of 8 Years
stage (m)
10
15
20
25
30
1900
1920
1940
1960
1980
2000
year


Periodogram
periodogram
100000
10000
1000
100
10
1
0.1
0.0
0.2
0.4
0.6
0.8
1.0
1.2
frequency cycles/year

TECHNICAL REPORTS

Statistics Department

University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.

2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosciences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.

3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.

4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11. No. 1, 185-214.

5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.

6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.

7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.

8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.

9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.

10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.

11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.

12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.

14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.

15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.

16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.

17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review.

18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.

19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).

20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting, 1985, Vol. 4, 251-262.

21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.

22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.

23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics. (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984.

25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.

26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.

27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference. (L. Le Cam, ed.) Wadsworth, 1985.

29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.

30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES. 1985 Vol 3 pp. 1-13.

31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.

32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.

33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.

34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.

35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science. Feb 1986, Vol. 1, No. 1, 3-39.

36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.

37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly. Feb 1986, Vol. 93, No. 2, 123-125.

38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.

39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.

40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.

41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.

42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.

43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?

44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.

45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.

46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.

47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data.

48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.

49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.

50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada. January, 1986.

51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.

53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.

54. BLACKWELL, D. (November 1985). Approximate normality of large products.

55. FREEDMAN, D. A. (December 1985, revised Dec. 1986). As others see us: A case study in path analysis. Prepared for the Journal of Educational Statistics.

56. LE CAM, L. and YANG, G. L. (January 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies.

57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.

58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.

59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.

60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.

61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.

62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.

64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.

65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.

66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data.

68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).

69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.

72. LEHMANN, E.L. (July 1986). Statistics - an overview.

73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.

74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.

75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.

76. O'SULLIVAN, F. (September 1986). Relative risk estimation.

77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.

78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.

79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks.

80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.

81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.

82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.

83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.

85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory.

86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.

87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.

88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.

89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.

91. DIACONIS, P and FREEDMAN, D.A. (March 1987). A finite version of de Finetti's theorem for exponential families, with uniform asymptotic estimates.

92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.

92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.

93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.

94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.

95. RIZZARDI, F. (Aug 1987). Two-Sample t-tests where one population SD is known.

96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in *Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.*

97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for cancer.

98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.

99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.

100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.

101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators.

102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.

103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.

104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer.

105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.

106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.

107. CHENG, C-S. (August 1987). Some orthogonal main-effect plans for asymmetrical factorials.

108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.

109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \le k \le n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.

110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.

113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.

114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.

115. BICKEL, P.J. and RITOV, Y. (September 1987). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (September 1987). Unachievable information bounds in non and semiparametric models.

117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.

118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.

119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.

120. DONOHO, D. (October 1987). Geometrizing rates of convergence, II.

121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. To appear in *Statistics a Guide to the Unknown*.

122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.

123. DONOHO, D.L. and MACGIBBON, B. (November 1987). Minimax risk for hyperrectangles.

124. ALDOUS, D. (November 1987). Stopping times and tightness II.

125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.

126. DALANG, R.C. (December 1987). Optimal stopping of two-parameter processes on hyperfinite probability spaces.

127. Same as No. 133.

128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.

129. SMITH, D.L. (December 1987). Vapnik-Červonenkis classes.

130. STONE, C.J. (November 1987). Uniform error bounds involving logspline models.

131. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.

132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes

133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.

134. DUBINS, L.E. and SCHWARZ, G. (December 1987). A sharp inequality for martingales and stopping-times.

135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.

136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.

137. DIACONIS, P. and FREEDMAN, D.A. (January 1988). On the consistency of Bayes estimates.

138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.

139. BERAN, R. (January 1988). Refining simultaneous confidence sets.

140. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.

141. BRILLINGER, D.R. (January 1988). Two reports on trend analysis: a) An Elementary Trend Analysis of Rio Negro Levels at Manaus, 1903-1985 b) Consistent Detection of a Monotonic Trend Superposed on a Stationary Time Series

142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (February 1988). Randomization in the two-armed bandit problem.

144. DABROWSKA, D.M. and DOKSUM, K.A. (February 1988). Graphical comparisons of cumulative hazards for two populations.

145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.

146. BICKEL, P.J. and RITOV, Y. (February 1988). Estimating integrated squared density derivatives.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.