Inference for Time Series with Mixed Spectrum

by

Shean-Tsong Chiu[1]

Technical Report 87-8, May 1987

[1]Mathematical Sciences Department, Rice University, Houston, Texas 77251.

Inference for Time Series with Mixed Spectrum

By Shean-Tsong Chiu

Rice University, USA

SUMMARY

The old and important problem of estimating the discontinuous (mixed) spectrum of a series containing periodic components was considered in this paper. Most nonparametric spectral estimation procedures were developed for estimating smooth spectra and did not provide satisfactory results in estimating mixed spectra. A nonparametric estimation procedure was proposed for estimating discontinuous spectra. The procedure first finds a robust filter, which is insensitive to the presence of periodic components, to prewhiten the noise series and then uses a feature preserving smoother on the residual periodograms to estimate the discontinuous spectrum of the filtered series. The procedure was applied to some simulated data, and the results were compared with the classical kernel estimates and the autoregressive spectral estimates. The proposed procedure performs much better than the classical methods in estimating mixed spectra. The proposed procedure was also applied to three real data sets, including the famous Canadian lynx data. The proposed procedure was extended to estimate high-dimensional spectra. The problem of testing the significance of periodic components was discussed, and a testing procedure was also suggested.

KEYWORDS: MIXED SPECTRUM; PERIODIC COMPONENTS; PERIODOGRAM; KERNEL SPECTRAL ESTIMATE; FEATURE PRESERVING SPECTRAL ESTIMATE; ROBUST PREWHITEN; AUTOREGRESSIVE SPECTRAL ESTIMATE, AKAIKE'S CRITERION; SPATIAL-TEMPORAL PROCESSES; FREQUENCY-WAVENUMBER SPECTRUM; CANADIAN LYNX TRAPPINGS; IMAGE PROCESSING.

1. Introduction

When analyzing a time series, one interesting and important question often asked is that whether the series contains periodic components. In order to answer this question, we need a procedure to estimate the discontinuous (mixed) spectrum of a series containing periodic components. We also need a procedure to test the significance of the periodic components. The problem of estimating a mixed spectrum has been studied extensively, and several estimation procedures have been proposed. Most of these methods try to use a continuous function to approximate a discontinuous spectrum and do not provide satisfactory results.

In this paper, we are interested in a series which might contain periodic components. The series of interest can be modeled as

$$X(t) = S(t) + \epsilon(t), \quad t=1, \ldots, T,$$

where $\epsilon(t)$ is a stationary series which satisfies the mixed condition of Brillinger (1975) (Assumption 2.6.1) that all moments of $\epsilon(t)$ exist and

$$\sum_{u_1, \ldots, u_{k-1}=-\infty}^{\infty} | c_{\epsilon,k}(u_1, \ldots, u_{k-1}) | < \infty, \tag{1}$$

where $c_{\epsilon,k}(u_1, \ldots, u_{k-1})$ is the kth order cummulant function of $\epsilon(t)$. Under this assumption, the series $\epsilon(t)$ has a bounded and uniformly continuous power spectrum

$$f_\epsilon(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} c_\epsilon(u)\exp(-i\lambda u),$$

where $c_\epsilon(u) = c_{\epsilon,2}(u)$ is the autocovariance function of $\epsilon(t)$. In the following discussion, we should call $\epsilon(t)$ and $f_\epsilon(\lambda)$ the noise series and the noise spectrum, respectively. Without loss of generality, we assume $\epsilon(t)$ has mean zero. The function $S(t)$ consists of periodic components,

$$S(t) = \sum_{i=1}^{K} R_i(t)\cos\{\omega_i t+\phi_i(t)\},$$

where K is the number of periodic components and $\omega_i \neq 0$ is the frequency of the ith component. The amplitudes $R_i(t)$ and the phase $\phi_i(t)$ are functions that change slowly over time t. We should allow K to be zero to include the case that $S(t) \equiv 0$. We call a periodic component pure when the amplitude and phase functions are constant, and call it modulated otherwise.

For a series which contains pure periodic components, it is possible to rigorously define the spectrum of $X(t)$ by assuming that the phases are independent random variables distributed uniformly over $(-\pi, \pi)$. Then $X(t)$ has a discontinuous (mixed) spectrum which is equal to the noise spectrum plus delta functions at the frequencies of the periodic components. However, we would prefer to view the series $X(t)$ as a non-stationary series with mean function $S(t)$. The spectrum of the series $X(t)$ is therefore not rigorously defined. In this paper, the word "spectrum" is used loosely; the spectrum of the series $X(t)$ means that

$$f_X(\lambda) = I_S(\lambda) + f_\epsilon(\lambda),$$

where

$$I_S(\lambda) = |d_S(\lambda)|^2/(2\pi T)$$

is the periodogram of the series $S(t)$ at frequency λ and

$$d_S(\lambda) = \sum_{t=1}^{T} S(t)\exp(-i\lambda t)$$

is the Fourier transform of $S(t)$. The periodograms $I_X(\lambda)$ and $I_\epsilon(\lambda)$ of the series $X(t)$ and $\epsilon(t)$ are defined similarly. The Fourier transform and the periodogram of a series are usually evaluated at the Fourier frequencies $\lambda_j=2\pi j/T$, $j=0,1,...$, by the Fast Fourier Transform [Cooley and Tukey (1965)].

According to the definition, the spectrum $f_X(\lambda)$ has substantial amplitude at frequencies close to ω_i's. Therefore, from the spectrum, one can obtain important information about the number, the amplitudes and the approximate frequencies of the periodic components. In practice, the spectrum is usually unknown and has to be estimated from the observed series.

One approach in studying series with periodic components is to estimate the periodic components first and then estimate the noise spectrum from the residual series. For references of this approach see Hannan (1973), Bolt and Brillinger (1979), Hasan (1982), Whittle (1952) and Walker (1973). These methods usually require some prior information which can be obtained from a nonparametric spectral estimate.

We briefly review two classical nonparametric estimates, the kernel and the autoregressive spectral estimates, in the next section. The classical methods are originally developed for estimating continuous spectra and have fundamental difficulties in estimating mixed spectra. We propose a procedure to estimate discontinuous spectra. The detail of the procedure is given in Sections 3 and 4. In Section 5 the proposed procedure and the two classical methods are applied to some simulated series. The proposed procedure performs much better than the classical methods in estimating mixed spectra. The problem of testing the significance of periodic components is discussed in Section 7. We also apply the proposed procedure to estimate the spectra of three sets of real data, including the famous Canadian lynx data. The results are given in Section 8. In Section 9 the proposed procedure is extended to estimate high-dimensional spectra.

2. Two Classical Spectral Estimators

Among nonparametric estimators, the kernel method is the one used most often. The procedure estimates $f_X(\lambda)$ by a weighted average of the periodograms at the Fourier frequencies close to λ. Readers are referred to Brillinger (1975) for detail information about the properties of the kernel estimator.

The motivation of using the kernel method is based on the the fact that, when the series satisfies the mixed condition of (1), the periodograms $I_X(\lambda_j)$ in a small window are approximately independent and identically distributed. However this result does not hold when the window contains peaks. The procedure will smooth the peaks, and overestimate the spectrum at frequencies close to the peaks. The biases can be reduced by using a smaller window, but this will increase the variance of the estimate and produce a very erratic spectrum. To determine a proper bandwidth is usually a difficult job.

Another procedure which attracts a lot of attention from engineers is the maximum entropy method proposed in Burg (1967). Lacoss (1971) showed that this method is essentially equivalent to the autoregressive spectral estimate of Parzen (1969). Parzen (1983) gave a review on the autoregressive spectral estimate. The autoregressive procedure were generalized to the autoregressive-moving average spectral estimate in Gray and Woodward (1986) and Morton and Gray (1984).

The procedure fits the series with an autoregressive process, and estimates the spectrum by

$$\hat{f}_{AR}(\lambda) = (2\pi)^{-1}\hat{\sigma}^2 / |A(\lambda)|^2,$$

where $\hat{\sigma}^2$ is the estimated residual variance and

$$A(\lambda) = 1 + \sum_{u=1}^{m} \hat{\phi}(u)\exp(-i\lambda u)$$

is the transfer function of the filter with the estimated coefficients $\hat{\phi}(u)$'s, and m is the order of the fitted autregressive process. The problem of determining the order is similar to the problem of selecting the bandwidth for the kernel estimate. Several criteria have been proposed for selecting the order [Akaike (1974), Parzen (1974, 1977) and Hannan and Quinn (1979)]. The most widely used criterion is the Akaike's criterion [Akaike (1974)] which is defined by

$$\text{AIC}(m) = \log\hat{\sigma}^2 + (2m)/T, \quad m \geq 1 \ .$$

When the series contains strong periodic components, the Akaike's criterion tends to choose a high order autoregressive process which will produce a lot of false peaks. On the other hand, when the series contains weak peaks, the Akaike's criterion tends to choose a small order, and the spectral estimate will fail to show the peaks. Therefore it is not proper to use the Akaike's criterion to choose the order in estimating mixed spectra, and one still has the problem of selecting a proper order for the autoregressive process. The other criteria have similar problems.

Since the autoregressive procedure usually gives a smoother (but not more accurate) estimate and makes it easier to locate the peaks, the autoregressive spectral estimate is often called "high-resolution" method by engineers. The autoregressive spectral estimator has been widely used in the engineering community for estimating mixed spectra.

From the discussion above, we see that these two procedures have fundamental difficulties. The kernel estimate requires the assumption that the spectrum is a smooth function; but a mixed spectrum is not smooth at the peaks. The autoregressive procedure tries to approximate a discontinuous function with a continuous function. We should note that both procedures are originally designed for estimating smooth spectra. We use a different approach to estimate mixed spectra. The basic idea is to use discontinuous functions to estimate discontinuous spectra, thus avoiding the fundamental difficulties of the classical methods.

3. The Feature Preserving Smoother

The proposed procedure estimates the spectrum $f_X(\lambda)$ from the periodograms, $I_X(\lambda_j)$, at the Fourier frequencies $\lambda_j = (2\pi j)/T$, $j=1, \ldots, [(T-1)/2]$. We observe that, when the series contains periodic components, $2I_X(\lambda)/f_\epsilon(\lambda)$ has approximately a noncentral chi-square distribution with 2 degrees of freedom and the noncentrality parameter $\{2I_S(\lambda)/f_\epsilon(\lambda)\}$. Since $I_S(\lambda)$ has substantial amplitude only at frequencies around the peaks, the periodograms $I_X(\lambda_j)$ can be viewed, approximately, as a sequence of independent and exponentially distributed random variables which contains some outliers (peaks). Except at the peaks, the mean function $\mu(j)$ of the sequence $I_X(\lambda_j)$ is a smooth function which is approximately equal to $f_\epsilon(\lambda_j)$.

Chiu (1987) considered the problem of estimating the mean function of a sequence of independent random variables, and proposed a feature preserving smoothing procedure for the situations in which the mean function might change abruptly. We apply the feature preserving procedure to smooth the periodograms and obtain an estimate of the mixed spectrum. We briefly describe the procedure here; readers are referred to Chiu (1987) for more detail.

Suppose that we have a sequence of independent observations $Y(1), \ldots, Y(N)$ with mean $\mu(1), \ldots, \mu(N)$, and we are interested in estimating the mean function $\mu(t)$. We assume that the mean function changes rather slowly except at some abrupt jumps. We estimate $\mu(t_0)$ from the observations in a window with length l. The window is usually centered at t_0. Intuitively, if we knew the locations of the abrupt changes, we would only use the observations in the same region of t_0 to estimate $\mu(t_0)$. Though, in practice, the locations of the abrupt changes are usually unknown, we can classify the window into different regions, and then estimate $\mu(t_0)$ by a weighted average over the points in the same region of t_0. The basic idea of the classification procedure is to merge the neighboring regions with minimum "distance" until there are only a few regions. The merger process is terminated when the distances between neighboring regions are too large. We now describe the procedure.

Step 1. At the beginning, each point is a region.

Step 2. The distances between neighboring regions are computed. The distance between two neighboring regions R_1 and R_2 is defined as

$$d = |\overline{Y}_1 - \overline{Y}_2| / \{(1/n_1) + (1/n_2)\}^{1/2},$$

where $\overline{Y}_k$ is the sample mean of the region R_k and n_k is the number of points in R_k.

Step 3. Find the minimum distance $d_{\min}$ among all neighboring regions.

Step 4. If the number of regions is bigger than L, go to Step 6, otherwise go to Step 5. The number L is chosen according to the complexity of the mean function. We let $L=5$ in this paper.

Step 5. When there are few regions, we compute the statistic $D_K = d_{\min}/s$, where s^2 is the mean squares within regions,

$$s^2 = \sum_{k=1}^{K} \left[\sum_{t=t_k}^{t_{k+1}-1} \left\{ Y(t) - \overline{Y}_k \right\}^2 / (T-K) \right],$$

K is the number of regions, and the kth region $R_k = \{t_k, \ldots, t_{k+1}-1\}$. We note that D_K is scale and location invariant. The statistic D_K tends to be large when the means of the regions are different. So if D_K is bigger than c_K, we stop merging regions and go to Step 7, otherwise go to Step 6. The critical values c_K are determined according to how much chance of finding a false discontinuity one is willing to tolerate.

Step 6. The neighboring regions with minimum distance are merged. If there is only one region now, go to Step 7, otherwise go to Step 2 and repeat the merger process.

Step 7. Estimate $\mu(t_0)$ from the observations which are in the same region of t_0.

Though the value of c_K can be set by trial-and-error, we would prefer to set the value as the upper α point of the null distribution of D_K. Due to the complexity of the procedure, it is difficult to obtain the exact null distribution. However, we could find approximate values by Monte Carlo methods. We note that the periodograms of a series with smooth spectrum are asymptotically exponentially distributed.

This smoothing procedure requires that the mean function changes slowly over t except at the abrupt changes. The procedure might find discontinuities for processes with rapidly changed (but continuous) mean functions. In practice, the spectrum of a series usually changes quite rapidly. Therefore, the feature preserving smoother might produce some undesired discontinuities. In order to solve this problem, we propose a robust prewhitening procedure in the next section. The procedure finds a filter such that the filtered series has a near constant noise spectrum.

4. A Robust Prewhitening Procedure

As discussed in Section 3, the kernel estimate might have substantial bias when the assumption of constant spectrum is violated. Prewhitening is often recommended for reducing the bias. The idea was proposed in Press and Tukey (1956). The objective of prewhitening is to filter a series such that the filtered series has a near constant spectrum. The assumption of constant spectrum is thereby better approximated. The spectrum of the original series is estimated by

$$\hat{f}(\lambda) = |A(\lambda)|^{-2}\hat{f}_R(\lambda),$$

where $A(\lambda)$ is the transfer function of the filter and $\hat{f}_R(\lambda)$ is the kernel estimate of the spectrum of the filtered (residual) series. We could obtain a near constant residual spectrum $f_R(\lambda)$ if we could find a filter with $|A(\lambda)|^{-2}$ close to $f_X(\lambda)$. From this, we see that $|A(\lambda)|^{-2}$ can be viewed as a rough estimate of the spectrum. We note that it is not required to actually filter the data to obtain the residual periodograms [Brillinger (1975), p 159]. The residual periodogram is approximately equal to the original periodogram times $|A(\lambda)|^2$. One general procedure of determining a filter is by fitting an autoregressive scheme to the data.

For series with mixed spectrum, it is difficult (if not impossible) to obtain a near constant residual spectrum. The rough estimate, $|A(\lambda)|^{-2}$, has serious biases caused by the presence of the peaks. It usually overestimates the spectrum at the frequencies close to the peaks.

Our objective is different from the one of the classical prewhitening. We would like to find a filter with $|A(\lambda)|^{-2}$ close to the continuous noise spectrum $f_\epsilon(\lambda)$ [not $f_X(\lambda)$], thus the residual noise spectrum will be near constant.

Since the periodogram $I_X(\lambda)$ has substantial amplitude only at the frequencies close to the peaks, the periodograms at these frequencies can therefore be viewed as outliers. Any procedure which can provide a "robust" estimate that is insensitive to these outliers will achieve our goal. We remark that the purpose and the meaning of this "robust" filter is different from the robust filter studied in Kleiner *et al.* (1979) and Martin and Yohai (1986). We now describe in detail such a procedure.

Step 1. Fit the data with an autoregressive scheme of order m and compute $\tilde{f}(\lambda)=\hat{\sigma}^2|A(\lambda)|^{-2}$, where $\hat{\sigma}^2$ is the estimated residual variance. The order $m=4$ or 5 will be sufficient for most applications. The selection of the order will be discussed later.

Step 2. Define the modified periodogram at the Fourier frequencies λ_j, $j=1, \ldots, T-1$, as

$$\tilde{I}(\lambda_i) = \begin{cases} I(\lambda_i) & I(\lambda_i)/\tilde{f}(\lambda_i) < c \\ c\tilde{f}(\lambda_i) & I(\lambda_i)/\tilde{f}(\lambda_i) \geq c \end{cases},$$

where $c = -\log(\alpha)$, α determines the portion of periodograms to be modified. The procedure is not very sensitive to the value of α. We let $\alpha=0.05$ in this paper.

Step 3. Compute the modified autocovariance function $\tilde{c}(u)$ from the modified periodograms by the inverse Fourier transform.

Step 4. From the modified autocovariance function, find the coefficients of the autoregressive processes, and compute $\tilde{f}(\lambda) = (2\pi)^{-1}\tilde{\sigma}^2/|A(\lambda)|^2$, where $\tilde{\sigma}^2 = \tilde{c}(0)/(1-\alpha)$ is a robust estimate for the residual variance.

Step 5. Repeat the iteration from Steps 2 to 4 until the estimates of the coefficients converge. In our experience, the differences between the consecutive estimates are usually smaller than 0.01 after 5 iterations.

Step 6. Obtain a rough estimate $\hat{f}_{PW}(\lambda)=\tilde{f}(\lambda)$ for the noise spectrum $f_\epsilon(\lambda)$.

We will show in the Appendix that, when $X(t)$ is an autoregressive Gaussian process, the robust procedure provides a consistent estimate for the parameters and the spectrum. We apply the feature preserving smoother to the approximate residual periodograms $I_R(\lambda_j) = I_X(\lambda_j)/\hat{f}_{PW}(\lambda_j)$, and obtain an estimate $\hat{f}_R(\lambda)$ for the residual spectrum. The final estimate of $f_X(\lambda)$ is $\hat{f}_{FP}(\lambda) = \hat{f}_R(\lambda)\hat{f}_{PW}(\lambda)$.

5. Some Simulated Examples

We apply the proposed procedure to five examples and compare the results with the classical kernel and the autoregressive spectral estimates. We first describe the examples.

Example 1, Series with a continuous spectrum. A realization of 200 points of the process

$$X(t) = Z_1(t) + Z_2(t),$$

where

$$Z_1(t) = 0.3Z_1(t-1) + 0.83Z_1(t) + \epsilon_1(t)$$

and

$$Z_2(t) = -0.9Z_2(t-1) + 0.83Z_2(t) + \epsilon_2(t)$$

are two second order autoregressive processes, and $\epsilon_1(t)$ and $\epsilon_2(t)$ are two independent white Gaussian series with mean zero and variance 1. The spectrum of $X(t)$ has two modes around frequencies 0.17 and 0.28.

Example 2, Series with a strong periodic component. A realization of 200 points of the process

$$X(t) = (20)^{1/2}\cos\{2\pi(0.201)t\} + Z(t),$$

where

$$Z(t) = 0.92Z(t-1) + \epsilon(t)$$

is a first order autoregressive process and $\epsilon(t)$ is a white Gaussian series with mean zero and variance 4.

Example 3, Series with a moderate periodic component. A realization of 200 points of the process

$$X(t) = 1.2\cos\{2\pi(0.2t)\} + Z(t),$$

where $Z(t)$ is the same autoregressive process (but different realization) as the one in Example 2.

Example 4. Series with two periodic components at close frequencies. A realization of 200 points of the process

$$X(t) = (20)^{1/2}\cos\{2\pi(0.2t)\} + (3)^{1/2}\cos\{2\pi(0.215t)\} + Z(t),$$

where $Z(t)$ is a first order autoregressive process as defined in (). The variance of the white series is 1.6. This example is similar to Example 3 of Gray and Woodward (1986). The amplitude of the noise spectrum at frequency 0.2 is about $1/(2\pi)$, which is the amplitude of the spectrum of the white noise series used in Gray and Woodward (1986). All methods considered in Gray and Woodward (1986) fail to separate the two close peaks at frequencies 0.2 and 0.215. Except for the proposed procedure, we do not know of any other nonparametric estimate that can separate these two peaks.

For each example, The resultant estimates for Examples 1 to 4 are shown in Figures 1 to 4, respectively. In each figure the periodograms are shown in plot A, the feature preserving, the classical kernel and the autoregressive spectral estimates are shown in plots B, C and D, respectively. For the feature preserving estimates, the window length is 21, L=5 and the critical values used are c_2=5.99, c_3=8.62, c_4=4.97 and c_5=6.06. These are the approximate 99th percentiles of the null distribution of the statistics D_2 to D_5, respectively. These values are obtained from the empirical distributions of 4000 simulations. The weighting function is proportional to a triangular kernel function $w(t) = 11-|t|$, $|t| \leq 10$. The value of α used in finding the robust filters is 0.05. The orders of the robust filters are 4 for Example 1, and 2 for Examples 2 to 4.

The classical kernel estimates use the same window length and weighting function as those used in the feature-preserving estimates. The orders of the prewhitening filters are 4 for all examples. The orders of the autoregressive spectral estimates in Examples 1 to 4 are 7, 17, 3 and 6, respectively, which are selected according to Akaike's criterion. The coefficients of the autoregressive processes are obtained from the Yule-Walker equation.

In the figures, the solid lines are the estimates and the dashed lines are the true noise spectra. The spectra are plotted in log scale. The heights of the vertical lines in Figures 2 and 3 indicate the amplitudes of the periodograms $I_S(\lambda)$ at frequencies 0.201 and 0.2, respectively. The two short vertical lines in Figure 4 show the locations of the peaks.

From Example 1, we see that the feature preserving procedure and the classical kernel procedure give similar estimates of the continuous spectrum. In Figures 2 to 4, it can be clearly seen that the feature preserving procedure gives much better estimates of mixed spectra than the two classical procedures do. The proposed procedure not only gives good estimates for the noise spectra but also accurately estimates the heights and the locations of the peaks.

The AR(3) spectral estimate in Example 3 fails to show the moderate peak. On the other hand, the AR(17) spectral estimate in Example 2 produces a lot of false peaks. These phenomena are consistent with the remark in Section 2 about the behavior of Akaike's criterion.

6. Select the Order for the Robust Prewhitening Filter

Though we feel that one can pick a proper order by trial-and-error in practice, it certainly would be helpful to have some guidance. We can modify Akaike's criterion and use it to choose the order. We defined the robust Akaike's criterion,

$$\mathrm{RAIC}(m) = \log\tilde{\sigma}^2 + (2m)/T, \quad m \geq 1,$$

where m is the order of the robust filter and $\tilde{\sigma}^2$ is the robust estimate of the residual variance, which was defined in Section 4. Figures 5.A to 5.D show the values of Akaike's criterion (denoted by *) and the robust Akaike's criterion (denoted by o) for Examples 1 to 4, respectively. The robust Akaike's criterion chooses order 4 for Example 1, and chooses order 1 (the correct order of the noise series) for Examples 2 to 4. From the figures, we observed that, for a series with mixed spectrum, the value of $\tilde{\sigma}^2$ might sharply increase as the order of the robust filter increases. This is an interesting and important feature of the robust prewhitening procedure. When the prewhitening procedure becomes seriously affected by the peaks, the robust estimate of the residual variance will sharply increase. Therefore, the order used should be smaller than the order at which the sharp increase occurs. According to the robust criterion, the orders of the prewhitening filters should not be bigger than 5 for Example 2, and 3 for Example 4.

When the series is not properly prewhitened, the feature preserving procedure might produce some false discontinuities. So, the proposed estimate can also be used as a diagnostic tool for checking whether the series is properly prewhitened. When the series is properly prewhitened, the "shapes" of the estimate $\hat{f}_{FP}(\lambda)$ and the rough estimate $\hat{f}_{PW}(\lambda)$ should be similar. To be more precise, the second order derivative of $\hat{f}_{PW}(\lambda)$ should be close to the second order derivative of $\hat{f}_{FP}(\lambda)$.

In order to illustrate how to use the feature preserving estimate as a diagnostic tool, we use a second order robust filter to prewhiten the series of Example 1, and plot $\hat{f}_{FP}(\lambda)$ (the solid line) and $\hat{f}_{PW}(\lambda)$ (the dashed line) in Figures 6. It can be seen that the two estimates are very different around the frequencies 0.1 and 0.2, where the estimate $\hat{f}_{FP}(\lambda)$ is discontinuous. This suggests that the second order robust filter does not properly prewhiten the series.

7. Test the Significance of Periodic Components

After finding possible peaks in the periodograms, one is often interested in determining whether the peaks are due to the periodic components. In this section, we assume that the noise is a Gaussian series. Fisher (1929) proposed a procedure for testing periodic components in a white series. The test is based on the statistic

$$F_T = \max_{1 \le j \le n} I_X(\lambda_j) / \sum_{j=1}^{n} I_X(\lambda_j),$$

where $n = [(T-1)/2]$, T is the length of the series. Fisher also gave the exact null distribution of F_T. Bloomfield (1976) derived the asymptotic null distribution of the statistic $Z_T = nF_T - \log n$, which is

$$\mathrm{pr}(Z_T < z) \approx \exp\{\exp(-z)\}.$$

Some studies related to this problem can be found in Hartley (1949), Whittle (1954), Norwroozi (1965), (1966), Shimshoni (1971), Lewis and Fieller (1979), Likes (1966), Kabe (1970), Siegel (1980), Priestley (1981) and Chiu (1986).

The problem becomes much more complicates for the case of a non-white noise series. In order to apply the Fisher's test, one must normalize the periodogram by the noise spectrum. However the noise spectrum is usually unknown in practice. Whittle (1954) suggested fitting an autoregressive process to the series, and then using Fisher's test to the periodograms normalized by the estimated autoregressive spectrum. Whittle also showed that, under the hypothesis that the noise series is an autoregressive Gaussian process, the test statistic has the same asymptotic null distribution as that for the Fisher's test. This approach has several drawbacks. As pointed out by Hannan (1961) (also see Example 3 in Section 6), the estimate around the frequencies of the peaks might be inflated by the peaks, and this will substantially reduce the power of the test. There is another more serious problem that the null hypothesis might be rejected due to the inadequacy of the autoregressive model instead of the presence of peaks. While we see no easy way to solve the second problem, the first problem might be solved by using a robust procedure, such as the one proposed in Section 5, to estimate the coefficients of the autoregressive process.

Hannon (1961) considered a kernel estimate which is not affected much by the periodogram being tested. The basic idea is to exclude the periodogram at frequency λ_j when one estimates the noise spectrum $f_\epsilon(\lambda_j)$. The kernel used by Hannon has zero weight at the center. Hannon applied Fisher's test to the periodograms normalized by the estimated spectrum and showed that the test statistic has asymptotically the same null distribution as that for Fisher's test. The estimation procedure was generalized and discussed in Bartlett (1967), Brillinger and Rosenblatt (1967), and Priestley (1964) and (1981). This test might fail to detect a weak peak which is close to a strong one.

In order to use the approximate distribution, one has to check whether the following two conditions are reasonable. (a) The noise spectrum is constant in the window. (b) The number of the periodograms in the window is big enough, such that the error introduced by substituting the estimated spectrum for the true one is negligible. The first condition would be approximately satisfied when one uses a small window, but this will decrease the number of the periodograms in the window. The condition (a) can be better approximated if we first apply the robust prewhitening filter to the series.

There are two basic problems in testing the significance of the peaks: (1) find an estimate for the noise spectrum which is insensitive to the presence of peaks, (2) find an approximate null distribution of the test statistic. The feature preserving procedure provides a solution for the first problem. For the second problem, we consider the statistics

$$F(\lambda_j) = I_R(\lambda_j)/\{\sum_{1 \le |i-j| \le l} I_R(\lambda_i)/(2l)\} \tag{2}$$

where $I_R(\lambda) = I_X(\lambda)/\hat{f}_{PW}(\lambda)$ is the approximate residual periodogram. The denominator in (2) is similar to the Hannan's estimate. Therefore, following the same argument in Hannan (1961), it can be shown that the maximum of $F(\lambda_j)$ has the same asymptotic null distribution as that for the Fisher's test. However, in practice, the condition (b) might not be able to satisfied, and we would like to consider another approximation.

We will show in the Appendix that, under the assumptions that the series is a Gaussian autoregressive process, the distribution of the maximum of $F(\lambda_j)$ is approximately equal to the distribution of the maximum of a sequence of independent F random variables with 2 and $2l$ degrees of freedom. Hence, we have

$$\mathrm{pr}\{\max_{1 \le j \le n} F(\lambda_j) < z\} \approx \{F_{2,2l}(z)\}^n .$$

When we use this test procedure in practice, we have to check whether $\hat{f}_{FP}(\lambda)$ and $\hat{f}_{PW}(\lambda)$ have similar shapes around the frequencies of the peaks. There are two other tests, the grouped periodogram test and the P(λ) test, proposed in Priestley (1962a) and (1962b).

8. Application to Some Real Examples

We now apply the feature preserving estimation procedure to to three sets of data; the variable-star data, the variation in the Earth's rotation rate and the Canadian lynx data. Except for the Canadian lynx data, the window length, the critical values and the value of L used by the smoothing procedure are the same as the values used in the simulated examples. Except for the variable-star data, a second order robust filter is used to prewhiten the series; the value of α is 0.05.

Data 1. Variable-star. The data set consists of the magnitudes of a variable star at midnight on 600 successive nights [from Whittaker and Robinson (1924), p. 349]. The 301th observation is corrected according to Bloomfield (1976), p. 91. This series contains two pure periodic components at close frequencies, and there is almost no noise in the series besides the rounded off errors [Bloomfield (1976)]. We apply the feature preserving smoother to the original periodograms to obtain the estimate. Figure 7 shows the periodograms and the estimated spectrum. The spectrum clearly shows the two major peaks. Besides the major peaks, the spectrum also shows several minor peaks. These minor peaks are the harmonics of the two major periodic components. The harmonics are caused by the roundoff errors. We note that the estimated spectrum is biased at frequencies below 0.03. This problem might be corrected by using a bigger L.

Data 2. Variations in the Earth's rotation rate. The rate of the Earth's rotation is not steady. In general the rate gradually slows down, but there are a lot of variations from year to year. The data set consists of the changes of day length from 1821 to 1970. This data set was published in Luo *et al.* (1977), and is also available in Wilson (1985). The series length is 150 and the unit is 0.00001 of a second. This series has a moderate linear trend. Based on the detrended series, Luo *et al.* (1977) claimed to detect 12 different periodicities, and tried to match these periods with the periods of the moon, the planets and the sunspots. The periodogram and estimated spectrum of the detrended series are plotted in Figure 8. We do not find any peak from the estimated spectrum.

Data 3. Canadian Lynx Data. The third data set is the annual lynx trappings in the MacKenzie River District of Northwest Canada for 114 consecutive years from 1821-1934. This data set has been analyzed extensively. For some references see Campbell and Walker (1977), Tong (1977), Priestley (1981), Lim and Tong (1983), Haggan *et al.* (1984) and Haggan and Oyetunji (1984). In order to compare our result with some available results, we use the transformed data $X(t)=\log_{10} Y(t)$, where $Y(t)$ is the original series.

The window length used by the smoothing procedure is 13, and the critical values are

c_2=7.03, c_3=9.98, c_4=5.98 and c_6=6.27. The critical values are the approximate 99th percentiles. The solid line in Figure 9 is the feature preserving estimate and the dashed line is the spectrum of the eleventh order autoregressive obtained in Tong (1977). The proposed procedure finds a peak at frequency 12/114, which is sharper than the peak of the AR(11) spectral estimate. Beside the peak, the proposed estimate is smoother than the autoregressive spectral estimate. We find the statistic $F(\lambda_{12})$=33.187 (l=4), and the approximate p-value is $1-F_{2,16}(33.187)^{56} \approx 0.00011$. We note that the estimates show two moderate modes around 21/114 and 33/114. We find $F(\lambda_{21})$=1.531 and $F(\lambda_{33})$=3.402, and $F_{2,16}(1.531)$=0.754 and $F_{2,16}(3.402)$=0.941. So these two peaks are far from being significant. We remark that these two modes are not located at the harmonics of the fundamental frequency. The autoregressive spectral estimate also shows a peak at frequency 45/114. However, based on the periodograms at the frequencies around 45/114, we do not see any evidence supporting the existence of a mode around that frequency. We have observed a similar phenomenon in Example 2 of Section 5, where a high order autoregressive spectral estimate tends to produce false peaks.

Complex demodulation [Bingham, Godfrey and Tukey (1967)] is a useful tool in studying series containing modulated periodic components. For detail see Bloomfield (1976) and Hasan (1983). Complex demodulation computes the instantaneous amplitude and phase of the periodic component at frequency λ. When the periodic component has a constant phase, the instantaneous phase should be a linear function of time. Figure 10 shows the instantaneous phase of the series demodulated at frequency $\lambda = (2\pi)/9.63 = 11.838(2\pi/114)$, which is the frequency estimated by Campell and Walker (1977). The smoothing kernel used here is a raised cosine window of length 25. From the plot, it can be seen that the instantaneous frequency varies around 11.838/114 with a quite regular pattern. This is consistent with the observations of Galbraith (1977) and Akaike (1977), that the series descends faster than it ascends.

9. Estimating Wavenumber Spectra of Spatial-Temporal Processes

The feature preserving smoother can also be used to estimate high-dimensional spectra. The situation we consider here is that we have an array of sensors which receive signals propagating through the array. The sensors are located at r_i, $i=1, \ldots, I$, in a two-dimensional space. We should assume that the signal sources are far away from the array, thus the signals can be viewed as plane waves. The records of the sensors can be modeled as

$$X(r_i,t) = \sum_{j=1}^{J} S_j\{t-\tau_j(r_i)\} + \epsilon(r_i,t).$$

The sensors receive the common signals $S_j(t)$, but with different time delays $\tau_j(r_i)$, together with a stationary spatial-temporal noise process $\epsilon(r_i,t)$, which has a smooth frequency-wavenumber spectrum. The time delays can be written as $\tau_j(r_i) = r_i(\cos\omega_j,\sin\omega_j)/s_j$, where s_j and $-(\cos\omega_j,\sin\omega_j)$ are the speed and the direction of the jth propagating waves, respectively. The vector $\mathrm{k}_j = (\cos\omega_j,\sin\omega_j)/s_j$ is called the slowness vector of the jth signal. The Fourier transform of the ith sensor's record has the approximate relation

$$\mathrm{d}_X(r_i,\lambda) \approx \sum_{j=1}^{J} \mathrm{d}_j(\lambda)\exp\{-i\lambda\tau_j(r_i)\} + \mathrm{d}_\epsilon(r_i,\lambda), \tag{3}$$

where $\mathrm{d}_j(\lambda)$ is the Fourier transform of the jth signal $S_j(t)$. It can be seen from (3) that, at a fixed frequency λ, the spatial process $\mathrm{d}_X(r_i,\lambda)$ contains J periodic components. The process $X(r,t)$ has a mixed frequency-wavenumber spectrum with peaks at wavenumbers $\lambda\mathrm{k}_j$. We can estimate the number from the number of peaks in the spectrum and estimate the velocities of the propagating waves from the locations of the peaks.

In order to apply the feature preserving smoother to high-dimensional data, we need to give a proper definition of "neighboring". We say two regions R_1 and R_2 are neighbors if there exist $(m_1,n_1)\in R_1$ and $(m_2,n_2)\in R_2$ such that either $m_1=m_2$ and $|n_1-n_2|=1$, or $n_1=n_2$ and $|m_1-m_2|=1$. The rest of the procedure can then be applied without modification.

Now we consider a simple example. Suppose we have an array of sensors located on the lattice $\{(m,n),\ 1\leq m,n\leq 25\}$, and

$$\begin{aligned} X(m,n,t) &= 0.2\cos\{(0.1)2\pi t - 1.73m - n\} \\ &\quad + 0.15\cos\{(0.1)2\pi t + 1.01m - 1.01n\} + \epsilon(m,n,t) \quad t=1, \ldots, 300, \end{aligned}$$

is the record of the sensor at (m,n), where $\epsilon(m,n,t)$ is a white Gaussian spatial-temporal process with mean zero and variance 100. Figure 11.A is the perspective plot of the wavenumber periodogram at frequency $(0.1)2\pi$.

We use a 5 by 5 window to estimate the wavenumber spectrum at frequency $(0.1)2\pi$. The value of L is 5 and the critical values used are c_2=9.58, c_3=7.40, c_4=6.26 and c_5=5.71. The critical values are the approximate 99th percentiles of the statistics D_K's obtained by Monte Carlo methods. The weighting function is proportional to

$$w(i,j) = \{1-\cos(2\pi i/6)\}\{1-\cos(2\pi j/6)\}, \quad 1 \leq i,j \leq 5.$$

The resultant estimate is shown in Figure 11.B. Figure 11.C shows the classical kernel estimate, which uses the same weighting function and window size as the ones used by the feature preserving smoother. This example clearly demonstrates the advantage of the proposed procedure.

10. Some Other Application

The smoothing procedure proposed in this paper can be used to estimate the mean function of a process that might contain some abrupt changes. One important application of the feature preserving smoother is in the field of image processing. (In fact, to smooth a noisy image is the original motivation of developing the smoothing procedure.) Images usually contain some edges at which the gray level changes abruptly. The edges provide important information about the image. A lot of efforts have been spend in searching for smoothers which can produce sharp edges.

Estimating a discontinuous regression function is another application. The feature preserving procedure was applied to estimate the mean function of a sequence of independent random variables in Chiu (1987). The procedure can be extended to handle non-equally spaced data. Finally we remark that the procedure can also be used to detect outliers.

Acknowledgement

I would like to thank Professor David Brillinger for directing my interest into the problem. The generous supports from Professors James Thompson and David Scott are also gratefully appreciated. This research was supported in part by ARO Grant DAAG 29-85-K-0212 and ONR Grant N 0001-485-K-0100.

Appendix

We consider autoregressive processes

$$X(t) = a(1)X(t-1) + \cdots + a(m)X(t-m) + \epsilon(t)$$

where $\epsilon(t)$ is a white Gaussian series with mean zero and variance σ^2. The coefficients satisfy the condition that the roots of $A(z)=0$ lie outside the unit circle where

$$A(z)=1+a(1)z+\cdots+a(m)z^m.$$

Let $c(u)$ and $f(\lambda)$ denote the autocovariance function and the spectrum of $X(t)$, respectively. Also define the modified sample autocovariance,

$$\hat{c}(u) = (1-\alpha)^{-1}T^{-1}\sum_{i=0}^{T-1}\tilde{I}(\lambda_j)\exp(i\lambda_j u),\ u = 0,\pm1,\ldots, \tag{4}$$

where

$$\tilde{I}(\lambda_i) = \begin{cases} I(\lambda_i) & I(\lambda_i)/f(\lambda_i) < c \\ cf(\lambda_i) & I(\lambda_i)/f(\lambda_i) \geq c \end{cases},$$

$c = -\log(\alpha)$, $0<\alpha<1$, and $I(\lambda)$ is the periodogram of $X(t)$. Now we have

Theorem 1. Suppose that $X(t)$, $t=0,\ldots,T-1$ is an autoregressive Gaussian process which satisfies the assumption above, then $\hat{c}(u)$ converges to $c(u)$ almost surely for $u = 0,\pm1,\ldots$.

Proof: From the results in Section 3.8 of Brillinger (1975), $X(t)$ can be represented as a moving-average process,

$$X(t) = \sum_{u=0}^{\infty} b(t-u)\epsilon(u)$$

with

$$\sum_{u=0}^{\infty}(1+|u|^l)|b(u)| < \infty$$

for all $l \geq 0$. Application of Theorem 6.2.2 of Priestley (1981) yields

$$\tilde{I}(\lambda) = f(\lambda)\{\sigma^2/(2\pi)\}^{-1}\tilde{I}_\epsilon(\lambda) + R(\lambda),$$

where

$$\tilde{I}_\epsilon(\lambda) = \begin{cases} I_\epsilon(\lambda) & I_\epsilon(\lambda) < c\sigma^2/(2\pi) \\ c\sigma^2/(2\pi) & I_\epsilon(\lambda) \geq c\sigma^2/(2\pi) \end{cases},$$

and

$$\mathrm{E}[|R(\lambda)|^2] = O(1/T^2)$$

uniformly in λ. Therefore,

$$(1-\alpha)\hat{c}(u) = T^{-1}\sum_{j=0}^{T-1} f(\lambda_j)\tilde{I}_\epsilon(\lambda_j)\exp(i\lambda_j u) + o(1), \tag{5}$$

where the error term converges to zero uniformly and almost surely. Since $\{\sigma^2/(2\pi)\}^{-1}I_\epsilon(\lambda_j)$ are independent and exponentially distributed random variables with mean 1, the right hand side of (5) converges to

$$(1-\alpha)c(u) = \int_{-\pi}^{\pi}(1-\alpha)f(\lambda)\exp(i\lambda u)d\lambda$$

almost surely, and this finishes the proof.

In practice, the spectrum $f(\lambda)$ is usually unknown. However, if we substitute a consistent estimate $\hat{f}(\lambda)$ for $f(\lambda)$ in the definition of $\hat{c}(u)$ in (4), the result of the theorem still holds. Under the assumptions of Theorem 1, $f(\lambda,\hat{\theta})$ is a consistent estimate of $f(\lambda)$, where $f(\lambda,\theta)$ denotes the power spectrum of the process when θ is the true parameter, and $\hat{\theta}$ is a consistent estimate of θ obtained from the original (unmodified) autocovariance function. From the theorem, it can be easily seen that the estimates of the coefficients and the spectrum obtained from $\hat{c}(u)$ are consistent.

We next prove a theorem about the asymptotic distribution of the test statistic max $F(\lambda_j)$

Theorem 2. Under the assumptions of Theorem 1 and letting $n=[(T-1)/2]$, then

$$\mathrm{pr}\{\max_{1\leq j\leq n} F(\lambda_j)<z\} \approx \{F_{2,2l}(z)\}^n .$$

Proof: Following the same argument in Theorem 1,

$$F(\lambda_j) = F_\epsilon(\lambda_j) + o(1)$$

where

$$F_\epsilon(\lambda_j) = I_\epsilon(\lambda_j) / \sum_{1 \le |i-j| \le l} I_\epsilon(\lambda_i),$$

and the error term converges to zero uniformly and almost surely. Note that $I_\epsilon(\lambda_j)$, $j=1, \ldots, [(T-1)/2]$, is a sequence of independently exponential random variables with a common mean $\sigma^2/(2\pi)$, we establish the theorem by applying the result in the Appendix of Chiu (1987).

References

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Contr.*, **AC-19**, 716-723.

[2] Akaike, H. (1977). Discussion on "Some comments on the Canadian lynx data" by H. Tong. *J. Roy. Statist. Soc. Ser. A,* **140,** 456.

[3] Barlett, M. S. (1967). Inference and stochastic processes. *J. Roy. Statis. Soc. Ser A,* **130,** 457-477.

[4] Bingham, C., Godfrey, M. D. and Tukey, J. W. (1967). Mordern techniques in power spectrum estimation. *IEEE Trans. Audio Electroacoust,* **AU-15,** 56-66.

[5] Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction,* J. Wiley, New York.

[6] Bolt, B. A. and Brillinger D. R. (1979). Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series. *Geophys. J. R. astr. Soc.*, **59**, 593-603.

[7] Brillinger, D. R. (1975). *Time Series Data Analysis and Theory,* Holt, Rinehart, and Winston, New York.

[8] Brillinger, D. R. and Rosenblatt, M. (1967). Asymptotic theory of *k*-th order spectra. In *Spectral Analysis of Time Series* Ed. B. Harris, 153-188. Wiley, New York.

[9] Burg, J. P. (1967). Maximum entropy spectral analysis. Paper presented at 37th Annual International Meeting, Soc. of Explor. Geophys., Oklahoma City, Okla..

[10] Campbell, M. J. and Walker, A. M. (1977). A survey of statistical work on the Mackenzie river series of annual Canadian lynx trappings for the years 1821-1934 and a new analysis. *J. Roy. Statis. Soc. A*, **140**, 411-431.

[11] Chiu, S. T. (1986). Detecting periodic components in a white time series. submitted to *JRSS*. (revising) Also as Technical Report 86-23, Department of Mathematical Science. Rice University, Houston.

[12] Chiu, S. T. (1987). A feature preserving smoother with application to the coal-mining disaster data of Britain. submitted to *Biometrika*. (revised) Also as Technical Report 87-1, Department of Mathematical Sciences, Rice University, Houston.

[13] Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, **19**, 297-301.

[14] Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proc. Roy. Soc. Ser. A.* **125**, 54-59.

[15] Galbraith, R. F. (1977). Discussion on "Some comments on the Canadian lynx data" by H. Tong. *J. Roy. Statist. Soc. Ser. A*, **140**, 453.

[16] Gray, H. L.. and Woodward, W. A. (1986). A new ARMA spectral estimator. *J. Amer. Statist. Assoc.*, **81**, 1100-1108.

[17] Grenander. U. and Rosenblatt, M. (1957). *Statistical analysis of stationary time series*, J. Wiley, New York.

[18] Haggan, V.. Heravi, S. M. and Priestley, M. B. (1984). A study of the application of state-dependent models in non-linear time series analysis.*J. Time Ser. Anal.*, **5**, 69-102.

[19] Haggan, V. and Oyetunji, O. B. (1984). On the selection of subset autoregressive time series models. *J. Time Ser. Anal.*, **5**, 103-113.

[20] Hannan, E. J. (1961). Testing for a jump in the spectral function. *J. Roy. Statist. Soc. Ser. B*, **23**, 394-404.

[21] Hannon, E. J. (1973). The estimation of frequency. *J. appl. Prob.*, **10,** 510-519.

[22] Hannon, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *J. Roy. Statis. Soc., Ser. B,* **41,** 190-195.

[23] Hartley, H. O. (1949). Tests of significance in harmonic analysis. *Biometrika,* **36,** 194-201.

[24] Hasan, T. (1982). Nonlinear time series regression for a class of amplitude modulated cosinusoids. *J. Time Ser. Anal.,* **3**, 109-122.

[25] Hasan, T. (1983). Complex demodulation: some theory and applications, in *Time Series in the Frequency Domain, handbook of statistics Vol. 3* ed. D.R. Brillinger and P.R. Krishnaiah, 125-156. North-Holland, New York.

[26] Kabe, D. G. (1970). Testing outliers from an exponential population. *Metrika,* **15,** 15-18.

[27] Kleiner, B., Martin, R. D. and Thomson, D. J. (1979). Robust estimation of power spectra. *J. Roy. Statis. Soc. Ser. B,* **41,** 313-351.

[28] Lacoss, R. T. (1971). Data adaptive spectral analysis method. *Geophysics,* **36,** 661-675.

[29] Lewis, T. and Fieller, N. R. J. (1979). A recursive algorithm for null distributions for outliers: I. Gamma samples. *Technometrics,* **21,** 371-376.

[30] Likes, J. (1966). Distribution of Dixon's statistics in the case of an exponential population. *Metrika,* **11,** 46-54.

[31] Lim, K. S. and Tong, H. (1983). A sttistical approach to difference-delay equation modelling in ecology - two case studies. *J. Time Ser. Anal.,* **4,** 239-267.

[32] Luo, S., Liang, S., Ye, S., Yan, S. and Li, Y (1977). Analysis of periodicity in the irregular rotation of the earth. *Chinese Astronomy.* **1,** 221-227.

[33] Kimber, A. C. (1983). Trimming in gamma samples. *Appl. Statist.,* **32,** 7-14.

[34] Martin, R. D. and Yohai, V. J. (1986). Influence functionals for time series. *Ann. Statis.,* **14,** 781-840.

[35] Morton, M. J., and Gray, H. L. (1984). The G-spectral estimator. *J. Amer. Statist. Assoc.,*

79, 692-701.

[36] Nowroozi, A. A. (1965). Eigenvibrations of the Earth after the Alaskan earthquake. *J. Geophys. Res.*, **70**, 5145-5156.

[37] Nowroozi, A. A. (1966). Terrestrial spectroscopy following the Rat Island earthquake. *Bull. seism. Soc. Am.*, **56,R 1269-1288.**

[38] Parzen, E. (1969). Multiple time series modelling. in *Multivariate Analysis II*, ed. P.R. Krishnaiah, 389-409. Academic Press, New York.

[39] Parzen, E. (1974). Some recent advances in time series modeling. *IEEE Trans. Autom. Contr.*, **AC-19**, 723-730..

[40] Parzen, E. (1977). Multiple time series: determining the order of approximating autoregressive schemes. in *Multivariate Analysis IV*, 283-295. North-Holland, Amsterdam.

[41] Parzen, E. (1983). Autoregressive spectral estimation. in *Time Series in the Frequency Domain, handbook of statistics Vol. 3* ed. D.R. Brillinger and P.R. Krishnaiah, 221-247. North-Holland, New York.

[42] Press, H. and Tukey, J. W. (1956). *Power spectral methods of analysis and their application to problems in airplane dynamics.* Bell Telephone System Monograph 2606.

[43] Priestley, M. B. (1962a). The analysis of stationary processes with mixed spectra-I. *J. Roy. Statis. Soc. Ser. B*, **24**, 215-233.

[44] Priestley, M. B. (1962b). The analysis of stationary processes with mixed spectra-II. *J. Roy. Statis. Soc. Ser. B*, **24**, 511-529.

[45] Priestley, M. B. (1964). Estimation of the spectral density function in the presence of harmonic components. *J. Roy. Statis. Soc. Ser. B*, **26**, 123-132.

[46] Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press, London.

[47] Siegel, A. F. (1980) Testing for periodicity in a time series. *J. Amer. Statist. Assoc.*, **75**, 345-348.

[48] Shimshoni, M. (1971) On Fisher's test of significance in harmonic analysis. *Geophys. J. R. astr. Soc.*, **23**, 373-377.

[49] Tong, H. (1977). Some comments on the Canadian lynx data. *J. Roy. Statis. Soc. A*, **140**, 432-436.

[50] Walker, A. M. (1973). On the estimation of a harmonic component in a time series with stationary dependent residuals. *Adv. Appl. Prob.*, **5**, 217-241.

[51] Whittaker, E. T. and Robinson, G. (1924). *The Calculus of Observations.* Blackie abd Son, London.

[52] Whittle, P. (1952). The simultaneous estimation of a time series harmonic component and covariance structure. *Trabajos de Estadistica*, **3**, 43-57.

[53] Whittle, P. (1954) The statitical analysis of a seiche record. *Sears Foundation Journal of Marine Research*, **13**, 76-100.

[54] Wilson, G. T. (1985). Variations in the Earth's rotation rate, 1820-1970. In *Data; A collection of problems from many fields for the student and research worker.* 133-134. Springer-Verlag, New York.

Caption

Figure 1. The results of Simulation 1; continuous spectrum. A. The periodograms. B. The feature preserving spectral estimate. C. The classical kernel spectral estimate. D. The 7-th order autoregressive spectral estimate.

Figure 2. The results of Simulation 2; continuous spectrum plus a large peak. A. The periodograms. B. The feature preserving spectral estimate. C. The classical kernel spectral estimate. D. The 17-th order autoregressive spectral estimate.

Figure 3. The results of Simulation 3; continuous spectrum plus a mild peak. A. The periodograms. B. The feature preserving spectral estimate. C. The classical kernel spectral estimate. D. The third order autoregressive spectral estimate.

Figure 4. The results of Simulation 4; continuous spectrum plus two peaks. A. The periodograms. B. The feature preserving spectral estimate. C. The classical kernel spectral estimate. D. The 6-th order autoregressive spectral estimate.

Figure 5. The Akaike's criterions (denoted by *) and the robust Akaike's criterions (denoted by o) for Examples 1 to 4 (plotted in A to D, respectively).

Figure 6. Comparison of $\hat{f}_{FP}(\lambda)$ and $\hat{f}_{PW}(\lambda)$ of Example 1. The order of the robust filter is 2.

Figure 7. The periodograms and the feature preserving spectral estimate for the variable-star series.

Figure 8. The periodograms and the feature preserving spectral estimate for the series of the Earth's rotation rate.

Figure 9. The periodograms, the feature preserving spectral estimate (solid line) and the 11-th order autoregressive spectral estimate (dashed line) [from Tong (1977)] of the Canadian lynx series.

Figure 10. The instantaneous phase of the Canadian lynx series demodulated at frequency $\lambda = (2\pi)/9.63$.

Figure 11. A. The wavenumber periodograms at frequency $(0.1)2\pi$. B. The feature preserving spectral estimate. C. The classical kernel spectral estimate.

Fig 1

Fig. 1.A

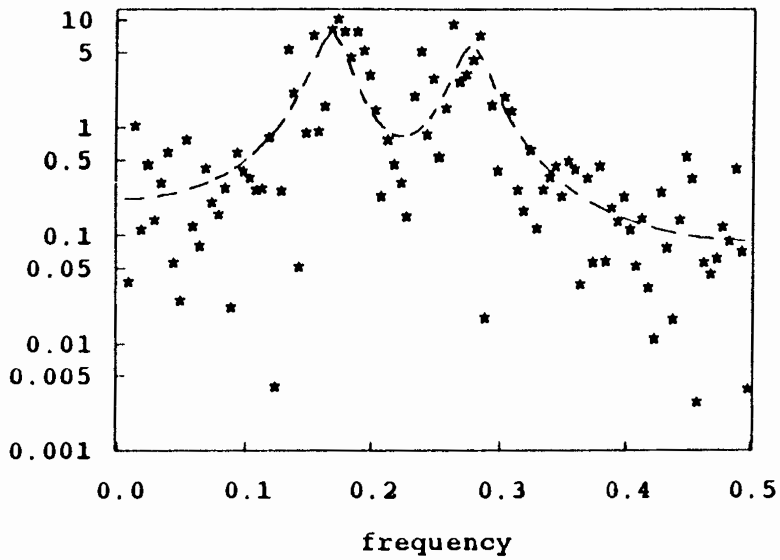


Fig. 1.B

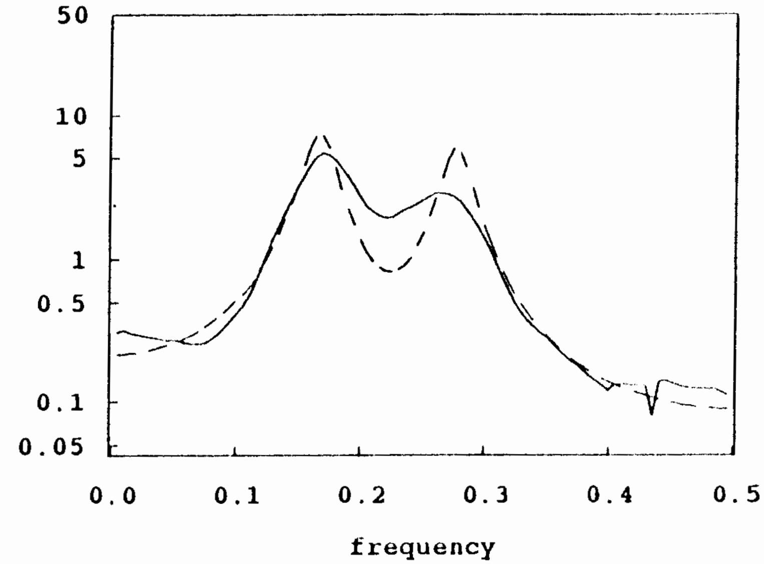


Fig. 1.C

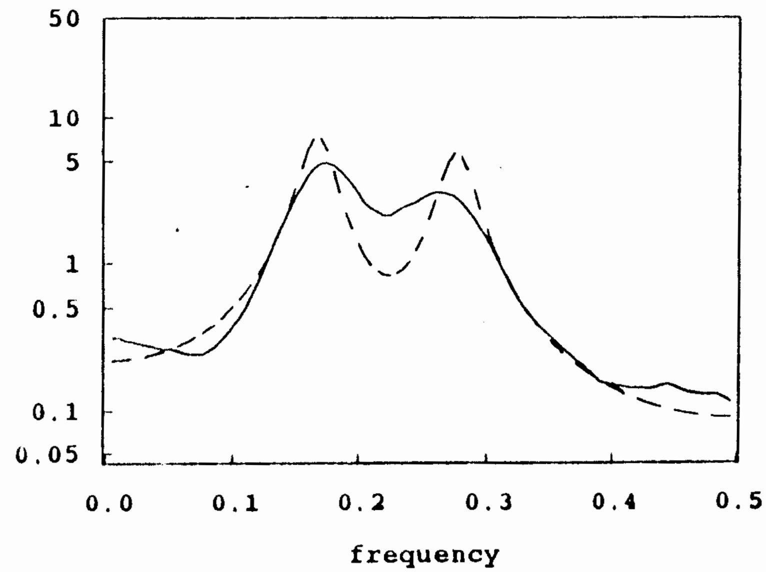


Fig. 1.D

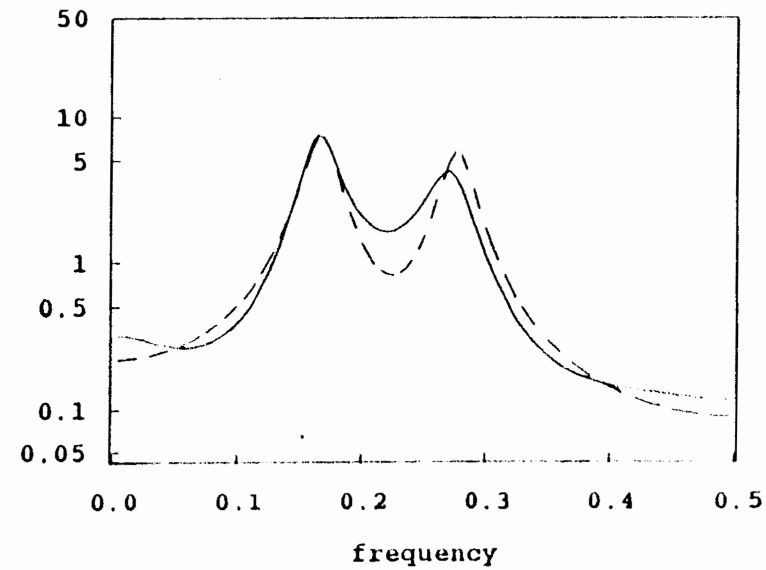


Fig. 2.A

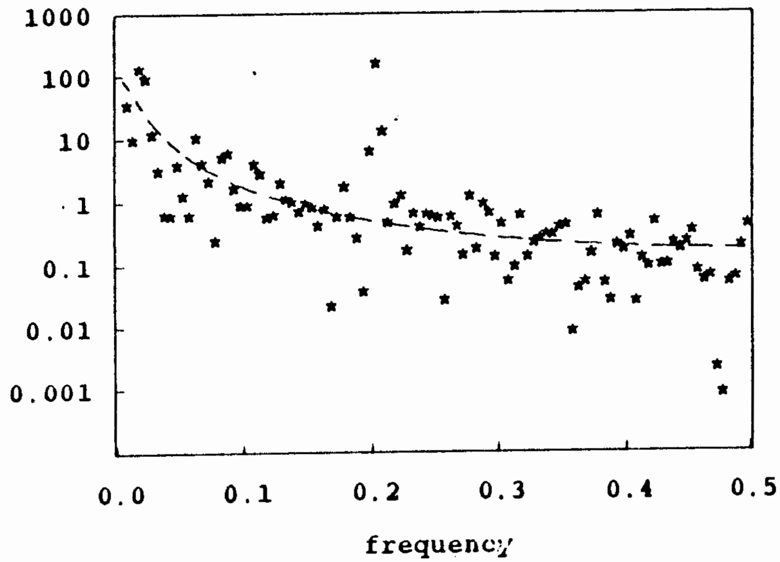


Fig. 2.B

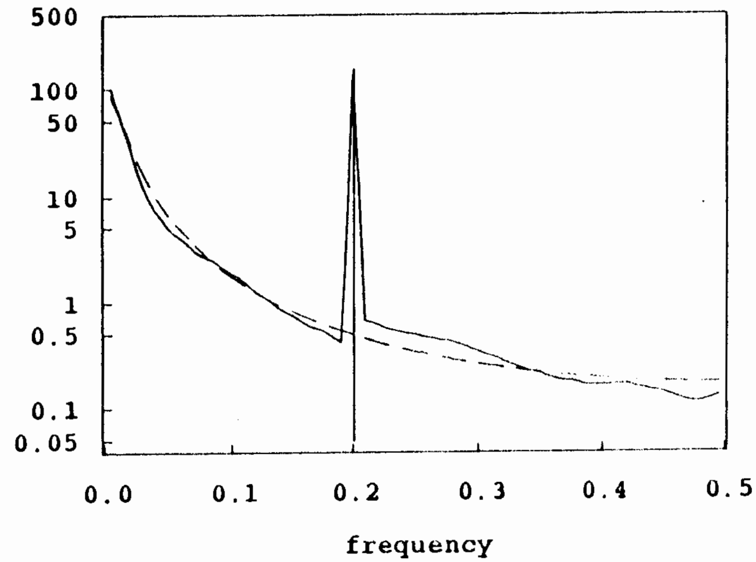


Fig. 2.C

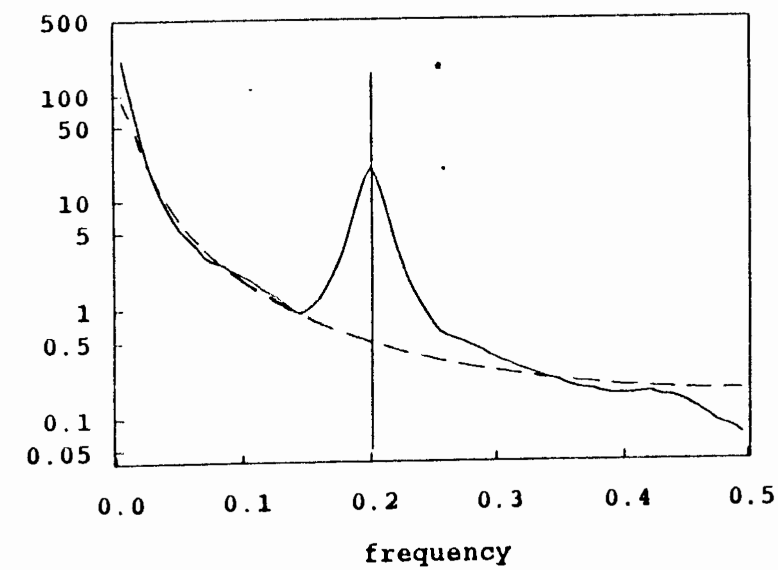


Fig. 2.D

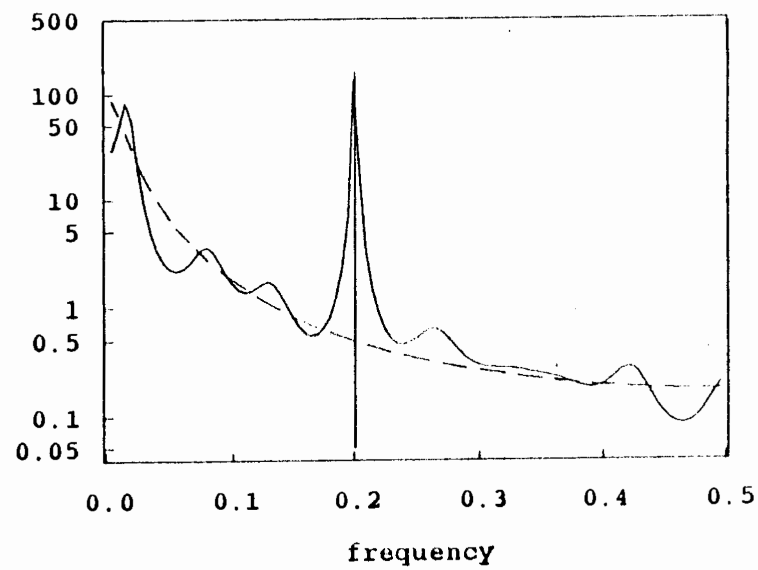


Fig. 3.A

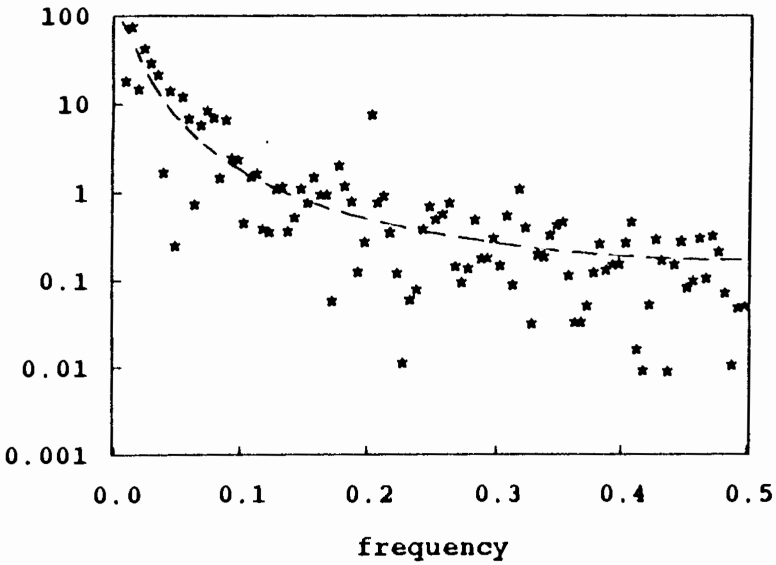


Fig. 3.B

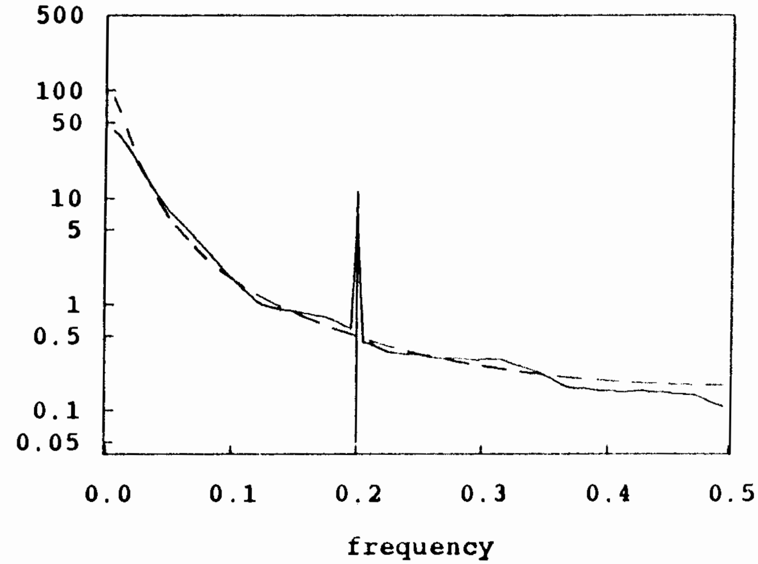


Fig. 3.C

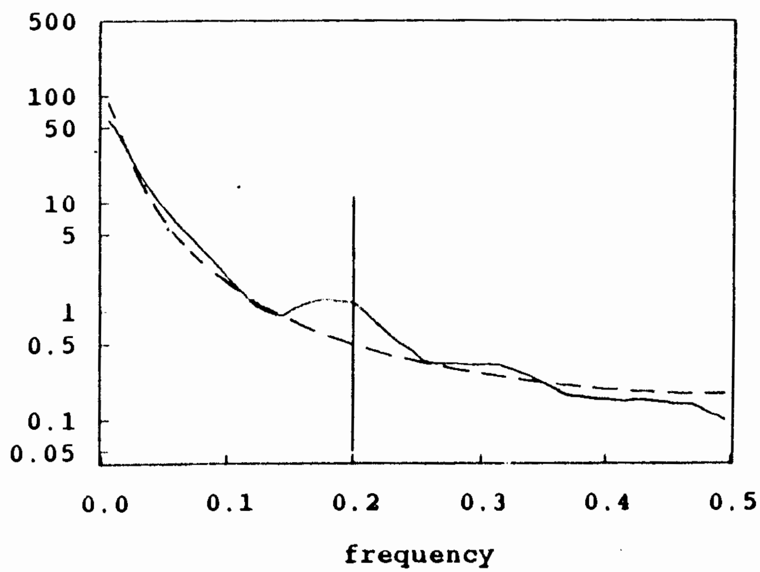


Fig. 3.D

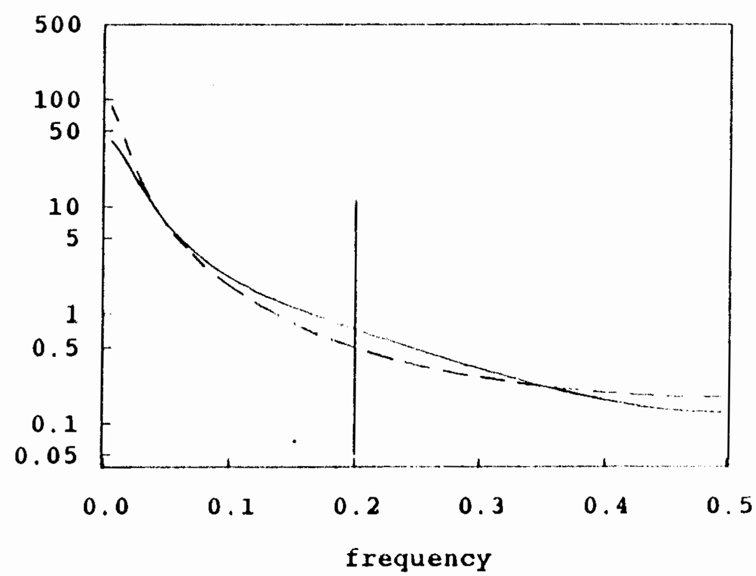


Fig. 4.A

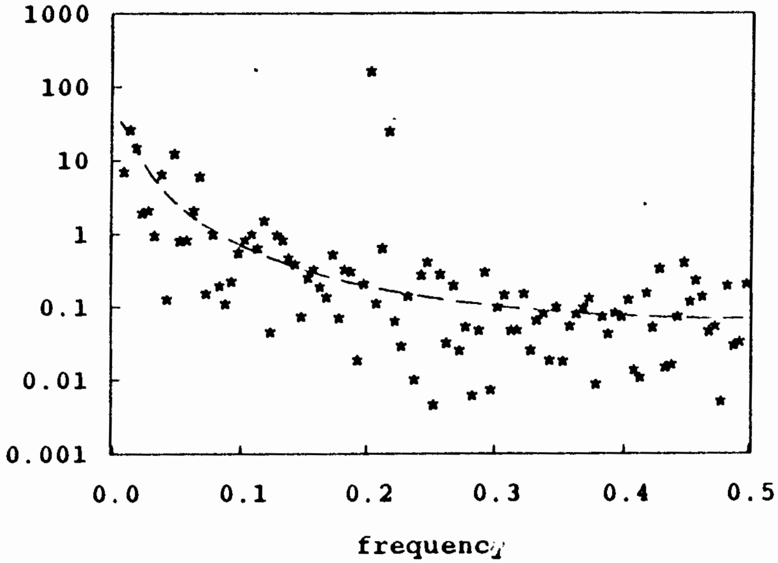


Fig. 4.B

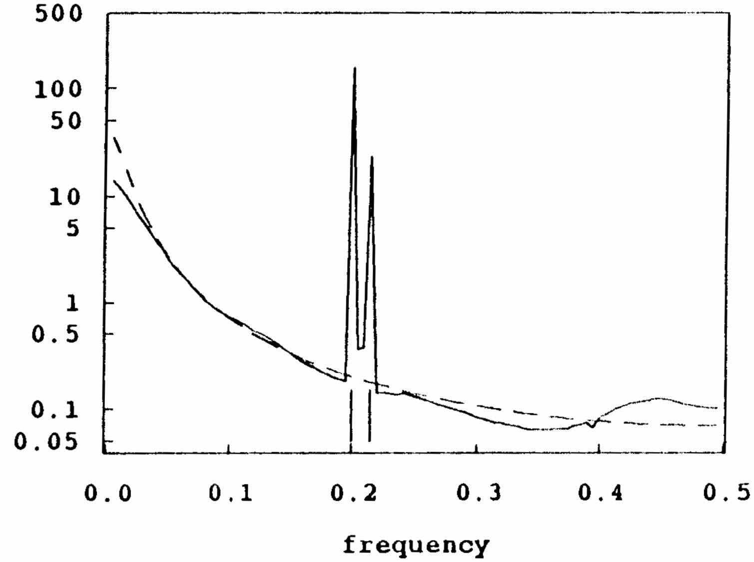


Fig. 4.C

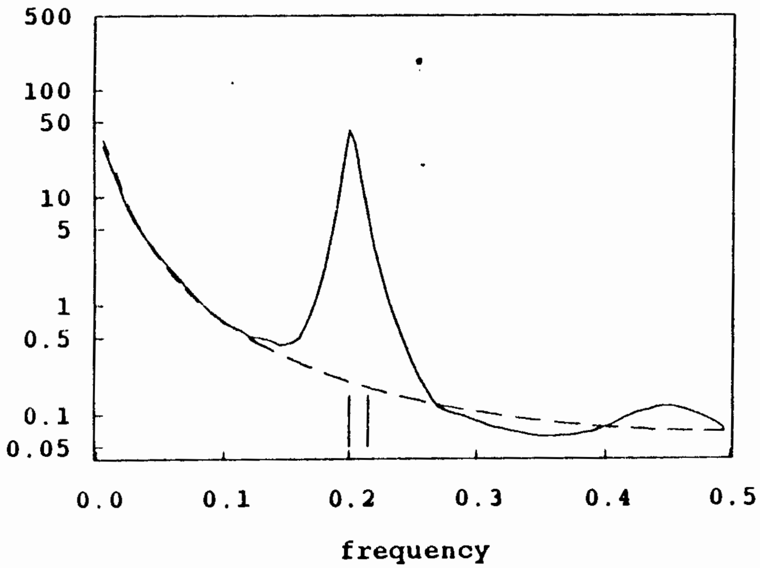


Fig. 4.D

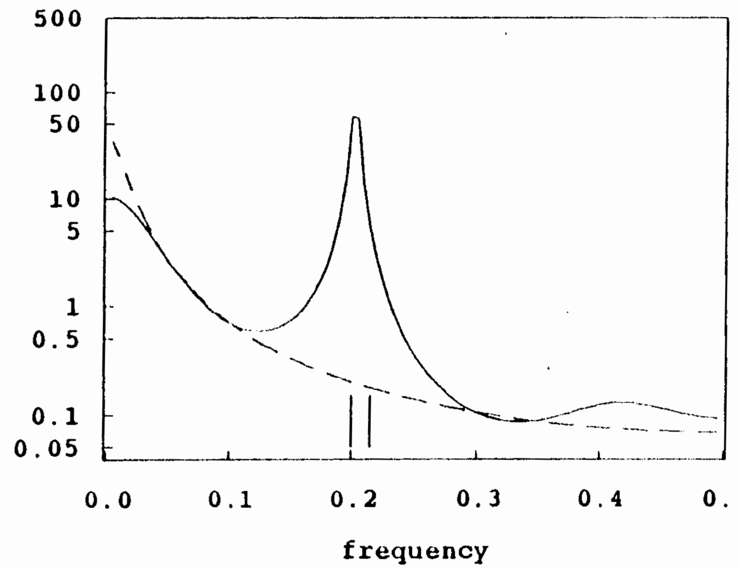


Fig. 5.A

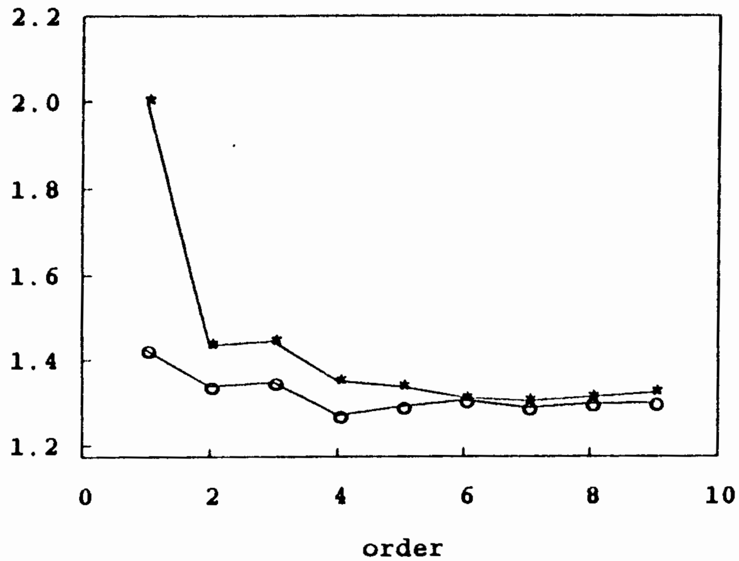


Fig. 5.B

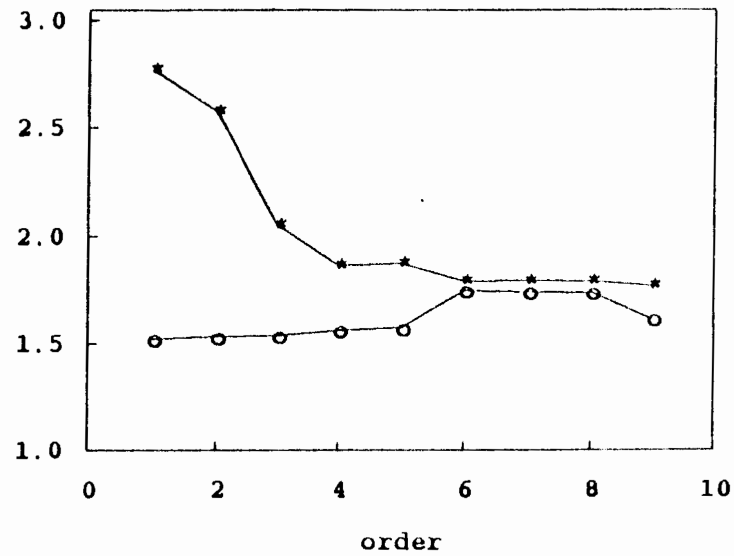


Fig. 5.C

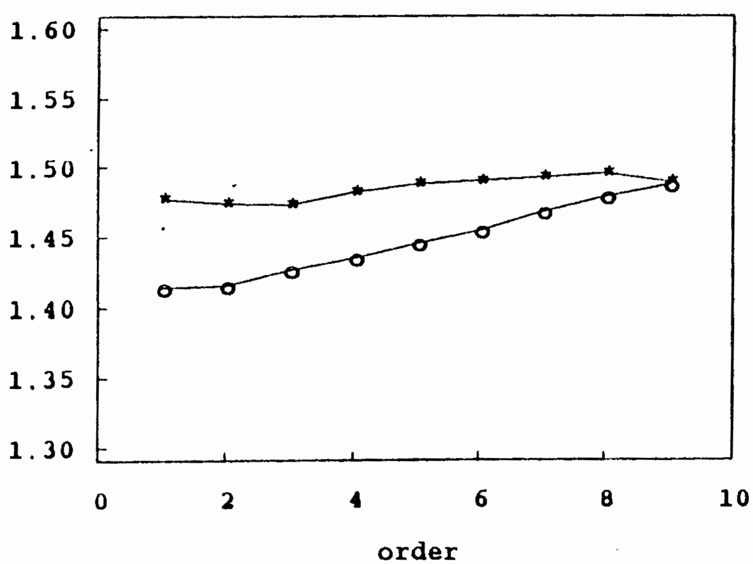
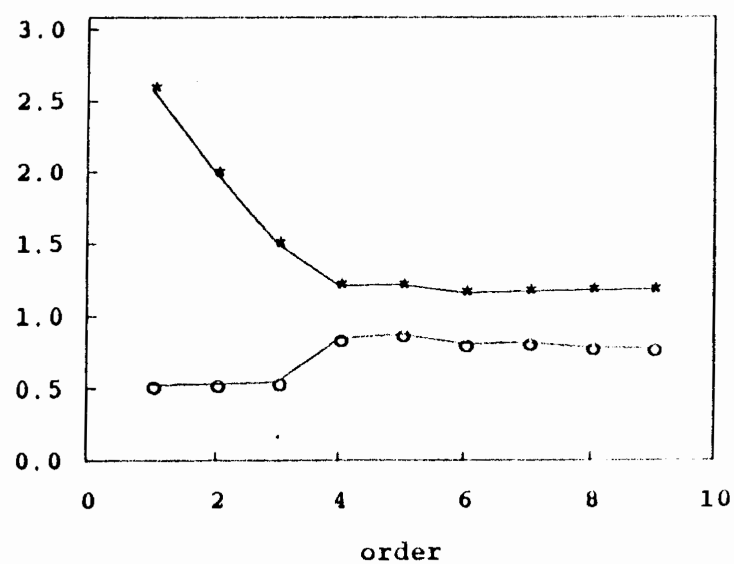
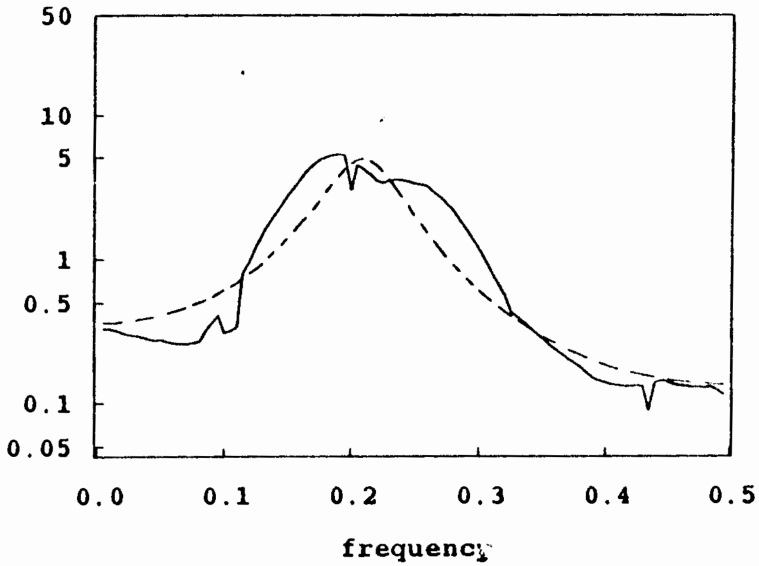
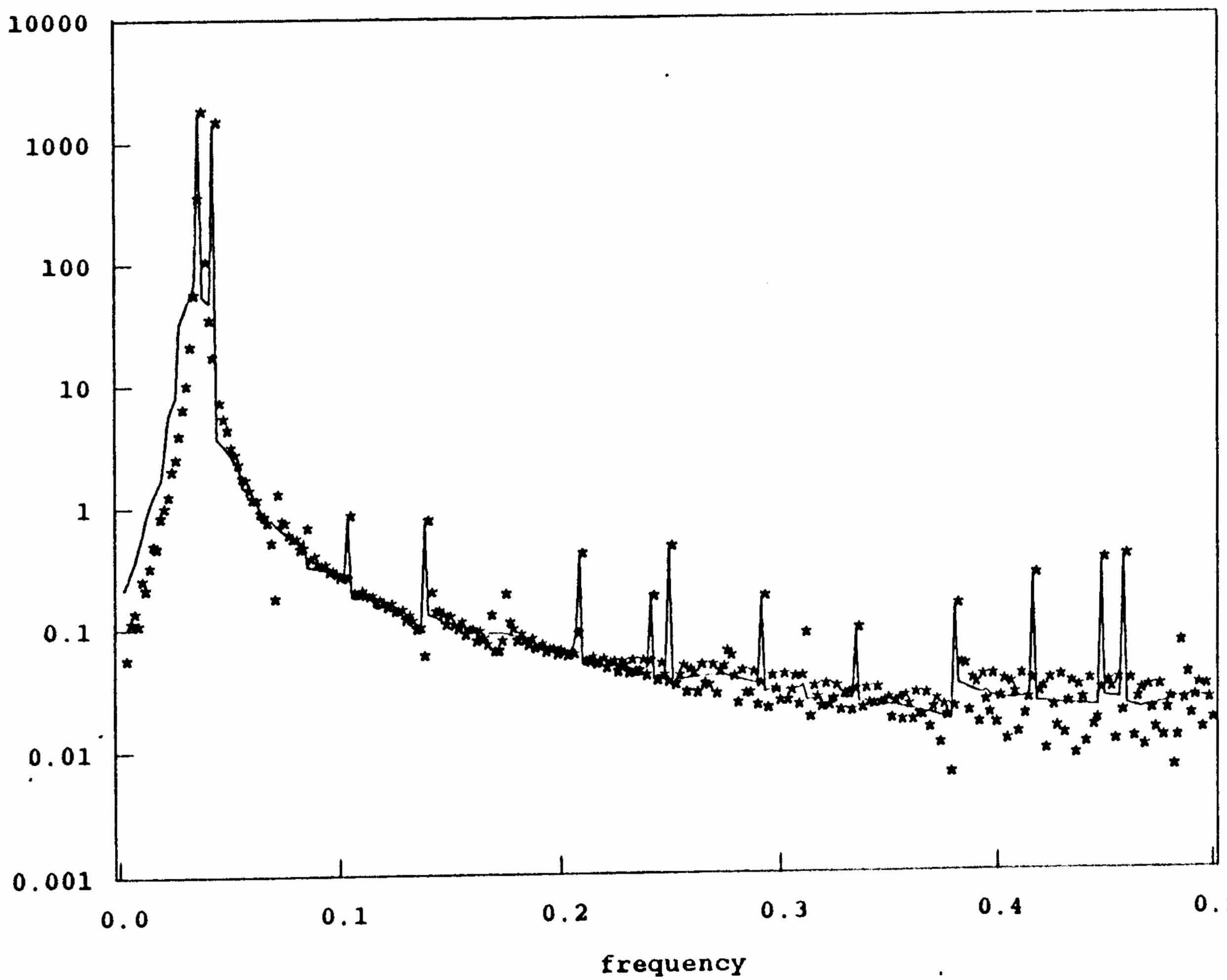


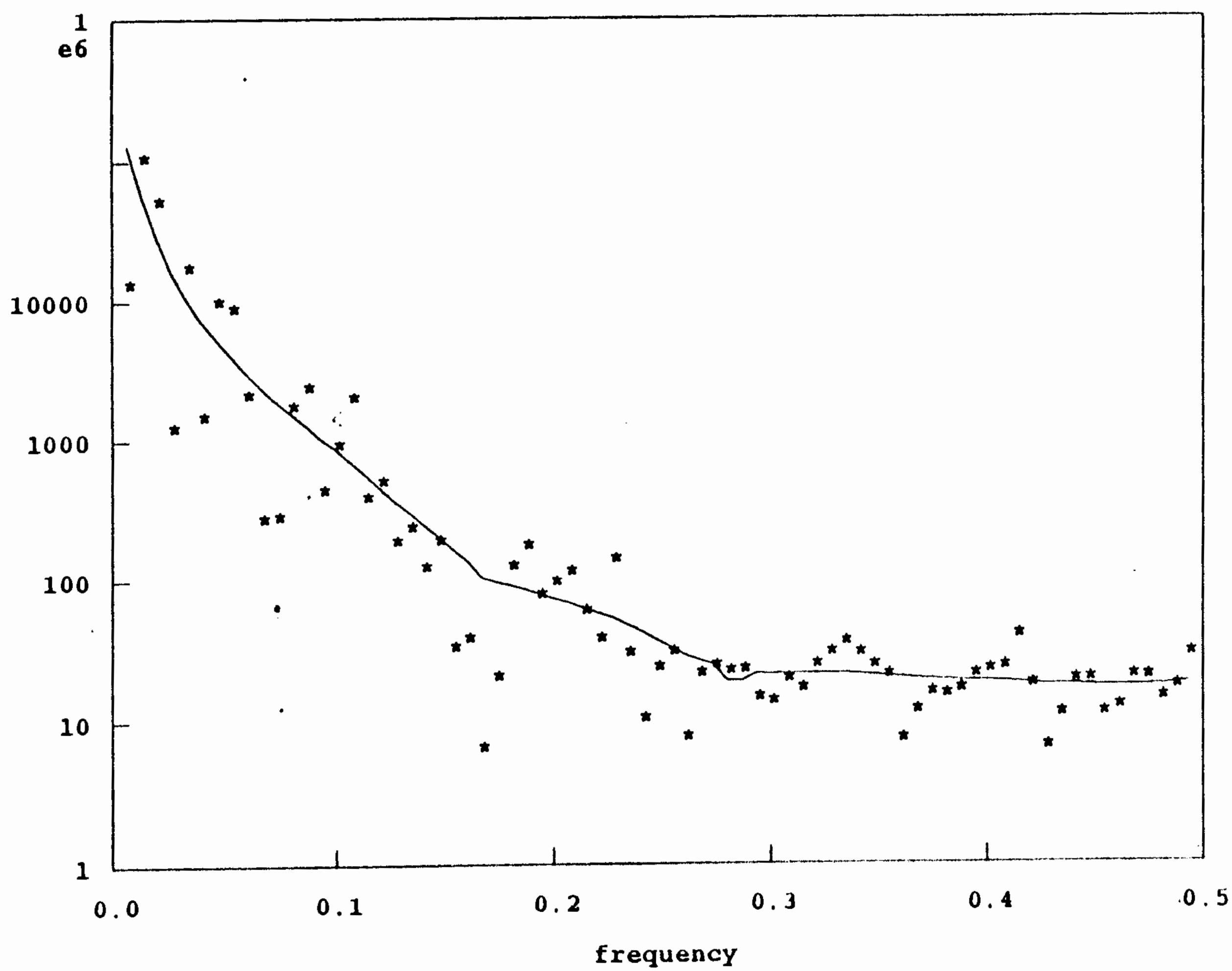
Fig. 5.D

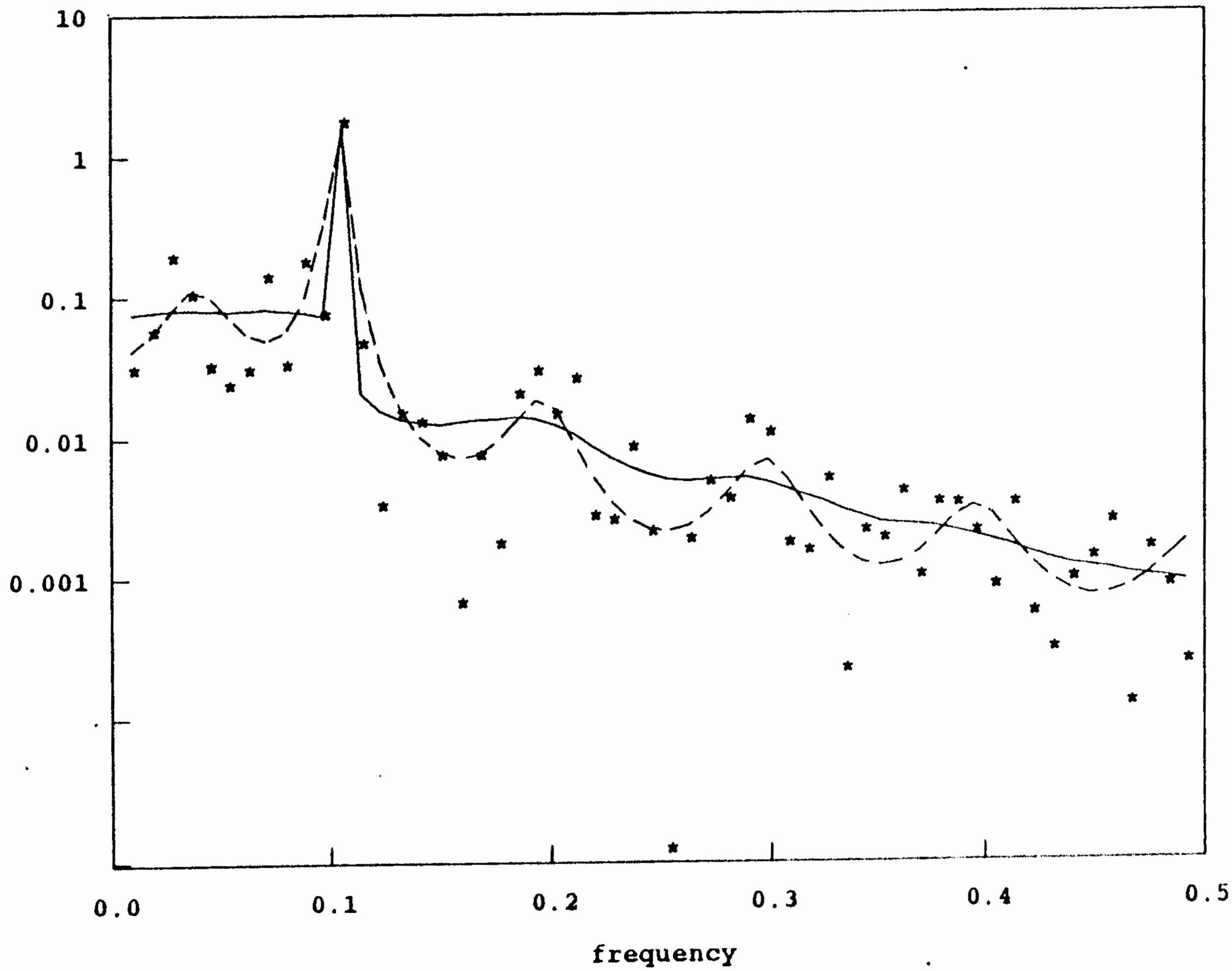




50
10
5
1
0.5
0.1
0.05
0.0
0.1
0.2
0.3
0.4
0.5
frequency


10000
1000
100
10
1
0.1
0.01
0.001
0.0
0.1
0.2
0.3
0.4
0.5
frequency


1
e6
10000
1000
100
10
1
0.0
0.1
0.2
0.3
0.4
0.5
frequency


10
1
0.1
0.01
0.001
0.0
0.1
0.2
0.3
0.4
0.5
frequency

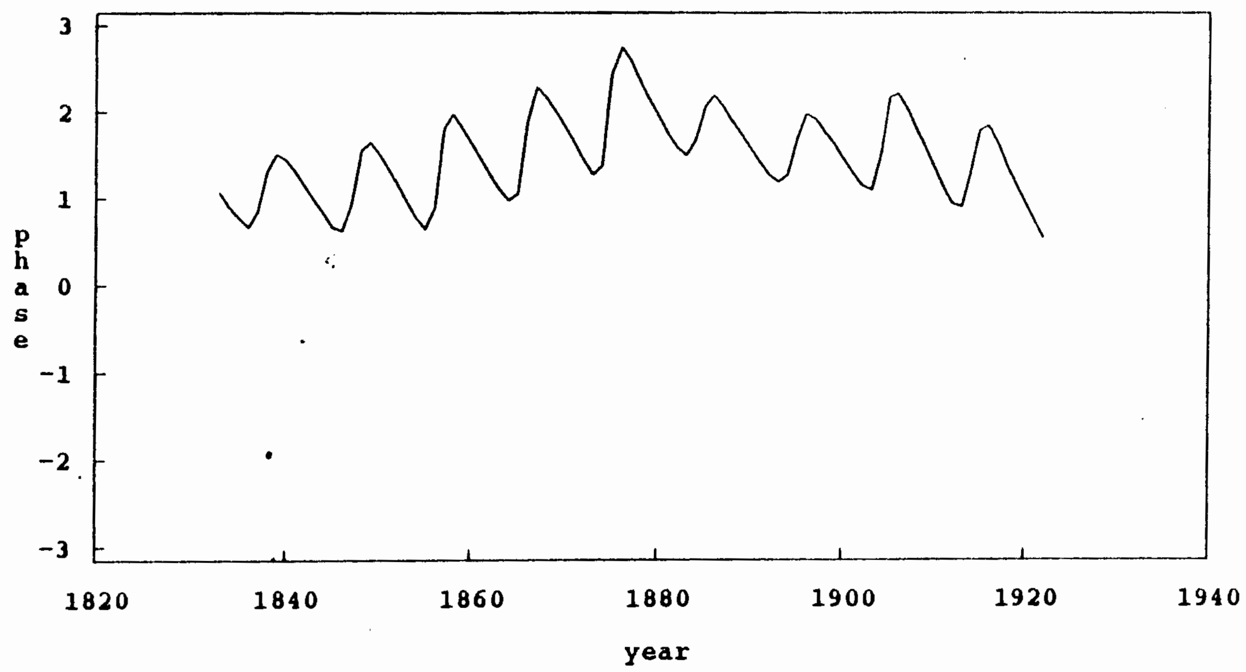

3
2
1
0
-1
-2
-3
phase
1820
1840
1860
1880
1900
1920
1940
year

Fig. 11.A

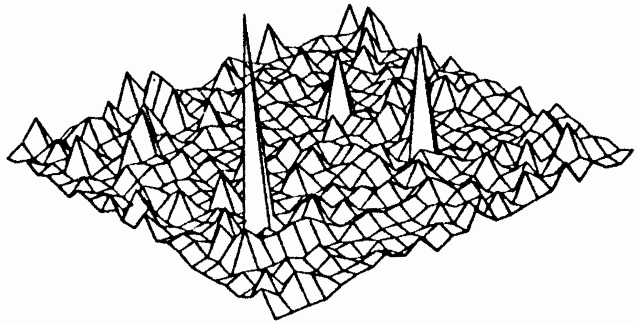


Fig. 11.B

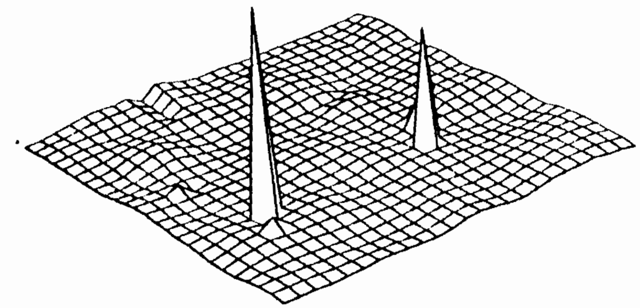


Fig. 11.C