FOURIER INFERENCE: SOME METHODS FOR THE ANALYSIS OF ARRAY AND NONGAUSSIAN SERIES DATA

BY

DAVID R. BRILLINGER

TECHNICAL REPORT NO. 45
JULY 1985

*RESEARCH PARTIALLY SUPPORTED BY
NATIONAL SCIENCE FOUNDATION GRANT DMS-8316634

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

Fourier Inference: Some Methods for the Analysis of Array and NonGaussian Series Data

DAVID R. BRILLINGER

University of California, Berkeley

Fourier inference is a collection of analytic techniques and philosophic attitudes,for the analysis of data, wherein essential use is made of empirical Fourier transforms. This paper sets down some basic results concerning the finite Fourier transforms of stationary process data and then, to illustrate the approach, uses those results to develop procedures for: i) estimating cloud and storm motion, ii) passive sonar and iii) fitting finite parameter models to nonGaussian time series via bispectral fitting. This last procedure is illustrated by an analysis of a stretch of Mississippi River runoff data. Examples i),ii) refer to data having the form $Y(x_j,y_j,t)$ for $j = 1,\cdots,J$ and $t = 0,\cdots,T-1$ say, and view that data as part of a realization of a spatial-temporal process. Such data has become common in geophysics generally and in hydrology particularly. The goal of this paper is to present some new statistical procedures pertinent to problems in the water sciences, equally it is to illustrate the genesis of those procedures and how their properties may be approximated.

INTRODUCTION AND SUMMARY

Statistical inference is concerned with making statements going beyond the data at hand. Fourier inference is the part that proceeds in this connection making essential use of Fourier transforms. Making use of such transforms is often found to simplify the study of a scientific problem, both philosophically and analytically. The latter results in part from the nice mathematical, statistical and computational properties of the Fourier transform. The Fourier transform isolates effects and often allows a problem to be replaced by one involving independent identically distributed observations, i.e. those with which the vast majority of statistical techniques are concerned.

Fourier inference is one concern of this paper. A second is the statistical analysis of array, or network, data. The data take the form

$$Y(x_j, y_j, t) \qquad (1)$$

for $j = 1, \ldots, J$ and $t = 0, \ldots, T-1$ with the index t viewed as "time" and the points (x_j, y_j) viewed as the (planar) locations of an array (or network) of sensors. For fixed j, $Y(x_j, y_j, t)$ $t = 0, \ldots, T-1$ is a stretch of time series data. For fixed t, $Y(x_j, y_j, t)$ $j = 1, \ldots, J$ is a sampling of a spatial field. Examples to be expanded upon later in the paper include: the contemporaneous rates at which rain is falling as recorded at a network of gauges, the fluctuating pressure levels being measured by an array of sonar sensors and the measured runoff rate at a station on a river. The particular scientific problems to be considered in connection with these examples are respectively: estimation of direction and velocity of storm movement from data at two time points, estimation of the direction of an energy source from data at a few sensors,and the fitting of a finite parameter model to the runoff. The procedure presented for this last is novel, making use of both second- and third-order information.

Array data are essential to the study of phenomena moving and varying in time and space. Study of such data, and corresponding processes, adds insight to the marginal space and time cases. Array data allow estimation of parameters important to other scientific problems, such as frequency-wavenumber spectra which appear in expressions for exceedance probabilities , see for example Forristall _et al_. (1978). Array data have become common in geophysics generally and in hydrology particularly.

Section 2 of the paper sets down basic notations, definitions and statistical properties of the Fourier transforms of (large) segments of stationary processes. Section 3 illustrates how these properties may be invoked to build analyses and to suggest techniques for making statistical inferences. The cases of model fitting and estimation will be concentrated upon. Section 4 is concerned with the estimation of cloud and storm motion, Section 5 with passive sonar and Section 6 with improved fitting of a model for river runoff by making use of both the power spectrum and bispectrum. This last case is illustrated with a preliminary analysis of monthly observations of Mississippi River runoff. The paper concludes with Discussion.

The work edited by Brillinger and Krishnaiah (1983) contains papers surveying a broad variety of aspects of Fourier inference.

SOME BASIC CONCEPTS AND RESULTS

This section serves to introduce the notation, the assumptions and the properties of the Fourier transform as a tool in the analysis of random process data.

By time series data will be meant a succession of values Y(0), Y(1),..., Y(T-1) with T the length. The series may be vector-valued, in which case boldface notation will be employed, $\mathbf{Y}(t)$ t = 0,...,T-1 . The Fourier transform of this last will be denoted by

$$\mathbf{d}_Y^T(\lambda) = \sum_{t=0}^{T-1} \mathbf{Y}(t) \exp\{-i\lambda t\} \quad , -\infty<\lambda<\infty \tag{2}$$

It will typically be computed for a set of discrete frequencies by some fast algorithm. See Heideman <u>et al</u>. (1984) for a review and references.

In many circumstances time series data may be usefully viewed as part of a realization $\mathbf{Y}(t,\omega)$ t = 0,±1,±2,... of a stochastic process. (Here ω is a random variable and $Y(\cdot,\omega)$ measurable in ω.) This allows parameters to be defined through which analysis and discussion may be carried out. Parameters that are of particular importance for this paper arise when the process $\mathbf{Y}(\cdot)$ is stationary (that is joint probability distributions are unchanged by time translation) and include the (matrix-valued) <u>covariance function</u>

$$\mathbf{c}_{YY}(u) = \text{cov}\{\mathbf{Y}(t+u),\mathbf{Y}(t)\} \tag{3}$$

t,u = 0,±1,... and the <u>spectral density matrix</u>

$$\mathbf{f}_{YY}(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} \mathbf{c}_{YY}(u) \exp\{-i\lambda u\} \tag{4}$$

for $-\infty<\lambda<\infty$. In the case that the process $\mathbf{Y}(\cdot)$ has mean $\mathbf{0}$, this last may be connected to the statistic (2.1) via

$$f_{YY}(\lambda) = \lim_{T\to\infty} (2\pi T)^{-1} E\left\{ d_Y^T(\lambda) \overline{d_Y^T(\lambda)}^{\tau} \right\} \tag{5}$$

(" " denoting matrix transpose, "—" complex conjugate. When the process is real-valued, one has the power spectrum. Letting $f_{jk}(\lambda)$ denote the entry in row j, column k of $f_{YY}(\lambda)$, one can define the coherence of the j-th and k-th components of $Y(\cdot)$ at frequency λ as

$$|R_{jk}(\lambda)|^2 = |f_{jk}(\lambda)|^2/f_{jj}(\lambda)f_{kk}(\lambda) \tag{6}$$

It is a useful descriptive parameter and also appears in many expressions for sampling variability.

There are corresponding definitions in the case of a spatial process, $Y(x,y)$ $x,y = 0,\pm1,\cdots$. Here x,y are the coordinates of location and for example in the stationary case one has the covariance function

$$c_{YY}(u,v) = \text{cov}\{Y(x+u,y+v),Y(u,v)\} \tag{7}$$

$u,v,x,y = 0,\pm1,\cdots$ and the power spectrum

$$f_{YY}(\alpha,\beta) = (2\pi)^{-2} \sum_u \sum_v c_{YY}(u,v) \exp\{-i(\alpha u+\beta v)\} \tag{8}$$

$-\infty < \alpha,\beta < \infty$

with (α,β) referred to as the wavenumber.

A spatial-temporal process has coordinates of both space and time. Its Fourier transform has coordinates of both wavenumber and frequency. In a variety of applications, a spatial-temporal process will be observed only at irregularly spaced positions $r_j = (x_j,y_j)$, $j = 1,\cdots,J$.

The development of the techniques/analyses presented in this paper will make substantial use of central limit theorems for empirical Fourier transforms. The distribution of a statistic like (2) will be approximated by a (complex) normal distribution. In the case that the process $\underset{\sim}{Y}(\cdot)$ is Gaussian, the distribution is exactly Gaussian. In the general case the result holds for processes that are stationary and mixing. Mixing here means that values of the process that are far apart in time (or space) are statistic- independent or nearly so. Precise formulations of mixing conditions leading to the desired central limit theorems may be found in Hannan (1970), Brillinger (1983) for example.

Quoting from the last reference, the sort of results one has, include:

(i) For $0<\lambda<\pi$, $\underset{\sim}{d}_Y^T(\lambda)$ is asymptotically normal with mean $\underset{\sim}{0}$ and covariance matrix $2\pi T \underset{\sim}{f}_{YY}(\lambda)$, as $T\to\infty$,

(ii) For $0<\lambda_1< \cdots <\lambda_K<\pi$, $\underset{\sim}{d}_Y^T(\lambda_1),\ldots,\underset{\sim}{d}_Y^T(\lambda_K)$ are asymptotically independent,

(iii) For $\lambda_k^T = 2\pi s_k^T/T$, with $0<\lambda_k^T<\pi$ and the s_k^T distinct integers, $k=1,\ldots,K$ $\underset{\sim}{d}_Y^T(\lambda_1^T),\ldots,\underset{\sim}{d}_Y^T(\lambda_K^T)$ are asymptotically independent complex normals with mean $\underset{\sim}{0}$ and covariance matrix $2\pi T \underset{\sim}{f}_{YY}(\lambda)$. Usually one takes $\frac{T}{k}$ as T ∞.

These results suggest many procedures pertinent to the analysis of time series data. One great simplification that they suggest follows from (iii), namely, to treat the $\underset{\sim}{d}_Y^T(\lambda_k^T)$ as if they were a sample from a normal distribution. As a preliminary example, one is led to estimate $\underset{\sim}{f}_{YY}(\lambda)$ by

$$\underset{\sim}{f}_{YY}^T(\lambda) = K^{-1}\sum_k (2\pi T)^{-1}\underset{\sim}{d}_Y^T(\lambda_k^T)\,\overline{\underset{\sim}{d}_Y^T(\lambda_k^T)}^{\tau} \qquad (9)$$

and to approximate the distribution of this statistic by a complex Wishart with parameters K and $\underset{\sim}{f}_{YY}(\lambda)$.

From the standpoint of developing empirical procedures and understanding their properties, the <u>Skorokhod representation</u> is an exceedingly valuable tool. It allows one to write the result (iii) above as

$$\underset{\sim}{d}_Y^T(\lambda_k^T) = \underset{\sim}{\mathcal{L}}_k + o_{a.s.}(T^{1/2}); \quad k = 1,\cdots,K \tag{10}$$

where the $\underset{\sim}{\mathcal{L}}_k$ are independent complex normals with mean $\underset{\sim}{0}$ and covariance matrix $2\pi T\ \underset{\sim}{f}_{YY}(\lambda)$. It allows one to obtain approximating distributions of functions of Fourier transforms by elementary manipulations. The abstract theorem is in Skorokhod (1956). (Here $T^{-1/2}o_{a.s.}(T^{1/2})$ tends to 0 with probability 1.

The above are first- and second-order results. In the bispectral estimation procedure to be presented, third-order results are employed. Suppose the real-valued series Y(t) has mean c_Y and <u>third-order moment function</u>

$$c_{YYY}(u,v) = E\{[Y(t+u) - c_Y][Y(t+v) - c_Y][Y(t) - c_Y]\} \tag{11}$$

$t,u,v = 0,\pm 1,\cdots$. Then the <u>bispectrum</u> of the process Y(.) at bifrequency (λ,μ) is given by

$$f_{YYY}(\lambda,\mu) = (2\pi)^{-2} \sum_u \sum_v c_{YYY}(u,v) \exp\{-i(\lambda u+\mu v)\} \tag{12}$$

$-\infty \quad ,\mu \quad \infty$.

In the case that Y(.) has mean 0, analogously to expression (5),

$$f_{YYY}(\lambda,\mu) = \lim_{T\to\infty} (2\pi)^{-2}T^{-1}\ E\left\{d_Y^T(\lambda)\ d_Y^T(\mu)\ \overline{d_Y^T(\lambda+\mu)}\right\}. \tag{13}$$

Fourier transforms at frequencies λ,μ,ν such that $\lambda+\mu+\nu = 0$ are seen to be statistically dependent in a special way. The quantity

$$I^T_{YYY}(\lambda,\mu) = (2\pi)^{-2}T^{-1}\, d^T_Y(\lambda)\, d^T_Y(\mu)\, \overline{d^T_Y(\lambda+\mu)} \tag{14}$$

is the <u>third-order periodogram</u>. Some of its statistical properties are developed in Brillinger and Rosenblatt (1967), in particular one has

$$E\, I^T_{YYY}(\lambda,\mu) \sim f_{YYY}(\lambda,\mu) \tag{15}$$

$$\text{var}\, I^T_{YYY}(\lambda,\mu) \sim \frac{T}{2\pi} f_{YY}(\lambda)\, f_{YY}(\mu)\, f_{YY}(\lambda+\mu) \tag{16}$$

provided that $0 < \mu < \lambda < \lambda+2\mu < \pi$. Further the asymptotic covariance of third-order periodograms at distinct bifrequencies is negligeable and the periodogram may be smoothed to estimate the bispectrum, with the estimate asymptotically normal. Details may be found in Brillinger and Rosenblatt (1967). Bispectral analysis is useful in dealing with nonGaussian processes and with nonlinear systems.

The specific procedures discussed later in this paper are concerned with the estimation of a finite dimensional parameter. The estimate is generally normal and an expression may be set down for the covariance matrix of its asymptotic distribution. The rigorous development of such results generally proceeds in the following two step fashion: (i) an empirical loss function $Q_T(\theta)$ is recognized. The estimate $\hat{\theta}$ is the value of θ minimizing $Q_T(\theta)$. As $T \to \infty$, $Q_T(\theta)$ tends in probability to the function $Q(\theta)$ having a unique minimum at $\theta = \theta_o$. Then, under some regularity conditions, $\hat{\theta}$ tends in probability to the "true" value θ_o; (ii) it follows that one can act as if $\hat{\theta}$ is near θ_o and set down the following Taylor series expansion

$$Q'_T(\hat{\theta}) = Q'_T(\theta_o) + (\hat{\theta} - \theta_o) \cdot Q''_T(\theta_*) \tag{17}$$

for θ_* between $\hat{\theta}$ and θ_o . So one can write

$$\hat{\theta} - \theta_o = -Q_T'(\theta_o) \cdot Q_T''(\theta_*)^{-1} \tag{18}$$

Often it is the case that $\sqrt{T}\, Q_T'(\theta_o)$ is asymptotically normal with mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{A}$ and the case that $Q_T''(\theta_*) \sim \underset{\sim}{B}^{-1}$. One can then conclude that $\sqrt{T}(\hat{\theta} - \theta_o)$ is asymptotically normal with mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{B}^{\tau}\underset{\sim}{A}\underset{\sim}{B}$.

The loss function may be a sum of squares, it may be a negative log likelihood function, it may be a negative log likelihood function corresponding to a multivariate normal with parametrized mean and covariance functions. In a broad variety of situations the computation of $\hat{\theta}$, for given Q_T , may be carried out by iteratively reweighted least squares - a technique that has much to be said in its favour, see Green (1984).

The above two-step procedure, of first showing consistency then making a Taylor expansion, is what was employed to formally develop the approximate distributions suggested for the examples presented in sections 4, 5, 6 of this paper.

To end this section of technical results, it may be mentioned that the Fourier transforms basic property of converting convolution into multiplication holds in an approximate sense in the finite case. In particular if $Y(t) = \Sigma\, a(u)X(t-u)$, then $d_Y^T(\lambda) \doteq A(\lambda) d_X^T(\lambda)$ with $A(\lambda) = \Sigma\, a(u) \exp\{-i\lambda u\}$. (This approximation is made explicit in Brillinger (1975a) Lemma 3.4.1, Theorems 4.5.2, 4.5.3 for example.)

SOME PARTICULAR TECHNIQUES

This section presents details concerning the implementation of regression analyses, Gaussian fitting and bispectral fitting by Fourier means. There is also brief mention of some variants in the computations.

Regression. To begin consider the case of a linear time invariant model such as

$$Y(t) = \mu + \sum_u a(u)\, X(t-u) + \varepsilon(t) \tag{19}$$

with $X(t)$, $Y(t)$ $t = 0,\dots,T-1$ data available and with $\varepsilon(t)$ a stationary mixing noise process. Taking the (finite) Fourier transform of each side of this relationship yields

$$d_Y^T(\lambda) \doteq A(\lambda)\, d_X^T(\lambda) + d_\varepsilon^T(\lambda) \tag{20}$$

for $0<\lambda<\pi$, where the approximate conversion of convolution to multiplication and the negligeability of the Fourier transform of a constant when $0<\lambda<\pi$, have been used. Consider now K frequencies, of the form $2\pi s/T$, near λ. For many filters $a(\cdot)$, $A(\cdot)$ is smooth so $A(2\pi s/T) \doteq A(\lambda)$ and one has

$$\begin{aligned} d_Y^T\left(\frac{2\pi s}{T}\right) &\doteq A(\lambda)\, d_X^T\left(\frac{2\pi s}{T}\right) + d_\varepsilon^T\left(\frac{2\pi s}{T}\right) \\ &\doteq A(\lambda)\, d_X^T\left(\frac{2\pi s}{T}\right) + \mathfrak{t}_s \end{aligned} \tag{21}$$

with the $\mathfrak{t}_s$ a sample of size K from a complex normal of mean 0 and variance $2\pi T\, f_{\varepsilon\varepsilon}(\lambda)$. The Skorokhod representation is used at the last step here. In other words

expression (21) has the essential form of a traditional regression relationship (with the minimal difference that the quantities appearing are complex-valued.) One is led directly to the estimate

$$A^T(\lambda) = \left[\sum_s d_Y^T\left(\frac{2\pi s}{T}\right) \overline{d_X^T\left(\frac{2\pi s}{T}\right)}\right]\left[\sum_s d_X^T\left(\frac{2\pi s}{T}\right) \overline{d_X^T\left(\frac{2\pi s}{T}\right)}\right]^{-1} \tag{22}$$

and to approximate the distribution of this last by a complex normal with mean $A(\lambda)$ and variance

$$f_{\varepsilon\varepsilon}(\lambda) \; [\sum_s d_X^T(\tfrac{2\pi s}{T}) \overline{d_X^T(\tfrac{2\pi s}{T})}\,]^{-1} \tag{23}$$

An estimate of $f_{\varepsilon\varepsilon}(\lambda)$, test statistics, residual plots and the like all suggest themselves. Many details may be found in Brillinger (1975a)for the time series case. Results follow in a similar manner for spatial and spatial-temporal cases, the essence is to recognize the appearance of an additive noise that one is willing to assume stationary and mixing.

<u>Nonlinear Regression</u>. Consider a model of the form

$$Y(t) = S(t|\theta) + \varepsilon(t) \tag{24}$$

where θ is a finite dimensional parameter and $\varepsilon(\cdot)$ a stationary mixing noise process. For example one might have,

$$S(t|\theta) = \sum_k \alpha_k \exp\{-\beta_k t\} \cos(\gamma_k t + \delta_k) \tag{25}$$

where $\theta = \{\alpha_k, \beta_k, \gamma_k, \delta_k \;\; k = 1,\cdots,K\}$. Taking the Fourier transform gives

$$d_Y^T(\tfrac{2\pi s}{T}) = d_S^T(\tfrac{2\pi s}{T}|\theta) + d_\varepsilon^T(\tfrac{2\pi s}{T}) \tag{26}$$

for $2\pi s/T$ near λ and the $d_\varepsilon^T(\frac{2\pi s}{T})$ approximately complex normals having mean mean 0 and variance $2\pi T\, f_{\varepsilon\varepsilon}(\lambda)$. The functional form of $d_S^T(\frac{2\pi s}{T}|\theta)$ being known up to the finite dimensional parameter θ, the problem is now one of nonlinear regression.

The particular case of the model (25) is studied in some detail in Bolt and Brillinger (1979) and Hasan (1982). The large sample distributions of the estimates are there obtained by the two step procedure described in section 2.

Gaussian Fitting. This is a procedure, essentially introduced in Whittle (1953), for the estimation of a finite dimensional parameter of a stationary process by maximizing a corresponding Gaussian likelihood. It may be motivated and described as follows: let $Y(\cdot)$ be a stationary mixing series with power spectrum, $f_{YY}(\lambda|\theta)$, containing the finite dimensional parameter θ. For s an integer write

$$I_s^T = (2\pi T)^{-1}|d_Y^T(\frac{2\pi s}{T})|^2 \quad , \quad f_s = f_{YY}(\frac{2\pi s}{T}|\theta) \quad . \tag{27}$$

I_s^T is the second-order periodogram at frequency $2\pi s/T$. Treating the $d_Y^T(\frac{2\pi s}{T})$ as exactly independent complex normals with mean 0 and variance $2\pi T f_s$ leads to, up to a constant, a negative log likelihood of

$$Q_T(\theta) = \sum_s(\log f_s + I_s^T/f_s) \tag{28}$$

The Gaussian estimate minimizes this $Q_T(\theta)$. (It is important to point out that while the specific form of the criterion (28) was motivated by Gaussian considerations, the resulting estimate may be investigated in its own right.) Properties of this estimate are developed in Hannan (1969), Brillinger (1975b) and Dzhaparidze and Yaglom (1983), for example. The estimating equations obtained by differentiation of expression (28) are

$$\sum_s (I_s^T - f_s) \frac{\partial f_s}{\partial \theta}/f_s^2 = \underset{\sim}{0} \tag{29}$$

Examination of these equations indicates that the estimate may be computed by iteratively reweighted least squares as follows: given the estimate at the previous iteration (nonlinear), regress I_s^T on f_s employing weight $\hat{f}_s^{-2}$ evaluated at the estimate of the previous iteration. Iterate til convergence.

Standard error estimates are a by product of this iterative procedure. They are appropriate in the case that fourth-order spectra of the series are negligeable; otherwise formulas like those of Section VII of Brillinger (1974) need to be employed.

An example of Gaussian fitting of runoff data is given in section 6 of this paper.

Bispectral Fitting. Gaussian fitting makes use of second-order information and statistics. When a process is nonGaussian, such a fitting procedure cannot be expected to be efficient. The procedure about to be described seeks to obtain improved estimates by incorporating third-order information. Suppose the series $Y(\cdot)$ has bispectrum $f_{YYY}(\lambda,\mu|\theta)$ also depending on θ. Write

$$I^T_{r,s} = (2\pi)^{-2}T^{-1}\, d^T_Y\left(\frac{2\pi r}{T}\right)d^T_Y\left(\frac{2\pi s}{T}\right)\overline{d^T_Y\left(\frac{2\pi(r+s)}{T}\right)} \tag{30}$$

$$f_{r,s} = f_{YYY}\left(\frac{2\pi r}{T},\frac{2\pi s}{T}\Big|\theta\right) \tag{31}$$

Bispectral estimates formed by smoothing the third-order periodograms (30) are asymptotically normal and independent of corresponding second-order quantities (see Brillinger and Rosenblatt (1967).) This suggests setting down $Q_T(\theta)$ that is the sum of the second-order term (30) and a term resulting from acting as if the third-order spectral estimates Gaussian. When this $Q_T(\theta)$ is differentiated, with respect to θ, the following system of estimating equations is obtained,

$$\sum_s (I^T_s - f_s)\frac{\partial f_s}{\partial\theta}/f_s^2 + \frac{2\pi}{T}\sum_r\sum_s (I^T_{r,s} - f_{r,s})\frac{\partial f_{r,s}}{\partial\theta}/f_r f_s f_{r+s} = \underset{\sim}{0} \tag{32}$$

The first term on the left here is the second-order one, (29) . The weights occurring in the second term correspond to the variance of the third-order periodogram as given by (16).

Examination of these equations indicates that, once again, the estimates may be computed by iteratively reweighted least squares. The regression formulation involves both the I_s^T and the $I_{r,s}^T$. Handle the I_s^T as before. Now, at the same time regress the $I_{r,s}^T$ $f_{r,s}$ employing weight $\frac{2\pi}{T}/\hat{f}_r \hat{f}_s \hat{f}_{r+s}$ evaluated at the estimate of the previous iteration. (Preparing a computer program to do this for the example presented in section 6 below did not prove enormously difficult.)

It is important to validate models and fits. In the present case the third-order fit may be examined by the standardizedquantities

$$\frac{2\pi}{T}|I_{r,s}^T - \hat{f}_{r,s}|^2/\hat{f}_r \hat{f}_s \hat{f}_{r+s} \tag{33}$$

where $\hat{f}_r = f_{YY}(\frac{2\pi r}{T}|\hat{\theta})$ and $\hat{f}_{r,s} = f_{YYY}(\frac{2\pi r}{T},\frac{2\pi s}{T}|\hat{\theta})$. An example is presented in section 6.

The asymptotic distribution of the estimates of bispectral fitting, may be worked out by the technique described in section 2. Their asymptotic variance is found to involve spectra of order up to 6. Hence the standard error estimates coming from iteratively reweighted least squares will not be appropriate generally. They will be appropriate when the higher-order spectra are negligeable relative to those of order 2. The standard errors presented in section 6 are those from iteratively reweighted least squares.

Some Elementary Modifications. It would be remiss not to point out that practical application of Fourier techniques often requires elementary preprocessing of the data. In the case that the spectrum of a process contains neighboring peaks or has a substantial dynamic fall-off, it can be crucial to taper the data prior to evaluating its Fourier transform. All that this involves is multiplying the (mean-corrected) data by a function that tapers smoothly to 0 at the boundaries of the region for which data are available and is near 1 elsewhere.

A second potent modification is prewhitening. Here there is preliminary model fitting or data processing in order to make the spectral functions more nearly constant in λ. This can lead to estimates that are substantially less biased. In the case of a bivariate process, rephasing (also known as alignment) can be crucial and it is an entirely elementary prewhitening operation. One simply shifts the time argument of one series to make the two series more nearly coherent.

In the next sections specific examples of the uses of the above tools to build analyses are presented. As part of the construction of these analyses, one seeks out stationary mixing noise processes in the situation to drive the stochastic analysis of the data.

THE ESTIMATION OF CLOUD AND STORM MOTION

Leese et al. (1970, 1971) were concerned with the determination of cloud motion from sequential pictures obtained via a geosynchronous satellite. Estimates of speeds and directions of movement were obtained by cross-correlating a picture, at various translations, with a picture taken 24 minutes earlier. If $Y(x,y,t)$ denotes the grey level of the picture element at location (x,y) at time t, then the average correlation of $Y(x+u,y+v,s)$ with $Y(x,y,0)$ across the picture, (really part of the picture), is estimated and the translation $(\hat{u}, \hat{v})$ at which this is maximized is determined. The speed of motion is then estimated by $(\hat{u}^2 + \hat{v}^2)^{1/2}/s$ and the direction estimated by $\tan^{-1}(\hat{u},\hat{v})$. The spatial array is a lattice, so the fast Fourier transform was employed in the computations. The results obtained were compared with estimates derived by manual methods and good agreement was found. Some other methods for tracking motion via pictures are described in Aggarwal et al. (1981). Examples of carrying out the computations optically are presented in Bohm et al. (1981).

Various researchers have concerned themselves with the problem of determining storm motion from rain-gauge data. We mention, initially, the papers of Felgate and Read (1975) and Shaw (1983). These workers proceed by cross-correlating the individual gauge time series in triples of gauges. Then (temporal) lags at which maximal correlation occurs are read off and used with the coordinates of the gauges to estimate parameters of interest.

Marshall (1980,1983) investigated consecutive 2-min rainfall amounts for two English storms. The networks had 19 and 36 gauges respectively (with not all operating continuously) and were arrayed irregularly. Marshall spatially interpolated the data to obtain values on a regular grid (lattice) and then looked for translations giving the maximum correlation, in the general manner of Leese *et al*. (1971). In this section a method is proposed that simultaneously makes use of all gauges (not just in triples), that does not require spatial interpolation (with its accompanying loss of information) and that is sufficiently formal that uncertainty measures may be provided along with the estimates evaluated. First the lattice case is discussed however.

To begin, here is a motivation for the use of the correlation made by Leese *et al*. . (This will later provide motivation for the procedure in the irregular array case.) To simplify the notation for the moment, let $Y_t(x,y) = Y(x,y,t)$ and suppose that data are available at $t = 0,1$ and $x = 1,\cdots,m$; $y = 1,\cdots,n$. Then the nearness of a translation of the second picture to the first may be measured by

$$\sum_x \sum_y \left[Y_1(x+u,y+v) - Y_0(x,y)\right]^2 \qquad (34)$$

with the summation over $1 \leq x,x+u \leq m$, $1 \leq y,y+v \leq n$. The translation may be estimated by minimizing this quantity. Supposing "stationarity" and that Y_1 and Y_0 have the same mean, one sees that minimizing (34) comes down essentially to finding the (u,v) that maximizes the sample correlation of $Y_1(\cdot+u,\cdot+v)$ and $Y_0(\cdot,\cdot)$. The estimate obtained in this fashion may not be expected to be efficient generally because expression (34) is a simple sum of squares. It ignores correlations between terms appearing. In order to develop an "efficient" estimate, turn to Fourier inference employing the results of Sections 2 and 3.

Consider the following formal structure: let $S(x,y)$ denote a stationary spatial process, corresponding to the signal at $t = 0$. Let

$$Y(x,y,t) = S(x+\chi t, y+\psi t) + \varepsilon(x,y,t) \tag{35}$$

with (χ,ψ) corresponding to the rate of translation of $S(\cdot)$. Suppose further that $\{\varepsilon_0(x,y), \varepsilon_1(x,y)\}$ is a stationary spatial process, the two components having coherence 0. Then, from (35) and the discussion in Section 3 on Fourier transforms of convolutions, with $\alpha_j = 2\pi j/m$ and $\beta_k = 2\pi k/n$ $T = (m,n)$

$$d_Y^T(\alpha_j,\beta_k;t) \doteq \exp\{i(\alpha\chi + \beta\psi)t\} d_S^T(\alpha_j,\beta\ ) + d_\varepsilon^T(\alpha_j,\beta_k;t) \tag{36}$$

for (α_j,β_k) near (α,β) and with the $d_\varepsilon^T(\alpha_j,\beta_k;t)$ approximately independent complex normal with mean 0 and variance $(2\pi)^2 mnf_{\varepsilon\varepsilon}(\alpha,\beta;t)$. Set $d_t^T(\alpha,\beta) = d_Y^T(\alpha,\beta;t)$. Then the variates

$$d_1^T(\alpha_j,\beta_k) - \exp\{i(\alpha\chi + \beta\psi)\} d_0^T(\alpha_j,\beta_k) \tag{37}$$

will be approximately independent complex normals having mean 0 and variance $(2\pi)^2 mn[f_{\varepsilon\varepsilon}(\alpha,\beta;1) + f_{\varepsilon\varepsilon}(\alpha,\beta;0)]$. Restricting, for the moment, consideration to Fourier values in the neighborhood of (α,β) one is led to estimate (χ,ψ) by minimizing

$$\sum_j \sum_k |d_1^T(\alpha_j,\beta_k) - \exp\{i(\alpha\chi + \beta\psi)\} d_0^T(\alpha_j,\beta_k)|^2$$

$$\text{or} \quad f_{11}^T(\alpha,\beta) - 2\,\mathrm{Re}\Big(\exp\{i(\alpha\chi + \beta\psi)\} f_{01}^T(\alpha,\beta)\Big) + f_{00}^T(\alpha,\beta) \tag{38}$$

where the $f^T(\alpha,\beta)$ denote spectral estimates at wavenumber (α,β) . Turning to the construction of an estimate involving all Fourier frequencies, let $\phi^T(\alpha,\beta) = \arg\{f_{10}^T(\alpha,\beta)\}$. Noting expression (38) and the dependence of the complex normal variances on (α,β) one is led to consider estimating (χ,ψ) by maximizing , for large P,

$$\sum_{p,q=1}^{P} \cos\left(\frac{2\pi p}{P}\chi + \frac{2\pi q}{P}\psi - \phi^T\left(\frac{2\pi p}{P},\frac{2\pi q}{P}\right)\right) w\left(\frac{2\pi p}{P},\frac{2\pi q}{P}\right) \tag{39}$$

for some weight function $w(\cdot)$. Now the variance of the asymptotic distribution of $\phi^T(\alpha,\beta)$ is proportional to $[|R_{10}(\alpha,\beta)|^{-2} - 1]$, with $|R_{10}(\cdot)|^2$ the coherence of the processes $Y_1(\cdot)$ and $Y_0(\cdot)$. This suggests taking $w(\cdot)$ in (39) to be

$$w^T(\alpha,\beta) = [\, |R_{10}^T(\alpha,\beta)|^{-2} - 1]^{-1} \tag{40}$$

based on estimated coherences. Note that with $w(\cdot) = |f_{10}^T(\cdot)|$, expression (39) is essentially the sample cross-correlation function and one has the Leese <u>et al</u>. estimation procedure.

An expression for the variance of the estimate constructed from (39) may be set down directly. One has $\phi^T(\alpha,\beta) \sim \alpha\chi + \beta\psi$ with an asymptotic variance $[|R_{10}(\alpha,\beta)|^{-2} - 1]/2K$, K being the number of periodogram values averaged in forming the spectral estimates. This gives the asymptotic covariance matrix of (χ,ψ) , by generalized least squares, as $\pi \underset{\sim}{A}^{-1}\underset{\sim}{B}\underset{\sim}{A}^{-1}/KP$ where

$$\underset{\sim}{A} = \iint [\alpha\ \beta]^{\tau} [\alpha\ \beta]\ w(\alpha,\beta)\ d\alpha\ d\beta$$
$$\underset{\sim}{B} = \iint [\alpha\ \beta]^{\tau} [\alpha\ \beta]\ w(\alpha,\beta)^2 [\, |R_{10}(\alpha,\beta)|^{-2} - 1]\ d\alpha\ d\beta \tag{41}$$

from which results for the Leese et al. procedure and the "efficient" procedure may be obtained by choice of $w(\cdot)$.

The development, just provided, is a direct extension to the spatial case of results developed by Hannan and Thomson (1973) and Hamon and Hannan (1974) using results developed for spatial series in Brillinger (1970,1974). Related work by Hannan for the time series case may be found in: Hannan (1975), Cameron and Hannan (1978,1979), Hannan (1983). A point that has been emphasized by Hannan in the time series case is that, in the Fourier approach it can be essential to rephase the series, that is realign them to put them approximately in phase, before commencing spectral computations. Such a "prewhitening" operation is called for in the present case as well. One means of estimating the realignment translation is via the values maximizing the cross-correlation of the two pictures.

In connection with the processes for which the above estimation procedure may prove useful, note that expression (36′) was basic. It involved replacement of the Fourier transform of the translated signal, $S(\cdot)$, by a simple multiple of the untranslated signal's Fourier transformation. This replacement may be expected to be reasonable for a broad class of processes, including transients, and was noted in Section 2.

Turn now to the case of an irregular array, and proceed by setting down an analog of expression (34) . Let the coordinates of the array sensors be denoted $\underset{\sim}{r}_j = (x_j, y_j)$. Let translations be denoted by $\underset{\sim}{\rho} = (u,v)$ and let Δ be small. Now the nearness of a translation, $\underset{\sim}{\rho}$, of the second image to the first may be measured by

$$\sum_{|\underset{\sim}{r}_j-\underset{\sim}{r}_k-\underset{\sim}{\rho}|<\Delta} |Y_1(\underset{\sim}{r}_j) - Y_0(\underset{\sim}{r}_k)|^2 \Big/ \sum_{|\underset{\sim}{r}_j-\underset{\sim}{r}_k-\underset{\sim}{\rho}|<\Delta} 1 \tag{42}$$

with the summation over the available data. The unknown translation may be estimated by minimizing (42) with respect to $\underset{\sim}{\rho}$, for a given Δ. Expression (42) might be generalized to the form

$$\sum_{j,k} w^T(\underset{\sim}{r}_j-\underset{\sim}{r}_k-\underset{\sim}{\rho}) \; |Y_1(\underset{\sim}{r}_j) - Y_0(\underset{\sim}{r}_k)|^2 \Big/ \sum w^T(\underset{\sim}{r}_j-\underset{\sim}{r}_k-\underset{\sim}{\rho}) \tag{43}$$

with $w^T(\cdot)$ a weight function concentrated near 0. (Masry (1983) considers a related time series covariance function estimate.)

The estimates obtained in this fashion are ordinary least squares, and hence may not be expected to be efficient generally. In the estimation of the covariance function, at lag u, of an ordinary time series one issue that arises is whether to divide the sum of lagged sample products by $T-|u|$ or by T. It seems to be the case that the latter choice is better in a variety of situations, particularly when the population covariance function is tending to 0 as $|u| \to \infty$. In the present situation, this leads to consideration of expressions (42), (43) multiplied by say $(1 - |\underset{\sim}{\rho}|/T)$ with T measuring the extent of the array. For larger $\underset{\sim}{\rho}$, the sum (42) has fewer terms and hence greater variability. The multiplier reduces the variability.

It is to be noted that this last procedure has not made use of a Fourier transform. There has been some study of the Fourier transforms of irregularly distributed observations, see Brillinger (1972) and Dunsmuir and Robinson (1981), but setting down a Fourier procedure here would be premature.

Two further hydrology references are Johnson and Bras (1979) and Amorocho (1981).

PASSIVE SONAR

Turn now to a class of situations exemplified by passive sonar. Let the data available be $Y(x_j, y_j, t)$, $j=1,\cdots,J$ and $t = 0,\cdots,T-1$. In contrast with the assumptions of the previous section, in the present case J will be assumed moderate and T large. Suppose that a "wave " is moving across the array from the far field and that it is desired to estimate the velocity of the wave and the direction from which it is coming. If the wave may be viewed as plane, then a model for the situation might be

$$Y(x,y,t) = \rho \cos(\alpha x+\beta y+\gamma t+\delta) + \varepsilon(x,y,t) \tag{44}$$

with the direction of travel specified by $\alpha = |\underset{\sim}{k}|\cos\phi$, $\beta = |\underset{\sim}{k}|\sin\phi$, where $\underset{\sim}{k} = (\alpha,\beta)$ is the wavenumber , and with the (phase) velocity given by $\gamma/|\underset{\sim}{k}|$. In what follows γ will be thought of as known. This comes about either from the collected data having been narrow-band temporally filtered at frequency γ, or from γ having been precisely estimated, T being large. The principal unknowns are (α,β).

A traditional means of estimating (α,β) is <u>beamforming</u>. Here one determines (α,β) to maximize

$$\left| \sum_j \sum_t Y(x_j,y_j,t) \exp\{-i(\alpha x_j+\beta y_j+\gamma t)\} \right|^2 \tag{45}$$

(see for example Knight *et al*. (1981). Now, we investigate this problem by the method of Fourier inference, in particular we determine the large sample distribution of the beamformed (or least squares) estimate and the maximum likelihood estimate.

The details of the results may be found in Brillinger (1985). The point of presenting it here is to: i) set down some results of practical importance and ii) to show how those results follow directly via Fourier inference.

Let $\underset{\sim}{Y}(t)$, $\underset{\sim}{\varepsilon}(t)$ denote the J vectors $[Y(x_j,y_j,t)]$, $[\varepsilon(x_j,y_j,t)]$ respectively. Let

$$\underset{\sim}{Y}_k = T^{-1}\sum_{t=0}^{T-1} \underset{\sim}{Y}(t) \exp\{-i2\pi kt/T\} \tag{46}$$

with a similar definition for $\underset{\sim}{\varepsilon}_k$, k being an integer. Suppose that the temporal frequency γ has the form $2\pi k'/T$, k' an integer $\neq 0$. From (44) one sees that

$$Y_{j,k'} \doteq \frac{\rho}{2} e^{i\delta} \exp\{i(\alpha x_j+\beta y_j)\} + \varepsilon_{j,k'} \tag{47}$$

while for $k \neq k'$, $Y_{j,k} \doteq \varepsilon_{j,k}$. Supposing one takes K frequencies of the form $2\pi k/T$ near γ, from the discussion in Section 2, the corresponding $\underset{\sim}{\varepsilon}_k$ are approximately independent complex normal variates with mean $\underset{\sim}{0}$ and covariance matrix $\frac{2\pi}{T}\underset{\sim}{f}_{\varepsilon\varepsilon}(\gamma)$. Now ordinary least squares estimates of α, β are seen to correspond to minimizing

$$\sum_j |Y_{j,k'} - \frac{\rho}{2} e^{i\delta} \exp\{i(\alpha x_j+\beta y_j)\}|^2 \tag{48}$$

with respect to α, β, δ, ρ or , asymptotically, to maximizing

$$|\sum_j Y_{j,k'} \exp\{-i(\alpha x_j+\beta y_j)\}|^2 \tag{49}$$

with respect to α, β . Let $\underset{\sim}{B}$ denote the J vector $[\exp\{i(\alpha x_j+\beta y_j)\}]$ and

$$\underset{\sim}{S} = \sum_{k\neq k'} \underset{\sim}{Y}_k\overline{\underset{\sim}{Y}}_k^{\tau} \tag{50}$$

(This last is proportional to an estimate of $\frac{2\pi}{T}\underset{\sim}{f}_{\varepsilon\varepsilon}(\gamma)$. Then the generalized least squares estimate corresponds to minimizing

$$\overline{(\underset{\sim}{Y}_{k'} - \underset{\sim}{B}\frac{\rho}{2} e^{i\delta})}^{\tau}\underset{\sim}{S}^{-1}(\underset{\sim}{Y}_{k'} - \underset{\sim}{B}\frac{\rho}{2} e^{i\delta}) \tag{51}$$

After some algebra, (see Brillinger (1985)), this last is seen to correspond to chosing (α,β) to maximize

$$|(\bar{\underset{\sim}{B}}^{\tau}\underset{\sim}{S}^{-1}\underset{\sim}{Y}_{k'})|^2/(\bar{\underset{\sim}{B}}^{\tau}\underset{\sim}{S}^{-1}\underset{\sim}{B}) \tag{52}$$

and bears a direct relationship to the Capon (1969) high resolution spectral estimate. In practice it is convenient to prepare contour plots of the quantity (52) as a function of (α,β).

The large sample distributions of the ordinary and generalized least squares estimates may be derived directly from expression (47) and the asymptotic normality of the $\underset{\sim}{\varepsilon}_k$. Let V_j denote the principal value of $\log Y_{j,k'}$. Then (47) leads to

$$V_j \doteq \log\frac{\rho}{2} + i\delta + i\alpha x_j + i\beta y_j + \frac{1}{\rho}\zeta_j \tag{53}$$

where $\underset{\sim}{\zeta} = [\zeta_j]$ is asymptotically complex normal with mean $\underset{\sim}{0}$, covariance matrix $\frac{2\pi}{T}\underset{\sim}{f}_{\varepsilon\varepsilon}(\gamma)$, asymptotically independent of the $\underset{\sim}{\varepsilon}_k$, $k \neq k'$. Let $\underset{\sim}{V} = [V_j]$ and let $\underset{\sim}{X}$ denote the $J \times 3$ matrix $[1\ x_j\ y_j]$. Then the ordinary least squares and the maximum likelihood estimates of $(\log\frac{\rho}{2} + i\delta, i\alpha, i\beta)$ in the model obtained from the above by replacing the $\underset{\sim}{\varepsilon}_k$ and $\underset{\sim}{\zeta}$ by normal variates (as is possible through the Skorokhod representation) are given by $(\underset{\sim}{X}^{\tau}\underset{\sim}{X})^{-1}\underset{\sim}{X}^{\tau}\underset{\sim}{V}$ and $(\underset{\sim}{X}^{\tau}\underset{\sim}{S}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}^{\tau}\underset{\sim}{S}^{-1}\underset{\sim}{V}$ respectively. Their covariance matrices may be estimated by $\hat{\rho}^{-2}(\underset{\sim}{X}^{\tau}\underset{\sim}{X})^{-1}\underset{\sim}{X}^{\tau}\underset{\sim}{S}\underset{\sim}{X}(\underset{\sim}{X}^{\tau}\underset{\sim}{X})^{-1}$ and $\hat{\rho}^{-2}(\underset{\sim}{X}^{\tau}\underset{\sim}{S}^{-1}\underset{\sim}{X})^{-1}(K-2)/(K-J-1)$ respectively. For this last to be reasonable, one needs $K > J+1$.

BISPECTRAL FITTING OF MISSISSIPPI RIVER RUNOFF 1861-1880

The bispectral fitting procedure introduced in Section 3 was employed in a preliminary study of river runoff, such data being often described as nonGaussian (eg. Lawrance and Kottegoda (1977).) The data available was monthly runoff at Eads Bridge, St. Louis fom January 1861 to Sept. 1961. In the present study only the first twenty years of data were used. Figure 1 presents the data in the form of Chernoff faces, Chernoff (1973). In this display the respective monthly values correspond to different features of the faces,eg. July corresponds to separation of the eyes (see the figure caption for the other correspondences.) Displays such as this are proving useful in throwing up suprising aspects of data. In the present case one notes that large year to year variation is present. Figure 2 is the traditional plot of the data. Examination of this figure suggests that the series is neither time reversible, (that is Y(-t) has the same distribution as Y(t)), nor symmetrically distributed. These are both necessary properties of Gaussian processes.

Seasonal variation is a pronounced feature of runoff data. Its nature is reasonably well understood. In the present study seasonal variation was"removed"by subtracting from individual monthly values, the average level for that month across the whole data set. Figure 3 presents an estimate of the density function of the first twenty years data, monthly means removed. (The estimate was computed via the procedure "density" of Becker and Chambers (1984).) Figure 4 is a normal probability plot for the same data. There are substantial indications of nonGaussianity. Figure 5 is a plot of a month's value versus the previous month's. Again nonGaussianity is suggested. In summary, this data seems a plausible candidate for bispectral fitting.

Let $Y(t)$ denote the seasonally adjusted value at time t, where t indexes the monthly values from 1861 through 1880. An autoregressive of order 2 was fit to this data by the method of Gaussian estimation, as described in Section 3. This process is described by

$$Y(t) + \alpha_1 Y(t-1) + \alpha_2 Y(t-2) = \varepsilon(t) \qquad (54)$$

where $\varepsilon(\cdot)$ is a white noise series with mean 0 and variance σ^2. The power spectrum of this process is given by

$$f_{YY}(\lambda) = \frac{\sigma^2}{2\pi}|1 + \alpha_1 \exp\{-i\lambda\} + \alpha_2 \exp\{-i2\lambda\}|^{-2} . \qquad (55)$$

The estimates of the parameters, and corresponding standard error estimates, are given in Table 1. The value of α_2 appears negligeable, but it will be retained for the analyses,as doing so causes no difficulty. Figure 6 is a plot of the second-order periodogram (27) and the corresponding fitted power spectrum as determined from expression (55). In order to assess the goodness-of-fit more formally an exponential probability plot of the $I_s^T/\hat{f}_s$ values was prepared. This is given as Figure 7. There is no suggestion of substantial departure from fit.

In expression (28) s ran from 1 to $T/2$. The autoregressive of order 2 was taken to begin. As indicated it fit reasonably well so no higher orders were studied. The exponential distribution of the periodogram follows from the complex normality of the Fourier values, see Theorem 5.2.6, Brillinger (1975a).

Next bispectral fitting of the model (54), with the additional assumption that $E\,\varepsilon(t)^3 = \gamma\sigma^3$, was carried out. The bispectrum of the process is given by

$$f_{YYY}(\lambda,\mu) = \frac{\gamma\sigma^3}{(2\pi)^2 A(\lambda) A(\mu) \overline{A(\lambda+\mu)}} \qquad (56)$$

where $A(\lambda) = 1 + \alpha_1 \exp\{-i\lambda\} + \alpha_2 \exp\{-i2\lambda\}$. The estimates of the parameters, and corresponding standard error estimates, are given in Table 2. It is to be noted that the standard errors have become smaller with the addition of the third-order information in the cases of α_1 and α_2 . That the estimate of σ and its s.e. are essentially the same for both fits probably results from the fact that σ is a second-order parameter, the new third-order information appears in γ . It is noteworthy too that the estimate of γ is 6.9 times its standard error, confirming evidence of nonGaussianity. Figure 8 is a contour plot of the modulus of the estimated bispectrum (estimate formed by averaging 15 periodograms.) Figure 9 is the corresponding fitted form, evaluated from (56). Figures10 and 11 are corresponding perspective plots. There is real agreement between the estimate and fit. In order to examine the goodness-of-fit in a more sensitive manner, standardized residuals were computed. Figures 12 and 13 are contour and perspective plots of the log quantities (33) . If the model is reasonable, then the distribution of these will be approximately exponential with mean 1 . There is no strong evidence of departures.

In summary, one can say that the bispectral fitting procedure has proved itself feasible, but that the estimates of second-order parameters were not dramatically improved, although a further parameter has been able to be estimated. The validations of the model provided by the contour and perspective plots do seem important.

DISCUSSION

The intent of this paper has been to set out some fundamental properties of the empirical Fourier transform and to illustrate how those properties could be used to build statistical analyses of some specific data sets. The processes involved may concern time functions, space functions or spatial-temporal functions (or even point processes.) The statistics computed may be linear, quadratic, cubic or more complicated. The analyses may be linear or nonlinear. The situation may be modeled via a finite dimensional parameter or not. Use of the Fourier transform transcends these issues. It converts convolution (filtering) into multiplication and it converts serial and spatial dependence into approximate independence and it does this latter in a fashion that traditional statistical procedures can often then be invoked. It is useful for both fitting and validation problems.

There are a host of other problems that can be approached via Fourier inference. These include: kriging, detection of change, analysis of extremes, fitting state space models and extrapolation/forecasting. There are many statistical procedures that have useful Fourier implementations. These include: discriminant analysis, principal components, empirical Bayes, Stein estimation and penalized likelihood fitting. The properties of Fourier procedures may be studied when there is long range dependence present in the process and when the model is false. There is insufficient space to do these things here, but hopefully the way forward to doing them is clear. Shumway (1984) is one reference that may be mentioned.

Acknowledgments. This work has been supported by the National Science Foundation Grant DMS-8316634.

I would like to thank R. Dahlhaus, P. Guttorp, R. Ihaka, F. O'Sullivan and D. Woolhiser for helpful comments. Julian Faraway carried out some preliminary computations with the bispectral fitting method.

REFERENCES

Aggarwal, J.K., L.S. Davis and W.N. Martin, Correspondence processes in dynamic scene analysis, Proc. IEEE, 69, 562-572, 1981.

Amoroche, J, Stochastic modelling of precipitation in space and time, Proc. Internat. Symp. on Rainfall-Runoff Modelling (Ed. V. P. Singh), Water Resources, Littleton, 3-20, 1981.

Arking, A., Lo, R. C. and A. Rosenfeld, A Fourier approach to cloud motion estimation, J. Applied Met. 17, 735-744, 1978.

Becker, R. A. and Chambers, J. M. , S, Wadsworth, Belmont, 1984.

Bohm, H., A.W. Lohmann, and G. Weigelt, Method for measuring the average motion of a large number of biological objects, Optics Letters, 6, 162-164, 1981.

Bolt, B.A. and D.R. Brillinger, Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series, Geophys. J. R. Astr. Soc., 59, 593-603, 1979.

Brillinger, D.R., The frequency analysis of relations between stationary spatial series, Proc. Twelfth Bien. Sem. Canadian Math. Congr. (Ed. R. Pyke), Canadian Math. Congress, Montreal, 39-81, 1970.

Brillinger, D.R., The spectral analysis of stationary interval functions, Proc. Sixth Berkeley Symp. Math. Statist. Prob. (Eds. L.M. LeCam, J. Neyman and E.L. Scott), Univ. Calif. Press, Berkeley, 483-513, 1972.

Brillinger, D. R., The Fourier analysis of stationary processes, Proc. IEEE, 62, 1628-1643, 1974.

Brillinger, D.R., Time Series: Data Analysis and Theory, Holt, Rinehart and Winston, New York, 1975a.

Brillinger, D.R., Statistical inference for stationary point processes, Stochastic Processes and Related Topics 1 (Ed. M.L. Puri), Academic, New York, 55-99, 1975b.

Brillinger, D. R., The finite Fourier transform of a stationary process, Handbook of Statistics 3 (Eds. D.R. Brillinger and P.R. Krishnaiah), North-Holland, Amsterdam, 21-37, 1983.

Brillinger, D.R., A maximum likelihood approach to frequency-wavenumber analysis, to appear in IEEE Trans. Acoust. Speech Sig. Proc., 1985.

Brillinger, D.R. and P.R. Krishnaiah, Eds., Handbook of Statistics 3, North-Holland, Amsterdam, 1983.

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

Cameron, M.A. and E.J. Hannan, Measuring the properties of plane waves, Math. Geology, 10, 1-22, 1978.

Cameron, M.A. and E.J. Hannan , Transient signals, Biometrika, 66, 243-258, 1979.

Capon, J., High-resolution frequency-wavenumber spectrum analysis, Proc. IEEE, 58, 760-770, 1970.

Chernoff, H., The use of faces to represent points in k-dimensional space graphically, J. Amer. Statist. Soc., 68, 361-368, 1973.

Dunsmuir, W. and P.M. Robinson, Parametric estimators for stationary time series with missing observations, Adv. Appl. Prob., 13, 129-146.

Dzhaparidze, K.O. and A.M. Yaglom, Spectrum parameter estimation in time series analysis, Developments in Statistics 4 (Ed. P.R. Krishnaiah), Academic, New York, 1-96, 1983.

Felgate, D.G. and D.G. Read, Correlation analysis of the cellular structure of storms observed by raingauges, J. Hydrol., 24, 191-200, 1975.

Forristall, G. Z., Ward, E. G., Borgman, L. E. and Cardone, V. J., Storm wave kinematics, Proc. Offshore Tech. Conf. , 1503-1514, 1978.

Green, P.J., Iteratively reweighted least squares for maximum likelihood estimation, J. R. Statist. Soc. B , 46, 149-192, 1984.

Hamon, B.V. and E.J. Hannan, Spectral estimation of time delay for dispersive and non-dispersive systems, J. Royal. Stat. Soc. C, 23, 134-142, 1974.

Hannan, E.J., The estimation of mixed autoregressive-moving average systems, Biometrika, 56, 579-592, 1969.

Hannan, E.J., Multiple Time Series, J.Wiley, New York, 1970.

Hannan, E.J., Measuring the velocity of a signal, Perspectives in Probability and Statistics (Ed. J. Gani), Academic, New York, 227-237, 1975.

Hannan, E.J., Signal estimation, Handbook of Statistics 3 (Eds. D.R. Brillinger and P.R. Krishnaiah), North-Holland, Amsterdam, 111-124, 1983.

Hannan, E.J. and P.J. Thomson, Estimating group delay, Biometrika, 60, 241-253, 1973.

Hasan, T., Nonlinear time series regression for a class of amplitude modulated cosinusoids, J. Time Series Analysis, 3, 109-122, 1982.

Heideman, M. T., Johnson, D. H. and Burrus, C. S., Gauss and the history of the fast Fourier transform, IEEE ASSP Mag. , 14-21, October 1984.

Johnson, E. R. and R. L. Bras, Real-time estimation of velocity and covariance structure of rainfall events using telemetered raingage data - a comparison of methods, J. Hydrol., 44, 97-123, 1979.

Knight, W.C., R.G. Pridham and S.M. Kay, Digital signal processing for sonar, Proc. IEEE, 69, 1451-1506, 1981.

Lawrance, A.J. and N.T. Kottegoda, Stochastic modelling of riverflow series, J. R. Statist. Soc. A, 140, 1-38 , 1977.

Leese, J.A., C.S. Novak and B.B. Clark, An automated technique for obtaining cloud motion from geosynchronous satellite data using cross correlation, J. Appl. Meteor., 10, 118-132, 1971.

Leese, J.A., C.S. Novak and V.R. Taylor, The determination of cloud pattern motions from geosynchronous satellite image data, Pattern Recognition, 2, 279-292, 1970.

Marshall, R.J., A spatial-temporal model of storm rainfall, J. Hydrol., 62, 53-62, 1983.

Marshall, R.J., The estimation and distribution of storm movement and storm structure, using a correlation analysis technique and rain-gauge data, J. Hydrol., 48, 19-39, 1980.

Masry, E., Nonparametric covariance estimation from irregularly-spaced data, Adv. Appl. Prob., 15, 113-132, 1983.

Shaw, S.R., An investigation of the cellular structure of storms using correlation teohniques, J. Hydrol., 62, 63-79, 1983.

Shumway, R. H., Time series in the soil sciences: is there life after kriging? Proc. Internat. Soc. Soil Sci. and Soil Sci. Soc. America Workshep on Soil Spatial Variability, Centre for Agricultural Publ., Wageningen, 1984.

Skorokhod, A.V., Limit theorems for stochastic processes, Theor. Probability Appl., 1, 261-290, 1956.

Whittle, P., The analysis of multiple stationary time-series, J. R. Statist. Soc. B, 15, 125-139, 1953.

Table 1

Second-order fit only

parameter	α_1	α_2	σ
estimate	−.61743	−.00551	3561.3
s.e.	.06511	.06511	163.3

Table 2

Second- and third-order fit

parameter	α_1	α_2	σ	γ
estimate	−.63408	−.01181	3563.4	1.2443
s.e.	.05388	.04982	163.3	.1802

Figure Captions

Figure 1: Jan. - area of face, Feb. - shape of face, Mar. - length of nose, Apr. - location of mouth, May - curve of smile, June - width of mouth, July - separation of eyes, Aug. - angle of eyes, Sept. - shape of eyes, Oct. - width of eyes, Nov. - location of eyes, Dec. - location of pupil (See Becker and Chambers (1984).)

Figure 2: Mississippi River runoff, 1861-1880, monthly means removed.

Figure 3: Density estimate, Mississippi River runoff, 1861-1880, monthly means removed.

Figure 4: Normal probability plot, data of Figure 2.

Figure 5: Scatter diagram of successive monthly values, data of Figure 2.

Figure 6: Log periodogram and fitted autoregressive of order 2, data of Figure 2.

Figure 7: Exponential probability plot of standardized periodogram values, data of Figure 2.

Figure 8: Log modulus of estimated bispectrum, data of Figure 2.

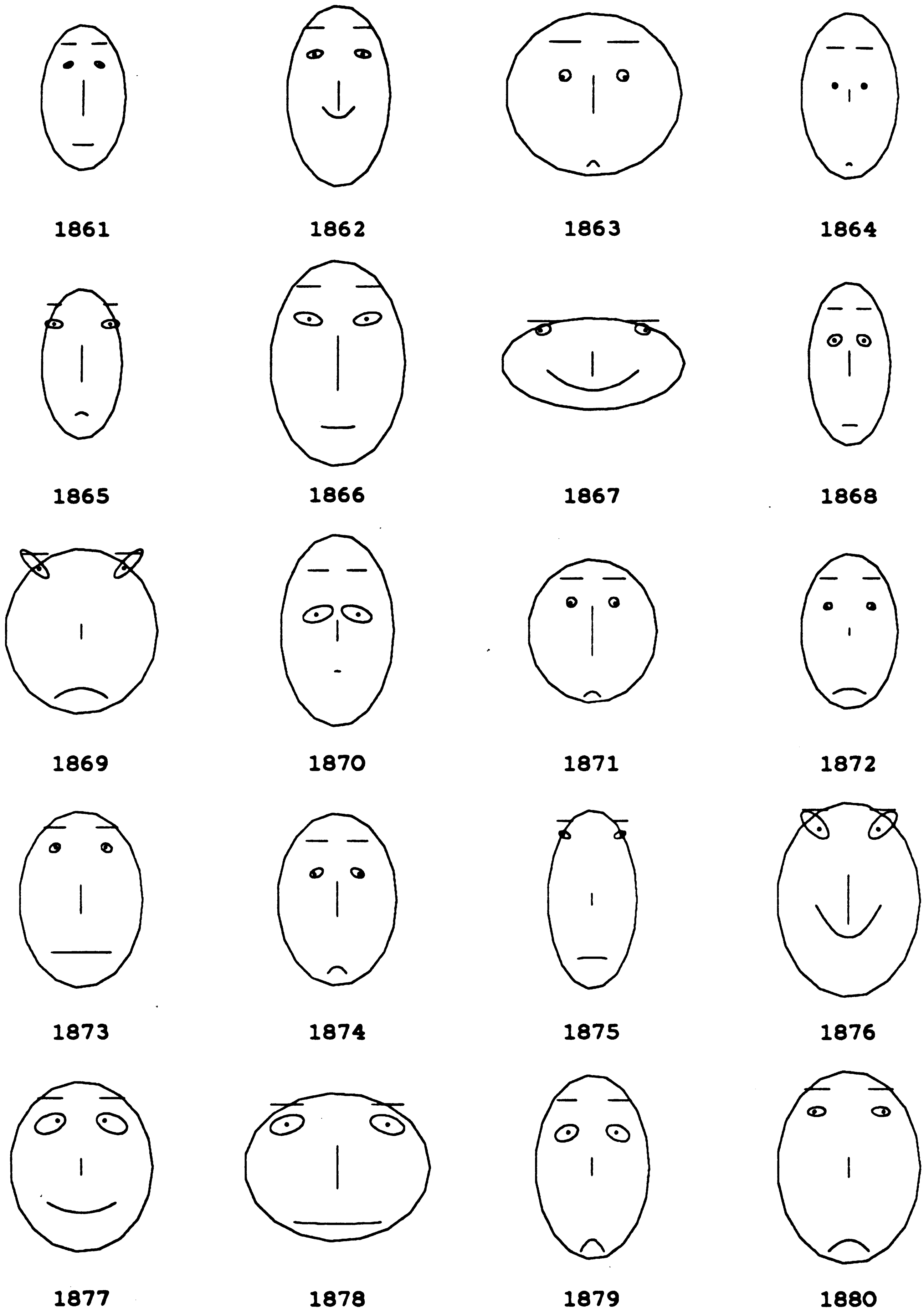
Figure 9: Log modulus of fitted bispectrum.

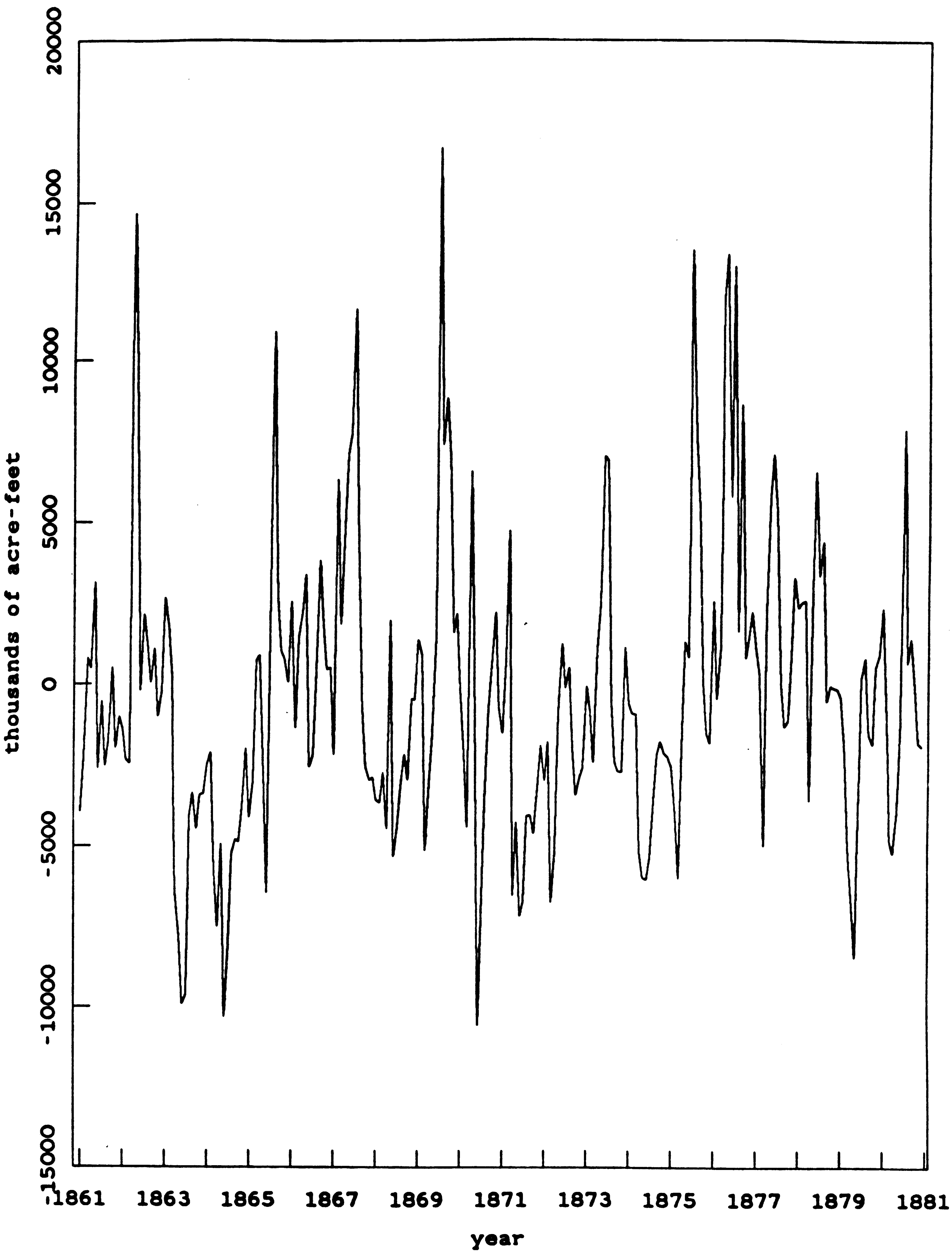
Figure 10: Perspective plot corresponding to Figure 8.

Figure 11: Perspective plot corresponding to Figure 9.

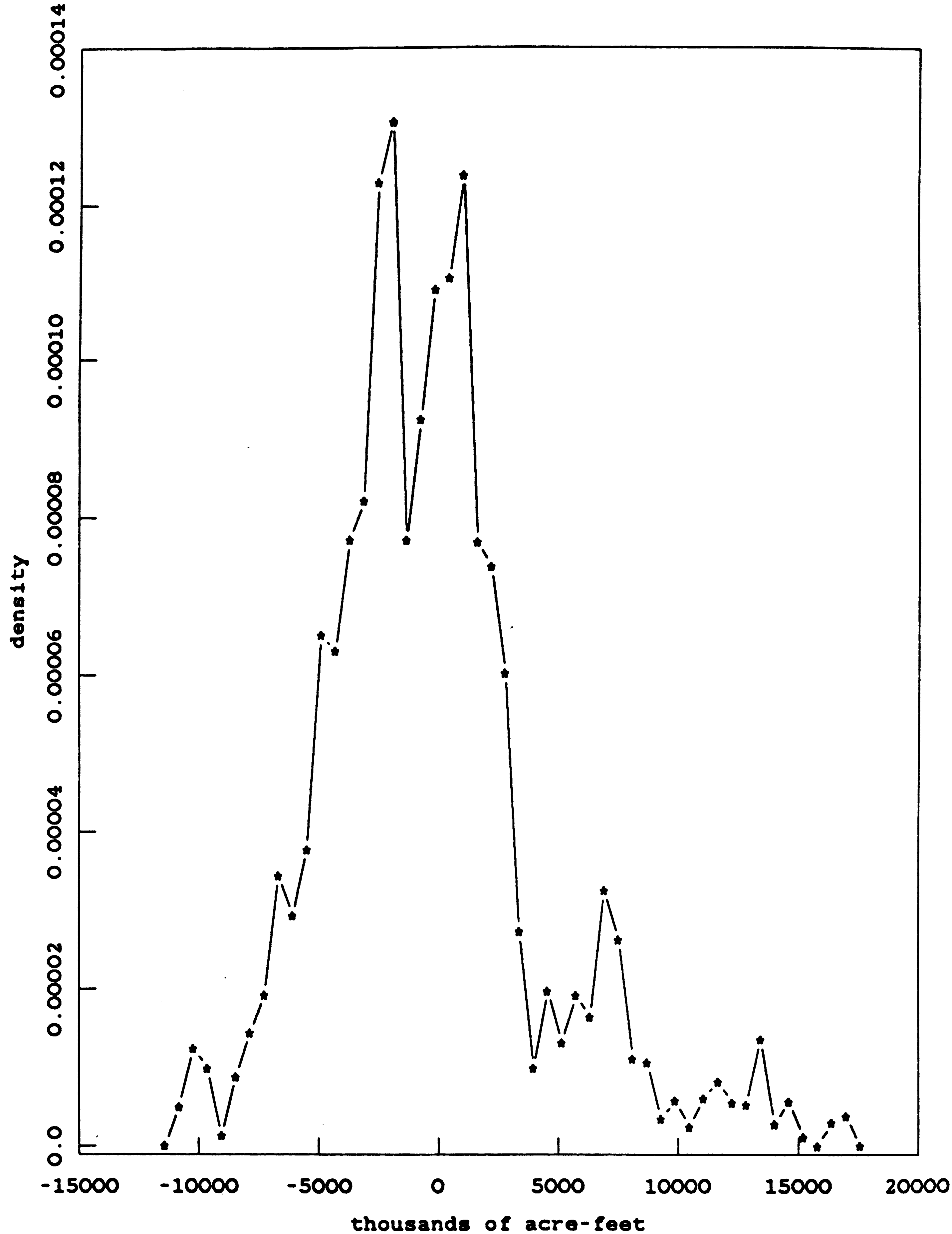
Figure 12: Contour plot of log modulus-squared of standardized residual.

Figure 13: Perspective plot corresponding to Figure 12.

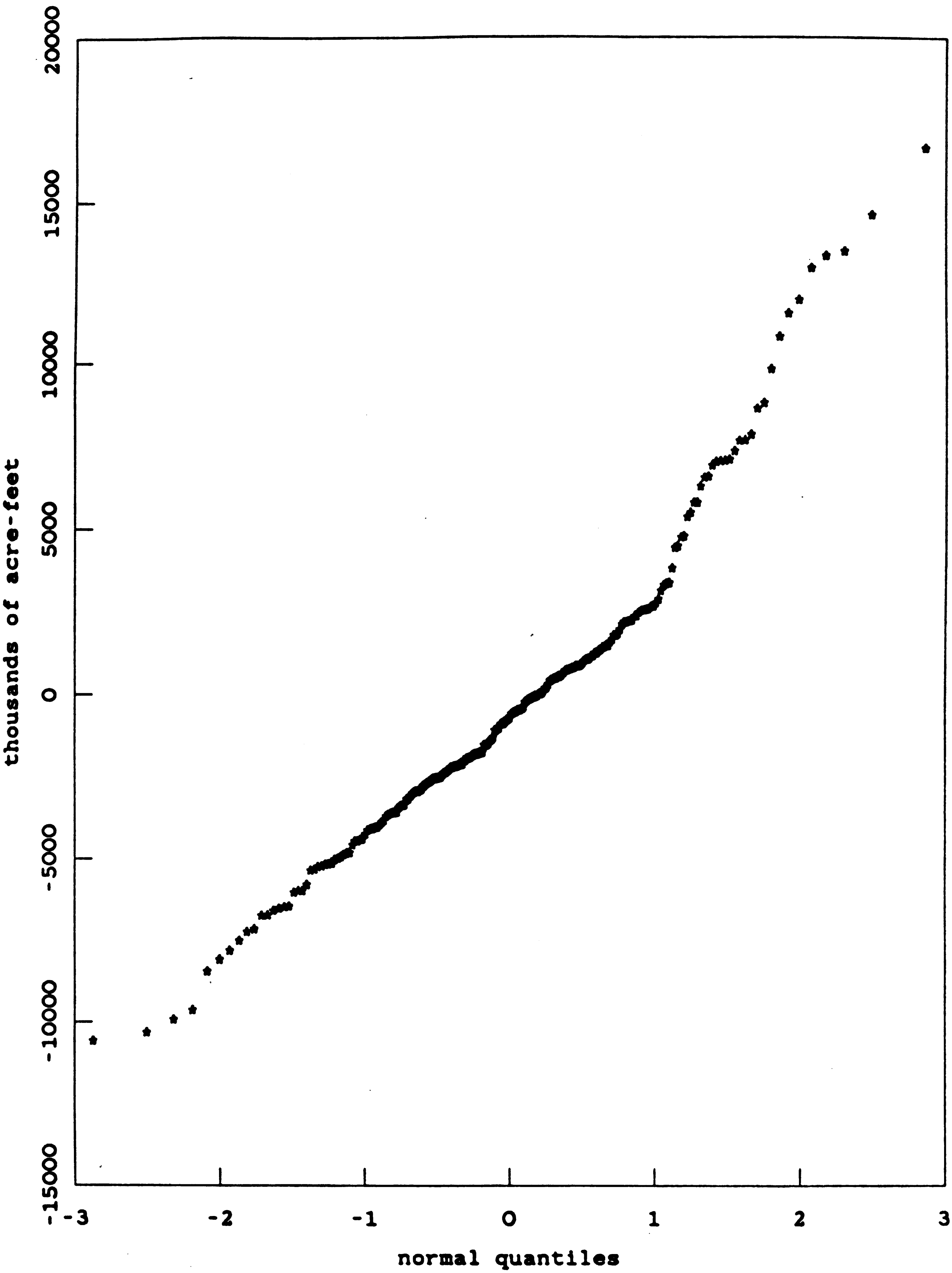

1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880



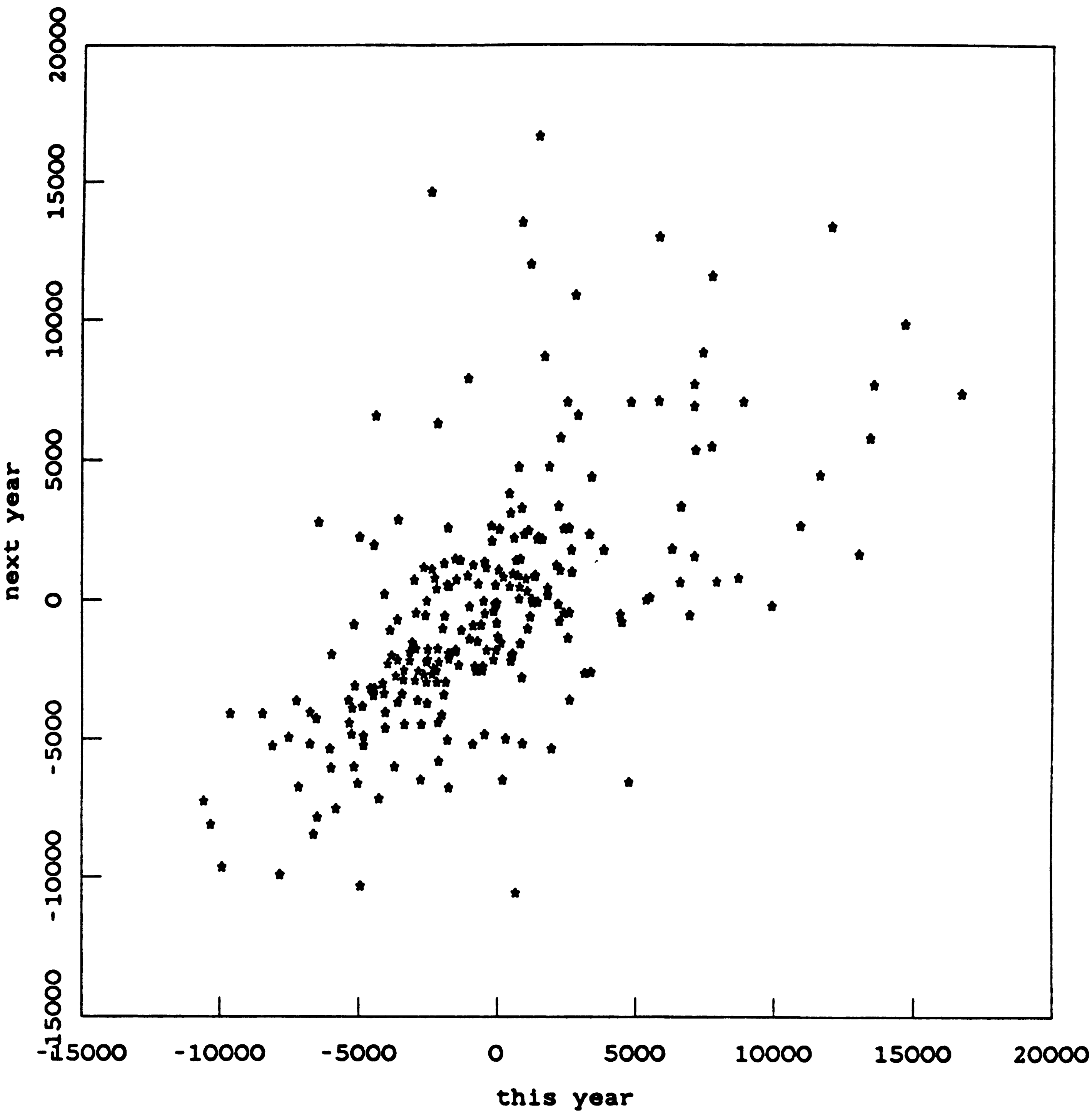
20000
15000
10000
5000
0
-5000
-10000
-15000
thousands of acre-feet
1861
1863
1865
1867
1869
1871
1873
1875
1877
1879
1881
year



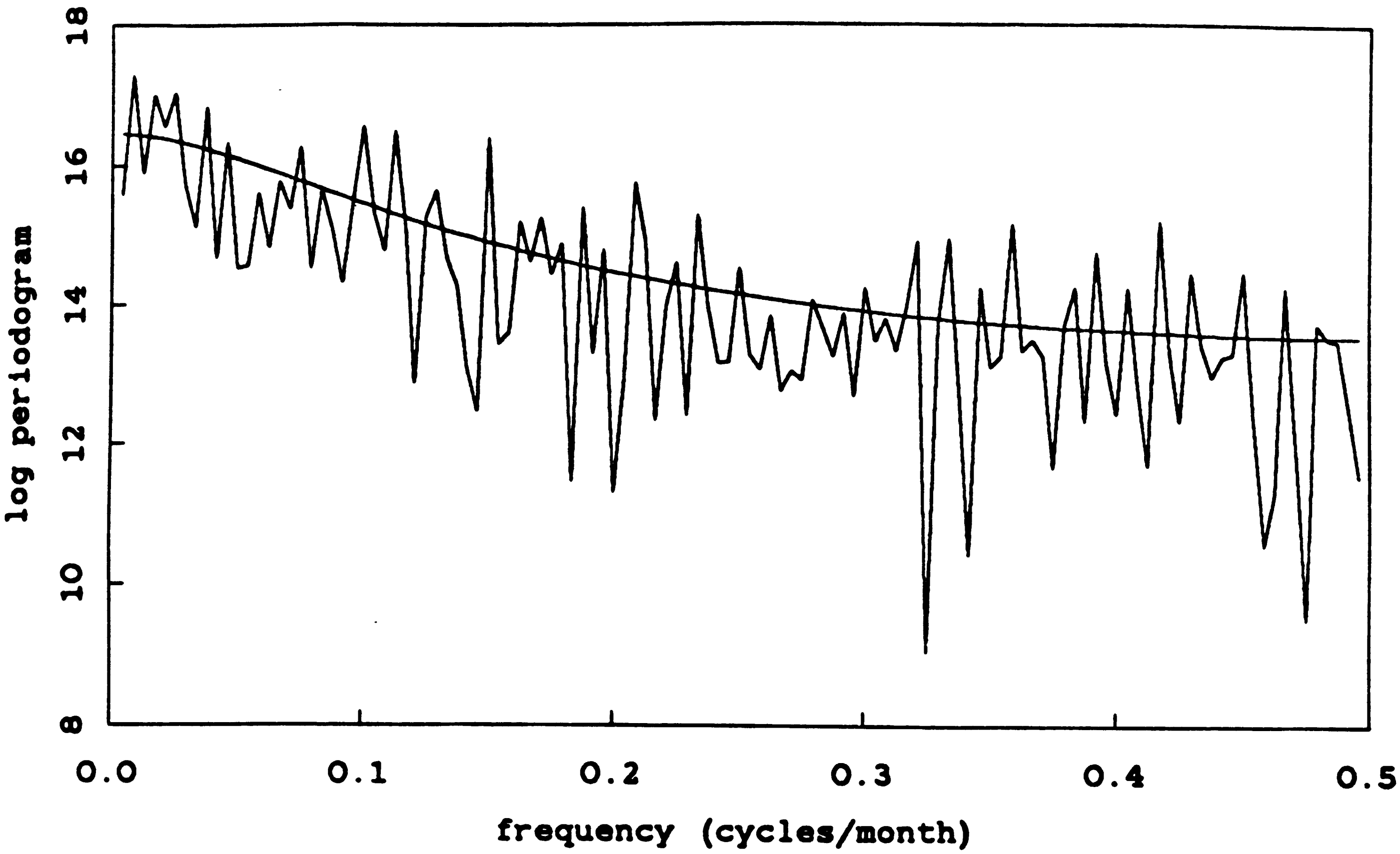
density
0.0
0.00002
0.00004
0.00006
0.00008
0.00010
0.00012
0.00014
-15000
-10000
-5000
0
5000
10000
15000
20000
thousands of acre-feet



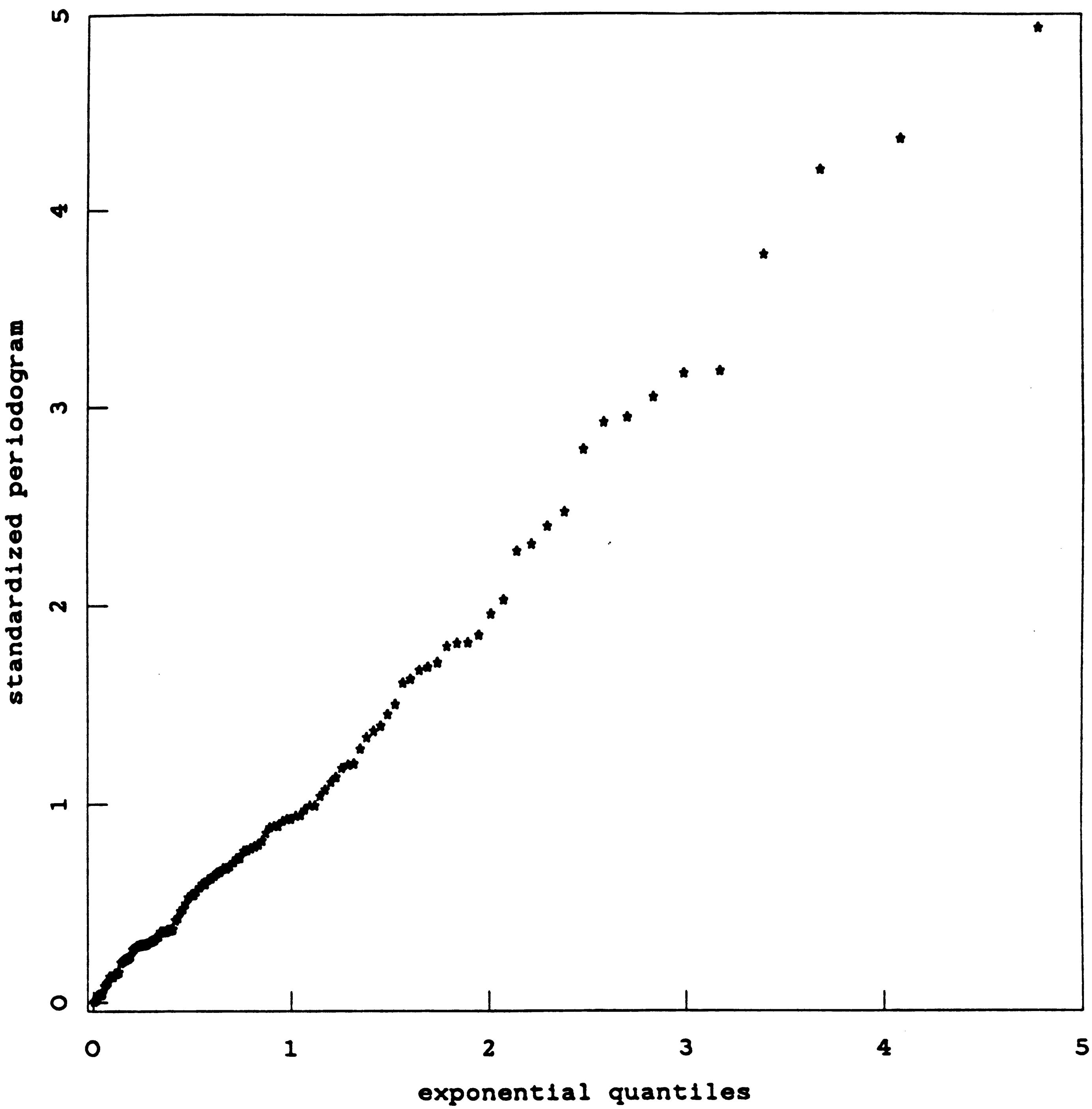
20000
15000
10000
5000
0
-5000
-10000
-15000
thousands of acre-feet
-3
-2
-1
0
1
2
3
normal quantiles

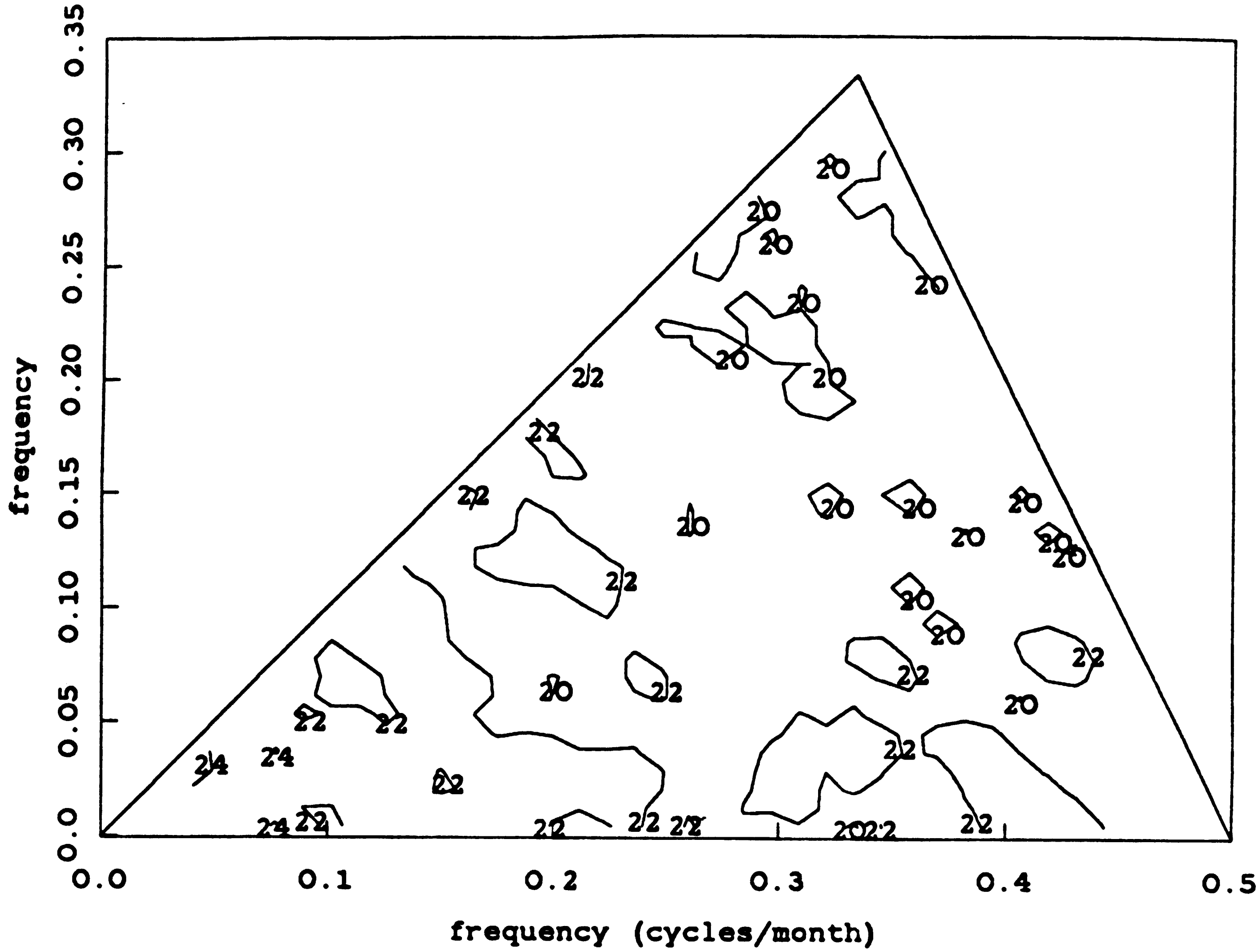


next year
20000
15000
10000
5000
0
-5000
-10000
-15000
-15000
-10000
-5000
0
5000
10000
15000
20000
this year

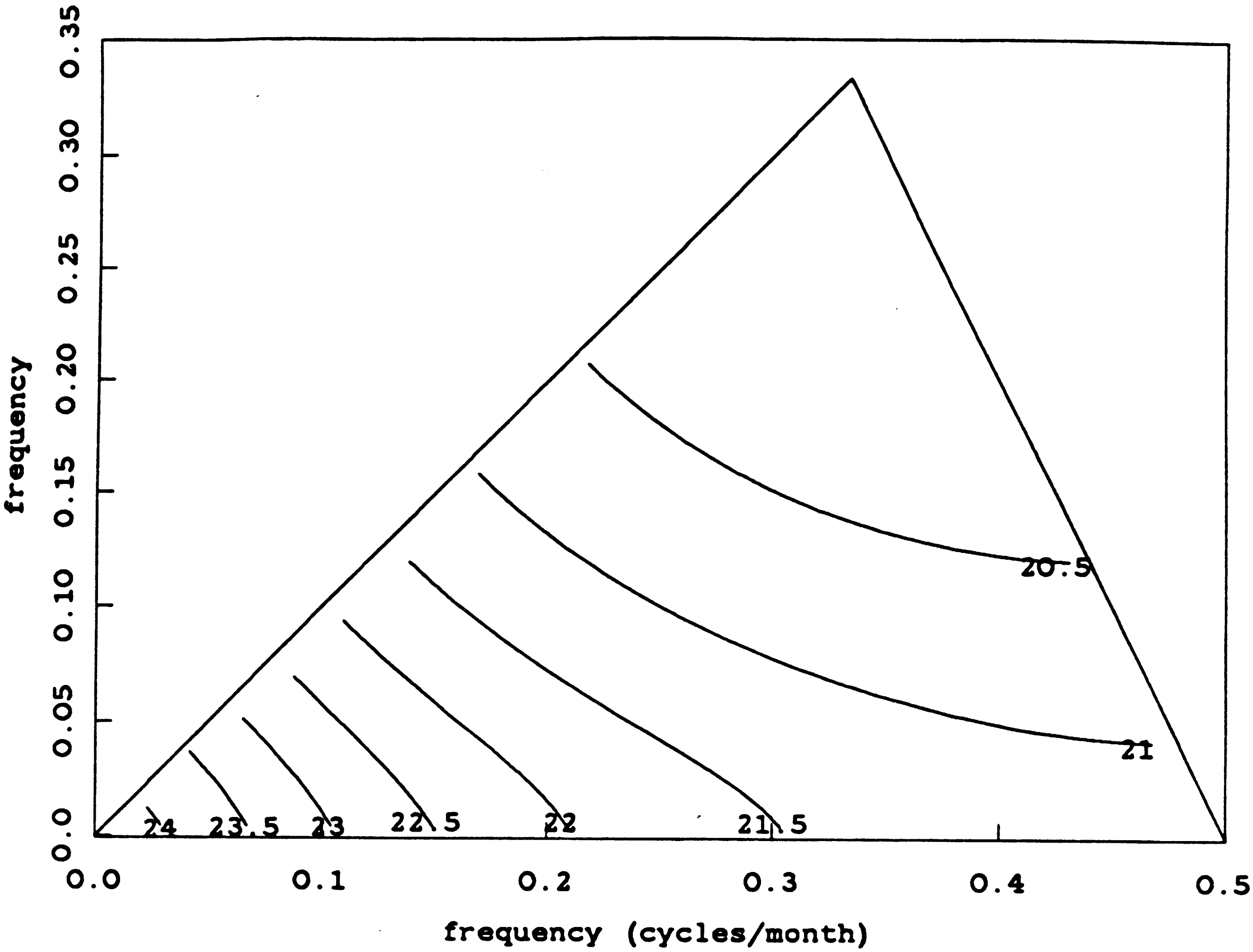


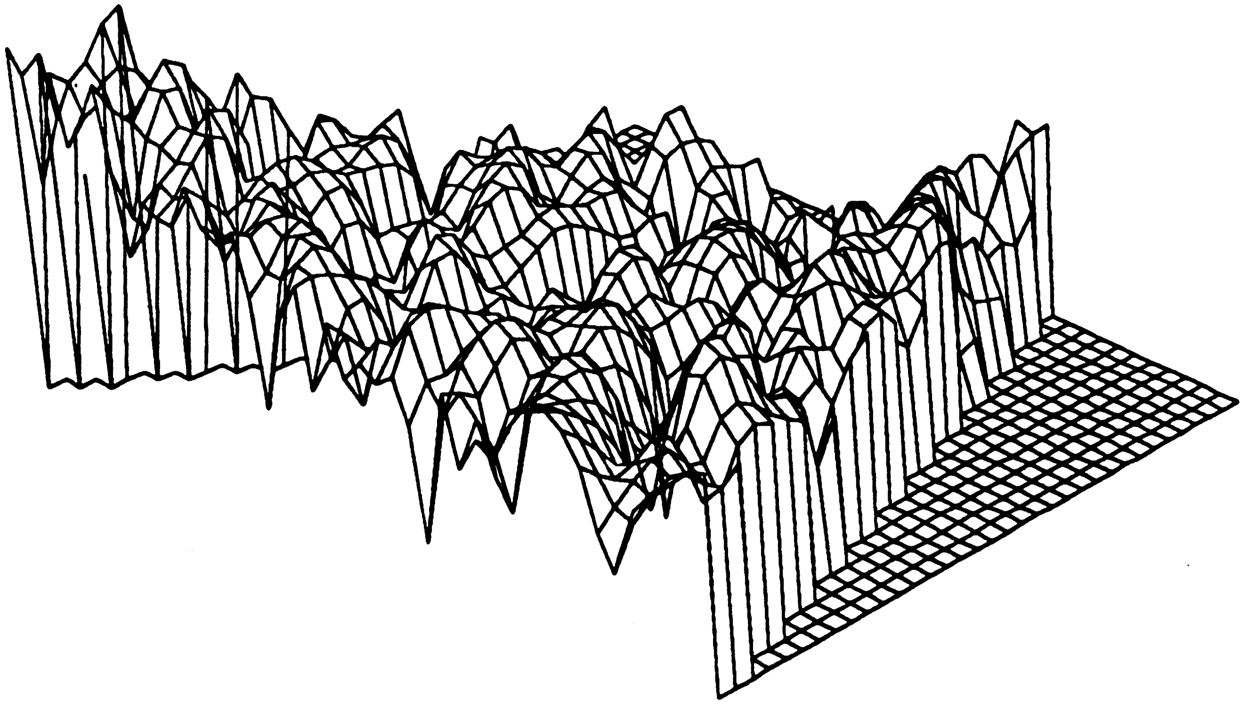
18
16
14
12
10
8
log periodogram
0.0
0.1
0.2
0.3
0.4
0.5
frequency (cycles/month)

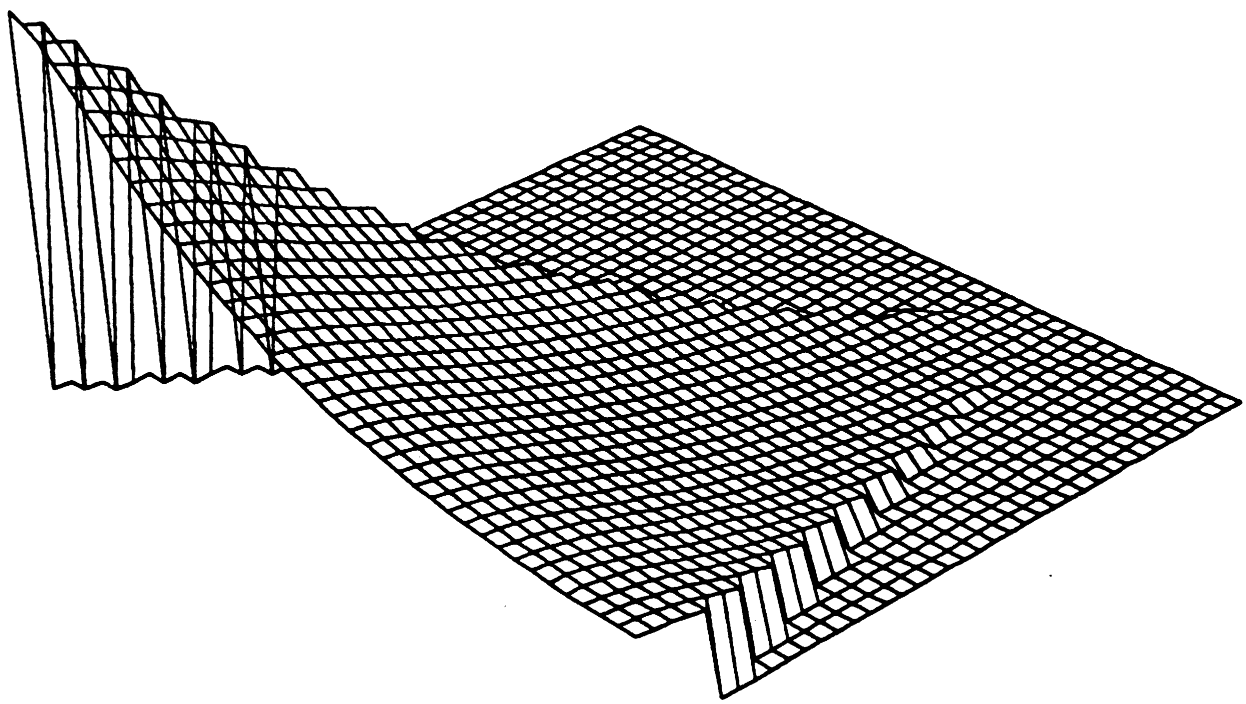

standardized periodogram
exponential quantiles
0
1
2
3
4
5

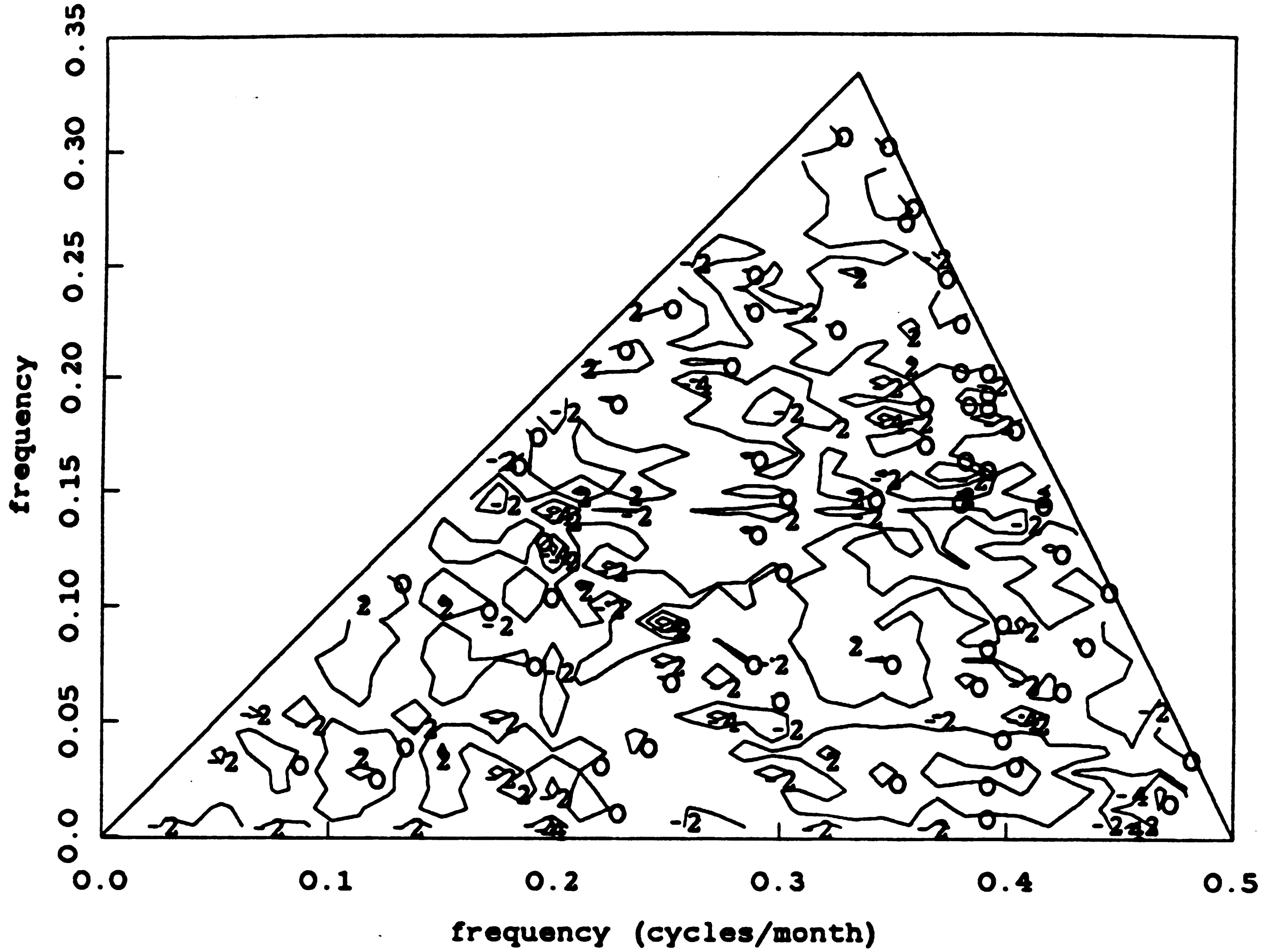


frequency
0.0
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.0
0.1
0.2
0.3
0.4
0.5
frequency (cycles/month)


0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.0
frequency
20.5
21
24
23.5
23
22.5
22
21.5
0.0
0.1
0.2
0.3
0.4
0.5
frequency (cycles/month)







frequency
0.0
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.0
0.1
0.2
0.3
0.4
0.5
frequency (cycles/month)