STATISTICAL ASPECTS OF STATIONARY PROCESSES WITH LONG-RANGE DEPENDENCE

JAN BERAN

Dept. of Statistics

University of North Carolina

Chapel Hill, NC 27599-3260

Abstract

The assumption of independence is often only an approximation to the real correlation structure. Models with short-range memory are well known and often used in practice. However even for supposedly i.i.d. high-quality data slowly decaying correlations may occur. If not taken into account, they have disastrous effects on tests and confidence intervals. Stationary processes with a pole of the spectrum at the origin can be used for modelling such data. In this paper a review is given of results on statistical inference for such long-memory processes.

STATISTICAL ASPECTS OF STATIONARY PROCESSES WITH LONG-RANGE DEPENDENCE

Jan Beran

Dept. of Stat., Univ. of North Carolina, Chapel Hill, NC 27599-3260

1. Introduction

The assumption of independence is in most cases only an approximation to the real correlation structure. Large correlations for small lags can be detected quite easily if the data set is large enough. Models with short-range memory (like ARMA-models, Markov processes) are well known and often used in practice. They have a smooth spectrum and exponentially decaying correlations. Already for these models interval estimates and prediction can differ considerably from the iid case though asymptotic rates of convergence remain the same. Yet there is strong evidence that even for supposedly independent identically distributed high-quality data the correlations may decay hyperbolically i.e. like $|k|^{-\alpha}$ with $\alpha \in (0,1)$, implying non-summability of the correlations and a pole of the spectrum at zero. Although single correlations may be small, the consequences for classical tests and confidence intervals are disastrous. Most estimates have a slower rate of convergence so that assuming independence or some kind of short-range dependence leads to underrating uncertainty by a factor which tends to infinity as the sample size tends to infinity.

Obviously many applied statisticians and natural scientists had been aware of this danger long before suitable stochastic models were known (see e.g. Newcomb (1886), K.Pearson (1902), Jeffreys (1939), Student (1927), Fairfield Smith (1938), Whittle (1956); see also Hampel et al.(1986) and Hampel(1987)). In particular the so-called Hurst effect (see Hurst 1951) - well known to hydrologists - can be explained by slowly decaying correlations. One of the data sets which led to the discovery of this effect is the record of Nile river minima. The periodogram in log-log-coordinates (figure 1) has a negative slope near the origin which is typical for long-range dependence. Figure 2 shows the periodogram of measurements on the 1 kg check standard weight done by the National Bureau of Standards, Washington. It is an example of high-precision measurements where in spite of the effort made to keep conditions as constant as possible long-range dependence occurs.

The simplest stationary processes with slowly decaying correlations or equivalently with a pole of the spectrum at zero are increments of self-similar processes, in the Gaussian case so-called fractional Gaussian noise. These models were first introduced to statistics by Mandelbrot and co-workers (Mandelbrot and Wallis 1968, 1969; Mandelbrot and van

Ness 1968). A stochastic process Y_t is called self-similar with self-similarity parameter H if $L(Y_{ct}) = L(c^H Y_t)$. For stationary increments $X_i = Y_i - Y_{i-1}$ we then have

$$R_k = cov(X_i, X_{i+k}) = \sigma^2(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H})/2, \tag{1}$$

where $\sigma^2 = var(X_i)$. The spectral density is of the form

$$f(x) = 2b(H,\sigma^2)(1-\cos\lambda)\sum_{j=\infty}^{\infty}|2\pi j + \lambda|^{2H-2}, \qquad x \in [-\pi,\pi], \tag{2}$$

with $b(H,\sigma^2) = \sigma^2 \sin(\pi H)\Gamma(2H+1)$. For $H \in (1/2,1)$ f has a pole at zero and $\sum_{j=\infty}^{\infty} R_k = \infty$. Note that from (1) one obtains

$$var(\bar{X}_n) = \sigma^2 n^{2H-2}. \tag{3}$$

More generally a stationary process is said to exhibit long-range dependence if

$$f(x) \sim_{|x|\to 0} L_1(x)|x|^{1-2H}, \qquad H \in (1/2,1), \tag{4}$$

where $L_1(.)$ is slowly varying for $|x| \to 0$ or equivalently (under weak regularity conditions on $L_1(.)$) if

$$R_k \sim_{|k|\to\infty} L_2(k)|k|^{2H-2}, \qquad H \in (1/2,1), \tag{5}$$

with $L_2(.)$ slowly varying for $|k| \to \infty$.

Theory of statistical inference for such processes is still in its infancy. Nevertheless for some basic situations suitable methods are sufficiently known nowadays to be used in practice (see Kuensch (1986)). We give a short (and necessarily incomplete) review of results in this field. The basic model will be fractional Gaussian noise though many of the results hold for any stationary process with (4) and (5).

2. Estimation

At first we consider estimation of $\mu = E(X_i)$ and $\sigma^2 = var(X_i)$. The arithmetic mean does not lose more than 2% of efficiency compared with the best linear unbiased estimator (BLUE). The asymptotic efficiency can be given explicitely:

Theorem 1 (Adenstedt 1974):

If (4) holds with $L_1(x) = const$ then

$$eff(\bar{X}_n, \hat{\mu}_{BLUE}) = \pi(2H-1)H/(B(3/2-H, 3/2-H)\sin\pi(H-1/2)). \tag{6}$$

Robust estimation of μ can be done without losing efficiency under the Gaussian model:

Theorem 2 (Beran 1985):

Define T_n by $\sum_{i=1}^{n} \psi(X_i - T_n) = 0$. If X_i is Gaussian with spectrum (4) then, under weak regularity conditions on ψ, T_n is asymptotically equivalent to $\bar{X}_n$.

In contrast to $\bar{X}_n$ the classical scale estimator $s^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ is very bad. It has a large bias and loses much efficiency. For $H \geq 3/4$ the efficiency is even equal to zero.

Several heuristic proposals have been made how to estimate H and σ^2, the best known being the so-called R/S-statistic first introduced by Hurst in hydrology (Hurst 1951) and further investigated by Mandelbrot and his co-workers (Mandelbrot and Wallis 1969, Mandelbrot and Taqqu 1979). Under the Gaussian model it loses much efficiency compared with maximum likelihood type methods. Asymptotic normality of the maximum likelihood estimator was proved by Yajima(1985). For the fractional ARMA-model this problem was also treated by Li and McLeod(1986). Efficient approximation to maximum likelihood has been considered by several authors. In Whittle's approximation (see Fox and Taqqu 1986, Beran 1986) the inverse covariance matrix of $(X_1, ..., X_n)$ is replaced by the two-sided infinite dimensional matrix $(a_{k,l})_{k,l=-\infty,\infty}$ with $a_{k,l} = a_{k-l} = (2\pi)^{-1} \int_{-\pi}^{\pi} f^{-1}(x) \cos kx dx$. One then has to minimize (with respect to H and σ^2)

$$\sum_{k,l=1}^{n} a_{k-l}(C_k - R_k) + (2\pi)^{-1} \int_{-\pi}^{\pi} \log f(x) dx, \tag{7}$$

where $C_k = n^{-1} \sum_{i=1}^{n-|k|} (X_i - \bar{X}_n)(X_{i+k} - \bar{X}_n)$. A discretized version of (7), the so-called HUB00-estimator, was proposed by Graf(1983). Graf also wrote a FORTRAN program for his estimator. A third possibility is to replace the inverse covariance matrix by the one-sided infinite dimensional matrix $(a_{kl})_{k,l \geq 1}$. This results in minimizing

$$n^{-1} \sigma_\varepsilon^2 \sum_{l=2}^{n} (X_l - \bar{X}_n - \sum_{k=1}^{l-1} b_k (X_{l-k} - \bar{X}_n))^2 + \log \sigma_\varepsilon^2, \tag{8}$$

where $X_i = \sum_{j=1}^{\infty} b_j X_{i-j} + \varepsilon_i$ is the infinite AR-representation of $(X_i)_{i \in Z}$ and $\sigma_\varepsilon^2 = var(\varepsilon_i)$. All three estimators have asymptotically the same distribution as the exact maximum likelihood estimator (for parts of the proof see Fox and Taqqu 1986, Beran 1986, Dahlhaus 1986):

Theorem 3:

If $\hat{\underline{\theta}}$ is defined by one of the approximate maximum likelihood methods described above, then $n^{\frac{1}{2}}(\hat{\underline{\theta}} - \underline{\theta})$ is asymptotically normally distributed with zero mean and covariance matrix

$V = 2D^{-1}$ where

$$D_{ij} = (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f(x) \frac{\partial}{\partial \theta_j} \log f(x) dx. \tag{9}$$

More generally theorem 3 holds for $\underline{\hat{\theta}} = (\hat{\sigma}^2, \hat{H}, \hat{\theta}_3, \cdots, \hat{\theta}_M)$ where the additional parameters θ_i $i \geq 3$ characterize the short range dependence structure. Note that $\frac{1}{2}D$ is the Fisher information matrix (for the proof see Dahlhaus 1986). Hence $\underline{\hat{\theta}}$ is asymptotically efficient.

Considering a class of nested models with spectral density (4) and $f = f_{\underline{\theta}}$ one can choose the dimension of $\underline{\theta}$ applying a version of Akaike's criterion (Beran 1986): Choose the model for which

$$AIC(M) = \sum_{k=-(n-1)}^{n-1} a_k(\underline{\hat{\theta}}) C_k^* + \sum_{|k| \geq n} a_k(\underline{\hat{\theta}}) R_k(\underline{\hat{\theta}}) + (2\pi)^{-1} \int_{-\pi}^{\pi} \log f_{\underline{\hat{\theta}}}(x) dx + \frac{2M}{n} \tag{10}$$

is minimal. Here $C_k^* = n(n-|k|)^{-1} C_k$ is the unbiased estimator of R_k and $\underline{\hat{\theta}}$ is one of the approximate maximum likelihood estimators given above. For Markov processes this criterion was derived by Künsch(1981). There the second term vanishes. By an analogous technique as in Shibata(1976) (10) can be shown to be inconsistent. It overestimates asymptotically the number of parameters with positive probability.

In practice parametric models are only approximations. Therefore it is important to know how $\underline{\hat{\theta}}$ behaves under deviations from the model and to find robust alternatives. Here we have at least three kinds of possible deviations: nongaussianity, other form of the spectrum at high frequencies or an other function $L_1(.)$ in (4), nonstationarity.

All three approximate maximum likelihood estimators are defined via quadratic forms. This makes them rather sensitive to deviations from Gaussianity. Fox and Taqqu(1986) even found an example where the rate of convergence is slower than $n^{-\frac{1}{2}}$. In practice we may hope that such extreme cases do not occur. Yet it would be useful to find methods which protect us against them. Data cleaning and transformations will prevent the worst. More robust estimators can be obtained for instance by huberizing the equations obtained from (8) by differentiation. A central limit theorem analogous to theorem 3 then holds.

The second kind of deviation has to be expected for most data sets, in particular if we use rather simple models - like fractional Gaussian noise - which only reflect the long-range aspects of the data. At least for some few lags significant correlations often occur. To avoid bias in $\hat{H}$ one has either to choose the right model or to bound the influence of periodogram ordinates at high frequencies. Consistency can be achieved even if the true spectrum differs from the model spectrum outside a neighbourhood of zero, however at the cost of a slower rate of convergence (Geweke and Porter-Hudak 1983). A compromise was proposed by Graf (Graf 1983; Graf, Hampel and Tacier 1984). His HUBINC-estimator bounds the influence of high frequency ordinates in a way which guarantees approximate consistency and rate of convergence $n^{-\frac{1}{2}}$. Under fractional Gaussian noise the bias is zero and efficiency is above 74% for all $H \in (0,1)$.

A negative slope of the spectrum sometimes indicates a trend in the data rather than long-range dependence. The question arises if and how these two models can be distinguished. Bhattacharya et al.(1983) showed that for $Y_i = X_i + m_i$ with X_i weakly dependent and m_i a slowly decreasing trend the R/S-estimator of H is asymptotically between 1/2 and 1. This is a strong argument against the R/S-statistic, at least if the possibility of trend can not be excluded a priori. A possibility how to discriminate trend from long-range dependence was proposed by Künsch (1986). He proved that for Y_i the periodogram $I_n(2\pi j/n)$ has asymptotically a non-central χ^2_2-distribution with non-centrality parameter going to zero uniformly for $|j| \geq \varepsilon n^{\frac{1}{2}}$ whereas for a stationary process with spectrum (4) and $L_1(x) = const$ it has (up to a constant) asymptotically x^{1-2H} times a central χ^2_2-distribution. A consequence of this result is that in contrast to R/S the approximate maximum likelihood estimators discussed above are consistent for Y_i. In practice it might be a good idea to leave out the first few periodogram ordinates though some experience is needed to choose a reasonable ε. Apart from the formal results the decision between trend and long-range dependence will also depend on the specific context and the field of application.

Finally one should not forget that a negative slope of the periodogram at zero also occurs for certain kinds of ARMA-processes. However if the slope is due to long-range dependence then in most cases an ARMA-model which fits well to the data has many parameters so that a model selection criterion like (10) or consistent versions of (10) will exclude the ARMA-model.

3. Tests and Confidence Intervals

Several tests for Hurst effect, in particular their power for distinguishing between fractional Gaussian noise and an AR(1)-process, were considered recently by Davies and Harte(1987). A possibility how to tell trend from long-range dependence was already mentioned above.

Given a stationary process with long-range dependence it is obvious from theorem 3 how to construct confidence intervals for $\hat{\underline{\theta}}$. By fitting Pearson curves to simulated moments of $\hat{\underline{\theta}}$ Graf(1983) obtained finite sample corrections. More general methods like small sample asymptotics have not yet been investigated in this context.

In the examples $\hat{H}$ is significantly larger than 1/2: For the Nile river data the 95% confidence interval for $\hat{H}$ is [0.788,0.913], for the NBS precision measurements of the 1 kg check standard weight it is [0.674,0.806]. In the first case the high value of $\hat{H}$ is no surprise. Long-range dependence is a well known phenomenon in hydrology. More surprising is the result for the second data set. Even on the 99% level $\hat{H}$ is significantly larger than 1/2, although every precaution had been made to guarantee independence and constant conditions.

Confidence intervals for the mean have to take into account the variability of $\hat{\underline{\theta}}$. If (4) holds then $var(\bar{X}_n) \sim_{n\to\infty} c(n)n^{2H-2}$ where $c(n) = 2\pi^{-1}L_1(n^{-1})\Gamma(-2H)\sin\pi(\frac{1}{2}-H)$. For a parametric model L_1 depends on $\underline{\theta}$ so that $c(n) = c(n;\underline{\theta})$. Hence a test statistic applicable to any Gaussian process with (4) can be defined by

$$T = (\bar{X}_n - \mu)c(n;\hat{\underline{\theta}})^{-\frac{1}{2}}n^{1-\hat{H}}, \tag{11}$$

where $\hat{\underline{\theta}}$ is some reasonable estimate. With the HUBINC-estimator of $\underline{\theta} = (\log b(H,\sigma^2), H)$, b defined by (2), fractional Gaussian noise as the basic model and setting $T = 0$ if $\hat{H} \notin (0,1)$ a good approximation to the distribution of T can be obtained (Beran 1986) which is also applicable to rather short series ($n \geq 64$):

$$P(T \leq u) \approx (1 - \Phi((1-H)n^{\frac{1}{2}}/\sigma)) + (1 - \Phi(Hn^{\frac{1}{2}}/\sigma)) + \int_{-H}^{1-H}\int_{-\infty}^{\infty} \Phi(ug(n,\underline{\theta};\underline{z}))\phi_n(\underline{z})dz_1dz_2, \tag{12}$$

where $\phi_n(\underline{z})$ is the two-dimensional normal density with zero mean and covariance matrix $n^{-1}V$, V as in theorem 3 and $g(n,\underline{\theta};\underline{z}) = (c(n;\underline{\theta}+\underline{z})/c(n;\underline{\theta}))^{\frac{1}{2}}n^{z_1}$. In particular for small values of n finite sample corrections of V based on simulations turned out to improve the approximation. Note that the first two terms refer to the case where $\hat{H} \notin (0,1)$. They vanish asymptotically and can be neglected except if n is small and H is near 1.

Applied to the NBS measurements the confidence intervals are about twice as large as under the assumption of independence. This reflects the additional uncertainty under long-range dependence with unknown H.

The method can be generalized to prediction intervals for the arithmetic mean of future observations.

4. Concluding Remarks

Long-range dependence can - if not taken into account - completely invalidate statistical inference. Methods for testing long-range dependence, fitting suitable models and testing location are sufficiently known nowadays to be used in practice.

More research is needed both to deal with more complex situations and to refine the methods in use. Central limit theorems for processes with long-range dependence are rather different from the classical type of theorems so that many standard results in statistics do not hold. The classical methods should be investigated under this aspect and new methods which also perform well under long-range dependence should be developed.

References

Adenstedt R.K. (1974). On large sample estimation for the mean of a stationary random sequence. *Ann. Statist.* **2**, 1095-1107.

Beran J. (1986). Estimation, testing and prediction for self-similar and related processes. PhD Thesis, ETH Zürich.

Beran J. and Künsch H.R. (1985). Location estimators for processes with long-range dependence. Research Report No.40, Seminar für Statistik, ETH Zürich.

Bhattacharya R.N., Gupta V.K. and Waymire E. (1983). The Hurst effect under trends. *J. Appl. Prob.* **20**, 649-662.

Dahlhaus R. (1986). Efficient parameter estimation for self-similar processes. preprint, Universität Essen.

Davies R.B. and Harte D.S. (1987). Tests for Hurst effect. *Biometrika* **74**, 95-102.

Fairfield Smith H. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *J. Agric. Sci.* **28**, 1-23.

Fox R. and Taqqu M.S. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian Time series. *Ann. Statist.* **14, Vol.2**, 517-532.

Geweke J. and Porter-Hudak S. (1983). The estimation and application of long memory time series models. *J. Time Series Anal.* **4**, 221-238.

Graf H.P. (1983). Long-range correlations and estimation of the self-similarity parameter. PhD thesis, ETH Zürich.

Graf H.P., Hampel F.R. and Tacier J. (1984). The problem of unsuspected serial correlations. *In Robust and Nonlinear Time Series Analysis. J. Franke, W.Härdle, R.D. Martin (eds), Lecture Notes in Statistics* **26**, Springer, New York, 127-45.

Hampel F.R. (1987). Data analysis and self-similar processes. *Invited paper, 46th Session of the ISI, Tokyo. to appear* , .

Hampel F.R., Ronchetti E.M., Rousseeuw P.J. and Stahel W.A. (1986). *Robust statistics. The approach based on influence functions..* New York: Wiley.

Hurst H.E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Engineers* **116**, 770-799.

Jeffreys H. (1939). *Theory of probability.* Oxford: Clarendon Press.

Künsch H.R. (1986). Statistical aspects of self-similar processes. *Invited paper, Proc. First World Congress of the Bernoulli Society, Tashkent* **Vol.1**, 67-74.

Künsch H.R. (1981). Thermodynamics and statistical analysis of Gaussian random fields. *Z. Wahrsch. verw. Gebiete* **58**, 407-427.

Künsch H.R. (1986). Discrimination between monotonic trends and long-range dependence. *J. Appl. Prob.* **23**, 1025-1030.

Li W.K. and McLeod A.I. (1986). Fractional time series modelling. *Biometrika* **73**, 217-221.

Mandelbrot B.B. and van Ness J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review* **10, No.4**, 422-437.

Mandelbrot B.B. and Taqqu M.S. (1979). Robust R/S analysis of long run serial correlation. *Proceedings 42nd Session of the ISI, Manila* **Book 2**, 69-100.

Mandelbrot B.B. and Wallis J.R. (1968). Noah, Joseph and operational hydrology. *Water Resources Research* **4, No.5**, 909-918.

Mandelbrot B.B. and Wallis J.R. (1969). Computer experiments with fractional Gaussian noises. *Water Resources Research* **5, No.1**, 228-267.

Newcomb S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *Amer. J. Math.* **8**, 343-366.

Pearson K. (1902). On the mathematical theory of errors of judgement, with special reference to the personal equation. *Philos. Trans. Roy. Soc. Ser. A* **198**, 235-299.

Shibata R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63**, 117-126.

Student (1927). Errors of routine analysis. *Biometrika* **19**, 151-164.

Whittle P. (1951). *Hypothesis testing in time series analysis.* New York: Hafner.

Yajima Y. (1985). On estimation of long-memory time series models. *Austral. J. Statist.* **27**, 303-320.

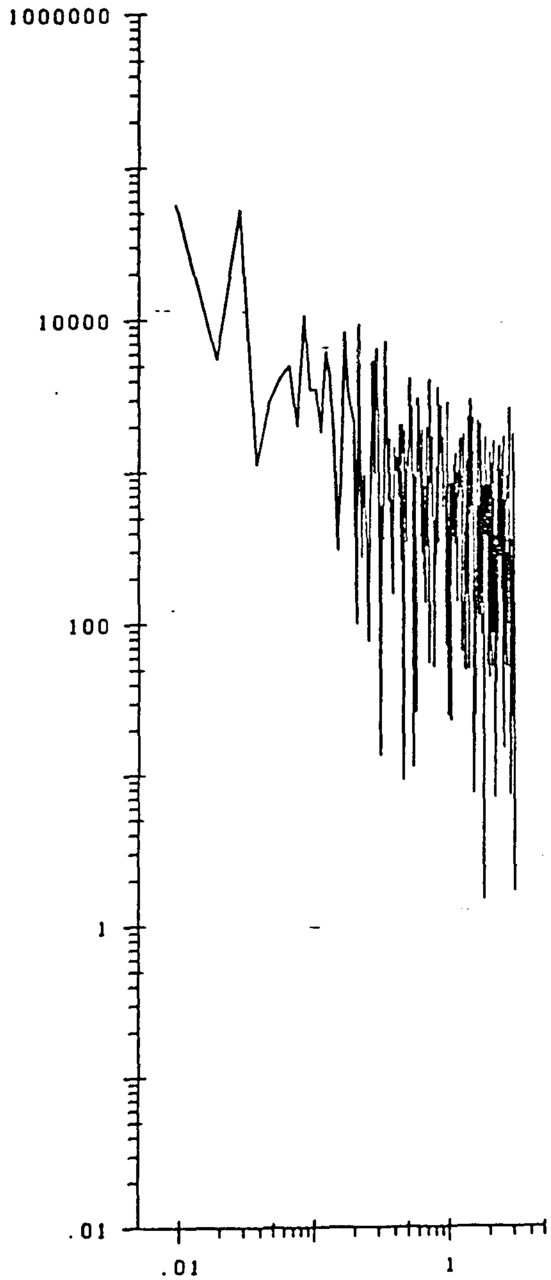

Figure 1

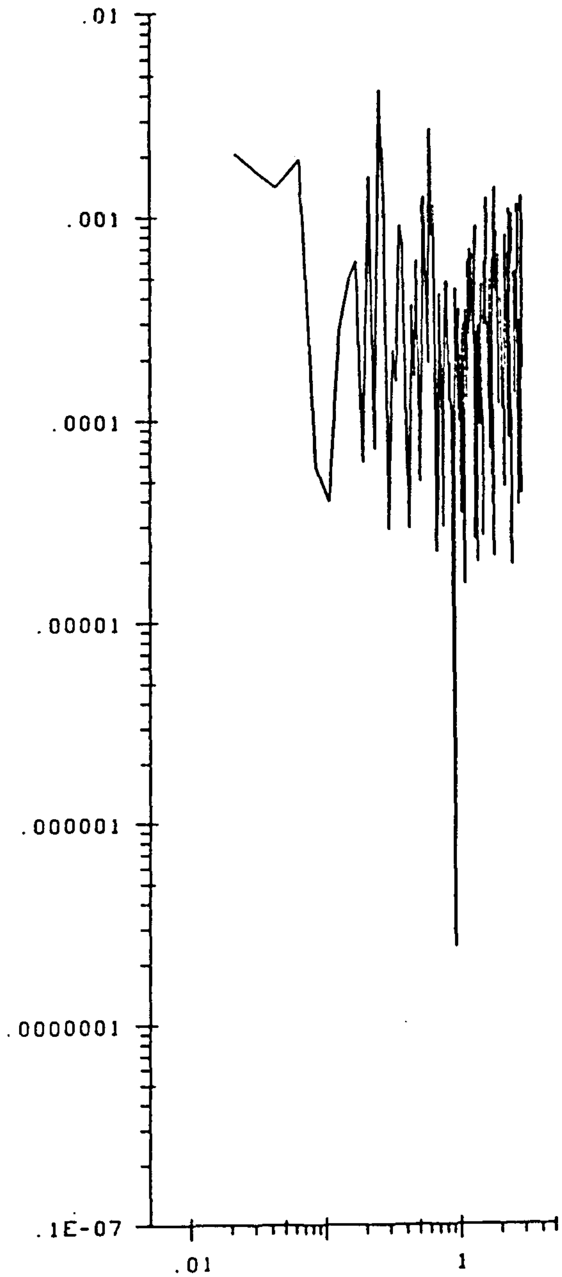

Figure 2