Statistical Science
1992, Vol. 7, No. 4, 404–427

Statistical Methods for Data with Long-Range Dependence

Jan Beran

Abstract. It is well known to applied statisticians and scientists that the assumption of independence is often not valid for real data. In particular, even when all precautions are taken to prevent dependence, slowly decaying serial correlations frequently occur. If not taken into account, they can have disastrous effects on statistical inference. This phenomenon has been observed empirically by many prominent scientists long before suitable mathematical models were known. Apart from some scattered early references, mathematical models with long-range dependence were first introduced to statistics by Mandelbrot and his coworkers (Mandelbrot and Wallis, 1968, 1969; Mandelbrot and van Ness, 1968). Since then, long-range dependence in statistics has gained increasing attention. Parsimonious models with long memory are stationary increments of self-similar processes with self-similarity parameter $H \in (1/2,1)$, fractional ARIMA processes and other stationary stochastic processes with non-summable correlations. In the last decade, many results on statistical inference for such processes have been established. In the present paper, a review of these results is given.

Key words and phrases: Long-range dependence, fractional Gaussian noise, fractional ARIMA, self-similar, point estimation, interval estimation, prediction.

1. INTRODUCTION

It is well known to experienced statisticians and scientists that the assumption of independence is in most cases only an approximation to the real dependence structure. For example, to mention just a few classical references, Box, Hunter and Hunter (1978) call this assumption "the declaration of independence," Mosteller and Tukey (1977) discuss the unreliability of the $\sigma/\sqrt{n}$ rule for the sample mean in a chapter called "Hunting out the real uncertainty: How $\sigma/\sqrt{n}$ can mislead." Student (1927) writes (also see Jeffreys 1939, p. 298) "After considerable experience, I have not encountered any determination which is not influenced by the date on which it is made; from which it follows that a number of determinations of the same thing made on the same day are likely to lie more closely together than if repetitions had been made on different days." Scheffé (1959) discusses how even small correlations can have strong effects on statistical inference. Jeffreys (1939) writes: "Internal correlation habitually produces such large departures from the usual rule that the standard error of the mean is $n^{-1/2}$ times that of one observation that the rule should never be definitely adopted until it has been checked. In a series of observations made by the same observer, and arranged in order of time, internal correlations is the normal thing, and at the present state of knowledge hardly needs a significance test any longer."

In practice, large correlations for small lags can already be detected quite easily for moderately large data sets. Models with short-range memory (like ARMA models, Markov processes) are well known and often used in practice. In particular, Box and Jenkins (1970) made ARMA models and related techniques popular among practioners. These models have a smooth spectrum and exponentially decaying correlations. Already for these models, interval estimates and prediction intervals can differ considerably from the iid case, though asymptotic rates of convergence remain the same. For example, the standard deviation of the sample mean $\overline{X}_n$ is proportional, but not necessarily equal, to $\sigma/\sqrt{n}$. Yet, there is strong empirical evidence that often even for supposedly independent identically distributed high-quality data the correlations may

Jan Beran is a Research Fellow and Statistical Consultant, Biostatistics Department, ISPM, University of Zurich, Sumatrastr. 30, 8006 Zurich, Switzerland.

decay hyperbolically, that is, like $|k|^{-\alpha}$ with $\alpha \in (0, 1)$, implying nonsummability of the correlations. Although single correlations may be small, the consequences of this kind of (long-range) dependence for classical tests and confidence intervals can be disastrous. Most estimates and test statistics have a slower rate of convergence so that assuming independence or some kind of short-range dependence leads to underrating uncertainty (measured by the size of confidence intervals) by a factor which tends to infinity as the sample size tends to infinity. For example, the standard deviation of $\overline{X}_n$ decays at a rate of $n^{-\alpha/2}$, $\alpha \in (0, 1)$, instead of $n^{-1/2}$. Thus, if $\alpha = 0.4$, one needs approximately 100,000 observations to achieve the same precision (standard deviation) of $\overline{X}_n$ as from 100 independent observations drawn from a population with the same variance.

Many applied statisticians and natural scientists had been aware of this danger, even long before suitable stochastic models were known. Best known is the occurrence of long-range dependence in geophysics and hydrology (for a review, see Lawrance and Kotegoda, 1977). In particular, the so-called Hurst effect (Hurst, 1951) can be explained by slowly decaying correlations. However, there are many other fields of application where this type of correlation occurs. As early as 1895 the astronomer Newcomb discussed the phenomenon of long-range dependence in astronomical data sets and called it "semi-systematic" errors. He also proposed a heuristic explanation by superposition of independent random errors and constant systematic errors. Karl Pearson (1902) observed slowly decaying correlations in simulated astronomical observations. An example of spatial long memory in agriculture is discussed by Smith (1938); (also see Whittle, 1956). He analyzed 40 uniformity trials. The variance of the mean yield as a function of n = number of plots turned out to be proportional to $n^{-\alpha}$ (with $0 < \alpha < 1$) instead of n^{-1}. Further examples are discussed, for instance, by Student (1927) for chemical data, Jeffreys (1939) for astronomical data, Smith (1938) and Whittle (1956, 1962) for agricultural data, Cox and Townsend (1948) for textile engineering data, Granger (1966, 1980), Mandelbrot (1969, 1973), Carlin, Dempster and Jonas (1985), Carlin and Dempster (1989) and Porter-Hudak (1990) for economical data, Mandelbrot and Wallis (1969) for data from biology, geophysics, meteorology and hydrology, Damerau and Mandelbrot (1973) for linguistic data, Burrough (1981) and Graf (1983) for environmental data, Graf, Hampel and Tacier (1984) for high-quality physical measurements, Haslett and Raftery (1989) for meteorological data and Beran, Sherman, Taqqu and Willinger (1992) for telecommunication data. Though most known examples are time series, long-range dependence is not restricted to this type of data. Several of the examples above include spatial data with long memory (also see Matheron, 1973; Haslett and Raftery, 1989; Solo, 1989; Gay and Heyde, 1990). Dependence along more general structures, such as abstract graphs, might be worth looking into. For more references, see Mandelbrot and Wallis (1969), Mandelbrot (1983), Cox (1984), Hampel et al. (1986), Künsch (1987), Hampel (1987), Beran (1988) and Haslett and Raftery (1989).

The simplest models with long memory are stationary processes with correlations decaying hyperbolically. A typical data set where such a model seems to fit is the record of the Nile river minima (Figure 1). This data played a key role in the discovery of long-range dependence in hydrological data by the famous hydrologist Hurst (1951). The plot of the data reveals several interesting features: At first sight the data might seem nonstationary, in particular parts of the data seem to have local trends or periodicities and the expected value seems to be changing slowly. A closer look at the whole series however shows that both, trends and periodicities, change with time in an irregular way and the overall mean seems to be constant. Such behaviour is typical for stationary processes with long memory (for the definition see below). For the Nile river minima, the correlations $\rho_k = \text{corr}(X_i, X_{i+k})$ decay approximately like $|k|^{-\alpha}$ (as $|k| \to \infty$) with α equal to 0.3 (see Section 4).

Due to the vast number of examples from hydrology and geophysics, long-range dependence is recognized by most hydrologists and geophysicists to be the rule rather than the exception. The phenomenon however occurs in many other areas of application. Even in situations where every precaution was taken to prevent dependence between the observations, slowly decaying correlations often occur. A typical example of such high-quality data measured under ideal circumstances, are the measurements of the 1-kg check standard weight provided to us by the U.S. National Bureau of Standards Washington (Figure 2). 289 high-precision measurements on the 1-kg check standard weight were made between 1963 and 1975, under conditions that were kept as constant as possible. In spite of the ideal

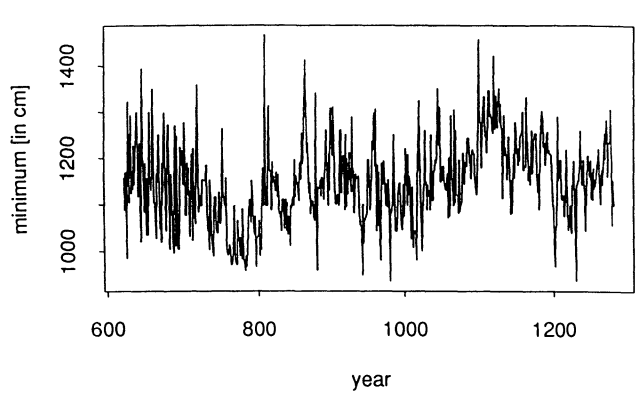

FIG. 1. *Nile River minima.*

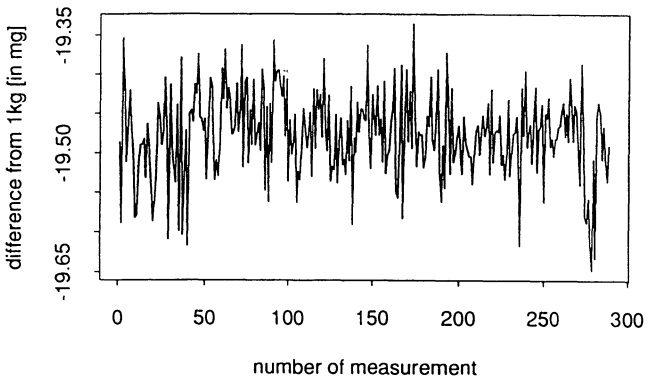

FIG. 2. *NBS precision measurements on the 1-kg check standard weight.*

circumstances, the correlations seem to decay with a rate approximately proportional to $|k|^{-\alpha}$ with α equal to 0.8 (see Section 4).

Several authors also discussed possible physical reasons for the occurrence of long-range dependence and derived physical models justified in their specific contexts (e.g., Cox and Townsend, 1948; Whittle, 1962; Mandelbrot, 1971; Klemes, 1974; Cassandro and Jona-Lasinio, 1978; Granger, 1980; Cox, 1984). In many situations, it seems rather difficult to construct a useful and sufficiently simple physical model. Yet, often the processes emerging from such physical models turn out to be long-memory processes of the above type. The two best known classes of stationary processes with slowly decaying correlations are increments of self-similar processes (in the Gaussian case so-called fractional Gaussian noise) and fractional ARIMA processes.

Self-similar processes and the corresponding increment processes were first introduced to statistics by Mandelbrot and co-workers (Mandelbrot and van Ness, 1968; Mandelbrot and Wallis, 1968, 1969). Brownian motion is self-similar and was known for a long time. Kolmogorov (1940) introduces fractional Brownian motion. Lamperti (1962) points to the fact that normalized sums of random variables converge to self-similar processes. For extensive surveys on self-similar processes see Verwaat (1987) and Taqqu (1988). A stochastic process $(Y_t)_{t\in R+}$ is called self-similar with self-similarity parameter H, if for any $c>0$ the stochastic process $(Y_{ct})_{t\in R+}$ is equal in distribution to the process $(c^H Y_t)_{t\in R+}$. If Y_t has stationary increments $X_i = Y_i - Y_{i-1}$ $(i \in N)$, then the covariances $R_k = \operatorname{cov}(X_i, X_{i+k}) = \int_{-\pi}^{\pi} \exp(ikx) f(x)\,dx$ are of the form

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

where $\sigma^2 = \operatorname{var}(X_i)$. The spectral density is given by

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

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

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

For $H = 1/2$ the X_i's are uncorrelated. The case $0 < H < 1/2$ where the spectrum is zero at zero and $\sum_{j=-\infty}^{\infty} R_k = 0$ is less interesting for statistical applications. It can however occur after overdifferencing.

Fractional ARIMA models were introduced by Granger and Joyeux (1980) and Hosking (1981). They are a natural generalization of standard ARIMA(p, d, q) models defined in Box and Jenkins (1970). By allowing $d = H - 1/2$ to assume any value between $-1/2$ and $1/2$, a fractional ARIMA process is defined by

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t. \tag{4}$$

Here, the ε_t's are iid zero mean normal random variables, B denotes the backshift operator, $\Phi(B)$ defines the AR-part and $\Theta(B)$ defines the MA-part of the process and $(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k}(-B)^k$ is the fractional difference operator. The spectral density is then of the form

$$f(x) = \frac{|\Theta(e^{ix})|^2}{|\Phi(e^{ix})|}\,|1-e^{ix}|^{-2d}, \quad x \in [-\pi,\pi]. \tag{5}$$

For $d = 0$, we obtain the usual ARIMA models. Long memory occurs for $d > 0$.

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

$$f(x) \sim_{|x|\to 0} L_1(x)|x|^{1-2H}, \quad H \in (1/2, 1), \tag{6}$$

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

$$R_k \sim_{|k|\to\infty} L_2(k)|k|^{2H-2}, \quad H \in (1/2, 1), \tag{7}$$

with $L_2(\cdot)$ slowly varying for $|k| \to \infty$. For simplicity we will assume that $L_1(0) = \lim_{x\to 0} L_1(x)$ exists and $0 < L_1(0) < \infty$.

A generalization of fractional ARIMA models was recently proposed by Gray, Zhang and Woodward (1989); (also see Hosking, 1981). They propose to model persistent cyclic behaviour by poles of the spectrum at nonzero frequencies. Equations (4) and (5) are generalized in the following way:

$$\Phi(B)(1-2uB+B^2)^{\lambda} X_t = \Theta(B)\varepsilon_t, \tag{4b}$$

(where $\lambda = d/2$, $|u| \le 1$) and

$$f(x) = \frac{|\Theta(e^{ix})|^2}{|\Phi(e^{ix})|}\,|1-2ue^{ix}+e^{2ix}|^{-2\lambda}, \quad x \in [-\pi,\pi]. \tag{5b}$$

Note that for $u = 1$ we obtain the original definitions (4) and (5).

The development of a theory of statistical inference for long-memory processes has become a very active field of research in the last decade. Though many prob-

lems are still unsolved, for some basic situations suitable methods are sufficiently known nowadays to be used in practice. This paper gives a review of recent results. Most methods have been developed for time series data. In this review, we therefore focus on these methods. There would certainly be a need for more methodological developments for the case of spatial data with long memory, or even data with a more general index variable. Typical examples from agronomy are discussed, for instance, in Smith (1938) and Whittle (1956, 1962). Other areas of application would be, for example, environmental sciences such as hydrology and meteorology. Some proposals of spatial models with long-range dependence have been made by Whittle (1962) and Renshaw (see his contribution to the discussion of Haslett and Raftery, 1989).

Not much seems to be known about statistical inference for spatial processes with long memory. A method of estimation for spatial models with long memory has been proposed recently by Haslett and Raftery (1989); (also see Künsch's contribution to the discussion of this paper). However, in their spatial model (and the corresponding data), long memory occurs only in time, not in the spatial index. More research would be needed on models, statistical inference and data examples where long memory occurs in the spatial index. Also, multivariate models with long-range dependence might prove useful in practical applications. Some research in this direction has been proposed by Li (contribution to the discussion on Haslett and Raftery, 1989). A similar problem has been considered by Hui and Li (1988). They defined fractional differenced periodic processes where $d = H - 1/2$ varies with the season. To my knowledge, no general theory on multivariate modelling with long memory is known in the literature. Possible alternative models for long-range dependence and "intermediate" models in the sense of "medium dependence" have been suggested by Tong, Künsch and Tjostheim, and Jones (see their contributions to the discussion of Haslett and Raftery, 1989). Developing such alternative models and the corresponding statistical methods is certainly an interesting area for future research. For a list of references on the probabilistic theory of long-memory processes, we refer the reader to Taqqu (1985, 1988) and Vervaat (1987).

2. POINT ESTIMATION

2.1 Location and Scale Estimation

At first, we consider estimation of $\mu = E(X_i)$ and $\sigma^2 = \text{var}(X_i)$. Rather surprisingly, it turns out that, in spite of slowly decaying correlations (7), the sample mean does not lose much efficiency compared to the best linear unbiased estimator (BLUE). An explicit formula for the asymptotic efficiency (given by the ratio of the two asymptotic variances) was derived by Adenstedt (1974):

$$eff(\overline{X}_n, \hat{\mu}_{BLUE}) = \frac{\pi(2H-1)H}{B(3/2-H, 3/2-H)\sin\pi(H-1/2)}. \tag{8}$$

Numerically, this is above 0.98 for all $H \in [1/2, 1]$ (Beran and Künsch, 1985; Samarov and Taqqu, 1988). To calculate the BLUE one would have to know or estimate all covariances. For most practical purposes, an efficiency loss of 2% does not matter so that the sample mean is not only much easier to calculate but also a sufficiently accurate estimate of μ.

A surprising consequence of long-range dependence is that robust estimation of μ can be done without losing efficiency under the Gaussian model (Beran, 1986, 1991). All M-estimators T_n, defined by $\Sigma_{i=1}^n \psi(X_i - T_n) = 0$, turn out to be asymptotically equivalent to the sample mean in the sense that $\text{var}(\overline{X}_n)/\text{var}(T_n) \to 1$ and $\text{var}(\overline{X}_n)^{-1/2}(\overline{X}_n - T_n) \to 0$ in probability (as $n \to \infty$). This is very much different from the situation of iid observations. There the asymptotic variance of T_n is equal to $E(\psi^2(X-\mu))/E^2(\psi'(X-\mu))$, which is larger than σ^2 for all nonlinear functions ψ.

Another interesting consequence of (7) was noted by Percival (1985). Define $\overline{X}_n(k) = m^{-1}\Sigma_{i=1}^m X_{ik}$, where k and m are integers such that $m \leqslant n/k$ and $m + 1 > n/k$, that is $\overline{X}_n(k)$ is the sample mean based only on the observations at time points $k, 2k, \ldots, mk$. For any fixed integer k, the relative asymptotic efficiency of $\overline{X}_n(k)$ compared to $\overline{X}_n(1)$, $eff(\overline{X}_n(k), \overline{X}_n(1)) = \lim_{n\to\infty}\text{var}(\overline{X}_n(1))/\text{var}(\overline{X}_n(k))$, turns out to be equal to one. This is very much in contrast to the iid case, where $eff(\overline{X}_n(k), \overline{X}_n(1))$ is equal to $1/k$. On the other hand, the deficiency of $\overline{X}_n(k)$ with respect to $\overline{X}_n(1)$, as defined by Hodges and Lehmann (1970), turns out to be infinite, even if (7) holds.

In contrast to $\overline{X}_n$, the classical scale estimator $s^2 = (n-1)^{-1}\ \Sigma_{i=1}^n\ (X_i - \overline{X}_n)^2$ is a bad estimator of σ^2. It has a large bias and loses much efficiency. For $H \geq 3/4$, the efficiency is even equal to zero, because its rate of convergence is slower than $n^{-1/2}$. An efficient $n^{1/2}$-consistent estimator of the scale will be discussed in Section 2.4.

2.2 Regression and Analysis of Variance

Parametric regression models of the form

$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{9}$$

where the errors ε_i are generated by a stationary process with slowly decaying correlations (7), have been investigated in Yajima (1988, 1991) and Künsch, Beran and Hampel (1992). The asymptotic distribution of the least squares estimator $\hat{\beta}_{LSE}$ and its efficiency com-

pared to the best linear unbiased estimator $\hat{\beta}_{BLUE}$ depends not only on the correlation structure but also on the design matrix $X = (x_{i,j}; i = 1, \ldots, n; j = 1, \ldots, p)$. As in the classical theory by Grenander and Rosenblatt (1957), the asymptotic theory can be characterized in terms of the so-called regression spectrum (as defined by Grenander and Rosenblatt). Since the spectral density of ε_t has a pole at zero, but is assumed to be continuous otherwise, the efficiency of $\hat{\beta}_{LSE}$ only depends on the behaviour of the regression spectrum near the origin. In particular, when estimating a polynomial trend, the rate of convergence of both estimators is slower than under independence by a factor proportional to $n^{H-1/2}$. The efficiency of $\hat{\beta}_{LSE}$ is less than one and is only a function of H and $L_1(0)$ (Yajima, 1988). As we saw in the previous section for the case $p = 1$, $x_{i,1} \equiv 1$, the actual efficiency loss is not necessarily large. On the other hand, when estimating seasonal components, the rate of convergence is the same as under independence and $\hat{\beta}_{LSE}$ is asymptotically efficient (Yajima, 1991).

A closer look at typical questions arising in analysis of variance reveals an interesting dichotomy between constants and contrasts. For standard complete random designs, least squares estimates of contrasts have asymptotically the same (conditional and unconditional) variance as under independence (Künsch, Beran and Hampel, 1992). For finite samples, the classical estimate of the variance of a contrast turns out to be unbiased, even in the presence of long-range correlations. This is very much in contrast to the estimation of constants, where the variance decays to zero with a slower rate than under independence [see (3) and Section 3.3], and confirms the experience of applied statisticians that constants are much more difficult to estimate than contrasts. Standard programs for ANOVA can therefore be used as long as the results refer to contrasts and the validity of the inference. Another question is how much efficiency can be gained by estimating the covariances and/or choosing suitable designs. For example, by using a blocked randomized design instead of complete randomization, the variance of the least squares estimator can be decreased considerably (see Künsch, Beran and Hampel, 1992). At the same time, the efficiency gain of the BLUE compared to least squares becomes negligible.

Nonparametric regression of the form

$$Y_i = g(i/n) + \varepsilon_i, \tag{10}$$

with errors ε_i having correlations (7), was considered by Hall and Hart (1989). The optimal rate of convergence of kernel estimators of g turns out to be $n^{(4H-4)/(6-2H)}$, which is slower than the optimal rate of convergence $n^{-2/5}$ for weakly dependent observations. For twice differentiable functions g, the optimal bandwidth is found to be proportional to $n^{(2H-2)/(6-2H)}$.

2.3 U-Statistics

Limit theorems for the empirical distribution function, U-statistics and von Mises statistics for long-range dependent processes were considered by Dehling and Taqqu (1989). They consider observations $Y_i = G(X_i)$, $i = 1, \ldots, n$, where X_i is a stationary Gaussian process satisfying (7) and G is a function of Hermite rank $m \geq 1$. Here G is said to have Hermite rank m, if $E(G(X_i)H_k(X_i)) = 0$ for $0 \leq k \leq m - 1$ and $E(G(X_i)H_m(X_i)) \neq 0$, where H_k is the kth Hermite polynomial. In particular, if $m = 1$ and $F_n(x)$ is the empirical distribution function $F_n(x) = \Sigma_{i=1}^n 1\{Y_i \leq x\}$, then $L_2(n)^{-1/2}n^{1-H}(F_n(x) - F(x))$ converges weakly, in the space of all cadlag-functions (functions which are right continuous everywhere with existing limit from the left) on $[-\infty, \infty]$ equipped with the supremum norm, to a constant $c(x)$ times a standard normal variable. Thus, in contrast to the case of independent or weakly dependent observations, the limit of the empirical process is, up to a deterministic constant, one random variable instead of a stochastic process. The consequences for goodness of fit tests (Beran and Ghosh, 1990, 1991) are discussed below in Section 3.2. For a law of the iterated logarithm for F_n and its application to U- and von Mises statistics see Dehling and Taqqu (1988). Also, for applications to the chi-squared goodness of fit test and the Cramer-von Mises-Smirnov goodness of fit, see Dehling and Taqqu (1990).

2.4 Estimation of H

As we saw in the previous sections, the rate of convergence of many standard statistics is determined by the value of the parameter H. For reliable statistical inference, it is therefore important to obtain a good estimate of H from the data. 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). It has some good robustness properties, in particular with respect to long-tailed distributions (see Mandelbrot and Taqqu, 1979), however under the Gaussian model it loses much efficiency compared with maximum likelihood type methods. A Bayesian approach to estimating H is discussed in Carlin, Dempster and Jonas (1985) and Carlin and Dempster (1989). Asymptotic normality of the maximum likelihood estimator was proved by Yajima (1985) in a special case. The general proof was given by Dahlhaus (1989). Efficient approximation to maximum likelihood has been considered by several authors. In Whittle's approximation (Beran, 1986; Fox and Taqqu, 1986) the inverse covariance matrix of $(X_1, \ldots, X_n)$ is replaced by the two-sided infinite dimensional matrix $(a_{k,l})_{k,l=-\infty,\infty}$ with $a_{k,l} = a_{k-l} =$

$(2\pi)^{-1}\int_{-\pi}^{\pi} f^{-1}(x)\cos(k-l)x\,dx$. One then has to minimize (with respect to H and σ^2)

$$(11)\qquad \sum_{k,l=1}^{n} a_{k-l}(C_k - R_k) + (2\pi)^{-1}\int_{-\pi}^{\pi} \log f(x)\,dx,$$

where $C_k = n^{-1}\sum_{i=1}^{n-|k|}(X_i - \overline{X}_n)(X_{i+k} - \overline{X}_n)$. A discretized version of (11), the so-called HUB00-estimator, was proposed by Graf (1983). Heuristically, his approach is based on the fact that the periodogram ordinates $I(\omega) = (2\pi n)^{-1}|\sum_{j=1}^{n} X_j \exp(ij\omega)|^2$ are asymptotically exponentially distributed and independent for different frequencies (also see Yajima, 1989). To be precise, this limit theorem actually only holds for frequencies above $\varepsilon n^{-\gamma}$ for any fixed $\gamma < 1/2$ and $\varepsilon > 0$ (Künsch, 1987). However, due to the averaging in (11) (or in the corresponding discrete version), the periodogram ordinates for frequencies below $\varepsilon n^{-\gamma}$ have asymptotically no effect on the limit distribution of the estimator. Graf also wrote a FORTRAN program for his estimator. This program seems to be rather fast, so that excessive CPU time does not seem to be a problem. The typical CPU times (on a CDC Cybers 721/722/174 computer) reported are 0.26 for $n = 64$, 0.45 for $n = 128$, 0.83 for $n = 256$, 1.68 for $n = 512$ and 3.23 for $n = 1{,}024$. Note that a "quick and dirty" way to get an approximate solution of (11) can be obtained easily for models (such as, e.g., fractional Gaussian noise) where H and the variance are the only parameters to be estimated:

1. Use the following parametrization: $f(\lambda; H, \sigma_\varepsilon^2) = \sigma_\varepsilon^2 f(\lambda; H, 1)$ where $\int_{-\pi}^{\pi} \log f(\lambda; H, 1)\,d\lambda = 0$. Note that σ_ε^2 is the variance of the innovation in the infinite AR representation of the process.
2. Find the minimum of the function $g(H) = \sum_{i=1}^{n^*} I(\lambda_i)/f(\lambda_i; H, 1)$ (n^* = integer part of $(n-1)/2$, $\lambda_i = 2\pi i/n$) to obtain $\hat{H}$. Obviously, this can be done by simply calculating $g(H)$ on a sufficiently fine grid of H-values (and perhaps plotting $g(H)$ against H). The scale parameter is then estimated by $\hat{\sigma}_\varepsilon^2 = g(\hat{H})$.

If more than two parameters have to be estimated, then computational problems become more serious. For a discussion of these issues, see Hosking (1984), Carlin (1987) and Haslett and Raftery (1989).

A third possibility of approximating the maximum likelihood equations is to replace the inverse covariance matrix by a one-sided infinite dimensional matrix. This results in minimizing

$$(12)\qquad n^{-1}\sigma_\varepsilon^{-2}\sum_{i=2}^{n}\Big(X_i - \overline{X}_n - \sum_{k=1}^{i-1} b_k(X_{i-k} - \overline{X}_n)\Big)^2 + \log\sigma_\varepsilon^2,$$

where $X_i - \mu = \sum_{j=1}^{\infty} b_j(X_{i-j} - \mu) + \varepsilon_i$ is the infinite AR representation of $(X_i)_{i\in Z}$ and $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_i)$. Essentially, (12) minimizes the sum of the squared residuals $r_i = X_i - \overline{X}_n - \sum_{j=1}^{i-1} b_j(X_{i-j} - \overline{X}_n)$ $(i = 1, \ldots, n)$. Note that the residuals r_i are estimates of the corresponding innovations $\varepsilon_i = X_i - \mu - \sum_{j=1}^{\infty} b_j(X_{i-j} - \mu)$ $(i = 1, \ldots, n)$. The variance of the innovations, $\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2$ is then estimated by $n^{-1}\sum_{j=2}^{n} r_j^2$. This is analogous to corresponding methods for autoregressive models. The advantage of (12) is that it is easy to obtain a more robust version (robust in the time domain) by a simple modification of (12) [see the equations (15a) and (15b) below]. A problem with approximation (12) is that the residuals r_i can be a poor approximation to the innovations ε_i, in particular for i close to n and large values of H. In particular, this implies that the residuals r_i are heteroscedastic which is not taken into account in (12). Although this has no influence on the asymptotic distribution, it can lead to a considerable bias and efficiency loss, if n is not very large. Following Haslett and Raftery (1989), (12) can be improved by replacing r_i by $r_i^* = (X_i - \hat{X}_i)v_i^{-1/2}$ where $\hat{X}_i$ is the best linear prediction of X_i given $X_{i-1}, \ldots, X_1$ and v_i is the conditional variance of $X_i - \hat{X}_i$. The residuals r_i^* can be calculated exactly by the Durbin-Levinson recursion (e.g., Hosking, 1982).

For all three estimators of $\theta = H$, or more generally of a parameter vector $\underline{\theta} = (\theta_1, H, \theta_3, \ldots, \theta_M)$ (where θ_1 is a scale parameter and the additional parameter θ_i, $i \geq 3$ characterize the short range dependence structure), a central limit theorem holds, with the same asymptotic covariance matrix as for the maximum likelihood estimator (Beran, 1984, 1986; Fox and Taqqu, 1986; Dahlhaus, 1989), that is $n^{1/2}(\hat{\underline{\theta}} - \underline{\theta})$ is asymptotically normally distributed with zero mean and covariance matrix $V = 2D^{-1}$ where

$$(13)\qquad D_{ij} = (2\pi)^{-1}\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_i}\log f(x)\,\frac{\partial}{\partial\theta_j}\log f(x)\,dx.$$

Note that $1/2D$ is the asymptotic Fisher information matrix (Dahlhaus, 1989) so that $\hat{\underline{\theta}}$ is asymptotically efficient.

Considering a class of nested models with spectral density $f(x) = f(x; \underline{\theta})$ and estimating θ by the approximate maximum likelihood method (11), one can choose the dimension of $\underline{\theta}$ by applying a version of Akaike's criterion (Beran, 1986, 1989b): Choose the model for which

$$(14)\qquad AIC(M) = \sum_{k=-(n-1)}^{n-1} a_k(\hat{\underline{\theta}})C_k^* + \sum_{|k|\geq n} a_k(\hat{\underline{\theta}})\hat{R}_k + (2\pi)^{-1}\int_{-\pi}^{\pi}\log f(x;\hat{\underline{\theta}})\,dx + \frac{2M}{n}$$

is minimal. Here $C_k^* = n(n-|k|)^{-1}C_k$ is the unbiased and $\hat{R}_k$ a consistent estimator of R_k. For Markov processes this criterion was derived by Künsch (1981). There the second term vanishes. By an analogous technique as in Shibata (1976), (14) can be shown to be

inconsistent. It overestimates asymptotically the number of parameters with positive probability. This means that even asymptotically we have no guarantee that we do not use an unnecessarily complicated model. Consistency is however not necessarily the most important property. In practice the true model, if there is any at all, will often by very complicated. The best we can hope is to obtain a reasonable approximate model. In many applications, the relevant criterion for the quality of such a model is its predictive power. In the sense of predictions, the AIC was shown to be optimal in the context of short-memory linear models (Shibata, 1980). I would suspect that an analogous result holds here, though it would need a thorough investigation.

In practice, parametric models are only approximations. Therefore, it is important to know how $\hat{\underline{\theta}}$ behaves under deviations from the model and to find alternative robust methods of estimation. Here, we have at least three kinds of possible deviations: nongaussianity, another form of the spectrum at high frequencies or another function $L_1(\cdot)$ in (6), nonstationarity.

All three approximate maximum likelihood estimators are defined via quadratic forms. This makes them rather sensitive to deviations from Gaussianity. The central limit theorem still holds for certain types of non-Gaussian processes (Avram, 1988; Giraitis, 1989; Giraitis and Surgailis, 1989; Terrin and Taqqu, 1991). However, there are also examples of non-Gaussian processes where the rate of convergence is slower than $n^{-1/2}$ (cf. Fox and Taqqu, 1985). In practice, we may hope that such extreme cases do not occur. Yet it would be useful to find methods that 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 (12) by differentiation. That is, instead of minimizing (12) one can solve the robustified "normal equations"

$$\text{(15a)} \qquad \sum_{l=2}^{n} \psi\left(\sigma_\varepsilon^{-2} r_l \frac{\partial}{\partial \theta_j} r_l\right) = 0 \quad (j = 2, \ldots, M)$$

and

$$\text{(15b)} \qquad \sum_{l=2}^{n} \chi(r_l/\sigma_\varepsilon) = 0,$$

where the ψ- and the χ-function are suitably chosen bounded functions and $\underline{\theta} = (\sigma_\varepsilon^2, H, \theta_3, \ldots, \theta_M)$. An analogous central limit theorem also holds for such estimators. In view of the results by Martin and Yohai (1986) for (short-memory) linear processes, it seems most likely that robustness can only be achieved if the ψ- and χ-functions are redescending. A more detailed analysis of the robustness properties of such estimators would be needed to decide which functions give at the same time robust and efficient (under the Gaussian model) estimates.

The second kind of deviation, in other words, deviations from the assumed shape of the spectrum, has to be expected for most data sets, in particular if we use rather simple models – like fractional Gaussian noise – that only reflect the long-range aspects of the data. For short lags, the correlations are typically different from those described by, for example, fractional Gaussian noise. If one uses a wrong model, then the maximum likelihood estimator (or any of the approximate MLE's) of H will be biased. This is the case because the maximum likelihood estimator uses periodogram ordinates in the whole frequency range $[0, \pi]$ (see, e.g., the contributions to the discussion of Haslett and Raftery, 1989, by Smith and Dempster; also see Graf, 1983; Graf et al., 1984). If, for example, the model has a flat spectrum at high frequencies but the periodogram has high peaks at high frequencies, the bias can be very large.

To avoid bias in $\hat{H}$, one has either to choose the right (or a better) model, for instance by introducing additional parameters, 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. This can be done however only at the cost of a slower rate of convergence, by using only m periodogram ordinates for the estimation of H with $m < (n-1)/2$ and $m/n \to 0$ $(n \to \infty)$ (Geweke and Porter-Hudak, 1983). It is interesting to note that this approach has analogies with the estimation of the tail of a distribution of the form $f(x) = ax^{-2c}(1 + bx^c + o(x^c))$ $(x \to \infty, a, c > 0)$ (see R. Smith's contribution to the discussion of Haslett and Raftery, 1989; Hall, 1982; Hall and Welsh, 1984, 1985). In both cases (estimation of the pole of a spectrum at zero and tail estimation for a distribution), consistency can be achieved only at the cost of a slower rate of convergence. Hall (1982) and Hall and Welsh (1984, 1985) also give an optimal choice for m and the corresponding optimal rate of convergence. It might be worth considering if analogous optimally results could be derived for the estimation of H. A compromise was proposed by Graf (Graf, 1983; Graf, Hampel and Tacier, 1984; for an explicit formula, see Beran, 1989a). His HUBINC-estimator bounds the influence of high-frequency ordinates in a way which guarantees approximate consistency under deviations from the ideal shape of the spectrum and rate of convergence $n^{-1/2}$.

Under fractional Gaussian noise the bias is zero and efficiency is above 74% for all $H \in (0, 1)$. The estimator is defined as follows: Let $a = 1 - 2H$, $\underline{\theta} = (a, \log b)$, $q_n(a, \lambda) = 2\pi\lambda^{2H-1}[\sin\{\pi H\}\Gamma\{2H+1\}]^{-1}E[I\{\lambda\}]$, $\lambda_{k,n} = 2\pi k/n$, $u(\lambda) = \min[3.5, 0.5 + 0.375\pi/\lambda]$, $l(\lambda) = \max[0, 0.6\{\lambda/\pi - 0.36\}^{1/2}]$, $z_k = I(\lambda_{k,n})$ and $w(\underline{\theta}, z, \lambda) = [zb^{-1} \cdot \lambda^{-a} q_n^{-1}\{a\ \lambda\}]_{l(\lambda)}^{u(\lambda)}$ where $[]_{l(\lambda)}^{u(\lambda)}$ denotes truncation by $l(\lambda)$ from below and by $u(\lambda)$ from above. Furthermore, let $mn(\lambda)$ be equal to the mean of a standardized exponen-

tial random variable that is is truncated by $l(\lambda)$ and $u(\lambda)$. Also we define n^* to be the largest integer below $n/2 - 1/2$, $\delta_1(a, \lambda) = \log\lambda + (\partial/\partial a)\log\, q_n(a, \lambda)$, $\delta_2(z, \lambda) = w(\underline{\theta}, z, \lambda) - mn(\lambda)$ and $\psi^{(n)}(z, \lambda, \underline{\theta}) = [\delta_1\{a, \lambda\}\delta_2\{z, \lambda\}, \delta_2\{z, \lambda\}]^t$. The HUBINC estimator of $\underline{\theta}$ is the solution of

$$\sum_{k=1}^{n^*} \psi^{(n)}(z_k, \lambda_{k,n}, \underline{\theta}) = 0. \tag{16}$$

Note that if we choose $l(\lambda) = 0$ and $u(\lambda) = \infty$, then we obtain Graf's approximate maximum likelihood estimator HUB00. Graf wrote a FORTRAN program to solve (16). As for the approximate maximum likelihood estimator, CPU time does not seem to be a problem. The typical CPU times are of the order of magnitude give above for the approximate MLE.

A negative slope of the logarithm of the spectrum near zero 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, Gupta and Waymire (1983) showed that for the nonstationary sequence $Y_i = X_i + m_i$ with X_i stationary and weakly dependent and m_i a slowly decreasing trend, the R/S-estimator of H is asymptotically between 1/2 and 1. This is an argument against using the R/S-statistic, at least if the possibility of trend can not be excluded a priori. A way 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 noncentrality parameter going to zero uniformly for $|j| > \varepsilon n^{-\gamma}$ (for any fixed $\varepsilon > 0, \gamma < 1/2$), whereas for a stationary process with spectrum (6) 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. They are based on averages of periodogram ordinates divided by the spectral density. The effect of periodogram ordinates for frequencies below $\varepsilon n^{-\gamma}$ is therefore asymptotically negligible. 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 (for related results, also see Hurvich and Beltrao, 1992). 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. Since we always deal with finite data sets, it is in principle not possible to decide whether the spectrum of the underlying process has a pole at zero or if it is continuous at the origin with negative slope in its neighbourhood. Also, transient phenomena can not be excluded a priori. However, if the underlying process in fact has long memory, then in most cases short-memory models (e.g., ARMA) will not give a good fit unless we use many (asymptotically an infinite number of) parameters. Model selection criteria that penalize lack of parsimony [e.g., (14)] will therefore tend to exclude such models. In contrast to that, the stationary processes defined by (6) [or (7)] describe long-range dependence by one single parameter H.

3. TESTS AND CONFIDENCE INTERVALS

3.1 Tests for Long-Range Dependence

Several tests for the so-called 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). They consider tests based on the R/S-statistic, tests derived from likelihood ratio tests and locally optimal tests which maximize the derivative of the power function at $H = 1/2$. In cases where one has a clearly specified null hypothesis and alternative, one might also derive other types of tests, such as exact likelihood ratio tests or likelihood ratio tests using an approximation to the likelihood (see Section 2.4), Wald tests, score tests and tests based on Bayes factors. The exact behaviour of these tests in the context long-memory processes is an open problem.

Given a stationary process with long-range dependence, it is obvious from the central limit theorem in Section 2.4 how to construct confidence intervals for $\underline{\theta}$. For example, for the model of fractional Gaussian noise, an approximate 95%-confidence interval for H is given by $\hat{H} \pm 1.96 V_{11}^{1/2} n^{-1/2}$ where $V = 2D^{-1}$ and D is the 2×2-matrix defined by (13) and (2). 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.

3.2 Goodness of Fit Tests

First consider testing the (composite) null hypothesis $H_0: f(x) \equiv f(x; \underline{\theta})$ against the alternative $H_A: f(x) \not\equiv f(x; \underline{\theta})$, where f is the true spectral density and $f(x; \underline{\theta})$ is the spectral density of a parametric model. In other words, we want to test if the spectral density of our parametric model is equal to the true spectral density in the whole frequency range $[0, \pi]$. A straightforward way to test this hypothesis is to look at the estimated residuals obtained from filtering the data $X_i = \sum_{k=1}^{\infty} \psi_k X_{i-k} + \varepsilon_i$ using the estimated coefficients $\psi_k(\hat{\underline{\theta}})$. A portmanteau statistic based on an increasing number of correlations of the residual process can then be used to construct a consistent omnibus test (Beran, 1992). The distribution of the test statistic under the composite hypothesis (i.e., θ unknown) turns out to be the same as under

the simple hypothesis (θ known). Obviously, if one suspects deviation from the model spectrum in a specific direction, then other more powerful tests can be constructed. In particular, if one wants to test a simple hypothesis against a simple alternative (e.g., fractional Gaussian noise versus AR(1)) one can construct likelihood ratio type tests (see section 3.1).

The effect of long-range dependence on goodness of fit tests for a (marginal) distribution, in particular testing normality, was considered in Beran and Ghosh (1990, 1991). The observation that, in practice, goodness of fit tests for a distribution almost always reject the null hypothesis for sufficiently large data sets, turns out to be not necessarily only due to the null hypothesis being almost never true but also due to the use of wrong rejection regions valid only under independence. For example, suppose that we test the simple null hypothesis that our data come from a standard normal distribution. We use the Kolmogorov-Smirnov test together with a corresponding table of critical values obtained under the assumption of independence. If in fact there is long-range dependence in the underlying process, then the probability of rejecting the null hypothesis tends to one (as n tends to infinity), even if the marginal distribution of the process is actually standard normal. The case of a simple hypothesis is different from the case of a composite hypothesis. In the former, classical goodness of fit tests reject the null hypothesis asymptotically with probability one, even if the null hypothesis is true. If the mean and variance are estimated from the data (composite hypothesis), rejection under the null hypothesis with asymptotic probability one only occurs for $H > 5/6$. The reason for the different behaviour for $H < 5/6$ is that, in this case, goodness of fit statistics can be written asymptotically as functionals of so-called Hermite processes of rank 3 and higher. For these processes the central limit theorem holds for $H < 1 - 1/(2 \cdot 3) = 5/6$ (e.g., Taqqu, 1979; Breuer and Major, 1983).

3.3 Tests for Location and Prediction Intervals

Confidence intervals for the mean have to take into account not only the slower rate of convergence, but also the variability of $\hat{\underline{\theta}}$. If (6) holds, then $\mathrm{var}(\overline{X}_n)/(c(n)n^{2H-2})$ converges to one (as $n \to \infty$) where $c(n) = 2\pi^{-1}L_1(n^{-1})\Gamma(-2H)\sin\pi(1/2 - H)$. 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 (6) can be defined by

$$T = (\overline{X}_n - \mu)c(n;\underline{\theta})^{-1/2}n^{1-\hat{H}}, \tag{17}$$

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, 1989a) that is also applicable to rather short series (approximately $n \geq 60$):

$$\begin{aligned} P(T \leq u) \approx\ & (1 - \Phi((1 - H)n^{1/2}/\sigma)) \\ & + (1 - \Phi(Hn^{1/2}/\sigma)) \\ & + \int_{-H}^{1-H}\int_{-\infty}^{\infty} \Phi(ug(n, \underline{\theta}; \underline{z}))\varphi_n(\underline{z})\, dz_1\, dz_2, \end{aligned} \tag{18}$$

where $\varphi_n(\underline{z})$ is the two-dimensional normal density with zero mean and covariance matrix $n^{-1}V$, V as in section 2.4 and $g(n, \underline{\theta}; \underline{z}) = (c(n; \underline{\theta} + \underline{z})/c(n; \underline{\theta}))^{1/2}n^{z_1}$. 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.

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

Prediction of single future observations is discussed by Granger and Joyeux (1980), Hosking (1981) and Peiris and Perera (1988) for the case where all parameters of the model are known. For fractional ARMA processes, predictions and prediction intervals follow by standard calculations from (5).

4. TWO DATA EXAMPLES

We apply some of the methods described in the previous sections to the two data sets introduced in Section 1.

The Nile River data are plotted in Figure 1. The periodogram is plotted in log-log-coordinates in Figure 3. The approximate maximum likelihood estimate HUB00 and the HUBINC estimate of H are 0.837 and 0.847 respectively. The corresponding approximate

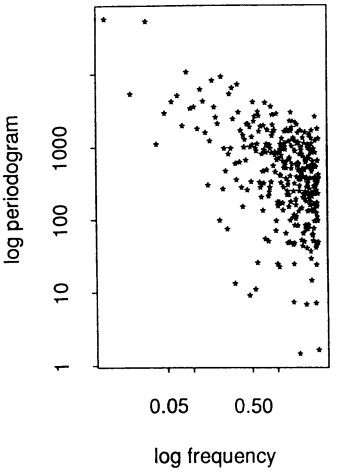

FIG. 3. *Periodogram (in log-log coordinates) of the Nile River minima.*

95%- and 99%-confidence intervals for H [obtained from (13)] are [0.786, 0.888] and [0.770, 0.904] for the approximate MLE and [0.788, 0.906] and [0.770, 0.925] for the HUBINC estimate. In view of Figure 3, it is not very surprising that there is such strong evidence for long-range dependence. The question arises how well a simple model like fractional Gaussian noise actually fits the data. This model was proposed by Mandelbrot and Wallis (1968, 1969) for the Nile River data. Its spectrum is flat for high frequencies so that it only can model long-memory properties. Using the generalized portmanteau goodness of fit statistic described in Section 3.2, we obtain an approximate P-value of 0.71. Thus, although the spectrum of fractional Gaussian noise is described by one parameter only (apart from the variance), it indeed provides a very good fit to the observed spectrum. If instead we try to fit an autoregressive model we need to use many more parameters. Using the Akaike criterion, the best ARMA-model turns out to be an AR(7)-model. The P-value for the goodness of fit test is approximately equal to 0.75. So, the quality of the fit is practically the same as for fractional Gaussian noise. However, we need seven parameters instead of one. Moreover, the order of the AR-model needed seems to increase with the sample size. For the first $n = 50$, 100, 400 and 660 observations, the selection criterion chooses the orders 1, 2, 4 and 7, respectively.

As second example we consider the NBS data described in Section 1. The differences of the measurements (in micrograms) from 1 kg are plotted in serial order in Figure 2. The periodogram in log-log coordinates is given in Figure 4. The dates when the observations were made are actually not exactly equidistant. These irregularities, however, mainly influence the spectrum at high frequencies. If we want to calculate confidence intervals for the mean, then we are only interested in the behaviour of the spectrum near the origin (see Section 3.3). Under these circumstances it seems particularly desirable that the estimate of H does not depend strongly on the high-frequency periodogram ordinates. Therefore, an estimator like the HUBINC estimate is more adequate than one of the approximate maximum likelihood estimates. The HUBINC estimate of H is equal to 0.602, which is significantly larger than 0.5, even at the level of significance 1% (Graf, 1983; Graf et al., 1984). The 95%- and 99%-confidence intervals for the mean, based on (18), are [−19.4886, −19.4649] and [−19.4937, −19.4599], respectively. They are more than twice as large as confidence intervals based on Student's t-test. We also computed the confidence intervals based on the first 128 observations. The length of the intervals is 0.04582 and 0.06882 for confidence levels 95% and 99% respectively. Comparing this to the length of the intervals based on all 289 observations, we see that for such small values of n the decreasing fluctuation of $\hat{H}$ has a considerable effect on the intervals so that their length decreases faster than $n^{\hat{H}-1}$. Also, already for $n = 128$ the confidence intervals are more than twice as large than the confidence intervals based on the t-test. This illustrates that even for relatively small sample sizes and weak long-range dependence the effect of such dependence on statistical inference is very strong. Here, this effect is even stronger for $n = 128$, because the estimate of H based on the first 128 observations turns out to be slightly higher ($\hat{H} = 0.634$) than the estimate for the whole series.

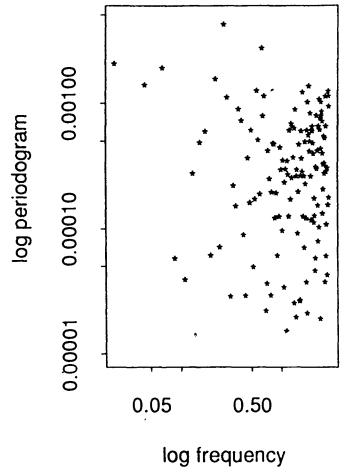

FIG. 4. *Periodogram (in log-log coordinates) of the NBS 1-kg check standard weights.*

5. CONCLUDING REMARKS

Long-range dependence is often encountered in practice, not only in hydrology and geophysics but in all fields of statistical applications. If not taken into account, it can completely invalidate statistical inference. For many standard situations, new statistical methods as well as properties of classical techniques are sufficiently known nowadays to be used in practice.

More research is needed, however, 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 methods that also perform well under long-range dependence should be developed. Also, in many applications spatial long-range dependence as well as multivariate time series (or multivariate spatial data) with long-range dependence occur. The development of statistical methods for such data will certainly be a rewarding task for future research.

ACKNOWLEDGMENTS

The author would like to thank Frank Hampel, Hans Künsch, Werner Stahel, Hanspeter Graf, Stamatis Cambanis and Emanuel Parzen for many inspiring discussions on the subject of long-range dependence and data analysis. Thanks go also to the U.S. National Bureau of Standards, Washington, for providing me with the data set used in this paper. Finally, I would like to thank the Editor and the referees for their very helpful and constructive comments.

REFERENCES

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

AVRAM, F. (1988). On the central limit theorem for sums of Wick powers. Preprint.

BERAN, J. (1984). Maximum likelihood estimation for Gaussian processes with long-range dependence. In *Abstracts Book, 16th European Meeting of Statisticians.* Marburg, Germany.

BERAN, J. (1986). Estimation, testing and prediction for self-similar and related processes. Ph.D. dissertation, ETH, Zürich.

BERAN, J. (1988). Statistical aspects of stationary processes with long-range dependence. Mimeo Series 1743, Dept. Statistics, Univ. North Carolina, Chapel Hill.

BERAN, J. (1989a). A test of location for data with slowly decaying serial correlations. *Biometrika* **76** 261–269.

BERAN, J. (1989b). Akaike's model choice criterion, long-range dependence and approximate maximum likelihood estimation. Technical Report 96, Dept. Statistics, Texas A&M Univ.

BERAN, J. (1991). M-estimators of location for data with slowly decaying serial correlations. *J. Amer. Statist. Assoc.* **86** 704–708.

BERAN, J. (1992). A goodness of fit tests for time series with long-range dependence. *J. Roy. Statist. Soc. Ser. B* **54** 749–760.

BERAN, J. and GHOSH, S. (1990). Goodness of fit tests and long-range dependence. *Directions in Robust Statistics and Diagnostics, Part I. IMA Vol. Math. Appl.* **33** 21–33. Springer, New York.

BERAN, J. and GHOSH, S. (1991). Slowly decaying correlations, testing normality, nuisance parameters. *J. Amer. Statist. Assoc.* **86** 785–791.

BERAN, J. and KUNSCH, H. R. (1985). Location estimators for processes with long-range dependence. Research Report 40, Seminar für Statistik, ETH, Zürich.

BERAN, J., SHERMAN, R., TAQQU, M. S. and WILLINGER, W. (1992). Variable-bit-rate video traffic and long-range dependence. Unpublished manuscript.

BHATTACHARYA, R. N., GUPTA, V. K. and WAYMIRE, E. (1983). The Hurst effect under trends. *J. Appl. Probab.* **20** 649–662.

BOX, G. E. P., HUNTER, W. G. and HUNTER, J. S. (1978). *Statistics for Experimenters.* Wiley, New York.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

BREUER, P. and MAJOR, P. (1983). Central limit theorems for non-linear functionals of Gaussian fields. *J. Multivariate Anal.* **13** 425–441.

BURROUGH, P. A. (1981). Fractal dimensions of landscapes and other environmental data. *Nature* **294** 240–242.

CARLIN, J. B. (1987). Seasonal analysis of economic time series. Ph.D. dissertation, Dept. Statistics, Harvard Univ.

CARLIN, J. B. and DEMPSTER, P. (1989). Sensitivity analysis of seasonal adjustments: Empirical case studies (with discussion). *J. Amer. Statist. Assoc.* **84** 6–32.

CARLIN, J. B., DEMPSTER, P. and JONES, A. B. (1985). On methods and models for Bayesian time series analysis. *J. Econometrics* **30** 67–90.

CASSANDRO, M. and JONA-LASINIO, G. (1978). Critical behaviour and probability theory. *Adv. in Phys.* **27** 913–941.

COX, D. R. (1984). Long-range dependence: A review. In *Statistics: An Appraisal* (H. A. David and H. T. David, eds.) 55–74. Iowa State Univ. Press.

COX, D. R. and TOWNSEND, M. W. H. (1948). The use of the correlogram in measuring yarn irregularity. *Proc. Roy. Soc. Edinburgh Sect. A* **63** 290–311.

DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766.

DAMERAU, F. J. and MANDELBROT, B. B. (1973). Tests of the degree of word clustering in samples of written English. *Linguistics* **102** 58–75.

DAVIES, R. B. and HARTE, D. S. (1987). Tests for Hurst effect. *Biometrika* **74** 95–102.

DEHLING, H. and TAQQU, M. S. (1988). The functional law of the iterated logarithm for the empirical process of some long-range dependent sequences. *Statist. Probab. Lett.* **7** 81–85.

DEHLING, H. and TAQQU, M. S. (1989). The empirical process of some long-range dependent sequences with application to U-statistics. *Ann. Statist.* **17** 1767–1783.

DEHLING, H. and TAQQU, M. S. (1991). Bivariate symmetric statistics of long-range dependent observations. *J. Statist. Plann. Inference* **28** 153–165.

FOX, R. and TAQQU, M. S. (1985). Noncentral limit theorems for quadratic forms in random variables having long-range dependence. *Ann. Probab.* **13** 428–446.

FOX, R. and TAQQU, M. S. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532.

GAY, R. and HEYDE, C. C. (1990). On a class of random field models which allows long range dependence. *Biometrika* **77** 401–403.

GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.* **4** 221–237.

GIRATIS, L. (1989). Central limit theorems for polynomial forms. *Lietuvos Matematikos Rinkinys* **29** 266–289.

GIRAITIS, L. and SURGAILIS, D. (1989). CLT for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate. Preprint.

GRAF, H. P. (1983). Long-range correlations and estimation of the self-similarity parameter. Ph.D. dissertation, ETH, Zürich.

GRAF, H. P., HAMPEL, F. R. and TACIER, J. (1984). The problem of unsuspected serial correlations. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26.** Springer, New York.

GRANGER, C. W. J. (1966). The typical spectral shape of an economic variable. *Econometrica* **34** 150–161.

GRANGER, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *J. Econometrics* **14** 227–238.

GRANGER, C. W. J. and JOYEUX R. (1980). An introduction to long-range time series models and fractional differencing. *J. Time Ser. Anal.* **1** 15–30.

GRAY, H. L., ZHANG, N. and WOODWARD, W. A. (1989). On generalized fractional processes. *J. Time Ser. Anal.* **10** 233–257.

GRENANDER, V. and ROSENBLATT, M. (1957). *Analysis of Stationary Time Series.* Wiley, New York.

HALL, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44** 37–42.

HALL, P. and HART, J. D. (1989). Nonparametric regression with long-range dependence. Technical Report 88, Dept. Statistics, Texas A&M Univ.

HALL, P. and WELSH, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12** 1079–1084.

HALL, P. and WELSH, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13** 331–341.

HAMPEL, F. R. (1987). Data analysis and self-similar processes. *Proceedings of the 46th Session of ISI, Book 4*, Tokyo.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics. The Approach Based on Influence Functions.* Wiley, New York.

HASLETT, J. and RAFTERY, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource (with discussion). *J. Roy. Statist. Soc. Ser. C* **38** 1–21.

HODGES, J. L. and LEHMANN, E. L. (1970). Deficiency. *Ann. Math. Statist.* **41** 783–801.

HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176.

HOSKING, J. R. M. (1982). Some models of persistence in time series. In *Times Series Analysis: Theory and Practice 1* (O. D. Anderson, ed.) 641–654. North-Holland, Amsterdam.

HOSKING, J. R. M. (1984). Modelling persistence in hydrological time series using fractional differencing. *Water Resources Research* **20** 1898–1908.

HUI, Y. V. and LI, W. K. (1988). On fractionally differenced periodic processes. Unpublished manuscript, Chinese Univ. Hong Kong and Univ. Hong Kong.

HURST, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* **116** 770–779.

HURVICH, C. M. and BELTRAO, K. I. (1992). Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series. Preprint.

JEFFREYS, H. (1939). *Theory of Probability.* Clarendon Press, Oxford.

KLEMES, V. (1974). The Hurst phenomenon: A puzzle? *Water Resources Research* **10** 675–688.

KOLMOGOROV, A. N. (1940). Wienersche Spiralen und einige andere interessante Kurven in Hlbertschen Raum. *Comptes Rendus (Doklady) Académie des Sciences de l'URSS (N.S.)* **26** 115–118.

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. Probab.* **23** 1025–1030.

KÜNSCH, H. R. (1987). Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society, Tashkent* (Yu. Prohorov and V. V. Sazonov, eds.) **1** 67–74. VNU Science Press, Utrecht.

KÜNSCH, H., BERAN, J. and HAMPEL, F. (1992). Contrasts under long-range correlations. To appear.

LAMPERTI, J. W. (1962). Semi-stable stochastic processes. *Trans. Amer. Math. Soc.* **104** 62–78.

LAWRANCE, A. J. and KOTEGODA, N. T. (1977). Stochastic modelling of riverflow time series (with discussion). *J. Roy. Statist. Soc. Ser. A* **140** 1–47.

LI, W. K. and MCLEOD, A. I. (1986). Fractional time series modelling. *Biometrika* **73** 217–221.

MANDELBROT, B. B. (1969). Long-run linearity, locally Gaussian process, H-spectra and infinite variance. *Internat. Econom. Rev.* **10** 82–111.

MANDELBROT, B. B. (1971). When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models. *Reviews of Economics and Statistics* **LIII** 225–236.

MANDELBROT, B. B. (1973). Le problème de la réalité des cycle lents et le syndrome de Joseph. *Economie Appliquée* **26** 349–365.

MANDELBROT, B. B. (1983). *The Fractal Geometry of Nature.* Freeman, New York.

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

MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437.

MANDELBROT, B. B. and WALLIS, J. R. (1968). Noah, Joseph and operational hydrology. *Water Resources Research* **4** 909–918.

MANDELBROT, B. B. and WALLIS, J. R. (1969). Computer experiments with fractional Gaussian noises. *Water Resources Research* **5** 228–267.

MARTIN, R. D. and YOHAI, V. J. (1986). Influence functionals for times series (with discussion). *Ann. Statist.* **14** 781–818.

MATHERON, G. (1973). The intrinsic random functions and their applications. *Adv. in Appl. Probab.* **5** 439–468.

MOSTELLER, F. and TUKEY, J. W. (1977). *Data Analysis and Regression: A Second Course in Statistics.* Addison-Wesley, Reading, Mass.

NEWCOMB, S. (1895). Astronomical constants (The elements of the four inner planets and the fundamental constants of astronomy). Supplement to the American Ephemeris and Nautical Almanac for 1897, US Government Printing Office, Washington, DC.

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.

PEIRIS, M. S. and PERERA, B. J. C. (1988). On prediction with fractionally differenced ARIMA models. *J. Time Ser. Anal.* **9** 215–220.

PERCIVAL, D. B. (1985). On the sample mean and variance of a long memory process. Technical Report 69, Dept. Statistics, Univ. Washington.

PORTER-HUDAK, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *J. Amer. Statist. Assoc.* **85** 338–344.

RAMSAY, F. (1974). Characterization of the partial autocorrelation function. *Ann. Statist.* **2** 1296–1303.

SAMAROV, A. and TAQQU, M. S. (1988). On the efficiency of the sample mean in long-memory noise. *J. Time Ser. Anal.* **9** 191–200.

SCHEFFÉ, H. (1959). *The Analysis of Variance.* Wiley, New York.

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

SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statistics* **8** 147–164.

SMITH, H. F. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science* **28** 1–23.

SOLO, V. (1989). Intrinsic random functions and the paradox of 1/f noise. Preprint.

STUDENT. (1927). Errors of routine analysis. *Biometrika* **19** 151–164.

TAQQU, M. S. (1979). Convergence of iterated processes of arbitrary Hermite rank. *Z. Wahrsch. Verw. Gebiete* **50** 53–83.

TAQQU, M. S. (1985). A bibliographic guide to self-similar processes and long-range dependence. In *Dependence in Proba-*

bility and Statistics (E. Eberlein and M. S. Taqqu, eds.) 137–165. Birkhäuser, Boston.

TAQQU, M. S. (1988). Self-similar processes. In *Encyclopedia of Statistical Sciences* 8 352–357. Wiley, New York.

TERRIN, N. and TAQQU, M. S. (1991). Convergence in distribution of sums of bivariate Appell polynomials with long-range dependence. *Probab. Theory Related Fields* **90** 57–81.

VERWAAT, W. (1987). Properties of general self-similar processes. *Bull. Inst. Internat. Statist.* **52** 199–216.

WHITTLE, P. (1956). On the variation of yield variance with plot size. *Biometrika* **43** 337–343.

WHITTLE, P. (1962). Topographic correlation, power-law covariance functions, and diffusion. *Biometrika* **49** 304–314.

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

YAJIMA, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16** 791–807.

YAJIMA, Y. (1989). A central limit theorem of Fourier transforms of strongly dependent stationary processes. *J. Time Ser. Anal.* **10** 375–383.

YAJIMA, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Statist.* **19** 158–177.

Comment: Short-Range Consequences of Long-Range Dependence

Arthur P. Dempster and Jing-Shiang Hwang

We welcome Jan Beran's informative sketch of the history of long-range dependence in many fields of applied statistical science, and likewise his review of the results of several decades of work by mathematical statisticians, mainly on asymptotic sampling theory of various robust as well as normality-based efficient estimators.

Our experience has been with applications of the models, most recently in Hwang (1992) and Dempster and Hwang (1992), to simultaneous estimation of employment time series of 51 U.S. states (including DC) given short input time series of $n = 48$ months. Since our data are fixed, we have emphasized issues related to modeling both time series of sampling error, which a priori have no long-range dependence (ignoring biases that cannot be assessed from our data), and underlying true time series, which appear empirically to have long-range dependence with parameter H close to 1 (but not greater than 1 because nonstationarity of unemployment and employment rate series is a priori implausible).

For inference about the true series, we have emphasized Bayesian thinking, and associated computational issues related to likelihoods of our fixed data, always under assumptions of normality, which appear generally to be reasonable in our case study. Although our theoretical approach to statistical inference is very different from that of Beran, we agree with his opening remarks about dangers from behaving as though traditional ways of thinking about level and variability of underlying short-memory stationary time series models continue to hold in the presence of stationary long-memory models. We direct our brief comments to exposing a few basic small n distinctions between inferences appropriate in situations characterized by short-range dependence and those with long-range dependence. We begin by exhibiting artificially generated pseudorandom "time series" that render in graphical form the main points about estimating the mean and variance of fractional Gaussian noise (fGn) data. We have found it convenient to use alternative notation τ^2 and d in place of σ^2 and H, where $d = 2H - 1$ and τ is chosen so that the spectral density

$$f(\lambda) \sim \tau^2 \lambda^{-d}$$

for λ close to zero. On this scale, $-1 < d < 1$ defines the range for fGn, but $0 < d < 1$ is the range of interest for long-range phenomena, with $d = 0$ corresponding to white noise and $d = 1$ marking the upper boundary where the spectral density first becomes nonintegrable at zero frequency. We use the same frequency domain conventions as Beran, namely, that $-\pi < \lambda < \pi$ and that $f(\lambda)$ is scaled so that σ^2 is its average value with respect to uniform measure. Appropriate roles for the alternative scale parameters τ^2 and σ^2 are elaborated below.

Figure 1 displays four series, each of length $n = 64$, simulated from four different fGn models with $d = 0.8$, 0.9, 0.99, 0.999. Part of the reason for the near coincidence of the curves apart from their levels is that all four were generated from innovations based on the same 64 normal pseudorandom values. In addition, however, the similarity implies covariance structures with remarkably similar forecast operators and resid-

Arthur P. Dempster is Professor and Jing-Shiang Hwang is Postdoctoral Fellow at Department of Statistics, Harvard University, Cambridge, Massachusetts 02138.