Exploratory Spectral Analysis of Hydrological Time Series

A.I. McLeod

Department of Statistical & Actuarial Sciences
The University of Western Ontario
London, Ontario N6A 5B7
Canada

K.W. Hipel

Department of Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada N2L 3C5

December 1994

Abstract:

Current methods of estimation of the univariate spectral density are reviewed and some improvements are suggested. It is suggested that spectral analysis may perhaps be best thought of as another exploratory data analysis (EDA) tool which complements rather than competes with the popular ARIMA model building approach. A new diagnostic check for ARMA model adequacy based on the nonparametric spectral density is introduced. Two new algorithms for fast computation of the autoregressive spectral density function are presented. A new style of plotting the spectral density function is suggested. Exploratory spectral analysis of a number of hydrological time series is performed and some interesting periodicities are suggested for further investigation. The application of spectral analysis to determine the possible existence of long memory in riverflow time series is discussed with long riverflow, treering and mud varve series. A comparison of the estimated spectral densities suggests the ARMA models fitted previously to these datasets adequately describe the low frequency component. The software and data used in this paper are available by anonymous ftp from fisher.stats.uwo.ca in the directory pub\mhts.

Key words: AIC-Bayes, autoregressive spectral density estimation, diagnostic checks for ARMA models, exploratory data analysis, fast Fourier transform, Hurst coefficient, long-memory time series, periodogram smoothing, riverflow time series, spectral density plots

Introduction Spectral analysis includes many useful methods based on the Fourier analysis of the time series. The most fundamental is the estimation of the spectral density function. The textbooks by Brillinger (1981) and Priestley (1981) offer definitive accounts of the theory of spectral analysis. An excellent introduction is given by the monograph of Bloomfield (1976). The recent monograph of Percival and Walden (1993) which devoted entirely to contemporary univariate spectral density estimation methods is also an excellent reference with an emphasis on applications. For an state-of-the-art introduction to additional spectral methods see Brillinger and Krishnaiah (1983).

The spectral analysis of time series provides a useful exploratory data analysis tool for examining time series data. Indeed exploratory data analysis (EDA) is often characterized by the four R's: Revelation, Re-expression, Residuals and Resistance. Spectral analysis satisfies the first three requirements and at least partially the fourth. *Revelation:* Spectral analysis can provide an intuitive frequency based description of the time series and indicate interesting features such as long memory, presence of high frequency variation and cyclical behaviour. *Re-expression:* Often the time series are transformed either with a power transformation to stabilize the variance or by differencing or other filtering to remove nonstationary features. *Residuals:* Often the time series data are prewhitened by fitting either trend models or simple parametric models such as autoregressive or autoregressive-moving average models and the residuals from these models are analyzed. *Resistance:* Standard spectral methods do not exhibit resistance in the usual EDA sense which would mean being insensitive to very large outliers. However spectral analysis methods do possess a certain degree of robustness since the normal distribution need not be assumed.

Tukey has pointed out on many occasions that exploratory spectral analysis may reveal features of the data which are missed by using low order ARMA models. For example, Tukey (1978) notes that "the discovery of phenomena is one of the major tasks of science and more phenomena have been discovered by detecting narrowish bumps in the spectrum than have been discovered by fitting ARMA or ARIMA models."

To apply spectral analysis we need a time series, denoted by, $z_t, t = 1, 2, \ldots$ which satisfies two theoretical conditions. The first condition is that the time series is covariance stationary which means that $\mathrm{E}(z_t) = \mu$ and $\mathrm{cov}(z_t, z_{t-k}) = \gamma_k$. Both μ and γ_k must exist and be independent of k. The sequence γ_k is called the sample autocovariance function (TACVF). The second requirement, known as ergodicity, is that the sample estimators of μ and γ_k must converge as the length of the sampled series increases. Given n successive observations, $z_t, t = 1, \ldots, n$, of a time series the sample estimator for μ is simply the sample mean, $\bar{z} = \sum z_t/n$ and for the TACVF, the sample autocovariance function (SACVF) is given as

$$c_k = \frac{1}{n} \sum_{\ell=k+1}^{n} (z_t - \bar{z})(z_{t-k} - \bar{z}) \quad \text{for } k \geq 0,$$

and for $k < 0$, $c_k = c_{-k}$. Necessary and sufficient conditions for a general covariance stationary process to be ergodic for the mean and TACVF are given by Hannan (1970, Theorem 6, p.210). One necessary condition for this is that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \gamma_k = 0. \tag{1}$$

A simple example of a stationary non-ergodic time series is the symmetrically correlated time series which has TACVF given by

$$\begin{aligned} \gamma_k &= \gamma_0, \quad \text{if k} = 0, \\ &= \gamma, \quad \text{if k} \neq 0, \end{aligned} \tag{2}$$

where $\gamma < \gamma_0$. Ergodicity is a theoretical condition which is not normally possible to verify in practice.

In the next section, the basic properties of the spectral density function of a time series are reviewed. In §3 of estimating the spectral density function are discussed. In §4, it is shown that the spectral density function provides the best tool to define what is meant by long memory in time series. The application of spectral methods to annual riverflow and other hydrological data is discussed. A new style of plotting the spectral density function is recommended.

Spectral analysis is a component in a PC time series package developed by McLeod and Hipel. A student version of this package is available via anonymous ftp as mentioned in the Abstract.

2 Spectral Analysis Primer

2.1 Spectral Density Function

Spectral analysis can be regarded as the development of a Fourier analysis for stationary time series. Just as in classical Fourier analysis a real function $z(t)$ is represented by a Fourier series, in spectral analysis the autocovariance function of a stationary time series has a frequency representation in terms of a Fourier transform. This representation was first given by Herglotz (1911) who showed that any positive-definite function, such as the autocovariance function, γ_k, of a stationary time series can be represented as

$$\gamma_k = \int\limits_{(-\pi,\pi]} e^{i\omega k} dP(\omega), \tag{3}$$

where $P(\omega)$ is the spectral distribution function. In the case where a spectral density function exists, we have $dP(\omega) = p(\omega)d\omega$ and Herglotz's equation, eq. (3), can be written

$$\gamma_k = 2 \int_0^{\pi} p(\omega) \cos(\omega k) d\omega. \tag{4}$$

The function $p(\omega), \quad -\pi \leq \omega \leq \pi$ is called the spectral density function and shares many properties of the probability density function. In addition, note that $p(\omega)$ is symmetric, $p(\omega) = p(-\omega)$. For mathematical convenience the units of ω are in radians per unit time, which is known as angular or circular frequency. In practice, however, it is more convenient to work in units of cycles per unit time, which is related to ω by the equation $\omega = 2\pi f$, where f is now in cycles per unit time.

2.2 Periodogram

A natural estimator of the spectral density function given n observations $z_1, \ldots, z_n$ from a covariance stationary time series, is given by the periodogram,

$$I(f_j) = \frac{1}{n}\Big|\sum_{t=1}^{n} z_t e^{-2\pi f_j (t-1)}\Big|^2, \tag{5}$$

where $f_j = j/n, \quad j = [-(n-1)/2], \ldots, 0, \ldots, [n/2]$, where $[\bullet]$ denotes the integer part function. Since $I(f_j) = I(-f_j)$ the periodogram is symmetric about 0, and so when the periodogram or spectral density is plotted we only plot the part where $f_j > 0$. When $f_j = 0$, $I(0) = n\bar{z}^2$, where $\bar{z} = \sum_{t=1}^{n} z_t/n$. This component, $I(0)$, is usually very large due to a non-zero mean and so is ignored in the periodogram and spectral plots.

In the case where the spectral density function exists, the expected value of $I(f_j)$ can be shown to be approximately equal to $p(f_j)$, where $p(f)$ is the spectral density function. In fact, in large samples, $I(f_j)$ for $j = 1, \ldots, [(n-1)/2]$ are statistically independent and exponentially distributed with mean $p(f_j)$.

Schuster (1898) developed the periodogram for searching for periodicities in time series. Sometimes $I(f_j)$ is plotted against its period $1/f_j$ but this is not very satisfactory because, since $I(f_j)$ is calculated at equi-spaced frequencies, the low-frequency part is too spread out. A new style of plotting is suggested in §4 which shows the periodicities on the x-axis.

2.3 Frequency Interpretation

Another important property for the interpretation of the periodogram is that it can be shown that $I(f_j)$ is proportional to the square of the multiple correlation between the observed data sequence $z_1, \ldots, z_n$ and a sinusoid having frequency f_j. Specifically, consider the regression,

$$z_t = A_0 + A_j \cos(2\pi f_j) + B_j \sin(2\pi f_j) + e_t, \tag{6}$$

where e_t is the error term. Then the least-squares estimates of A_0, A_j and B_j are given by

$$A_0 = \frac{1}{n}\sum_{t=1}^{n} z_t$$

$$A_j = \frac{2}{n}\sum_{t=1}^{n} z_t \cos(2\pi f_j)$$

$$B_j = \frac{2}{n}\sum_{t=1}^{n} z_t \sin(2\pi f_j).$$

The multiple correlation coefficient, R_j^2, can be shown to be given by

$$R_j^2 = A_j^2 + B_j^2.$$

In terms of the periodogram we have then

$$I(f_j) = \frac{n}{2} R_j^2.$$

Thus $p(f)$ can be interpreted as measuring the strength of a random sinusoidal component having a period $1/f$ in the data sequence. Time series exhibiting cycles or oscillatory behaviour will have a peak in the spectral density function at the frequency which corresponds to the cycle period. For example, if there is a ten-year cyclical component, then there will be a peak in the spectral density function at $f = 0.1$. The sharpness of the peak depends on how closely the period $1/f$ appears in the data sequence, and the relative size of the peak depends on the relative amplitude of the cycle in the time series. The units of f_j are cycles per unit time. The period corresponding to f_j is $T_j = 1/f_j$.

2.4 ANOVA Decomposition

Taking $k = 0$ in eq. (4) we obtain for $\mathrm{var}(z_t) = \gamma_0$,

$$\mathrm{var}(z_t) = 2 \int_0^{2\pi} p(\omega) d\omega. \tag{7}$$

It can be shown that spectral analysis provides an anova like decomposition of a time series into its frequency components and this fact is illustrated in eq. (7). The sample analogue of eq. (7) is

$$\sum_{t=1}^{n} (z_t - \bar{z})^2 = 2 \sum_{j=1}^{[(n-1)/2]} I(f_j) + I(f_{[n/2]}),$$

where the last term $I(f_{[n/2]})$ is omitted when n is an odd.

The spectral density function can be derived by taking the inverse Fourier transformation to eq. (4) which yields,

$$p(f) = \sum_{k=-\infty}^{\infty} \gamma_k e^{-2\pi f k}, \quad |f| \leq 0.5, \tag{8}$$

The sample analogue of this formula also holds, viz.

$$I(f) = \sum_{k=-(n-1)}^{n-1} c_k e^{-2\pi f k},$$

where c_k denotes the sample autocovariance function given by

$$c_k = \frac{1}{n} \sum_{\ell=k+1}^{n} (z_t - \bar{z})(z_{t-k} - \bar{z}) \quad \text{for } k \geq 0,$$

and for $k < 0$, $c_k = c_{-k}$.

2.5 Aliasing

There is an upper limit to the highest frequency that can be observed in the time series. This upper limit, which is 0.5 cycles per unit time or π radians per unit time, is called the Nquist frequency. This upper limit arises because of the discrete time nature of our time series. There is no such upper limit in the continuous time case. To see why aliasing occurs, let $\tilde{f}$ denote any frequency in the interval $[0, 0.5]$ and let $\ddot{f} = \tilde{f} + 0.5$. Then it is easily shown for all integer t that $\cos(\ddot{f}) = \cos(\tilde{f})$ and $\sin(\ddot{f}) = \sin(\tilde{f})$. The frequencies $\tilde{f}$ and $\ddot{f}$ are said to be *aliases.* Aliased frequencies, such as $\ddot{f}$, are observationally indistinguishable from frequencies in the range $[0, 0.5]$.

2.6 Spectral Density and ARMA Model

The ARMA(p, q) model, may be written in operator notation as,

$$\phi(B)z_t = \theta(B)a_t,$$

where, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p, \theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q, a_t$ is white noise with variance σ_a^2 and B is the backshift operator on t, is said to be not redundant if and only if $\phi(B) = 0$ and $\theta(B) = 0$ have no common roots. Due to stationarity and invertibility, all roots of the equation $\phi(B)\theta(B) = 0$ are assumed to be outside the unit circle. When $q = 0$ this model is referred to as the autoregression of order p or AR(p). Fitting high order autoregressive models to estimate the spectral density will be discussed in §3.

The theoretical spectral density for time series generated by the ARMA(p, q) model is given by

$$p(f) = \sigma_a^2 |\psi(e^{2\pi i f})|^2,$$

where

$$\psi(B) = \frac{\theta(B)}{\phi(B)}.$$

3 Spectral Density Estimation Methods

3.1 Periodogram Smoothing

This is the frequency domain approach because in this approach one works with the Fourier transform of the data sequence rather than the original data sequence. It is the most natural approach to estimating $p(f)$ although the other approach by high-order autoregressive modelling will often produce a more accurate overall estimate in the sense of the integrated mean-square error criterion. On the other hand, periodogram smoothing is better suited to revealing bumps and narrow peaks in the spectrum which is often the principal goal of exploratory spectral analysis in the first place.

The *Discrete Fourier Transform* (DFT) of the sequence $z_1, \ldots, z_n$, is defined as

$$Z_j = \sum_{t=1}^{n} z_t e^{-2\pi i t f_j}, \quad j = -[(n-1)/2], \ldots, 0, \ldots, [n/2],$$

where $[\bullet]$ denotes the integer part and $f_j = j/n$. The frequencies f_j are referred to as the Fourier frequencies. The functions $\cos(2\pi f_j t)$ and $\sin(2\pi f_j t)$ are orthogonal with respect to the usual inner product when evaluated at the Fourier frequencies.

The periodogram is given by

$$I(f_j) = \frac{1}{n}|Z_j|^2.$$

It follows from the orthogonality mentioned above that

$$\sum_{j=-[(n-1)/2]}^{[n/2]} I(f_j) = \sum_{t=1}^{n} z_t^2.$$

Thus $I(f_j)$ can be interpreted as analysis of variance of the data. $I(f_j)$ shows the amount of variation due to each frequency component. Since the periodogram is symmetric about zero, only positive frequencies need be considered.

The periodogram smoothing approach to the estimation of $p(f)$ is based on the following two large-sample results:

$$< I(f_j) > \approx p(f_j),$$

and

$$\begin{aligned} \mathrm{cov}(I(f_j), I(f_k)) &\approx p^2(f_j), \quad \text{when} j = k, \\ &\approx 0, \quad \text{when} j \neq k. \end{aligned}$$

From the above two equations we see that although $I(f_j)$ is an unbiased estimator of $p(f_j)$, it is not consistent. If $p(f)$ is assumed to be a smooth function of f, an estimator with smaller mean-square error can be obtained by averaging values of the periodogram.

The periodogram smoother may be written

$$\hat{p}(f_j) = \sum_{i=-q}^{i=q} w_i I(f_{j+i}),$$

where q is the half-length of the smoother and the weights w_i satisfy the following conditions:

(i) $w_i \geq 0$,

(ii) $w_i = w_{-i}$,
(iii) $\sum_{i=-q}^{q} w_i = 1$.

The *Rectangular* window uses $w_i = (2q+1)^{-1}, \quad i = -q, \ldots, 0, \ldots, q$. As shown by Hamming (1977, §5.8), the *Modified Rectangular*,

$$w_i = 1/(2q), \quad \text{if } |\mathrm{i}| < \mathrm{q},$$

$$= 0.5/(2q), \quad \text{if } |\mathrm{i}| = \mathrm{q},$$

is generally even better. These smoothers can be iterated several times to produce smoothers which yield very smooth and generally more desirable estimates of the spectral density. These iterated smoothers correspond to a single non-iterated smoother whose weights, w_i, form a symmetric bell-shaped like function with peak at w_0. It is sometimes of interest to plot these weights, w_i. Since the weights are symmetric, it is sufficient to plot the weights for $i = 0, ..., q$. When the weights are plotted on the same graph as the spectral density, it is convenient to rescale the weights since it is just the shape and width of the weight function that we are primarily interested in.

The equivalent degrees of freedom, denoted by edf, is given by

$$\mathrm{edf} = \frac{2}{\sum\limits_{i=-q}^{q} w_i^2}.$$

For the Rectangular smoother edf $= 4q + 2$. An approximate 95% confidence interval for $p(f_j)$ is

$$\left(\frac{\mathrm{edf}\ \hat{p}(f_j)}{\chi^2_{0.975}(\mathrm{edf})} \frac{\mathrm{edf}\ \hat{p}(f_j)}{\chi^2_{0.025}(\mathrm{edf})} \right),$$

where $\chi^2_{0.025}(\mathrm{edf})$ and $\chi^2_{0.975}(\mathrm{edf})$ denote the 2.5% and the 97.5% points of the χ^2-distribution on edf degrees of freedom. The percentage point for fractional degrees of freedom are obtained by linear interpolation. See Bloomfield (1976, §8.5) and Brillinger (1981, §5.7) for proofs.

These confidence intervals may be used to provide an informal ARMA diagnostic check. If the model is adequate then its spectral density function should lie mostly within limits. It should be noted that since these limits are not simultaneous, some allowance should be made for small departures of the fitted ARMA spectral density.

3.2 Autoregressive Spectral Density Estimation

Here $p(f)$ is estimated by fitting an autoregression,

$$\phi(B)(z_t - \mu) = a_t,$$

where $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and a_t is white noise with variance σ_a^2. Then $p(f)$ can be determined from the equation

$$p(f) = \frac{\sigma_a^2}{|\phi(e^{2\pi i f})|^2}.$$

This is a popular method since it seems to produce an estimate of the spectral density which often has lower integrated mean square error than the periodogram smoothing approach.

Akaike (1979) addressed the problem of choosing the model order p by suggesting that the weighted average of all spectral functions of all autoregressions of orders $p = 0, 1, \ldots, K$, where K is some upper limit, should be used. Each spectral density in the average is weighted according to the quasi-likelihood or Bayes posterior given by

$$e^{-\frac{1}{2}\mathrm{AIC}}.$$

This weighted average estimate is itself equivalent to a special AR(K) model with coefficients determined by the weighting. This model is referred to as the AR-AIC-Bayes filter. Typically, the AR-AIC-Bayes filter produces a spectral density which shows more features similar to the estimate produced by periodogram smoothing than simply fitting a fixed order AR model.

As recommended by Percival and Walden (1993, p.414–417) the AR model parameters are estimated using the Burg algorithm rather than the standard Yule-Walker estimates. The reason for this is that the standard Yule-Walker estimates are now known to be severely biased in some situations.

3.3 Algorithms For Fast Autoregressive Spectral Density Computation

The spectral density function is normally interpreted by plotting it over the interval $(0, \frac{1}{2})$. This requires its evaluation at a large number of values. We provide two new faster algorithms for this computation. Fast AR spectrum evaluation is of interest when p is large. In applications, large values of p are not uncommon particularly with the AR-AIC-Bayes filter and also in other applications such as in Percival and Walden (1993, p.522). The two algorithms given can easily be extended to the evaluation of the spectral density function of ARMA models.

The direct method evaluates the spectral density function, $f(\lambda)$ given the parameters $\phi_1, \ldots, \phi_p$ and σ_a^2 directly and in general requires $O(np)$ floating point operations (flops) in addition to $O(np)$ complex exponential evaluations.

This method may be improved by avoiding the calculation of the large number of complex exponentials. First, note that the evaluation of the complex exponential is equivalent to one sine and cosine function evaluation. So if $f(\lambda)$ is evaluated at equally spaced values throughout $[0, \frac{1}{2}]$ then the necessary trigonometric functions may be evaluated recursively using the sum of angles formulae for sine and cosine functions. This technique has been previously used in the evaluation of the discrete Fourier transform at a particular frequency, see Robinson (1967, pp. 64–65). Given the autoregressive parameters $\phi_1, \ldots, \phi_p$ and innovation variance σ_a^2, the recursive algorithm for spectral density computation may be summarized as follows:
Step 1: Initializations.
Select M the number of equi-spaced frequencies. Typically, $M \leftarrow 256$ is adequate. Set $k \leftarrow 1$, $s_\beta \leftarrow 0$, $c_\beta \leftarrow 1$, $s_m \leftarrow \sin(\pi/(m-1))$ and $c_m \leftarrow \cos(\pi/(m-1))$.
Step 2: Compute $p(k/2(m-1))$
(a) Set $j \leftarrow 1$, $s_\alpha \leftarrow s_\beta$, $c_\alpha \leftarrow c_\beta$, $s \leftarrow s_\alpha$ $c \leftarrow c_\alpha$ $A \leftarrow 0$ and $B \leftarrow 1$.

(b) $A \leftarrow A - s\phi_j$ and $B \leftarrow A - c\phi_j$
(c) $j \leftarrow j + 1$
(d) If $j > p$ then $p(k/2(m-1)) \leftarrow \sigma_a^2/(A^2 + B^2)$ and go to *Step 3.* Otherwise if $j \leq p$ then $t \leftarrow c$, $c \leftarrow c_\alpha t - s_\alpha s$, $s \leftarrow s_\alpha t + c_\alpha s$, and return to *(b)*
Step 3: Increment for next k
Set $k \leftarrow k + 1$. Terminate if $k > m$. Otherwise if $k \leq m$ then $t \leftarrow c_\beta$, $c_\beta \leftarrow c_m t - s_m s_\beta$, $s_\beta \leftarrow s_m t + c_m s_\beta$, and return to *Step 2*

In many situations an even faster method is based on the fast Fourier transform (FFT). Given a sequence $\{\psi_k\}, \quad k = 0, \ldots, M$ the discrete Fourier transform is a sequence $\{\Psi_\ell\}, \quad \ell = 0, \ldots, M$ given by

$$\Psi_\ell = \sum_{k=0}^{M} \psi_\ell e^{\frac{i2\pi k\ell}{M}}.$$

When $M = 2^q$ for some positive integer q, the discrete Fourier transform may be evaluated using an algorithm known as the FFT which requires only $O(M \log_2(M))$ flops when M is a power of 2. The FFT may be applied to evaluate $f(\lambda)$ by setting $\psi_0 = -1$, $\psi_k = \phi_k, \quad k = 1, \ldots, p$ and $\psi_k = 0, \quad k > p$. To evaluate $f(\lambda)$ at $N = 2^r$ equi-spaced points on $[0, \frac{1}{2}]$, set $M = 2N$ and apply the FFT. Then $f(\ell/(2N)) = \sigma_a^2/(\Psi(\ell)\bar{\Psi}(\ell)), \quad \ell = 0, \ldots, N$, where $\bar{\Psi}(\ell)$ denotes the complex conjugate of $\Psi(\ell)$. In many situations, $N = 256$ is adequate.

The FFT method is the most expedient and convenient method to use with higher level programming languages such as *Mathematica* in which it is best to vectorize the calculations. In such languages the FFT is provided as a system level function. For example, in *Mathematica* (Wolfram, 1991) the following function evaluates the spectral density with $\sigma_a^2 = 1$ at 256 equi-spaced points:

```
Arspec[phi_] := 1./(512*Re[#1*Conjugate[#1]] & )[
 Take[Fourier[Join[{1}, -phi, Table[0, {511 - Length[phi]}]]], 256]]
```

In the above *Mathematica* function the parameter `phi` is the vector $(\phi_1, \ldots, \phi_p)$.

The recursive algorithm may be preferable to the FFT method in certain situations where the number of frequencies at which the spectral density is to be evaluated is not a power of 2. This is the case, for example, when spectral methods are used to estimate the parameters of ARMA models (Hannan, §VI.5).

The three algorithms outlined above were programmed in Fortran. The FFT method used the algorithm of Monro (1976). Timings for $N = 256$ equi-spaced evaluations of $f(\lambda)$ on $[0, \frac{1}{2}]$ on a 286/7 PC are shown in Table I. For faster computers, the times will be reduced. Counting the number of flops the recursive method requires $O(Np)$ while the FFT method requires $O(N \log_2(N))$. This suggests as a rough estimate, for $p > 8$ it may be expected that the FFT will be faster. In fact, as shown in Table I, the FFT method is faster for $p \geq 10$.

TABLE I

TIMINGS IN SECONDS FOR AR(p) SPECTRUM.

method	$p=1$	$p=5$	$p=10$	$p=30$	$p=60$	$p=120$
direct	0.54	2.14	4.01	11.86	23.62	47.01
recursive	0.33	0.88	1.59	4.34	8.46	16.81
FFT	1.16	1.21	1.26	1.27	1.26	1.32

4 The Hurst Phenomenon from the Spectral Viewpoint

Long memory time series were first suggested by Barnard (1956) as a possible cause of the Hurst effect. However, Hipel and McLeod (1978) have demonstrated that the Hurst effect can also be due to the finite, albeit long, length of the geophysical time series studied by Hurst.

Barnard (1956) pointed out that the autocovariance function of such a long-memory time series is not summable and several authors, for example Hipel and McLeod (1978), have used this as the definition of long memory. In fact the spectral viewpoint provides a much better definition and understanding of long memory in time series. As first suggested by Hosking (1981), we will say that a covariance-stationary time series exhibits long memory if and only if

$$\lim_{f\to 0} p(f) = +\infty. \tag{9}$$

From eq. (8), it follows that the autocovariance function is not summable. However for the same reason it also follows that any time series such that

$$\lim_{f\to f_0} p(f) = \infty$$

also has a TACVF which is not summable. Thus summability of the TACVF is a necessary but not a sufficient condition for long memory. Note however that the divergence of $\sum \gamma_k$ is equivalent to (9).

Thus the symmetrically correlated process of eq. (2), is a long-memory process which is stationary but not ergodic. The ARIMA(p,d,q) models of Box and Jenkins (1976) when $d \geq 1$ are also long memory models in the sense that the sample spectral density function tends to get indefinitely large at very low frequencies as the sample size increases although these models are non-stationary as well. The first stationary model suggested was Mandelbrot's fractional Gaussian noise (FGN) model described by Mandelbrot and Van Ness (1968) and advocated for hydrological time series by Mandelbrot and Wallis (1969). A more flexible approach to long-memory models was initiated by Granger and Joyeux (1980) and Hosking (1981) who suggested what is now referred to as the fractional ARMA model. This provides a comprehensive family of stationary and ergodic models which generalize the usual ARMA model. Beran (1992) gives a recent review of long-memory time series models and several other researchers have enthusiastically recommended long-memory models for various types of geophysical data.

4.1 Exploratory Spectral Analysis of Hydrological Time Series

Hipel and McLeod (1978) previously analyzed some of the longest available stationary hydrological time series to see if there was a long-memory effect which could not be accounted for by short-memory ARMA models. Some of the data from that study is described in Table II. These datasets include the longest available annual unregulated riverflows as well as long time series on water level, mud varves and treerings.

If the long-memory hypothesis is correct then one would expect to see high power in the lowest frequencies and the longer the series length, the more pronounced this effect would be expected to become. It is of interest to compare the two methods of exploratory spectral estimation described in §3. With periodogram smoothing it is expected that the estimates of the spectral density will be more variable but perhaps better able to resolve the peak at the zero frequency. In the autoregressive approach we will set K quite large so that a good approximation is obtained for any possible long memory effect. The spectral plots are shown in Figures 1–20. Panel (a) in each Figure refers to the periodogram smooth and Panel (b) to the AR-AIC-Bayes Filter.

The plotting style for the spectral density shows the periods instead of frequencies on the x-axis. But the scaling on the x-axis is equi-spaced on the frequency scale. This scale desirable because every interval on the frequency scale corresponds to the same number of periodogram estimates. All that changes is just the particular labels used on the axis. These labels are more immediately informative since in most applications most practioners think in terms of the period not of the frequency. The y-axis shows the log to the base 2 of the spectral density function estimate. Cleveland (1994, p.122) pointed out that it is often preferable in statistical graphics to use a log to the base 2 transformation for one of the axes since it is much easier to interpret. When log to the base 2 is used a change of one unit means a doubling of the value in the orginal domain. Similarly, an increase of 0.5 on a log to the base 2 scale, implies a 40% increase. On the other hand fractional powers of ten are much harder to interpret.

In the bottom left-hand corner of each plot of the periodogram smooth the shape of the smoother is inset. Only the right half of the bell shaped smoother is shown. The ordinate scale is just chosen for convenience and is not important. The width of the bell shows the number of frequencies involved and its shape indicates the relative weights, w_i, in the smooth.

In general the two spectral estimation methods, periodogram smoothing and the AR-AIC-Bayes Filter provide quite similar estimates of the spectral density. In general, both methods show that their is high power near the zero frequency which is consistent with the long-memory hypothesis. Some interesting differences between these methods occur for MSTOUIS, NEUMUNAS, BIGCONE, NARAMATA, and NAVAJO. In each of these cases, the periodogram smooth suggests a peak close to zero but not exactly at zero whereas the AR-AIC-Bayes Filter has the peak at zero. In some cases the periodogram smooth suggests interesting possible periodicities in the time series such as in the case of the GOTA or the NINEMILE.

TABLE II

HYDROLOGICAL TIME SERIES.

Code Name	Description	n
danube	Danube River, Orshava, Romania, 1837–1957, annual flow	120
gota	Gota River, Sjotorp-Vanersburg, Sweden, 1807–1957, annual flow	150
mstouis	Mississippi River, St. Louis, Missouri, 1861–1957, annual flow	96
neumunas	Neumunas River, Smalininkai, Russia, 1811–1943, annual flow	132
ogden	St. Lawrence River, Ogdensburg, N.Y., 1860–1957, annual flow	96
rhine	Rhine River, Basel, Switzerland, 1807–1957, annual flow	150
minimum	Nile River, Rhoda, Egypt, 622–1469, annual minimum level	848
espanola	Espanola, Ontario, -471– -820, mud varve	350
bigcone	Big cone spruce, Southern California, 1458–1966, treering width	509
bryce	Ponderosa pine, Bryce, Utah, 1340–1964, treering width	625
dell	Limber pine, Dell, Montana, 1311–1965, treering width	655
eaglecol	Douglas fir, Eagle Colorado, 1107–1964, treering width	858
exshaw	Douglas fir, Exshaw, Alberta, 1460–1965, treering width	506
lakeview	Ponderosa pine, Lakeview, Oregon, 1421–1964, treering width	544
naramata	Ponderosa pine, Naramata, B.C., 1451–1965, treering width	515
navajo	Douglas fir, Belatakin, Arizona, 1263–1962, treering width	700
ninemile	Douglas fir, Nine Mile Canyon, Utah, 1194–1964, treering width	771
snake	Douglas fir, Snake River Basin, 1281–1950, treering width	669
tioga	Jeffrey pine, Tioga Pass, California, 1304–1964, treering width	661
whitemtn	Bristlecone pine, California, 800–1963, treering width	1164

4.2 Comparison with the ARMA Models

For comparison, Panel (c) in each Figure shows the fitted spectral density for the ARMA models estimated by Hipel and McLeod (1978). The specific ARMA model is indicated on the plot. These ARMA models were fitted using the approximate likelihood algorithm of McLeod (1977). In many cases, such as for MINIMUM and ESPANOLA, the best fitting ARMA has even higher spectral mass near zero than either of the spectral methods. Only in the case of the NINEMILE does the spectral density function from the fitted ARMA model seem to be drastically out of line. In this case, after refitting, it was found that an ARMA(9,1) model with $\phi_2 = \ldots = \phi_8 = 0$ provided a better fit. The new diagnostic plot is shown in Figure 21.

The period of about ten years for the NINEMILE series is very close to the annual sunspot numbers. Using a pre-whitening analysis, it is found that there is a rather large negative correlation, -0.121 (± 0.06 sd) between treering width and sunspot activity in the previous year. It is known that high sunspot activity is strongly related to the incidence of melanoma in human males in Conneticut two years later (Andrews and

Hertzberg, 1984, p.199), so it is plausibe that high sunspot activity may be damaging to tree growth at this particular site.

[Figures 1-21 About Here]

4.3 Conclusions

Exploratory spectral analysis of our datasets has perhaps uncovered a new phenomenon in the case of NINEMILE. Some of the other data sets also show suggestive periodicities as well.

Exploratory spectral analysis shows that for most of the datasets analyzed there is indeed a large spectral mass near zero. However comparison of the estimated spectral densities with those of the best fitting ARMA model indicates that this phenomenon can easily be modelled satisfactorily by a low order ARMA model. In situations such as this where there are two possible models, a so called long memory model and an ARMA model, one could appeal to the principle of parsimony or to Occam's razor. The principle of parsimony advocates choosing the model with the fewest number of parameters. On the other hand Occam's razor indicates the conceptually simplest model should be chosen. For most practicing hydrologists, Occam's razor would probably suggest that the long-memory model should be discarded since there is little a priori justification. From the standpoint of the principle of parsimony there seems very little to gain by switching to the more complex long-memory models since low order ARMA models seem to do quite well.

References

Akaike, H. (1979), A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika* 66(2), 237–242.

Andrews, D.F. and Herzberg, A.M. (1984), *Data: A Collection of Problems from Many Fields for the Student and Research Worker*, New York: Springer-Verlag.

Barnard, G.A. (1956), Discussion on Methods of using long term storage in reservoirs, *Proc. Inst. Civil Eng.* 1, 552–553.

Beran, J. (1992), Statistical methods for data with long-range dependence, *Statistical Science* 7(4) 404–427.

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

Box, G.E.P. and Jenkins, G.M. (1975) *Time Series Analysis: Forecasting and Control*, (2nd edn). San Francisco: Holden-Day.

Brillinger, D.R. (1981), *Time Series: Data Analysis and Theory*, Expanded Edition, San Francisco: Holden-Day.

Brillinger, D.R. and Krishnaiah, P.R. (1983), *Time Series in the Frequency Domain*, Amsterdam: North Holland.

Cleveland, W.S. (1994), *The Elements of Graphing Data*, Revised Edition, Summit: Hobart Press.

Granger, C.W.J. and Joyeux, R. (1980), An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1(1), 15–29.

Hamming, R.W. (1977), *Digital Filters*, New Jersey: Prentice-Hall.

Hannan, E.J. (1970), *Multiple Time Series*, New York: Wiley.

Herglotz, G. (1911), Über potenzreihen mit positivem reellem teil im einheitskreis. *Sitzgxber. Sachs Akad. Wiss.* 63, 501–511.

Hipel, K.W. & McLeod, A.I. (1994), *Time Series Modelling of Water Resources and Environmental Systems*, Amsterdam: Elsevier.

Hosking, J.R.M. (1981), Fractional differencing, *Biometrika*, 68(1), 165–176.

Mandelbrot, B.B. and Van Ness, J.W. (1968), Fractional Brownian motion, fractional noises and applications. *SIAM Review* 10(4), 422–437.

Mandelbrot, B.B.and Wallis, J.R. (1969), Some long-run properties of geophysical records. *Water Resources Research* 5(2), 321–340.

McLeod, A.I. (1977), Improved Box-Jenkins estimators. *Biometrika* 64(3), pp.531–534.

Monro, D.M. (1976) Algorithm AS 97. Real discrete fast Fourier transform. *Applied Statistics* 25(2), 173–180.

Percival, D.B. and Walden, A.T. (1993) *Spectral Analysis for Physical Applications*, Cambridge University Press.

Priestley, M.B. (1981), *Spectral Analysis and Time Series*, New York: Academic Press.

Schuster, A. (1898), On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terr. Magn.* 3, 13–41.

Robinson, E.A. (1967). *Multichannel Time Series Analysis with Digital Computer Programs*, San Francisco: Holden-Day.

Tukey, J. W. (1978), Comment on a Paper by H.L. Gray et al. in *The Collected Works of John W. Tukey, Volume II, Time Series: 1965–1984*. Ed. D.R. Brillinger. Monterey: Wadsworth.

Wolfram, S. (1991) *Mathematica: A System for Doing Mathematics by Computer*, (2nd edn). Redwood City: Addison-Wesley.