Stochastic Hydrol. Hydraul. 6 (1992) 270–288

Stochastic Hydrology and Hydraulics

Estimation of periodicities in hydrologic data

A. R. Rao and **G. D. Jeong**

School of Civil Engineering, Purdue University, W. Lafayette, IN 47907, USA

Fi-John Chang

Department of Agricultural Engineering, National Taiwan University, Taipei, Taiwan, 100764 R.O.C.

Abstract: Periodicites in hydrologic data are frequently estimated and studied. In some cases the periodic components are subtracted from the data to obtain the stochastic components. In other cases the physical reasons for the occurrence of these periodicities are investigated. Apart from the annual cycle in the hydrologic data, periods corresponding to the 11 year sunspot cycle, the Hale cycle and others have been detected.

The conclusions from most of these studies depend on the reliability and robustness of the methods used to detect these periodicities. Several spectral analysis methods have been proposed to investigate periodicities in time series data. Several of these have been compared to each other. The methods by Siddiqui and Wang and by Damsleth and Spjotvoll, which are stepwise procedures of spectrum estimation, have not been evaluated.

Two of the methods of spectral analysis proposed by Siddiqui and Wang and by Damsleth and Spjotvoll are investigated in this study by using generated and observed data. Siddiqui and Wang's method is found to be superior to the Damsleth and Spjotvoll's method.

Key words: Periodicities, hydraulic cycle, spectral methods

1 Introduction

Many time series in hydrology are periodic. Many others have been suggested to be potentially periodic. One of the principal approaches to investigate such time series is to estimate the spectrum of the series. A time series can theoretically be completely described by harmonic functions although such a description would include too many coefficients. Hence, only the significant harmonic components and frequencies that are present in a time series are estimated. Stepwise procedures for estimation of the dominant frequencies and corresponding harmonic components has been proposed by several investigators. The purpose of the present study is to compare the relative merits of two of these stepwise procedures which have been developed by Siddiqui and Wang (1984) and by Damsleth and Spjotvoll (1982). The fast Fourier transform (FFT) is a computer algorithm for calculating the discrete Fourier transforms which is frequently used to calculate the sample spectrum. The FFT is also used in the present study.

A review of the literature revealed that the performance of the procedures of the stepwise procedures by Siddiqui and Wang (1984) and by Damsleth and Spjotvoll (1982) have not

been compared. Also there are no other stepwise spectral estimation procedures proposed since 1984.

Estimation of spectra of time series is an active area of investigation. Of the more important developments in spectrum estimation are the maximum entropy Spectral analysis (Burg, 1967; Marple, 1980) Autoregressive-Moving Average Spectrum (Cadzow, 1981, Friedlander, 1982-1, 1982-2, Kay 1980, Kay and Marple, 1981), estimation of Harmonics from a covariance function by Pisarenko (1973). There is a considerable body of literature in which these methods are discussed and compared. For example Beamish and Priestley (1981) compared the autoregressive and window spectral methods. Rao et al. (1984) compared the maximum entropy and maximum likelihood spectral estimation methods. Durgunoglu and Rao (1985) compared several methods of spectral estimation although none of which belonged to the stepwise method discussed herein. As the performance of stepwise procedures to estimate the spectrum have not been investigated, they were investigated and the results are reported herein.

After the significant harmonics are determined, they are used to either remove the periodicity in the time series or to incorporate them in models. If the periodic component is eliminated from the time series the residual stochastic component is modelled. The efficiency of modeling the stochastic component depends on the estimation of periodicity. A robust algorithm to estimate the periodic component would thus be very useful in stochastic hydrology.

The procedure adopted in the comparison of these methods is as follows. The methods are used to estimate the significant harmonic components. These harmonic components are used to regenerate the series. The residuals between the observed and regenerated series and the goodness of the regeneration of the original series are used to select the better method. Both synthetic data whose periodic characteristics are known and observed data are used in the study.

This paper is organized in five sections. The methods and estimation procedures are given in the second section. The data used in this study are presented in the third section. Results are discussed in the fourth section. A set of conclusions is given in the last section.

2 The methods and estimation procedures

For a time series x_t (t = 1, ..., n) showing periodic structure, the following model is usually considered

$$x_t = a_0 + \sum_{j=1}^{m} \left(a_j \cos \frac{2\pi jt}{n} + b_j \sin \frac{2\pi jt}{n} \right) \tag{1}$$

Where

$$a_0 = \frac{1}{n} \sum_{t=1}^{n} x_t$$

$$a_j = \frac{2}{n} \sum_{t=1}^{n} x_t \cos \frac{2\pi jt}{n}$$

$$b_j = \frac{2}{n} \sum_{t=1}^{n} x_t \sin \frac{2\pi jt}{n}$$

with j = 1, 2, ..., m. The squared amplitude at frequency j/n is $p_j^2 = a_j^2 + b_j^2$.

The time series can be completely described when n is an odd number and m = (n-1)/2. However, in natural time series, only a few harmonic cycles may significantly contribute to

the structure of the time series. In other words, a few harmonic components might adequately fit the time series.

A well known decision rule for determining the significant harmonics was formulated by Fisher (1929). This rule declares a cycle k to be different from zero if

$$\frac{p_k^2}{\sum_{j=1}^{m} p_j^2} > c \tag{2}$$

where c is a constant determined so that the maximum probability of claiming a spurious cycle to have amplitude different from 0 is less than some preassigned significance level ϵ. However, if a time series has more than one harmonic component, the denominator in equation (2) tends to be large. Kashyap and Rao (1976) and Bolviken (1982) have shown that Fisher's method has very low power in such cases. To improve the sensitivity of the test in such circumstances, Bolviken (1982) modified Fisher's test.

In Bolviken's modification, $\hat{p}_1 \leq \hat{p}_2 \leq ..... \leq \hat{p}_m$ are an ordered set of $\hat{p}_j$'s. Let $\hat{p}_j$ be the expected value of p_j. In Bolviken's method it is claimed that p_k is significantly different from zero if

$$\frac{\hat{p}_k^2}{\sum_{j=1}^{m-a} \hat{p}_j^2} > c \tag{3}$$

where "a" is a trimming number and c is a rejection constant. For a=0 this method is equivalent to Fisher's method. If a>0, equation (3) consists of trimming some of the largest $\hat{p}_j^2$ values in estimating the residual variance.

From the above tests, let L be the number of amplitudes (p_j) that is considered to be different from zero. Given values of λ_1, λ_2, ... λ_L and fitting the regression equation

$$\hat{x}_t = \hat{\mu} + \sum_{j=1}^{L} \left(\hat{a}_j \cos 2\pi\lambda_j t + \hat{b}_j \sin 2\pi\lambda_j t\right)$$

by the least squares method, the coefficients $\hat{a}_j$ and $\hat{b}_j$ are obtained. This method can be applied successfully when the frequencies are well separated and actual value of L is not too large. However, when two or more frequencies are close together, the estimation may turn out to be very unstable and may give wrong results (Damsleth and Spjotvoll (1982)). Two methods which do not have these disadvantages have been developed. The procedures of these two methods are briefly described next. The fast Fourier transform (FFT) method is also discussed in the last portion of this section.

2.1 High resolution frequency analysis (HRFA) of times series

This method was developed by Siddiqui and Wang (1984), and it has been satisfactorily applied to three geological times series. The procedure of this method is described by the following steps:

Let x(t), $-n \leq t \leq n$, be a time series that has been read at equal intervals of time and has 2n+1 observations.

1. Compute $|\Delta Z_n (f_j)|$ where

$$\Delta Z_n(f_j) = \sum_{t=-n}^{n} e^{-2\pi i f_j t} \frac{\sin \pi \Delta f t}{\pi t} x(t) \quad -0.5 \le f \le 0.5$$

with $\Delta f = 0.005$ and $f_j = j\Delta f$, $\quad j=1, 2, \ldots (2\Delta f)^{-1}$.

2. Find the local maximum of $|\Delta Z_n(f_j)|$ in each frequency range $(f_j \pm 0.0025)$. This is achieved by performing a more refined scanning of the interval (for example, with $\Delta f = 0.0001$).

3. After the $\hat{f}_j$ which corresponds to the local maximum of $|\Delta Z_n(f_j)|$ in each frequency range is chosen,

   a. compute $\bar{x}$, a_j, and b_j, $j = 1, 2, \ldots, 100$. Where

$$\bar{x} = \frac{1}{2n+1} \sum_{t=-n}^{n} x(t)$$

$$a_j = \frac{2}{2n+1} \sum_{t=-n}^{n} [x(t)-\bar{x}] \cos 2\pi \hat{f}_j t$$

$$b_j = \frac{2}{2n+1} \sum_{t=-n}^{t=n} [x(t) - \bar{x}] \sin 2\pi \hat{f}_j t$$

   The squared amplitude in each frequency $\hat{f}_j$ is computed by $c_j^2 = a_j^2 + b_j^2$, $j = 1, 2, \cdots 100$. The subscripts j are so chosen that $c_1^2 \ge c_2^2 \ge \cdots \ge c_{100}^2$, and the frequencies are subscripted accordingly so that c_1^2 corresponds to frequency $\hat{f}_1$, c_2^2 to $\hat{f}_2$ and so on.

   b. Let $S^2 = \sum_{t=-n}^{n} [x(t) - \bar{x}]^2$ be the total sum of squares of the x(t) series, and

$$g = \frac{2n+1}{2} \frac{c_1^2}{S^2}$$

   g corresponds to the test statistic proposed by Fisher (1929). For large N, the value of g_α may be computed approximately by the formula

$$g_\alpha = \frac{1}{N} [\ln N - \ln(-\ln(1-\alpha))]$$

   c. Compare g and g_α. If $g < g_\alpha$, reject $\hat{f}_1$, considering it as *spurious,* and go to step 4.

   If $g \ge g_\alpha$, then $\hat{f}_1$ is accepted as a significant frequency and the computation proceeds to the next step.

   d. Compute $x_1(t) = x(t) - a_1 \cos 2\pi \hat{f}_1 t - b_1 \sin 2\pi \hat{f}_1 t$ for $t = -n, \ldots, 0, 1, \ldots, n$. Use $x_1(t)$ as an input time series and go to step 1.

The computations are continued in this fashion until either all the 100 frequencies pass the test and are considered to be significant, in which case L = 100, or $\hat{f}_{L+1}$ $0 \leq L \leq 99$ does not pass the test and $\hat{f}_{L+1}$, $\hat{f}_{L+2}$, $\hat{f}_{L+3}$, ... $\hat{f}_{100}$ are declared spurious.

4. Fit the regression equation

$$\hat{x}(t) = \hat{\mu} + \sum_{j=1}^{L} (\hat{\alpha}_j \cos 2\pi\hat{f}_j\, t + \hat{\beta}_j \sin 2\pi\hat{f}_j\, t)$$

by the least squares method and obtain the residual series $\hat{e}(t) = x(t) - \hat{x}(t)$. These four steps complete a single cycle of this procedure. The residual series $\hat{e}(t)$ is now subjected to steps 1 to 4 to obtain any additional significant frequencies that might have been missed in the first cycle. However, in the cases studied herein, this recycle procedure did not detect any additional significant frequency.

2.2 A stepwise estimation procedure (SEP)

This method was developed by Damsleth and Spjotvoll (1982), based on the theory developed by Anderson (1971) which demonstrated that the presence of a significant frequency λ will induce large values of the amplitudes of the harmonics close to it. Thus the large values among the amplitudes $\hat{p}_1$, ..., $\hat{p}_m$ indicate regions in which the true frequencies may be found. This fact led Damsleth and Spjotvoll to develop the following procedure.

1. Calculate the harmonic coefficients and test their significance.

   a. Use the $\{x_t\}$ series as input.

   b. Calculate the harmonic coefficients (i.e. calculate a_0, a_j, b_j, and $\hat{p}_j$, for j=1, ...m by using equation 1).

   c. Choose the largest amplitude, say $\hat{p}_k$, and test whether or not it is significant by using the equation 3 with a given c. The value of c is found from the Bolviken's tables (1983) for a given significance level ϵ and a parameter a.

   d. If there are no significant amplitudes, go to step 5. Otherwise go to the step 2.

2. Minimize $\sum_{t=1}^{n} [x_t - \mu - \alpha \cos 2\pi\lambda t - \beta \sin 2\pi\lambda t]^2$ with respect to μ, α, β, and λ, using k/n as a starting value for λ.

3. Remove the effect of the frequency closest to k/n by calculating

$$x_t' = x_t - \hat{\alpha} \cos 2\pi\hat{\lambda} t - \hat{\beta} \sin 2\pi\hat{\lambda} t$$

where $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\lambda}$ are the minimizing values obtained in step 2.

4. Provided a > 0, reduce the trimming factor a by one. Go to step 1 with the series $\{x_t'\}$ obtained in step 3.

5. Using the obtained frequencies, estimate the coefficients $\{\alpha_j\}$ and $\{\beta_j\}$ simultaneously.

2.3 The fast fourier transform

The well known fast Fourier transform (FFT) is used for calculating discrete Fourier transform (DFT) of a series. The Fourier transform (X_k) of a finite series x_t, t = 0, 1, 2, ..., N-1 is defined as

$$X_k = \frac{1}{N} \sum_{t=0}^{N-1} x_t \, e^{\frac{-i(2\pi kt)}{N}} \qquad k = 0, 1, 2, ..., N-1.$$

For computing X_k, it is necessary to make N multiplications of the form $x_t \, e^{\frac{-i(2\pi kt)}{N}}$ for each of N values of X_k. Consequently calculation of the full sequence X_k would require N^2 multiplications. Instead of calculating the discrete fourier transform of the original sequences, the FFT works by partitioning the full sequence x_t into a number of shorter sequences. The FFT thus reduces the number of operations to N $\log_2$ N. For analyzing long time series, the FFT thus provides an enormous saving in computer processing time. The basic theory of FFT is as follows:

A series x_t, t = 1, 2, ..., N-1, (where N is even), is partitioned into two half-series y_t and z_t, where

$$y_t = x_{2t}; \quad z_t = x_{2t+1}; \quad t = 0, 1, 2, ... \left(\frac{N}{2} - 1\right)$$

The series y_t and z_t each consists of N/2 values and hence have Fourier transforms

$$Y_m^{(\frac{N}{2})} = \frac{2}{N} \sum_{t=0}^{(\frac{N}{2}-1)} y_t \, e^{-i\frac{(4\pi tm)}{N}}$$

$$Z_m^{(\frac{N}{2})} = \frac{2}{N} \sum_{t=0}^{(\frac{N}{2}-1)} z_t \, e^{-i\frac{(4\pi tm)}{N}} \qquad m = 0, 1, 2, ... \left(\frac{N}{2} - 1\right).$$

But $X_m^{(N)}$ and $Y_m^{(\frac{N}{2})}$, $Z_m^{(\frac{N}{2})}$ are related, since

$$X_m^{(N)} = \frac{1}{N} \sum_{t=0}^{N-1} x_t \, e^{-\frac{i2\pi tm}{N}}$$

$$X_m^{(N)} = \frac{1}{N} \left\{ \sum_{t=0}^{(\frac{N}{2}-1)} x_{2t} \, e^{-\frac{i2\pi(2t)m}{N}} + \sum_{t=0}^{(\frac{N}{2}-1)} x_{2t+1} \, e^{-\frac{i2\pi(2t+1)m}{N}} \right\}$$

$$= \frac{1}{2}\left\{ Y_m^{(\frac{N}{2})} + e^{-\frac{i2\pi m}{N}} Z_m^{(\frac{N}{2})} \right\} \quad \text{for } m = 0, 1, 2, \ldots, \frac{N}{2} - 1 \tag{4}$$

Consequently,

$$X_m^{(N)} = \frac{1}{2}\left\{ Y_{m-\frac{N}{2}}^{(\frac{N}{2})} + e^{-\frac{i2\pi m}{N}} Z_{m-\frac{N}{2}}^{(\frac{N}{2})} \right\} \quad \text{for } m = \frac{N}{2}, \frac{N}{2} + 1, \ldots, N-1 \tag{5}$$

The Fourier transform for the series x_t is obtained from the Fourier series of the half-series y_t and z_t. Similarly, if N is a power of 2, the half-sequences y_t and z_t are partitioned into quarter-sequences, and so on, until eventually the last sub-sequences have only one term each.

3 Data used in this study

Three different types of data are used in this study to test the methods. The first set consists of 10 generated harmonic time series. The second set consists of data generated by 6 different Autoregressive moving average (ARMA) models. The theoretical spectra are known for both of these series. The third set consists of 11 observed time series that describe hydrologic, climatic, sunspot and economic phenomena. Among the observed series, the sunspot and the wheat index series are well known and much studied. Details of these data sets are presented next.

3.1 Harmonic time series

Ten sets of simulated data are generated in this category. The data are generated by the formula

$$x_t = \mu + \sum_{j=1}^{L} (a_j \cos 2\pi\lambda_j t + b_j \sin 2\pi\lambda_j t) + \epsilon_t \tag{6}$$

where μ is a constant, a_j, b_j are coefficients, λ_j is the frequency, and ϵ_t is a normal random variate with zero mean and a given variance. The first five time series are generated by using a single low frequency ($\lambda_j = 0.01$), and the last five are generated by using two frequencies which are close to each other ($\lambda_j - \lambda_i \leq 0.0002$). Table 1 gives the coefficients, frequencies, means, variances and the means and variances of ϵ_t of these data. These data were used to investigate the ability of the methods to identify the frequencies in time series which have a single low frequency component or which have two frequency components which are close to each other.

3.2 ARMA time series

Six different ARMA models were used to generate the test data. The models used to generate the data are given below:

Table 1. The coefficients for generation of harmonic time series

No.	a_j	b_j	f_i	mean	variance	ϵ_t
1	15.30	10.20	0.01	47	168	no noise
2	15.30	10.20	0.01	47	205	N(0, 0.2V1)
3	15.30	10.20	0.01	46	240	N(0, 0.4V6)
4	15.30	10.20	0.01	47	276	N(0, 0.6V1)
5	15.30	10.20	0.01	46	346	N(0, V1)
6	15.30	10.20	0.0552	46	341	no noise
7	15.30	10.20	0.0552	45	416	N(0, 0.2V6)
8	15.30	10.20	0.0552	45	487	N(0, 0.4V6)
9	15.30	10.20	0.0552	44	559	N(0, 0.6V6)
10	15.30	10.20	0.0552	44	701	N(0, V6)

V1 = 167.78 V6 = 340.59

Table 2. The results of harmonic time series

Case	Variance σ^2	HRFA σ_ϵ^2	HRFA $\sigma_\epsilon^2/\sigma^2$	SEP σ_ϵ^2	SEP $\sigma_\epsilon^2/\sigma^2$
1	168	0.605	0.004	5.28	0.0315
2	205	33.4	0.163	43.7	0.213
3	240	71.7	0.299	83.3	0.347
4	275	105	0.382	122.	0.442
5	345	171	0.497	187.	0.541
6	341	11.2	0.0328	7.10	0.0208
7	416	79.8	0.192	81.2	0.195
8	487	147	0.302	166	0.340
9	559	279	0.50	234	0.418
10	701	419	0.597	386	0.550

Model No.	Model
Model 1:	$x_t = 0.4\, x_{t-1} + 0.45\, x_{t-2} + \epsilon_t$
Model 2:	$x_t = 2.7607\, x_{t-1} - 3.8106\, x_{t-2} + 2.6535\, x_{t-3} - 0.9238\, x_{t-4} + \epsilon_t$
Model 3:	$x_t = -1.7\, x_{t-1} - 2.4\, x_{t-2} - 1.634\, x_{t-3} - 0.878\, x_{t-4} - 0.168\, x_{t-5} + \epsilon_t$
Model 4:	$x_t = -0.25\, x_{t-1} - 0.5\, x_{t-2} - \epsilon_{t-1} - 0.75\, \epsilon_{t-2} + \epsilon_t$
Model 5:	$x_t = 0.95\, \epsilon_{t-1} + \epsilon_t$
Model 6:	$x_t = y_t + z_t$
	$y_t = 0.5\, y_{t-1} + \epsilon_t$
	$z_t = 0.4 \cos(0.2\pi t) + 0.7 \cos(0.5\pi t) + \cos(0.67\pi t)$

In each model, ϵ_t is a zero-mean, uncorrelated normal sequence with unit variance.

From each of these models, series consisting of 50, 100, and 200 samples were generated. Sequences consisting of more than these number of points were generated. The initial 200 points in each set were ignored so that the transient effects of the initial conditions are not significant. The variance of the series for each model with different sample sizes is given in Table 4. The notation used in Tables 3 and 4 and in the figures is as follows: case 1.50 indicates model 1 with sample size 50.

The true spectral density functions of models 1-6 are known. These are listed below.

a. *Model 1:* A single spectral peak at f=0 cycle per time (CPT) and little power elsewhere.

b. *Model 2:* Two sharp, close spectral peaks near f=0.1 CPT.

c. *Model 3:* A sharp spectral peak near f=0.32 CPT.

d. *Model 4:* A wide band spectrum with a single peak near f=0.25 CPT.

e. *Model 5:* A high spectrum value at f=0 which slowly decreases to zero at f=0.5.

f. *Model 6:* A mixed harmonic and AR(1) process.

The theoretical spectra of these ARMA models are calculated by the well known formula,

$$p(f) = 2\sigma_a^2 \frac{|1 - \theta_1 e^{-i2\pi f} - \cdots - \theta_q e^{-i2\pi qf}|^2}{|1 - \phi_1 e^{-i2\pi f} - \cdots - \phi_p e^{-i2\pi pf}|^2} \quad 0 \le f \le 0.5$$

where σ_a^2 is the series variance, θ_1, ..., θ_q are moving average coefficients, and ϕ_1, ..., ϕ_p are autoregressive coefficients of ARMA models. The intensity of the spectrum at frequency f_j is calculated by using:

$$I(f_j) = \frac{N}{2}(a_j^2 + b_j^2) \quad j = 1, 2, ..., q$$

where N is the number of data points in the series, a_j and b_j are the significant harmonic coefficients which are picked up from each method. The intensity $I(f_j)$ is compared with the theoretical spectrum.

3.3 Observed hydrologic time series

Eleven observed time series are used in this study. The observed data series consist of sunspot number (SN1) series; series of double (Hale) sunspot cycle (DSN1); and three drought area index (DAI) series and three flood area index (FAI) series from India. The three DAI series are the percentage areas of India corresponding to mean monsoon index for a given year. These data were collected and analyzed by Bhalme and Mooley (1981) who studied the dependence between DSN1 and DAI and FAI series by using the spectrum, cross spectrum and harmonic dial analysis. The severity of drought or flooding increases from DAI1 to DAI3 or FAI1 and FAI3.

The SN1 series are coded as being positive or negative according to the polarity associated with sunspot numbers and these constitute the double sunspot series. The details of the derivation of double sunspot series are found in Mitchell et al. (1979). All the series are

Table 3. The peak theoretical and sample intensities of ARMA time series

Case	True SDF	FFT	HRFA	%	SEP	%
1.50	128	128	130	1.0	78	0.61
1.100	206	206	172	0.83	142	0.69
1.200	487	487	338	0.69	329	0.68
2.50	58840	58840	55990	0.95	58839	1.0
2.100	99150	99150	95786	0.97	96440	0.98
2.200	139800	139800	139100	0.99	106719	0.76
3.50	846	846	783	0.93	567	0.67
3.100	935	935	820	0.88	839	0.90
3.200	5584	5584	4615	0.83	658	0.12
4.50	185	185	167	0.90	59	0.32
4.100	365	365	280	0.77	246	0.67
4.200	957	957	512	0.54	62	0.07
5.50	83	83	73	0.88	0	0.00
5.100	154	154	110	0.71	24	0.16
5.200	380	380	201	0.53	0	0.0
6.50	95	95	87	0.92	0	0.0
6.100	179	179	162	0.91	33	0.19
6.200	359	359	176	0.49	55	0.15

Table 4. The results of ARMA time series

	Variance	HRFA		SEP	
Case	σ^2	σ_ϵ^2	$\sigma_\epsilon^2/\sigma^2$	σ_ϵ^2	$\sigma_\epsilon^2/\sigma^2$
1.50	2.61	0.165	0.0632	1.03	0.394
1.100	2.08	0.371	0.178	0.656	0.315
1.200	2.45	0.713	0.291	0.795	0.325
2.50	1200	0.473	0.0004	0.0986	0.00008
2.100	1000	5.29	0.0053	27.40	0.0274
2.200	702	21.0	0.030	33.00	0.0471
3.50	17.3	0.228	0.0132	5.70	0.33
3.100	9.44	0.616	0.0652	0.997	0.106
3.200	28.1	3.41	0.1213	24.8	0.882
4.50	3.77	0.278	0.0737	2.58	0.683
4.100	3.69	0.826	0.224	1.21	0.327
4.200	4.81	2.18	0.453	4.49	0.935
5.50	1.69	1.202	0.120	1.69	1.00
5.100	1.56	0.41	0.263	1.31	0.844
5.200	1.91	0.981	0.514	1.91	1.0
6.50	1.93	0.132	0.0684	1.93	1.0
6.100	1.81	0.423	0.234	1.47	0.813
6.200	1.80	0.780	0.433	1.53	0.847

annual observations starting from 1891 A.D. and ending in 1979 A.D. with 89 observations. The statistical characteristics of these data are given in Table 5.

Table 5. Statistical characteristics of observed times series

Series	Period	N	Mean	Variance	Skewness Coefficient	Source
SN1	1891-1979	89	55.87	1973	1.04	Schonweise (1978)
DSN1	" "	"	5.39	5066	0.32	Mitchell et al. (1979)
DAI1	" "	"	27.20	432	0.85	
DAI2	" "	"	10.58	191	1.82	Bhalme and
DAI3	" "	"	2.67	49	4.16	Mooley (1981)
FAI1	" "	"	24.97	342	0.73	
FAI2	" "	"	11.58	128	1.23	
FAI3	" "	"	4.12	34	1.77	
BWP I	1727-1869	143	99.19	328	1.28	Anderson (1971)
CEAR	" "	"	34.13	20.43	0.30	Met. Magazine (1971)
CEAAT	" "	"	9.16	0.42	-0.35	Lamb (1977)

From the elementary statistics given in Table 5 the DSN1 series may be inferred to be approximately normally distributed. The SN1, DAI and FAI series are highly skewed. A 22--year cycle is present in the DSN1 series (Mitchell et al. (1979)) and in FAI series (Bhalme and Mooley (1981)) whereas the DAI series show a weak quasi-periodicity of 2.7 - 3 years (Bhalme and Mooley (1981)). The sunspot series SN1 has a well known periodicity of 11 years.

Other observed series are the Beveredge Wheat Price Index (BWPI), Central England Annual Rainfall (CEAR) and Central England Annual Air Temperature (CEAAT) series. All these series start from 1727 A.D. and end in 1869 with 143 observations. The statistical characteristics of these series are also given in Table 5. These data are not periodic but can be fitted by small order AR models. These data are also approximately normally distributed.

4 Results and discussion

Three methods (HRFA, SEP, and FFT) are used to estimate the significant spectral intensities and corresponding frequencies of the time series. The series are reconstructed by using the significant components. These are called the regenerated series. The theoretical spectra and the estimated spectra obtained by each method were also computed. In the case of FFT, the regeneration is exact. The variance, residual variance, and the ratio of the variance of the series to the residual variance were computed as a measure of the goodness of fit between the computed and observed or generated data. The results from the three different types of time series are separately presented below.

4.1 Harmonic time series

The results for the harmonic time series are shown in Table 2 and examples of the results are given in Figures 1 and 2. These results indicate that the FFT method can perfectly fit the observed data in all cases as it should. However, when the frequency of a series is not equal to an integer times the inverse of the sample size (i.e. j/N, N = the sample size, j = an integer between 1 to N), the estimated amplitudes and frequencies are spread over a small range. Both the HRFA and SEP methods can fit the series very well when the series do not have noise (cases 1 and 6, Table 2). The goodness of fit decreases as the noise increases,

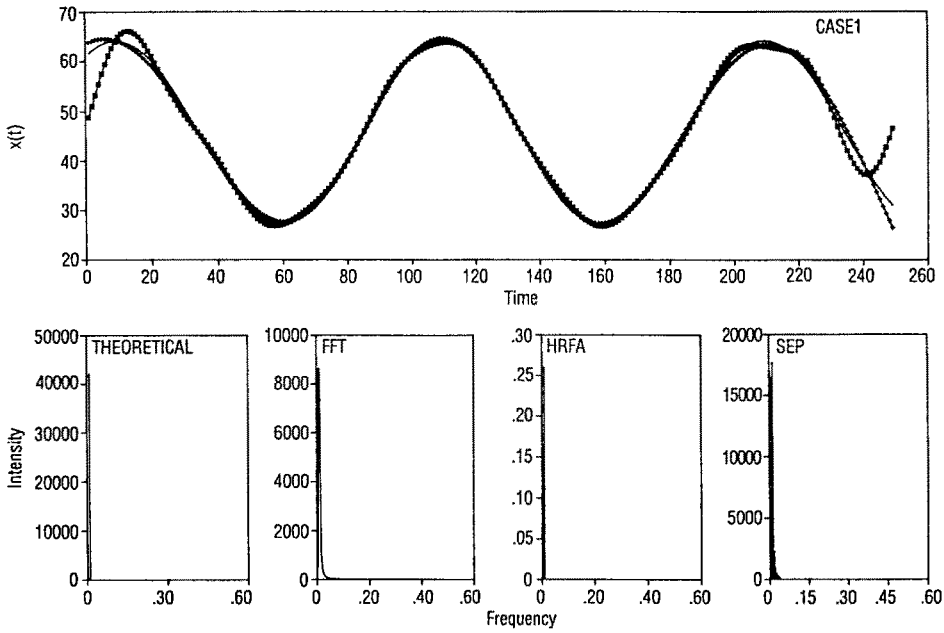

Figure 1. Original and reconstructed time series for case 1 in Table 1 (top). Original series and those reconstructed by using FFT (solid line), HFRA (+) and SEP (*) estimates are shown. Theoretical and FFT, HRFA and SEP spectra are shown in the bottom row.

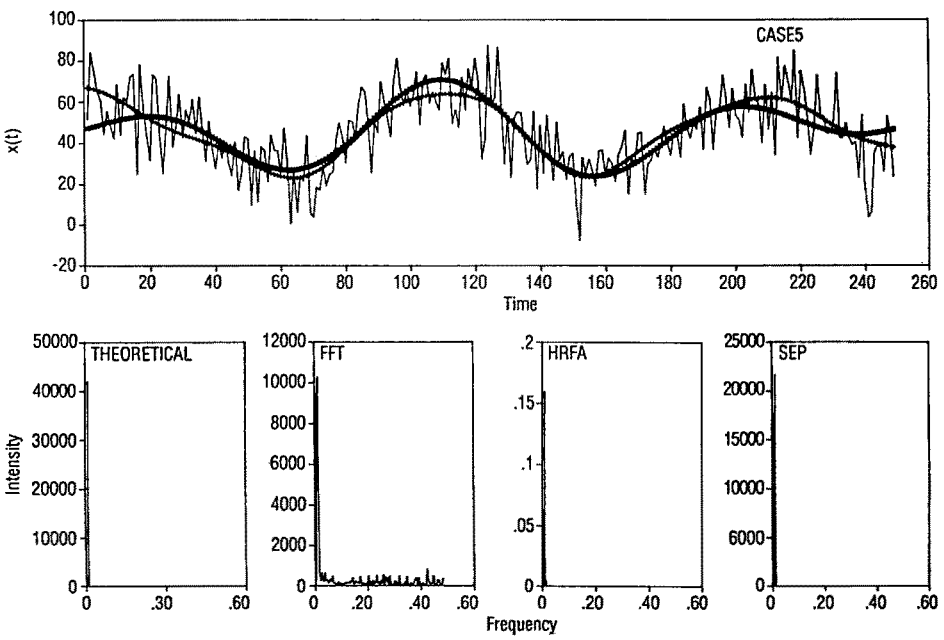

Figure 2. Original and reconstructed time series for case 5 in Table 1 (top). Original series and those reconstructed by using FFT (solid line), HFRA (+) and SEP (*) estimates are shown. Theoretical and FFT, HRFA and SEP spectra are shown in the bottom row.

although the computed series can still match the harmonic trend in the original data. In all cases, both methods are able to pick up the significant frequencies in the series. The HRFA and SEP methods can identify low frequencies (0.01), but not two very close frequencies ($\lambda_j - \lambda_i \leq 0.0002$). An example of the difficulty in identifying two low frequencies is shown in Figure 2. The HRFA can identify two close frequencies which have a difference greater than 0.005. However, this result can be modified by simply selecting a smaller value of Δf. The SEP method can identify two close frequencies which have a difference that is larger than 1/N, N = number of sample points. Therefore when the sample size is large, the SEP method can identify two close frequencies.

4.2 ARMA time series

The results of analysis of ARMA time series are shown in Tables 3 and 4 and Figures 3 through 5. From the results in Table 3, it is seen that the HRFA method, in most cases, can explain about 80% of the peak intensity of the spectrum. The SEP method can explain about 65% of the peak intensity of the AR series, and *it fails* to explain the variance of the MA process (Case 5). Table 4 shows the series variance (σ^2), the residual variance (σ_ϵ^2) and the ratio of residual variance to series variance ($\sigma_\epsilon^2/\sigma^2$) in the HRFA and SEP methods. The HRFA method has smaller residual variance and the ratio ($\sigma_\epsilon^2/\sigma^2$) than the SEP method in all except case 2.50. The HRFA method can estimate the spectra of the ARMA process better than the SEP method. Comparing the values of $\sigma_\epsilon^2/\sigma^2$ between AR models 1, 2, 3 and ARMA models 4 and 5, the AR models have smaller values of $\sigma_\epsilon^2/\sigma^2$ than ARMA models which shows that both methods can characterize periodicities in AR processes better than in ARMA processes.

Figures 3 through 5 are examples of the goodness of fit between the generated data and simulated results for AR and ARMA models. The theoretical spectrum and estimated spectra from each method are also shown in these figures. The FFT method misses the two peaks present in the data in Figure 3. The HRFA and SEP estimates indicate these peaks. The HRFA estimates are superior to the SEP estimates in Figure 4 where the FFT has a spurious second peak. The FFT and HRFA estimates indicate a wideband spectrum in Figure 5 whereas the SEP estimate shows a single spike and thereby completely misrepresents the true spectrum.

4.3 Observed time series

The results are shown in Table 6 and Figures 6 through 10. From the results in Table 6, it is found that the SEP method fails to represent adequately the time series in most cases, whereas the HRFA method is able to explain at least a part of the variance in most cases. In fact, the HRFA method can reduce the ratio of variances to values smaller than 10% for the DSN and the SN series. For the series CEAAT, CEAR, DAI1, DAI3 and FAI2, the HRFA method fails to detect any periodicity. However, these series are close to being random and their intensities seem equally distributed in the frequency domain. In other words, these series do not have significant periodicities.

5 Conclusions

Based on tests on 10 harmonic times series, 6 ARMA time series, and 11 observed time series on two methods of estimation of significant harmonic component, we find that the HRFA method is superior to the SEP method. However, in both the methods, the algorithm had to be changed to obtain good results.

Table 6. The results of observed hydrologic time series

Series	Data num.	Variance σ^2	HRFA σ_ϵ^2	HRFA $\sigma_\epsilon^2/\sigma^2$	SEP σ_ϵ^2	SEP $\sigma_\epsilon^2/\sigma^2$
BWPI	143	330.0	49.7	0.151	304.0	0.922
CEAAT	143	0.419	0.306	0.73	0.419	1.0
CEAR	143	20.6	17.7	0.862	20.6	1.0
DAI.1	89	437.0	305.0	0.698	437.0	1.0
DAI.2	89	193.0	21.6	0.112	193.0	1.0
DAI.3	89	49.9	49.9	1.0	49.9	1.0
DSN.1	89	5120.0	112.0	0.0218	3890.0	0.76
FAI.1	89	346.0	29.9	0.0864	346.0	1.0
FAI.2	89	130.0	88.8	0.685	130.0	1.0
FAI.3	89	34.7	18.4	0.531	34.7	1.0
SN.1	89	2000.0	112.0	0.056	903.0	0.452

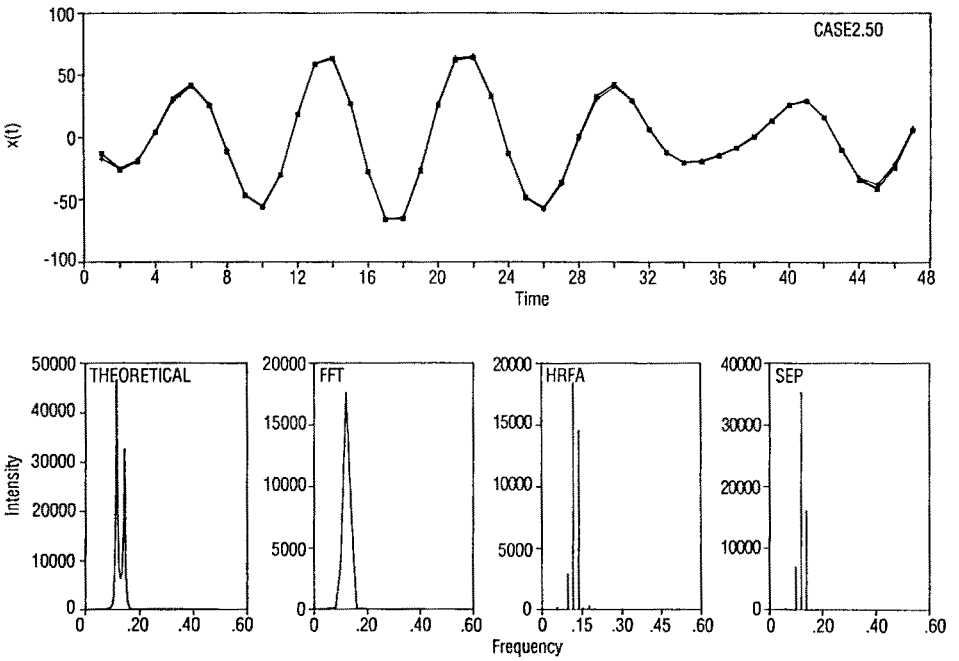

Figure 3. Original and reconstructed times series for Model 2 with 50 samples (top). Original series and those reconstructed by using FFT (solid line), HFRA (+) and SEP (*) estimates are shown. Theoretical, FFT, HRFA and SEP spectra are shown in the bottom row.

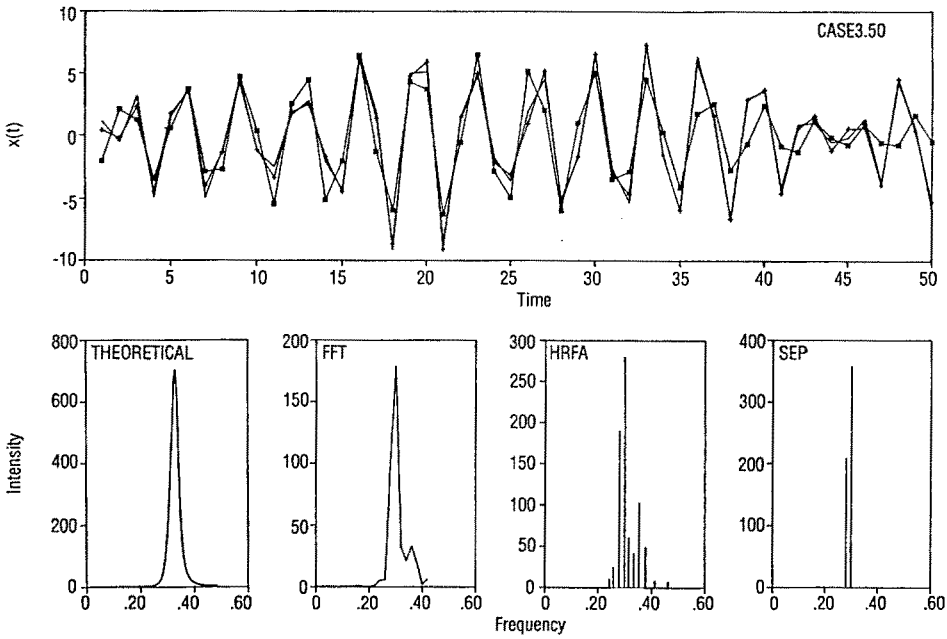

Figure 4. Original and reconstructed times series for Model 3 with 50 samples (top). Original series and those reconstructed by using FFT (solid line), HFRA (+) and SEP (*) estimates are shown. Theoretical, FFT, HRFA and SEP spectra are shown in the bottom row.

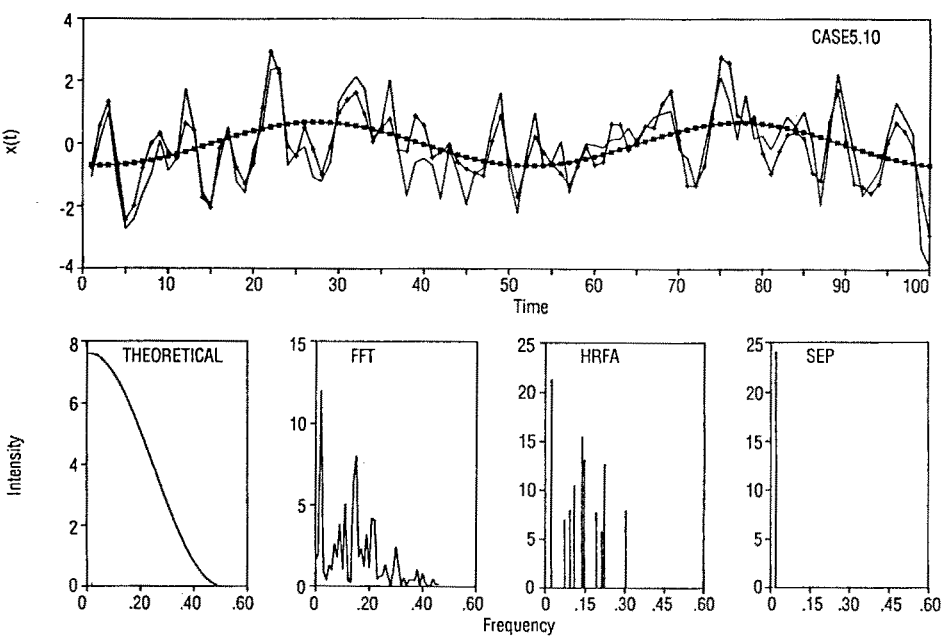

Figure 5. Original and reconstructed times series for Model 5 with 100 samples (top). Original series and those reconstructed by using FFT (solid line), HFRA (+) and SEP (*) estimates are shown. Theoretical, FFT, HRFA and SEP spectra are shown in the bottom row.

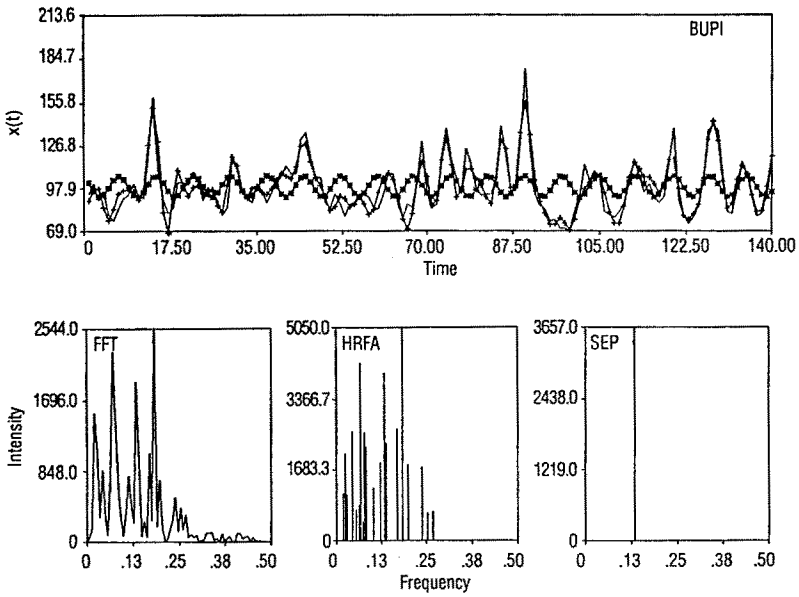

Figure 6. Original and reconstructed Beveridge Wheat Price Index series (top). The original series and those reconstructed by FFT (solid line), HRFA (+) and SEP (*) estimates are shown. FFT, HRFA and SEP spectra are shown in the bottom row.

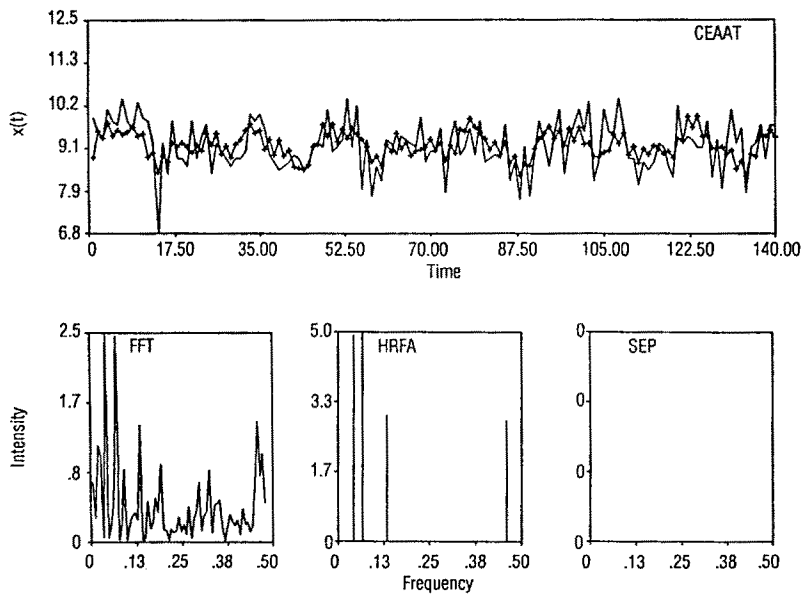

Figure 7. Original and reconstructed Central England Annual Air Temperature series (top). The original series and those reconstructed by FFT (solid line), HRFA (+) and SEP (*) estimates are shown. FFT, HRFA and SEP spectra are shown in the bottom row.

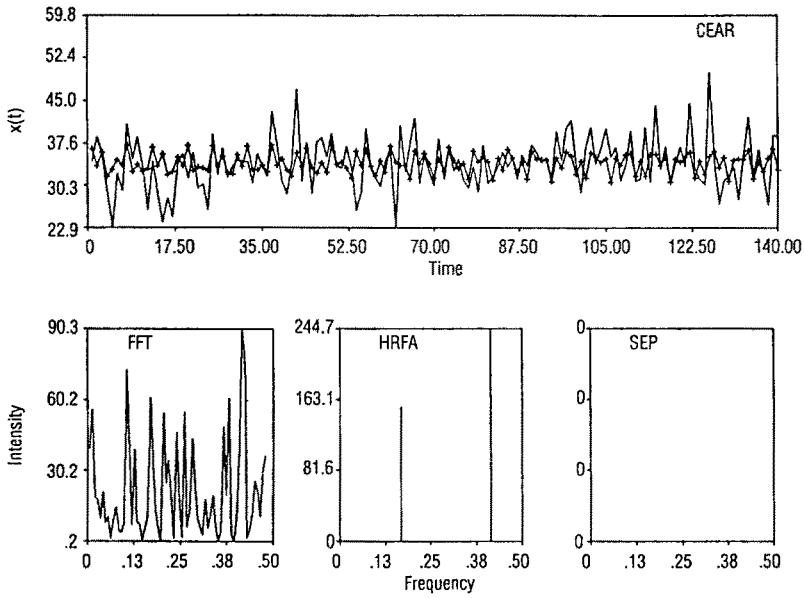

Figure 8. Original and reconstructed Central England Annual rainfall series (top). The original series and those reconstructed by FFT (solid line), HRFA (+) and SEP (*) estimates are shown. FFT, HRFA and SEP spectra are shown in the bottom row.

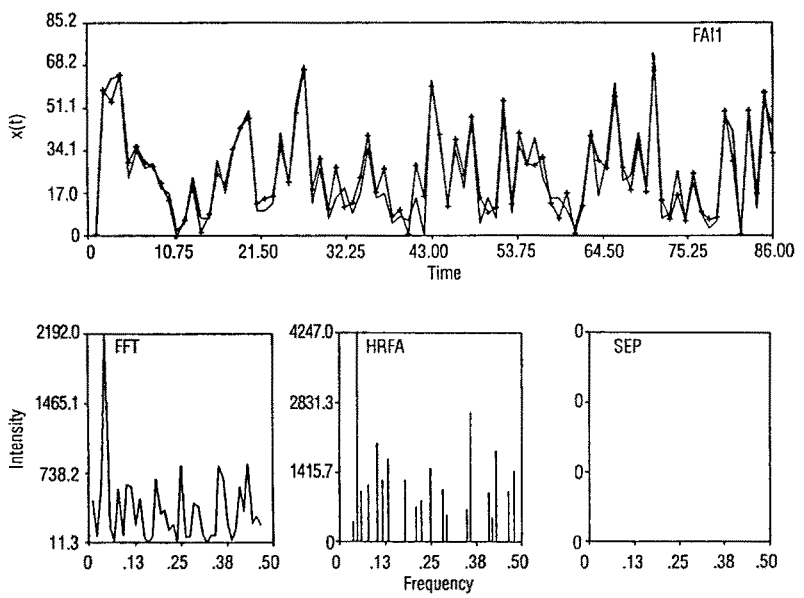

Figure 9. Original and reconstructed Flood Area Index (FAI1) series (top). The original series and those reconstructed by FFT (solid line), HRFA (+) and SEP (*) estimates are shown. FFT, HRFA and SEP spectra are shown in the bottom row.

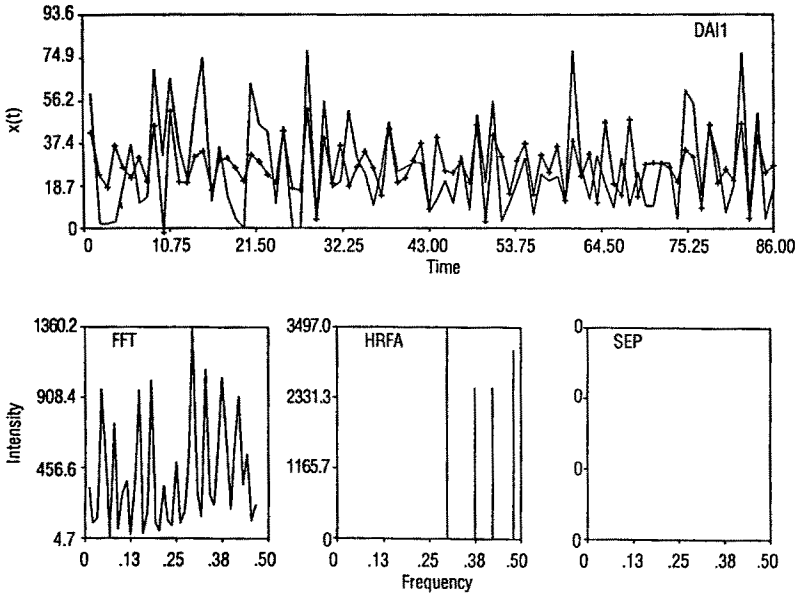

Figure 10. Original and reconstructed Drought Area Index (DAI1) series (top). The original series and those reconstructed by FFT (solid line), HRFA (+) and SEP (*) estimates are shown. FFT, HRFA and SEP spectra are shown in the bottom row.

References

Anderson, T. W. 1971: The statistical analysis of time series. New York, John Wiley

Beamish, N.; Priestley, M. B. 1981: A study of autoregressive and window spectral estimation. Appl. Statist. 30, pp. 41-58

Bhalme, H. N.; Mooley, D. A. 1981: Cyclic fluctuations in the flood area and relationship with double (Hale) sunspot cycle. J. Am. Met. Soc., pp. 1041-1048

Bolviken, E. 1983: New tests of significance in periodogram analysis. Scandinavian Journal of Statistics 10, pp. 1-9

Burg, J. P. 1967: Maximum entropy spectral analysis. In: Proc. 37th Meeting, Society of Exploration Geophysics, Oklahoma City, Oklahoma

Cadzow, J. A. 1981: Autoregressive moving average spectral estimation: A model equation error procedure. IEEE Trans. on Geoscience and Remote Sensing, Vol. G.E. 19, pp. 24-28

Damsleth, E.; Spjotvoll, E. 1982: Estimation of trigonometric components in time series. Journal of the American Statistical Association 77, pp. 381-387

Durgunoglu, A.; Rao, A. R. 1985: ARMA spectral analysis of hydrologic time series. CE-HSE-85-12, School of Civil Engineering, Purdue University, W. Lafayette, IN 47907, pp. 207

Fisher, R. A. 1929: Test of significance in periodogram analysis. Proceedings of the Royal Society, Ser. A140, pp. 411-431

Friedlander, B. 1982: A recursive maximum likelihood algorithm for ARMA spectral estimation. IEEE Trans. on Information Theory IT-28, pp. 639-646

Friedlander, B. 1982: System identification technique for adaptive signal processing. Circuits Systems Signal Process, Vol. 1, No. 1

Jenkins, G. M.; Watts, D. G. 1968: Spectral analysis and its applications. Holden-Day, San Francisco, California

Kashyap, R. L.; Rao, A. R. 1976: Dynamic stochastic models from empirical data. Academic Press, New York

Kay, S. M. 1980: A new ARMA spectral estimator. IEEE Trans. Acoust. Speech and Signal Process ASSP-28, 1980, pp. 585-588

Kay, S. M.; Marple, S. L., Jr. 1981: Spectrum analysis – A modern perspective. Proceedings of the IEEE 69, No. 11, pp. 1380-1419

Lamb, H.H. 1979: Climate: Present, past and future. Vol. 2, Climatic History and the Future, London, Methuen, pp. 572-579

Marple, S. L., Jr. 1980: A new autoregressive spectrum analysis algorithm. IEEE Trans. Acoust. Speech and Signal Process ASSP-28, pp. 441-454

Mitchell, J. M.; Stockton, C. W.; Meko, D. M. 1979: Evidence of a 22-year rhythm of drought in the western U.S. related for the Hale solar cycle since the 17th century. In: Solar Terrestrial Influences on P.T.O Weather and Climate, ed. by B. M. McCormac and T. A. Seliga, D. Reidel Pub. Co., Dordrecht, Holland

Pisarenko, V. F. 1973: The retrieval of harmonics from a covariance function. Geophysical J. Royal Astro. Soc. 33, pp. 347-366

Rao, A. R.; Durgunoglu, A. 1988: AR and ARMA spectral estimation. Stochastic Hydrology and Hydraulics 2, pp. 35-50

Rao, A. R.; Padmanabhan, G.; Kashyap, R. L. 1984: A comparative analysis of recently developed methods of spectral analysis. In: Frontiers in Hydrology, ed. W. H. C. Maxwell and L. R. Beard, Water Resources Publications, Littleton, Colorado, pp. 127-149

Schonweise, C. D. 1978: On the problem of statistical climate modeling. Ach. Met. Geoph. Biokl. Series B16, pp. 105-120

Siddiqui, M. M.; Wang, C. C. 1984: High-resolution frequency analysis of geological time series. Journal of Geophysical Research 89, No. D5, pp. 7195-7201