NILU TR : 11/84
REFERENCE : E-8414
DATE : SEPTEMBER 1984

SPECTRAL ANALYSIS OF THE TIME SERIES

Alena Moldanova

NORWEGIAN INSTITUTE FOR AIR RESEARCH
P.O. BOX 130, N-2001 LILLESTRØM
NORWAY

ISBN--82-7247-488-3

TABLE OF CONTENT

SPECTRAL ANALYSIS OF TIME SERIES

NILU are in the possession of long time series of measurements f.i. concentrations of sulphur dioxids and sulphate in aerosols from Birkenes and Bear Island, and it is possible to investigate the more complicated structures in the data. In the following text we shall restrict ourselves to the search for the periodicity.

THE PERIODICITY

The problems concerning the periodicity are easily understood when one recalls the wave theory from physics.

Periodicity in a given data sequence can be investigated by performing a spectral analysis of the data. It is usually used for the analysis of stationary time series, but, as shown in Andel [1] or Granger & Hatanaka [2], the same methods can be applied also when the stationarity is not granted.

The real periodic time series is modelled through

$$Z(t)=a_o + \sum_{i=1}^{r} a_i \cos(\omega_i t + \phi_i^1) + b_i \sin(\omega_i t + \phi_i^2) + X_t \qquad (1)$$

$$t = \ldots, -1, 0, 1, \ldots$$

where a_o represents the mean value of the series, a_i and b_i are the amplitudes of the cosine respective sine component of the wave in the frequency ω_i, $i=1,\ldots,r$. r is the number of cycles present in the spectrum of $\{Z_t\}$, and $\{X_t\}$ is some unknown stationary series. The least square estimates of the coefficients in (1) are given by

$$\hat{a}_o = \bar{Z}, \quad \hat{a}_i = \frac{2}{N} \sum_{t=1}^{N} Z_t \cos\omega_i t, \quad \hat{b}_i = \frac{2}{N} \sum_{t=1}^{N} Z_t \sin\omega_i t,$$

$i = 1,\ldots,r$, N denotes the length of the series.

The relation (1) can be rewritten for the real series as

$$Z_t = a_o + \sum_{i=1}^{r} a_i \cos(\omega_i t + \phi_i) + X_t. \qquad (2)$$

We shall assume that both the $\{Z_t\}$ and the $\{X_t\}$ are real series, and that X_t is normally distributed with zero mean and variance $\sigma^2 > 0$ for $t=\ldots,-1, 0, 1,\ldots$ However, $\{Z_t\}$ need not be stationary, and therefore its special density function $f(\lambda)$, $-\pi \leq \lambda \leq \pi$, need not be defined. The spectral density function provides a description of the strength of each frequency band. Its value is zero for all frequencies not present in the spectrum of a given series, and it is nonzero otherwise, with largest values in the most important frequency bands.

Under certain assumptions about the form of non-stationarity we may introduce another function $b(\lambda)$, which will serve as an counterpart of the spectral density for the purposes of investigating the periodicity. Since $\{Z_t\}$ is a real series, we shall consider λ from the interval $\langle 0,\pi \rangle$. We shall refer to both $b(\lambda)$ and $f(\lambda)$ as spectral densities.

When investigating the periodicity of $\{Z_t\}$, we shall actually compute an estimate of a spectral density composed of the spectral density of $\{X_t\}$ and of the periodic component. There are several estimators of the spectral density available, we shall use two of them, namely, the periodogram and the Parzen's estimator. The periodogram $I_N(\lambda)$ is defined by the relation

$$I_N(\lambda) = \frac{1}{2\pi N} \sum_{t=1}^{N} Z_t e^{-it\lambda}, \quad -\pi \leq \lambda \leq \pi.$$

For the real series this relation can be rewritten as

$$I_N(\lambda) = \frac{1}{2\pi} \left(C_o + 2 \sum_{k=1}^{N-1} C_k \cos(k\lambda)\right), \quad 0 \leq \lambda \leq \pi$$

where

$$C_k = \frac{1}{N} \sum_{t=1}^{N-k} X_t X_{t+k} \quad , \; k = 0, \ldots, N-1.$$

If the model (2) holds, then for large N the value of the periodogram in the points (or in the frequencies) $\lambda_1, \ldots, \lambda_r$ will be large, while it will be considerably smaller in other frequencies. The periodogram can be used for a test of the following hypothesis (cf. Hannan [3], VII§6). Let us suppose that there is only one frequency ω present in the spectrum of $\{Z_t\}$. Then the relation (2) will stand as

$$Z_t = a_o + a_1 \cos(\omega t + \phi) + X_t, \quad t = 1, \ldots, N.$$

We want to test the hypothesis that $a_1 = 0$, that is, that there is no periodicity present in the considered time series and the Z_t is normally distributed with zero mean and variance $\sigma^2 > 0$, $t = 1, \ldots, N$. Let us consider the values of the periodogram in the points

$$\lambda_s = \frac{2\pi s}{N} \quad , \quad s = 1, \ldots, m, \quad m = \frac{N-1}{2}$$

(we can always remove one observation to obtain N odd). Let us arrange the values $I_N(\lambda_1), \ldots, I_N(\lambda_m)$ in descending order of magnitude, let V_1 denote the largest value, etc., V_m the smallest one. Let

$$W = \frac{V_1}{V_1 + V_2 + \ldots + V_r} .$$

If V_1 is much larger then the sum of all values, then W will tend to 1, otherwise, W will tend to 1/m. Therefore, the values of W close to 1 will indicate a possible departure from the hypothesis. It can be shown (cf. Andel [1],VII§3) that for a large n the distribution of W under the null hypothesis can be approximated in a following manner: for all z < 0

$$\lim_{m \to \infty} P\left(W > \frac{z + \ln m}{m}\right) = 1 - \exp(-e^{-z}). \qquad (3)$$

We may extent this procedure to the case where there are more than one frequency present in the spectrum of $\{Z_t\}$. When V_1 was shown to be significant, we may omit this largest value and compute

$$W^{(1)} = \frac{V_2}{V_2 + V_3 + \ldots + V_m}$$

and again use the approximation (3) for the distribution of $W^{(1)}$, with m-1 instead of m (cf. Andel [1], Whittle [4]). If V_2 shows to be significant, we may repeat this procedure as long as necessary. But we should notice, that a test carried out in this way, is only approximate:

1. the null hypothesis is an ideal one and may not be in accordance with the type of the data,
2. when repeating the test procedure we accumulate several non-simultaneous test, and therefore the significance level is barely proximate,
3. repeating of the test procedure also violates the power of the test.

As an estimator of spectral density the periodogram has some disadvantages, e.g., its variance is not decreasing for a large sample size. Therefore we often use other estimators. The most widely used is one of the trimmed estimators - the Parzen's estimator (cf. Granger & Hatanaka [2], Hannan [3]). Its length is recommended to be between N/5 and N/6. We have used this estimator in the programme SPEDENCOR (user MOLDAN).

THE DEMODULATION

The demodulation process is described in Granger & Hatanaka [2], chapter 10.

Let us suppose that our process is not stationary in such way that $\{Z_t\}$ can be represented as

$$Z_t = a_o + \sum_{i=1}^{r} a_i(t)\cos(\omega_i t + \phi_i(t)) + X_t,$$

r is number of the frequencies present in the spectrum,
$t = 1,\ldots,N$,

where $a_i(t)$ and $\phi_i(t)$ respectively are affecting the amplitude phase of the wave in the frequency ω_i. An estimate of these changes can be obtained by using the demodulation process. The first step is a multiplication of the series $\{Z_t\}$ by a special function. Then in the second step we use a low-pass filter, that is, a filter which cuts off all but the frequencies near zero. The quality of demodulation depends of course upon the filter used. To carry out a demodulation on the frequency ω, $\delta<\omega<\pi-\delta$, δ being the lowest frequency not cut off by the filter, we generate the new series $\{P_t\}$ and $\{Q_t\}$ such that

$$P_t = Z_t \cos\omega t$$
$$Q_t = Z_t \sin\omega t, \quad t = 1,\ldots,N,$$

and apply a filter F, which leaves only the frequencies in the band $\langle 0,\delta \rangle$, on $\{P_t\}$ and $\{Q_t\}$. We obtain the series $\{R_t\}$ and $\{S_t\}$ such that

$$R_t = F(P_t)$$
$$S_t = F(Q_t) \quad t = k+1,k+2,\ldots,N-k,$$ k denotes the length of the filter F.

Then an estimator of the amplitude of the frequency ω at time t will be the function $2(R_t^2 + S_t^2)^{1/2}$ and an estimator of the phase at time t will be the function p(t) satisfying the relation tan(p(t)) = S(t)/R(t).

The choice of the filter is important in order to obtain reasonable results. We should like to use a filter as narrow as possible to have δ very close to zero, since we demodulate the whole band of frequencies $(\omega+\delta, \omega-\delta)$.The construction of the filter requires many observations, and this might be critical, since we are not always in abundance of data. Also, it is recommended in literature to use two successive filters, i.e. first to apply a filter F_1 on the whole series and then on the new (and shorter) series to apply a filter F_2. This procedure should diminish the leakage problems, i.e., a leak of a strong frequency into its surroundings and its multiples.

PROGRAMMES

There are the following programmes available on the volume RAIN (user MOLDAN):

- SPEDENCOR to carry out a spectral analysis of the data
- DEMOD to provide a demodulation on a given frequency ω
- ESTPER to estimate the coefficients of any particular frequency.

All the three programmes are based on the theory described above.

SPEDENCOR uses a straightforward algorithm to calculate the values of the periodograms in the points

$$\lambda_s = \frac{s}{2\pi N}, \quad s = 1,\dots,m, \; m = \frac{N-1}{2},$$

according to the relation (2), and the Parzen's estimator of the spectral density is of length 366. Its length should be between N/5 and N/6. To meet this consideration the variable MM (the length of Parzen's estimator) in the DATA statement is to be changed.

The test of significance of the 35 largest values of the periodogram, as described above, is included.

When the programme fails because of large amounts of data, the part calculating the periodogram can be omitted, or a double precision could be introduced for all variables. This will, however, cause a very long execution time. There exist better algorithms for this, for example the Fast Fourier Transform.

To use (MOLDAN) SPEDENCOR, the format statement 630 should be changed. The programme reads one sample a time into a vector, and then selects only the wanted item from the vector. It is indicated by the answer to "variable number". Also, the dimension of the variable VAL should be altered, accordingly. All arrays begin at index 1. SPEDENCOR can be used for data sets up to 2100 data points, if more is necessary, follow the instructions above.

Programme (MOLDAN)DEMOD uses one filter of the length 200, which is suitable for the data consisting of about 2000 samples. For them the δ equals roughly .03, so that this filter can be used for demodulation of waves in the frequencies ω from the interval $(.03, \pi-.03)$.

To use DEMOD the names of input (UNIT=2) and output (UNIT=3) files should be altered, as well as the input format 300 and the value of variable OM, which specifies the required frequency.

Programme (MOLDAN)ESTPER computes the least square estimates of the coefficients in equation (1) after the formulas described below (1). To use ESTPER, the dimensions of the data arrays should be altered, the names of the input (UNIT=2) and output (UNIT=3) files, and also the DATA statements LA (the required frequencies) and NLA (number of the required frequencies). The input format 10 should be rewritten.

The programmes DEMOD and ESTPER may be used for prewhitening or recolouring the spectra, as well as for a reconstruction of the particular waves or their linear combinations.

SOME RESULTS

The programmes described above were used on the time series of SO_2 and SO_4 in aerosol from Bear Island (26.7.1977-28.2.1983) and Birkenes (1.1.1977-30.12.1983). First, all available data were analysed, but the series from Birkenes, consisting of 2555 data points, appeared to be beyond the power of SPEDEN-COR. Therefore, and also for the sake of comparability of the results, the time interval available at Bear Island was also used for Birkenes. To obtain the values of the periodogram in the frequencies corresponding to the annual periodicity, only 5 years (1825 data points with the 29.2.1980 omitted) were analysed from Bear Island. As the series for SO_2 measurements are similar to SO_4 with respect to periodicity, only the results of SO_4 analysis are presented in the tables. Also, SO_2 measurements are less exact than the SO_4 ones, which causes larger values of the spectral density. The spectral densities for SO_4 and SO_2 from Birkenes and Bear Island are presented here.

In the interpretation of the results of the spectral decomposition there are several crucial points we should be aware of. Several frequencies appear strongly in the spectrum without any meaningful interpretation, only because they are multiples of some other strong frequency, observable by the means of the periodogram or not. This is visible in Table 2, where the results of the decomposition of the series of SO_4 from Bear Island are listed and grouped with respect to the exact multiplicity, that is, in one group there are all significantly high multiples of one frequency listed.

When the $\{X_s\}$ is not a periodical series described by (1), there may be some frequencies present in the spectrum of the $\{Z_s\}$ which are due to the form of $\{X_s\}$. That is because there is always at least one frequency present in the spectrum of most of the aperiodical models, (e.g. ARMA). The frequency band depends on the coefficients of the model considered.

Since the results of the significance test are not exact, only the ranks of the largest values of the periodogram are given. However, the tests seem to indicate that even the 20th strongest frequency is significantly stronger than the remaining ones on the significance level at least 10%.

THE FILES CONTAINING SOME OF THE RESULTS

On the volume RAIN (user MOLDAN):

SO2BEAR:SYMB SO4BEAR:SYMB SO2BIRK:SYMB SO4BIRK:SYMB	The results of the spectral analysis from the period July 27, 1977 - Feb. 28, 1983.

DSO4:SYMB - The periodogram of the SO4 Bear Island data obtained by using double precision arithmetics.

MM2:SYMB - The periodogram of the precipitation amount from Bear Island (the same period), all days included.

On user MOLDAN:

BIRMOD:SYMB BRMOD:SYMB	Demodulation of the SO_4 Birkenes series on two frequencies

SO2DENSPLOT:SYMB BEARSO2DENS:SYMB SO4DENSPLOT:SYMB BEARSO4DENS:SYMB	Files for FILESHOW, they contain the plots of the spectral densities from Birkenes and Bear Isl.

IVER:SYMB - Ny Ålesund series of SO_2, missing data are substituted by zeros

IVOUT:SYMB - Spectral analysis of the data from IVER

NYAALDENS:SYMB - File containing the plot of spectral density from IVER, to use with FILESHOW.

REFERENCES

(1) Andel, J. Statisticka analyza casovych rad. Praha, SNTL, 1976. (Statistical analysis of the time series).

(2) Granger, C.W.J. Hatanaka, M. Spectral analysis of economic time series. Princeton, New Jersey, Princeton Univ.. Press, 1964.

(3) Hannan, E.J. Multiple time series. New York, Wiley, 1976.

(4) Whittle, P. The Simultaneous Estimation of Time Series, Harmonic Compounds and Covariance Structure. Trabajos de Estadistica, 3, 43-57 (1952).

Time-series compendium:

Box, G.P.E. Jenkins, W.G. Time series analysis, forecasting and control. 2nd. rev. ed. San Francisco, Holden Day, 1976.

Table 1: Spectral analysis of the SO_4 July 26, 1977 - Feb. 28, 1983.

	Bear Island		Birkenes	
Rank	Period	Time (days)	Time (days)	Period
1	.0185	340.5	408.6	.0154
2	.0154	408.6	185.7	.0338
3	.0400	157.2	40.9	.1538
4	.0031	2043.0	19.3	.3260
5	.0215	291.9	18.9	.3321
6	.3260	19.3	16.6	.3783
7	.0707	88.8	340.5	.0185
8	.0953	65.9	2043.0	.0031
9	.0338	185.7	60.0	.1047
10	.0554	113.5	20.0	.3137
11	.1445	43.5	136.2	.0461
12	.0369	170.3	33.5	.1876
13	.3137	20.0	12.2	.5136
14	.1538	40.9	37.1	.1692
15	.0492	127.7	145.8	.0431
16	.3629	17.3	681.0	.0092
17	.1384	45.4	16.2	.3875
18	.6820	9.6	255.4	.0246
19	.1076	58.4	170.3	.0369

Table 2: Bear Island - Analysis of 5 years of SO_4 in aerosol.
February 1978-February 1983 (29.2.1980 removed) - 1825 analyses.
The rank of the value of the periodogram is given in brackets.
Time in days.

Groups:				
365.0(1)	182.5(2)	45.6(15		
365.0(1)	121.7(5)	17.4(9)		
365.0(1)	{38.8(8)}*	19.2(6)	{9.6(24)}*	
152.1(7)	86.9(3)	65.2(10)	43.5(13)	10.9(16)
Single frequencies:				
	1825 (5)			
	38.8(8)*			
	18.6(11)			
	41.5(12)			
	114.1(14)			
	45.6(15)			

*These frequencies are not exact multiples, but very near ones.

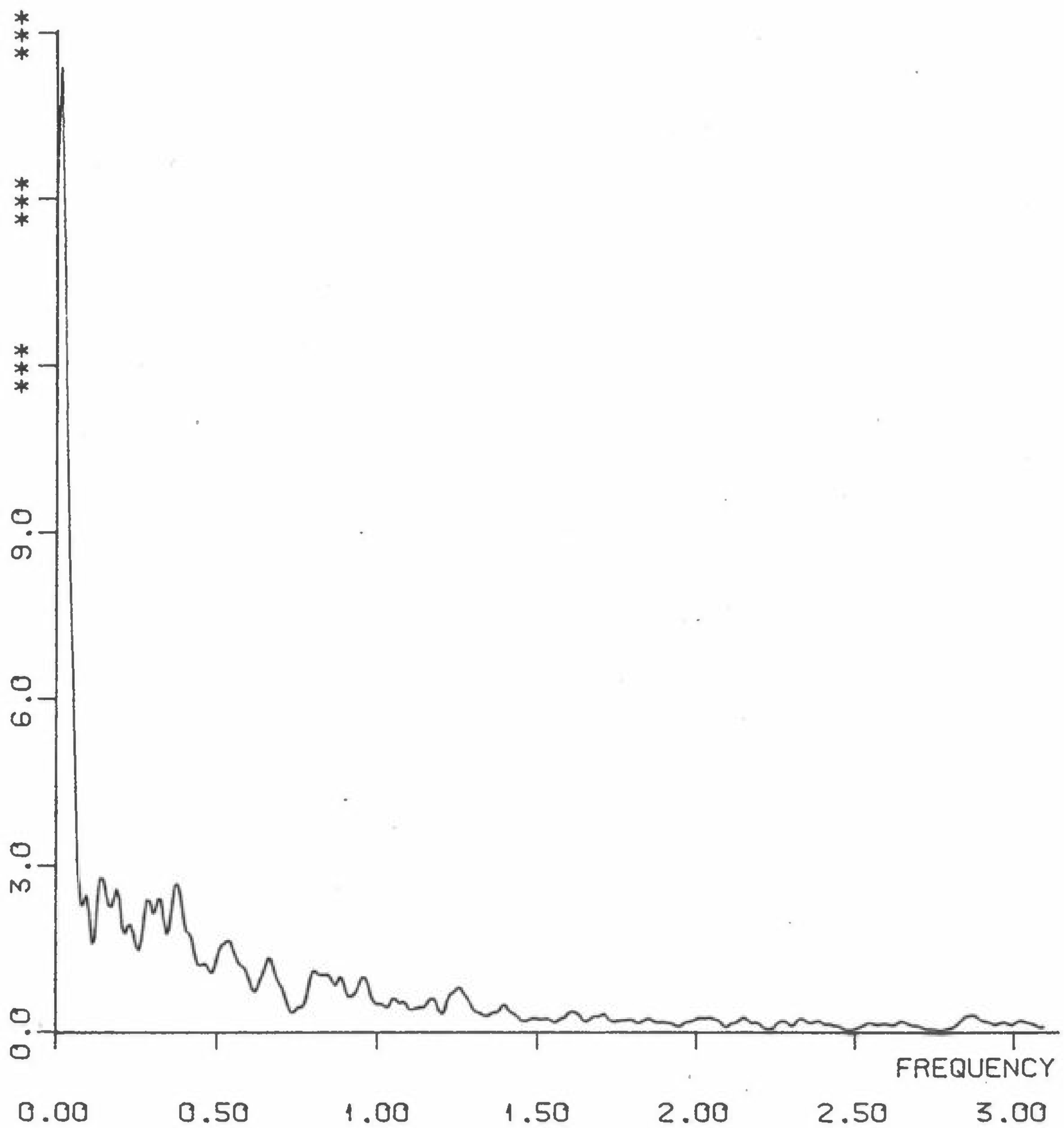

Figure 1: SO2 BIRKENES -PARZENS EST. OF SPECTRAL DENSITY. JULY 77-FEB 83

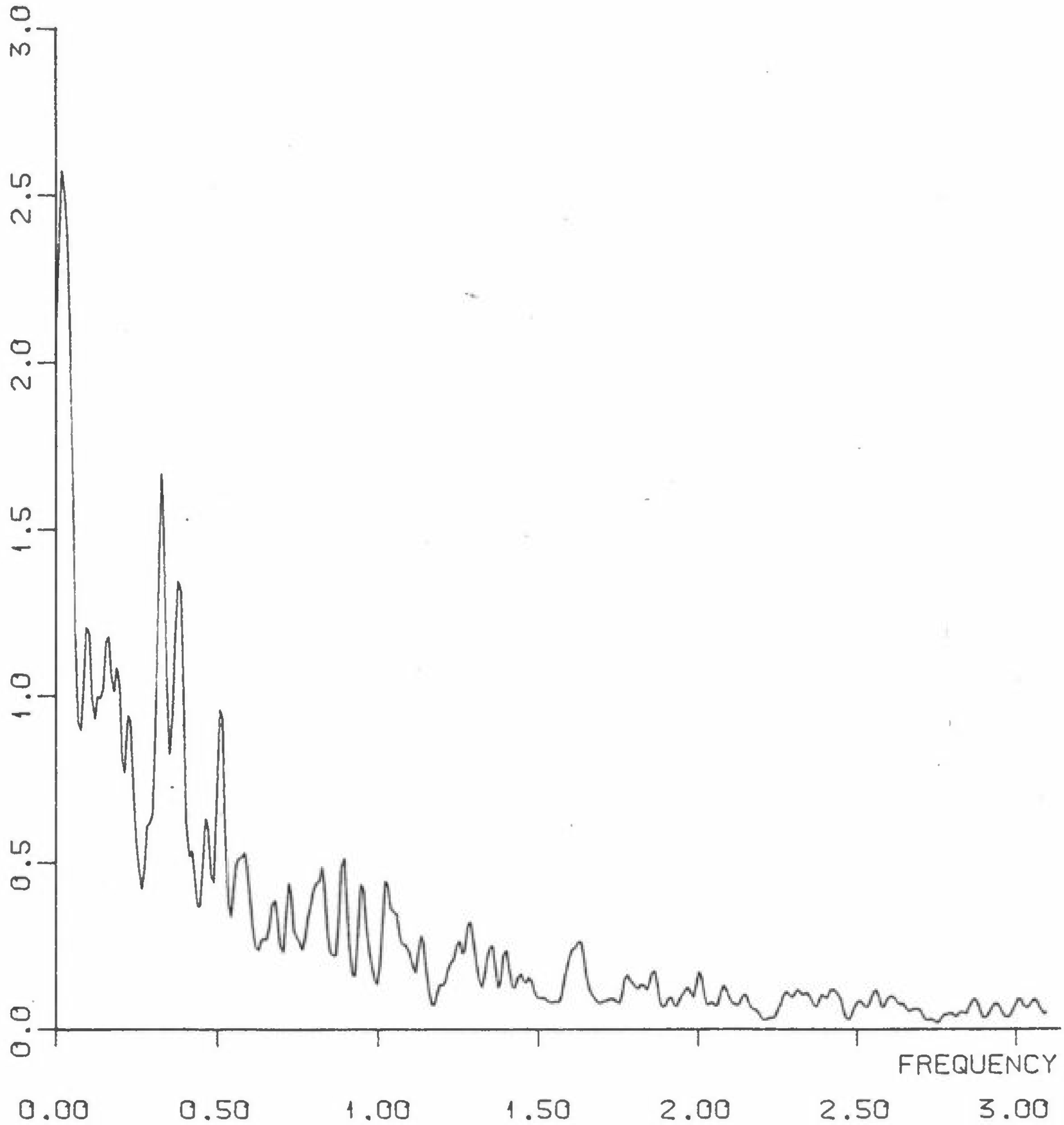

Figure 2: SO4 BIRKENES -PARZENS EST. OF SPECTRAL DENSITY, JULY 77-FEB 83

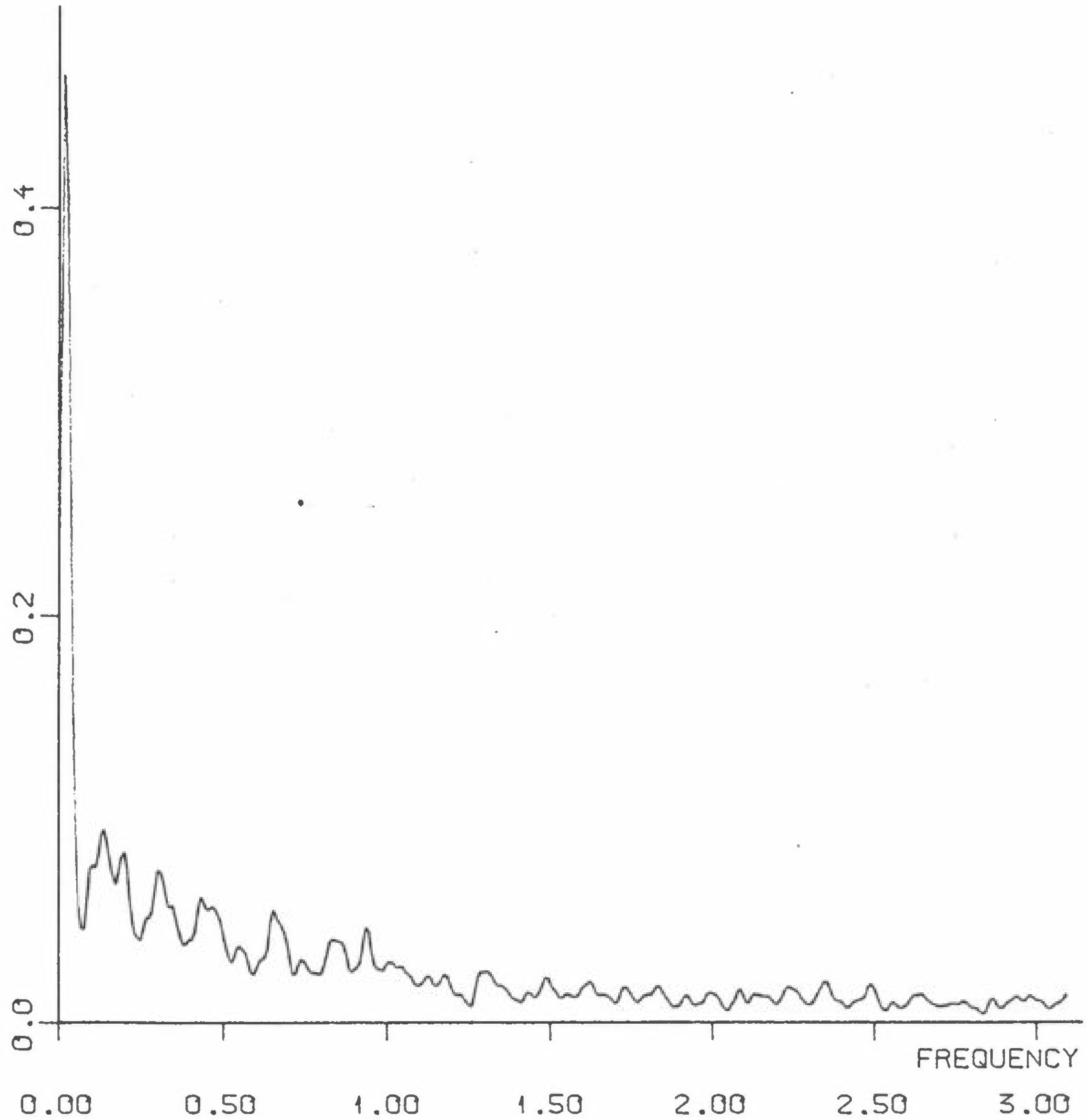

Figure 3: SO2 BEAR ISLAND-PARZENS EST. OF SPECTRAL DENSITY, 26.7.77-28.2.83

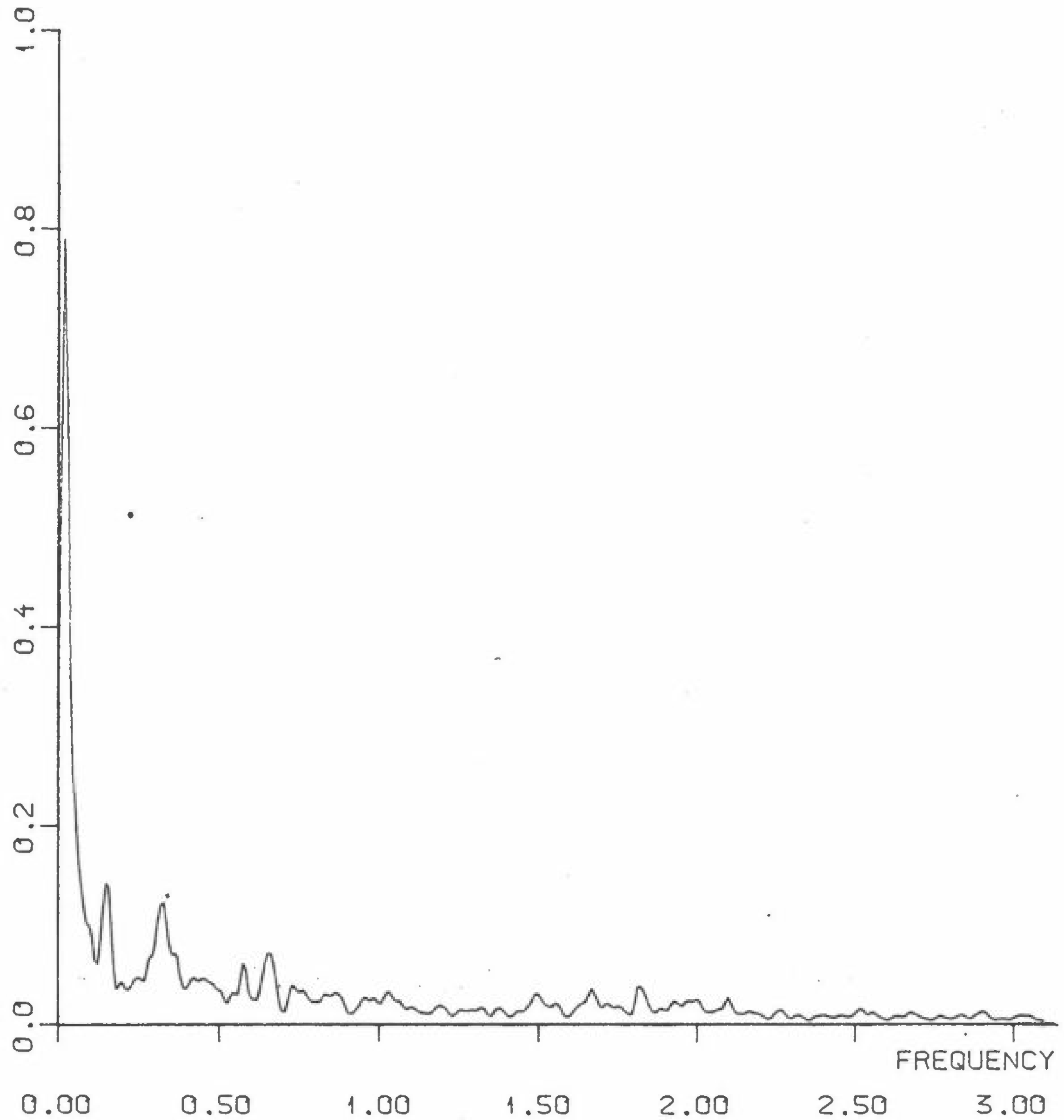

Figure 4: SO4 BEAR ISLAND-PARZENS EST. OF SPECTRAL DENSITY, 26.7.77-28.2.83

```
C  DEMOD - PROGRAM FOR DEMODULATION OF PERIODIC  PROCESS ON THE FREQ. OMEGA
       REAL X(2200),Y(2200),R(2200),S(2200),GAMA(2200),ME
C X.. INPUT FILE ,GAMMA..THE COMPUTED VALUE OF THE PROCESS
C Y.. OUTPUT FILE
C R.S,X,Y .. TO SAVE THE SPACE,WE DO NOT STORE THE ORIGINAL VALUES AND THE
C  FIELDS ARE REWRITTEN IN THE COURSE OF THE COMPUTATION. WHEN THE SECOND
C  FILTER IS INCLUDED INTO COMPUTATIONS, THE NAMES OF OUTPUT ARAYS SHOULD
C  BE ALSO ALTERED.
C OM.. VALUE OF THE MODELLED FREQUENCY, ME..MEAN
       DATA OM/.325/
       DATA PI/3.141552653592/
       OPEN (2,FILE='BIRKJEM')
       OPEN(3,FILE='BRMOD')
       ME=00.0
       I=1
100    READ(2,300,END=200) X(I)
       ME=ME+X(I)
       I=I+1
       GOTO 100
200    ME=ME/I
       N=I
       WRITE(3,*) 'MEAN',ME,'  N.OBS.',N,'   MODEL.FREQ.',OM
       DO 207 I=1,N
       R(I)=(X(I)-ME)*SIN(OM*I)
       S(I)=(X(I)-ME)*COS(OM*I)
207    CONTINUE
C THIS LENGTH OF THE FILTER IS SUITABLE FOR LARGE DATA SETS(F.EX. 2000)
C FILETR 1
       DO 202 J=200,N-200
       Y(J)=0.
       X(J)=0.
       DO 201 I=1,400
          Y(J)=Y(J)+R(J-200+I)/400
          X(J)=X(J)+S(J-200+I)/400
201    CONTINUE
202    CONTINUE
C FILTER 2
C        DO 203 J=5,N-5
C        R(J)=0.
C        S(J)=0.
C        DO 204 K=1,5
C          R(J)=R(J)+Y(J-3+K)/5
C          S(J)=S(J)+X(J-3+K)/5
C 204    CONTINUE
C 203    CONTINUE
C
C BE AWARE THAT THE RTANSFORM IS NOT EXACTLY ARCTANGENS (PI-SHIFT).
       WRITE(3,221)
       DO 220 I=200,N-200
       R(I)=2*SQRT(X(I)*X(I)+Y(I)*Y(I))
       IF(X(I)*Y(I).LE.0.) S(I)=ATAN(-X(I)/Y(I))-PI
       IF(Y(I)*X(I).GE.0.) S(I)=ATAN(-X(I)/Y(I))
       GAMA(I)=2*(Y(I)*SIN(OM*I)+X(I)*COS(OM*I))
       WRITE(3,222) I,R(I),S(I),GAMA(I),I
220    CONTINUE
221    FORMAT(3X,'T',8X,' A(T)',10X,'B(T)',6X,'GAMA(T)',)
222    FORMAT(I6,3F12.4,I4.0)
300    FORMAT(80X,/,50X,F10.3,30X)
       CLOSE(UNIT=2)
       CLOSE(UNIT=3)
       END
```

```
C     ESTPER - PROGRAM FOR ESTIMATING PARAMETERS OF PERIODIS PROCES
        REAL X(365),A(10),B(10),LA(10),Y
C X...DATA ARRAY,A..COEFS AT THE cos,B..COEFS AT THE sin IN THE EXPRESSION
C Z(t)=SUMA(Ai*COS(LAi)+ Bi*SIN(LAi)); MEAN IS SUBTRACTED FROM Z(t).
C    i IS INDEX
C LAi = LA(I)  ARE REQUIRED FREQUENCIES AS SPECIFIED IN DATA STATEMENT.
C UNIT 2..INPUT DATA, UNIT 3..OUTPUT DATA; NLA IS NUMBER OF REQUIRED FRQS.
C FORMAT STATEMENT IS LABELED 100.
        OPEN(2,FILE='SUMA')
        OPEN(3,FILE='TEOR')
        DATA LA/.01721,.03443,.05164,.32707,.36150,.13771,.34428,.15493,
     ..65414,.18936/
        DATA NLA/10/
        ME=0.0
        I=0
50      READ(2,10,END=100) Y
        I=I+1
        X(I)=Y
        ME=ME+Y
        GOTO 50
100     N=I
        ME=ME/N
        DO 140 J=1,NLA
        B(NLA)=0.0
140     A(NLA)=0.0
        DO 150 I=1,N
        DO 155 J=1,NLA
         A(J)=A(J)+COS(I*LA(J))*(X(I)-ME)
         B(J)=B(J)+SIN(I*LA(J))*(X(I)-ME)
155     CONTINUE
150     CONTINUE
        WRITE(3,*) ' KOEFS OF THE FUNCTION  SUMA(A(K)*COS(T*LAMBDA(K)) +
     . B(K)(SIN(T*LAMBDA(K)) '
        WRITE(3,*)'LAMBDAS',LA
        WRITE(3,*)'           A(K)                    B(K)'
        DO 160 J=1,NLA
         B(J)=2*B(J)/N
        A(J)=2*A(J)/N
        WRITE(3,*) A(J),B(J)
160     CONTINUE
        WRITE(3,*)
        WRITE(3,*) 'ESTIMATE OF THE OBSERVED VALUES'
        DO 170 J=1,N
        Z=0.0
        DO 175 K=1,NLA
        Z=Z+A(K)*COS(J*LA(K))+B(K)*SIN(J*LA(K))
175     CONTINUE
        WRITE(3,500) Z+ME
170     CONTINUE
500     FORMAT(F10.4)
10      FORMAT(F10.3)
        CLOSE(UNIT=2)
        CLOSE(UNIT=3)
C
C
        END
```

NORSK INSTITUTT FOR LUFTFORSKNING (NILU)
NORWEGIAN INSTITUTE FOR AIR RESEARCH

(NORGES TEKNISK-NATURVITENSKAPELIGE FORSKNINGSRÅD)

POSTBOKS 130, 2001 LILLESTRØM (ELVEGT. 52), NORGE

RAPPORTTYPE Teknisk rapport	RAPPORTNR. TR 11/84	ISBN-82-7247-488-3	
DATO SEPTEMBER 1984	ANSV. SIGN. Odd F. Skogvold	ANT. SIDER 20	PRIS kr. 20,-
TITTEL Spectral analysis of time series		PROSJEKTLEDER Jan Schaug NILU PROSJEKT NR. E-8414	
FORFATTER(E) Alena Moldanova		TILGJENGELIGHET A OPPDRAGSGIVERS REF.	
OPPDRAGSGIVER (NAVN OG ADRESSE) NILU			
3 STIKKORD (à maks. 20 anslag) Tidsserier	Spektral analyse	Bakgrunnskonsen-trasjoner	
REFERAT (maks. 300 anslag, 7 linjer) Rapporten gir en kort teoretisk bakgrunn for spektralanalysen med eksempler fra SO_2 og partikulært SO_4 fra Birkenes og Bjørnøya.			

TITLE Spectral analysis of time series
ABSTRACT (max. 300 characters, 7 lines) The report gives a short introduction to spectral analysis and presents some results from Birkenes and Bear Island.

* Kategorier: Åpen - kan bestilles fra NILU A
Må bestilles gjennom oppdragsgiver B
Kan ikke utleveres C