INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392-1

SMR.380/3

COLLEGE ON THEORETICAL AND EXPERIMENTAL RADIOPROPAGATION SCIENCE

6 - 24 February 1989

Statistical Analysis of Ionospheric Parameters Including Correlation Studies with other Geophysical and Solar Data

Alberto E. Giraldez

Lab. Ionosferico de La Armada (LIARA) -SENID
Argentina

STATISTICAL ANALYSIS OF IONOSPHERIC PARAMETERS
INCLUDING CORRELATION STUDIES WITH OTHER
GEOPHYSICAL AND SOLAR DATA

Alberto E. GIRALDEZ

Lab. Ionosferico de La ARMADA (LIARA)-SENID
Av. Libertador 327-1638 Vte. Lopez (Bs. As.)
ARGENTINA

INDEX

1989

SPECTRA VERSUS TIME DOMAIN ANALYSIS

Time Domain

Analysis of geophysical parameters is generally made in space-time domain, at last during the first steps.
The reason is that almost every geophysical observing system (active and passive), records a signal (voltage, current) as a function of time and /or geometric position in space.
During this first stage of data adquisition, there are two concepts which we have to bear in mind, namely:
-The final result obtained from the equipment is a set of values, preprocessed by the observing system itself with certain degree of complexity, which geophysisists refer to as EXPERIMENTAL VALUES. In fact those values are the result of applying to the received signal, a transformation process, based on physical models which assign a correspondence between the actual received signal and the associated geophysical parameter.
-A distinction is being made between continuous and discrete sets of data, in fact this is an arbitrary and nonsense distinction, as there does not exist any continuous observing system.
In the best case the data adquisition is made in a continuous manner (passive detectors), but the hardware has an integration time, and process the input signal in small blocks, giving a mean value for each integration time period. In fact the continuous-discrete data dicotomy should be aplied in a broader sense, according to the physical parameter under study, rather than to the record aspect.
(eg.1. An Ionosonde with a 15 min. sounding schedule is almost continuous for studies of seasonal variations, discrete for tidal studies and almost useless for short period gravity waves studies.
eg.2. A standard VHF-UHF Field Strenght receiver, with an integration time constant in the miliseconds range is a good instrument for studies of fading for voice and low speed digital data transmissions, but dangerous for fading studies of high density data transmissions.)

Depending on the objectives, the data should be handled in different ways. The first step, whichever the objectives were, is the data validation study.
This process enables the user to asses the validity and confidence of the data set. At this point radiopropagation studies and ionospheric physical studies take different roads. The radiopropagation field strenght data must be checked with most probable distribution tests, while ionospheric physical data is correlated with other solar and geophysical data.

Auto and Cross correlation analyses are going to be described in detail in the corresponding sections, being a good validation test for experimental data.

Spectral Domain

In the case you are looking for periodic structures in your data record, spectral analysis is a good one, but not the only processing tool you have at hand. (periodic structures can also be detected by autocorrelation and harmonic function fitting, as shall be explained later). Basically the spectral analysis is a set of operations performed on the data record, which depends on its characteristics.
The success or failure on getting good, physically meaning spectral results depends not only on applying a Discrete or Fast Fourier Transform (DFT and FFT respectively from now on) algoritm to a data record, but also on the filtering and windowing of the data previous to the FT operation, and on a judicious and critical study of the spectral results.
Spectral data processing provides us with additional information concerning uncovered periodicities, relative intensities of different periodic waves and power spectral density or energy distribution in the operation bandwidth. Simultaneous "ghost lines" are generated in the process, and ALIASING of waves outside the operation bandwidth can happen.
The warnings above mentioned have the intention to avoid the reader's disappointment after having a lot of time and effort in implementing an FT algoritm. Get the first results, and realize that nonconclusive or meaningless results were obtained.
Next section is devoted to FT use in detail, and the items already mentioned will be explained, in connection with geophysics. This lecture has not the intention of being an FT handbook, but the application of what the reader can learn on FT books to our specific field of interest.

FOURIER SERIES AND FOURIER TRANSFORMS

The usefulness of the F.T. is based on the FS definition and properties, which states that any continuous and nondivergent function extending from $-\infty$ to $+\infty$ can be decomposed in an infinite series of harmonic terms, thus being equivalent to the original function. Mathematically can be stated as:

$$S(f)=\int_{-\infty}^{+\infty} s(t).e^{-j2\pi ft}.dt \qquad (1)$$

s(t) = waveform to be decomposed into sum of sinusoids
S(f) = Fourier Transform of s(t)
$j=\sqrt{(-1)}$

From this definition, all the FT analysis is built up, but this lecture will not enter that field farther than what is needed to understand geophysical data processing.
As already mentioned, experimental data is discrete in nature then formula (1) is useless in our case.
For our purposes, its discrete equivalent is appropriate and is stated as:

$$S(f_k)=\sum_{i=0}^{N-1} (s(t_i).e^{(-j2\pi t_i.f_k)}.(t_{i+1}-t_i)) \qquad (2)$$

$$(k=0,1,2,...,N-1)$$

However, it is easily seen that if there are N data points in the series s(ti), and we wish to determine the N sinusoids approximation, there are N.N multiplications to be performed.
Under certain conditions, that number can be reduced to N.log(N) by the application of the Fast Fourier Transform algoritm (FFT).
For a clear and full description of FT theory, the book "The Fast Fourier Transform" written by Oran Bringham (Prentice Hall N.Y. 1974) is highly recommended.
In what follows, a series of fundamental properties of FT theory are stated, with a comment related to their meaning for geophysical data processing.

a) If g(t) is integrable in the sense

$$\int_{-\infty}^{+\infty}|g(t)|.dt < \infty \qquad (3)$$

then its Fourier Transform G(f) exists and satisfies the inverse FT.

b) Parseval's theorem

If $g(t)$ fulfills the a) condition, and the FT pair is defined as:

$$G(f)=\int_{-\infty}^{+\infty} g(t).e^{(-j2\pi ft)}.dt \equiv \text{F.T. of } g(t)$$

$$g(t)=\int_{-\infty}^{+\infty} G(f).e^{(+j2\pi ft)}.df \equiv \text{Antitransformed}$$

then: $$\int_{-\infty}^{+\infty} g(t)^2\, dt = \int_{-\infty}^{+\infty} |G(f)|^2\, df \qquad (4)$$

which means that the total energy is conserved on passing from one space to his transformed space. This is a fundamental relation that enables comparison of associated energy to different waves.

c) Linearity

$$F\{a(t)+b(t)\} = F\{a(t)\} + F\{b(t)\} \qquad (5)$$

where F indicates FT of the time series a(t) and b(t). This means that it is possible to filter, or separate independent processes in a given record.

d) Time Scaling

If k is a real constant greater than zero,

$$F\{g(k.t)\} = G(f/k)/|k| \qquad (6)$$

Frequency Scaling

$$F^{-1}\{G(kf)\} = g(t/k)/|k| \qquad (7)$$

This property has great importance in the windowing process, as contractions in one space produce expansion in the associated one, affecting the bandpass effectiveness.

e) Time Shifting

$$F\{g(t-to)\} = F\{g(t)\}.e^{(-j2\pi fto)} \qquad (8)$$

This property has important consequences in the interpretation of spectra results, as it means that two data records might have the same spectral lines pattern and amplitude, but different associated phases, due to a different starting time of the measurements respect to the phenomena under study.

f) Convolution

The convolution has an important role in FT theory. Its definition is:

$$c(t)=\int_{-\infty}^{+\infty} a(\tau).b(t-\tau).d\tau = a(t)*b(t) \qquad (9)$$

where c(t) is the convolution function of a(t) and b(t). The importance in this case is because the FT of a convolution product is directly the product of the FT of the functions:

$$F\{a(t)*b(t)\} = F\{a(t)\}.F\{b(t)\} = A(f).B(f) \qquad (10)$$

this is of key importance in the understanding of truncation effects on a time series and filter effects on spectra.

g) Correlation

The correlation integral,

$$c(t)=\int_{-\infty}^{+\infty} a(\tau).b(t+\tau).d\tau \qquad (11)$$

has identical properties to the convolution only in the case b(t) is an even function.

In general, the FT of the correlation is:

$$F\{\ c(t)\ \} = A(f).B^*(f) \qquad (12)$$

where $B^*(f)$ = complex conjugate of B(f)

Some geophysical observation techniques, as radar, performs on-line autocorrelation processing of the signal and after that FT of the result. This means in fact that the result is not a spectra but a power spectra of the signal. Correlation will be annalyzed in another section, out of the context of FT.

Fourier Series (FS)

A periodic function $g(t)$ with period To, expressed as a FS is given by:

$$g(t)=ao/2 + \sum_{n=1}^{\infty} [an.\cos(2\pi nt/To) + bn.\sin(2\pi nt/To)\ ] \qquad (13)$$

the coefficients are given by:

$$an=(2/To).\int_{-To/2}^{To/2} y(t).\cos(2\pi nt/To).dt \qquad (n=0,1,2,\ldots) \qquad (14)$$

$$bn=(2/To).\int_{-To/2}^{To/2} y(t).\sin(2\pi nt/To).dt \qquad (n=1,2,\ldots) \qquad (15)$$

or in complex notation:

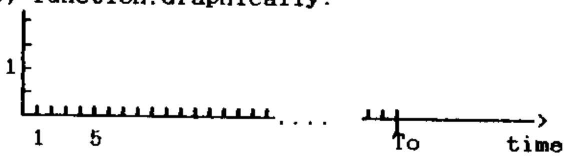
$$y(t)=\sum_{-\infty}^{+\infty} \alpha n.e^{j2\pi nt/To} \qquad (16)$$

where $\alpha n= (an-j.bn)/2$

$$\alpha n=(1/To).\int_{-To/2}^{To/2} y(t).e^{-j2\pi nt/To}.dt \qquad (17)$$

$$(n=0,\pm1,\pm2,\ldots)$$

This formalism is the most general mathematical formulation of FT , but yet there is a problem, it depends on integrals to be solved, and as we have time series of data (or space data record) which are not continuous and we have no analytic expression which describe them, no advance seems to have been made up to now. But a sudden jump ahead is made if we think each data point as the convolution of an impulse distribution function (Dirac Delta function) with a continuous function, which is unknown to us except at certain points, those where the δ is nonzero.
The following example uses the above mentioned ideas in a practical case.
Suppose you have a two days long record of foF2 parameter, taken from an ionosonde, with a 15 min. schedule rutine operation.
It means you have a set of 4 points/hour during 48 hrs, then 192 data points.
We shall call this the Total number of points (N).
The total duration of the record is 2 days or 48 hours or 2880 minutes, as you prefer, and this number will be called the duration or fundamental period of the record (To).
Now, we have to construct a Delta which becomes nonzero at the location of each data point of the "continuous" unknown foF2(t) function. Graphically:



We have to define the starting point, it can be called 0 or 1 dependindg on your preferences, remember we have N points, thus our counter should go from 0 to N-1 ,starting at 0 or from 1 to N ,starting at 1. We shall name the counter as 'n'. In this example we will use t=0 as starting point. Then the set of Delta functions is:

The first data is at	t=0 thus	$\delta(t)$	describes it
" second " "	t=1 "	$\delta(t-15)$	" "
third	t=2	$\delta(t-2x15)$	
fourth	t=3	$\delta(t-3x15)$	
....	...	...	
any	t=n	$\delta(t-nx15)$	
last	t=N-1	$\delta(t-(N-1)x15)$	

Here the time have been adopted in minutes ,and you see that 15 min=2 days/192 points or 15=Total Time (To)/N

The δ can be generally stated as $\delta(t-n.To/N)$
Then the set of values can be described as the convolution of a "continuous foF2(t)" and a $\delta(t-n.To/N)$.

Going back again to the general description of FT theory, it is easily demonstrated that:

$$F\{ \sum_{-\infty}^{+\infty} \delta(t-n.T) \} = (1/T).\sum_{-\infty}^{+\infty} \delta(f-n/T) \qquad (18)$$

and using the Delta function definition:

$$a(t).\delta(t-to)=a(to)\delta(t-to) \qquad (19)$$

we obtain :

$$g(t)=a(t)*b(t) \rightarrow G(f)=A(f).B(f) \qquad (20)$$

when b(t) is the Delta Function:

$$G(f)=A(f).(1/To).\sum_{-\infty}^{+\infty} \delta(f-n/To)$$

$$G(f)=(1/To).\sum_{-\infty}^{+\infty} A(n/To).\delta(f-n/To) \qquad (21)$$

The remaining step from the theoretical point of view, to get the discrete FT, is the demonstration that the FT of a continuous function is the same that of a set of samples of that function, provided that the sampling frequency is double than the higher frequency component in the function.

As stated in the Nyquist sampling rate. In case this condition is not fulfilled (the sampling frequency is smaller than the highest frequency in the record), the aliasing effect is observed. Aliasing is a redistribution of the higher frequency portion of the band into the available bandwidth, "aliasing" two or more spectral lines on a single "name" one. Graphically, the effect is as a folding of the full spectra in as many parts as necessary to fit into the observable bandwidth. This effect is always present, but is important only when the amount of energy associated with lines outside the observabled band is important.
In geophysics it is not always easy to assert the existence or not of considerable Aliasing in a given record. The best way to elucidate it is the following:

1)-Obtain the Fourier Spectra of the record.
2)-Reconstruct the "original" record using Inverse FT.
If the original and reconstructed records are equal, there is no Aliasing effect, thus no higher frequency lines are being missed in your adopted sampling frequency.
On the other hand, if they differ ,the reason is that the coefficients of the FT obtained are not purely corresponding to their associated frequency. They are contaminated (added) with other higher frequencies (outside the visible band) coefficients which have been aliassed with them in a folding process, thus breaking the relative coefficient weight equilibrium.

In order not to extend the text, the step by step procedure to demonstrate that DFT and IFT are equivalent, provided the Nyquist condition is fulfilled, will not be made. Now we go to the DFT expression and FFT asociated procedure.

Discrete Fourier Transform (DFT)

The DFT of a record containing N points, equally spaced, which we shall refer as x(k) (k=1,2,..N or k=0,1,2,...,N-1) is a series of harmonic terms whose coefficients indicate the weight of each harmonic component in the equivalent representation. Each term of the series will be represented as X(n). Then:

$$X(n)=(1/N).\sum_{k=0}^{N-1} x(k).e^{-j2\pi nk/N} \qquad (22)$$

(n=0,1,2,...,N-1)

It is very usefull at this stage to define and get familiar with the "buiding block" of FT, namely:

$$\text{a Phasor in complex space} \equiv \quad W=e^{-j2\pi/N} \qquad (23)$$

The rest of the terms are built by rotation of this phasor in n.k steps.
Formula (22) is the FT hart and its implementation in machine code is very easy. The expected results are:

$$ao= (1/N).\sum_{k=1}^{N} x(k)$$

$$an=(2/N).\sum_{k=1}^{N} x(k).\cos(2\pi nk/N)$$

$$bn=(2/N).\sum_{k=1}^{N} x(k).\sin(2\pi nk/N) \qquad (24)$$

as you all know. At this stage, we can introduce the FFT ideas and to define the similarities and differences with DFT.

Fast Fourier Transform (FFT)

As it is observed in the preceding formulas, for each an and bn N products between the data points and trigonometric functions must be made. Also N terms of the transform are being calculated thus N.N multiplications are needed, together with the 2.N sin and cos function values. This is a considerable number of operations to be performed, even for a fast computer.
The FFT algoritm becomes so popular just for its hability in reducing considerably the amount of multiplications .
Let us see how it is being done.
The best way of realizing it is by means of a practical example. Suppose we have a four points data record (x(0), x(1),x(2),x(3)) then :
N=4
W=exp(-j2π/4)
we define for simplicity -----> $W.\alpha.\beta=\exp[(-j2\pi/4).\alpha.\beta]$
the DFT is proportional to:

X(0)=x(0).W.0.0 + x(1).W.0.1 + x(2).W.0.2 + x(3).W.0.3
X(1)=x(0).W.1.0 + x(1).W.1.1 + x(2).W.1.2 + x(3).W.1.3
X(2)=x(0).W.2.0 + x(1).W.2.1 + x(2).W.2.2 + x(3).W.2.3
X(4)=x(0).W.3.0 + x(1).W.3.1 + x(2).W.3.2 + x(3).W.3.3

here we see the following:

1)the first term on the right is identical with the four eq. then three redundant multiplications are being performed.
2) W.0.0=W.0.1=W.0.2=W.0.3=W.1.0=W.2.0=W.3.0 = 1
$W.\alpha.\beta=W.\beta.\alpha$
3) as α and β are factors, the important value is their product result thus W.2.3 ≡ W6 and so on. we can see then that:

the four roots of 1 in the complex plane are:

```
        1  j  -1  -j  1  j  -1  -j  1  j  -1  -j  1
        W0
           W1 W2  W3
                      W4     W6
                                        W9
        |||               |||          |||            |||
X(0)  ▪
X(1)  ▪  ▪  ▪  ▪
X(2)  ▪     ▪      ▪      ▪
X(3)  ▪         ▪         ▪        ▪
```

This table shows in the upper line a cyclic repetition of the four possible rooth values (or harmonic values). The next four lines encolumns the values of the phasors which determine the fourier coefficients. A dotted field indicates starting of the rooths cycle. The last four lines show the number of cycles necessary to determine each FT coefficient ,and a small block is in column with the used value.

From this table and the set of equations of the preceding page,you can see :

x(2)W2=x(2)W6
x(2)W0=x(2)W4
x(3)W0=-x(3)W6
etc.

All these cases are reductions in the number of multiplications to be performed,as well as the repetition in the possible values of the harmonic functions.
Another point which si clear in the table is the sequence of W's utilization for building up the sequence of Fourier coefficients. The first (X(0)) uses only one , the next (X(1)) uses ALL the different values, the next SKIPS one value each time and the last SKIPS two values each time.
Some authors prefer to say that a permutation in the roots order must be done for each coefficient.
This permutation and reordering is what is mentioned as the scrambling process in the FFT algoritm.
This process is simple for data records of 2^n points length and is the standard FFT algoritm. For data points other than 2^n the algoritm is not as simple, also more time consuming but several algoritms have been developed for special cases. FFT algoritms are available as routine library in math. software. As Apendix A you will see a FT algoritm which uses some of the FFT properties to get FT of an arbitrary number of points, being thus an intermediate speed algoritm, written in Quick Basic.

3.-Truncation of Time Series

In the previous section we have seen that the limited length of the record introduces disturbances in the spectra, named ALIASING ,which is an inherent property of the DFT, due to its theoretical continuous and infinite character.
There exists also another disturbing effect due to the finite length of the record, in some sense similar to the aliasing, but this can be minimized by appropriate data handling. This is the windowing or filtering effect.

The aliasing, as already mentioned, is the folding of the infinite series of FT coefficcients over and over the working bandwidth. Its effect is important only when the data record sampling rate is "slower" than part of the existing waves.
These "fast" waves cannot be appropriately identified and are taken as part of slower ones. Remember that the FT theory is based on the assumptions that: a)The record lenght is equal to the fundamental wave period, and b)the sampling rate is fast enough as to pick up the highest frequency, according to the NYQUIST FREQUENCY.
This effect is then inherent to the theoretical derivation and cannot be eliminated.
The sequence of additions arising in a FT series corresponding to N data points is ,once obtained the (an) and (bn) Fourier coefficients :

```
F(f)                   first fold                second fold
 "                      ┌----->------------->----┐
ao            <----- aN.cos Nwf                   +
 +                      +
a1.cos wf   <-- a(N-1).cos (N-1)wf <-a(N+1).cos (N+1)wf
 +                      +                         +
a2.cos 2wf <-- a(N-2).cos (N-2)wf <-a(N+2).cos (N+2)wf
 +                      +                         +
 :                      :                         :
 :                      :                         :
 +                      +                         +
a(N/2-2)cos(N/2-2)wf <-- idem (N/2+2) <-- idem (N+N/2-2)
 +                      +                         +
a(N/2-1)cos(N/2-1)wf <-- idem (N/2+1) <-- idem (N+N/2-1)
 +                      +                         +
a(N/2)cos(N/2)wf        ^     <-------------- idem (N+N/2)
 ¦                                                ¦
 └-->-------------------┘                         └----->
```

* only cosine terms have been printed, sine terms must be added.

Filters and windows
Wheighting functions are ALWAYS used in the analysis of data records. These functions, depending on their particular application are commonly known as FILTERS or WINDOWS.
We shall present here only three of them, the most commonly used . They are : Rectangular (BoxCar) , Hanning and Kaiser-Bessel, describing their main application and time-frequency specifications.
A filter is a device that transmits a signal that is the result of covolving the input signal with the response function of the filter (h(t)). In the frequency domain this corresponds to a complex multiplication of the frequency spectrum of the signal by the frequency response function of the filter.
The most important filter characteristics are, in the frequency domain:

Centre frequency: Is defined as the arithmetic or geometric mean value of the lower and upper frequency limits.
geometric mean= √(fl.fu)=% constant bandwidth
arithmetic mean= fo= (fl+fu)/2 = constant bandwidth

Bandwidth: is defined as the width of the working frequency scale. The extreme values are defined by the lenght of the record (fl=1/T) and the sampling rate (fu= 1/2.Ts). The 3 dB bandwidth (Half power in amplitude) gives information about its hability to separate components of similar amplitude, and this determines the resolution of the analysis.

Ripple: The ammount of ripple in the passband of the filter, characterizes the uncertainty with wich the amplitude of a given signal can be determined.

Selectivity: Is a descriptor which indicates the ability of a filter to separate components of widely different levels. The basic parameter of selectivity is the shape factor. (The ratio of the filter bandwidth at an attenuation of 60 dB, to its 3 dB bandwidth.

Figures 1 and 2 show the above mentioned parameters.

Fig.1

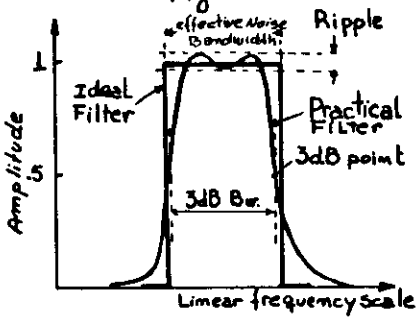
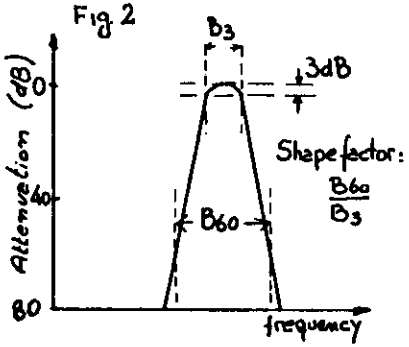


Fig.2



Windowing
FT analysis is made in blocks of data ie. each FT calculation is a transform of a record of finite length. The signal is thus limited to a window. What happens with the signal outside the window is unknown and does not enter explicitly in the transform calculation, but impose conditions to the results.
Individual window types will emphasize parts of the signal in different ways, thus giving different results (different spectra).
The three windows above mentioned are shown in Figure 3 with a table of its characteristics.

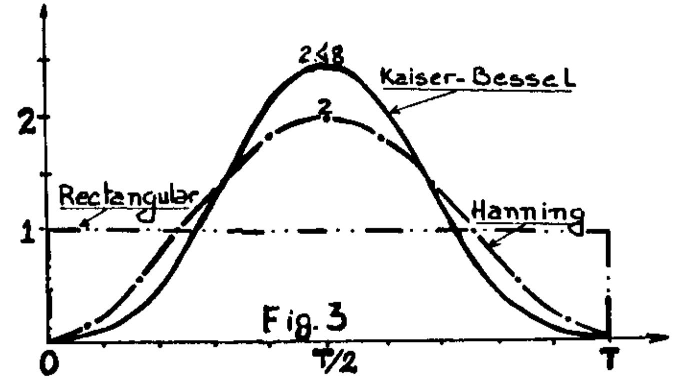


Fig. 3

Time Windows

	max. amp.	min. amp.	3dB limit
Rectangular	1	1	1.0 T
Hanning	2	0	0.5 T
Kaiser-Bessel	2.48	0	0.38 T

Frequency Windows

	3dB Bandwidth	Ripple	Highest Sidelobe	60dB Bandw.	Shape Factor
Rectangular	0.87 δf	3.92dB	-13.3dB	665 δf	750
Hanning	1.44 δf	1.42dB	-31.5dB	13.3 δf	9.2
K.-Bessel	1.71 δf	1.02dB	-66.6dB	6.1 δf	3.6

We shall see now those filters in detail.

Rectangular

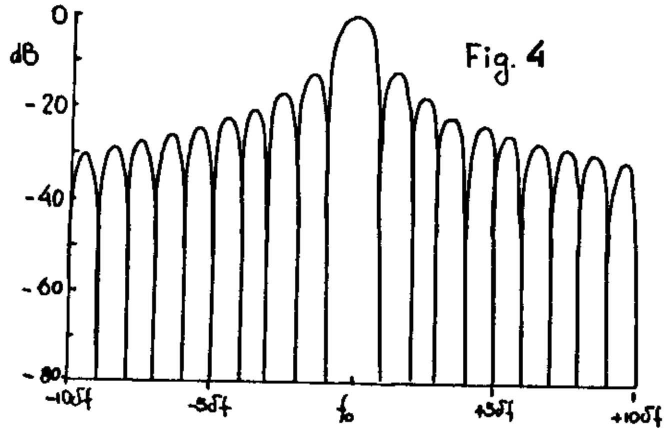
This filter/window ,also called Flat or BoxCar is in fact no weighting at all on the finite time record. It is defined as:

w(t)=1 for $0 \le t < T$ (T=record length) (25)
w(t)=0 elsewere

The filter characteristic given by the Integral FT of this window is:

$$W(jf)= (1- \exp(-jfT))/ (jf) \qquad (26)$$

Figure 4 is ist Fourier Transform.



The filter has a mainlobe which is twice the width of the line-filter spacing ($\delta f=2\pi n/T$) and an INFINITE number of sidelobes with widths equal to the line filter spacing. For the analysis of deterministic/harmonic signals this is a poor filter because it has :

1.- Very poor selectivity,due to the wide 60dB bandwidth.
2.-Relatively large (3.9dB) ripple in the passband,which means an amplitude variation of 2.38 times between maxima and minima in the working bandwidth (the "flat top" area).

At first sight,it seems that the BoxCar is a poor quality filter. This is not always true. For example,if you have a sinusoid ,which coincides with the central frequency of the filter (the window is exactly one period of the wave) you will get after transforming,the real maximum amplitude (filter factor = 1) and zero amplitude at all integer δf values,getting an exact result .Then this is a good choice. The worst case is when the frequency of the sinusoid coincides with a crossover frequency between two adjacent filters (eg. window lenght = 1½ wave periods). The result will be a decreased maximum value on 3.9dB,while all other sidelobes contribute with an appreciable power (leakage). The practical use of the rectangular window is for analyzing transients with shorter duration than the record lenght (T).

Due to the flat and unit amplification factor in the time domain,all parts of the signal are equally weighted. In the frequency domain the bandwidth of the signal is greater than the bandwidth of the filter,because the signal is shorter than T , and therefore the filter characteristics will have no influence on the calculated spectrum of the transient signal.

Hanning weighting

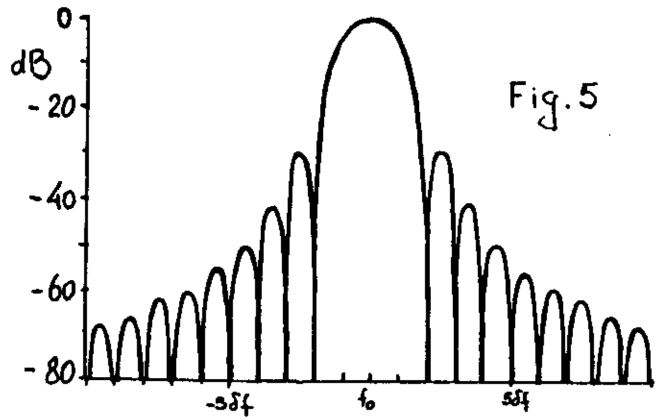
This filter already shown in a previous Figure,is a smooth function defined as:

$w(t)= [1 - \cos (2\pi t/T)] = 2.\sin^2(\pi t/T)$ for $0 \le t < T$
w(t)= 0 elsewere (27)

its FT expression is:

$$W(f)= \sin(\pi fT)/(\pi f) \qquad (28)$$

Figure 5 is its F.T.



The mainlobe is 4δf,double the width of the rectangular window. Sidelobes are more attenuated,and the fall off rate is much faster than for BoxCar weighting. This means that the 60dB bandwidth is much narrower,giving better selectivity. The ripple is only 1.4 dB.

Defining the noise bandwidth as the equivalent BoxCar window (total area in space-time) ,Hanning window has 50% greater noise bandwidth than BoxCar. Then, power spectrum values for broadband random signals will therefore be 1.5 times higher when analyzed using Hanning instead of Rectangular filters. Leakage for a single sinusoid as described in the previous filter text, is greately reduced. The Hanning window thus performa better than rectangular one with respect to selectivity, passband ripple and leakage, and should be used in most cases where continuous signals are analyzed.

Kaiser-Bessel weighting

This filter responds to the expression:

$$w(t)=1-1.24\cos(2\pi t/T)+0.244\cos(4\pi t/T)-0.00305\cos(6\pi t/T) \quad \text{for } 0 \le t < T$$
$$w(t) = 0 \text{ elsewhere} \qquad (29)$$

The FT of this filter is shown in Figure 6.

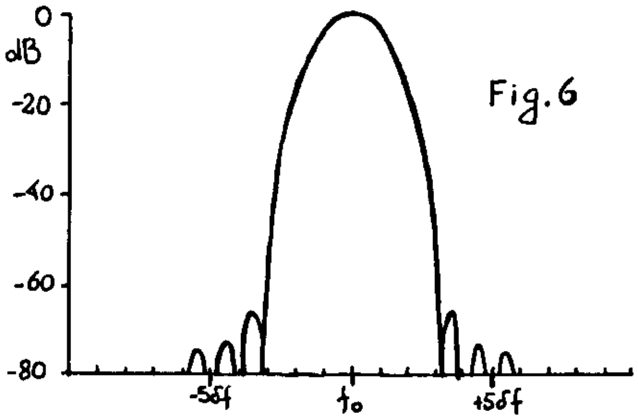


Fig.6

As it is observed, this is superior to other filters in selectivity performance. The 60 dB bandwidth is only 6.1 times the line spacing ,due to the extremely low level of the highest sidelobe, which is found to be at -67 dB. For harmonic signals analysis, the only difference between "best case" and "worst case" is its maximum amplitude error (ripple in the passband of -1 dB), presenting almost no leakage, due to the sharp sidelobes reduction. Because of its good selectivity, should be used as window when separation of closely spaced frequency components with widely different amplitude levels is required.

For analysis of periodic signals, the Kaiser-Bessel window is probably the best choice. The disadvantage, in comparison with Hanning is speed and that a uniform weighting of the time signal cannot be achieved by standard overlap analysis, in aplications which require real time processing. Also has a wider noise bandwidth (1.8 δf), and overestimation of random signals gives rise to enhanced uniform leakage at resonances as well as at antiresonances .

Autoregressive Methods

Autoregressive methods are the statistical face of data processing. They are closely related with spectral methods, and that association will be shown in the following paragraphs. A detailed analysis of this topic can be found in "Spectral Analysis in Geophysics" M.Bäth, Elsevier Sci. 1974. The Autocorrelation function is defined as:

$$C11(\tau) = \int_{-\infty}^{\infty} f1(t).f1(t+\tau).dt \qquad (30)$$

It has its maximum for $\tau=0$. This function, so defined for the one-dimensional case can be extended to multi-dimensional cases as:

$$C11(\alpha,\beta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f1(x,y).f1(x+\alpha,y+\beta).dxdy \qquad (31)$$

The cross-correlation function is defined as:

$$C12(\tau) = \int_{-\infty}^{\infty} f1(t).f2(t+\tau).dt \qquad (32)$$

The generalization for more than one dimension is similar to the auto-correlation case. Here f1(t), f2(t), etc. are real functions if we assign them to geophysical data values. Their use in the "discrete data world" of geophysical data records does not present any problem in their discretization. Simply the integrals are replaced by additions. The problem is that the numerical values are not normalized and the Auto and Cross correlation values will depend on the record length. The solution is to redefine them as normalized correlations by dividing the above expressions by their value at $\tau = 0$. This way the maximum Cij value is one, and different record lengths can be compared .
At this point I want to make a distinction in between two different uses of correlation expressions.
The first and immediate one is to use these expressions to determine, in the auto-correlation case, if the record has periodicities. Starting at $\tau = 0$ (C11(0) = 1) , the coefficient will start decreasing as τ grows. If it grows again and reaches a relative maxima this indicates the record has periodic structures, and the fundamental period is the τ value for the maxima. It might happen that several τ values (not equally spaced) present peaks, then harmonic analysis will help in determining the spectral structure, with an apriori information of the expected waves frequencies.

The cross-correlation uses two different time (or space) series with arbitrary relative shifts. For each shift value (τ) there is a corresponding value of the correlation. This function is very usefull in determining the paralelism between two time series. This has a great practical significance in any geophysical observation of propagating waves by means of an array of stations. The time shift which maximizes the cross-correlation would correspond to the most likely phase shift between correlated stations.

For several geophysical phenomena the auto-correlation often decreases exponentially with increasing time lag. This behaviour is connected with Markovian processes, and once you get this result, without even doing the spectral analysis, be sure that the Power Spectral Density of the record has a frequency dependence of the type $f\exp(-3)$. This is connected with the other use of correlation expressions mentioned in the previous page.

This second use is connected with the association in between correlation functions and the corresponding Power Spectra of the working time series. The demonstrations of the expressions which follow will not be given, but you can find them in the books already recommended.

The following simbols will be used:
f1(t) , f2(t) = time series
F1(f) , F2(f) = Fourier Transforms of the time series
C11(τ) = Auto-Correlation function of f1(t) (lag=τ)
C12(τ) =Cross-Correlation of f1(t) with f2(t)
E11(f) = $|F1(f)|^2$ = Power spectrum of f1(t)
E12(f) = $F1(f)^*.F2(f)$ =Cross Power spectrum f1(t)and f2(t)
* = the star symbol indicates convolution

List of correlation properties and associations between correlation functions and F.T.:

$$C11(\tau) = C11(-\tau) = f1(\tau)*f1(-\tau) \qquad (33)$$
$$C12(\tau) = C21(-\tau) = f1(-\tau)*f2(\tau)$$
E11(f) is the F.T. of C11(τ) or: C11(τ) <-> E11(f)
C12(τ) <-> E12(f)
$$E21(f) = E12(-f) = E21^*(-f)$$

Unlike the Power spectrum E11, which is always real and positive, the Cross-Power E12 is in general complex. Then |E12(f)| can be used as a measure of the cross power. It is usefull to split E12(f) into its co-spectrum (real part) and quadrature spectrum (imaginary part) :

$$E12(f) = P12(f) - j.Q12(f) = F1^*(f)\ F2(f) \qquad (34)$$

where $P12(f)$ is the co-spectrum and $Q12(f)$ is the quadrature spectrum (abreviated as quad-spectrum).
It is easy to demonstrate that:

$P12(f) = P21(f)$

$Q12(f) = -Q21(f)$

$$P12(f) = \frac{1}{2} \int [\, C12(\tau) + C21(\tau) \,] \cos f\tau \, d\tau \qquad (35)$$

$$Q12(f) = \frac{1}{2} \int [\, C12(\tau) - C21(\tau) \,] \sin f\tau \, d\tau \qquad (36)$$

The phase lag ϕ of F1 with respect to F2 is obtained from:

$$\tan \phi = -Q12(f)/P12(f) \qquad (37)$$

By analogy, in the case of autopower:

$E11(f) = P11(f)$

$E22(f) = P22(f)$

$Q11(f) = Q22(f) = 0$

Another definition which is always used is : Coherence

$$|\, \gamma 12(f) \,| = |\, E12(f) \,| / \sqrt{[\, E11(f)\, E22(f) \,]} \qquad (38)$$

Theoretically, the coherence function would be equal to 1, independent of frequency. It is not so in practical applications due to the windowing and smoothing effects.

As mentioned in this and the previous page, the association between correlation functions and power spectra is the second aspect of correlation use. As they are an associated Fourier pair, if you have a graphical idea of simple functions and their associated FT, the Auto-Regressive (AR) aspect of a data record gives you an idea of the Fourier power spectra you will obtain, and also of the physical phenomena involved.

We shall see this with a few examples:

1) You have a data record with no evident periodicities, and performs its auto-correlation and draws it on the computer screen. The figure shows a single peak at $\tau=0$ and a sharp decrease without further secondary peaks for other τ values (exponential decrease).
You can immediately deduce that the power spectra will have no defined lines, the shape will be "bell type" with a main slope of f exp(-2).
This is the already mentioned case of Markov process. If you calculate the Power Spectral Density ($F(f)/f$), and draws it in log-log scale, a line (slope -3) will describe that spectra (the well known Garret-Munk spectra).
Reasons: wave instability saturation-Resonant interactions (RIA) or associated to other physical reason.

2) The same procedure than before, and you see at your screen a ($\sin \alpha / \alpha$) figure. Successive maxima, separated by deep valleys.
In the spectral domain it means a constant signal level in all the bandwidth (white noise, even in a limited band). Then your record does not corresponds to discrete waves travelling, but to a uniform distribution of energy all over the band.
Reasons: noise level saturation-Data record excessively short- insufficient sampling rate speed-or many other reasons.

3) At your screen you see a wavelike pattern with successive maxima and minima . If maxima are equally spaced, in the spectral domain you have a single wave of period τ and frequency $1/\tau$. If maxima are not equally spaced, a set of waves are present in the record.

These are extreme cases (nature is not simple), and mixed cases are common, but with this procedure you have a starting point and some idea of the record content.

It must be remembered that using this method, the FT is the Power Spectra of the record, and the phase information have been lost. If we are interested both in power and phase of a given record transform, the FT of the original record (filtered, windowed etc) must be obtained, and the power calculated in the standard way ($a^2 + b^2$).

Reliability of Harmonic Analysis

As we have seen ,harmonic analysis is a valuable tool in geophysics,we can get a lot of information concerning the behaviour of the parameter under study from its spectrum. Now it is time to mention the limitations,concerning geophysical applications.

Most geophysical processes are non-stationary ,and also the media is dispersive (earth,atmosphere,ionosphere). Thus,a single harmonic wave produced in such a media,will propagate and suffer the dispersive effects.These effects are mainly the change of amplitude and speed.The observation of that wave at different locations or at the same place and during a period of time provides us with the required data record.Easily we can deduce that our data record,supposing it contains information of a single harmonic source will give different results depending the location, direction and variability of the media during the measuring period.

The recording instrument bandwidth and data sampling rate acts as filters on the recorded information,distorting its content as already discussed.

The spectra of such a record ,even being a "single harmonic" source,will be considerably distorted.Reconstructions of the signal can be attempted,provided the filter characteristics are well known.

The above discussed case is not commonly encountered. Single harmonic sources in nature are infrequent.The most commom case is that the source is non stationary (impulsive,short duration and band limited rather than single oscillation).

In this case,the dispersive characteristics of the media not only produce attenuation and speed variations,but also filtering of the signal.

This last effect can be described as follows.The source produce a disturbance (pressure,density temperature ,etc) with a given structure in time (sudden peak,gaussian,or any shape you can imagine).As stated in Fourier theory,it is possible to reconstruct that original shape (at the source location) with a Fourier series.

As the disturbance propagates,EACH of its Fourier original components suffer a DIFFERENT attenuation and velocity changes.The result is that the original phase relations among the different terms changes continuously ,as well as the relative importance of the Fourier coefficient.Putting all this together we can say that the original waveform is in permanent change.A record of that disturbance constructed at a distant point,during a time period,"freezes" the evolution at each data point,but not in between data points.The resulting record is a sequence of different geophysical situations related by the time evolution of the source and the media,or by the time evolution of the source and the variable distance to the source.

When using FT methods for geophysical data records study,as in our case,we have to bear in mind that a rigurous stationary theory is being used for a non-stationary process study.

FT method assumes that there are not frequency changes along the sample.As this is not an infrequent case in geophysics,a good practice is to work with a sliding window over the record.This means to make partitions of the record, and calculate FT of each part.Those partitions can partially overlap,then we obtain a series of FT showing the evolution of the frequency structure with time. This is called moving-window spectra.

Two comments on this practical approach.first,we have to be sure that the partition is sufficiently wide as to contain the longer period we are looking for at least one time,and second , From the FT properties we saw that time shifting does not alter the power spectra,but the phase is altered considerably,then phase comparison between different spectra is meaningless,except if partition is made in such a way that all of them start with the same phase.(in general this can be accomplished for only one frequency of the set).

ANNEX A

(1)

```
10 REM -------NOMBRE DEL PROGRAMA : "FT-IIQB.BAS "-------------------
40  PRINT "NUMERO DE DATOS"
50 INPUT m: H = m
80 z = INT(m / 2) + 1
90 DIM x(m + 1), u(30)
   DIM Yu1(m + 1), AU1(m + 1), BU1(m + 1), axu1(z), BSu1(z)
   DIM Yu2(m + 1), AU2(m + 1), BU2(m + 1), Axu2(z), BSu2(z)
100 DIM WC(m + 1), WS(m + 1)
    DIM AIu1(z), BIu1(z), CU1(z), pu(z), du(m + 1)
    DIM AIu2(z), BIu2(z), Cu2(z)
   DIM u1(250), u2(250)

'       ********* PIDE DATOS PARA EL CALCULO ***********
'----------------------------------------------------------------------

    INPUT "nombre del archivo 1 y extension"; NOMBRE$
    j = 0: sumhm = 0: sumh0 = 0: sumu = 0
    OPEN "i", #1, (NOMBRE$)
    WHILE NOT EOF(1)
    INPUT #1, u
       u(j) = u: sumu = sumu + u

       x(j) = j
       j = j + 1
    WEND
'    m = j - 1
    CLOSE #1
    medu1 = sumu / m



310 REM *********** CALCULA PERIODO/LONG ONDA/INTRVALO DE MUESTRAS**********
'       ************CALCULA COEFICIENTES FOURIER DE PRIMERA VUELTA**********
    CLS
340 ZZ = 2 * 3.14159 / x(m - 1)
       FOR r = 0 TO m - 1
          WC(r) = COS(x(r) * ZZ)
          WS(r) = SIN(x(r) * ZZ)
       NEXT r
'   ------DETERMINA ALGORITMO EN BASE A NRO.DE DATOS PAR O IMPAR --------
                          n = m / 2
                 IF INT(n) = n THEN SA$ = "PAR"
              IF SA$ = "PAR" THEN GOTO 820
'---------- CALCULO PARA NRO.IMPAR DE DATOS -------------------------

                    n = (m - 1) / 2
         FOR k = 0 TO n
           axu1(k) = 0: BSu1(k) = 0'BORRA COEFICIENTES DE CORRIDAS ANTERIORES
```

(2)

```
          axu1(k) = 0: BSu1(k) = 0'BORRA COEFICIENTES DE CORRIDAS ANTERIORES
         NEXT k
       FOR k = 1 TO n
          FOR r = 0 + CO TO H - 1 - FI
               j = k * x(r - CO) - INT(k * x(r - CO) / x(m - 1)) * x(m - 1)
                  Y1(r) = u1(r) - aedu1
                    AIu1(k) = Y1(r) * WC(j)
                     axu1(k) = axu1(k) + AIu1(k)
                    BIu1(k) = Y1(r) * WS(j)
                     BSu1(k) = BSu1(k) + BIu1(k)
640       NEXT r
650 REM ***********CALCULA COEF.FOURIER NRO IMPAR DE DATOS*****************
               AU1(k) = axu1(k) / (2 * n + 1)
               BU1(k) = BSu1(k) / (2 * n + 1)
               CU1(k) = SQR(AU1(k) ^ 2 + BU1(k) ^ 2)
720          IF k > 1 THEN GOTO 750'AQUI DEFINE PUNTO DE PARTIDA
                        BAJu = CU1(1):  MAXu = CU1(1)
750                 IF CU1(k) < BAJu THEN BAJu = CU1(k)
              IF CU1(k) > MAXu THEN MAXu = CU1(k)

800     NEXT k
810 GOTO 1150


820 '---------------CALCULO PARA NRO.PAR DE DATOS ---------------------
830 FOR k = 0 TO n
    axu1(k) = 0: BSu1(k) = 0
    Axu2(k) = 0: BSu2(k) = 0
850 NEXT k                        'BORRA COEFICIENTES EXISTENTES EN MEMORIA
860 FOR k = 1 TO n
870     FOR r = 0 + CO TO H - 1 - FI
880          j = k * x(r - CO) - INT(k * x(r - CO) / x(m - 1)) * x(m - 1)
                Yu1(r) = u1(r) - aedu1
                Yu2(r) = u2(r) - aedu2
             AIu1(k) = Yu1(r) * WC(j)
             AIu2(k) = Yu2(r) * WC(j)
             axu1(k) = axu1(k) + AIu1(k)
             Axu2(k) = Axu2(k) + AIu2(k)
               BIu1(k) = Yu1(r) * WS(j)
               BIu2(k) = Yu2(r) * WS(j)
               BSu1(k) = BSu1(k) + BIu1(k)
               BSu2(k) = BSu2(k) + BIu2(k)
990     NEXT r
1000  REM ********* CALCULA COEF.FOURIER PARA NRO. PAR DE DATOS ****
1010            AU1(k) = axu1(k) / m
                AU2(k) = Axu2(k) / m
1020            BU1(k) = BSu1(k) / m
                BU2(k) = BSu2(k) / m
1030              CU1(k) = SQR(AU1(k) ^ 2 + BU1(k) ^ 2)
                  Cu2(k) = SQR(AU2(k) ^ 2 + BU2(k) ^ 2)
```

(3)

```
1070 IF k > 1 THEN GOTO 1100
1080  BAJu = CU1(1): MAXu = CU1(1)
1100 IF CU1(k) < BAJu THEN BAJu = CU1(k)
1110 IF CU1(k) > MAXu THEN MAXu = CU1(k)
1140 NEXT k
1150 REM ******** DETERMINA VALORES MAXIMO Y MINIMO **********
1190 DELAMPu = INT(MAXu) - INT(BAJu) + 2
1200  ventyu = DELAMPu
1210 FACTIZQu = ventyu / DELAMPu: factderu = ventyu / DELAMPu
1220 REM ********** DIBUJA ESPECTRO EN PANTALLA ********************
1230  CLS :    SCREEN 2
1240 WINDOW (0, 0)-(m, 1.2 * ventyu)
1250 LINE (0, 0)-(z - 1, ventyh), , B: LINE (z + 1, 0)-(m, ventyh), , B
     FOR j = 0 TO n - 1 STEP 10
     LINE (j, 0)-(j, ventyu / 20): LINE (j + z, 0)-(j + z, ventyu / 20)
     NEXT j
1270  FOR k = 1 TO n - 1
1300 LINE (k, (SQR(CU1(k)) - BAJu) * FACTIZQu)-((k + 1), (SQR(CU1(k + 1)) - BAJu) * FACTIZQu)
     LINE (k + z, (SQR(Cu2(k)) - BAJu) * factderu)-((k + 1 + z), (SQR(Cu2(k + 1)) - BAJu) * factderu)
1320 NEXT k
1330 LOCATE 1, 1: PRINT "NRO DATOS ="; m, "nombre file ="; NOMBRE$
1340 LOCATE 3, 5: PRINT "POWER SPECTRA"
1350 LOCATE 3, 43: PRINT "POWER SPECTRAL DENSITY"
1360  GOTO 1590        'LPRINT "QUIERE TABLA IMPRESA? (S/N)"
1370 GA$ = ""
1380 WHILE GA$ <> "S" AND GA$ <> "N": GA$ = INKEY$: WEND
1390 IF GA$ = "N" THEN GOTO 1430
1400 FOR k = 0 TO n
1410 LPRINT "A("; k; ")="; a(k), "B("; k; ")="; B(k), "C("; k; ")="; C(k)
1420 NEXT k
1430 REM ********** PONE LOS FLAGS EN NEUTRO Y CONTROLA FLUJO ********
1440 GA$ = "": LPRINT "USA LOS MISMOS DATOS? (S/N)"
1450 WHILE GA$ <> "S" AND GA$ <> "N": GA$ = INKEY$: WEND
1460 IF GA$ = "N" THEN GOTO 1580
1470 FOR I = 0 TO H - 1
1480 Y(I) = D(I)
1490 NEXT I
1500 LPRINT "NRO TOT.DE DATOS "; H, "NRO DATOS ANTERIOR "; m
1510 LPRINT "NRO DATOS QUE ELIMINA AL COMIENZO"
1520 INPUT CO
1530 LPRINT "NRO DATOS QUE ELIMINA AL FINAL   "
1540 INPUT FI
1550 m = H - CO - FI
1560 LPRINT "NTRO DATOS ACTUALES "; m
1570 GOTO 340
1580 SCREEN 0
1590 LOCATE 2, 54: INPUT "GRABA (S/N)"; GRAB$
1600 IF GRAB$ = "N" THEN GOTO 1730
1610 '

'OPEN "O", #1, (NOMBRE$)


'               NEXT j
'CLOSE #1

1730 END
```