APPENDIX C

TIME SERIES ANALYSIS METHODS

Da tempo e mare non s'impara niente.[1]
(Guccini 1983)

C.1. FOURIER-TYPE METHODS

Within the work made during this PhD course, in some cases (see e.g. Chs. 4, 5 and 6) the timing analysis of data series was needed in order to investigate the presence of periodicities hidden in them. In this Appendix the methods applied to this task in the present Thesis work will be described. The algorithms which were used can be divided into two broad classes: in this Section the methods related to the Fourer Transforms will be presented, while in the next one the methods applying the least-squares fitting technique will be described. It is to be recalled that, in the following, only a synthetic description of all these methods will be given: for a more complete treatment, which is in any case beyond the scopes of this Thesis, the reader is referred to the original papers.

C.1.1. The Discrete Fourier Transform (DFT)

The Discrete Fourier Transform, or DFT, method allows one to analyze the presence of sinusoidal periodicities, contained within a data series, thanks to the study of the power spectrum obtained from the application of the Fourier Transforms on the data themselves.

According to the theory, however, it is not possible to obtain the DFT of a function $f(t)$ sampled in a finite number N of points: what can actually be computed is

[1] Tr.: *From time and sea nothing is learnt.*

just the convolution $D(\nu)$ (i.e. the **raw** power spectrum; Fig. C.1a) of the true DFT, i.e. the function $F(\nu)$, and the so-called **spectral window** $W(\nu)$; an example of this latter function is reported in Fig. C.1b. The spectral window is the DFT of the sampling function $s(t)$; this function is equal to 1 in correspondence of the times at which a measurements were made, and 0 elsewhere. In practice, therefore, the problem is due to the fact that the function $f(t)$ is known in a finite and limited number of points only. This produces the following effects:

- if T is the time span of the entire observation, the time frequency resolution in the DFT will be of the order $\Delta\nu \sim 1/T$;
- the discrete sampling produced by the data does not allow identifying periodicities with frequency larger than the **Nyquist critical frequency** ν_c, which is equal to $1/(2D)$, where D is the time separating two consecutive data points;
- the finite sampling time (that is, the total observation time) T and the discrete sampling produce spurious frequencies in the spectral window $W(\nu)$: this induces the creation of spurious frequency peaks in the raw spectrum $D(\nu)$ also. These peaks can be either daily *aliases* or higher-order harmonics of the "true" periodicity.

All this therefore defines a spectrum indicated as $D(\nu)$ in which, besides the peak corresponding to the frequency of the real modulation of the phenomenon, one can find its harmonics, that is, signals with frequency $\nu = \nu_{true} \pm k$, and the *aliases*, which are peaks with frequency ν equal to $k - \nu_{true}$, where k is in both cases a positive integer number. To this it should be added that, as already stressed in Ch. 4, if the signal is not sinusoidal, harmonics with frequency substantially different with respect to that of the true period gain strength; moreover, if flickering is present in the light curve, the low-frequency peaks are increased (*red noise effect*). Besides, if the number of data points is not high, the spectral window heavily biases all of the raw spectrum. It has also to be remarked that the presence of more periodicities in the same data set produces a further increase of the noise in $D(\nu)$.

Therefore, if the sampling of the light curve is not dense and accurate enough, the raw spectrum appears quite complex (see e.g. Fig. 4.4).

A more thorough and complete description of the Fourier Transforms can be found, for example, in Press et al. (*1992*).

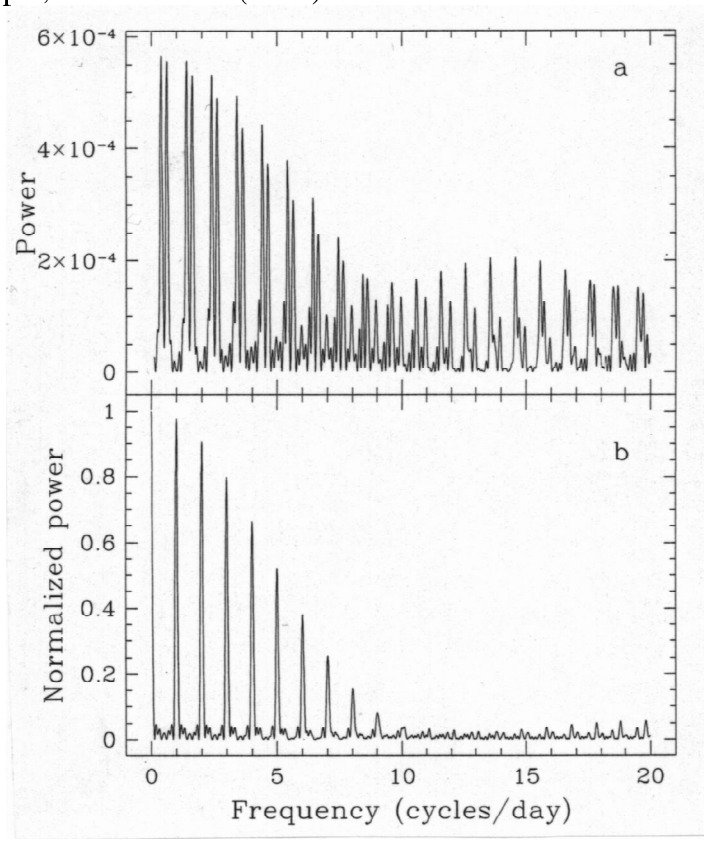

Fig. C.1. a Raw power spectrum $D(\nu)$ and **b** normalized spectral window $W(\nu)$ referring to the *V*-band data presented in Ch. 4.

C.1.2. The CLEAN algorithm

There is in any case the possibility of subtracting the spectral window $W(\nu)$ from the raw spectrum: this can be done by using the *CLEAN* method (Roberts et al. 1987). It was originally created to be applied in radio astronomy to reduce the noise in radio maps (Högbom 1974); subsequently, it has been modified so that it could be applied in the monodimensional field of time series. It consists in cleaning (hence its name) the power spectra from the secondary peaks due to the temporal incompleteness of the data set.

The algorithm, once that the functions $D(\nu)$ and $W(\nu)$ are known, works according to the following guidelines. First, it determines, in the raw power spectrum, the

frequency which is associated with the higher power peak; then it multiplies the spectral window $W(\nu)$ by a gain factor g, the value of which ranges between 0 and 2, and overlaps it to the above mentioned peak and subtracts this modified spectral window to $D(\nu)$. In such a way the secondary peaks associated by aliasing to the main one are also canceled out. The whole procedure is applied N_c times, and each time the subtracted frequency and the power associated with it are listed in the *CLEAN*ed spectrum $B(\nu)$, represented in the example of Fig. C.2a. At the end of the N_c iterations, what remains of $D(\nu)$ is a residual spectrum $R(\nu)$, illustrated in Fig. C.2b, in which the highest peak R_{max} (this parameter is given in the output of the program) indicates the goodness of the *CLEAN*ing procedure of the raw spectrum. The user therefore, by giving N_c e g as input information to the program (having care to remember that the product $g \cdot N_c$ has to be lower than the total number of sampled frequencies), can characterize the best *CLEAN*ed spectrum $B(\nu)$ corresponding to the lowest value of R_{max}.

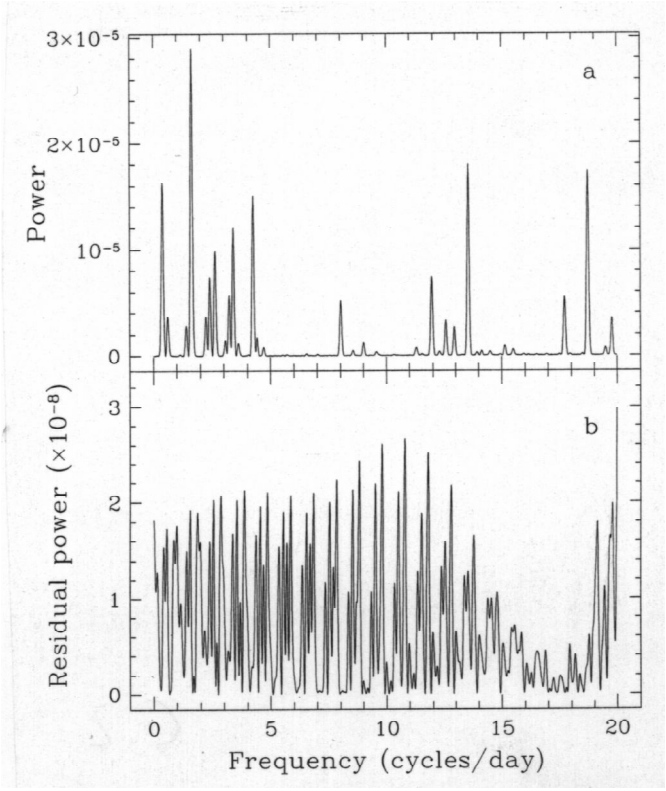

Fig. C.2. a *CLEANed* power spectrum $B(\nu)$ and **b** residual spectrum $R(\nu)$ referring to the *V*-band data presented in Ch. 4.

The parameter g should not be larger than 1: in such a way the work of the algorithm is more gradual. It is to be stressed that the results obtained with the

CLEAN algorithm have always to be checked with the use of other methods, particularly if more periodicities are present in the data or if strong noise effects of different nature affect the spectral window (e.g., flickering, poor data sampling, etc.).

C.2. LEAST-SQUARES FITTING METHODS

The methods described in this paragraph are instead based on the evaluation of the root mean square between the observed data and a theoretical trial curve of period P. In this way an error parameter Θ is then associated with every period contained in the interval $[P_1, P_2]$: the smaller Θ, the better the fit performed on the data assuming the corresponding periodicity $P(\Theta)$. It is therefore possible to evaluate which is the period that best describes the analyzed time series.

Within the framework of this Thesis, several methods of this kind have been applied: they are briefly reviewed below following an alphabetical order.

C.2.1. Barning's method

This method (Barning 1963) consists in fitting to the time series a linear combination of a sine and a cosine, both with frequency $f = 2/P$, where P is the trial period. Therefore, once the relative errors of the least-squares fit are calculated for each trial frequency, these errors are transformed in a probability percentage that each frequency has of being the true one. Then, once that the frequency and the amplitude of the modulation which best fits the data are calculated, this very modulation is subtracted from the time series; in this way it is possible to study the existence of other, possibly hidden, periodicities weaker than the main one. This method therefore allows to highlight a series of periodic modulations with different amplitudes within a same data set, assuming that they have a sinusoidal character. It should be stressed again that the frequencies analyzed here, and obtained as a result of the algorithm procedure, are corresponding to the **half** of the trial period P.

C.2.2. Dworetsky's method

With this method (Dworetsky 1983) the quantity to be minimized is the distance between adjacent points of the curve folded with every trial period P; it is therefore also called *string-length method*. By definition, such algorithm does not possess limitations in its applications which can be induced by the shape of the light curve. It is used in the cases in which the data are few, form unevenly sampled sets and are collected across long time intervals.

C.2.3. PDM method

Once that an interval of trial periods has been selected, this algorithm (Stellingwerf 1978) consists in dividing every period in N_b sub-intervals, each of them divided in turn into N_c parts, having care of choosing $N_b > N_c$ and $N_b \cdot N_c < N$, where N is the total number of points in the data set. This method has the task of minimizing the dispersion of the data contained in every phase sub-interval along the phase axis; hence the name Phase Dispersion Minimization (PDM). It is a least-squares technique; however, it does not perform the fit of a curve to the data points. Rather, it fits the points to an average curve by means of each sub-interval. The method illustrated here is useful in the case of ill- and irregularly sampled sinusoidal modulations.

C.2.4. Renson's method

This method also (Renson 1978) does not use any hypothesis on the light curve shape; it works by minimizing the sum of the ratios between the squares of the magnitude differences between adjacent points on the light curves, once they are phased with every trial period P, and those of the differences between the phases corresponding to such data points. In this way the algorithm tries to reduce the dispersion along both the axis of phases and the axis of magnitudes. This method suffers from aliasing and the trueness of the periods that turn out to be the most

probable should subsequently be checked by plotting the light curves folded with such periods. This method should be applied when the data set is not large.

C.2.5. Schöneich-Lange's method

This method (Schöneich & Lange 1981) has an approach similar to that of the Barning's method, in the sense that it uses a trial function with sinusoidal shape and with period P; however, this period is not considered in its entirety, but rather it is divided into N phase sub-intervals, and an error parameter is calculated on each of them. The best periodicity will thus be the one for which the total error of the fitting procedure turns out to be the lowest. The algorithm is not indicated for modulations whose shape is substantially different from a sinusoid.

C.2.6. Sterken's method

The present method (Sterken 1977), similarly to the Barning's and Schöneich-Lange's methods, uses a trial function in the form of linear combination of a sine and a cosine of period P, and calculates an error parameter on the whole curve (and not, as in the case of the Schöneich-Lange's method, just on parts of it). Of course, in this case also the most likely period will be the one for which the fit error is the lowest; this algorithm also is mainly indicated when the shape of the light curve is sinusoidal.

REFERENCES OF APPENDIX C

- Barning F.J.M., 1963, BAN, 17 (n. 1), 22
- Dworetsky M.M., 1983, MNRAS, 203, 917
- Högbom J.A., 1974, A&AS, 15, 417
- Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., 1992, Numerical Recipes. Cambridge Univ. Press, Cambridge
- Renson P., 1978, A&A, 63, 125
- Roberts D.H., Lehár J., Dreher J.W., 1987, AJ, 93, 968
- Schöneich W., Lange D., 1981, Astron. Nachr., 302, 153
- Stellingwerf R.F., 1978, ApJ, 224, 953
- Sterken C., 1977, A&A, 57, 361