SPECTRAL CONTENTS OF ASTRONOMICAL UNEQUALLY SPACED TIME-SERIES: CONTRIBUTION OF TIME-FREQUENCY AND TIME-SCALE ANALYSES

Sylvie Roques and Carole Thiebaut

Observatoire Midi-Pyrénées
14, avenue Edouard Belin, 31400-Toulouse, FRANCE

ABSTRACT

We analyze how time-scale and time-frequency methods can be suited to study unequally spaced astronomical time-series, in terms of recovering precise spectral contents of a pulsating star. In many situations in astronomy, non-uniform sampling proves to be the natural way in which the signal is available for investigation and the spectral information of interest is lost. We propose to study two kinds of methods which enables us to treat this type of signals: first, an approach with global wavelet spectra leading to a wavelet scale that can be compared to the Fourier period. Second, a time-frequency analysis (matching pursuit) providing a progressive constructive process allowing us to identify the fine structure of the driven frequencies and to control the error propagation. We apply, compare and discuss these methods for simulations and for luminosity observations of variable stars. This allows us to conclude that the matching pursuit is the more efficient method, in particular independent of the quality of sampling.

1. INTRODUCTION AND ASTROPHYSICAL CONTEXT

The non-uniform sampling problem arises in many astronomical fields [1,2], particulary in stellar physics when one observes the luminosity of variable stars (asteroseismology). The frequencies deduced from these luminosity variations represent an important source of information. In particular, they are a constraint for the stellar evolution models because the fine structure of the vibration modes and their frequency separations may yield physical parameters, for example the rotation period of the star or the composition of its layers [3,4]. Of course, observations have to be long enough to obtain the best possible resolution on the spectra. Astronomers know the hard cost of obtaining such complete observations: the lack of information is essentially due to diurnal cuts, poor weather conditions or equipment malfunctions. This leads to unobservability of the star under study on some samples or time intervals.

Generally, the astronomical gapped data are of two types: first, evenly spaced time series separated by wide gaps [5] (typically the problem of day/night alternation for observations during several 24 hours and a sampling rate of about 50 s); many different methods have been proposed to deal with these problem, and in particular AR models allowing prediction in the gaps [6] combined with observing campaigns with telescopes distributed over several longitudes. Second, strongly unequally spaced time-series with samples missing almost everywhere when the data under study represent several years of observations with a mean sampling rate of 1 day (here, telescope failures or weather conditions are the main cause of the gaps [7]); this second case is the subject of this paper. Of course, problems of this kind arise not only in the treatment of astronomical signals, but in that case it is of capital importance to solve them to be able to carry out a physical interpretation of the observations, no other experimental alternative being possible. Moreover, as we search for oscillations characteristic of the structural properties of the star (i.e. arising almost everywhere in the signal), one understands the necessity of getting, at the same time, information about the lifetime of a given peak of the spectrum.

In this context, the wavelet analysis [8] and the time-frequency analysis [9] have the ability to decompose the signal into contributions localized both in time and in scale (or frequency), and thus are specially attractive to analyse such astronomical data [10]. Indeed, on the power spectrum or on the periodogram, the intervals including low amplitude peaks are hard to identify by eye because each feature in the spectrum is accompanied by sidelobes whose nature closely depends on the noise and on the irregular distribution of the data. As they can be of substantial amplitude, they can lead to the confusion of features due to oscillations with those arising from the segmented nature of the observing window.

We adapt here two types of approach: 1) a global wavelet transform and the associated wavelet spectrum [11], and 2) a matching pursuit decomposition [12]. The results are also compared to those given by a periodogram [13]. As the intrinsic idea of these methods requires that the signal is regularly sampled, we deliberately use a simple linear interpolation (sampling rate equal to one day) recommended by De Waele & Broesen [14].

2. THE PROPOSED APPROACHES

The global wavelet transform used here corresponds to a continuous wavelet approach [8] allowing us to define a global wavelet spectrum as the square of the modulus of the wavelet coefficients for each scale [11]. This leads to an *equivalent Fourier period* (which can be derived analytically for each wavelet function) which can be easily compared to the Fourier power spectrum or to the periodogram [13].

The matching pursuit algorithm [12] allows us to choose, in a redundant dictionary of time-frequency waveforms, a set of vectors that match the signal as well as possible, thanks to iterated one-dimensional projections. The dictionary is defined as a family of time-frequency functions obtained by dilating, modulating and translating a single real even function $k_\nu(t) \in L^2(R)$. The *atoms* (elements) of the dictionary are defined like a wavelet, but all the parameters (dilation scale, translation and frequency modulation), indexed by ν, can vary at the same time.

The light curve $\psi(t)$ is approximated with a single vector k_{ν_0} chosen in the dictionary such that $|\langle \psi(t), k_{\nu_0}(t) \rangle|$ is as large as possible. The light curve is then decomposed into the form $\psi(t) = \langle \psi(t), k_{\nu_0}(t) \rangle k_{\nu_0}(t) + R\psi(t)$ where $R\psi(t)$ is the residual vector after approximating $\psi(t)$ in the "direction" $k_{\nu_0}(t)$. The main idea is to sub-decompose the residue $R\psi(t)$, by finding a vector $k_{\nu_1}(t)$ that matches it as well as possible, as it was done for $\psi(t)$. Each time, the procedure is repeated on the residue that is obtained, and a convergence criterion exists. Finally, the signal is decomposed into $\sum_{i=0}^{\infty} \langle R^i \psi(t), k_{\nu_i}(t) \rangle k_{\nu_i}(t)$ where the atoms $k_{\nu_i}(t)$ are the ones that match the signal structures as well as possible. We can then build a hierarchy of coherent structures $\{k_{\nu_i}(t)\}$ yielding a time-frequency energy distribution of the signal, that will be represented in the further diagrams.

3. APPLICATION EXAMPLES

The study is conducted on several astronomical data. Only one will be represented on the figures: S Persei [15]. S Persei is a pulsating semi-regular variable star of "type C": its light curve is characterized by a simple wave and is sometimes interrupted by various irregularities. Its known period is 822 days. The S Persei light curve is irregulary sampled and the same sampling has been used to build a simulated signal. The signals and their periodograms are presented in Figs. 1, 2 & 5. Note that the sampling of this star observation presents large gaps at the beginning of the run. We also define a "sampling" signal for which each value is equal to the time step of the S Persei data (the wider is the gap the larger is the value). It is normalized so that its variance is the same as the relevant simulated signal.

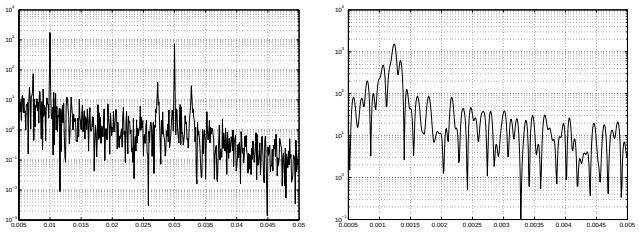

Fig. 1. Periodograms of the simulated signal (left) and of S Persei light curve (right) in a logarithmic scale.

Simulated Data: The simulated signal is the sum of two cosines of periods 100 and 33.3 days. No noise has been added, the aim being to analyze the effect of the non-uniform sampling and how the methods answer this question. In Fig. 1 (left) the two frequencies are visible and the harmonics essentially correspond to the annual cycle of the observations at 365 days.

In Figure 2, we present the Wavelet Power Spectrum (WPS) and the Global Wavelet Spectrum (GWS) obtained for the simulation. In the GWS, the 95% confidence level that could be obtained for a white noise is quasi superposed with the X axis (the linear interpolation induces errors comparable to noise). In the WPS the continuous white line indicates the cone of influence (important edge effects).

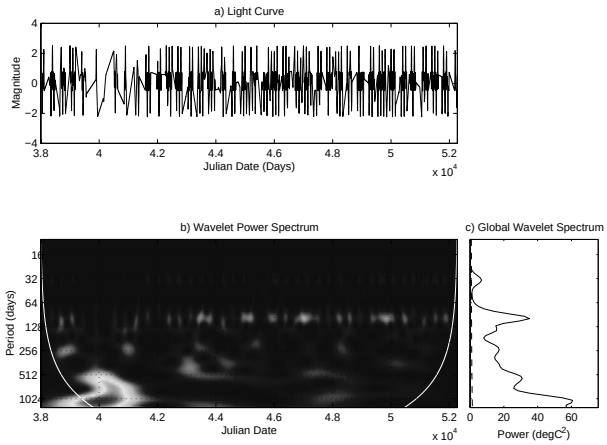

Fig. 2. Simulated signal (top), Wavelet Power Spectrum (left) and Global Wavelet Spectrum of the simulated signal (right) with a Morlet wavelet.

In the WPS, one can identify the two periods. A large zone due to the bad sampling around t=40000 *modified* Julian days (MJD) appears in the low frequencies and the 100 days period does not appear clearly at this time. The GWS, highlights the same problem. Fig. 3 presents the GWS of the simulated signal compared with that of the sampling signal: the irregular sampling is responsible only for the peak at 256 days.

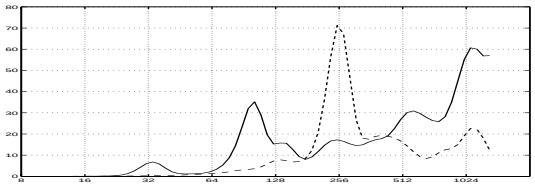

Fig. 3. Global Wavelet Spectrum of the simulated signal (continuous line) compared to the one of the sampling signal (dashed line).

The Matching Pursuit results presented here use the free graphical interface developed in our Institute [16]. The linearly interpolated simulated signal is decomposed on 100 atoms with a Spline 0 window (Fig. 4). The long atoms represent the most coherent structures of the signal. The peaks which do not correspond to star oscillations but are artifacts due to the sampling appear localized on a short time or in a large frequency range (vertical atoms). We can clearly identify a long atom at 0.03 day^{-1} and several shorter atoms at 0.01 day^{-1} (the harmonic frequencies correspond to the annual cycle). As these atoms are among the most energetic in the decomposition, the simulated frequencies are thus perfectly highlighted.

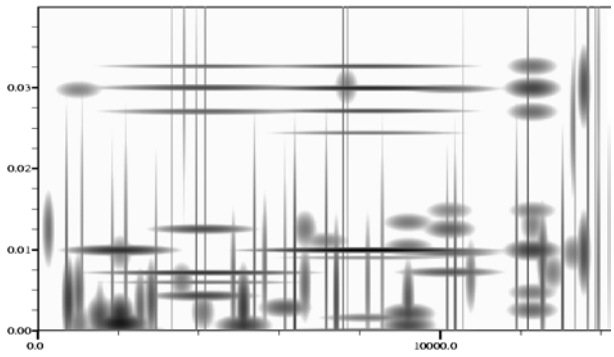

Fig. 4. Time-frequency decomposition of the simulated signal. Each stain represents the energy density of each atom. X axis: time; Y axis: frequency.

Real data: We analyze the real light curve of the variable star S Persei with the same methods (see Fig. 5). The periodogram (Fig. 1) presents a noisy behaviour, a large peak centered on the frequency 0.00124 day^{-1} (806.4 days) and two other important ones at 0.00134 day^{-1} (746.3 days) and at 0.00106 day^{-1} (943.4 days). There is also a fourth at 0.0008 day^{-1} (1250 days). The known period of S Persei (822 days) is thus not correctly identified.

Fig. 5 presents the WPS and the GWS of S Persei obtained with the same wavelet (Morlet) as for the simulated signal. In the WPS, one can clearly identify a period at 826 days at the end and at the beginning of the time interval. In

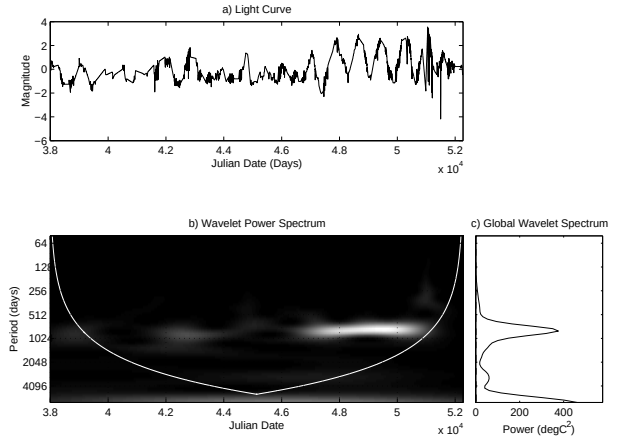

Fig. 5. S Persei light curve (top), Wavelet Power Spectrum (left) and Global Wavelet Spectrum (right) with a Morlet wavelet. The 95% confidence level is not visible.

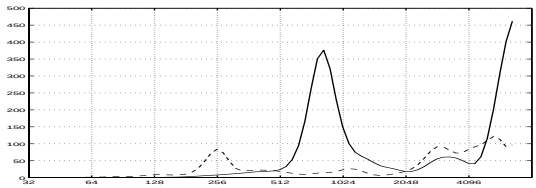

Fig. 6. Global Wavelet Spectrum of S Persei signal (continuous line) compared to the one of the sampling signal (dashed line).

the GWS, this period is also observable, together with some other ones after 3000 days, but the highest one is outside the cone of influence. The GWS of the corresponding "sampling" signal is presented Fig. 6. The superposition with that of S Persei explains its peak centered at 3000 days, due to the sampling, which is thus not relevant.

The Matching Pursuit analysis of S Persei is performed on 100 atoms with a Spline 0 window. No long lifetime frequency can be identified by eye in the diagram (see Fig. 7), but the first atom of the decomposition at 0.001221 day^{-1} (819 days), representing the most energetic structure of the signal, approaches at best the known oscillating period of this star (822 days).

4. DISCUSSION AND CONCLUSION

The periodogram provides well resolved spectra but is very sensitive to the uneven sampling; here in particular, peaks due to the annual cycle of observation appear large for the simulated signal. The S Persei periodogram presents a noisy behaviour and a central lobe perturbated by harmonic frequencies.

The 95% confidence level associated to the WPS and

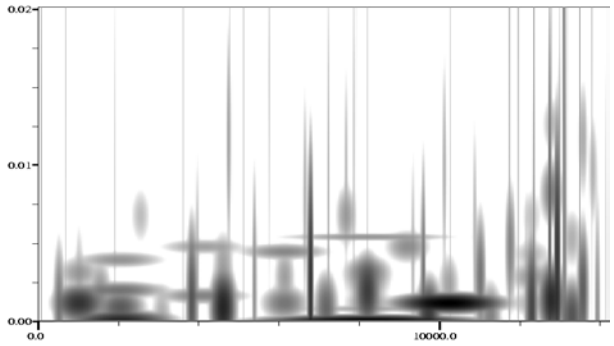

Fig. 7. Time-frequency decomposition of S Persei signal. X axis: time; Y axis: frequency.

the cone of influence are efficient tools to check if the highlighted frequency is meaningful. But this method reveals problems in the field of low frequencies: in the GWS, peaks of high amplitude are not always explained by those of the sampling signal. This restriction prevents from carrying out an associated error analysis.

As for the periodogram, the Matching Pursuit analysis is sensitive to the annual cycle of observation, but the atoms hierarchy provides information on the dominating frequencies of the signal if they are not easily indentifiable on the diagram. Let us define an associated error as the quadratic reconstruction error after decomposition on 400 atoms. Table 1 presents an error analysis for simulated signals built according to several variable stars data with samplings of different quality (column 1 of Table 1). Here, the standard deviation of the corresponding sampling signal is used to measure this quality. One can note that the reconstruction error is small and in particular not correlated with the quality of sampling. This can be explained by the choice of atoms of different lifetime and frequency offered by the large dictionary, which can solve the problem of a very lacunar sampling (see for example the error reconstruction for T Camelopardis compared to its sampling).

In conclusion, let us recall that the known period of S Persei at 822 days is found at 806 days by the periodogram, at 826 days by the Global Wavelet Spectrum and at 819 days by the Matching Pursuit. This last method also permits to conduct an error analysis. The Matching Pursuit algorithm is thus well suited for spectral investigation of irregularly sampled variable stars signals. Finally, let us specify that we plan to investigate a comparison with the results that can be obtained by the interpolation technique proposed by Strohmer [17] in a forthcoming work.

Star Name	Stand. Deviation	Reconst. Error
AC Herculis	2.17	3.71%
S Persei	23.16	11.22%
RV Tauri	46.07	8.53%
T Camelopardis	114.60	10.98%

Table 1. Reconstruction Error versus signal sampling

5. REFERENCES

[1] M. Carbonnel, R. Olivier and J.L. Ballester, "Power spectra of gapped time series: a comparison of several methods", Astron. Astroph., No 264, p. 350, 1992.
[2] S. Roques, B. Serre, N. Dolez, "Band-limited interpolation applied to rapidly oscillating stars time series", Mon. Not. R. Astr. Soc., No. 308, p. 876, 1999.
[3] D. Koester and G. Chanmugan, "Physics of white dwarfs stars", Reports on Progress in Physics, No 53, p. 837, 1990.
[4] P. Bradley, "Theoretical models for asteroseismology of DA white dwarfs", Astroph. J., No 468, p. 350, 1996.
[5] G.G. Fahlman and T.J. Ulrych, "A new method for estimating the power spectrum of gapped data", Mon. Not. R. Astr. Soc., no 199, p. 53, 1982.
[6] S. Roques, A. Schwarzenberg-Czerny and N. Dolez, "Parametric spectral analysis applied to gapped time-series of variable stars", Baltic Astronomy, Vol. 9(1), p. 463, 2000.
[7] C. Catala and 26 authors, "Short term spectroscopy variability in the pre-main sequence Herbig Ae star AB Aurignae during the MUSICOS 96 campaign", Astron. Astrophys., No 345, p. 884, 1999.
[8] I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, 1992.
[9] P. Flandrin, *Time-frequency/Time-scale analysis*, Academic Press, 1998.
[10] J.D. Scargle, "Wavelet Methods in Astronomical Time Series Analysis", Applications of Time Series Analysis in Astronomy and Meteorology, Chapman & Hall, p. 226, 1997.
[11] C. Torrence and G.B. Compo, "A Practical Guide to Wavelet Analysis", Bull. Am. Meteo. Soc., Vol. 79(1), p. 61, 1997.
[12] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries", IEEE Trans. Signal Process., 41(12), p. 3397, 1993.
[13] J.H. Horne and S.L. Baliunas, "A prescription for period analysis of unevenly sampled time series", Astrophys. J., no 301, p. 757, 1986.
[14] De Waele S. and Broersen P.M.T., "Error measures for resampled irregular data", IEEE Trans. Instr. Measurements, vol. 49, no 2, p. 216, 2000.
[15] http://vsnet.kusastro.kyoto-u.ac.jp/vsnet/index.html
[16] http://webast.ast.obs-mip.fr/people/fbracher
[17] T. Strohmer, "Computationally Attractive Reconstruction of Band-Limited Images from Irregular Samples", IEEE Trans. Image Proc., 6(4), p. 540, 1997.