

Time-Scale and Time-Frequency Analyses of Irregularly Sampled Astronomical Time Series

C. Thiebaut

Centre d'Etude Spatiale des Rayonnements, 9 avenue du Colonel Roche - Boite postale 4346, 31028 Toulouse Cedex 4, France
Email: carole.thiebaut@cnes.fr

S. Roques

Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées, 14 avenue Edouard Belin, 31400 Toulouse, France
Email: sylvie.roques@obs-mip.fr

Received 27 May 2004; Revised 21 January 2005

We evaluate the quality of spectral restoration in the case of irregular sampled signals in astronomy. We study in details a time-scale method leading to a global wavelet spectrum comparable to the Fourier period, and a time-frequency matching pursuit allowing us to identify the frequencies and to control the error propagation. In both cases, the signals are first resampled with a linear interpolation. Both results are compared with those obtained using Lomb's periodogram and using the *weighted wavelet Z-transform* developed in astronomy for unevenly sampled variable stars observations. These approaches are applied to simulations and to light variations of four variable stars. This leads to the conclusion that the matching pursuit is more efficient for recovering the spectral contents of a pulsating star, even with a preliminary resampling. In particular, the results are almost independent of the quality of the initial irregular sampling.

Keywords and phrases: astronomical time series, irregular sampling, time-scale methods, time-frequency methods, wavelets, matching pursuit.

1. INTRODUCTION

Nonuniform sampling problems arise in many astronomical fields [3, 22], particularly in Stellar physics when one observes the light curves of variable stars (asteroseismology) or spectroscopic variabilities. The frequencies deduced from the light variations of such stars represent an important source of information. In particular, they can help constrain stellar evolution models, because the structure of the vibration modes and their frequency separations may yield physical parameters of the star, such as the rotation period or the composition of its layers [1, 16]. Another field of application concerns the development of automatic classifiers for variable stars, where the period is a very discriminating parameter [27]. Of course, observations have to cover a long enough time span for the best possible resolution of the power density spectra. The difficulty in obtaining such complete observations is well known: the lack of information is essentially due to diurnal cuts, poor weather conditions, or equipment malfunctions. Generally, such astronomical data

are of two types. First, evenly spaced time series separated by wide gaps [9] (typically day/night alternation for observations of short-period stars taken over several days). In that case, many different methods have been proposed to deal with this problem, for example, autoregressive models which predict data for the gaps [21] combined with observing campaigns with telescopes at several different longitudes [18]. Second, unequally spaced time series with samples missing almost everywhere. Here, data under study are from several years of observations (long-period stars) with a mean sampling rate of a few days (here, telescope failures or bad weather conditions are the main causes of the gaps [4]). This second case is considered in this paper. Of course, problems of this kind do not arise only when processing astronomical signals. In an astrophysical context, it is of capital importance to solve them in order to carry out a physical interpretation of the observations, no other experimental alternative is possible.

Another problem often arises in complement: searching for oscillations that are characteristics of the structural properties of the star (i.e., these that arise almost everywhere in the signal) thanks to a frequency analysis. One then understands the necessity of getting information about the lifetime of a given peak of the resulting power density spectrum.

In this context, wavelet analysis [6] and time-frequency analysis [10], which have the ability to decompose the signal into contributions localized both in time and in scale (or frequency), are thus especially attractive to obtain this information. These analyses are widely described for simulated data for example in Szatmáry et al. [26] and in variable star research in Kiss and Szatmáry [14]. These authors conclude that such methods of decomposition, labeled with a scale and a position parameter, provide interpretable visual representations of astronomical data, as an alternative to the standard spectral analysis.

Unfortunately, wavelet and time-frequency analyses are generally not directly applicable to the particular case of irregularly sampled data [24], thus one often uses standard techniques like periodograms [13]. However, with such a simple spectral technique, intervals including low-amplitude peaks are hard to identify in the processed data because each feature is contaminated by noise and convolved by a function whose nature closely depends on the irregular distribution of the data. As the corresponding aliases can be of substantial amplitude, they can lead to the confusion of features due to real oscillations with those arising from the segmented nature of the observing window. If one deals only with signals whose spectra are dominated by a small number of components at discrete frequencies, nonlinear deconvolution methods like the widely used CLEAN technique are efficient [20]. Obviously, in the absence of any a priori information on the spectrum to be recovered, the case considered here, this technique becomes unreliable.

Time-frequency methods are efficient, particularly when an examination by eye of the periodogram leads to a failure (see Section 4). If data sampling is not equidistant, one must first resample the signal to build a regular sampling, before applying a time-frequency method. Several techniques exist to do this and the associate errors are widely discussed in the literature. de Waele and Broersen [8] divide them into simple and complex methods. In particular, they conclude that linear interpolation is a robust resampling method although it provides a signal whose standard deviation is biased with a systematic error. However, this error can be corrected by replacing the standard deviation obtained with linear interpolation by the value given with the method they propose: the nearest neighbor resampling. Foster [11] proposed a rescaled wavelet technique called *weighted wavelet Z-transform* (WWZ). It is developed specifically for unevenly sampled data in the context of observation of variable stars: here, the wavelet is rescaled to satisfy admissibility condition on such irregular sampling. One can for example read the paper of Haubold [12] for an interesting analysis of this method. In this paper, our results will also be compared with those obtained by the WWZ technique.

The signals under consideration in this paper being irregularly sampled, we have opted for a processing method in two stages: (1) we have resampled the data by using a linear interpolation (with a sampling rate typically equal to one day and without applying any additional smoothing or filtering—contrary to preprocessing techniques recommended by Buchler et al. [2]), (2) we have then applied

two appreciably different types of time-frequency analyses: a global wavelet transform and the associated wavelet spectrum [28], which is described in Section 2, and a matching pursuit decomposition [17] developed in Section 3. The results are discussed and compared (Section 4) to those found using a periodogram and WWZ, for simulated signals and for light curves of four variable stars: T Camelopardis—a Mira variable star of period 373.2 days, S Persei—a Type C semiregular star of period 822 days, AC Herculis—a Type A RV Tauri variable star of period 75.01 days, and RV Tauri—a Type B RV Tauri star of period 78.73 days. These periods are those specified in the fourth edition of the General Catalog of Variable Stars, see <http://www.sai.msu.su/groups/cluster/gcvs/gcvs>. This leads us in particular to forecast a chaotic light curve of AC Herculis as predicted by Kollath et al. [15].

2. GLOBAL WAVELET SPECTRA

We recall that a wavelet decomposition is an expansion of an arbitrary function into smoothed localized contributions labeled by a scale and a time parameter. Its aim is to expand a signal into a series of coefficients of specified energy and then to capture fine and coarse features at different scales. Moreover, it provides an easily interpretable visual representation of the signal (see, e.g., the book of Daubechies [6] for more details). Wavelets are generated by a function $\psi(t)$ named *the analyzing wavelet*. This function should have a finite energy, and its integral should vanish. These two conditions mean that the wavelet should oscillate like a short wave. The analyzing wavelet is *the mother* of the wavelet family. The wavelet family $\{\psi\}$ is generated by translating and dilating the analyzing wavelet. Then, one can write any function as a linear combination of the elements of the family.

Here we consider the continuous wavelet transform of a real signal $s(t)$ with respect to the analyzing wavelet $\psi(t)$. The wavelet transform is defined as the function

$$C(b, a) = \frac{1}{\sqrt{a}} \int \psi^* \left(\frac{t-b}{a} \right) s(t) dt \quad (1)$$

on the time-scale plane. Here, a is a dilation scale and b a translation parameter; the asterisk denotes the complex conjugate. The qualitative information given by the visual output supplements the information obtained by inspection of the signal itself, or its Fourier transform. The wavelet transform displays information in a wide range of scale parameters on a single picture.

The choice of the analyzing wavelet is generally guided by a compromise between time and frequency resolutions and by its ability to capture localized features of the signal. As we are essentially interested in wavelet power spectra, the wavelet we used here is the Morlet wavelet (a complex exponential modulated by a Gaussian)

$$\psi_0(t) = \pi^{-1/4} \exp(i\omega_0 t) \exp\left(-\frac{t^2}{2}\right), \quad \text{with } \omega_0 = 6, \quad (2)$$

which offers high frequency resolution because it is very well localized in frequencies. In contrary, using derivative of Gaussian wavelet would result in a good time localization, but a poor one in frequency.

The global wavelet transform used in our analysis corresponds to a continuous wavelet approach allowing the definition of a global wavelet spectrum as the square of the modulus of the wavelet coefficients for each scale together with statistical significance tests [28]. The aim of the method proposed by these authors is to provide quantitative tools associated with wavelet analysis. This leads in particular 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.

3. MATCHING PURSUIT ALGORITHM

The matching pursuit algorithm, introduced by Mallat and Zhang [17], allows us to choose, in a given redundant finite dictionary of time-frequency waveforms, a set of vectors that match the signal as well as possible. The dictionary \mathbf{D} is defined as a family (not a basis) of time-frequency functions obtained by dilating, modulating, and translating a single real even function $k(t) \in L^2(\mathbb{R})$. The *atoms* (elements) of the dictionary are defined by

$$k_\nu(t) = \frac{1}{\sqrt{a}} k\left(\frac{t-b}{a}\right) e^{i\omega t}, \quad (3)$$

where a is the dilation scale, b the translation parameter, and ω a frequency modulation. One defines $\nu = (a, b, \omega)$ as the atom index in the dictionary. Note that the factor $1/\sqrt{a}$ normalizes the $L^2(\mathbb{R})$ norm of $k_\nu(t)$ to unity. If the window $k(t)$ is Gaussian, the joint time-frequency localization of all the atoms is a minimum, and in this case $k_\nu(t)$ is a *Gabor function*. Note that the family $\{k_\nu\}$ is not a wavelet family in the sense that a given dilation allows several analyzing frequency values. It can be seen as a superposition of a wavelet transform and a short-term Fourier transform. In particular, the underlying family is nonorthogonal. In practice, the atoms of the family are thus oscillating functions modulated by window functions. They are generated by two mother functions that satisfy localization properties: a window function $k(t)$ (a “Spline 0” in this paper—see Figure 1) and a wavelet-like function. For each value of the atom index ν , one has a new atom of the family, until obtaining a complete collection of atomic waveforms.

A matching pursuit algorithm computes adaptive signal representations: it expands any signal into a set of atoms selected among the redundant dictionary \mathbf{D} , to match its components as well as possible, through iterated one-dimensional projections.

In our particular case, it is the resampled light curve $s(t)$, from which the mean value has been subtracted after resampling so that it becomes zero mean, which is approximated by a single vector k_{ν_1} chosen from the dictionary \mathbf{D} such that $|\langle s(t), k_{\nu_1}(t) \rangle|$ is as large as possible. Here, $\langle \cdot, \cdot \rangle$ denotes the

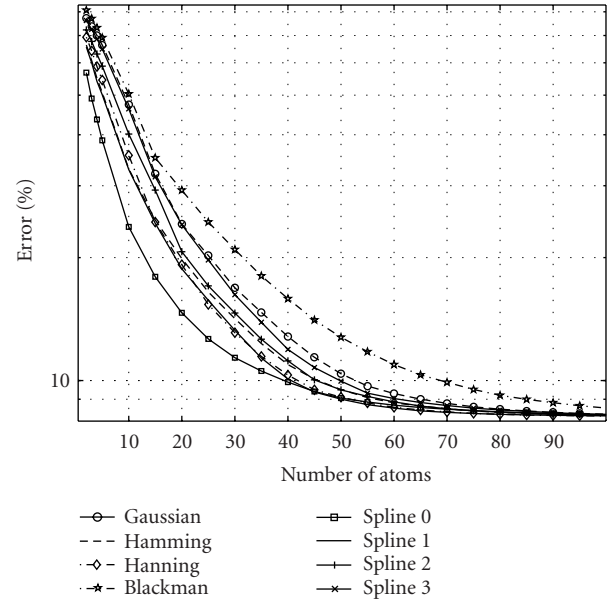


FIGURE 1: Reconstruction error versus number of atoms of the matching pursuit decomposition for each shape of the even real function $k(t)$ (in logarithmic scale).

scalar product in \mathbb{R} . In practice, one has to compute a scalar product for all the values of ν . The vector k_ν giving the largest one is k_{ν_1} .

The light curve is then decomposed into the form

$$s(t) = \langle s(t), k_{\nu_1}(t) \rangle k_{\nu_1}(t) + R s(t), \quad (4)$$

where $R s(t)$ is the residual vector after approximating $s(t)$ in the “direction” $k_{\nu_1}(t)$. Clearly, $k_{\nu_1}(t)$ is orthogonal to $R s(t)$, and hence, one has the relation

$$\|s(t)\|^2 = |\langle s(t), k_{\nu_1}(t) \rangle|^2 + \|R s(t)\|^2, \quad (5)$$

where $\|\cdot\|$ denotes the Euclidian norm. Note that if the family $\{k_\nu\}$ is nonorthogonal, the vector $k_{\nu_1}(t)$ is orthogonal to the residual $R s(t)$. This important property allows the construction of the algorithm: the main idea of the matching pursuit is to subdecompose the residue $R s(t)$, by finding a vector $k_{\nu_2}(t)$ that matches it as well as possible, as was done for $s(t)$. Each time, the procedure is repeated on the obtained residue:

$$\begin{aligned} s(t) &= \langle s(t), k_{\nu_1}(t) \rangle k_{\nu_1}(t) + R s(t), \\ R s(t) &= \langle R s(t), k_{\nu_2}(t) \rangle k_{\nu_2}(t) + R^2 s(t), \\ &\vdots \\ R^n s(t) &= \langle R^n s(t), k_{\nu_n}(t) \rangle k_{\nu_n}(t) + R^{n+1} s(t). \end{aligned} \quad (6)$$

It is easy to determine a convergence criterion of the algorithm, by examining the decrease of the norm of the residue.

Finally, the signal is decomposed into

$$s(t) = \sum_{i=1}^{\infty} \langle R^i s(t), k_{v_i}(t) \rangle k_{v_i}(t), \quad (7)$$

where the atoms $k_{v_i}(t)$ are the ones that match the signal structures as well as possible. We can then build a hierarchy of the main signal structures ($k_{v_1}(t), k_{v_2}(t), \dots, k_{v_n}(t)$) yielding a time-frequency energy distribution.

An energy conservation theorem results from (5):

$$\|s(t)\|^2 = \sum_{i=1}^{\infty} |\langle R^i s(t), k_{v_i}(t) \rangle|^2. \quad (8)$$

The energy density of $s(t)$ in the time-frequency plane (t, u) is defined by

$$(Es)(t, u) = \sum_{i=1}^{\infty} |\langle R^i s(t), k_{v_i}(t) \rangle|^2 (Wk_{v_i})(t, u), \quad (9)$$

where $(Wk_{v_i})(t, u)$ is the Wigner-Ville distribution [29] defined as follows:

$$(Wk_v)(t, u) = \int_{-\infty}^{\infty} k_v\left(t + \frac{\tau}{2}\right) k_v^*\left(t - \frac{\tau}{2}\right) e^{-iu\tau} d\tau. \quad (10)$$

In practice, it is this energy density Es (9) that is represented in the time-frequency diagrams. Note that this density does not include the interference terms of the Wigner-Ville distributions, because it is computed from an atomic decomposition of $s(t)$.

Decomposition onto orthonormal basis or the method of Coifman and Wickerhauser [5] selects in a global way the basis that is best adapted to the signal properties. The results associated with these methods are hardly interpretable. On the contrary, the matching pursuit decomposition is a constructive process which allows to detect and characterize the time-frequency components one by one, from the highest energetic one to the lowest.

In a matching pursuit diagram, we can choose to select only some atoms representing the structures of interest. In our case, these are the most coherent ones. The corresponding atoms appear as long (in time) elements in the time-frequency plane (horizontal atoms). As the noise of the time series does not correlate well with any long lifetime dictionary element, its information is diluted and then subdecomposed in several “stains” localized in a short time interval. The peaks, even of large amplitude, which do not correspond to star oscillations but are artifacts due to the sampling or to the gaps of the observing window or corresponding to highly transient phenomena appear in the time-frequency diagram localized in a very short time period, in a large frequency range (vertical atoms). A simple operation, keeping only the long lifetime atoms (whatever their frequency), allows us to eliminate spurious information (e.g., corresponding to noise).

The shape of the even real function $k(t)$ can be chosen according to various criteria, while still in agreement with the mathematical conditions authorizing the matching pursuit decomposition. These criteria depend on the adaptability of $k(t)$ to the studied signal and on some oversensitivity to errors. Eight different window shapes have been considered: Gaussian, Hamming, Hanning, Blackman, Spline 0, Spline 1, Spline 2, and Spline 3. Their description can be found in Oppenheim and Schaffer [19] and in de Boor [7]. Figure 1 presents the evolution of the quadratic reconstruction error versus the number of atoms used to decompose a simulated signal with these eight different windows. The tested signal is a regularly spaced 1200-point-long signal, the sum of two cosines with periods of 33.3 and 100 days. The quadratic reconstruction error is defined as the norm of the residual signal between the simulated and the reconstructed ones divided by the norm of the simulated signal. The Blackman window always gives the largest error. The Gaussian and Spline 3 windows also show large errors whatever the number of atoms. Up to 50 atoms, the Spline 0 window is clearly the best. For a decomposition with more than 50 atoms, all the windows, except the Blackman window, present similar reconstruction errors. Following this, as the signals we analyze here are of quite constant amplitude, the adaptation of $k(t)$ to the data leads us to choose the Spline 0 window. Indeed, the rectangular shape of this window does not create an amplitude modulation.

4. APPLICATION EXAMPLES AND THE COMPARISON WITH THE PERIODOGRAM AND WITH THE WWZ

Although we have analyzed the light curves of four stars, we focus on only two of them (AC Herculis and RV Tauri—see <http://www.kusastro.kyoto-u.ac.jp/vsnet/index.html>) and the results are summarized in Table 1.

The data set of the irregularly sampled observations spans JD 2 440 000–2 450 000 for AC Herculis and JD 2 432 223–2 452 270 for RV Tauri (see Figure 2a). JD is the Julian day number, number of days that have elapsed since noon of January 1, 4713 B.C. of our civil calendar. In addition, we have created two artificial signals (sum of two cosines with periods of 100 and 33.3 days) with the same nonequidistant sampling scheme as those of AC Herculis and RV Tauri, and without added noise. The choice not to add noise is linked to the concern of analyzing only the effect of the nonuniform sampling and how the methods address this question.

The periodograms from these variable star observations and those from the simulated signals made from the same irregular sampling are presented in Figures 3 and 4. In these figures (top), the two 100-day and 33.3-day periods are visible. The clearly identifiable aliases essentially correspond to the annual cycle of the observations, proving that one finds it effective in the sampling.

We used also the WWZ and experimented it with different values for the parameter c [11] defining the trade-off between time resolution and frequency resolution. After accepting the default parameters proposed by Foster

TABLE 1: Results for the four variable stars (column 1). The sampling quality is defined by the mean number of observations per day (column 2). The reconstruction error computed for the matching pursuit analysis is presented (column 3). The known period and those found with the different analyses are indicated (columns 4–8): periodogram (P_{Perio}), GWS (P_{GWS}), matching pursuit decomposition (P_{MP}), and weighted wavelet Z-transform (P_{WWZ}). In the WWZ, $c = 0.005$ for data of S Per. and T Cam., $c = 0.04$ for AC Her, and $c = 0.0125$ for RV Tau. In brackets: double periods are also found and are corresponding to the periods indicated in the fourth edition of the General Catalog of Variable Stars.

Star	O/day	Error	P_{GCVS}	P_{Perio}	P_{GWS}	P_{MP}	P_{WWZ}
AC Her.	0.757	3.71%	75.01	37.74 _[75,19]	37.00 _[75,00]	37.73 _[75,19]	38.24 _[71,43]
S Per.	0.112	11.22%	822.00	806.40	826.00	819.00	831.90
RV Tau.	0.190	8.53%	78.73	39.26	39.00	39.29 _[8,77]	38.8
T Cam.	0.051	10.98%	373.20	371.75	373.00	372.30	371.10

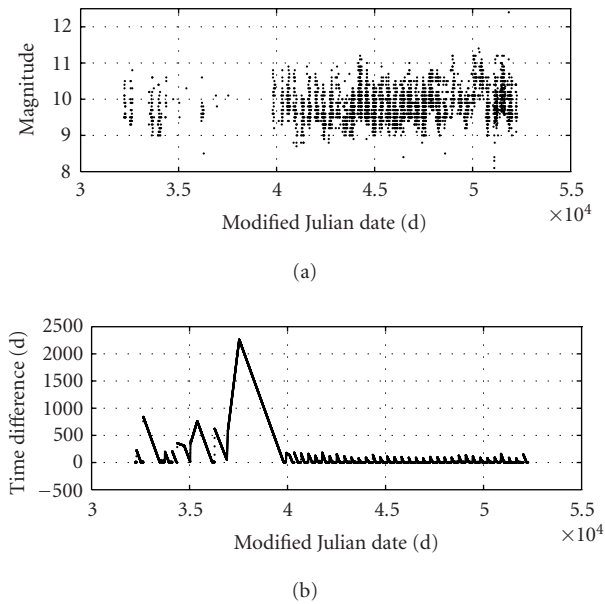


FIGURE 2: RV Tauri sampling signal (modified Julian date = JD-2 400 000.5).

($c = 0.0125$) for time and frequency grids, we settled on $c = 0.04$ for time sampling of AC Herculis and kept $c = 0.0125$ for time sampling of RV Tauri as better compromises in the sense that they were the first lowest values for which we obtained the target periods. WWZ applied on these test data can reveal peaks at frequencies 100.5 and 33.2 days for time sampling of AC Herculis and at frequencies 100.0 and 32.9 days for time sampling of RV Tauri. The same constants will be used for application to the observations.

In order to perform time-frequency and time-scale analyses, the data (simulated and real) have also been linearly interpolated with a time step equal to one day. As the known periods of the studied variable stars are more than 50 days, Shannon's theorem is satisfied. The so-interpolated signals are presented in Figures 5 and 6 (simulated) and Figures 7 and 8 (real). Note that contrary to the case of AC Herculis, whose missing samples are well distributed throughout the observation, RV Tauri data contain very large gaps at the beginning of the run.

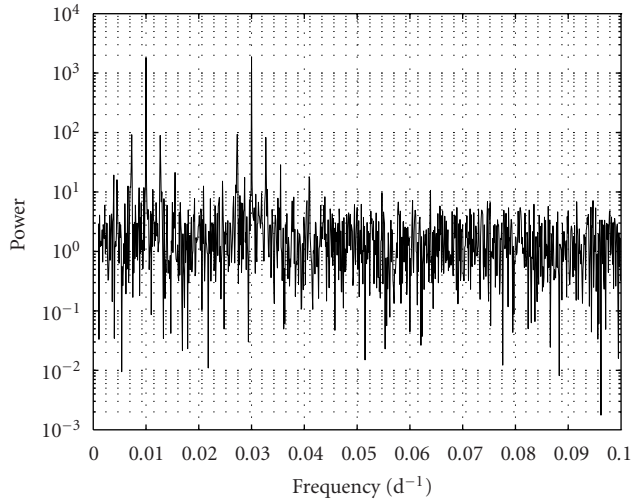
We then define two “sampling” signals for which each value is equal to the time step of the AC Herculis and RV Tauri data, respectively. Let t_i be the time when the data are available. The sampling signal values are $t_{i+1} - t_i$ (the wider the gap, the larger the value—see Figure 2b). The “sampling” signals are normalized so that their variances are the same as the corresponding simulated signals. These sampling signals are used in the two following sections: their wavelet transforms are compared to those of simulated and real signals.

4.1. Results from the simulations

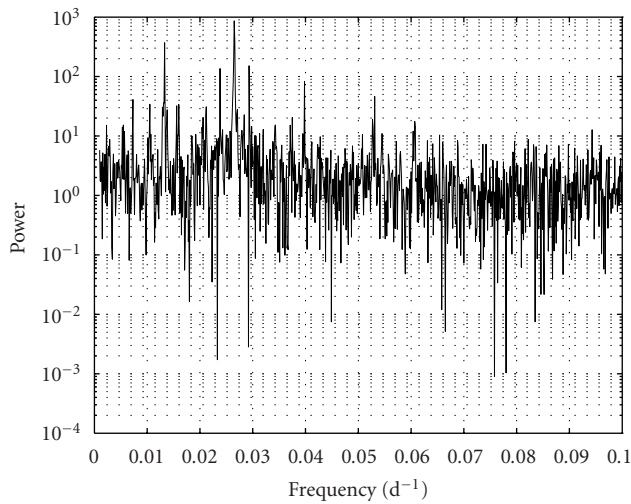
The wavelet power spectrum (WPS) and the global wavelet spectrum (GWS) obtained for both simulations are presented in Figures 5 and 6. The Torrence and Compo [28] approach provides statistical tools for establishing the validity of the results. This allows us to show in the GWS the 95% confidence level that would be obtained for white noise, which is the noise present in periodic variable stars observations as those studied here (the natural assumption of red noise is not adequate here because these data were obtained from different observers). This level (shown by the dashed line) is quasisuperposed on the x -axis (the linear interpolation induces errors comparable to white noise). In the WPS, the continuous white line indicates the cone of influence (zone where the edge effects are important). Information outside this cone is not relevant.

From an examination by eye of the WPS from the simulated signal on AC Herculis sampling (Figure 5b), one can identify the two simulated periods at 100.5 and 33.1 days. The corresponding GWS (Figure 5c) shows these two periods which are over the 95% confidence level. In the WPS from the simulated signal from the RV Tauri sampling (Figure 6b), one is also able to identify the two periods also at 100.5 and 33.1 days, although in a more indistinct way, and a third one at 230 days. The large zone due to the bad sampling at the beginning of the run, around $t = 36\,000$ modified Julian days (MJD = JD-2 400 000.5), appears as a large stain in the low frequencies. Although it is inside the cone of influence, it overshadows the presence of both periods. The GWS (Figure 6c) highlights the same problem.

Figures 9 and 10 present the GWS from simulated signals compared with those of the corresponding sampling signals. The purpose of this comparison is to discriminate, from the



(a)

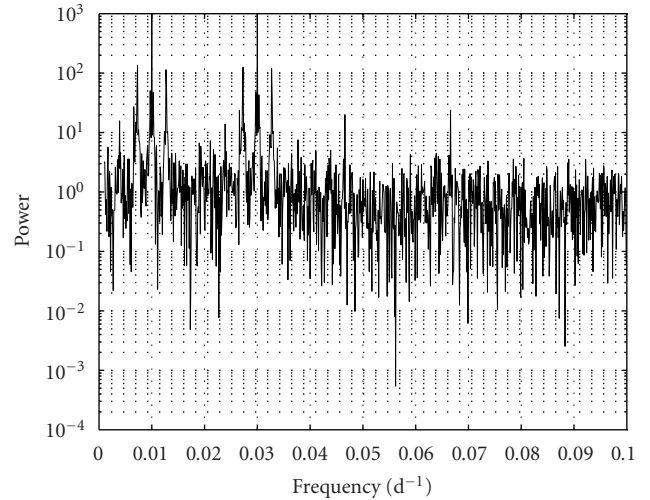


(b)

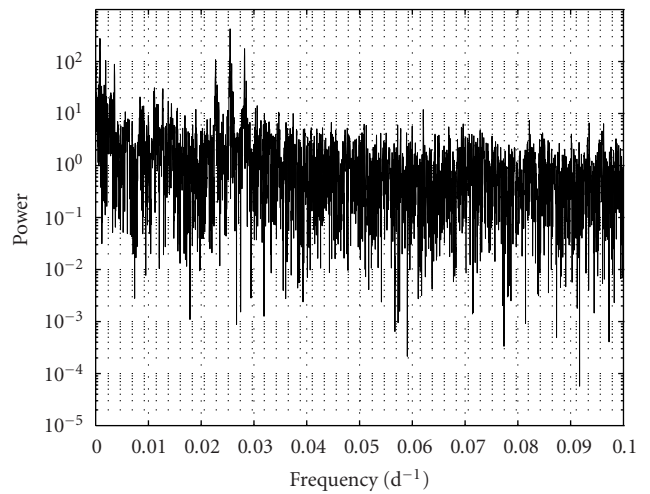
FIGURE 3: Periodograms from the AC Herculis light curve (b) and from the corresponding simulated signal (a), presented on a logarithmic scale.

results of the GWS, if the presence of a peak of significant amplitude can be explained or not by the irregularity of the sampling. In these simple cases, we can verify that the irregular sampling of RV Tauri is responsible for the peaks above roughly 500 days, detected in the GWS (Figure 10). Their nonvalidity was already confirmed by their position outside the cone of influence. However, the GWS from the sampling signal does not explain the 230-day period (see Section 5). Note also that for AC Herculis (Figure 9), the two structures visible at 365 days and 183 days obviously correspond to the annual cycle of the observations.

To complete our analysis, we examine the matching pursuit decomposition of the same simulations. The results presented here use the free graphical user interface developed at our institute (<http://webast.ast.obs-mip.fr/people/fbracher>)



(a)



(b)

FIGURE 4: Periodograms from the RV Tauri light curve (b) and from the corresponding simulated signal (a), presented on a logarithmic scale.

based on the LastWave software of Bacry available at <http://www.cmap.polytechnique.fr/~bacry/LastWave/>.

The linearly interpolated simulated signals are decomposed into functions from the dictionary \mathbf{D} with a Spline 0 window. The energy density of the 100 first atoms $(k_{\nu_1}(t), k_{\nu_2}(t), \dots, k_{\nu_{100}}(t))$ is shown on Figures 11 and 12. The long atoms represent the most coherent structures of the signal. The peaks which do not correspond to Stellar oscillations but are artifacts due to the sampling appear localized on a short time or cover a large frequency range (vertical atoms).

In the decomposition of the signal built from the AC Herculis sampling (Figure 11), we can identify two long atoms at exactly 0.03 day^{-1} and 0.01 day^{-1} , characteristic of the two frequencies introduced into the simulations. These are labeled 1 and 2 in the decomposition (cf. (6)).

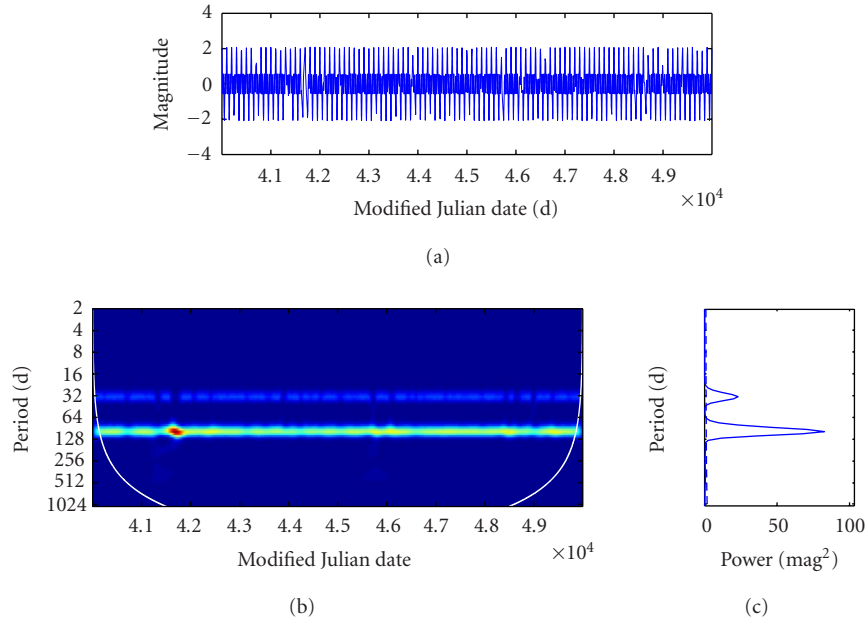


FIGURE 5: (a) Simulated signal built from the AC Herculis sampling (light curve). (b) Wavelet power spectrum. The continuous white line indicates the cone of influence. (c) Global wavelet spectrum of the simulated signal with a Morlet wavelet. The 95% confidence level that could be obtained for a white noise is shown by the dashed line.

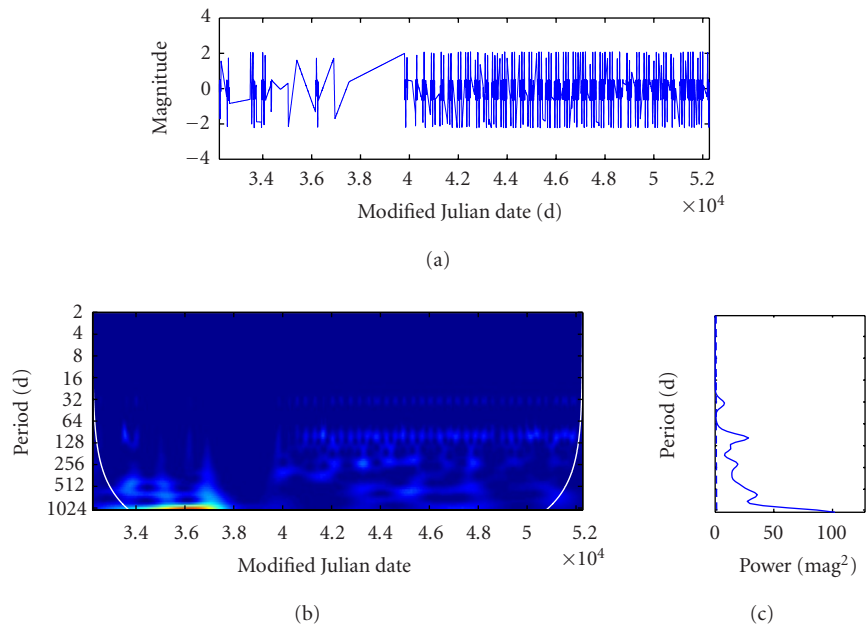


FIGURE 6: (a) Simulated signal built from the RV Tauri sampling (light curve). (b) Wavelet power spectrum. The continuous white line indicates the cone of influence. (c) Global wavelet spectrum of the simulated signal with a Morlet wavelet. The 95% confidence level that could be obtained for a white noise is shown by the dashed line.

The decomposition of the signal built from the RV Tauri sampling (Figure 12) presents an atom at 0.03 day^{-1} (with aliases corresponding to the annual cycle) in the second part of the data. Another atom appears at 0.01 day^{-1} with the same lifetime and aliases. As these atoms

are among the most energetic in the decomposition (numbers 2 and 3—number 1, the very first, being a low-frequency atom of short lifetime centered on $t = 4000 \text{ MJD}$), the simulated frequencies are also perfectly highlighted here.

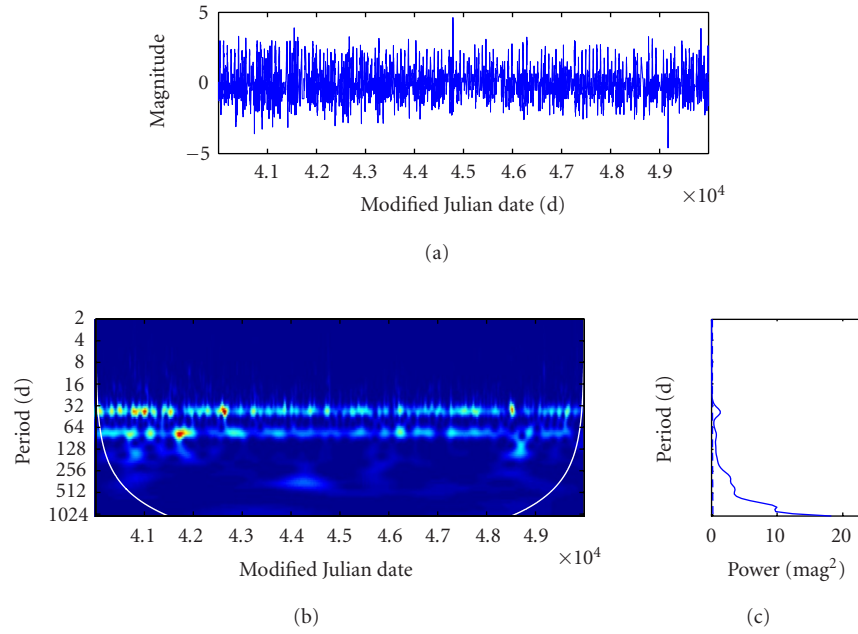


FIGURE 7: (a) AC Herculis light curve. (b) Wavelet power spectrum. The continuous white line indicates the cone of influence. (c) Global wavelet spectrum of the simulated signal with a Morlet wavelet. The 95% confidence level that could be obtained for a white noise is shown by the dashed line.

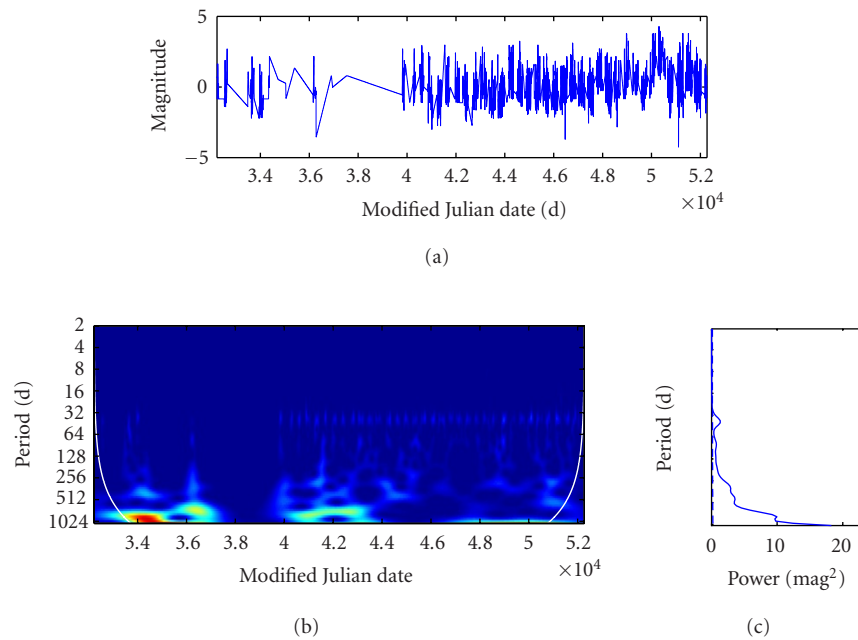


FIGURE 8: (a) RV Tauri light curve. (b) Wavelet power spectrum. The continuous white line indicates the cone of influence. (c) Global wavelet spectrum of the simulated signal with a Morlet wavelet. The 95% confidence level that could be obtained for a white noise is shown by the dashed line.

4.2. Application to the observations

The analysis of the AC Herculis and the RV Tauri light curves was conducted with the same methods. AC Herculis periodogram (Figure 3b) presents four important peaks

above 30 mag²: the first frequency (in terms of largest power) is centered at 0.0265 day⁻¹ (37.736 days), the following one at 0.0133 day⁻¹ (75.188 days), and the two others at 0.0398 day⁻¹ (25.126 days) and 0.0531 day⁻¹ (18.832 days). The first

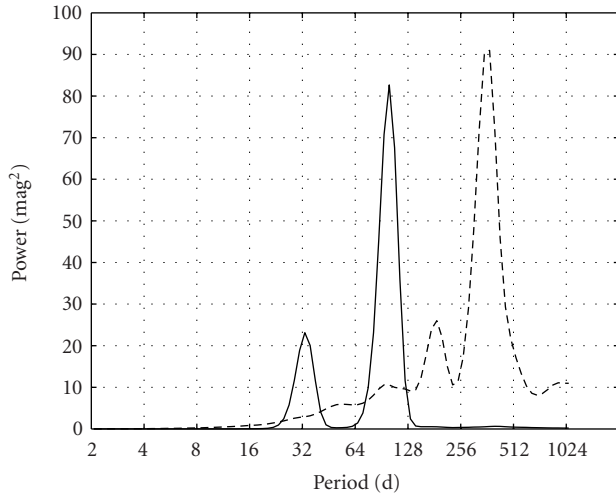


FIGURE 9: Global wavelet spectrum of the simulated signal made from the AC Herculis sampling (solid line) compared to the GWS from the corresponding sampling signal (dashed line).

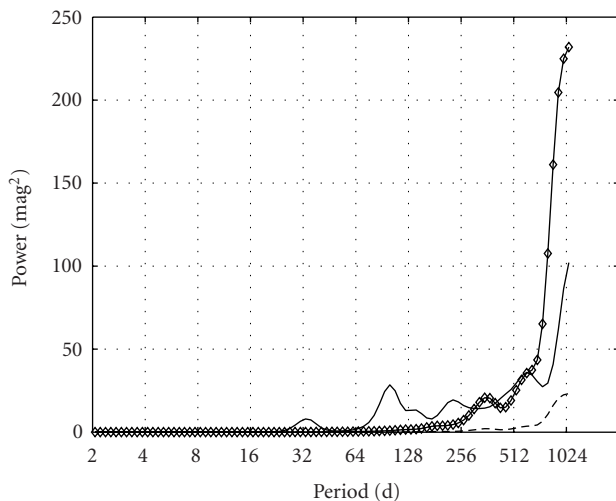


FIGURE 10: Global wavelet spectrum of the simulated signal made from the RV Tauri sampling (solid line) compared to the GWS from the corresponding sampling signal (dashed line). The solid line with diamonds represents the dashed line multiplied by a factor 10, for clarity.

two frequencies present aliases corresponding to the annual cycle of the observations. The known period of AC Herculis ($T_0 = 75.01$ days) is correctly identified, although it appears less energetic than the one at $T_0/2$, and its harmonics at $T_0/3$ and $T_0/4$ are also well detected. Finding half the known period, revealed by the wavelet analysis and the matching pursuit as well, will be discussed further (Section 4).

The RV Tauri periodogram (Figure 4b) reveals very noisy behavior, but the results indicate a first peak at $0.025471 \text{ day}^{-1}$ (39.26 days), corresponding to roughly half of the known period of this star, and a second one at 0.00082 day^{-1}

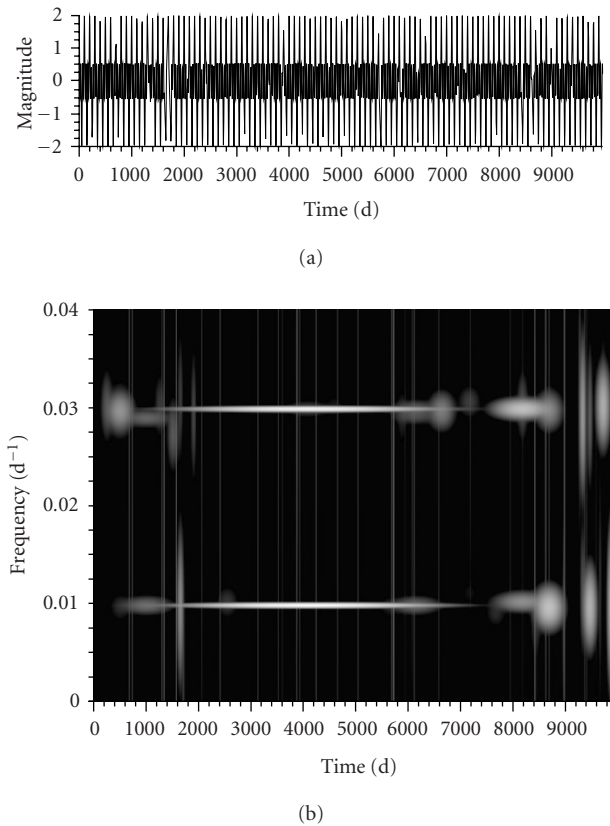


FIGURE 11: Time-frequency decomposition (b) of the simulated signal (a) made from the AC Herculis sampling. Each grey level represents the energy density of each atom.

(1219.5 days) which is the known long period of RV Tauri. Both peaks present aliases corresponding to the annual cycle of the observations.

WWZ analysis for AC Herculis reveals a high-amplitude peak at 38.24 days and a less energetic one at 71.43 days. For RV Tauri, one can surprisingly identify the most prominent peak at 1299 days and two others of same amplitude at 502.51 and 38.8 days (as expected). In both cases, the value of the constant c was chosen as for the simulated signals: $c = 0.04$ for AC Herculis and $c = 0.0125$ for RV Tauri. However, the heuristic choice of c makes it difficult to find evidence for exact periods. For example, if one chooses a lower or a higher value for c , periods and/or amplitudes appear to be slightly different. It seems that this point is not discussed in the literature.

Figures 7 and 8 present the WPS and the GWS from AC Herculis and RV Tauri, respectively, obtained with the same wavelet (Morlet) as for the simulated signals.

In AC Herculis WPS (Figure 7b), one is able to determine two periods: 37 days and 75 days. Both of them are present throughout the observation. Once again, the 37-day period appears just before (in terms of power) the 75-day period. In the corresponding GWS, these two periods appear over the 95% confidence level and two other less energetic peaks also appear above this level at 141 and 374 days. Note that the

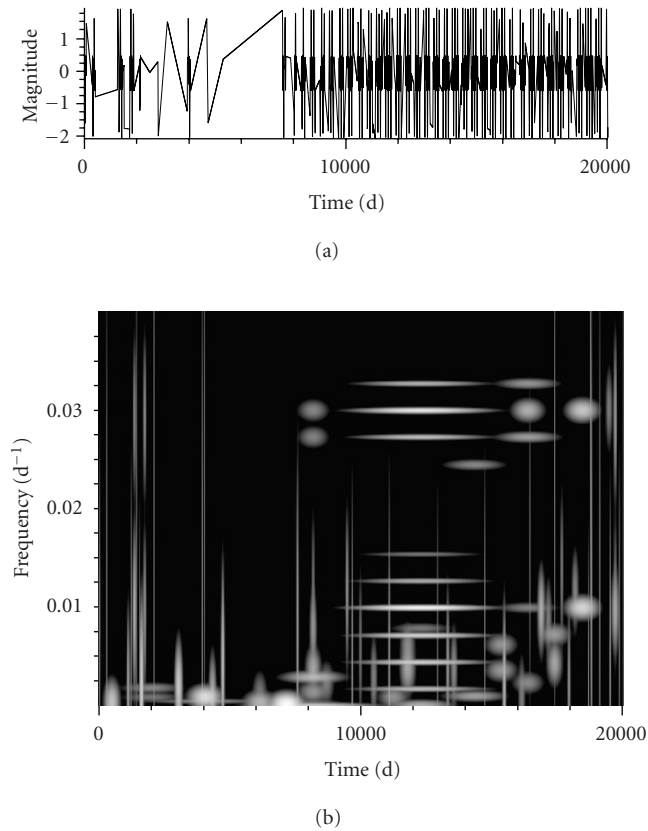


FIGURE 12: Time-frequency decomposition (b) of the simulated signal (a) made from the RV Tauri sampling. Each grey level represents the energy density of each atom.

harmonics at $T_0/3$ and $T_0/4$ detected in the periodogram do not appear in the WPS. The corresponding GWS is analyzed in Figure 13, superposed on that of the “sampling” signal. In fact, the nature of the sampling explains the last peak (374 days) and even the 141-day one, which are not real.

In the RV Tauri WPS (Figure 8b), only a single period at 39 days in the second part of the time interval can be identified. In the corresponding GWS (Figure 8c), it is also observable, together with some others above 500 days, but essentially outside the cone of influence. The GWS from RV Tauri is presented in Figure 14. The superposition on that of the corresponding sampling signal explains the peaks above 500 days, due to the sampling, which confirms that they are not real (as was already revealed by the cone of influence).

The matching pursuit analyses of AC Herculis and RV Tauri are computed with a Spline 0 window and keeping the first 100 atoms of the decomposition, as was done for the simulated signals (Figures 15 and 16). Several long lifetime atoms appear in the AC Herculis decomposition: the two most energetic ones are located at the same frequency: 0.0265 day^{-1} (37.73 days) and the third oscillates at 0.0133 day^{-1} (75.19 days). AC Herculis known period (75.01 days) is thus well determined, but as in the wavelet analysis, half the known period appears first in terms of energy.

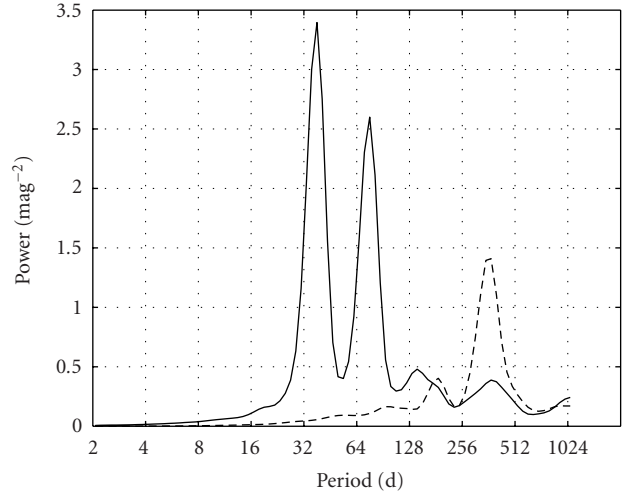


FIGURE 13: Global wavelet spectrum of the AC Herculis signal (solid line) compared to the GWS from the corresponding sampling signal (dashed line).

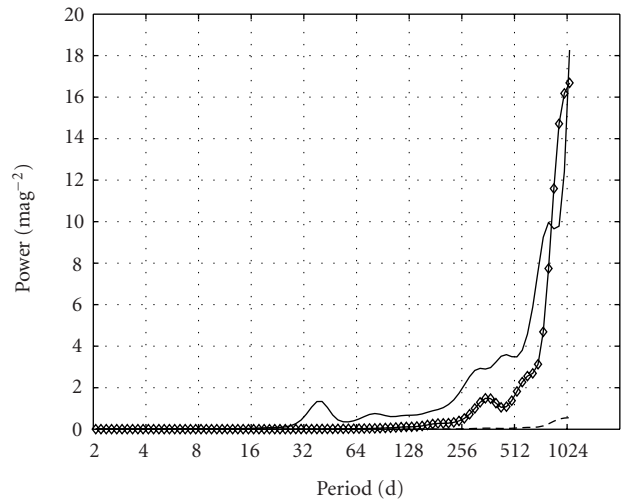


FIGURE 14: Global wavelet spectrum of the RV Tauri signal (solid line) compared to the GWS from the corresponding sampling signal (dashed line). The solid line with diamonds represents the dashed line multiplied by a factor 20, for clarity.

The harmonics revealed in the periodogram are also present with a high energy in the decomposition (atom 10, frequency: 0.0398 day^{-1} (25.12 days); atom 12, frequency: 0.0531 day^{-1} (18.83 days)).

Among the frequencies that can be identified in the RV Tauri diagram (Figure 16), the first long lifetime atom (order 4) is at 0.0255 day^{-1} (39.29 days) and is centered around 12 000 days. It represents the most energetic structure of the signal and is half the known oscillating period of this star (78.73 days). Two atoms (orders 15 and 20) are centered at 0.0227 day^{-1} (44.02 days) and 0.0282 day^{-1} (35.46 days). They correspond to aliases due to the annual cycle of observations. The $0.012695 \text{ day}^{-1}$ frequency (78.77 days) is in a more distant position (order 28) from an energy point of view.

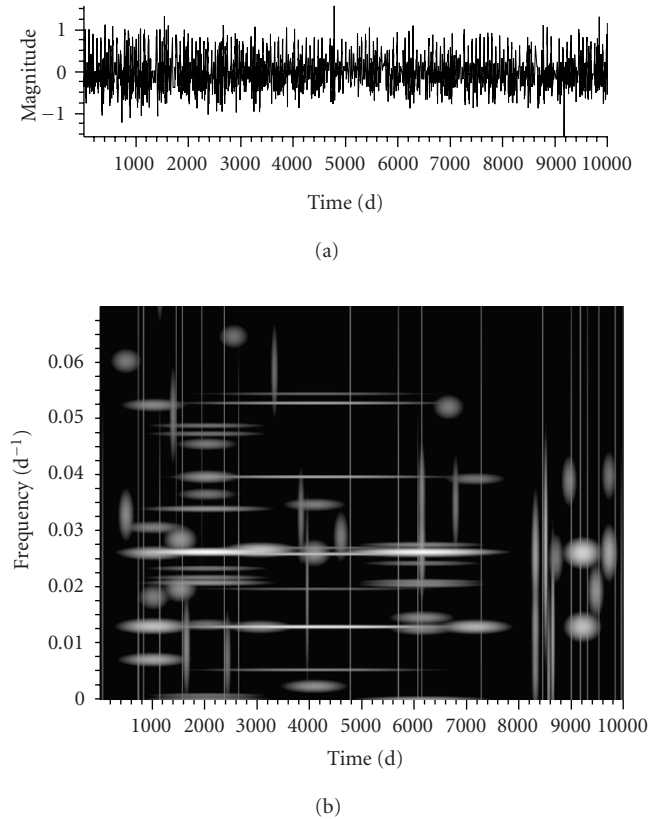


FIGURE 15: Time-frequency decomposition (b) of the AC Herculis signal (a).

Figure 17 represents the results of a matching pursuit decomposition of S Persei detailed in another paper [23], showing that this type of method can be applied to longer period stars. In that particular case, no long lifetime frequency can be identified by eye in the diagram, but the first atom of the decomposition is at 0.00122 day^{-1} (819 days) and centered around 10 000 days. This atom is not as long as the strong apparent periodicity in the last part of the data because the energy of a longer lifetime atom (double lifetime) would be smaller (it would take into account the noise on both sides of this part of the light curve). While neither the periodogram nor the wavelet analysis was able to find S Persei known period (822 days), the matching pursuit analysis offers the possibility of clearly detecting it, since it appears as the first atom of the decomposition.

5. DISCUSSION

Lomb's periodogram provides well-resolved power density spectra but it is oversensitive to irregular sampling; in some of our cases (e.g., S Persei), the periodogram was too noisy to be analyzed without a preliminary resampling. Moreover, although there exists a large literature concerning the statistical properties of the periodogram, its intrinsic nature does not allow us to implement a reconstruction error analysis comparable to the error provided by matching pursuit analysis.

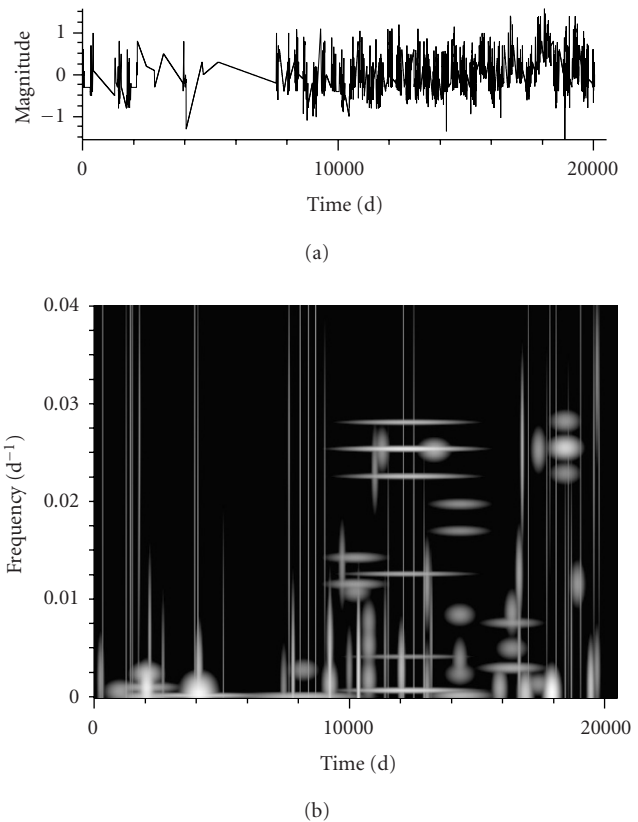


FIGURE 16: Time-frequency decomposition (b) of the RV Tauri signal (a).

The application of WWZ to the real data leads to the identification of known periods in the four stars. However, in the case of RV Tauri, the prominent period is 1299 days and not 38.8 days as expected. Different choices of c parameter did not allow us to improve the result. Concerning AC Herculis, S Persei and T Camelopardis, the most obvious features are consistent oscillations, as they are indicated in the fourth edition of the General Catalog of Variable Stars. We were not able to compute associated errors since, as indicated by Foster [11], their determination is "extraordinary complex."

The 95% confidence level associated with the WPS and the cone of influence are efficient tools to check if the highlighted frequencies are meaningful. However, in the regions of interpolated large gaps, neither the cone of influence nor the 95% confidence level is unable to discriminate spurious frequencies. But this method reveals problems for low frequencies: in the GWS, high-amplitude peaks are not always explained by those from the sampling signal. The poorest the sampling quality, the more important the problems at low frequencies. These problems are due to the linear interpolation, but other interpolation methods (e.g., cubic spline) lead to the same problem. This is what prevents us from carrying out an associated error analysis.

As for the periodogram, the matching pursuit analysis is oversensitive to the annual cycle of the observations, but the atom hierarchy provides quantitative information on the

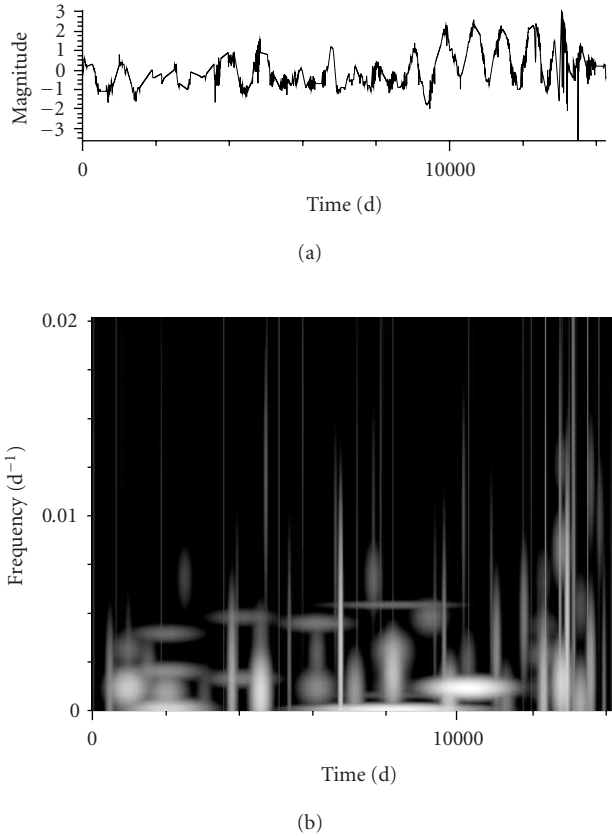


FIGURE 17: Time-frequency decomposition (b) of the S Persei signal (a).

dominating frequencies of the signal. We can easily define the associated error as the quadratic reconstruction error after decomposition on a large number of atoms (e.g., 400 in our case). Table 1 presents this error for simulated signals (third column) associated with the four variable stars presented in Section 1. The mean number of observations per day is used to estimate the quality of the sampling (second column in Table 1). One can note that the reconstruction error is not correlated with the sampling quality. This is explained by the large choice of atoms of different lifetimes and frequencies offered by the dictionary, which can solve the problem caused by a very incomplete sampling (see, e.g., the error reconstruction for T Camelopardis, compared to its sampling). Table 1 also presents the four variable stars periods and the ones found by the periodogram (P_{perio}), the GWS (P_{GWS}), the matching pursuit decomposition (P_{MP}), and the weighted wavelet Z -transform (P_{WWZ}).

We have noted throughout the paper the fact that for AC Herculis, we systematically found half the known period for this star. For RV Tauri, half the known period was also found. However, with the matching pursuit decomposition, we were also able to identify the known period (78.73 days). These two periods also found by Zsoldos [30] are characteristics of the double-wave shape of RV Tauri-like stars.

The matching pursuit decomposition appears to be particularly suitable for a deeper analysis of AC Herculis. Indeed, this star's light curve is supposed to be chaotic [15]. The matching pursuit analysis is probably the first step to use, towards a simple nonlinear dynamical analysis. Based on certain physical properties of such variable stars, a relevant atom selection allows us to reconstruct a signal characteristic of their structural properties. Note that the WWZ, also highlighting two periods, would not allow us to select atoms in the same simple manner. Our ongoing work should provide support for the results of Kollath et al. [15]. In the particular case of AC Herculis, our work could also explain why the found period (37.73 days) turns out to be half the period indicated in the literature (75.01 days) and why these periods are so close from an energetics point of view. In fact, we are probably facing a period-doubling phenomenon: if the star oscillates with a stable fundamental period, say $T_0 = 37.73$ days, when some parameters vary, a period-doubling bifurcation may occur, leading to another stable period of $2T_0 = 75.01$ days. Both periods can be observed, with a variable amplitude which depends on the run.

6. CONCLUSION

We have used a wavelet analysis and a matching pursuit decomposition to investigate the role of irregular sampling with linear interpolation in the determination of the spectral contents of variable star's light curves and we have compared them with results given by a Lomb's periodogram and by weighted wavelet Z -transform (which is a time-scale method that can be compared in its principle to Lomb's periodogram). The proposed algorithm is composed of two steps: first, a preprocessing is done by interpolating the irregularly sampled light curve; second, a time-scale or a time-frequency analysis is applied on resampled data.

Of course, the ability to analyze the frequency content of a data set, together with its time dependence, is by itself a powerful tool. To analyze unevenly spaced signals, stationary methods, although technically suited, fail to determine frequencies; nonstationary methods are better suited even at the price of a preliminary resampling. This incites a better use of time-scale or time-frequency methods. In this context, the results yielded by the matching pursuit analysis are significantly better in terms of ability not only to recover the right frequency (see Table 1), but also to conduct an error analysis and remain independent of the sampling quality. It is important to notice at this stage that if one had only kept the well-sampled parts of the signals, the results would of course have been immediate. Unfortunately, this procedure is impossible in the case of systematic astronomical observations which produce a large amount of data. This work is the first attempt to apply a matching pursuit algorithm to light curves of variable stars. We attached attention to interpolation procedure which at first sight could appear very simple. This is also the first time in astronomy where a periodogram is compared to global wavelet power spectra (i.e., 2D spectra), and to WWZ technique. Lastly, this is the first step to a chaotic

behavior analysis in the continuity of Kiss and Szatmáry's work [14].

The known period of AC Herculis at 75.01 days was found (in second position) at 75.19 days by the periodogram and the matching pursuit analysis, at 75.00 days by the GWS and at 71.43 by the WWZ. Harmonics at $T_0/2$, $T_0/3$, and $T_0/4$ have been clearly identified by the periodogram and the matching pursuit analysis. The two last (resp., three) harmonics do not appear in the GWS (resp., WWZ). The period of RV Tauri at 78.73 days was not found neither by the periodogram nor by the global wavelet spectrum nor by the weighted wavelet Z -transform. However, half of this period could be found at 39.26 days by the periodogram, at 39.00 days by the GWS, and at 38.80 by the WWZ. The matching pursuit decomposition reveals both periods: 39.29 days and 78.77 days. This last method also permits us to conduct an error analysis.

The matching pursuit algorithm thus appears well suited for spectral investigation of irregularly sampled variable stars signals. This study, moreover, offers the new benefit of simply requiring a linear interpolation of the data and allows us to propose a simple guideline for processing such signals:

- (1) resample the signal with a linear interpolation;
- (2) choose a time step compatible with the searched frequency range;
- (3) save this new signal as a column ascii file (zero mean);
- (4) download the matching-pursuit graphical user interface *guimauve* at webast.ast.obs-mip.fr/people/fbracher (Linux version preferred) and install it with the rpm command;
- (5) execute the command *guimauve*;
- (6) open the signal file, set the right time step (menu Signal), decompose the signal (menu Matching), choose the atom number to be investigated (100 by default) and the window (Gaussian by default). Options are activated by the mouse and the scroll bar;
- (7) information on atoms are written at the bottom of the window: ordered in hierarchy, lifetime of the oscillation ("octave" defines the length of the atom as 2^{octave}), time and frequency. The energy is quantitatively accessible thanks to the menu File and Save Decomposition. A reconstruction is possible from an atom or parameter selection.

Processing a simple linear interpolation before applying a time-frequency analysis offers advantages over a Fourier transform or a periodogram from nonresampled data. However, we plan to investigate, in a forthcoming work, a comparison of the matching pursuit results with linear interpolation of the data versus the interpolation technique proposed by Strohmmer [25].

ACKNOWLEDGMENTS

Wavelet Matlab software was provided by C. Torrence and G. Compo, and is available at paos.colorado.edu/research/

wavelets/. The corresponding article [28] is also downloadable at this URL. Matching pursuit software uses a free graphical user interface developed in our institute and available for Linux (and Windows for a simplest version) at <http://webast.ast.obs-mip.fr/people/fbracher/>. The code is based on LastWave software of E. Bacry (<http://www.cmap.polytechnique.fr/~bacry/LastWave/>). Weighted wavelet Z -transform was performed using the computer program WWZ developed by the American Association of Variable Star Observers (<http://www.aavso.org>). The authors are grateful to the staff of the Variable Star Network (VSNET) for their online database. The authors wish to thank Hervé Carfantan and Laurent Koechlin for helpful discussions on early versions of this paper and for their comments and suggestions. We gratefully thank Natalie Webb for her essential corrections.

REFERENCES

- [1] P. A. Bradley, "Theoretical models for asteroseismology of DA white dwarf stars," *Astrophysical Journal*, vol. 468, pp. 350–368, 1996.
- [2] J. R. Buchler, Z. Kollath, T. Serre, and J. Mattei, "Nonlinear analysis of the light curve of the variable star R scuti," *Astrophysical Journal*, vol. 462, pp. 489–504, 1996.
- [3] M. Carbonell, R. Oliver, and J. L. Ballester, "Power spectra of gapped time series: a comparison of several methods," *Astronomy & Astrophysics*, vol. 264, pp. 350–360, 1992.
- [4] C. Catala, J.-F. Donati, T. Böhm, et al., "Short-term spectroscopic variability in the pre-main sequence Herbig Ae star AB Aur during the MUSICOS 96 campaign," *Astronomy & Astrophysics*, vol. 345, pp. 884–904, 1999.
- [5] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 713–718, 1992.
- [6] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pa, USA, 1992.
- [7] C. de Boor, *A Practical Guide to Splines*, Springer, New York, NY, USA, 1978.
- [8] S. de Waele and P. M. T. Broersen, "Error measures for resampled irregular data," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 2, pp. 216–222, 2000.
- [9] G. G. Fahlman and T. J. Ulrych, "A new method for estimating the power spectrum of gapped data," *Monthly Notices of the Royal Astronomical Society*, vol. 199, pp. 53–65, 1982.
- [10] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego, Calif, USA, 1998.
- [11] G. Foster, "Wavelets for period analysis of unevenly sampled time series," *Astronomical Journal*, vol. 112, no. 4, pp. 1709–1729, 1996.
- [12] H. J. Haubold, "Wavelet analysis of the new solar neutrino capture rate data for the Homestake experiment," *Astrophysics and Space Science*, vol. 258, no. 1-2, pp. 201–218, 1997.
- [13] J. H. Horne and S. L. Baliunas, "A prescription for period analysis of unevenly sampled time series," *Astrophysical Journal*, vol. 302, no. 2, pp. 757–763, 1986.
- [14] L. L. Kiss and K. Szatmáry, "Period-doubling events in the light curve of R Cygni: Evidence for chaotic behaviour," *Astronomy & Astrophysics*, vol. 390, no. 2, pp. 585–596, 2002.
- [15] Z. Kollath, J. R. Buchler, T. Serre, and J. Mattei, "Analysis of the irregular pulsations of AC Herculis," *Astronomy & Astrophysics*, vol. 329, no. 1, pp. 147–155, 1998.

- [16] D. Koester and G. Chanmugam, "Physics of white dwarf stars," *Reports on Progress in Physics*, vol. 53, no. 7, pp. 837–915, 1990.
- [17] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [18] R. E. Nather, D. E. Winget, J. C. Clemens, C. J. Hansen, and B. P. Hine, "The whole earth telescope—a new astronomical instrument," *Astrophysical Journal*, vol. 361, pp. 309–317, September 1990.
- [19] A. Oppenheim and R. Schaffer, *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.
- [20] D. H. Roberts, J. Lehar, and J. W. Dreher, "Time series analysis with CLEAN. I. Derivation of a spectrum," *Astronomical Journal*, vol. 93, no. 4, pp. 968–988, 1987.
- [21] S. Roques, A. Schwarzenberg-Czerny, and N. Dolez, "Parametric spectral analysis applied to gapped time-series of variable stars," *Baltic Astronomy*, vol. 9, no. 1, pp. 463–477, 2000.
- [22] S. Roques, B. Serre, and N. Dolez, "Band-limited interpolation applied to the time series of rapidly oscillating stars," *Monthly Notices of the Royal Astronomical Society*, vol. 308, no. 3, pp. 876–886, 1999.
- [23] S. Roques and C. Thiebaut, "Spectral contents of astronomical unequally spaced time-series: contribution of time-frequency and time-scale analyses," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '03)*, vol. 6, pp. 473–476, Hong Kong, China, April 2003.
- [24] J. D. Scargle, "Wavelet methods in astronomical time series analysis," in *Applications of Time Series Analysis in Astronomy and Meteorology*, T. S. Rao, M. B. Priestley, and O. Lessi, Eds., pp. 226–248, Chapman & Hall, New York, NY, USA, 1997.
- [25] T. Strohmmer, "Computationally attractive reconstruction of band-limited images from irregular samples," *IEEE Trans. Image Processing*, vol. 6, no. 4, pp. 540–548, 1997.
- [26] K. Szatmáry, J. Vinkó, and J. Gál, "Application of wavelet analysis in variable star research. I. Properties of the wavelet map of simulated variable star light curves," *Astronomy and Astrophysics Supplement Series*, vol. 108, pp. 377–394, 1994.
- [27] C. Thiebaut, M. Boer, and S. Roques, "Steps towards the development of an automatic classifier for astronomical sources," in *Astronomical Data Analysis II*, J.-L. Starck and F. D. Murtagh, Eds., vol. 4847 of *Proceedings of SPIE*, pp. 379–390, Waikoloa, Hawaii, USA, August 2002.
- [28] C. Torrence and G. B. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological Society*, vol. 79, no. 1, pp. 61–78, 1998.
- [29] J. Ville, "Théorie et applications de la notion de signal analytique," *Cables et Transmission*, vol. 2A, no. 1, pp. 61–74, 1948 (French).
- [30] E. Zsoldos, "RV Tauri and the RVB phenomenon. I. Photometry of RV Tauri," *Astronomy and Astrophysics Supplement Series*, vol. 119, no. 3, pp. 431–437, 1996.

and analysis of astrophysical signals with time-frequency methods. With the support of a postdoctoral fellowship from the French Space Agency (CNES), she started working on optical observations of space debris in October 2003 for the CNES. Since September 2004, she is working on onboard data processing in CNES.

S. Roques was born in Narbonne (Aude), France, in 1956. She attended the University Paul Sabatier at Toulouse, where she received the M.S. degree in solid-state physics in 1980 and a Ph.D. degree in electron microscopy in 1982. She carried on her research at the "Centre National de la Recherche Scientifique" (CNRS), the French National Science Council. She is the Director of Research at the CNRS since 2000. She has published about 75 papers in electron microscopy, optics, astronomy, and image processing. Her current research deals with the reconstruction problems in signal and image processing with particular reference to universe sciences applications. She is presently the Managing Director of "Laboratoire d'Astrophysique" of Midi-Pyrénées Observatoire (Toulouse, France).



C. Thiebaut was born in North of France in 1977. She received an Engineering degree of electronics and signal processing in 2000 from the Institut National Polytechnique de Toulouse (ENSEEIH). She has got her Ph.D. degree in September 2003 in signal and image processing with a Fellowship from the Centre National de la Recherche Scientifique (CNRS). During her three years in the Center for the Study of Radiation in Space (CESR), she has led studies on astronomical image analysis, object detection, automatic classification using neural networks,

