

Analyses temps-échelle et temps-fréquence de séries temporelles astronomiques échantillonnées irrégulièrement

Carole THIEBAUT¹, Sylvie ROQUES² et Michel BOER¹

¹CESR-OMP-CNRS, 9 av. du Colonel Roche, BP 4346, 31028 Toulouse Cedex 4

²LAT-OMP-CNRS, 14 av. Edouard Belin – 31400 Toulouse

carole.thiebaut@cesr.fr, roques@ast.obs-mip.fr, michel.boer@cesr.fr

Résumé – Nous analysons comment des méthodes temps-échelle et temps-fréquence sont appropriées à l'étude de séries temporelles astronomiques échantillonnées irrégulièrement, en terme de recouvrement spectral précis d'une étoile variable. Nous proposons deux types de méthodes qui nous permettent de traiter ce type de signaux : une première approche associée à un spectre global en ondelettes comparable à un spectre de Fourier, une seconde approche temps-fréquence (poursuite adaptée – « matching pursuit ») associée à un processus de construction progressive nous permettant d'identifier les structures significatives du signal étudié. Nous appliquons ces méthodes à des simulations et à des courbes de lumière (évolution de la luminosité de l'étoile en fonction du temps) d'étoiles variables. Nous discutons et comparons les résultats obtenus. Cette étude nous permet de conclure que la poursuite adaptée est la méthode la plus efficace, en particulier, elle est indépendante de la qualité de l'échantillonnage des données étudiées.

Abstract – We analyse 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. We propose to study two kinds of methods which enables us to treat this type of signals: first, an approach with global wavelet spectra that can be compared to the Fourier spectrum. Second, a time-frequency analysis (matching pursuit) providing a progressive constructive process allowing us to identify the fine structure of the suited signal. 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 sampling quality.

1. Introduction et contexte

L'échantillonnage non-uniforme est un problème qui apparaît dans beaucoup de domaines de l'astronomie, notamment lors de l'observation des étoiles variables. Le manque de données est essentiellement dû à l'alternance jour/nuit, aux conditions d'observation (mauvais temps) et aux problèmes techniques des moyens d'observation. Les signaux associés à l'évolution de la luminosité de l'étoile en fonction du temps, appelés aussi courbes de lumière, sont donc échantillonnés irrégulièrement avec un taux d'échantillonnage moyen de quelques jours. Les fréquences extraites de ces observations sont souvent une importante source d'informations pour les astronomes [10].

Le périodogramme modifié présenté dans [4, 8] est une technique standard utilisée pour extraire les fréquences des signaux échantillonnés irrégulièrement. Ces signaux doivent être assez longs pour obtenir une bonne résolution spectrale. La durée de vie des fréquences présentes dans le signal est également une information importante pour l'analyse des courbes de lumière. Contrairement au périodogramme, les analyses temps-fréquence [3] et temps-échelle [1] sont des outils qui peuvent répondre à cette attente. Mais ces méthodes s'appliquent sur des signaux échantillonnés régulièrement. Il existe des méthodes de ré-échantillonnage complexes largement répandues dans la littérature [9], mais l'interpolation linéaire est une méthode robuste et facile à mettre en œuvre [2]. Les courbes de lumière sont donc interpolées linéairement avec un pas d'interpolation égal à un jour (les étoiles variables étudiées ont des périodes de plus de 50 jours).

Deux types d'approches sont proposées : une transformée en ondelettes et le spectre global associé [11], une décomposition Matching Pursuit [6] (Poursuite Adaptative). Les résultats sont comparés à ceux du périodogramme pour des signaux présentant des échantillonnages de différentes qualités (voir le nombre d'observations/jour). La section suivante présente les approches proposées. Les résultats sont exposés dans la section 3 et la discussion et les conclusions font l'objet de la dernière partie de cet article.

2. La méthode proposée

La décomposition en ondelettes correspond à l'expansion d'une fonction arbitraire sur des contributions localisées caractérisées par une échelle et un paramètre temporel. La transformée globale en ondelettes [11] que nous considérons ici correspond à une analyse en ondelettes continue [1], ce qui nous permet de définir un spectre global d'ondelettes comme le carré du module des coefficients en ondelettes pour chaque échelle.

L'algorithme de poursuite adaptative, développée par Mallat & Zhang [6] nous permet de choisir dans un dictionnaire redondant de formes d'ondes temps-fréquence, un ensemble de vecteurs qui correspondent le mieux au signal, grâce à un processus itératif de projections unidimensionnelles. Le dictionnaire est composé d'une famille de fonctions temps-fréquence générées par dilatation, translation et modulation d'une fonction analysante, réelle et unique $k_{\nu}(t)$. Dans notre cas, la courbe de lumière $s(t)$, après ré-échantillonnage et mise à moyenne nulle, est approximée par un vecteur $k_{\nu 0}$ du dictionnaire tel que $|\langle s(t), k_{\nu 0}(t) \rangle|$ soit la

plus grande possible. Le signal $s(t)$ est donc décomposé sous la forme $s(t) = \langle s(t), k_{v0}(t) \rangle \cdot k_{v0}(t) + R_s(t)$ avec $R_s(t)$ le vecteur résiduel après l'approximation de $s(t)$ dans la direction k_{v0} . On sous-décompose ensuite le vecteur résidu de la même manière : on trouve ainsi un second vecteur $k_{v1}(t)$. La procédure est réitérée sur le résidu obtenu et un critère de convergence existe. Le signal est finalement décomposé en $\sum_{i=0}^{\infty} \langle R^i s(t), k_{vi}(t) \rangle \cdot k_{vi}(t)$ avec $k_{vi}(t)$ les atomes qui correspondent le mieux aux structures du signal. La hiérarchie des structures cohérentes $\{k_{vi}(t)\}$ nous permet de construire la distribution d'énergie temps-fréquence du signal.

3. Les applications

Les résultats sont comparés à ceux du périodogramme pour des signaux présentant des échantillonnages de différentes qualités. Les courbes de lumière étudiées sont celles de quatre étoiles variables de différents types obtenues par le réseau des observateurs japonais [12] :

- *T Camelopardis* (T Cam) : P=373.2 jours, étoile de type Mira, 0.051 observations/jour.
- *S Persei* (S Per) : P=822 jours, étoile de type Semi-régulière, 0.112 observations/jour.
- *AC Herculis* (AC Her) : P=75.01 jours, étoile de type RV Tauri, 0.757 observations/jour
- *RV Tauri* (RV Tau) : P=78.73 jours, étoile de type RV Tauri, 0.190 observations/jour.

Les caractéristiques de ces objets sont issues du « General Catalog of Variable Stars » [13]. Les résultats relatifs à la variable S Per sont présentés dans [7]. Nous présentons ici les résultats relatifs à la variable AC Her pour la période d'observation JD 2 440 000-2 450 000. La courbe de lumière associée est caractérisée par une double oscillation caractéristique des variables du type RV Tauri. L'échantillonnage d'AC Her est de meilleure qualité que celui de S Per.

Nous construisons des signaux simulés associés à chaque courbe de lumière réelle.

3.1 Application aux signaux simulés

Les signaux synthétiques sont la somme de deux cosinus de période 100 et 33.3 jours et leur échantillonnage est celui de la variable associée. Les données synthétiques sont volontairement non bruitées pour uniquement analyser l'effet de l'échantillonnage irrégulier et comment les méthodes proposées y répondent. La figure 1 (gauche) présente le périodogramme du signal simulé associé à AC Her (voir Fig.2 a) pour la courbe de lumière) : les deux fréquences mises en jeu sont visibles. Les pics situés de part et d'autre de ces deux fréquences correspondent au cycle annuel d'observation (la différence de fréquence est égale à $1/365 \text{ jour}^{-1}$).

Le spectre de puissance en ondelettes et le spectre global associé obtenus pour le signal simulé sont présentés Fig.2. Pour le spectre global, le niveau de confiance à 95% obtenu pour un bruit blanc est superposé avec l'axe des périodes. Pour le spectre de puissance, la ligne continue blanche

correspond au cône d'influence qui nous indique la zone dans laquelle les effets de bord ne sont plus négligeables. Les deux périodes du signal simulé sont visibles sur les deux spectres.

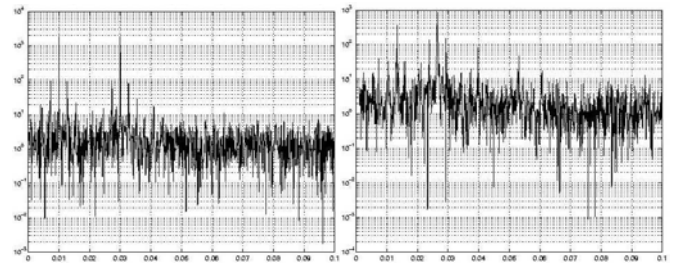


FIG.1 : Périodogramme du signal simulé associé à AC Her (gauche) et de la courbe de lumière d'AC Her (droite) en échelle logarithmique.

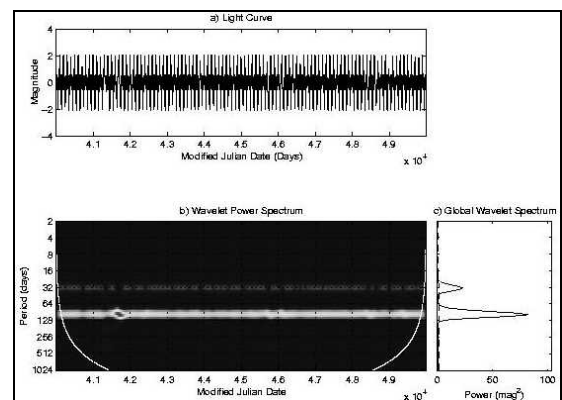


FIG.2 : a) Signal simulé, b) Spectre en puissance d'ondelettes du signal simulé, c) Spectre global en ondelettes associé.

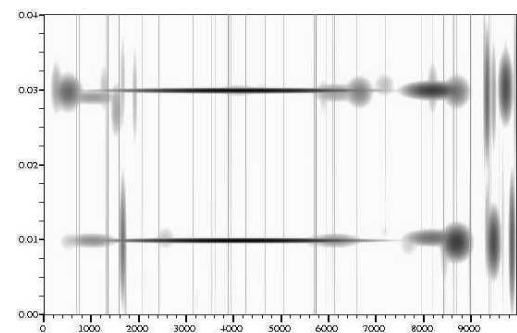


FIG.3 : Décomposition temps-fréquence du signal simulé associé à AC Her. Abscisse : temps (jours). Ordonnée : fréquence (jour^{-1}).

La décomposition temps-fréquence associée à la poursuite adaptée est présentée Fig.3. Les résultats de cette approche sont obtenus avec l'interface graphique publique [14] développée dans notre institut. Le signal est décomposé sur 100 atomes avec une fenêtre Spline 0. Les longs atomes représentent les structures cohérentes du signal étudié. Les pics qui ne correspondent pas aux oscillations des étoiles, mais à des artefacts, sont localisés sur un intervalle de temps restreint et une large gamme de fréquence (atomes verticaux).

Nous identifions deux longs atomes à 0.01 et 0.03 jour^{-1} caractéristiques des deux fréquences du signal simulé. Leur ordre d'apparition dans la hiérarchie est 1 et 2.

3.2 Application aux données réelles

Nous analysons la courbe de lumière d'AC Her de la même façon. Le périodogramme de cette étoile (Fig.1, droite) présente quatre pics importants au-dessus de 30 mag^2 : la première fréquence (en terme de puissance) est centrée sur 0.0265 jour^{-1} (37.736 jours), la suivante sur 0.0133 jour^{-1} (75.188 jours) et les deux autres sur les fréquences 0.0398 jour^{-1} (25.126 jours) et 0.0531 jour^{-1} (18.832 jours). Les deux premières fréquences présentent des pics voisins liés au cycle annuel d'observation. La période connue d'AC Her ($T_0=75.01$ jours) est correctement identifiée, même si elle apparaît moins énergétique que $T_0/2$. Ses harmoniques à $T_0/3$ et $T_0/4$ sont détectés. Le fait de trouver une demi-période plus énergétique que la période connue sera discuté dans la dernière section.

Le spectre en puissance et le spectre global en ondelettes associé sont présentés Fig.4. Le spectre en puissance met en évidence deux périodes à 37 et 75 jours. Elles sont toutes les deux présentes tout au long de l'observation. La période à 35 jours apparaît, cette fois encore, plus énergétique. Dans le spectre global associé, nous identifions deux autres fréquences au-dessus du niveau de confiance à 95%, situées à 141 et 374 jours. Notons que les harmoniques à $T_0/3$ et $T_0/4$ n'apparaissent pas dans cette approche. Nous définissons un signal « échantillonnage » dont chaque valeur est égal au pas temporel des données d'AC Her (soit t_i les instants auxquels nous disposons de données pour AC Her, les valeurs du signal « échantillonnage » correspondant sont égales à $t_{i+1}-t_i$). Nous superposons (Fig. 5) le spectre global des données d'AC Her et du signal « échantillonnage » associé pour discriminer les pics qui peuvent être expliqués par l'échantillonnage irrégulier. Les deux pics présents à 141 et 374 jours peuvent être expliqués par l'échantillonnage. Dans le cas de la variable S Per [10], le signal « échantillonnage » ne pouvait pas toujours expliquer les pics présents dans les basses fréquences.

L'analyse temps-fréquence de la courbe de lumière d'AC Her est réalisée avec une fenêtre Spline 0. La décomposition sur les 100 premiers atomes est présentée Fig.6. Nous observons plusieurs atomes de longue durée : les deux plus énergétiques possèdent la même fréquence 0.0265 jour^{-1} (37.73 jours) et le troisième plus énergétique est centrée sur 0.0133 jour^{-1} (75.19 jours). La période connue d'AC Her est bien identifiée, mais la demi-période correspondante est plus énergétique comme pour les deux autres approches. Les harmoniques détectés dans le périodogramme sont également présents dans la décomposition temps-fréquence avec une forte valeur d'énergie (atome 10 : fréquence 0.0398 jour^{-1} (25.12 jours) ; atome 12, fréquence 0.0531 jour^{-1} (18.83 jours)).

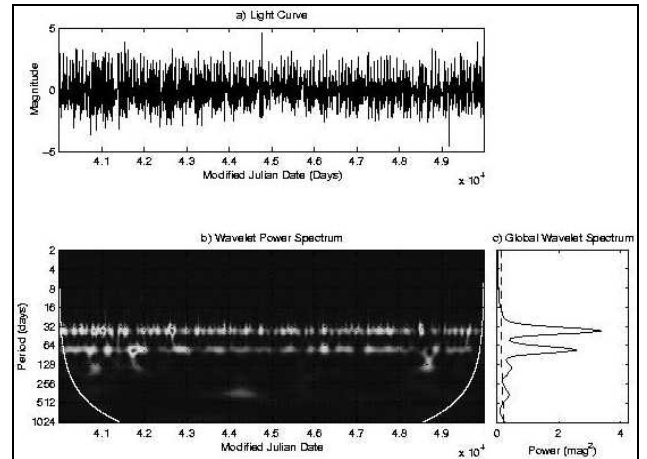


FIG.4 : a) Courbe de lumière d'AC Her, b) Spectre en puissance d'ondelettes du signal réel, c) Spectre global en ondelettes associé.

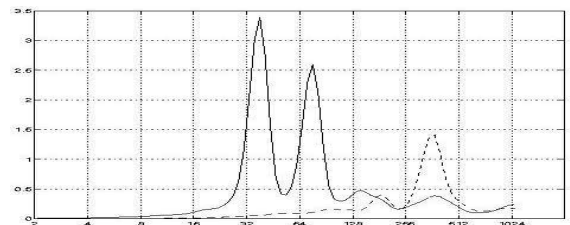


FIG.5 : Spectre global en ondelettes du signal réel (courbe continue) et du signal « échantillonnage » associé (courbe pointillée).

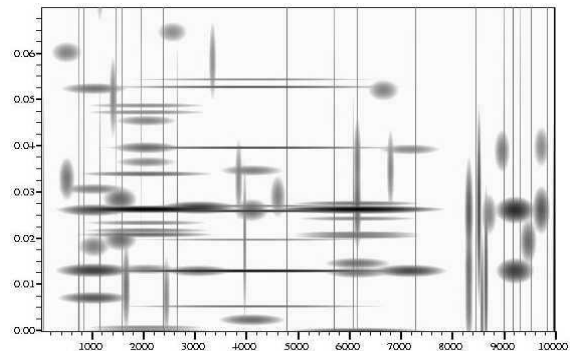


FIG.6 : Décomposition temps-fréquence du signal simulé associé à AC Her. Abscisse : temps (jours). Ordonnée : fréquence (jour^{-1}).

4. Discussion et conclusion

Le périodogramme fournit des spectres bien résolus mais il est très sensible à l'échantillonnage irrégulier ; pour certaines variables, le spectre est trop bruité pour pouvoir identifier les fréquences du signal sans un ré-échantillonnage préliminaire (voir le spectre bruité de S Per [10]).

Le niveau de confiance à 95% et le cône d'influence associés respectivement au spectre global d'ondelettes et au spectre de puissance sont des outils efficaces pour vérifier si

les fréquences détectées sont réelles. Mais cette approche présente quelques problèmes dans les basses fréquences : dans le spectre global en ondelettes, les pics de forte amplitude ne sont pas toujours expliqués par ceux du signal «échantillonnage». Plus la qualité de l'échantillonnage est faible, plus ces problèmes dans les basses fréquences sont importants.

Comme pour le périodogramme, la poursuite adaptée est sensible au cycle annuel d'observation, mais la hiérarchie des atomes nous fournit une information quantitative sur les fréquences dominantes du signal. Nous définissons une erreur associée à cette approche comme l'erreur de reconstruction quadratique après décomposition du signal sur un large nombre d'atomes (400 dans notre cas). La table 1 (dernière ligne) présente ces erreurs pour les signaux simulés correspondants aux quatre variables présentées. Nous notons que cette erreur n'est pas corrélée avec la qualité de l'échantillonnage des données étudiées (l'erreur de reconstruction de T Cam est très faible alors que le signal est très dégradé). La table 1 présente également les périodes connues des quatre variables (Période_{GCVS}) et celles obtenues par les différentes approches : Période_{Péριο} pour le périodogramme, Période_{GWS} pour le spectre global en ondelettes et Période_{MP} pour la poursuite adaptative.

Dans cet article, nous avons trouvé que la moitié de la période connue d'AC Her est toujours plus énergétique que la période connue. Cette étoile est supposée chaotique d'après l'étude [5]. La poursuite adaptative est probablement le premier pas vers une analyse dynamique non-linéaire. Une sélection appropriée des atomes de la décomposition, basée sur les propriétés physiques de l'étoile étudiée, nous permet de reconstruire un signal caractéristique de ces propriétés physiques. Une étude en cours de développement à ce sujet sur des variables du type Mira, RV Tauri et semi-régulières pourra apporter une contribution aux résultats de [5]. Nous avons en particulier déjà trouvé un exposant de Lyapunov maximal égal à 0.00092 pour AC Herculis, confirmant l'hypothèse du comportement chaotique. Dans ce cas particulier, ce comportement explique aussi pourquoi la période la plus énergétique mise en évidence est la moitié de la période connue d'AC Her. En fait, nous sommes probablement face à un problème de doublage de période : si l'étoile oscille avec une période fondamentale stable, $T_0=37.73$ jours, lors de la variation de paramètres, une bifurcation de double période peut apparaître, menant à une autre période stable $2T_0=75.01$ jours. Les deux périodes peuvent être observées avec une amplitude variable et dépendant du moment d'observation.

Nous avons utilisé une analyse en ondelettes et une décomposition par poursuite adaptée pour comprendre le rôle de l'échantillonnage irrégulier dans la détermination du contenu spectral des courbes de lumière relatives à des objets variables. Nous avons comparé les résultats de ces méthodes avec ceux du périodogramme. La possibilité de connaître la durée de vie de ce contenu spectral fournie par ces deux

analyses est un atout important. Dans le cas de signaux non-stationnaires, c'est une évidence, et pour les signaux quasi-stationnaires échantillonnés irrégulièrement, pour lesquelles les méthodes stationnaires échouent, c'est également adapté. Les résultats de l'analyse temps-fréquence sont les meilleurs en terme de capacité à retrouver les bonnes fréquences. De plus, cette analyse nous a permis de définir une erreur associée et cette erreur ne dépend pas de la qualité de l'échantillonnage des données étudiées.

Nous rappelons que la période connue d'AC Her égale à 75.01 jours est déterminée (en seconde position) à 75.19 jours par le périodogramme et la poursuite adaptée, et à 75.00 jours par le spectre global d'ondelettes. Les harmoniques $T_0/2$, $T_0/3$ et $T_0/4$ ont été clairement identifiés par le périodogramme et la décomposition temps-fréquence. Les deux derniers harmoniques ne sont pas détectés par l'analyse temps-échelle.

TAB.1 : Tableau des résultats pour les quatre variables

	AC Her	S Per	RV Tau	T Cam
Obs./jour	0.757	0.112	0.190	0.051
Période _{GCVS}	75.01	822.00	78.73	373.20
Période _{Péριο}	37.74	806.40	39.26	371.75
Période _{GWS}	37.00	826.00	39.00	373.00
Période _{MP}	37.73	819.00	39.29	372.30
Erreur MP	3.71%	11.22%	8.53%	10.98%

References

- [1] Daubechies I., *Ten Lectures on wavelets*, SIAM, Philadelphia, 1992.
- [2] De Waele S. & Broersen P.M.T., *IEEE Trans. Instr. Measurements*, 49 (2), 216, 2000.
- [3] Flandrin P., *Time-Frequency/Time-scale analysis*, Academic Press, 1998.
- [4] Horne J.H. & Baliunas S.L., *Astroph. Journ.*, 301, 757, 1986.
- [5] Kollath Z. et al., *Astron. & Astroph.*, 329, 147 1998.
- [6] Mallat S. & Zhang Z., *IEEE Trans. Signal Proc.*, 41(12), 3397, 1993.
- [7] Roques S. & Thiebaut C., to appear in Proc. of ICASSP'03, Hong Kong, 2003.
- [8] Scargle J.D., *Astrophys. Journ.*, 263(2), 835-853, 1982.
- [9] Strohmmer T., *IEEE Trans. Image Proc.*, 6(4), 540, 1997.
- [10] Thiebaut C. et al., *Astron. Data Analysis II*, Starck & Murtagh Eds, Proc. SPIE, 4847, 379, 2002.
- [11] Torrence C. & Compo G.B., *Bull. Am. Meteo. Soc.*, 79(1), 61, 1997.
- [12] <http://vsnet.kusastro.kyoto-u.ac.jp/vsnet/index.html>.
- [13] <http://www.sai.msu.su/groups/cluster/gcvs/>.
- [14] <http://webast.ast.obs-mip.fr/people/fbracher>.