

Sur l'utilisation de rapports de périodogrammes pour la détection de raies dans un bruit de densité spectrale partiellement inconnue

Sophia SULIS, David MARY, Lionel BIGOT

Laboratoire Lagrange,
Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Blvd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France
Sophia.Sulis@oca.eu

Résumé – Nous proposons une méthode de détection de raies dans un bruit de densité spectrale (DSP) partiellement inconnue. L'originalité de cette méthode est de ne pas poser un modèle paramétrique sur la DSP mais de s'appuyer sur des simulations du processus sous l'hypothèse nulle. Nous montrons que l'utilisation de rapports de périodogrammes permet de calibrer la procédure de test, qui devient asymptotiquement indépendante de la DSP. L'intérêt de l'approche est illustré sur des séries temporelles synthétiques et réelles concernant la détection d'exoplanètes.

Abstract – We propose a detection method operating when the Power Spectral Density (PSD) of the noise is partially unknown. Our method does not rely on a parametric model of the PSD but uses instead numerical simulations of the process under the null hypothesis. We show that using ratios of periodograms allows an efficient calibration of the testing procedure, which becomes asymptotically independent of the PSD of the noise. We illustrate the good behavior of the method on synthetic and real data in the framework of extrasolar planet detection.

1 Introduction

Cette étude est motivée par la détection de planètes extrasolaires. Les futures missions spatiales (e.g PLATO [1]) permettront d'avoir accès à de longues séries temporelles d'observation d'étoiles cibles (mois, années) avec un pas d'échantillonnage très fin ($< \text{min}$). À l'aide de ces nouveaux instruments, il sera possible d'extraire des données de mesures de flux (photométrie) ou de vitesses radiales (spectroscopie) provenant de l'étoile cible. Dans ce dernier cas, considéré ici, la signature d'une planète tellurique apparaîtra comme une composante périodique essentiellement sinusoidale, de très faible amplitude et se superposant à un bruit.

La recherche moderne d'exoplanètes via cette technique est marquée par deux développements qui modifient substantiellement l'aspect méthodologique de la détection. Premièrement, le bruit des détecteurs à bord sera si faible et la durée d'acquisition si longue que la détection sera essentiellement limitée par le "bruit" que constitue le signal de l'étoile elle-même – un processus aléatoire très complexe qui dépend de l'étoile. Deuxièmement, les avancées en modélisation hydrodynamique (HD) à 3D et en puissance de calcul permettent de modéliser de façon très précise ce "bruit stellaire" [2]. Ces progrès ouvrent la perspective de recourir à des simulations HD pour estimer de façon fiable les statistiques du bruit sous l'hypothèse nulle (planète absente), sans avoir à construire des modèles paramétriques *ad hoc*, nécessairement très complexes pour être réalistes, et qui rendraient le contrôle statistique de la procédure de test problématique.

La recherche de composantes périodiques dans des séries temporelles est un des problèmes les plus anciens et les plus féconds dans le domaine de la détection. Schuster [3] a posé en 1898 les bases statistiques du problème de détection en intro-

duisant le périodogramme :

$$P(\nu) := \frac{1}{N} \left| \sum_{j=1}^N X(t_j) e^{-i2\pi\nu t_j} \right|^2, \quad (1)$$

où ν est la fréquence et $X(t)$ une série obtenue en N dates t_1, \dots, t_N . Ce périodogramme est à la base de la grande majorité des méthodes de détection existantes. Celles-ci reposent essentiellement sur la recherche d'une ou quelques valeurs de P anormalement élevées sous l'hypothèse nulle \mathcal{H}_0 , et les statistiques des données sous \mathcal{H}_0 conditionnent le contrôle des erreurs de première espèce (fausses alarmes). Or, ces statistiques ne sont quasiment jamais connues en pratique. Même dans le cas simple d'un processus $X(t)$ régulièrement échantillonné (i.e., à des dates $t_j : \forall j \in \{1, \dots, N-1\}, t_{j+1} - t_j = \Delta_t$ avec $\Delta_t \in \mathbb{R}^{+*}$), stationnaire de second ordre et gaussien sous \mathcal{H}_0 , le réglage de la significativité du test nécessite d'en estimer la densité spectrale de puissance (DSP). Le contrôle du bruit statistique de cette estimation constitue donc un point critique pour l'interprétation des tests en pratique.

Un test de référence, celui de Fisher [4], se place dans le cadre où $X(t_j = j\Delta_t)$ est issu d'un bruit blanc (indépendant et identiquement distribué, i.i.d.) gaussien de variance inconnue. Dans ce cas, ce test est uniformément plus puissant (UPP) contre les alternatives "monochromatiques" (une seule composante sinusoidale sous \mathcal{H}_1) [5]. De nombreuses variantes existent dans le cas "polychromatique" [6], chap. 6, et [5], [7]).

Dans le cadre de cette étude, nous supposons que sous \mathcal{H}_1 une seule sinusoïde d'amplitude, fréquence et phase inconnues se superpose à un bruit de DSP essentiellement inconnue. Dans un tel cas, les approches de détection existantes reposent sur une étape d'estimation de la DSP à l'aide des données, par exemple en choisissant *a priori* un modèle paramétrique de la DSP de type ARMA ([8], [9]), ou encore en utili-

sant un périodogramme moyennant les données sur des fenêtres paramétrées [6], ou sur une sélection multi-critères des coefficients spectraux/cepstraux [10]. En astronomie, la détection est souvent faite en combinant plusieurs méthodes (préblanchissement, suppression de dérives locales, ... [11]).

Notre approche se démarque de ces travaux sur le point suivant : nous choisissons de ne poser aucun modèle paramétrique *a priori* sur la DSP sous \mathcal{H}_0 et de ne nous fier pour son estimation qu'aux simulations. Nous supposons que nous disposons pour cela d'un ensemble de réalisations indépendantes du processus sous \mathcal{H}_0 sous la forme de L séries $X_\ell(t_j)$, $j = 1, \dots, N$, $\ell = 1, \dots, L$, obtenues ici par simulations HD et échantillonnées sur la même grille que les données. La moyenne empirique des périodogrammes de ces séries simulées est alors utilisée pour calibrer le périodogramme des données (et, de fait, toute la procédure de détection). Notons que l'erreur de modèle ne disparaît pas avec cette approche, elle existe à travers : i) les biais possibles des simulations HD ; nous négligerons cet effet ici ; ii) le bruit stochastique d'estimation lié à la valeur finie de L . En effet, une série simulée de 10^2 jours nécessite typiquement 3 mois de calculs sur 120 coeurs. Ce coût de calcul signifie que L est et restera pour plusieurs années de l'ordre de 10^2 au plus, ce qui pose la question de l'étude de l'impact du bruit d'estimation en l'absence de tout biais.

La Sec. 2 présente le modèle et rappelle les propriétés asymptotiques du périodogramme. La Sec. 3 présente les tests considérés. Le contrôle des fausses alarmes est étudié en Sec. 4., et la Sec. 5 présente des résultats numériques.

2 Modèle statistique

Nous considérons les hypothèses :

$$\begin{cases} \mathcal{H}_0 : X(t_j) = \epsilon(t_j) \\ \mathcal{H}_1 : X(t_j) = A \sin(2\pi f t_j + \phi) + \epsilon(t_j) \end{cases} \quad (2)$$

où la série temporelle $X(t_j)$, $j = 1, \dots, N$ est issue d'un échantillonnage régulier sur une durée totale $T = N\Delta_t$, $\epsilon(t)$ est un bruit gaussien stationnaire du second ordre, de DSP notée $S(\nu)$. L'amplitude A , la fréquence f et la phase ϕ sont inconnues. Dans l'application astrophysique considérée, les composantes stellaires et planétaires sont respectivement modélisées par $\epsilon(t)$ et par le terme harmonique.

Le périodogramme de Schuster (Eq.(1)) peut être vu comme un estimateur non-paramétrique de la DSP. Dans la suite, nous nous limiterons aux fréquences de Fourier $\{\nu_k = \frac{k}{T}\}$, $k = 0, \dots, \frac{N}{2}$, en supposant N pair sans perte de généralité.

Rappelons quelques propriétés remarquables de cet estimateur. Sous certaines conditions sur les cumulants, le périodogramme est un estimateur asymptotiquement non biaisé mais inconsistant de la DSP (cf [12], Chap. 5, théorèmes 5.2.1 et 5.2.4). Cependant, asymptotiquement, le périodogramme évalué à deux fréquences (de Fourier) différentes ν_k et ν_n est non seulement décorrélié, mais aussi indépendant et distribué selon (cf [12], Chap. 5, théorème 5.2.6) :

$$P(\nu_k) \sim \begin{cases} \frac{S(\nu_k)}{2} \chi_2^2 & \text{pour } k = 1, \dots, \frac{N}{2} - 1, \\ S(\nu_k) \chi_1^2 & \text{pour } k = 0 \text{ et } k = \frac{N}{2}. \end{cases} \quad (3)$$

Supposons maintenant que l'on dispose de L séries temporelles indépendantes $X_\ell(t_j)$, issues du processus sous \mathcal{H}_0 et échantillonnées sur la même grille. Un estimateur consistant de la DSP de ce processus peut être obtenu par :

$$P_{moy}(\nu) := \frac{1}{L} \sum_{\ell=1}^L \frac{1}{N} \left| \sum_{j=1}^N X_\ell(t_j) e^{-i2\pi\nu t_j} \right|^2. \quad (4)$$

En vertu des théorèmes ci-dessus et de l'indépendance des $X_\ell(t_j)$, $\ell = 1, \dots, L$, on a en effet $\lim_{L \rightarrow \infty} \mathbb{E}[P_{moy}(\nu)] = S(\nu_k)$ et

$\lim_{L \rightarrow \infty} \text{var}[P_{moy}(\nu)] = \lim_{L \rightarrow \infty} S(\nu_k)^2 / L = 0$. De plus, les séries $X_\ell(t_j)$ étant indépendantes, on a asymptotiquement en N et sous les hypothèses des théorèmes ci-dessus :

$$P_{moy}(\nu_k) \sim \begin{cases} \frac{S(\nu_k)}{2L} \chi_{2L}^2 & \text{pour } k = 1, \dots, \frac{N}{2} - 1, \\ \frac{S(\nu_k)}{L} \chi_L^2 & \text{pour } k = 0 \text{ et } k = \frac{N}{2}. \end{cases} \quad (5)$$

Définissons enfin le "périodogramme rapport" comme le périodogramme P normalisé (ou calibré) par un périodogramme moyen obtenu via L réalisations indépendantes sous \mathcal{H}_0 :

$$P_r(\nu_k) := \frac{P(\nu_k)}{P_{moy}(\nu_k)}. \quad (6)$$

Rappelons que pour deux variables aléatoires indépendantes $V_1 \sim \chi_{d_1}^2$ et $V_2 \sim \chi_{d_2}^2$, on a : $\frac{V_1/d_1}{V_2/d_2} \sim F(d_1, d_2)$, avec F une loi de Fisher-Snedegor de densité φ_F [13]. On déduit de (3) et (5) que sous \mathcal{H}_0 , P_r se distribue asymptotiquement suivant :

$$P_r(\nu_k) \sim \begin{cases} \frac{S(\nu_k) \chi_{2L}^2 / 2L}{2S(\nu_k) \chi_{2L}^2} = \frac{\chi_{2L}^2 / 2}{\chi_{2L}^2 / 2L} \sim F(2, 2L), & k = 1, \dots, \frac{N}{2} - 1, \\ \frac{\chi_L^2}{\chi_L^2 / L} \sim F(1, L) & , k = 0 \text{ et } k = \frac{N}{2}. \end{cases} \quad (7)$$

3 Statistiques de tests considérées

Nous illustrons ici l'effet de (6) via l'approche de Fisher [4], mais cette calibration peut clairement être appliquée à toute statistique de test basée sur le périodogramme P (e.g., [5],[7]).

Par souci de concision et toujours sans perte de généralité nous ne considérons que les fréquences $\nu_0 < \nu_k < \nu_{\frac{N}{2}}$. Nous noterons de plus dans la suite les statistiques d'ordre de P par $\min_k P(\nu_k) = P_{(1)} < P_{(2)} < \dots < P_{(\frac{N}{2}-1)} = \max_k P(\nu_k)$.

3.1 Test de Fisher

Le test de Fisher est défini par :

$$T_{Fi} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \gamma, \quad \text{avec} \quad T_{Fi} := \frac{P_{(\frac{N}{2}-1)}}{\sum_{k=1}^{\frac{N}{2}-1} P_{(k)}}, \quad (8)$$

et $\gamma \in \mathbb{R}^+$ est un seuil qui contrôle le taux de fausses alarmes. Dans le cas où ϵ est un bruit blanc de variance σ^2 dans le modèle (2), avec $S(\nu) = \sigma^2 = Cte$, on montre facilement avec (3) que le dénominateur de (8) est à une constante près un estimateur asymptotiquement non biaisé de σ^2 (et donc de S) [6]. Or, dans le cas où la variance est connue, le test du Rapport de Vraisemblance Généralisé (RVG) correspond à $P_{(N/2-1)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \gamma$ [14]. Le test de Fisher peut donc être vu comme une calibration du RVG par la variance (ou la DSP) estimée sous \mathcal{H}_0 .

Soulignons que le test (8) est très utilisé en astronomie [15], probablement en raison de sa simplicité et de son caractère UPP pour un bruit blanc [15], même si cette hypothèse n'est pas toujours valide. L'étude des effets d'une DSP "colorée" (cf Sec. 5) sur ses performances nous a mené à proposer les tests suivants.

3.2 Tests basés sur le périodogramme rapport

D'après (7), le périodogramme rapport P_r (6) est, pour N grand, identiquement distribué : le rapport supprime la dépendance en fréquence liée à la DSP sous \mathcal{H}_0 . D'après la discussion en Sec. 3.1, il est naturel de proposer $P_r(\nu_k)$ comme statistique de test à chaque fréquence et de rechercher la composante d'amplitude maximale. En utilisant la même notation que précédemment pour les statistiques d'ordre de P_r , on a :

$$T_r \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \gamma, \quad \text{avec} \quad T_r := P_{r(\frac{N}{2}-1)}. \quad (9)$$

Le périodogramme rapport (6) étant asymptotiquement i.i.d. (sur $0 < k < \frac{N}{2}$), on peut aussi suivre l'approche de Fisher en appliquant le test (8) à P_r au lieu de P :

$$T_{F_{i,r}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \gamma, \quad \text{avec} \quad T_{F_{i,r}} := \frac{P_{r(\frac{N}{2}-1)}}{\sum_{k=1}^{\frac{N}{2}-1} P_{r(k)}}. \quad (10)$$

4 Distributions asymptotiques sous \mathcal{H}_0 et taux de fausses alarmes associés

Le contrôle précis des fausses alarmes constitue un point critique. Sous \mathcal{H}_0 , les statistiques de test (8), (9) et (10) font toutes intervenir le maximum d'un ensemble de $\frac{N}{2} - 1$ variables aléatoires $T_{F_i}(n)$, $T_r(n)$ et $T_{F_{i,r}}(n)$ dont les définitions et les distributions sont, en utilisant (5) et (7) pour $0 < n < \frac{N}{2}$:

$$\begin{aligned} T_{F_i}(n) &:= \frac{P(n)}{\sum_{k=1}^{\frac{N}{2}-1} P(k)} \underset{as, ni, i.i.d.}{\sim} \frac{\chi_2^2 S(n)/2}{\sum_{k=1}^{N/2-1} \chi_2^2 S(k)/2}, \\ T_r(n) &:= P_r(n) \underset{as, i.i.d.}{\sim} F(2, 2L), \\ T_{F_{i,r}}(n) &:= \frac{P_r(n)}{\sum_{k=1}^{\frac{N}{2}-1} P_r(k)} \underset{as, ni, i.i.d.}{\sim} \frac{F(2, 2L)}{\sum_{k=1}^{N/2-1} F(2, 2L)}, \end{aligned} \quad (11)$$

où nous avons mis en évidence le caractère asymptotique en N (as), l'indépendance et l'équidistribution ($i.i.d.$) ou non (n).

En ce qui concerne T_{F_i} , les $T_{F_i}(n)$ ne sont pas i.i.d. ; la statistique de test dépend de plus de la DSP S (non connue). Ceci rend impossible le contrôle *a priori* du taux de fausse alarme.

Dans le cas de T_r , la densité des $T_r(n)$ est donnée par :

$$\varphi_F(x, 2, 2L) = \frac{1}{B(1, L)} \cdot \frac{1}{L} \cdot (1 + \frac{x}{L})^{-L-1} = (1 + \frac{x}{L})^{-L-1}, \quad (12)$$

où on a utilisé l'expression générale de φ_F [13] et la valeur de la fonction Beta incomplète $B(1, L) = \int_0^1 (1-t)^{L-1} dt = \frac{1}{L}$. On peut vérifier que $\varphi_F(\gamma, 2, 2L)$ est normée à 1. La fonction de répartition $\Phi_F(\gamma, 2, 2L) := \Pr(F(\gamma, 2, 2L) \leq \gamma)$ s'obtient en utilisant l'expression générique $F(\gamma, d_1, d_2) = I_{\frac{d_1 \gamma}{d_1 \gamma + d_2}}(\frac{d_1}{2}, \frac{d_2}{2})$ [13] ou par intégration directe de $\varphi_F(\gamma, 2, 2L)$ et mène à :

$$\Phi_F(\gamma, 2, 2L) = 1 - (\frac{L}{\gamma + L})^L. \quad (13)$$

La probabilité de fausse alarme asymptotique s'obtient alors facilement grâce à l'indépendance asymptotique des $T_r(n)$:

$$\begin{aligned} P_{FA}(T_r, \gamma) &:= \Pr(\max_n T_r(n) > \gamma | \mathcal{H}_0) = 1 - \prod_n \Pr(T_r(n) > \gamma | \mathcal{H}_0) \\ &= 1 - (\Phi_F(\gamma, 2, 2L))^{\frac{N}{2}-1} = 1 - (1 - (\frac{L}{\gamma + L})^L)^{\frac{N}{2}-1}, \end{aligned} \quad (14)$$

qui est une expression remarquablement simple et, comme nous allons le voir en Sec. 5, fiable pour de petites valeurs de N ($N \gtrsim 10^3$ dans l'exemple considéré).

Enfin, dans le cas de $T_{F_{i,r}}$, les $T_{F_{i,r}}(n)$ sont asymptotiquement ni.i.d. et suivent une loi de rapport d'une variable F sur une somme de variables F . À notre connaissance il n'existe pas de description analytique de la loi résultante. Soulignons que pour l'application visée (où $N \approx 10^7$), une approche par simulations de Monte Carlo est numériquement hors de portée.

5 Résultats numériques

Dans le but de faire des simulations de type Caractéristique Opérationnelle de Réception (CORE) et d'analyser le comportement des tests, il est utile de connaître la DSP théorique sous \mathcal{H}_0 . De plus, il n'est pas possible de faire des CORE sur des séries où $N \approx 10^7$ pour L grand. On utilise donc ici dans un premier temps un bruit synthétique à N petit (1024, 10^4 simulations). Un important travail de simulations est en cours sur le mésocentre de l'OCA afin de simuler des données solaires obtenues sur 18 ans par l'instrument GOLF du satellite SOHO. Nous présenterons de premiers résultats en fin de section.

Le bruit stellaire $\epsilon(t)$ généré sous \mathcal{H}_0 que nous considérons d'abord est issu d'un processus AR : la courbe bleue en haut de la Fig. 1 représente sa DSP. Cette forme est « représentative » d'un bruit stellaire dans le sens où les basses fréquences sont plus énergétiques en raison de divers phénomènes (convection, activité magnétique) ; localement, à plus hautes fréquences, certaines zones peuvent être "contaminées" par des modes de pulsation (la "bosse" autour de 4 mHz ici). Les lignes verticales indiquent où le signal est injecté sous \mathcal{H}_1 : nous avons choisi deux cas extrêmes et opposés : signatures de même amplitude mais situées dans des régions spectrales respectivement très (f_1) et peu (f_2) énergétiques de la DSP.

La vignette de droite sur la même ligne évalue la fiabilité du contrôle de la P_{FA} pour le test T_r par l'Eq. (14), dans le cas où $N = 1024$, pour $L = 5$ (rouge : Eq.(14), vert : observé), $L = 20$ (bleu : Eq.(14), cyan : observé) et $L = 100$ (noir : Eq.(14), magenta : observé). L'accord est clairement satisfaisant.

Les deux vignettes de la ligne 2 comparent les CORE obtenues pour les tests T_{F_i} (8, magenta), T_r (9, noir) et $T_{F_{i,r}}$ (10, vert). Dans le cas des tests impliquant P_r , le style de ligne code la valeur de L (voir légende). La CORE de gauche correspond à f_1 et celle de droite à f_2 . Dans le cas de f_1 (dans une "bosse" de la DSP), le test classique de Fisher T_{F_i} apparaît plus puissant que les autres. Dans le cas de f_2 (dans un creux de la DSP), les performances sont inversées. Pour comprendre cet effet, il est intéressant d'observer sur quelles fréquences se distribuent sous \mathcal{H}_0 les maxima des tests $\max_n T_{F_i}(n)$ (Fig. 1, ligne 4, pa-

nel du haut) et $\max_n T_r(n)$ (panel du bas). Les fausses alarmes sur le test non calibré T_{Fi} sont majoritairement dans la zone de forte énergie de la DSP, $[0 - 1]$ mHz (et idem pour la probabilité de détection, non montrée). Dans le cas de T_r en revanche, le taux de fausse alarme est uniforme en fréquence. Les zones de forte énergie de la DSP du bruit ont pour effet de focaliser le test sur ces zones. Quand le signal sous \mathcal{H}_1 s'y trouve, cet effet agit comme une connaissance *a priori* favorable ; par contre, les signaux en dehors de ces zones sont indétectables. A l'inverse, les tests T_r et $T_{Fi,r}$ permettent une bonne détectabilité sur toute la plage de fréquences, avec des performances proches de la valeur asymptotique pour $L \approx 10^2$.

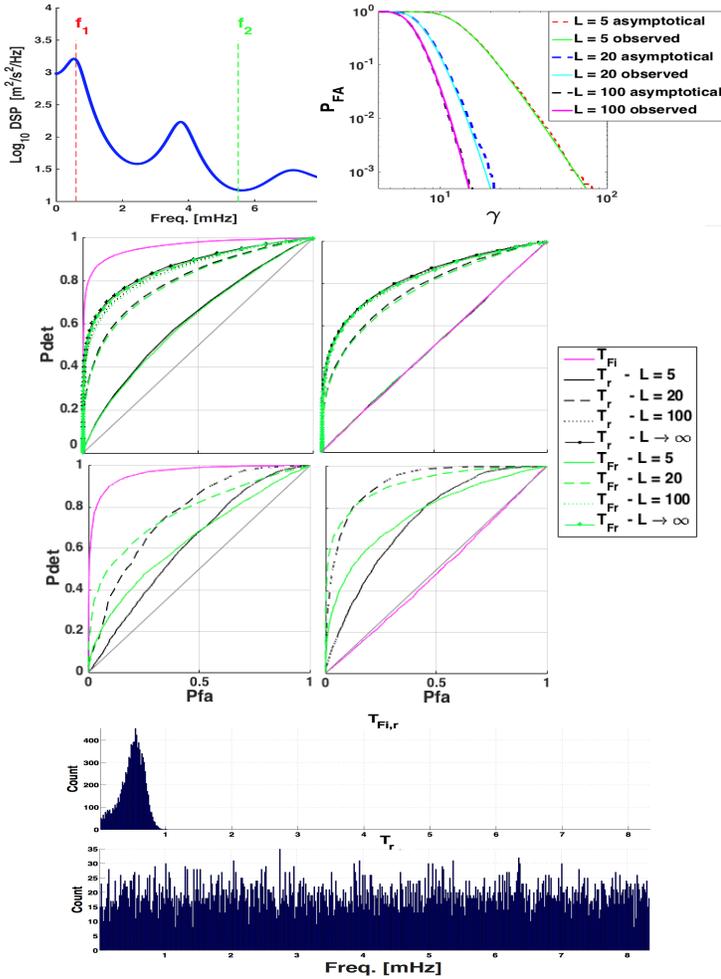


FIGURE 1 – Ligne 1 : DSP synthétique (gauche) ; Évaluation de la fiabilité de (14), droite ; Ligne 2 : CORE sur AR ; Ligne 3 : CORE sur données GOLF ; Ligne 4 : Pour processus AR, estimation de la distribution en fréquence sous \mathcal{H}_0 de $\max_n T_{Fi,r}(n)$ (haut), $\max_n T_r(n)$ (bas).

La ligne 3 de la Fig. 1 présente des résultats dans le cas où les données et le processus simulé sous \mathcal{H}_0 sont obtenus à partir de données réelles (GOLF) : dans ce cas, le dénominateur de (6) est construit grâce à des blocs pris aléatoirement dans les données. Ces résultats constituent encore un cadre idéal

lisé puisqu'en pratique la calibration sera obtenue par simulations HD. Les triplets $(A; f; \phi)$ simulés sous \mathcal{H}_1 sont $(A_1 = 0, 3 \text{ m/s}; f_1 = 7, 5 \text{ mHz}; 0^\circ)$ et $(A_2 = 1 \text{ m/s}; f_2 = 7, 5 \text{ mHz}; 0^\circ)$. Courbes de gauche : cas où f_1 est dans une "bosse" de la DSP solaire ; droite : f_2 est dans un "creux". Les paramètres sont $N = 1000$, $\Delta_t = 20 \text{ s}$, et le nombre de simulations de Monte Carlo est 1272 ($L = 20$) et 4452 ($L = 5$). Les résultats sont très comparables à ceux obtenus sur le bruit AR. Notons enfin que dans tous les cas : 1) $T_{Fi,r}$ et T_r (vert et noir) donnent des résultats similaires, mais seul T_r permet de contrôler précisément la P_{FA} avec (14) ; 2) les performances des tests basés sur le périodogramme rapport augmentent logiquement avec L , l'estimation de la DSP devenant plus précise.

Remerciements :

Nous remercions Thales Alenia Space, la région PACA ainsi que le projet TITAN du défi MASTODONS du CNRS pour le cofinancement de ces travaux.

Références

- [1] H. Rauer et al., "The PLATO 2.0 mission," *Experimental Astronomy*, vol. 38, pp. 249–330, 2014.
- [2] L. Bigot et al., "The diameter of the CoRoT target HD 49933," *Astronomy and Astrophysics*, vol. 534, no. 3, 2011.
- [3] A. Schuster, "On the investigation of hidden periodicities," *Journal of Geophysical Research*, vol. 3, pp. 13, 1898.
- [4] R. A. Fisher, "Tests of Significance in Harmonic Analysis," *Royal Soc. of London Proc. Series A*, vol. 125, pp. 54–59, 1929.
- [5] A. Siegel, "Testing for periodicity in a time series," *Journal of the American Stat. Assoc.*, vol. 75, no. 370, pp. 345–348, 1980.
- [6] M.B. Priestley, *Spectral Analysis and Time Series*, Academic Press, San Diego, Calif., 1981.
- [7] S.-T. Chiu, "Detecting periodic components in a white gaussian time series," *Journal of the Royal Statistical Society. Series B*, vol. 51, no. 2, pp. 249–259, 1989.
- [8] L.B. White, "Detection of sinusoids in unknown colored noise using ratios of AR spectrum estimates," *Proc. Information, Decision and Control*, pp. 257–262, 1999.
- [9] J. Gryca, "Detection of multiple sinusoids buried in noise via balanced model truncation," *Instrumentation and Measurement Tech. Conf. IEEE*, vol. 2, pp. 1353–1358, 1998.
- [10] C. Zheng, "Detection of multiple sinusoids in unknown colored noise using truncated cepstrum thresholding and local signal-to-noise-ratio," *Applied Acoustics*, pp. 809–816, 2012.
- [11] A. P. Hatzes, "The Radial Velocity Detection of Earth-mass Planets in the Presence of Activity Noise : The Case of α Centauri B b," *Astrophysical Journal*, vol. 770, pp. 133, 2013.
- [12] D.R. Brillinger, *Time Series : Data Analysis and Theory*, Holden Day, San Francisco, 1981.
- [13] M. Abramowitz and I. Stegun, *Spectral Analysis and Time Series*, Nat. Bureau of Standards - U.S. Gov. Printing Office., 1972.
- [14] S. Kay, *Fundamentals of Statistical signal processing. Vol II : Detection theory.*, Prentice-Hall, Inc, 1998.
- [15] A. Schwarzenberg-Czerny, "The distribution of empirical periodograms," *MNRAS*, vol. 301, pp. 831–840, 1998.