



**HAL**  
open science

# Modélisation des fréquences cardiaques instantanées durant un marathon et estimation de leurs paramètres fractals

Jean-Marc Bardet, Véronique Billat, Imen Kammoun

► **To cite this version:**

Jean-Marc Bardet, Véronique Billat, Imen Kammoun. Modélisation des fréquences cardiaques instantanées durant un marathon et estimation de leurs paramètres fractals. *Journal de la Société Française de Statistique*, 2009, 150 (1), pp.101-126. hal-00343361v2

**HAL Id: hal-00343361**

**<https://hal.science/hal-00343361v2>**

Submitted on 3 Aug 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Modélisation des fréquences cardiaques instantanées durant un marathon et estimation de leurs paramètres fractals

Jean-Marc Bardet

*Université Paris 1, SAMOS-MATISSE-CES, 90 rue de Tolbiac, 75013 Paris, France.*

Véronique Billat

*UBIAE Laboratoire INSERM 902 et Université d'Evry., Genopole, Boulevard F. Mitterrand, Evry Cedex, France.*

Imen Kammoun

*Université Paris 1, SAMOS-MATISSE-CES, 90 rue de Tolbiac, 75013 Paris, France.*

**Résumé.** Pour pouvoir caractériser et interpréter des signaux physiologiques comme ceux constitués par les fréquences cardiaques instantanées (ou les durées entre deux battements de coeur successifs), il est intéressant de trouver des paramètres qui restent constants au milieu des fluctuations. Depuis une quinzaine d'années, de nombreux papiers ont considéré le paramètre de Hurst, paramètre de longue mémoire, comme une telle constante caractéristique. Des raffinements ont été également proposés en prenant en compte les comportements en basses et hautes fréquences et en associant à chacun de ces régimes des paramètres de type Hurst. Cependant ces résultats posent plusieurs problèmes. En premier lieu, ils utilisent la méthode DFA pour estimer ce ou ces paramètres. Or, il est clair que cette méthode n'est pas la plus efficace et est très sensible à d'éventuelles tendances. En second lieu, et c'est encore plus préjudiciable par rapport aux résultats obtenus, le paramètre de type Hurst pour les basses fréquences, est souvent estimé à des valeurs supérieures à 1, borne supérieure admissible pour le paramètre d'un processus longue-mémoire stationnaire. Nous proposons des solutions à ces deux problèmes. Tout d'abord, nous utilisons un estimateur construit à partir d'une analyse par ondelettes, ce qui a plusieurs avantages: des vitesses de convergence optimales dans un cadre semi-paramétrique pour l'estimation du paramètre de Hurst, une robustesse à d'éventuelles tendances polynomiales et une estimation convergente des deux paramètres fractals lorsque l'on impose des conditions de localisation en fréquence pour l'ondelette mère. Ensuite, nous proposons comme modèle pour ces données un processus formé par les accroissements d'un mouvement brownien fractionnaire multi-échelle avec deux régimes non nuls (un régime basse fréquence et un autre haute fréquence). La méthode d'analyse par ondelettes permet d'estimer les deux paramètres fractals (à valeurs dans  $\mathbb{R}$  et non simplement dans  $(0, 1)$ ) associés à ces deux régimes, mais également de réaliser un test d'adéquation du modèle. Des simulations permettent de vérifier le bon comportement de ces statistiques. Appliqué aux données de fréquences cardiaques instantanées relevées chez des athlètes courant le marathon, un tel modèle est accepté lorsqu'on l'applique de façon différenciée aux trois périodes de course (début, milieu et fin de course) alors qu'il ne l'est pas si l'on considère l'ensemble de la course. On montre ainsi une augmentation au cours de la course du paramètre basse fréquence, ce qui va dans le même sens que les résultats obtenus par Peng *et al.* (1995): le coeur en début de course fonctionne comme celui de personnes en bonne santé, alors qu'en fin de course son comportement est proche de celui de personnes

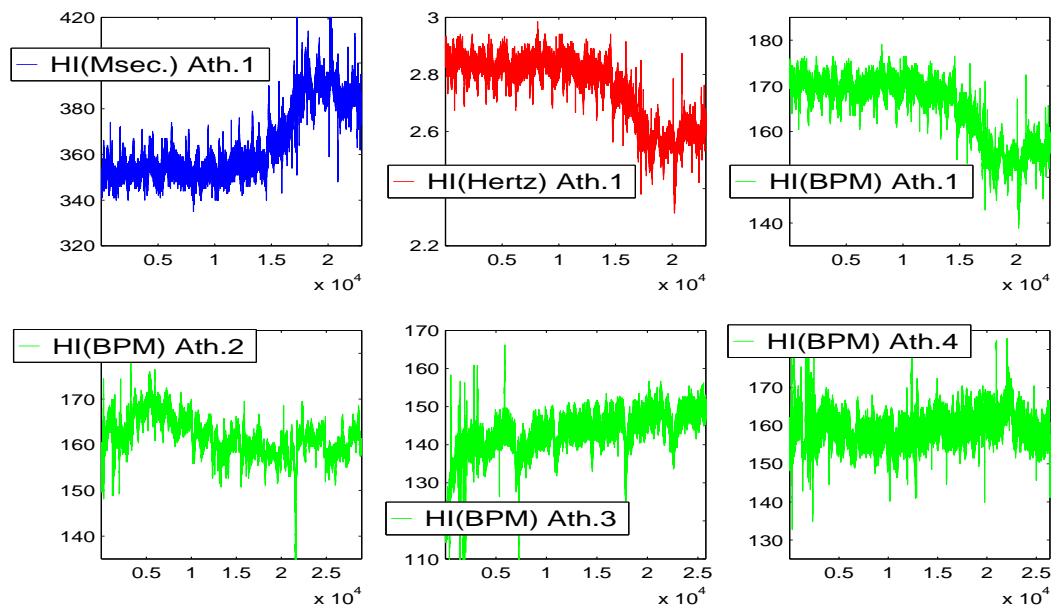
E-mail: bardet@univ-paris1.fr, veronique.billat@wanadoo.fr, imen.kammoun@gmail.com

ayant des dysfonctionnements cardiaques. Une telle évolution qui n'est pas corrélée avec la valeur même des fréquences cardiaques, est à associer avec la fatigue apparaissant en fin de course. La détection d'une trop grande valeur du paramètre fractal basses fréquences pourrait donc être essentiel pour garantir la santé de coureurs amateurs.

*Mots clés:* Série de fréquences cardiaques instantanées; Analyse par ondelettes; Detrended Fluctuation Analysis; Paramètre de Hurst; Processus longue mémoire; Bruit gaussien fractionnaire.

## 1. Introduction

Le but de cet article est de proposer un modèle et une méthode d'identification de ce modèle pour des données de durées entre battements de coeur successifs recueillis auprès d'athlètes courant un marathon (ici le Marathon de Paris 2004). La Figure 1 propose plusieurs exemples de telles données (le nombre total d'athlètes dont nous avons obtenu des relevés fiables est de 9 sur les 50 enregistrés, la moyenne de leurs fréquences cardiaques instantanées étant de 162 battements par minutes).



**Fig. 1.** En haut, battements cardiaques pour l'Athlète 1 en ms, Hertz et BPM (Battements Par Minute), et, en bas, pour les Athlètes 2, 3 et 4 en BPM.

Les données de fréquence cardiaques sont évidemment un enjeu important en médecine et une nombreuse littérature existe sur ce sujet. En particulier, les problèmes d'arythmies cardiaques sont décelées à partir de tels relevés (souvent sur 24h ou parfois plus). En général ce sont les grands changements de fréquences (ou de durées inter-battements) qui sont décelés et étudiés. Un paramètre de type "fractal" est désormais fréquemment utilisé pour caractériser l'erratisme des trajectoires. Le terme fractal recoupe en général deux acceptions: le

fait que le processus soit stationnaire et à longue-mémoire (ce qui se mesure souvent par le paramètre de Hurst) et le fait que le processus puisse avoir un comportement local de type höldérien (ce qui se mesure souvent par un exposant de Hölder).

Les premiers papiers de Gough (1992, 1993) ont observé que les fluctuations des battements de coeur de fœtus semblent suivre un comportement de type “fractal” sans être très explicite sur ce que cela signifie (si ce n’est le fait que l’on obtienne une loi de puissance pour la fonction DFA, voir ci-dessous). Peng *et al.* (1993, 1995) ont mis en évidence le comportement complexe de la dynamique des battements de coeur, qui semble pouvoir être caractérisée par un processus longue-mémoire pour des sujets en bonne santé. Dans tous ces papiers, les données ont été enregistrées à partir d’électro-cardiogrammes mesurant les durées successives entre deux battements pendant une longue période (24h). Par ailleurs, les résultats sont différents pour les sujets en bonne santé et des sujets atteints par des maladies cardiaques: le paramètre de longue-mémoire est une caractéristique pertinente du fonctionnement cardiaque. Les papiers Absil *et al.* (1999) et Krstacic *et al.* (2002) confirment ces résultats.

Plusieurs autres papiers ont également été écrits pour étendre les études précédentes à d’autres cas ou bien pour présenter des améliorations techniques. Ainsi, ont été aussi étudiées les distinctions entre les battements de coeur pendant le sommeil et l’éveil (voir Kantelhardt *et al.*, 2002, ou Staudacher *et al.*, 2005), entre hommes et femmes (voir Pikkujäämsä *et al.*, 2001), entre hommes sur terre et cosmonautes (voir Ivanov *et al.*, 1999), entre sujets en bonne santé et patients traités en unités intensives de réanimation (voir Yeh *et al.*, 2006), et beaucoup d’autres...

Dans tous les papiers précédemment cités, la Detrended Fluctuation Analysis (DFA dans la suite) introduite par Gough (1993) et Peng *et al.* (1994) est utilisée pour estimer le paramètre de longue mémoire (sauf dans Peng *et al.*, 1993, où la méthode d’estimation repose sur les moments du premier ordre). La DFA est une extension de la méthode des variances agrégées pour les séries chronologiques avec tendance (voir aussi Taqu *et al.*, 1995). Dans une partie de ces études, la DFA est appliquée seulement à de “grandes” tailles de fenêtre ce qui conduit à considérer l’estimation d’un paramètre de type Hurst (longue-mémoire). En s’intéressant aussi aux petites tailles de fenêtre (distinguant les hautes et les basses fréquences), deux exposants de type Hurst sont estimés (voir par exemple Peng *et al.*, 1995, ou Karasik *et al.*, 2002) mettant en évidence deux régimes physiologiques différents. De même des analyses spectrales de tels signaux ont mis en évidence un comportement double: une phase dite “orthosympathique” pour les fréquences inférieures comprises entre 0.04 et 0.15 Hz et une phase dite “parasymphique” pour les fréquences supérieures entre 0.15 et 0.5 Hz (voir à ce sujet les conclusions du groupe de travail de l’European Soc. Cardiology and the North American Society of Pacing and Electrophysiology, 1996).

Pour estimer le ou les paramètres fractals, la méthode d’analyse par ondelettes a aussi été appliquée aux données de battements de coeur dans différents travaux tels ceux d’Ivanov *et al.* (1996, 1999, 2001), Havlin *et al.* (1999) ou Goldberger *et al.* (2002), avec en vue deux améliorations par rapport à la méthode DFA: les tendances polynomiales n’influent sur l’estimation (dès que le nombre de moments de l’ondelette s’annulant est suffisant) et les processus multifractaux peuvent aussi être étudiés.

Enfin, les battements de coeur de sujet pratiquant des exercices physiques ont également été étudiés. A travers la littérature, on peut citer Karasik *et al.* (2002) (utilisant la méthode DFA et montrant d’importantes différences entre l’exercice et le repos suivant la bande d’échelles choisies), Martinis *et al.* (2004) (utilisant la méthode R/S et montrant que le paramètre de Hurst augmente avec l’intensité de l’exercice), Wesfreid *et al.* (2005) (util-

isant une analyse par ondelette pour distinguer le comportement multifractal entre des vitesses libres et des vitesses constantes pour des coureurs courant le 10000m).

L'étude statistique qui suit va tenter de répondre à divers questions et problèmes que posent les différents travaux précédemment cités:

- Que mesure-t-on? Une régularité locale, un comportement de type longue mémoire ou autre chose?
- La méthode utilisée pour estimer le (ou les paramètres) est-elle efficace et robuste?
- Comment expliquer que le paramètre estimé soit souvent proche de 1 voire supérieur à 1, ce qui a priori n'est pas acceptable pour un exposant de Hölder ou un paramètre de Hurst d'un processus longue mémoire (dont les moments d'ordre 2 existent clairement pour les séries de fréquences cardiaques)?
- Quel modèle proposer pour modéliser ces données de battements cardiaques et peut-on tester l'ajustement à un tel modèle?
- Y-a-t-il une particularité aux données relevées durant un effort d'endurance comme un marathon?

Nous proposons trois étapes majeures permettant, nous semble-t-il, de répondre à ces questions et problèmes:

- (a) En premier lieu (voir Section 2), on découpe les données suivant une méthode adaptative de détection de ruptures en moyenne et variance (on reprend ici un algorithme proposé dans Lavielle, 1999). On verra ainsi qu'un tel découpage se résume pour la plupart des coureurs à trois moments de leur course: le début, le milieu et la fin de course, ce qui semble assez naturel. Et nous verrons que les comportements sur chacune de ces phases sont différents.
- (b) On l'a dit, la méthode DFA (Detrended Fluctuation Analysis) est souvent utilisée pour estimer le ou les paramètres fractals dans la plupart des travaux cités précédemment. Or cette méthode n'est pas la plus efficace et n'est pas robuste en cas de présence de tendance pour estimer le paramètre de longue mémoire (voir Bardet et Kammoun, 2008). Les méthodes semi-paramétriques comme celles du minimum de contraste introduite par Robinson (1995) ou de log-régression du périodogramme (voir Moulines et Soulier, 2003) sont théoriquement et empiriquement plus intéressantes. Il en est de même avec la méthode d'analyse par ondelettes (voir par exemple Abry *et al.*, 2003, ou Moulines *et al.*, 2007), que nous allons utiliser dans la Section 3. Enfin, l'utilisation d'une ondelette mère dont le support de la transformée de Fourier est compact permet de localiser le comportement du processus en fréquence. On observe ainsi un log-variogramme des coefficients d'ondelettes en fonction des échelles comportant deux parties clairement linéaires. Ceci confirme l'hypothèse de deux régimes physiologiques qui fonctionnent clairement de manière différente pendant un exercice d'endurance comme le marathon.
- (c) Pour modéliser un tel log-variogramme pour des données biomécaniques, Bardet et Bertrand (2007a,b) avaient proposé d'utiliser un processus gaussien à accroissements stationnaires appelé mouvement brownien fractionnaire multi-échelle, c'est-à-dire un processus dont la densité spectrale est une loi de puissance dont les paramètres

dépendent de la bande de fréquence sur laquelle on se trouve. Pour modéliser les données de battements de coeurs par un processus stationnaire, nous utiliserons le processus constitué par les accroissements d'un mouvement brownien fractionnaire multi-échelle à deux régimes non nuls sur chacune des bandes  $[\omega_0, \omega_1]$  (appelé intervalle basses fréquences) et  $[\omega_1, \omega_2]$  (appelé intervalle hautes fréquences) avec  $0 < \omega_0 < \omega_1 < \omega_2 < \infty$ , la représentation spectrale du processus est donc la même que celle d'un bruit gaussien fractionnaire de paramètres respectifs  $H_0$  (basses fréquences) et  $H_1$  (hautes fréquences), à la différence près que ces deux paramètres pouvant prendre toutes valeurs dans  $\mathbb{R}$  et non seulement des valeurs dans  $(0, 1)$ . La méthode d'analyse par ondelettes développée dans Bardet et Bertrand (2007b) permet d'estimer ces paramètres, après avoir utilisé une méthode adaptative inspirée de Bardet *et al.* (2008) pour sélectionner les deux zones de fréquences  $[\omega_0, \omega_1]$  et  $[\omega_1, \omega_2]$ . Enfin, un test d'adéquation de type Chi-deux découlera de ces estimations.

Lorsque l'on applique cette méthodologie aux données cardiaques, on observe un phénomène très intéressant: le test d'adéquation est accepté dans chacune des trois phases des trajectoires (début de course, milieu de course et fin de course) alors qu'il ne l'est pas si on l'applique sur l'intégralité de chaque trajectoire. On montre également qu'il y a une augmentation significative du paramètre basses fréquences au fil de la course (le paramètre hautes fréquences restant lui constant). Ceci ne pouvait pas être observé en utilisant la méthode DFA ou une méthode de type variations quadratiques (ces deux méthodes pouvant être vues comme des méthodes d'ondelettes avec une ondelette mère définie comme une combinaison linéaire finie de fonctions indicatrices, voir à ce sujet Vedel *et al.*, 2008): la propriété de localisation en fréquence de l'ondelette mère est essentielle pour permettre une estimation convergente, ce que montrent également des simulations.

En devenant ainsi d'une certaine manière de plus en plus "longue mémoire", la trajectoire en fin de course ressemble à celles de certains malades du coeur, alors qu'elle ressemblait à celles d'individus en bonne santé en début de course (les comparaisons sont obtenues à partir des résultats de Peng *et al.*, 1993). Ainsi, une trop grande valeur du paramètre fractal basses fréquences peut être représentatif de l'apparition d'une fatigue du coeur et témoigner d'un danger alors même que les valeurs des fréquences cardiaques instantanées ne paraissent pas alarmantes. Les fluctuations pourraient donc s'avérer plus informatives que les valeurs, et la mise en place de détection séquentielle d'une trop grande régularité (trop de "longue mémoire") cardiaque pourrait permettre de prévenir des risques cardiaques (comme cela arrive malheureusement trop fréquemment lors d'épreuves telles que le marathon).

## 2. Détection des différentes phases de courses

Nous avons donc commencé par appliquer une méthode de détection adaptative des changements en moyenne et variance de la trajectoire de chacun des coureurs. Cette méthode a été développée et programmée par M. Lavielle (voir Lavielle, 1999, Lavielle et Moulines, 2000, Lavielle et Ludeña, 2001) et de nouveaux développements non asymptotiques ont été apportés dans Birgé et Massart (2007). Le principe de la méthode est le suivant:

On suppose que l'on observe  $(Y(i), i = 1, \dots, n)$  et qu'il existe  $\tau = (\tau_1, \tau_2, \dots, \tau_{K-1})$  avec  $0 = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < n = \tau_K$  tels que pour tout  $j \in \{1, 2, \dots, K\}$ , la loi de  $Y(i)$  dépend d'un paramètre  $\theta_j \in \Theta \subset \mathbb{R}^d$  (avec  $d \in \mathbb{N}$ ) pour tout  $\tau_{j-1} < i \leq \tau_j$ . Donc,  $K$  est le nombre de segments entre les différents instants ordonnés  $\tau = (\tau_1, \tau_2, \dots, \tau_{K-1})$ . On

supposera pour commencer que  $K$  est connu. Soit maintenant une fonction de contraste

$$U_\theta(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1})),$$

avec  $\theta \in \mathbb{R}^d$  et que l'on applique à chaque vecteur  $(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1}))$  pour tout  $j \in \{0, 2, \dots, K-1\}$ . L'exemple classique d'une telle fonction est celui-ci:

$$U_\theta(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1})) = -2 \log L_\theta(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1})),$$

où  $L_\theta$  est la vraisemblance (on suppose le modèle statistique dominé). Ainsi, pour tout  $j \in \{0, 2, \dots, K-1\}$ , on peut définir:

$$\hat{\theta}_j = \underset{\theta \in \Theta}{\operatorname{Argmin}} U_\theta(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1})).$$

Soit

$$\hat{G}(\tau_1, \dots, \tau_{K-1}) = \sum_{j=0}^{K-1} U_{\hat{\theta}_j}(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1})).$$

En conséquence, un estimateur  $(\hat{\tau}_1, \dots, \hat{\tau}_{K-1})$  peut être défini comme:

$$(\hat{\tau}_1, \dots, \hat{\tau}_{K-1}) = \underset{0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < n}{\operatorname{Argmin}} \hat{G}(\tau_1, \dots, \tau_{K-1}). \quad (1)$$

Le principe d'une telle estimation est très général (il peut par exemple être appliqué dans le cas de séries temporelles ayant des tendances polynomiales) et on pourra trouver différents articles donnant le comportement asymptotique de l'estimateur  $(\hat{\tau}_1, \dots, \hat{\tau}_{K-1})$  suivant les hypothèses de dépendance faites sur  $Y$  (voir par exemple Bai, 1998, Bai et Perron, 1998, Lavielle, 1999, Lavielle et Moulines, 2000, ou Lavielle et Ludeña, 2001).

Dans le cas particulier des données de battements de coeur, nous avons choisi de déterminer les changements éventuels en moyenne et variance. D'autres choix auraient été possibles, en particulier celui des changements de la régularité locale (voir par exemple dans Bardet et Kammoun, 2008) ou des changements de la densité spectrale (voir Bertrand *et al.*, 2009). Choisir les changements en moyenne et variance est intéressant dans un premier temps, car cela ne présage pas de l'existence d'un ou deux paramètres de type fractal. Ainsi, pour tout  $j \in \{0, 1, \dots, K-1\}$ , on considère le modèle général:

$$Y(i) = \mu_j + \sigma_j \varepsilon_i \text{ for all } i \in \{\tau_j + 1, \dots, \tau_{j+1}\},$$

où  $\theta_j = (m_j, \sigma_j) \in \mathbb{R} \times (0, \infty)$  et  $(\varepsilon_i)$  est une suite de variables aléatoires centrées et de variance 1, non nécessairement indépendantes.

Dans un tel cadre, une fonction de contraste naturelle est:

$$U_{\theta_j}(Y(\tau_j + 1), \dots, Y(\tau_{j+1})) = \sum_{\ell=\tau_j+1}^{\tau_{j+1}} \frac{(Y(\ell) - m_j)^2}{\sigma_j^2},$$

et ainsi l'estimateur bien connu de  $\theta_j$  est:

$$\hat{\theta}_j = (\hat{m}_j, \hat{\sigma}_j) = \left( \frac{1}{\tau_{j+1} - \tau_j} \sum_{\ell=\tau_j+1}^{\tau_{j+1}} Y(\ell), \frac{1}{\tau_{j+1} - \tau_j} \sum_{\ell=\tau_j+1}^{\tau_{j+1}} (Y(\ell) - \hat{m}_j)^2 \right).$$

Donc l'estimateur  $(\hat{\tau}_1, \dots, \hat{\tau}_{K-1})$  peut être déduit de (1).

Quand le nombre de ruptures est inconnu, on utilise un nouveau contraste  $V$  construit comme la somme du précédent contraste  $U$  et d'une fonction de pénalisation ( $pen$ ) croissante en le nombre de ruptures  $K$ , soit, plus précisément,

$$\hat{V}(\tau_1, \dots, \tau_{K-1}, K) = \hat{G}(\tau_1, \dots, \tau_{K-1}) + \beta \times pen(K),$$

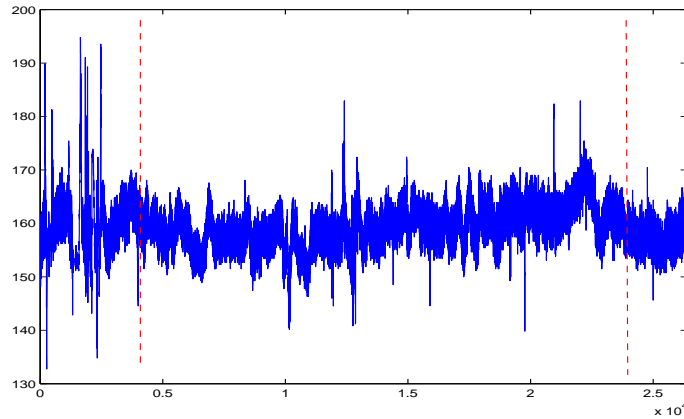
avec  $\beta > 0$ . En conséquence, en minimisant  $V$  par rapport à  $\tau_1, \dots, \tau_{K-1}, K$ , un estimateur  $\hat{K}$  est aussi obtenu, estimateur dont la valeur peut varier en fonction du paramètre de pénalisation  $\beta$ .

Pour les données de fréquences cardiaques, nous avons choisi  $pen(K) = K$ . Par ailleurs, si  $\hat{G}_K = \hat{G}(\hat{\tau}_1, \dots, \hat{\tau}_{K-1})$ , pour  $K = K_1, \dots, K_{MAX}$  on peut définir

$$\beta_i = \frac{\hat{G}_{K_i} - \hat{G}_{K_{i+1}}}{K_{i+1} - K_i} \text{ et } l_i = \beta_i - \beta_{i+1} \text{ avec } i \geq 1.$$

Ainsi, la valeur  $\hat{K}$  retenue est la plus grande valeur des  $K_i$  telle que  $l_i \gg l_j$  pour  $j > i$ .

Lorsque l'on applique cette méthode de détection adaptative des ruptures, on trouve pour presque tous les coureurs  $\hat{K} = 3$ , soit 3 phases de course de tailles différentes: le début (environ 20min en général), la fin (environ 20min également en général) et le milieu de la course (le reste du temps). La Figure 2) propose un exemple d'un tel découpage. Pour 2 coureurs, nous n'avons pu distinguer le début ou la fin de course de la partie centrale et  $\hat{K} = 2$ .



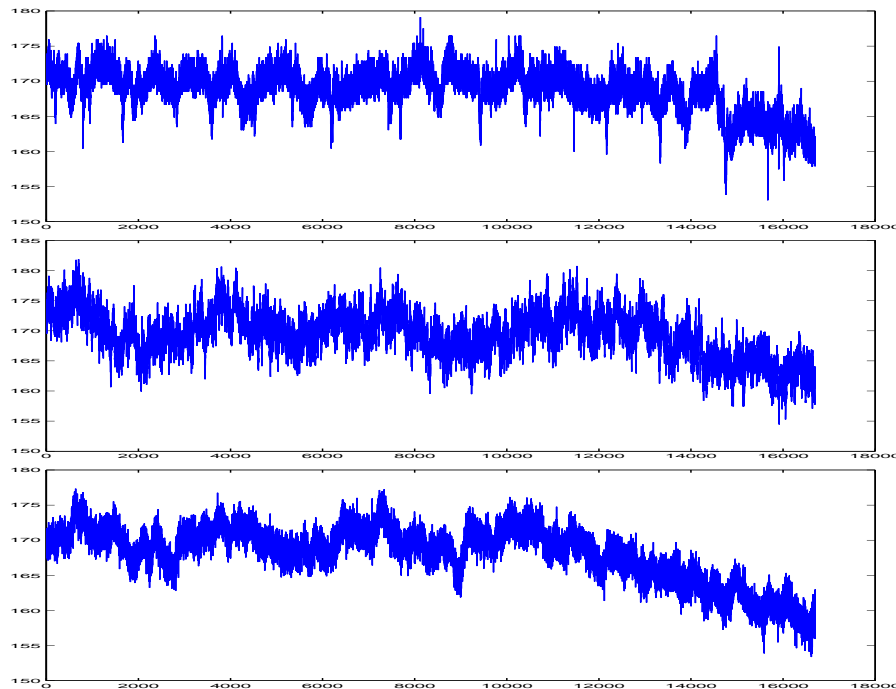
**Fig. 2.** Détermination des différentes phases de course pour un athlète

On considérera désormais les données de fréquences cardiaques sur chacune des phases obtenues. On fera l'approximation que ces données sont stationnaires même si un test non-paramétrique tel celui de Priestley-Subba Rao montre que ceci n'est pas vrai pour presque toutes les séries de données (seul le milieu de course de l'athlète 8 passe le test). Nous allons nous intéresser à un paramètre de type fractal qui pourrait être caractéristique du signal sur chacune de ses phases.



### 3. Un premier modèle et l'estimation par analyse par ondelettes

Quand on étudie les données de fréquences cardiaques sur chacune des phases (mais aussi sur l'intégralité des données), on remarque que le corrélogramme est typique de celui d'un processus à forte mémoire, c'est-à-dire qu'il présente une décroissance très lente des valeurs absolues de l'autocorrélation empirique. Par ailleurs, l'hypothèse de normalité de la distribution des données semble plausible (ne sachant pas vraiment la dépendance entre les données on ne peut valider une telle intuition par un test). On aurait donc envie, dans une première étape, de modéliser les données dans chaque phase de course par un bruit gaussien fractionnaire (qui est un processus gaussien à longue mémoire). Cette intuition peut également sembler acceptable lorsque l'on compare le graphe d'une trajectoire durant une phase de course avec celui d'un bruit gaussien fractionnaire simulé avec un paramètre proche de 1 (voir la Figure 3).



**Fig. 3.** De haut en bas, comparaison entre le graphe des données de battements de cœur durant le milieu de course (Athlète 4) avec celui d'un bruit gaussien fractionnaire simulé ( $H = 0.99$ ) et celui des accroissements d'un mouvement brownien fractionnaire multi-échelle à deux régimes ( $H_0 = 1.25$  en basses fréquences et  $H_1 = -0.15$  en basses fréquences). Une tendance linéaire par morceaux a été rajoutée sur les données simulées pour améliorer la ressemblance visuelle.

Rappelons juste quelques éléments théoriques concernant le bruit gaussien fractionnaire (BGF). En premier lieu, on dit qu'un processus stationnaire  $Y = \{Y(k), k \in \mathbb{N}\}$  possédant des moments d'ordre 2 est à longue mémoire ou fortement dépendant si:

$$\sum_{k \in \mathbb{N}} |r_Y(k)| = \infty \quad \text{avec} \quad r_Y(k) = \mathbb{E}[Y(0)Y(k)].$$

Un cas particulier de processus longue mémoire est celui dont la covariance se comporte comme

$$r_Y(k) \sim k^{2H-2}L(k), \text{ quand } k \rightarrow \infty,$$

avec  $L(k)$  une fonction à variations lentes (*i.e.*  $\forall t > 0, L(xt)/L(x) \rightarrow 1$  quand  $x \rightarrow \infty$ ) et  $H \in (\frac{1}{2}, 1)$  est appelé le paramètre de Hurst.

On peut relier la notion de processus longue mémoire à celle de processus autosimilaire de la façon suivante. Tout d'abord, on dit que  $X = \{X(t), t \geq 0\}$  est un processus autosimilaire à accroissements stationnaires lorsque  $\forall c > 0$ :

$$(X(ct))_t \stackrel{L}{\underset{d}{\sim}} c^H(X(t))_t,$$

où  $H$  est appelé le paramètre d'autosimilarité. Si on considère le processus agrégé  $\{X(t), t \geq 0\}$  tel que  $X(k) = \sum_{i=1}^k Y(i)$  où  $Y$  est un processus autosimilaire, alors sous des conditions assez faibles (par exemple  $Y$  est gaussien ou est un processus linéaire causal), on peut montrer que pour  $k \rightarrow \infty$ , la loi de  $\{X(t), t \geq k\}$  est celle d'un processus autosimilaire (voir Doukhan *et al.*, 2003, pour plus de détails).

Le BGF est un exemple de processus longue mémoire gaussien. Plus précisément,  $Y^H = \{Y^H(k), k \in \mathbb{N}\}$  est un BGF, lorsque

$$r_{Y^H}(k) = \frac{\sigma^2}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}) \text{ pour tout } k \in \mathbb{N},$$

avec  $H \in (0, 1)$  et  $\sigma^2 > 0$ . On vérifie par une formule de Taylor que pour  $H \in (\frac{1}{2}, 1)$ ,

$$r_{Y^H}(k) \sim \sigma^2 H(2H-1)k^{2H-2}, \text{ quand } k \rightarrow \infty.$$

Pour un BGF centré, le processus agrégé correspondant, noté ici  $X^H$ , est appelé le mouvement brownien fractionnaire (MBF) et  $X^H$  est l'unique processus gaussien  $H$ -autosimilaire et à accroissements stationnaires, vérifiant donc

$$\text{Var}(X^H(k)) = \sigma^2 |k|^{2H} \quad \forall k \in \mathbb{N}.$$

Il est clair que  $Y^H(k) = X^H(k) - X^H(k-1)$ : la suite des accroissements d'un MBF constitue un BGF.

Avant de pouvoir tester si les données de battements de coeur peuvent être modélisées par un BGF, commençons par dire quelques mots sur les estimateurs possibles de  $H$ . En premier lieu, on trouvera un large éventail de tels estimateurs dans le livre de Doukhan *et al.* (2003). Il est bien clair que pour estimer le paramètre  $H$  d'un BGF à partir d'une de ses trajectoires, il est préférable d'utiliser une méthode paramétrique, en particulier la méthode du contraste de Whittle qui est numériquement bien plus rapide que celle du maximum de vraisemblance (on trouvera les principaux résultats asymptotiques dans Fox et Taqqu, 1987). Cependant, ici, nous aimerions travailler dans un cadre semi-paramétrique (nous verrons ci-dessous pourquoi) et en plus avec l'ajout éventuel d'une tendance au processus. La méthode d'estimation utilisant une analyse par ondelettes offre tous les avantages vis-à-vis d'un tel cahier des charges. De plus un test d'adéquation sera facilement implémentable à partir de cette estimation. Présentons maintenant succinctement cette méthode, qui a été introduite par Flandrin (1992), puis développée par Bardet *et al.* (2000) ou Abry *et al.* (2003).

On considère donc  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction mesurable appartenant à  $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ , appelée ondelette mère. Pour  $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}$  et  $\lambda = (a, b)$ , on définit la famille de fonctions  $\psi_\lambda$  par

$$\psi_\lambda(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t}{a} - b\right).$$

Les paramètres  $a$  et  $b$  sont appelés l'échelle et le décalage de la transformée en ondelette (ici continue). Soit  $d_Z(a, b)$  le coefficient d'ondelette du processus  $Z = \{Z(t), t \in \mathbb{R}\}$  pour l'échelle  $a$  et le décalage  $b$ :

$$d_Z(a, b) = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} \psi\left(\frac{t}{a} - b\right) Z(t) dt = \langle \psi_\lambda, Z \rangle_{L^2(\mathbb{R})}.$$

Lorsque l'on considère une série chronologique plutôt qu'un processus en temps continu, on peut choisir d'approcher l'intégrale par une somme de Riemann (d'autres méthodes sont possibles, mais l'approximation par une somme de Riemann permet de retrouver l'approximation obtenue par l'algorithme de Mallat, équivalent de la FFT pour les ondelettes) et on obtient ainsi le coefficient d'ondelette approché  $e_Z(a, b)$ . L'erreur commise lors de l'approximation de  $d_Z(a, b)$  par  $e_Z(a, b)$  a été évaluée dans Bardet(2002) et Bardet et Bertrand (2007b). Si elle est préjudiciable à la vitesse de convergence, elle n'empêche cependant pas la convergence présentée ci-dessous. On suppose par ailleurs que la fonction  $\psi$  est telle qu'il existe  $M \in \mathbb{N}^*$  vérifiant,

$$\int_{\mathbb{R}} t^m \psi(t) dt = 0 \text{ for all } m \in \{0, 1, \dots, M\}; \quad (2)$$

On dit alors que,  $\psi$  a ses  $M$  premiers moments qui s'annulent.

Effectuer une analyse en ondelette d'un processus revient à regarder le comportement des coefficients d'ondelettes. Ceci peut être fait pour des processus longue mémoire ou autosimilaire. On obtient en particulier que:

- Si  $Z$  est un processus longue mémoire stationnaire de paramètre de Hurst  $H$ ,

$$\mathbb{E}(d_Z^2(a, b)) = \text{Var}(d_Z(a, b)) \sim C(\psi, H) a^{2H-1} \text{ quand } a \rightarrow \infty.$$

- Si  $Z$  est un processus  $H$ -autosimilaire à accroissements stationnaires,

$$\mathbb{E}(d_Z^2(a, b)) = \text{Var}(d_Z(a, b)) \sim K(\psi, H) a^{2H+1} \text{ pour tout } a > 0$$

avec  $C(\psi, H)$  et  $K(\psi, H)$  deux constantes positives dépendant seulement de  $\psi$  et de  $H$  (ces résultats ont été prouvés dans Flandrin, 1992 et Abry *et al.*, 1998). Ainsi dans les deux cas, la variance des coefficients d'ondelettes est une loi de puissance en  $a$ , et une log-log-régression permet d'obtenir un estimateur de  $H$ . Lorsque l'on considère une trajectoire  $(Z(1), \dots, Z(N))$  du processus, cette log-log-régression se fera à partir d'un estimateur "naturel" de la variance, la variance empirique des coefficients d'ondelettes, par exemple *i.e.*

$$S_N(a) = \frac{1}{[N/a]} \sum_{i=1}^{[N/a]} e_Z^2(a, i). \quad (3)$$

On peut alors également tracer le graphe des  $(\log a_i, \log S_N(a_i))_{1 \leq i \leq \ell}$  en choisissant une famille d'échelles  $(a_i)_i$  appropriée et le coefficient directeur de la droite de régression par moindres carrés fournit l'estimateur  $\widehat{H}_{OND}$  de  $H$ . En particulier, pour un processus stationnaire à longue mémoire on a :

$$\widehat{H}_{OND} = \frac{\widehat{\theta}_N^{(1)} + 1}{2}$$

$$\text{où } \begin{pmatrix} \widehat{\theta}_N^{(1)} \\ \widehat{\theta}_N^{(2)} \end{pmatrix} = (Z'Z)^{-1}Z' \begin{pmatrix} \log S_N(a_1) \\ \vdots \\ \log S_N(a_m) \end{pmatrix} \text{ et } Z = \begin{pmatrix} \log a_1 & 1 \\ \vdots & \vdots \\ \log a_m & m \end{pmatrix}. \quad (4)$$

Dans un cadre semi-paramétrique général de processus longue mémoire stationnaires gaussiens ou linéaires, Moulines *et al.* (2007) et Roueff et Taqu (2008) ont montré que  $\widehat{H}_{OND}$  converge vers  $H$  suivant une vitesse de convergence optimale (par rapport à un critère minimax) dès que les échelles sont choisies de manière adéquate. Veitch *et al.* (2003) et Bardet *et al.* (2008) ont proposé une sélection adaptative de ces échelles et montré que la vitesse de convergence était conservée (à un logarithme près).

Dans les cadres respectifs des FBM et des processus de Rosenblatt, qui sont autosimilaires et à accroissements stationnaires, Bardet (2002) et Bardet et Tudor (2008) ont déterminé la vitesse de convergence de  $\widehat{H}_{OND}$  vers  $H$ . Ces travaux s'étendent facilement à un cadre semi-paramétrique (voir par exemple la généralisation aux processus multi-échelles dans Bardet et Bertrand, 2007).

Enfin, que ce soit pour les processus longue mémoire ou autosimilaires, les résultats de convergence de  $\widehat{H}_{OND}$  vers  $H$  sont conservés même si une tendance polynomiale est rajoutée au processus (cela a été remarqué pour la première fois dans Abry *et al.*, 1998). Pour cela, il suffit que les  $d$  premiers moments de  $\psi$  s'annulent, où  $d$  est le degré de la tendance polynomiale. Aussi, si on considère une ondelette annulant tous ses moments, comme c'est le cas notamment si l'on choisit une ondelette de Lemarié-Meyer, alors toute tendance polynomiale est annulée.

Pour terminer, nous dirons que l'analyse par ondelette de tels processus a un avantage supplémentaire: on peut facilement construire un test d'adéquation de type Chi-deux à partir de la log-log-régression précédente. Pour cela, il suffit d'utiliser une log-log-régression par moindres carrés pseudo-généralisés (la matrice de covariance de la log-log-régression peut être estimée), puis de définir une distance entre les points et cette droite de régression (pour plus de détails, voir dans le cadre du FBM, Bardet, 2002, et dans le cadre du mouvement brownien fractionnaire multi-échelle, Bardet et Bertrand, 2007b; notons que dans un tel cadre les échelles  $a_i$  doivent avoir une vitesse supérieures à  $N^{1/3}$  pour que l'approximation des  $d_Z(a, b)$  par les  $e_Z(a, b)$  ne crée pas un biais sur l'estimation de  $H$ ).

Evoquons maintenant la méthode DFA (pour Detrended Fluctuation Analysis) que les physiologistes utilisent beaucoup. Cette méthode est une version pour des séries à tendances de la méthode des variances agrégées. Plus précisément, elle consiste à agréger le processus (qui est supposé être de type longue mémoire + tendance) et le diviser dans des fenêtres de même taille, puis de détendancionaliser ce processus à partir de régressions linéaires dans chaque fenêtre, de calculer la variance des résidus, et d'estimer  $H$  à partir d'une log-log-

**Table 1.** Valeurs de  $H$  estimées, avec la méthode DFA ( $\hat{H}_{DFA}$ ) et la méthode d'analyse par ondelettes ( $\hat{H}_{OND}^{(1)}$ ) avec  $\psi$  ondelette mère de Daubechies d'ordre 6.

	Course entière		Début de course		Milieu de course		Fin de course	
	$\hat{H}_{DFA}$	$\hat{H}_{OND}^{(1)}$	$\hat{H}_{DFA}$	$\hat{H}_{OND}^{(1)}$	$\hat{H}_{DFA}$	$\hat{H}_{OND}^{(1)}$	$\hat{H}_{DFA}$	$\hat{H}_{OND}^{(1)}$
Ath1	0.96	1.03	1.05	1.03	1.13	1.35	0.71	1.22
Ath2	1.12	1.27	0.91	0.90	1.19	1.42	1.53	1.39
Ath3	1.17	1.10	0.74	0.78	1.19	1.29	-	-
Ath4	1.16	1.26	-	-	1.15	1.23	1.22	1.12
Ath5	1.18	1.24	1.24	1.15	1.19	1.22	1.43	1.49
Ath6	1.28	1.37	1.34	1.03	1.18	1.39	1.41	1.50
Ath7	1.18	1.32	1.09	1.18	1.23	1.29	1.14	1.22
Ath8	1.28	1.23	1.34	1.00	1.05	1.28	-	-
Ath9	1.36	1.37	0.89	0.93	1.27	1.48	1.33	1.37

régression des différentes variances résiduelles obtenues pour différentes tailles de fenêtre. Cependant, cette méthode pose problème. Si, pour un processus gaussien stationnaire longue mémoire, la vitesse de convergence de  $\hat{H}_{DFA}$  (estimateur de  $H$  par DFA) reste intéressante, elle l'est moins (à des constantes près) que la simple méthode des variances agrégées, ou que les méthodes du log-périodogramme, du contraste de Whittle local ou d'analyse par ondelettes (voir Bardet et Kammoun, 2008 pour plus de détails). Le point très négatif de cette méthode est celui de la détendancialisaiton par régression linéaire dans chaque fenêtre: ceci ne peut avoir d'effet positif que si la tendance est une constante. Dès que la tendance est un polynôme de degré supérieur, ou bien une simple fonction constante par morceaux, les effets de cette étape sont très nocifs et l'estimateur  $\hat{H}_{DFA}$  converge vers tout à fait autre chose que  $H$ ... (voir encore Bardet et Kammoun, 2008).

Voici maintenant les résultats dans la Table 1 de l'application des deux méthodes d'estimation (DFA et ondelette) pour le paramètre de longue mémoire sur les données de battements de coeur. Dans un premier temps, on a l'habitude d'utiliser une ondelette mère  $\psi$  relative à une analyse multirésolution. Ici,

$\hat{H}_{OND}^{(1)}$  : estimateur  $\hat{H}_{OND}$  de la formule (4) avec  $\psi$  une ondelette de Daubechies d'ordre 6.

Notons également que l'on a choisi pour la méthode DFA des tailles de fenêtres comprises entre  $N^{0.4}$  et  $N/10$  (où  $N$  est la longueur de la série), et pour la méthode par ondelette des échelles  $a_i = i N^{0.4}$  où  $i = 1, \dots, 20$ .

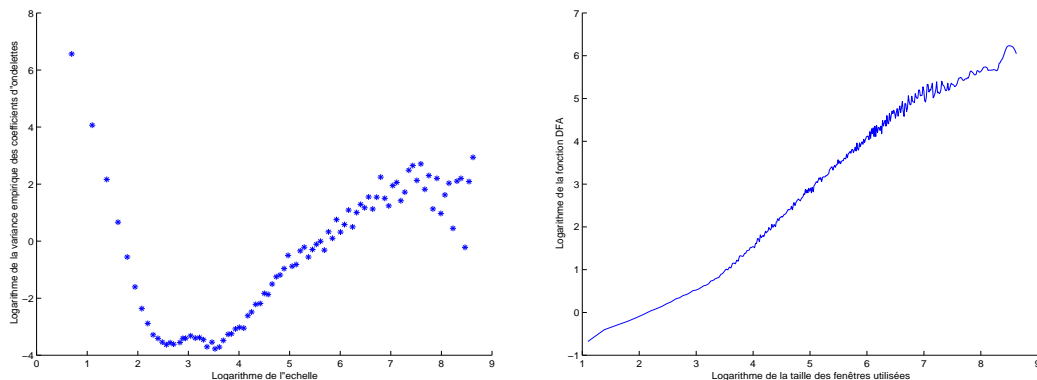
Deux problèmes importants découlent de ces résultats.

- (a) En premier lieu, les estimateurs  $\hat{H}_{DFA}$  et  $\hat{H}_{OND}^{(1)}$  sont souvent plus grands que 1. Pourtant un BGF et même plus généralement un processus longue-mémoire de variance finie ne peut seulement être défini que si  $H < 1$ . Pour obtenir un processus dont l'estimation d'un paramètre fractal serait supérieur à 1, au moins une des trois hypothèses vérifiées par le BGF doit être remise en cause:
- l'hypothèse que le processus est stationnaire;
  - l'hypothèse que le processus est gaussien;

- l’hypothèse que le paramètre de Hurst du processus soit bien le  $H$  estimé.

Dans la suite, un nouveau modèle sera proposé et nous ferons le choix de conserver les deux premières hypothèses qui ne nous semblent pas devoir être remises en question mais de remplacer la troisième par un modèle paramétrique plus complexe.

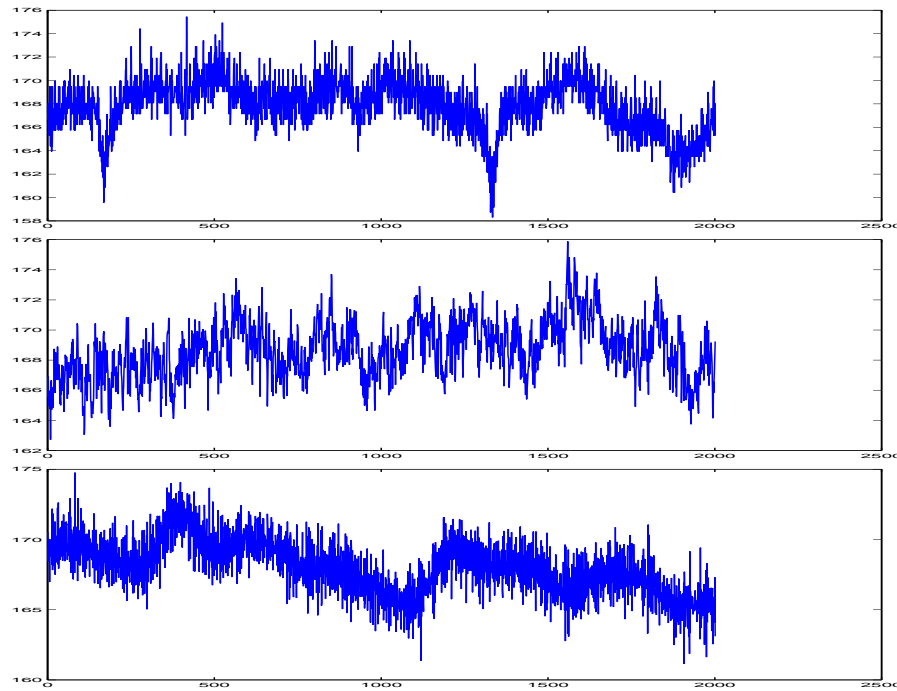
- (b) Le second problème concerne le test d’adéquation évoqué plus haut. En effet le test n’est jamais accepté, que l’on considère la série en son intégralité (souvent de l’ordre de 25000 données) ou en ses différentes phases. On peut donc dire que les données ne se comportent pas comme un BGF. Une explication de ce rejet peut se déduire de la Figure 4: quand on considère le résultat de la log-log-régression, les points  $(\log a_i, \log S_N(a_i))_{1 \leq i \leq \ell}$  sont clairement alignés pour de “grandes” échelles mais ce n’est pas le cas pour toutes les échelles. Sur la même figure, on remarque également le même type de phénomène avec l’estimateur DFA. Ainsi les données de battements de coeur se comportent à peu près comme un BGF pour des “grandes échelles” (ou des petites fréquences), mais le comportement est différent pour des échelles plus petites. On remarquera aussi que l’adéquation d’un BGF aux données n’est finalement pas très bonne lorsque l’on opère un “zoom” sur les graphes de la Figure 3. On obtient ainsi les graphes de la Figure 5, où il apparaît que localement le BGF de paramètre 0.99 ne “remplit” pas suffisamment l’espace (la dimension de Hausdorff est trop proche de 1) contrairement à la série des données de fréquence cardiaques instantanées (la dimension de Hausdorff est beaucoup plus proche de 2).



**Fig. 4.** Pour le milieu de course de l’Athlète 4, graphes des log-log regressions des: 1/ variances des coefficients d’ondelettes en fonction de l’échelle avec l’ondelette de Daubechies d’ordre 6 (à gauche); 2/ fonctions DFA en fonction des tailles de fenêtres utilisées (à droite).

#### 4. Un nouveau modèle pour modéliser les fréquences cardiaques: le bruit gaussien localement fractionnaire

Les Figures 4 et 3 nous permettent finalement d’arriver à la même conclusion: pour “bien” modéliser les données de fréquences cardiaques instantanées, il faudrait un processus qui se comporte comme un BGF de paramètre proche de 1 et même plus grand que 1 pour les grandes échelles (ou petites fréquences) et comme un BGF de paramètre proche de 0



**Fig. 5.** De haut en bas, comparaison entre le graphe des données de 2000 battements de coeur durant le milieu de course (Athlète 4) avec celui d'un bruit gaussien fractionnaire simulé ( $H = 0.99$ ) et celui des accroissements d'un mouvement brownien fractionnaire multi-échelle à deux régimes ( $H_0 = 1.25$  en basses fréquences et  $H_1 = -0.15$  en basses fréquences).

(de manière à obtenir une dimension de Hausdorff du graphe proche de 2) pour les petites échelles (ou les hautes fréquences). On remarque également que sur la Figure 4, deux zones de linéarité apparaissent, une pour les grandes échelles (basses fréquences) et une autre pour les petites échelles (hautes fréquences).

Le modèle que nous avons choisi pour modéliser les fréquences cardiaques s'inspire des résultats déjà obtenus dans Bardet et Bertrand (2007b), article dans lequel déjà apparaissait la nécessité de distinguer un comportement différent entre les hautes et les basses fréquences. On voit donc qu'en restreignant la bande de fréquence on obtient bien une zone de linéarité (soit un comportement de même type que le BGF) alors qu'en dehors de cette zone d'autres phénomènes se produisent. Pour arriver à construire un modèle qui puisse se comporter de cette manière, on rappelle la représentation spectrale d'un MBF  $X^H = \{X^H(t), t \in \mathbb{R}\}$  de paramètres  $H$  et  $\sigma^2$ :

$$X^H(t) = \sigma C(H) \int_{\mathbb{R}} \frac{e^{it\xi} - 1}{|\xi|^{H+1/2}} \widehat{W}(d\xi) \quad \text{pour } t \in \mathbb{R}$$

où  $W(d\xi)$  est une mesure brownienne  $\widehat{W}(d\xi)$  est sa transformée de Fourier au sens des distributions et  $C^2(H) = \pi^{-1}H\Gamma(2H)\sin(\pi H)$ . Aussi peut-on définir une généralisation  $X_\rho = \{X_\rho(t), t \in \mathbb{R}\}$  de ce processus appelé mouvement brownien fractionnaire multi-

échelle défini dans Bardet et Bertrand (2007a) qui est tel que:

$$X_\rho(t) = \int_{\mathbb{R}} \frac{e^{it\xi} - 1}{\rho(\xi)} \widehat{W}(d\xi) \quad \text{pour } t \in \mathbb{R}$$

où la fonction  $\rho : \mathbb{R} \rightarrow [0, \infty)$  est une fonction paire telle qu'avec  $0 < \omega_0 < \omega_1 < \omega_2 < \infty$ :

- $\rho(\xi) = \frac{1}{\sigma_0} |\xi|^{H_0+1/2}$  pour  $|\xi| \in [\omega_0, \omega_1[$  avec  $H_0 \in \mathbb{R}$ ,  $\sigma_0 > 0$ ;
- $\rho(\xi) = \frac{1}{\sigma_1} |\xi|^{H_1+1/2}$  pour  $|\xi| \in [\omega_1, \omega_2[$  avec  $H_1 \in \mathbb{R}$ ,  $\sigma_1 > 0$ ;
- $\rho(\xi) = 0$  pour  $|\xi| \notin [\omega_0, \omega_2]$ .

Les principaux avantages d'une telle généralisation du MBF sont les suivantes:

- (a) C'est un modèle paramétrique permettant beaucoup plus de souplesse en modélisation que le MBF;
- (b)  $X_\rho$  à la même représentation spectrale qu'un MBF dans une bande de fréquences et qu'un autre MBF dans une autre bande de fréquences;
- (c) Dans ces bandes de fréquences, les paramètres  $H_0$  et  $H_1$ , que l'on appellera respectivement paramètres fractals basses et hautes fréquences ne sont pas restreints à être dans  $[0, 1]$  mais ils peuvent prendre n'importe quelle valeur de  $\mathbb{R}$ .

Aussi, à partir d'une telle définition d'un mouvement brownien fractionnaire bi-échelle, on peut en déduire le modèle suivant pour les données de fréquences cardiaques, modèle appelé bruit gaussien fractionnaire multi-échelle (ici bi-échelle):

$$Y_\rho(t) = X_\rho(t+1) - X_\rho(t) = 2 \cdot \mathcal{R}e \left( \int_{\mathbb{R}} \frac{e^{it\xi} \sin(\xi/2)}{\rho(\xi)} \widehat{W}(d\xi) \right) \quad \text{pour } t \in \mathbb{R}.$$

Remarquons que  $Y_\rho = \{Y_\rho(t), t \in \mathbb{R}\}$  est un processus gaussien stationnaire et que la fonction  $4 \sin^2(\xi/2) \rho^{-2}(\xi)$  est aussi appelée la densité spectrale de  $Y_\rho$ .

Il est possible de simuler avec un temps de calcul relativement réduit une trajectoire de  $(Y_\rho(1), Y_\rho(2), \dots, Y_\rho(N))$  en utilisant un algorithme utilisant le plongement de la matrice de covariance de  $Y_\rho$  dans une matrice circulante (on trouvera des détails sur cette méthode de simulation dans Doukhan *et al.*, 2003). Les Figures 3 et 5 permettent de comparer une simulation de la trajectoire d'un processus  $Y_\rho$  avec  $H_0 = 1.25$  et  $H_1 = -0.15$  avec une série de fréquences cardiaques instantanées. De près comme de loin la ressemblance entre ces trajectoires semble très grande. A ce stade, on peut déjà être satisfait d'un tel modèle. Voyons maintenant une méthode d'estimation des paramètres fractals (qui nous permettra de comprendre pourquoi on a choisi  $H_0 = 1.25$  et  $H_1 = -0.15$ ) dont découlera également un test d'adéquation pour un tel modèle.

## 5. Estimation des paramètres fractals du modèle, test d'adéquation et application aux données de fréquences cardiaques instantanées

### 5.1. Définition des estimateurs et du test

Notre but pour commencer est d'obtenir une estimation des paramètres fractals  $H_0$  et  $H_1$  à partir d'une trajectoire connue

$$(Y_\rho(1), Y_\rho(2), \dots, Y_\rho(N))$$



du bruit gaussien fractionnaire bi-échelle défini ci-dessus. Pour ce faire, comme dans Bardet et Bertrand (2007a et b), nous allons à nouveau utiliser une analyse par ondelettes, avec une propriété de localisation en fréquence. Cela se traduit par le fait que l'ondelette mère  $\psi : \mathbb{R} \mapsto \mathbb{R}$  est une fonction  $\mathcal{C}^\infty$  qui vérifie les propriétés suivantes:

**Définition des ondelettes mères vérifiant la propriété de localisation en fréquence**

- pour tout  $s \geq 0$ ,  $\int_{\mathbb{R}} |t^s \psi(t)| dt < \infty$ ;
- la transformée de Fourier  $\widehat{\psi}(\xi)$  de  $\psi$  est une fonction paire dont le support est  $[-\beta, -\alpha] \cup [\alpha, \beta]$  avec  $0 < \alpha < \beta$ .

L'ondelette de Lemarié-Meyer est un exemple d'une telle ondelette mère. En utilisant les résultats de Bardet et Bertrand (2009), pour tout  $a > 0$ ,  $(d_{Y_\rho}(a, b))_{b \in \mathbb{R}}$  est un processus stationnaire gaussien centré et

$$\mathbb{E}(d_{Y_\rho}^2(a, \cdot)) = 8a \int_{\alpha/a}^{\beta/a} |\widehat{\psi}(au)|^2 \frac{\sin^2(u/2)}{\rho^2(u)} du.$$

Aussi pour  $i = 0, 1$ , si l'on choisit une échelle  $a$  telle que  $[\frac{\alpha}{a}, \frac{\beta}{a}] \subset [\omega_i, \omega_{i+1}]$ , i.e.  $a \in [\frac{\beta}{\omega_{i+1}}, \frac{\alpha}{\omega_i}]$ , alors:

$$\mathbb{E}(d_{Y_\rho}^2(a, \cdot)) = 8\sigma_i^2 \int_{\alpha}^{\beta} |\widehat{\psi}(u)|^2 \frac{\sin^2(u/2a)}{(u/a)^{2H_i+1}} du \simeq K(\psi, H_i, \sigma_i) a^{2H_i-1} (1 + O(a^{-2})),$$

pour  $a$  suffisamment grand. avec  $K(\psi, H_i, \sigma_i) = \left(2\sigma_i^2 \int_{\alpha}^{\beta} \frac{|\widehat{\psi}(u)|^2}{(u)^{2H_i-1}} du\right)$  ne dépendant que de  $\psi$ ,  $H_i$  et  $\sigma_i$ . Remarquons qu'une telle propriété ne peut être vérifiée que si le support de  $\psi$  est tel que:

$$\frac{\beta}{\alpha} < \min \left\{ \frac{\omega_1}{\omega_0}, \frac{\omega_2}{\omega_1} \right\}.$$

En utilisant les résultats de Moulines *et al* (2007) ou Bardet *et al* (2007), on montre que pour  $a \in [\frac{\beta}{\omega_{i+1}}, \frac{\alpha}{\omega_i}]$  et en supposant que  $a$  donc  $\frac{\beta}{\omega_{i+1}}$  est suffisamment grand, la variance empirique renormalisée  $S_N(a)/a^{2H_i-1}$ , où  $S_N$  est définie par (3) et calculée à partir de  $(Y_\rho(1), \dots, Y_\rho(N))$ , converge vers  $K(\psi, H_i, \sigma_i)$  suivant un théorème de la limite centrale avec une vitesse de convergence en  $\sqrt{N/a}$ . Pour un tel théorème, on doit avoir les relations:

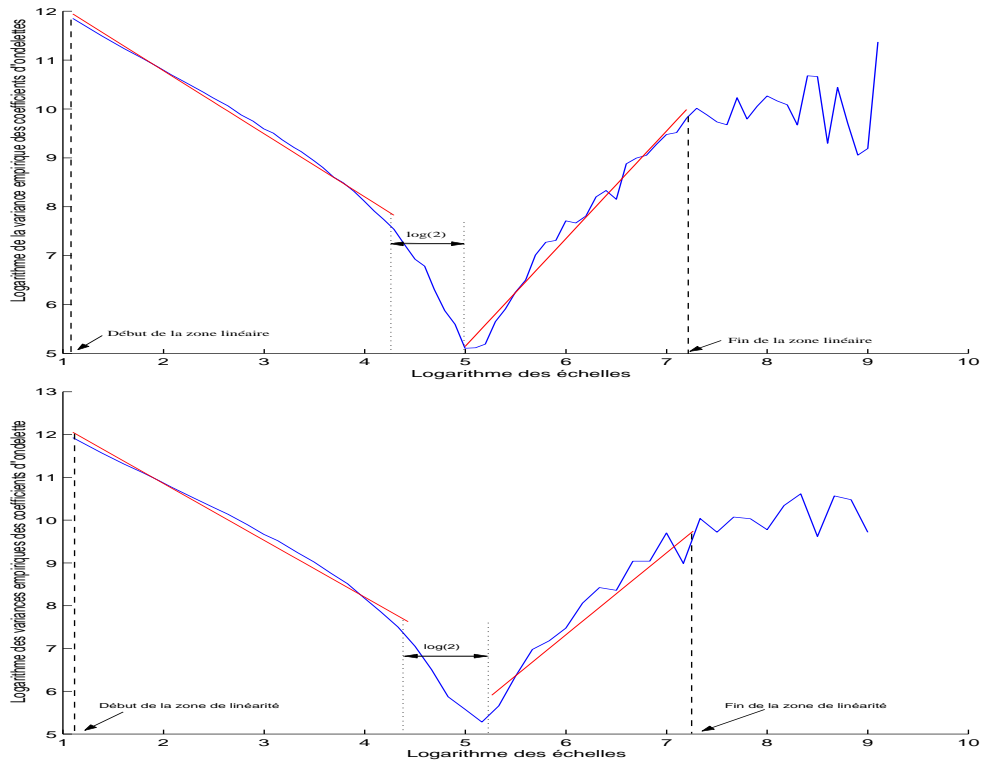
$$N/a \rightarrow \infty \quad \text{et} \quad N/a^5 \rightarrow 0.$$

De plus, pour  $\ell$  tel que  $\ell a \leq \frac{\alpha}{\omega_i}$ , une log-log-régression des  $(i a, S_N(i a))_{1 \leq i \leq \ell}$  fournit un estimateur de  $H$  noté  $\widehat{H}_{OND}^{(2)}$ , donc ici

$\widehat{H}_{OND}^{(2)}$  : estimateur  $\widehat{H}_{OND}$  de la formule (4) avec  $\psi$  une ondelette vérifiant la propriété de localisation en fréquence.

On montre donc que  $\widehat{H}_{OND}^{(2)}$  suit un théorème de la limite centrale avec une vitesse de convergence en  $\sqrt{N/a}$  (ce qui revient à un supremum de vitesse de convergence en  $N^{2/5}$ ). De plus, comme précédemment, un test d'adéquation du Chi-deux peut-être construit à partir d'une distance entre les points  $(\log(ia), \log(S_N(ia)))_{1 \leq i \leq \ell}$  et la droite de régression par moindres carrés pseudo-généralisés (on trouvera plus de détails sur la construction de ce test dans Bardet et Bertrand, 2007b).

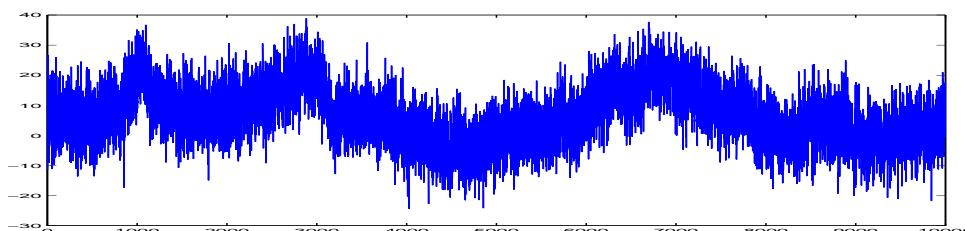
La Figure 8 donne deux exemples pour des données de fréquences cardiaques instantanées de graphes logarithmiques des variances empiriques des coefficients d'ondelettes avec une ondelette  $\psi$  vérifiant la propriété de localisation en fréquence. Le nombre d'échelles choisies est bien supérieur à celui de la Figure 4. A comparer ces graphes pour l'Athlète 4, on remarque que la propriété de localisation en fréquence de l'ondelette mère permet de diminuer notablement la zone intermédiaire entre les deux zones de linéarité (qui est bien plus importante avec l'ondelette mère de Daubechies d'ordre 6).



**Fig. 6.** Le graphe logarithmique des variances empiriques des coefficients d'ondelettes pour le milieu de course de l'Athlète 2 (en haut) et l'Athlète 4 (en bas), où l'ondelette  $\psi$  est une ondelette de type Lemarié-Meyer avec  $\beta/\alpha = 2$

**Remarque:** Il peut y avoir désormais une ambiguïté sur le modèle estimé (chose qui n'était pas possible lorsque l'on considérait un bruit gaussien fractionnaire et un mouvement

brownien fractionnaire). En effet, lorsque l'on obtient  $H_{OND,0}^{(2)} \simeq 1.25$  et  $H_{OND,1}^{(2)} \simeq -0.15$ , on peut penser que le modèle sous-jacent est un bruit gaussien fractionnaire bi-échelle (stationnaire) avec paramètres fractals 1.25 et  $-0.15$ . Mais on pourrait aussi penser à un mouvement brownien fractionnaire bi-échelle (à accroissements stationnaires) avec paramètres 0.25 et  $-1.15$ . La Figure 7 présente la simulation d'une trajectoire d'un tel processus et effectivement on ne voit guère de différence avec celle du bruit gaussien fractionnaire bi-échelle. Mais comme on a fait l'hypothèse à peu près vérifiée que la moyenne et la variance restaient constantes sur chaque phase de course, on préférera modéliser avec un processus stationnaire qu'un processus à accroissements stationnaires.



**Fig. 7.** Trajectoire d'un mouvement brownien fractionnaire bi-échelle de paramètres  $H_0 = 0.25$ ,  $H_1 = -1.15$ ,  $\omega_0 = 0.001$ ,  $\omega_1 = 0.05$  et  $\omega_2 = 10$ .

## 5.2. Estimation adaptative des zones de linéarité

La principale difficulté pour un définir de tels estimateurs et tests est de déterminer les "bonnes" zones de fréquence  $[\omega_0, \omega_1]$  et  $[\omega_1, \omega_2]$ . Nous allons diviser en deux étapes cette détermination que nous allons rendre automatique.

- (a) La première étape consiste à sélection le début de la zone de linéarité des petites échelles et la fin de la zone de linéarité des grandes échelles (on se reportera aux graphes de la Figure 8 pour une meilleure compréhension). Pour ce faire, nous allons utiliser un algorithme mis en place dans Bardet *et al.* (2008). On fixe  $\ell$  (dans les applications, on choisira  $\ell = 10$  à 15 suivant le nombre de données disponibles). Ensuite, on calcule la somme des carrés des résidus de la régression linéaire simple des  $(\log(i a), \log(S_N(i a)))_{1 \leq i \leq \ell}$  en faisant varier  $a$  dans la grille

$$\left\{ \left[ \left( \frac{N}{\ell} \right)^0 \right], \left[ \left( \frac{N}{\ell} \right)^{\frac{1}{\ell \log N}} \right], \left[ \left( \frac{N}{\ell} \right)^{\frac{2}{\ell \log N}} \right], \dots, \left[ \left( \frac{N}{\ell} \right)^{\frac{\ell \log N}{\ell \log N}} \right] \right\}.$$

En regardant le graphe de la Figure 8, on conçoit bien que la somme des carrés résiduels va avoir aux moins deux minima (correspondant à peu près au début des zones de linéarité). Le premier minima va nous permettre de définir le début de la zone de linéarité des petites échelles, noté  $\hat{a}_{min}$ . On posera donc  $\hat{\omega}_2 = \frac{\beta}{\hat{a}_{min}}$ .

Pour déterminer la fin de la zone de linéarité des grandes échelles, on reprend le même procédé, mais "à l'envers": on considère la sommes des carrés résiduels de la régression linéaire simple des  $(\log(a/i), \log(S_N(a/i)))_{1 \leq i \leq \ell}$  en faisant varier  $a$  dans la grille

$$\left\{ \left[ \ell \left( \frac{N}{\ell} \right)^0 \right], \left[ \ell \left( \frac{N}{\ell} \right)^{\frac{1}{\ell \log N}} \right], \left[ \ell \left( \frac{N}{\ell} \right)^{\frac{2}{\ell \log N}} \right], \dots, \left[ \ell \left( \frac{N}{\ell} \right)^{\frac{\ell \log N}{\ell \log N}} \right] \right\}.$$

On peut encore s'attendre à obtenir au moins deux minima; le dernier des minima va correspondre à notre estimation de la fin de la zone de linéarité des grandes échelles, noté  $\hat{a}_{max}$ . On posera donc  $\hat{\omega}_0 = \frac{\alpha}{\hat{a}_{max}}$ .

- (b) On travaille désormais dans la bande d'échelles comprises entre  $\hat{a}_{min}$  et  $\hat{a}_{max}$ . Pour déterminer une estimation de  $\omega_1$ , on utilise la même méthode que celle présentée dans Bardet et Bertrand (2007b). On va ici encore minimiser la somme des carrés de résidus, mais cette fois ce seront les résidus de deux régressions linéaires, avec la condition qu'entre la plus grande échelle de la première zone de linéarité et la plus petite de la seconde zone le rapport doit être de  $\beta/\alpha$  (c'est le  $\log(2)$  qui apparaît sur les graphes de la Figure 8 car on a utilisé ici  $\beta/\alpha = 2$ ). Pour la double régression qui minimise la somme des carrés de résidus, on note  $\hat{a}_1$  la plus grande échelle pour la zone de linéarité des petites échelles. Alors  $\omega_1 = \frac{\alpha}{\hat{a}_1}$ .

Les résultats obtenus dans Bardet *et al.* (2008) et Bardet et Bertrand (2007b) assurent les convergences en probabilité de  $\hat{\omega}_i$  vers  $\omega_i$  pour  $i = 0, 1, 2$ . Une fois les  $\hat{\omega}_i$  calculés, il ne reste plus qu'à estimer les deux paramètres fractals  $H_0$  et  $H_1$  soit  $\hat{H}_{OND,0}^{(2)}$  et  $\hat{H}_{OND,1}^{(2)}$  obtenu par l'analyse par ondelettes dans les bandes d'échelles  $[\tilde{a}_{min}, \tilde{a}_1]$  et  $[\tilde{a}_2, \tilde{a}_{max}]$ , respectivement (ces bandes d'échelles sont choisies comme dans Bardet et Bertrand, 2007b, de telle manière que  $[\tilde{a}_{min}, \tilde{a}_1] \subset [\hat{a}_{min}, \hat{a}_1]$  et  $[\tilde{a}_2, \tilde{a}_{max}] \subset [2\hat{a}_1, \hat{a}_{max}]$ ). En reprenant la même preuve que Bardet et Bertrand (2007b), on peut montrer que  $\hat{H}_{OND,0}^{(2)}$  et  $\hat{H}_{OND,1}^{(2)}$  convergent vers  $H_0$  et  $H_1$  suivant un théorème de la limite centrale à la vitesse  $\sqrt{N/\tilde{a}}$ , ce qui conduit encore à un supremum de vitesse de convergence en  $N^{2/5}$ . Un test d'adéquation de type Chi-deux portant sur les deux zones de linéarité et décrit comme précédemment peut aussi être mis en place.

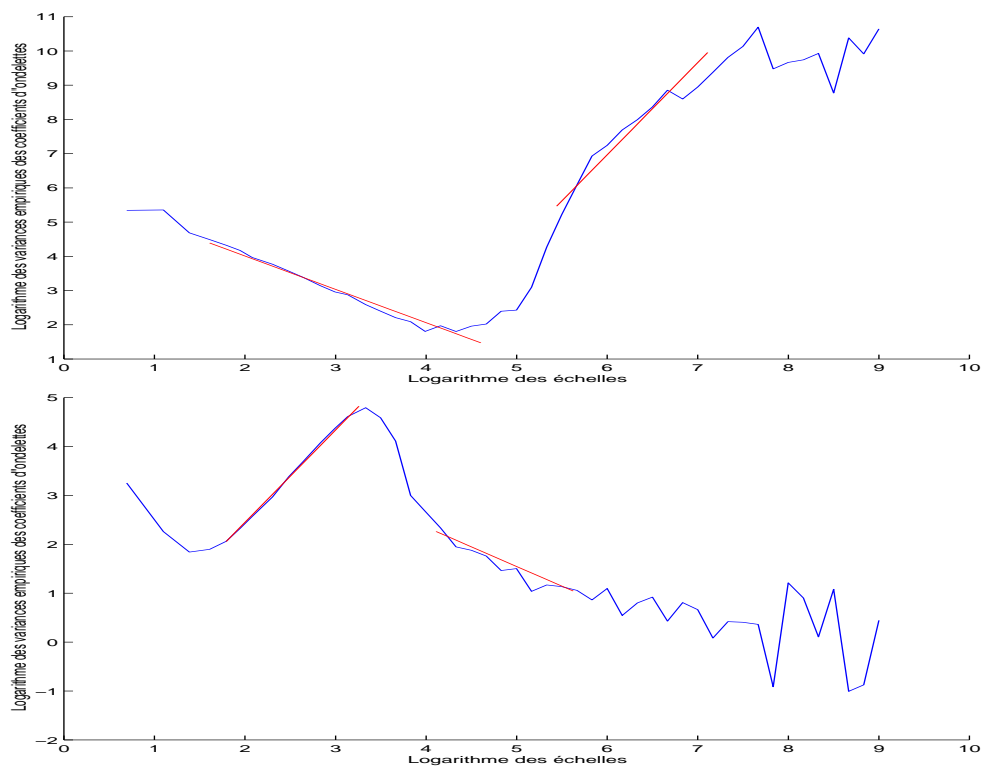
**Remarque:** Toute la démarche proposée (estimation et test) l'a été dans le cadre d'un modèle paramétrique, le bruit gaussien fractionnaire multi-échelle. Nous l'avons développée dans le cas particulier de deux régimes non nuls, mais cela se généralise clairement pour un nombre  $K$  de régimes. Cependant cela reste paramétrique, et une méthode de type maximum de contraste aurait également pu être envisagée. Mais toute cette démarche et les résultats asymptotiques évoqués sont également valables pour la classe de modèles semi-paramétriques telles que  $\rho(\xi) = \sigma_i^{-1}|\xi|^{H_i+1/2}$  pour  $|\xi| \in [\omega_i, \omega_{i+1}[$  pour  $i = 0, 1, \dots, K$  et avec  $0 < \omega_0 < \omega_1 < \dots < \omega_{K+1} < \infty$ , et pour  $|\xi| \in [0, \omega_0[ \cup ]\omega_{K+1}, \infty[$ , on ne fixe pas  $\rho$  à 0 comme précédemment, mais on demande juste que  $\rho$  vérifie

$$\int_{\mathbb{R}} (1 \wedge |\xi|^2) \frac{1}{\rho^2(\xi)} d\xi < \infty \quad , \quad \rho(\omega_0^-) \neq \rho(\omega_0^+) \quad \text{et} \quad \rho(\omega_{K+1}^-) \neq \rho(\omega_{K+1}^+)$$

(ces deux dernières conditions permettent l'estimation de  $\omega_0$  et  $\omega_2$ ; des conditions plus faibles sur la dérivée première pourrait également les remplacer, mais cela nous emmènerait trop loin). On voit donc par là que la méthode proposée s'applique à des processus gaussiens dont le log-variogramme des coefficients d'ondelettes présente deux zones de linéarité sans tenir compte de ce qui se passe ailleurs; cela nous semble particulièrement adapté aux données de fréquences cardiaques.

### 5.3. Résultats de simulation et applications aux données de fréquences cardiaques instantanées

**Résultats de simulations:** En premier lieu, nous avons effectué des simulations de bruits gaussiens fractionnaires bi-échelles dont nous avons ensuite estimé les paramètres fractals après avoir estimé les fréquences  $\omega_i$ . Pour ne pas trop surcharger le papier, nous ne présentons qu'un choix de deux types de bruits gaussiens fractionnaires bi-échelles: un qui mime les données cardiaques en milieu de course ( $H_0 = 1.25$ ,  $H_1 = -0.15$ ,  $\omega_0 = 0.001$ ,  $\omega_1 = 0.05$  et  $\omega_2 = 3$ ) et un autre avec des paramètres très différents ( $H_0 = -0.5$ ,  $H_1 = 1.5$ ,  $\omega_0 = 0.01$ ,  $\omega_1 = 0.3$  et  $\omega_2 = 3$ ). La Figure 8 un exemple du graphe de log-log régression obtenu pour chacun de ces processus.



**Fig. 8.** Deux exemples de graphes logarithmiques des variances empiriques des coefficients d'ondelettes pour des bruits gaussiens fractionnaires bi-échelles, le premier avec  $H_0 = 1.25$ ,  $H_1 = -0.15$ ,  $\omega_0 = 0.001$ ,  $\omega_1 = 0.05$  et  $\omega_2 = 3$  (en haut) et le second avec  $H_0 = -0.5$ ,  $H_1 = 1.5$ ,  $\omega_0 = 0.01$ ,  $\omega_1 = 1$  et  $\omega_2 = 3$  (en bas), où l'ondelette  $\psi$  est une ondelette de type Lemarié-Meyer avec  $\beta/\alpha = 2$

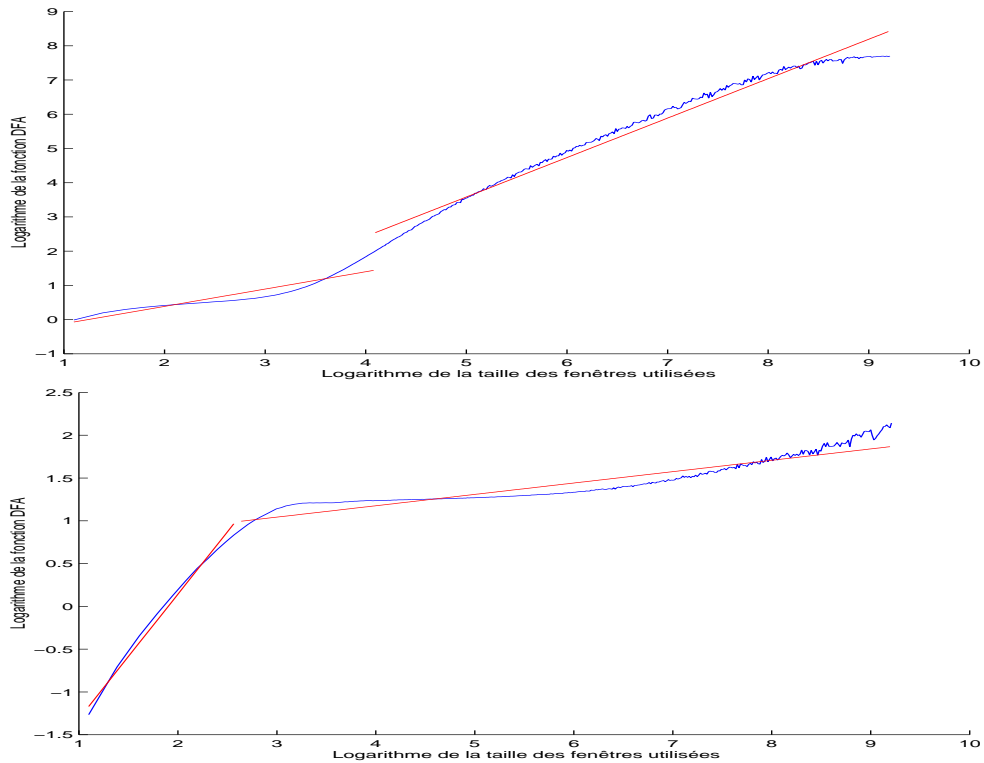
La Table 2 présente le tableau des résultats numériques pour ces simulations après 20 replications indépendantes. On s'aperçoit que la méthode d'estimation par ondelettes est

**Table 2.** Estimation des différents paramètres pour deux bruits gaussiens fractionnaires bi-échelles de taille 10000 à partir de 20 replications

	MFGN				
	$H_0$	$H_1$	$\omega_0$	$\omega_1$	$\omega_2$
Valeur	1.25	-0.15	0.001	0.05	3
Moyenne Estimateur Ondelette	1.23	-0.02	0.007	0.05	1.73
Ecart-type Estimateur Ondelette	0.22	0.09	0.09	0.04	0.52
Moyenne Estimateur DFA	1.46	0.33	-	0.03	-
Ecart-type Estimateur DFA	0.04	0.02	-	0.004	-
Valeur	-0.5	1.5	0.01	0.3	3
Moyenne Estimateur Ondelette	-0.17	1.36	0.01	0.20	1.83
Ecart-type Estimateur Ondelette	0.09	0.05	0.005	0.02	0.19
Moyenne Estimateur DFA	0.14	1.45	-	0.07	-
Ecart-type Estimateur DFA	0.01	0.02	-	0.003	-

convergente pour les différents paramètres du modèle même si l'écart-type de ces estimations est parfois assez élevé. Nous avons également mis en place une méthode de détection automatique de la rupture des deux zones de linéarité pour la méthode DFA (méthode fondée également sur une minimisation de la somme des carrés des résidus dans les 2 zones). On s'aperçoit alors que les estimateurs du paramètre fractal hautes fréquences ainsi que celui de la fréquence de coupure  $\omega_1$  ne sont pas convergents, le paramètre fractal basses fréquences étant lui correctement estimé (dans tous les cas, les écarts-type sont très faibles). Si on augmente la taille de l'échantillon (par exemple en passant à une taille de 100000) on n'améliore pas du tout la situation, voire même elle empire. On comprendra mieux la situation en observant les graphes de log-régression de la fonction DFA pour les deux bruit gaussiens fractionnaires bi-échelles précédents (voir la Figure 9 avec  $N = 100000$ ). On s'aperçoit alors qu'il n'y pas vraiment de claire zone de linéarité car les deux régimes "fractionnaires" du processus agissent l'un sur l'autre et ne sont pas "isolés" comme c'est le cas avec la méthode d'estimation par ondelettes (un tel phénomène avait déjà été identifié pour la méthode des variations quadratiques, dans Bardet et Bertrand, 2007a). Dans le cas de processus avec deux régimes "fractionnaires" comme le bruit gaussien fractionnaire bi-échelle, il est donc préférable de ne pas utiliser la méthode DFA pour estimer les paramètres.

**Résultats de l'application des méthodes d'estimation et du test d'adéquation aux données:** Nous avons appliqué les deux méthodes adaptatives (ondelette avec propriété de localisation en fréquence et DFA) pour estimer les paramètres fractals basses fréquences  $H_0$  et hautes fréquences  $H_1$ , ainsi que la fréquence de coupure  $\omega_1$ . Les résultats sont récapitulés dans la Table 3. Les estimations sur la course entière sont également présentées même si elles ne sont pas vraiment exploitables (si ce n'est de manière négative, pour marquer en particulier que la méthode DFA ne peut être appliquée sans précaution, ce qui est un peu moins vrai avec la méthode d'ondelettes). On s'aperçoit que l'estimation  $H_{OND,0}^{(2)}$  du paramètre fractal basses fréquences augmente au fil de la course (il passe d'une moyenne de 1.20 à 1.40 en milieu de course, puis 1.55 en fin de course), alors qu'il reste presque constant avec la méthode DFA (entre 1.25 et 1.30). Grâce à l'estimation par ondelettes, on observe aussi que le paramètre fractal hautes fréquences reste constant tout au long de la course (entre  $-0.2$  et  $-0.1$ ) et que la fréquence de coupure semble ne pas non plus



**Fig. 9.** Deux exemples de graphes logarithmiques des fonctions DFA pour des bruits gaussiens fractionnaires bi-échelles, le premier avec  $H_0 = 1.25$ ,  $H_1 = -0.15$ ,  $\omega_0 = 0.001$ ,  $\omega_1 = 0.05$  et  $\omega_2 = 3$  (en haut) et le second avec  $H_0 = -0.5$ ,  $H_1 = 1.5$ ,  $\omega_0 = 0.01$ ,  $\omega_1 = 1$  et  $\omega_2 = 3$  (en bas), et  $N = 100000$  dans les deux cas.

évoluer. En revanche, et comme on pouvait s’y attendre depuis le résultat des simulations, la méthode DFA n’est absolument pas utilisable pour estimer le paramètre fractal hautes fréquences et la fréquence de coupure.

Par ailleurs, le test d’adéquation a été également appliqué. Les résultats sont les suivants: le test est rejeté 7 fois sur 9 lorsque l’on considère la course en son intégralité, mais entre une et trois fois si on considère chacune des phases de course. Ceci permet de penser que la modélisation par un bruit gaussien fractionnaire bi-échelle est réellement adéquate pour ce type de données.

## 6. Conclusion

Voici les différents points à retenir de notre étude:

- (a) En premier lieu, la méthode DFA n’est pas une méthode efficace pour estimer les deux paramètres fractals des séries de fréquences cardiaques instantanées, en particulier lorsqu’une tendance est présente.

**Table 3.** Estimation des différents paramètres des différentes phases de course, avec les deux méthodes.

		Course entière			Début de course			Milieu de course			Fin de course		
		$H_0$	$H_1$	$\omega_1$	$H_0$	$H_1$	$\omega_1$	$H_0$	$H_1$	$\omega_1$	$H_0$	$H_1$	$\omega_1$
Ath1	$H_{OND}^{(2)}$	0.98	-0.18	0.03	1.40	-0.12	0.05	1.46	-0.17	0.05	1.70	-0.18	0.05
	$H_{DFA}$	0.96	0.86	0.01	1.17	0.97	0.05	1.09	0.71	0.01	1.25	0.34	0.02
Ath2	$H_{OND}^{(2)}$	1.23	-0.19	0.05	1.35	-0.16	0.06	1.60	-0.14	0.07	1.59	-0.15	0.07
	$H_{DFA}$	1.11	1.05	0.01	0.90	1.19	0.02	1.58	0.74	0.05	1.26	0.72	0.03
Ath3	$H_{OND}^{(2)}$	1.40	-0.13	0.07	1.24	-0.08	0.05	1.49	-0.14	0.07	-	-	-
	$H_{DFA}$	1.15	0.81	0.05	1.32	0.81	0.03	1.22	0.65	0.04	-	-	-
Ath4	$H_{OND}^{(2)}$	1.60	-0.15	0.06	-	-	-	1.25	-0.15	0.05	1.33	-0.14	0.07
	$H_{DFA}$	1.18	0.71	0.04	-	-	-	1.17	0.64	0.03	1.23	0.92	0.03
Ath5	$H_{OND}^{(2)}$	1.59	-0.14	0.06	0.99	-0.13	0.05	1.36	-0.14	0.07	1.88	-0.11	0.08
	$H_{DFA}$	1.28	0.71	0.05	1.33	0.79	0.05	1.23	0.66	0.05	1.46	1.11	0.05
Ath6	$H_{OND}^{(2)}$	0.99	-0.20	0.03	1.11	-0.15	0.07	1.70	-0.15	0.06	1.75	-0.10	0.06
	$H_{DFA}$	1.27	1.08	0.03	1.43	1.02	0.05	1.25	0.92	0.03	1.36	1.28	0.1
Ath7	$H_{OND}^{(2)}$	1.51	-0.14	0.07	1.62	-0.20	0.05	1.42	-0.15	0.07	1.42	-0.15	0.07
	$H_{DFA}$	1.24	0.77	0.03	1.22	0.66	0.03	1.26	0.78	0.03	1.25	0.79	0.05
Ath8	$H_{OND}^{(2)}$	1.07	-0.25	0.03	0.50	-0.10	0.04	1.24	-0.20	0.05	-	-	-
	$H_{DFA}$	1.31	0.74	0.04	1.33	1.03	0.09	1.06	0.74	0.02	-	-	-
Ath9	$H_{OND}^{(2)}$	1.53	-0.18	0.05	1.05	-0.12	0.09	1.27	-0.23	0.04	1.25	-0.22	0.04
	$H_{DFA}$	1.37	0.82	0.02	1.20	0.81	0.01	1.25	0.83	0.01	1.22	0.94	0.02

- (b) En second lieu, il est important de découper une course à pied de longue haleine comme un marathon en différentes phases. Les modes de fonctionnement physiologiques semblent différents.
- (c) La modélisation par un bruit gaussien fractionnaire bi-échelle semble très fructueuse dans le cas présent des fréquences cardiaques, mais sans doute aussi dans de nombreux autres.
- (d) Totalemment appropriée à ce modèle (puisqu'elle fonctionne aussi avec échelles ou fréquences...), la méthode d'analyse par ondelettes montre ici l'étendue de ses possibilités de robustesse, d'estimation et même de test.
- (e) Vis-à-vis du fonctionnement du coeur en situation de course d'endurance, on montre que le paramètre fractal basses fréquences semble être caractéristique de l'apparition d'une fatigue. Cela se marque par une augmentation de la régularité des battements du coeur. Numériquement, on peut montrer que pour les 9 athlètes considérés on obtient que  $\hat{H}_{dbut} \simeq 1.2$ ,  $\hat{H}_{milieu} \simeq 1.4$  et  $\hat{H}_{fin} \simeq 1.55$ . On peut comparer ces résultats avec ceux obtenus dans Goldberger *et al.* (2002) et Peng *et al.* (1993, 1994): pour des patients dont on a mesuré les durées d'interbattements pendant 24h, ils ont obtenu (avec la méthode DFA; il serait intéressant d'appliquer la méthode d'analyse par ondelettes dans ce cas...) que pour les patients en bonne santé  $H = 1.01 \pm 0.16$  alors que pour des patients présentant un dysfonctionnement cardiaque  $H = 1.24 \pm 0.22$ . Il y a donc quelque chose de pathologique à avoir des battements de coeur trop réguliers (d'ailleurs n'est pas plus régulier qu'un coeur qui ne bat plus...) et la fin du marathon peut être chez certains une source d'inquiétude réelle.
- (f) Enfin, la méthode adaptative que nous avons choisie nous permet d'estimer  $\omega_1$  la fréquence de coupure. Cette fréquence n'évolue pas au fil de la course et reste à



peu près constante à une valeur de 0.05. Si on traduit cela en Hertz, il faut prendre en compte la durée moyenne des inter-battements cardiaques et on obtient environ  $0.13Hz$ : c'est très proche des  $0.15Hz$  fournis par la référence médicale de l'European Soc. Cardiology and the North American Society of Pacing and Electrophysiology (1996). Ce paramètre physiologique distinguant les deux régimes de contrôle nerveux du coeur ne semble donc pas être modifié durant une course comme le marathon.

De nombreuses perspectives s'ouvrent au regard de cette première étude. Tout d'abord, il faudrait tester les résultats obtenus ici sur d'autres jeux de données (pendant des courses à pied, mais aussi pour des sujets au repos, durant leurs sommeils...). Ensuite, puisqu'une augmentation du paramètre fractal basses fréquences semble aller de pair avec l'apparition d'une fatigue du coeur, il serait intéressant de pouvoir détecter (séquentiellement) lorsque cette fatigue est trop importante (voire dangereuse): on aurait là un moyen de prévenir des accidents cardiaques... Enfin, il serait intéressant de corrélérer les évolutions de ce paramètre avec celles éventuelles d'autres variables: vitesse instantanée du coureur, fréquence pulmonaire, taux d'acide lactique,... qui constituent les nouveaux jeux de données recueillis par le laboratoire INSERM 902 du Génomole de l'Université d'Evry.

## References

- [1] Abry, P., Flandrin, P., Taqqu, M. et Veitch, D. (2003) Self-similarity and long-range dependence through the wavelets lens. *Theory and applications of Long-Range Dependence*, 527-556, Birkhäuser.
- [2] Abry, P., Veitch, D. et Flandrin, P. (1998) Long-range dependent: revisiting aggregation with wavelets. *J. Time Ser. Anal.*, 253-266.
- [3] Absil, P.A., Sepulchre, R., Bilge, A. et Gérard, P. (1999). Nonlinear analysis of cardiac rhythm fluctuations using DFA method. *Physica A*, 235-244.
- [4] Bai, J. (1998) Least squares estimation of a shift in linear processes. *J. of Time Series Anal.*, 5, 453-472.
- [5] Bai, J. et Perron, P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, 66, 47-78.
- [6] Bardet, J.M. (2002) Statistical study of the wavelet analysis of fractional Brownian motion. *IEEE Trans. Inform. Theory.*, 48, 991-999.
- [7] Bardet, J.M. et Bertrand, P. (2007a) Definition, properties and wavelet analysis of multiscale fractional Brownian motion. *Fractal*, 15, 73-87.
- [8] Bardet, J.M. et Bertrand, P. (2007b) Identification of the multiscale fractional Brownian motion with biomechanical applications. *Journal of Time Series Analysis*, 28, 1-52.
- [9] Bardet, J.M. et Bertrand, P. (2009) Estimation of the spectral density of a process observed at random times. Preprint HAL.
- [10] Bardet, J.M., Bibi, H. et Jouini, A. (2008) Adaptive wavelet based estimator of the memory parameter for stationary Gaussian processes. *Bernoulli*, 14, 691-724.

- [11] Bardet, J.M. et Kammoun, I. (2008) Detecting changes in the fluctuations of a Gaussian process and an application to heartbeat time series. Preprint HAL.
- [12] Bardet, J.M. et Kammoun, I. (2008) Asymptotic Properties of the Detrended Fluctuation Analysis of Long Range Dependence Processes. *IEEE Trans. Inform. Theory.*, 54, 2041-2052.
- [13] Bardet, J.M., Lang, G., Moulines, E. et Soulier, P. (2000) Wavelet estimator of long-range dependent processes. *Statistical Inference for Stochastic Processes*, 3, 85-99.
- [14] Bardet, J.M. et Tudor, C. (2008) Hurst index estimation for the Rosenblatt process: wavelets and chaos expansion. Preprint HAL.
- [15] Bertrand, P.R., Teyssière, G., Boudet et Chamoux, A. (2009). . Preprint HAL.
- [16] Birgé, L. et Massart, P. (2007) Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields*, 138, 33-73.
- [17] Flandrin, P. (1992) Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. on Inform. Theory*, 38, 910-917.
- [18] Fox, R. et Taqqu, M.S. (1986) Large-sample properties of parameter estimates for strongly dependent Gaussian time series. *Ann. Statist.*, 14, 517-532.
- [19] Goldberger, A.L., Amaral, L.A.N., Hausdorff, J.M., Ivanov, P.C., Peng, C.K. et Stanley, H.E. (2002) Fractal Dynamics in Physiology: Alterations with Disease and Aging. *PNAS*, 99, 2466-2472.
- [20] Gough N.A.J. (1992). Fractals, chaos and fetal heart rate. *Lancet*, 339, 182-183.
- [21] Gough N.A.J. (1993). Fractal analysis of fetal heart rate variability. *Physiol. Meas.*, 14, 309-315.
- [22] Havlin, S., Amaral, N.A., Ashkenazy, Y., Goldberger, A.L., Ivanov, P.Ch., Peng, C.K. et Stanley, H.E. (1999). Application of statistical physics to heartbeat diagnosis. *Physica A*, 274, 99-110.
- [23] K. Hu, P.C. Ivanov, Z. Chen, P. Carpena et H.E. Stanley (2001). Effect of trends on detrended fluctuation analysis, *Physical Review E*, 64, 011114.
- [24] Ivanov, P.C., Rosenblum, M.G., Peng, C.K., Mietus, J., Havlin, S., Stanley, H.E. et Goldberger, A.L. (1996). Scaling Behaviour of Heartbeat Intervals obtained by Wavelet-Based Time-Series Analysis. *Nature*, 383, 323-327.
- [25] Ivanov, P.C., Rosenblum, M.G., Nunes Amaral, L.A., Struzik, Z.S., Havlin, S., Goldberger, A.L. et Stanley, H.E. (1999). Multifractality in human heartbeat dynamics. *Nature*, 399, 461-465.
- [26] Ivanov, P.Ch., Bunde, A., Amaral, L.A.N., Havlin, S., Fritsch-Yelle, J., Baevsky, R.M., Stanley, H.E. et Goldberger, A.L. (1999). Sleep-wake differences in scaling behavior of the human heartbeat: Analysis of terrestrial and long-term space flight data. *Europhys. Lett.*, 48, 594-600.

- [27] Ivanov, P.C., Amaral, L.A.N., Goldberger, A.L., Havlin, S., Rosenblum, M.G., Stanley, H.E. et Struzik, Z. (2001). From  $1/f$  Noise to Multifractal Cascades in Heartbeat Dynamics. *Chaos*, 11, 641-652.
- [28] Kantelhardt, J.W., Ashkenazy, Y., Ivanov, P.Ch., Bunde, A., Havlin, S., Penzel, T., Peter, H.J. et Stanley, H.E. (2002). Characterization of sleep stages by correlations of magnitude and sign of heartbeat increments. *Phys. Rev. E*, 65, 051908.
- [29] Karasik, R., Sapir, N., Ashkenazy, Y., Ivanov, P. Ch., Dvir, I., Lavie, P. et Havlin, S. (2002). Correlation differences in heartbeat fluctuations during rest and exercise. *Physical Review E*, 66, 062902.
- [30] Krstacic, G., Krstacic, A., Martinis, M., Vargovic, E., Knezevic, A., Smalcelj, A., Jembrek-Gostovic, M., Gamberger, D. et Smuc, T. (2002). Non-linear analysis of heart rate variability in patients with coronary heart disease. *Computers in Cardiology*, 673-675.
- [31] Lavielle, M. (1999) Detection of multiple changes in a sequence of random variables. *Stoch. Process Appl*, 79-102. <http://www.math.u-psud.fr/lavielle/>.
- [32] Lavielle, M. et Moulines, E. (2000) Least-squares estimation of an unknown number of shifts in a time series. *J. of Time Series Anal.*, 21, 33-59.
- [33] Lavielle M. et Ludeña, C. (2000) The multiple change-points problem for the spectral distribution. *Bernoulli* 6, 845-869.
- [34] Martinis, M., Knezevic, H., Krstacic G. et Vargovic, E. (2004). Changes in the Hurst exponent of heartbeat intervals during physical activities, *Physical review. E*, 70, 012903.
- [35] Moulines, E., Roueff, F. et Taqqu, M.S. (2007), On the spectral density of the wavelet coefficients of long memory time series with application to the log-regression estimation of the memory parameter. *J. Time Ser. Anal.*, 28, 155-187.
- [36] Peng, C.K, Mietus, J., Hausdorff, J., Havlin, S., Stanley, H.E. et Goldberger, A.L. (1993). Long-Range Anticorrelations and Non-Gaussian Behavior of the Heartbeat. *Phys. Rev. Lett.*, 70, 1343-1346.
- [37] Peng, C.K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E. et Goldberger, A.L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49, 1685-1689.
- [38] Peng, C.K., Havlin, S., Stanley, H.E. et Goldberger, A.L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, 5, 82.
- [39] Pikkujäämsä, S.M., Mäkikallio, T.H., Airaksinen, K.E. et Huikuri, H.V. (2001). Determinants and interindividual variation of R-R interval dynamics in healthy middle-aged subjects. *Am. J. Physiol. Heart Circ. Physiol.*, 280, 1400-1406.
- [40] Moulines, E. et Soulier, P. (2003). Semiparametric spectral estimation for fractional processes. In P. Doukhan, G. Oppenheim et M.S. Taqqu editors. *Theory and applications of long-range dependence*, 251-301, Birkhäuser, Boston.
- [41] Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence, *The Annals of statistics*, 23, 1630-1661.

- [42] Roueff, F. et Taqqu, M.S. (2008). Asymptotic normality of wavelet estimators of the memory parameter for linear processes. Preprint arXiv:0805.0778.
- [43] Taqqu, M.S., Teverovsky, V. et Willinger, W. (1995). Estimators for long-range dependence: an empirical study. *Fractals*, 3, 785-798.
- [44] Staudacher, M., Telsner, S., Amann, A., Hinterhuber, H., et Ritsch-Marte, M. (2005). A new method for change-point detection developed for on-line analysis of the heartbeat variability during sleep. *Physica A*, 349, 582-596.
- [45] Vedel, B., Wendt, H., Abry, P. et Jaffard, S. On the impact of the number of vanishing moments on the dependence structures of compound Poisson motion and fractional Brownian motion in multifractal time. Preprint.
- [46] Veitch, D., Abry, P. et Taqqu, M.S. (2003). On the Automatic Selection of the Onset of Scaling, *Fractals*, 11, 377-390.
- [47] Wesfreid, E., Billat, V. et Meyer, Y. (2005) Multifractal analysis of heartbeat time series in human races. *Appl. Comput. Harmon. Anal.*, 18, 329-335.
- [48] Yeh, R.G., Shieh, J.S., Han, Y.Y., Wang Y.J. et Tseng S.C. (2006). Detrended fluctuation analyses of short-term heart rate variability in surgical intensive care units. *Biomed Eng.- Appl*, 18, 67-72.
- [49] Task force of the European Soc. Cardiology and the North American Society of Pacing and Electrophysiology (1996), Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Circulation* 93 (5), 1043-1065.