



**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. 2008. hal-00343361v1

**HAL Id: hal-00343361**

**<https://hal.science/hal-00343361v1>**

Preprint submitted on 1 Dec 2008 (v1), last revised 3 Aug 2009 (v2)

**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 leur paramètre de fractalité

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.*

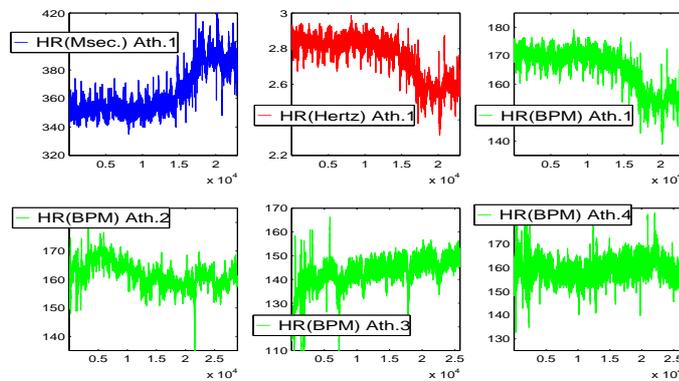
**Summary.** Pour pouvoir caractériser et interpréter des signaux physiologiques comme ceux constitués par les fréquences cardiaques instantanées (ou les durées entre deux battements successifs), il est intéressant de trouver des paramètres qui restent constants au milieu des fluctuations. Depuis une quinzaine d'années, plusieurs papiers ont considéré le paramètre de Hurst, paramètre de longue mémoire, comme une telle constante caractéristique. Cependant leurs résultats posent deux problèmes. En premier lieu, ils utilisent la méthode DFA pour estimer ce paramètre. 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 aux résultats obtenus, le paramètre de Hurst, souvent estimé à des valeurs supérieures à 1, borne supérieure admissible pour un tel paramètre, n'est pas une constante représentative de ces signaux, ce qui signifie que les processus longue mémoire ne sont peut-être pas de bons modèles. Nous proposons des solutions à chacun de 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, une robustesse à des tendances polynomiales et un graphe de log-regression extrêmement informatif. Ensuite, un nouveau modèle semi-paramétrique, appelé processus Gaussien localement fractionnaire, caractérisé par un paramètre, appelé paramètre de fractalité locale, pouvant prendre toute valeur réelle, est défini. La méthode d'analyse par ondelette permet d'estimer ce paramètre mais également de réaliser un test d'adéquation du modèle. 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 du paramètre de fractalité locale au cours de la course, 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 ayant des dysfonctionnements cardiaques. Une telle évolution, que la DFA ne permet pas d'observer et 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 de fractalité locale pourrait donc être essentiel pour garantir la santé de coureurs amateurs.

E-mail: bardet@univ-paris1.fr

**Keywords:** 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.** Battements cardiaques pour l'Athlète 1 en ms, Hertz et BPM (en haut), et pour les Athlètes 2, 3 et 4 en BPM (en bas)

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 fréquence (ou de durée inter-battements) qui sont décelés et étudiés. Cependant, 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). La régularité des trajectoires formées par ces données peut donc s'avérer intéressante à étudier.

Quoique celui-ci puisse aussi bien être relatif à une régularité locale comme à un comportement à long terme, un paramètre de type "fractal" est désormais fréquemment utilisé pour caractériser l'erraticisme des trajectoires. Ainsi, lorsque l'on considère une série chronologique stationnaire dont la série des autocovariances est non-sommable on dit que la série est à longue mémoire et le paramètre de Hurst mesure en quelque sorte la dépendance. Depuis son introduction par Hurst (1951) pour des données hydrologiques, cette propriété de longue mémoire (ou de forte dépendance) a été mise en évidence dans de nombreux champs thématiques: télécommunications, informatique, économie et finance, biomécanique... Ainsi, assez

naturellement, de nombreuses recherches ont été menées dans cette direction pour décrire les fluctuations de signaux physiologiques en fonction du temps (voir par exemple Peng *et al.*, 1993, Goldberger, 2001 et Goldberger *et al.*, 2002). Par ailleurs, pour un processus à accroissements stationnaires, l'exposant de Hölder local permet de mesurer la régularité locale de la trajectoire. Lorsque le processus est en plus autosimilaire (invariant par changement d'échelles), son paramètre d'autosimilarité est cet exposant. Le mouvement brownien fractionnaire (voir plus loin un rappel de sa définition) est l'exemple sans doute le plus célèbre d'un tel processus. On trouvera dans Peng *et al.* (1995) une analyse en ces termes des données de battements cardiaques. Lorsque le processus est dit multifractal, plusieurs paramètres peuvent être représentatifs de la régularité de la trajectoire. Des applications aux données de battements cardiaques peuvent se trouver dans Ivanov *et al.* (1999) ou Absil *et al.* (1999) (le caractère multifractal, et plus précisément bifractal pouvant s'expliquer par les deux comportements précédemment cités).

L'étude statistique qui suit va tenter de répondre à divers questions et problèmes laissés en plan par les études effectués sur ces données. Les problèmes issus de ces études sont les suivants:

- Que mesure-t-on? Une régularité locale ou un comportement de type longue mémoire?
- Quelle est la méthode utilisée pour estimée 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 processus autosimilaire ou à longue mémoire (dont les moments d'ordre 2 existent clairement)?
- Quel modèle proposer pour modéliser ces données de battements cardiaques et peut-on tester l'ajustement d'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 et exhibent des propriétés invisibles lorsque l'on prend la série de données en son intégralité.
- (b) La méthode DFA (Detrended Fluctuation Analysis) est souvent utilisée pour estimer le paramètre "fractal" dans la plupart des travaux cités précédemment (voir aussi Absil *et al.*, 1999, ou Ivanov *et al.*, 1999). Or, cette méthode qui est une extension pour des séries avec tendance de la méthode des variances agrégées, n'est pas une méthode efficace et robuste 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 bien 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, qui a de plus l’avantage de s’adapter au modèle proposé ci-dessous.

- (c) Un nouveau modèle, appelé bruit gaussien localement fractionnaire, est proposé pour modéliser les données dans chacune des phases. Il constitue une généralisation semi-paramétrique du bruit gaussien fractionnaire (BGF en abrégé, obtenu à partir des accroissements du mouvement brownien fractionnaire): sa représentation harmonisable (spectrale) est la même que celle du BGF dans une certaine bande de fréquences, c’est-à-dire une fonction puissance dépendant d’un paramètre appelé paramètre de fractalité locale. Ce paramètre est donc analogue au paramètre de Hurst du BGF sauf qu’il peut prendre toutes valeurs réelles dès que la bande fréquence est finie et ne comprend pas 0. La méthode d’analyse par ondelettes permet d’estimer ce paramètre et également de proposer un test d’adéquation.

Lorsque l’on applique cette méthodologie à nos données cardiaques, on observe un phénomène très intéressant: le test d’adéquation est accepté dans chacune des trois phases des trajectoires (alors qu’il ne l’est pas si on l’applique sur l’intégralité de chaque trajectoire) et on montre qu’il y a une augmentation significative du paramètre de fractalité locale. Ceci ne pouvait pas être observé en utilisant la méthode DFA. En devenant ainsi d’une certaine manière de plus en plus “régulière”, 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 de ce paramètre de fractalité locale peut être représentatif de l’apparition d’une fatigue du coeur. Une trop grande “régularité” de la trajectoire des fréquences cardiaques instantanées peut témoigner d’un danger alors même que les valeurs de ces fréquences ne paraissent pas alarmantes. Les fluctuations pourraient s’avérer plus informatives que les valeurs, et la mise en place de détection séquentielle d’une trop grande régularité cardiaque pourraient permettre de prévenir des risques cardiaques (comme cela arrive malheureusement trop fréquemment lors d’épreuves telles le marathon).

## 2. Détection des différentes phases de courses

Nous avons donc commencer 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 (on trouvera des détails sur ce qui suit dans Lavielle, 1999, Lavielle et Moulines, 2000, et surtout Lavielle et Ludenä, 2001). 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  $\tau = (\tau_1, \tau_2, \dots, \tau_{K-1})$ . 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: is

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

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:

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

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:

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

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 Ludenä, 2001).

Dans le cas particulier des données de battements de coeur, nous avons choisi de déterminer les changements éventuels en moyenne et variance (un autre choix, celui des changements de régularité, a été fait dans Bardet et Kammoun, 2007). Ainsi, pour tout  $j \in \{0, 1, \dots, K-1\}$ , on considère le modèle général:

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

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.

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,

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

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

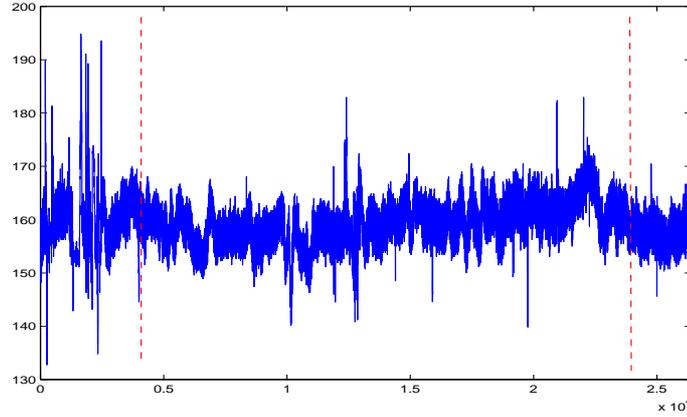
pénalisation  $\beta$ .

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

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

Ainsi, la valeur  $\widehat{K}$  retenu 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  $\widehat{K} = 3$ , ce que l'on pourrait qualifier de 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  $\widehat{K} = 2$ .



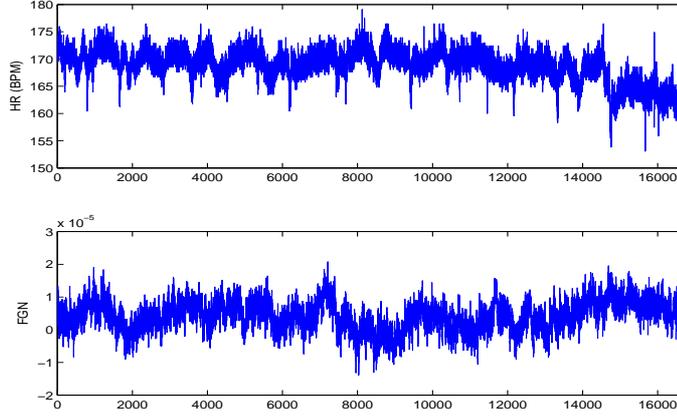
**Fig. 2.** Détermination des différentes phases de course pour un athlète

Lorsque l'on considère les données de fréquences cardiaques sur chacune des phases obtenues, on obtient à peu près un comportement stationnaire. Pour essayer de déterminer si un changement a lieu durant une de ses phases, nous allons nous intéresser à un paramètre de type fractal qui pourrait en être caractéristique.

### 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 une décroissance très lente 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). Ainsi, le bruit gaussien fractionnaire pourrait être un modèle plausible pour les données de fréquences cardiaques dans chaque phase de course.



**Fig. 3.** Comparaison entre le graphe des données de battements de coeur durant le milieu de course (Athlète 4) et celui d'un bruit gaussien fractionnaire simulé ( $H = 0.99$ )

Rappelons juste quelques éléments théoriques concernant le 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 (LM) ou fortement dépendant si:

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

Un cas particulier de processus LM 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 LM 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  $\forall c > 0$ :

$$(X(ct))_t \stackrel{\mathcal{L}}{=} c^H (X(t))_t,$$

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 LM gaussien. Plus précisément,  $Y^H = \{Y^H(k), k \in \mathbb{N}\}$  est un BGF, lorsque

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

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 FGN 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 nos données de battements de coeur peuvent être modélisées par un BGF, commençons par dire quelques mots sur les estimateurs de  $H$ . En premier lieu, disons que l'on trouvera un large éventails de tels estimateurs dans le livre de Doukhan *et al.* (2003). Il est bien clair que si l'on voulait estimer stricto sensu le paramètre  $H$  d'un BGF à partir d'une de ses trajectoires, on préférerait utiliser une méthode paramétrique, en particulier le méthode du contraste de Whittle (dont 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 Abry *et al.* (2003) et Bardet *et al.* (2000). Dans Wesfreid *et al.* (2005), une analyse multifractale des données de battements cardiaques utilisant le maximum des modules de la transformée en ondelettes a été appliquée pour la première fois.

On considère donc  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction 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 "shift" 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 shift  $b$ :

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

Lorsque l'on considère une série chronologique plutôt qu'un processus en temps continu, une somme de Riemann peut remplacer l'intégrale et on obtient ainsi le coefficient d'ondelette approché  $e_Z(a, b)$ . On suppose par ailleurs que la fonction  $\psi$  est telle qu'il existe  $M \in \mathbb{N}^*$  vérifiant,

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

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 LM ou autosimilaire. On obtient en particulier que:

- Si  $Z$  est un processus LM 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 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 lo-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  $H$ .

Dans un cadre semi-paramétrique général de processus LM stationnaires Gaussien ou linéaires, Moulines *et al.* (2007) et Roueff et Taqqu (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 faite aux processus multi-échelles dans Bardet et Bertrand, 2007).

Enfin, que ce soit pour les processus LM 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. Une manière de toujours annuler une telle tendance est de considérer une ondelette annulant tous ses moments, comme c'est le cas notamment si l'on choisit une ondelette de Lemarié-Meyer.

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

**Table 1.** Application du premier estimateur  $\widehat{H}_{OND}^{(1)}$  par la méthode d'analyse par ondelette sur les données de battements de coeur des différents athlètes

	<i>Phases</i>			
	Course entière	Début de course	Milieu de course	Fin de course
Ath1	0.8931	1.1268	1.1064	1.2773
Ath2	1.1174	0.7871	1.0916	0.8472
Ath3	1.0208	1.0315	1.1797	-
Ath4	0.9273	-	1.0407	0.7925
Ath5	1.0986	1.3110	1.0113	1.3952
Ath6	1.0769	1.5020	1.1597	1.3673
Ath7	1.0654	1.4237	1.1766	1.0151
Ath8	0.9568	1.6600	0.9699	1.1948
Ath9	0.9379	1.5791	0.9877	0.7263

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; notons que dans un tel cadre les échelles  $a_i$  doivent avoir une vitesse supérieures à  $N^{1/3}$ ).

Les physiologistes qui ont pour la première fois voulu estimer un paramètre de type Hurst pour des données de battements de coeur ont choisit d'utiliser une autre méthode: la méthode DFA (pour Detrended Fluctuation Analysis) introduite par Peng *et al.* (1994). Cette méthode est une version pour des séries à tendances de la méthode des variances agrégés. Plus précisément, elle consiste à agréger le processus (qui est supposé être de type LM + tendance) et le diviser dans des fenêtres de tailles différentes, puis de détendancionaliser ce processus à partir de régressions linéaires dans chaque fenêtre, de calculer alors la variance des résidus, et d'estimer  $H$  à partir d'une log-log-régression des différentes variances résiduelles obtenues pour différentes tailles de fenêtre.

Disons le tout net: cette méthode est un non-sens statistique. Si, pour un processus gaussien stationnaire LM, la vitesse de convergence de  $\widehat{H}_{DFA}$  (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 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 détendancionalisation 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  $\widehat{H}_{DFA}$  converge vers tout à fait autre chose que  $H$ ... (voir encore Bardet et Kammoun, 2008).

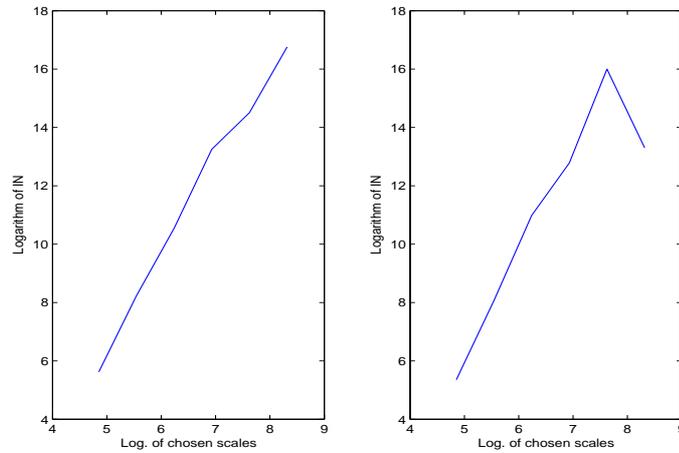
On présente dans la Table 1 les différents calculs de  $\widehat{H}_{OND}^{(1)}$  (on a choisit ici une ondelette mère de Daubechie d'ordre 4 et des échelles de longueur  $a_i = i \cdot N^{0.4}$  avec  $i = 1, \dots, 6$ ) sur nos données de battements de coeur pour les 9 athlètes, pour la série entière des données d'un athlète, puis pour chacune de ses phases de course.

Deux problèmes importants découlent de ces résultats.

- (a) En premier lieu, les estimations de  $\widehat{H}_{OND}^{(1)}$  sont souvent plus grandes que 1. Pourtant un BGF ne peut seulement être défini que pour  $H \in (0, 1)$ . 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 soit être remise en cause:
- l'hypothèse que le processus est stationnaire;
  - l'hypothèse que le processus est gaussien;
  - l'hypothèse que seulement deux paramètres ( $H$  et  $\sigma^2$ ) suffisent à définir le processus.

Dans la suite, un nouveau modèle sera proposé et nous ferons le choix de conservé les deux premières hypothèses mais de remplacer la troisième par une hypothèse semi-paramétrique.

- (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  $a_i \leq a_m$ , mais pas suffisamment pour  $a_i \geq a_m$ . Ainsi nos données de battements de coeur se comportent comme un BGF pour des "petites" échelles (ou des hautes fréquences) mais pas pour des "grandes échelles" (ou des petites fréquences)

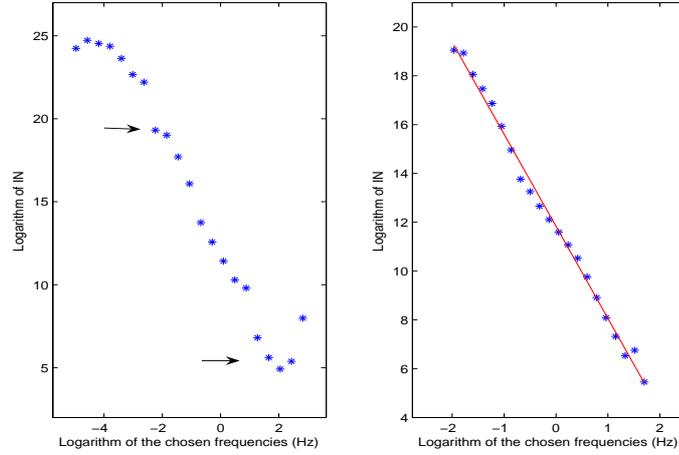


**Fig. 4.** The log-log graph of the variance of wavelet coefficients relating to the HR series observed during all the race and in the end of race (Ath2)

**Remarque:** Cette dernière conclusion obtenue à partir de l'examen de la Figure 4 confère un autre avantage de la méthode d'analyse par ondelette. En effet, cette figure permet de contrôler visuellement ce que l'on fait et de pouvoir questionner le modèle choisi. Par exemple, dans Bardet et Bertrand (2007), du fait de la présence de 2 zones de linéarité de ce graphe  $(\log a_i, \log S_N(a_i))_{1 \leq i \leq \ell}$ , on a développé un modèle de type multiéchelle pour être en meilleure adéquation avec des données biomécanique.

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

Le modèle que nous avons choisi pour modéliser les fréquences cardiaques s'inspire des résultats déjà obtenu dans Bardet et Bertrand (2007). Contrairement à ce qui a été fait précédemment nous allons désormais travailler avec des fréquences plutôt que des échelles (avec la simple définition que l'un est l'inverse de l'autre). Si on reprend, le même type de graphe que celui de la Figure 4 mais en fréquence (ce qui inverse le sens de la courbe) et avec beaucoup plus de fréquences choisies, on obtient la Figure 5 qui suit:



**Fig. 5.** Le log-log graphe des variances empiriques des coefficients d'ondelettes pour les battements de coeur de l'Athlète 6 lors de la phase d'arrivée en choisissant une bande de fréquence [0.01 12] (à gauche) et de [0.2 4] (à droite).

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 définir une généralisation  $X_\rho = \{X_\rho(t), t \in \mathbb{R}\}$  de ce processus que nous appellerons "mouvement brownien localement fractionnaire" 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 que:

- $\rho(\xi) = \frac{1}{\sigma} |\xi|^{H+1/2}$  pour  $|\xi| \in [\omega_0, \omega_1]$  avec  $H \in \mathbb{R}$ ,  $\sigma > 0$  and  $0 < \omega_0 < \omega_1 < \infty$ ;
- $\int_{\mathbb{R}} (1 \wedge |\xi|^2) \frac{1}{\rho^2(\xi)} d\xi < \infty$ .

La preuve de l'existence d'un tel processus, dont les accroissements sont stationnaires, peut se trouver dans Cramér et Leadbetter (1967). Les principaux avantages d'une telle généralisation du MBF sont les suivantes:

- (a) C'est un modèle semi-paramétrique permettant beaucoup plus de souplesse en modélisation;
- (b)  $X_\rho$  se comporte comme un MBF seulement dans une bande de fréquence;
- (c) Dans cette bande, le paramètre  $H$ , que l'on appellera "paramètre de fractalité locale" n'est pas restreint à être dans  $[0, 1]$  mais il peut prendre n'importe quelle valeur de  $\mathbb{R}$ .

Aussi, à partir d'une telle, peut-on déduire le modèle suivant pour les données de fréquences cardiaques:

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

Remarquons que  $Y_\rho = \{Y_\rho(t), t \in \mathbb{R}\}$  est un processus gaussien stationnaire et que la fonction  $(2 \sin(\xi/2) \rho^{-1}(\xi))^2$  est aussi appelée la densité spectrale de  $Y_\rho$ .

Notre but est maintenant d'obtenir une estimation du paramètre de fractalité locale  $H$  à partir d'une trajectoire

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

connue où  $\Delta_N \rightarrow 0$  et  $N\Delta_N \rightarrow \infty$  quand  $N \rightarrow \infty$ . Il est bien clair que cela signifie également qu'une trajectoire  $(X_\rho(\Delta_N), \dots, X_\rho(N\Delta_N))$  est observée. La particularité de l'analyse par ondelette proposée est de considérer une ondelette mère vérifiant les propriétés suivantes:  $\psi : \mathbb{R} \mapsto \mathbb{R}$  est une fonction  $\mathcal{C}^\infty$  telle que:

- 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  $0 < \alpha < \beta$ .

Alors, en utilisant les résultats de Bardet et Bertrand (2007), pour tout  $a > 0$  tel que  $[\frac{\alpha}{a}, \frac{\beta}{a}] \subset [\omega_0, \omega_1]$ , i.e.  $a \in [\frac{\beta}{\omega_1}, \frac{\alpha}{\omega_0}]$ ,  $(d_{X_\rho}(a, b))_{b \in \mathbb{R}}$  est un processus stationnaire gaussien et

$$\mathbb{E}(d_{X_\rho}^2(a, \cdot)) = K(\psi, H, \sigma) \cdot a^{2H+1},$$

avec  $K(\psi, H, \sigma) > 0$  ne dépendant que de  $\psi, H$  et  $\sigma$ . Remarquons qu'une telle propriété ne peut être vérifiée que si le support de  $\psi$  est tel que:

$$\frac{\beta}{\alpha} < \frac{\omega_1}{\omega_0}.$$

**Table 2.** Valeurs de  $H$  estimées, avec la méthode DFA et la méthode d’analyse par ondelettes sur la bande de fréquences  $[0.2, 4]$ . On note par (\*) les séries pour lesquelles le test d’adéquation est rejeté. Enfin, les p-values en bas du tableau proviennent d’un test de comparaison des deux échantillons  $(\hat{H}_{DFA})_{1,\dots,9}$  and  $(\hat{H}_{OND}^{(2)})_{1,\dots,9}$  pour les différentes séries

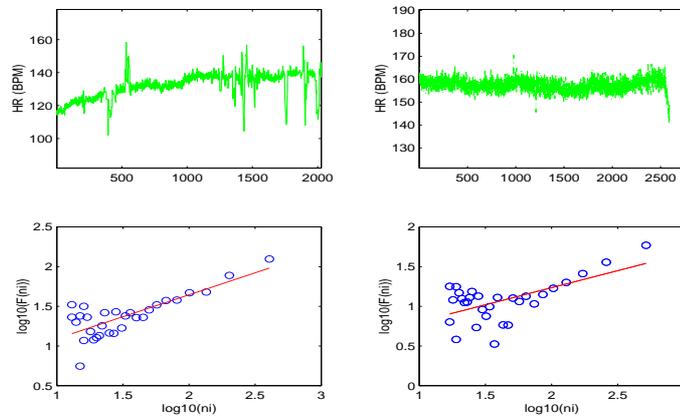
	Course entière		Début de course		Milieu de course		Fin de course	
	$\hat{H}_{DFA}$	$\hat{H}_{OND}^{(2)}$	$\hat{H}_{DFA}$	$\hat{H}_{OND}^{(2)}$	$\hat{H}_{DFA}$	$\hat{H}_{OND}^{(2)}$	$\hat{H}_{DFA}$	$\hat{H}_{OND}^{(2)}$
Ath1	0.928	1.288*	1.032	1.192	1.060	1.214	0.429	1.400
Ath2	1.095	1.268*	0.905	0.973	1.126	1.108	1.240	1.452*
Ath3	1.163	1.048	0.553	0.898	1.130	1.172	-	-
Ath4	1.193	0.916*	-	-	1.098	1.249*	1.172	1.260
Ath5	1.239	1.110	1.267	1.117*	1.133	1.205	1.273	1.348*
Ath6	1.247	1.084*	1.237	1.106	1.091	1.172	1.436	1.338
Ath7	1.155	1.095	0.850	1.295	1.182	1.186*	1.129	1.209
Ath8	1.258	1.011	1.304	1.128*	0.995	1.134	1.122	1.247
Ath9	1.243	1.429*	0.820	1.019	1.127	1.535*	1.250	1.238*
<i>p-value</i>	0.6414		0.3723		0.0225		0.1260	
<i>F-stat</i>	0.23		0.85		6.38		2.65	

De plus, pour  $a \in [\frac{\beta}{\omega_1}, \frac{\alpha}{\omega_0}]$ , la variance empirique  $S_N(a)$  définie par (3) et calculée à partir de  $(X_\rho(\Delta_N), \dots, X_\rho(N\Delta_N))$  converge vers  $\mathbb{E}(d_{X_\rho}^2(a, \cdot))$  suivant un théorème de la limite centrale avec une vitesse de convergence en  $\sqrt{N\Delta_N}$ . Ainsi lorsque  $(a_1, \dots, a_\ell) \in [\frac{\beta}{\omega_1}, \frac{\alpha}{\omega_0}]^\ell$ , une log-log-régression des  $(a_i, S_N(a_i))_{1 \leq i \leq \ell}$  fournit une estimation ode  $H$  (et l’estimateur ainsi construit,  $\hat{H}_{OND}^{(2)}$  suit un théorème de la limite centrale avec une vitesse de convergence en  $\sqrt{N\Delta_N}$ ). De plus, comme précédemment, un test d’adéquation du Chi-deux peut-être construit à partir d’une distance entre les points  $(\log(a_i), \log(S_N(a_i)))_{1 \leq i \leq \ell}$  et la droite de régression par moindres carrés pseudo-généralisés.

La principale difficulté pour un définir de tels estimateur et test est de déterminer la “bonne” zone de fréquence  $[\omega_0, \omega_1]$  (car  $\omega_0$  et  $\omega_1$  sont concrètement a priori inconnus). La solution la plus simple, telle celle que retrace la Figure 5 consiste à choisir une bande de fréquence très large puis de sélectionner graphiquement la meilleure zone de linéarité du nuage de points  $(\log(a_i), \log(S_N(a_i)))_{1 \leq i \leq \ell}$ . Ceci pourrait être fait pour les 9 coureurs et pour les différentes phases de course. On peut ainsi obtenir une bande de fréquences commune à l’ensemble de ces graphes: c’est cette démarche que nous avons choisie ici, du fait que l’on peut espérer trouver en cette bande de fréquence une sorte de constante humaine. Cependant, une méthode plus “statistique” pourrait être également mise en place à partir d’un estimateur adaptatif (pour chaque série de données) de cette meilleure zone de linéarité, comme cela avait déjà été entrepris dans Bardet et Bertrand (2007). Nous avons donc appliqué la méthode évoquée en utilisant une ondelette de Lemarié-Meyer “dilatée” (qui vérifie les propriétés évoquées ci-dessus) telle que  $\beta = 2\alpha = 4\pi/3$ , d’abord dans la bande de fréquence  $[0.01, 12]$  (voir Figure 5). On en déduit ensuite une zone commune  $[\bar{\omega}_0, \bar{\omega}_1] = [0.2, 4]$  pour séries entières et partielles de données. Ceci permet d’obtenir les estimations de  $H$  par les deux méthodes, soit  $\hat{H}_{DFA}$  et  $\hat{H}_{OND}^{(2)}$  obtenu par l’analyse par ondelettes dans la bande de fréquences choisie. Les résultats sont récapitulés dans la Table 2.

**Commentaires des résultats obtenus:**

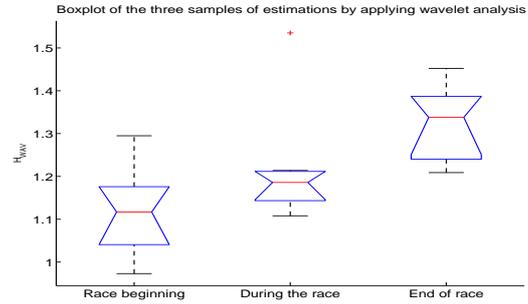
- (a) Les différentes estimations sont en général comprises entre 1 et 1.5. Seule la méthode DFA conduit parfois à des valeurs différentes. C'est le cas en particulier du début de course de l'Athlète 3 ou de la fin de course de l'Athlète 1. La Figure 6) examine ces deux séries, et l'on peut observer que les séries semblent entachées d'une tendance (un polynôme de degré 2 serait approprié). Or, comme cela a été montré théoriquement et numériquement dans Bardet et Kammoun (2008), la méthode DFA n'est pas du tout robuste dans un tel cadre. Ceci confirme que la méthode DFA peut conduire à des résultats non fiables.



**Fig. 6.** Les résultats de la méthode DFA appliquée au début de course de l'Athlète 3 (gauche) ou de la fin de course de l'Athlète 1 (droite)

- (b) En second lieu, comparons les résultats entre les deux méthodes. Si l'on considère les données relevées sur l'ensemble de la course, le test de comparaison ne marque pas de différences significatives entre les échantillons  $(\hat{H}_{DFA})_{1,\dots,9}$  et  $(\hat{H}_{OND}^{(2)})_{1,\dots,9}$  (on obtient une p-value de 0.64). En revanche, la même comparaison sur la phase de milieu de course (qui est la part en temps la plus importante de la course) n'aboutit pas au même résultat: il y a clairement une différence significative entre  $(\hat{H}_{DFA})_{1,\dots,9}$  et  $(\hat{H}_{OND}^{(2)})_{1,\dots,9}$  sur cette portion de course (on obtient une p-value de 0.02). En revanche, si les estimations sont encore assez (mais non significativement) différentes en fin de courses, elles sont proches en début de course.
- (c) On peut aller plus avant encore dans l'étude des valeurs des différentes estimations de  $H$ . En effet, la partie droite de la Figure 7) trace les différents "boxplots" obtenus pour les valeurs de  $(\hat{H}_{OND}^{(2)})_{1,\dots,9}$  pour les 3 phases de course. Il semble apparaître une augmentation significative de la valeur de  $\hat{H}_{OND}^{(2)}$  au fil du temps de la course. Une analyse de la variance (voir la partie gauche de la Figure 7) confirme qu'il y a une évolution de  $\hat{H}_{OND}^{(2)}$ : la p-value est de 0.0158. Cette évolution n'apparaît cependant pas si on s'en tient à la méthode DFA puisque dans ce cas la p-value est de 0.85!

	$\hat{H}_{DFA}$	$\hat{H}_{OND}^{(2)}$
<i>p-value</i>	0.8570	0.0158
<i>F-stat</i>	0.16	5.27



**Fig. 7.** Test de comparaison des 3 échantillons  $(\hat{H}_{DFA})_{1,\dots,9}$  et  $(\hat{H}_{OND}^{(2)})_{1,\dots,9}$  (à gauche) et boxplots des 3 échantillons  $(\hat{H}_{OND}^{(2)})_{1,\dots,9}$  relatifs au début, milieu et fin de course (à droite)

- (d) Enfin, vis-à-vis des résultats obtenus par le test d'adéquation, on remarquera que celui est bien plus souvent accepté lorsqu'on l'applique sur les différentes phases de course plutôt que sur la course entière. Ainsi, l'intérêt de ce découpage et la qualité de la modélisation par un bruit gaussien localement fractionnaire sont confirmés par ces tests (on voit aussi l'amélioration qu'apporte  $\hat{H}_{OND}^{(2)}$  par rapport à  $\hat{H}_{OND}^{(1)}$  en revenant sur les résultats de la Table 1 pour laquelle, rappelons-le, le test n'était jamais accepté).

## 5. Conclusion

Voici les différents points à retenir de notre étude:

- En premier lieu, il ne faut pas utiliser la méthode DFA pour estimer un paramètre de type fractal. Cette méthode est à bannir!
- 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.
- La modélisation par un bruit gaussien localement fractionnaire nous semble très fructueuse dans le cas présent des fréquences cardiaques, mais sans doute aussi dans de nombreux autres. La possibilité de définir le paramètre de fractalité locale dans  $\mathbb{R}$  doit s'avérer très intéressante dans de nombreux cas concrets.
- Liée intimement 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.
- Enfin, vis-à-vis du fonctionnement du coeur en situation de course d'endurance, on montre que le paramètre de fractalité locale 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  $\hat{H}_{dbut} \simeq 1.1$ ,  $\hat{H}_{milieu} \simeq 1.2$  et  $\hat{H}_{fin} \simeq 1.3$  et l'on vérifie une telle hypothèse par un test de Fisher (dont la p-value est de 0.65). 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 (malheureusement avec la méthode DFA! il faudrait voir également ce que donne la méthode d'analyse par ondelettes...) 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.

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 de fractalité locale 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 paramètres 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 Génompol 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. *J. Physica A: Statistical mechanics and its applications*, 235-244.
- [4] Bagger, M., Petersen, P.H. et Pedersen, P.K. (2003) Biological Variation in variables associated with exercise training. *Int. J. Sports Med*, 433-440.
- [5] Bai, J. (1998) Least squares estimation of a shift in linear processes. *J. of Time Series Anal.*, 5, 453-472.
- [6] Bai, J. et Perron, P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, 66, 47-78.
- [7] Bardet, J.M. (2002) Statistical study of the wavelet analysis of fractional Brownian motion. *IEEE Trans. Inform. Theory.*, 48, 991-999.
- [8] Bardet, J.M., Bertrand, P. (2007) Identification of the multiscale fractional Brownian motion with biomechanical applications. *Journal of Time Series Analysis*, 28, 1-52.
- [9] 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.
- [10] Bardet, J.M. et Kammoun, I. (2007) Detecting changes in the fluctuations of a Gaussian process and an application to heartbeat time series. Prépublication HAL.
- [11] 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.

- [12] 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.
- [13] Bardet, J.M. et Tudor, C. (2008) Hurst index estimation for the Rosenblatt process: wavelets and chaos expansion. Preprint HAL.
- [14] Billat, V., Wesfreid, E., Cottin, F., Kapfer, C. et Koralztein, J.P. (2003) Fractal analysis of speed and physiological oscillations in long- and middle-distance running: Effect of training. *International Journal of Computer Science in Sport*, 16-30. <http://www.billat.net/>.
- [15] Cramér, H. et Leadbetter, M.R. (1967) Stationary and related stochastic processes. Sample function properties and their applications. Wiley and Sons.
- [16] Flandrin, P. (1992) Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. on Inform. Theory*, 38, 910-917.
- [17] Fox, R. and Taqqu, M.S. (1986) Large-sample properties of parameter estimates for strongly dependent Gaussian time series. *Ann. Statist.*, 14, 517-532.
- [18] Goldberger, A.L. (2001) Heartbeats, hormones, and health : is variability the spice of life? *Am J Respir Crit Care Med*, 163, 1289-1290.
- [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] Ivanov, P.C., Amaral, L.A.N., Goldberger, A.L., Havlin, S., Rosenblum, M., Struzik, Z. et Stanley, H.E. (1999). Multifractality in human heartbeat dynamics. *Nature*, 399, 461-465.
- [21] 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/>.
- [22] 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.
- [23] Lavielle M. et Ludenä, C. (2000) The multiple change-points problem for the spectral distribution. *Bernoulli* 6, 845-869.
- [24] 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.
- [25] Moulines, E., Roueff, F. et Taqqu, M.S. (2008), On the spectral density of the wavelet coefficients of long memory time series with application to the log-regression estimation of the memory parameter. Preprint ArXiv.
- [26] 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.

- [27] Peng, C.K., Mietus, J., Hausdorff, J., Havlin, S., Stanley H.E. and Goldberger A.L. Long-Range Anticorrelations and Non-Gaussian Behavior of the Heartbeat. *Phys. Rev. Lett.*, 70, 1343-1346.
- [28] 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.
- [29] 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.
- [30] Pikkujämsä, 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.
- [31] Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence, *The Annals of statistics*, 23, 1630–1661.
- [32] Roueff, F. et Taqqu, M.S. (2008). Asymptotic normality of wavelet estimators of the memory parameter for linear processes. Preprint arXiv:0805.0778.
- [33] Veitch, D., Abry, P. et Taqqu, M.S. (2003). On the Automatic Selection of the Onset of Scaling, *Fractals*, 11, 377–390.
- [34] Wesfreid, E., Billat, V. et Meyer, Y. (2005) Multifractal analysis of heartbeat time series in human races. *Appl. Comput. Harmon. Anal.*, 18, 329-335.
- [35] 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.