

# Statistique, physiologie et mesures de stress

Pierre Raphaël BERTRAND\*

Laboratoire de Mathématiques,  
UMR 6620  
CNRS et Université Blaise Pascal  
(Université Clermont-Ferrand 2)  
FRANCE.

## Résumé

Nous présentons des applications de la statistique des séries temporelles et de la statistique des processus à différents signaux physiologiques. On rappelle d'abord l'utilisation de modèles fractals pour la position du centre de gravité d'un individu humain mesurée en laboratoire. Ce signal est bien modélisé par un mouvement brownien fractionnaire multi-échelle avec différents indices de Hurst selon les bandes de fréquence. On montre ensuite la transposition de ce type de modèle au signal rythme cardiaque pour un individu hors laboratoire (marathonien ou travail posté). Sortir du laboratoire pose un nouveau problème, celui de la segmentation du signal pour suivre la variation des conditions environnementales. Enfin, nous présentons les applications actuellement en cours pour les mesures de stress (ou bien-être) pour des sujets relevant du spectre de l'autisme avec des signaux physiologiques multivariés.

## Abstract :

We present some applications of time series analysis and statistics of stochastic processes to physiological signals. We first recall the use of fractal models for the position of the center of gravity of a human being measured inside the laboratory. This signal is modelled by a multiscale fractional Brownian motion with different Hurst indices according the different frequency bands. We then give the transposition of this model to a signal for human beings outside laboratory (heart rate of Marathon runner or shift worker). Going outside the laboratory (or hospital) addresses new questions, namely the segmentation of the signal to follow the variation of environmental conditions. Finally, we present a work in progress for measuring the stress (or wellbeing) of a subject on the autism spectrum, based on multivariate physiological signals.

---

\*Travaux effectués avec les soutiens de l'ANR-12-BS01-0016-01 projet "Do Well B." (2013-2016), de Digitéo projet "PHYSIOSTAT" (2009-2011) et de l'INRIA (2008-2011).

## Introduction

Cet article présente diverses applications de la statistique des séries temporelles en physiologie et santé. Je me restreindrai à des travaux auxquels j'ai collaboré et qui ont majoritairement été effectués avec des équipes clermontoises. L'histoire est un peu anecdotique mais elle illustre à la fois l'efficacité des mathématiques dans les applications et les opportunités de collaborations pluridisciplinaires qu'offrent des universités de taille moyenne comme l'Université Blaise Pascal (Clermont-Ferrand 2) et l'Université d'Auvergne (Clermont-Ferrand 1).

Spécialiste des applications des probabilités et des statistiques aux données réelles, j'ai été recruté à Clermont-Ferrand en 1996. Durant les années 1990, la principale applications des modèles stochastiques était la *finance mathématique*. Je travaillais donc sur la modélisation des séries financières, cours d'action, taux d'intérêt, et divers produits financiers sophistiqués. Mais, j'étais venu au laboratoire de mathématiques de Clermont-Ferrand pour collaborer sur toute application des modèles probabilistes à des données réelles, je pensais renforcer la collaboration en mécanique aléatoire existant entre le laboratoire de mathématique (UMR 6620) et l'INSIS (Institut des Sciences de l'Ingénierie et des Systèmes, UMR 6602). Le hasard décida autrement et je fus amené à collaborer avec des spécialistes de physiologie, médecine du travail, puis psychologie.

Le plan de ce chapitre suit cette évolution chronologique. Dans la première section, je présenterai les applications pour l'équilibre humain en biomécanique. Dans la seconde section, je présenterai des applications au rythme cardiaque en physiologie de l'exercice (sport) ou médecine du travail. Enfin, la dernière section présentera les applications actuellement en cours pour les mesures de stress (ou bien-être), dont l'autisme, avec des signaux physiologiques multivariés. Ces applications à l'autisme intègrent les travaux précédents, en y ajoutant de nouveaux outils statistiques.

## 1 Biomécanique et régulation de l'équilibre humain

Le début relève du hasard : j'avais été contacté en 1998 par deux collègues de l'UFR STAPS de Clermont-Ferrand effectuant leur recherche au LIMOS (UMR 6158 CNRS) pour analyser la régulation de l'équilibre humain à partir de données mesurées sur des étudiants. Il s'agissait de reproduire et d'améliorer le traitement statistique proposé par Collins et de Lucas, professeurs à Boston University, Department of Biomedical Engineering (USA).

### 1.1 Description des données

À l'aide d'une plateforme de force on mesure la position du centre de pression qui correspond à la projection du centre de gravité d'un individu à l'équilibre en position debout "*immobile*", pendant une minute à la fréquence 100 Hz. Typiquement, pour un sujet adulte (18 – 60 ans),

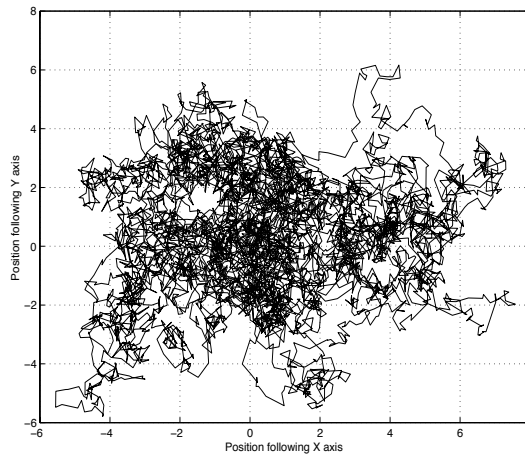
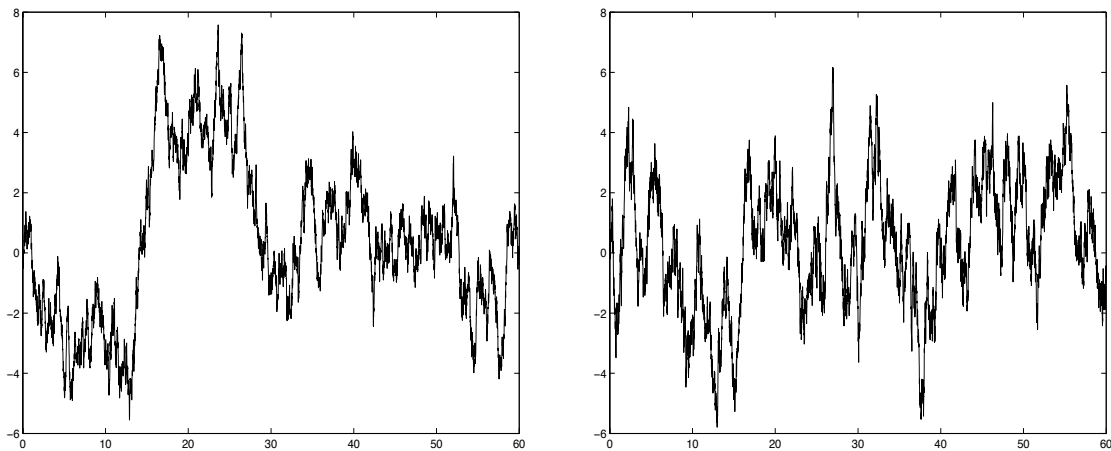


FIGURE 1 – Trajectoire du centre de gravité

FIGURE 2 – Trajectoire  $X(t)$  et  $Y(t)$ 

on obtient une trajectoire du centre de pression  $(X(t), Y(t))$  représentée sur la figure 1, où les unités sont en millimètre, l'axe des X représente la direction avant/arrière, tandis que l'axe des Y représente la direction latérale. Quand on omet l'aspect temporel, on obtient un nuage de point qui peut être décrit par une ellipse de confiance contenant 90% des points. Diverses études ont montré que la surface de cette ellipse est minimale pour des adultes (18 – 60 ans), tandis qu'elle est plus importante pour des enfants de 5–6 ans puis recommence à croître chez les personnes âgées. L'interprétation standard est que l'équilibre serait moins bien contrôlé par les jeunes enfants ou les personnes âgées.

On peut également représenter les deux trajectoires  $X(t)$  et  $Y(t)$  en fonction du temps  $t$  en seconde. On obtient ainsi deux séries temporelles illustrées par la figure 2.

## 1.2 Modélisation mathématique

Nous décrivons ci-dessous la problématique de la modélisation des séries physiologiques en s'appuyant sur l'exemple de l'équilibre humain en biomécanique.

### Problématique générale

La problématique générale de ce chapitre est la suivante : on observe un phénomène physiologique dépendant du temps et aléatoire, comme sur la figure 2. Le caractère aléatoire se traduit par une variation inter-sujet, mais aussi par une fluctuation au cours du temps. On le décrit donc par un modèle probabiliste. La dépendance temporelle implique la modélisation par une série temporelle (c.à-d. dépendant d'un temps discontinu), ou par un processus stochastique (c.à-d. dépendant d'un temps continu, bien qu'observé à des instants discontinus). On souhaite ensuite estimer les paramètres du modèle, ce qui nécessite des techniques de statistique des séries temporelles ou de statistique des processus stochastiques. Les paramètres du modèle sont supposés correspondre à l'état physiologique de l'individu tandis que leurs variations sont censées refléter les variations de l'état du sujet.

### L'exemple de la biomécanique ...

Collins et de Lucas [10, 11] ont initié la modélisation des mouvements du centre de gravité par un mouvement brownien fractionnaire (MBF) avec deux régimes différents selon les échelles de temps. À haute fréquence, il s'agissait d'un MBF avec un indice de Hurst  $H > 0.5$  donc persistant et n'assurant pas l'équilibre, tandis qu'à basse fréquence il s'agissait d'un MBF avec un indice de Hurst  $H < 0.5$  donc anti-persistant et assurant l'équilibre. L'idée physiologique sous-jacente était que différents systèmes sensoriels et neuro-musculaires (oreille interne, vision, muscles, plante des pieds) interfèrent dans l'équilibre humain et que ces différents systèmes correspondent à différentes bandes de fréquence.

### ... et la nécessité d'adapter les modèles mathématiques

Les spécialistes des diverses applications utilisent divers modèles mathématiques pour décrire leurs données et améliorer la compréhension des phénomènes. Ce travail de modélisation crée un besoin de collaboration avec les mathématiciens, en particulier pour la mise au point des estimateurs statistiques. Dans la section suivante, nous présentons un rappel sur le mouvement brownien fractionnaire et ses différentes propriétés probabilistes. Ceci montre que le MBF ne peut pas correspondre à l'intuition physiologique sous-jacente sans une adaptation du modèle mathématique.

### 1.3 Mouvement brownien fractionnaire

Le mouvement brownien fractionnaire (MBF) a été introduit de manière abstraite par Kolmogorov (1940) comme "*spirale de Wiener dans un espace de Hilbert*" [20], puis popularisé à partir des années 1960 par Mandelbrot pour modéliser les cours du coton, les crues du Nil et diverses applications, cf. [21]. Le lecteur curieux pourra consulter les notes historiques [24, Chap. 14] complétées par l'autobiographie de Mandelbrot [22].

A un facteur multiplicatif  $\sigma^2$  prêt, le MBF est entièrement caractérisé par un unique paramètre  $H \in ]0, 1[$ , appelé indice de Hurst en hommage à l'hydrologue qui étudia les crues du Nil [18] et le MBF est noté  $B_H$ . Il existe plusieurs définitions et/ou représentations du MBF. On peut définir le MBF

- soit comme un processus gaussien centré de structure de covariance

$$\text{cov}[B_H(s), B_H(t)] = \frac{\sigma^2}{2} \left\{ |s|^{2H} + |t|^{2H} - |t - s|^{2H} \right\} \quad \text{pour tout } (s, t) \in \mathbb{R}^2, \quad (1)$$

où

$$\sigma^2 = \text{Var}[B_H(1)] = \mathbb{E}[B_H(1)^2]$$

est un facteur d'échelle ;

- soit comme une moyenne mobile cf. [21] ;
- soit par sa représentation en série aléatoire d'ondelettes [23] ;
- soit par sa représentation harmonisable (ou spectrale) [20] :

$$B_H(t) = \int_{\mathbb{R}} \frac{(e^{it\xi} - 1)}{|\xi|^{H+1/2}} dW(\xi), \quad \text{pour tout } t \in \mathbb{R}, \quad (2)$$

L'intégrale (2) requiert une définition mathématique élaborée :

- $dW(\xi)$  est une mesure de Wiener complexe tel que le processus  $B_H(t)$  soit à valeurs réelles cf. [24],
- l'intégrale  $\int_{\mathbb{R}} f(\xi) dW(\xi)$  est une variable aléatoire gaussienne centrée définie pour toute fonction  $f \in \mathbb{L}^2(\mathbb{R}, \mathbb{C})$ ,

La représentation (2) est donnée à une constante multiplicative près, elle est définie sous la condition d'intégrabilité

$$\int_{\mathbb{R}} \left| \frac{(e^{it\xi} - 1)}{|\xi|^{H+1/2}} \right|^2 d\xi < \infty$$

qui est équivalente à  $H \in ]0, 1[$ .

On montre que toutes ces définitions correspondent au même objet : en langage mathématique, on a égalité en distribution des différentes représentations (à une constante multiplicative prêt). La figure 3 donne deux simulations numériques de trajectoires de MBF correspondant aux deux

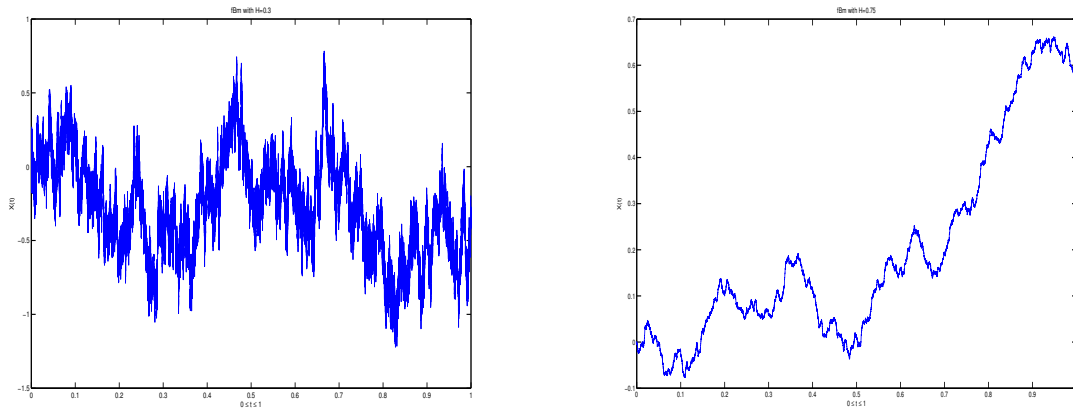


FIGURE 3 – Simulation de trajectoires d'un FBM : **a)** indice de Hurst  $H = 0,3$ ; **b)** indice de Hurst  $H = 0,75$ .

indices de Hurst  $H = 0.3$  ou  $H = 0.75$ . Quand  $H = 1/2$ , le MBF correspond au mouvement brownien (aussi appelé processus de Wiener). De ce fait, le MBF peut être considéré comme une généralisation du mouvement brownien. L'indice de Hurst  $H$  pilote plusieurs propriétés importantes du MBF :

- La **rugosité** ou **régularité höldérienne des trajectoires correspond à l'indice de Hurst  $H$** . Grosso modo, pour un accroissement de temps  $\epsilon$ , on a

$$B_H(t + \epsilon) - B_H(t) \simeq Cte \times \epsilon^H, \quad (3)$$

donc plus l'indice de Hurst  $H$  est proche de 0 plus la trajectoire est irrégulière, et quand  $H$  s'approche de 1, la trajectoire devient "*presque*" continument différentiable. La variation de la régularité des trajectoires en fonction de l'indice de Hurst  $H$  est illustrée par la figure3. En toute rigueur la formule (3) doit être remplacée par la loi du logarithme itéré :

$$\limsup_{\epsilon \rightarrow 0^+} \frac{|B_H(t + \epsilon) - B_H(t)|}{\sqrt{2\epsilon^{2H} \ln(\ln(1/\epsilon))}} = 1 \text{ presque sûrement,} \quad (4)$$

pour tout  $t \in \mathbb{R}$ . Au facteur  $\sqrt{\ln(\ln(1/\epsilon))}$  prêt, on retrouve la formule (3).

- **La corrélation des accroissements dépend de l'indice de Hurst  $H$ , ainsi que leur persistance ou anti-persistance :**

Le MBF est un processus à accroissements stationnaires. Pour un pas de temps  $\Delta t$  et un décalage  $j \times \Delta t$  avec  $j$  nombre entier, la formule (1) permet de calculer la corrélation des

accroissements :

$$\begin{aligned} r(j) &:= \text{corr} \left[ B_H((j+1)\Delta t) - B_H(j\Delta t), B_H(\Delta t) - B_H(0) \right] \\ &= \frac{\sigma^2}{2} \times \{|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}\}. \end{aligned} \quad (5)$$

Si  $H > 1/2$ ,  $r(j) > 0$  pour tout entier  $j \in \mathbb{N}$ , les accroissements sont positivement corrélés, donc le processus est persistant. Si  $H < 1/2$ ,  $r(j) < 0$  pour tout entier  $j \in \mathbb{N}$ , les accroissements sont négativement corrélés, donc le processus est anti-persistant. Si  $H = 1/2$ ,  $r(j) = 0$  pour tout entier  $j \in \mathbb{N}$ , le MBF est un mouvement brownien et ses accroissements sont alors indépendants. Si  $H > 1/2$ , on retrouve l'*effet Joseph*<sup>1</sup> décrit par Mandelbrot et illustré par la figure 3 b), tandis que la figure 3 a) montre des oscillations beaucoup moins longues autour de la moyenne.

- Le MBF est **auto-similaire d'ordre  $H$** , c.à-d. pour tout réel  $\lambda > 0$ , on a

$$\{X(\lambda t), t \in \mathbb{R}\} \stackrel{\text{loi}}{\sim} \{\lambda^H X(t), t \in \mathbb{R}\}. \quad (6)$$

Ceci signifie qu'à toutes les échelles de temps, le MBF a la même structure statistique, ou que si on zoome d'un facteur  $\lambda$ , on voit la même allure de courbe (à un facteur multiplicatif  $\lambda^H$  prêt).

- Le MBF admet une **densité spectrale en loi de puissance en  $-(2H + 1)$** , c.à-d.

$$f(\xi) = \sigma^2 \times |\xi|^{-(2H+1)},$$

qui correspond à une fonction affine en coordonnées log-log, c.à-d.

$$\log f(\xi) = -(2H + 1) \log |\xi| + Cte.$$

On peut alors estimer l'indice de Hurst par régression linéaire de la densité spectrale estimée à différentes fréquences.

## 1.4 Mouvement brownien fractionnaire multi-échelle pour l'équilibre humain

Le MBF se caractérise par son invariance d'échelle (ou autosimilarité). Ceci revient à dire que pour tout pas de temps  $\Delta t$  auquel on observe un MBF, on a le même indice de Hurst, donc

---

1. Il s'agit d'une référence à l'ancien testament : Pharaon faisait le rêve récurrent de sept vaches grasses suivies de sept vaches maigres. Joseph, alors premier ministre, interpréta le rêve comme sept années de bonnes récoltes suivies de sept mauvaises années et proposa de construire des greniers. Mandelbrot expliqua que les récoltes dans l'ancienne Egypte étaient directement liées au niveau des crues du Nil, fit le lien avec les travaux de l'hydrologue Hurst qui avait montré que la série des crues du Nil correspondait à un indice  $H = 0,75$  et retrouva ainsi de longs cycles de bonnes années suivies de mauvaises. Puis il appliqua le même modèle aux prix du coton dans la vallée du Mississippi.

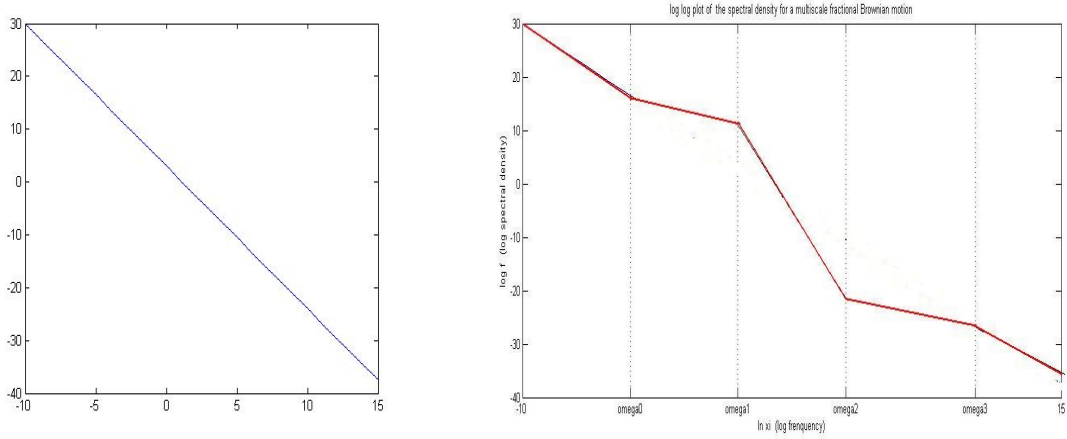


FIGURE 4 – Densité spectrale (coordonnées log-log) : a) pour un MBF ; b) pour un MBF multi-échelle.

la même structure de corrélation des accroissement donnée par (5), puis la même propriété de persistance ou d'anti-persistance. Un tel modèle ne permet pas de rendre compte de l'interprétation sous-jacente en biomécanique d'indices de Hurst différents selon les pas de temps [10, 11]. Pour pallier ce défaut, dans [3, 4] nous avons relaxé l'hypothèse d'autosimilarité et introduit le mouvement brownien fractionnaire multi-échelle défini par

$$X(t) = \int_{\mathbb{R}} (e^{it\xi} - 1) \cdot f^{1/2}(\xi) dW(\xi), \quad \text{pour tout } t \in \mathbb{R}, \quad (7)$$

où

- $dW(\xi)$  est une mesure de Wiener complexe tel que le processus  $X(t)$  soit à valeurs réelles ;
- la densité spectrale  $f(\xi)$  est une fonction paire, positive, telle que

$$f(\xi) = \sigma_i^2 |\xi|^{-(2H_i+1)} \quad \text{pour tout } |\xi| \in [\omega_i, \omega_{i+1}). \quad (8)$$

pour une famille de fréquences  $\omega_0 = 0 < \omega_1 < \dots < \omega_K < \omega_{K+1} = \infty$  et des familles de nombres réels  $(\sigma_i, H_i) \in (\mathbb{R}_+, \mathbb{R})$  pour  $i = 0, \dots, K$ .

En bref, le processus  $X$  défini par (7) a une densité spectrale affine par morceaux avec  $K$  ruptures situées aux fréquences  $(\omega_k)$  avec  $\omega_1 < \dots < \omega_K$ . Pour  $K = 0$  (zéro rupture fréquentielle), on retrouve un MBF. Ceci est illustré par la figure 4 a) et b). L'intégrale stochastique du terme de droite de (7) n'est définie que si l'intégrand est dans  $\mathbb{L}^2(\mathbb{R}, \mathbb{C})$ . La formule (7) est donc définie sous la condition

$$\int_{\mathbb{R}} |e^{it\xi} - 1|^2 \cdot f(\xi) d\xi < \infty, \quad (9)$$

qui est équivalente à  $H_K > 0$  et  $H_0 < 1$ . Aucune condition supplémentaire n'est imposée sur les autres indices de Hurst, on a donc  $H_i \in \mathbb{R}$ , pour  $0 < i \leq K$ . Ceci correspond aux constatations



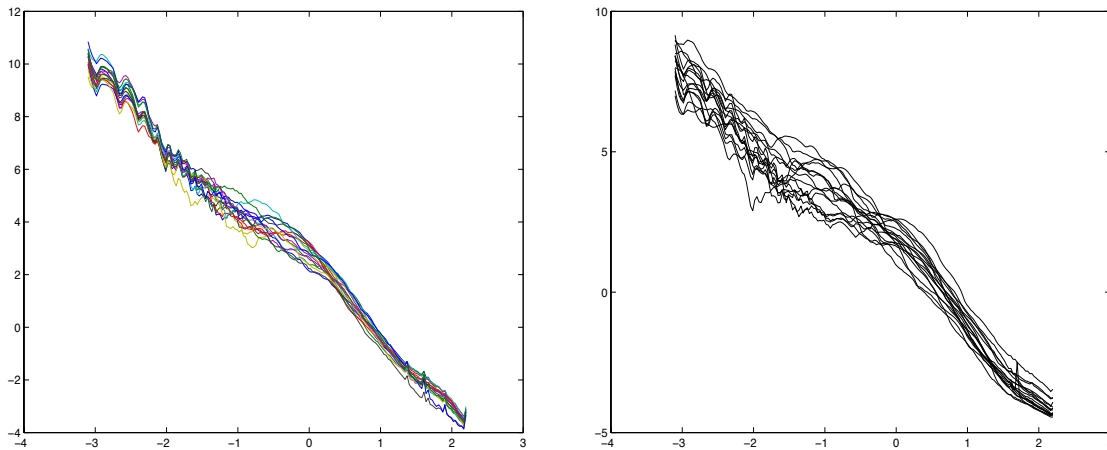


FIGURE 5 – Densité spectrale (log-log) du mouvement avant/arrière  $X(t)$  et du mouvement latéral  $Y(t)$  pour 16 sujets.

empiriques sur certains jeux de données physiologiques, mais s'avère impossible dans le cadre d'une modélisation par un MBF, pour lequel on a toujours  $0 < H < 1$ .

### 1.5 Etude statistique du mouvement brownien fractionnaire multi-échelle

Conformément à la problématique générale décrite section 1.2, on souhaite ensuite estimer les paramètres du modèle (7). La densité spectrale à une fréquence donnée peut être estimée en utilisant une analyse en ondelette, cf. [16, 4, 5]. On a appliqué cet algorithme aux données recueillies sur 16 sujets pendant une minute, représentées dans la figure 5. On constate visuellement que les densités spectrales sont affines par morceaux avec un ou deux coudes. On peut également construire un test d'adéquation qui rejette le modèle avec  $K = 0$  rupture fréquentielle, et accepte, au risque  $\alpha = 5\%$ , le modèle avec  $K = 1$  rupture fréquentielle pour 12 sujets (resp. 13 sujets) sur 16 pour  $X(t)$  (resp.  $Y(t)$ ), ou le modèle avec  $K = 2$  ruptures fréquentielles dans les autres cas, cf. [7, 4].

## Conclusion

La démarche exposée dans cette section est exemplaire du travail du mathématicien appliqué. Un modèle mathématique relativement sophistiqué est employé par des spécialistes d'une autre discipline, correspondant à des interprétations spécifiques aux domaine et posant la question du choix des bons outils statistiques. L'étude fine met en évidence la nécessité d'adapter le modèle pour mieux coller aux interprétations intuitives. Enfin, le modèle ainsi adapté s'avère transposable à de nouvelles applications.

Ainsi, nous avons proposé un modèle stochastique et construit des estimateurs statistiques

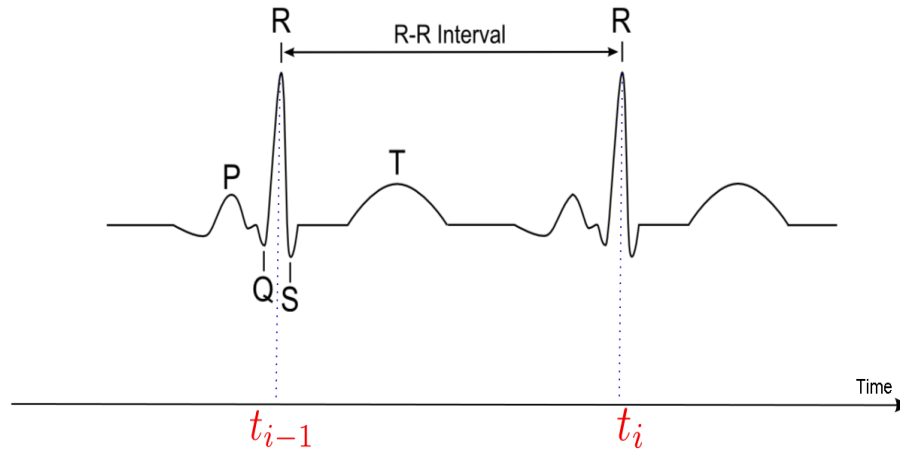


FIGURE 6 – Electro cardiogramme (ECG) et intervalle RR

permettant de valider l'intuition des spécialistes de physiologie de différents comportements de type fractionnaire selon les différentes bandes de fréquence. Ces différents comportements sont reliés par les spécialistes aux différents systèmes physiologiques en œuvre pour la régulation de l'équilibre humain (vision, oreille, muscles, plante des pieds). Nous avons également développé des estimateurs statistiques semi-paramétriques dans le cadre de ce nouveau modèle.

## 2 Régulation du rythme cardiaque et stress physiologique

Le même phénomène de régulation par différents systèmes neurophysiologiques, se traduisant mathématiquement par des comportements différents selon les différentes bandes de fréquence du signal, se retrouve dans l'analyse du rythme cardiaque. On peut en effet mesurer la durée de chaque battement cardiaque qui correspond à l'intervalle RR (en seconde) cf. figure 6, puis le traduire en fréquence cardiaque instantanée selon l'équation  $HR = 60/RR$  ( $HR = \text{"Heart Rate"}$ ). On obtient alors la série des fréquences cardiaques instantanées illustrée par la figure 7. On note une grande variabilité du rythme cardiaque, appelée variabilité sinusale ou HRV (*"Heart Rate Variability"*) par les cardiologues [26]. Le rythme cardiaque d'un sujet sain paraît ainsi aussi irrégulier et aléatoire que le cours des actifs financiers. Selon la jolie formule de Golberger [17],

*"Heartbeats [...] and health : is variability the spice of life ?"*

De fait, la régulation du rythme cardiaque résulte de l'interaction entre deux systèmes antagonistes, le système parasympathique ou vagal qui agit comme un frein et le système orthosympathique qui agit comme un accélérateur. Ces deux systèmes sont associés aux bandes de fréquences LF (*Low Frequency* :  $[0.04Hz, 0.15Hz]$ ) et HF (*High Frequency* :  $[0.15Hz, 0.5Hz]$ ). La bande LF correspond au système orthosympathique et la bande HF au système parasympathique.

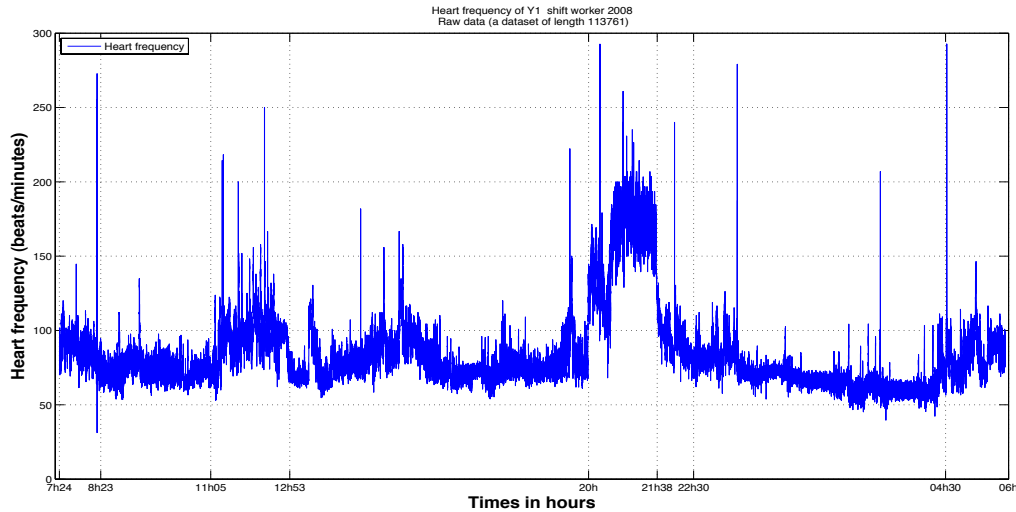


FIGURE 7 – Rythme cardiaque pour un travailleur posté pendant 24H

## 2.1 Modélisation mathématique

Pour le rythme cardiaque, nous avons travaillé sur des données fournies dans le cadre de deux collaborations : l'une en physiologie de l'exercice, pour des marathoniens<sup>2</sup> avec V. BILLAT (INSERM et Évry Génopole), l'autre avec le service de médecine du travail du CHU Clermont-Ferrand (cf. figure 7). Dans les deux cas, il s'agissait de données recueillies en condition écologique ou *in vivo*, c.à-d. en dehors d'un laboratoire (ou hôpital), tandis que les données sur l'équilibre humain étaient mesurées en laboratoire sur une plateforme de force, avec des conditions expérimentales spécifiées. Sortir du laboratoire (ou de l'hôpital) pour traiter des données en condition de vie réelle, entraîne une nouvelle adaptation du modèle.

En effet, dans un laboratoire les conditions expérimentales restent fixes et on peut modéliser la série par un processus stationnaire, ou à accroissements stationnaires. Hors laboratoire, les conditions peuvent varier. Il est alors nécessaire de passer d'un modèle gaussien stationnaire à un modèle gaussien localement stationnaire introduit par Dahlhaus [15]. D'après le théorème de Cramér [12], tout processus gaussien admet une représentation spectrale. On en déduit que tout processus gaussien localement stationnaire  $X$  admet la représentation

$$X(t) = \mu(t) + \int_{\mathbb{R}} e^{it\xi} \cdot f(t, \xi)^{1/2} dW(\xi), \quad \text{pour tout } t \in \mathbb{R}, \quad (10)$$

où

- $dW(\xi)$  est une mesure de Wiener complexe tel que le processus  $X(t)$  soit à valeurs réelles ;

---

2. Cette collaboration a été financée par le contrat DIM-Digitéo PHYSIOSTAT ("*Détection des ruptures des états stables physiologiques par l'approche fractale des signaux*"), 2009-2011.

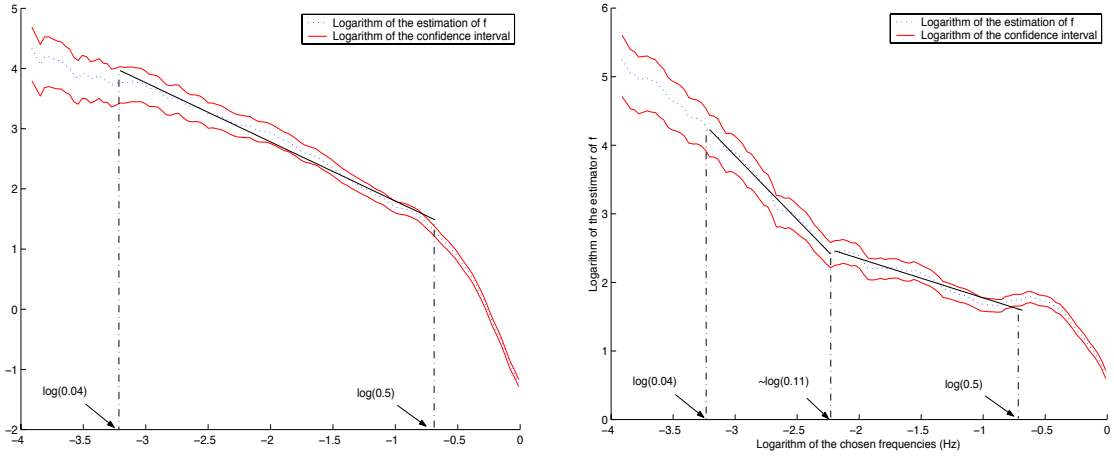


FIGURE 8 – Densité spectrale sommeil/travail

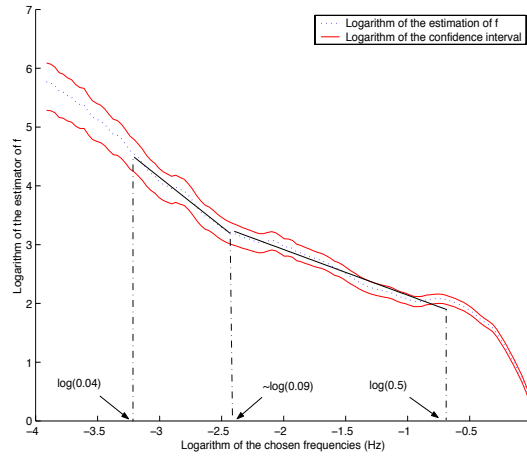


FIGURE 9 – Densité spectrale repos ou loisir

- la densité spectrale  $\xi \mapsto f(t, \xi)$  est une fonction paire et positive, constante par morceaux par rapport au temps  $t$ , ce qui revient à dire qu'il existe une partition  $\tau_1, \dots, \tau_K$  telle que  $f(t, \xi) = f_k(\xi)$  pour  $t \in [\tau_i, \tau_{i+1} [$  ;
- la fonction  $t \mapsto \mu(t)$  correspond à l'espérance du processus, elle est constante par morceaux pour une partition  $\tilde{\tau}_1, \dots, \tilde{\tau}_L$  avec  $\mu(t) = \mu_\ell$  si  $t \in [\tilde{\tau}_\ell, \tilde{\tau}_{\ell+1} [$ .

Si la segmentation est connue, on peut estimer la densité spectrale sur chacun des segments. On utilise l'analyse en ondelette [16], on obtient alors différentes densités spectrales selon le type d'activité (sommeil, travail, repos ou loisir), illustré par la représentation en coordonnées log-log des figures 8 et 9 (reprises de [5]).

On remarque que la densité spectrale est affine par morceaux sur les bandes LF (orthosympathique) HF (parasympathique) et reste affine sur la bande VLF (*Very Low Frequency*) avec la

même pente que sur la bande LF, la même constatation est faite par [14]. On remarque également que la forme de la densité spectrale varie avec le type d'activité.

## 2.2 Identification de la segmentation d'un processus localement stationnaire

Dans l'exemple précédent, on connaissait les périodes d'activité de l'individu (travail, repos ou loisir, sommeil) et on pouvait vérifier une modification du comportement de la densité spectrale dans les bandes LF et HF. On a ainsi mis en évidence une densité affine par morceaux sur les bandes HF, LF et VLF ainsi que des ruptures de la densité spectrale selon le comportement de l'individu. Ceci suggère que des phénomènes de fatigue neuro-physiologique se traduisent par des ruptures sur la moyenne  $\mu(t)$ , ou la densité spectrale dans les bandes HF ou LF. On est donc amené à chercher des ruptures sur la moyenne d'une série temporelle aléatoire. En effet, on montre dans la section 2.2.2 que les variations de la densité spectrale dans la bande HF ou LF entraîne des ruptures sur la moyenne de deux séries dérivées de celles des coefficients d'ondelette. Nous présentons ci-dessous un rappel sur la détection de rupture.

### 2.2.1 Détection de rupture

La segmentation des séries temporelles ainsi que la détection de rupture en temps réel sont deux branches spécifiques des statistiques appliquées, souvent considérées comme une sous section de l'économétrie, resp. des statistiques industrielles ou du contrôle de qualité. Les premiers travaux (en temps réel ou "*sequential analysis*" en anglais) remontent à A.Wald et furent effectués pendant la seconde guerre mondiale puis récapitulés dans [27]; la détection rapide de rupture redevint un sujet chaud pendant la guerre froide, cf. [25], avant d'être appliqué dans l'industrie.

La détection de rupture *a posteriori* dans différents modèles stochastiques a été étudiée depuis les années 1950. Une importante littérature existe sur le sujet, cf. les monographies [6, 9, 13]. Cependant, la prise en compte de la complexité des algorithmes et du temps de calcul informatique est récente. On peut la dater du début du 21ème siècle simultanément à l'explosion de la taille des jeux de données due aux progrès technologiques et informatiques. Ce phénomène est popularisé sous le nom de "*Big Data*"<sup>3</sup>. Pour fixer les idées, donnons un modèle simple avec des ruptures :

### Modèle le plus simple de ruptures

Soit  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  une série temporelle de taille  $n$ , c.à-d. une suite de variables aléatoires gaussiennes indépendantes indexées par le temps  $t = 1, 2, \dots, n$  pour laquelle il existe une segmentation  $\tau = (\tau_1, \dots, \tau_K)$  avec  $0 = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_K < \tau_{K+1} = n$ . L'entier  $K$

---

3. comme le confirment deux numéros spéciaux intitulés "*The data deluge*" the Economist February 2010, et "*Big Data*" La Recherche, N° 482, décembre 2013.

correspond au nombre de ruptures et peut être nul. Pour  $k = 0, \dots, K$ , si  $t \in [\tau_k, \tau_{k+1}[$ , alors  $X_t \in \mathcal{N}(\mu_k, \sigma_k^2)$ , où  $\mathcal{N}(\mu, \sigma)$  est la loi normale de moyenne  $\mu$  et de variance  $\sigma^2$ .

La question statistique est d'identifier les points de ruptures  $(\tau_k)_{k=1, \dots, K}$ , de déterminer leur nombre  $K$  et d'estimer les valeurs  $(\mu_k)_{k=0, \dots, K}$  de la moyenne entre les instants de rupture.

L'état de l'art des années 1990 consistait à utiliser la méthode des moindres carrés (éventuellement pénalisée) couplée à l'algorithme de la programmation dynamique, ce qui aboutit à une complexité en temps de calcul et en quantité de mémoire de l'ordre de  $\mathcal{O}(n^2)$ . Dans [8], nous avons proposé une méthode en deux étapes, appelée FDPV, dont la complexité en temps de calcul et en quantité de mémoire sont toutes deux d'ordre  $\mathcal{O}(n)$ . L'algorithme FDPV a été codé en Java<sup>4</sup> et traite le jeu de données correspondant à la figure 10 (de taille  $n = 130.000$ ) en 250 millisecondes avec un processeur Intel Core i7-3770 cadencé à une fréquence de 3.4 GHz.

### 2.2.2 Analyse temps-fréquence dans les bandes HF et LF par une méthode d'ondelette

Suivant la recommandation des cardiologues, on souhaite détecter les variations de l'énergie spectrale dans les bandes HF et LF. Cette approche est corroborée par les analyses statistiques menées dans [5, 14] qui montre que la densité spectrale est affine par morceaux (en coordonnées log-log) sur les bande HF et LF. Soient  $\psi_1$  and  $\psi_2$  deux ondelettes dont les supports fréquentiels sont concentrés dans les bandes HF resp. LF. Pour un processus stochastique, par exemple donné par (10), on définit les coefficients d'ondelette

$$d_i(b) = \int_{\mathbb{R}} \psi_i(t - b) X(t) dt,$$

Supposons pour simplifier que le processus est stationnaire, on peut alors appliquer la formule de Fubini stochastique, cf. [1], et on obtient

$$d_i(b) = \int_{\mathbb{R}} \widehat{\psi}_i(\xi) f(\xi)^{1/2} dW(\xi), \quad \text{puis} \quad \mathbb{E}[|d_i(b)|^2] = \int_{\mathbb{R}} |\widehat{\psi}_i(\xi)|^2 f(\xi) dW(\xi).$$

Le résultat reste valable pour des processus gaussiens localement stationnaires. On construit ensuite deux nouvelles séries

$$Z_1(b) = \log(|W_1(b)|^2), \quad \text{et} \quad Z_2(b) = \log(|W_2(b)|^2). \quad (11)$$

On remarque que

$$Z_1(b) = \log\left(\int_{\mathbb{R}} |\widehat{\psi}_1(\xi)|^2 f(\xi) d\xi\right) + \zeta_1(b),$$

et

$$Z_2(b) = \log\left(\int_{\mathbb{R}} |\widehat{\psi}_2(\xi)|^2 f(\xi) d\xi\right) + \zeta_2(b),$$

---

4. un prototype Matlab fut réalisé en 2010 par Pierre Bertrand et une nouvelle version Java a été codé en 2013 par Guillaume Paugam, ingénieur informatique travaillant dans le cadre du projet ANR "Do Well B."

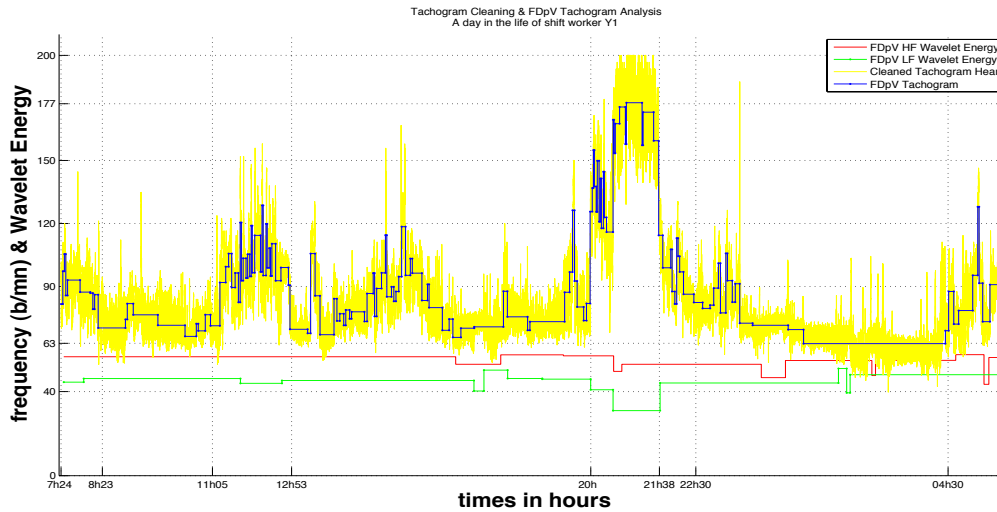


FIGURE 10 – Rythme cardiaque pour un travailleur posté pendant 24H

où  $\zeta_1(b)$  et  $\zeta_2(b)$  sont des variables de loi  $\ln \chi_1^2$ . Donc, les ruptures sur la moyenne des séries  $Z_1(b)$ ,  $Z_2(b)$  correspondent aux ruptures au cours du temps sur la densité spectrale dans les bandes HF et LF, cf. [1, 8, 19].

### 2.2.3 Résultats obtenu sur un travailleur posté et un marathonien

En résumé, à partir de la série des intervalles RR, nous calculons deux nouvelles séries : celle des modules coefficients d'ondelette HF et LF, que nous segmentons ensuite par la méthode FdPv de détection a posteriori de rupture sur la moyenne. Cette méthode est appliquée à la série RR du travailleur posté de la figure 7. Les données ont été préalablement nettoyées par un algorithme décrit dans [19], puis traduit en Java et amélioré en collaboration avec Guillaume Paugam. On obtient ainsi la figure 10. On a appliqué la même méthode à un marathonien illustré par la figure 11. Dans les deux cas, on obtient une segmentation plus fine qui permet à la fois d'analyser le signal et de le compresser. Ainsi le signal de la figure 11 a) est de taille 45.000 et il est compressé à une taille environ 1.000 dans la figure 11 b). On distingue également la période d'échauffement du marathonien avant le début de la course à 8h45, ainsi qu'une fatigue physique apparaissant après 3H de courses. De plus, on peut classifier les divers états selon le stress physique [19, 2].

## 3 Stress psychologique et autisme

L'ensemble des travaux exposés dans les sections précédentes met en évidence la possibilité de mesurer l'impact des régulations neuro-psychologiques par des méthodes de statistique des

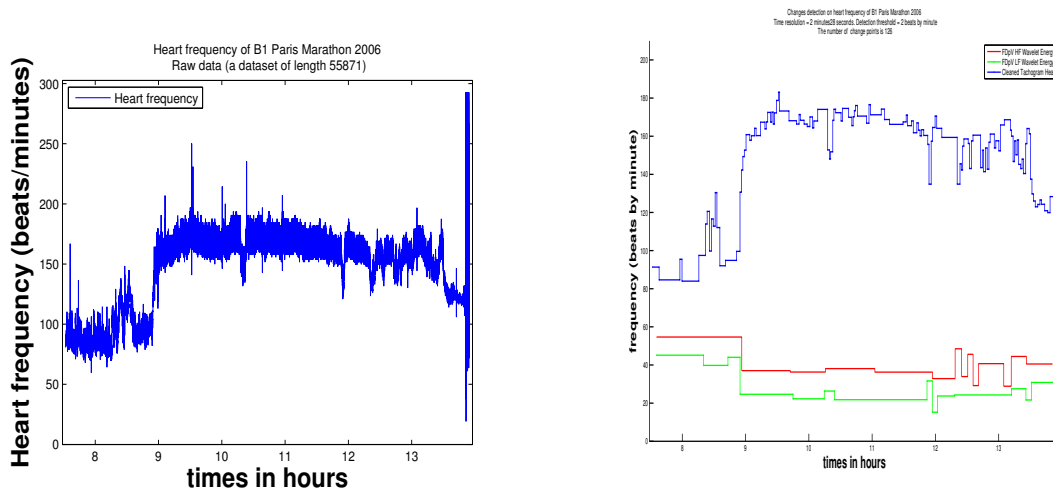


FIGURE 11 – a) signal brut d'un marathonien ; b) signal segmenté par FDpV + module des coefficients d'ondelette HF et LF

séries appliquées à des signaux physiologiques. En collaboration avec le laboratoire de psychologie (UMR 6024), nous nous sommes intéressé à l'autisme chez les adolescents (12-22 ans). Notre objectif à long terme est la création d'un Système de suivi Personnel de la Santé (SPS), pour la détection précoce des comportements difficiles (par exemple, accès de colère) de personnes relevant du spectre de l'autisme, à l'extérieur de l'hôpital. En effet, les progrès technologiques récents permettent actuellement la surveillance de signaux physiologiques grâce à des systèmes embarqués et non invasifs (rythme cardiaque, conductance de la peau, pression artérielle, fréquence respiratoire, etc...). On peut donc, avec une instrumentation légère et bien supportée, enregistrer plusieurs paramètres physiologiques et leur évolution au cours du temps. Ceci permet de s'affranchir d'une instrumentation lourde et intrusive telle que l'IRM ou la biopsie, qui ne sont praticables qu'en laboratoire ou à l'hôpital. On applique également les méthodes décrites dans les sections 1 et 2 pour le rythme cardiaque ou le mouvement du poignet.

Cette perspective a été validée par l'ANR (Agence Nationale pour la Recherche) qui finance le projet " *Do Well B. : Design Of Well Being monitoring systems* ", littéralement « Conception de Systèmes de Monitoring du Bien-être ». Le projet « Do Well B. » a démarré en 2013, c'est un projet pluridisciplinaire impliquant quatre équipes :

- des probabilistes et statisticiens du Laboratoire de Mathématique de Clermont-Ferrand ;
- des chercheurs dans le domaine de l'autisme (LAPSCO, UMR 6024) du Laboratoire de Psychologie Clermont-Ferrand ;
- des physiologistes et cardiologues du CHU Clermont-Ferrand ;
- des spécialistes dans le traitement des signaux audiovisuels de l'École Nationale d'Ingénieurs de Tunis (ENIT, Tunisie).



Nous avons à ce jour conçu les protocoles expérimentaux qui vont être soumis à la Commission de Protection des Personnes (CPP). Nous pourrons ensuite procéder aux tâches suivantes : recrutement des participants, passation du protocole expérimental, recueil des données, exploitation et interprétation de ces dernières en concertation avec les psychologues et spécialistes de physiologie. Dans un second temps, les résultats seront soumis pour publication dans des revues de statistiques, traitement du signal et des revues de psychologie ou médecine. Les prochains challenges de probabilité et statistique seront

- de passer de la détection a posteriori de rupture sur les paramètres structurels des séries physiologiques, à la détection en temps réel ;
- de développer une méthode de détection de rupture sur la pente de la densité spectrale.

## Conclusion

Nous avons présenté des exemples d'applications de la statistique à des séries physiologiques et mis en évidence l'interaction entre application et modélisation mathématique.

Le développement technologique actuel et les avancées dans la miniaturisation des systèmes électroniques permettent l'introduction d'appareils de mesures physiologiques de plus en plus fiables et précis. Nous assistons aux prémices d'un développement massif de capteurs capables de faire des mesures aussi précises que dans un centre médical spécialisé, offrant en outre la possibilité d'être manipulables par le grand public sans l'intervention d'un professionnel. De plus, des solutions logicielles ont été introduites permettant de donner des résumés statistiques de certains signaux physiologiques. Le développement de tels capteurs intelligents et connectés ("*smart device*") suppose la mise au point préalable d'algorithmes et de modèles mathématiques qui ouvriront ensuite sur de nouvelles applications.

## Remerciements

1. Cet ensemble de travaux a été fait avec les soutiens 1) Digitéo projet "PHYSIOSTAT" (2009-2011) ; 2) INRIA qui m'a accueilli en délégation (2008-2011), programme BIO, médecine numérique, application en e-santé, ou TIC et santé et 3) ANR projet n°12-BS01-0016-02 intitulé «*Do Well B. : Design of Well Being Monitoring System*» (2013-2016).
2. Je cite librement des parties des travaux de thèses que j'ai co-encadrées : Arnaud Mouzat (1999-2004), Medhi Fhima (2008-2011), Nadia Khalfa (2010-...).
3. Je remercie également tous mes collègues, co-auteurs ou relecteurs<sup>5</sup> mathématiciens : (ordre alphabétique) A. Ayache (Lille 1), N. Azzaoui, A. Benassi, J-M. Bardet (Paris 1), S. Da-

---

5. quand aucune affiliation n'est précisée, il s'agit de collègues clermontois.

- chian, M.E. Dury, G. Fleury, A. Guillin, D. Malouche (Tunis), J. Lévy-Vehel (INRIA Saclay) ;
4. et non-mathématiciens : M. Dabonneville et P. Vaslin (LIMOS, UMR 6158 CNRS et Université Clermont-Ferrand 2), V. Billat (INSERM et Évry Université), R. Ghazi, M. Jaidane (ENIT, Tunis), Gil Boudet, A. Chamoux et F. Dutheil (CHU Clermont-Fd), C. Auxiette, P. Chambres et M. Mermilod (Lapsco, UMR 6024 CNRS et Université Clermont-Ferrand 2) ;
  5. ainsi que Guillaume Paugam (ingénieur informatique du projet ANR "Do Well B. ").

## Références

- [1] Ayache, A., Bertrand, P.R. (2011), "Discretization error of wavelet coefficient for fractal like process", *Advances in Pure and Applied Mathematics*, vol. 2, No. 2 : 297-321.
- [2] Azzaoui N, Guillin A, Dutheil F, Boudet G, Chamoux A, Perrier C, Schmidt J, Bertrand PR : Classifying heartrate by change detection and wavelet methods for emergency physicians. *ESAIM Proceedings* 2014.
- [3] Bardet, J.M., Bertrand, P. (2007a), "Definition, properties and wavelet analysis of multiscale fractional Brownian motion", *Fractals*, 15, p. 1-15.
- [4] Bardet, J.M., Bertrand, P. (2007b), "Identification of the multiscale fractional Brownian motion with biomechanical applications", *Journal of Time Series Analysis*, 28, p. 1-52.
- [5] Bardet, J.M., Bertrand, P.R. (2010), "A nonparametric estimator of the spectral density of a continuous-time Gaussian process observed at random times", *Scandinavian J. of Statistics*. Vol. 37 : 458-476.
- [6] Basseville, M., and Nikiforov, I. (1993). *Detection of abrupt changes : theory and application*. Prentice Hall, Englewood Cliffs, NJ.
- [7] Bertrand, P., Bardet, J.M., Dabonneville, M., Mouzat, A., Vaslin, P. (2001), "Automatic determination of the different control mechanisms in upright position by a wavelet method", *Proceedings of the I.E.E.E. Engineering in Medicine and Biology Society*, T. 23, vol. 1-4, 1163-1166.
- [8] Bertrand, P.R., Fhima M., Guillin, A. (2011), "Off-line detection of multiple change points with the Filtered Derivative with p-Value method", *Sequential Analysis*, 30 (2) : 172-206.
- [9] Brodsky, B. E., and Darkhovsky, B. S. *Nonparametric methods in change-point problems*, vol. 243 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993.
- [10] Collins J.J., de Luca C.J. (1993). Open-loop and closed-loop control of posture : A random walk analysis of center-of-pressure trajectories, *Exp. Brain Research*, 9, 308-318.

- [11] Collins J.J., de Luca C.J. (1994). Random walking during quiet standing, *Physical Review Letters*, 73 : 764–767.
- [12] Cramér, H. and Leadbetter, M. R. (1967). *Stationary and related stochastic processes. Sample function properties and their applications*. Wiley and Sons.
- [13] M. Csörgő and L. Horvath (1997) *Limit Theorem in Change-Point Analysis*, J.Wiley & Sons.
- [14] Cysarz D., von Bonin, D., Brachmann, P., Buetler, S. Edelauser, F., Laederach-Hofmann, K. and Heusser P. (2008). Day-to-night time differences in the relationship between cardio-respiratory coordination and heart rate variability. *Physiol. Meas.* **29** :1281-1291.
- [15] R. Dahlhaus (1997) Fitting time series models to nonstationary processes, *Ann. Stat.*, **25**.
- [16] Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. on Inform. Theory*, 38, p. 910-917.
- [17] Goldberger, A.L. (2001). Heartbeats, hormones and health : is variability the spice of life ? *Am. J. Crit. Care Med.*, 163, 1289–1290.
- [18] Hurst,H. (1951) Long Term Storage Capacity of Reservoirs, Transactions of the American Society of Civil Engineers, 116, 770–799.
- [19] Khalfa, N., Bertrand, P.R., Boudet, G., Chamoux, A., Billat, V. (2012). “Heart rate regulation processed through wavelet analysis and change detection. Some case studies”, *Acta Biotheoretica*. Vol. 60, Issue 1-2, 109–129.
- [20] Kolmogorov, A.N. (1940) : Wiener'sche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. C. R. (Doklady) Acad. URSS (N.S.) 26, 115–118
- [21] Mandelbrot, B. and Van Ness J. (1968). Fractional Brownian motion, fractional noises and applications. *SIAM review* 10, p.422-437.
- [22] Mandelbrot, B. (2014) *La forme d'une vie*. Flammarion.
- [23] Meyer, Y., Sellan, F., Taqqu, M. S. : Wavelets, generalized white noise and fractional integration : the synthesis of fractional Brownian motion, *J. Fourier Anal. Appl.* 5(5) 465–494, (1999)
- [24] Samorodnitsky, G. and Taqqu M.S. (1994), *Stable non-Gaussian Random Processes*, Chapman and Hall.
- [25] Shiryaev A.N. (2010) *Quickest Detection Problems : Fifty Years Later*, *Sequential Analysis* Volume 29, Issue 4, 2010
- [26] 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.
- [27] Wald, A. (1947). *Sequential Analysis*. New York : John Wiley and Sons. ISBN 0-471-91806-7.