

Identification de la composante déterministe d’une dynamique du monde réel

D. J. De Oliveira^{1,2}, M. E. D. Gomes¹, L. A. Aguirre¹ & C. Letellier¹

¹ Lab. de Modelagem Análise e Controle de Sistemas Não-Lineares (MACSIN), Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Belo Horizonte, MG, 31270-901, Brazil

² CORIA UMR 6614 — Université de Rouen, BP. 12, 76801 Saint-Etienne du Rouvray cedex
deoliveira@coria.fr

Résumé. Les systèmes dynamiques du monde réel résultent le plus souvent d’une combinaison entre une composante déterministe — qui peut être de basse dimension — et une composante aléatoire. Lorsque la part de la composante aléatoire est très importante devant celle de la composante déterministe, l’étude de ces systèmes se révèle le plus souvent assez ardue. Il est alors particulièrement important de pouvoir séparer les deux composantes. Pour cela, nous avons développé une procédure reposant sur un modèle NARMAX entrée-sortie ; elle sera testée sur des modèles numériques simples et sur un électrocardiogramme de 24 heures.

Abstract. Dynamical systems from the real world are most often a combination of a deterministic component — which may be low dimensional — and a stochastic component. When the “weight” of the stochastic component is important comparing to those of the deterministic component, the investigation of such a system is rather difficult. It is thus relevant to distinguish both components. In order to do that, we developed a procedure based on an input-output NARMAX model ; this procedure will be tested using simple numerical models and a 24 hours electrocardiogram.

Notre motivation réside dans la compréhension de la dynamique sous-jacente à l’activité cardiaque. Plus particulièrement, nous souhaitons déterminer s’il existe une composante de nature déterministe au sein de la variabilité cardiaque, notamment à l’échelle de la journée. Tout d’abord, un électrocardiogramme sur 24 heures s’étudie le plus souvent sous la forme d’un tachogramme, soit la succession des intervalles de temps entre les ondes R des pulsations cardiaques (Fig. 1). Ces tachogrammes résultent des interactions cardiovasculaires et des transitions entre différents états physiologiques liés à l’activité humaine (sommeil, digestion, activité physique ou intellectuelle, etc.) [1]. Le problème majeur qui se présente est que la variabilité du rythme cardiaque se développe sur des échelles de temps très différentes. Si la variabilité à l’échelle du battement est très réduite chez le sujet sain (il a peu d’arythmies), il existe des fluctuations qui peuvent être importantes à échelles de temps beaucoup plus longues et pouvant aller jusqu’à celles des rythmes circadiens. Malheureusement, une étude sérieuse de ces fréquences nécessiterait des enregistrements sur plusieurs jours consécutifs, voire plusieurs semaines. En effet, une certaine régularité des tachogrammes se retrouve chez un sujet sain qui prend soin de répéter à l’identique son activité sur 24 heures.

La variabilité à très basse fréquence étant liée à l’activité humaine, elle ne peut être, par essence, déterministe : elle ne peut donc pas être modélisée par un modèle déterministe. Pour cela, nous cherchons à appliquer une variation de l’un des paramètres du modèle par une modulation basse fréquence directement extraite de la dynamique originale. Par ailleurs, il y a des fluctuations sur de courtes échelles de temps qui résultent au mieux d’une dynamique déterministe de haute dimension. Tout le problème se résume alors à l’identification d’une composante déterministe qui sera excitée par une composante aléatoire.

1 Technique de modélisation

L’origine complexe de l’activité cardiaque nécessite d’utiliser une procédure de modélisation reposant sur plusieurs blocs. Dans ce qui suit, nous considérons que la composante très basse fréquence — à l’échelle

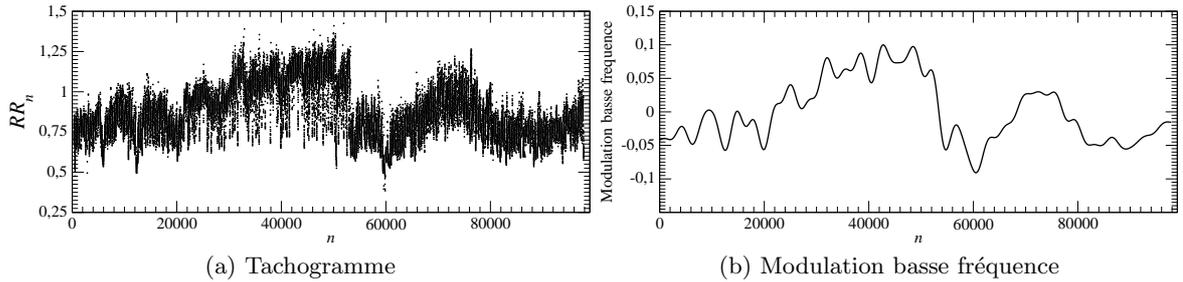


Fig.1. Tachogrammes sur 24 heures (plus de 90000 battements) d'un sujet sain. Les fluctuations à très basse fréquence — que nous appelons *modulation basse fréquence* — sont la signature de l'impact de l'activité humaine sur la variabilité cardiaque.

de la journée — ne peut que résulter d'une composante aléatoire liée à l'activité humaine. Nous utiliserons donc une modulation basse fréquence comme entrée du modèle, non pas comme une excitation en entrée du modèle mais plutôt comme une variation de l'un de ses paramètres. En effet, pour son identification, nous n'avons recours qu'à un jeu de données assez limité sur lequel la modulation basse fréquence est presque constante : le modèle ne peut donc y être sensible. Ensuite, il y a clairement une composante haute fréquence dont l'origine est un système complexe de haute dimension qui, à l'heure actuelle, ne peut être prise en compte que par un modèle stochastique. Pour cette composante, nous utiliserons la procédure introduite par McSharry et ses collègues [2]. Elle consiste à identifier deux fréquences caractéristiques — lorsqu'elles existent — dans le signal original et à construire un signal aléatoire dont le spectre de Fourier est la somme de deux distributions Gaussiennes $S(f) = S_1(f) + S_2(f)$ telles que

$$S_i(f) = \frac{\sigma_i^2}{\sqrt{2\pi c_i^2}} e^{-\frac{(f - f_i)^2}{2c_i^2}}$$

où f_i désignent les fréquences choisies et c_i leur déviation standard. Les puissances associées à chacune des fréquences sont données par les quantités σ_i^2 . Les phases peuvent être distribuées aléatoirement ou corrélées à la phase du signal original. Enfin, la composante déterministe est obtenue à l'aide d'une technique de modélisation globale discrète. Le modèle est du type NARMAX (Non-linear Auto-Regressive Moving Average eXogeneous inputs) [3]. La procédure de modélisation peut se résumer selon le schéma représenté Fig. 2. Comme nous le verrons, lors de l'identification du modèle NARMAX, le signal de sortie est le signal original (ECG ou une dynamique déterministe à laquelle on appliquera une modulation basse fréquence). Ensuite, lorsque le modèle est intégré, nous obtenons un signal synthétique en sortie. Dans tous les cas, le signal d'entrée est un signal aléatoire construit selon la procédure de McSharry et ses collègues [2].

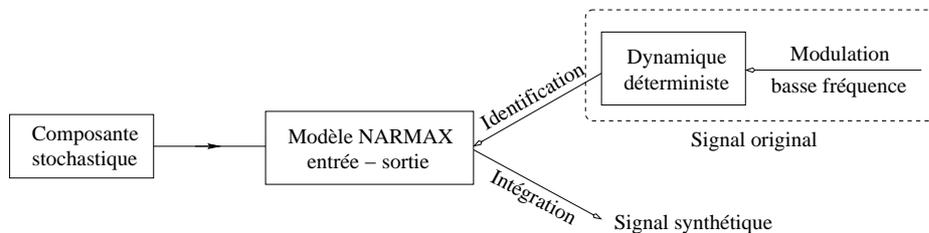


Fig.2. Le modèle NARMAX est estimé à l'aide d'une entrée du type composante stochastique. Le signal de sortie est le signal original (ECG ou dynamique déterministe lentement modulée par variation d'un paramètre) lors de l'identification du modèle et le signal synthétique lors de l'itération du modèle.

2 Identification de composante déterministe simple

Notre procédure est testée en utilisant la fonction logistique

$$x_n = \mu x_{n-1}(1 - x_{n-1}) \quad (1)$$

comme système dynamique déterministe. Un bruit additif est ajouté à la dynamique de manière à simuler la contamination par du bruit de la dynamique sous-jacente. Le bruit est construit selon la procédure de McSharry [2]. Aucune fréquence caractéristique ne se dégageant du spectre de Fourier d'une fonction logistique, les deux fréquences sont posées arbitrairement. Ceci résulte du fait que la dynamique symbolique est presque complète, c'est-à-dire que la presque totalité des séquences possibles sont réalisées par une trajectoire. Pour un paramètre $\mu = 3,89$, le comportement asymptotique est chaotique (Fig. 3).

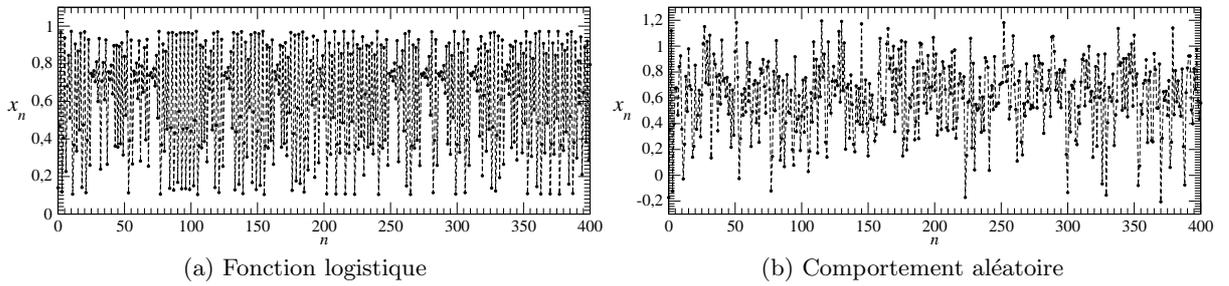


Fig.3. Comportement chaotique solution de la fonction logistique pour $\mu = 3,89$ et comportement aléatoire obtenu par la procédure de McSharry *et al.*

Nous commençons par estimer un modèle à partir de l'évolution de la fonction logistique bruitée à 5%. Par rapport au schéma (Fig. 2), il n'y a pas de modulation basse fréquence. Le modèle NARMAX ainsi obtenu est de la forme :

$$\begin{aligned} y_k = & 0,0232 - 3,8008y_{k-1}^2 + 3,7869y_{k-1} - 0,0238y_{k-20}y_{k-11} + 0,0171y_{k-24}y_{k-17} \\ & - 0,0771y_{k-5}y_{k-1} + 0,0484y_{k-5}y_{k-3} + 0,0736y_{k-2}y_{k-1} - 0,0394y_{k-3}y_{k-2} \\ & - 0,02y_{k-14}y_{k-4} + 0,019y_{k-1}y_{k-22} - 0,0157y_{k-2}y_{k-19} \end{aligned} \quad (2)$$

où les termes en y_{k-n} désignent la composante déterministe et les termes en u_{k-n} la composante aléatoire. On notera la présence de seulement deux termes de poids relativement faible pour la composante aléatoire. Le modèle a, par conséquent, majoritairement retenu une composante déterministe. Nous pouvons remarquer que le modèle contient un sous-système de la forme

$$y_k = 3,7869y_{k-1} - 3,8008y_{k-1}^2 = 3,7869y_{k-1}(1 - 1,00367y_{k-1})$$

soit une forme particulièrement proche de la fonction logistique (1). Le reste des termes, y compris les termes déterministes, sont introduits pour tenir compte de la composante aléatoire.

Une intégration du modèle (2) permet de retrouver la forme caractéristique d'une parabole lorsqu'une application de premier retour est tracée (Fig. 4c). Elle est favorablement comparée à la fonction logistique bruitée (Fig. 4b) : toutes deux diffèrent clairement de la composante aléatoire (Fig. 4a). Une validation fine du modèle peut se faire à l'aide d'une dynamique symbolique. Puisque nous avons une dynamique unimodale — la fonction logistique, une dynamique à deux symboles suffit : « 0 » pour la branche croissante et « 1 » pour la branche décroissante de la parabole. Dans un premier temps, nous utilisons une partition telle que le point critique soit en $x_c = 0,5$, ce qui correspond à la partition théorique de la fonction logistique. La série temporelle $\{y_k\}$ est donc transformée en une suite de symboles (des 0 et des 1). Deux choses sont recherchées : la population d'orbites périodiques et les probabilités de réalisation de ces orbites périodiques.

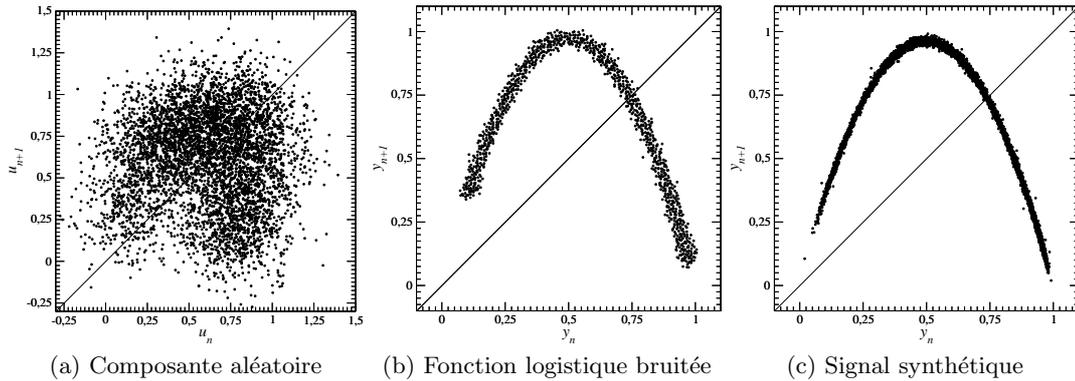


Fig.4. Comportement (c) obtenu par itération du modèle (2) excité par le signal aléatoire (a). Le paramètre μ est fixé à 3,89 et 5% de bruit additionnel est injecté (b).

Pour $\mu = 3,89$, la population d'orbites périodiques est caractérisée par la séquence principale (10011). Comme cela peut être facilement vérifié à l'aide d'un diagramme de bifurcations, le couple d'orbites de période 3 — respectivement codées par (101) et (100) — fait partie de la population d'orbites. En effet, la fenêtre de période 3 est observée pour $\mu \in [3,825 ; 3,86]$, soit une valeur de μ inférieure à celle imposée à la fonction logistique (3,89). La composante aléatoire — dont la phase est corrélée à l'évolution de la fonction logistique bruitée — se caractérise par une dynamique symbolique complète, c'est-à-dire que la séquence principale est (100000); en d'autres termes, toutes les séquences symboliques sont réalisées (Tab. 1). Toutefois, il existe une différence importante avec une dynamique déterministe caractérisée par une dynamique symbolique complète : les différentes séquences symboliques ne sont pas équiprobables comme c'est le cas pour les dynamiques déterministes [4]. Un bruit blanc serait également caractérisé par des séquences symboliques équiprobables. Notons que les probabilités de réalisation de la composante aléatoire (Fig. 5a) et de la fonction logistique bruitée diffèrent (Fig. 5b). Le modèle obtenu a une dynamique caractérisée par une séquence principale (1001), soit une population d'orbites légèrement plus développée que celle de la dynamique originale. La composante aléatoire tend donc à développer la dynamique comme cela est habituellement observé. Une entropie de Shannon [5] est calculée : alors qu'une valeur de 2,73 est trouvée pour le bruit, l'entropie du signal synthétique (2,00) est très proche de celle du signal original (2,06). Nous pouvons considérer le modèle comme validé.

Tab.1. Population d'orbites périodiques instables extraites de la fonction logistique ($\mu = 3,89$), des différents modèles et de la composante aléatoire. Un point est reporté lorsque l'orbite est identifiée. Seules les orbites de période inférieure à 7 sont reportées.

Orbites extraites	Composante aléatoire	Fonction logistique	Modèle (2)	Modèle (3)	Orbites extraites	Composante aléatoire	Fonction logistique	Modèle (2)	Modèle (3)
1	•	•	•	•	10011	•	•	•	•
10	•	•	•	•	100111	•		•	
1011	•	•	•	•	100110	•		•	
101110	•	•	•	•	1001	•		•	
101111	•	•	•	•	1000	•			
10111	•	•	•	•	100010	•			
10110	•	•	•	•	100011	•			
101	•	•	•	•	10001	•			
100	•	•	•	•	10000	•			
100101	•	•	•	•	100001	•			
10010	•	•	•	•	100000	•			

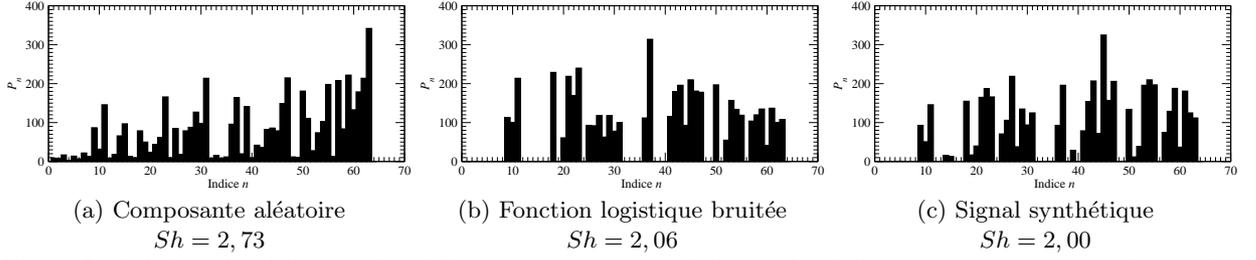


Fig.5. Probabilités de réalisation des différentes séquences symboliques des différentes composantes et du signal synthétique obtenu par intégration du modèle NARMAX (2).

Nous appliquons maintenant notre procédure d'identification lorsqu'une dérive sur le paramètre μ de la fonction logistique est introduite. La modulation (Fig. 1b) est extraite par filtre passe-bas de l'électrocardiogramme enregistré sur 24 heures (Fig. 1a). Nous retrouvons des fenêtres périodiques (période 3) typiquement observée pour $\mu \in [3,825 ; 3,86]$. Le modèle obtenu à partir de cette dynamique a la forme

$$\begin{aligned}
 y_k = & 0.0643 + 3.6376y_{k-1} - 3.7268y_{k-1}^2 - 0.0143y_{k-19}y_{k-16} + 0.0775y_{k-13}y_{k-12} \\
 & + 0.0142y_{k-14}y_{k-11} - 0.0493y_{k-12}y_{k-8} - 0.0358y_{k-13}y_{k-7} - 0.0183y_{k-9}y_{k-4} \\
 & - 0.0471u_{k-7} + 0.0843u_{k-7}y_{k-1} + 0.0245u_{k-4}u_{k-1}
 \end{aligned} \quad (3)$$

Comme précédemment, nous retrouvons un sous-système du type

$$y_k = 3.6376y_{k-1} - 3.7268y_{k-1}^2 = 3.6376y_{k-1} (1 - 1,0245y_{k-1}) . \quad (4)$$

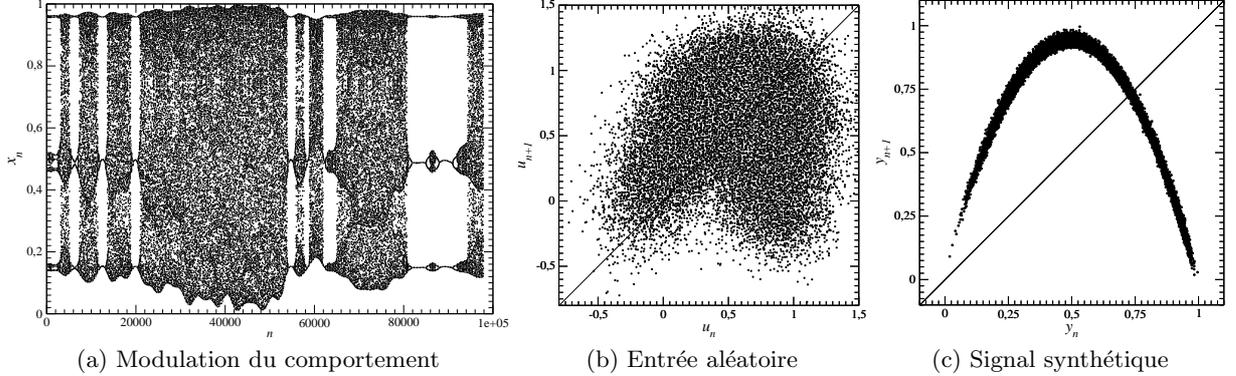


Fig.6. Modulation du comportement de la fonction logistique par une dérive appliquée sur le paramètre μ . La dérive est celle extraite de l'électrocardiogramme avec un coefficient de normalisation pour maintenir μ sur l'intervalle $[3.79 ; 3.99]$. L'entropie de Shannon estimée à partir des données synthétiques est de 1.92.

Si la structure algébrique de la fonction logistique est encore une fois clairement identifiée, bien que la valeur du paramètre soit moins bien estimée qu'en l'absence de modulation basse fréquence (3,6376 contre 3,7879 pour 3,89). Ceci provient du fait que le modèle est estimé sur une courte fenêtre (500 points) qui est ici choisie autour de $n = 60000$ (Fig.6a), soit une valeur pour laquelle la dynamique est moins développée ($\mu < 3,89$). Toutefois, lorsque les deux termes prépondérants sont regroupés sous la forme (4), il est possible de faire varier le paramètre en facteur et de retrouver le diagramme de bifurcations de la fonction logistique. En injectant la modulation basse fréquence, nous retrouvons alors une évolution très semblable à celle obtenue Fig. 6a).

3 Application à un tachogramme sur 24 heures

Nous appliquons enfin notre procédure de modélisation au tachogramme de la Fig. 1a et dont l'application de premier retour est représentée Fig. 7b. Etant donné que la dynamique est maintenant non stationnaire, il est nécessaire de calculer l'entropie de Shannon à l'aide de la dynamique symbolique, qui permet l'utilisation de la totalité des 90000 points [5]. Une entropie assez faible (0,045) est obtenue. La composante aléatoire construite selon la procédure de McSharry et ses collègues présente une variabilité plus lente (Fig. 7a) mais plus développée comme le révèle l'entropie assez élevée (0,920). Le modèle NARMAX entrée-sortie présente une variabilité plus proche de celle du tachogramme, comme le montre l'entropie relativement voisine de celle du tachogramme (0,137 contre 0,045).

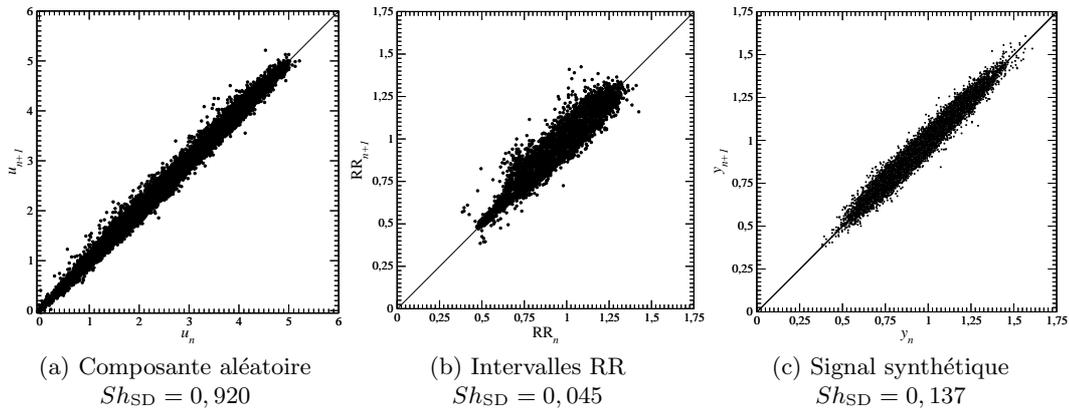


Fig.7. Application de premier retour calculée sur les intervalles RR du tachogramme de la Fig. 1.

La procédure de modélisation par bloc, en combinant des composantes déterministes et des composantes aléatoires permet d'aborder sous un nouvel angle la modélisation de l'activité cardiaque. Toutefois, cela reste difficile en raison des multiples échelles de temps en jeu. Nos premiers résultats montrent qu'un composante déterministe agit sur la variabilité cardiaque. Un travail plus affiné sur la validation devrait permettre de mieux identifier le rôle et l'importance de cette composante.

Remerciements : Le séjour de Dair est financé par le CAPES. La collaboration est partiellement financée par le CNRS et le CNPq.

Références

1. R. W. DE BØER, J. M. KAREMAKER & J. STRACKEE, Hemodynamic fluctuations and baroreflex sensitivity in human : a beat-to-beat model, *American Journal of Physiology*, **253**, 680-689 (1987).
2. P. E. MCSHARRY, G. CLIFFORD, L. TARASSENKO, L. A. SMITH, Method for Generating an Artificial RR Tachogram of a Typical Healthy Human over 24-Hours, *Computers in Cardiology IEEE*, **29**, 225-228 (2002).
3. L. A. AGUIRRE & S. A. BILLINGS, Identification of models for chaotic systems from noisy data : implications for performance and nonlinear filtering, *Physica D*, **85**, 239-258 (1995).
4. J. GODELLE & C. LETELLIER, Symbolic sequence statistical analysis for free liquid jets, *Physical Review E*, **62** (6), 7973-7981, 2000.
5. H. RABARIMANANTSOA, C. LETELLIER, L. ACHOUR, A. CUVELIER & J.-F. MUIR, Diagrammes de récurrence et entropie de Shannon pour l'étude des asynchronismes en ventilation non invasive, 9^{ème} Rencontres du Non Linéaire, 8-10 Mars 2006, ce volume.