

Biais dans l'analyse de dynamiques lentes/rapides à forte composante linéaire : cas de la détection de non-linéarité

Elise Roulin¹, Ubiratan Santos Freitas² & Christophe Letellier¹

¹ CORIA UMR 6614 — Université de Rouen, BP. 12, 76801 Saint-Etienne du Rouvray cedex

² ADIR Association — EA GRHV 3830, Hôpital de Bois-Guillaume, France

roulin@coria.fr

Résumé. La détection de non-linéarité par prédiction à un pas en avant repose sur l'avantage à utiliser un modèle non-linéaire plutôt qu'un modèle linéaire. Lorsque la probabilité p d'avoir de meilleurs résultats avec un modèle non-linéaire est supérieure à 0,99, un titrage du bruit peut être appliqué. Lorsque $p \approx 0,5$, le modèle non-linéaire n'est pas plus — ni moins — performant que le modèle linéaire. Nous traitons ici en détails le cas où $p \approx 0$, c'est-à-dire le cas où le modèle linéaire est significativement plus performant que le modèle non-linéaire. Une dynamique caricaturale correspondante à ce cas est construite. Nous proposons ainsi une explication possible à des résultats obtenus pour des situations de fibrillations auriculaires lors d'une étude précédente de dynamiques cardiaques.

Abstract. Detecting nonlinearities by one-step-ahead prediction is based on the advantage to use a nonlinear model rather than a linear model. When the probability p to get better results with a nonlinear model is greater than 0.99, a noise titration can be applied. When $p \approx 0.5$, the nonlinear model was not more performant than the linear model. We here investigate in details the case $p \approx 0$, that is, when the linear model is significantly more performant than the linear model. A caricature corresponding to this case is built. We thus explain results obtained in a previous study where atrial fibrillation was found with nearly zero p -value.

1 Introduction

Dans le domaine de la biomédecine, les enjeux liés à l'identification d'un comportement chaotique à partir d'une série temporelle expérimentale sont nombreux, aussi bien d'un point de vue physiologique que thérapeutique. Toutefois, malgré maintes tentatives, la preuve irréfutable de la présence de chaos au sein de dynamiques biomédicales n'a jamais été apportée [1], car c'est la propriété de sensibilité aux conditions initiales qui est le plus souvent testée (plus grand exposant de Lyapunov, etc.) alors que la question — nécessaire — du déterminisme est rarement abordée. À ce jour, la seule preuve satisfaisante de l'existence d'un déterminisme sous-jacent serait l'obtention d'un modèle global reproduisant la dynamique et vérifiant par ailleurs la sensibilité aux conditions initiales [2], mais un tel modèle n'a encore jamais été obtenu à partir d'enregistrements biologiques [3].

Pour pallier ce manque, Poon & Barahona ont proposé une technique de titrage du bruit, en affirmant pouvoir quantifier la chaoticité de la dynamique sous-jacente à partir d'une série temporelle [4]. Malheureusement, dans certains cas, cette technique se révèle incapable de distinguer une dynamique aléatoire d'une dynamique déterministe [5]. Elle ne peut donc pas être utilisée comme indicateur de présence de chaos. Toutefois, elle peut contribuer à l'identification d'une composante non-linéaire. Elle a ainsi permis d'analyser, à partir d'enregistrements électrocardiographiques de patients souffrant de différentes pathologies cardiaques, l'impact des types de pathologies sur la non-linéarité du rythme cardiaque [3]. Cette technique a également été utilisée pour distinguer des dynamiques ventilatoires spontanées [6] ou assistées mécaniquement [7]. Néanmoins, cette technique doit être appliquée avec certaines précautions relatives à l'échantillonnage de la série temporelle analysée, le paramétrage des modèles utilisés pour la prédiction, le choix de l'observable, etc. Ces précautions seront explicitées dans la section 2.

Parmi les dynamiques biologiques ou biomédicales observées, certaines peuvent se ramener à des oscillations de relaxation [10]. Ceci a été observé à l'échelle cellulaire, que ce soit au niveau de la cellule

unique [12,13] ou des niveau des tissus [14], mais également à l'échelle de l'organisme, considérant les rythmes physiologiques [8,9] ou biologiques [11], ou encore à un niveau comportemental, macroscopique et collectif [15]. Ces oscillations présentent, la plupart du temps, une augmentation rapide de l'amplitude, suivie d'une relaxation lente de retour à la valeur de base. Typiquement, la dynamique lente peut être reproduite par un processus linéaire, tandis que la dynamique rapide est contrôlée par un processus non-linéaire, pilotant notamment les fluctuations de l'amplitude. Ces comportements sont délicats à analyser de par leur structure dynamique intrinsèque. Afin de pouvoir mieux comprendre ces phénomènes, nous construisons une dynamique artificielle, reposant sur des oscillations périodiques dont l'amplitude est modulée par une fonction logistique. La série temporelle résultante se présente comme un cas « pathologique » du point de vue du traitement du signal, dans la mesure où la non-linéarité agit par impulsions très brèves. Nous montrons ici que la technique de détection de non-linéarité [4] échoue à détecter la composante non-linéaire de la dynamique ainsi construite.

2 Principe de la détection de non-linéarité

La technique de titrage du bruit [4] est conditionnée par une détection de non-linéarité, basée sur une estimation de modèles polynomiaux auto-régressifs utilisés pour la prédiction à un pas en avant. Précisons que, de manière générale, la validation par prédiction à un pas en avant souffre de certaines faiblesses comme cela a été récemment montré [16]. Suivant cette détection, et seulement si les données révèlent effectivement une composante non-linéaire, le titrage du bruit est appliqué. Ces deux étapes sont maintenant détaillées.

2.1 Détection de non-linéarité

Une série temporelle $\{y_k\}_{k=1}^N$ ($k \in \mathbb{N}$) est analysée à l'aide d'une part, de modèles linéaires et, d'autre part, de modèles non-linéaires. Ces modèles n'ont pas pour vocation de reproduire la dynamique globale mais seulement d'être optimaux pour une prédiction à un pas en avant. Chacun de ces modèles est estimé sur une fenêtre de la série temporelle. La forme générale de ces modèles est la suivante :

$$\begin{aligned} \tilde{y}_n(d, M) &= a_0 + a_1 y_{n-1} + a_2 y_{n-2} + \dots + a_{\kappa+1} y_{n-1}^2 + a_{\kappa+2} y_{n-1} y_{n-2} + \dots + a_{M-1} y_{n-\kappa}^d \\ &= \sum_{m=0}^{M-1} a_m z_m(n) \end{aligned} \quad (1)$$

où $\{z_m(n)\}_{m=1}^M$, $M \in \mathbb{N}$, est la base fonctionnelle composée de toutes les combinaisons distinctes des coordonnées décalées $\{y_{n-k}\}_{k=1}^{\kappa}$, $\kappa \in \mathbb{N}$, jusqu'au degré maximum de non-linéarité permis d . Il y a alors $M = \frac{(\kappa+d)!}{(d!\kappa!)}$ termes. Des deux paramètres d et M découle donc l'ordre κ du polynôme utilisé pour la modélisation. Pour un modèle linéaire, $d = 1$, ce qui implique $\kappa = M - 1$.

Ainsi, chaque famille de modèles peut être décrite univoquement avec les paramètres M — nombre maximal d'éléments de la base fonctionnelle — et d — degré de non-linéarité maximal des modèles de la famille. Les M termes ne contribuent pas tous de la même façon à reproduire la dynamique et, même si l'ajout de termes supplémentaires au développement polynomial contribue en général à améliorer la qualité prédictive du modèle, il est possible qu'un trop grand nombre de termes finisse par réduire cette qualité [2]. La recherche du modèle optimal parmi ceux de la « famille » définie par les paramètres M et d commence donc par la sélection des termes les plus pertinents, selon la méthode de minimisation de l'erreur de prédiction

$$\varepsilon(M, d)^2 = \frac{\sum_{n=1}^N (\tilde{y}_n(M, d) - y_n)^2}{\sum_{n=1}^N (y_n - \bar{y})^2} \quad (2)$$

où $\bar{y} = 1/N \sum_{n=1}^N y_n$, et où $\varepsilon(M, d)^2$ est une variance normalisée des erreurs résiduelles. L'inclusion d'un $m^{\text{ème}}$ terme au modèle (1) induit une réduction de ε^2 ; réduction qui peut être exprimée comme un pourcentage de l'erreur maximale $\max(\varepsilon^2)$, conduisant au rapport de réduction d'erreur (ou *ERR pour*

error reduction ratio) [19]. Les coefficients a_m sont ensuite estimés en utilisant la méthode des moindres carrés dans le but de minimiser l'erreur de prédiction ϵ^2 . À l'issue de ce processus, nous disposons d'un modèle polynomial auto-régressif linéaire, et d'un second modèle qui est non-linéaire, les deux modèles étant paramétrés par un même couple (M, d) .

La probabilité p pour que le meilleur modèle non-linéaire soit plus performant pour décrire la dynamique à un pas en avant que le meilleur modèle linéaire est calculée. Si la détection de non-linéarité est concluante, en autorisant une marge d'erreur de 1%, c'est-à-dire si la valeur de p est plus grande que 0,99, le titrage de bruit est appliqué.

2.2 Titrage du bruit

Le titrage en lui-même utilise une distribution gaussienne de bruit blanc pour quantifier la contribution de la composante non-linéaire à la dynamique sous-jacente aux données expérimentales. Pour titrer le bruit d'une série temporelle, un bruit blanc ν_n , de même déviation standard que y_n et d'amplitude A ($0 \leq A \leq 1$) croissante, est ajouté aux données, jusqu'à ce que la non-linéarité ne soit plus détectée, à un certain degré de confiance près. Tant que la détection de non-linéarité demeure concluante, le paramètre A est augmenté jusqu'à ce que la valeur de p passe sous le seuil préalablement défini de 0,99. Le titrage est ici terminé et la déviation standard du bruit blanc alors ajouté correspond à la limite de bruit de l'échantillon testé (LB).

2.3 Recommandations pour une utilisation optimale

Nous avons précédemment montré que les résultats obtenus avec la détection de non-linéarité et le titrage du bruit dépendaient fortement de certains choix relatifs à l'obtention des modèles d'une part, et à la série temporelle analysée, d'autre part [20]. Ainsi l'ensemble de la gamme des paramètres d et M ne peut être utilisée. En effet, la structure des modèles ne doit être ni trop petite, ni trop grande, faute de quoi les résultats risquent d'être biaisés. Aussi, le degré de non-linéarité doit être au moins de 3 et le nombre de termes au moins égal à 20 ou 30. Il est raisonnable de ne pas dépasser une centaine de termes constituant le modèle. Par ailleurs, la limite de bruit est sensible à la réalisation du bruit utilisé pour le titrage; aussi, il est nécessaire de considérer une valeur moyenne (correctement définie du point de vue statistique) de la limite de bruit [20].

Si plusieurs variables sont disponibles pour l'étude d'une dynamique, le résultat peut dépendre de la variable choisie [17] et du taux d'échantillonnage utilisé [20]. Enfin, lorsque la probabilité p est voisine de 0,50, ceci indique que le choix entre un modèle linéaire et un modèle non-linéaire n'est pas évident, et que la pertinence de ces modèles est à peu près équivalente. Lorsqu'elle est supérieure à 0,99, l'utilisation d'un modèle non-linéaire conduit à de meilleures prédictions qu'un modèle linéaire, et l'on peut conclure que la dynamique sous-jacente présente une composante non-linéaire. Lorsqu'elle est au voisinage de 0, en principe, la dynamique sous-jacente est principalement linéaire, mais cela n'a jamais été vérifié. C'est ce que nous allons tester dans ce qui suit.

3 Dynamique lente/rapide à modulation chaotique d'amplitude

3.1 Motivation

Le recours à la détection de non-linéarité pour l'analyse de dynamiques cardiaques a permis de révéler des dynamiques de nature très différentes entre des patients souffrant de fibrillation auriculaire ($p \approx 0$), d'insuffisance cardiaque congestive ($p \approx 1$) et un groupe de sujets sains ($p \approx 0.75$) [3]. De manière surprenante, la détection de non-linéarité indiquait que la dynamique cardiaque était très fortement linéaire chez les patients souffrant de fibrillation auriculaire, comme le révélait le p voisin de zéro (Fig. 1). Aucun autre exemple — y compris impliquant des cycles limites — n'avaient conduit à des valeurs de p aussi faibles.

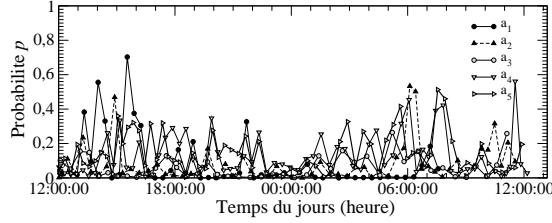


FIG. 1. Probabilités p calculées à partir des $\Delta RR_n = RR_{n+1} - RR_n$ pour les 5 patients souffrant de fibrillation auriculaire. Paramètres de modélisation : $(d, M) = (3, 50)$. D'après [3].

3.2 Le modèle

On considère ici la série temporelle correspondant à un comportement périodique, modulé en amplitude par une fonction logistique itérée en régime chaotique. Le comportement périodique se ramène à un signal triangulaire pour lequel la décroissance est linéaire et lente par rapport à la croissance qui est ici très rapide puisque l'amplitude maximale est atteinte en un seul pas de temps (une itération). L'amplitude maximale du $i^{\text{ème}}$ cycle est donnée par

$$A_i = \frac{n}{20} \left(1 + \frac{y_i}{10}\right) \quad (3)$$

où n définit le nombre d'itérations dans une oscillation et y_i est une solution de la fonction logistique

$$y_{n+1} = \mu y_n (1 - y_n). \quad (4)$$

La série temporelle $\{x_k\}$ est alors construite selon le processus par morceaux

$$x_{k+1} = \begin{cases} x_k - \frac{A_i}{n} & \text{si } x_k > 0 \\ A_i & \text{si } x_k \leq 0. \end{cases} \quad (5)$$

où $i = \lfloor \frac{k}{n} \rfloor$. À chaque cycle i correspond une seule valeur A_i ; la période des oscillations demeure donc constante, et ne dépend pas de la valeur de l'amplitude, qui elle, varie à chaque cycle.

Une série temporelle typique est représentée Fig. 2.

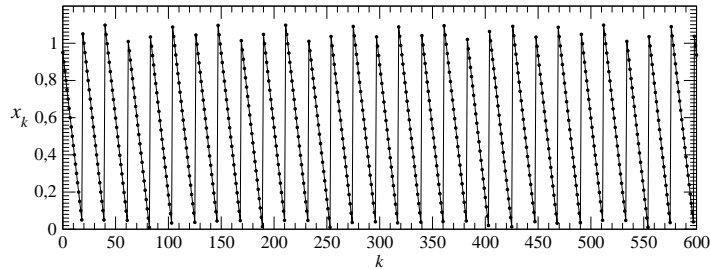


FIG. 2. Série temporelle de la dynamique lente-rapide à modulation chaotique d'amplitude. Valeurs des paramètres : $\mu = 3,9$ et $n = 20$.

Cette série temporelle est gouvernée la plupart du temps par un processus linéaire défini par la première équation du processus (5). Sur de très brefs instants, le processus non-linéaire déterminant les fluctuations de l'amplitude agit. Du point de vue de la série temporelle, le processus est principalement linéaire. Toutefois, si une application de premier retour à la section de Poincaré prise en tout point du signal est calculée, nous retrouvons une parabole caractéristique de la fonction logistique (Fig. 3). De ce fait, nous nous attendons à ce que l'analyse d'une série temporelle $\{x_k\}$, solution de ce processus de relaxation, conduise à l'identification d'une dynamique purement linéaire.

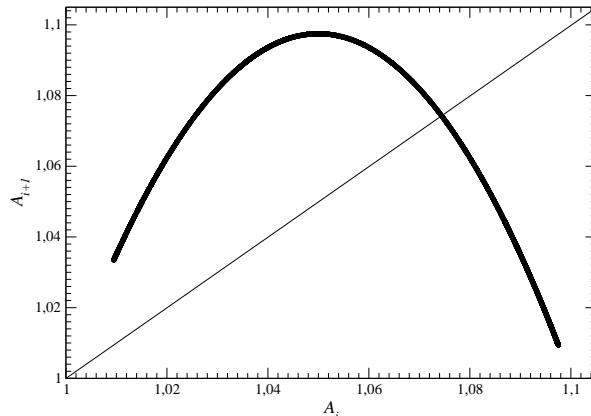


FIG. 3. Série temporelle de la dynamique lente-rapide à modulation chaotique d'amplitude. Valeurs des paramètres : $\mu = 3,9$ et $n = 20$.

3.3 Résultats

La technique de détection de non-linéarité décrite section 2 est alors appliquée à la série temporelle $\{x_k\}$. Les paramètres de modélisation sont fixés à $(d, M) = (3, 50)$, et une moyenne sur 5 détections a été réalisée pour chaque valeur de p . Nous observons que les valeurs de la probabilité p pour que le meilleur modèle non-linéaire soit plus performant pour la prédiction à un pas en avant que le meilleur modèle linéaire sont toujours nulles, et ce, quel que soit le nombre α d'oscillations de relaxation du signal contenues dans la fenêtre considérée. Nous avons effectué les tests pour $n = 20$ et $\alpha \geq 12$, soit une fenêtre d'environ 240 points pour l'estimation des modèles. Même pour des valeurs du nombre α d'oscillations à 30, la valeur de p reste à zéro. Ceci signifie que la composante non-linéaire agit trop sporadiquement pour qu'elle puisse être détectée sur la base d'une erreur de prédiction à un pas en avant. Précisons que typiquement, un modèle linéaire ne fournit une mauvaise prédiction que lors d'un retour brutal à l'amplitude maximale, c'est-à-dire pour un point sur n . Plus n est grand, moins cette erreur pèse sur la statistique. De ce fait, puisqu'avec $n = 20$, la technique échoue déjà à détecter la composante non-linéaire, *a fortiori*, ceci reste toujours vrai pour des n plus grands.

La série temporelle que nous avons construite se présente alors comme un cas « pathologique », dans la mesure où, la non-linéarité agissant par impulsions très brèves, la technique de détection de non-linéarité échoue à détecter la composante non-linéaire pourtant clairement présente dans la dynamique, et ne met en évidence que le processus linéaire. Notons cependant que si le problème est abordé du point de vue d'une application de premier retour, la probabilité p calculée présente une valeur toujours égale à 1, comme c'est le cas lors d'une étude de la fonction logistique, révélant alors clairement la présence d'une composante non-linéaire sous-jacente à la dynamique. Cet exemple, certes caricatural, montre qu'il n'est pas équivalent de travailler à partir d'une série temporelle représentative de l'évolution du système dans l'espace des phases, et de travailler avec une série temporelle associée à une section de Poincaré ou une application de premier retour. Une telle différence avait déjà été relevée dans le cas d'une analyse par entropie de Shannon [17].

4 Conclusion

Outre les précautions d'utilisation déjà mentionnées de la détection de non-linéarité, nous montrons ici que la recherche d'une composante non-linéaire par détection de non-linéarité selon la technique du titrage du bruit échoue lorsque la série temporelle considérée résulte d'une dynamique où la non-linéarité agit par impulsions brèves. Notre dynamique caricaturale se révèle être un excellent cas d'école pour tester la robustesse de toute technique d'analyse. Par ailleurs, la non-équivalence entre une analyse selon une trajectoire dans l'espace des phases ou selon une section de Poincaré est encore confirmée ici, pour

l’outil de détection de non-linéarité associé au titrage du bruit. Aussi, nous concluons naturellement que l’étude de toute dynamique devrait se faire préférentiellement dans une section de Poincaré.

Références

1. L. GLASS, Introduction to controversial topics in nonlinear science : Is the heart rate chaotic?, *Chaos*, **19**, 028501 (2009).
2. C. LETELLIER, L. A. AGUIRRE & U. S. FREITAS Frequently Asked Questions about global modelling, *Chaos*, **19**, 023103, 2009.
3. U. S. FREITAS, E. ROULIN, J.-F. MUIR & C. LETELLIER, Identifying determinism underlying heart rate : the right task?, *Chaos*, **19**, 028505 (2009).
4. C.-S. POON & M. BARAHONA, Titration of chaos with added noise, *Proceedings of the National Academy of Sciences (USA)*, **98**, 7107-7112 (2001).
5. U. S. FREITAS, E. ROULIN & C. LETELLIER, Failure for distinguishing colored noise from chaos by the “Noise titration” technique, *Physical Review E*, **79**, 035201 (2009).
6. M.-N. FIAMMA, C. STRAUS, S. THIBAUT, M. WYSOCKI, P. BACONNIER & T. SIMILOWSKI, Effects of hypercapnia and hypocapnia on ventilatory variability and the chaotic dynamics of ventilatory flow in humans, *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, **292**, R1985-R1993 (2007).
7. L. MANGIN, M.-N. FIAMMA, C. STRAUS, J.-P. DERENNE, M. ZELTER, C. CLERICI & T. SIMILOWSKI, Source of human ventilatory chaos : Lessons from switching controlled mechanical ventilation to inspiratory pressure support in critically ill patients, *Respiratory Physiology & Neurobiology*, **161** (2), 189-196 (2008).
8. B. VAN DER POL, The heartbeat considered as a relaxation oscillation and an electrical model of the heart, *Philosophical Magazine*, **6**, 763-775 (1928).
9. B. VAN DER POL, Biological rhythms considered as relaxation oscillations, *Acta Medica Scandinavica*, **103** (S108), 76 (1940).
10. K. KLOTTER, General properties of oscillating systems, *Cold Spring Harbor Symposia on Quantitative Biology*, **25**, 185-187 (1960).
11. J. S. BARLOW, A phase-comparatormodel for the diurnal rhythm of emergence of *Drosophila*, *Annals of the New-York Academy of Science*, **98**, 788-805 (1962).
12. J. TYSON & S. KAUFFMAN, Control of mitosis by a continuous biochemical oscillation : synchronisation : spatially inhomogeneous oscillations, *Journal of Mathematical Biology*, **1**, 289-310 (1975).
13. M. R. GUEVARA, L. GLASS & A. SHRIER, Phase locking, period-doubling bifurcations, and irregular dynamics in periodically stimulated cardiac cells, *Science*, **214**, 1350-1353 (1981).
14. J. P. KEENER, F. C. HOPPENSTEADT & J. RINZEL, Integrate-and-fire models of nerve membrane response to oscillatory input, *SIAM Journal of Applied Mathematics*, **41**, 503-517 (1981).
15. W. LIU, D. XIAO & Y. YI, Relaxation oscillations in a class of predator-prey systems, *Journal of Differential Equations*, **188**, 306-331 (2003).
16. M. P. DAFILIS, N. C. SINCLAIR, P. J. CADUSCH & D. T. J. LILEY, Re-evaluating the performance of the nonlinear prediction error for the detection of deterministic dynamics, *Physica D*, **240**, 695-700 (2011).
17. C. LETELLIER, Estimating the Shannon entropy : recurrence plots versus symbolic dynamics, *Physical Review Letters* **96**, 254102 (2006).
18. M. BARAHONA & C.-S. POON, Detection of nonlinear dynamics in short noisy time series, *Nature*, **381**, 215-217 (1996).
19. L. A. AGUIRRE & S. A. BILLINGS, Improved structure selection for nonlinear models based on term clustering, *International Journal of Control*, **62** (3), 569-587 (1995).
20. E. ROULIN, U. S. FREITAS & C. LETELLIER, Guidelines for improving detection of nonlinearity and noise titration, *Physical Review E*, à paraître.