

FRÉDÉRIC BEAULIEU

**RÉDUCTION NON-LINÉAIRE DU BRUIT  
ET  
DÉTECTION DE NON-STATIONNARITÉ**

Mémoire  
présenté  
à la Faculté des études supérieures  
de l'Université Laval  
pour l'obtention  
du grade de maître ès sciences (M.Sc.)

Département de physique  
FACULTÉ DES SCIENCES ET DE GÉNIE  
UNIVERSITÉ LAVAL

AVRIL 2000

# Résumé

Le bruit et la non-stationnarité d'une série temporelle observée peuvent limiter l'information que l'on peut extraire à propos de la dynamique du phénomène observé. À cet effet, d'une part, deux méthodes non-linéaires de réduction du bruit, celles de Schreiber et GHKSS, sont étudiées systématiquement au moyen de signaux contaminés artificiels. On montre que celles-ci fonctionnent mieux dans le cas d'un bruit de mesure que dans le cas d'un bruit dynamique. D'autre part, en ce qui concerne la non-stationnarité, plusieurs méthodes de détection sont présentées, dont une toute nouvelle basée sur le flot d'information (version cumulant) entre le passé et le futur. Le potentiel et la robustesse de cette dernière sont établis au moyen de cinq exemples artificiels, dont notamment l'intermittence. Enfin, le problème qui consiste à prédire l'avènement d'une crise épileptique à partir d'un électroencéphalogramme est réexaminé du point de vue de cette nouvelle méthode.

# Avant-propos

L'auteur tient à remercier tout spécialement le directeur de ses travaux de recherche, en l'occurrence le professeur Louis J. Dubé, pour son enthousiasme contagieux, sa franche cordialité, sa grande compétence, son remarquable professionnalisme et enfin une première chance accordée dans le monde de la recherche (été 1996). Ce fut une source d'inspiration constante au cours de ces quatre années d'interaction.

L'auteur est également très reconnaissant envers le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), pour l'octroi d'une bourse d'études supérieures ÉS A. Ce support financier a rendu possible la réalisation de ce projet de recherche dans les meilleures conditions.

Enfin, il serait bien injuste de passer sous silence le plaisir enrichissant de côtoyer d'une part les autres membres du groupe de recherche, soit Philippe Després, Bernard Doyon et Babak Pourbohloul, et d'autre part les nombreux amis et professeurs qui travaillent au département de physique de l'Université Laval, pour n'en nommer que quelques uns : Nicolas Hô, Patrick Pilot, Pierre Mathieu, etc...

# Table des matières

<b>Résumé</b>	<b>1</b>
<b>Avant-propos</b>	<b>2</b>
<b>Introduction</b>	<b>10</b>
<b>1 Recouvrement : coordonnées à délai</b>	<b>14</b>
1.1 Théorie . . . . .	15
1.2 Pratique . . . . .	18
<b>2 Réduction non-linéaire du bruit</b>	<b>21</b>
2.1 Signaux et bruits artificiels . . . . .	22
2.2 Critères de performance . . . . .	25
2.3 Méthode de Schreiber . . . . .	26
2.3.1 Description de la méthode . . . . .	26
2.3.2 Résultats . . . . .	27
2.4 Méthode GHKSS . . . . .	29
2.4.1 Description de la méthode . . . . .	29
2.4.2 Résultats . . . . .	32

<i>Table des matières</i>	4
<b>3 Non-stationnarité</b>	<b>51</b>
3.1 Importance du problème . . . . .	51
3.2 Survol de la littérature . . . . .	54
3.2.1 Kennel . . . . .	54
3.2.2 Witt, Kurths et Pikovsky . . . . .	56
3.2.3 Yu, Lu et Harrison . . . . .	56
3.2.4 Manuca et Savit . . . . .	57
3.2.5 Schreiber . . . . .	60
3.2.6 Cao, Mees et Judd . . . . .	61
3.2.7 Hively, Gailey et Protopopescu . . . . .	62
<b>4 Flot d'information</b>	<b>64</b>
4.1 Flot d'information d'un signal stationnaire . . . . .	64
4.1.1 Approche formelle . . . . .	64
4.1.2 Approche pratique . . . . .	68
4.2 Mesure de non-stationnarité . . . . .	75
4.3 Exemples artificiels . . . . .	76
4.3.1 Dynamique segmentée . . . . .	77
4.3.2 Application triangulaire asymétrique . . . . .	80
4.3.3 Application du boulanger généralisée . . . . .	82
4.3.4 Processus auto-régressif du premier ordre AR(1) . . . . .	84
4.3.5 Intermittence . . . . .	86
4.4 Précurseur d'une crise d'épilepsie . . . . .	89
4.4.1 Modèle du contrôleur épileptique . . . . .	91
4.4.2 Données épileptiques de Schiff . . . . .	93

<i>Table des matières</i>	5
<b>Conclusion</b>	<b>99</b>
<b>A Réduction du bruit : réseau de neurones</b>	<b>110</b>
<b>B Matrice de covariance</b>	<b>113</b>
<b>C Implémentation du déplacement de Bernoulli</b>	<b>118</b>
<b>D Fonction d'autocorrélation</b>	<b>120</b>
<b>E Contrôle du chaos en 1D</b>	<b>125</b>

# Liste des tableaux

2.1	Méthode de Schreiber – Application logistique – Bruit de mesure . . .	37
2.2	Méthode GHKSS – Application logistique – Bruit de mesure . . . . .	37
2.3	Méthode de Schreiber – Application logistique – Bruit dynamique . .	40
2.4	Méthode GHKSS – Application logistique – Bruit dynamique . . . . .	40
2.5	Méthode de Schreiber – Application de Hénon – Bruit de mesure . . .	44
2.6	Méthode GHKSS – Application de Hénon – Bruit de mesure . . . . .	44
2.7	Méthode de Schreiber – Application de Hénon – Bruit dynamique . .	47
2.8	Méthode GHKSS – Application de Hénon – Bruit dynamique . . . . .	47
2.9	Méthode de Schreiber – Application de Mackey-Glass – Bruit de mesure	50
2.10	Méthode GHKSS – Application de Mackey-Glass – Bruit de mesure .	50
4.1	Écart-type $\sigma_\eta$ du bruit dynamique selon l'intervalle de la Fig. 4.12. . .	92

# Table des figures

2.1	Spectres en puissance illustrant les propriétés d'un signal chaotique (attracteur de Hénon) contaminé . . . . .	23
2.2	Filtrages non-linéaires – Application logistique – Bruit de mesure gaussien de 2.5% . . . . .	35
2.3	Filtrages non-linéaires – Application logistique – Bruit de mesure uniforme de 10% . . . . .	36
2.4	Filtrages non-linéaires – Application logistique – Bruit dynamique uniforme de 1% . . . . .	38
2.5	Filtrages non-linéaires – Application logistique – Bruit dynamique gaussien de 3% . . . . .	39
2.6	Filtrages non-linéaires – Attracteur de Hénon – Bruit de mesure uniforme de 1% . . . . .	41
2.7	Filtrages non-linéaires – Attracteur de Hénon – Bruit de mesure gaussien de 5% . . . . .	42
2.8	Pour la réduction de bruit GHKSS de la Fig. 2.7d, variation du niveau de bruit et de $r_{\text{dyn}}$ au fil des itérations. . . . .	43
2.9	Pour la réduction de bruit GHKSS de la Fig. 2.7d, variation du nombre moyen de voisins et de la valeur rms des corrections scalaires au fil des itérations. . . . .	43

<i>Table des figures</i>	8
2.10 Filtrages non-linéaires – Attracteur de Hénon – Bruit dynamique gaussien de 2% . . . . .	45
2.11 Filtrages non-linéaires – Attracteur de Hénon – Bruit dynamique uniforme de 4% . . . . .	46
2.12 Filtrages non-linéaires – Application de Mackey-Glass – Bruit de mesure gaussien de 2.5% . . . . .	48
2.13 Filtrages non-linéaires – Application de Mackey-Glass – Bruit de mesure uniforme de 10% . . . . .	49
4.1 Graphes à délai de la dynamique segmentée . . . . .	77
4.2 Flot d'information $m(10, r)$ vs $r$ de la dynamique segmentée . . . . .	79
4.3 Pour la dynamique segmentée, décomposition de la mesure $m(10, 1)$ en ses trois ordres . . . . .	80
4.4 Pour la dynamique segmentée, performance de la mesure de non-stationnarité en fonction de $n$ . . . . .	81
4.5 Pour la dynamique segmentée, performance de la mesure de non-stationnarité en fonction de $l_w$ . . . . .	81
4.6 Pour la dynamique segmentée, performance de la mesure de non-stationnarité en fonction de $t_w$ . . . . .	82
4.7 Mesure de non-stationnarité pour deux séries temporelles issues de l'application triangulaire asymétrique, l'une stationnaire, l'autre non stationnaire. . . . .	83
4.8 Mesure de non-stationnarité pour deux séries temporelles issues de l'application du boulanger généralisée, l'une stationnaire, l'autre non stationnaire. . . . .	84

4.9	Mesure de non-stationnarité pour deux séries temporelles issues d'un processus AR(1), l'une stationnaire, l'autre non stationnaire. . . . .	85
4.10	(a) Mesure de non-stationnarité pour deux séries temporelles très courtes issues d'un processus AR(1), l'une stationnaire, l'autre non stationnaire. (b) $m_{\text{avg}}(\text{NS}) - m_{\text{avg}}(\text{S})$ . . . . .	87
4.11	Mesure de non-stationnarité pour une série temporelle issue de l'application logistique où se manifeste un chaos intermittent. . . . .	88
4.12	Modèle du contrôleur épileptique et mesure de non-stationnarité . . .	93
4.13	Mesure de non-stationnarité pour les ISIs du patient ME . . . . .	95
4.14	Zoom de la Fig. 4.13. . . . .	96
4.15	Mesure de non-stationnarité pour les ISIs du patient RO1 . . . . .	97
4.16	Zoom de la Fig. 4.15. . . . .	97
A.1	Flot d'information $m(10, r)$ vs $r$ . . . . .	111
A.2	Attracteur de Hénon : (a) contaminé par un bruit dynamique gaussien, (b) filtré au moyen d'un réseau de neurones. . . . .	112
D.1	Fonctions d'autocorrélation numériques et théoriques de la dynamique segmentée . . . . .	124

# Introduction

En science, il est reconnu que l'observation constitue une activité fondamentale, et il est fréquent que celle-ci se résume à la mesure d'une observable à intervalles de temps réguliers. Dans ces conditions courantes d'observation, l'information que l'on peut extraire à propos de la dynamique du phénomène observé est entièrement contenue dans une série temporelle. Le problème consiste alors à récupérer cette information. Les exemples d'une telle situation foisonnent, et proviennent de domaines très variés : suivi de l'intensité d'un faisceau laser au cours du temps, évolution de la concentration d'une substance lors d'une réaction chimique, enregistrement d'un électroencéphalogramme, progression d'un indice sur un marché boursier, etc... Les enjeux qui gravitent autour de cette science de l'analyse des séries temporelles sont donc très importants.

Cette problématique n'est certes pas nouvelle. Des méthodes traditionnelles sont à la disposition de la communauté scientifique, que l'on pense à l'analyse de Fourier ou encore à l'analyse en termes de processus stochastiques. Ainsi, lorsqu'une superposition de fréquences est à l'origine d'un signal, la préséance de l'analyse de Fourier (ou analyse linéaire) s'avère indiscutable. D'autre part, si le système sous observation est dominé par une influence externe aléatoire, un processus stochastique tel un modèle ARMA [GSS91] suffira à décrire efficacement le signal mesuré.

Au début des années 80, de nouvelles méthodes ont fait leur apparition. Ces méthodes, issues de la dynamique non-linéaire et de la théorie du chaos déterministe,

ont considérablement élargi l'horizon de cette science de l'analyse des séries temporelles (Réfs. [GSS91], [OSY94], [Sch99] et [HKS99]). Deux concepts importants expliquent cette introduction de la dynamique non-linéaire, de concert évidemment avec la croissance exponentielle de la puissance informatique :

1. *Un système chaotique de basse dimension peut être à l'origine de signaux complexes*, qui s'apparentent à des signaux stochastiques. Cette similitude est particulièrement importante dans l'espace de Fourier et au niveau de la prédiction à moyen terme et à long terme.<sup>1</sup> En d'autres termes, un signal compliqué ne provient pas nécessairement d'un système compliqué de très haute dimension et/ou d'une source stochastique.
2. *À partir de la mesure simultanée de seulement quelques observables indépendantes d'un système déterministe, voire d'une seule, il est possible de recouvrir l'espace des phases complet du système*, au sens que nous spécifierons bientôt (espace des phases reconstruit). C'est la théorie mathématique du recouvrement, dont les théorèmes sous-tendent de nombreuses applications pratiques de la dynamique non-linéaire.

Il importe toutefois de demeurer lucide relativement à ces nouvelles méthodes non-linéaires, dont certaines seront étudiées dans le présent mémoire. Les deux points précédents ont provoqué un certain engouement dans la communauté scientifique, en raison d'un renouveau des techniques disponibles. Le danger réside en l'identification hâtive du caractère chaotique d'un signal, en vue d'utiliser les outils à la mode (voir par exemple la Réf. [The95]). Il faut garder à l'esprit, comme le souligne Schreiber [Sch99], que l'ensemble des systèmes pouvant être à l'origine d'un signal est conceptuellement balayé par le degré de non-linéarité et par la proportion aléatoire/déterminisme, la composante déterministe étant grandement affectée par le nombre de degrés de liberté. Le chaos déterministe de basse dimension n'est

---

<sup>1</sup>La sensibilité aux conditions initiales, qui caractérise un système chaotique, ne ruine pas les prédictions à court terme [OSY94].

pas, a priori, l'alternative la plus probable. Néanmoins, si l'on adopte une attitude très pragmatique, et que le but visé consiste à réaliser un certain objectif et non nécessairement de tirer des conclusions définitives, alors rien n'interdit l'emploi du paradigme chaotique dans certaines circonstances.

Quoi qu'il en soit, peu importe le type d'analyse envisagé, toute série de données réelles, c.-à-d. provenant de l'observation, présente obligatoirement les quatre caractéristiques suivantes avec lesquelles il faudra composer :

1. C'est le résultat d'un échantillonnage (fréquence d'échantillonnage).
2. Il n'y a qu'un nombre fini de données.
3. Il y a présence de bruit, ce qui masque en partie la vraie dynamique, celle qui intéresse le scientifique.<sup>2</sup>
4. La série temporelle est potentiellement non stationnaire, c.-à-d. que ses propriétés dynamiques changent d'une manière effective au cours du temps.

Ces caractéristiques peuvent entraver sérieusement l'analyse, ou en d'autres termes, limiter la quantité d'information que l'on peut récupérer à propos de la dynamique du phénomène observé. L'effet limitatif des deux premières peut généralement être bien contrôlé par un ajustement adéquat des conditions d'observation, pour autant qu'une certaine latitude expérimentale le permette. À l'opposé, les deux dernières caractéristiques (bruit et non-stationnarité), beaucoup plus compromettantes, nécessiteront souvent un travail supplémentaire avant l'analyse. Et à ce niveau, les méthodes traditionnelles linéaires doivent être complètement repensées dans le cas des séries temporelles chaotiques.

---

<sup>2</sup>La précision nécessairement finie des instruments de mesure contribue à ce bruit (bruit de mesure). Et même si l'origine déterministe du signal ne fait aucun doute, il faut réaliser qu'un système n'est jamais parfaitement isolé, et ce malgré toutes les précautions que pourrait prendre un expérimentateur. L'influence désordonnée de l'environnement extérieur prend souvent la forme d'une faible injection stochastique dans la dynamique (bruit dynamique).

L'objet du présent mémoire consiste justement à présenter des méthodes, provenant ou inspirées de la dynamique non-linéaire, pour traiter efficacement le bruit et la non-stationnarité d'une série temporelle. Il s'agit donc, en un sens, d'une étude fondamentale pour l'analyse des séries temporelles, en particulier celles qui sont chaotiques.

Compte tenu de sa grande importance, le recouvrement au moyen de coordonnées à délai constitue le point de départ de ce mémoire. Ainsi, le chapitre premier discute de cette extraordinaire technique, qui est à la base de nombreuses méthodes présentées dans les chapitres suivants.

Au second chapitre, deux méthodes non-linéaires de réduction du bruit seront étudiées, soit les méthodes bien-établis de Schreiber [Sch93] et GHKSS [GHK<sup>+</sup>93]. L'objectif caché et ultime de cette étude consiste à situer les performances d'une nouvelle méthode de réduction du bruit par rapport aux algorithmes présentés dans ce chapitre. Cette nouvelle méthode est basée sur l'utilisation d'un réseau de neurones et a été proposée très récemment par Deco. Une introduction à cette approche novatrice est présentée à l'annexe A.<sup>3</sup> En préparation à cette éventuelle comparaison [DBDa], des données exclusivement artificielles ont été utilisées, afin d'être en mesure de jauger plus précisément les performances.

Au troisième chapitre, l'attention est portée vers l'épineux problème de la non-stationnarité d'une série temporelle. Suivant une discussion des subtilités entourant cette problématique, on y retrouve un survol des méthodes proposées dans la littérature pour détecter la non-stationnarité d'un signal. Enfin, au quatrième et dernier chapitre, une nouvelle approche au problème de la non-stationnarité est présentée, approche due à Dubé et coauteurs [DBDb] et basée sur le concept de flot d'information. Afin de valider cette toute récente méthode de détection, de nombreux exemples artificiels issus de la littérature sont étudiés, et l'application importante qui consiste à prédire l'avènement d'une crise épileptique à partir d'un électroencéphalogramme est reconsidérée du point de vue de cette nouvelle approche.

---

<sup>3</sup>Le quatrième chapitre doit être lu au préalable.

# Chapitre 1

## Recouvrement : coordonnées à délai

De loin l'idée la plus géniale dans le contexte de recouvrement est l'utilisation de coordonnées à délai, pour reconstruire l'espace des phases à partir d'un unique signal scalaire. En effet, d'un point de vue expérimental, les deux situations suivantes sont monnaie courante :

1. Tout ce dont l'on dispose est un signal scalaire.
2. Même s'il est possible pour un expérimentateur de mesurer simultanément plusieurs quantités, il s'avère très difficile, voire impossible, de s'assurer de l'indépendance des observables.<sup>1</sup>

Selon un papier de Packard et coauteurs [PCFS80] datant de 1980, Ruelle serait à l'origine de l'astuce des coordonnées à délai.<sup>2</sup> Quoi qu'il en soit, Takens [Tak81]

---

<sup>1</sup>À supposer que l'on puisse certifier l'indépendance, il demeure encore la possibilité où l'on ne dispose pas de suffisamment d'observables simultanées pour reconstruire pleinement l'espace des phases. Dans ce cas, la technique des coordonnées à délai, appliquée à chaque observable, peut permettre de combler l'information manquante.

<sup>2</sup>Dès 1979, Mackey et Glass utilisaient des graphes à délai pour visualiser les solutions de leur équation différentielle à délai.

fournissait en 1981 une preuve mathématique légitimant en quelque sorte une telle pratique.

## 1.1 Théorie

Nous allons examiner brièvement le théorème de Takens, dans une version ultérieure plus explicite donnée par Sauer et coauteurs [SYC91]. Nous discuterons ensuite de l'applicabilité de ce théorème dans un contexte expérimental. Rappelons d'abord la définition d'un système dynamique, définition restreinte ici aux systèmes de dimension finie, i.e. excluant les équations aux dérivées partielles :

**Définition 1 (Système dynamique de dimension finie)** *Soit  $U$  l'ensemble  $\mathcal{R}^d$  ou un sous-ensemble ouvert de  $\mathcal{R}^d$ , et soit  $G$  l'ensemble  $\mathcal{R}$  (équations différentielles ordinaires) ou l'ensemble  $\mathcal{Z}$  (équations aux différences). Un système dynamique consiste en un couple  $(U, \phi)$ , où  $U$  est appelé l'espace des phases, et  $\phi$  une application continue (règle générale un difféomorphisme) appelée le flot,*

$$\begin{aligned}\phi : U \times G &\rightarrow U \\ (\mathbf{x}, t) &\mapsto \phi(\mathbf{x}, t) = \phi_t(\mathbf{x})\end{aligned}$$

et vérifiant les deux propriétés suivantes :

$$\begin{aligned}\phi_{s+t}(\mathbf{x}) &= \phi_s(\phi_t(\mathbf{x})) \\ \phi_0(\mathbf{x}) &= \mathbf{x}.\end{aligned}$$

Également utile pour apprécier le théorème de recouvrement est la notion de capacité («box-counting dimension») d'une fractale, l'une des nombreuses variantes de la dimension fractale :

**Définition 2 (Capacité)** *Soit un sous-ensemble  $S \subset \mathcal{R}^d$ , et soit  $N(\epsilon)$  le nombre minimum d'hypercubes à  $n$  dimensions nécessaire pour couvrir  $S$ . La capacité  $D_0$  de*

$S$  est définie par

$$D_0 = \lim_{\epsilon \rightarrow 0} \frac{\ln N(\epsilon)}{\ln(1/\epsilon)}, \quad (1.1)$$

si une telle limite existe.

On peut maintenant énoncer le théorème de recouvrement au moyen de coordonnées à délai («Fractal Delay Embedding Prevalence Theorem») :<sup>3</sup>

**Théorème 1 (Recouvrement au moyen de coordonnées à délai)** *Soit  $\phi$  le flot d'un système dynamique continu sur un sous-ensemble ouvert  $U \subset \mathcal{R}^d$ , et soit  $A$  un sous-ensemble compact<sup>4</sup> de  $U$  de capacité  $D_0$ . Soit  $m \in \mathcal{N}$  tel que  $m > 2D_0$ , et soit  $\tau \in \mathcal{R}$  tel que  $\tau > 0$ . Supposons que  $A$  contient :*

1. *Au plus un nombre fini de points d'équilibre.*
2. *Aucune orbite périodique de période  $\tau$  ou  $2\tau$ .*
3. *Au plus un nombre fini d'orbites périodiques de période  $3\tau, 4\tau, \dots, m\tau$ . Supposons de plus, pour chacune de ces orbites périodiques, que la matrice jacobienne de l'application de Poincaré ait des valeurs propres distinctes.*

Alors pour presque toute fonction continûment différentiable  $h : U \rightarrow \mathcal{R}$ , l'application des coordonnées à délai

$$\begin{aligned} \mathbf{H}(h, \phi, \tau) : U &\rightarrow \mathcal{R}^m \\ \mathbf{x} &\mapsto (h(\mathbf{x}), h(\phi_{-\tau}(\mathbf{x})), h(\phi_{-2\tau}(\mathbf{x})), \dots, h(\phi_{-(m-1)\tau}(\mathbf{x}))) \end{aligned}$$

est :

1. *une injection sur  $A$ ,*
2. *une immersion sur chaque sous-ensemble compact  $C$  d'une variété continûment différentiable contenue dans  $A$ .*

---

<sup>3</sup>Le théorème qui suit s'applique aux systèmes dynamiques continus. Un théorème tout à fait analogue existe pour les systèmes dynamiques discrets, cf. [SYC91].

<sup>4</sup>C'est-à-dire fermé et borné.

Donnons quelques précisions sur ce théorème, sans entrer profondément dans les détails techniques mathématiques :

1. L'expression «pour presque toute fonction on observe une certaine propriété» est équivalente ici à l'expression «la fonction a une certaine propriété avec probabilité un», au sens d'une mesure dans un espace de fonctions («prevalence»). En d'autres termes, si par extrême infortune la fonction choisie n'a pas la propriété, une perturbation arbitrairement petite présente la propriété avec probabilité un.
2. Le fait que  $\mathbf{H}$  soit une immersion sur  $C$  signifie que l'application linéaire  $\mathbf{DH}(\mathbf{x}_0)$  (matrice jacobienne de  $\mathbf{H}$  au point  $\mathbf{x}_0 \in C$ ) est injective  $\forall \mathbf{x}_0 \in C$  (correspondance un à un entre deux espaces tangents, l'un centré en  $\mathbf{x}_0$  et l'autre centré en  $\mathbf{H}(\mathbf{x}_0)$ ). De façon plus explicite, si  $\delta \mathbf{x} = \mathbf{x} - \mathbf{x}_0$  et  $\|\delta \mathbf{x}\| \ll 1$ , alors  $\delta \mathbf{H}(\mathbf{x}) = \mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{x}_0) \approx \mathbf{DH}(\mathbf{x}_0)\delta \mathbf{x}$ . En d'autres termes,  $\mathbf{H}$  préserve la structure différentielle de toute variété contenue dans  $A$ .

Nonobstant les quelques conditions plus fines de ce théorème, rarement vérifiées en pratique, quel en est le message ? À partir d'une seule observable  $y(t) = h(\mathbf{x}(t))$  d'un système dynamique évoluant sur un sous-ensemble invariant  $A \subset \mathcal{R}^d$  de capacité  $D_0$  (e.g. attracteur étrange), on peut recouvrir  $A$  dans un espace des phases reconstruit au moyen de *vecteurs à délais*

$$\mathbf{y}(t) = (y(t), y(t - \tau), y(t - 2\tau), \dots, y(t - (m - 1)\tau)), \quad (1.2)$$

pour autant que la condition  $m > 2D_0$  soit satisfaite. (On appelle communément  $m$  la dimension de recouvrement et  $\tau$  le délai de recouvrement.) La correspondance un à un entre  $A$  et  $\mathbf{H}(A)$  (bijection) permet d'importer dans l'espace des phases reconstruit le déterminisme d'origine, et par conséquent d'y réaliser des prédictions [Cas87]. De plus, le fait que  $\mathbf{H}$  préserve la structure différentielle de toute variété contenue dans  $A$  présente deux conséquences très importantes :

1. La dimension fractale de tout sous-ensemble fermé de  $A$  demeure inchangée sous l'action de  $\mathbf{H}$ .

2. Les exposants de Lyapunov sont reproduits dans l'espace des phases reconstruit.<sup>5</sup>

On comprend donc que ce théorème de recouvrement, au moyen de coordonnées à délai, est à la base de l'étude des séries temporelles chaotiques.

## 1.2 Pratique

Nous avons déjà mentionné les quatre caractéristiques d'un signal expérimental (échantillonnage, longueur finie, bruit, non-stationnarité). Les deux premières signifient que l'on dispose d'une série de données de la forme  $\{y_t\}_{t=1}^N$  (fréquence d'échantillonnage constante), où  $N$  est la longueur de la série temporelle. La reconstruction revient alors à peupler un espace des phases reconstruit de  $N_{\tau,m} = N - (m - 1)\tau$  vecteurs à délai  $\mathbf{y}_t = (y_t, y_{t-\tau}, \dots, y_{t-(m-1)\tau})$ , où  $\tau$  est ici un entier. La question est maintenant de savoir si sous de telles contraintes, le théorème de recouvrement demeure valide. La réponse formelle semble être négative [Sch99]. Quoi qu'il en soit, cela n'arrête pas pour autant de nombreux scientifiques d'utiliser avec succès la technique des coordonnées à délai dans des réalisations pratiques (cf. [OSY94]). Encore une fois, c'est un pragmatisme rigoureux qu'il convient d'adopter.

Si la série temporelle était d'une longueur et d'une précision infinies (ce qui implique sans bruit), le choix de la fréquence d'échantillonnage et du délai  $\tau$  serait sans importance (si l'on fait abstraction des quelques choix de  $\tau$  prohibés par le théorème). Le fait que la longueur de la série temporelle soit en pratique finie modifie les règles du jeu. Pour un signal chaotique par exemple, il importe d'abord de s'assurer, autant que possible, que la fréquence d'échantillonnage et la longueur de la série temporelle sont telles que l'on couvre de façon satisfaisante l'attracteur étrange. Ensuite, il faut

---

<sup>5</sup>Néanmoins, il apparaît également des exposants de Lyapunov superflus, artefacts du processus de recouvrement et dont l'origine n'est pas dynamique. À ce sujet, on peut consulter les références [EKRC86], [BBA90], [Par92] et plus récemment [STY98].

réaliser que la qualité du recouvrement dépend en pratique du choix des paramètres de recouvrement  $m$  et  $\tau$ . La condition  $m > 2D_0$  est une condition suffisante mais non nécessaire. Il se peut très bien qu'une valeur inférieure soit tout à fait adéquate.

Comment choisir  $m$  et  $\tau$  en pratique? La littérature fournit une réponse conventionnelle : on choisit  $\tau$  comme le premier zéro de la fonction d'autocorrélation (indépendance linéaire) ou comme le premier minimum de l'information mutuelle moyenne entre  $y_t$  et  $y_{t+\tau}$  [FS86]. Le but visé est d'obtenir deux coordonnées suffisamment indépendantes l'une de l'autre mais pas trop. Une fois  $\tau$  déterminé, on choisit  $m$  comme la dimension minimum où le pourcentage de faux voisins dans l'espace des phases reconstruit tombe à zéro (attracteur reconstruit déplié) [KBA92]. Rappelons qu'a priori,  $D_0$  n'est pas connu, quoiqu'il serait possible d'estimer une borne inférieure via la dimension de corrélation  $D_2$  [OSY94].

La vraie réponse est que l'optimisation d'un recouvrement est grandement tributaire de l'application visée. Par exemple, pour réaliser une réduction de bruit, l'expérience démontre sans équivoque qu'il est préférable de conserver  $\tau$  à l'unité et de prendre une dimension de recouvrement  $m$  très élevée («over-embedding»). Mais au fait, qu'en est-il du bruit par rapport au théorème de recouvrement? Ses effets sont très néfastes, et c'est pourquoi nous devons le réduire suffisamment avant d'inférer des propriétés de l'attracteur étrange d'origine à partir de celles de l'attracteur reconstruit. Étonnamment, nous utiliserons au chapitre suivant la reconstruction pour améliorer les révélations de la reconstruction. Les résultats positifs que l'on obtient justifient une telle démarche.

Un dernier point : dans de très nombreuses applications (méthode des faux voisins, prédiction, réduction du bruit, spectre des exposants de Lyapunov, dimension de corrélation, contrôle du chaos, etc...), une tâche répétitive que l'on accomplit dans l'espace des phases reconstruit consiste à rechercher les proches voisins d'un point. Le temps de calcul dépend donc de façon cruciale de l'algorithme conçu à cet effet. En particulier, la méthode directe qui consiste à balayer l'ensemble des points s'avère sou-

vent inadéquate. Le progiciel TISEAN [HKS99] offre une pléiade de routines inhérentes aux méthodes non-linéaires d'analyse des séries temporelles. On y retrouve notamment une méthode efficace de recherche des voisins, dite assistée par boîtes et très élégamment implémentée par Hegger en C. Nous avons utilisé cet algorithme. Enfin, dans le présent mémoire, le voisinage d'un point, e.g.  $\mathbf{y}_t$ , sera défini via un rayon  $\epsilon$  (norme maximale), noté  $\mathcal{U}_\epsilon^{(t)}$ , et les proches voisins seront au nombre de  $|\mathcal{U}_\epsilon^{(t)}|$ .

## Chapitre 2

# Réduction non-linéaire du bruit

La présence de bruit pose un sérieux handicap à l'extraction d'information sur le système au moyen de l'espace des phases reconstruit. En effet, le système ne peut être scruté à toutes les échelles, ce qui compromet par exemple le calcul de la dimension fractale de l'attracteur étrange [KS93]. De plus, le déterminisme pur d'origine est réduit, ce qui compromet par exemple la qualité des prédictions. Il y a donc tout intérêt à chercher à éliminer le bruit, ou du moins à le réduire.

Un signal chaotique est typiquement accompagné d'un spectre en puissance qui couvre densément l'étendue en fréquence, et ce même si le signal est propre (voir l'exemple de la Fig. 2.1). Par conséquent, règle générale, le spectre en puissance ne permet pas de séparer efficacement le bruit (composante stochastique) de la composante déterministe chaotique. Les méthodes linéaires de filtrage, qui souvent procèdent en associant les très hautes fréquences au bruit, doivent donc être remplacées. Durant la dernière décennie, plusieurs méthodes non-linéaires ont été proposées (la référence [KS93] est une revue critique de plusieurs de ces méthodes), méthodes qui utilisent avantageusement les propriétés géométriques du signal dans l'espace des phases reconstruit. Parmi ce groupe, nous avons retenu deux algorithmes. Le premier, extrêmement simple, provient de Schreiber [Sch93]. Le second, qui semble être

l'un des plus performants, est celui de GHKSS [GHK<sup>+</sup>93].

La stratégie non-linéaire revient donc à abandonner l'espace de Fourier pour l'espace des phases reconstruit comme lieu de travail.

## 2.1 Signaux et bruits artificiels

Afin d'étudier les performances des filtres non-linéaires dans un environnement bien contrôlé, nous avons utilisé des séries temporelles bruitées de façon artificielle et provenant de systèmes dynamiques discrets bien connus, fonctionnant en régime chaotique. Chaque application retenue s'exprime directement en terme de coordonnées à délai :

$$x_t = f(a_1, \dots, a_l; x_{t-1}, \dots, x_{t-d}), \quad (2.1)$$

où  $d$  est la dimension et  $l$  le nombre de paramètres de l'application. Le premier système est l'application logistique<sup>1</sup> ( $d = l = 1$ ),

$$x_t = f(r; x_{t-1}) = rx_{t-1}(1 - x_{t-1}), \quad (2.2)$$

le second l'application de Hénon<sup>2</sup> [H76] ( $d = l = 2$ ),

$$x_t = f(a, b; x_{t-1}, x_{t-2}) = 1 + bx_{t-2} - ax_{t-1}^2, \quad (2.3)$$

et enfin le dernier, l'application de Mackey-Glass ( $d = k + 1$ ,  $l = 5$ ),

$$\begin{aligned} x_t &= f(a, b, c, k, t_f; x_{t-1}, x_{t-k}, x_{t-k-1}) \\ &= \frac{1}{2k + bt_f} \left[ (2k - bt_f)x_{t-1} + at_f \left( \frac{x_{t-k-1}}{1 + x_{t-k-1}^c} + \frac{x_{t-k}}{1 + x_{t-k}^c} \right) \right]. \end{aligned} \quad (2.4)$$

Cette dernière application se résume à une discrétisation de l'équation différentielle à délai dite de Mackey-Glass [MG77] (délai de rétroaction :  $t_f$ , nombre de pas durant

<sup>1</sup>Nous avons choisi  $r = 3.8$  (à l'intérieur d'une fenêtre chaotique).

<sup>2</sup>Nous avons choisi  $a = 1.29$  et  $b = 0.3$  (attracteur étrange).

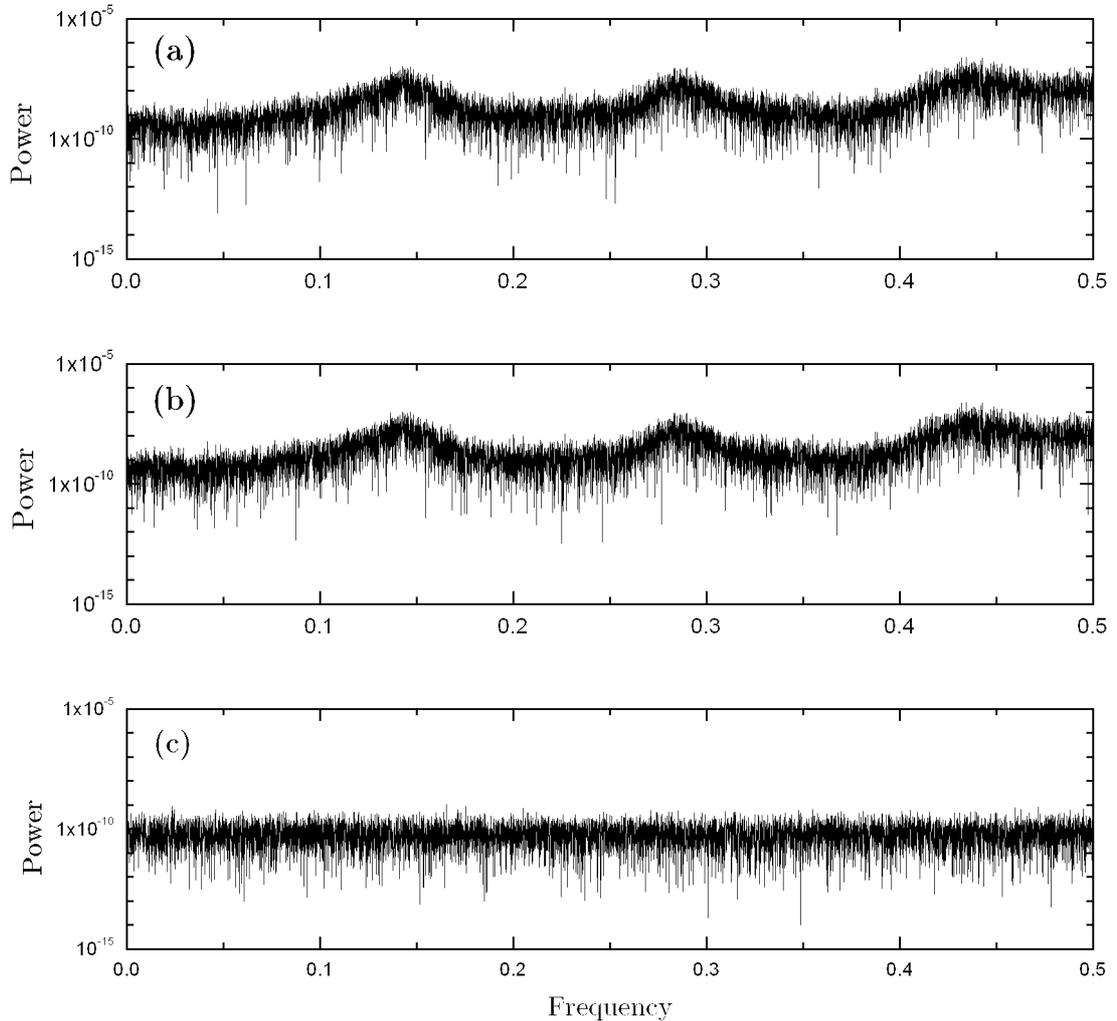


FIG. 2.1 – Spectre en puissance (FFT) pour trois séries temporelles de  $2^{14}$  points : (a) variable  $x$  de l’attracteur de Hénon, (b) signal précédent additionné de 10% de bruit de mesure de distribution uniforme et (c) bruit de mesure de distribution uniforme. Il est clair que dans le présent cas, typique des signaux chaotiques bruités, l’utilisation d’un filtre passe-bas pour éliminer le bruit est à déconseiller, puisqu’une composante importante du signal d’origine serait perdue. Un bruit au spectre plat, tel celui montré en (c), est dit blanc, pour des raisons évidentes. Notons au passage qu’il existe des filtres linéaires plus performants que le simple filtre passe-bas, e.g. les filtres de type Wiener. Ceux-ci demeurent tout de même moins efficaces que les filtres non-linéaires dans le cas des signaux chaotiques [GHK<sup>+</sup>93].

$t_f : k$ ).<sup>3</sup>

Chaque application requiert  $d$  conditions initiales  $(x_{2-d}, \dots, x_1)$ , et par convention, nous ne retenons que la dernière dans notre série temporelle de travail, qui sera scalaire :  $\{x_t\}_{t=1}^N$ . Dans notre banque de séries temporelles artificielles, nous utilisons invariablement  $N = 2^{14} = 16\,384$ .

On distingue deux types de bruit, le *bruit de mesure ou d'observation* et le *bruit dynamique*. Afin d'éviter toute confusion, nous réserverons le symbole  $\{x_t\}$  à une série sans bruit, et affecterons le symbole  $\{s_t\}$  à un signal bruité. Qui plus est, en aucun cas nous ne bruitons les conditions initiales, i.e.  $s_t = x_t \forall t \in [2 - d, 1]$ . Ainsi donc, pour un bruit de mesure :

$$s_t = x_t + \eta_t, \quad (2.8)$$

où  $\eta_t$  est la valeur prise au temps  $t$  par une variable stochastique  $\eta$  de *moyenne nulle* et de variance  $\sigma_\eta^2$ . Pour un bruit dynamique :

$$s_t = f(a_1, \dots, a_l; s_{t-1}, \dots, s_{t-d}) + \eta_t. \quad (2.9)$$

---

<sup>3</sup>Nous avons choisi des valeurs classiques pour les différents paramètres :  $a=0.2$ ,  $b=0.1$ ,  $c=10$ ,  $t_f=17$  et  $k=170$ . Il est important de garder à l'esprit que la dimension de l'application de Mackey-Glass est de  $k+1$ , i.e.  $k+1$  variables sont nécessaires pour décrire l'état du système et en déterminer l'évolution subséquente. Si l'on écrit l'application sous la forme  $\mathbf{x}_t = \mathbf{F}(\mathbf{x}_{t-1})$ , ce fait devient évident. Pour ce faire, on introduit  $k+1$  variables

$$x_t^i \equiv x_{t-i} \quad i \in [0, k], \quad (2.5)$$

et le système prend alors la forme standard :

$$x_t^0 = f(a, b, c, k, t_f; x_{t-1}^0, x_{t-1}^{k-1}, x_{t-1}^k) \quad (2.6)$$

$$x_t^i = x_{t-1}^{i-1} \quad i \in [1, k]. \quad (2.7)$$

Avec  $k=170$ , on se retrouve donc avec un système de dimension relativement élevée. Néanmoins, l'attracteur étrange correspondant est de dimension fractale  $D_1 \approx 2.1$  (dimension d'information évaluée au moyen de la conjecture de Kaplan-Yorke) [OSY94].

Dans notre étude, la variable aléatoire  $\eta$  est de distribution uniforme ou gaussienne. Dans le cas gaussien, nous restreignons les valeurs possibles à l'intervalle  $]-G\sigma_\eta, G\sigma_\eta]$ .

Pour un bruit de mesure, on définit le niveau de bruit  $\kappa$  de la façon suivante :

$$\kappa = \sqrt{\frac{\langle \eta_t^2 \rangle}{\langle x_t^2 \rangle}}, \quad (2.10)$$

où  $\langle \bullet \rangle = \frac{1}{N} \sum_{t=1}^N \bullet$ . Ainsi, si l'on veut ajouter du bruit de mesure à un niveau  $\kappa$ , la recette suivante est commode :

$$\eta_t = \left( \frac{\sqrt{\langle x_t^2 \rangle} \kappa}{\sigma_\xi} \right) \xi_t, \quad (2.11)$$

où  $\xi_t$  est une variable aléatoire de *moyenne nulle* et de variance  $\sigma_\xi^2$  (nombres pseudo-aléatoires générés par l'ordinateur). Par extension, nous utiliserons également  $\kappa$  pour quantifier le bruit dynamique et la recette (2.11) pour le générer.

Pour le bruit de mesure, la restriction permissive  $G = 100.0$  et les quatre niveaux de bruit suivants ont été considérés : 1.0%, 2.5%, 5.0% et 10.0%. Pour le bruit dynamique, nous avons resserré considérablement l'intervalle d'acceptation à  $G = 2.0$ , de manière à éviter la divergence de l'orbite, et étudié les niveaux de bruit de 1.0%, 2.0%, 3.0% et 4.0%. Pour l'application logistique, le dernier niveau de bruit dynamique s'est avéré intraitable (divergence de l'orbite).

## 2.2 Critères de performance

Afin de quantifier les performances d'une réduction de bruit, il importe de se doter de critères objectifs. L'origine artificielle de nos données rend cette tâche beaucoup plus aisée. En toute continuité, nous allons désigner par  $\{\tilde{s}_t\}$  la série résultant d'une réduction de bruit. Ainsi, dans le cas où la dynamique est connue exactement et exprimable en terme de coordonnées à délai, on peut facilement évaluer l'*erreur dynamique*

(violation du déterminisme) :

$$E_{\text{dyn}}\{s_t\} = \left[ \frac{1}{N-d} \sum_{t=d+1}^N (s_t - f(a_1, \dots, a_l; s_{t-1}, \dots, s_{t-d}))^2 \right]^{1/2}. \quad (2.12)$$

D'autre part, pour un bruit de mesure où l'on a accès à la série originale sans bruit  $\{x_t\}$ , on peut calculer l'*erreur observationnelle* <sup>4</sup>:

$$E_0\{s_t\} = \left[ \frac{1}{N} \sum_{t=1}^N (s_t - x_t)^2 \right]^{1/2}. \quad (2.13)$$

Les deux critères dérivés suivants permettent de chiffer l'amélioration obtenue par la réduction de bruit :

$$r_{\text{dyn}} = \frac{E_{\text{dyn}}\{s_t\}}{E_{\text{dyn}}\{\tilde{s}_t\}} \quad (2.14)$$

$$r_0 = \frac{E_0\{s_t\}}{E_0\{\tilde{s}_t\}} \quad (2.15)$$

Ces deux quantités complémentaires s'interprètent facilement : une valeur supérieure à l'unité indique un effet bénéfique. Pour un bruit dynamique, seul  $r_{\text{dyn}}$  sera utilisé.

## 2.3 Méthode de Schreiber

### 2.3.1 Description de la méthode

Cette méthode, d'ordre zéro, consiste à remplacer la coordonnée centrale ( $m$  impair) d'un vecteur à délais  $\mathbf{s}_n = (s_n, s_{n-\tau}, \dots, s_{n-(m-1)\tau})$  par la coordonnée centrale moyenne de ses proches voisins, i.e.

$$\tilde{s}_{n-(m-1)\tau/2} = \frac{1}{|\mathcal{U}_\epsilon^{(n)}|} \sum_{\mathbf{s}_t \in \mathcal{U}_\epsilon^{(n)}} s_{t-(m-1)\tau/2}, \quad (2.16)$$

sous l'espoir d'un bruit de mesure qui localement sera de moyenne nulle. On ne corrige que la coordonnée centrale essentiellement pour la raison suivante : la sensibilité aux conditions initiales, qui se manifeste dans les deux directions du temps sur

<sup>4</sup>Le niveau de bruit n'est que l'erreur observationnelle normalisée par rapport à  $\sqrt{\langle x_t^2 \rangle}$ .

un attracteur chaotique (exposants de Lyapunov positifs et négatifs), rend les coordonnées extrêmes plus susceptibles de différer *dynamiquement* entre voisins que les centrales. Ainsi, remplacer une coordonnée extrême par la moyenne correspondante sur les proches voisins risque fort bien de lessiver du contenu dynamique, le risque étant beaucoup moindre au centre.

Le rayon  $\epsilon$  du voisinage doit être choisi suffisamment grand pour couvrir les effets du bruit, i.e. recueillir d'authentiques proches voisins catapultés par le bruit, tout en demeurant adéquatement réduit. Une bonne valeur de départ est 2 à 3 fois la valeur rms du bruit, i.e.  $\sqrt{\langle \eta_t^2 \rangle}$ . Pour les itérations subséquentes, la valeur rms des corrections apportées à l'itération précédente représente un bon ordre de grandeur pour  $\epsilon$ . Une utilisation normale de cet algorithme se traduira donc par la décroissance de  $\epsilon$  au fil des itérations, jusqu'à l'inaction éventuelle de l'algorithme, en raison de l'absence de proches voisins autres que le point lui-même.

### 2.3.2 Résultats

Les performances d'une réduction de bruit composée de plusieurs itérations dépendent de façon non triviale des différents paramètres. Il est difficile, à notre avis, de prétendre à la réduction optimale. Quoiqu'il en soit, on peut parvenir à de très bons résultats, comme ceux que nous allons présenter ici. En effet, le programme que nous avons conçu pour réduire le bruit (Schreiber ou GHKSS) comporte une interface très commode. À chaque itération, on peut tenter une réduction et visualiser ses effets (e.g.  $r_0$  et  $r_{\text{dyn}}$ ) avant de l'accepter. Ceci permet de réaliser une certaine optimisation, et conduit à de très bonnes réductions. Pour les trois applications étudiées dans ce travail, nous avons travaillé avec un délai de recouvrement  $\tau = 1$ , et ce aussi bien pour la méthode de Schreiber que la méthode GHKSS. Nos explorations numériques ont confirmé que ce choix était de loin le meilleur.

Nous allons visualiser l'effet des réductions de bruit au moyen du plan  $(x_{t-\tau}, x_t)$

dans l'espace des phases reconstruit, ou en d'autres mots, au moyen d'un graphe à délai. (Le délai de visualisation peut être différent du délai de réduction.) Pour l'application logistique et celle de Hénon, un délai de visualisation  $\tau = 1$  s'avère un choix incontournable, et dans les deux cas, l'attracteur reconstruit est complètement déplié dans le plan  $(x_{t-1}, x_t)$ . Pour l'application de Mackey-Glass, nous avons choisi un délai de visualisation  $\tau = 110$  (premier minimum de l'information mutuelle moyenne) pour la visualisation. Selon la méthode des faux voisins, l'attracteur reconstruit serait déplié dès  $m_{\min} = 3$ . Par conséquent, la visualisation 2D s'avère suffisamment claire (l'orbite ne se recoupe pas trop). Il faut également se rappeler que l'attracteur étrange est quasiment plat ( $D_1 \approx 2.1$ ). Enfin, l'origine de cette application, soit la discrétisation d'une équation différentielle, explique le fait que les points tracent une véritable trajectoire. On peut dans ce cas considérer la série temporelle comme résultant de l'échantillonnage d'un système dynamique continu.

La partie (c) des Figs. 2.2 à 2.13 montre l'attracteur filtré obtenu au moyen de l'algorithme de Schreiber, que l'on peut comparer au cas contaminé en (a), ou encore au cas sans bruit en (b). Dans tous les cas, l'amélioration est palpable. Les tableaux 2.1, 2.3, 2.5, 2.7 et 2.9 présentent les performances réalisées. Dans le cas de l'application logistique et de celle de Hénon, une dimension de recouvrement variant de 11 à 3 a été utilisée. Pour Mackey-Glass, où seul le bruit de mesure a été étudié,  $m$  a été essentiellement fixé à 25. Pour le bruit de mesure, la dimension de recouvrement doit effectivement demeurer élevée. Pour le bruit dynamique,  $m$  est règle générale diminué progressivement, et nous avons observé qu'il est souvent efficace de réduire  $m$  de 2 à chaque itération.

## 2.4 Méthode GHKSS

### 2.4.1 Description de la méthode

L'algorithme GHKSS s'insère dans la famille des méthodes par projections locales, et en constitue vraisemblablement le plus illustre membre, pour un certain nombre de raisons [GHK<sup>+</sup>93]. *Cette approche repose sur l'hypothèse que dans l'espace des phases reconstruit, les points s'inscrivent relativement bien sur une variété de basse dimension.* Le bruit tend à disperser les points hors de cette structure, et la stratégie consiste à les y projeter de retour. Bien entendu, il est impératif de travailler avec une dimension de recouvrement suffisamment élevée, non seulement pour révéler correctement la variété, mais également la diffusion due au bruit. Une dimension de recouvrement supérieure à celle minimale requise pour déplier l'attracteur ( $m_{\min}$ ) est donc nécessaire («over-embedding»), afin que la projection curative soit possible.

Ignorer cette hypothèse de travail peut conduire à des résultats fallacieux. Une expérience fort illustrative à ce sujet a été réalisée dans la référence [HKS99]. En effet, une réduction de bruit GHKSS appliquée à des nombres aléatoires gaussiens fait surgir un semblant de structure. Prudence s'impose.

L'algorithme requiert donc la spécification de la dimension de l'hypothétique variété. Cette dernière est sondée localement, dans le voisinage de chaque point, en recherchant  $Q$  contraintes linéaires (dynamique locale) que satisfont les proches voisins. L'intersection de ces  $Q$  hyperplans définit ainsi le sous-espace linéaire de dimension  $m - Q$  associé localement à la variété. La méthode de Cawley-Hsu-Sauer (Réfs. [CH92a], [CH92b] et [Sau92]) s'affaire alors essentiellement à la projection orthogonale du point sur le sous-espace. La méthode de Schreiber-Grassberger [SG91], quant à elle, ne retient que la correction centrale de cette projection. Comme chaque mesure scalaire entre dans la composition de  $m$  vecteurs à délais, ceci évite les multiples corrections plus ou moins à l'unisson.<sup>5</sup> Ces deux méthodes sont dites à un pas («one-step

<sup>5</sup>Le choix de la coordonnée centrale repose sur les mêmes raisons que celles avancées lors de la

method»). Celles à deux pas ajoutent à l'estimation de la dynamique locale un calcul de la correction en minimisant à la fois celle-ci et le non-respect de la dynamique locale.

L'algorithme GHKSS fusionne habilement ces deux étapes en une seule, tout en prenant soin d'éviter de larges corrections aux composantes extrêmes d'un vecteur à délais. Il en résulte une méthode très efficace. Pour ce faire, on troque d'abord la métrique euclidienne pour une métrique truquée, où l'accent est mis sur les composantes centrales, viz

$$P_{ij} = \begin{cases} 1 & i = j \neq 1, m \\ \delta \ll 1 & i = j = 1, m \\ 0 & \text{sinon.} \end{cases} \quad (2.17)$$

Sous la métrique diagonale  $\mathbf{P}$ , le produit scalaire de deux vecteurs s'écrit :

$$(\mathbf{v}|\mathbf{w}) = \mathbf{v} \cdot \mathbf{P}\mathbf{w} = \mathbf{P}\mathbf{v} \cdot \mathbf{w}. \quad (2.18)$$

On formule ensuite, localement, le problème de minimisation suivant : on cherche une correction  $\boldsymbol{\theta}_t$  pour chacun des  $|\mathcal{U}_\epsilon^{(n)}|$  voisins de  $\mathbf{s}_n$ , un ensemble de  $Q$  vecteurs de contrainte  $\mathbf{a}_q^{(n)}$  et  $Q$  scalaires  $b_q^{(n)}$  de telle sorte que :

$$\sum_{\mathbf{s}_t \in \mathcal{U}_\epsilon^{(n)}} \mathbf{P}^{-1}\boldsymbol{\theta}_t \cdot \boldsymbol{\theta}_t = \min, \quad (2.19)$$

sous les  $|\mathcal{U}_\epsilon^{(n)}|Q$  contraintes suivantes (dynamique locale) :

$$\mathbf{a}_q^{(n)} \cdot (\mathbf{s}_t + \boldsymbol{\theta}_t) + b_q^{(n)} = 0 \quad \text{où} \quad \mathbf{s}_t \in \mathcal{U}_\epsilon^{(n)}, \quad (2.20)$$

et sous les  $Q^2$  contraintes d'orthogonalité suivantes :

$$\mathbf{a}_q^{(n)} \cdot \mathbf{P}\mathbf{a}_{q'}^{(n)} = \delta_{qq'}. \quad (2.21)$$

La solution, par la méthode des multiplicateurs de Lagrange, nous donne directement 

---

discussion de l'algorithme de Schreiber.

une correction pour  $\mathbf{s}_n$  (calcul laborieux mais direct) :<sup>6</sup>

$$\theta_{n,i} = \frac{1}{R_{ii}} \sum_{q=1}^Q \left[ \sum_{j=1}^m R_{jj} e_{q,j}^{(n)} (\xi_j^{(n)} - s_{n,j}) \right] e_{q,i}^{(n)}, \quad (2.22)$$

où  $R_{ij} = \delta_{ij}/\sqrt{P_{ii}}$  et où  $\boldsymbol{\xi}^{(n)}$  représente le centre de masse des proches voisins, i.e.

$$\boldsymbol{\xi}^{(n)} = \frac{1}{|\mathcal{U}_\epsilon^{(n)}|} \sum_{\mathbf{s}_t \in \mathcal{U}_\epsilon^{(n)}} \mathbf{s}_t. \quad (2.23)$$

Les  $Q$  vecteurs propres de la matrice  $\mathbf{R}\mathbf{C}^{(n)}\mathbf{R}$  associés aux plus petites valeurs propres sont notés  $\mathbf{e}_q^{(n)}$ , où  $\mathbf{C}^{(n)}$  est la matrice de covariance (consulter l'annexe B) :

$$C_{ij}^{(n)} = \left[ \frac{1}{|\mathcal{U}_\epsilon^{(n)}|} \sum_{\mathbf{s}_t \in \mathcal{U}_\epsilon^{(n)}} s_{t,i} s_{t,j} \right] - \xi_i^{(n)} \xi_j^{(n)}. \quad (2.24)$$

Pour une présentation alternative de la méthode GHKSS, on peut consulter le chapitre 10 de la référence [KS97].

Il est possible de corriger en partie pour les effets non-linéaires (variété courbe) négligés par l'approximation linéaire locale (Réfs. [HKS99] et [KS93]). Pour ce faire, on prend soin d'emmagasiner l'ensemble des corrections  $\boldsymbol{\theta}_n$  avant de les appliquer. Pour chaque point, on calcule ensuite la correction moyenne  $\langle \boldsymbol{\theta} \rangle_{\mathcal{U}_\epsilon^{(n)}}$  de ses proches voisins. Pour un bruit de mesure typique et une variété exclusivement linéaire, on s'attend à ce que cette correction moyenne soit nulle. La légère non-linéarité présente dans  $\mathcal{U}_\epsilon^{(n)}$  entraîne une tendance vers le centre de courbure locale, et par le fait même une valeur non nulle pour  $\langle \boldsymbol{\theta} \rangle_{\mathcal{U}_\epsilon^{(n)}}$ . Remplacer la correction  $\boldsymbol{\theta}_n$  par  $\boldsymbol{\theta}_n - \langle \boldsymbol{\theta} \rangle_{\mathcal{U}_\epsilon^{(n)}}$  élimine cette tendance malfaisante.

Comme nous l'avons déjà souligné, une correction pour chaque vecteur à délais implique  $m$  corrections pour chaque mesure scalaire  $s_n$ , corrections possiblement conflictuelles. On peut à nouveau se servir de la métrique truquée ici, pour accorder plus

<sup>6</sup>Bien que la solution complète propose une correction pour chaque proche voisin incluant le point lui-même, seule la correction destinée à ce dernier est retenue, et le processus répété pour l'ensemble des points de l'espace des phases reconstruit.

d'importance aux corrections provenant de vecteurs à délais où la mesure scalaire joue le rôle d'une composante centrale. La correction apportée à chaque mesure devient finalement :

$$\tilde{s}_n = s_n + \frac{1}{\text{Tr}\mathbf{R}^{-1}} \sum_{i=1}^m \frac{\theta_{n+i-1,i}}{R_{ii}}. \quad (2.25)$$

Enfin, il demeure toujours possible qu'une correction scalaire soit anormalement élevée, et ce pour un certain nombre de raisons [KS93]. Il est recommandé de réduire une telle correction, afin d'éviter d'éventuelles distorsions. Une façon de repérer ces anomalies consiste à calculer la moyenne et l'écart-type de l'ensemble des corrections scalaires. Toute correction située à disons plus de 10 écart-types de la moyenne est ensuite revue à la baisse.

## 2.4.2 Résultats

Les paramètres importants de la méthode GHKSS comprennent la dimension de recouvrement  $m$ , le rayon du voisinage  $\epsilon$  et le nombre de contraintes  $Q$ . Comme il est souligné dans la référence [GHK<sup>+</sup>93], de meilleurs résultats sont obtenus si l'on permet aux différents paramètres de varier d'une itération à l'autre, comme dans le cas de la méthode de Schreiber. À la lumière de nos expériences numériques, cette remarque est des plus importantes. De plus, dans le cadre de l'algorithme GHKSS, il est préférable de contrôler plus finement le voisinage que la seule spécification de  $\epsilon$ , afin d'utiliser des linéarisations locales de qualité. L'approche retenue consiste en un balayage de plusieurs valeurs comprises entre  $\epsilon_{\min}$  et  $\epsilon_{\max}$  et en l'imposition d'un nombre minimum de proches voisins. Notre routine affiche à l'écran le pourcentage de points corrigés pour chaque valeur croissante de  $\epsilon$ .

À quel moment faut-il arrêter l'itération du processus de réduction ? La réponse semble être la suivante : le plus tôt possible dès que l'on voit que l'on ne gagne plus que des poussières à effectuer une itération additionnelle. Pour l'algorithme de Schreiber et dans le cas d'un bruit de mesure, rapidement le processus stagne et on

n'arrive plus à réduire le bruit davantage, i.e. à augmenter à la fois  $r_0$  et  $r_{\text{dyn}}$ . C'est pratiquement sans danger. La situation est beaucoup plus risquée dans le cas d'un bruit dynamique ou encore lors de l'utilisation de l'algorithme GHKSS : l'acharnement peut conduire à des distorsions importantes et faire apparaître des structures dont l'origine n'est pas dynamique («over-cleaning»). Il faut donc savoir s'arrêter au bon moment. Évidemment, le suivi est beaucoup plus facile dans le cadre de séries artificielles. Pour des données expérimentales, il faudra travailler plus fort [KS93]. Pour les deux algorithmes, quelques itérations seulement suffisent habituellement.

Parallèlement à l'observation de  $r_0$  et  $r_{\text{dyn}}$ , d'autres quantités méritent d'être considérées lors du suivi du processus de réduction. Le nombre moyen de proches voisins peut par exemple guider le choix de  $\epsilon$  dans la méthode de Schreiber. Plus important encore est la valeur rms des corrections scalaires : lors de l'utilisation de l'algorithme GHKSS, celle-ci doit décroître rapidement, voire exponentiellement, au fil des itérations. Dans le cas contraire, on ne converge peut-être pas vers une authentique variété de basse dimension, et donc l'hypothèse de départ de la méthode n'est pas vérifiée. Pour des données réelles, cette remarque est de la plus haute importance.

La partie (d) des Figs. 2.2 à 2.13 montre l'attracteur filtré obtenu au moyen de l'algorithme GHKSS, que l'on peut comparer au cas contaminé en (a), au cas sans bruit en (b), ou encore au résultat de l'algorithme de Schreiber en (c). Les tableaux 2.2, 2.4, 2.6, 2.8 et 2.10 présentent les performances réalisées. On remarque d'emblée que la méthode GHKSS est plus performante que celle de Schreiber.

Les Figs. 2.8 et 2.9 présentent le détail d'une réduction de bruit GHKSS typique (bruit de mesure), soit celle de la Fig. 2.7(d). On observe d'abord que c'est au cours de la première itération que le niveau de bruit subit la plus grande diminution, alors que les dernières itérations n'ont que très peu d'effet (décroissance exponentielle). La courbe de  $r_{\text{dyn}}$  est quant à elle plutôt d'allure sigmoïde. La valeur rms des corrections scalaires diminue bel et bien de façon quasi exponentielle. Quant au nombre moyen de proches voisins dans l'espace des phases reconstruit, sa décroissance systématique

traduit le fait que moins il y a de bruit, plus on peut resserrer le rayon du voisinage afin d'obtenir des estimations de la dynamique locale plus précises.

Pour le choix de  $m$  et  $Q$  (nombre de contraintes), rappelons qu'il est préférable de choisir une dimension de recouvrement supérieure à celle minimale requise, et de prendre  $Q$  tel qu'une variété à  $m - Q$  dimensions puisse contenir ou presque l'attracteur étrange. Nous avons également observé qu'il peut être avantageux de diminuer progressivement  $m$  au fil des itérations, de même que la différence  $m - Q$ . Ainsi, dans le cas de l'application logistique ( $m_{\min} = 1$ ) ou de l'application de Hénon ( $m_{\min} = 2$ ), on peut prendre par exemple  $m = 9$  et  $Q = 7$ , et un nombre moyen de voisins adéquat se situera généralement dans l'intervalle  $[20, 80]$ . Pour l'application de Mackey-Glass, nous avons fixé  $m = 75$ . Il semble qu'une valeur supérieure pour la dimension de recouvrement aurait conduit à de meilleures réductions de bruit, mais le temps de calcul devenait beaucoup trop long (nombre moyen de voisins au-dessus de 500 pour de bonnes réductions).

Il est important de mentionner que la méthode de Schreiber et celle de GHKSS ont été conçues pour traiter un bruit de mesure, et bien qu'en principe elles ne devraient pas entraîner de résultats aberrants dans le cas d'un bruit dynamique, ce n'est pas là leur spécialité. On observe en effet (cf. Figs. 2.4, 2.5, 2.10 et 2.11) que ces deux algorithmes ne parviennent pas à supprimer les régions de l'espace des phases visitées par le système uniquement en raison du bruit dynamique. Il apparaît également plusieurs distorsions localisées.

Dernière observation : il semble que les deux méthodes de réduction de bruit soient plus performantes dans le cas d'un système continu que dans le cas d'un système discret (observer  $r_{\text{dyn}}$ ). En fait, comme le souligne Schreiber [Sch93], plus la fréquence d'échantillonnage est élevée et plus il est facile de nettoyer un signal. Pour un système dynamique discret, la fréquence d'échantillonnage est limitée à la «fréquence intrinsèque» du système, alors que pour un système dynamique continu, une telle limite n'existe pas a priori.

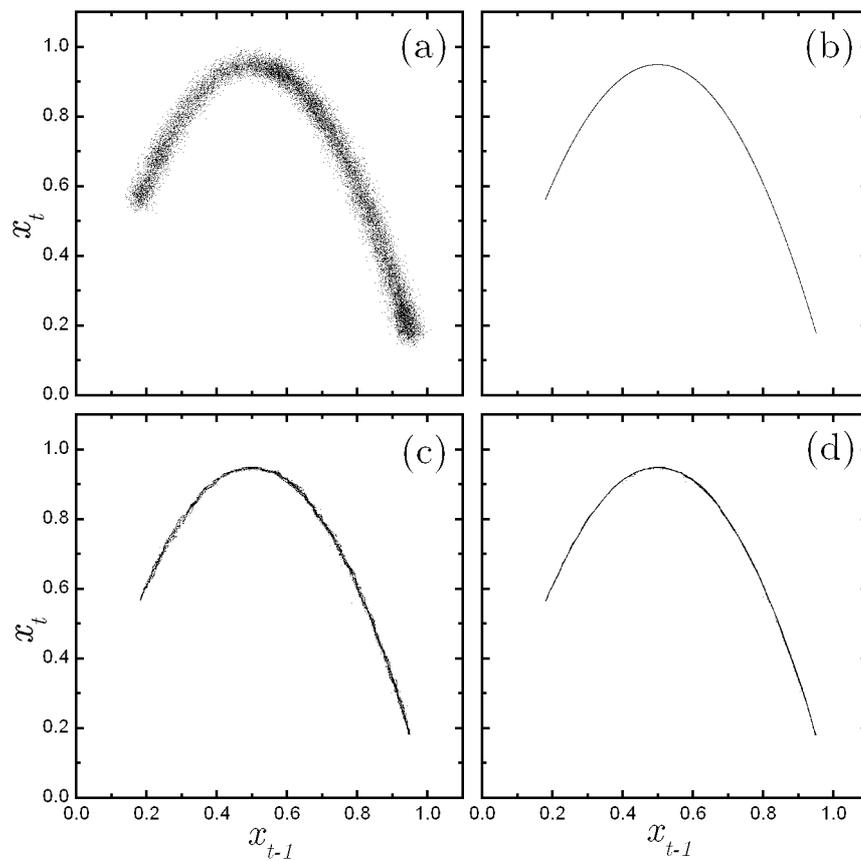


FIG. 2.2 – Graphes à délai de 16 383 points illustrant les performances de deux filtres non-linéaires. (a) Application logistique contaminée par un *bruit de mesure gaussien de 2.5%*. (b) Application logistique sans bruit. (c) Réduction de bruit Schreiber. (d) Réduction de bruit GHKSS.

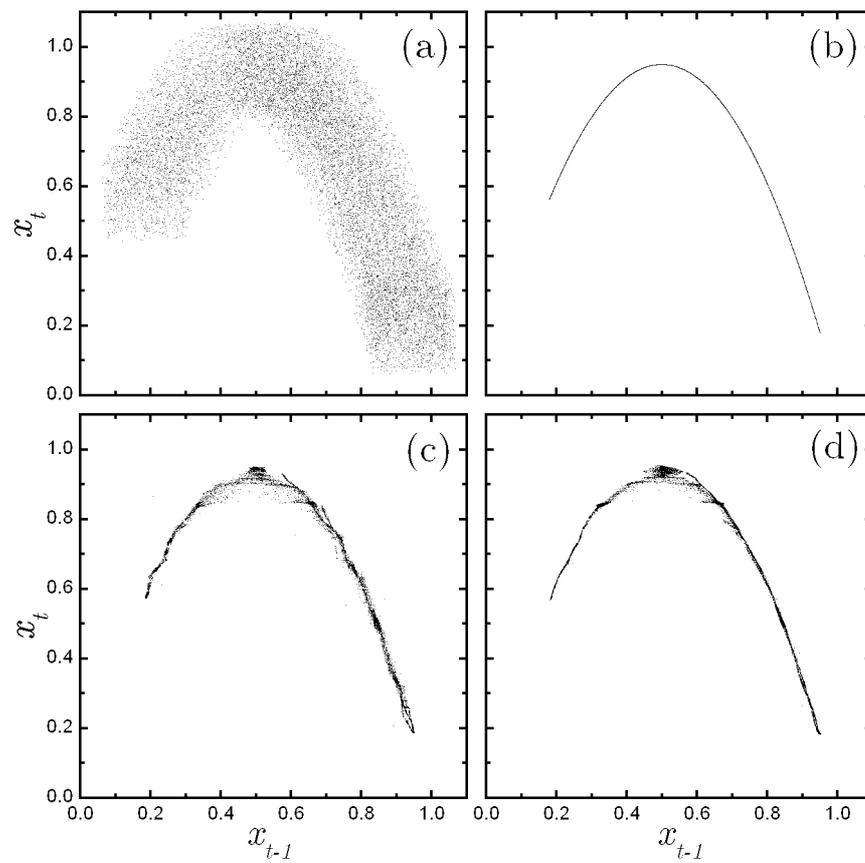


FIG. 2.3 – Graphes à délai de 16 383 points illustrant les performances de deux filtres non-linéaires. (a) Application logistique contaminée par un *bruit de mesure uniforme de 10%*. (b) Application logistique sans bruit. (c) Réduction de bruit Schreiber. (d) Réduction de bruit GHKSS.

TAB. 2.1 – Méthode de Schreiber – Application logistique – Bruit de mesure

Distribution	$\kappa$ initial (%)	$\kappa$ final (%)	$r_0$	$r_{\text{dyn}}$	Nombre d'itérations
gaussienne	0.998	0.273	3.66	8.27	3
gaussienne	2.49	0.752	3.32	6.73	2
gaussienne	4.98	1.56	3.2	11.3	3
gaussienne	9.93	3.43	2.92	11.2	3
uniforme	1.0	0.248	4.04	9.89	3
uniforme	2.51	0.588	4.26	12.8	3
uniforme	5.01	1.37	3.65	12.1	3
uniforme	9.97	3.33	3.01	11.0	2

TAB. 2.2 – Méthode GHKSS – Application logistique – Bruit de mesure

Distribution	$\kappa$ initial (%)	$\kappa$ final (%)	$r_0$	$r_{\text{dyn}}$	Nombre d'itérations
gaussienne	0.998	0.155	6.44	17.6	5
gaussienne	2.49	0.454	5.49	18.5	4
gaussienne	4.98	1.28	3.91	18.7	5
gaussienne	9.93	3.0	3.33	19.0	5
uniforme	1.0	0.149	6.72	14.8	4
uniforme	2.51	0.396	6.32	18.0	4
uniforme	5.01	1.27	3.95	17.1	4
uniforme	9.97	2.93	3.44	16.4	3

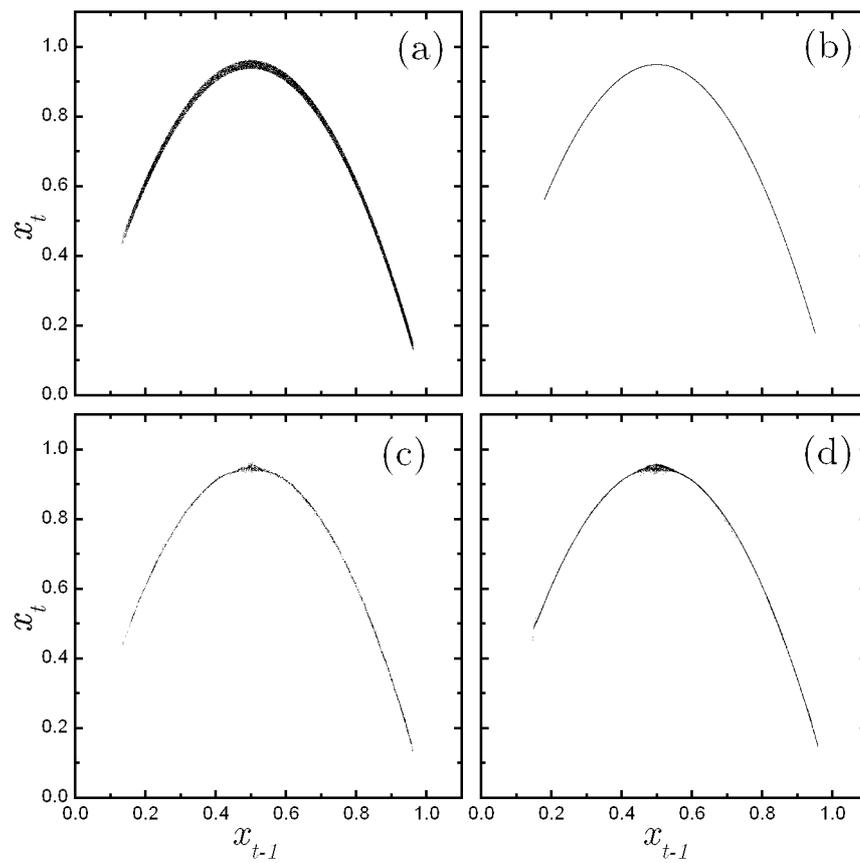


FIG. 2.4 – Graphes à délai de 16 383 points illustrant les performances de deux filtres non-linéaires. (a) Application logistique contaminée par un *bruit dynamique uniforme de 1%*. (b) Application logistique sans bruit. (c) Réduction de bruit Schreiber. (d) Réduction de bruit GHKSS.

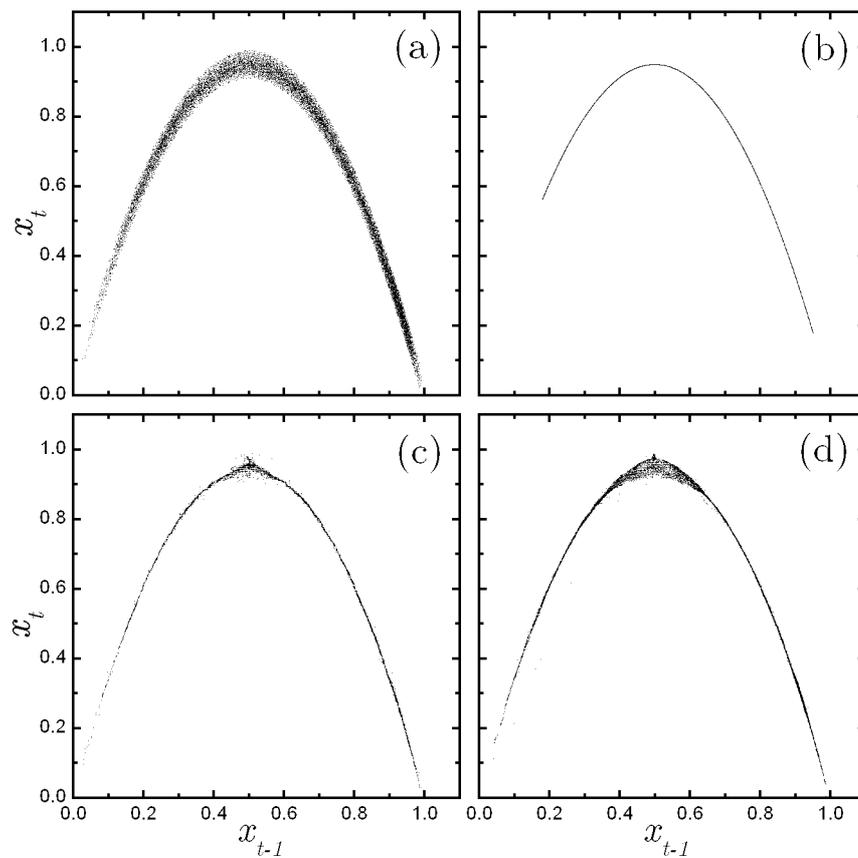


FIG. 2.5 – Graphes à délai de 16 383 points illustrant les performances de deux filtres non-linéaires. (a) Application logistique contaminée par un *bruit dynamique gaussien de 3%*. (b) Application logistique sans bruit. (c) Réduction de bruit Schreiber. (d) Réduction de bruit GHKSS.

TAB. 2.3 – Méthode de Schreiber – Application logistique – Bruit dynamique

Distribution	$\kappa$ initial (%)	$r_{\text{dyn}}$	Nombre d'itérations
gaussienne	1.0	3.33	5
gaussienne	2.0	3.22	5
gaussienne	3.0	3.08	4
uniforme	1.0	3.02	5
uniforme	2.0	3.06	4
uniforme	3.0	2.95	5

TAB. 2.4 – Méthode GHKSS – Application logistique – Bruit dynamique

Distribution	$\kappa$ initial (%)	$r_{\text{dyn}}$	Nombre d'itérations
gaussienne	1.0	3.91	4
gaussienne	2.0	3.42	6
gaussienne	3.0	2.6	3
uniforme	1.0	3.71	3
uniforme	2.0	3.24	5
uniforme	3.0	2.85	4

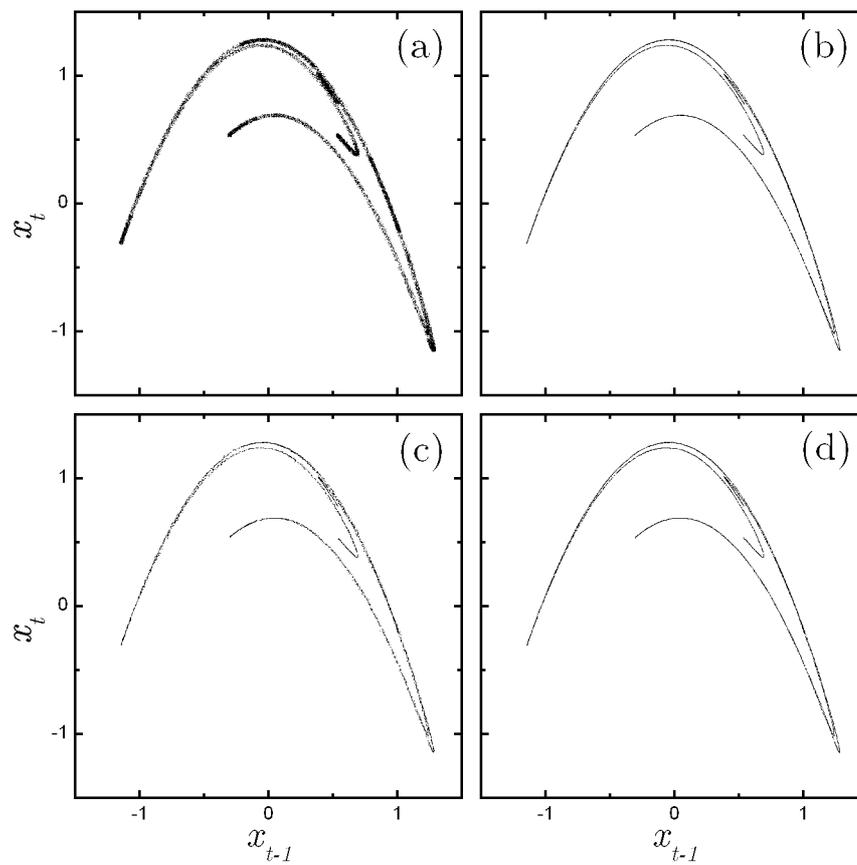


FIG. 2.6 – Graphes à délai de 16 383 points illustrant les performances de deux filtres non-linéaires. (a) Attracteur de Hénon contaminé par un *bruit de mesure uniforme de 1%*. (b) Attracteur de Hénon sans bruit. (c) Réduction de bruit Schreiber. (d) Réduction de bruit GHKSS.

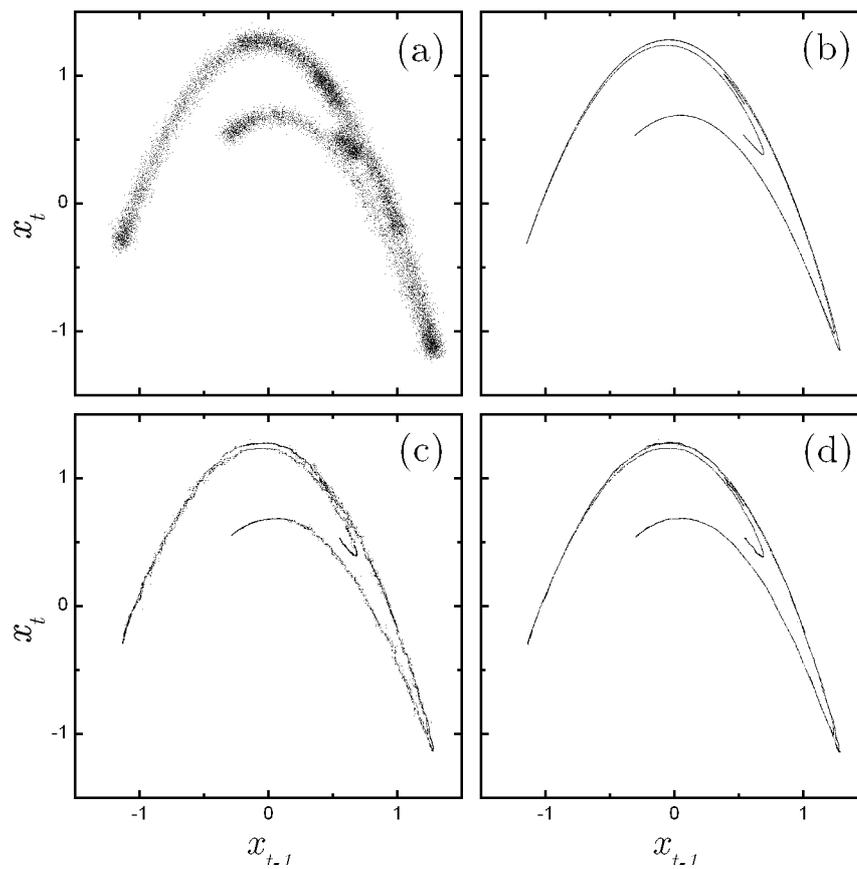


FIG. 2.7 – Graphes à délai de 16 383 points illustrant les performances de deux filtres non-linéaires. (a) Attracteur de Hénon contaminé par un *bruit de mesure gaussien de 5%*. (b) Attracteur de Hénon sans bruit. (c) Réduction de bruit Schreiber. (d) Réduction de bruit GHKSS.

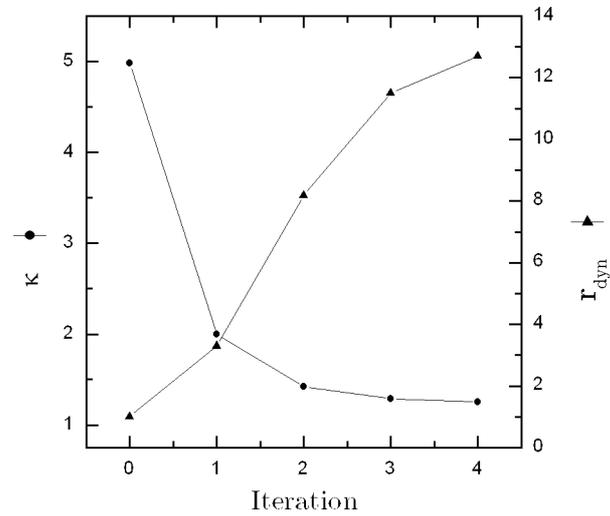


FIG. 2.8 – Pour la réduction de bruit GHKSS montrée à la Fig. 2.7d (Hénon, bruit de mesure gaussien de 5%), variation du niveau de bruit  $\kappa$  et de  $r_{\text{dyn}}$  au fil des itérations.

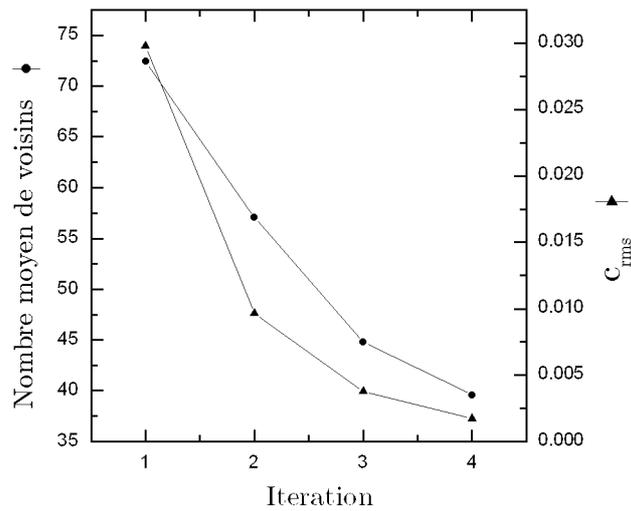


FIG. 2.9 – Pour la réduction de bruit GHKSS montrée à la Fig. 2.7d (Hénon, bruit de mesure gaussien de 5%), variation du nombre moyen de voisins et de la valeur rms des corrections scalaires au fil des itérations.

TAB. 2.5 – Méthode de Schreiber – Application de Hénon – Bruit de mesure

Distribution	$\kappa$ initial (%)	$\kappa$ final (%)	$r_0$	$r_{\text{dyn}}$	Nombre d'itérations
gaussienne	0.998	0.421	2.37	3.84	2
gaussienne	2.49	0.916	2.72	6.23	3
gaussienne	4.98	1.65	3.02	6.08	2
gaussienne	9.93	3.48	2.87	7.43	3
uniforme	1.0	0.407	2.46	3.86	3
uniforme	2.51	0.772	3.25	5.85	3
uniforme	5.01	1.47	3.42	6.39	3
uniforme	9.98	3.51	2.86	6.71	3

TAB. 2.6 – Méthode GHKSS – Application de Hénon – Bruit de mesure

Distribution	$\kappa$ initial (%)	$\kappa$ final (%)	$r_0$	$r_{\text{dyn}}$	Nombre d'itérations
gaussienne	0.998	0.293	3.41	8.38	3
gaussienne	2.49	0.757	3.3	10.3	3
gaussienne	4.98	1.25	3.99	12.7	4
gaussienne	9.93	3.58	2.8	12.3	4
uniforme	1.0	0.297	3.37	8.22	4
uniforme	2.51	0.74	3.39	9.19	3
uniforme	5.01	1.22	4.1	13.3	5
uniforme	9.98	3.94	2.55	11.7	5

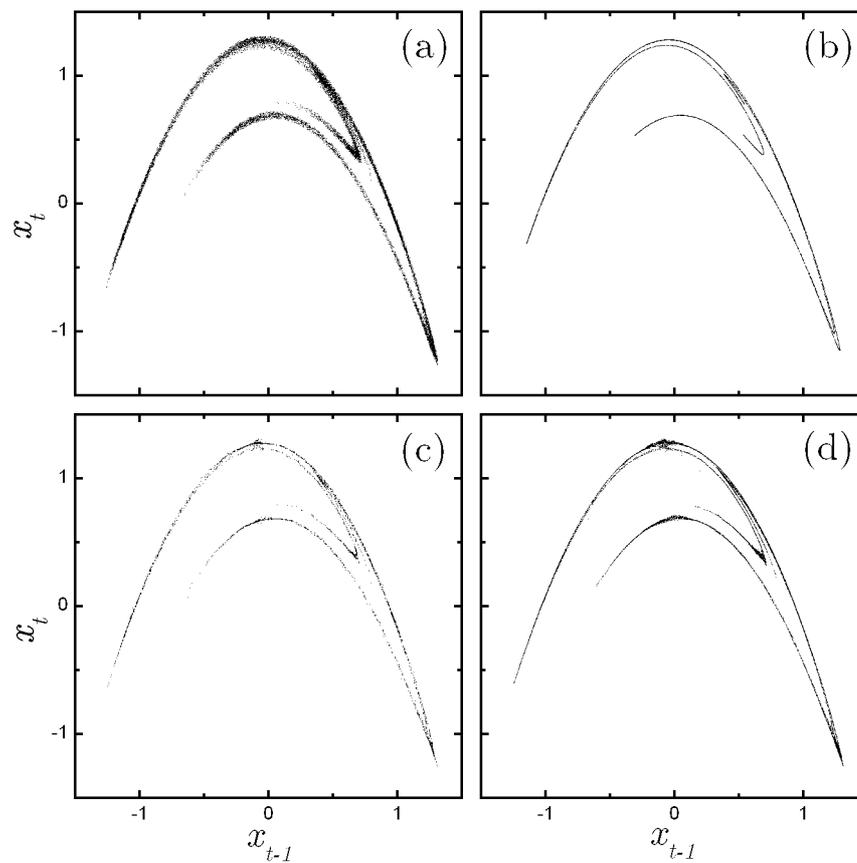


FIG. 2.10 – Graphes à délai de 16 383 points illustrant les performances de deux filtres non-linéaires. (a) Attracteur de Hénon contaminé par un *bruit dynamique gaussien de 2%*. (b) Attracteur de Hénon sans bruit. (c) Réduction de bruit Schreiber. (d) Réduction de bruit GHKSS.

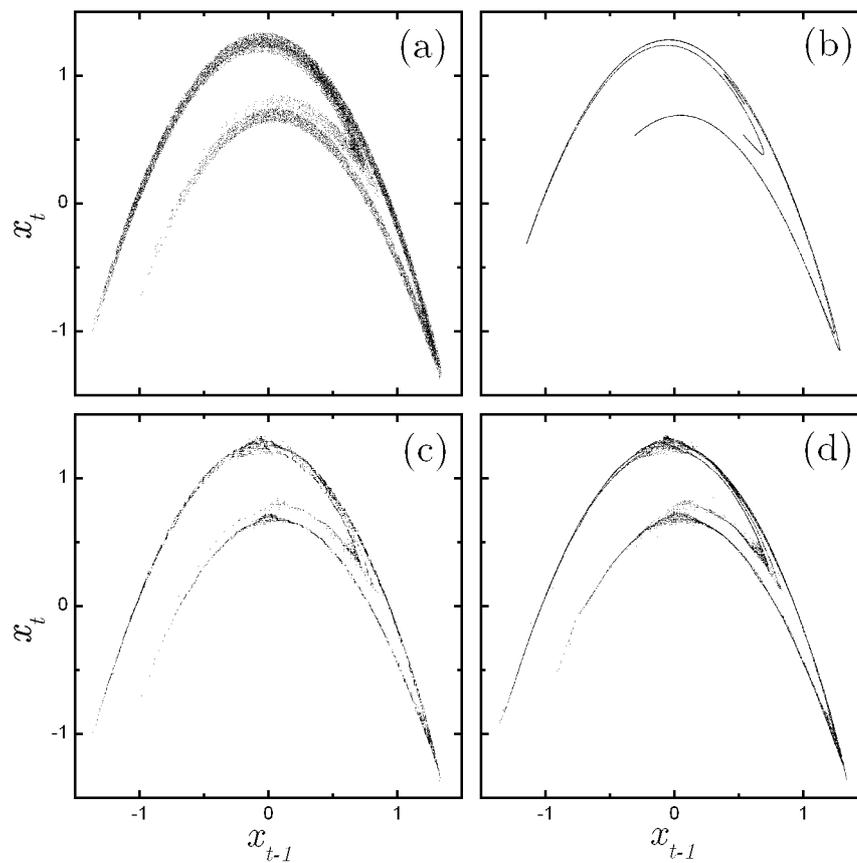


FIG. 2.11 – Graphes à délai de 16 383 points illustrant les performances de deux filtres non-linéaires. (a) Attracteur de Hénon contaminé par un *bruit dynamique uniforme de 4%*. (b) Attracteur de Hénon sans bruit. (c) Réduction de bruit Schreiber. (d) Réduction de bruit GHKSS.

TAB. 2.7 – Méthode de Schreiber – Application de Hénon – Bruit dynamique

Distribution	$\kappa$ initial (%)	$r_{\text{dyn}}$	Nombre d'itérations
gaussienne	1.0	2.06	4
gaussienne	2.0	2.41	5
gaussienne	3.0	2.61	4
gaussienne	4.0	2.74	4
uniforme	1.0	2.19	4
uniforme	2.0	2.59	4
uniforme	3.0	2.66	4
uniforme	4.0	2.85	4

TAB. 2.8 – Méthode GHKSS – Application de Hénon – Bruit dynamique

Distribution	$\kappa$ initial (%)	$r_{\text{dyn}}$	Nombre d'itérations
gaussienne	1.0	3.33	3
gaussienne	2.0	3.22	3
gaussienne	3.0	3.27	4
gaussienne	4.0	3.06	4
uniforme	1.0	3.24	3
uniforme	2.0	3.5	4
uniforme	3.0	3.46	4
uniforme	4.0	3.14	4

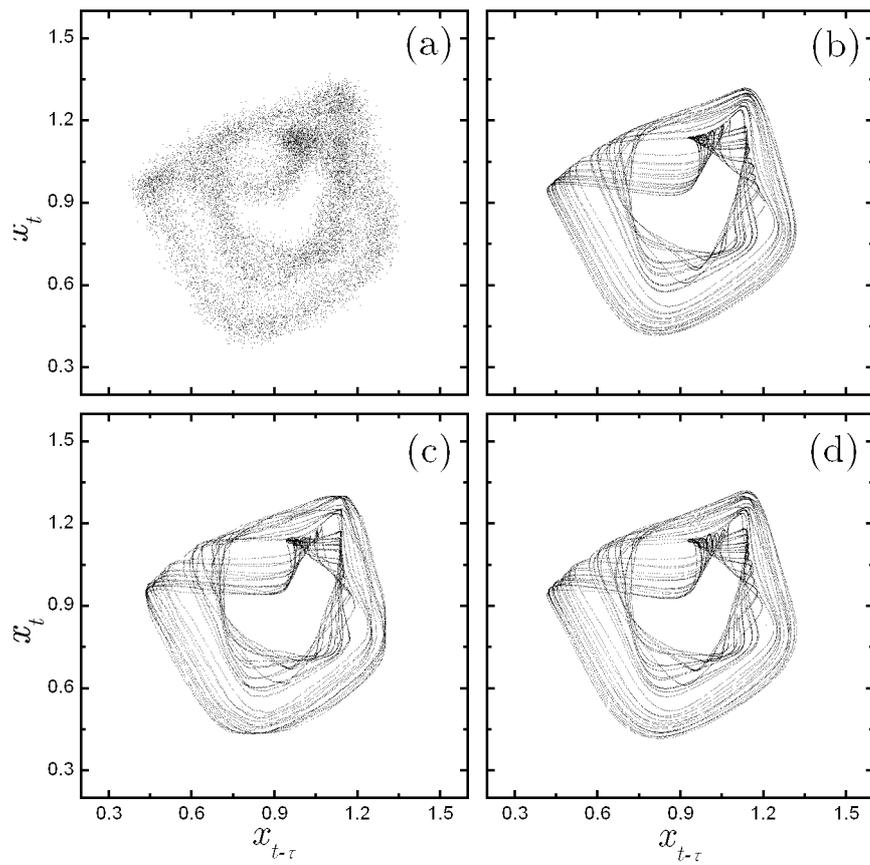


FIG. 2.12 – Graphes à délai ( $\tau=110$ ) de 16 274 points illustrant les performances de deux filtres non-linéaires. (a) Application de Mackey-Glass contaminée par un *bruit de mesure gaussien de 2.5%*. (b) Application de Mackey-Glass sans bruit. (c) Réduction de bruit Schreiber. (d) Réduction de bruit GHKSS.

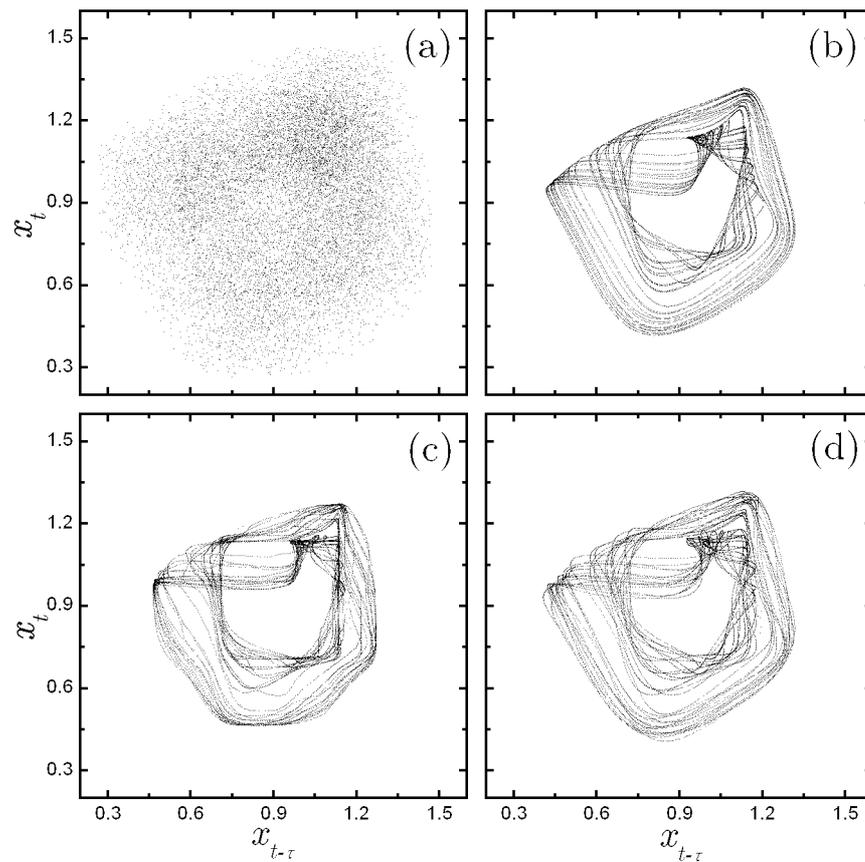


FIG. 2.13 – Graphes à délai ( $\tau=110$ ) de 16 274 points illustrant les performances de deux filtres non-linéaires. (a) Application de Mackey-Glass contaminée par un *bruit de mesure uniforme de 10%*. (b) Application de Mackey-Glass sans bruit. (c) Réduction de bruit Schreiber. (d) Réduction de bruit GHKSS.

TAB. 2.9 – Méthode de Schreiber – Application de Mackey-Glass – Bruit de mesure

Distribution	$\kappa$ initial (%)	$\kappa$ final (%)	$r_0$	$r_{\text{dyn}}$	Nombre d'itérations
gaussienne	0.998	0.364	2.74	16.5	2
gaussienne	2.49	0.641	3.90	23.2	2
gaussienne	4.98	1.21	4.12	25.6	3
gaussienne	9.93	2.51	3.98	25.7	4
uniforme	1.00	0.306	3.28	17.5	2
uniforme	2.51	0.558	4.49	33.1	2
uniforme	5.01	0.973	5.15	37.4	2
uniforme	9.97	1.90	5.28	41.5	3

TAB. 2.10 – Méthode GHKSS – Application de Mackey-Glass – Bruit de mesure

Distribution	$\kappa$ initial (%)	$\kappa$ final (%)	$r_0$	$r_{\text{dyn}}$	Nombre d'itérations
gaussienne	0.998	0.171	5.83	27.3	3
gaussienne	2.49	0.410	6.08	26.6	3
gaussienne	4.98	0.784	6.37	26.3	3
gaussienne	9.93	1.47	6.81	26.2	3
uniforme	1.00	0.168	5.97	49.2	4
uniforme	2.51	0.405	6.19	41.6	3
uniforme	5.01	0.780	6.42	42.6	3
uniforme	9.97	1.53	6.54	42.0	3

# Chapitre 3

## Non-stationnarité

Il a été mentionné précédemment qu'un système naturel n'est jamais parfaitement isolé. Dans certaines circonstances, plus particulièrement dans le cadre d'une expérience bien contrôlée réalisée en laboratoire, l'effet net des perturbations externes peut être ramené à la présence d'un faible bruit dynamique. Des techniques, certes encore à parfaire, sont actuellement disponibles pour traiter des signaux contaminés par ce type de bruit, comme il a été vu au chapitre précédent. Or dans de nombreux autres cas, l'influence externe est à ce point dirigée et importante qu'elle imprègne à la série temporelle observée des corrélations à long terme, ou encore une non-stationnarité manifeste.

### 3.1 Importance du problème

La présence de non-stationnarité engendre des complications évidentes dans l'analyse d'une série temporelle. En effet, presque toutes les méthodes d'analyse linéaires et non-linéaires requièrent, souvent implicitement, un certain niveau de stationnarité pour leur bon fonctionnement [Sch99]. Un premier exemple concerne l'estimation de la dimension de corrélation de l'attracteur étrange à l'origine d'une série temporelle

[OSY94]. Dans ce cas, il est impératif que la probabilité de «tirer» une paire de vecteurs à délais dont la distance est inférieure à  $\epsilon$  demeure constante au cours du temps, ou en d'autres termes soit stationnaire, et ce pour toute distance  $\epsilon$ .

Un second exemple provient de Timmer [Tim98] et concerne l'utilisation de la méthode des données synthétiques («surrogate data») [TEL<sup>+</sup>92] pour établir le caractère non-linéaire du processus à l'origine d'un signal. En bref, on y rappelle que le rejet de l'hypothèse nulle de linéarité *et stationnarité* peut être expliqué par un processus stochastique linéaire mais *non stationnaire*. Il faut donc au préalable s'assurer d'une certaine stationnarité avant d'associer directement le rejet de l'hypothèse nulle à un processus non-linéaire. Et comme le souligne avec insistance Timmer, le processus non-linéaire identifié par le test peut être aussi bien stochastique que déterministe. Attention aux conclusions hâtives.

Ainsi, il semble que la première étape de presque toute analyse d'une série temporelle réelle devrait consister à tester s'il y a stationnarité ou non. Devant un constat de non-stationnarité, quatre alternatives s'offrent alors :

1. Dans certains cas, la non-stationnarité peut prendre la forme d'une simple tendance que l'on peut retrancher efficacement («detrending»).
2. Une approche plus générale consiste à découper convenablement la série temporelle en plusieurs fenêtres, de telle sorte qu'à l'intérieur de chacune la dynamique puisse être considérée quasi stationnaire.
3. Abandonner l'analyse conventionnelle et reprendre l'expérience en tentant de mieux isoler le système sous observation, si cela est possible.
4. Étudier de près la non-stationnarité, c.-à-d. s'intéresser aux changements dynamiques qui s'opèrent au cours du temps. Là est peut-être le phénomène scientifique de première importance. Un exemple incontournable concerne l'analyse de l'électroencéphalogramme (EEG) d'un patient épileptique : l'avènement d'une crise est associé à une perte de complexité dynamique.

L'une des difficultés majeures ici concerne la définition même du concept de stationnarité *pour une série temporelle*. C'est un problème ouvert, et ce malgré l'existence de définitions asymptotiques, qui ne sont applicables formellement qu'aux systèmes sous-jacents. Ainsi, pour un processus stochastique  $X_t$ , on distingue deux types de stationnarité :

1. *Stationnarité forte* : Pour tout ensemble  $\{t_1, t_2, \dots, t_n\}$  et pour tout  $\tau$ , les densités de probabilité  $p(x_1, t_1; x_2, t_2; \dots; x_n, t_n)$  et  $p(x_1, t_1 + \tau; x_2, t_2 + \tau; \dots; x_n, t_n + \tau)$  sont identiques. Ceci implique entre autres que la densité de probabilité 1D est indépendante du temps.
2. *Stationnarité faible* : Les moments du premier et du second ordre (moyenne, variance) sont indépendants du temps et la fonction d'autocorrélation ne dépend que du temps relatif. Il est à noter que ce type de stationnarité est insuffisant pour les méthodes non-linéaires d'analyse des séries temporelles [Sch97]. La stationnarité forte englobe la stationnarité faible.

Dans ce cadre conceptuel, la série temporelle est vue comme une réalisation d'un processus stochastique. Or la stationnarité telle que définie ci-haut se rapporte formellement au processus et non à une réalisation. Qui plus est, en pratique, les probabilités de transition ne sont pas connues, et on ne peut qu'en estimer un nombre fini à partir de la série temporelle.

Dans l'optique mathématique des systèmes dynamiques, la stationnarité sera plutôt associée au flot, i.e. constance des paramètres. Encore là, même un système dynamique stationnaire peut générer une série temporelle qui à toutes fins pratiques sera non stationnaire. Le cas type est celui de l'intermittence, où un mouvement régulier est entrecoupé de fréquentes bouffées chaotiques, et ce malgré, par exemple, la constance du paramètre de l'application logistique, peu avant la fenêtre de période 3 [Str94]. Ainsi, comme le souligne Schreiber [Sch99], le fait que le système demeure inchangé pendant la prise des mesures n'est pas une condition suffisante, ni même nécessaire, pour la stationnarité d'une série temporelle.

Principalement au cours des cinq dernières années, plusieurs groupes de recherche ont proposé des algorithmes de détection de non-stationnarité pour les séries temporelles, algorithmes que l'on peut diviser en deux grandes familles :

1. *Test binaire* : l'algorithme se contente d'identifier la série temporelle comme stationnaire ou non stationnaire.
2. *Résolution temporelle* : l'algorithme permet de suivre les changements dynamiques qui s'opèrent au cours du temps.

Pour concevoir leur algorithme, chaque groupe de recherche a convenu d'examiner la non-stationnarité selon un certain angle, et il est à se demander si un peu de subjectivité ne s'est pas introduit dans le domaine, profitant de l'absence d'une définition commune. Dans un cadre pratique, il vaut peut-être mieux se réjouir devant la richesse des outils disponibles, car chaque algorithme peut être particulièrement adapté à certaines situations.

## 3.2 Survol de la littérature

Au chapitre suivant, une nouvelle méthode de détection de non-stationnarité sera présentée. Afin de situer celle-ci par rapport aux méthodes déjà existantes, il convient d'examiner ce que propose la littérature au problème de la non-stationnarité. Ce survol n'est pas exhaustif, mais il est très représentatif de ce qui a été suggéré plus récemment. Les trois premières méthodes appartiennent à la famille des tests binaires, alors que les suivantes permettent une résolution temporelle de la non-stationnarité.

### 3.2.1 Kennel

La référence [Ken97] propose un test statistique pour diagnostiquer la non-stationnarité d'une série temporelle essentiellement déterministe. L'idée est de surveiller indirectement une éventuelle dérive de la distribution des points dans l'espace des

phases reconstruit. Pour ce faire, on regroupe chaque vecteur à délais  $\mathbf{s}$  à son plus proche voisin  $\mathbf{s}^{\text{nn}}$ , et on calcule pour chaque paire le temps qui les sépare en terme d'indices temporels :

$$D = |T(\mathbf{s}^{\text{nn}}) - T(\mathbf{s})|, \quad (3.1)$$

où  $T$  dénote l'indice temporel du point, i.e.  $T(\mathbf{s}_t) = t$ . La non-stationnarité entraîne les voisins à être plus particulièrement rapprochés dans le temps, alors que dans le cas contraire, l'indice temporel d'un proche voisin est a priori indépendant de celui du point de référence. On peut donc élaborer un test statistique qui compare la distribution observée de  $D$  à celle prévue dans le cadre d'une série temporelle stationnaire.

Dans un test statistique classique, on présume les observations *indépendantes*. L'origine dynamique des données ici oblige à prendre certaines précautions en ce sens. La première consiste à imposer une valeur minimale  $D_{\text{min}}$ , de telle sorte qu'un proche voisin soit réellement indépendant du point de référence. La seconde s'affaire à ne compter qu'une seule fois la valeur de  $D$  associée à deux segments de trajectoire composés de proche voisins. L'auteur démontre clairement l'importance de ces corrections.

La distribution attendue pour  $D$  peut être générée facilement à l'ordinateur, ou si la série temporelle est suffisamment longue, estimée de façon analytique. De cette distribution hypothétique on extrait ensuite la médiane  $D^*$  afin de performer un test unilatéral pour une proportion [LM93]. Pour ce faire, on évalue la proportion  $p$  des valeurs observées telles que  $D < D^*$ . Pour l'hypothèse nulle associée à la stationnarité, on a trivialement  $p_0 = 0.5$ . La comparaison de  $p$  et  $p_0$  selon une règle de décision permet de rejeter ou non l'hypothèse nulle à un certain niveau de signification.

Ce test statistique de non-stationnarité, conçu soigneusement par Kennel, fonctionne apparemment très bien. Néanmoins, sa principale lacune repose sur le fait qu'il ne permet pas de suivre les changements dynamiques au cours du temps.

### 3.2.2 Witt, Kurths et Pikovsky

Witt et coauteurs [WKP98] proposent un test de stationnarité basé principalement sur le concept de stationnarité tel que défini en statistique mathématique. Tout s'articule donc autour d'un processus stochastique.

Ainsi, l'approche de Witt, Kurths et Pikovsky consiste à combiner un test vérifiant l'indépendance temporelle de la densité de probabilité 1D à un test vérifiant l'équivalent pour le spectre en puissance. On teste donc une stationnarité un peu plus sévère que la version faible. Le point de départ, classique, revient à découper la série temporelle en fenêtres qui ne se chevauchent pas. Chaque test est doté d'une statistique ( $\chi^2$  modifié) qui permet de comparer les fenêtres entre elles. (Une distribution attendue est estimée pour chacune des deux statistiques.)

Cette méthode doit être utilisée avec vigilance. En particulier, dans le cas d'un système déterministe, les auteurs recommandent d'abandonner le premier test en raison d'une distribution attendue inadéquate. De façon plus générale, on note que plus la longueur de corrélation du processus est importante et plus le bon fonctionnement du test combiné requiert des fenêtres suffisamment longues. Malgré ces subtilités, de bons résultats sont obtenus, et on arrive même à différencier divers régimes dynamiques dans le cas d'un système de haute dimension.

### 3.2.3 Yu, Lu et Harrison

Il n'y a rien de fondamentalement neuf dans le papier de Yu et coauteurs [YLH98] par rapport à celui de Kennel [Ken97] : on troque essentiellement le test statistique pour une simple inspection visuelle. Il y a toutefois quelques légères modifications. La séparation temporelle entre deux vecteurs à délais prend maintenant la forme

$$\mathcal{D} = \begin{cases} T(\mathbf{s}_j) - T(\mathbf{s}_i) = j - i & \text{si } j > i \\ N_{\tau,m} + T(\mathbf{s}_j) - T(\mathbf{s}_i) = N_{\tau,m} + j - i & \text{si } j < i, \end{cases} \quad (3.2)$$

ce à quoi les auteurs se réfèrent sous l'expression d'indice temporel entre deux points. Rappelons que  $N_{\tau,m}$  désigne simplement le nombre de vecteurs à délais. Cette subtilité entraîne la distribution de  $\mathcal{D}$  sous l'hypothèse de stationnarité à être uniforme plutôt que triangulaire. Dans l'accumulation des différentes valeurs de  $\mathcal{D}$ , on ne s'intéresse plus seulement au plus proche voisin, mais bien à tous les voisins contenus dans un voisinage de rayon  $\epsilon$  (stabilité statistique accrue). La première précaution de Kennel demeure, alors que la seconde est abandonnée. La distribution observée de  $\mathcal{D}$  est normalisée en utilisant  $K + 1$  intervalles. On obtient alors la distribution  $\rho_\epsilon(\xi)$ , où  $\xi \in [0, K]$  est l'indice temporel normalisé.

Les auteurs suggèrent de considérer plusieurs valeurs de  $\epsilon$ , et de porter l'ensemble des distributions ainsi générées sur un même graphique, appelé graphique espace-indice temporel («space time-index plot») ou graphique  $\epsilon$ - $\xi$ . Pour les mêmes raisons que celles avancées par Kennel, la non-stationnarité aura pour effet de dévier les distributions de l'uniformité, et plus spécifiquement, celles-ci seront en forme de V. Plus  $\epsilon$  est élevé et plus la distribution correspondante se rapproche de la distribution uniforme, pour des raisons évidentes. Cette procédure appliquée à des séries artificielles et expérimentales donne de très bons résultats. Néanmoins, sa profonde similarité avec le test statistique de Kennel lui confère les mêmes points faibles, entre autres l'impossibilité d'étudier en temps réel les changements dynamiques.

Note : très récemment, les mêmes auteurs ont raffiné leur méthode, en introduisant notamment l'indice temporel croisé («cross-time index») et un test du  $\chi^2$ . Voir la référence [YLH99].

### 3.2.4 Manuca et Savit

L'approche proposée par Manuca et Savit au problème de la non-stationnarité est a priori convaincante et semble solide [MS96]. L'hypothèse de travail, relativement souple, consiste à décomposer le système en une dynamique interne stationnaire  $\mathbf{x}$ ,

sujette à une force externe  $\mathbf{a}$  – entendu au sens large de paramètre(s) – potentiellement non stationnaire :

$$\dot{\mathbf{x}} = \mathbf{G}(\mathbf{a}; \mathbf{x}) \quad (3.3)$$

$$\dot{\mathbf{a}} = \mathbf{H}(\mathbf{a}, t). \quad (3.4)$$

La dynamique de la force externe (éq. (3.4)) est dite métadynamique.

La série temporelle à analyser est décomposée en  $W$  fenêtres qui ne se chevauchent pas. De façon standard, on procède alors à un recouvrement. L'idée est d'utiliser une mesure de proximité dynamique («measure of dynamical closeness») afin de comparer les fenêtres entre elles. À l'intérieur de chacune, on suppose la force externe quasi constante et donc la dynamique quasi stationnaire. La statistique retenue par les auteurs, symétrique, est la suivante (fenêtres  $i$  et  $j$ ) :

$$d_{ij}^{(m)}(\epsilon) = \max \left\{ d_{i \leftarrow j}^{(m)}(\epsilon), d_{j \leftarrow i}^{(m)}(\epsilon) \right\}, \quad (3.5)$$

où

$$d_{i \leftarrow j}^{(m)}(\epsilon) = \sqrt{\left| R_{ii}^{(m)}(\epsilon) - R_{ij}^{(m)}(\epsilon) \right|}. \quad (3.6)$$

La quantité  $R_{ij}^{(m)}(\epsilon)$  mesure indirectement l'amélioration de la prédiction à un pas dans la fenêtre  $j$  utilisant la fenêtre  $i$ , ou vice versa, en passant d'une dimension de recouvrement de  $m - 1$  à  $m$ . L'évaluation de  $R_{ij}^{(m)}(\epsilon)$  se fait au moyen de l'intégrale de corrélation croisée («cross-correlation integral») :

$$C_{ij}^{(m)}(\epsilon) = \text{Probabilité} \left( \|\mathbf{x}_i^{(m)} - \mathbf{x}_j^{(m)}\| < \epsilon \right), \quad (3.7)$$

où  $\mathbf{x}_i^{(m)}$  est un vecteur à délais tiré de la fenêtre  $i$ . Le rapport

$$S_{ij}^{(m)}(\epsilon) = \frac{C_{ij}^{(m+1)}(\epsilon)}{C_{ij}^{(m)}(\epsilon)} \quad (3.8)$$

est identifié à la probabilité conditionnelle que deux points – l'un de la fenêtre  $i$  et l'autre de la fenêtre  $j$  – soient à une distance inférieure à  $\epsilon$  en  $m + 1$  dimensions, sachant que tel est le cas en  $m$  dimensions. Ceci est relié à la capacité de prédire le

prochain scalaire dans la fenêtre  $i$  donné les  $m$  précédents, utilisant l'information de la fenêtre  $j$  (ou vice versa). Par conséquent, le rapport

$$R_{ij}^{(m)}(\epsilon) = \frac{S_{ij}^{(m)}(\epsilon)}{S_{ij}^{(m-1)}(\epsilon)} \quad (3.9)$$

a bien la signification avancée plus haut. Si l'un désire travailler en 1D, il est alors nécessaire de remplacer  $R$  par  $C$  dans l'équation (3.6), pour des raisons évidentes. Autrement, la quantité  $R$  demeure celle avec la meilleure puissance discriminatoire.

Une valeur élevée de  $d_{ij}^{(m)}(\epsilon)$  est donc associée à une dissimilitude dynamique entre les deux fenêtres (forces externes différentes). Évidemment, il convient de normaliser adéquatement la mesure de proximité et de choisir un seuil approprié. Néanmoins, étant donné que  $d_{ij}^{(m)}(\epsilon)$  n'est qu'approximativement une métrique<sup>1</sup> et que son évaluation repose sur un nombre limité de données, il est préférable de se doter d'un procédé plus robuste de comparaison. La nouvelle statistique est calculée de la façon suivante :

$$M_{ij} = \sum_{k=1}^W \Theta(|d_{ik} - d_{jk}| - \sqrt{2}), \quad (3.10)$$

où  $\Theta$  est la fonction de Heaviside. On compare donc les deux fenêtres par rapport à l'ensemble des fenêtres.

Cet ensemble couvre un certain nombre de métaétats différents, et donc de dynamiques quasi stationnaires différentes. Manuca et Savit proposent un algorithme, basé sur  $M_{ij}$ , qui permet de regrouper les fenêtres en amas, où à l'intérieur de chacun d'eux les fenêtres présentent une dynamique similaire, ou en d'autres termes, la même force externe (dans les limites de l'analyse). Cette démarche présente trois intérêts majeurs :

1. À l'intérieur de chaque amas, on peut rabouter les fenêtres membres et étudier la dynamique sous des conditions statistiques considérablement améliorées.

---

<sup>1</sup>Si  $\mathbf{d}$  n'est pas une métrique, certaines inconsistances peuvent s'y abriter, nonobstant les limitations statistiques. Par exemple,  $d_{ij}$  et  $d_{ik}$  pourraient être très proches de zéro en coexistant avec une valeur élevée de  $d_{jk}$ .

2. En assignant un symbole à chaque amas, on a accès à une métadynamique symbolique. La non-stationnarité se visualise alors directement.
3. En choisissant quelques amas de référence pour former un espace des métaphases et en «projetant» chaque fenêtre dans cet espace, on reconstruit en quelque sorte l'évolution de la force externe. Ceci peut révéler certaines caractéristiques de la non-stationnarité, e.g. dérive lente de paramètres.

Il est important de rappeler que l'analyse est réalisée pour une dimension de recouvrement  $m$ , à une résolution de  $\epsilon$ , et à une échelle de temps fixée par la longueur de chaque fenêtre. Ce qui peut apparaître non stationnaire sous un certain angle peut devenir stationnaire sous un nouveau point de vue (stationnarité conditionnelle). Si le but visé est d'établir la stationnarité fondamentale de la série temporelle, la stratégie revient à diminuer progressivement le nombre de fenêtres et augmenter la dimension de recouvrement. Si à la limite perdure une non-stationnarité, on doit s'y résigner.

Dans l'article de Manuca et Savit, plusieurs séries temporelles, artificielles et expérimentales, ont été analysées selon leur méthode. Les résultats sont très positifs, et surpassent de beaucoup ceux obtenus en utilisant uniquement la fonction d'autocorrélation. Particulièrement intéressant est le dernier exemple, où survient une catastrophe dans un système mécanique (à la fin de la série temporelle). Si la fonction d'autocorrélation permet d'anticiper la catastrophe 5 secondes auparavant, l'espace des métaphases recule cette détection à 3 minutes. Soulignons enfin que cette méthode peut facilement être utilisée en temps réel.

### 3.2.5 Schreiber

Schreiber (Réfs. [Sch97] et [Sch99]) préconise une approche qui n'est pas sans rappeler celle de Manuca et Savit [MS96]. La série temporelle est découpée en plusieurs fenêtres, qui pourraient se chevaucher. La comparaison des fenêtres entre elles est réalisée au moyen de l'erreur de prédiction croisée  $\gamma_{ij}^{(m)}(\epsilon)$  («cross-prediction error»).

La fenêtre  $i$  est utilisée comme base pour effectuer des prédictions élémentaires (approximation constante de la dynamique locale au moyen des proches voisins) dans la fenêtre  $j$ , et l'erreur de prédiction rms qui en résulte est précisément  $\gamma_{ij}^{(m)}(\epsilon)$ .

Dans la référence [Sch97], Schreiber suggère de visualiser directement la matrice  $\gamma_{ij}$  au moyen d'un dégradé de gris. Nous trouvons difficile l'interprétation d'un tel graphique. Dans la référence [Sch99], on invoque plutôt la détermination d'amas de dynamiques similaires au moyen de  $\gamma_{ij}$ , et de leur utilisation pour construire un espace des paramètres dans lequel on peut suivre les changements dynamiques. Tout comme celle de Manuca et Savit [MS96], la méthode de Schreiber (version amas) semble prometteuse.

### 3.2.6 Cao, Mees et Judd

Le papier de Cao et coauteurs [CMJ97], simple et clair, est intéressant. On n'y présente pas de mesure de non-stationnarité proprement dite, mais on est en mesure de produire un diagramme de bifurcation à partir d'une série temporelle. Pour parvenir à ce résultat, on modélise la dynamique non stationnaire par un modèle régressif dépendant du temps («Time-dependent regressive (TDR) model») :

$$s_t = f(t; s_{t-\tau_1}, s_{t-\tau_2}, \dots, s_{t-\tau_m}). \quad (3.11)$$

Le temps joue le rôle du paramètre de bifurcation. Le modèle spécifique utilisé par les auteurs consiste en un réseau d'ondelettes («wavelet network»), ce qui s'apparente aux fonctions radiales.<sup>2</sup> En utilisant des séries artificielles assez courtes (quelques centaines de données), les auteurs démontrent la valeur prédictive de leur TDR. Le

---

<sup>2</sup>Les fonctions radiales («radial basis functions») sont couramment utilisées pour construire une estimation globale  $\mathbf{y}_{t+1} = \hat{\mathbf{f}}(\mathbf{y}_t)$  de la dynamique dans l'espace des phases reconstruit (Réfs. [Cas87] et [Sch99]). Une fonction radiale est de la forme  $\Phi(\|\mathbf{y} - \mathbf{y}^{(c)}\|)$ , où  $\mathbf{y}^{(c)}$  est le centre de la fonction et où  $\Phi$  sera par exemple une gaussienne.

diagramme de bifurcation est ensuite obtenu en itérant le modèle pour chaque valeur de  $t$ , diagramme relativement juste.

S'il est suffisamment précis, le diagramme de bifurcation révèle directement les changements de régime du système. Leurs résultats suggèrent également que cette approche est robuste par rapport au bruit de mesure et au bruit dynamique. Néanmoins, seule l'application d'Ikeda a été considérée, et il reste à voir comment se comporte une telle procédure en présence de données réelles. À surveiller.

### 3.2.7 Hively, Gailey et Protopopescu

L'approche de Hively et al. [HGP99] consiste à scinder la série temporelle en deux segments inégaux : le premier, beaucoup plus court, servira de base pour comparer avec une fenêtre glissante qui balaira le second. Il s'agit bien entendu d'une comparaison entre régimes dynamiques, possiblement différents en raison de la présence de non-stationnarité. Pour réaliser cette comparaison, les auteurs suggèrent d'utiliser la densité invariante [BS93] dans l'espace des phases reconstruit.

Afin de capturer plus d'information dynamique, on emploie plutôt l'*espace des phases connecté*, peuplé de vecteurs à  $2m$  dimensions  $\mathbf{Y}(t) = (\mathbf{y}(t), \mathbf{y}(t+1))$ . Cet espace est ensuite recouvert d'hypercubes afin d'y évaluer une densité invariante discrétisée, à la fois pour le segment de base et pour la fenêtre glissante. Les deux densités sont alors comparées directement au moyen d'un test du  $\chi^2$  ([LM93] et [PFTV92]).

Pour tester leur méthode, Hively et al. utilisent des données artificielles (système de Lorenz [Str94]) et des données réelles (EEG). En parallèle au test du  $\chi^2$ , on utilise également le premier minimum de l'information mutuelle, la dimension de corrélation et l'entropie de Kolmogorov pour réaliser les comparaisons dynamiques. Ces mesures non-linéaires, couramment utilisées pour détecter la non-stationnarité, constituent des descriptions sommaires de la dynamique, et sont donc, selon les auteurs, moins

précises que la comparaison directe des densités invariantes dans l'espace des phases connecté. Ce fait est particulièrement marqué lors de la discrimination entre régimes chaotiques très proches, et les résultats du groupe de recherche illustrent la remarquable sensibilité de leur test basé sur la densité invariante.

# Chapitre 4

## Flot d'information

Dans le présent chapitre, une nouvelle méthode de détection de non-stationnarité pour une série temporelle sera présentée et étudiée. Cette approche novatrice, qui permet une résolution temporelle de la non-stationnarité, repose sur le concept de flot d'information, concept qui sera introduit dès la prochaine section.

### 4.1 Flot d'information d'un signal stationnaire

#### 4.1.1 Approche formelle

Afin de caractériser la dynamique inhérente à un signal stationnaire, il peut être fort révélateur de s'intéresser à la transmission d'information du passé vers le futur, ou en d'autres termes, à la question de savoir de quelle façon le passé conditionne le futur. Le cadre naturel pour aborder une telle démarche s'avère indubitablement l'utilisation d'une dynamique symbolique (Réfs. [BS93] et [OSY94]), de concert évidemment avec les concepts de la théorie de l'information (Réfs. [BS93] et [DO96]).

Suivant une telle approche, l'étape première consiste donc à encoder la dynamique de la série temporelle sous une forme symbolique. Pour ce faire, on morcelle la

région visitée  $A$  de l'espace des phases reconstruit (e.g. attracteur étrange) selon une partition  $\beta$ , qui consiste en  $R$  sous-ensembles disjoints  $B_i$  de  $A$ , i.e.

$$\beta = \{B_i\}_{i=1}^R, \quad \cup_{i=1}^R B_i = A \quad \text{et} \quad B_i \cap B_j = \emptyset \text{ si } i \neq j. \quad (4.1)$$

En assignant à chaque «boîte»  $B_i$  le symbole  $i$  (alphabet de  $R$  symboles), on peut donc transformer la série temporelle originale  $\{y_t\}$  en une séquence de symboles  $\{i_t\}$ , appelée itinéraire.

On introduit ensuite la théorie de l'information. Soit  $p_i^\beta$  la probabilité d'observer le symbole  $i$  de la partition  $\beta$  en suivant l'itinéraire, ou en des termes plus précis, en tirant au hasard un symbole de l'itinéraire. Dans le cas d'un attracteur étrange (série temporelle chaotique), cette probabilité est reliée directement à la mesure naturelle [OSY94] du sous-ensemble  $B_i$ , obtenue au moyen de la densité invariante. On définit alors l'*entropie de Shannon* associée à l'itinéraire de la façon suivante :

$$H^\beta = - \sum_{i=1}^R p_i^\beta \log_2 p_i^\beta \quad (\text{bit}). \quad (4.2)$$

Cette quantité, toujours positive, mesure l'incertitude inhérente à l'apparition d'un symbole lors du suivi de l'itinéraire. En effet, si par exemple seul le symbole  $j$  est possible, i.e.  $p_i^\beta = \delta_{ij}$ , alors il n'y a pas d'incertitude et l'entropie de Shannon est nulle. D'autre part, l'incertitude maximale régnera si chaque symbole est équiprobable, et de façon cohérente,  $H^\beta$  sera également maximal (valeur maximale :  $\log_2 R$ ).<sup>1</sup>

Supposons que l'itinéraire est tel, ou enfin que la dynamique est telle, que seulement  $R_n$  blocs de  $n$  symboles consécutifs peuvent être obtenus<sup>2</sup>, et soit  $p_k^{\beta,n}$  la probabilité d'observer le bloc  $k \in [1, R_n]$ . Par extension, on introduit l'*entropie de bloc* de la dynamique symbolique :

$$H^\beta(n) = - \sum_{k=1}^{R_n} p_k^{\beta,n} \log_2 p_k^{\beta,n}. \quad (4.3)$$

---

<sup>1</sup>Pour  $p \in [0, 1]$ , la fonction  $p \ln p$  demeure finie, est négative et s'annule aux deux extrémités. Par conséquent, seule l'utilisation d'un alphabet infini peut entraîner la divergence de l'entropie de Shannon.

<sup>2</sup>Si tous les blocs sont possibles, alors naturellement  $R_n = R^n$ .

Si l'on désigne ensuite par  $p_{k,i}^{\beta,n,r}$  la probabilité d'observer le bloc  $k$  et  $r$  intervalles de temps plus tard le symbole  $i$ , on peut définir l'*entropie conditionnelle* associée à la prédiction à  $r$  pas donnée la réalisation d'un bloc de longueur  $n$  :

$$H^\beta(n+r|n, \dots, 1) = - \sum_{k=1}^{R_n} \sum_{i=1}^R p_{k,i}^{\beta,n,r} \log_2 \left( \frac{p_{k,i}^{\beta,n,r}}{p_k^{\beta,n}} \right). \quad (4.4)$$

Comme le conditionnement réduit l'entropie, un fait intuitif qui peut être démontré aisément [DO96], et que l'entropie est une quantité positive, l'inégalité suivante tient :

$$0 \leq H^\beta(n+r|n, \dots, 1) \leq H^\beta. \quad (4.5)$$

En fait, le cas maximal est atteint lorsqu'il y a indépendance statistique totale, i.e.  $p_{k,i}^{\beta,n,r} = p_k^{\beta,n} p_i^\beta$ , et la divergence n'est possible que lors de l'utilisation d'un alphabet infini (partition infinitésimale). À l'opposé, dès que le bloc détermine parfaitement le symbole qui se trouve  $r$  pas plus loin, i.e.  $p_{k,i}^{\beta,n,r} = p_k^{\beta,n} \delta_{ij(k)}$  où  $j = j(k)$ , l'entropie conditionnelle de l'équation (4.4) devient nulle.

Afin d'obtenir une quantité plus intrinsèque à la dynamique, il est préférable dans un premier temps de se débarrasser de la dépendance par rapport à longueur  $n$  du bloc passé. On peut corriger cette lacune en considérant tout le passé, et pour  $r = 1$  (prédiction à un pas), on obtient ainsi l'*entropie de source* :

$$h^\beta = \lim_{n \rightarrow \infty} H^\beta(n+1|n, \dots, 1). \quad (4.6)$$

Enfin, le choix d'une partition ne devrait pas influencer la caractérisation de la dynamique, d'où l'utilisation de l'*entropie de Kolmogorov-Sinai* :

$$h = \sup_{\beta} h^\beta. \quad (4.7)$$

Le supremum se rapporte à l'ensemble des partitions possibles, et une partition pour laquelle le supremum est atteint est appelée une *partition génératrice*. L'utilisation d'une partition génératrice finie ( $R$  fini) peut être associée à l'utilisation d'un instrument de mesure de précision limitée. Dans cette perspective, l'entropie

de Kolmogorov-Sinai chiffre de façon formelle l'incertitude entourant une prochaine mesure lorsque tout le passé est connu avec une certaine précision.

Pour un signal périodique,  $h$  est nul, alors que pour un processus stochastique,  $h$  diverge. Entre ces deux extrêmes, c'est le domaine du chaos déterministe, que l'on définit souvent justement par la condition  $0 < h < \infty$ . En outre, comme l'entropie de Kolmogorov-Sinai est reliée à la sensibilité aux conditions initiales, elle permet également de quantifier l'«intensité chaotique» d'un système.<sup>3</sup> L'entropie de Kolmogorov-Sinai constitue donc certainement une excellente mesure pour discriminer plusieurs dynamiques différentes.

La recherche du supremum, ou encore d'une partition génératrice, représente toutefois une tâche fort ardue, et ce même pour des systèmes artificiels. Fort heureusement, il existe une approche plus systématique. Celle-ci consiste à couvrir l'espace des phases d'une grille d'hypercubes, tous d'arête  $\epsilon$ , et de prendre la limite  $\epsilon \rightarrow 0$  (partition infinitésimale) :

$$h = \lim_{\epsilon \rightarrow 0} h^{\beta(\epsilon)}. \quad (4.8)$$

Selon Beck et Schlögl [BS93], les deux définitions devraient converger dans la plupart des cas.<sup>4</sup>

Comme l'entropie de Kolmogorov-Sinai correspond à l'information qui manque en moyenne pour spécifier un prochain état, donné le passé complet, il permet d'étudier le flot d'information du passé vers le futur. Maintenant, pour une étude plus complète, il faudra peut-être considérer le cas général d'une prédiction à  $r$  pas, et pour ce faire, on introduit l'*entropie de Kolmogorov-Sinai généralisée* [DSS97c] :

$$h_r = \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} H^{\beta(\epsilon)}(n+r|n, \dots, 1). \quad (4.9)$$

L'étude de  $h_r$  en fonction de  $r$  pourrait donc permettre, a priori, de discriminer encore plus efficacement divers processus.

<sup>3</sup>Pour une certaine classe d'applications 1D, l'entropie de Kolmogorov-Sinai coïncide en fait avec l'exposant de Lyapunov. Voir la référence [BS93] pour plus de détails.

<sup>4</sup>Évidemment, pour des données réelles, la limite  $\epsilon \rightarrow 0$  est tout aussi inaccessible.

En fait, le gain concerne davantage les situations pratiques que théoriques. En effet, Deco et al. [DSS97c] ont démontré que pour un système chaotique,  $h_r$  croît linéairement avec  $r$ , et la pente de cette relation correspond précisément à  $h$ , i.e.  $h_r = hr$ . Par conséquent, l'entropie de Kolmogorov-Sinai décrit complètement le flot d'information d'un système chaotique. De plus, la présence d'une composante stochastique fera également diverger  $h_r$ , peu importe  $r$ . Par ailleurs, dans le cadre de la caractérisation d'une série temporelle stationnaire, Schittenkopf et Deco [SD97] parviennent à différencier efficacement un bruit pur, un signal chaotique, un signal chaotique contaminé par un bruit et enfin un processus de Markov [Gar85] au moyen d'une variante de l'entropie de Kolmogorov généralisée (« $\epsilon$ -information flow»). L'astuce consiste à omettre la limite  $\epsilon \rightarrow \infty$ , et à observer la dépendance par rapport à  $\epsilon$  de pair avec celle par rapport à  $r$ .

### 4.1.2 Approche pratique

Plus récemment encore, Deco et al. [DSS97b] ont proposé une approche beaucoup plus simple à l'étude du flot d'information d'un processus stationnaire, approche particulièrement adaptée à l'étude pratique des séries temporelles. L'idée consiste à examiner directement la dépendance statistique entre le passé et des valeurs futures, en invoquant le concept probabiliste de *cumulant* («cumulant-based information flow»).

Considérons un bloc de  $n$  points consécutifs et le point situé  $r$  intervalles d'échantillonnage plus tard. Pour un processus stationnaire, ce dispositif conceptuel est complètement décrit par un vecteur de variables aléatoires continues

$$\mathbf{X} = (X_1, \dots, X_n, X_{n+r}), \quad (4.10)$$

dont la distribution de probabilité sera donnée par la fonction de densité  $p(\mathbf{x})$ .<sup>5</sup> L'indépendance statistique entre le «passé»  $(X_1, \dots, X_n)$  et le «futur»  $X_{n+r}$  se traduit

---

<sup>5</sup>La convention suivante est utilisée : la lettre majuscule est réservée à la variable aléatoire et la lettre minuscule désigne une valeur possible. Remarque : dans le cas d'une fréquence

par la relation bien-connue suivante :

$$p(\mathbf{x}) = p(x_1, \dots, x_n)p(x_{n+r}). \quad (4.11)$$

Introduisons la *fonction caractéristique* de la distribution de probabilité associée à  $p(\mathbf{x})$ , au moyen de la notation  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n, \alpha_{n+r})$  :

$$\Phi(\boldsymbol{\alpha}) = \langle \exp(i\boldsymbol{\alpha} \cdot \mathbf{X}) \rangle \quad (4.12)$$

$$= \int p(\mathbf{x}) \prod_{q=1}^{n, n+r} \exp(i\alpha_q x_q) dx_q, \quad (4.13)$$

où  $\langle \rangle$  désigne une moyenne et  $\prod_{q=1}^{n, n+r}$  signifie un produit qui couvre  $q = 1, \dots, n, n+r$ . Les distributions de probabilité marginales associées à  $p(x_1, \dots, x_n)$  et  $p(x_{n+r})$  possèdent également des fonctions caractéristiques :<sup>6</sup>

$$\Phi_P(\alpha_1, \dots, \alpha_n) = \left\langle \exp \left( i \sum_{q=1}^n \alpha_q X_q \right) \right\rangle \quad (\text{passé}) \quad (4.16)$$

$$\Phi_F(\alpha_{n+r}) = \langle \exp(i\alpha_{n+r} X_{n+r}) \rangle \quad (\text{futur}). \quad (4.17)$$

On peut se convaincre aisément que la relation d'indépendance (4.11) se traduit de la façon suivante en termes des fonctions caractéristiques, i.e. dans l'espace de Fourier :

$$\Phi = \Phi_P \Phi_F. \quad (4.18)$$

Nous allons maintenant développer les fonctions caractéristiques en terme des cumulants, et étudier l'incidence de la relation (4.18) sur ces derniers. Ceci nous permettra de construire la mesure de corrélation statistique cherchée. Nous développerons  $\Phi_P$ .  
d'échantillonnage inutilement élevée, il pourrait être plus efficace d'introduire un délai  $\tau$  et ainsi de considérer le vecteur  $(X_1, \dots, X_{(n-1)\tau+1}, X_{(n+r-1)\tau+1})$ .

<sup>6</sup>Fonctions de densité marginales :

$$p(x_1, \dots, x_n) = \int p(\mathbf{x}) dx_{n+r} \quad (4.14)$$

$$p(x_{n+r}) = \int p(\mathbf{x}) dx_1 \cdots dx_n. \quad (4.15)$$

On note d'abord, trivialement :

$$\exp\left(i\sum_{q=1}^n\alpha_q x_q\right) = \prod_{q=1}^n \exp(i\alpha_q x_q) \quad (4.19)$$

$$= \sum_{q_1, \dots, q_n=0}^{\infty} \frac{i^{(q_1+\dots+q_n)}}{q_1! \dots q_n!} \alpha_1^{q_1} \dots \alpha_n^{q_n} x_1^{q_1} \dots x_n^{q_n}. \quad (4.20)$$

Les *moments* de la distribution de probabilité décrite par  $p(x_1, \dots, x_n)$  sont définis de la façon suivante :

$$M_{q_1 \dots q_n} = \int x_1^{q_1} \dots x_n^{q_n} p(x_1, \dots, x_n) dx_1 \dots dx_n. \quad (4.21)$$

La fonction caractéristique  $\Phi_P$  peut donc s'écrire

$$\Phi_P(\alpha_1, \dots, \alpha_n) = \sum_{q_1, \dots, q_n=0}^{\infty} \frac{i^{(q_1+\dots+q_n)}}{q_1! \dots q_n!} M_{q_1 \dots q_n} \alpha_1^{q_1} \dots \alpha_n^{q_n}. \quad (4.22)$$

Ainsi, à un facteur près, les coefficients d'un développement en série de puissances de la fonction caractéristique ne sont que les moments de la distribution de probabilité. Il est intéressant de remarquer que la fonction de densité, la fonction caractéristique et l'ensemble complet des moments décrivent tout aussi bien la distribution de probabilité. En effet, la connaissance de tous les moments permet de construire la fonction caractéristique (cf. éq. (4.22)), alors qu'une transformée de Fourier inverse de cette dernière permet d'obtenir la fonction de densité :

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^n} \int \exp\left(-i\sum_{q=1}^n\alpha_q x_q\right) \Phi_P(\alpha_1, \dots, \alpha_n) d\alpha_1 \dots d\alpha_n. \quad (4.23)$$

Poursuivons : si l'on pose  $Q = q_1 + \dots + q_n$ , l'équation (4.22) peut être présentée comme suit :

$$\Phi_P = \sum_{q_1, \dots, q_n=0}^{\infty} \frac{i^Q}{Q!} M_{q_1 \dots q_n} \alpha_1^{q_1} \dots \alpha_n^{q_n} \underbrace{\frac{Q!}{q_1! \dots q_n!}}. \quad (4.24)$$

Nombre de façons de placer  $Q$  objets discernables dans  $n$  boîtes (pas d'ordre dans les boîtes,  $q_j$  objets dans boîte  $j$ ).

On écrit ensuite :

$$\Phi_P = \sum_{Q=0}^{\infty} \frac{i^Q}{Q!} \left( \sum'_{q_1, \dots, q_n} \frac{Q!}{q_1! \cdots q_n!} M_{q_1 \dots q_n} \alpha_1^{q_1} \cdots \alpha_n^{q_n} \right), \quad (4.25)$$

où le prime affecté à la somme signifie respect de la contrainte

$$q_1 + \cdots + q_n = Q. \quad (4.26)$$

Si l'on introduit une nouvelle notation pour les moments, en l'occurrence

$$\mathcal{M}_{l_1 \dots l_Q} = \langle X_{l_1} \cdots X_{l_Q} \rangle \quad \text{où } l_j \in [1, n] \forall j, \quad (4.27)$$

on peut établir l'équivalence suivante :

$$\sum'_{q_1, \dots, q_n} \frac{Q!}{q_1! \cdots q_n!} M_{q_1 \dots q_n} \alpha_1^{q_1} \cdots \alpha_n^{q_n} = \delta_{Q0} + \sum_{l_1, \dots, l_Q=1}^n \mathcal{M}_{l_1 \dots l_Q} \alpha_{l_1} \cdots \alpha_{l_Q}. \quad (4.28)$$

Le delta de Kronecker  $\delta_{Q0}$  couvre le cas  $Q = 0$  pour lequel la somme du côté droit est inexistante.<sup>7</sup> Pour se convaincre de cette équivalence, on observe :

$$M_{q_1 \dots q_n} = \mathcal{M}_{\underbrace{1 \dots 1}_{q_1} \dots \underbrace{n \dots n}_{q_n}}. \quad (4.29)$$

En fait, la dernière égalité s'étend à tout ensemble de  $Q$  indices de  $\mathcal{M}$  comprenant  $q_1$  indice(s) 1,  $q_2$  indice(s) 2, ...,  $q_n$  indice(s)  $n$ . (cf. éq. (4.21) et (4.27)). Le nombre d'ensembles *distincts*<sup>8</sup> de  $Q$  indices qui ont cette propriété est de  $Q!/(q_1! \cdots q_n!)$ . Pour un ensemble d'indices  $\{q_1, \dots, q_n\}$  de  $M$  conforme à (4.26), la somme du côté droit de l'équation (4.28) couvre tous les ensembles distincts d'indices  $\{l_1, \dots, l_Q\}$  de  $\mathcal{M}$  correspondants. De plus, cette même somme garantit le traitement de tous les ensembles  $\{q_1, \dots, q_n\}$  satisfaisant la contrainte (4.26). La fonction caractéristique peut donc prendre la forme suivante :

$$\Phi_P = 1 + \sum_{Q=1}^{\infty} \frac{i^Q}{Q!} \sum_{l_1, \dots, l_Q=1}^n \mathcal{M}_{l_1 \dots l_Q} \alpha_{l_1} \cdots \alpha_{l_Q}. \quad (4.30)$$

<sup>7</sup>Le cas  $Q = 0$  correspond à  $q_1 = \cdots = q_n = 0$  ( $M_{0 \dots 0} = 1$ , cf. éq. (4.21)).

<sup>8</sup>La permutation de deux indices identiques ne crée pas un nouvel ensemble.

De façon alternative, on peut considérer la *fonction caractéristique des cumulants*, définie simplement

$$\Psi_P = \ln \Phi_P. \quad (4.31)$$

Le développement de  $\Psi_P$  en série de puissances fait apparaître non plus les moments, mais bien plutôt les *cumulants*<sup>9</sup>  $\mathcal{K}_{l_1 \dots l_Q}$  de la distribution de probabilité :

$$\Psi_P(\alpha_1, \dots, \alpha_n) = \sum_{Q=1}^{\infty} \frac{i^Q}{Q!} \sum_{l_1, \dots, l_Q=1}^n \mathcal{K}_{l_1 \dots l_Q} \alpha_{l_1} \cdots \alpha_{l_Q}. \quad (4.32)$$

Le lecteur prêtera attention à l'absence ici d'un terme unitaire associé à  $Q = 0$ , puisque  $\Phi_P(0, \dots, 0) = 1 \Rightarrow \Psi_P(0, \dots, 0) = 0$ . La connaissance de l'ensemble des cumulants détermine la distribution de probabilité. En regard de (4.32), les cumulants s'obtiennent au moyen de la relation suivante :

$$\mathcal{K}_{l_1 \dots l_Q} = (-i)^Q \left. \frac{\partial^Q \Psi_P}{\partial \alpha_{l_1} \cdots \partial \alpha_{l_Q}} \right|_{\alpha_{l_j}=0 \ \forall j}. \quad (4.33)$$

Si chaque variable statistique  $X_1, \dots, X_n$  est de moyenne nulle, c.-à-d.

$$\langle X_1 \rangle = \dots = \langle X_n \rangle = 0, \quad (4.34)$$

on trouve, après une brève séance d'algèbre, la relation entre les cumulants et les moments :

$$\mathcal{K}_{l_1} = \mathcal{M}_{l_1} = 0 \quad (4.35)$$

$$\mathcal{K}_{l_1 l_2} = \mathcal{M}_{l_1 l_2} \quad (4.36)$$

$$\mathcal{K}_{l_1 l_2 l_3} = \mathcal{M}_{l_1 l_2 l_3} \quad (4.37)$$

$$\mathcal{K}_{l_1 l_2 l_3 l_4} = \mathcal{M}_{l_1 l_2 l_3 l_4} - \mathcal{M}_{l_1 l_2} \mathcal{M}_{l_3 l_4} - \mathcal{M}_{l_1 l_3} \mathcal{M}_{l_2 l_4} - \mathcal{M}_{l_1 l_4} \mathcal{M}_{l_2 l_3}. \quad (4.38)$$

Maintenant, en terme des fonctions caractéristiques des cumulants, la relation d'indépendance (4.18) devient

$$\Psi(\alpha_1, \dots, \alpha_n, \alpha_{n+r}) = \Psi_P(\alpha_1, \dots, \alpha_n) + \Psi_F(\alpha_{n+r}), \quad (4.39)$$

---

<sup>9</sup>Lors de la convolution de deux distributions de probabilité, les cumulants ont la propriété fort intéressante de s'additionner, d'où leur nom. Les moments ne présentent pas cette propriété.

où

$$\Psi(\boldsymbol{\alpha}) = \sum_{Q=1}^{\infty} \frac{i^Q}{Q!} \sum_{l_1, \dots, l_Q=1}^{n, n+r} \mathcal{K}_{l_1 \dots l_Q} \alpha_{l_1} \cdots \alpha_{l_Q} \quad (4.40)$$

$$\Psi_F(\alpha_{n+r}) = \sum_{Q=1}^{\infty} \frac{i^Q}{Q!} \sum_{l_1, \dots, l_Q=n+r}^{n+r} \mathcal{K}_{l_1 \dots l_Q} \alpha_{l_1} \cdots \alpha_{l_Q}. \quad (4.41)$$

La vérification de la relation (4.39) pour toute valeur de  $\alpha_1, \dots, \alpha_n, \alpha_{n+r}$  (variables indépendantes) force l'égalité des coefficients apparaissant devant un même groupe de puissances  $\alpha_1^{q_1} \cdots \alpha_n^{q_n}$ , ou de façon équivalente, devant un même groupe  $\alpha_{l_1} \cdots \alpha_{l_Q}$ . En effet, à l'instar d'un moment  $\mathcal{M}_{l_1 \dots l_Q}$ , un cumulants  $\mathcal{K}_{l_1 \dots l_Q}$  est invariant sous permutation de ses indices. Par conséquent, l'indépendance statistique, entre le bloc de  $n$  points consécutifs et le point situé  $r$  intervalles d'échantillonnage plus tard, se traduit par le fait qu'un grand nombre de cumulants doivent être nuls :

$$\mathcal{K}_{l_1 \dots l_Q} = 0 \quad \text{si} \quad \exists l_j = n+r \quad \text{et} \quad \exists l_k \neq n+r \quad \text{où} \quad j, k \in [1, Q]. \quad (4.42)$$

En fait, la condition de l'équation (4.42) correspond aux groupes  $\alpha_{l_1} \cdots \alpha_{l_Q}$  absents du côté droit de l'équation (4.39). Ceci suggère la construction directe d'une mesure de corrélation statistique (linéaire et non-linéaire) :

$$\sum_{Q=1}^{\infty} \sum'_{l_1, \dots, l_Q} \mathcal{K}_{l_1 \dots l_Q}^2, \quad (4.43)$$

où le prime signifie respect de la condition de l'équation (4.42). Manifestement, cette mesure est positive, et sera nulle si et seulement si l'indépendance statistique (4.11) est présente. Enfin, comme l'objectif est d'étudier le flot d'information de tout le passé vers le futur, il serait plus approprié de prendre la limite  $n \rightarrow \infty$ .

En pratique cependant, le calcul exact de cette mesure de corrélation statistique est inaccessible, pour des raisons évidentes. Si le processus est ergodique, les cumulants pourront être estimés via des moyennes temporelles. Ensuite, il sera nécessaire de tronquer les cumulants d'ordre  $Q$  supérieur à un certain  $Q_{\max}$ . Les auteurs de la référence [DSS97b] suggèrent à cet effet le compromis  $Q_{\max} = 4$ . Enfin, il faudra se contenter d'une valeur finie pour la longueur du bloc passé, e.g.  $n = 10$ .

D'autres modifications méritent d'être apportées. D'abord, afin de réduire le temps de calcul, il est préférable de ne sommer que les cumulants différents. D'autre part, afin de mieux balancer l'impact des différents ordres de cumulants, il vaut mieux pondérer adéquatement la contribution de chaque ordre, ce qui conduit finalement à la mesure suivante :

$$m(n, r) = \underbrace{\frac{1}{\mathcal{N}_2} \sum_{l_1=1}^n \mathcal{K}_{l_1 n+r}^2}_{\text{Ordre 2 (linéaire)}} + \underbrace{\frac{1}{\mathcal{N}_3} \sum_{l_1=1}^n \sum_{l_2=l_1}^{n, n+r} \mathcal{K}_{l_1 l_2 n+r}^2}_{\text{Ordre 3 (non-linéaire)}} + \underbrace{\frac{1}{\mathcal{N}_4} \sum_{l_1=1}^n \sum_{l_2=l_1}^{n, n+r} \sum_{l_3=l_2}^{n, n+r} \mathcal{K}_{l_1 l_2 l_3 n+r}^2}_{\text{Ordre 4 (non-linéaire)}}, \quad (4.44)$$

où  $\mathcal{N}_Q$  représente le nombre de cumulants différents d'ordre  $Q$  qui deviennent nuls sous indépendance statistique :

$$\mathcal{N}_2(n) = n \quad (4.45)$$

$$\mathcal{N}_3(n) = \frac{n}{2}(n+3) \quad (4.46)$$

$$\mathcal{N}_4(n) = \frac{n}{6}(n^2 + 6n + 11). \quad (4.47)$$

Il est à noter qu'au moyen de la transformation simple

$$X' = X - \langle X \rangle, \quad (4.48)$$

on peut obtenir des variables statistiques de moyenne nulle, et donc utiliser les équations (4.35) à (4.38) pour calculer facilement les cumulants d'intérêt. Toutefois, on prendra note que le fait de tronquer les cumulants d'ordre supérieur ne permet pas de conclure, sur la seule base de  $m(n, r) = 0$ , à l'indépendance statistique (4.11), même si cette dernière demeure l'explication la plus plausible.

Pour un signal périodique de période  $T$ , il n'est pas difficile de se convaincre que  $m(n, r + T) = m(n, r)$ , ou en d'autres termes, que la périodicité est transposée au profil de corrélation statistique. Pour un signal chaotique, la mesure  $m(n, r)$  devrait décroître très rapidement avec  $r$ , en principe d'autant plus rapidement que le temps

de Lyapunov est court.<sup>10</sup> Il faut réaliser dans ce cas que la dimension de recouvrement minimale et la présence de bruit affectent également le profil de corrélation.

Le profil  $m(n, r)$  vs  $r$  permet donc, dans une certaine mesure, de caractériser le flot d'information du processus stationnaire à l'origine d'une série temporelle. Il fournit donc une statistique utile qui pourra servir à discriminer plusieurs dynamiques différentes.<sup>11</sup> À l'annexe A, la mesure  $m(n, r)$  est employée dans une nouvelle méthode de réduction du bruit par réseau de neurones, dont il a déjà été fait mention. Dans la prochaine section, on s'en sert plutôt pour détecter la non-stationnarité.

## 4.2 Mesure de non-stationnarité

Comme le profil  $m(n, r)$  vs  $r$  permet de classifier différents processus stationnaires, son évaluation dans une fenêtre glissante peut permettre de suivre les changements dynamiques qui s'opèrent à l'intérieur d'une série temporelle non stationnaire. C'est l'idée essentielle de notre nouvelle méthode de détection de non-stationnarité [DBDb].

Il est plus commode, quoiqu'un peu moins discriminant, d'extraire une seule quantité représentative du flot d'information que de considérer le profil entier. À cet effet, on peut fixer de façon convenable un horizon de prédiction  $r_{\max}$ , et ne retenir qu'une corrélation statistique moyenne :

$$m_{\text{avg}}(n, r_{\max}) = \frac{1}{r_{\max}} \sum_{r=1}^{r_{\max}} m(n, r). \quad (4.49)$$

Cette moyenne constitue en quelque sorte une mesure de non-stationnarité, puisque des variations *significatives*<sup>12</sup> de  $m_{\text{avg}}(n, r_{\max})$  lors du déplacement de la fenêtre révèlent la présence de non-stationnarité.

<sup>10</sup>Le temps de Lyapunov n'est que l'inverse de l'exposant de Lyapunov maximal.

<sup>11</sup>Dans la référence [DSS97b], la mesure  $m(n, r)$  permet d'identifier facilement l'ordre du processus de Markov à l'origine d'un signal.

<sup>12</sup>Significatif au sens de non attribuable simplement aux fluctuations statistiques inhérentes à l'emploi d'une fenêtre de longueur finie.

De façon à ce que les variations de  $m_{\text{avg}}(n, r_{\text{max}})$  ne soient tributaires que d'un changement du flot d'information et non d'un simple changement d'échelle, il convient de standardiser les données à l'intérieur de la fenêtre avant d'y évaluer les différents cumulants. Si l'on désigne par  $\bar{x}$  la moyenne et par  $s_x$  l'écart-type des données de la fenêtre, la standardisation consiste uniquement en la transformation suivante :

$$x \mapsto \frac{x - \bar{x}}{s_x}. \quad (4.50)$$

Ce faisant, la mesure de non-stationnarité  $m_{\text{avg}}$  est sans dimension.

La longueur de la fenêtre glissante sera notée  $l_w$  (nombre de données) et le pas de translation  $t_w$ . Règle générale, pour un suivi acceptable, on aura  $t_w \ll l_w$ , i.e. chevauchement. Autre note : la valeur calculée pour  $m_{\text{avg}}(n, r_{\text{max}})$  est naturellement inscrite à la fin de la fenêtre dans le temps.

D'emblée on réalise le compromis inévitable entre la résolution temporelle de la non-stationnarité et la précision statistique. Il est vrai que diminuer le pas de translation  $t_w$  peut permettre d'augmenter la résolution temporelle sans pour autant amplifier les fluctuations statistiques. Mais ce n'est pas là le paramètre qui a le plus d'importance. La longueur de la fenêtre glissante prime. Ainsi, diminuer  $l_w$  correspond à exiger un suivi plus précis dans le temps. Or le prix à payer est une qualité statistique amoindrie en ce qui concerne l'évaluation des cumulants. La conséquence directe est une courbe davantage brisée pour la mesure de non-stationnarité. À l'opposé, l'augmentation de  $l_w$  (suivi plus grossier) se traduit par une courbe plus lisse de résolution temporelle réduite.

### 4.3 Exemples artificiels

Dans cette section, le comportement de notre mesure de non-stationnarité sera étudié au moyen de cinq groupes de séries temporelles artificielles non stationnaires. Les quatre premiers sont issus directement de la littérature sur la non-stationnarité.

Ceci permettra d'illustrer, dans un environnement bien contrôlé, le potentiel de notre approche, la robustesse de cette nouvelle mesure de non-stationnarité par rapport aux différents paramètres ( $n$ ,  $r_{\max}$ ,  $l_w$  et  $t_w$ ), de même que les limites de la méthode.

### 4.3.1 Dynamique segmentée

La première série temporelle non stationnaire provient d'une dynamique segmentée, exemple utilisé par Manuca et Savit («alternating dynamics») [MS96] :

1. Nombres aléatoires de distribution uniforme sur l'intervalle  $[0, 1]$  (Fig. 4.1a).
2. Application de la tente (Fig. 4.1b) :

$$x_{t+1} = \begin{cases} 2x_t & \text{si } x_t \leq 1/2 \\ 2(1 - x_t) & \text{si } x_t > 1/2. \end{cases} \quad (4.51)$$

3. Déplacement de Bernoulli («binary shift map») (Fig. 4.1c) :

$$x_{t+1} = 2x_t \bmod 1. \quad (4.52)$$

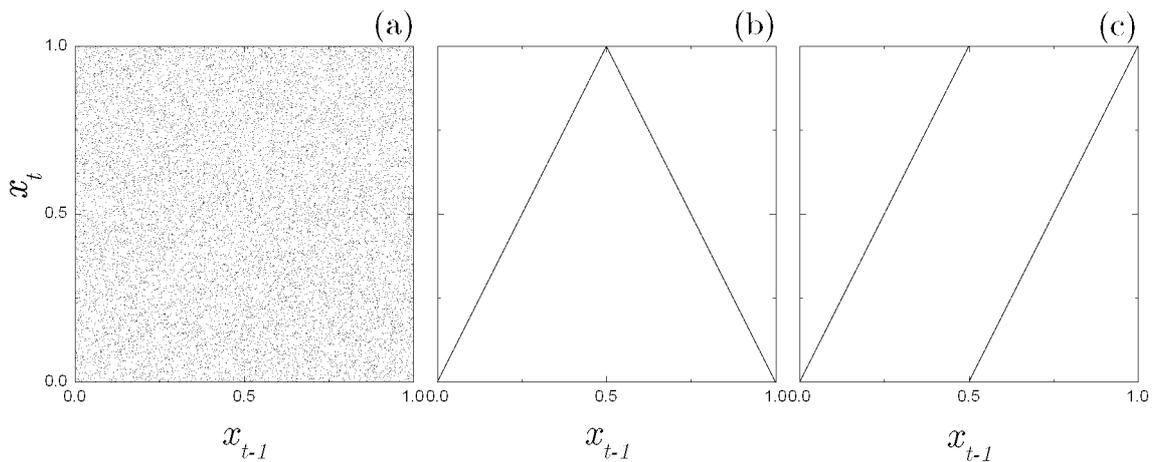


FIG. 4.1 – Graphes à délai (un pour chaque segment de la dynamique segmentée). (a) Nombres aléatoires de distribution uniforme. (b) Application de la tente. (c) Déplacement de Bernoulli.  $N=16\ 384$  pour chaque segment.

Ainsi, le premier tiers de la série temporelle est aléatoire, le second régi par la dynamique de l'application de la tente, et enfin le dernier sous la férule du déplacement de Bernoulli (consulter l'annexe C pour l'implémentation du déplacement de Bernoulli). Chaque segment comprend 16 384 points. La particularité intéressante est le fait que les trois segments présentent la même distribution de probabilité 1D, uniforme sur l'intervalle  $[0, 1]$  (consulter l'annexe D). Dans l'article de Manuca et Savit auparavant cité, on soutient de plus que la fonction d'autocorrélation est commune aux trois segments, et donc qu'une analyse purement linéaire ne peut discerner les trois dynamiques différentes. Si l'on ne néglige pas le caractère fondamentalement discret de la présente dynamique, cette dernière affirmation est fautive. La fonction d'autocorrélation du déplacement de Bernoulli se démarque des deux autres, ce qui est démontré à l'annexe D. Par conséquent, les cumulants d'ordre deux, qui ne sont que des fonctions de corrélation, suffisent à différencier le segment du déplacement de Bernoulli des deux précédents (analyse linéaire).

L'exemple de la dynamique segmentée en est un très simple, dont la non-stationnarité n'est pas très difficile à déceler. Toutefois, cet exemple artificiel s'avère fort commode pour mettre en évidence, de façon claire, de nombreuses caractéristiques de notre mesure de non-stationnarité.

Examinons d'abord l'évolution du flot d'information complet, i.e. le profil  $m(n, r)$  vs  $r$ , lors de la translation de la fenêtre glissante. La Fig. 4.2 fait ressortir directement les trois segments dynamiques différents. De plus, il apparaît clair qu'une prédiction à un pas est suffisante, i.e. on peut prendre simplement  $r_{\max} = 1$ .

Trois ordres ( $Q=2, 3$  et  $4$ ) de cumulants contribuent à la mesure  $m_{\text{avg}}$  (les trois termes dans l'équation (4.44)), et il peut être intéressant d'observer chaque ordre séparément, ce que permet la Fig. 4.3. On note que dans le présent cas, chaque ordre distingue son propre segment. En effet, l'ordre 2 ne fait ressortir que le déplacement de Bernoulli, un fait déjà souligné, l'ordre 3 ne distingue que l'application de la tente, et enfin l'ordre 4 ne semble sensible qu'au déterminisme. La superposition des trois

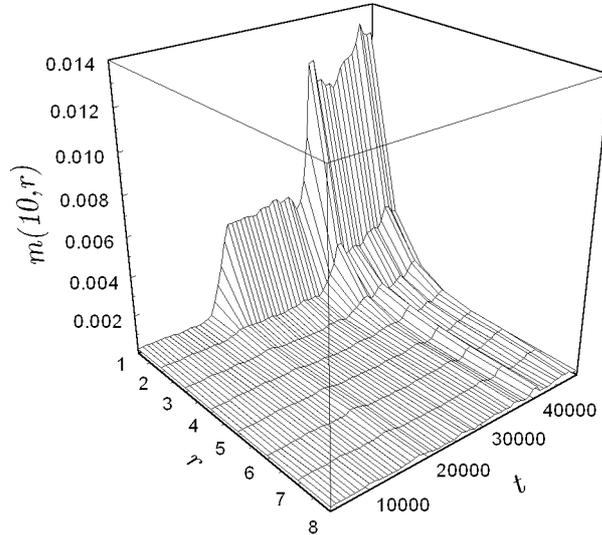


FIG. 4.2 – Flot d'information  $m(10, r)$  vs  $r$  de la dynamique segmentée.  $l_w=4000$  et  $t_w=10$ .

ordres génère une mesure efficace de non-stationnarité, et chaque ordre est important. L'ajout d'ordres supérieurs se traduirait certainement par une efficacité accrue de détection, mais d'une part, le temps de calcul augmenterait de façon vertigineuse ( $\mathcal{N}_Q(n) \sim n^{Q-1}$ ), et d'autre part, l'importance d'un ordre décroît lorsque  $Q$  augmente, et ce autant du point de vue numérique (cf. Fig. 4.3) que du point de vue de la signification statistique [Gar85]. Le compromis  $Q_{\max} = 4$  semble donc adéquat.

Examinons ensuite la dépendance de notre mesure de non-stationnarité par rapport aux trois paramètres ( $n$ ,  $l_w$  et  $t_w$ ) qui affectent ses performances de détection ( $r_{\max}$  a été convenablement fixé à l'unité). Un seul paramètre sera varié à la fois. Les Figs. 4.4 à 4.6 illustrent, dans le présent cas, la grande robustesse de notre méthode. En effet, sur un vaste domaine des paramètres, la mesure de non-stationnarité réalise efficacement son objectif. De plus, ces figures illustrent concrètement les considérations de la section 4.2 sur le compromis entre résolution temporelle et précision statistique. Comme la valeur de  $m_{\text{avg}}$  est placée à la fin de la fenêtre dans le temps, il en résulte un certain délai de réponse par rapport aux changements dynamiques qui s'opèrent.

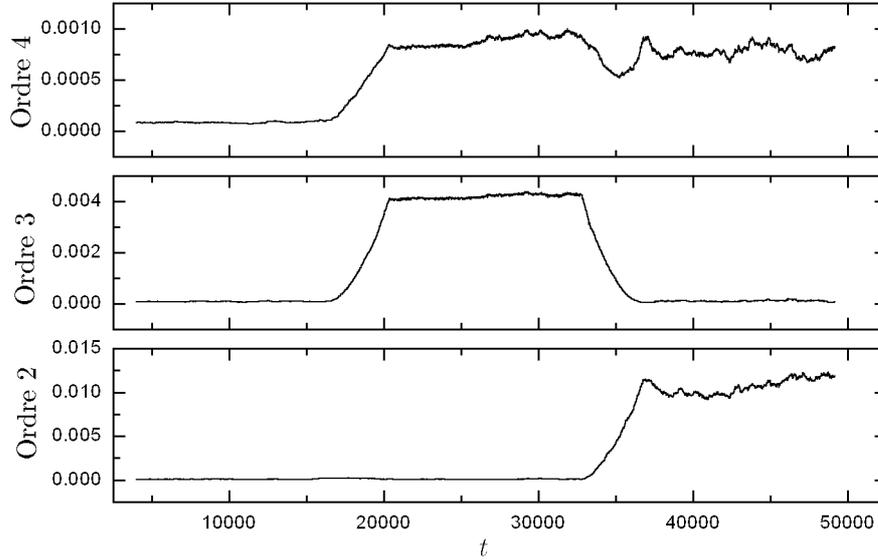


FIG. 4.3 – Pour la dynamique segmentée, décomposition de la mesure  $m(10, 1)$  en ses trois ordres.  $l_w=4000$  et  $t_w=10$ .

Dans le cas de la dynamique segmentée, cela se traduit par un décalage entre le début d'un nouveau segment dynamique et l'atteinte du plateau correspondant pour  $m_{\text{avg}}$ . Évidemment, diminuer à la fois  $l_w$  et  $t_w$  aident à réduire ce délai.

### 4.3.2 Application triangulaire asymétrique

Le second groupe est issu de l'application triangulaire asymétrique :

$$x_{t+1} = \begin{cases} x_t/w & \text{si } x_t \leq w \\ (1 - x_t)/(1 - w) & \text{si } x_t > w. \end{cases} \quad (4.53)$$

Le cas  $w = 1/2$  correspond à l'application de la tente. En variant le paramètre d'asymétrie  $w$ , on peut générer une série temporelle non stationnaire. Exemple utilisé par Witt et coauteurs [WKP98]. La Fig. 4.7, où l'on retrouve un cas stationnaire et un autre non-stationnaire (dérive adiabatique du paramètre), est convaincante. On remarque toutefois que l'écart entre les deux courbes s'estompe, i.e. devient non

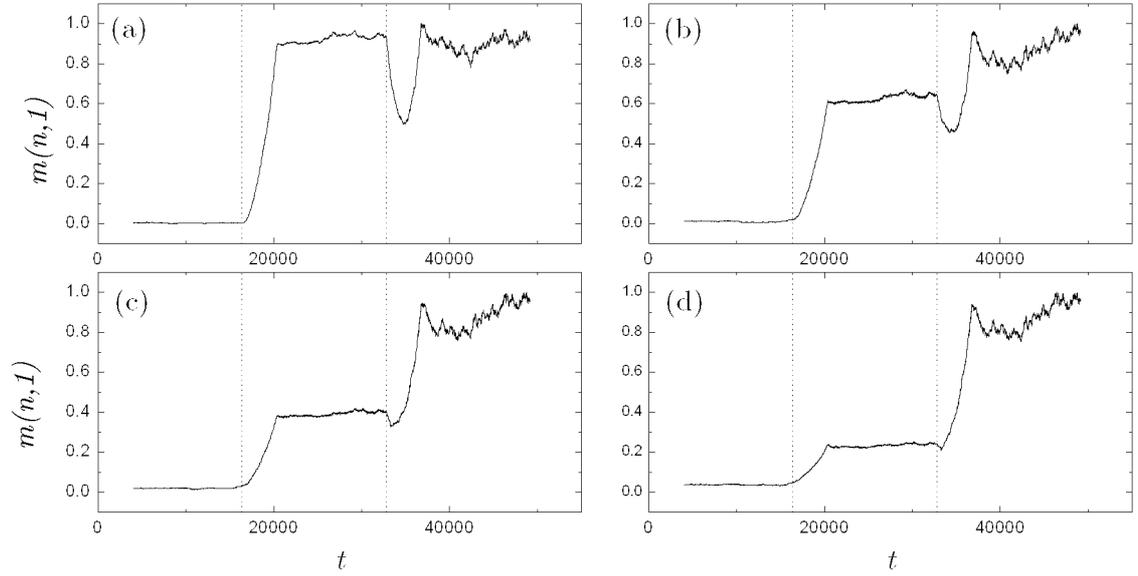


FIG. 4.4 – Mesure de non-stationnarité  $m(n, 1)$ , normalisée par rapport à sa valeur maximale, pour la dynamique segmentée.  $l_w=4000$  et  $t_w=10$ . (a)  $n=1$ . (b)  $n=5$ . (c)  $n=10$ . (d)  $n=20$ .

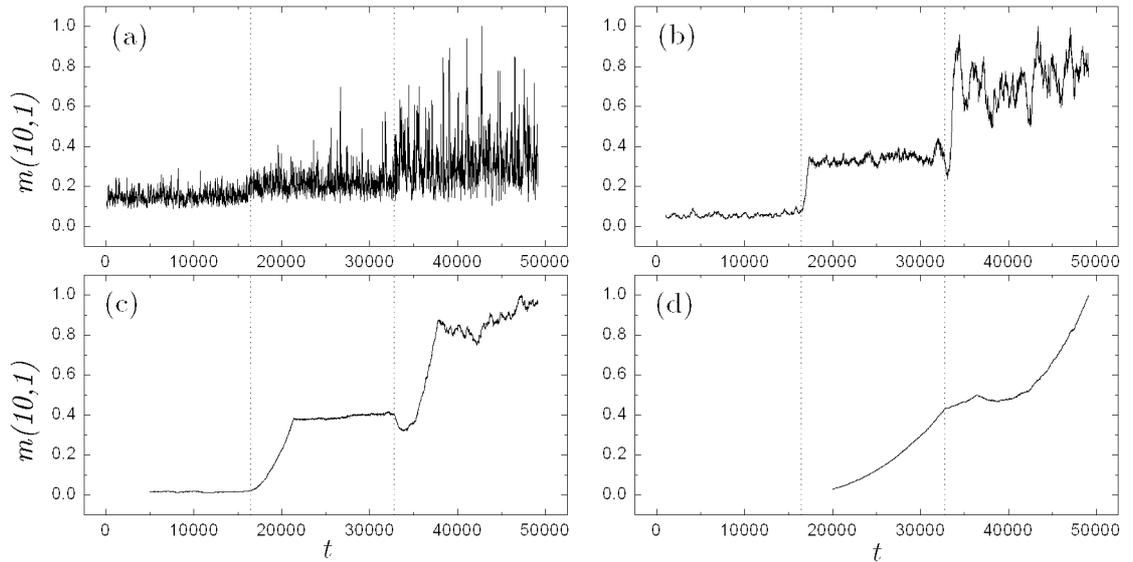


FIG. 4.5 – Mesure de non-stationnarité  $m(10, 1)$ , normalisée par rapport à sa valeur maximale, pour la dynamique segmentée.  $t_w=10$ . (a)  $l_w=100$ . (b)  $l_w=1000$ . (c)  $l_w=5000$ . (d)  $l_w=20\ 000$ .

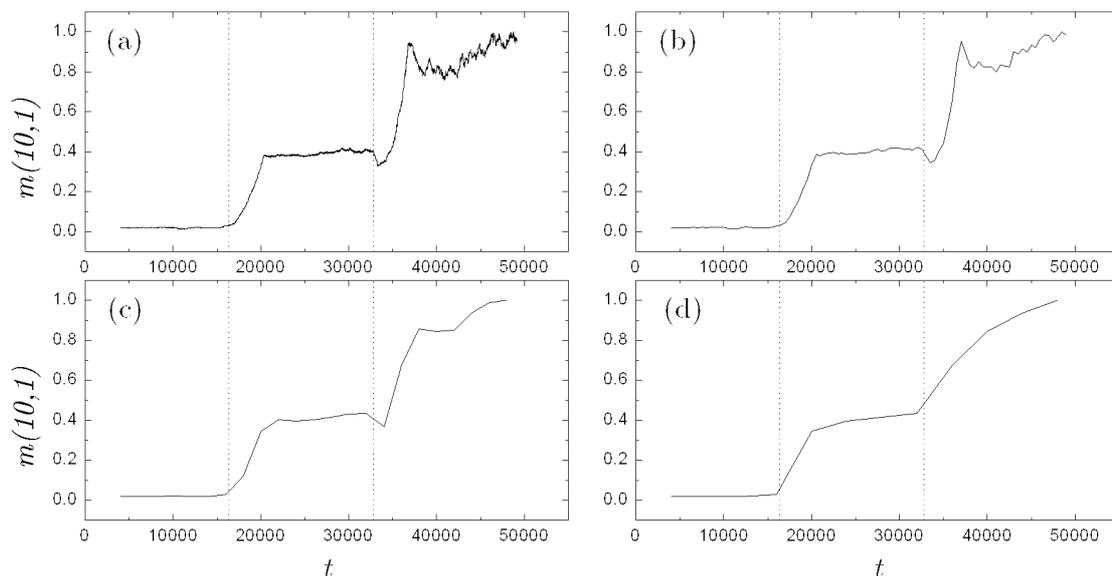


FIG. 4.6 – Mesure de non-stationnarité  $m(10,1)$ , normalisée par rapport à sa valeur maximale, pour la dynamique segmentée.  $l_w=4000$ . (a)  $t_w=10$ . (b)  $t_w=500$ . (c)  $t_w=2000$ . (d)  $t_w=4000$ .

significatif, bien avant le temps où le paramètre  $w$  est le même pour les deux séries ( $t = 32\,000$ ).

### 4.3.3 Application du boulanger généralisée

Le troisième groupe se rapporte à l'application du boulanger généralisée :

$$u_{t+1} = \begin{cases} \beta u_t & \text{si } v_t \leq \alpha \\ 1/2 + \beta u_t & \text{si } v_t > \alpha \end{cases} \quad (4.54)$$

$$v_{t+1} = \begin{cases} v_t/\alpha & \text{si } v_t \leq \alpha \\ (v_t - \alpha)/(1 - \alpha) & \text{si } v_t > \alpha, \end{cases} \quad (4.55)$$

où  $v_t \in [0, 1]$  et  $\alpha, \beta \in ]0, 1[$ . Les deux exposants de Lyapunov de cette application peuvent être calculés analytiquement (voir e.g. chapitre 4 de la référence [OSY94]) :

$$\lambda_1 = \alpha \ln\left(\frac{1}{\alpha}\right) + (1 - \alpha) \ln\left(\frac{1}{1 - \alpha}\right) > 0 \quad (4.56)$$

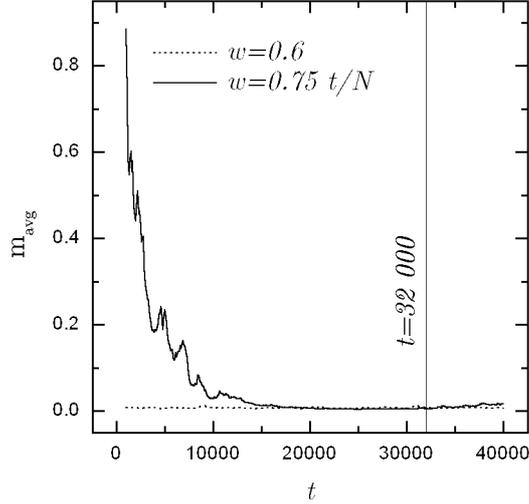


FIG. 4.7 – Mesure de non-stationnarité  $m_{\text{avg}}$  pour deux séries temporelles issues de l'application triangulaire asymétrique ( $x_1 = 0.1$ ), l'une stationnaire, l'autre non stationnaire.  $n = 5$ ,  $r_{\text{max}} = 5$ ,  $l_w = 1000$  et  $t_w = 100$ .

$$\lambda_2 = \ln \beta < 0. \quad (4.57)$$

On observe que si l'on varie le paramètre  $\beta$ , la valeur de l'exposant de Lyapunov maximal demeure inchangée. En faisant dériver ce paramètre de façon adiabatique, on peut donc construire une série temporelle dont la non-stationnarité sera subtile. L'idée provient de Schreiber [Sch97], qui suggère également de « mesurer » l'observable  $w_t = u_t + v_t$ , en prenant soin de retrancher la moyenne mobile et de normaliser par rapport à la variance mobile.<sup>13</sup> La série temporelle finale est donc  $\{s_t\}_{t=1}^N$  :

$$s_t = \frac{w_t - \langle w \rangle_k}{\sqrt{\langle (w - \langle w \rangle_k)^2 \rangle_k}}, \quad (4.58)$$

où  $\langle \bullet \rangle_k$  dénote ici une moyenne calculée sur les indices  $t' = t - k, \dots, t + k$ . Yu et coauteurs ont repris cet exemple dans leur travail [YLH98].

<sup>13</sup>Ceci ajoute à la difficulté de détection de la non-stationnarité. Ainsi, faisant abstraction des fluctuations dues à l'emploi d'une fenêtre glissante de longueur finie, la moyenne, la variance et l'exposant de Lyapunov maximal demeurent essentiellement constants.

Nous avons généré deux séries temporelles de  $N = 40\,000$  points, avec  $\alpha = 0.4$ ,  $u_1 = v_1 = 0.1$  et  $k = 50$ . Pour la première, stationnaire,  $\beta = 1/2$ . Pour la seconde, non stationnaire,  $\beta = t/N$ .<sup>14</sup> La Fig. 4.8 est sans équivoque : notre mesure détecte la non-stationnarité présente dans la seconde série temporelle. L'allure de la courbe suggère également la dérive d'un paramètre, et ce message est robuste par rapport aux différents paramètres de  $m_{\text{avg}}$ .

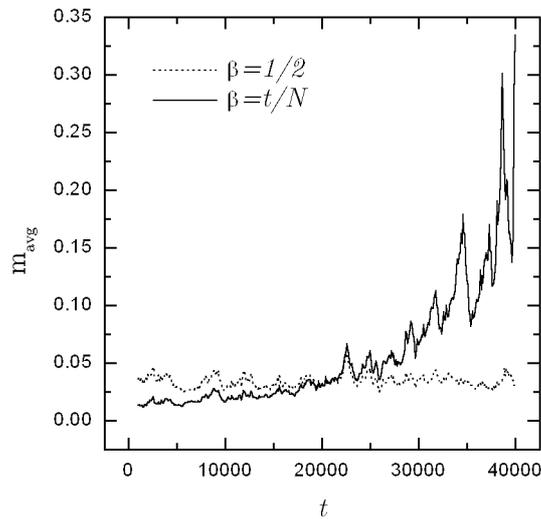


FIG. 4.8 – Mesure de non-stationnarité  $m_{\text{avg}}$  pour deux séries temporelles issues de l'application du boulanger généralisée (voir texte pour détails), l'une stationnaire, l'autre non stationnaire.  $n = 5$ ,  $r_{\text{max}} = 5$ ,  $l_w = 1000$  et  $t_w = 100$ .

#### 4.3.4 Processus auto-régressif du premier ordre AR(1)

Le quatrième groupe s'insère dans la famille des processus stochastiques auto-régressifs. Nous nous limiterons au premier ordre :

$$x_t = a(t-1)x_{t-1} + \xi_t, \quad (4.59)$$

<sup>14</sup>La valeur de  $\beta$  au temps  $t$  est celle utilisée pour calculer  $u_{t+1}$ .

où  $\xi_t$  est la valeur prise à l'itération  $t$  par une variable aléatoire de distribution gaussienne. Si l'on permet au coefficient de varier dans le temps, i.e.  $a(t) \neq a_0$ , le processus stochastique (4.59) génère alors une série temporelle non stationnaire. Exemple utilisé par Witt et coauteurs [WKP98]. La Fig. 4.9 est éloquent : la dérive linéaire du coefficient est palpable.

Witt et coauteurs [WKP98] ont examiné le comportement de leur test de stationnarité sous des conditions beaucoup plus extrêmes que celles de la Fig. 4.9. En effet, ils ont considéré les longueurs très courtes de  $N = 2000$  et  $4000$ , avec une dérive linéaire de  $a$  soit de  $0.5$  à  $0.6$ , soit de  $0.5$  à  $0.7$ . Ils ont obtenu des résultats variables, et leur conclusion, évidente, est que la détection de non-stationnarité n'est réalisée que pour des séries temporelles suffisamment longues dans le cas d'une stationnarité très faible, ou suffisamment non stationnaires dans le cas de séries très courtes.

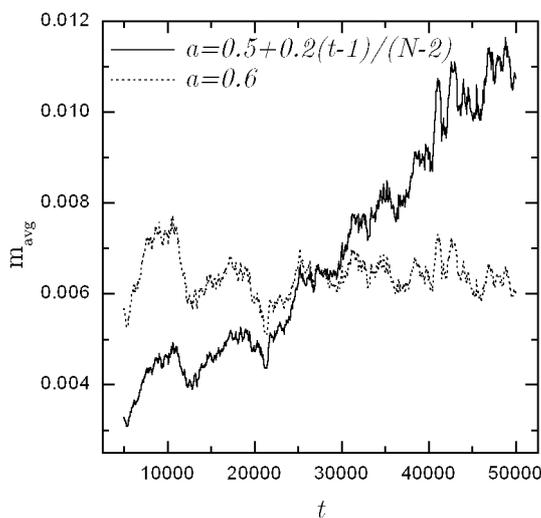


FIG. 4.9 – Mesure de non-stationnarité  $m_{\text{avg}}$  pour deux séries temporelles issues d'un processus AR(1), l'une stationnaire, l'autre non stationnaire.  $n = 10$ ,  $r_{\text{max}} = 5$ ,  $l_w = 5000$  et  $t_w = 50$ .

Quel est le comportement de notre mesure de non-stationnarité dans de telles conditions ? Et bien nous obtenons également des résultats variables, selon le point

de vue adopté. Si la courbe de  $m_{\text{avg}}$  à elle seule doit révéler la non-stationnarité, ce que l'on exige en fait en pratique, alors notre mesure s'avère plutôt inefficace. La Fig. 4.10a illustre ce point. À l'opposé des exemples présentés jusqu'à maintenant, la courbe ne suggère pas immédiatement la présence de non-stationnarité. Les fluctuations statistiques prédominantes masquent tout message en ce sens. Néanmoins, si l'on utilise la série temporelle stationnaire correspondante, on arrive à faire ressortir la dérive linéaire du coefficient. Soit  $m_{\text{avg}}(\text{NS})$  la mesure associée à la série non stationnaire et  $m_{\text{avg}}(\text{S})$  l'équivalent pour la série stationnaire. Le graphe de  $m_{\text{avg}}(\text{NS}) - m_{\text{avg}}(\text{S})$  vs  $t$  est très prêt d'une relation linéaire (cf. Fig. 4.10b).

### 4.3.5 Intermittence

Le dernier exemple artificiel de cette section s'avère un peu différent de ses congénères. La non-stationnarité n'est pas induite par concaténation de dynamiques différentes ou encore par la dérive lente d'un paramètre. Le système dynamique ne change point au cours du temps. Pourtant, a posteriori, la série temporelle déploie une non-stationnarité manifeste.

Cet exemple a déjà été évoqué au chapitre précédent. Il s'agit du phénomène d'intermittence [Str94]. L'application logistique permet d'observer ce phénomène très aisément :

$$x_t = f(r; x_{t-1}) = rx_{t-1}(1 - x_{t-1}). \quad (4.60)$$

Lorsque le paramètre  $r$  franchit le cap  $r_0 = 1 + \sqrt{8} = 3.828427\dots$ ,  $f^3$  subit une bifurcation col-noeud d'où émergent deux orbites périodiques, l'une stable et l'autre instable. Ceci explique l'apparition soudaine de la fenêtre de période 3 sur le diagramme de bifurcation, suite au chaos. Précisément à  $r = r_0$ , le graphe de  $f(r_0, x)$  vs  $x$  est tangent à la diagonale en trois points (fusion des deux orbites périodiques), et pour  $r$  tout juste inférieur à cette valeur, les trois points de tangence disparaissent au profit de trois «canaux» très étroits. Ceux-ci retiennent littéralement l'orbite lors de

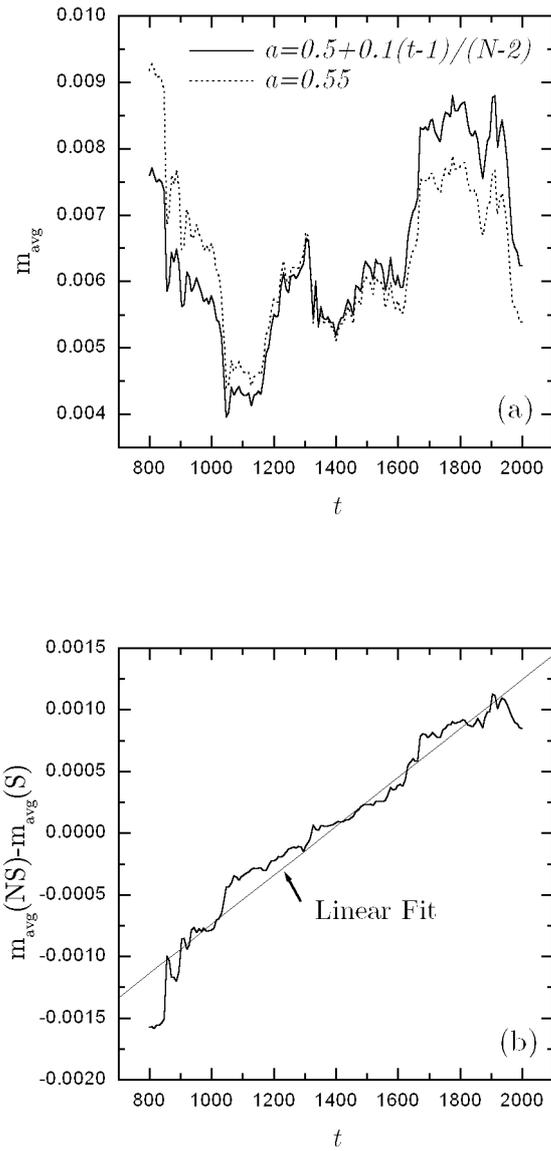


FIG. 4.10 – (a) Mesure de non-stationnarité  $m_{\text{avg}}$  pour deux séries temporelles issues d'un processus AR(1), l'une stationnaire, l'autre non stationnaire.  $n = 10$ ,  $r_{\text{max}} = 5$ ,  $l_w = 800$  et  $t_w = 8$ . (b)  $m_{\text{avg}}(\text{NS}) - m_{\text{avg}}(\text{S})$ .

son passage, bien que celle-ci finisse par s'en y échapper pour y revenir ultérieurement. L'orbite prend alors la forme d'un mouvement régulier (fantôme de l'orbite de période 3) entrecoupé de fréquentes bouffées chaotiques. C'est l'intermittence.

Lorsque  $r$  diminue et s'éloigne de  $r_0$ , les épisodes périodiques se raccourcissent jusqu'à ce que règne éventuellement un chaos total (scénario de route vers le chaos par intermittence). Afin d'équilibrer les épisodes périodiques et chaotiques, le choix  $r = 3.828285$  est adéquat. Et comme la durée de chaque épisode est relativement courte, il est préférable de restreindre la longueur de la série temporelle sous étude.

La Fig. 4.11 présente la résolution temporelle de la non-stationnarité, telle que réalisée par la mesure  $m_{\text{avg}}$ , d'une orbite de longueur  $N = 2048$  où se manifeste un chaos intermittent. Le suivi des épisodes dynamiques qui se succèdent est très satisfaisant. Il est à noter toutefois que le second ordre de cumulants suffit pour accomplir cette tâche. En fait, les trois ordres ont étonnamment des tracés très similaires.

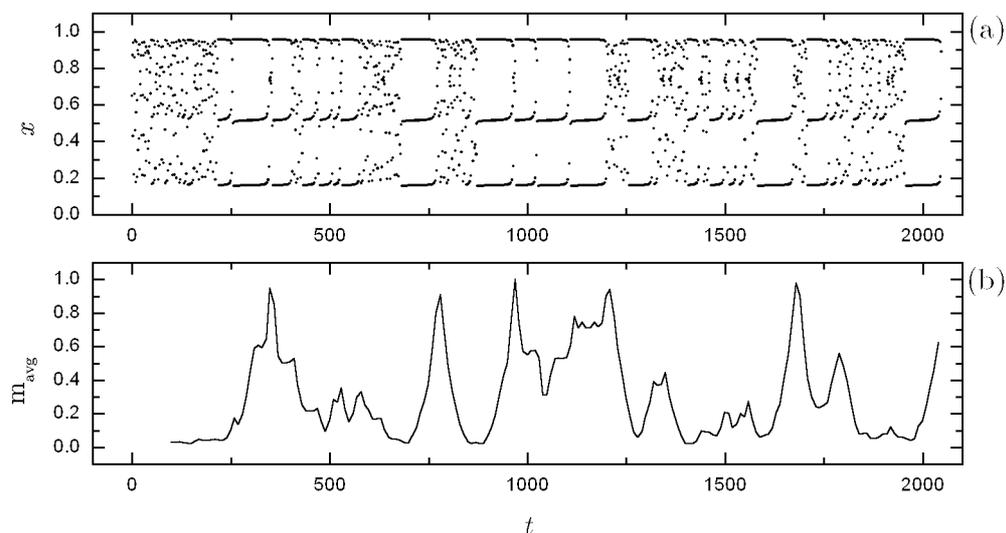


FIG. 4.11 – Intermittence. (a) Orbite de l'application logistique pour  $r = 3.828285$ . Condition initiale :  $x_1 = 0.65$ . (b) Mesure de non-stationnarité  $m_{\text{avg}}$  normalisée par rapport à sa valeur maximale.  $n = 10$ ,  $r_{\text{max}} = 10$ ,  $l_w = 100$  et  $t_w = 10$ .

## 4.4 Précurseur d'une crise d'épilepsie

S'il est un défi de détecter la non-stationnarité d'une série temporelle, ou encore de suivre simultanément les changements dynamiques, il en est un encore plus grand de prédire l'avènement de certains épisodes dynamiques. En fait, conjointement à l'objectif de valider l'hypothèse de stationnarité avant l'analyse, l'objectif d'anticiper un «bouleversement dynamique» canalise l'effort de nombreux groupes de recherche. Il faut comprendre que l'enjeu est de taille, puisque des applications importantes attendent une méthode efficace. L'une d'entre elles est au coeur de nombreuses investigations : la prédiction de l'avènement d'une crise épileptique à partir de l'analyse en temps réel d'un EEG (recherche de signes précurseurs). L'épilepsie frappe environ 1% de la population [Sch98].

La crise d'épilepsie provient d'une décharge, c.-à-d. d'un brusque et intense accès d'activité, d'un groupe plus ou moins important de neurones cérébraux qui deviennent synchrones (déclin de la complexité dynamique). Si cette activité anormale origine d'une lésion localisée du cerveau, le foyer épileptogène, l'épilepsie est dite partielle ou encore focale. Environ 20% des patients souffrant d'une telle forme d'épilepsie ne peuvent être soulagés par l'usage de médicaments anticonvulsivants [QALG]. Pour ce groupe infortuné, l'intervention chirurgicale (résection de la zone épileptogène) devient l'ultime solution. Peut-on envisager un remède moins drastique, moins risqué et moins coûteux [Gla97] ?

Pour diagnostiquer et étudier l'épilepsie, on se sert de l'électroencéphalographie, qui consiste à mesurer, au moyen d'électrodes posées sur le cuir chevelu, les variations de potentiel électrique qui se produisent au niveau du cortex. Dans certaines circonstances, e.g. délimitation précise de la zone épileptogène avant l'ablation, on a plutôt recours à des électrodes intracrâniennes (moins de bruit). Ce procédé invasif d'investigation est alors appelé électrocorticographie. Quoiqu'il en soit, l'électroencéphalogramme (EEG), ou encore l'électrocorticogramme (ECoG), donne accès à la dynamique du

cerveau, un indescriptible réseau de neurones. On peut donc projeter l'utilisation des méthodes de détection de non-stationnarité pour analyser ce signal très spécial, en vue de prédire la venue de crises épileptiques.<sup>15</sup> Se dessine alors une nouvelle modalité de traitement pour l'épilepsie focale : une électrode implantée de façon permanente au foyer épileptogène achemine son signal (activité électrique macroscopique du foyer) à un dispositif miniaturisé, qui s'affaire à surveiller le développement d'une crise et à contrecarrer celle-ci le cas échéant.

Le renouveau non-linéaire dans l'analyse des séries temporelles a insufflé de nouveaux espoirs cliniques au problème de la prédiction d'une crise épileptique à partir d'un EEG, et une véritable course anime actuellement le domaine. Si les méthodes linéaires traditionnelles permettent d'anticiper la crise au mieux quelques secondes auparavant, les méthodes non-linéaires semblent repousser ce seuil de détection à plusieurs minutes [LE98]. L'hypothèse du chaos de basse dimension à l'origine des signaux électriques cérébraux s'avère pourtant fort discutable (voir par exemple [The95]). Néanmoins, le véritable objectif ici concerne davantage la prédiction effective des crises épileptiques que l'obtention de conclusions scientifiques définitives, bien que celles-ci demeurent extrêmement intéressantes en soi.

Ainsi, deux groupes de recherche, en l'occurrence Lehnertz-Elger [LE98] et Martinerie et al. [MALVQ<sup>+</sup>98], concentrent leurs efforts sur la dimension de corrélation  $D_2$  [OSY94], qu'ils estiment quantifier adéquatement la complexité de la dynamique cérébrale. La chute de celle-ci lors de l'installation de la crise épileptique s'accompagne donc d'une diminution correspondante pour  $D_2$ . Toutefois, Lehnertz et Elger se défendent bien de calculer une dimension au sens strict du terme ; on parle plutôt avec prudence d'une définition opérationnelle. Il n'en demeure pas moins que cette

---

<sup>15</sup>Dans le cas de l'épilepsie focale, la synchronisation d'une masse critique de neurones conduisant à la crise semble se former par un recrutement progressif de proche en proche [MALVQ<sup>+</sup>98]. Ce recrutement préictal (avant la crise) devrait en principe laisser des traces décelables sur l'EEG, d'où la possibilité d'anticiper la crise épileptique.

définition opérationnelle leur permet d'étudier la dynamique spatio-temporelle de l'épilepsie focale au moyen de plusieurs électrodes intracrâniennes implantées dans la zone épileptogène [LE95], d'étudier l'influence d'un médicament anticonvulsivant (carbamazépine) sur cette dynamique [LE97], et enfin d'anticiper la crise épileptique plusieurs minutes à l'avance [LE98].

D'autres approches ont récemment été proposées. L'une d'entre elles compare le spectre en puissance (estimé à partir d'une transformée de Fourier discrète lissée) d'un segment de référence préictal à celui d'une fenêtre glissante au moyen de l'entropie renormalisée [KWT98] ou encore de l'entropie de Kullback-Leibler [QALG]. Une autre approche s'articule autour de l'entropie de Tsallis [CDL<sup>+</sup>98]. Enfin, certains groupes de recherche présentés au chapitre précédent ont jaugé la puissance de leurs tests de non-stationarité (résolution temporelle) au moyen d'EEGs (Réfs. [HGP99] et [WKP98].)

Notre mesure de non-stationarité constitue-t-elle un bon précurseur d'une crise d'épilepsie? C'est à cette question importante que nous tenterons d'apporter maintenant quelques éléments de réponse.

#### 4.4.1 Modèle du contrôleur épileptique

Comme modèle simpliste de route vers la crise épileptique, on peut considérer un scénario pour lequel il y a transition progressive du chaos (fonctionnement normal du cerveau) vers un mouvement périodique (synchronisation des neurones). L'idée ici est d'actionner un mécanisme de contrôle du chaos en présence d'un bruit dynamique. Au fur et à mesure que l'amplitude de ce bruit diminue, le contrôleur épileptique parvient de mieux en mieux à stabiliser l'orbite du système autour d'une orbite périodique instable. Bien entendu, ce scénario un peu naïf ne constitue qu'une pâle imitation

d'une dynamique considérablement plus complexe.<sup>16</sup>

Le système dynamique utilisé est l'application logistique, et on retrouve à l'annexe E la description de la technique de contrôle en 1D. On prendra note toutefois que le contrôleur épileptique tente de stabiliser l'orbite périodique instable en tout point de l'espace des phases, et n'attend pas d'être au voisinage de la cible. La grandeur de la perturbation demeure tout de même limitée à une valeur maximale peu élevée.

La Fig. 4.12 présente ce scénario d'installation graduelle d'une crise épileptique, accompagné de la mesure de non-stationnarité  $m_{\text{avg}}$ . L'amplitude du bruit dynamique est diminuée par paliers (cf. Tab. 4.1). La perte de complexité se matérialise très bien dans la croissance de  $m_{\text{avg}}$ . C'est ce type de comportement que l'on aimerait obtenir dans le cas du suivi d'un EEG d'un patient épileptique :  $m_{\text{avg}}$  grimpe avant l'avènement de la crise (précurseur).

TAB. 4.1 – Écart-type  $\sigma_\eta$  du bruit dynamique selon l'intervalle de la Fig. 4.12.

Intervalle	$\sigma_\eta$
I	$1.5 \times 10^{-2}$
II	$1.0 \times 10^{-2}$
III	$0.75 \times 10^{-2}$
IV	$0.5 \times 10^{-2}$
V	$0.3 \times 10^{-2}$

---

<sup>16</sup>Il n'en demeure pas moins que plusieurs groupes de recherche semblent mettre en évidence la présence d'orbites périodiques instables (UPOs pour «Unstable Periodic Orbits») dans la dynamique de réseaux neuronaux, et ce même à l'échelle macroscopique de l'EEG [SFN<sup>+</sup>98]. On commence même à suggérer un lien très étroit entre ces UPOs et l'épilepsie, d'où l'idée d'un anti-contrôle du chaos qui consiste à dévier la trajectoire loin de ces UPOs, au moyen de fines perturbations [LVQMAV97]. L'application d'un faible bruit électrique au foyer épileptogène, à titre d'anti-contrôleur pour traiter l'épilepsie, n'est peut-être pas une idée insensée.

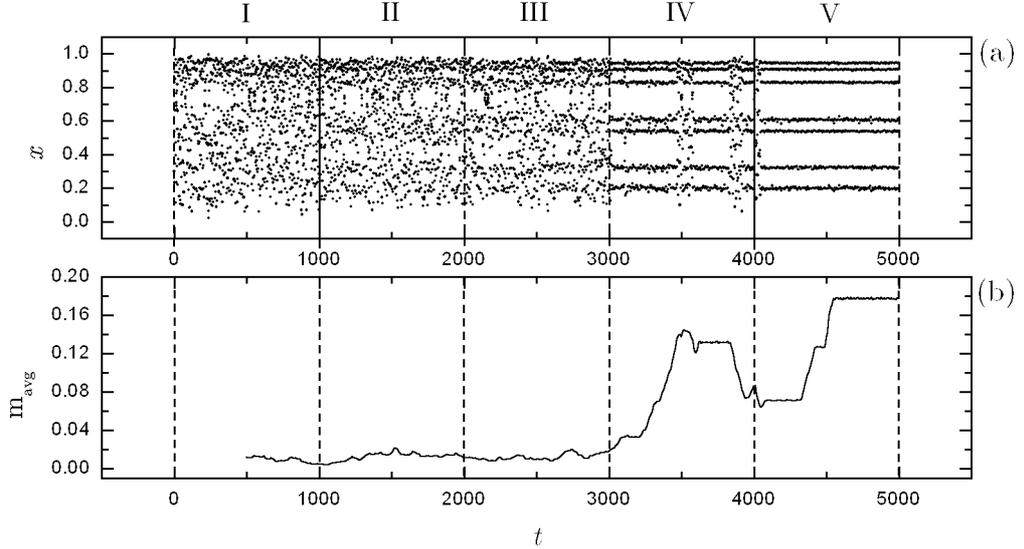


FIG. 4.12 – Contrôleur épiléptique en présence de bruit dynamique (voir Tab. 4.1).

Paramètres pour  $m_{\text{avg}}$  :  $n = 10$ ,  $r_{\text{max}} = 10$ ,  $l_w = 500$  et  $t_w = 10$ .

#### 4.4.2 Données épiléptiques de Schiff

L'électroencéphalogramme d'un patient épiléptique est règle générale caractérisé par l'émergence de pics répétés. Ainsi, dans ce cas, plutôt que d'analyser directement le signal brut de l'EEG, signal contaminé, il peut être plus efficace et plus fiable de ne retenir que la succession des pics. Pour ce faire, on traduit l'EEG en une série d'intervalles entre pics successifs («interspike interval» ou ISI), et l'attention est alors portée exclusivement sur la dynamique des ISIs. Si  $T_{i-1}$  et  $T_i$  correspondent aux temps de deux pics successifs, alors  $\text{ISI}(i) = T_i - T_{i-1}$ .

Dans le cas où un système dynamique est couplé à un modèle du type «integrate-and-fire» pour produire des ISIs, il semble exister une correspondance un à un entre les vecteurs à délais constitués d'ISIs et les états du système dynamique sous-jacent, pour autant que la dimension des vecteurs de reconstruction soit suffisamment grande (Réfs. [Sau94], [Sau95] et [CS99]). Selon ce qui a été vu au chapitre premier, ce fait implique que les ISIs peuvent être prédits à partir du passé. La question est

maintenant de savoir si pour des ISIs provenant d'un EEG, on peut en faire autant.

L'article de Scott et Schiff [SS95] aborde cette importante question dans le cas d'EEGs préictals. Dans cette étude, on a d'abord procédé à l'enregistrement d'un EEG qui couvre environ une heure de données et qui se termine tout juste avant le début de la crise, et ce pour quatre patients. Ces EEGs ont ensuite été convertis en ISIs au moyen d'un algorithme de détection automatique des pics. Il est à noter que cette conversion a été revue et corrigée à la main pour pallier à certaines défaillances de l'algorithme. Enfin, en utilisant d'une part un modèle ARIMA pour réaliser des prédictions linéaires et d'autre part l'espace des phases reconstruit pour réaliser des prédictions non-linéaires élémentaires, Scott et Schiff arrivent à la conclusion que pour trois des quatre patients, les ISIs peuvent être prédits de façon satisfaisante. Cependant, parmi ces trois, un seul requiert véritablement l'approche non-linéaire pour les prédictions.

Or comme le soulignent Scott et Schiff, la question la plus intéressante demeure la suivante : les ISIs préictals peuvent-ils permettre de prédire l'avènement de la crise épileptique ? Bien que l'échec de notre mesure de non-stationnarité à déceler des signes précurseurs ne permettrait pas de conclure ici par la négative, la réussite suggérerait fortement l'affirmative. En fait, on a toutes les raisons de croire que des signes précurseurs peuvent être identifiés à partir d'ISIs, puisque tel semble être le cas à partir de l'EEG. Notre mesure  $m_{\text{avg}}$  pourra-t-elle mettre en évidence des signes avant-coureurs de la crise épileptique à partir d'ISIs préictals ?

Deux des quatre patients de l'étude de Scott et Schiff seront analysés ici au moyen de  $m_{\text{avg}}$ , en l'occurrence ME et RO1 (données aimablement fournies par S.J. Schiff). Les Figs. 4.13 à 4.16 présentent certains résultats de cette analyse.

Pour ME, celui dont les ISIs nécessitent des prédictions non-linéaires, on observe effectivement à la Fig. 4.13b ou à la Fig. 4.14b (zoom) une montée progressive de  $m_{\text{avg}}$  jusqu'au début de la crise, montée similaire à celle de la Fig. 4.12b. Cette courbe relativement lisse est obtenue en utilisant une longueur appréciable pour la fenêtre

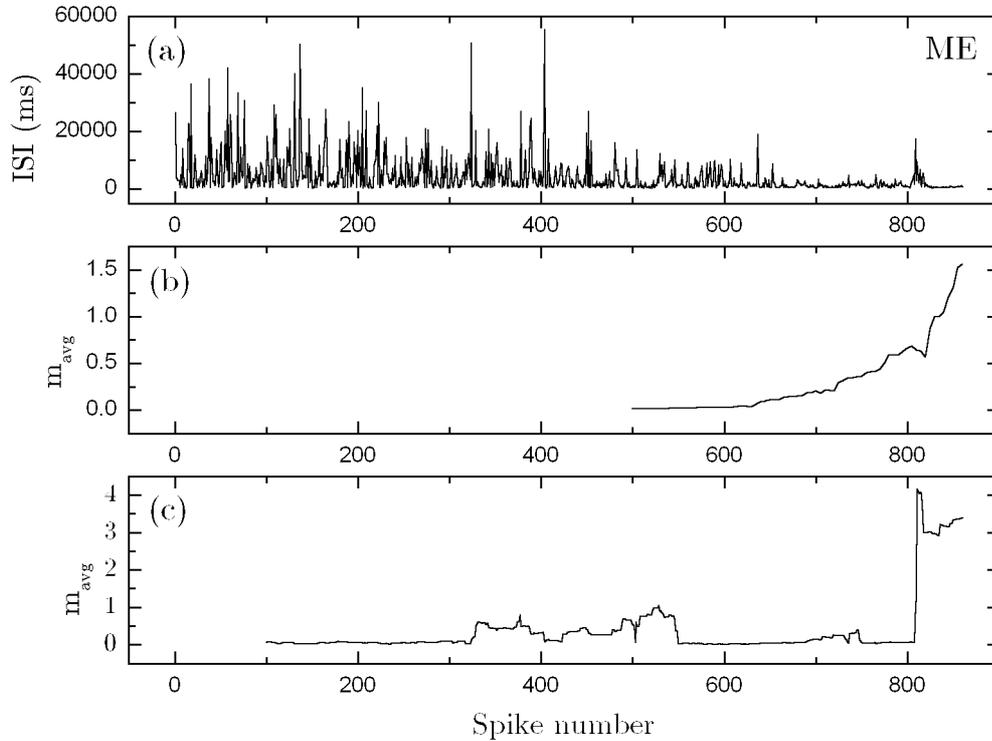


FIG. 4.13 – Patient ME. (a) ISIs. (b) Mesure  $m_{\text{avg}}$  ( $n = 5$ ,  $r_{\text{max}} = 1$ ,  $l_w = 500$  et  $t_w = 5$ ). (c) Mesure  $m_{\text{avg}}$  ( $n = 5$ ,  $r_{\text{max}} = 1$ ,  $l_w = 100$  et  $t_w = 1$ ).

glissante, soit  $l_w = 500$ . Chaque ordre de cumulants renvoie ici le même message de l'imminence d'une crise (courbes analogues). De façon certes un peu arbitraire, on peut fixer le seuil de détection à  $i \approx 750$ , ce à quoi correspond une prédiction d'environ 3 minutes.

Si l'on diminue  $l_w$  afin d'obtenir une meilleure résolution temporelle, certains éléments nouveaux apparaissent (cf. Fig. 4.13c). D'abord, on note que la croissance de  $m_{\text{avg}}$  vers le sommet de la crise s'avère considérablement plus abrupte. Ensuite, un certain plateau se dessine vers  $i \approx 325-550$ . S'agit-il d'un signe précurseur (prédiction d'environ 18 minutes) ou de simples fluctuations statistiques ?

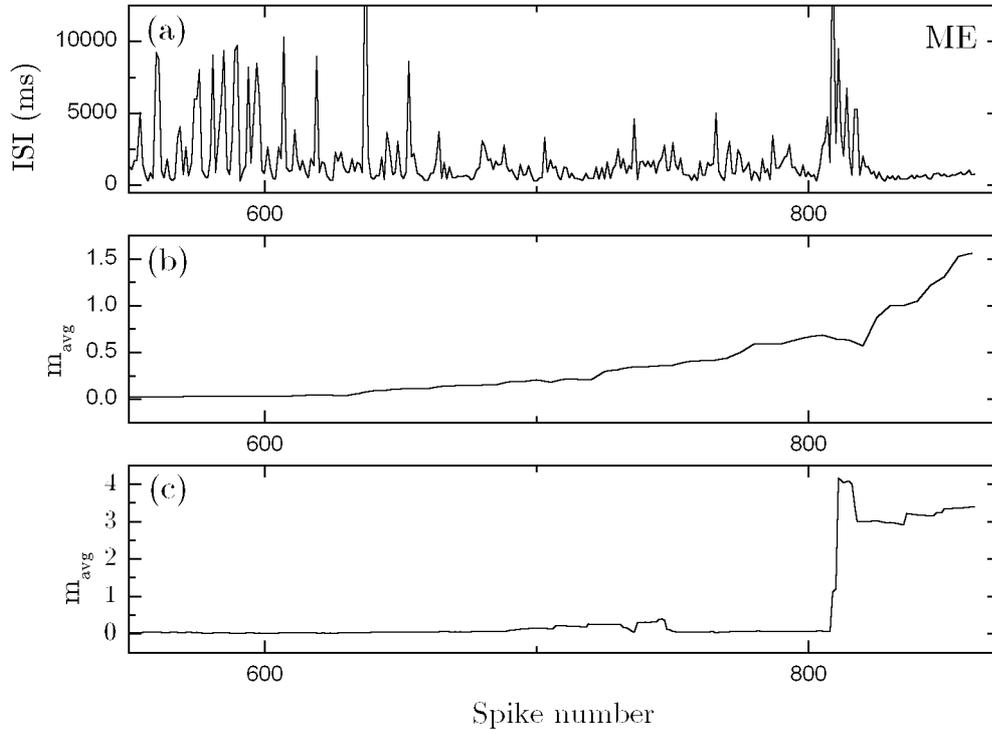


FIG. 4.14 – Zoom de la Fig. 4.13.

Malheureusement, dans sa forme actuelle, notre méthode de détection de non-stationnarité ne permet pas de répondre à une telle question. Éventuellement, il sera nécessaire de joindre un support statistique, probablement au moyen de la méthode des données synthétiques («surrogate data») [TEL<sup>+</sup>92], pour trancher dans de telles situations. Ce perfectionnement fera l'objet de travaux futurs.

Pour le patient RO1, il fut nécessaire d'introduire un délai  $\tau$  (cf. sous-section 4.1.2) pour obtenir une forme intéressante pour la courbe de  $m_{\text{avg}}$  (Fig. 4.15b ou 4.16b). Encore là, ce sommet atteint par la mesure de non-stationnarité vers  $i \approx 650$  constitue-t-il un signe précurseur (prédiction d'environ 15 minutes)? Bien que le message livré par les Figs. 4.13 à 4.16 ne dépende pas de façon cruciale du choix des différents paramètres pour  $m_{\text{avg}}$ , il est clair qu'un certain biais (recherche de signes

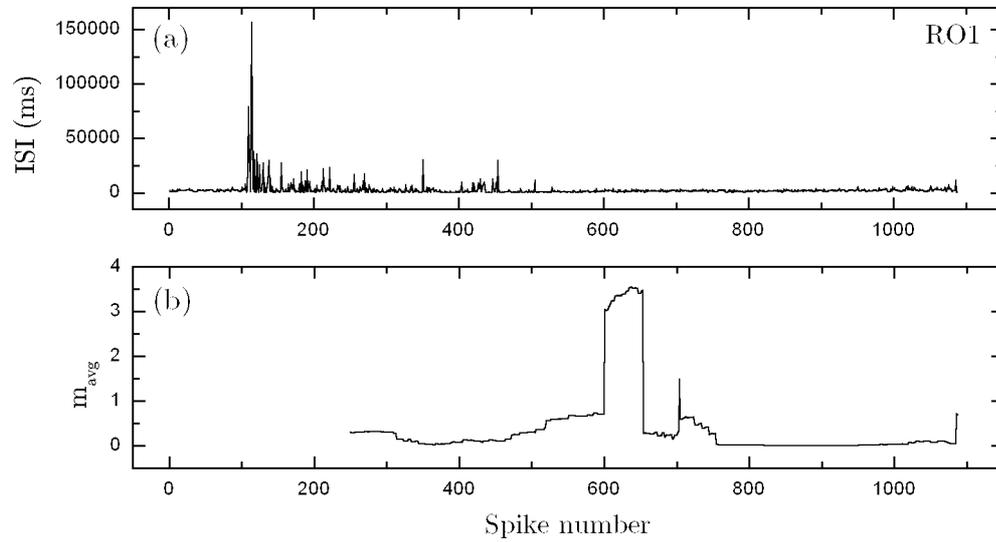


FIG. 4.15 – Patient RO1. (a) ISIs. (b) Mesure  $m_{\text{avg}}$  ( $\tau = 10$ ,  $n = 5$ ,  $r_{\text{max}} = 1$ ,  $l_w = 250$  et  $t_w = 1$ ).

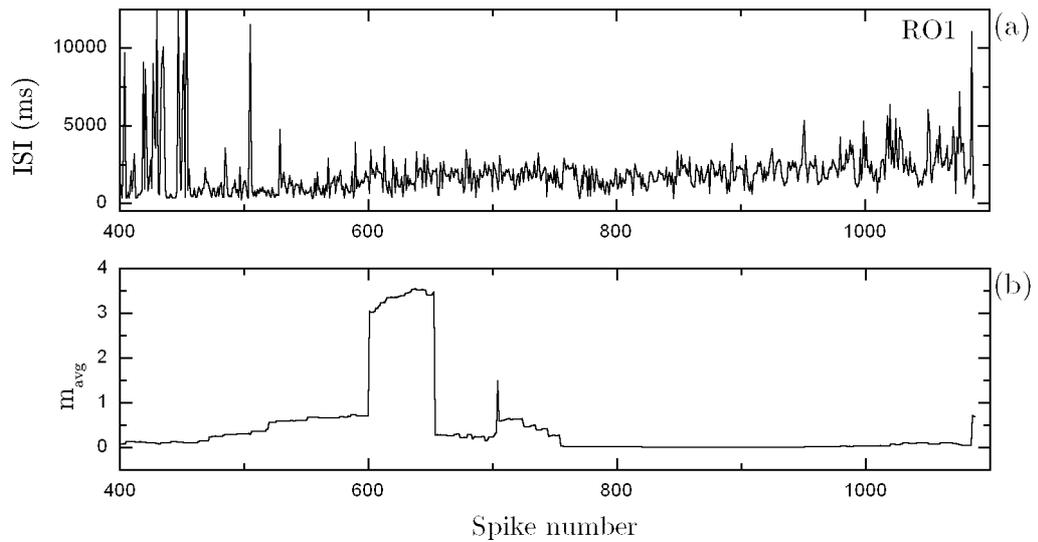


FIG. 4.16 – Zoom de la Fig. 4.15.

précurseurs) a conditionné leur sélection plutôt qu'une détermination systématique. En particulier, le choix de  $\tau$  n'est motivé ici que par le résultat obtenu, et un procédé méthodique de sélection fait défaut. Un effort a été tenté en ce sens au moyen de la fonction d'autocorrélation, mais sans succès immédiat.

Compte tenu des quelques points litigieux dont il a été discuté, il faut réaliser la teneur plus spéculative des résultats présentés ici avec les données épileptiques de Schiff. Somme toute, ces résultats fragmentaires demeurent très encourageants.

# Conclusion

Le présent mémoire abordait l'important problème de l'analyse des séries temporelles réelles à son niveau le plus fondamental, c.-à-d. celui qui concerne le traitement des caractéristiques potentiellement nuisibles du signal. Ainsi, des méthodes s'insérant dans le nouveau courant non-linéaire ont été présentées pour réduire le bruit d'une part, et pour détecter la non-stationnarité d'autre part.

Dans un premier temps, deux méthodes non-linéaires de réduction du bruit (Schreiber et GHKSS), particulièrement adaptées au cas des signaux chaotiques, ont été étudiées en détail. Ces deux méthodes n'utilisent pas la dynamique exacte du système, et sont par conséquent utilisables en pratique. La référence [KSH<sup>+</sup>93] présente à cet effet l'étude de deux signaux expérimentaux (écoulement de Taylor-Couette et laser NMR). Au mieux, ces deux méthodes permettent de réduire le niveau de bruit d'un ordre de grandeur, et les niveaux de bruit étudiés dans ce mémoire sont représentatifs du domaine où ces algorithmes sont véritablement efficaces. Aussi, même si l'optimisation des réductions de bruit a été discutée avec insistance, il n'en demeure pas moins que ces méthodes sont très robustes. En d'autres mots, l'obtention de bonnes performances ne dépend pas de façon critique des valeurs choisies pour les différents paramètres. En fait, le principal danger réside plutôt dans l'itération acharnée et aveugle de l'algorithme de réduction. Enfin, selon notre étude, la méthode GHKSS se révèle plus performante que celle de Schreiber, ce qui est corroboré par la comparaison des Réfs. [Sch93] et [GHK<sup>+</sup>93].

Néanmoins, il a été observé que ces deux méthodes performant beaucoup mieux dans le cas d'un bruit de mesure, pour lequel elles ont été conçues, que dans le cas d'un bruit dynamique. Rappelons que si le bruit de mesure modélise la précision finie de nos instruments d'observation, le bruit dynamique modélise quant à lui le fait qu'un système réel n'est jamais parfaitement isolé de son environnement, et ce malgré toutes les précautions de l'expérimentateur. Le traitement du bruit dynamique est donc tout aussi important, et de par sa nature, il semble nécessaire d'avoir plutôt recours à une méthode qui se concentre davantage sur l'aspect dynamique que sur l'aspect géométrique du signal contaminé. Ceci pourrait permettre de supprimer plus efficacement les régions de l'espace des phases visitées uniquement en raison de l'action du bruit dynamique. À cet égard, la nouvelle méthode proposée par Deco et al. (présentée brièvement à l'annexe A) constitue une innovation prometteuse. En effet, selon des résultats préliminaires, cette méthode pourrait être particulièrement performante pour réduire le bruit dynamique. Une publication prochaine [DBDa], où l'on compare directement cette méthode à celles de Schreiber et GHKSS au moyen des résultats de ce mémoire, permettra de confirmer ou non cette première tendance. À surveiller attentivement.

Dans un second temps, le subtil problème de la non-stationnarité d'une série temporelle a été abordé. D'emblée le caractère équivoque du concept même de la non-stationnarité d'un signal a été relevé, ce qui explique en partie la pluralité des approches. Nombre de ces approches ont été présentées sommairement dans ce mémoire, d'une part pour situer une nouvelle méthode de détection présentée plus loin, d'autre part pour offrir un panorama des outils actuellement disponibles, car rappelons que chacun peut être particulièrement adapté à certaines circonstances. Qui plus est, l'utilisation de plusieurs méthodes peut permettre, par exemple, de consolider un diagnostic de stationnarité.

Il a été également souligné que deux grands objectifs motivent l'étude de la non-stationnarité d'une série temporelle. Le premier revient à valider l'hypothèse de la

stationnarité du signal (test binaire), hypothèse incontournable pour la plupart des méthodes d'analyse. Le second objectif consiste à suivre les changements dynamiques qui s'opèrent au cours du temps (résolution temporelle de la non-stationnarité), et de façon ultime, à prédire l'avènement de certains événements dynamiques. Dans cette perspective, la non-stationnarité n'est pas considérée comme une caractéristique nuisible pour l'analyse, mais bien au contraire, comme le phénomène de première importance.

Ainsi, c'est le concept de flot d'information entre le passé et le futur, formellement défini à l'origine dans le cadre de la théorie de l'information, qui sous-tend une nouvelle méthode de détection de non-stationnarité, méthode qui a été présentée en détail dans ce mémoire. Une approche pratique pour sonder le flot d'information consiste à examiner directement la dépendance statistique entre un bloc de points consécutifs (le passé) et des points futurs, au moyen d'une mesure de corrélation exprimée en termes de cumulants. Comme c'est la dynamique du processus à l'origine de la série temporelle qui modèle cette dépendance statistique, l'évaluation de cette mesure de corrélation dans une fenêtre glissante engendre une véritable mesure de non-stationnarité. Le potentiel de cette nouvelle méthode de détection (résolution temporelle) a été démontré au moyen de cinq exemples artificiels (applications), dont quatre sont tirés directement de la littérature sur le sujet. En outre, ces exemples ont permis d'établir la robustesse de cette mesure de non-stationnarité par rapport aux différents paramètres qui peuvent affecter les performances de détection.

L'une des applications principales de l'analyse détaillée de la non-stationnarité d'un signal est sans contredit la prédiction de l'avènement d'une crise épileptique à partir d'un électroencéphalogramme. L'espoir clinique engendré par l'introduction récente des méthodes non-linéaires dans ce domaine a été mentionné, et à cet égard, la mesure de non-stationnarité présentée dans ce mémoire semble tout aussi prometteuse, du moins selon une première analyse de données épileptiques fournies par S.J. Schiff.

Néanmoins, cette étude de données réelles a également contribué à mettre en relief certaines lacunes de cette nouvelle approche, lacunes que de futurs travaux de recherche devront résorber. Premièrement, un support statistique adéquat fait cruellement défaut, afin d'établir la signification des variations observées de la mesure de non-stationnarité. Il faut envisager ici l'utilisation éventuelle de la méthode des données synthétiques («surrogate data») [TEL<sup>+</sup>92]. Deuxièmement, l'introduction d'un délai pour pallier à une fréquence d'échantillonnage trop élevée demeure une opération ad hoc, qu'il conviendra de systématiser. Enfin, il convient de souligner un certain inconfort théorique qui pourrait affecter la fiabilité de la méthode : l'emploi de l'hypothèse ergodique pour évaluer les différents cumulants à partir d'un segment stationnaire. Cette hypothèse n'est sûrement pas vérifiée dans tous les cas.

Quoi qu'il en soit, cette nouvelle méthode de détection de non-stationnarité fonctionne très bien, et certaines de ses qualités, fort intéressantes en pratique, lui confèrent certains avantages sur des méthodes alternatives. D'abord, le cadre théorique est très général, ce qui confère à la méthode une grande polyvalence. En particulier, l'hypothèse du chaos déterministe de basse dimension n'est pas utilisée, ce dont ne peuvent se prévaloir plusieurs méthodes non-linéaires, e.g. les travaux de Kennel [Ken97], de Yu et al. [YLH99] et Hively et al. [HGP99]. Ensuite, cette mesure de non-stationnarité est très rapide à calculer à l'ordinateur, ce qui facilite grandement un suivi en temps réel. Ceci pourrait être un facteur décisif lors de l'éventuelle conception du dispositif miniaturisé qui surveillera le développement d'une crise au foyer epileptogène. Or à ce niveau, autant le potentiel de notre mesure de non-stationnarité que le potentiel du traitement en lui-même demeurent à confirmer.

# Bibliographie

- [AMS<sup>+</sup>88] A.M. Albano, J. Muench, C. Schwartz, A.I. Mees, and P.E. Rapp. Singular-value decomposition and the Grassberger-Procaccia algorithm. *Phys. Rev. A*, **38** :3017–3026, 1988.
- [BBA90] P. Bryant, R. Brown, and H. Abarbanel. Lyapunov exponents from observed time series. *Phys. Rev. Lett.*, **65** :1523–1526, 1990.
- [BGMP97] L. Battiston, C. Gabrieli, T.A. Minelli, and A. Pascolini. Non-linear time series tests for future EEG applications. In *Analysis of the Electrical Activity of the Brain*, chapter 4. John Wiley & Sons, 1997.
- [BK86] D.S. Broomhead and G.P. King. Extracting qualitative dynamics from experimental data. *Physica D*, **20** :217–236, 1986.
- [BS93] C. Beck and F. Schlögl. *Thermodynamics of Chaotic Systems : an Introduction*. Cambridge Nonlinear Science Series. Cambridge University Press, 1993.
- [Cas87] M. Casdagli. Nonlinear prediction of chaotic time series. *Physica D*, **59** :845–866, 1987.
- [CDL<sup>+</sup>98] A. Capurro, L. Diambra, D. Lorenzo, O. Macadar, M.T. Martin, C. Mostaccio, A. Plastino, J. Pérez, E. Rofman, M.E. Torres, and J. Velluti. Human brain dynamics : the analysis of EEG signals with Tsallis information measure. *Physica A*, **265** :235–254, 1998.

- [CH92a] R. Cawley and G.-H. Hsu. Local-geometric-projection method for noise reduction in chaotic maps and flows. *Phys. Rev. A*, **46** :3057–3082, 1992.
- [CH92b] R. Cawley and G.-H. Hsu. SNR performance of a noise reduction algorithm applied to coarsely sampled chaotic data. *Phys. Lett. A*, **166** :188, 1992.
- [CMJ97] L. Cao, A. Mees, and K. Judd. Modeling and predicting non-stationary time series. *Int. J. Bifurcation and Chaos*, **7** :1823–1831, 1997.
- [CS99] R. Castro and T. Sauer. Reconstructing chaotic dynamics through spike filters. *Phys. Rev. E*, **59** :2911–2917, 1999.
- [Dav98] M.E. Davies. Nonlinear noise reduction through Monte Carlo sampling. *Chaos*, **8** :775–781, 1998.
- [DBDa] G. Deco, F. Beaulieu, and L.J. Dubé. Prediction and noise filtering in dynamical systems : A neural network approach. (in preparation).
- [DBDb] L.J. Dubé, F. Beaulieu, and G. Deco. A novel measure of nonstationarity in dynamical systems. (in preparation).
- [DO96] G. Deco and D. Obradovic. *An Information-Theoretic Approach to Neural Computing*. Perspectives in Neural Computing. Springer, 1996.
- [DSS97a] G. Deco, C. Schittenkopf, and B. Schürmann. Determining the information flow of dynamical systems from continuous probability distributions. *Phys. Rev. Lett.*, **78** :2345–2348, 1997.
- [DSS97b] G. Deco, C. Schittenkopf, and B. Schürmann. Dynamical analysis of time series by statistical tests. *Int. J. Bifurcation and Chaos*, **7** :2629–2652, 1997.
- [DSS97c] G. Deco, C. Schittenkopf, and B. Schürmann. Information flow in chaotic symbolic dynamics for finite and infinitesimal resolution. *Int. J. Bifurcation and Chaos*, **7** :97–105, 1997.

- [EKRC86] J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, and S. Ciliberto. Liapunov exponents from time series. *Phys. Rev. A*, **34** :4971–4979, 1986.
- [FS86] A.M. Fraser and H.L. Swinney. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, **33** :1134–1140, 1986.
- [Gar85] C.W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer Series in Synergetics. Springer, second edition, 1985.
- [GHK<sup>+</sup>93] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber. On noise reduction methods for chaotic data. *Chaos*, **3** :421–435, 1993.
- [Gla97] J. Glanz. Mastering the nonlinear brain. *Science*, **277** :1758–1760, 1997.
- [GSS91] P. Grassberger, T. Schreiber, and C. Schaffrath. Nonlinear time sequence analysis. *Int. J. Bifurcation and Chaos*, **1** :521–547, 1991.
- [H76] M. Hénon. A two-dimensional mapping with a strange attractor. *Comm. Math. Phys.*, **50** :69, 1976.
- [HGP99] L.M. Hively, P.C. Gailey, and V.A. Protopopescu. Detecting dynamical change in nonlinear time series. *Phys. Lett. A*, **258** :103–114, 1999.
- [HKS99] R. Hegger, H. Kantz, and T. Schreiber. Practical implementation of nonlinear time series methods : The TISEAN package. *Chaos*, **9** :413–435, 1999.
- [KBA92] M. Kennel, R. Brown, and H. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A*, **45** :3403–3411, 1992.
- [Ken97] M.B. Kennel. Statistical test for dynamical nonstationarity in observed time-series data. *Phys. Rev. E*, **56** :316–321, 1997.

- [KS93] E.J. Kostelich and T. Schreiber. Noise reduction in chaotic time-series data : A survey of common methods. *Phys. Rev. E*, **48** :1752–1763, 1993.
- [KS97] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge Nonlinear Science Series. Cambridge University Press, 1997.
- [KSH<sup>+</sup>93] H. Kantz, T. Schreiber, I. Hoffmann, T. Buzug, G. Pfister, L.G. Flepp, J. Simonet, R. Badii, and E. Brun. Nonlinear noise reduction : A case study on experimental data. *Phys. Rev. E*, **48** :1529–1538, 1993.
- [KWT98] K. Kopitzki, P.C. Warnke, and J. Timmer. Quantitative analysis by renormalized entropy of invasive electroencephalograph recordings in focal epilepsy. *Phys. Rev. E*, **58** :4859–4864, 1998.
- [LE95] K. Lehnertz and C.E. Elger. Spatio-temporal dynamics of the primary epileptogenic area in temporal lobe epilepsy characterized by neuronal complexity loss. *Electroencephalography and clinical Neurophysiology*, **95** :108–117, 1995.
- [LE97] K. Lehnertz and C.E. Elger. Neuronal complexity loss in temporal lobe epilepsy : effects of carbamazepine on the dynamics of the epileptogenic focus. *Electroencephalography and clinical Neurophysiology*, **103** :376–380, 1997.
- [LE98] K. Lehnertz and C.E. Elger. Can epileptic seizures be predicted ? Evidence from nonlinear time series analysis of brain electrical activity. *Phys. Rev. Lett.*, **80** :5019–5022, 1998.
- [LM93] S. Lessard and Monga. *Statistique : concepts et méthodes avec exercices et corrigés*. Les Presses de l'Université de Montréal, 1993.
- [LVQMAV97] M. Le Van Quyen, J. Martinerie, C. Adam, and F.J. Varela. Unstable periodic orbits in human epileptic activity. *Phys. Rev. E*, **56** :3401–3411, 1997.

- [MALVQ<sup>+</sup>98] J. Martinerie, C. Adam, M. Le Van Quyen, M. Baulac, S. Clemenceau, B. Renault, and F.J. Varela. Epileptic seizures can be anticipated by non-linear analysis. *Nature Med.*, **4** :1173–1176, 1998.
- [MG77] M.C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, **197** :287, 1977.
- [MS96] R. Manuca and R. Savit. Stationarity and nonstationarity in time series analysis. *Physica D*, **99** :134–161, 1996.
- [OGY90] E. Ott, C. Grebogi, and J.A. Yorke. Controlling chaos. *Phys. Rev. Lett.*, **64** :1196–1199, 1990.
- [OSY94] E. Ott, T. Sauer, and J.A. Yorke. *Coping with Chaos : Analysis of Chaotic Data and The Exploitation of Chaotic Systems*. Wiley Series in Nonlinear Science. John Wiley & Sons, 1994.
- [Par92] U. Parlitz. Identification of true and spurious Lyapunov exponents from time series. *Int. J. Bifurcation and Chaos*, **2** :155–165, 1992.
- [PCFS80] N.H. Packard, J.P. Crutchfield, J.D. Farmer, and R.S. Shaw. Geometry from a time series. *Phys. Rev. Lett.*, **45** :712–715, 1980.
- [PFTV92] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C : The Art of Scientific Computing*. Cambridge University Press, second edition, 1992.
- [QALG] R.Q. Quiroga, J. Arnhold, K. Lehnertz, and P. Grassberger. Kullback-Leibler and renormalized entropy : Applications to EEGs of epilepsy patients.
- [Sau92] T. Sauer. A noise reduction method for signals from nonlinear systems. *Physica D*, **58** :193, 1992.
- [Sau94] T. Sauer. Reconstruction of dynamical systems from interspike intervals. *Phys. Rev. Lett.*, **72** :3811–3814, 1994.

- [Sau95] T. Sauer. Interspike interval embedding of chaotic signals. *Chaos*, **5** :127–132, 1995.
- [Sch93] T. Schreiber. Extremely simple nonlinear noise-reduction method. *Phys. Rev. E*, **47** :2401–2404, 1993.
- [Sch97] T. Schreiber. Detecting and analyzing nonstationarity in a time series using nonlinear cross-predictions. *Phys. Rev. Lett.*, **78** :843–846, 1997.
- [Sch98] S.J. Schiff. Forecasting brain storms. *Nature Med.*, **4** :117–118, 1998.
- [Sch99] T. Schreiber. Interdisciplinary application of nonlinear time series methods. *Phys. Rep.*, **308**, 1999.
- [SD97] C. Schittenkopf and G. Deco. Identification of deterministic chaos by an information-theoretic measure of the sensitive dependence on the initial conditions. *Physica D*, **110** :173–181, 1997.
- [SFN<sup>+</sup>98] P So, J.T. Francis, T.I. Netoff, B.J. Gluckman, and S.J. Schiff. Periodic orbits : A new language for neuronal dynamics. *Biophys. J.*, **74** :2776–2785, 1998.
- [SG91] T. Schreiber and P. Grassberger. A simple noise-reduction method for real data. *Phys. Lett. A*, **160** :411, 1991.
- [Sha81] R. Shaw. Strange attractors, chaotic behavior, and information flow. *Z. Naturforsch.*, **36a** :80–112, 1981.
- [SS95] D.A. Scott and S.J. Schiff. Predictability of EEG Interictal Spikes. *Biophys. J.*, **69** :1748–1757, 1995.
- [Str94] S.H. Strogatz. *Nonlinear dynamics and chaos : with applications to physics, biology, chemistry and engineering*. Addison-Wesley, 1994.
- [STY98] T.D. Sauer, J.A. Tempkin, and J.A. Yorke. Spurious lyapunov exponents in attractor reconstruction. *Phys. Rev. Lett.*, **81** :4341–4344, 1998.

- [SYC91] T. Sauer, J.A. Yorke, and M. Casdagli. Embedology. *J. Stat. Phys.*, **65** :579–616, 1991.
- [Tak81] F. Takens. Detecting strange attractors in turbulence. In *Lectures Notes in Mathematics*, volume **898**. Springer, 1981.
- [TEL<sup>+</sup>92] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J.D. Farmer. Testing for nonlinearity in time series : the method of surrogate data. *Physica D*, **58** :77–94, 1992.
- [The95] J. Theiler. On the evidence for low-dimensional chaos in an epileptic electroencephalogram. *Phys. Lett. A*, **196** :335–341, 1995.
- [Tim98] J. Timmer. Power of surrogate data testing with respect to nonstationarity. *Phys. Rev. E*, **58** :5153–5156, 1998.
- [WKP98] A. Witt, J. Kurths, and A. Pikovsky. Testing stationarity in time series. *Phys. Rev. E*, **58** :1800–1810, 1998.
- [YLH98] D. Yu, W. Lu, and R.G. Harrison. Space time-index plots for probing dynamical nonstationarity. *Phys. Lett. A*, **250** :323–327, 1998.
- [YLH99] D. Yu, W. Lu, and R.G. Harrison. Detecting dynamical nonstationarity in time series data. *Chaos*, **9** :865–870, 1999.

# Annexe A

## Réduction du bruit : réseau de neurones

Dans cette annexe, nous allons présenter, dans ses grandes lignes, une nouvelle méthode de réduction du bruit [DBDa]. La section 4.2 est préalable à la compréhension de l'exposé qui va suivre.

L'idée de cette nouvelle méthode consiste à isoler de façon explicite la composante déterministe de la composante stochastique d'un signal contaminé, afin d'utiliser l'estimation obtenue de la composante déterministe pour générer un signal plus pur. Pour réaliser cette extraction, il est nécessaire de sélectionner un modèle approprié pour la composante déterministe, de même qu'identifier le type de bruit gaussien présent (dynamique ou observationnel). Un réseau de neurones permet l'apprentissage de la composante déterministe selon un modèle mathématique et un type de bruit donnés. Le flot d'information (version cumulant) est ensuite employé comme statistique dans une série de tests d'hypothèses, afin de déterminer le modèle mathématique et le type de bruit qui décrivent le mieux le signal contaminé (coïncidence du profil  $m(n, r)$  vs  $r$ ). Dans le cas d'un bruit de mesure, l'ordre du processus de Markov correspondant au modèle déterministe sélectionné peut être identifié à la dimension minimale de

recouvrement. Les détails mathématiques de cette méthode seront présentés dans la référence [DBDa].

Voyons un exemple. Il s'agit de filtrer l'attracteur de Hénon contaminé par du bruit dynamique gaussien (cf. Fig. A.2a). Le flot d'information du signal contaminé apparaît à la Fig. A.1 sous le vocable «original». En considérant une dimension de recouvrement  $m = 2$ , un réseau de neurones a appris la composante déterministe selon l'hypothèse d'un bruit dynamique d'une part, et selon l'hypothèse d'un bruit observationnel d'autre part. La Fig. A.1 montre le flot d'information pour chacune de ces deux hypothèses (bruit ajouté à la composante déterministe). De façon très claire, le type de bruit est correctement identifié, et l'itération du modèle déterministe correspondant génère un attracteur beaucoup plus net (cf. Fig. A.2b). En particulier, on observe que cette méthode parvient à supprimer efficacement les régions de l'espace des phases visitées uniquement en raison du bruit dynamique.

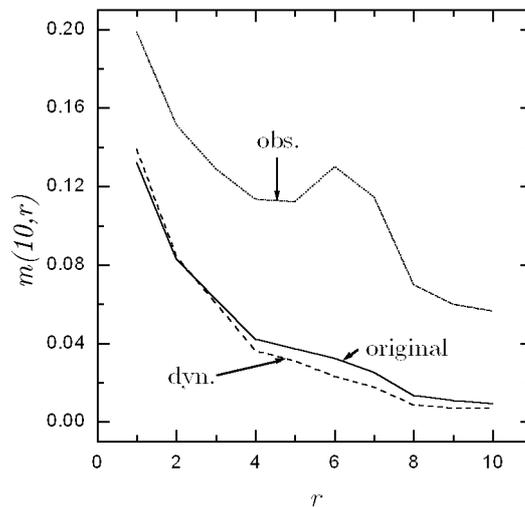


FIG. A.1 – Flot d'information  $m(10, r)$  vs  $r$ .

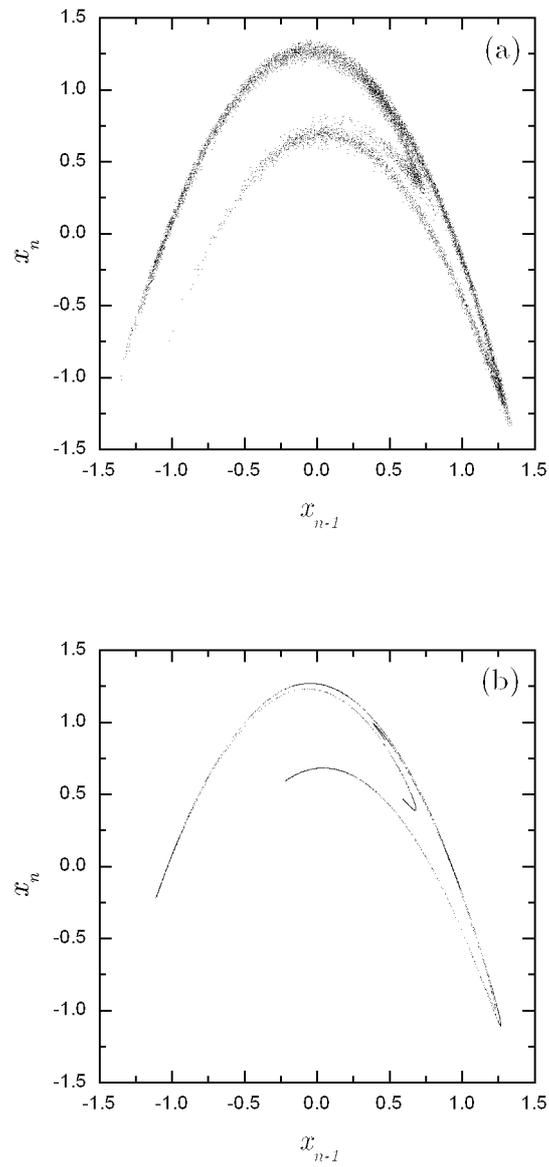


FIG. A.2 – Attracteur de Hénon : (a) contaminé par un bruit dynamique gaussien, (b) filtré au moyen d'un réseau de neurones.

## Annexe B

### Matrice de covariance

Considérons une série temporelle scalaire  $\{y_t\}_{t=1}^N$  issue de la mesure d'une observable d'un système dynamique, et un recouvrement approprié en dimension  $m$  ( $\tau = 1$ ) :  $\mathbf{y}_t = (y_t, y_{t-1}, \dots, y_{t-(m-1)})$ . Focalisons maintenant notre attention sur l'un des points dans l'espace de recouvrement, en l'occurrence  $\mathbf{y}_n$ , dont le voisinage rappelons-le est noté  $\mathcal{U}_\epsilon^{(n)}$ . Les proches voisins de  $\mathbf{y}_n$  sont au nombre de  $|\mathcal{U}_\epsilon^{(n)}|$ , et pour des raisons de commodité, seront identifiés de la façon suivante :  $\{\mathbf{y}_k^{(n)}\}_{k=1}^{|\mathcal{U}_\epsilon^{(n)}|}$ . Le centre de masse de ces voisins s'évalue comme suit :

$$\boldsymbol{\xi}^{(n)} = \langle \mathbf{y}_k^{(n)} \rangle = \frac{1}{|\mathcal{U}_\epsilon^{(n)}|} \sum_{k=1}^{|\mathcal{U}_\epsilon^{(n)}|} \mathbf{y}_k^{(n)}, \quad (\text{B.1})$$

et sa  $i^{\text{ième}}$  composante sera notée  $\xi_i^{(n)}$ . De façon similaire, la  $i^{\text{ième}}$  composante de  $\mathbf{y}_k^{(n)}$  sera notée  $y_{ki}^{(n)}$ . On peut enfin construire la matrice  $|\mathcal{U}_\epsilon^{(n)}| \times m$  suivante :

$$\mathbf{Y}^{(n)} = \frac{1}{\sqrt{|\mathcal{U}_\epsilon^{(n)}|}} \begin{pmatrix} y_{11}^{(n)} - \xi_1^{(n)} & y_{12}^{(n)} - \xi_2^{(n)} & \dots & y_{1m}^{(n)} - \xi_m^{(n)} \\ y_{21}^{(n)} - \xi_1^{(n)} & y_{22}^{(n)} - \xi_2^{(n)} & \dots & y_{2m}^{(n)} - \xi_m^{(n)} \\ \vdots & \vdots & \vdots & \vdots \\ y_{|\mathcal{U}_\epsilon^{(n)}|1}^{(n)} - \xi_1^{(n)} & y_{|\mathcal{U}_\epsilon^{(n)}|2}^{(n)} - \xi_2^{(n)} & \dots & y_{|\mathcal{U}_\epsilon^{(n)}|m}^{(n)} - \xi_m^{(n)} \end{pmatrix}. \quad (\text{B.2})$$

De façon équivalente :

$$Y_{ki}^{(n)} = \frac{1}{\sqrt{|\mathcal{U}_\epsilon^{(n)}|}} (y_{ki}^{(n)} - \xi_i^{(n)}). \quad (\text{B.3})$$

À un facteur multiplicatif près, chaque rangée de  $\mathbf{Y}^{(n)}$  correspond à un proche voisin de  $\mathbf{y}_n$ , dans le repère du centre de masse. La matrice  $m \times m$  symétrique  $\mathbf{C}^{(n)} = [\mathbf{Y}^{(n)}]^T \mathbf{Y}^{(n)}$  est appelée *matrice de covariance*. Examinons la signification de cette matrice. On observe d'abord :

$$C_{ij}^{(n)} = \sum_{k=1}^{|\mathcal{U}_\epsilon^{(n)}|} Y_{ki}^{(n)} Y_{kj}^{(n)} \quad (\text{B.4})$$

$$= \frac{1}{|\mathcal{U}_\epsilon^{(n)}|} \sum_{k=1}^{|\mathcal{U}_\epsilon^{(n)}|} (y_{ki}^{(n)} - \xi_i^{(n)})(y_{kj}^{(n)} - \xi_j^{(n)}) \quad (\text{B.5})$$

$$= \frac{1}{|\mathcal{U}_\epsilon^{(n)}|} \sum_{k=1}^{|\mathcal{U}_\epsilon^{(n)}|} (y_{ki}^{(n)} y_{kj}^{(n)} + \xi_i^{(n)} \xi_j^{(n)} - \xi_i^{(n)} y_{kj}^{(n)} - \xi_j^{(n)} y_{ki}^{(n)}) \quad (\text{B.6})$$

$$= \langle y_{ki}^{(n)} y_{kj}^{(n)} \rangle + \xi_i^{(n)} \xi_j^{(n)} - \xi_i^{(n)} \langle y_{kj}^{(n)} \rangle - \xi_j^{(n)} \langle y_{ki}^{(n)} \rangle. \quad (\text{B.7})$$

Or par définition (cf. (B.1))  $\langle y_{ki}^{(n)} \rangle = \xi_i^{(n)}$  et on obtient donc finalement [LM93] :

$$C_{ij}^{(n)} = \langle y_{ki}^{(n)} y_{kj}^{(n)} \rangle - \langle y_{ki}^{(n)} \rangle \langle y_{kj}^{(n)} \rangle \quad (\text{B.8})$$

$$= \text{Cov}(y_{ki}^{(n)}, y_{kj}^{(n)}). \quad (\text{B.9})$$

La covariance d'une variable avec elle-même est la variance. En effet :

$$\text{Cov}(y_{ki}^{(n)}, y_{ki}^{(n)}) = \langle (y_{ki}^{(n)})^2 \rangle - \langle y_{ki}^{(n)} \rangle^2 \quad (\text{B.10})$$

$$= \sigma^2(y_{ki}^{(n)}). \quad (\text{B.11})$$

De plus, la covariance de deux variables indépendantes est nulle. Ceci découle simplement du fait que l'on a dans ce cas :  $\langle y_{ki}^{(n)} y_{kj}^{(n)} \rangle = \langle y_{ki}^{(n)} \rangle \langle y_{kj}^{(n)} \rangle$ . Par conséquent, si la matrice de covariance n'est pas diagonale, il en résulte que les  $m$  composantes des proches voisins, dans le présent repère du centre de masse et considérées comme des variables aléatoires, ne sont pas indépendantes. Or un changement de base peut fournir  $m$  nouvelles composantes, combinaisons linéaires des anciennes, qui seront cette fois statistiquement indépendantes. Il s'agit à cet effet de diagonaliser  $\mathbf{C}^{(n)}$ .

Pour ce faire, nous utiliserons la décomposition en valeurs singulières de la matrice  $\mathbf{Y}^{(n)}$ , décomposition possible si  $|\mathcal{U}_\epsilon^{(n)}| \geq m$  [AMS<sup>+</sup>88] :

$$\mathbf{Y}^{(n)} = \mathbf{V} \mathbf{\Sigma} \mathbf{U}^T, \quad (\text{B.12})$$

où  $\mathbf{V}$  est une matrice  $|\mathcal{U}_\epsilon^{(n)}| \times m$  dont les colonnes sont orthogonales, i.e.

$$(\mathbf{V}^T \mathbf{V})_{ij} = \delta_{ij}. \quad (\text{B.13})$$

$\Sigma$  est une matrice diagonale  $m \times m$ , i.e.  $\Sigma_{ij} = \sigma_i \delta_{ij}$ , et enfin  $\mathbf{U}$  est une matrice orthogonale  $m \times m$ . Utilisant cette décomposition, on calcule :

$$\mathbf{C}^{(n)} = \mathbf{U} \Sigma^T \mathbf{V}^T \mathbf{V} \Sigma \mathbf{U}^T \quad (\text{B.14})$$

$$= \mathbf{U} \Sigma \mathbf{1} \Sigma \mathbf{U}^T \quad (\text{B.15})$$

$$= \mathbf{U} \Sigma^2 \mathbf{U}^{-1}, \quad (\text{B.16})$$

où  $\mathbf{1}$  est la matrice identité. Pour obtenir la dernière égalité, nous avons utilisé le caractère orthogonal de  $\mathbf{U}$ , i.e.  $\mathbf{U}^T = \mathbf{U}^{-1}$ . On reconnaît la diagonalisation de la matrice de covariance : les colonnes de  $\mathbf{U}^T$  sont les vecteurs propres de  $\mathbf{C}^{(n)}$ , et les  $\sigma_i^2$  les valeurs propres correspondantes. Comme la matrice de covariance est symétrique, ses valeurs propres sont réelles et ici non-négatives. Comme  $\mathbf{U}^T$  est orthogonal, les vecteurs propres de la matrice de covariance forment une base orthonormée. Les directions associées sont appelées *axes principaux*. De notre discussion précédente concernant la covariance, on note que les composantes des proches voisins selon ces axes principaux sont bien des variables indépendantes statistiquement, de variance  $\sigma_i^2$  selon l'axe  $i$ . Si un  $\sigma_i^2$  est nul, c'est donc dire que localement (proche de  $\mathbf{y}_n$ ), la dynamique reconstruite ne s'étend pas selon cet axe. Soit  $l \leq m$  le nombre d'axes principaux visités ( $\sigma_i^2 > 0$ ) pour une série temporelle exempte de bruit. L'entier  $l$  correspond alors à la dimension minimale d'un sous-espace linéaire qui renferme tous les proches voisins de  $\mathbf{y}_n$  [GHK<sup>+</sup>93], i.e. à la dimension de recouvrement minimale  $m_{\min}$ .

La correspondance un à un entre l'espace des phases original et celui reconstruit permet d'importer dans ce dernier le déterminisme du premier. Concrètement, cela se traduit de la façon suivante :

$$\mathbf{y}_n = \tilde{\mathbf{F}}(\mathbf{y}_{n-1}), \quad (\text{B.17})$$

où  $\tilde{\mathbf{F}}$  est une hypothétique application non-linéaire, de la forme suivante :

$$\tilde{F}_i(\mathbf{y}) = \begin{cases} \tilde{f}(\mathbf{y}) & \text{si } i = 1 \\ (\mathbf{y})_{i-1} & \text{sinon.} \end{cases} \quad (\text{B.18})$$

En terme des mesures de la série temporelle, cette application non-linéaire est équivalente à la contrainte :

$$y_n = \tilde{f}(y_{n-1}, y_{n-2}, \dots, y_{n-m}). \quad (\text{B.19})$$

Si la dimension de recouvrement correspond à celle minimale requise, i.e  $m = m_{\min}$ , alors les  $m$  composantes des vecteurs à délais sont des variables indépendantes. Pour  $m = m_{\min} + 1$ , la contrainte (B.19) conserve à  $m_{\min}$  le nombre de composantes indépendantes. Cette contrainte peut être réécrite sous forme implicite :

$$f(y_n, y_{n-1}, y_{n-2}, \dots, y_{n-m}) = y_n - \tilde{f}(\mathbf{y}_{n-1}) = 0. \quad (\text{B.20})$$

En fait, le nombre de composantes indépendantes de  $\mathbf{y}_n$  sera constant  $\forall m \geq m_{\min}$  (recouvrement adéquat), et en général pour  $m = m_{\min} + Q$ , ce sera l'existence de  $Q$  contraintes

$$f_q(y_n, y_{n-1}, y_{n-2}, \dots, y_{n-m}) = 0, \quad q = 1, \dots, Q \quad (\text{B.21})$$

qui s'en chargera.

L'approche généralement retenue consiste à déterminer une approximation locale de ces contraintes. En d'autres termes, dans le voisinage  $\mathcal{U}_\epsilon^{(n)}$  de  $\mathbf{y}_n$ , les contraintes peuvent être écrites sous la forme affine suivante :

$$f_q^{(n)}(\mathbf{y}_k^{(n)}) \approx \mathbf{a}_q^{(n)} \cdot \mathbf{y}_k^{(n)} + b_q^{(n)} = 0, \quad (\text{B.22})$$

où  $\mathbf{a}_q^{(n)}$  est un vecteur constant et  $b_q^{(n)}$  une constante. Il ne s'agit que d'une simple linéarisation :

$$f_q^{(n)}(\mathbf{y}_k^{(n)}) = f_q^{(n)}(\mathbf{y}_n - (\mathbf{y}_n - \mathbf{y}_k^{(n)})) \quad (\text{B.23})$$

$$\approx f_q^{(n)}(\mathbf{y}_n) + \nabla f_q(\mathbf{y}_n) \cdot (\mathbf{y}_n - \mathbf{y}_k^{(n)}), \quad (\text{B.24})$$

et par conséquent on peut identifier :

$$\mathbf{a}_q^{(n)} = \nabla f_q(\mathbf{y}_n). \quad (\text{B.25})$$

Chacune des  $Q$  contraintes locales correspond ainsi à un hyperplan (sous-espace linéaire de dimension  $(m-1)$ ) dans l'espace de recouvrement à  $m$  dimensions. L'intersection des  $Q$  hyperplans définit le sous-espace linéaire de dimension  $m-Q$  renfermant les proches voisins de  $\mathbf{y}_n$ . Ce sous-espace linéaire sur lequel vit le voisinage de  $\mathbf{y}_n$  est précisément celui engendré par les vecteurs propres dominants ( $\sigma_i^2 > 0$ ) de la matrice de covariance  $\mathbf{C}^{(n)}$ , introduite précédemment. De plus, en raison des propriétés d'un gradient,  $\mathbf{a}_q^{(n)}$  est orthogonal à l'hyperplan  $q$ , et donc au sous-espace linéaire abritant les voisins.

En présence de bruit, aucune des valeurs propres  $\sigma_i^2$  de la matrice de covariance ne sera identiquement nulle [BK86]. Le bruit, généralement associé à un signal de très haute dimension, peut ainsi donner l'impression que les voisins de  $\mathbf{y}_n$  habitent pleinement l'espace linéaire de dimension  $m$ . Or pour un niveau de bruit raisonnable, on s'attend à ce que pour les directions associées à l'exploration due au bruit, les valeurs propres  $\sigma_i^2$  soient faibles. Ceci suggère immédiatement une méthode de réduction du bruit : projeter  $\mathbf{y}_n$  (bruité) sur le sous-espace vectoriel associé aux vecteurs propres *dominants* (les  $\sigma_i^2$  les plus élevés) de  $\mathbf{C}^{(n)}$ . Ce sous-espace vectoriel devrait être de dimension  $m-Q$ , et si l'on note par  $\boldsymbol{\theta}_n$  la correction apportée à  $\mathbf{y}_n$ , l'Ansatz suivant est tout désigné :

$$\boldsymbol{\theta}_n = \sum_{q=1}^Q \mu_{nq} \mathbf{a}_q^{(n)}. \quad (\text{B.26})$$

Enfin, si l'on utilise plutôt la métrique truquée  $\mathbf{P}$  de l'algorithme GHKSS (cf. section 2.4), la correction prend plutôt la forme :

$$\boldsymbol{\theta}_n = \sum_{q=1}^Q \mu_{nq} \mathbf{P} \mathbf{a}_q^{(n)}, \quad (\text{B.27})$$

forme que l'on voit apparaître lors de la résolution du problème de minimisation ( $\mu_{nq}$  est un multiplicateur de Lagrange).

## Annexe C

# Implémentation du déplacement de Bernoulli

L'arithmétique flottante de l'ordinateur rend l'implémentation directe du déplacement de Bernoulli inappropriée. Au niveau binaire, chaque itération revient à une translation des bits de la mantisse vers la gauche, conjuguée à l'abandon du bit qui était situé à l'extrême gauche. Le problème provient du fait que l'ordinateur comble automatiquement le vide créé à l'extrême droite par un bit nul (du moins en  $\mathbb{C}$ ). Après un certain nombre d'itérations, on arrive donc inévitablement à zéro, pour y rester.

En représentation binaire, le déplacement de Bernoulli se résume donc à déplacer la virgule vers la droite sur la condition initiale, et à abandonner les bits qui passent à gauche de la virgule. Rappelons que multiplier par deux revient à décaler les bits vers la gauche. Soit en effet

$$x = \sum_{i=-\infty}^{\infty} b_i 2^i \tag{C.1}$$

et

$$2x = \sum_{i=-\infty}^{\infty} b'_i 2^i. \tag{C.2}$$

On a

$$2x = \sum_{i=-\infty}^{\infty} b_i 2^{i+1} \quad (\text{C.3})$$

$$= \sum_{i=-\infty}^{\infty} b_{i-1} 2^i, \quad (\text{C.4})$$

et donc  $b'_i = b_{i-1}$ . Une condition initiale en représentation flottante ne renferme qu'un nombre fini de bits ; rapidement ceux-ci sont effacés par le déplacement de Bernoulli. On comprend donc que pour générer une série temporelle d'une longueur appréciable, il faut présenter à l'application une condition initiale extrêmement précise (des milliers de bits), ce que ne permet pas l'ordinateur.

Afin de contourner ce problème, l'astuce est de générer cette condition initiale très fine au fil des itérations, en choisissant de façon aléatoire le nouveau bit à droite (probabilité 1/2 d'être nul) à chaque itération.<sup>1</sup> Pour ce faire, la mantisse de la condition initiale (nombre flottant) est stockée dans un nombre entier, et on peut alors manipuler directement les bits utilisant les opérateurs du C («bitwise operators»). Il est à noter toutefois que l'implémentation naïve qui consiste à remplacer 2 par 1.999999999999 dans l'application fonctionne tout aussi bien.

---

<sup>1</sup>Évidemment, l'infiniment précise condition initiale où tous les bits à droite d'un certain bit sont nuls conduira inévitablement au point fixe zéro. Néanmoins, dans notre approche, cette possibilité a une probabilité excessivement faible de se réaliser.

## Annexe D

# Fonction d'autocorrélation

La fonction d'autocorrélation d'un signal scalaire  $\{x_t\}$  (temps discret) est définie de la façon suivante [BS93] :

$$C(\tau) = \lim_{N \rightarrow \infty} \left( \frac{1}{N} \sum_{t=1}^N x_t x_{t+\tau} \right) - \bar{x}^2, \quad (\text{D.1})$$

où  $\tau$  est un entier positif et  $\bar{x}$  la moyenne temporelle, i.e.

$$\bar{x} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N x_t. \quad (\text{D.2})$$

Dans le cas d'une application unidimensionnelle ( $x_{t+1} = f(x_t)$ ) ergodique<sup>1</sup> de densité invariante  $\rho(x)$ , on peut calculer la fonction d'autocorrélation d'une façon alternative :

$$C(\tau) = \int \rho(x) x f^\tau(x) dx - \langle x \rangle^2, \quad (\text{D.3})$$

où  $\langle x \rangle$  est la moyenne d'ensemble, i.e.

$$\langle x \rangle = \int \rho(x) x dx. \quad (\text{D.4})$$

Dans cette annexe, nous allons calculer de façon analytique la fonction d'autocorrélation pour des nombres aléatoires de distribution uniforme ( $C_R$ ), pour le

---

<sup>1</sup>L'ergodicité permet de substituer les moyennes temporelles par des moyennes d'ensemble.

déplacement de Bernoulli ( $C_B$ ) et enfin pour l'application de la tente ( $C_T$ ), i.e. les trois segments de la dynamique segmentée. Dans les trois cas,  $x_t$  est restreint à l'intervalle  $[0, 1]$ , et la densité invariante<sup>2</sup> est uniforme sur cet intervalle, i.e.  $\rho(x) = 1$  [BS93]. Par conséquent, la moyenne temporelle et  $C(0)$  sont communs aux trois «dynamiques» :  $\bar{x} = \langle x \rangle = \frac{1}{2}$  et

$$C(0) = \overline{x^2} - \bar{x}^2 = \langle x^2 \rangle - \langle x \rangle^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}. \quad (\text{D.5})$$

Pour des nombres aléatoires, il n'y a aucune corrélation entre les valeurs passées et les valeurs futures :

$$C_R(\tau > 0) = \int_0^1 \int_0^1 xx' dx dx' - \langle x \rangle^2 = \langle x \rangle^2 - \langle x \rangle^2 = 0. \quad (\text{D.6})$$

À l'annexe C, il a été mentionné que le déplacement de Bernoulli ( $x_{t+1} = 2x_t \bmod 1$ ) revient à décaler les bits vers la gauche (multiplication par 2) et à abandonner celui qui passe à gauche de la virgule (modulo 1). Pour  $\tau$  itérations, il y a  $2^\tau$  parties entières différentes ( $2^\tau$  intervalles de longueur  $2^{-\tau}$ ) qui peuvent être retranchées à la suite des  $\tau$  décalages vers la gauche. Si l'on utilise la base binaire, on peut donc écrire directement :

$$C_B(\tau) = \sum_{b_1, \dots, b_\tau=0}^1 \int_{0, b_1 \dots b_\tau}^{0, b_1 \dots b_\tau + 2^{-\tau}} x(2^\tau x - b_1 \dots b_\tau) dx - \frac{1}{4}, \quad (\text{D.7})$$

ce qui est équivalent à :

$$C_B(\tau) = \sum_{k=0}^{2^\tau-1} \int_{k2^{-\tau}}^{(k+1)2^{-\tau}} x(2^\tau x - k) dx - \frac{1}{4}. \quad (\text{D.8})$$

En performant l'intégration et en réarrangeant quelque peu les termes, on obtient :

$$C_B(\tau) = \frac{1}{3 \cdot 2^{2\tau+1}} \sum_{k=0}^{2^\tau-1} [(2-k)(k+1)^2 + k^3] - \frac{1}{4}. \quad (\text{D.9})$$

Enfin, en se rappelant la relation suivante

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}, \quad (\text{D.10})$$

---

<sup>2</sup>Pour les nombres aléatoires, il serait plus juste de parler de densité de probabilité.

on obtient le résultat cherché :

$$C_B(\tau) = \frac{1}{3 \cdot 2^{\tau+2}}. \quad (\text{D.11})$$

Le calcul pour l'application de la tente s'avère un peu plus délicat. Rappelons cette application :

$$x_{t+1} = \begin{cases} f_0(x_t) = 2x_t & \text{si } x_t \in I_0 := [0, \frac{1}{2}] \\ f_1(x_t) = 2(1 - x_t) & \text{si } x_t \in I_1 := ]\frac{1}{2}, 1]. \end{cases} \quad (\text{D.12})$$

En base binaire,  $1 - x$  revient à inverser tous les bits si  $x \in [0, 1]$ . Pour se convaincre, il faut d'abord se rappeler que le nombre 1 peut être représenté de deux façons équivalentes,

$$1 = 2^0 \equiv \sum_{i=1}^{\infty} 2^{-i} = \frac{\frac{1}{2}}{1 - \frac{1}{2}} \text{ (série géométrique) } = 1, \quad (\text{D.13})$$

et que  $1 - b_i$  inverse le bit  $b_i$ . Comme l'inversion des bits d'un nombre compris dans l'intervalle  $I_1$  produit un nombre compris dans l'intervalle  $I_0$ , l'application de la tente se résume à un déplacement de Bernoulli avec inversion des bits si  $x \in I_1$ .

S'il se produit un nombre pair d'inversions lors de l'itération répétée de l'application de la tente, alors l'effet résultant se confond avec l'utilisation du déplacement de Bernoulli. S'il se produit un nombre impair d'inversions, la seule différence consiste en l'inversion des bits du résultat final. Quelle est la condition, exprimée en fonction des bits de la condition initiale  $x_0 = 0, b_1 \cdots b_\tau \cdots$ , pour obtenir un nombre impair d'inversions à la suite de  $\tau$  itérations? Étonnamment, cette condition est toute simple :  $b_\tau = 1$ .

Pour démontrer cette condition, on peut introduire une fonction d'inversion, en l'occurrence  $g_\tau(b_1, \dots, b_\tau)$ , qui sera nulle s'il n'y a pas d'inversion résultante après  $\tau$  itérations et non nulle autrement. On a de façon triviale :

$$g_1(b_1) = b_1 \quad (\text{D.14})$$

et

$$g_2(b_1, b_2) = b_1 b_2 + [1 - b_1] b_2 = b_2. \quad (\text{D.15})$$

On se convainc aisément de la relation de récurrence suivante :

$$g_\tau(b_1, \dots, b_\tau) = b_1 [1 - g_{\tau-1}(1 - b_2, \dots, 1 - b_\tau)] + [1 - b_1] g_{\tau-1}(b_2, \dots, b_\tau). \quad (\text{D.16})$$

À partir de celle-ci, on trouve directement  $g_3(b_1, b_2, b_3) = b_3$ , et en poursuivant de la même manière, on obtient le cas général  $g_\tau(b_1, \dots, b_\tau) = b_\tau$ , d'où la condition énoncée plus haut.

S'il y a une inversion résultante après  $\tau$  itérations, on a  $x_\tau = 1 - (2^\tau x_0 - b_1 \cdots b_\tau)$ . Si l'on pose  $k = b_1 \cdots b_\tau$ , la condition  $b_\tau = 1$  est équivalente à  $k$  impair, et on peut alors écrire :

$$C_T(\tau) = \sum_{k=0}^{2^\tau-1} (-1)^k \int_{k2^{-\tau}}^{(k+1)2^{-\tau}} x \left( 2^\tau x - 2 \left\lfloor \frac{k+1}{2} \right\rfloor \right) dx - \frac{1}{4}, \quad (\text{D.17})$$

où  $\lfloor x \rfloor$  désigne la partie entière de  $x$ . En performant l'intégration et en réarrangeant quelque peu les termes, on obtient d'abord :

$$\begin{aligned} C_T(\tau) &= \frac{1}{3 \cdot 2^{2\tau}} \left[ \sum_{k=0}^{2^\tau-1} (-1)^k (3k^2 + 3k + 1) \right. \\ &\quad \left. + \frac{3}{2} \sum_{k \text{ impair}} (k+1)(2k+1) - \frac{3}{2} \sum_{k \text{ pair}} k(2k+1) \right] - \frac{1}{4}. \end{aligned} \quad (\text{D.18})$$

Pour  $\tau > 0$ , on a :

$$C_T(\tau > 0) = \frac{1}{3 \cdot 2^{2\tau}} \left[ 3 \cdot 2^{\tau-1} + \sum_{k=1}^{2^\tau-2} (-1)^k (3k^2 + 3k + 1) - 6 \sum_{k=1}^{2^{\tau-1}-1} k \right] - \frac{1}{4}. \quad (\text{D.19})$$

Le résultat final est donc :

$$C_T(\tau) = C_R(\tau) = \frac{1}{4} \delta_{\tau 0}. \quad (\text{D.20})$$

La Fig. D.1 présente la vérification numérique des résultats analytiques de cette annexe.

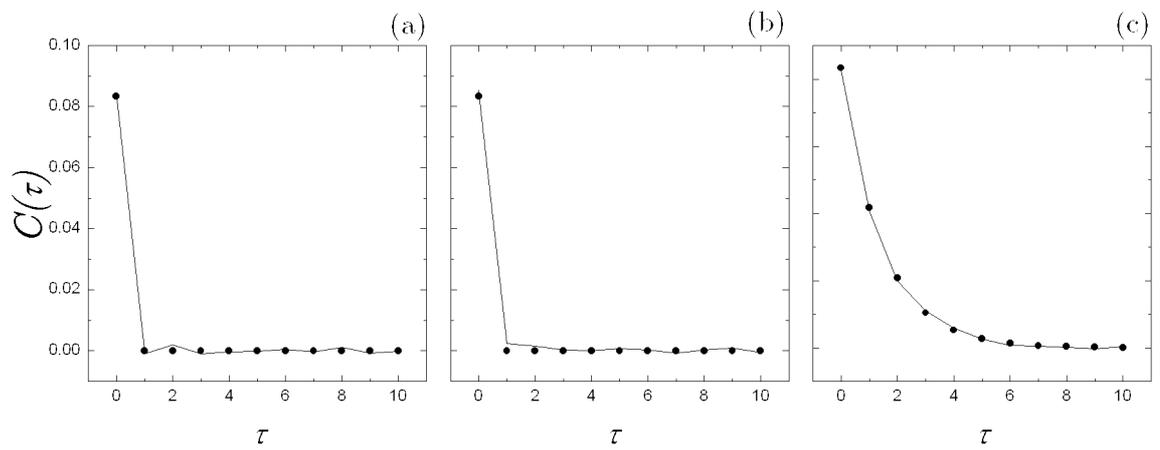


FIG. D.1 – Fonctions d'autocorrélation numériques (-) et théoriques (●). (a) Nombres aléatoires de distribution uniforme. (b) Application de la tente. (c) Déplacement de Bernoulli.  $N=16\ 384$  pour chaque segment.

# Annexe E

## Contrôle du chaos en 1D

Ce n'est que récemment [OGY90] que l'idée d'exploiter la richesse d'un système chaotique a réellement pris son envol, l'idée de contrôler le chaos. Cette richesse se traduit par l'existence d'une infinité d'orbites périodiques instables contenues de façon dense à l'intérieur d'un attracteur chaotique. Autant d'orbites périodiques signifie autant de régimes disponibles pour un système, d'où l'extraordinaire versatilité d'un système chaotique. Le principe de contrôle se résume en fait à stabiliser une orbite périodique instable, e.g. celle qui conduit à une performance accrue, au moyen de perturbations fines d'un paramètre  $r$  du système. La démarche de contrôle est donc essentiellement la suivante :

1. Pour une valeur nominale du paramètre de contrôle  $r = r_0$ , pour laquelle le système est chaotique, rechercher des orbites périodiques instables et sélectionner celle qui entraîne le comportement désiré.
2. Toujours pour  $r = r_0$ , démarrer le système. En raison du caractère ergodique de l'attracteur chaotique, après un certain temps transitoire, l'orbite du système entre dans le voisinage d'un des membres de l'orbite périodique instable sélectionnée.
3. Actionner alors le mécanisme de contrôle afin de forcer la trajectoire du système

à suivre l'orbite périodique instable. Pour ce faire, à chaque itération, on applique une perturbation fine adéquate  $\delta r$  sur  $r_0$ . Les perturbations ne sont pas cumulatives.

Dans cette annexe, seul le contrôle pour une application 1D sera considéré, et l'application logistique servira d'exemple :

$$x_{t+1} = rx_t(1 - x_t). \quad (\text{E.1})$$

La technique de contrôle en 1D s'avère fort simple. Soit une application 1D à un paramètre telle l'application logistique :

$$x_{t+1} = f(r; x_t). \quad (\text{E.2})$$

Par hypothèse, cette application est dissipative et présente un attracteur chaotique pour un segment de l'axe  $r$  contenant  $r_0$ . Cet attracteur renferme une pléiade d'orbites périodiques instables, dont notamment celle-ci de période  $p$  :  $x_i^* = f^p(r_0; x_i^*)$  où  $i = 1, \dots, p$  et  $x_{i+1 \bmod p}^* = f(r_0; x_i^*)$ .<sup>1</sup> Dans un voisinage réduit de  $x_i^*$ , i.e.  $|x_t - x_i^*| < \epsilon$  où  $\epsilon$  est petit, et lorsque l'on applique une petite perturbation du paramètre de contrôle  $|\delta r_t| \ll |r_0|$ , on a par linéarisation :

$$x_{t+1} - x_{i+1}^* \approx \frac{\partial f(r_0; x_i^*)}{\partial x}(x_t - x_i^*) + \frac{\partial f(r_0; x_i^*)}{\partial r} \delta r_t, \quad (\text{E.3})$$

où  $\delta r_t = r_t - r_0$  (perturbation appliquée pour déterminer  $x_{t+1}$ ). On cherche  $\delta r_t$  de façon à stabiliser l'orbite périodique, i.e.  $x_{t+1} - x_{i+1}^* = 0$ . Insérant cette condition dans l'équation (E.3), on trouve la perturbation nécessaire :

$$\delta r_t = \frac{\frac{\partial f(r_0; x_i^*)}{\partial x}(x_t - x_i^*)}{-\frac{\partial f(r_0; x_i^*)}{\partial r}}. \quad (\text{E.4})$$

En regard des hypothèses de travail données auparavant, cette perturbation (E.4) doit être fine afin d'être pleinement efficace car sinon, les termes  $\mathcal{O}(\delta r_t^2)$  devraient

<sup>1</sup>Afin d'alléger la notation, nous omettrons par la suite d'indiquer  $\bmod p$ .

être considérés dans l'équation (E.3) avant d'y calculer  $\delta r_t$ . Or si cela était nécessaire, on imagine sans peine que le calcul de la perturbation stabilisatrice pourrait devenir fort laborieux, à moins que les dérivées d'ordre supérieur impliquant  $r$  soient toutes nulles.

Il y a une seconde raison à limiter la grandeur de la perturbation du paramètre de contrôle. En effet, on peut envisager qu'une perturbation excessive modifie profondément le régime du système, e.g. disparition de l'attracteur chaotique et de l'orbite périodique instable, et qu'au fil des itérations successives, le mécanisme de contrôle (formule linéaire (E.4)) n'arrive plus à stabiliser au moyen de petites perturbations. Le système peut alors s'emballer, et demeurer bien loin du régime performant envisagé à  $r \simeq r_0$ . Nous donnerons enfin une troisième raison, soit la philosophie de la technique de contrôle OGY [OGY90] : composer avec le système dans son régime chaotique et ne chercher qu'à le perturber très légèrement pour obtenir le comportement désiré. Au niveau algorithmique, on impose donc une borne  $|\delta r| < \Delta$  pour l'application de la perturbation. Si cette dernière excède en valeur absolue la borne, on a alors le choix de choisir  $|\delta r| \in [0, \Delta]$ .

Dans le cas spécifique de l'application logistique, la formule linéaire de contrôle (E.4) prend la forme suivante :

$$\delta r_t = \frac{r_0(2x_i^* - 1)(x_t - x_i^*)}{x_i^*(x_i^* - 1)}. \quad (\text{E.5})$$

Pour l'application logistique, il serait possible d'écrire aisément une formule de contrôle en considérant les termes d'ordre supérieur dans le développement de Taylor. La raison est que  $\partial^k f / \partial r^k = 0 \forall k > 1$ , et seul  $\delta r_t$  apparaît (pas d'autres puissances).