

Comparaison de systèmes et arbres de régression

Systems Comparison and Regression Trees

par Anne-Emmanuelle BADEL* Olivier MICHEL* et Alfred HERO**

* Laboratoire de Physique (URA 1325 CNRS),

École Normale Supérieure de Lyon,
46 allée d'Italie, 69364 Lyon Cedex 07, France,

** Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48109-2122, USA

résumé et mots clés

Nous proposons ici une application de la méthode de modélisation non linéaire non paramétrique de systèmes dynamiques, présentée dans un précédent article. L'approche proposée dans le cadre de ce travail repose sur une partition récursive de l'espace d'état du système, conduisant à un arbre de régression. Ce modèle fournit une estimation de l'histogramme d -dimensionnel de l'espace des états du système : nous montrons comment l'utilisation de distances ou de divergences entre lois de probabilité permet alors de quantifier les différences dynamiques entre systèmes. Cette approche est illustrée sur deux exemples : la détection de changements de modèles autorégressifs dans une série temporelle et la détection de la présence éventuelle d'un soliton de type « breather » susceptible d'apparaître dans le comportement d'une chaîne d'oscillateurs couplés soumis à un potentiel extérieur.

Distances, f -divergences, statistique, comparaison, modélisation, non linéaire, non paramétrique, systèmes dynamiques, arbres de régression, espace des états.

abstract and key words

A method for non parametric modelling of dynamical systems was presented in a previous paper. This work intends to propose a new approach for addressing the problem of dynamical systems comparison and detection of abrupt changes. The algorithm that is presented here, relies upon both d -dimensional histogram and regression tree estimation, and the use of f -divergences. Illustrations on different non linear systems are provided.

Distances, f -divergences, statistics, comparison, modelling, non linear, non parametric, dynamical systems, regression Trees, state space.

1. introduction

La comparaison de la dynamique de deux systèmes, éventuellement non linéaires, constitue une part importante des études qui leur sont consacrées. Cela nécessite d'une part de disposer de modèles décrivant correctement la dynamique des systèmes et d'autre part de définir une mesure quantifiant les différences entre les estimations des modèles associés à ces systèmes. Dans cet article, nous proposons d'utiliser la modélisation non linéaire non paramétrique à l'aide des arbres de régression proposée dans [14, 2] pour comparer des systèmes dynamiques. Cette approche est basée sur la représentation de la dynamique d'un système

dans son espace des états réel ou reconstruit. Soient \underline{X}_n et \underline{Y}_n respectivement le vecteur d'état de ce dernier observé à la date¹ n . L'équation dynamique d'évolution du système s'écrit de façon générale :

$$\begin{cases} \underline{X}_{n+1} = f(\underline{X}_n + \underline{\xi}_n) \\ \underline{Y}_n = g(\underline{X}_n) + \underline{\varepsilon}_n \end{cases} \quad (1)$$

où $\underline{\xi}_n$ est un bruit d'état et $\underline{\varepsilon}_n$ un bruit d'observation. $f : \mathbb{R}^p \rightarrow \mathbb{R}^p$ décrit l'évolution dynamique du système tandis que $g : \mathbb{R}^p \rightarrow \mathbb{R}^q$ traduit l'observation de l'état. Dans la suite, p sera supposé faible et q sera le plus souvent égal à 1 : l'espace

1. Tous les systèmes et les séries temporelles étudiés ici seront considérés comme discrets que ce caractère soit une propriété intrinsèque ou la conséquence d'un échantillonnage.

des états considéré sera dans ce cas obtenu par la méthode de plongement [2, 25, 17, 22].

Le modèle fourni par les arbres de régression [14, 2] permet de décrire à l'aide d'un histogramme d -dimensionnel pour un espace des états de dimension d la dynamique d'un système dans l'hypothèse où il est observé en régime stationnaire au sens où le système reste localisé dans une partie bornée de son espace des états (et où les régimes transitoires sont exclus). Cet histogramme est construit par une partition récursive de l'espace des états au sein desquelles l'ensemble des réalisations du vecteur d'état ne présente plus de structure dynamique propre (tout au moins à la précision à laquelle il est observé). La mesure des différences de comportement dynamique est alors obtenue à l'aide de divergences entre lois de probabilité [4, 16, 8].

La suite de cet article est organisée de la façon suivante : l'ensemble des notions et outils utilisés dans le cadre de cette étude est rappelé dans la section 2. Ainsi la définition des f -divergences de I. Csiszár [4, 16, 8] et certaines de leurs propriétés sont développées, tout comme est rappelée la méthode d'estimation des fonctions de distribution à partir des arbres de régression.

Les deux parties suivantes concernent les résultats d'une telle comparaison sur deux exemples : la détection de « ruptures » (ou changements brusques) de modèles autorégressifs dans une série temporelle et la détection de solitons de type « breather » susceptibles d'apparaître dans le comportement d'une chaîne d'oscillateurs couplés soumis à un potentiel extérieur.

2. divergences entre systèmes modélisés par les arbres de régression

Le travail présenté dans cet article repose sur une hypothèse importante. Elle suppose qu'il est possible d'observer le système dans son espace des états d -dimensionnel (les problèmes liés à l'estimation et éventuellement à la reconstruction de cet espace par la méthode des retards seront abordés plus loin). Cela restreint évidemment le cadre de cette étude aux systèmes possédant un faible nombre de degrés de liberté (on notera désormais d_0 ce nombre). Chaque observation, qui correspond à un point dans l'espace des états, sera considérée comme une réalisation d'un processus stochastique à d ($d > d_0$) dimensions. La comparaison de deux systèmes consiste donc ici à comparer les « trajectoires » formées des différentes positions des observations au sein de l'espace des états et donc à comparer les fonctions de distributions des réalisations du vecteur d'état.

2.1. comparaison de distributions de probabilité

De nombreuses mesures ont été proposées pour quantifier les écarts entre lois de probabilité [4, 16]. L'approche retenue est basée sur la fonction de vraisemblance entre deux lois de probabilité. Soit un processus de densité de probabilité p qu'on souhaite évaluer à l'aide d'un processus de densité de probabilité q . Le rapport de vraisemblance est alors défini par :

$$V \triangleq \frac{q}{p} \quad (2)$$

Le fait que la fonction logarithme soit strictement croissante permet d'utiliser plus généralement la log-vraisemblance :

$$\ln V \triangleq \ln \left(\frac{q}{p} \right) \quad (3)$$

L'espérance mathématique conditionnée par p : E_p conduit à définir la divergence de Kullback-Leibler [12] entre les deux lois :

$$KL(P, Q) = -E_p [\ln V] = -\int_{\mathcal{H}} p \ln \frac{q}{p} d\mu \quad (4)$$

où μ est la mesure définie sur l'espace de probabilités \mathcal{H} sur lequel sont définies les deux lois P et Q .

Cette divergence appartient à la classe générale des f -divergences introduites par I. Csiszár [8] ou S.M. Ali et S.D. Silvey [1]. Soit un couple de fonctions (f, g) respectivement fonction réelle convexe continue sur \mathbb{R}_+ et fonction réelle croissante sur \mathbb{R} , on définit :

$$\begin{aligned} d(P, Q) &= g \left(E_p \left[f \left(\frac{q}{p} \right) \right] \right) \\ &= g \left(\int_{\mathcal{H}} p f \left[\frac{q}{p} \right] d\mu \right) \end{aligned} \quad (5)$$

Si $f(x) = -\ln x$ et $g(x) = x$, $d(P, Q)$ est la divergence de Kullback-Leibler. De nombreuses quantités, correspondant à la multitude de choix possibles pour les fonctions f et g (cf. [4, 16] et tableau 1 pour quelques exemples), mesurent l'écart entre les deux lois de probabilité en évaluant la dispersion (relativement à l'une des lois) du rapport de vraisemblance entre ces deux lois.

Nous pouvons noter les cas particuliers suivants :

- distance variationnelle de Kolmogorov ou distance de Matusita pour $\alpha = 1$:

$$V(P, Q) = \int_{\mathcal{H}} |q - p| d\mu \quad (6)$$

- divergence d'Hellinger ou distance de Matusita pour $\alpha = 2$:

$$H(P, Q) = \sqrt{\int_{\mathcal{H}} (\sqrt{q} - \sqrt{p})^2 d\mu} \quad (7)$$

- distance de Bhattacharyya ou distance de Chernoff pour $\alpha = \frac{1}{2}$:

$$B(P, Q) = -\ln \int_{\mathcal{H}} \sqrt{pq} d\mu \quad (8)$$

Tableau 1. – Différentes divergences de la classe des f -divergences de Csiszár en fonction du choix des fonctions f et g .

Divergence	fonction f	fonction g	Expression
Matusita	$f(x) = 1 - x^{\frac{1}{\alpha}} ^{\alpha}$	$g(x) = x^{\frac{1}{\alpha}}$	$M_{\alpha}(P, Q) = \int_{\mathcal{H}} \left q^{\frac{1}{\alpha}} - p^{\frac{1}{\alpha}} \right ^{\alpha} d\mu$
Kullback-Leibler	$f(x) = -\ln x$	$g(x) = x$	$KL(P, Q) = \int_{\mathcal{H}} p \ln \frac{p}{q} d\mu$
Chernoff	$f(x) = x^{1-\alpha}$	$g(x) = -\ln(x)$	$C_{\alpha}(P, Q) = -\ln \int_{\mathcal{H}} p^{\alpha} q^{1-\alpha} d\mu$
χ^2	$f(x) = (x - 1)^2$	$g(x) = x$	$\chi^2(P, Q) = \int_{\mathcal{H}} \frac{(p - q)^2}{p} d\mu$

Il est difficile (voir [4]) de donner des recommandations précises à propos du choix d'une divergence ou d'une autre. D'autre part, il a été montré, dans un contexte paramétrique, que les f -divergences sont toutes localement équivalentes; la métrique quadratique différentielle qui peut leur être associée coïncide en effet avec la métrique définie par la matrice d'information de Fisher [5, 18]. Dans le cadre général non paramétrique, les divergences de Kullback-Leibler et de Hellinger sont les plus documentées et les plus utilisées. Il a été montré qu'elles permettent de résoudre des problèmes appliqués (détection, classification, codage, etc [4]).

Dans cet article, nous avons le plus souvent utilisé la divergence de Rényi [19, 20] égale, à un coefficient multiplicatif près, à la distance de Chernoff. Ce coefficient a été choisi pour obtenir la divergence de Kullback-Leibler comme limite de la divergence de Rényi. Historiquement le coefficient a été choisi égal à $\frac{1}{\alpha - 1}$ soit :

$$R_{\alpha}(P, Q) = \frac{1}{\alpha - 1} \ln \int_{\mathcal{H}} p^{\alpha} q^{1-\alpha} d\mu \quad (9)$$

dont la limite quand α tend vers 1 est la divergence de Kullback-Leibler [3] :

$$\lim_{\alpha \rightarrow 1} R_{\alpha}(P, Q) = KL(P, Q) \quad (10)$$

Il arrive parfois que ce coefficient multiplicatif soit choisi égal à : $\frac{1}{\alpha(\alpha - 1)}$. Cela permet de retrouver $KL(Q, P)$ et $KL(P, Q)$ quand α tend respectivement vers 1 et 0. Dans la suite, seule la première définition est retenue. Ainsi, si la distribution Q est uniforme, il y a coïncidence des valeurs prises par la divergence de Rényi $R(P, Q)$ et l'information de Rényi associée à la distribution P [19, 20]. Un raisonnement identique montre que, si Q est uniforme, la divergence $KL(P, Q)$ est égale à l'entropie de Shannon (ou de Boltzmann-Gibbs) de la distribution P .

Enfin, nous avons volontairement restreint notre étude au choix de $0 < \alpha < 1$. Cette restriction garantit l'existence de la divergence de Rényi quand l'une ou l'autre des distributions passe par zéro (le choix de α en dehors de cet intervalle conduit alors à une divergence numérique).

Finalement les seules divergences retenues dans ce travail sont les suivantes :

- la distance variationnelle de Kolmogorov, qui s'exprime comme une distance en norme L^1 des fonctions de densité p et q :

$$V(P, Q) = \int_{\mathcal{H}} |q - p| d\mu \quad (11)$$

- la divergence de Hellinger, très souvent utilisée, qui fournit une base de comparaison :

$$H(P, Q) = \sqrt{\int_{\mathcal{H}} (\sqrt{q} - \sqrt{p})^2 d\mu} \quad (12)$$

- la divergence de Kullback-Leibler, dérivée de la fonctionnelle d'entropie de Shannon :

$$KL(P, Q) = \int_{\mathcal{H}} p \ln \frac{p}{q} d\mu \quad (13)$$

- la divergence de Rényi, dérivée de la fonctionnelle d'entropie de Rényi :

$$R_{\alpha}(P, Q) = \frac{1}{\alpha - 1} \ln \int_{\mathcal{H}} p^{\alpha} q^{1-\alpha} d\mu \quad (14)$$

Les entropies de Shannon et de Rényi sont les deux seules formes d'entropie additive (voir [19]). L'interprétation des résultats de divergence ou d'information, dans le cadre de la physique, est en effet beaucoup plus délicate si l'additivité n'est pas vérifiée (voir par exemple [11, 24] pour l'introduction et l'usage d'entropie non additive).

2.2. mise en œuvre

2.2.1. estimation des lois de probabilité

On ne dispose couramment que d'une série temporelle expérimentale notée x_n . L'espace des états du système peut alors être reconstruit par la méthode de plongement ou méthode des retards².

2. Voir [2] pour un rappel plus détaillé de la méthode.

Cette méthode proposée par H. Whitney [25], reprise par N.H. Packard, J.P. Crutchfield, J.D. Farmer et R. Shaw [17] et justifiée par F. Takens [22] pour les séries infinies chaotiques consiste à construire les vecteurs :

$$\underline{\mathbf{X}}_n = [x_n \ x_{n-\tau} \ \dots \ x_{n-(d-1)\tau}]^T \quad (15)$$

où d et τ sont respectivement la dimension et le retard de reconstruction. F. Takens [22] a prouvé l'équivalence (à un difféomorphisme près³) de l'espace constitué de l'ensemble des vecteurs $\underline{\mathbf{X}}_n$ avec l'espace des états réel, pour des séries infinies avec la condition suffisante $d \geq 2D_2 + 1$ (D_2 est la dimension de corrélation [9]). Dans le cas des séries finies, le choix du retard (qui est indifférent dans le cas des séries infinies) devient critique [13] : généralement τ coïncide avec le premier zéro de la fonction d'autocorrélation ou le premier minimum de l'information mutuelle. Une discussion plus poussée sur le choix des paramètres de reconstruction d et τ est présentée dans [2].

Un algorithme récursif de partition de l'espace des états éventuellement reconstruit a été proposé dans [14, 2, 15]. Il permet d'estimer l'histogramme d -dimensionnel (pour un espace de dimension d) associé à la fonction de densité de probabilité des réalisations du vecteur d'état. Les cellules de l'histogramme sont définies récursivement, jusqu'à ce que la distribution des observations soit uniforme (estimée telle au sens de la loi de χ^2) au sein de ces cellules, ou que ces dernières soient pratiquement vides. On montre que cette estimation de l'histogramme, en « bins » de taille variable, dirigée par les données, conduit à une description plus robuste et plus concise que les approches classiques recourant à un maillage uniforme de l'espace des états.

On notera dans la suite Π l'ensemble des cellules de la partition finale.

2.2.2. comparaison entre lois de probabilité

La méthode proposée conduit à une partition dépendant directement des données : l'ensemble des cellules, s'il est ainsi plus restreint, diffère suivant les données utilisées pour son élaboration. Nous proposons alors d'utiliser les données (vecteurs d'état d -dimensionnels) issues du système S_2 , par exemple, pour estimer la probabilité : $p_k^{(S_2|S_1)}$, probabilité qu'une réalisation du vecteur d'état du système S_2 se trouve dans la k -ième cellule de la partition induite par les données du système S_1 . La vraisemblance de cette estimation s'exprime alors par :

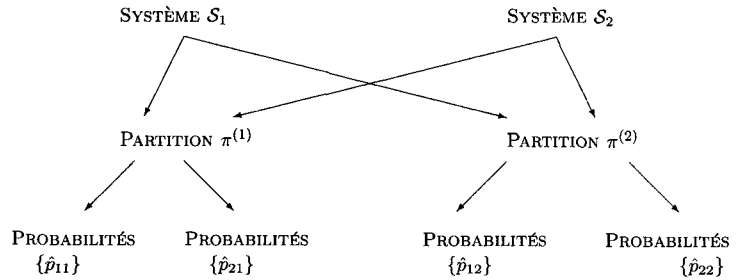
$$\frac{p_k^{(S_2|S_1)}}{p_k^{(S_1|S_1)}} \quad (16)$$

où k indice la cellule considérée. Dans la suite de cet article, on note :

$$\{ \hat{p}_{ij} \} = \left\{ p_k^{(S_i | S_j)} \right\} \quad (17)$$

3. Une fonction f est un homéomorphisme si elle est bijective et continue ainsi que son inverse f^{-1} ; si de plus elle est différentiable, il s'agit d'un difféomorphisme.

où $(i, j) \in \{1, 2\}^2$. Le schéma suivant indique l'ensemble des quatre distributions de probabilité qui peuvent ainsi être estimées :



Les systèmes sont alors comparés à l'aide des f -divergences introduites précédemment, appliquées non plus aux fonctions de densité de probabilité mais aux distributions de probabilité estimées :

$$g \left(E_{\hat{p}_{jj}} \left[f \left(\frac{\hat{p}_{ij}}{\hat{p}_{jj}} \right) \right] \right) \quad (18)$$

L'objectif poursuivi est de comparer les dynamiques des systèmes dont sont issues les données observées; l'approche proposée doit par conséquent être indépendante :

- de l'amplitude des signaux,
- de leur « phase initiale ».

Deux sinusoides de même fréquence⁴ doivent être confondues dans l'approche proposée, indépendamment de leur amplitude respective. Si les systèmes sont identiques mais asynchrones, l'algorithme doit là encore les confondre. Pour répondre à ces contraintes, les observations dans l'espace d -dimensionnel sont centrées, normalisées et exprimées dans leur base de vecteurs propres. Ainsi, l'approche proposée est invariante par translation, homothétie et rotation comme l'illustrent les figures 1(a) à 1(d).

La mise en œuvre pratique de la comparaison de deux systèmes consiste à choisir un des deux systèmes comme référence au sens où la partition de l'espace des états sera construite à partir des données de ce système. Les histogrammes associés à chacun des deux systèmes sont ensuite estimés sur cette unique partition de l'espace. La comparaison des deux systèmes est alors mesurée en évaluant l'une des divergences précédemment introduites entre les deux distributions de probabilité ainsi obtenues. L'inconvénient de cette méthode est de privilégier le système de référence : seules les réalisations de ce système conditionnent la construction de la partition. Pour s'affranchir de ce problème, la méthode adoptée consiste à construire les partitions relatives aux deux systèmes, à évaluer les deux divergences calculables sur ces deux partitions et à « symétriser » en prenant la moyenne des deux valeurs obtenues (éventuellement après avoir comparé les résultats obtenus séparément avec l'une ou l'autre des partitions).

4. τ est supposé être identique pour les deux systèmes.

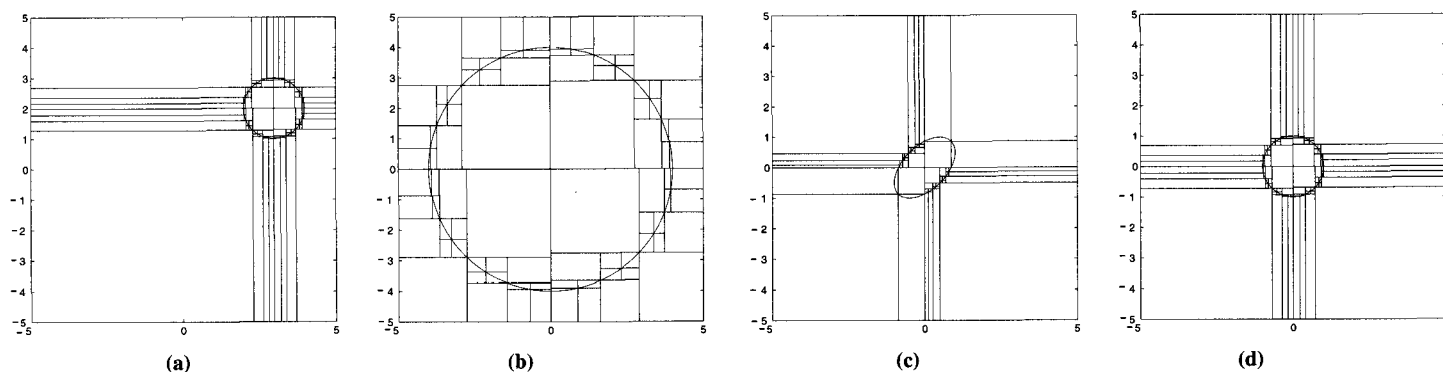


Figure 1. – Espaces des états constitué de deux sinusoides et partitions associées :

(a) les données ne sont pas centrées,

(b) les données ne sont pas normalisées,

(c) les données ne sont pas ré-exprimées dans leur base propre,

(d) les données sont centrées, normalisées et exprimées dans leur base propre.

Quelles que soient les données initiales, les influences de la moyenne (a), de l'amplitude (b) et de la phase (c) sont supprimées pour conduire à l'espace (d).

Par construction, certaines divergences (par exemple, les divergences de Kulback-Leibler ou de Rényi) ne sont pas des expressions symétriques [3] par rapport aux deux distributions de probabilité; cela conduit à « privilégier » l'une des deux distributions, qui apparaît, de fait, comme une distribution de référence. Dans le cadre de problèmes pour lesquels une distribution de référence s'impose, cela n'est pas gênant. Dans un contexte plus général, il peut sembler souhaitable de disposer d'une mesure de dissimilarité entre distributions qui ne privilégie aucune des distributions. Il est aisé de symétriser l'expression donnée par l'équation (5) vis-à-vis de $P = p_1$ et $Q = p_2$ en prenant la moyenne arithmétique des divergences obtenue en considérant successivement p_1 et p_2 comme distribution de référence :

$$d_s(p_1, p_2) = \frac{1}{2} \left[g \left(E_{p_1} \left[f \left(\frac{p_2}{p_1} \right) \right] \right) + g \left(E_{p_2} \left[f \left(\frac{p_1}{p_2} \right) \right] \right) \right] \quad (19)$$

Cette mesure ne s'exprime pas sous la forme d'une f -divergence; cependant il est facile de vérifier que les principales propriétés énoncées dans [4] sont conservées. Il existe, dans la méthode proposée, une seconde source de dissymétrie entre les rôles joués par p_1 et p_2 . En effet, la partition utilisée pour estimer ces distributions est une partition elle-même estimée à l'aide d'un ensemble d'observations (séries temporelles d'apprentissage) de distribution de probabilité p_2 ou p_1 . A nouveau, la symétrie de la divergence est restaurée en calculant la moyenne des divergences (19) estimées sur chacune des partitions soit, à l'aide des notations précédemment introduites :

$$d(\mathcal{S}_1, \mathcal{S}_2) = \frac{1}{2} [d_s(p_{11}, p_{21}) + d_s(p_{12}, p_{22})] \quad (20)$$

3. détection de changements de modèles

3.1. présentation du problème

Dans ce paragraphe, l'intérêt des outils présentés dans la section précédente est illustré dans le cadre du problème de la détection d'une rupture de modèle. L'exemple envisagé ici est constitué d'une suite de q modèles autorégressifs d'ordre maximal p consécutifs :

$$x_n = \begin{cases} \sum_{i=1}^p a_i^{(0)} x_{n-i} + \epsilon_n & \text{si } n < N_0 \\ \sum_{i=1}^p a_i^{(1)} x_{n-i} + \epsilon_n & \text{si } N_0 < n < N_1 \\ \dots \\ \sum_{i=1}^p a_i^{(q)} x_{n-i} + \epsilon_n & \text{si } n > N_{q-1} \end{cases} \quad (21)$$

L'ensemble des informations utilisées pour les simulations numériques, présentées au paragraphe suivant, sont rassemblées dans le tableau 2. Les p coefficients du j -ième modèle autorégressif $\{a_i^{(j)}\}_{i=1, \dots, p}$ y sont portés en première ligne. Sur la seconde ligne sont portées les valeurs complexes des pôles du modèle considéré. t_{j_1, j_2} et d_{j_1, j_2} sont respectivement les dates auxquelles se produisent les ruptures (ou passages du modèle j_1 au modèle j_2) et la distance cepstrale entre ces derniers. En remarquant que la variance du bruit d'entrée ϵ_n est identique pour tous les modèles considérés dans cette simulation, et que ces derniers sont tous stables, causaux et à minimum de phase (voir [4, 10]), la distance

Tableau 2. – Description des séries temporelles formées d'une suite de q modèles autorégressifs d'ordre maximal p consécutifs.

Série	$a_i^{(0)}$ $p_i^{(0)}$	$t_{0,1}$ $d_{0,1}$	$a_i^{(1)}$ $p_i^{(1)}$	$t_{1,2}$ $d_{1,2}$	$a_i^{(2)}$ $p_i^{(2)}$	$t_{2,3}$ $d_{2,3}$	$a_i^{(3)}$ $p_i^{(3)}$
x	$-0, 2; 0, 1$ $-0, 43; 0, 23$	3000 1,96	$0, 8; -0, 9$ $0, 95e^{\pm 1,14j}$				
y	$0, 8; -0, 9$ $0, 95e^{\pm 1,14j}$	3000 1,97	$-0, 4; 0, 3; -0, 2$ $0, 61e^{\pm 1,46j}; -0, 54$				
z	$0, 8; -0, 9$ $0, 95e^{\pm 1,14j}$	3000 1,38	$0, 3; -0, 1$ $0, 32e^{\pm 1,08j}$	6000 1,36	$0, 7; -0, 9$ $0, 95e^{\pm 1,19j}$	9000 0,94	$0, 5; -0, 4$ $0, 63e^{\pm 1,16j}$

cepstrale entre les modèles j_1 et j_2 s'exprime par :

$$d_{j_1, j_2} = 2 \sum_{k=1}^L (c_k^{j_1} - c_k^{j_2})^2 \quad (22)$$

où

$$c_k^j = \frac{-1}{k} \sum_{i=1}^p (p_i^{(j)})^k \quad (23)$$

Les résultats consignés dans le tableau 2 ont été obtenus avec $L = 10$ (voir [4] pour une discussion du choix de L). Ces distances cepstrales sont données afin de fournir une base de comparaison avec d'autres algorithmes de détection. Notons cependant que seuls la simplicité de mise en oeuvre des modèles AR et le souci d'illustrer notre approche à l'aide d'exemples simples et dans le contexte bien documenté de la détection de rupture (voir [7, 6]) justifient les simulations suivantes.

La dynamique de la série temporelle envisagée est estimée par une reconstruction de l'espace des états par la méthode des retards. Le retard est égal au pas d'échantillonnage et la dimension de reconstruction à $(p+1)$, où p désigne l'ordre maximal des modèles autorégressifs. On considère donc l'ensemble des vecteurs :

$$\underline{\mathbf{X}}_n = [x_n \ x_{n-1} \ \dots \ x_{n-p}]^T \quad (24)$$

Soit T la date à laquelle est envisagée la recherche de changement. La méthode proposée consiste à comparer deux systèmes notés $\mathcal{S}_1(T)$ et $\mathcal{S}_2(T)$ obtenus à partir des deux sous-ensembles de M réalisations consécutives définis par :

$$\begin{cases} \mathcal{S}_1(T) = \{\underline{\mathbf{X}}_n\} & n = T - M, T - M + 1, \dots, T - 1 \\ \mathcal{S}_2(T) = \{\underline{\mathbf{X}}_n\} & n = T + 1, T + 2, \dots, T + M \end{cases} \quad (25)$$

avec $M \ll N$, où N est le nombre total de points de la série temporelle. Les différences entre ces deux systèmes sont mesurées à l'aide de la divergence $d(\mathcal{S}_1(T), \mathcal{S}_2(T))$ entre leurs lois de probabilité estimées (voir paragraphe précédent). Cette analyse est répétée pour différentes valeurs de T telles que $M+1 \leq T \leq N-M$. Seules les divergences de Rényi (19) sont étudiées ici.

3.2. résultats obtenus sur quelques exemples

D'après la formule (19), la divergence de Rényi entre $\mathcal{S}_1(T)$ et $\mathcal{S}_2(T)$ est :

$$\begin{aligned} R_\alpha(T) &= R_\alpha(\mathcal{S}_1(T), \mathcal{S}_2(T)) \\ &= \frac{1}{4(\alpha - 1)} \ln \left[\frac{\sum_i p_{11,i}^\alpha p_{12,i}^{1-\alpha}}{\sum_i p_{12,i}^\alpha p_{11,i}^{1-\alpha}} \right] \left[\frac{\sum_i p_{21,i}^\alpha p_{22,i}^{1-\alpha}}{\sum_i p_{22,i}^\alpha p_{21,i}^{1-\alpha}} \right] \end{aligned} \quad (26)$$

Exemple 1 : série x

Pour la série x , les résultats obtenus avec la divergence de Rényi pour des valeurs de α comprises entre 0 et 1 sont présentés aux figures 2(a) et 2(b) pour des fenêtres d'analyse respectivement de $M = 500$ et $M = 1000$ points.

La divergence entre les systèmes $\mathcal{S}_1(T)$ et $\mathcal{S}_2(T)$ définis par (25) présente un maximum pour la date T correspondant à la date du changement de modèles autorégressifs : il s'agit de la date où $\mathcal{S}_1(T)$ et $\mathcal{S}_2(T)$ sont les plus dissemblables. La largeur du « pic » pour lequel les divergences mesurées sont importantes correspond à la largeur des fenêtres d'analyse comme le montre la comparaison des figures 2(a), 2(b) et 2(c).

Soit T_R la date du changement de modèles.

– Si $T \leq T_R - M$, $\mathcal{S}_1(T)$ et $\mathcal{S}_2(T)$ décrivent la dynamique du même modèle autorégressif : $R_\alpha(\mathcal{S}_1(T), \mathcal{S}_2(T))$ est faible.

– Si $T_R - M \leq T \leq T_R$, $\mathcal{S}_2(T)$ est constitué de la superposition des dynamiques correspondant à deux modèles autorégressifs distincts tandis que $\mathcal{S}_1(T)$ ne décrit celle que du premier : quand T augmente, le nombre de points issus du deuxième modèle dans la fenêtre d'analyse croît et augmente la dissemblance de $\mathcal{S}_2(T)$ et $\mathcal{S}_1(T)$: $R_\alpha(\mathcal{S}_1(T), \mathcal{S}_2(T))$ est une fonction croissante de T .

– Si $T = T_R$, $\mathcal{S}_1(T)$ et $\mathcal{S}_2(T)$ décrivent la dynamique de deux modèles distincts et le maximum de dissemblance est atteint, ce qui se traduit par le maximum de $R_\alpha(\mathcal{S}_1(T), \mathcal{S}_2(T))$.

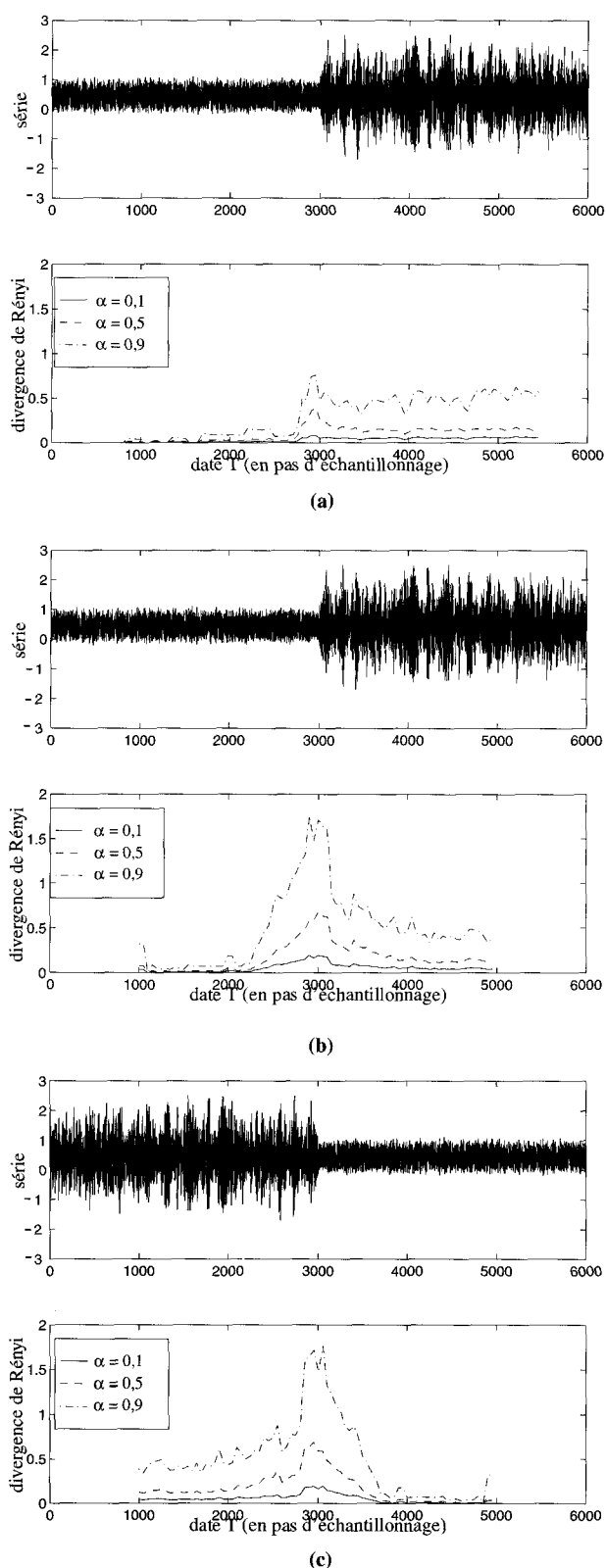


Figure 2. – Évolution des divergences de Rényi pour $\alpha = 0,1$, $\alpha = 0,5$ et $\alpha = 0,9$ en fonction de la date T séparant les deux espaces comparés (obtenus avec $M = 500$ points (a) et $M = 1000$ points (b)) pour la série temporelle x_n définie dans le tableau 2 ainsi que pour la série x_{-n} avec $M = 1000$ points (c).

– Si $T_R \leq T \leq T_R + M$, $S_1(T)$ comprend la superposition des dynamiques correspondant aux deux modèles autorégressifs tandis que $S_2(T)$ ne décrit celle que du second : quand T augmente, la dissemblance du système S_1 avec S_2 diminue du fait de la diminution du nombre de points du premier modèle dans la fenêtre d'analyse : $R_\alpha(S_1(T), S_2(T))$ est une fonction décroissante de T .

– Si $T \geq T_R + M$, $S_1(T)$ et $S_2(T)$ décrivent la dynamique d'un même modèle autorégressif : $R_\alpha(S_1(T), S_2(T))$ est faible.

La valeur moyenne de la divergence entre $S_1(T)$ et $S_2(T)$ issus d'un même modèle dépend du modèle autorégressif : elle est plus importante ici par exemple pour le deuxième modèle que pour le premier comme le prouvent la figure 2.

Les résultats obtenus à partir de la série x_{-n} (cf. figure 2(c)) présentent les mêmes caractéristiques, ce qui confirme que la technique employée détecte des différences de dynamique.

Exemple 2 : série z

L'étude de la série z définie dans le tableau 2 montre que plusieurs changements peuvent être détectés. La seule réserve concerne la taille de la fenêtre d'analyse : il est clair que le nombre de points de cette dernière doit nécessairement être inférieur à la durée (estimée en nombre de points) entre deux changements.

La figure 3(b) permet de vérifier que la détection de rupture ne dépend pas des seules fluctuations de l'amplitude ou de la puissance du signal observées directement sur les séries temporelles. La même expérience est reconduite sur le cas où la variance du bruit d'excitation est choisie de manière à ce que l'énergie du signal de sortie soit constante pour chacun des modèles successifs. Si la distinction des « pics » est un peu plus difficile, les ruptures n'en restent pas moins apparentes.

Exemple 3 : série y

Le dernier exemple abordé montre que la technique proposée permet aussi de détecter des ruptures entre deux modèles d'ordre différent : la série y considérée est formée d'un modèle autorégressif d'ordre 2 suivi d'un modèle autorégressif d'ordre 3.

Quelques remarques s'imposent ici.

– Les résultats de détection de rupture, présentés dans ce paragraphe, semblent indiquer que le choix optimal du paramètre α de la divergence de Rényi est $\alpha \simeq 1$ (nous nous sommes volontairement restreints à $0 < \alpha < 1$). Il est intéressant de « normaliser » les résultats précédents par la valeur maximale atteinte par la fonction divergence choisie. Ainsi $R_\alpha(S_1(T), S_2(T))$ est remplacé par

$$\frac{R_\alpha(S_1(T), S_2(T))}{\max_T [R_\alpha(S_1(T), S_2(T))]}.$$

L'opportunité de choisir α très proche de 1, ce qui fait tendre R_α vers la divergence de Kullback-Leibler (ou une variante symétrisée), disparaît alors. Il n'existe à notre connaissance pas de critères simples permettant de choisir la valeur optimale de α dans le cas de la divergence de Rényi entre lois de probabilités estimées.

– Dans les simulations présentées, les deux fenêtres de signal utilisées sont glissantes et adjacentes (à la date T). Aucun modèle n'est utilisé *a priori*. Aucune hypothèse n'est nécessaire sur la stationnarité du signal dans l'une ou l'autre des fenêtres. De ce fait, la comparaison avec d'autres méthodes (voir e.g. [6]) à

Comparaison de systèmes et arbres de régression

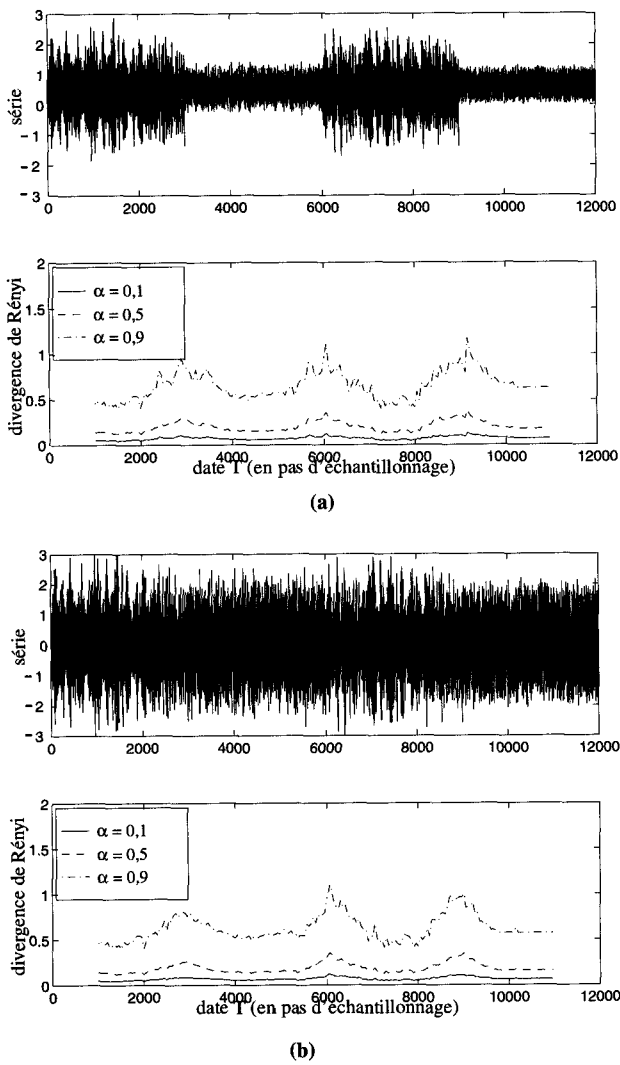


Figure 3. – Évolution des divergences de Rényi pour $\alpha = 0,1$, $\alpha = 0,5$ et $\alpha = 0,9$ en fonction de la date T séparant les deux espaces comparés (obtenus avec $M = 1000$ points (a)) pour la série temporelle z définie dans le tableau 2. (b) présente pour la même série mais avec une variance du bruit choisie pour obtenir $E(z_n^2)$ constant ($M = 1000$ points).

fenêtres glissantes construites sur l'hypothèse de signaux à spectre rationnel est difficile. Ces dernières utilisent, de fait, les filtres adaptés aux signaux étudiés, quand l'approche étudiée dans cet article est totalement non paramétrique.

– Rappelons cependant l'hypothèse forte utilisée ici : toutes les réalisations du vecteur d'état restent confinées dans un hyper-volume à d dimensions, de taille finie. Ainsi, la présence d'une instabilité, si elle est observable (temps court, limitation par une non linéarité, ou rupture de modèle à partir d'un seuil donné [23, 2]) n'est perçue à travers l'algorithme que comme un régime dynamique particulier.

Dans le paragraphe suivant, l'intérêt de cette approche non paramétrique est illustré dans le cadre de la détection de solitons, dont la dynamique non linéaire n'est pas connue analytiquement.

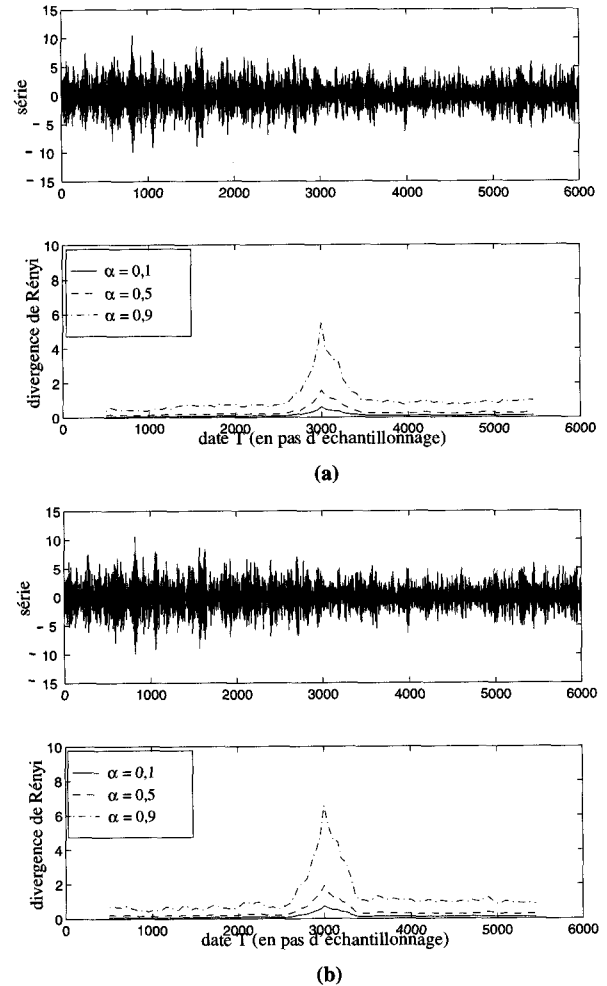


Figure 4. – Évolution des divergences de Rényi pour $\alpha = 0,1$, $\alpha = 0,5$ et $\alpha = 0,9$ en fonction de la date T séparant les deux espaces comparés (obtenus avec $M = 500$ points) pour la série y définie dans le tableau 2 avec une reconstruction en dimension $d = 4$ (a) et $d = 3$ (b).

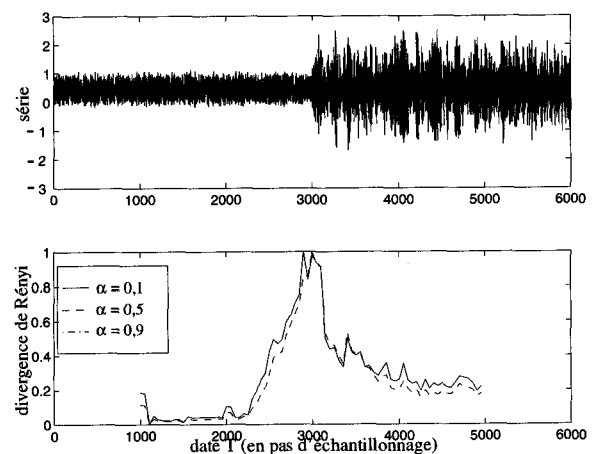


Figure 5. – Évolution des divergences de Rényi pour $\alpha = 0,1$, $\alpha = 0,5$ et $\alpha = 0,9$ en fonction de la date T séparant les deux espaces comparés (obtenus avec $M = 1000$ points) pour la série temporelle x définie dans le tableau 2. Les résultats obtenus sont normalisés par la valeur maximale atteinte par la divergence.

4. détection de « breathers »

4.1. description du problème

L'intérêt de la méthode proposée est illustré dans ce paragraphe par son application au problème de détection de comportement dynamique de type soliton. L'existence de solitons traduit des effets collectifs et l'existence de structures spatiales cohérentes liées à un grand nombre de degrés de liberté. Ce sont des ondes solitaires localisées spatialement et présentant une stabilité exceptionnelle au sens où leur forme et leur vitesse ne varient pas lors de leur propagation, y compris lors de collisions avec d'autres solitons. Correspondant à un maximum d'énergie localisée, les solitons sont également considérés comme des quasi-particules présentant des propriétés ondulatoires. Les solitons de type breather, auxquels on s'intéresse ici, sont une des solutions de l'équation de Sine-Gordon :

$$\frac{\partial^2 \phi}{\partial t^2} - c_0^2 \frac{\partial^2 \phi}{\partial x^2} + \omega_0^2 \sin \phi = 0 \quad (27)$$

Cette solution correspond au cas où l'interaction entre solitons peut être modélisée par un potentiel attractif. Les données utilisées ici proviennent d'une simulation numérique⁵ sur un modèle d'oscillateurs couplés : une chaîne de N oscillateurs de masse m couplés à l'aide de ressorts de raideur k et soumis à un potentiel de confinement V . L'Hamiltonien lié à un oscillateur indexé par i s'écrit donc :

$$H_i = \frac{1}{2} m \dot{x}_i^2 + \frac{k}{2} (x_i - x_{i+1})^2 + V(x_i) \quad (28)$$

Lorsque le potentiel V est un potentiel de Morse :

$$V(x) = \frac{1}{2} \omega_0^2 (e^{-\alpha x} - 1)^2 \quad (29)$$

des breathers peuvent se développer de manière locale en temps et en espace tandis que ces structures particulières n'existent pas avec un potentiel harmonique :

$$V(x) = \frac{1}{2} \omega_0^2 x^2 \quad (30)$$

Dans la suite, trois types de données⁶ seront utilisés correspondant à un potentiel harmonique sur une évolution temporelle de 1024 pas et à un potentiel de Morse sur une évolution temporelle de 1024 ou 4096 pas.

5. Ces données ont été fournies par M. Peyrard, professeur au laboratoire de Physique de l'Ecole Normale Supérieure de Lyon.

6. Le nombre d'oscillateurs étant fini, des conditions aux limites périodiques seront appliquées : le comportement du $(N + 1)$ -ième oscillateur est identique à celui du premier ($x_{N+1} = x_1$). Les différents paramètres ont les valeurs suivantes : $\frac{1}{2} \omega_0^2 = 0,03$, $m = 300$, $k = 0,08$ et $\alpha = 4,5$.

4.2. comparaison des dynamiques locales

Les breathers sont « statiques » au sens où ils ne se déplacent pas mais ils oscillent localement. De plus, lorsqu'ils apparaissent, ils n'affectent pas l'ensemble de la chaîne mais un nombre limité d'oscillateurs. Leur localisation en temps et en espace conduit à l'utilisation de la méthode proposée dans les paragraphes précédents, en considérant les comportements dynamiques locaux d'un groupe d'oscillateurs voisins. La modélisation reposant directement sur les réalisations du système dans son espace des états donne une représentation exhaustive de la dynamique locale du système : la partition de l'espace induit par les données en est une signature.

Les espaces des états sont obtenus en considérant les vecteurs dont les composantes sont les amplitudes prises par un ensemble d'oscillateurs consécutifs ou voisins sur la chaîne d'oscillateurs :

$$S_q = \{\mathbf{X}_i = [x_{q,i} \ x_{q+1,i} \ \dots \ x_{q+d-1,i}]^T, \quad (31)$$

$$\text{pour } i = 1, \dots, N\} \quad (32)$$

$x_{q,i}$ étant la position de l'oscillateur q à la date i . Ainsi si on retient la dimension $d = 2$ pour l'espace des états, ce dernier sera constitué par le comportement de deux oscillateurs contigus. Sur chaque espace des états pouvant être constitué de cette manière est estimée la partition à l'aide des algorithmes présentés précédemment.

Un groupe d'oscillateurs voisins (ou de manière équivalente un espace S_{ref}) est choisi comme référence et les divergences entre lois de probabilité entre cette référence et les différents systèmes d'oscillateurs consécutifs : $d(S_{ref}, S_i)$ sont estimées. Les données utilisées (correspondant à 1024 pas temporels pour 256 oscillateurs) sont visualisées par les figures 6(a) pour le potentiel harmonique et 6(b) pour le potentiel de Morse. L'existence de breathers dans le cas du potentiel de Morse est visible grâce aux nettes variations de teinte traduisant des différences importantes d'amplitude. *A contrario*, l'apparence homogène des amplitudes en temps et en espace pour le potentiel harmonique révèle l'absence de ces structures particulières.

Différents types de divergences sont appliqués dans le contexte décrit aux paragraphes précédents avec des espaces des états de dimension 2. Les figures suivantes présentent les courbes $d(S_{ref}, S_i) = f(i)$ pour ces diverses divergences :

- Distance variationnelle de Kolmogorov : voir figure 7.
- Divergence d'Hellinger : voir figure 8.
- Divergence de Kullback-Leibler : voir figure 9.
- Divergences de Rényi : voir figure 10.

Les résultats obtenus avec ces quatre divergences illustrent la capacité de la méthode proposée à détecter la présence de structures localisées dans l'espace : $d(S_{ref}, S_i)$ est nulle quel que soit S_i en l'absence de breathers (cas du potentiel harmonique) et présente des variations notables lorsque ces structures peuvent exister sur certaines parties de la chaîne d'oscillateurs (cas du potentiel de Morse). Les différentes divergences employées conduisent au même type de conclusions même si les distances variationnelle de Kolmogorov ou d'Hellinger, par exemple, donnent lieu à des fluctuations plus importantes, visibles notamment sur le cas du potentiel harmonique.

Comparaison de systèmes et arbres de régression

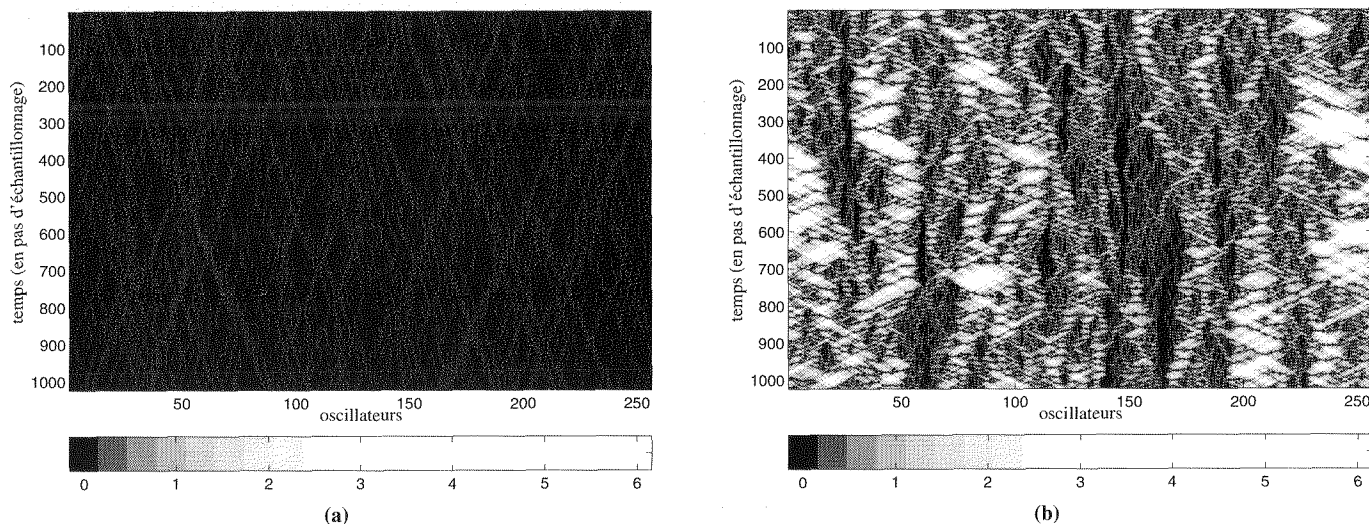


Figure 6. – Données liées à un potentiel harmonique (a) et à un potentiel de Morse (b) pour 1024 pas temporels et 256 oscillateurs.

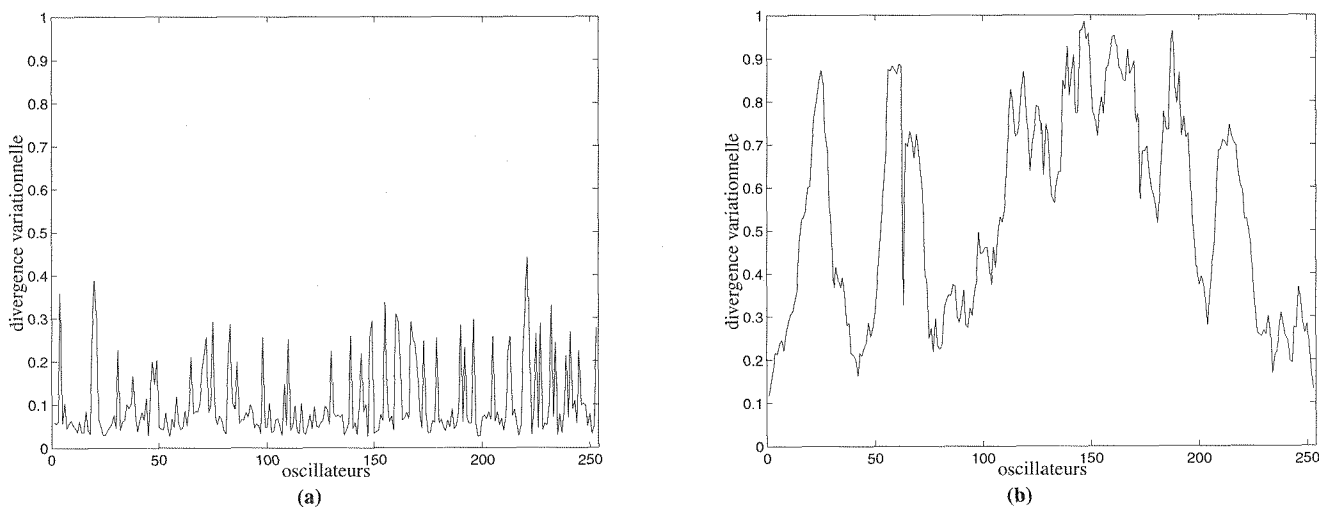


Figure 7. – Distance variationnelle de Kolmogorov entre le système S_1 formé par les deux premiers oscillateurs et les autres systèmes S_q constitués de deux oscillateurs successifs qu'il est possible de former pour les données relatives à un potentiel harmonique (a) et à un potentiel de Morse (b).

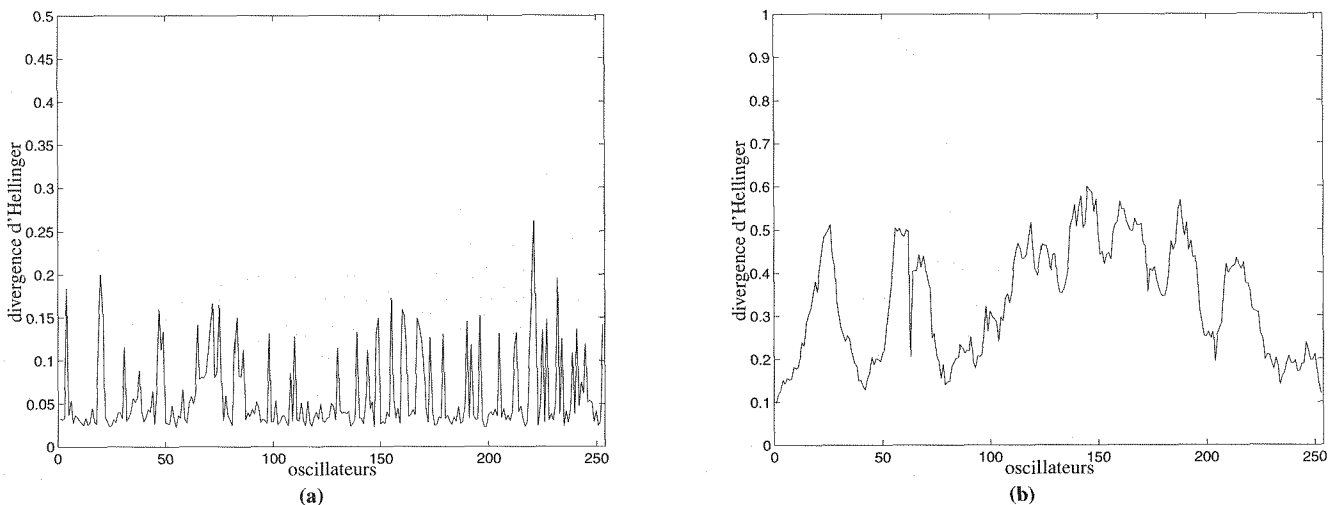


Figure 8. – Divergence d'Hellinger entre le système S_1 formé par les deux premiers oscillateurs et les autres systèmes S_q constitués de deux oscillateurs successifs qu'il est possible de former pour les données relatives à un potentiel harmonique (a) et à un potentiel de Morse (b).

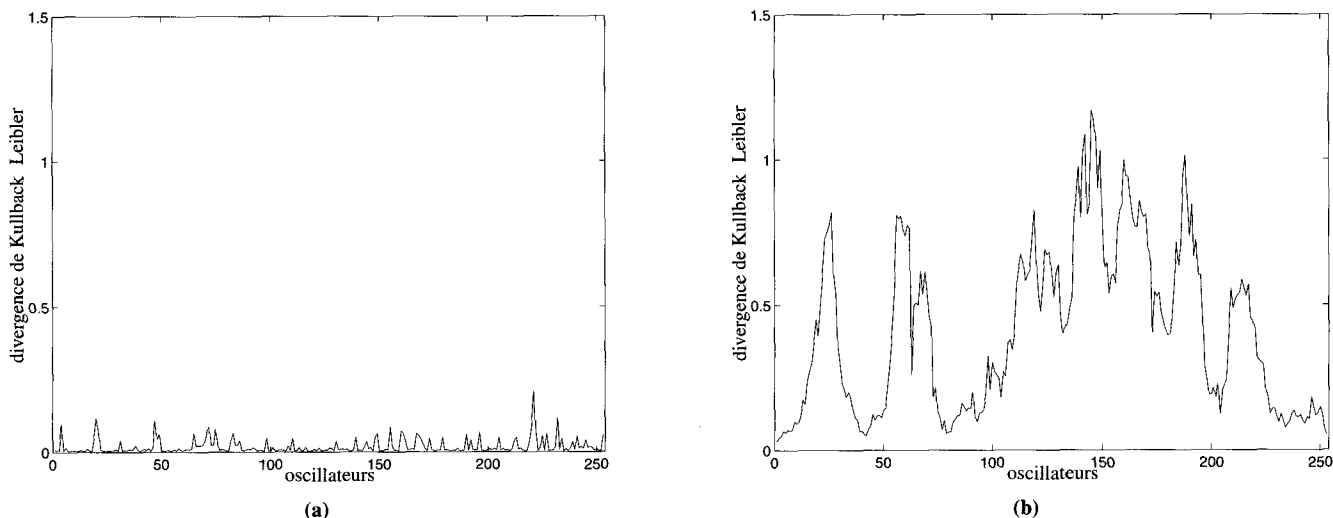


Figure 9. – Divergence de Kullback-Leibler entre le système S_1 formé par les deux premiers oscillateurs et les autres systèmes S_q constitués de deux oscillateurs successifs qu'il est possible de former pour les données relatives à un potentiel harmonique (a) et à un potentiel de Morse (b).

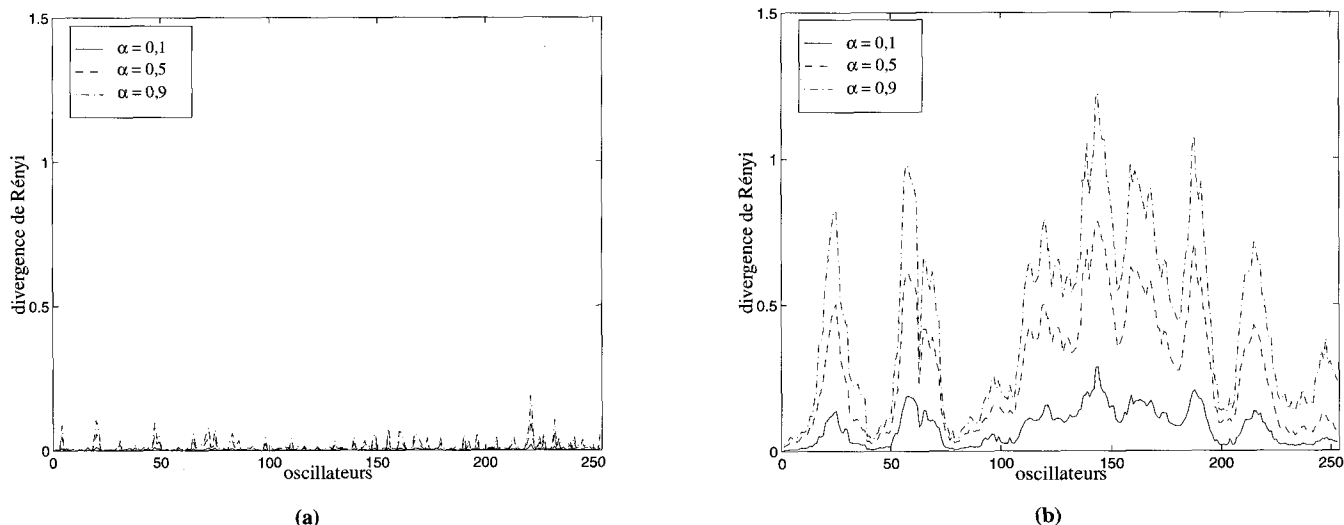


Figure 10. – Divergences de Rényi ($\alpha = 0,1$, $\alpha = 0,5$ et $\alpha = 0,9$) entre le système S_1 formé par les deux premiers oscillateurs et les autres systèmes S_q constitués de deux oscillateurs successifs qu'il est possible de former pour les données relatives à un potentiel harmonique (a) et à un potentiel de Morse (b).

Cette étude ne fournit malheureusement pas une réponse complète au problème de la détection : la seule conclusion qu'il est possible de tirer est la capacité de classer les espaces des états entre eux selon un critère de ressemblance au sens où les divergences entre leur distribution de probabilité sont faibles. Ensuite si on souhaite une véritable détection au sens où il serait possible d'affirmer avec certitude « il existe un breather à tel endroit », il faudrait savoir *a priori* si l'espace des états de référence correspond à ce critère. Cet aspect ne sera pas abordé ici.

L'influence de l'éventuelle symétrisation des divergences fera l'objet du paragraphe suivant tandis que la mesure des ressemblances et des dissemblances à partir du choix d'une autre référence ou avec une dimension de l'espace des états (ou un nombre d'oscillateurs consécutifs) plus grand sera étudiée aux paragraphes ultérieurs.

4.3. influence de la symétrisation des divergences

La figure 11 présente les résultats obtenus à partir des différentes combinaisons possibles permettant de construire une divergence (ici la divergence de Rényi) à partir des histogrammes empiriques estimés sur S_1 ou S_2 . Le tableau 3 présente les notations employées pour ces combinaisons (en utilisant les notations utilisées au paragraphe 2.2 pour les distributions de probabilité).

Quelque soit la quantité $R_1^{(1)}$, $R_2^{(1)}$, $R_1^{(2)}$ ou $R_2^{(2)}$ envisagée pour mesurer l'écart entre deux distributions de probabilité, le même comportement est observé : les distributions « proches » avec une quantité le restent avec une autre. Il est possible de montrer que la symétrisation ne modifiera donc pas les conclusions. Par contre,

Comparaison de systèmes et arbres de régression

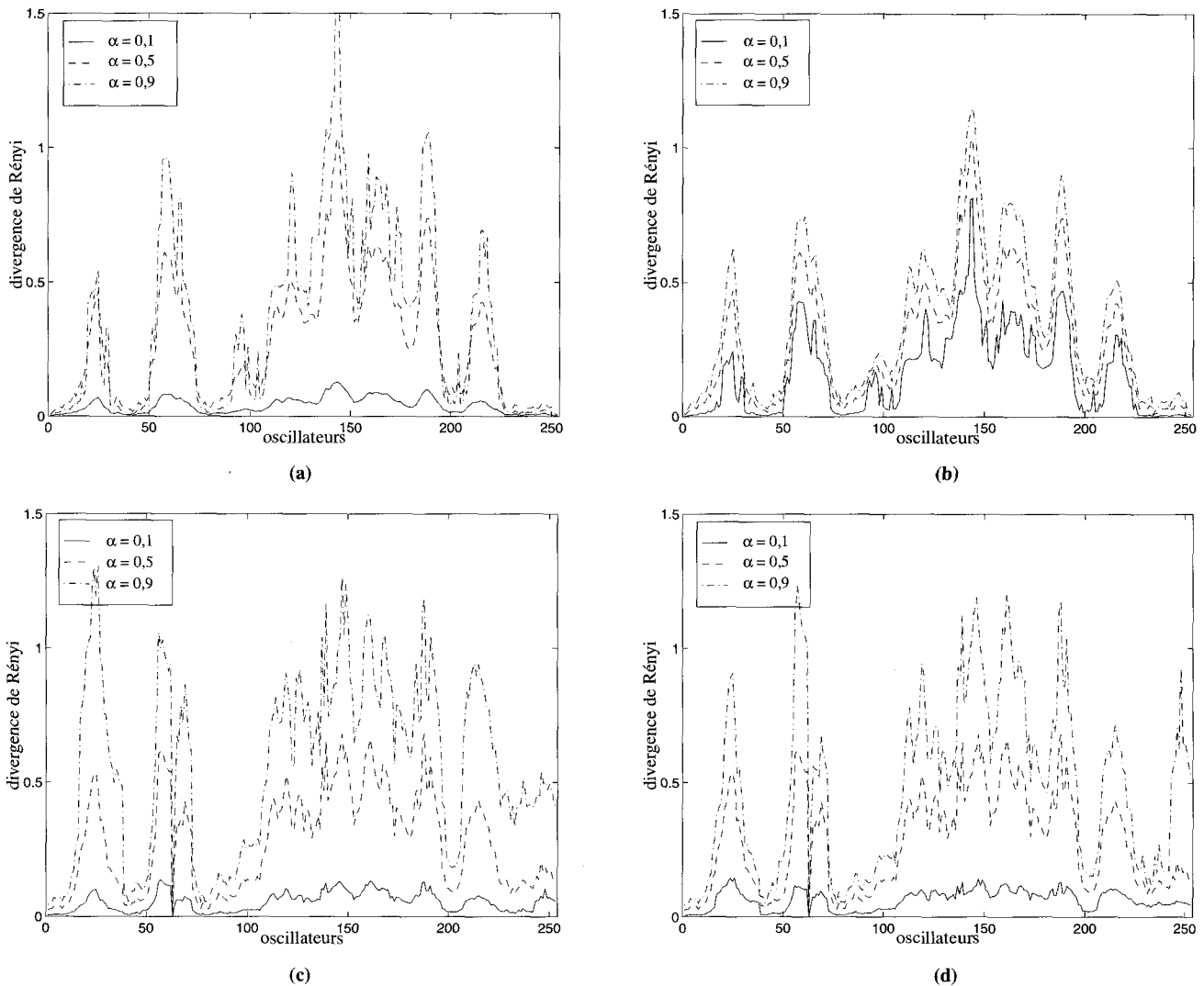


Figure 11. – Divergences de Rényi ($\alpha = 0,1$, $\alpha = 0,5$ et $\alpha = 0,9$) : $R_1^{(1)}$ (a), $R_2^{(1)}$ (b), $R_1^{(2)}$ (c) et $R_2^{(2)}$ (d) entre le système de référence S_1 formé par les deux premiers oscillateurs et les autres systèmes S_i constitués de deux oscillateurs successifs qu'il est possible de former pour les données relatives à un potentiel de Morse.

elle permet de ne privilégier aucune partition en considérant les quantités : $\frac{R_1^{(1)} + R_1^{(2)}}{2}$, $\frac{R_1^{(1)} + R_2^{(2)}}{2}$, $\frac{R_2^{(1)} + R_1^{(2)}}{2}$ et $\frac{R_2^{(1)} + R_2^{(2)}}{2}$ ou aucune distribution par rapport à l'autre avec $\frac{R_1^{(1)} + R_2^{(1)}}{2}$ ou $\frac{R_1^{(2)} + R_2^{(2)}}{2}$ ou encore l'expression totalement symétrisée : $\frac{R_1^{(1)} + R_2^{(1)} + R_1^{(2)} + R_2^{(2)}}{4}$.

La validité de l'approche dépend également du nombre de cellules que contient la partition : le cas extrême où la partition se réduit à une unique cellule conduit à une probabilité 1 pour la cellule et donc à une valeur nulle pour toutes les divergences proposées. Cette mesure de distance n'implique pas pour autant l'absence de différences au niveau dynamique. La symétrisation complète de l'expression permet de s'affranchir de ce type

d'artefacts dans la mesure où le risque d'obtenir une seule cellule pour chacune des deux partitions correspondant respectivement aux deux systèmes est beaucoup plus faible que celui d'avoir une cellule unique avec une seule partition.

Tableau 3. – Résumé des différentes divergences envisageables pour comparer deux systèmes avec la divergence de Rényi.

Partition	Probabilité p_i	Probabilité q_i	Notation de la divergence
$\pi^{(1)}$	$S_1 \rightarrow p_{11}$	$S_2 \rightarrow p_{21}$	$R_1^{(1)}$
$\pi^{(1)}$	$S_2 \rightarrow p_{11}$	$S_1 \rightarrow p_{21}$	$R_2^{(1)}$
$\pi^{(2)}$	$S_1 \rightarrow p_{12}$	$S_2 \rightarrow p_{22}$	$R_1^{(2)}$
$\pi^{(2)}$	$S_2 \rightarrow p_{22}$	$S_1 \rightarrow p_{12}$	$R_2^{(2)}$

4.4. influence du choix de l'espace de référence S_{ref}

Les figures 12(a), 12(b) et 12(c) correspondent à divers choix d'espace de référence : le système S_1 formé des deux premiers oscillateurs, le système S_{25} et le système S_{42} correspondant respectivement aux oscillateurs 25 – 26 et 42 – 43 et présentant des dynamiques respectivement différente et proche (d'après les résultats obtenus à partir de l'étude avec le système S_1 comme référence).

Les résultats confirment les observations initiales : les systèmes S_1 et S_{42} présentent des dynamiques similaires et la divergence entre les distributions de probabilité estimées est faible. Les valeurs prises par $d(S_1, S_i)$ et $d(S_{42}, S_i)$ sont elles-aussi similaires. *A contrario*, les systèmes S_1 et S_{25} , dont la divergence entre distributions de probabilité estimées est importante, correspondent à des dynamiques différentes et les valeurs de $d(S_1, S_i)$ et $d(S_{25}, S_i)$ sont complémentaires au sens où à une faible valeur de $d(S_1, S_i)$ correspond une valeur importante de $d(S_{25}, S_i)$ et réciproquement. Ainsi puisqu'un couple d'oscillateurs différent de S_1 ne présente pas également une dynamique notablement différente de S_{25} (de dynamique différente de S_1), les différents couples d'oscillateurs peuvent être classés en deux catégories.

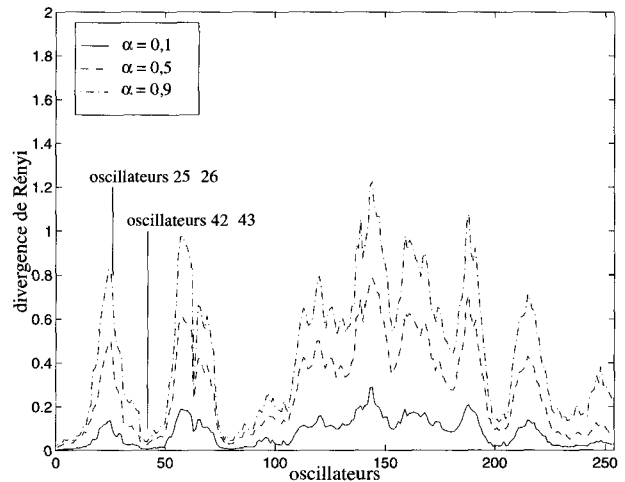
4.5. influence de la dimension de l'espace des états

Dans ce paragraphe, on considère des séries temporelles de 4096 points (cf. figure 13). L'étude porte sur l'influence de la dimension de l'espace des états (ou nombre d'oscillateurs voisins) pris en compte pour estimer la dynamique locale.

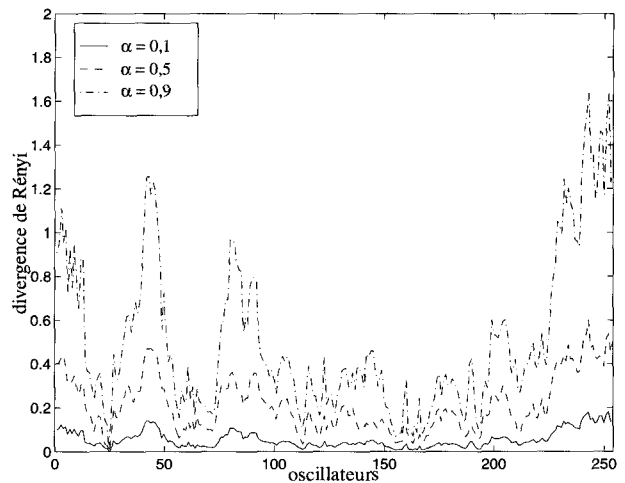
Soient L l'amplitude caractéristique des fluctuations temporelles observées sur les oscillateurs et N le nombre d'échantillons (temporels) considérés. La densité moyenne de réalisations d'un vecteur d'état de dimension d s'exprime alors par unité d'hyper-volume : $\delta = \frac{N}{L^d}$. Par conséquent, la probabilité d'observer une réalisation du vecteur d'état dans une cellule de dimension caractéristique l_0 varie comme $\left(\frac{l_0}{L}\right)^d$. Cet effet de dilution statistique conduit à une augmentation de la variance d'estimation de la distribution de probabilité quand d augmente. Sur les figures 14 et 15 sont représentées les divergences de Rényi pour α et d variables : dans le premier cas le nombre d'observations $N = 4096$ est fixé et dans le second la densité moyenne d'observation est maintenue constante (N varie comme $\left(\frac{L}{l_0}\right)^d$).

Les variations de $d(S_1, S_i) = f(i)$ présentent de nombreuses similitudes : les maxima obtenus concernent les mêmes oscillateurs (pour une même référence). Les valeurs numériques sont cependant différentes du fait du nombre de cellules de la partition qui varie également avec la dimension. On montre, par exemple, que la valeur maximale de la divergence de Rényi (19) vaut :

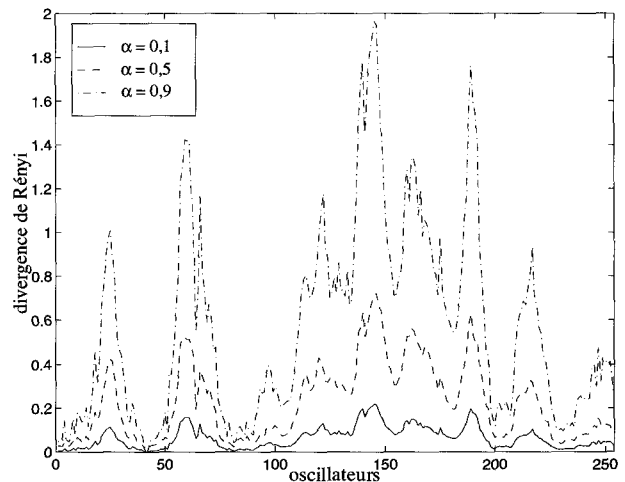
$$\mathcal{R}_{\max} = \frac{\ln \mathcal{N}_C}{2(1 - \alpha)} \quad (33)$$



(a)



(b)



(c)

Figure 12. – Divergences de Rényi ($\alpha = 0,1$, $\alpha = 0,5$ et $\alpha = 0,9$) totalement symétrisées entre un système S_i de référence constitué des deux oscillateurs voisins i et $i + 1$ (S_1 (a), S_{25} (b) et S_{42} (c)) et les autres systèmes S_i qu'il est possible de former pour les données relatives à un groupe de deux oscillateurs successifs relatifs à un potentiel de Morse.

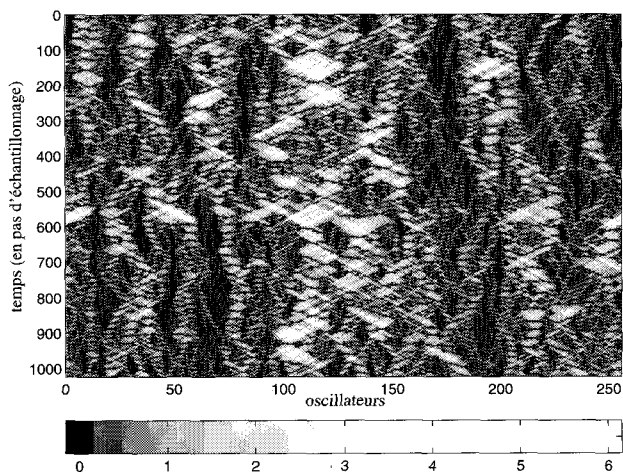


Figure 13. – Données relatives à un potentiel de Morse avec un plus grand nombre de pas temporels.

en notant \mathcal{N}_C le nombre total de cellules de la partition et en supposant que cette valeur maximale est obtenue pour les distributions P :

$$p_i = \begin{cases} 1 & \text{si } i = i_0 \\ 0 & \text{sinon} \end{cases} \quad (34)$$

et Q (uniforme) :

$$q_i = \frac{1}{\mathcal{N}_C} \quad (35)$$

Ainsi, pour $\alpha = 0,5$ par exemple, \mathcal{R}_{\max} prend les valeurs 2,7388; 4,0492; 5,3636 et 6,3183 respectivement pour les dimensions 2, 3, 4 et 5 (ces valeurs correspondent à la moyenne des quantités déterminées expérimentalement en fonction du nombre de cellules obtenues pour les partitions).

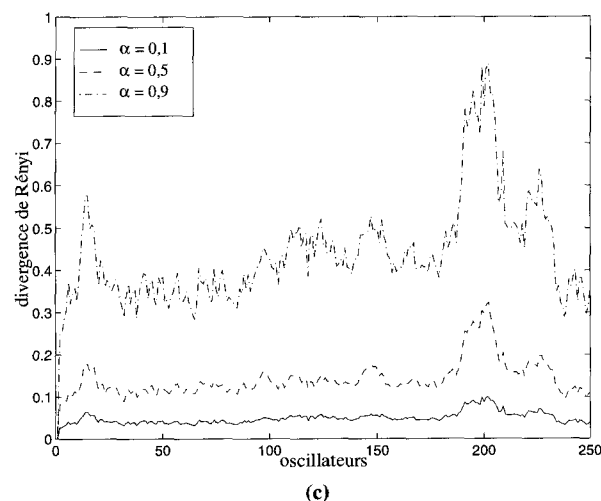
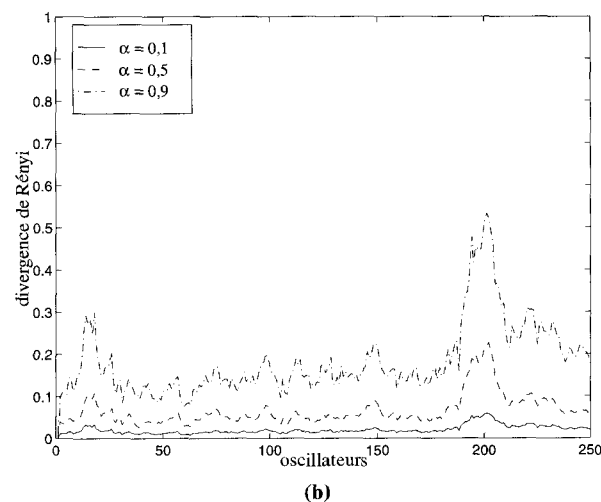
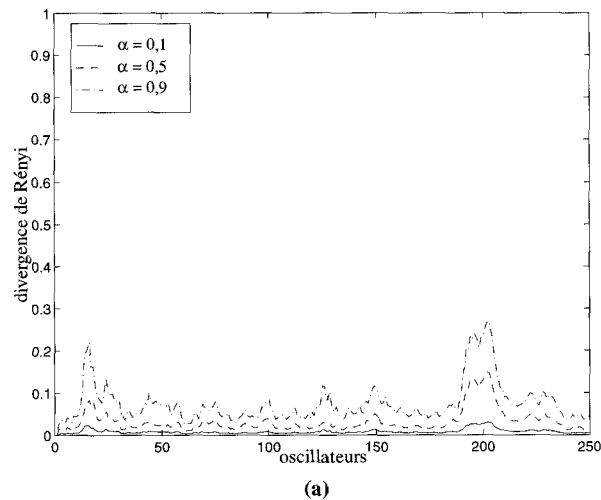
La comparaison des courbes relatives à la dimension 5 et aux autres montre l'influence de l'étendue spatiale des breathers : les différences de comportement dynamique pour les oscillateurs proches de l'indice 100 sont amplifiées en dimension 5.

L'augmentation du nombre de points conduit à masquer les effets de la présence d'un comportement dynamique local de type breather : la méthode compare les dynamiques moyennes et l'augmentation du nombre de points entraîne une dilution des comportements localisés. La comparaison des résultats obtenus avec 1024 et 4096 points illustre cet aspect.

5. conclusion

Les arbres de régression fournissent une modélisation intéressante des systèmes dynamiques surtout lorsqu'aucun *a priori* sur leur comportement n'est disponible et que les données constituent la seule information sur le système. L'utilisation de ce modèle en vue de comparer des dynamiques entre elles est particulièrement fructueuse : le plus souvent on ne dispose que de peu ou pas d'informations sur les données. L'histogramme d -dimensionnel

fourni par le modèle permet d'envisager une quantification des différences dynamiques à l'aide d'une estimation de distances ou de divergences entre lois de probabilité. Cette méthode a été



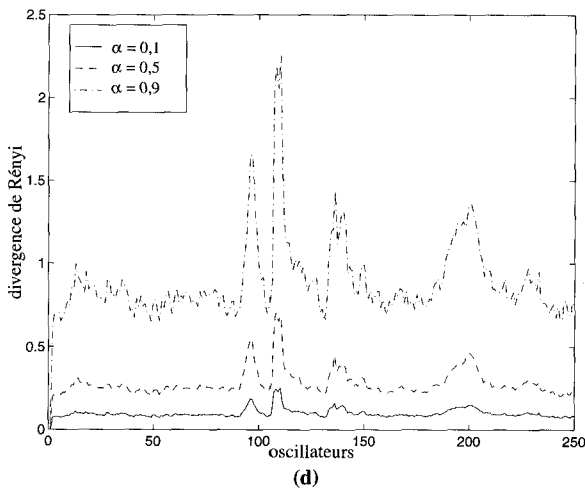


Figure 14. – Divergences de Rényi totalement symétrisées entre les systèmes S_1 formés des premiers oscillateurs successifs en dimension 2 (a), 3 (b), 4 (c) et 5 (d) et les autres systèmes S_q d'oscillateurs successifs qu'il est possible de former pour les données relatives à un potentiel de Morse.

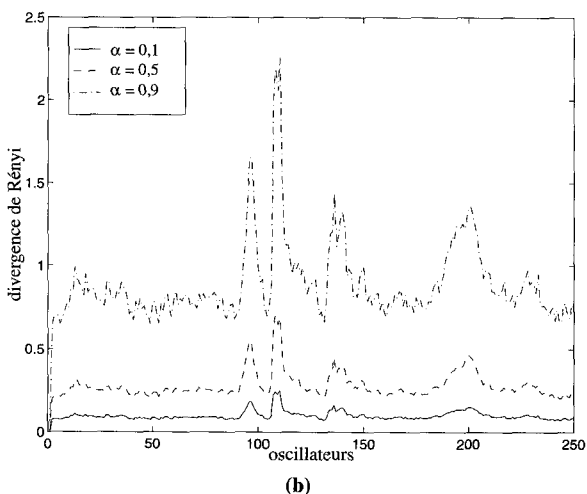
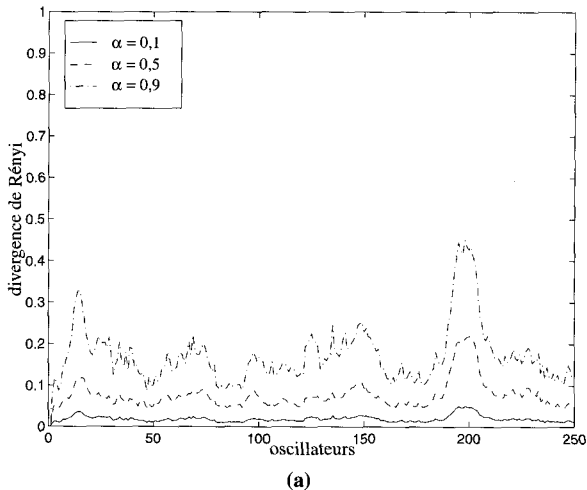


Figure 15. – Divergences de Rényi totalement symétrisées entre les systèmes S_1 formés des premiers oscillateurs successifs en dimension 4 (a) et 5 (b) à résolution constante et les autres systèmes S_q d'oscillateurs successifs qu'il est possible de former pour les données relatives à un potentiel de Morse.

appliquée sur deux exemples avec succès. La détection de changements de modèles autorégressifs est en effet obtenue sans connaître les modèles. Pour une réelle détection des modes « breather » dans la chaîne d'oscillateurs, il serait nécessaire de définir une référence correspondant à la présence ou à l'absence de tels solitons : l'approche proposée permet alors de quantifier l'écart d'un système à cette référence.

BIBLIOGRAPHIE

- [1] S.M. Ali, S.D. Silvey, "A General Class of Coefficients of Divergence of One Distribution from Another", *Journal of Royal Statistical Society*, vol. 28, 1966, pp. 131-142.
- [2] A.-E. Badel, O. Michel, A.O. Hero, "Modélisation Non Linéaire Non Paramétrique des Séries Temporelles à l'aide d'Arbres de Régression - Applications à l'Analyse de Séries Chaotiques", *Traitement du Signal*, vol. 14, no. 2, 1997, pp. 117-133.
- [3] A.-E. Badel, O. Michel, A.O. Hero, "Choosing the Right Rényi's Divergence Parameter to Compare Two Probability Distributions", rapport interne, 1997.
- [4] M. Basseville, "Distances Measures for Signal Processing and Pattern Recognition", *Signal Processing*, vol. 18, 1989, pp. 349-369.
- [5] M. Basseville, "Information : entropies, divergences et moyennes", IRISA, Publication interne No. 1020, 1996.
- [6] M. Basseville, A. Benhallam, C. Doncarli, M.F. Lucas, D. de Brucq, O. Colot, H. Rix, E. Thierry, F. Kauffmann, Fiches descriptives d'algorithmes de segmentation de signaux, *Traitement du Signal*, Vol.9, No.1, 1992.
- [7] M. Basseville, I.V. Nikiforov, "Detection of Abrupt Changes, Theory and Application", Prentice Hall Information and System Sciences Series, Thomas Kailath, Series Editor, 1993.
- [8] I. Csiszar, "Information-Type Measures of Divergence of Probability Distributions and Indirect Observations", *Studia Sci. Math. Hung.*, vol. 2, 1967, pp. 299-318.
- [9] P. Grassberger, I. Procaccia, "Characterization of Strange Attractors", *Physical Review Letters*, vol. 50, no. 5, 1983, pp. 346-349.
- [10] A.H. Gray, J.D. Markel, "Distance Measures for Signal Processing", *I.E.E.E. Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 5, pp. 380-391, 1976.
- [11] M.E. Havrda, F. Charvat, "Quantification method of classification processes : concept of α -structural entropy", *Kybernetika*, vol.3, pp. 30-34, 1967
- [12] S. Kullback, R.A. Leibler, "On Information and Sufficiency", *Annals of Mathematical Statistics*, vol. 22, 1951, pp. 79-86.
- [13] O. Michel, P. Flandrin, "Application of Methods Based on Higher Order Statistics for Chaotic Signal Analysis", *Signal Processing*, vol. 53, no.2, 1996.
- [14] O. Michel, A.O. Hero, "Tree-Structured Non-Linear Signal Modeling and Prediction", *ICASSP'95 Proceedings*, Detroit, Michigan, 1995.
- [15] O. Michel, A.O. Hero, A.-E. Badel, "Tree Structured Non-Linear Signal Modeling and Prediction", soumis à *I.E.E.E. Signal Processing*, Août 1997.
- [16] T. Nemetz, "Information-Type Measures and their Applications to Finite Decision Problems", *Carleton Mathematical Lecture Notes*, vol. 17, Carleton University, Ottawa, 1977.
- [17] N. H. Packard, J.P. Crutchfield, J.D. Farmer, R. Shaw, "Geometry from a Time Series", *Physical Review Letters*, vol. 45, no. 9, 1980, pp. 712-716.
- [18] C.R. Rao, "Differential Metrics in Probability Spaces", Institute of Mathematical Statistics, Lecture Notes - Monograph Series, vol. 10, S. Gupta Ed., pp. 217-240.
- [19] A. Rényi, "On Measures of Entropy and Information", *4th Berkeley Symposium*, vol. I, pp. 547-561, 1961.
- [20] A. Rényi, "Probability Theory", North Holland Publishing Company, 1970.

Comparaison de systèmes et arbres de régression

- [21] D. Ruelle, *Chaotic Evolution and Strange Attractors*, Cambridge University Press, 1989.
- [22] F. Takens, "Detecting Strange Attractors in Turbulence", *Lecture Notes in Mathematics*, vol. 898, 1981, pp. 366-381.
- [23] H. Tong, "Non Linear Time Series : a Dynamical System Approach", Oxford Science Publication, Oxford University Press, NY, 1990.
- [24] C. Tsallis, "Possible Generalization of Boltzmann-Gibbs Statistics", *Journal of Statistical Physics*, Vol.52, No.1-2, pp. 479-496, 1988.
- [25] H. Whitney, "Differentiable Manifolds", *Annals of Mathematics*, vol. 37, no. 3, 1936, pp. 645-680.

Manuscrit reçu le 3 juillet 1997.

LES AUTEURS

Anne-Emmanuelle BADEL



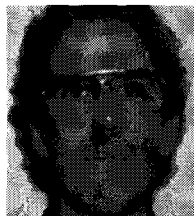
Ancienne élève de l'Ecole Normale Supérieure de Lyon; reçue à l'agrégation de Sciences Physiques (option Physique) en 1994. Doctorat en Sciences de l'ENS-Lyon, spécialité Physique (1998). Agrégé-Préparateur au département des Sciences de la Matière à l'ENS-Lyon depuis 1995. Domaines de recherche : arbres de régression, modélisation non linéaire et systèmes dynamiques.

Olivier MICHEL



Ancien élève de l'Ecole Normale Supérieure de Cachan; reçu à l'agrégation de Sciences Physiques (option Physique Appliquée) en 1986. Doctorat en Sciences de l'Université Paris XI-Orsay, spécialité Traitement du Signal(1991). Maître de Conférences au département des Sciences de la Matière à l'ENS-Lyon depuis 1991. Domaines de recherche : systèmes dynamiques et théorie de l'information, modélisation non linéaire, statistiques d'ordre élevé, analyse temps-fréquence, arbres de régression.

Alfred O. HERO



Alfred HERO est né a Boston, Massachusetts (Etats Unis) en 1955. Il a obtenu son PhD en *Electrical Engineering and Computer Science* (EECS) à Princeton University en 1984. Il est Professeur de EECS et de Biomedical Engineering à l'Université de Michigan, Ann Arbor (Etats Unis), depuis 1984. Il a été visiting scientist à M.I.T. Lincoln Laboratory en 1987, 1988 et 1989, au LSS et à l'ENSTA en 1991, et à Ford Scientific Research Laboratory en 1993. Il est Fellow de l'Institute of Electrical and Electronic Engineers

(IEEE) et il est membre de la American Statistical Association (ASA). Ses intérêts scientifiques sont principalement l'application des méthodes d'analyse statistique aux problèmes de traitement d'images, au traitement du signal et aux télécommunications.