

# Comparaison de systèmes modélisés par des arbres de régression

Anne-Emmanuelle Badel,<sup>◇</sup> Olivier Michel,<sup>◇</sup> et Alfred Hero<sup>\*</sup>

<sup>◇</sup> Laboratoire de Physique (URA 1325 CNRS), École Normale Supérieure de Lyon,  
46 allée d'Italie, 69364 Lyon Cedex 07, France  
tel : 04 72 72 85 59 - fax : 04 72 72 80 80

<sup>\*</sup> Department of EECS, University of Michigan  
Ann Arbor, MI 48109-2122, USA

e-mail : aebadel@physique.ens-lyon.fr, omichel@physique.ens-lyon.fr, hero@eecs.umich.edu

## RÉSUMÉ

Nous présentons une méthode permettant de comparer des systèmes dynamiques à faible nombre de degrés de liberté. À partir d'arbres de régression associés à une partition récursive de l'espace des états (éventuellement reconstruit) du système [1, 2], une mesure de distance entre lois de probabilité induites par les arbres de régression est proposée. L'intérêt de cette approche est illustré sur deux exemples : la détection de changement de modèle autorégressif puis la recherche de solitons (de type "breather") dans une chaîne d'oscillateurs couplés non linéairement.

## ABSTRACT

A method is presented for comparing dynamical systems with a low number of degrees of freedom. From a coarse grained partitioning of the state-space which is related to a recursive tree growing algorithm [1, 2], the use of distance between probability laws induced by the partition is suggested. Its interest is illustrated on two examples : autoregressive model changes and "breather" soliton detection in a set of non linearly coupled oscillators.

## 1 Introduction

La modélisation non linéaire non paramétrique proposée dans [1, 2] consiste à construire une partition récursive de l'espace des phases (éventuellement obtenu par plongement [3]) représentant la dynamique du système. Ce modèle peut être représenté à l'aide d'un arbre de régression. L'objet de cet article est l'utilisation de cette approche pour comparer les comportements dynamiques de différents systèmes : nous proposons la mise en évidence des différences de dynamique par une mesure de distances entre les distributions de probabilité associées aux partitions de leur espace des états respectif. L'intérêt de cette approche est illustré sur deux exemples : la détection de changement de modèle autorégressif dans une série temporelle, puis une application basée sur l'étude d'une chaîne d'oscillateurs couplés et soumis à un potentiel extérieur (potentiel harmonique ou potentiel de Morse). Dans cet exemple, l'objectif est la détection de la présence éventuelle d'un soliton de type "breather" susceptible de se développer dans le cas du potentiel de Morse.

## 2 Arbres de régression et distances

Dans la suite de cette étude, il est supposé que chaque système observé peut être assimilé à un système dynamique à petit nombre de degrés de liberté. Cette hypothèse fondamentale justifie d'en conduire l'étude dans son espace des états. En

l'absence de tout modèle de la dynamique du système considéré, nous proposons de caractériser cette dernière par la fonction de densité de probabilité (ddp) qu'a le vecteur d'état du système de se trouver dans un voisinage donné de l'espace. Il est important de souligner que seuls des régimes stationnaires (au sens où les phases transitoires ne sont pas considérées) relèvent de cette approche. Cette dernière est fondée sur l'estimation de l'histogramme  $d$ -dimensionnel de la distribution des observations dans l'espace des états. Aucune partition de l'espace n'est imposée *a priori* : une construction récursive de la partition tend à définir un ensemble de "bins" ou de cellules, au sein desquelles la fonction de densité de probabilité (ddp) des observations est uniforme. Le test d'uniformité de la ddp, ainsi que les critères de partitionnement utilisés, sont décrits plus largement dans [1, 2], et brièvement rappelés ci-après.

L'élaboration du modèle par un algorithme récursif peut être représentée à l'aide d'un arbre : à l'étape  $p$  de la récursion, une cellule  $C_p$  de l'espace des phases (de dimension  $d$ ) est subdivisée en  $2^d$  sous-cellules (appelées cellules enfants,  $C_{p+1,i=1\dots 2^d}$ ). Les distributions marginales sur chacun des  $d$  axes de l'espace sont estimées (pour la cellule  $C_p$ ); la construction des cellules enfants  $C_{p+1,i=1\dots 2^d}$  est ensuite définie à partir des valeurs obtenues pour l'estimation du médian de ces distributions. L'uniformité de la ddp sur  $C_p$  est testée à partir des  $2^d$  valeurs  $p_{p+1,i} = \frac{|C_{p+1,i}|}{|C_p|}$ , où  $|C|$  représente le nombre d'observations contenues dans  $C$ .

Si la distribution de probabilité  $p_{p+1,i=1\dots 2^d}$  est jugée uni-

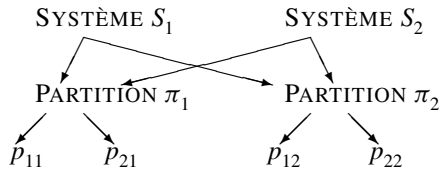
forme (e.g. au sens d'un test de  $\chi^2$ ),  $C_p$  est mémorisée : c'est une cellule de la partition finale ou "feuille terminale" et l'algorithme de partition est stoppé. Sinon, la partition, ainsi que l'ensemble des cellules  $C_{p+1,i=1\dots 2^d}$  sont mémorisées et analysées à l'étape  $p+1$  de l'algorithme de partitionnement récursif.

Le modèle ainsi obtenu fournit une distribution de probabilité associée à la dynamique du système : plutôt que de comparer directement les systèmes, il est possible de définir une distance entre les distributions de probabilité associées à chaque modèle.

Cependant, les partitions considérées étant totalement dirigées par les données (au sens où elles en dépendent directement par construction), une comparaison des systèmes par les histogrammes  $d$ -dimensionnels de leurs réalisations dans leur espace des états respectifs nécessite quelques précautions ; les cellules de la partition définie pour chacun des systèmes ne coïncident pas. Différentes approches sont alors possibles ; en particulier, il peut être envisagé de calculer la partition la plus fine définie par l'intersection des partitions associées à chaque système. Le coût algorithmique de cette approche en dimension  $d$  étant élevé, nous proposons ici d'utiliser l'arbre associé à chacune des partitions comme classifieur, et d'utiliser les distributions de probabilité

$$p_{ij,k} = P[x_i \in C_{j,k}]$$

où  $x_i$  est une observation du système  $S_i$  et  $C_{j,k}$  la cellule  $k$  de la partition de l'espace induite par les observations du système  $S_j$ . L'ensemble des distributions de probabilités qui peuvent être construites ainsi, est représenté par le schéma suivant :



$p_{11}$  est un estimateur de la ddp de  $S_1$  estimée à partir des données mesurées sur  $S_1$ .  $p_{21}$  peut être considéré comme un estimateur de la ddp des réalisations du vecteur d'état de  $S_1$  à partir des observations de  $S_2$ . On construit alors le rapport de vraisemblance de ce dernier :  $V = \frac{p_{21}}{p_{11}}$ . Différentes mesures de distances entre lois de probabilité de forme générale :

$$g \left( E_{p_{11}} \left[ f \left( \frac{p_{21}}{p_{11}} \right) \right] \right) \tag{1}$$

sont envisageables [4, 5, 6]. Nous nous sommes limités dans cette communication à l'étude de la distance de Rényi d'ordre  $\alpha$  (qui correspond au choix de  $g(x) = \frac{\log x}{\alpha-1}$  et  $f(x) = x^{1-\alpha}$ ) :

$$r_\alpha(S_1, S_2) = \frac{1}{\alpha-1} \log \left[ \sum_k p_{11,k}^\alpha p_{21,k}^{1-\alpha} \right] \tag{2}$$

Afin de ne privilégier aucun système par rapport à l'autre, une forme symétrisée de la distance précédente est utilisée :

$$r_\alpha^{(1)}(S_1, S_2) = \frac{1}{2(\alpha-1)} \log \left[ \sum_k p_{11,k}^\alpha p_{21,k}^{1-\alpha} \right] \left[ \sum_k p_{21,k}^\alpha p_{11,k}^{1-\alpha} \right] \tag{3}$$

De cette façon, deux distances  $r_\alpha^{(1)}(S_1, S_2)$  et  $r_\alpha^{(2)}(S_1, S_2)$  correspondant respectivement au choix de l'une ou l'autre des

partitions  $\pi_1$  ou  $\pi_2$  sont calculables. Afin de s'affranchir des dépendances liées au choix arbitraire d'une partition, nous préférons considérer une expression "totalement symétrique" obtenue par la moyenne des deux distances  $r_\alpha^{(1)}(S_1, S_2)$  et  $r_\alpha^{(2)}(S_1, S_2)$  calculées sur chacune des deux partitions :

$$R_\alpha(S_1, S_2) = \frac{1}{4(\alpha-1)} \log \left[ \sum_i p_{11,i}^\alpha p_{21,i}^{1-\alpha} \right] \left[ \sum_i p_{21,i}^\alpha p_{11,i}^{1-\alpha} \right] + \frac{1}{4(\alpha-1)} \log \left[ \sum_i p_{12,i}^\alpha p_{22,i}^{1-\alpha} \right] \left[ \sum_i p_{22,i}^\alpha p_{12,i}^{1-\alpha} \right] \tag{4}$$

### 3 Application à la détection de changements de modèles

On considère une série temporelle de  $N$  points, construite à l'aide d'une succession de modèles autorégressifs. L'ordre maximal  $m$  du modèle sera ici supposé connu ; nous illustrons dans cet exemple la capacité de l'approche proposée à détecter les dates de rupture, i.e. de changement de modèles. La dynamique de la série étudiée est analysée par plongement dans un espace de dimension  $(m+1)$  avec une valeur du retard prise égale au pas d'échantillonnage (voir [7] pour une discussion plus générale), ce qui conduit à construire l'ensemble des vecteurs :

$$\mathbf{X}_n = [x_n \ x_{n-1} \ \dots \ x_{n-m}]^T \tag{5}$$

Deux partitions de l'espace  $(m+1)$ -dimensionnel ainsi obtenu sont estimées à partir de deux sous-ensembles de  $N_0$  réalisations consécutives ( $N_0 \ll N$ ) :

$$\begin{aligned} S_1(T) : \{ \mathbf{X}_n \}, n = T - N_0, \dots, T - 1 \\ S_2(T) : \{ \mathbf{X}_n \}, n = T + 1, \dots, T + N_0 \end{aligned} \tag{6}$$

La distance  $d(T) = R_\alpha(S_1(T), S_2(T))$ , (cf (4)),  $N_0+1 \leq T \leq N - N_0$ , doit être maximale quand les dynamiques estimées de  $S_1$  et  $S_2$  à partir de  $N_0$  réalisations sont le plus dissemblables, soit pour la date  $T$  correspondant au changement de modèles :  $S_1(T)$  et  $S_2(T)$  sont alors estimés à partir de réalisations issues de deux modèles autorégressifs distincts.

La figure 1 présente les résultats obtenus pour la série temporelle suivante ( $m = 2$ ) :

$$x_n = \begin{cases} 0.8x_{n-1} - 0.9x_{n-2} + \varepsilon_n & \text{si } n = 3, \dots, 2999 \\ 0.3x_{n-1} - 0.1x_{n-2} + \varepsilon_n & \text{si } n = 3000, \dots, 5999 \\ 0.7x_{n-1} - 0.9x_{n-2} + \varepsilon_n & \text{si } n = 6000, \dots, 8999 \\ 0.5x_{n-1} - 0.4x_{n-2} + \varepsilon_n & \text{si } n = 9000, \dots, 12000 \end{cases} \tag{7}$$

$\varepsilon$  est un bruit blanc Gaussien centré de variance  $\sigma^2$ .

On montre alors que la largeur des "pics" pour lesquelles les distances mesurées sont significatives, est fonction de  $N_0$ , et que le maximum de distance obtenu est fonction de la différence entre les modèles AR successifs. Dans le cas représenté sur la figure 1(b), la variance du bruit d'excitation est choisie pour que l'énergie du signal  $x_n$  soit la même pour chacun des quatre modèles successifs. Bien que la distinction des "pics" soit plus difficile, les ruptures de modèles restent cependant apparentes.

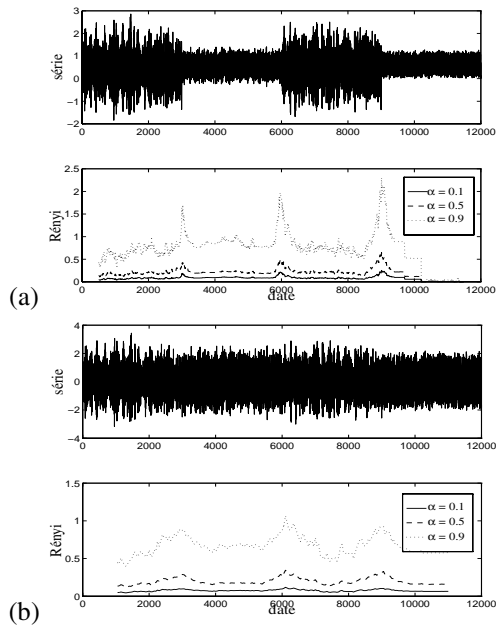


FIG. 1 — Évolution des distances de Rényi pour  $\alpha = 0.1$ ,  $\alpha = 0.5$  et  $\alpha = 0.9$  en fonction de la date  $T : d(T)$ . La série totale est constitué de  $N = 12000$  points et les espaces des états sont estimés avec  $N_0 = 500$  réalisations en dimension  $m + 1 = 3$ . (a) correspond à la série définie en (7) tandis que (b) présente les résultats obtenus à partir du modèle (7), pour lequel la variance du bruit d’excitation est choisie pour obtenir  $E(x_n^2)$  constant.

## 4 Détection de solitons

L’intérêt d’une modélisation s’appuyant directement sur les réalisations du système dans son espace des états tient à la spécificité de ce dernier de donner une représentation exhaustive de la dynamique du système. Ainsi, la partition (ou l’arbre de classification qui lui est associé) de l’espace des états, estimé à partir d’un ensemble d’observations, est la signature d’une certaine dynamique gouvernant le système.

Chaque ensemble d’observations est normalisé de sorte que les volumes d’espace des états “visités” par le système pour un ensemble d’observations donné se superposent : seules les caractéristiques dynamiques sont retenues.

Une orthogonalisation des données à l’aide d’une décomposition en valeurs singulières est opérée préalablement à la construction de la partition, afin de supprimer toute dépendance à une rotation éventuelle du système d’axes de l’espace des états.

La capacité de la méthode proposée ici à détecter des différences de comportements dynamiques est parfaitement illustrée par l’exemple suivant. On considère le champ spatio-temporel décrivant les positions d’un ensemble d’oscillateurs de masse  $m$ , constitué en chaîne. Les liens entre les oscillateurs sont des ressorts de raideur  $k$ , et l’ensemble est placé dans un champ de potentiel extérieur  $V(x)$ .

Dans le cas où la chaîne d’oscillateurs est placée dans un potentiel de Morse,

$$V(x) = \frac{1}{2} \omega_0^2 (e^{-\alpha x} - 1)^2 \quad (8)$$

les études théoriques montrent qu’il peut s’y développer des

solitons de type “breather”. Cependant, leur détection par une méthode systématique à partir de l’étude du champ spatio-temporel obtenu par simulation numérique<sup>1</sup> (avec conditions aux limites périodiques) n’est pas acquise.

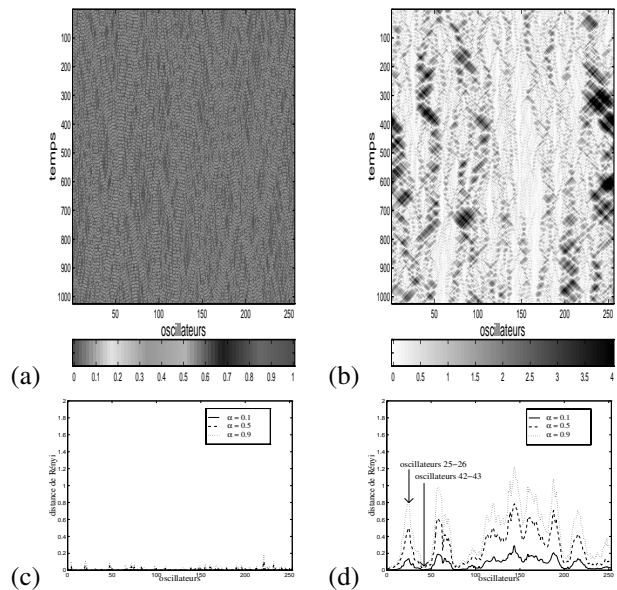


FIG. 2 — Données (a,b) et distances de Rényi ( $\alpha = 0.1$ ,  $\alpha = 0.5$  et  $\alpha = 0.9$ ) (c,d) entre les deux premiers oscillateurs et les autres groupes de deux oscillateurs successifs relatifs à un potentiel harmonique (a,c) et à un potentiel de Morse (b,d)

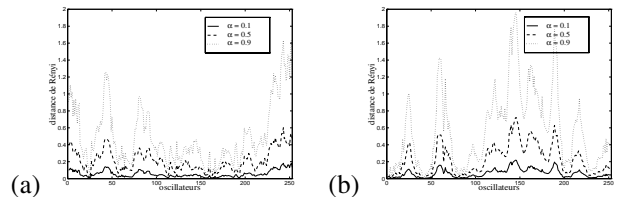


FIG. 3 — Distances de Rényi ( $\alpha = 0.1$ ,  $\alpha = 0.5$  et  $\alpha = 0.9$ ) entre un système  $S_i$  de référence constitué des deux oscillateurs voisins  $i$  et  $i + 1$  ( $S_{25}$  (a) et  $S_{42}$  (b)) et les autres systèmes  $S_j$  qu’il est possible de former à l’aide de deux oscillateurs successifs dans le cas du potentiel de Morse.

Nous proposons de reconstruire la dynamique spatiale locale (au sens où seul un ensemble de  $n$  oscillateurs adjacents est considéré simultanément), à partir des observations

$$S_q = \mathbf{X}_i = [x_{q,i}, \dots, x_{q+n-1,i}]^T, \text{ pour } i = 1, \dots, N$$

$x_{q,i}$  étant la position de l’oscillateur  $q$  à la date  $i$ . L’ensemble des dynamiques locales  $S_q$  constitue un ensemble d’espaces de états dont on souhaite comparer les dynamiques entre elles. Après avoir choisi une valeur  $q_{ref}$  de  $q$  définissant une référence, l’évolution des distances  $d(S_{q_{ref}}, S_q)$  en fonction de  $q$  permet de mesurer les écarts de comportements dynamiques entre la référence  $q_{ref}$  et les différents systèmes.

La figure (2) illustre l’intérêt de l’approche proposée : dans le cas d’un potentiel  $V(x)$  harmonique, où la dynamique des oscillateurs est bien connue et ne révèle aucune structure locale particulière, la fonction distance  $d(S_1, S_q) = f(q)$  est

<sup>1</sup>Données numériques fournies par Pr. M. Peyrard, Laboratoire de Physique, ENS-Lyon.

constante, égale à zéro. Dans le cas du potentiel de Morse,  $d(S_1, S_q)$  présente de fortes variations, révélant l'existence de dynamiques différentes en différents endroits de la chaîne d'oscillateurs : cette inhomogénéité conduit au partage du champ spatio-temporel en différentes zones suivant la similitude du comportement dynamique.

La figure 3 montre que ces résultats sont indépendant du choix de la référence choisie,  $S_1$  dans le cas de la figure 2.

**Remerciements** Nous souhaitons remercier M. Peyrard pour les données qu'il nous a fournies ainsi que pour les discussions stimulantes et enrichissantes concernant leur analyse.

## Références

- [1] O. Michel, A.O. Hero : "Tree-Structured Non-Linear Signal Modeling and Prediction.", *IEEE ICASSP'95*, Detroit, USA.
- [2] A.-E. Badel, O. Michel, A.O. Hero : "Arbres de Régression : Modélisation Non Paramétrique et Analyse des Séries Temporelles.", à paraître dans *Traitement du Signal*, 1997.
- [3] F. Takens : "Detecting Strange Attractors in Turbulence", *Lecture Notes in Mathematics*, vol. 898, pp.366-381, 1981.
- [4] M. Basseville : "Distance Measures for Signal Processing and Pattern Recognition.", *Signal Processing*, vol. 18, pp.349-369, 1989.
- [5] T. Nemetz : "Information Type Measures and their Applications to Finite Decision Problems", *Carleton Mathematical Lecture Notes*, no 17, Carleton University, Ottawa, 1977.
- [6] I. Csiszár : "Information-Type Measures of Divergence of Probability Distributions and Indirect Observations", *Studia Sci. Math. Hung.*, vol. 2, pp. 299-318, 1967.
- [7] H. Whitney : "Differentiable Manifolds", *Annals of Math.*, vol. 37, no.3, pp.645-680, 1936.