ARBRES DE RÉGRESSION POUR L'ANALYSE DE SÉRIES CHAOTIQUES

Anne-Emmanuelle Badel, □ *Olivier Michel* □
et Alfred Hero ◇

□ Laboratoire de Physique (URA 1325 CNRS)
Ecole 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

L'analyse non paramétrique des données (ou des séries) issues de modèles statistiques non additifs a été introduite par Sonquist et Morgan ([1]) puis reprise par Breiman, Friedman, Olshen et Stone ([2]). Nous présentons ici une méthode de prédiction non linéaire reposant sur une partition adaptative de l'espace des phases (reconstruit) du système étudié. La partition est obtenue par une méthode de segmentation récursive de l'espace, dirigée par un critère de maximum d'entropie. Nous illustrons l'intérêt de cette approche pour la prédiction de signaux chaotiques, pour lesquels les algorithmes de prédiction linéaire ne sont pas satisfaisants. Nous montrons que l'erreur de prédiction peut être utilisée pour déterminer les paramètres de reconstruction de l'espace des phases. Les résultats obtenus sont comparés à ceux issus d'autres tests.

Abstract

Non-parametric analysis for non-additive statistical data was introduced by Sonquist et Morgan ([1]) and developped by Breiman, Friedman, Olshen et Stone ([2]). This paper presents a non-linear prediction method based on an adaptative partition of the (reconstructed) phase space of the system under study. The partition is obtained from a recursive tiling of the phase space according to a maximum entropy principle. The attractive feature of this approach is illustrated by the prediction of chaotic time series for which the linear prediction algorithms are not efficient enough. We show that the prediction error can be used to determine the embedding reconstruction parameters. The obtained results are compared with results from other tests.

1. DESCRIPTION DE L'APPROCHE PAR LES ARBRES DE RÉGRESSION

Soit $x_k, k = 1, \ldots, N$ une série temporelle échantillonnée; l'espace des phases du système est reconstruit par la méthode des retards de Takens ([3] et [4]): on construit les vecteurs

$$\mathbf{X}_k = [x_k, x_{k-\tau}, \ldots, x_{k-\tau(d-1)}]^t$$

où d et τ sont respectivement la dimension et le retard de reconstruction. Les vecteurs $\mathbf{X}_k$ seront dans la suite considérés comme les réalisations d'un processus aléatoire stationnaire, de loi de probabilité P.

On définit les quantités suivantes:

- $\Pi = \{\pi_1, \ldots, \pi_L\}$ une partition de cet espace des phases en L cellules complémentaires
- $\{q_1, \ldots, q_L\}$ un ensemble de points représentatifs de chacune des cellules de la partition Π
- $Q : Q(\mathbf{X}) = q_L$ si $\mathbf{X} \in \pi_l$

À partir de ces définitions, on peut dire que la quantité $\mathbf{X}_q(k) \stackrel{\text{def}}{=} Q(\mathbf{X}_k)$ est une quantification de $\mathbf{X}_k$.

Notons que la fonction de densité de probabilité discrète $P_{\mathrm{X}}(q_j), j = 1, \ldots, L$ est égale à l'histogramme théorique $\left[P_{\mathrm{X}_k \in \pi_j}, j = 1, \ldots, L\right]$, de $\mathbf{X}$. La méthode des arbres de régression consiste à utiliser une procédure récursive pour obtenir une partition toujours plus dense $\Pi^l = \{\pi_1^l, \ldots, \pi_{L_l}^l\}$ de $\mathbf{R}^d$, de façon à minimiser (itérativement) la distorsion entre $\mathbf{X}_k$ et sa valeur quantifiée. Nous nous limiterons ici au cas de cellules rectangulaires dans $\mathbf{R}^p$.

Considérons une partition Π^l, de profondeur l, de l'espace des phases en L_l cellules. Cette partition peut être affinée en segmentant à leur tour les cellules π^l en 2^d sous-cellules (que nous appelerons *enfants* par la suite), selon un critère discuté plus loin. La distribution échantillonnée correspondant à l'ensemble des enfants d'une cellule n'est retenue que si elle s'écarte significativement d'une distribution uniforme. La mesure de distance retenue entre les distributions est basée sur un test de Chi-deux à 2^d degrés de liberté. Si la partition n'est pas retenue, la cellule π^l est alors considérée comme une feuille terminale de l'arbre. Voir la figure(1) pour une illustration de cette construction.

Choix du critère de partitionnement
Soient

- H_0 : hypothèse de distribution uniforme sur une cellule pour des vecteurs $\mathbf{X}$ à composantes i.i.d.

- N_l le nombre de réalisations dans la cellule π_l

Le nombre de bits nécessaires pour coder les N_l réalisations de la cellule π_l peut sous hypothèse H_0 se réécrire sous la forme

$$S_{\pi_l} = \sum_{\pi_{l+1}} S_{\pi_{l+1}} - \sum_{j=1}^{card(\pi_{l+1})} p_j log_2 p_j$$

où p_j est la probabilité de se trouver dans la cellule indicée par j. La distribution qui maximalise l'entropie de Shannon étant la distribution uniforme, c'est par rapport à cette distribution que seront comparées les distributions associées à la segmentation d'une cellule *parent*. Le gain d'information lié à la partition en k cellules vaut

$$\Delta = log_2 k + \sum_{i=1}^{k} p_i log_2 p_i$$

et s'annule dans le cas où l'hypothèse H_0 est effectivement vérifiée. Ainsi lorsque la distribution est localement uniforme, aucune structure intéressante n'est décelée: il n'est pas utile de poursuivre la segmentation de la cellule. Une cellule peut aussi être considérée comme une feuille terminale si elle ne contient plus assez de points pour que les tests statistiques utilisés demeurent pertinents.

Du fait que seules des cellules rectangulaires sont considérées, la partition est obtenue en appliquant un seuil pour chaque cellule *parent* suivant chacune des d dimensions de l'espace. On montre que le critère $\Delta = 0$ sous hypothèse H_0 conduit à choisir comme seuil le médian échantillonné ([5] et [6]). Il s'agit d'un estimateur asymptotiquement non biaisé du médian vrai et la distribution obtenue sur les enfants est au maximum d'entropie si les coordonnées de $\mathbf{X}$ sont indépendantes. D'autre part, cette règle de partition est optimale au sens de la minimalisation du taux de distorsion moyen de la quantification vectorielle résultante ([5] et [6]).

2. APPLICATION À LA PRÉDICTION

Soit un système dynamique de dimension d décrit par l'équation d'état: $\mathbf{X}_{n+1} = F_{\mathbf{X}_n}(\mathbf{X_n}) + \epsilon_\mathbf{n}$ où F_{X_n} est une fonction non-linéaire, elle même dépendant de $\mathbf{X}_n$. ϵ_n peut être considéré comme un bruit d'observation ou une perturbation de l'état déterministe. La partition de l'espace des phases (ou de façon équivalente l'arbre de régression) est estimée à partir d'une série d'apprentissage de N données. À chaque cellule ainsi créée est associée une valeur de sortie quantifiée par exemple selon:

$$Q(\mathbf{X}_k) = \text{médian}(\mathbf{X}(j+1)|\mathbf{X}(j) \in \pi_l)$$

et $\hat{\mathbf{X}}(k+1) = Q(\mathbf{X})$ Ce choix de $Q(\mathbf{X})$ permet de réaliser de la prédiction avec le médian des valeurs futures (après un temps τ) des points de la série d'apprentissage. On obtient ainsi une prédiction quantifiée sur une structure d'arbres optimale au sens du maximum d'entropie de la distorsion pour la quantification vectorielle obtenue ([5] et [6]).

La figure (2 a) présente l'arbre obtenu en dimension 3 pour un retard $\tau = 4$ en pas d'échantillonnage pour un système chaotique expérimental ([7]), ainsi que la prédiction (figure 2 b) à un pas (τ) obtenue à partir de cet arbre. Le manque de points pour avoir des tests statistiques pertinents (l'une des deux causes d'obtention d'une feuille terminale lors de la construction de l'arbre) permet de satisfaire le compromis nécessaire à une bonne prédiction: plus la taille moyenne des cellules est faible, meilleure seraient a priori l'estimation du médian et par conséquent la prédiction mais du fait du nombre fini des données, il faut une taille suffisante des cellules pour conserver la pertinence statistique de l'estimation.

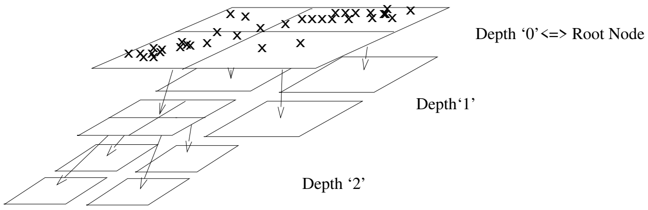

Figure 1: Pour un espace des phases reconstruit de dimension $d = 2$, le noeud initial est subdivisé en 4 sous-ensembles et cette distribution est jugée non uniforme: on poursuit la subdivision de chaque cellule. Parmi les subdivisions qui sont réalisées au niveau suivant, seule celle du coin gauche est jugée non uniforme et subdivisée à nouveau.

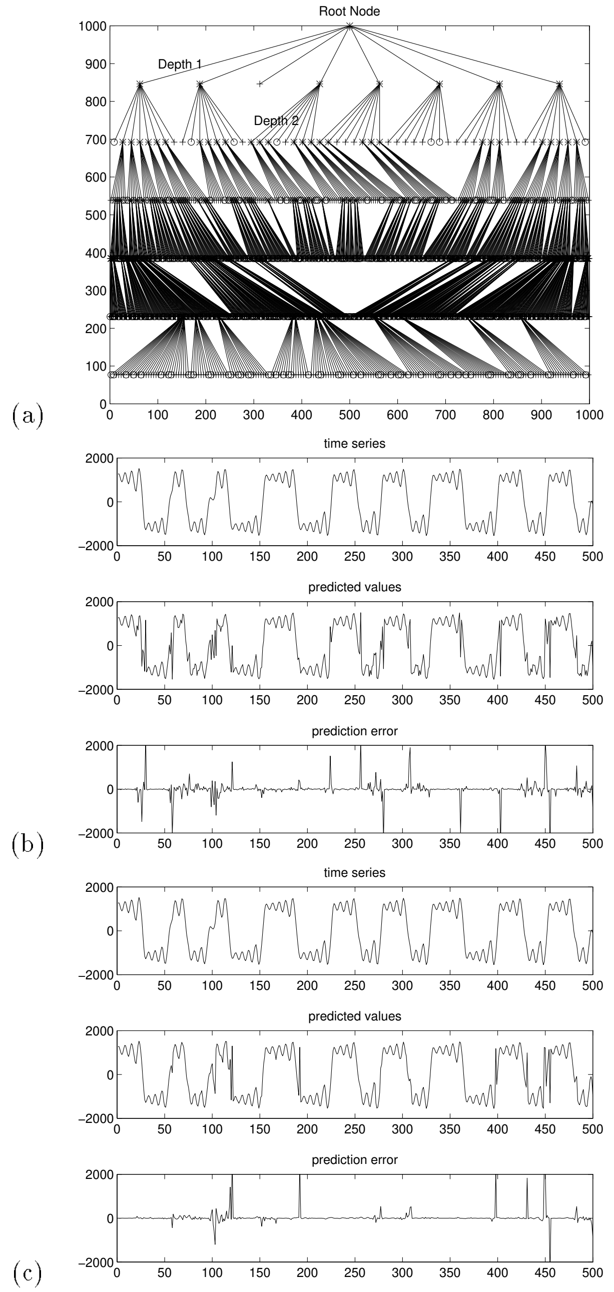

Figure 2: (a): Arbre estimé en dimension 3 et avec un retard de 4 en pas d'échantillonnage pour le système expérimental; prédiction à un pas et erreur de prédiction au médian à partir de cet arbre estimé (b) et par la méthode des plus proches voisins (c).

On montre que les performances obtenues en prédiction sont comparables à celles obtenues par l'algorithme classique de prédiction par la méthode de plus proches voisins ([8]). La figure (2 c) présente les résultats obtenus avec cette méthode pour la même reconstruction de l'espace des phases que pour la figure (2 b). Dans les deux cas, la série d'apprentissage comprend 8192 points.

3. ESTIMATION DE τ ET D À PARTIR DE L'ERREUR DE PRÉDICTION

La prédiction basée sur les arbres de régression peut être réalisée pour différentes valeurs de la dimension et du retard de reconstruction. L'étude de la variance de l'erreur de prédiction en fonction des valeurs du retard et de la dimension de reconstruction présente un minimum qui permet de choisir les paramètres de reconstruction τ et d. La variance V présentée ici est normalisée par l'énergie du signal:

$$V = \frac{\sum_{i=1}^{p}(e_i - < e_i >_i)^2}{\sum_{i=1}^{p}(x_i - < x_i >_i)^2}$$

où p est le nombre de prédictions à un pas réalisées (ici $p = 500$), e_i l'erreur de prédiction pour la $i^{ème}$ prédiction et x_i les valeurs de la série étudiée.

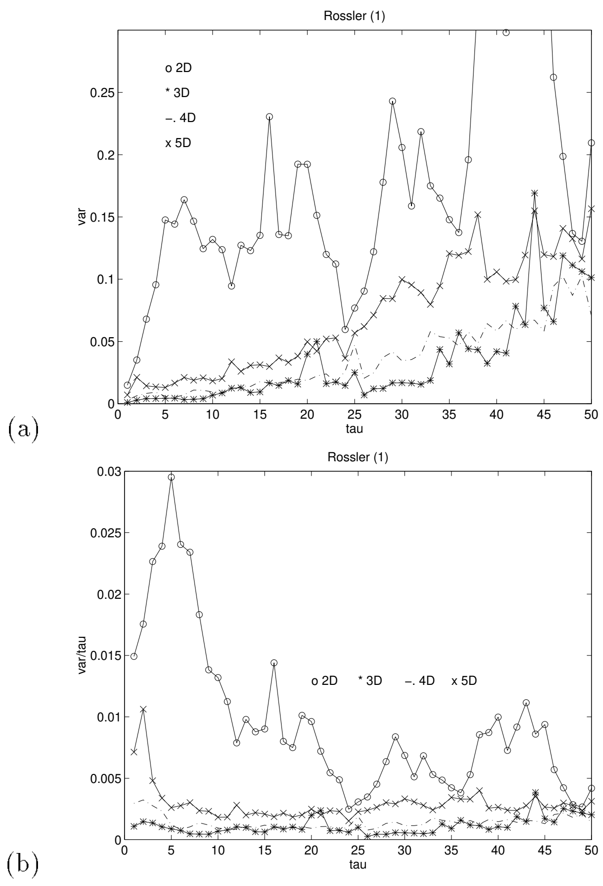

Figure 3: Variance d'erreur de prédiction normalisée par l'énergie du signal pour le système de Rössler non normalisée (a) et "normalisée" (b) par τ en utilisant la totalité des points disponibles.

L'algorithme proposé estime la variance d'erreur de prédiction à un pas en fonction de la dimension et du retard de reconstruction: les paramètres de reconstruction seront choisis de façon à minimiser cette variance d'erreur "normalisée" par la valeur du

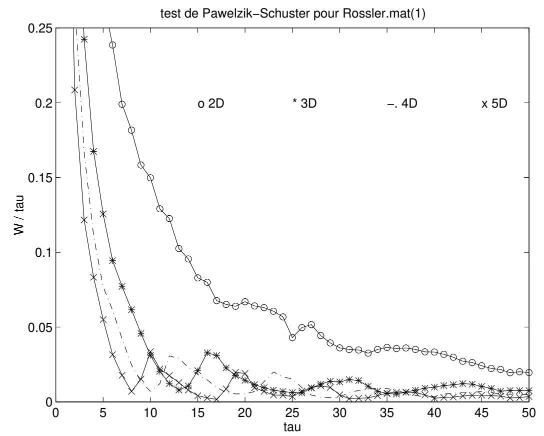

Figure 4: Test de Pawelzik-Schuster pour le système de Rössler

retard τ: $\frac{V}{\tau}$ afin de s'affranchir de la dépendance linéaire pour de faibles valeurs du retard.

Ces résultats sont comparés avec ceux proposés par Liebert, Pawelzik et Schuster ([9]): les paramètres de reconstruction sont choisis pour minimiser:

$$W = ln < (\prod_{k=1}^{p} \frac{dist^{\tau}_{d+1}(i,j(k,d))dist^{\tau}_{d}(i,j(k,d+1))}{dist^{\tau}_{d+1}(i,j(k,d+1))dist^{\tau}_{d}(i,j(k,d))})^{\frac{1}{2p}} >_i$$

où $dist^{\tau}_{m}(i,j(k,n))$ est la distance en dimension m du $k^{ième}$ plus proche voisin en dimension n.

Cette quantité permet d'évaluer le degré de violation des propriétés topologiques quand la dimension de reconstruction augmente d'une unité: lorsque les paramètres de reconstruction sont obtenus, les voisinages d'un point sont conservés en passant d'une dimension à la suivante. W/τ sera d'autant plus proche de 0 que les propriétés topologiques seront bonnes et donc que les paramètres seront optimaux, la "normalisation" par τ est due à la même volonté que dans le cas de la variance d'erreur de s'affranchir de la dépendance linéaire aux faibles valeurs de τ.
Les figures (3a et b) d'une part et (4) d'autre part montrent ce que donne chacun de ces algorithmes pour le système de Rössler synthétisé. La variance d'erreur (non "normalisée" par τ) est petite pour de faibles valeurs de τ du fait qu'alors les fortes corrélations entre composantes donnent lieu à une prédiction quasi-linéaire; à l'inverse pour des valeurs importantes de τ, la forte décorrélation conduit à des variances se rapprochant de l'énergie du signal. Quant à l'influence de la dimension d, la variance présente un minimum en fonction de d. Cette détermination des paramètres à l'aide de la variance donne pour Rössler: $\tau = 8$ et $d = 3$.

En ce qui concerne les résultats du test de Pawelzik, Liebert et Schuster, leur algorithme tend à préconiser: $\tau = 17$ et $d = 5$. La méthode de la variance d'erreur fournit donc des résultats conformes à ce qui avait été remarqué dans ([10]). Outre les performances de l'approche par les arbres en terme de prédiction non linéaire, ces résultats montrent que cette approche permet de déterminer les paramètres de reconstruction de l'espace des phases avec des performances au moins comparables à celles des méthodes usuelles, pour un coût de calcul moindre. Le comportement statistique de la quantité V fera l'objet de travaux futurs.

Références

[1] J.N. Sonquist and J.N. Morgan, "The detection of interaction effects," Monograph 35, Survey Research Center, Institute for Social Research, University of Michigan, 1964.

[2] L. Breiman, J.H.Friedman, R.A. Olshen, C.J. Stone, "Classification and Regression Trees," Wadsworth Advanced Books and Software, 1984.

[3] J.P. Eckmann and D. Ruelle, "Ergodic Theory of Chaos and Strange Attractors," *Rev. Mod. Phys.*, Vol. 57, No. 3, pp. 617–656, 1985.

[4] A.M. Fraser, "Information and Entropy in Strange Attractors," *IEEE Trans on Inf. theory*, vol.35, No.2, 1989, pp 245-262.

[5] O. Michel, A. Hero, "Tree Structured non-linear signal modeling and prediction," *ICASSP*, 1995.

[6] O. Michel, A. Hero, "Tree-based modeling of non-linear processes for prediction, detection, and classification," technical report in preparation, University of Michigan, 1995.

[7] T.P. Weldon, "An Inductorless Double-Scroll Chaotic Circuit," *Am. J. Phys.*, vol.58,No.10, pp. 936-941, 1990.

[8] J.D. Farmer, J.J. Sidorowitch, "Exploiting Chaos to Predict the Future and Reduce Noise," *Evolution, Learning and Cognition*, ed. Lee, Y.C. (World Scientific).

[9] W. Liebert, K. Pawelzik, H.G. Schuster, "Optimal Embeddings of Chaotic Attractors from Topological Considerations," *Europhysics Letters*, vol.14, no.6, pp 521-6, march 1991.

[10] O. Michel, P. Flandrin, "Higher Order Statistics for Chaotic Signal Analysis," in Digital Signal Processing Techniques and Applications, Academic Press, to be published in 1995.