

COMPARISON OF SEVERAL METHODS TO PREDICT CHAOTIC TIME SERIES

Anne-Emmanuelle Badel $\diamond$

Dominique Guégan $\triangle^+$

Ludovic Mercier $\triangle^*$

Olivier Michel $\diamond$

$\diamond$ Laboratoire de Physique, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France,

$\triangle$ ENSAE-CREST, (+ Timbre J120, * Timbre J340), 3 avenue Pierre Larousse, 92245 Malakoff Cedex, France

e-mail : aebadel@physique.ens-lyon.fr, guegan@ensae.fr, mercier@ensae.fr, omichel@physique.ens-lyon.fr

ABSTRACT

The aim of this paper is to compare different prediction methods for chaotic deterministic systems. We consider three different methods to evaluate the dynamics of the systems : the Nearest Neighbors, the Radial Basis Functions and the Regression Tree. We use a comparison criterion suited to chaotic systems : the prediction horizon. The optimal prediction horizon is discussed with respect to the sampling time step. We apply these methods to simulated chaotic system (Lorenz system), experimental chaotic system (Double-Scroll) and to intra-day series of exchange rates, namely DEM/FRF. We provide developments concerning the choice of the parameters involved in chaotic time series prediction.

1. INTRODUCTION

An interesting issue in the time series literature is the extent to which chaotic time series are predictable. We consider here sampled time series x_n (with a time step h) stemming from a continuous non linear dynamical system, defined by :

$$\begin{cases} \mathbf{X}_0 \in \mathbb{R}^m & \text{some initial conditions} \\ \mathbf{X}_{n+1} = \phi_h(\mathbf{X}_n) \end{cases} \quad (1)$$

where $\mathbf{X}_n \in \mathbb{R}^m$ stands for the current state of the system at time n in the phase space and ϕ the non linear function of the current state. A non linear deterministic dynamical system will be called chaotic [3] if the system is ergodic and very sensitive to initial conditions. This latter property makes medium term prediction impossible, but does not preclude any possibility of short term prediction¹. The purpose of this paper is to compare three non parametric methods for short term prediction of chaotic time series : Nearest Neighbors (NN), Radial Basis Functions (RBF) and Regression Tree (RT). There exists several criteria for comparing the quality of prediction, see e.g. [9, 2]. The adopted criterion in this paper is based on the horizon prediction Hh defined as the date at which the prediction errors departs from the actual value from more than a given distance related to the dynamics of the system. h stands for the time sampling step, and H is the number of those sampling steps that are predictable from one starting point.

The performances and reliability of the prediction methods are investigated for three kinds of series : a computer simulated Lorenz chaotic system, an electronic Double-Scroll experiment and an intra-day exchange rates time series. Although the chaotic

¹ Long term prediction may sometimes be obtained, given that the invariant measure of the system is known.

nature of these latter financial data is still an open problem, their complex behavior makes them both challenging and interesting to trial within the context of chaotic systems studies.

The remainder of this paper is organized as follows. The prediction methods and the comparison criterion that we used, are the purpose of Section 2. Section 3 details the time series used for testing the different methods. The reconstruction of the phase space by the Takens's embedding method for experimental time series is also recalled and a summary of the performed comparisons is presented. Section 4 is devoted to describe some problems related to the presence of additive observation noise. Results on the robustness of the prediction algorithm are presented and a link between prediction horizon and reconstruction parameters is proposed.

2. PREDICTION METHODS

A method to forecast chaotic time series consists in reconstructing the function ϕ_h in (1) by a certain $\hat{\phi}_h$. The predicted values $\hat{\mathbf{X}}_{N+1}, \dots, \hat{\mathbf{X}}_{N+H}$ are given by :

$$\begin{cases} \hat{\mathbf{X}}_{N+1} = \hat{\phi}_h(\mathbf{X}_N) \\ \hat{\mathbf{X}}_{N+i+1} = \hat{\phi}_h(\hat{\mathbf{X}}_{N+i}) \end{cases} \quad (2)$$

Only iterated predictions will be considered in this paper. We focus our attention upon three non parametric approaches: the celebrated Nearest Neighbor method (NN) which corresponds to a local approach, a semi-local method relying on the use of radial basis functions (RBF), and regression tree based approach making use of the existence of an invariant measure for the system.

2.1. Nearest neighbors (NN)

This method [1] is local and relatively easy to implement. If only one neighbor is used, we search among the learning set $\mathbf{X}_0^l, \dots, \mathbf{X}_{N-1}^l$ the point $\mathbf{X}_i^l$ which is the closest to $\mathbf{X}_N$ in the sense of a certain norm (for example, the Euclidean norm) then we define :

$$\hat{\phi}_h(\mathbf{X}_N) = \mathbf{X}_{i+1}^l. \quad (3)$$

It is possible to use more than one neighbor, to take more than one past value for each one or to consider various weights. Since the dynamics is assumed to be continuous in the phase space, we use K neighbors $\mathbf{X}_1^l, \dots, \mathbf{X}_K^l$ of $\mathbf{X}_N$, a weighting function w and a norm $\|\cdot\|$ of $\mathbb{R}^m$. Then we get :

$$\hat{\phi}_h(\mathbf{X}_N) = \sum_{k=1}^K w(\|\mathbf{X}_N - \mathbf{X}_k^l\|) \mathbf{X}_{k+1}^l. \quad (4)$$

Following B. Finkenstädt and P. Kuhbier [4], we assume that the importance of each neighbor is proportional to the exponential of its distance to $\underline{\mathbf{X}}_N$:

$$\hat{\phi}_h(\underline{\mathbf{X}}_N) = \frac{\sum_{k=1}^K e^{-(\|\underline{\mathbf{X}}_N - \underline{\mathbf{X}}_{i_k}^t\|)} \underline{\mathbf{X}}_{i_k}^t}{\sum_{k'=1}^K e^{-(\|\underline{\mathbf{X}}_N - \underline{\mathbf{X}}_{i_{k'}}^t\|)}} \quad (5)$$

2.2. Radial basis functions (RBF)

This semi-local method [1] approaches ϕ_h with a set of radial basis functions by $\hat{\phi}_h$ in the following way :

$$\hat{\phi}_h(\underline{\mathbf{X}}_N) = \sum_{c=1}^C \lambda_c w_c(\|\underline{\mathbf{X}}_N - \underline{\mathbf{Y}}_c\|). \quad (6)$$

where $\underline{\mathbf{Y}}_c$ are the centers of the radial functions which are chosen via a clustering method over the past values of $\underline{\mathbf{X}}_i$. The functions w_c are arbitrarily chosen radial functions like Gaussian or multi-quadratic function ; here we use multi-quadratic functions for computational efficiency :

$$w_c(r) = \frac{1}{\sqrt{r^2 + r_c^2}}, \quad (7)$$

where r_c is the radius of the centers $\underline{\mathbf{Y}}_c$. Their number is arbitrary but always smaller than N because we do not want to interpolate ϕ but to approximate it. The parameters λ_c are estimated using a simple regression on the past values once the centers have been chosen.

2.3. Regression trees (RT)

The driving idea is to estimate the probability of having an observation in a neighborhood (defined in the phase-space) of a state vector at a given time . This amounts to construct a partition of the phase space in order to get some constant probability within each cell: thus an m-dimensional histogram [7] (for an m-dimensional phase space) is obtained. The proposed construction is recursive and the derived structure can be represented by a tree : each non-uniform cell of the phase space is splitted in 2^m cells by choosing the marginal median on each axis. The uniformity of the estimated probability distribution within a cell is tested by the χ^2 test : if $\{p_i\}$ represents the set of empirical probabilities of a sub-partition (into $\frac{1}{2^m}$ sub-cells) of the cell under study, then the test measures the departure of $\{p_i\}$ from the uniform distribution. Since the aim is to obtain an uniform distribution, the stopping rule is the following : if the distribution is said uniform by the χ^2 test, the subdivision is stopped, otherwise, the splits are memorized, together with the "children cells" and the partition is iterated further. To achieve some prediction from this approach, the partition has to be constructed from a learning set and we affect a prediction value to each cell of the final partition : e.g. median or mean of future values of the set of points (from the learning set) belonging to this cell (see e.g. [8]).

3. APPLICATION TO TIME SERIES

3.1. The simulated Lorenz attractor

The continuous time representation for the Lorenz attractor we choose is :

$$\begin{cases} \frac{dx}{dt} = \sigma(y - x) \\ \frac{dy}{dt} = -xz + rx - y \\ \frac{dz}{dt} = xy - bz \end{cases} \quad (8)$$

Here we take the parameter values : $\sigma = 16$, $R = 45.92$, $b = 4$.

The prediction experiment has been performed from a learning set of 5000 consecutive points of the numerically integrated system (8). The horizon is estimated from iterated one-step forward prediction (i.e. the prediction length is set to the sampling step h^2) as depicted in Section 2. This experiment is reproduced 500 times, from times series obtained with different initial conditions, and the results are averaged over all the experiments to provide an estimate of Hh . The 4800 first points are used to form the learning set. The 200 next points will be used as reference values for estimating the prediction error, as the predictions are performed from the 4801-st point.

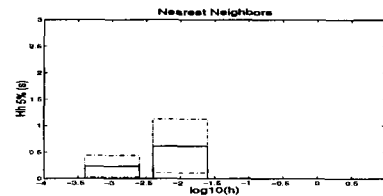


Figure 1. 5% prediction horizon Hh for the Lorenz attractor with the 3-NN method for sampling steps $h = 0.001$, $h = 0.01$, $h = 0.1$ and $h = 1$. Hh is estimated as the mean of 500 prediction horizons performed with a learning set of 4800 points. The dashed lines show the prediction horizon standard error.

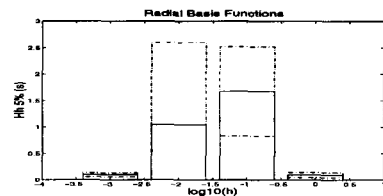


Figure 2. 5% prediction horizon Hh for the Lorenz attractor with the RBF method (100 centers) for sampling steps $h = 0.001$, $h = 0.01$, $h = 0.1$ and $h = 1$. Hh is estimated as the mean of 500 prediction horizons performed with a learning set of 4800 points. The dashed lines show the prediction horizon standard error.

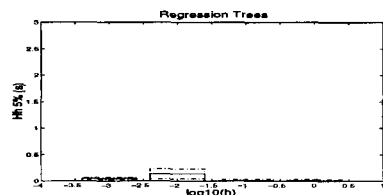


Figure 3. 5% prediction horizon Hh for the Lorenz attractor with the RT method for sampling steps $h = 0.001$, $h = 0.01$, $h = 0.1$ and $h = 1$. Hh is estimated as the mean of 500 prediction horizons performed with a learning set of 4800 points. The dashed lines show the prediction horizon standard error.

The figures (1), (2) and (3) show respectively for the NN, RBF and RT methods the prediction horizon at 1% (a) and 5% (b) accu-

²The sampling step is chosen as an integer multiple of the integration step, for sake of convenience

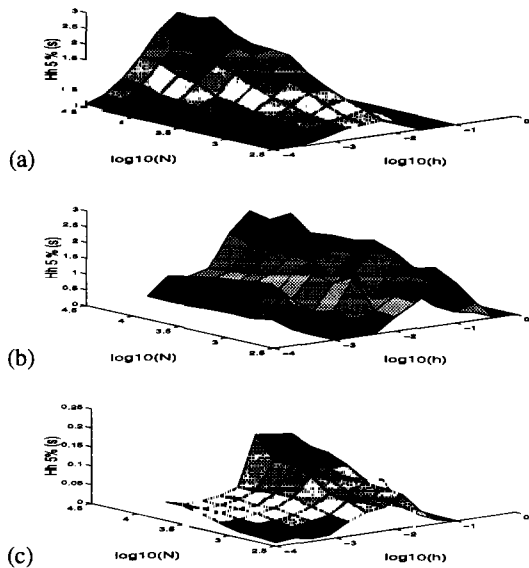


Figure 4. "Horizon Map" : evolution of prediction horizon (in seconds) with the number N of points of the learning set and the sampling step h for the Lorenz attractor with the following methods : NN (a), RBF (b) and RT (c).

racy estimated as the mean of 500 prediction experiments ; standard errors of the results are plotted in dashed lines.

It is interesting to study [5] the evolution of the prediction horizon with the number N of points that are used to build the learning set and with the sampling step h . On figure (4) the horizon-map obtained for varying N and h is presented, as obtained for the Lorenz attractor, by the three above mentioned methods. These graphs show the increase of the prediction horizon with the number of points in the learning set. It also evidences the existence of an optimal step h for which the prediction horizon is maximum.

3.2. Electronic Double-Scroll Experiment

The time series that we consider was recorded on an experimental electronic setup realizing a chaotic oscillator of the Chua's family [11]. The circuit admitted the well known Double-Scroll equation :

$$\begin{cases} \frac{dx}{dt} = \alpha(y - \gamma) \\ \frac{dy}{dt} = x - y + z \\ \frac{dz}{dt} = -\beta y \end{cases} \quad (9)$$

where

$$\gamma = \begin{cases} m_1 x + m_0 - m_1 & \text{if } x > 1 \\ m_0 x & \text{if } -1 < x < 1 \\ m_1 x - m_0 - m_1 & \text{if } x < -1 \end{cases} \quad (10)$$

with the parameter values : $\alpha = 9$, $\beta = \frac{100}{7}$, $m_0 = -\frac{1}{7}$ and $m_1 = \frac{2}{7}$. The sampling time for recording the series was $h = 69,4 \mu s$.

For the case of the Double-Scroll experiment, only one time series was recorded. As a consequence, the prediction horizon was computed from a phase space reconstruction based on Takens's method [10]. However, when dealing with finite duration observations (finite time series) the choice of both the time delay τ and the embedding dimension m is of major incidence on the accuracy of

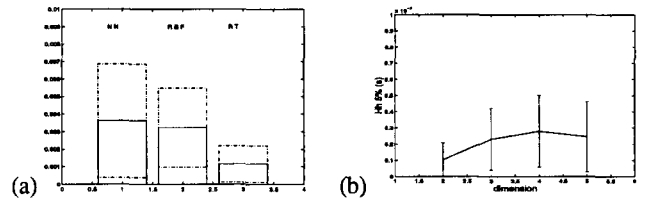


Figure 5. 5% prediction horizon for the the experimental Double-Scroll series for the three methods (a) : NN, RBF and RT with a 3-dimensional reconstruction and evolution of 5% prediction horizon with the value of reconstruction dimension for the RT method (b).

the reconstruction. Here, τ is set to match the sampling step h and $m = 2, 3, 4, 5$ (see e.g. [6], for discussion on the topic of setting these parameters).

The following figure (5) illustrates the behavior of the 5% prediction horizon with the three methods : NN, RBF and RT (a). The influence of the value of reconstruction dimension is illustrated in figure (5(a)).

3.3. Economical intra-day data

We have selected the Deutsch mark/French Franc exchange rates. Our empirical investigation covers the period between 01/09/94 and 31/08/95 with 420 134 data. The series we consider is the mean of bid and ask quotes without the outliers found during the first step. The same procedure as for the previous experimental data is used to reconstruct the dynamics from the single series we have.

Figure (6) shows the same kind of results as figure (5) for the Double Scroll experimental system when τ equals 600 seconds : 5% prediction horizon with the three methods : NN, RBF and RT (a) and evolution of 5% prediction horizon with the reconstruction dimension for the RT method.

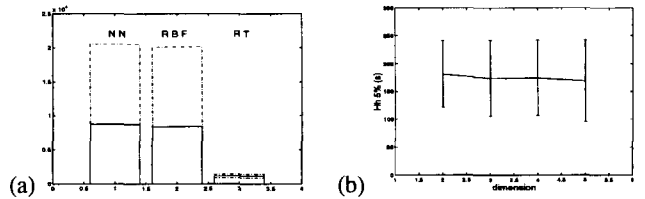


Figure 6. 5% prediction horizon for the DEM/FRF intra-day series for the three methods (a) : NN, RBF and RT with a 3-dimensional reconstruction and evolution of 5% prediction horizon with the value of reconstruction dimension for the RT method (b).

4. DISCUSSION OF THE RESULTS

4.1. Behavior of the methods in the presence of additive noise

Experimental data y_n are in general corrupted by noise. Here we restrict our study to the case of additive observation noise, i.e. to the case where the perturbations do not take part in the dynamical evolution of the system; such noise intervenes only in the observation equation of the dynamical system.

Let ϵ_n be an m -dimensional noise. Thus the dynamics of the experimental data are following the equations $\begin{cases} \mathbf{X}_{n+1} = \phi_h(\mathbf{X}_n) \\ y_n = x_n + \epsilon_n \end{cases}$. where x_n is a coordinate of the state vector $\mathbf{X}_n$.

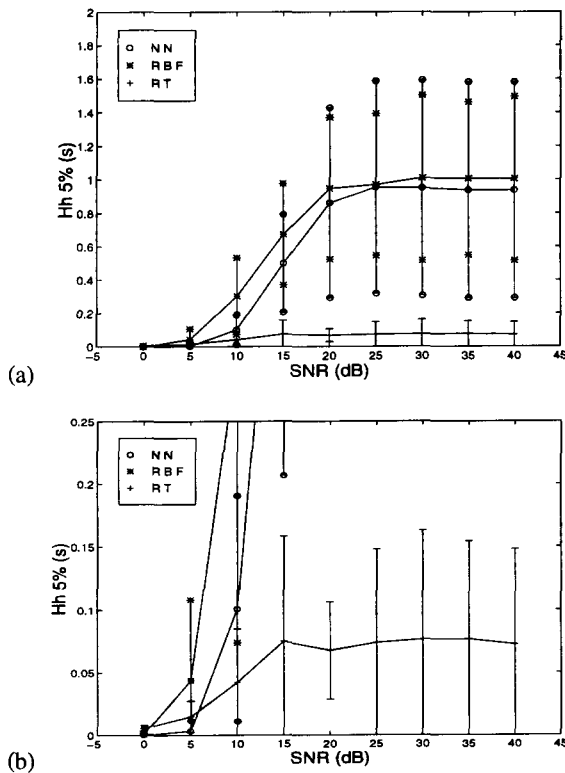


Figure 7. Influence of the noise on the 5% prediction horizon with the Lorenz attractor for the three methods. (b) shows details for low prediction horizon.

For testing whether the proposed prediction methods perform reasonably in the presence of additive noise, the initial data sets have been corrupted by an additive white zero-mean Gaussian noise (WGN). We adopted the following definition for the signal to noise ratio (SNR) :

$$SNR = 10 \log_{10} \frac{E[\|x_n - E[x_n]\|^2]}{\sigma^2} \quad (11)$$

σ^2 stands for the variance of the WGN. In the following, the influence of noise is illustrated on prediction horizon measurements computed for the corrupted Lorenz attractor, for SNR ranging from 0 to 40 dB, and $h = .01$.

Figures (7) show the influence of additional observation WGN on the 5% prediction horizon with the simulated Lorenz for a SNR of 0 to 40 dB. The three methods seem to be quite robust to noise until different level : 25 dB for NN, dB for RBF and 15 dB for RT.

4.2. Using Hh for estimating embedding parameters

The purpose of this section is to evaluate the advantage of using Hh for estimating the embedding parameters τ and m . Therefore, the prediction horizon Hh is estimated for different values of m . Figure (5(b)) evidences the existence of an optimal (with respect to the accessible horizon length) reconstruction dimension. It is interesting to note that the value $m = 4$ is the same value obtained with a different criterion : the prediction error variance normalized by the signal energy [8]. Figure 8 present the evolution of normalized one step prediction error variance with the reconstruc-

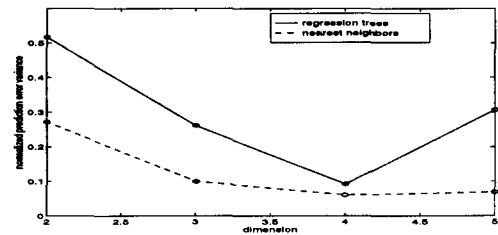


Figure 8. Normalized by the signal energy variance of the one step prediction error for the experimental Double Scroll system with the NN and RT methods

tion dimension for the NN and RT methods with the Double-Scroll experiment.

Acknowledgment. We are grateful to C. Dunis for giving us the data exchange rates that we use in this study.

REFERENCES

- [1] M. Casdagli. Nonlinear prediction in chaotic time series. *Physica D*, 35:335–359, 1989.
- [2] F.X. Diebold and R.S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263, 1995.
- [3] J.P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Review of Modern Physics*, 57(3):617–656, 1985.
- [4] B. Finkenstädt and P. Kuhbier. Forecasting nonlinear economic time series : A simple test to accompany the nearest neighbor approach. *Empirical economics*, 20:243–263, 1995.
- [5] D. Guégan and L. Mercier. Prediction in chaotic time series : Methods and comparisons using simulations. Working Paper INSEE 9616, 1996.
- [6] O. Michel and P. Flandrin. Higher order statistics for chaotic signal analysis. In C.L. Leondes, editor, *Computer Techniques and Algorithms in Digital Signal Processing*, volume 75 of *Control and Dynamic Systems, Advances in Theory and Applications*, pages 105–154. Academic Press, 1996.
- [7] O. Michel and A.O. Hero. Tree-structured non-linear signal modeling and prediction. In *ICASSP'95*, Detroit, Michigan, 1995. IEEE-ICASSP.
- [8] O. Michel, A.O. Hero, A.-E. Badel, and P. Flandrin. Tree based modeling, prediction and analysis of chaotic time-series. In *Proceedings of IEEE Workshop on Non Linear Signal and Image Processing*, volume I, pages 117–120, Halkidiki, Greece, June 1995.
- [9] G. Sugihara and R.M. May. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344:734–741, April 1990.
- [10] F. Takens. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898:366–381, 1981.
- [11] T.P. Weldon. An inductorless double-scroll chaotic circuit. *American Journal of Physics*, 58(10):936–941, 1990.