Local Optimal Metrics and Nonlinear Modeling of Chaotic Time Series

P. García*
Departamento de Física Aplicada, Facultad de Ingeniería, Universidad Central de Venezuela

J. Jiménez† and A. Marcano‡
Grupo de Dinámica de Sistemas Complejos e Inteligencia Artificial, Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela

F. Moleiro
Sección de Cardiologia Experimental, Instituto de Medicina Tropical, Facultad de Medicina, Universidad Central de Venezuela

(Received 19 July 1995)

We consider the problem of prediction and nonlinear modeling for chaotic time series and examine the effects of changing the local metric used to select nearest neighbors in the embedding space of delay register vectors. Analyzing simulated numerical data and real data, it is shown that the fit achieved for the case where the components of the metric tensor are constants over the whole attractor is improved by a proper selection of the local metric. Our results also suggest how deviations from the Euclidean case can be used as a tool to discriminate chaos from correlated noise.

PACS numbers: 05.45.+b

Among the reasons that deterministic nonlinear modeling techniques of complex data series (i.e., uncorrelated series with a flat Fourier spectrum) have received a great deal of attention is that they can be used to forecast, at least in the short term, the evolution of a chaotic system whose underlying dynamic is unknown [1–4]. In addition, from the cross-correlation function between observed values and those predicted through these techniques, it is possible to estimate the largest Lyapunov exponent of the dynamics [5], even for sparse series (those containing only of the order of 10^3 data points). Also, when the fit achieved using nonlinear modeling is better than the one obtained using probabilistic models, it is reasonable to assume that a deterministic mechanism governs the process under study [6–9], providing in this way a tentative criterion to discriminate between chaos and noise.

Most of these techniques can be grouped into two major classes: global and local ones. In global methods, the whole past information is used for predictions about the evolution of the system under study. Clearly, this has the disadvantage that if new information is taken into account, all the parameters of the model may change and then a long parameter estimation time may be required.

The local methods overcome this drawback by utilizing only part of the history. In fact, the basic idea which supports local techniques is that, if a deterministic mechanism governs the evolution of the data series $x_1, x_2, \ldots, x_N$, then, for sufficiently high values of d, any value x_p will be given by

$$x_p = f(x_{p-1}, x_{p-2}, \ldots, x_{p-d}) = f(X_{p-1}), \qquad (1)$$

where f is a continuous rule, and we have introduced the notation $X_i = (x_i, x_i - 1, \ldots, x_{i-d+1})$.

Therefore the future values of the data series x_{N+1}, $x_{N+2}, \ldots, x_{N+t}$ can be approximately predicted from the evolution $x_{N(q)+1}, x_{N(q)+2}, \ldots, x_{N(q)+t}$ of a set $\{X_{N(q)};\ 1 \leq q \leq k\}$ of k vectors which are close enough to the vector X_N.

In order to compare the closeness of a pair of vectors, the standard Euclidean metric is often used. Nevertheless, as has been shown by Murray [10], there are no reasons to choose the Euclidean metric by default.

For instance, let us assume that the metric arises from a metric tensor whose components are

$$g_{ij} = \delta_{ij} e^{2ic}, \qquad (2)$$

where δ_{ij} is the Kronecker delta and i, j run from 1 to d. In Ref. [10] it is shown that if the parameter c is varied to minimize the root mean square error of forecasts, then there is a single global minimum corresponding to a value of c close to the most negative Lyapunov exponent of the dynamics.

Now, in Ref. [10] only uniform metrics through the attractor are considered, and this is an assumption which is equivalent to supposing that metric properties do not change from one point to another on the attractor. As an extension we shall assume that there exists a metric tensor whose components $g_{ij}(X)$ $(1 \leq i, j \leq d)$ depend on the position in the embedding space in such a way that the nearest neighbor $X_{v(p)} = (x_{v(p)}, x_{v(p)-1}, \ldots, x_{v(p)-d+1})$ of a given point $X_p = (x_p, x_{p-1}, \ldots, x_{p-d+1})$ optimizes predictions (at zero order of approximation) one step into the future:

$$d_g(X_p, X_{v(p)}) = \text{minimum}$$
$$\text{when } |x_{p+1} - x_{v(p)+1}| = \text{minimum}, \qquad (3)$$

where $d_g(X_p, X_{v(p)})$ is the distance from X_p to $X_{v(p)}$ in the metric defined by $g_{ij}(X)$.

Naturally, if we search for a vector which gives an absolute minimum for the prediction error one step into the future, the prediction error for t units of time ahead is not necessarily minimized when $t > 1$. In other words, the fulfillment of the condition $|x_{p+1} - x_{v(p)+1}| = \text{minimum}$ does not imply the fulfillment of $|x_{p+t} - x_{v(p)+t}| = \text{minimum}$ for $t > 1$.

In order to have a quantitative indication of this effect, a number of numerical experiments were conducted on several chaotic systems, yielding results with the same general characteristic: If for each vector X_p we search for the vector $X_{v(p)}$ that minimizes the error $|x_{p+1} - x_{v(p)+1}|$ and we restrict the search procedure to the k nearest neighbors of X_p in the Euclidean metric, then there exists a critical value k_c such that the prediction quality t units of time ahead ($t > 1$) gets poorer for $k > k_c$. This can be seen in Fig. 1 where we show the behavior of the centered correlation between the series of predicted values and the observed values

$$C = \frac{\langle (x_{pr} - \langle x_{pr} \rangle)(x_{ob} - \langle x_{ob} \rangle) \rangle}{[\langle (x_{pr} - \langle x_{pr} \rangle)^2 \rangle \langle (x_{ob} - \langle x_{ob} \rangle)^2 \rangle]^{1/2}} \qquad (4)$$

as a function of t. The simulated data were generated by the Hénon map

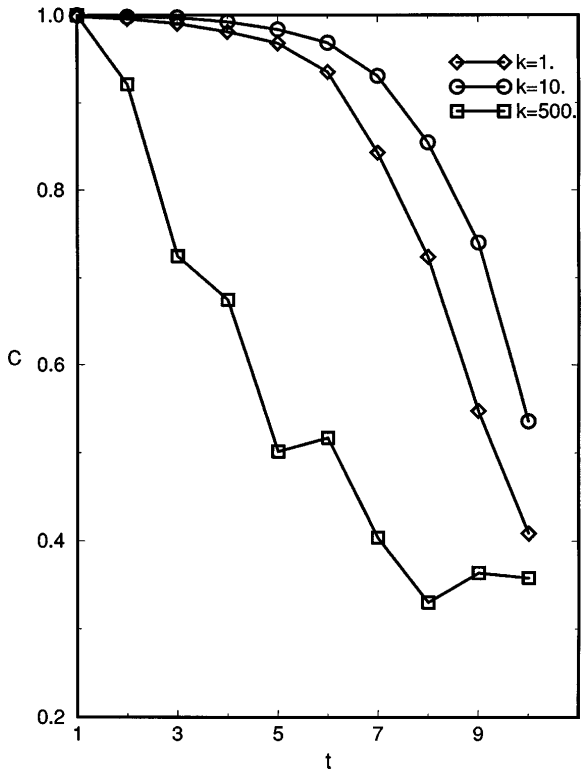

FIG. 1. Correlation function C as a function of the prediction interval t for 500 data points from the Hénon map. In the figure, □ corresponds to the case $k = 500$, where the searching procedure for each vector $X_{v(p)}$ is made on the whole register, o corresponds to the case $k = 10$, and ◇ corresponds to the Euclidean case $k = 1$. In all cases the embedding dimension is 2.

$$x_{n+1} = 1 - 1.4x_n^2 + 0.3x_{n-1}, \qquad (5)$$

and three different values of k were considered.

For that reason, hereafter we shall assume that the metric satisfies Eq. (3) with this restriction. In other words, in searching for the vector which minimizes the prediction error for $t = 1$, we restrict ourselves to small neighborhoods ($k \ll N$).

In such a case it makes sense to consider Eq. (1) up to first order

$$x_{p+1} - x_{v(p)+1} \cong -\nabla_p f \cdot \epsilon_{v(p)}, \qquad (6)$$

where $\epsilon_{v(p)} \equiv X_{v(p)} - X_p$ and $\nabla_p f$ is the gradient of f at X_p. It then follows that the condition on the error $|x_{p+1} - x_{v(p)+1}| = \text{minimum}$ can be approximated by the condition $|\nabla_p f \cdot \epsilon_{v(p)}| = \text{minimum}$.

Now, denoting by $f_{,ip}$ and $\epsilon_{iv(p)}$ ($1 \le i \le d$) the ith components of the vectors $\nabla_p f$ and $\epsilon_{v(p)}$, respectively, the last condition can be rewritten as

$$\left\{ \left[\sum_{i=1}^{i=d} f_{,ip}\, \epsilon^i_{v(p)} \right]^2 \right\}^{1/2} = \text{minimum}, \qquad (7)$$

which is then equivalent to

$$\left\{ \sum_{i=1}^{i=d} \sum_{j=1}^{j=d} g_{ij}(X_p) \epsilon^i_{v(p)} \epsilon^j_{v(p)} \right\}^{1/2} = \text{minimum}, \qquad (8)$$

where $g_{ij}(X_p)$ are the components of a symmetric tensor

$$g_{ij}(X_p) \equiv f_{,ip}\, f_{,jp}\ . \qquad (9)$$

From Eq. (9) it is clear that, up to first order of approximation, the left-hand side of Eq. (8) can be interpreted as the distance between the vectors in the neighborhood, and the nearest neighbor in the metric defined by Eq. (8) is precisely that which minimizes the prediction error.

An illustration of the performance of the local optimal metric approximation is given in Fig. 2, where we compare the behavior of the correlation coefficient C which is obtained by using a global optimal metric Eq. (2) with the corresponding one obtained by using the local optimal metric Eq. (9). The components $f_{,ip}$ can be numerically evaluated by rewriting Eq. (6) for $d + 1$ neighbors and solving the resulting linear system of equations.

Besides the practical interest of improving forecasting, the deviation of the local optimal metric from the Euclidean metric can be used as a tool to distinguish between chaos and correlated noise. By studying how the correlation C depends on the embedding dimension d, on the prediction interval t, and on the number of points N, it is possible to distinguish chaos from uncorrelated noise. Nevertheless, the problem of distinguishing chaos from correlated noise or combinations of determinism and randomness is a more difficult task [11].

In order to distinguish between colored noise and chaos, let us consider the quantity $P(k)$ operationally defined as follows.

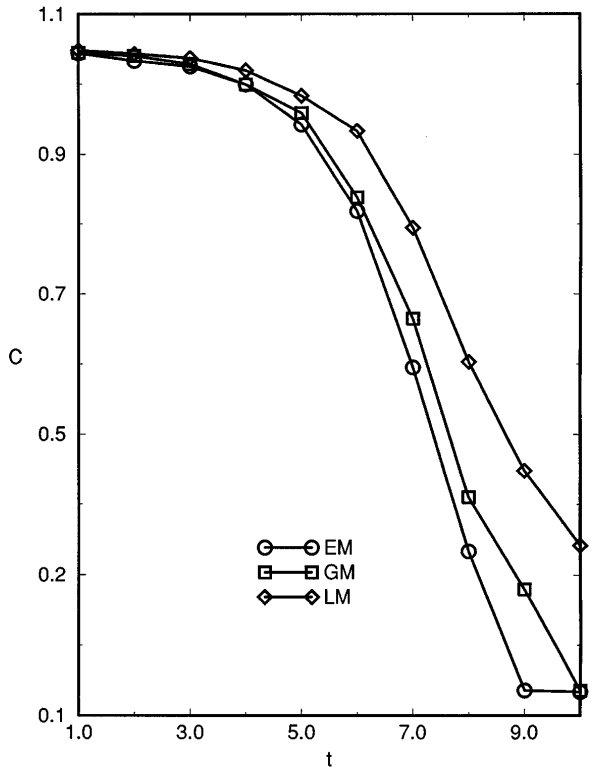

FIG. 2. C vs t for 500 data points from the Hénon map. The embedding dimension is 2. The symbols correspond to ∘, the Euclidean metric (EM) $k = 1$; □, the global optimal metric (GM); and ◇, the local optimal metric (LM) for the case $k = 10$.

(1) For each vector X_p we determine both the nearest neighbor in the Euclidean metric $X_{e(p)}$ and the best predictor $X_{v(p)}$ among the k nearest Euclidean neighbors.

(2) Running over the whole register we count how many times, $n(k)$, it happens that $X_{e(p)} = X_{v(p)}$.

(3) We define $P(k) \equiv n(k)/(N - d + 1)$.

$P(k)$ can be interpreted as the probability that the best predictor be equal to the nearest Euclidean neighbor. In other words, that the local optimal metric (for k neighbors) be the Euclidean metric.

Now, for signals which are of stochastic origin (correlated or not), there is no *a priori* reason to select any particular neighbor as the best predictor, and then the probability $p_i(k)$ that the best predictor be the ith neighbor ($1 \leq i \leq k$) must be independent of i. Since $p_1(k) = P(k)$, it then follows that $P(k) = p_i(k) = k^{-1}$. Furthermore, this result does not depend on the embedding dimension d. Then, if we represent the slopes m of the curves $\log[P(k)]$ vs $\log(k)$ as a function of d, we obtain a nearly constant $m \approx -1$.

On the other hand, whenever the data are chaotic, the k nearest neighbors are not equivalent. In fact, in order to minimize the prediction error we just need to consider those neighbors that lie on the stable submanifold of each point X_p. Because this is a subset of the k neighbors, then as $\log(k)$ is increased $\log[P(k)]$ decreases more slowly than for the noisy data. In other words, $P(k)$ must scale as k^{-a} with $a < 1$, at least for embedding dimensions equal to or greater than the critical value d_c which optimize predictions. This can be seen in Fig. 3 where we show the results for the Hénon map for the Lorenz system

$$\dot{x} = 16(y - x), \qquad \dot{y} = -xz + 45.92x - y,$$
$$\dot{z} = xz - 4z \tag{10}$$

for uncorrelated noise and for the autoregressive linear process

$$x_n = 0.4x_{n-1} + 0.9x_{n-2} - 0.4x_{n-3} - 0.3x_{n-4} + \epsilon_n\,, \tag{11}$$

where the ϵ_n are obtained from a Gaussian distribution of width 1.

Relying on the results obtained from numerical simulated data, we have applied the procedure sketched above to the series of heartbeat intervals (*R-R* intervals) from a group of healthy subjects. It has been suggested by some authors [12] that the normal heart rhythm shows features of deterministic chaos. However, it has been shown in Ref. [7] that a linear prediction method based on an autoregressive model yields better predictions than those obtained when a particular nonlinear deterministic model is used, which suggests that the series must by characterized by colored noise.

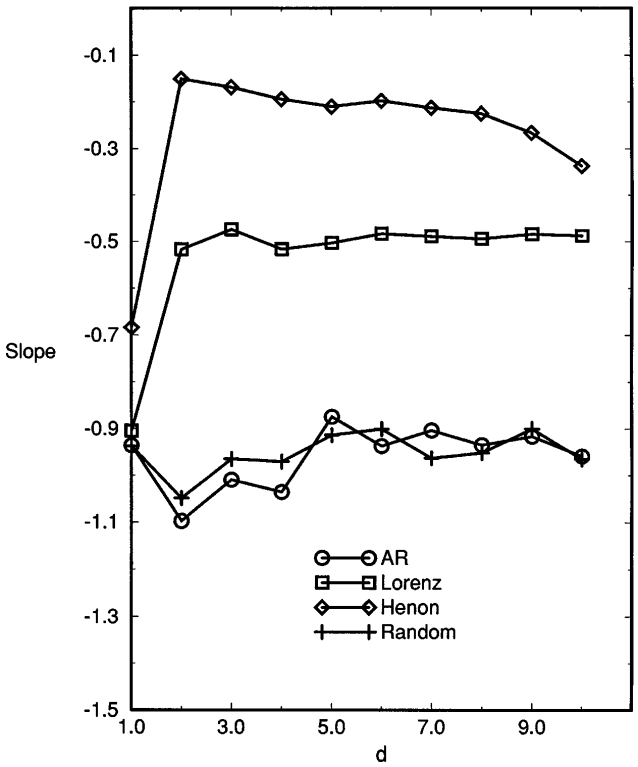

FIG. 3. The slope of $\log[P(k)]$ vs $\log(k)$ as a function of d. In the figure, (∘) corresponds to the AR model, (□) to the x component of the Lorenz system, (◇) to the Hénon map, and (+) to uncorrelated noise. In all cases $N = 1024$.

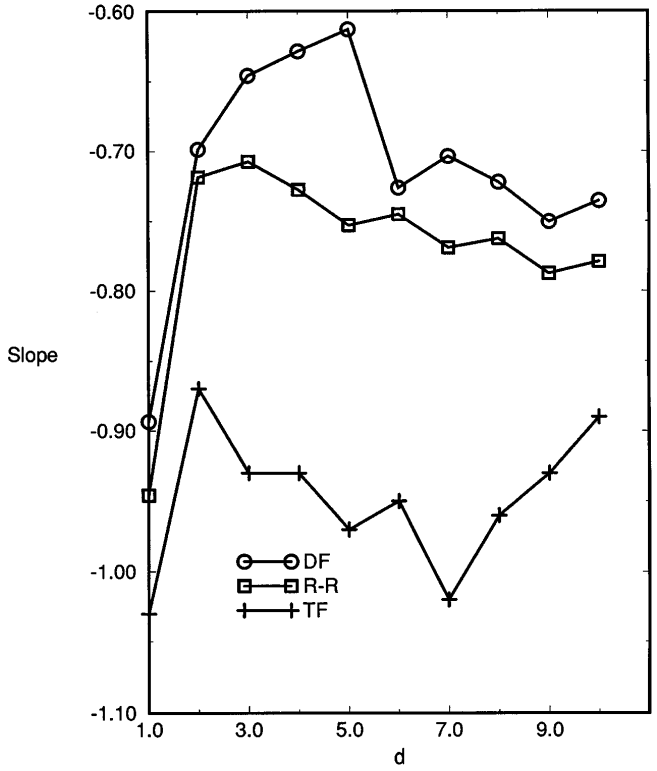

FIG. 4. The same as in Fig. 3 but for $N = 4096$ data points from a dripping faucet (DF) in a chaotic regime ($\circ$), $N = 2048$ *R-R* intervals ($\square$), and $N = 1024$ from thermal fluctuations (TF) ($+$) of the voltage across a resistor.

In order to examine the chaos evidence, electrocardiograms of 12 healthy adults were recorded for 30 minutes in supine position, and the peaks of the *R* waves were determined, yielding a series of 2048 consecutive *R-R* intervals which have an experimental error of about ± 0.003 seconds. For each subject, the corresponding series was analyzed by the method described previously.

In Fig. 4 we show the results for one patient, which are representative of the entire sample. We also show the results for the data from a dripping faucet in a chaotic regime and for experimental data corresponding to low-temperature thermal fluctuations of the voltage across a resistor. Comparison of the curves in Fig. 4 with the curves in Fig. 3 indicates that *R-R* intervals behave the same as chaotic data.

Finally, we would like to mention that even restricting ourselves to zero order approximations (in the sense that predictions are made by following the evolution of the nearest neighbor in the Euclidean, global, and also local optimal metrics), the extension to approximations of higher order can be easily implemented. An improvement in the prediction quality can also be obtained by properly using locally optimal metrics.

For instance, let us consider the Hénon map for which all derivatives can be exactly evaluated, $f_{1p} = -2.8x_p$; $f_{2p} = 0.3$, and let us denote by x_p^* the predicted value of x_p.

In a local linear approximation we obtain

$$\begin{aligned} x_{p+1}^* &= f(x_{l(p)}, x_{l(p)-1}) + \nabla_p f \cdot \epsilon_{l(p)} \\ &= 1 - 1.4x_{l(p)}^2 + 0.3x_{l(p)-1} \\ &\quad + (-2.8x_p)(x_p - x_{l(p)}) \\ &\quad + 0.3(x_{p-1} - x_{l(p)-1}), \end{aligned} \tag{12}$$

where $(x_{l(p)}, x_{l(p)-1})$ is a neighbor of (x_p, x_{p-1}).

Using the nearest neighbor in the Euclidean metric [i.e., $(x_{l(p)}, x_{l(p)-1}) = (x_{e(p)}, x_{e(p)-1})$], we obtain for the predictions and observed values correlation function $C_e = 0.973$. Nevertheless, if at this order of approximation the nearest neighbor in the locally optimal metric defined by

$$\begin{aligned} g_{ij}(x_p, x_{p-1}) &= f_{,1}^2 \delta_{1i}\delta_{1j} + 2f_{,1}f_{,2}\delta_{1i}\delta_{2j} + f_{,2}^2\delta_{2i}\delta_{2j} \\ &= 7.84x_p^2\delta_{1i}\delta_{1j} + 2(-0.84x_p)\delta_{1i}\delta_{2j} \\ &\quad + 0.09\delta_{2i}\delta_{2j} \end{aligned} \tag{13}$$

is used [i.e., $(x_{l(p)}, x_{l(p)-1}) = (x_{v(p)}, x_{v(p)-1})$], we obtain $C_{lm} = 0.999$ (the number of data points is $N = 80$, and the size of the neighborhood used is $k = 4$).

*Electronic address: pgarcia@dino.conicit.ve
†Postal address: A.P. 52120, Caracas 1050-A, Venezuela. Electronic address: jjimenez@dino.conicit.ve
‡Electronic address: almarcan@dino.conicit.ve

[1] J. D. Farmer and J. J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).
[2] D. M. Wolpert and R. C. Miall, Proc. R. Soc. London B **242**, 82 (1990).
[3] J. Jiménez, J. A. Moreno, and G. J. Ruggieri, Phys. Rev. A **45**, 3553 (1992).
[4] M. Giona, F. Lentini, and V. Cimaglli, Phys. Rev. A **44**, 3492 (1991).
[5] D. J. Wales, Nature (London) **350**, 485 (1991).
[6] G. Sugihara and R. May, Nature (London) **344**, 734 (1990).
[7] J. Lefebvre, D. A. Goodings, M. V. Kamath, and E. L. Fallen, Chaos **3**, 267 (1993).
[8] J. Jiménez, A. Marcano, J. A. Moreno, and G. Ruggieri, Phys. Lett. A **169**, 25 (1992).
[9] D. M. Rubin, Chaos **2**, 525 (1992).
[10] D. Murray, Physica (Amsterdam) **68D**, 318 (1993).
[11] M. Casdagli, J. R. Stat. Soc. B **54**, 303 (1992).
[12] J. E. Skinner, A. L. Goldberger, G. Mayer-Kress, and R. D. Ideker, Bio. Tech. **8**, 1018 (1990).