Forecasting on chaotic time series: A local optimal linear-reconstruction method

J. Jiménez, J. A. Moreno, and G. J. Ruggeri
Grupo de Dinámica de Sistemas Complejos e Inteligencia Artificial, Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Apartado 47906, Caracas 1041-A, Venezuela
(Received 8 April 1991)

An alternative forecasting technique for chaotic time series, based on the optimal association concept, is presented. The method is applied on series generated by the logistic and Hénon maps, and on experimental data corresponding to rainfall of a storm event. In all three cases the quality of the forecasts is analyzed in terms of the prediction interval, the length of the historic data available on the time series, and the dimension of the embedding space. It is shown that the method is capable of producing very satisfactory short-term forecasts for data sequences of small lengths as they often occur in real experiments. Our results also show that the present technique can be used to discriminate complex signals associated with deterministic chaos from those of random origin.

PACS number(s): 05.45.+b, 06.50.Dc

I. INTRODUCTION

Frequently, for both the scientist and engineer, the information available about the system under observation comes in the form of one, or possibly more, series of chronological data whose values in the future are to be predicted with some degree of approximation. A simple and well-known approach to this problem consists in trying to isolate tendencies, seasonalities, and characteristic frequencies present in the data, which are then incorporated in some dynamical model to simulate the main trends of the data. In many cases the model is subsequently refined through probabilistic considerations with the assumption that the time series is a realization of some underlying stochastic process [1].

As an alternative to this classical approach, a variety of techniques have been developed in recent years to tackle the problem of forecasting chaotic time series [2–10]. All these techniques show a characteristic behavior when used to forecast complex time series generated by some deterministic mechanism. On the one hand, they produce better results than the standard statistical models when used to perform short-term forecasting, and on the other hand, the accuracy of predictions falls as the prediction interval increases. This is in contrast with the results found for truly random time series where the accuracy does not depend on the prediction interval.

As it has been suggested in the literature [8,10], these facts could be used to distinguish, at least heuristically, deterministic complex signals from some signals which are of stochastic origin. We are referring, for instance, to uncorrelated time series or data which cannot be distinguished from white noise on the basis of its power spectrum.

In this paper we present a prediction method for deterministic chaotic time series, which also allows one to detect determinism in complex signals. Unlike other approaches to this problem, such as dimension measurement [11], which require large data sets [12], our method seems to produce sensible results even for relatively short-time series.

The method described here combines ideas and results already present in the literature related to this problem with some basic results familiar in the subject of associative memories and linear-reconstruction theory. As in previous forecasting methods for chaotic series, the historic data are used to construct a file of d-dimensional "delay-register" vectors $(x_i, x_{i+1} \dots x_{i+d-1})$, which, for a sufficiently high value of d, should lie in a geometrical support diffeomorphic to the attractor of the involved dynamical system [13,14]. Predictions are made under the assumption that the coefficients of the optimal linear association of a vector in terms of its near neighbors are preserved under the dynamical rule (see Sec. II).

We present the results of some numerical experiments in which the method is used to forecast series generated with the logistic map [15], the Hénon map [16], and for experimental data coming from a storm event previously reported in the literature [17]. In order to evaluate the quality of our forecasts, we use some estimators introduced previously in the literature. They are the normalized error used by Farmer and Sidorowich [2] and the correlation between the observed and predicted values of each series used by Sugihara and May [8]. In Sec. III we report how these quantities depend on the number of observations, prediction interval, and embedding dimension.

II. LOCAL OPTIMAL LINEAR ASSOCIATION

As in some of the previously mentioned forecasting techniques, the method that we present here is supported by the fact that, if a deterministic mechanism governs the evolution of the system, then, because the data vector will move on a submanifold of the embedding space according to a continuous dynamical rule, the future value of a data vector can approximately be predicted from the evolution of its nearest neighbors.

Naturally, the relative distances from the vector whose future value is being forecasted (predictee) to its neighbors will be changed by the dynamics. Now, if a set of neighbors, whose evolution is known, is chosen to define a neighborhood to the predictee, the question is how to locate the position of the evolved predictee within the neighborhood defined by the evolved nearest neighbors.

Basically, our prediction technique for chaotic time series consists in the establishment of an optimal linear association between the predictee and a set of its near-neighboring vectors, together with the assumption that the same linear association holds between the unknown predicted vector and the transformed vectors of the neighbors.

Let $\{X_1, X_2, \ldots, X_\epsilon\}$ be a set of n-dimensional vectors which span a subspace $\mathcal{L} \subset \mathbb{R}^n$. An arbitrary vector Y is uniquely expressible as the sum of two mutually orthogonal projections, one of which is the orthogonal projection onto the space $\mathcal{L}$. It is easy to see that this projection is the best linear combination of the X's which approximates Y in the sense of least squares. This approximation of Y defines its optimal linear association with the set of vectors X. It can be written as $\sum_\alpha C_\alpha V_\alpha$, where $C_\alpha = Y \cdot V_\alpha$. The dot represents scalar product, and V_α is a member of an orthonormal Gram-Schmidt basis for the subspace $\mathcal{L}$.

The essentials of the application of the concept of optimal linear association to the prediction of chaotic time series can be formulated as follows. According to Takens [14], if a time series $x_1, x_2, \ldots, x_N, \ldots$ corresponds to one of the components of a dynamical system φ, then it is a generic property that, for sufficiently high values of d, the delay-register vectors $(x_i, x_{i+1}, \ldots, x_{i+d-1})$ fall in a geometric support which is diffeomorphic to the attractor of the dynamical system. Then, denoting by ϕ such diffeomorphism, a functional relation of the form

$$(x_{i+j}, x_{i+j+1}, \ldots, x_{i+j+d-1}) = f^{0j}(x_i, x_{i+1}, \ldots, x_{i+d-1}) \quad (1)$$

exists where $f \equiv \phi_0 \varphi_0 \phi^{-1}$ is the composition of the involved functions and the index $0j$ denotes the jth iteration.

In other words, if we have a predictee X_p which is a member of a file of delay-register vectors, the "best predictor" j units of time later is the jth iteration of the underlying dynamical rule itself. Since in general the rule is not known, the prediction task consists in finding a suitable approximation to it, at least for small values of the iteration j. In this paper we show that using the local optimal linear-association method a very good approximation is obtained.

In order to construct an approximation to the register vector $\tilde{X}_p$ to which the predictee X_p is mapped by the dynamical rule, we proceed in the following way. Let $\{X_1, X_2, \ldots, X_k\}$ be a set of the vectors nearest to X_p (in the Euclidean distance), obtained from the file of available or historic delay-register vectors. The number k of nearest neighbors that we have chosen will be specified below. Let us denote by ξ_p the vector X_p relative to the centroid R of the neighbor X's. That is,

$$R \equiv (1/k) \sum_{i=1}^{k} X_i, \quad \xi_p \equiv X_p - R \ . \quad (2)$$

Let also ξ_i be the relative vector from X_i to the centroid (Fig. 1):

$$\xi_i \equiv X_i - R \ . \quad (3)$$

The set of coefficients $\{C_\alpha\}$ which defines the optimal linear association of ξ_p in terms of the ξ_i can be obtained, as above, by projecting ξ_p on the Gram-Schmidt basis constructed out of $\{\xi_i\}$.

To predict the vector $\tilde{X}_p$ to which X_p evolves after a time lag, we look at the vectors $\tilde{X}_i$, known to be the images of the nearest neighbors X_i under the dynamical rule after the same time lag. Let $\tilde{R}$ be the new centroid and $\tilde{\xi}_p$ the (unknown) relative vector corresponding to $\tilde{X}_p$:

$$\tilde{\xi}_p \equiv \tilde{X}_p - \tilde{R}, \quad \tilde{R} \equiv (1/k) \sum_{i=1}^{k} \tilde{X}_i \ . \quad (4)$$

Similarly $\tilde{\xi}_i$ will denote the relative vector (Fig. 1):

$$\tilde{\xi}_i \equiv \tilde{X}_i - \tilde{R} \ . \quad (5)$$

Our fundamental approximation is that the coefficients $\tilde{C}_\alpha$ which occur in the optimal linear association between $\tilde{\xi}_p$ ad the set $\{\tilde{\xi}_i\}$ are equal to C_α.

In other words, the vector $\tilde{\xi}_p$ is "reconstructed" according to

$$\tilde{\xi}_p = \sum_\alpha C_\alpha \tilde{V}_\alpha \ , \quad (6)$$

where the $\tilde{V}_\alpha$ are obtained from the $\tilde{\xi}_i$ through an orthogonalization process and the index α orders the vectors according to the distances from the corresponding X_i to X_p. It is shown in the Appendix that Eq. (6) is in fact true up to first order. In other words, to this order of approximation, the vector reconstructed according to Eq. (6) gives the exact result.

The procedure followed here has the additional advantage of avoiding the explicit calculation of the corresponding Jacobian matrix through a simple and stable algorithm. This should be compared, for instance, with local linear (least-squares) approximations to the dynamics where conflicting requirements have to be imposed. In fact, small neighborhoods are needed to ensure good local

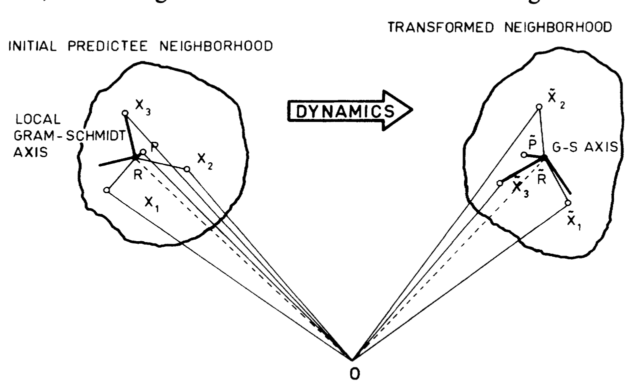

FIG. 1. Schematic diagram showing the principles of the present forecasting method. An optimal linear association between the predictee and a collection of near delay-register vectors is used to reconstruct an approximation to the image of the predictee.

approximations, whereas the stability of the linear relation between neighbors and their transformations under the dynamics increases with large neighborhoods.

It is important to remark that although Eq. (6) takes correctly into account the first-order terms, it contains also nonlinear terms. Because of this, we refer to Eq. (6) as a local linear reconstruction instead of a local linear approximation.

III. RESULTS

A number of numerical experiments were made to test the forecasting ability of the methodology described above. These were conducted on the logistic map:

$$x_{n+1}=4x_n(1-x_n)\ , \tag{7}$$

the Hénon map,

$$x_{n+1}=1-1.4x_n^2+0.3x_{n-1}\ , \tag{8}$$

and on a data record of a storm which took place in the city of Boston on October 25, 1980 [17]. This record consisted of 1990 measurements of the amount of rain falling at a specific site, taken at equally spaced intervals of 15 sec (see Fig. 2). We chose to analyze this particular data because it has been suggested by the authors in Ref. [17] that a low-dimensional chaotic attractor is involved.

In all cases the number k of neighbors was taken initially as twice the embedding dimension d. This number seemed to us sufficiently high so as to allow the Gram-Schmidt procedure an appropriate selection of a basis for the local subspace. Predictions were made using a fixed number of previous register vectors as history. The set of register vectors was updated after each prediction.

In order to have a numerical indication of the quality of our predictions, the following two quantities were computed for each run: the centered correlation between the series of predicted values and the observed values [8],

$$C=\frac{\langle (x_{\text{pred}}-\bar{x}_{\text{pred}})(x_{\text{obs}}-\bar{x}_{\text{obs}})\rangle}{[\langle (x_{\text{pred}}-\bar{x}_{\text{pred}})^2\rangle\langle (x_{\text{obs}}-\bar{x}_{\text{obs}})^2\rangle]^{1/2}}\ , \tag{9a}$$

and the normalized error [2]

$$E=\frac{\langle (x_{\text{pred}}-x_{\text{obs}})^2\rangle^{1/2}}{\langle (x_{\text{obs}}-\bar{x}_{\text{obs}})^2\rangle^{1/2}}\ . \tag{9b}$$

Some qualitative features of our results for the logistic map are contained in the scatter diagrams of Figs. 3(a) and 3(b), where predicted values are compared with the real ones at intervals of 1 and 5 units of time ahead, respectively. It is seen that the prediction quality gets poorer as the interval of prediction is increased. This is a general characteristic of chaotic signals and is due to the presence of positive Lyapunov exponents. This behavior was not observed when our method was used, for instance, in periodic time series contaminated with a cer-

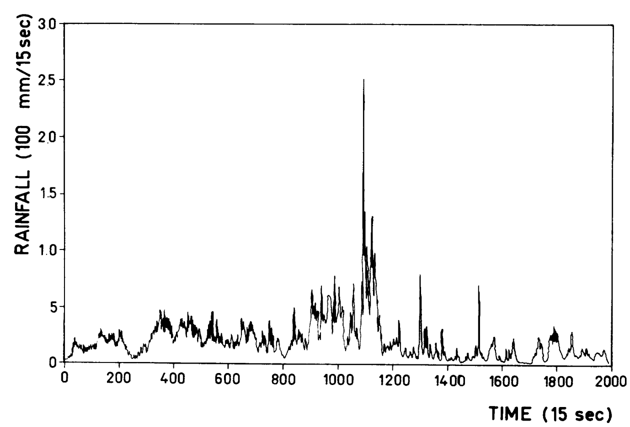

FIG. 2. Experimental data corresponding to a temporal record of a storm event which occurred in the city of Boston on October 25, 1980.

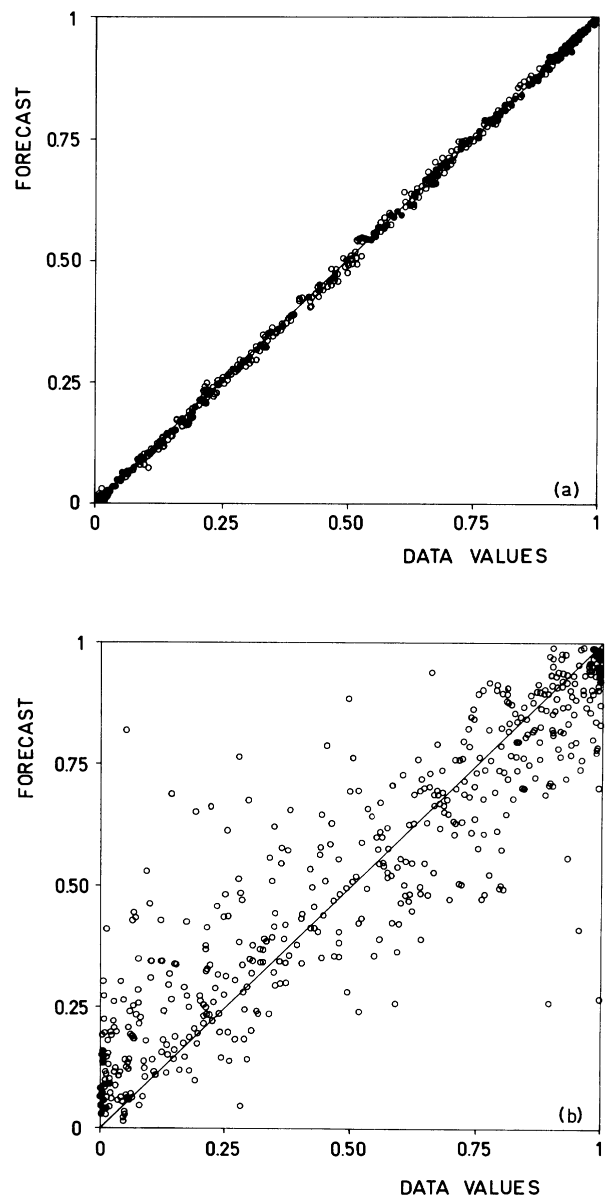

FIG. 3. (a) and (b) Scatter diagrams for 500 forecasts of a 1000-point realization of a logistic time series. The prediction intervals into the future are (a) one and (b) five time steps. The set of delay-register vectors employed is built out of 500 points in the series, for an embedding dimension of 2.

tain amount of noise [8]. This is due to the fact that for random signals the register vectors do not evolve through continuous rules, and then near neighbors can evolve, even at short term, toward rather distant vectors. The results are depicted in Fig. 4.

It is worth mentioning that if instead of our method, or similar ones, standard statistical techniques are used, short-term predictions made on chaotic time series are found to be poorer essentially because of the absence of correlation between the values of these series.

In Fig. 5 the behavior of the correlation coefficient between predicted and real values, for the data of Fig. 2, is compared with the corresponding coefficients for the logistic and intermittent maps [18],

$$x_{n+1}=1-1.7498x_n^2 , \tag{10}$$

contaminated with a 10% additive noise. It can be seen that although in all three cases the correlation coefficient decreases when the prediction interval increases, the curve falls slower for both the intermittent map and

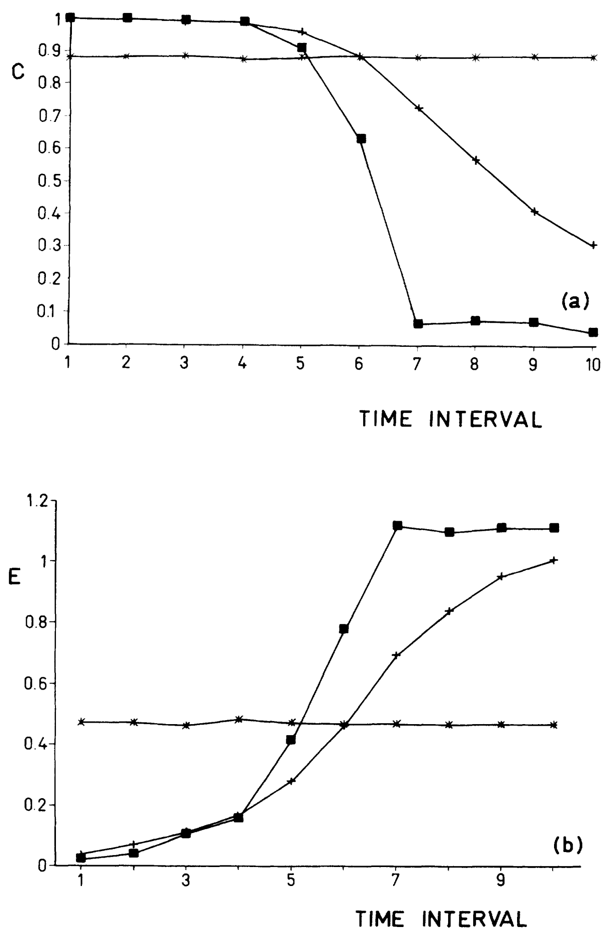

FIG. 4. (a) Prediction accuracy as measured by the centered correlation function C and (b) relative error E as a function of the time interval. In the figure ■ corresponds to the logistic map, + to the Hénon map, and * to a limit cycle of the logistic map with parameter 3.542 to which a 30% uncorrelated noise was added. In all cases the embedding dimension is 2. The set of delay-register vectors employed is built from 500 points in the series out of which 500 predictions are subsequently made.

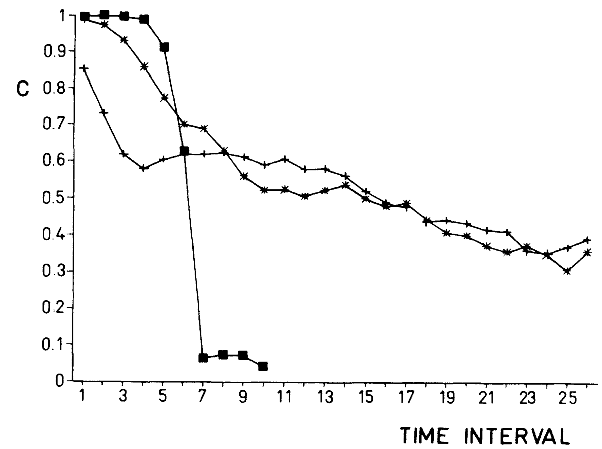

FIG. 5. Correlation function C as a function of the prediction interval for data from a storm event (+) compared to simulations over the logistic map (■) and the intermittent map (*) contaminated with a 10% additive noise. Both for the logistic and intermittent maps the embedding dimension is 2, the historic data contain 500 points, and we make 500 predictions. For the storm event the embedding dimension is 9 and the number of register vectors is 800.

storm data than for the chaotic series. For the intermittent map this result follows from the fact that when predicting the portion of the data corresponding to the laminar regime (period-3 cycle), the correlation coefficient is independent of the prediction interval. Similar observed behavior of the correlation coefficients for the intermittent map and storm data suggests that the chaotic behavior reported in this latter case [17] could be of the type of an intermittent phenomenon. Although this observed behavior does not *a priori* rule out other classes of deterministic dynamics, it is not compatible with the kind of results obtained, with the present method, for stochastic processes where the correlation coefficient remains very small for all forecasting intervals.

As expected, our results show a marked sensitivity to the number of points taken as the history to effect the predictions. As this number gets higher, the attractor is covered with a higher density and so the distance from the reference vector (from which the prediction is made) to its neighbors diminishes. This has the effect of improving our linear approximation. This can be seen from Fig. 6. This effect should be expected as long as the density of points on the attractor is not so high that the mean distance between them gets lower than the numerical resolution, in which case the predictions become dominated by the resolution noise [15]. Finally, this behavior is not observed for random data because, as there is no relation between a given vector and its neighbors, there should be no improvements in the quality of predictions when the number of points in the history increases.

Figure 7 shows how predictions depend on the embedding dimension d, keeping fixed both the volumes of data employed and the prediction interval. For small values of d, the register vectors do not lie in a support

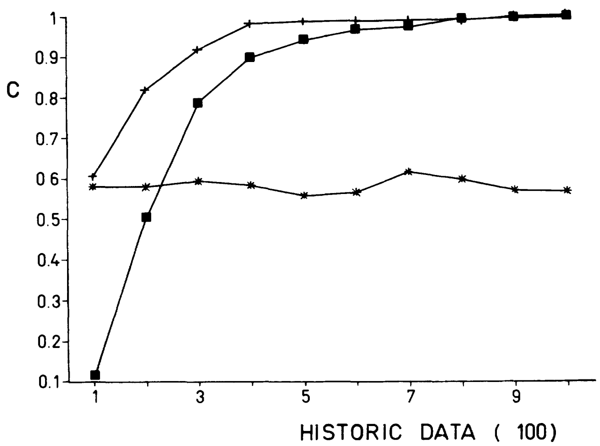

FIG. 6. Correlation function C as a function of the number of points taken as history. In the figure * corresponds to a superposition of a limit cycle of the logistic map, with parameter 3.542, and a 30% uncorrelated noise. The symbols ■ and + refer to the logistic and Hénon chaotic maps, respectively. In all cases the embedding dimension is 2 and the time interval is 5.

diffeomorphic to the attractor and consequently do not represent bijectively the underlying dynamics. This implies that increasing d (from small values to those which give a diffeomorphic reconstruction of the attractor) should improve the quality of the forecasting. On the other hand, the distance between a predictee and its nearest neighbor is prone to increase when, keeping fixed the used data volume, d is augmented. This fact has the effect of deteriorating the predictions. There exist then two competing effects when the value of d increases up to the critical value which corresponds to a diffeomorphic reconstruction of the attractor. Above this value, predictions deteriorate systematically. Again, because of the reasons already mentioned, this type of behavior is not observed for random series.

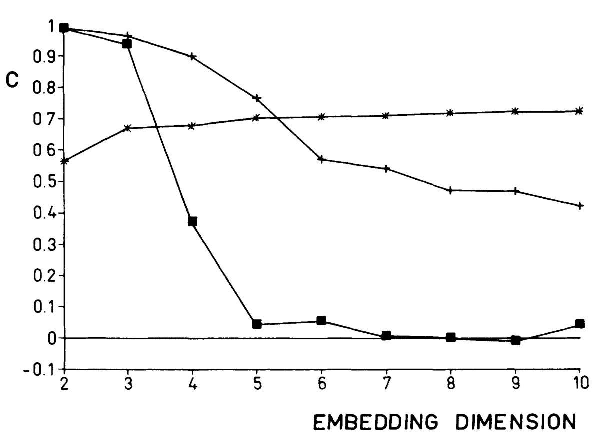

FIG. 7. Correlation function C is shown as a function of the embedding dimension d for predictions four steps into the future and a 100-point history. The symbols correspond to ■, logistic map with parameter 4; +, Hénon map; and *, logistic limit cycle, with parameter 3.542, contaminated with a 60% additive, uncorrelated noise.

IV. CONCLUSIONS

We have presented an alternative method of forecasting chaotic time series. It consists basically in establishing an optimal local linear relation between a delay-register vector and a set of its nearest neighbors. The approximation we use consists in taking the coefficients of this local relationship as constants during the desired prediction interval, so that the predicted vector can be easily reconstructed from the set of transformed nearest neighbors. We have applied the method to a number of interesting typical cases. In all cases the results for short-term predictions are satisfactory, even for history lengths as small as 300 values of the series.

The most important features of the method are, first of all, its geometrical simplicity, a fact that can be translated to a very simple algorithm and, correspondingly, a very time-efficient computing code that can be implemented in microcomputers. The forecasting accuracy we have found using the method described in this paper is as good as those found with other methods in similar examples [2,3,8–10]. These other methods are either more difficult to implement or are more time consuming since they require a substantially larger data set.

The good accuracy found in short-time series, as those which often occur in real measuring problems, makes the method highly suitable for processing experimental data. There is no need of preprocessing the data in order to rule out possible seasonalities or tendencies. Finally, the method makes it possible to discriminate random series from deterministic ones. This can be achieved by applying the method over the data set of interest and observing the behavior of the prediction quality for different values of the embedding dimension, the length of the history used to construct the file of delay-register vectors, and the prediction interval.

To conclude, let us mention a number of possible ways of improving some of the distinct aspects of the method. The forecasting accuracy could be improved by adjusting the coefficients of the local linear relation, after the evolution, in some more complicated and presumably more exact way. This, however, would require some amount of previous processing of the data before the method is applied. A more promising modification could be to establish a nonlinear local relation, instead of the linear one adopted here, but in this case it would not be proper to speak of nonlinear transformations which are optimal in an absolute sense.

ACKNOWLEDGMENTS

We are grateful to Dr. Henry Power and Beatriz Febres de Power for making the storm data available to us and for many interesting discussions related to it.

APPENDIX

In this appendix it will be shown that, up to first order, the vector $\tilde{\xi}_p$ can be reconstructed according to Eq. (6). In fact, the predictee $X_p = R + \xi_p$ is mapped by the dynamics to

$$\tilde{X}_p = f^{0j}(X_p) \simeq f^{0j}(R) + \mathcal{J}\xi_p \ , \qquad \text{(A1)}$$

where $\mathcal{J}$ denotes the Jacobian matrix. A similar expression holds for the neighbors of the predictee. It then follows that, up to first order, $\tilde{R} = f^{0j}(R)$. Since $\xi_p = \sum C_\alpha V_\alpha$, then

$$\tilde{X}_p \simeq \tilde{R} + \sum C_\alpha \mathcal{J} V_\alpha \ . \qquad \text{(A2)}$$

Let $\tilde{V}_1 = \tilde{\xi}_1$ be the vector taken as the first member of the Gram-Schmidt basis for the evolved neighborhood. Then $\tilde{V}_1 = f^{0j}(R + \xi_1) - \tilde{R} \simeq \mathcal{J} V_1$. The other vectors of the basis are constructed according to the recursive formula

$$\tilde{V}_\beta = \tilde{\xi}_\beta - \sum_{\alpha=1}^{\beta-1} (\tilde{V}_\alpha \cdot \tilde{\xi}_\beta)(\tilde{V}_\alpha / |\tilde{V}_\alpha|^2) \ , \qquad \text{(A3)}$$

which, up to first order, leads to

$$\tilde{V}_\beta = \mathcal{J} V_\beta + \sum_{\alpha=1}^{\beta-1} \{ [(V_\alpha \cdot \xi_\beta) | V_\alpha|^2] - [(\tilde{V}_\alpha \cdot \tilde{\xi}_\beta) / |\tilde{V}_\alpha|^2] \} \mathcal{J} V_\alpha \ . \qquad \text{(A4)}$$

For locally invertible transformations, in this order of approximation, the scalar products are invariant. Hence the term within the curly brackets in the right-hand side of (A4) is then strictly zero. In any other case this is a higher-order term. Therefore, $\tilde{V}_\beta = \mathcal{J} V_\beta$, for all β, up to first order. Substituting these results in Eq. (4), we see that Eq. (6) is exact up to this order.

[1] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods* (Springer-Verlag, New York, 1987).
[2] J. D. Farmer and J. J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).
[3] F. Varosi, C. Grebogi, and J. A. Yorke, Phys. Lett. A **124**, 59 (1987).
[4] M. Casdagli, Physica D **35**, 335 (1989).
[5] A. Fraser, Physica D **34**, 391 (1989).
[6] A. Fraser, IEEE Trans. Inf. Theory **IT-35**, 245 (1989).
[7] E. Kostelich and J. A. Yorke, Physica D **41**, 183 (1990).
[8] G. Sugihara and R. May, Nature **344**, 734 (1990).
[9] H. Abarbanel, R. Brown, and J. Kadtke, Phys. Rev. A **41**, 1782 (1990).
[10] D. M. Wolpert and R. C. Miall, Proc. R. Soc. London B **242**, 82 (1990).
[11] P. Grassberger and I. Proccacia, Phys. Rev. Lett. **50**, 346 (1983).
[12] D. Ruelle, Proc. R. Soc. London A **427**, 241 (1990).
[13] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[14] F. Takens, in *Dynamical Systems and Turbulence,* edited by D. Rand and L. S. Young (Springer-Verlag, Berlin, 1981).
[15] R. S. Shaw, Z. Naturforsch. A **36**, 80 (1981).
[16] M. Hénon, Commun. Math. Phys. **50**, 69 (1976).
[17] I. Rodriguez-Iturbe, B. Febres de Power, M. Sharifi, and K. Georgakakos, Water Resour. Res. **28**, 1667 (1989). These data were provided by Earl Williams of the Department of Meteorology of the Massachusetts Institute of Technology.
[18] J. P. Eckmann, Rev. Mod. Phys. **53**, 643 (1981).