Improvement of the local prediction of chaotic time series

Gong Xiaofeng[1,*] and C. H. Lai[1,2]
[1]*Department of Physics, National University of Singapore, Singapore 119260*
[2]*Department of Computational Science, National University of Singapore, Singapore 119260*
(Received 20 May 1999)

In this paper we explore the effect of *pseudofalse* neighbor points, which are true neighbor points in the reconstructed attractor, but which are considered not suitable to be used when local methods are adopted to predict the chaotic time series. In our approach, the ϵ^p neighbor points are used to reduce the influence of the pseudofalse neighbor points, thereby improving the performance of the local prediction of the chaotic time series. [S1063-651X(99)11111-5]

PACS number(s): 05.45.−a

I. INTRODUCTION

Predicting the evolution of physical systems from observations is one of the most pressing challenges of modern time series analysis. Since chaotic systems are frequently encountered in various fields ranging from physics and chemistry to biology and others, prediction of chaotic systems thus has great practical significance. Although deterministic chaos is characterized by positive Lyapunov exponents, and thus intrinsically associated with a loss of predictability, a great deal of research efforts which focus on the nonlinear prediction of chaotic time series [1–17] has shown that reliable short term predictions can still be achieved. Furthermore, in addition to the importance of the predictions themselves, the methodology based on the predictability of the systems under study has proved to be a powerful tool to analyze nonlinear systems in general. For example, from the cross-correlation function between observed values and the predicted values through these techniques, it is possible to estimate the largest Lyapunov exponent of the dynamics [13]. Also, when the fit or agreement achieved using nonlinear modeling is better than the one obtained using probabilistic models, it is reasonable to assume that there is a deterministic mechanism which governs the process under study, providing in this way a tentative criterion to discriminate between chaos and noise [4,14,15,17].

The predictions of chaotic time series are based on the reconstruction of strange chaotic attractors from an observed time series $x(t)$ [20]. It was shown [20] that an embedding of the attractors can be obtained by constructing a vector $X_t = (x(t), x(t+\tau), \ldots, x(t+(d-1)\tau))^T$ from time-delayed coordinates, where d is the embedding dimension and τ is the delay time. Then the dynamics on the attractor is a map $f: R^d \rightarrow R^d$ with $x_{t'} = f(x_t)$, where x_t is the current state and $x_{t'}$ is the future state. Thus if we find an approximation f_t of f, then we can use the f_t as a prediction function.

Most prediction techniques can be grouped into two major classes: global and local. In global methods, the whole past information is used for predictions about the evolution of the system under study. One example is the method of neural networks [16]. Clearly, this kind of approach has the disadvantage that if new information is taken into account then all the parameters of the model may change, and then a long parameter estimation time may be required. Local methods overcome this drawback by utilizing only part of the history. In particular, if X_{N+1} is needed, only the set of points of reconstructed space $X_N(q): 1 \leqslant q \leqslant K$, which are close enough to the predicting point X_N, are used to fit the local function f. We can fit a new function for each time $N+t$ and the prediction is given by the value of the fitting function at X_{N+t}, where X_{N+t} itself is also a predicted value. The approach allows for considerable flexibility in building a globally nonlinear model, while fitting a few parameters in each local patch.

However, two difficulties arise from this approach. One is in deciding how to choose suitable neighbor points, and the other is the question of how long into the predicted series we can trust. In this paper, we explore the existence of a class of neighbor points in phase space which are not suitable for use in local prediction — we call these *pseudofalse* neighbor points. We propose a method to choose alternative and more relevant neighbor points (ϵ^p neighbor points), and thus improve the performance of the local prediction. As a by-product, the prediction process, which adopts our procedure to select the neighbor points, will in many cases stop automatically (i.e., there will be no ϵ^p neighbor points around some predicted point in the phase space) when the predicted trajectory deviates significantly from the original one. Thus it can give us a modest warning of the reliability of the predicted series.

The paper is organized as follow: In Sec. II we will briefly review the local prediction method, and in Sec. III we will define the ϵ^p neighbor point. The improvement of the local prediction by using ϵ^p neighbor points instead of the ordinary ϵ neighbor points is described in Sec. IV Finally, conclusions and some discussions of this approach are presented.

II. LOCAL PREDICTION OF CHAOTIC TIME SERIES

The first step to establish the local prediction model is to transform the observed scalar signal $s_i, i = 1,2, \ldots, N$ with the sample interval δt, into M-dimensional time-delay vectors: $X_i = (s_i, s_{i-r}, \ldots, s_{i-(M-1)r})$, with the delay time τ

*Electronic address: scip7254@nus.edu.sg

$=r\delta t$. For sufficiently large M, the trajectory of reconstructed vectors X_i reflects the true state space evolution of chaotic systems.

Deterministic predictions assume that the trajectory governed by a deterministic continously mapping F. If the state at time j was similar to the present X_i (and thus close in the phase space), continuity of F guarantees that X_{j+1} will also be close to X_{i+1}. Let X_i be the present state. To predict X_{i+1}, we can find all points within a region in phase space of radius ϵ, i.e., the ϵ neighbors of $X_i: X_{i_k}, k=1,2,\ldots,K$, where the index k does not refer to time order, and denote the image of each vector X_{i_k} as X_{i_k+1}. The predicted value $\hat{X}_{i+1}$ is then estimated from these K images. Specifically, for each X_{i_k} we define the displacement $\Delta_{i_k}=X_{i_k}-X_i$; $\hat{X}_{i+1}$ is then computed as $\hat{X}_{i+1}=X_i+\Delta_i$, where Δ_i is determined by local interpolation. Different methods thus appear based on the different methods of local interpolation, where the local linear method is most attractive due to its simplicity. This approach works very well for low-dimensional chaotic systems. There are also many ways to improve the performance of the prediction, e.g., by reducing the weight given to near neighbors which are themselves neighbors in time; by allowing K to vary with X_i; by using regularization techniques; and by adopting other metrics instead of the normal Euclidean metric.

III. *PSEUDOFALSE* NEIGHBORS AND ϵ^p NEIGHBORS

One of the critical steps to establish the local prediction model is to find the neighbor points of a given point in the training data set. Our choice of neighbors is limited by the finite size of the data set, by the stochastic noise, and most importantly by the complex structure of the attractor. These limitation are the main source of errors in the analysis. Finding legitimate neighbors of a given point is one of the most critical tasks in obtaining reliable results. False neighbors can be caused by improper embedding, such as an insufficient large embedding dimension [21]. In this paper, we are not concerned with this kind of difficulty, since it cannot be altered after we reconstruct the state space using a particular embedding method. The question of how to perform an optimal reconstruction for the purpose of the prediction will be considered elsewhere.

While considering the particular problem of prediction, there is at one other kind of *false* neighbor. From the point of the view of attractor reconstruction, they are true neighbor points in the original phase space, but will greatly influence the performance of the local fitting of the dynamics. We illustrate this effect in Fig. 1, where the projection of the three-dimensional state space is described. Projected trajectories appears to cross near the point X_n. Suppose the predicting (reference) point is X_n, while X_{n_1} and X_{n_2} are two neighbor points of X_n, and $\|X_n-X_{n_1}\|<\|X_n-X_{n_2}\|$. Note that in the phase space, X_{n_1} is a true neighbor of X_n. It is evident that if we took this point into account for local fitting, it would decrease the fitting accuracy dramatically, since the deviation at the next time is large. For a chaotic system, this will lead to the fast amplification of error in future predictions because of the positive Lyapunov exponent. It is then clear that for the purpose of local prediction, X_{n_2} is much better than X_{n_1}; even X_{n_1} is closer to the reference point. We call such points pseudofalse neighbor points for local predictions. In our experience, the the error of the local prediction method always bursts at a certain region in the state space. (Ref. [10] also observes this phenomenon). We believe that an important reason for this is the existence of these pseudofalse neighbors. Pseudofalse neighbors frequently occur near saddle points in phase space. For example, in the Lorenz system, when using $x(t)$ to predict, almost all divergence of the predicted trajectory from the original one burst when the trajectory approaches $x(t)=0$, where near crossing of trajectories happens and the number of pseudofalse neighbors increases dramatically (see Fig. 2).

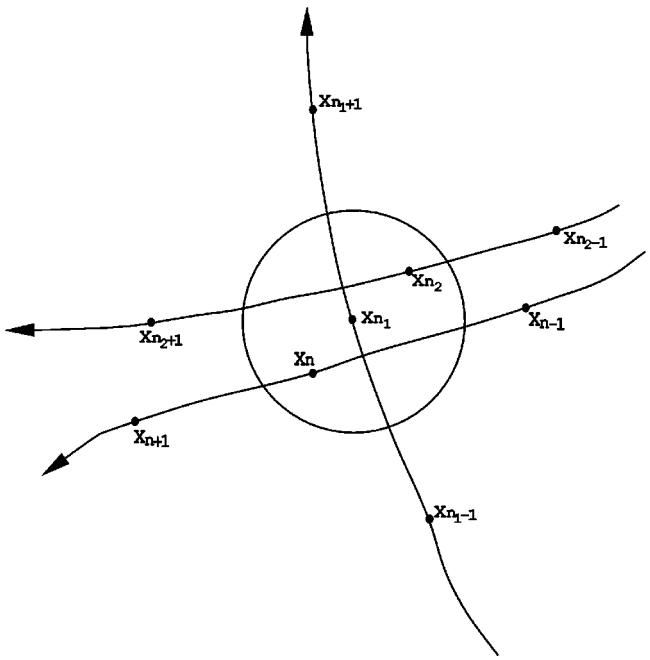

FIG. 1. A schematic view illustrating the effect of the pseudofalse neighbor point. In the figure, X_{n_1} is the pseudofalse neighbor point of X_n, when using local prediction methods.

IV. IMPROVEMENT OF THE LOCAL PREDICTION

Our approach to eliminate the influence of the pseudofalse neighbors is to choose the ϵ^p neighbor instead of ordinary ϵ neighbor to fit the local function. Suppose we have the scalar chaotic time series, $\{x(n), n=1,2,\ldots,N\}$. According to the Takens' theory [20], we can reconstruct the state space from the delayed coordinates. The state space can be written as $X_n=(x(n),x(n-\tau),\ldots,x(n-d(\tau-1)))^T$, where τ is the time lag, and d is the embedding dimension. It can be shown that $d\geqslant 2D_0+1$ (where D_0 is the box-counting dimension of the original attractor) is sufficient, but smaller embedding dimensions will suffice for the purposes of forecasting [18]. The value of τ can be determined by the first minimum of the mutual information [19] or the first zero value of the correlation function, and the embedding dimension d can be decided by the nearest-false neighbor method [22].

In the reconstructed state space, for a particular reference point X_N, the ordinary ϵ neighbor points $\{X_{N(j)}, j=1,2,\ldots\}$ satisfied $\|X_{N(j)}-X_N\|<\epsilon$ for $j=1,2,\ldots$. The ϵ^p

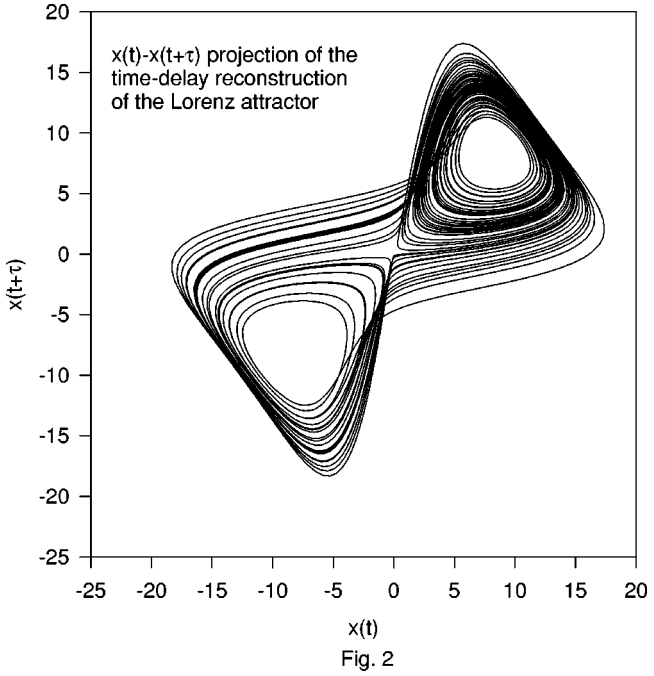

FIG. 2. The $x(t)-x(t+\tau)$ projection of the time-delay reconstruction of the Lorenz attractor described as Eq. (2), where $\tau=17$ and $d=3$.

neighbor points $\{X_{N(j,p)}, j=1,2,\ldots,p=1,2,\ldots\}$ are defined as those points satisfying

$$\|X_{N(j)-p}-X_{N-p}\|<\epsilon \tag{1}$$

for every j and p. Noted that when $p=0$, it is equivalent to the ordinary ϵ neighbor. This approach is in fact using a segment of the pattern included in the training data set to ensure that the predicted value does not diverge from the original one. (However, if the original time series generate a new pattern that is distinct from the training data set, then this method may give a wrong result. This is in general caused by insufficient training data, and cannot be resolved in the framework of the local prediction method). The main rationale behind this strategy is that the ϵ^p neighbor points lie near the reference point not only in state space, but also in tangent space. The similar idea is also used to identify the embedding dimension [26,21], where the directions of vectors in a neighborhood are examined to exclude the false crossing caused by the improperly reconstructed space. The pseudofalse neighbor points for prediction mentioned above can be excluded when adopting the ϵ^p neighbor points.

To illustrate the improvement of the performance of local prediction method by using ϵ^p neighbor points, we use the Lorenz model as an example. This model describes the Rayleigh-Bénard convection arising from the two-dimensional Navier-Stokes equation, which is formulated for a fluid slab of finite thickness subjected to gravity loading, heated from below and between the top cold and the bottom hot surface the temperature is held constant. The partial differential equations were transformed to a set of three ordinary differential equations [23]. Later Lorenz described the way he derived the model both physically and mathematically [24]. The Lorenz equations can be written in the following forms:

$$\begin{aligned}\dot{x}&=-\sigma x+\sigma y,\\ \dot{y}&=-xz+rx-y,\\ \dot{z}&=xy-bz,\end{aligned} \tag{2}$$

where x is the amplitude of the convection motion, y is the temperature between the ascending and descending currents, z is the distortion of the vertical temperature profile from linearity, and σ, r, and b are dimensionless parameters. The parameters are most commonly selected to be $\sigma=10, r=28$, and $b=8/3$ for a rich dynamical behavior [25].

The standard fourth-order Runge-Kutta method is used to solute the equations (discarding the transients). The time step is set to be $\delta=0.01$, and the x values are used to reconstruct the state space with the time lag $\tau=17$ and $d=3$. One of the projection of the reconstructed attractor is described in Fig. 2.

Since our main purpose is to show the improvement of the performance of the prediction, we adopt the simplest local prediction method, the locally constant approximation method. In order to predict one step into the future of the reference point X_N, we (1) find the set of ϵ^p neighbor points of the reference point $U(X_N,\epsilon^p)$ [the Euclidean metric is used in Eq. (1)]; (2) take the average of the values of the one-step future prediction of these points as the predicted state vector for the next time step,

$$\hat{X}_{N+1}=\frac{1}{\|U(X_N,\epsilon^p)\|}\sum_{X_{N(k)}\in U(X_N,\epsilon^p)} X_{N(k)}, \tag{3}$$

where $\|U(X_N,\epsilon^p)\|$ denotes the number of the points in the set of $U(X_N,\epsilon^p)$; and then (3) take first component of the vector as the scalar predicted value.

Randomly selected time series with the length of $N=10000$ are used as training set, and the prediction is made as far as $n=400$ time steps into the future. Three examples are shown in Figs. 3(a)–3(c). In making the predictions, we set $\epsilon=0.5$, and if there is no ϵ^p neighbor point of certain reference point and certain p, which means that there is no such pattern in the training data, (i.e., the predicted trajectory already diverge from the original one), we then stop the prediction process. (In actual implementation, because of the correlation between the temporally nearby points, an ϵ^p neighbor point is in most cases an ϵ^{p+1} neighbor point also, we increase p by $\Delta p=5$ to reduce the computing time). In Fig. 3(a), the case of $p=3\Delta p$, the prediction process is stopped around time step 110. This means that if we use $\epsilon^{p=15}$ neighbor points, then after 110 steps, the predicted result is already totally unreliable. No such indicator and information is available if we are using ordinary neighbor points.

To measure the quality of the prediction, the M step rms error

$$e(p)=\frac{1}{M}\sum_{k=1}^{k=M}[x(k)-x'(k,p)]^2, \tag{4}$$

where M is the total prediction step, $x(k)$ is the observed value, and $x'(k,p)$ is the predicted value using ϵ^p neighbor points. Furthermore, the cross-correlation between the series of predicted values and the observed values,

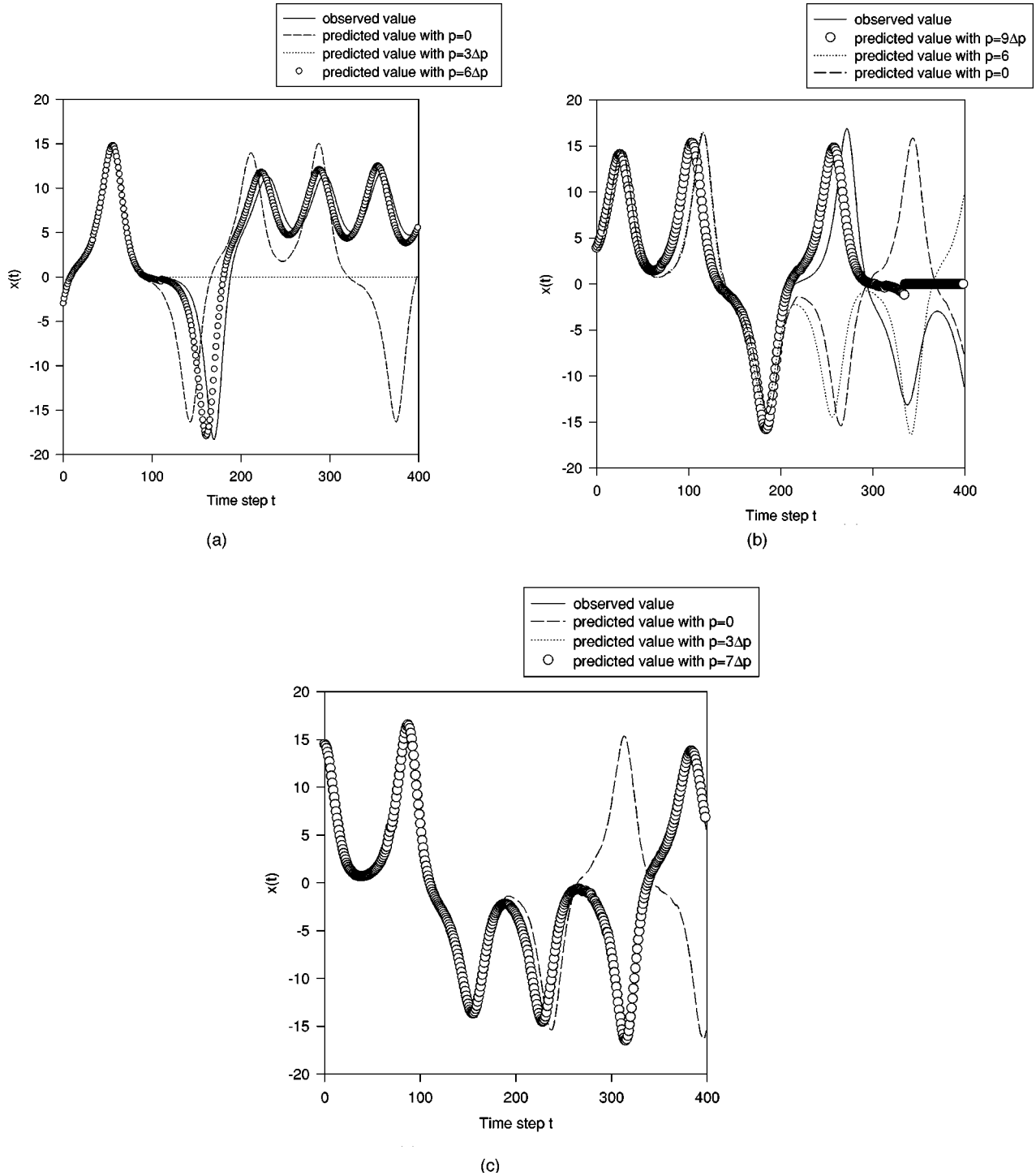

FIG. 3. Predicted and observed values of the x coordinate of the Lorenz system using different p. Three time series with length $N=10\,000$ are selected randomly from a long time series severed as the training set. 400 time steps are predicted. $\Delta p=5$ and $\epsilon=0.5$.

$$C(p)=\frac{\sum_{k=1}^{k=M}[x(k)-\overline{x}][x'(k,p)-\overline{x'}(p)]}{\sqrt{\sum_{k=1}^{k=M}[x(k)-\overline{x}]^2}\sqrt{\sum_{k=1}^{k=M}[x'(k,p)-\overline{x'}(p)]^2}}, \tag{5}$$

are also calculated as a function of p and prediction time k, where $x(k)$ represents the observed values, $\overline{x}$ is the average value of the observed value, $x'(k,p)$ represents predicted values using ϵ^p neighbor points, and $\overline{x'}(p)$ is the corresponding average value.

The results are shown in Figs. 4 and 5. It can be seen clearly that the modified local prediction method improved the performance greatly. Because choosing legitimate neighbor points is the inevitable step for local modeling, our method is thus valuable to more sophistical local prediction methods. Furthermore, it should be pointed out that it cannot be ascertained that this technique will improve the performance greatly for every system, since the rationale behind

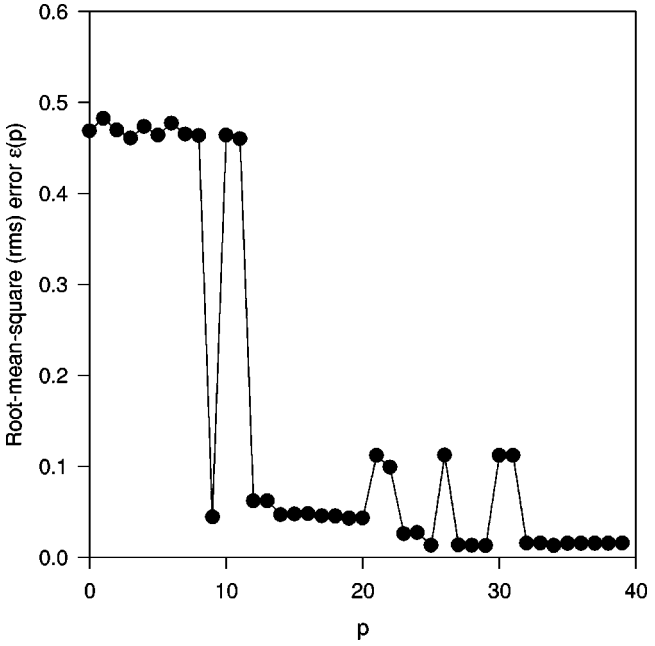

FIG. 4. Root-mean-square (rms) errors are described as a function of p. The training data set is the same as the one used in Fig. 3(b) $\delta p = 1$, and the predicted time step $M = 400$.

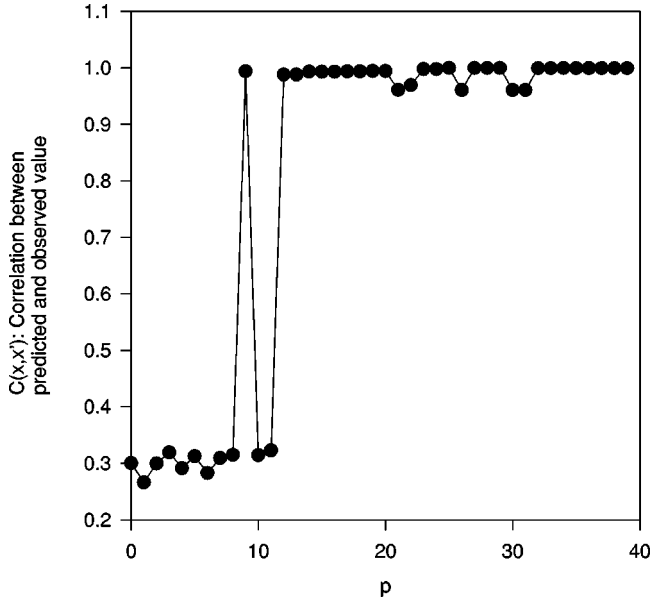

FIG. 5. The cross correlation of the predicted and observed values is described as a function of p. The training data set and the parameters are the same as in Fig. 4.

this technique is based on the existence of pseudofalse neighbor points as defined. However, we believe that the proposed techniques exploits more of the available information in the data sets, and is likely to lead to more satisfactory results. On the practical side, since recurrence is a fundamental characteristics of nonlinear dynamical systems, we can always expect to find pseudofalse neighbors for use in the local prediction.

V. DISCUSSION AND CONCLUSION

An interesting phenomenon is that when the parameter p is systematically increased, there seems to exist a critical value p_c (cf. Figs. 4 and 5), around which the prediction performance will change dramatically. Beyond p_c, there is effectively no change in the performance of the method. How the critical value p_c varies with different reference points, system dynamics, and difference local fitting method needs further study. In our simulation, there seems to be an optimal p value after averaging over the whole attractor. This may suggest the optimal time window for prediction. The problem however is that this ''optimal'' value is sensitive to the diameter of the neighborhood ϵ, and thus the effect of the finite length of the time series should be considered.

Another important problem is the influence of the noise. It is well known that any method based on tangent space is sensitive to noise [21], because both points used to form the tangent vector are inaccurate. In our method, however, this problem is not so serious. As we mentioned above, we should use ϵ^p neighbors to construct our predictor, where p always larger than 0 (ϵ^0 neighbors just the ordinary ϵ neighbors), while the noise can *kick off* the legitimate point from the set of ϵ^p neighbor, there is just little possibility to *kick* a wrong point into our ϵ^p neighbors when p is large. So the noise may decrease the length of the effective prediction, but will not raise the risk of absurd prediction.

In conclusion, we have explored the problem of the pseudofalse neighbor points in phase space for local prediction, and developed a method using the ϵ^p neighbor points instead of ordinary ϵ neighbor points. The improvement of the performance is significant, even for the simplest local constant approximation method. Essentially, this method extends the local prediction by replacing the close-by points by a close-by pattern, i.e., a series of points sequentia in time. This approach utilizes effective the temporal correlation in addition to the spatial correlation, while the ordinary local prediction method uses only the latter. Our results also indicate that there exists a critical value p_c in the time window. It is also important to note that the local prediction method has its intrinsic limitation as mentioned above, e.g., it cannot predict the new pattern but only the existing ones in the training data set. While our method can improve the performance of local prediction, it is not our purpose in this paper to overcome these intrinsic difficulties.

ACKNOWLEDGMENTS

X.G. gratefully acknowledges the research scholarship support at the National University of Singapore.

[1] *Time Series Production: Forecasting the Future and Understanding the Past*, edited by A.S. Weigend and N.A. Gershenfeld (Addison-Wesley, Reading, MA, 1994).
[2] J.D. Farmer and J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).
[3] M. Casdagli, Physica D **35**, 335 (1989).
[4] G. Sugihara and R.M. May, Nature **334**, 734 (1990).
[5] T. Sauer, *Time Series Production: Forecasting the Future and*

Understanding the Past (Ref. [1]), pp. 175–193.
[6] A.L. Lloyd and M.B. Gravenor, Proc. R. Soc. London Ser. B **256**, 157 (1994).
[7] J. Deppish, H.-U. Bauer, and T. Geisel, Phys. Lett. A **158**, 57 (1991).
[8] K. Pawelzik and H.G. Schuster, Phys. Rev. A **43**, 1808 (1991).
[9] E. Castillo and J.M. Gutiérrez, Phys. Lett. A **244**, 71 (1998).
[10] A.K. Alparslan, M. Sayar, and A.R. Atilgan, Phys. Rev. E **58**, 2640 (1998).
[11] M. Giona, F. Lentini, and V. Cimagalli, Phys. Rev. A **44**, 3496 (1991).
[12] R. Brown, Phys. Rev. E **47**, 3962 (1993).
[13] D.J. Walse, Nature **350**, 485 (1991).
[14] D.M. Rubin, Chaos **2**, 525 (1992).
[15] T. Ikeguchi and K. Ahara, Phys. Rev. E **55**, 2530 (1997).
[16] E.A. Wan, in *Time Series Production: Forecasting the Future and Understanding the Past* (Ref. [1]), p. 195.
[17] M. Barahona and Chi-Sang Poon, Nature **381**, 215 (1996).
[18] T. Sauer, J.A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[19] A.M. Fraser and H.L. Swinney, Phys. Rev. A **33**, 1134 (1986).
[20] F. Takens, in *Dynamical Systems and Turbulence*, edited by D.A. Rand and L.S. Young (Springer-Verlag, Berlin, 1981).
[21] H.D.I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, New York, 1996).
[22] M.B. Kennel, R. Brown, and H.D.I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).
[23] E.N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[24] E.N. Lorenz, Tellus, Ser. A **36**, 98 (1984).
[25] A.J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion* (Springer-Verlag, New York, 1983).
[26] D.T. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992).