Evaluation of nonlinearity and validity of nonlinear modeling for complex time series

Tomoya Suzuki,[1] Tohru Ikeguchi,[2] and Masuo Suzuki[3]

[1]*Department of Information Systems Design, Doshisha University, 1-3 Tatara-Miyakodani, Kyotanabe-city, Kyoto 610-0394, Japan*

[2]*Department of Information and Computer Sciences, Saitama University, 225 Shimo-Ohkubo, Sakura-ku, Saitama-city, Saitama 338-8570, Japan*

[3]*Department of Physics, Tokyo University of Science, 1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan*

(Received 16 October 2006; revised manuscript received 29 August 2007; published 1 October 2007)

Even if an original time series exhibits nonlinearity, it is not always effective to approximate the time series by a nonlinear model because such nonlinear models have high complexity from the viewpoint of information criteria. Therefore, we propose two measures to evaluate both the nonlinearity of a time series and validity of nonlinear modeling applied to it by nonlinear predictability and information criteria. Through numerical simulations, we confirm that the proposed measures effectively detect the nonlinearity of an observed time series and evaluate the validity of the nonlinear model. The measures are also robust against observational noises. We also analyze some real time series: the difference of the number of chickenpox and measles patients, the number of sunspots, five Japanese vowels, and the chaotic laser. We can confirm that the nonlinear model is effective for the Japanese vowel /a/, the difference of the number of measles patients, and the chaotic laser.

DOI: 10.1103/PhysRevE.76.046202 PACS number(s): 05.45.Tp, 02.50.−r

I. INTRODUCTION

Analysis methods based on the linear theory, such as frequency spectrum analysis, are used for the prediction and control of complex phenomena [1]. However, discoveries of deterministic chaos indicate that linear modeling is not necessarily suitable for the analysis of such complex behavior. When we consider a complex phenomenon, we are often required to determine whether the phenomenon is linear or nonlinear, or possibly chaotic, only on the basis of a time series. Evaluation of nonlinear predictability is one of the popular methods used to identify deterministic chaos from the viewpoint of a sensitive dependence on initial conditions [2]; if the time series is chaotic, the prediction accuracy decreases exponentially as a prediction step increases. Although nonlinear prediction methods are indispensable for nonlinear time series analysis, the nonlinearity of the time series does not always imply the validity of the application of a nonlinear model to the time series. This is because the nonlinear models are more complex and require a greater number of parameters than the linear model. This indicates that it is very important to consider both the evaluation of the nonlinearity of the time series and the validity of the application of nonlinear modeling to the time series. However, conventional studies (e.g., Ref. [3]) have discussed either the nonlinearity or the validity; therefore, these methods could not recognize the case in which the nonlinear modeling is excessive even if the data exhibit nonlinearity.

In this paper, we propose two measures. The first measure quantifies the nonlinearity of the time series by comparing nonlinear-prediction errors with an optimum linear-prediction error using the statistical inference of the cross-validation (CV) method [4]. Thus, we estimate how the nonlinearity of the time series contributes to the improvement in the prediction accuracy [5].

However, nonlinear modeling generally increases the modeling complexity. Thus, nonlinearity does not always imply that the model is efficient. In other words, the modeling efficiency depends on a trade-off between the complexity of the model and the accuracy of data fitting. In order to discuss the efficiency, we introduce an information criterion—namely, the minimum description length (MDL) [6–9]—to propose the second measure. By using the MDL, the second measure evaluates the advantage of a nonlinear model over an optimal linear model. In the contrast to conventional studies [3], our study evaluates the nonlinearity of the data and the validity of nonlinear modeling by comparing nonlinear modeling and linear modeling in terms of the measures.

To confirm the validity of the two proposed measures, we apply them to the Rössler system [10] with several observational noises. Hence, we confirm that the proposed measures are valid not only for testing the nonlinearity of the original data but also for quantifying the efficiency of nonlinear modeling in comparison with that of optimal linear modeling, even if the observed time series is contaminated with heavy observational noise. In addition, in terms of the application of the proposed measures to real data analysis, we analyze the NH_3-far infrared (FIR) laser [11], annual number of sunspots, the difference of the numbers of measles and chickenpox patients [2], and five Japanese vowels [12,13].

II. EVALUATION OF NONLINEARITY AND VALIDITY OF NONLINEAR MODELING

First, we provide an outline of the linear and nonlinear modeling employed in this study. To examine whether the original data exhibit nonlinearity, it is natural to compare the fitting accuracy of a linear model with that of a nonlinear model. By the Takens embedding theorem [14] and its extension [15], we denote a point in a reconstructed state space as $X(t)=\{x(t),x(t-\tau),\ldots,x(t-(d-1)\tau)\}$. Here, d is an embedding dimension and τ is the delay time. The construction of a model corresponds to the estimation of the function F in $x(t)=F(X(t-\tau))$. If we consider $X(t-\tau)$ as an input set and $x(t)$ as an output, then the linear modeling implies an approximation of F with a hyperplane—that is, an autoregressive (AR) model [1]—while the nonlinear modeling implies an approximation of F with a hypersurface.

In this study, we use the radial basis function (RBF) network [7–9,16] as the nonlinear component of the nonlinear model. The RBF(d,m) network is represented as follows:

$$x(t+\tau) = \sum_{i=1}^{d} a_i x(t-(i-1)\tau) + a_0 + \sum_{j=1}^{m} \lambda_j \Phi(|X(t)-C_j|) + \eta(t), \quad (1)$$

$$\Phi(|X(t)-C_j|) = \exp[-\alpha_j |X(t)-C_j|^2], \quad \eta(t) \sim N(0,\hat{\sigma}^2), \quad (2)$$

where Φ is an RBF that forms a hypersurface and C_j is the center point of Φ. These terms bend the hyperplane locally to form a hypersurface similar to pushing a soft film at C_j, and $\alpha_j = \frac{1}{d+1}\sum_{i=1}^{d+1}|C_j - C_i|$, where C_i is the ith nearest point of C_j; λ_j corresponds to the pressure required to bend the hyperplane locally. Thus, intuitively, the RBF model is a simple extension of the AR model. The modeling implies estimation of the sets $\{a_i\}_{i=0}^{d}$, $\{\lambda_j\}_{j=1}^{m}$, $\{\alpha_j\}_{j=1}^{m}$, and $\{C_j\}_{j=1}^{m}$. The parameters of the model are estimated by least-mean-squares error fitting. The fitting error corresponds to $\hat{\sigma}^2$ in Eq. (2). If $m=0$, Eq. (1) is reduced to the AR(d) model. Generally, a linear model is considerably simpler than a nonlinear model.

Here, let us consider again what is an optimal model. Although we use the RBF terms as the nonlinear component in Eq. (1), the chosen RBF model is merely an approximation. Then, this approximation can never be justified because accurate information regarding the nonlinearity class of the data is unavailable. In addition, even if the information is provided explicitly, estimation of an optimal model in the class requires a large amount of computation, and it possibly becomes an NP-complete problem. However, we can examine the advantage of the nonlinear model if we give priority to the optimization of AR(d) linear terms in Eq. (1) for fitting linear noise over the optimization of the RBF terms for fitting nonlinear noise that could not be fitted using linear terms.

The centers of the RBF network C_j $(j=1,\ldots,m)$ are determined to be the worst-fitted points by the optimal AR(d) model. If the model fitting at $X(T)$ is the worst, a new center point is located at $X(T)$ and a new RBF is set at $X(T)$. Then, the above process is repeated and the errors are calculated. The advantages of nonlinear modeling can be examined by estimating the reduction in the information criteria. Then, if we do not consider the modeling complexity—that is, the number of modeling parameters—we can use as many RBFs as required to build a complex hypersurface for fitting the data to the model. If the fitting accuracy improves due to an increase in the number of RBFs, it can be concluded that the original data are nonlinear.

To estimate the fitting errors obtained by each model, we used the CV method [4], which is one of the resampling schemes [17]. The original data $x(t)$ were divided into k $(k=1,2,\ldots,K)$ parts, where one of the parts was considered as testing data in order to estimate the modeling accuracy and the remaining parts were considered as learning data to be modeled by Eq. (1). To perform a test, we predicted the testing data by the AR(d) model or by the RBF(d,m) network estimated from the learning data of the other $K-1$ parts. Then, we estimated the prediction accuracies $\hat{\sigma}_k^2$ $(k=1,\ldots,K)$ by considering each part of the original data as the testing data and the remaining part as the learning data. Further, we considered their mean value as the final modeling error $\hat{\sigma}^2 = \frac{1}{K}\sum_{k=1}^{K}\hat{\sigma}_k^2$. By this method, we estimated a general fitting error without overfitting the original data. For simplicity, we set $K=2$ in our study. Using this method, we selected d^* for the optimum AR model and $m^*(d)$ for the RBF network.

To quantify the nonlinearity of the original data, we proposed the first measure as follows:

$$E(d) = \frac{\hat{\sigma}^2_{\mathrm{RBF}(d,m^*(d))}}{\hat{\sigma}^2_{\mathrm{AR}(d^*)}}, \quad (3)$$

where $\hat{\sigma}^2_{\mathrm{RBF}(d,m^*(d))}$ is the fitting error obtained by the RBF network for the case $m=m^*(d)$ and $\hat{\sigma}^2_{\mathrm{AR}(d^*)}$ is the fitting error obtained by the optimum AR model for the case $d=d^*$. In Eq. (3), we could use the same embedding dimension d for both the AR model and the RBF network in order to perform a simple comparison of the linear and nonlinear models. However, we compared each optimum model based on its complexity. When d is unsuitable to reconstruct an attractor from the observed time series, it is possible that the linear and nonlinear terms in Eq. (1) interfere with each other; for example, the nonlinear terms model the linearity of the time series. In such a case, nonuniform embedding [7–9] reduces the problem of interference between the linear and nonlinear models, because the nonuniform embedding accurately reconstructs an attractor in a state space. We intend to discuss this possibility in a future study.

However, if the time series does not have any nonlinearity, the optimum linear model must be more accurate than any nonlinear model. That is, the condition $\hat{\sigma}^2_{\mathrm{RBF}(d,m^*(d))} \geqq \hat{\sigma}^2_{\mathrm{AR}(d^*)}$ is satisfied. On the other hand, even if the observed time series has nonlinearity and the linear terms model nonlinearity, the linear model cannot be more accurate than the nonlinear model. Then, the condition $\hat{\sigma}^2_{\mathrm{RBF}(d,m^*(d))} \leqq \hat{\sigma}^2_{\mathrm{AR}(d^*)}$ is satisfied. Thus, by examining whether the measure $E(d)$ is less than 1, we can evaluate whether the original time series exhibits nonlinearity. In addition, even if we cannot completely eliminate the interference between the linear and nonlinear models, we can minimize it by changing d and determine the largest nonlinearity as $E(d)$.

Even if an original time series exhibits nonlinearity, it is not always feasible to employ a nonlinear model because it increases the modeling complexity [18]. Thus, nonlinearity does not always imply the validity of nonlinear modeling. To measure the balance between the complexity and the fitting error of a model, we adopted a major criterion—namely, the MDL [6–9]:

$$M(n) = N \ln \hat{\sigma}^2 + n \ln N, \quad (4)$$

where N is the data length and n is the number of modeling parameters. In this study, to estimate n, we removed ineffective terms whose contribution rates are less than 1[%]. Here, a modeling residual is denoted as $\hat{\sigma}_i$ when we remove the ith linear term of Eq. (1). A modeling residual is denoted as $\hat{\sigma}_j$ when we remove the jth nonlinear term of Eq. (1). Then, the contribution rate of the ith term is defined as the ratio of $\Delta\hat{\sigma}_i(=\hat{\sigma}_i-\hat{\sigma}\geqq 0)$ to $\Sigma_i^d \Delta\hat{\sigma}_i$. The contribution rate of the jth nonlinear term is defined as the ratio of $\Delta\hat{\sigma}_j$ to $\Sigma_j^m \Delta\hat{\sigma}_j$. If we do not introduce the contribution rate, it may be possible that all the terms of the higher-order AR model will be nonzero because the MDL counts each term equally regardless of its contribution. Then, the number of parameters of an optimum model is determined by minimizing the criteria. For the AR(d) model, $M(n_1(d))$ with $n_1(d)=d+1-\phi_1(d)$, where $\phi_1(d)$ is the number of linear terms removed from the model, is minimized in order to obtain the optimum embedding dimension d^*. For the RBF(d,m) network, $M(n_2(d,m))$ with $n_2(d,m)=(d+2)[m-\phi_2(d)]+d+1-\phi_1(d)$, where $\phi_2(d)$ is the number of nonlinear terms removed from the model, is minimized in order to obtain the optimum number of RBFs $m^*(d)$. To quantify the efficiency of nonlinear modeling and estimate d^*, we used a strategy that is different from that in Refs. [7–9]:

$$P(d) = \frac{M(1) - M(n_2(d,m^*(d)))}{M(1) - M(n_1(d^*))}, \quad (5)$$

where $M(1)$ is the value of MDL for the simplest model—that is, $d=0$ and $m=0$. Equation (5) represents the improvement ratio of each optimum model. If the measure $P(d)$ is greater than 1, the nonlinear modeling will be more suitable for the original time series from the viewpoint of the modeling efficiency.

The MDL considers that all the terms of an estimated model have the same weight. Then, even if the contribution rate of a term is small, it is counted equally. On the other hand, the CV method does not consider the number of terms for modeling and it positively adopts the terms that can improve the fitting accuracy of modeling, even if the size of the model increases. That is, the CV method directly confirms whether a hypersurface (nonlinear model) or a hyperplane (linear model) fits data more accurately without overfitting. In other words, the MDL estimates the efficiency of the model from the viewpoint of Occam's razor and the CV method reveals the hidden nonlinearity in the original data.

To summarize the proposed method, the nonlinearity is detected if $E(d)<1$ and the efficiency of nonlinear modeling is detected if $P(d)>1$. However, our method does not always work well to any original data. Even if the data have nonlinearity, it is possible that our method does not detect the nonlinearity—that is, a false-negative response. The false-negative response is caused by the NP-completeness of the problem of finding the true model even if we know a model class perfectly. Therefore, we do not aggressively insist on the linearity and the efficiency of linear modeling of the original data even if any nonlinearity is not detected. On the other hand, our method is robust to the false-positive response which misjudges that linear original data exhibit nonlinearity. If the time series does not have any nonlinearity, the nonlinear model must not be more accurate than the optimum linear model. That is, the condition $\hat{\sigma}^2_{\mathrm{RBF}(d,m^*(d))} \geq \hat{\sigma}^2_{\mathrm{AR}(d^*)}$ is satisfied. This means a robustness to the false-positive response.

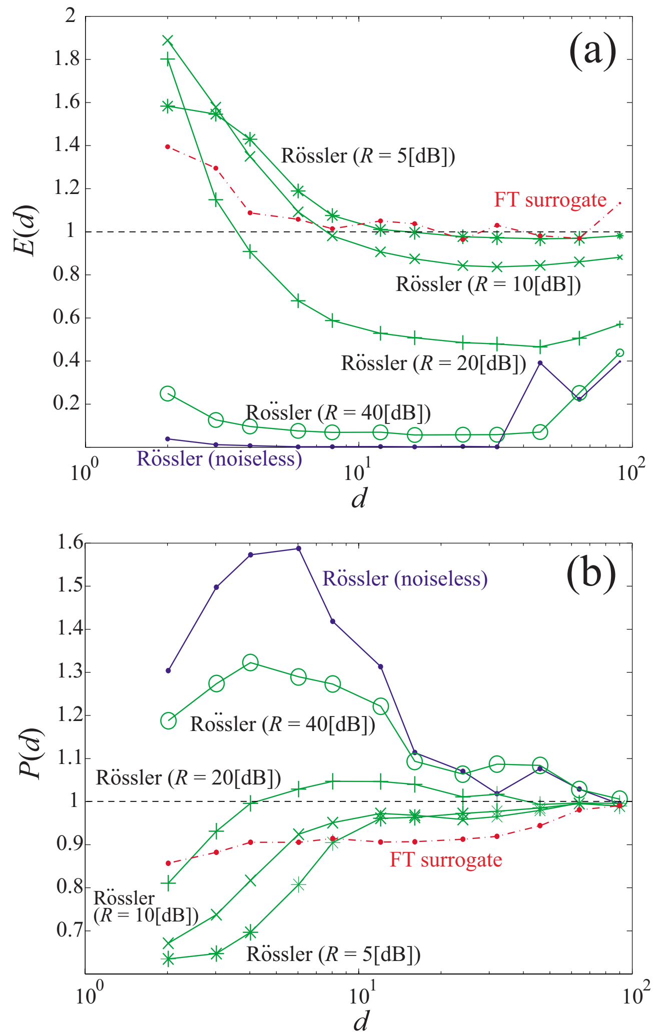

FIG. 1. (Color online) Estimation of (a) nonlinearity with Eq. (3) and (b) efficiency of nonlinear modeling with Eq. (5) for the Rössler model with and without observational noise and its FT surrogate.

III. APPLICATION OF THE PROPOSED MEASURES TO NONLINEAR SYSTEMS

To confirm the validity of the proposed measures, we applied them to the first variable of the Rössler equations [10]—namely, $\dot{x}=-y-z$, $\dot{y}=x+ay$ and $\dot{z}=b+z(x-c)$ ($a=0.36$ $b=0.4$ and $c=4.5$)—as a nonlinear time series, where we set the data length as $N=8000$. Then, we normalized the time series x and set $\tau=15$, which was decided by the autocorrelation function of the observed time series. We also used the Fourier-transformed (FT) surrogates [19] of the Rössler equations for producing a linear time series.

The results are shown in Fig. 1. For the linear data (FT surrogates: dash-dotted lines in Fig. 1), because $E(d)$ is always greater than 1 and $P(d)$ is always less than 1, the nonlinearity and validity of the nonlinear model are not confirmed. The solid lines with stars, crosses, pluses, and solid circles denote the results of the examination of the robustness of the proposed measures against observational noise. We added the Gaussian noise ϵ of the signal-to-noise ratio R [dB] to the Rössler equations. Here, we defined R [dB]$=10 \ln \sigma_x^2/\sigma_\epsilon^2$, where σ_x^2 and σ_ϵ^2 are the variances of x and ϵ.

In the case of nonlinear data, we observed regions where $E(d)<1$ and $P(d)>1$; the data exhibit nonlinearity, and nonlinear modeling is more effective than linear modeling. Moreover, we evaluated the optimum embedding dimension for modeling at which $P(d)$ is maximized. More importantly, when the Rössler data contain a large amount of observational noise, the proposed measures can estimate a degree of nonlinearity and the validity of the nonlinear model of the original data embedded in a d-dimensional state space according to the degree of the deviation of $E(d)$ and $P(d)$ from 1. It is natural that these measures come close to 1 as the amount of noise increases because the nonlinearity is reduced by linear noise.

It has been reported that when the least-squares parameter estimation method is used to model data that are contaminated with observational noise, the nonlinear model cannot be estimated accurately; the estimated nonlinear model tends to have some extra terms with considerably smaller coefficients [22,23]. One of the reasons for obtaining reasonable results as shown in Fig. 1 is that the contribution rates introduced above might function effectively. In a future study, we will examine the influence of model degeneracy on the proposed methods in detail and discuss methods to solve this problem from different viewpoints.

IV. APPLICATION TO REAL DATA ANALYSIS

To demonstrate the application of the proposed measures to real data, we analyzed the NH_3-FIR laser ($N=1000$) [11], the first difference of the number of chickenpox patients ($N=533$) [2], the first difference of the number of measles patients ($N=432$) [2], and the annual number of sunspots ($N=294$) by using $\tau=1$, which was decided by the autocorrelation of the data.

The results are shown in Fig. 2. We confirm that the data of the NH_3-FIR laser and measles patients exhibit nonlinearity because $E(d)<1$ and nonlinear modeling is more effective because $P(d)>1$. Moreover, the appropriate embedding dimensions are small with a maximum value of $P(d)$, as shown in Fig. 2(b). As d becomes large, $E(d)$ and $P(d)$ come close to 1. That is, the nonlinearity and validity of the nonlinear model are weakened. On the other hand, we could not confirm the nonlinearity of the data of the chickenpox patients because $E(d)\geq 1$ for all d. Moreover, the validity of nonlinear modeling is not confirmed because $P(d)\leq 1$ for all d, particularly at small values of d.

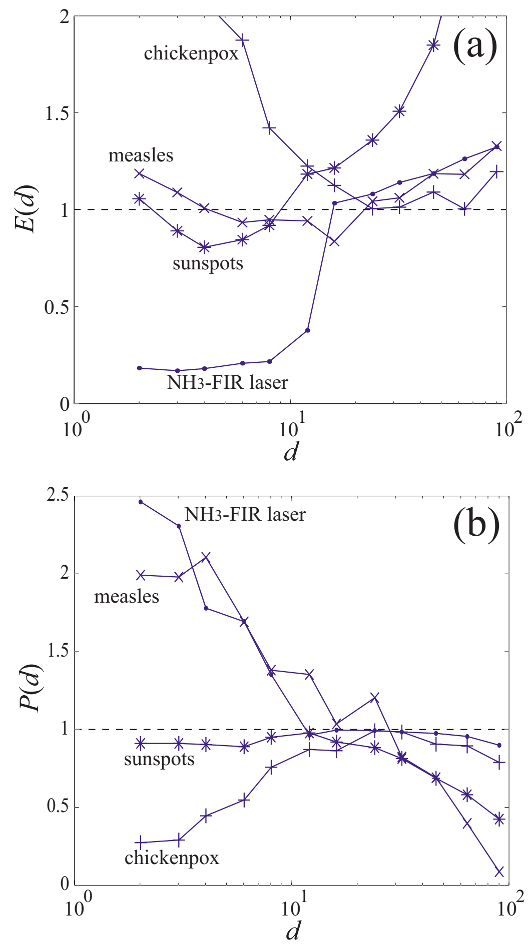

FIG. 2. (Color online) The same as Fig. 1, but for the real data described in the text.

Although the above results are consistent with conventional results, the application of the proposed measures to the sunspot numbers reveals an interesting result: that the time series exhibits nonlinearity because $E(d)<1$ $(2\leq d\leq 8)$. However, the validity of its nonlinear model is not confirmed because $P(d)$ is always less than 1.

Finally, we analyzed five Japanese vowels [13]—/a/, /i/, /u/, /e/, and /o/—as shown in Fig. 3. The data length was $N=8000$. The results are shown in Fig. 4. Conventional studies [12,13,20,21] reported that the Japanese vowels have a nonlinear fluctuation, which is an essential property of their naturalness. These studies indicate that the fluctuation is caused by nonlinear and possibly chaotic dynamics. In Fig. 4(a), we obtain the same results as those in Fig. 1 because $E(d)<1$, particularly at large values of d. However, we could not confirm the validity of the nonlinear model for the vowels /i/, /u/, /e/, and /o/ because $P(d)$ is always less than 1 for the data of these vowels, as shown in Fig. 4(b). Only in the case of the vowel /a/ is $P(d)$ greater than 1 at large values of d.

V. CONCLUSIONS

In this paper, we proposed two measures. The first measure [Eq. (3)] only considers the nonlinearity of data, but

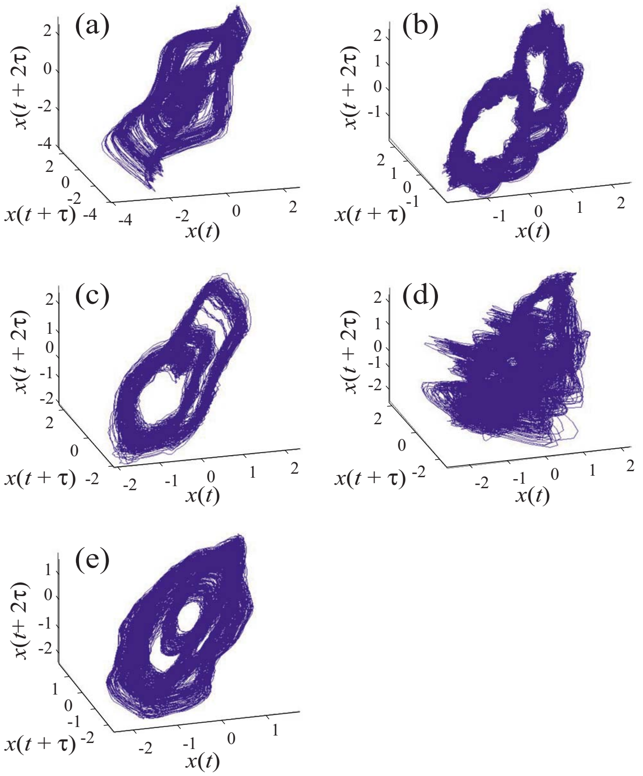

FIG. 3. (Color online) Reconstructions of the five Japanese vowels ($N=8000$): (a) /a/ ($\tau=4$), (b) /i/ ($\tau=15$), (c) /u/ ($\tau=13$), (d) /e/ ($\tau=10$), and (e) /g/ ($\tau=8$) [13].

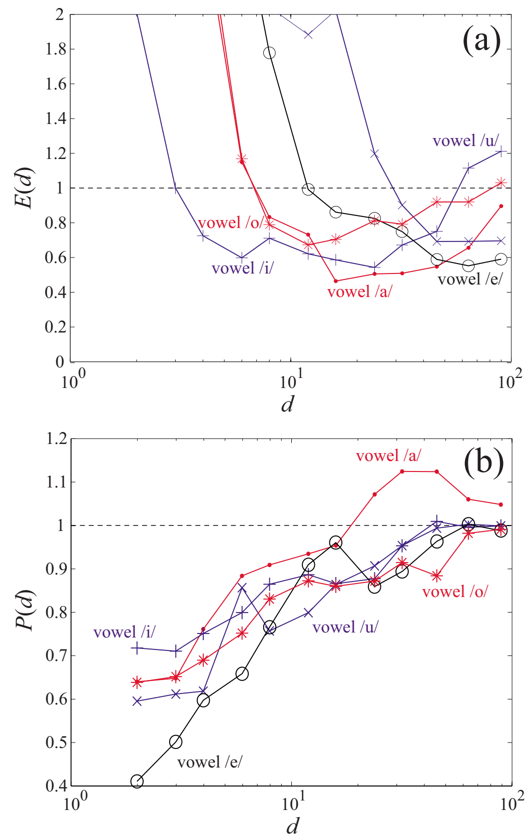

FIG. 4. (Color online) The same as Fig. 1, but for the five Japanese vowels.

does not consider the complexity of data, which allows us to use as many modeling parameters as required to fit a model to the data. If the fitting accuracy is improved, the original data exhibit nonlinearity. On the other hand, the second measure [Eq. (5)] considers both the complexity and efficiency of a model from the viewpoint of the information criteria. We note that these two measures have the possibility of a false-negative response because the problem to be solved or to find the true model belongs to NP-completeness; therefore, we do not aggressively insist on the linearity and the efficiency of linear modeling of the original data. However, as we showed in Sec. II, the measures are very robust to the false-positive response.

By numerical simulations, we confirmed the efficiency of the proposed measures. In addition, the measures can support the existence of observational noise and they are useful in determining the optimum embedding dimension for compact modeling. In fact, we discussed several examples of real data for which nonlinear modeling is not always suitable, even if the data exhibit nonlinearity.

Finally, we would like to raise an important future issue to discuss the availability of nonuniform embedding [7–9] for the proposed methods. Moreover, Ref. [22] reports that the so-called "error in variables" problem in which the least-squares parameter estimation for nonlinear modeling has a significant bias. The residual $\hat{\sigma}^2$ in Eq. (2) may not be estimated accurately. Therefore, the optimal model selected on the basis of the information criteria often tends to be overparametrized [23]. In the future, it is also important to investigate the influence of such problems on the proposed method in order to develop a highly effective algorithm.

ACKNOWLEDGMENTS

The authors thank George Sugihara for providing data on measles and chickenpox patients and Tetsuya Harada for providing data on Japanese vowels. The authors are also grateful to anonymous reviewers for providing several valuable comments. The research of T.I. was partly supported by Grants-in-Aid for Scientific Research (C) (No. 16300072 and No. 17500136) from the Japan Society for the Promotion of Science.

[1] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting* (Springer-Verlag, Berlin, 1996).
[2] G. Sugihara and R. M. May, Nature (London) **344**, 734 (1990).
[3] H. Tong, *Non-Linear Time Series: A Dynamical System Approach* (Oxford University Press, Oxford, 1993).
[4] D. M. Allen, Technometrics **13**, 469 (1971).
[5] T. Miyano *et al.*, Physica D **135**, 305 (2000).
[6] J. Rissanen, *Stochastic Complexity in Statistical Inquiry* (World Scientific, Singapore, 1989).
[7] K. Judd and A. Mees, Physica D **82**, 426 (1995).
[8] M. Small and K. Judd, Physica D **117**, 283 (1998).
[9] T. Nakamura, K. Judd, and A. Mees, Int. J. Bifurcation Chaos Appl. Sci. Eng. **13**, 1263 (2003).
[10] O. E. Rössler, Phys. Lett. **57A**, 397 (1976).
[11] *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld (Addison-Wesley, Boston, 1993).
[12] I. Tokuda *et al.*, J. Acoust. Soc. Am. **110**, 3207 (2001).
[13] T. Ikeguchi and K. Aihara, J. Intell. Fuzzy Syst. **5**, 33 (1997).
[14] F. Takens, in *Detecting Strange Attractors in Turbulence, in Dynamical System and Turbulence*, Lecture Notes in Mathematics, Vol. 898, edited by D. A. Rand and L. S. Young (Springer-Verlag, Berlin, 1981).
[15] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[16] A. Mees, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 669 (1993).
[17] B. Efron, Ann. Stat. **7**, 1 (1979).
[18] M. Small and C. K. Tse, Phys. Rev. E **66**, 066701 (2002).
[19] J. Theiler, Phys. Lett. A **155**, 480 (1991).
[20] T. Ifukube *et al.*, J. Acoust. Soc. Jpn. **47**, 903 (1991).
[21] T. Miyano *et al.*, Int. J. Bifurcation Chaos Appl. Sci. Eng. **10**, 1973 (2000).
[22] P. E. McSharry and L. A. Smith, Phys. Rev. Lett. **83**, 4285 (1999).
[23] K. Judd and T. Nakamura, Chaos **16**, 033105 (2006).