Global modeling of complex data series using the term-ranking approach and its application to voice synthesis

Chao Tao, Xiaojun Liu, and Jack J. Jiang

Phys. Rev. E **84**, 026205 — Published 15 August 2011

DOI: 10.1103/PhysRevE.84.026205

Global modeling of the complex data series using the term-ranking approach and its application to the voice synthesis

Chao Tao* and Xiaojun Liu

Key Lab of Modern Acoustics, Ministry of Education, Nanjing University, Nanjing 210093, People's Republic of China

Jack J. Jiang

Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin Medical School, Madison, WI 53792-7375, USA

*To whom correspondence should be addressed:

Dr. Chao Tao

Institute of Acoustics, Nanjing University,

22 Hankou Road,

Nanjing 210093, P. R. China

Tel: (86)-25-8359-2442

Email address: taochao@nju.edu.cn

Abstract

A term-ranking approach is proposed to globally model the underlying dynamics of chaotic series. The basic idea of this approach is to rank candidate bases before they are used to construct the global model. The ranked bases are involved in the global model one by one in a sequence from high to low until the best model is found. Simulations show that the model obtained by the term-ranking approach has a much longer prediction time, but fewer coefficients than the widely-used standard model. The proposed approach is also successfully applied to coding and synthesis of chaos-like voice data, showing promise for its use with truly noisy experimental data.

PACS numbers: 05.45.Gg, 05.45. Tp, 43.70.+i, 87.19.-j

I. Introduction

Global modeling of a complex data series has continuously received a great deal of attention because of its capabilities in forecasting the evolution of a chaotic system of unknown underlying dynamics [1–4], estimating the largest Lyapunov exponent [5] for sparse series, diagnosing chaos [6], determining the embedding dimension [7], and many other useful functions.

The general idea of global modeling is to choose a flexible functional form with appropriate coefficients to reproduce dynamic behavior of experimental data on the *whole* attractor. A widely used global modeling approach [1-4, 8-9] is to assume that the functional form is a linear superposition of basis functions (e.g. power function), that is [4]:

$$f(x_1, x_2, ..., x_K) = \sum_{l_1, l_2, ..., l_K = 0}^{D} \left(c_{l_1, l_2, ... l_K} \prod_{j=1}^{K} x_j^{l_j} \right), \qquad \sum_{j=1}^{K} l_j \le D. \tag{1}$$

where K is the dimension of the nonlinear system and D is the order of the power function. The number M of terms and coefficients in Eq. (1) is very large: $M = (D+K)!/D!K!$. Ideally (without any error), the exact values of coefficients $c_{l_1, l_2, ..., l_K}$ can be calculated using the least squares technique [3-4], chaos synchronization, and genetic algorithm [10-11]. If the model is not effective, the order D or the dimension K can be increased until a satisfactory model is found. The aforementioned process is called the standard approach and the global model created by this approach will be referred to as the standard model in this study.

A standard model usually must contain a large number of coefficients and terms to approach the unknown underlying dynamics of experimental data. However, due to calculation error and system noise in practice, the more coefficients the model has, the less accurate its coefficient value estimation is. Therefore, the standard approach often fails to describe an observed process, even if the degree and dimension are increased. In short, it can be said that failure of the standard approach is due to numerous inefficient terms and the dynamical effects of overparametrization [12-13].

In order to overcome the limitations of the standard approach, one must decide which function basis should be involved in the global model. Since the mid 1990s, many adaptations have improved the global model with a selective effect basis. Zeroing-and-refitting [14] is one of the easiest methods for model selection, which estimates the parameters of a large model and eliminates terms with sufficiently small parameters. Aguirre and Billings [15] proposed an alternative, simple method, where, instead of deleting specific model terms, entire term clusters are deleted based on behavior of the respective cluster parameters [15-16]. Besides term parameters, other measures of complexity for a given model are defined to address the structure selection problem, such as the entropy [17], the maximum description length [18-20], the minimum description length [7], Bayesian information criterion, and Akaike information criterion [21-22]. Error reduction ratio (ERR) is another practical method to optimize the model structure. By selecting the most relevant regressor according to regressor quality in an orthogonal space [6, 15, 23-25], ERR has been successfully applied to

experimental data. Finally, Aguirre and Letellier have presented a comprehensive review on modeling nonlinear dynamics and chaos [26]. Despite these advances, global modeling of a complex data series is still a challenge of nonlinear science.

Recently, a method was proposed to detect the underlying dynamical structure from short and noise time series [27]. Using this method, the intrinsic dynamical structure, i.e., those terms that actually exist in the original system, can be effectively detected, and global models with intrinsic terms could be constructed. Unfortunately, this method often requires specific embedding space or knowledge of the state vector in order to get intrinsic dynamics. Moreover, even for the intrinsic terms, each of them is not equal but has hierarchies from the aspect of global modeling. If all terms were treated equally, many inefficient terms would be inevitably included, causing model failure.

This study generalizes the concenpt of structure detection [27], and proposes a term-ranking principle to globally model a complex data series. Instead of determining which terms actually exist in the original system, the basis functions are graded according to their possible contribution to the predictability of global models. The purpose of the current study is not to build parcimonious models, but to obtain the model with the best predictability. Therefore, the ranked basis functions are involved in the global model in a one by one, high to low sequence until the longest prediction time has been achieved. The details of the term-ranking principle are presented in Section II. Then, we examine the validation of the proposed method for discrete and continuous data series, respectively. Finally, the term-ranking method is applied to coding and synthesis of real voice signals.

II. GLOBAL MODELING USING THE TERM-RANKING APPROACH

A. Global modeling using the term-ranking approach

It is assumed that experimental data are recorded in the form of a time series $\{s(n)\}_{n=1}^{N}$, where n represents the discrete time, and N is the number of data. Global modeling using the term-ranking principle includes three main steps:

The first step is ranking the candidate bases. If there are M candidate bases used to construct a global model, we write them as $\{\mathbf{\Phi}_{[1]},...,\mathbf{\Phi}_{[m]},...,\mathbf{\Phi}_{[M]}\}$, where the subscript [$m$] shows that the bases are not arranged in hierarchical order. In the standard approach to global modeling, all of the bases are treated equally and directly used to construct the global model. In contrast to this standard approach, the term-ranking approach ranks candidate basis functions before they are used in the global model. For the sake of distinction, we write the ranked bases as $\{\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_m, \ldots, \mathbf{\Phi}_M\}$. The basis function with a small subscript has a high ranking and one with a large subscript has a low ranking, i.e., the ranking of the basis $\mathbf{\Phi}_i$ is hgher than that of the basis $\mathbf{\Phi}_j$ if $i < j$. In subsection II-B, we will introduce how to rank the bases in detail.

The second step is to construct a global function by involving the ranked bases in a one by one, high to low ranking sequence. In order to approach the global dynamics of $s(n)$, a function

$$\mathbf{x}(n+1) = \mathbf{f}[\mathbf{x}(n)] \quad \text{with} \quad \mathbf{x}(n) = [x_1(n), x_2(n), ..., x_K(n)]^T, \tag{1}$$

is constructed using the ranked bases. The state vector $\mathbf{x}(n)$ can be reconstructed from the observed series $s(n)$ using the time-delay method, that is, $x_k(n) = s(n - (k-1)\tau)$ with k=1,

2, …, K, where K is the embedding dimension and τ is lag. The nonlinear function $\mathbf{f}[\mathbf{x}(n)]$ is a linear superposition of the *ranked* basis function $\mathbf{\Phi}_m$:

$$\mathbf{f}[\mathbf{x}(n)] = \mathbf{c}_1\mathbf{\Phi}_1[\mathbf{x}(n)] + \mathbf{c}_2\mathbf{\Phi}_2[\mathbf{x}(n)] + ... + \mathbf{c}_m\mathbf{\Phi}_m[\mathbf{x}(n)], \quad (2)$$

where m = 1, 2, …, M represents the number of terms and coefficients in the function. The values of coefficients $\mathbf{c}_m$ are calculated by the least squares technique, that is, their values are selected so as to minimize the one-step prediction error:

$$\varepsilon^2 = \frac{1}{N}\sum_{n=1}^{N}\left[\mathbf{x}(n+1) - \mathbf{f}\left[\mathbf{x}(n)\right]\right]^2 . \quad (3)$$

The third step is to find the best global model. Prediction ability could be closely associated with application of the global model, such as dimension estimation, entropy calculation, Lyapunov exponent calculation, forecasting, signal coding/decoding, and the chaos test. Therefore, in order to obtain a global model with the best predictability, we also use the prediction time T_{pred} as a criterion to quantitatively measure model quality in this study [4, 28]. With the known coefficients, a model phase orbit $\mathbf{y}(\mathbf{n})$ can be calculated with

$$\mathbf{y}(n+1) = \mathbf{f}[\mathbf{y}(n)] \quad \text{with} \quad \mathbf{y}(1) = \mathbf{x}(1) . \quad (4)$$

When comparing the model phase orbit $\mathbf{y}(\mathbf{n})$ with the original phase orbit $\mathbf{x}(\mathbf{n})$, a prediction time T_{pred} can be calculated. T_{pred} is the time interval in which a root-mean-squared prediction error $\sigma(n)$ (normalized to the standard deviation of the observable s) achieves a certain preset threshold σ_{th}. We choose the threshold value σ_{th}=0.05 for the simulation in Section III, unless otherwise specified. The prediction time

T_{pred} is presented in units of discrete time n in this study.

As shown in Eq.(2), graded basis functions are involved in the global model in a ranking sequence from high (small subscript) to low (large subscript). Once a new term $\mathbf{\Phi}_m$ is involved in the global model, the coefficients of Eq.(2) are estimated and the prediction time T_{pred} of the global model with m bases is also calculated. Finally, the model with the longest prediction time T_{pred} is determined as the optimal global model.

B. Term-ranking method

One pivotal step of the above global modeling approach is in grading the candidate bases. In this study, an algorithm developed from Barahona and Poon's nonlinear detection method [29-30] is used to rank the candidate bases [27]. A similar approach was also used to detect the intrinsic dynamic structure from noise and short chaotic series [27]. This term-ranking method can be described as follows:

Initially, all basis functions are not arranged in order hierarchically, and are written as $\{\mathbf{\Phi}_{[1]}, \ldots, \mathbf{\Phi}_{[m]}, \ldots, \mathbf{\Phi}_{[M]}\}$. One-step prediction functions are constructed with the sum of the first $[m]$ candidate bases,

$$\mathbf{g}_{(m)}(\mathbf{x}) = \mathbf{a}_0 + \mathbf{a}_1\mathbf{\Phi}_{[1]}(\mathbf{x}) + \ldots + \mathbf{a}_m\mathbf{\Phi}_{[m]}(\mathbf{x}). \quad (5)$$

Their coefficients are estimated by solving the least-squares problem $\mathbf{x}(n) = \mathbf{g}_{(m)}[\mathbf{x}(n-1)]$, where the subscript (m) represents the number of bases included in the function. With known coefficients, the one-step prediction series can be calculated by $\tilde{\mathbf{x}}(n) = \mathbf{g}_{(m)}[\mathbf{x}(n-1)]$. The one-step prediction error power $\varepsilon^2(m)$ of the function $\mathbf{g}_{(m)}$ is calculated by

$$\varepsilon(m) = \sqrt{\frac{\sum_{n=1}^{N} \left(\tilde{\mathbf{x}}(n) - \mathbf{x}(n)\right)^2}{\sum_{n=1}^{N} \left(\mathbf{x}(n) - \bar{\mathbf{x}}(n)\right)^2}}, \tag{6}$$

where $\bar{\mathbf{x}}(n)$ is the time average value of $\mathbf{x}(n)$. Involving the candidate bases in the prediction function one by one, the one-step prediction error power $\varepsilon^2(m)$ of a series of function $\mathbf{g}_{(m)}$ (m = 1, 2, …M) is calculated.

The only difference between $\mathbf{g}_{(m-1)}$ and $\mathbf{g}_{(m)}$ is the basis $\mathbf{\Phi}_{[m]}$ in the function $\mathbf{g}_{(m)}$, and therefore, the change of prediction error is highly related to $\mathbf{\Phi}_{[m]}$. Barahona and Poon revealed that if nonlinear terms can significantly reduce the one-step prediction error, this indicates that the underlying dynamics of the data could be nonlinear [29]. Tao et al. extrapolated on this idea [27] and claimed that the terms that can significantly decrease the one-step prediction error, and could be exist in the original system. This concept is generalized in the present study: the changes in prediction error of $\mathbf{g}_{(m)}$ and $\mathbf{g}_{(m-1)}$ actually imply the potential performance of the basis in a global model; therefore, the one-step prediction error difference between $\mathbf{g}_{(m)}$ and $\mathbf{g}_{(m-1)}$ may be used to rank the candidate bases $\mathbf{\Phi}_{[m]}$.

The parameter "information discrepancy" $V(m) = \log\left[\varepsilon(m-1)/\varepsilon(m)\right]$ is used to qualify the changes of one-step prediction error [27], with positive V meaning $\varepsilon(m) < \varepsilon(m-1)$. Moreover, a large V value indicates that its corresponding basis has a high potential in improving the global model. Therefore, the basis corresponding to the highest V value is *permanently* given the 1st (highest) ranking in all candidates and renamed as $\mathbf{\Phi}_1$. The left terms are *temporarily* ranked according to their V value. From high ranking

to low ranking, they are rewritten as $\mathbf{\Phi}_{[2]}$, $\mathbf{\Phi}_{[3]}$,…, $\mathbf{\Phi}_{[M]}$.

To continue, the above process is repeated. For the j-th step, there are j–1 basis functions, which have been permanently ranked in the first (j-1)-steps. They are written as $\mathbf{\Phi}_1$,…, $\mathbf{\Phi}_{j-1}$. Constructing the nonlinear one-step prediction function

$$\mathbf{g}_{(m-1)} = \mathbf{a}_0 + ... + \mathbf{a}_{j-1}\mathbf{\Phi}_{j-1} + \mathbf{a}_j\mathbf{\Phi}_{[j]} + ... + \mathbf{a}_{m-1}\mathbf{\Phi}_{[m-1]}, \tag{7}$$

$$\mathbf{g}_{(m)} = \mathbf{a}_0 + ... + \mathbf{a}_{j-1}\mathbf{\Phi}_{j-1} + \mathbf{a}_j\mathbf{\Phi}_{[j]} + ... + \mathbf{a}_{m-1}\mathbf{\Phi}_{[m-1]} + \mathbf{a}_m\mathbf{\Phi}_{[m]}. \tag{8}$$

The V value of the basis $\mathbf{\Phi}_{[m]}$ ($m = j, \ldots, M$) can be calculated by $V(m) = \log[\varepsilon(m-1)/\varepsilon(m)]$, where $\varepsilon(m-1)$ and $\varepsilon(m)$ are the one-step prediction errors calculated by $\mathbf{g}_{(m-1)}$ and $\mathbf{g}_{(m)}$. The basis function $\mathbf{\Phi}_{[m]}$ ($m = j, \ldots, M$) with the highest V value is permanently given the j^{th}-rank and renamed as $\mathbf{\Phi}_j$. The left basis functions are temporally ranked j+1 to M $\{\mathbf{\Phi}_{[j+1]}, \ldots, \mathbf{\Phi}_{[M]}\}$ according to their V value from large to small.

Finally, repeating the above process M times, the ranking procedure is finished, and all bases are permanently ranked $\{\mathbf{\Phi}_1, \mathbf{\Phi}_2, \ldots, \mathbf{\Phi}_M\}$.

III. NUMERICAL EXAMPLES

Numerical simulations are employed to examine the proposed global modeling approach.

A. Continuous System

The first example is modeling the global dynamics of continuous time series generated by the Rössler system, which is governed by the differential equations:

$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x - c). \end{aligned} \tag{9}$$

The observed time series is $s(n) = y(n\Delta t)$, where $n = 1, 2, \ldots N$ represents the discrete time and $\Delta t = 0.1$ is the sampling interval. The state vector $\mathbf{x}(n)$ is reconstructed from the scalar time series $s(n)$, where the embedding dimension $K = 3$ is determined by using the false nearest neighbor method [31] and the time lag $12\Delta t = 1.2$ is closed to the first zero of the autocorrelation function. It is known that the Rössler system with $a = 0.398$, $b = 2.0$ and $c = 4.0$ will generate chaotic behavior, as shown in Fig. 1 (a). Both the standard approach and the term-ranking approach were used to model globally dynamic behavior of the observable $\{s(n)\}$, with 1000 points used for this modeling.

Figure 1 (b) is the attractor generated by a standard global model. The global model with order $D = 7$ and dimension $K = 3$ is reconstructed from observed data using the standard approach. With the same initial state as the attractor shown in Fig.1(a), the attractor of the standard global model is plotted in Fig.1(b). From comparing Fig.1(b) with Fig.1(a), it is apparent that the attractor of the global model is evidently different from the original attractor. After a short transient stage, the trajectory of the global model is attracted to a limit cycle, i.e., the output of the reconstructed standard model is not chaotic, but periodic. The prediction time of this global model is $T_{\text{pred}} = 102$. The standard global model does not reflect the dynamical characteristics of the original system.

Figure 1 (c) presents the attractor of the global model obtained by the term-ranking approach with the order $D = 7$ and dimension $K = 3$. The state vector $\mathbf{x}(n)$ is also

reconstructed from the scalar time series $s(n)$ using the time-delay method with a time lag $12\Delta t$ = 1.2. All bases are initially involved in the one-step prediction function in a sequence from low power to high power, i.e., $x_1^0 x_2^0 x_3^0$, $x_1^1 x_2^0 x_3^0$, $x_1^0 x_2^1 x_3^0$, $x_1^0 x_2^0 x_3^1$, $x_1^2 x_2^0 x_3^0$, $x_1^1 x_2^1 x_3^0$, …, $x_1^0 x_2^0 x_3^D$. The V value of each candidate is calculated, as shown in Fig. 2(a). The third basis $\Phi_{[3]} = x_1^0 x_2^1 x_3^0$ has the largest V value and it is permanently ranked as 1st. The remaining bases are sorted again according to their V value. For the 62nd steps, the first 61 bases have been ranked. So only the V values of the remainnig 59 bases are calculated, as shown in Fig. 2(b). Finally, the basis with the highest V value of these 59 is permanently given the 62nd rank. This process is repeated 120 times, and all bases are ranked. The ranked terms are then involved one by one in the global model, arranging in a sequence from high to low. Figure 2(c) illustrates the prediction time T_{pred} as a function of the number of bases in the model with the solid (red) circles. The best model is obtained when the first 90 most highly ranked bases are involved in the model. Its prediction time is T_{pred}=2283. With the same initial state as the original attractor, the trajectory generated by the term-ranking model (the first 90 terms) is plotted in Fig.1(c).

It was argued that a bifurcation diagram is more sensitive indice than other tests, such as variance of residuals, prediction error, or a correlation test [15, 32]. Therefore, bifurcation diagrams are also used to assess the model quality. The parameter a of the Rössler system is varied from 0.32 to 0.52. Then, the global model is reconstructed from 1000 recorded time series. By plotting the intersection point [labeled as $y(N)$] of the trajectory and the Poincaré section $y(t) = 0$ in the delay embedding space, the bifurcation

diagram is illustrated in Fig.3. It reveals a number of dynamical regimes in the range of $0.32 \leq a \leq 0.52$, including period, bifurcation, and chaos. The term-ranking global model [Fig.3 (b)] has the same dynamical behavior as the original system [Fig.3(a)].

Prediction time is also illustrated as a function of parameter a in Fig.3 (c). In general, the term-ranking approach can provide a much longer prediction time than the standard model can. Particularly, the prediction time of the term-ranking model is always infinite [the infinite prediction time is expressed as $T_{\text{pred}} = 10000$ in Fig.3(c)] for all the periodic regimes, but is short for the chaotic regime. This is because the chaotic orbit is sensitively dependent on initial conditions, but the stable periodic orbit is not. The term-ranking model accurately reflects the difference between the periodic orbit and the chaotic attractor, while the standard model fails in this aspect, since its prediction horizon often falls short of even the stable periodic orbit.

When $0.43295 < a < 0.556$ for $b = 2$ and $c = 4$ [33], the Rössler system corresponds to the phase non-coherent attractor, which is associated with more complicated topologies. Therefore, global modeling of phase non-coherent attractor is more difficult than the phase coherent attractor. Fig.3, shows that the term-ranking model can still accurately reproduce dynamical characteristics of the original system with a much longer prediction time than the standard model.

The unstable periodic orbit is closely related to the topological characterization of the attractors and could be helpful for controlling dynamical systems [33]. Therefore, reconstruction of the global model from the unstable periodic orbit has important

scientific and practical values. The Rössler attractor with a = 0.398, b = 2, and c = 4 has 24 unstable periodic orbits up to period-11 [33] from which the global model is constructed. As shown in Fig.4, the term-ranking model provides a long prediction time and generally shows better performance than the standard one.

Finally, using model length [Fig.5(a)] and predictive ability [Fig.5(b)], we quantitatively compare the global models obtained by the standard and term-ranking approaches. The model dimension is fixed at K=3 but the order of power function is varied as D = 1, 2, 3, 4, 5, 6, 7, 8. The results obtained by the standard approach are given in Fig. 5 represented by empty (blue) circles. It is apparent that when the order is D = 1, the standard global model has only 4 bases, but its prediction time T_{pred} = 2 is also very short. The model cannot reproduce the dynamical behavior of the original data on the whole attractor. As suggested by the standard approach, increasing order D to 2, 3, 4 gradually increased prediction time to 17, 129, 734, and simultaneously, the number of bases in the model was increased to M = 10, 20, 35. When the order is increased to D = 5, 6, 7, 8, the number of bases increases dramatically to M = 56, 84, 120, 165. However, the added terms do not enlarge the prediction time T_{pred} but reduce it to 493, 257, 102, and 38.

The global model obtained by the term-ranking approach has fewer terms but longer prediction time than the standard model. The results obtained by this approach are given in Fig. 5 with the solid (red) circles. When $D \leq 4$, the term-ranking model has similar prediction time and term number as the standard model. Whereas, when $D > 4$, the

prediction time of the standard model no longer increases with order, while it keeps increasing for the term-ranking model. When order D = 8, the prediction time of the term-ranking model reaches T_{pred} = 2499, which is more than three times longer than the best standard model (T_{pred}=734. K = 3 and D = 4).

This numerical example shows that the term-ranking approach can provide a much better global model of continuous systems than the standard approach.

B. Discrete System

Discrete system modeling (or map modeling) designates the case when a signal is generated by difference equations. Much attention has been paid to map modeling [34], since its potential for not only the signal from the discrete system, but also the return map of the continuous system. In the second example, the performance of the proposed method will be examined in discrete systems. A chaotic signal is generated by a discrete ecological system [35]

$$\begin{aligned} x(n) &= 118\,y(n-1)\exp\big(-0.001y(n-1)-0.001x(n-1)\big), \\ y(n) &= 0.2x(n-1)\exp\big(-0.07y(n-1)-0.07x(n-1)\big) \\ &\quad + 0.8y(n-1)\exp\big(-0.05y(n-1)-0.025x(n-1)\big). \end{aligned} \tag{10}$$

As shown in Fig.6 (a), the trajectory of this discrete ecological system evolves in an attractor composed of fractals in several disconnected domains and visits each of them in a periodic manner as indicated by the arrows. Moreover, Eq. (10) is not in a polynomial form but in an exponential form. The strong periodicity and the exponential structure of this system make global modeling difficult [8]. In this example, it is supposed that both $x(n)$ and $y(n)$ are observed and 1000 points (N = 1000) are used for modeling.

Figure 6(b) gives the prediction time as a function of the number of terms in the model with $K = 2$ and $D = 11$. Using a standard approach, the longest prediction time is only $T_{\text{pred}} = 12$ (the blue empty circle), while using the term-ranking approach, the longest prediction time is $T_{\text{pred}} = 63$ (the red solid circles), which is more than 5 times that of the standard model.

Figure 7 quantitatively compares the performance of the standard approach and the term-ranking approach, where the model dimension is fixed at $K = 2$ and the order D is varied from 1 to 14. As shown in Fig. 7(a), using the standard approach, prediction time increases with order D only when $D \leq 8$. The maximum prediction time of the standard model is about $T_{\text{pred}} = 12$, however, using the term-ranking method, prediction time increases exponentially with D. The relationship between the prediction time T_{pred} and the order D can be approximated by $T_{\text{pred}} = a \times b^D$ with $a = 0.86166$ and $b = 1.44711$ [the green solid line in Fig. 7(a)]. When $D = 14$, the prediction time of the best model dramatically increases to $T_{\text{pred}} = 161$, which is more than 13 times that ($T_{\text{pred}} = 12$) of the best standard model.

A more important advantage of the term-ranking approach is that the model is much shorter than the standard model. For $K = 2$ and $D = 11$ [the case shown in Fig. 6(b)], the standard model has 78 coefficients. The best model obtained by the term-ranking method has only 20 coefficients. Figure 7(b) compares the number of coefficients of the model obtained by the standard approach (the blue empty circles) and term-ranking approach (the red solid circles). When $D = 14$, there are 120 coefficients in a standard model, while

there is only 26% of that, or 32 coefficients total, when using the term-ranking approach model. This example shows that the term-ranking approach can increase the predictive ability of the global model, and simultaneously decrease its length. The ecological system [Eq.(10)] is in an exponential form. Using a Taylor series, an exponential function can be expanded in a polynomial form $e^x = 1 + x + x^2/2!+x^3/3!....$ Resulting in $ye^{x+y} = y + xy + y^2 + x^2y/2 + xy^2 + y^3/2 + x^3y/6 + x^2y^2/2+ xy^3/2+ y^4/6 + \ldots.$ It is apparent by comparing this polynomial with the standard model, that many terms in the model do not belong in this equation. These terms are inefficient and do not make the global model structure more closely related to the unknown underlying dynamics. They decrease the accuracy of coefficient estimation. Therefore, even a great number of terms do not redeem the inefficiency of the standard model. The term-ranking approach, on the other hand, grades each term, and only involves the most efficient terms in the global model, which provides a much longer prediction time with fewer terms than the standard model.

IV. APPLICATION OF THE TERM-RANKING APPROACH TO SPEECH CODING AND SYNTHESIS

Voice production is a highly nonlinear process [36-39]. Rich nonlinear phenomena [40-46], including diplophonia [40-41], period doubling bifurcations [42], chaos, and spatiotemporal chaos [43-46] have been observed in voice. A linear prediction of voice signal is unlikely to precisely code chaos-like voice data and would often miss nonlinear information in the speech signal [47]. Therefore, a nonlinear method could be a natural method for voice coding and synthesis. The aforementioned numerical examples have

shown that the term-ranking approach simultaneously increases the prediction time and reduces the model length, which, from an informatics viewpoint means that the global model obtained by the term-ranking approach can generalize a complex data series using less information. This advantage promises that the proposed term-ranking approach could be applied to coding and synthesizing complex data, such as the chaos-like voice data.

Figures 8(a) and 9(a) illustrate the phase portrait and waveform of the voice signal. The voice data $\{x(n)\}_{n=1}^{N}$ is sampled at 8000 Hz and 16 bits data depth. A delay function

$$\begin{aligned} x(n) &= f\left[x(n-1), x(n-2), ..., x(n-K)\right] \\ &= c_1\phi_1\left[x(n-1), ..., x(n-k)\right] + ... + c_m\phi_m\left[x(n-1), ..., x(n-k)\right] \end{aligned} \tag{11}$$

is used to approach the global dynamics of voice data, where m = 1, 2, …, M and $\phi_m \in \{\prod_{k=1}^{K} y(n-k)^{d_k}\}$ with $\sum_{k=1}^{K} d_k$ = 0, 1, …, D. K = 16 and D = 2 are used in this study. The voice is the result of the joint effects of the vocal fold vibration and the vocal tract filter. The vibration of the vocal folds produces the glottal wave, which excites the end of the vocal tract. A series of partial reflections and transmissions of the acoustic wave happen in different parts of the vocal tract. This filtering effect occurring in the vocal tract can be represented by the delay terms according to the source filter theory of voiced sounds [48]. Therefore, in order to find the optimal model of a voice signal, a large range of K is used. The total term number is $M = (K + D)!/K!D!$=153 and 500 voice data points are used to construct the global model. As shown in Fig. 8(b) with blue empty circles, the prediction time of the standard model gradually increases when the first 128 terms are involved one by one. More terms do not increase the prediction time but rather

decrease it, and the maximum prediction time $T_{pred} = 18$ is obtained by the standard approach.

Using the term-ranking approach, the ranked terms are involved in the global model one by one in the ranking sequence from high to low, and the calculated prediction time is given in Fig. 8(b),represented with red solid circles. Highly-ranking terms rapidly increase the prediction time of the model, and yet the remaining low ranking terms hardly improve the predictive ability of the global model. Only when including the first 51 highest ranking terms, the term-ranking model surpasses the standard model in prediction time. In other words, the term-ranking approach can generalize the voice data using less information (i.e., model coefficients) than the standard one. Based on this factor, two kinds of voice coding and synthesis schemes are presented.

The first is a lossless scheme. Initially, a global model [Eq. (11)] of voice data $x(n)$ is created. The information, carried by the original voice series $x(n)$, can be precisely recovered by the model's coefficients c_m and the one-step prediction error [i.e., residual error $\xi(n) = x(n) - f[x(n-1),\ldots, x(n-K)]$:

$$x'(n) = f[x'(n-1),...,x'(n-K)] + \xi(n), \tag{12}$$

with the initial condition $x'(1) = x(1)$, …, $x'(K) = x(K)$. Clearly, the synthesized voice data $x'(n)$ is exactly equal to the original data $x(n)$, as shown in Fig. 9(b), and therefore, this is a lossless method.

The number of the global model coefficients are much fewer than the voice data and the residual error $\xi(n)$ is much smaller than the original data ($\|\xi(n)\| << \|x(n)\|$). For a

given resolution, fewer bits are needed to encode the residual error and the model coefficients than the original voice data. The model obtained by the term-ranking approach has fewer coefficients and smaller residual error than the standard model, thus it can encode voice data with fewer bits than the standard one. This point can be quantitatively measured by Bayesian information criterion (BIC) [21],

$$C_{BIC} = \ln(\varepsilon^2) + m\ln(N)/N, \tag{13}$$

where ε^2 is the error power, N is the data length, and m is the number of free coefficients in global model. A lower C_{BIC} value implies either fewer explanatory variables, better fit, or both. As shown in Fig. 8(c), the C_{BIC} value of the term-ranking model (the red solid circles) is smaller than that of the standard model (the blue empty circles). That is, the term-ranking approach could provide a more efficient lossless compression method of voice data than the standard approach.

The second is a lossy scheme. The term-ranking approach not only significantly decreases the one-step prediction error (residual error), but also increases the prediction time, and the lossy coding scheme is based on the short-term prediction capability of the global model.

In this scheme, a global model of the voice is built and the output of that model is used directly as the synthesized voice. However, because the initial sensitive properties of the chaos-like voice signal, the short-term prediction of the global model is soon divergent from the original data. Therefore, the original voice data is used to correct the prediction of the model every Δn step, that is,

$$x'(n) = (1 - \delta(n \bmod \Delta n)) f[x'(n-1), ..., x'(n-k)] + \delta(n \bmod \Delta n) x(n), \qquad (14)$$

where $\delta(n) = 1$ if $n = 0$ and $\delta(n) = 0$ if $n \neq 0$. The term-ranking model involving the first 46 terms with the highest ranking is used to reproduce the voice data. The result is given in Fig. 9(c), where $\Delta n = 3$. It is apparent that the synthesized waveform is quite close to the original one. For the sake of comparison, the standard model involving 46 terms is also used to synthesize the voice data, but because of its poor predictive ability, the synthesized waveform is significantly different from the original data and is finally divergent, as shown in Fig. 9(d).

The above examples indicate that the term-ranking approach is robust enough for application with experimental data. Voice coding and synthesis could be one potential application of the term-ranking global modeling approach.

V. CONCLUSION

In this study, a term-ranking principle is proposed to improve the global model and its predictability of complex time series. Our basic idea of global modeling is to rank the candidate bases before they are involved in the model. The ranked bases are then included in the global model one by one, in a high to low ranking sequence until the best model is found. Simulations show that the model obtained by the term-ranking approach has a much longer prediction time and much fewer coefficients than the widely used standard model. These numerical simulations display the advantages of the term-ranking approach in globally modeling the underlying dynamics of the complex data series.

The proposed method is also applied to real voice data, where the sampling error, measurement noise, and stochastic component are inevitable. The successful synthesis of the real voice signal promises that the term-ranking approach is robust enough for application to noisy experimental data. Therefore, it can be expected that the term-ranking approach could be widely used for the analysis and processing of chaos-like natural data, such as modeling, coding, compression, prediction, and so on.

Finally, the power function is used as the candidate bases in this study. However, it is found that the polynomial is sometimes not powerful enough to model some time series, such as the z variable of the Rössler system. Letellier and Aguirre et al. show that the main reason for this is due to the nature of the map between the original phase space and the embedding space induced by the observable signal [49]. They introduce "observability coefficients" to understand the difference between the system variables. (e.g., the y and z variables of the Rössler system) [49-52]. In order to model such a variable, the rational functions have to be used in the global model [53]. In addition, Ménard, Letellier et al have shown that a rational function can significantly improve the quality of global models for maps, because the models take advantage of the ability of rational functions to capture complicated structures that may be generated by maps [34]. Their method is also applied to experimental data from copper electrodissolution. Therefore, it would be worth extending the term-ranking approach to more general bases in the future.

Acknowledgement. This study was supported by NSF of China (No. 10904069 and No. 10874088), NSF of Jiangsu, China (No. SBK201021985), SRFDP 200802840032, and program for NCET. We thank James Madsen for his help in preparing the manuscript.

REFERENCES

[1] E. Baake, M. Baake, H.G. Bock, and K. M. Briggs, Phys. Rev. A **45**, 5524 (1992).

[2] R. Brown, N. F. Rulkov, and E. R. Tracy, Phys. Rev. E **49**, 3784 (1994).

[3] G. Gouesbet and C. Letellier, Phys. Rev. E **49**, 4955 (1994).

[4] B. P. Bezruchko and D. A. Smirnov, Phys. Rev. E **63**, 016207 (2000).

[5] D. J. Wales, Nature (London) **350**, 485 (1991).

[6] L. A. Aguirre and S. A. Billings, Int. J. Bif. and Chaos **5**, 449 (1995).

[7] Y. I. Molkov, D. N. Mukhin, E. M. Loskutov, and A. M. Feigin, G. A. Fidelin, Phys. Rev. E **80**, 046207 (2009).

[8] C. Tao, Y. Zhang, G. Du, J. J. Jiang, Phys. Lett. A **332**, 197 (2004).

[9] L. A. Aguirre and E. C. Furtado, Phys. Rev. E, **76**, 046219 (2007).

[10] C. Tao, Y. Zhang, G. Du and J. J. Jiang, Phys. Rev. E **69**, 036204 (2004).

[11] C. Tao, Y. Zhang, J. J. Jiang, Phys. Rev. E **76**, 016209 (2007).

[12] L. A. Aguirre and S. A. Billings, Physica D **80**, 26 (1995).

[13] E. M. A. M. Mendes and S. A. Billings, Int. J. Bif. and Chaos **8**, 535 (1998).

[14] J. B. Kadtke, J. Brush, and J. Holzfuss, Int. J. Bif. and Chaos **3**, 607 (1993).

[15] L. A. Aguirre and S. A. Billings, Int. J. Control **62**, 569 (1995).

[16] L. A. Aguirre, R. A. M. Lopes, G. F. V. Amaral, and C. Letellier, Phys. Rev. E **69**, 026701 (2004).

[17] J. P. Crutchfield and B. S. McNamara, Complex Systems **1**, 417 (1987).

[18] J. Rissanen, Stochastic Complexity in Statistical Inquiry, vol. 15 of Series in

Computer Science, World Scientific, Singapore, (1989).

[19] K. Judd and A. Mees, Physica D, 82, 426 (1995).

[20] M. Small and C. K. Tse, Phys. Rev. E **66**, 066701 (2002).

[21] G. E. Schwarz, Ann. Stat. **6**, 461 (1978).

[22] H. Akaike, IEEE Trans. Autom. Control **19**, 716 (1974).

[23] S. A. Billings, S. Chen, and M. J. Korenberg, Int. J. Control **49**, 2157 (1989).

[24] G. Kerschen, B. F. Feeny, J.-C. Golinval, Comput. Methods Appl. Mech. Eng. **192**, 1785 (2003).

[25] L. A. Aguirre and S. A. Billings, Physica D **85**, 239 (1995)

[26] L. A. Aguirre, C. Letellier, Mathematical Problems in Engineering 2009, 238960 (2009).

[27] C. Tao, X. Liu, and G. Du, Phys. Rev. E **81**, 046209 (2010).

[28] B. P. Bezruchko, T. V. Dikanev, D.A. Smirnov, Phys. Rev. E. **64**, 036210 (2001).

[29] M. Barahona and C.-S. Poon, Nature (London) **381**, 16, 215-217 (1996).

[30] C.-S. Poon, and C. K. Merrill, Nature (London) **389**, 492–495 (1997).

[31] M. B. Kennel, R. Brown, H. D. I. Abarbanel, Phys. Rev. A, **45**, 3403 (1992).

[32] C. Letellier, L. Le Sceller, P. Dutertre, and G. Gouesbet, Z. Fei, J. L. Hudson, J. Phys. Chem. **99**, 7016 (1995).

[33] C. Letellier, P. Dutertre, and B. Maheu, Chaos **5**, 271 (1995).

[34] O. Ménard, C. Letellier, J. Maquet, and G. Gouesbet, Phys. Rev. E. **62**, 6325 (2000).

[35] B. Cazelles and R. H. Ferriere, Nature **355**, 25 (1992).

[36] R. Laje and G. B. Mindlin, Phys. Rev. Lett. **91**, 258104 (2003).

[37] R. Laje, T. Gardner, G. B. Mindlin, Phys. Rev. E **64**, 056201 (2001).

[38] M. A. Trevisan, J. M. Mendez, and G. B. Mindlin, Phys. Rev. E **73**, 061911 (2006).

[39] M. A. Trevisan, B. Cooper, F. Goller, and G. B. Mindlin, Phys. Rev. E **75**, 031908 (2007).

[40] P. Mergell, H. Herzel, and I. R. Titze, J. Acoust. Soc. Am. **108**, 2996 (2000).

[41] J. Neubauer, P. Mergell, U. Eysholdt, and H. Herzel, J. Acoust. Soc. Am. **110**, 3179 (2001).

[42] W. Mende, H. Herzel., and K. Wermke, Phys. Lett. A **145**, 418 (1990).

[43] I. Tokuda, T. Riede, J. Neubauer, M. J. Owren, H. Herzel, (2002). J. Acoust. Soc. Am. **111**, 2908 (2002).

[44] Y. Zhang and J. J. Jiang, Phys. Rev. E **72**, 035201 (2005).

[45] A. Facchini, H. Kantz, and E. Tiezzi, Phys. Rev. E **72**, 021915 (2005).

[46] C. Tao and J. J. Jiang, Phys. Rev. E **77**, 061922 (2008).

[47] C. Tao, M. Jing, X. Xu and G. Du, Technical Acoustics 21, 175 (2002).

[48] M. A. Trevisan, M. C. Eguia, G. B. Mindlin, Phys. Rev. E **63**, 026216 (2001).

[49] C. Letellier, L. A. Aguirre, and J. Maquet, Phys. Rev. E **71**, 066213 (2005).

[50] L. A. Aguirre, IEEE Trans. Educ. **38**, 33 (1995).

[51] C. Letellier, J. Maquet, L. Le Sceller, G. Gouesbet, and L. A. Aguirre, J. Phys. A

31, 7913 (1998).

[52] C. Letellier, L. A. Aguirre, Phys. Rev. E **82**, 016204 (2010).

[53] C. Lainscsek, C. Letellier, and I. Gorodnitsky, Phys. Lett. A **314**, 409 (2003).

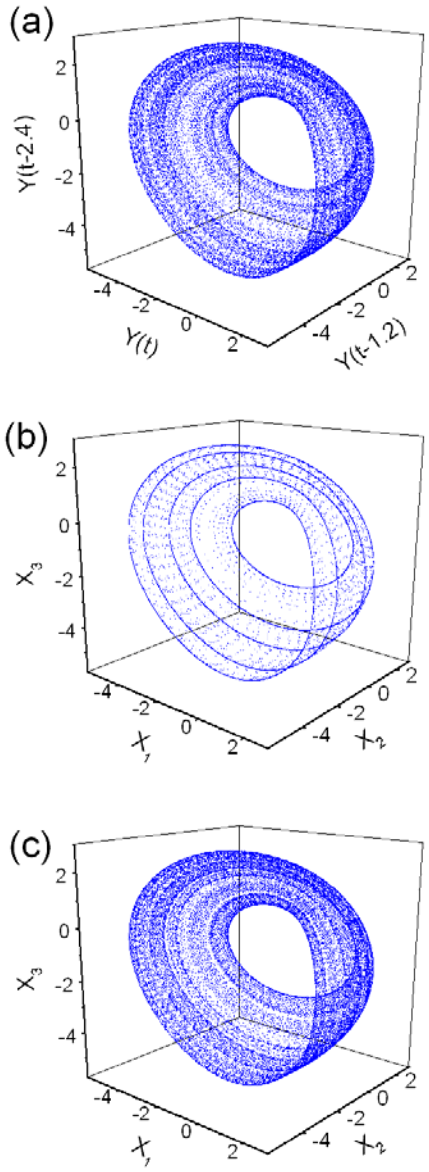

Figure. 1 (color online): Reconstruction of the Rössler attractor. (a) Delay representation of the original Rössler attractor with the lag $12\Delta t$ and the embedding dimension $K = 3$. (b) Attractor generated by the standard model with $K = 3$ and $D = 7$, where 120 coefficients are involved in the model. (c) Attractor generated by the global model obtained by the term-ranking approach with $K = 3$ and $D = 7$, where 90 terms are involved in the model.

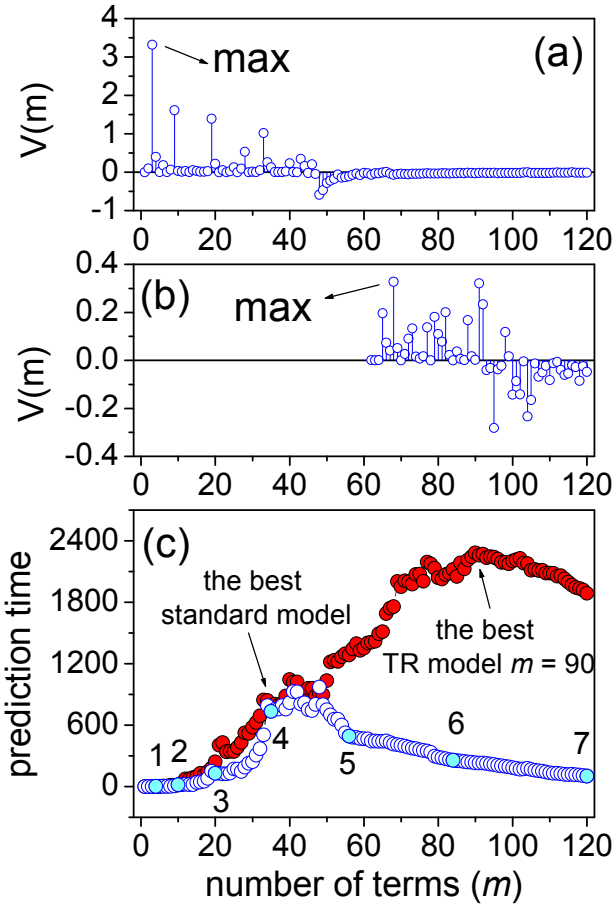

Figure 2 (color online): Global modeling of the Rössler attractor. (a) The $V(m)$ value of each candidate basis. (b) The $V(m)$ value of the left candidate bases, after the first 61 high ranking bases have been found. (c) The prediction time T_{pred} as a function of the number m of terms in the model with $K = 3$ and $D = 7$, where the empty (blue) scatters and the dark (red) scatters represent the results obtained by the standard approach and the term-ranking approach, respectively.

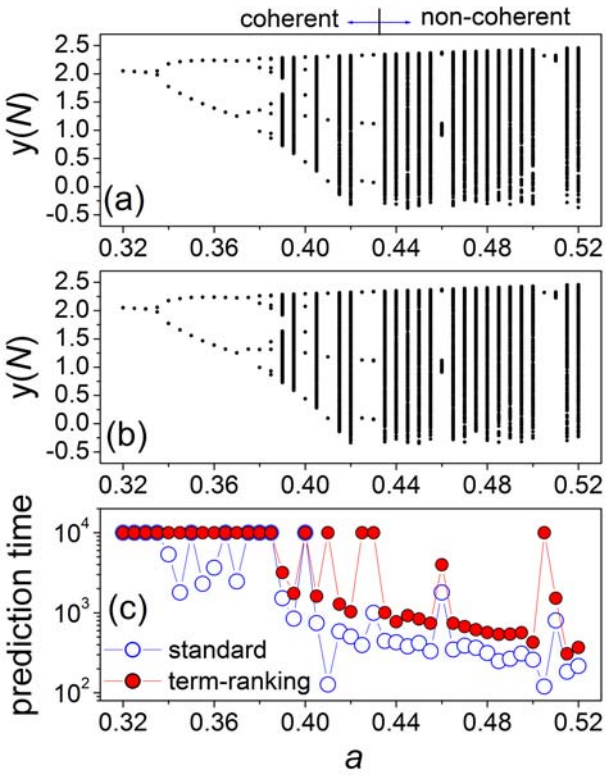

Figure 3 (color online): Revealing the dynamic characteristics of the Rössler system with $0.32 \leq a \leq 0.52$, $b = 2$ and $c = 4$. For $0.43295 < a < 0.556$, $b = 2$ and $c = 4$, the Rössler system corresponds to phase non-coherent attractor. (a) Bifurcation diagram of the Rössler system. (b) Bifurcation diagram of the global model reconstructed using the term-ranking method. (c) The prediction time as a function of the parameter a, where $T_{\text{pred}} = 10^4$ represents that the prediction time approach infinite.

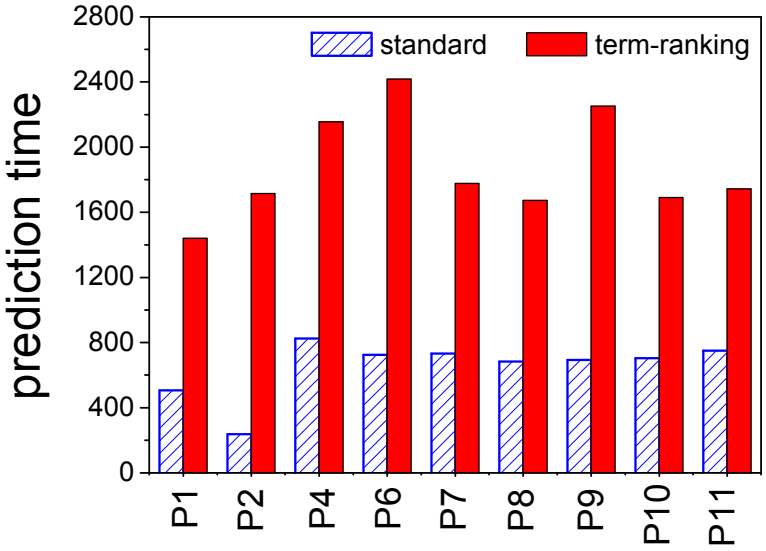

Figure 4 (color online): Reconstruction of global model from the unstable periodic orbit of the Rössler attractor with a = 0.398, b = 2, and c = 4, where there are 24 unstable periodic orbits up to period-11. The abscissa P1, …, P11 represents the unstable periodic orbits from period-1 to period-11.

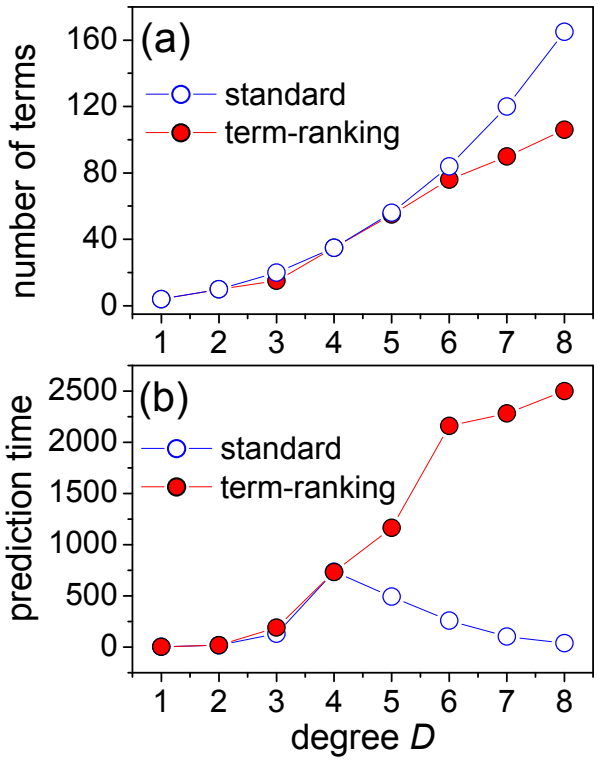

Figure 5 (color online): Comparison of the standard approach (empty blue circles) and the term-ranking approach (red solid circles) with $K = 3$ and $D = 1, 2, \ldots, 8$. (a) The number of terms (i.e. the number of coefficients) in the global model. (b) The prediction time T_{pred} as a function of degree D.

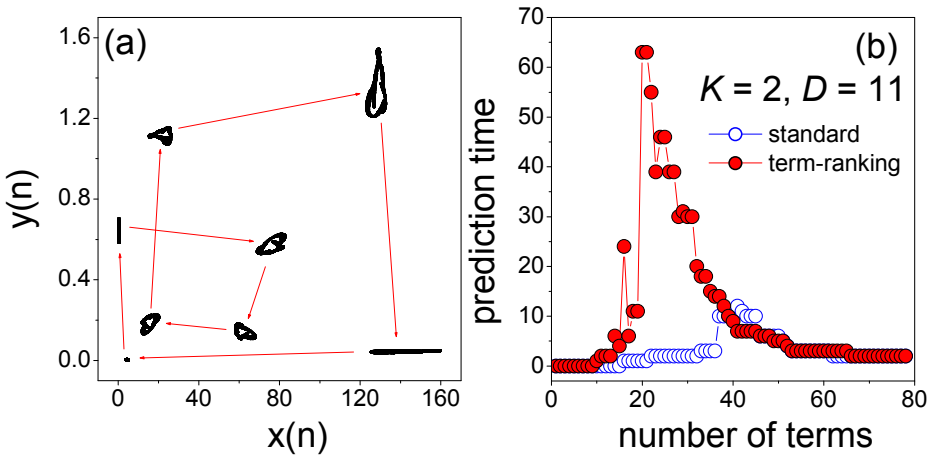

Figure 6 (color online): Global modeling of a discrete ecological system. (a) The attractor of the ecological system. (b) The prediction time T_{pred} as a function of the number of terms in the model with $K = 2$ and $D = 11$, where the empty (blue) circles and the solid (red) circles represent the results obtained by the standard approach and the term-ranking approach, respectively.

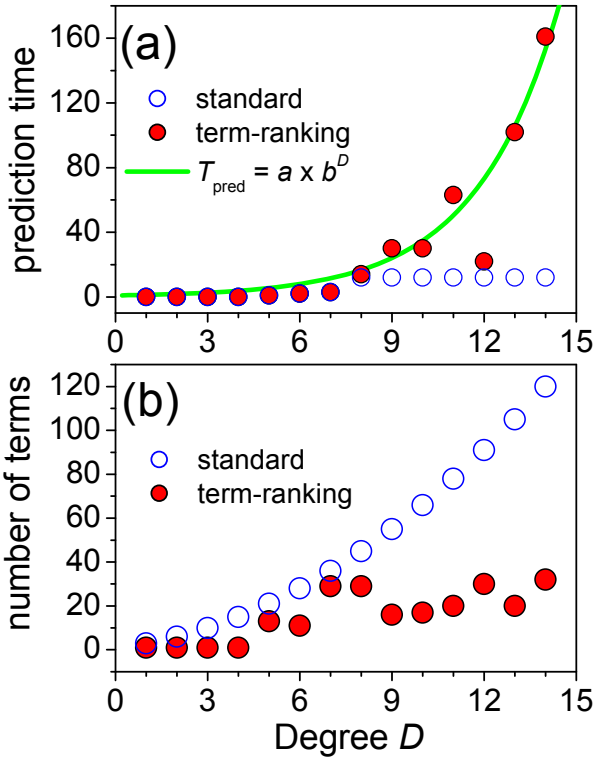

Figure 7 (color online): Comparison of the standard approach (empty blue circles) and the term-ranking approach (solid red circles) with $K = 2$ and $D = 1, 2, \ldots, 14$. (a) The prediction time T_{pred} as a function of degree D. The green solid line follows the relationship $T_{pred}=a\times b^D$ with $a = 0.86166$ and $b = 1.44711$. (b) The number of terms (i.e. the number of coefficients) in the global model.

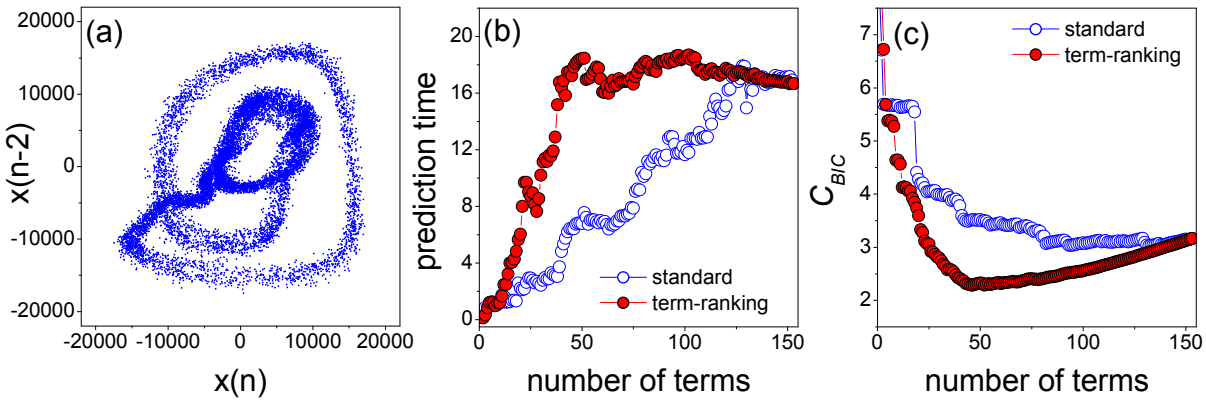

Figure 8 (color online): Global modeling of voice data series. (a) Delay representation of the attractor of the voice series. (b) The prediction time T_{pred} as a function of the number of terms in the model with $K = 16$ and $D = 2$, where the threshold value $\sigma_{\text{th}} = 0.2$ is used for the calculation of the prediction time. (b) Bayesian information criterion C_{BIC} as a function of the number of terms in the model. In (b) and (c), the empty (blue) scatters and the dark (red) scatters represent the results obtained by the standard approach and the term-ranking approach, respectively.

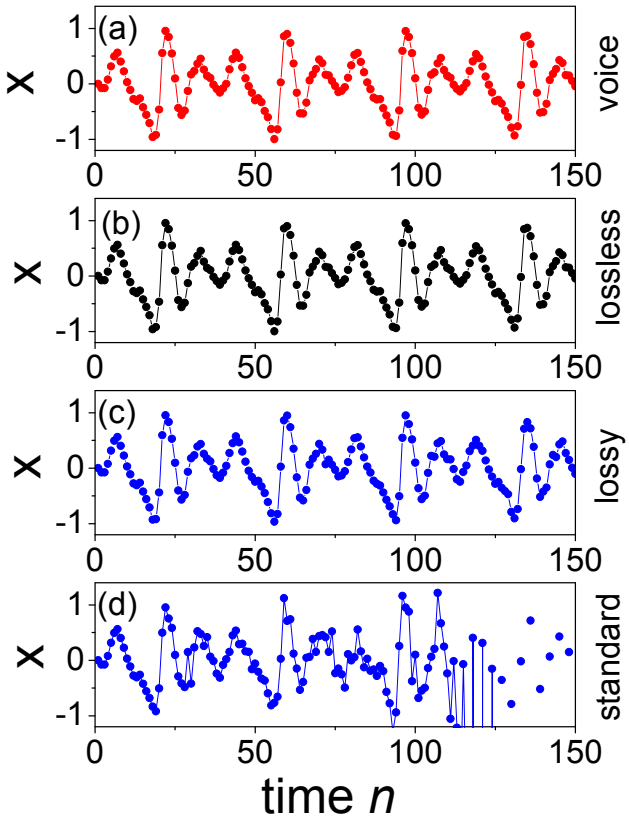

Figure 9 (color online): Synthesis of voice signal. (a) Original voice data. (b) The lossless synthesis of voice using the term ranking model. (c) The lossy synthesis of voice using the term ranking model. (d) The lossy synthesis of voice using the standard model.