Building dynamical models from data and prior knowledge: The case of the first period-doubling bifurcation

Luis Antonio Aguirre* and Edgar Campos Furtado
Laboratório de Modelagem, Análise e Controle de Sistemas Não-Lineares, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, M.G., Brazil
(Received 12 April 2007; published 23 October 2007)

This paper reviews some aspects of nonlinear model building from data with (gray box) and without (black box) prior knowledge. The model class is very important because it determines two aspects of the final model, namely (i) the type of nonlinearity that can be accurately approximated and (ii) the type of prior knowledge that can be taken into account. Such features are usually in conflict when it comes to choosing the model class. The problem of model structure selection is also reviewed. It is argued that such a problem is philosophically different depending on the model class and it is suggested that the choice of model class should be performed based on the type of *a priori* available. A procedure is proposed to build polynomial models from data on a Poincaré section and prior knowledge about the first period-doubling bifurcation, for which the normal form is also polynomial. The final models approximate dynamical data in a least-squares sense and, by design, present the first period-doubling bifurcation at a specified value of parameters. The procedure is illustrated by means of simulated examples.

DOI: 10.1103/PhysRevE.76.046219 PACS number(s): 05.45.−a, 05.10.−a, 02.70.−c, 07.05.Tp

I. INTRODUCTION

Model building from data was rather a mature field by the mid 1990s in what concerns linear discrete-time models [1]. Nonlinear discrete-time representations appeared in the 1980s with polynomials [2], radial basis functions [3], and multilayer perceptron neural networks [4]. Other model classes include fuzzy models [5] and wavelets [6]. The first works applied to nonlinear dynamics appeared in the late 1980s and early 1990s [7–11].

The aforementioned papers were followed by many others in which diverse aspects of model building were addressed in the context of particular model classes. Some of such references will be quoted throughout this paper, whenever appropriate.

In the field of nonlinear dynamics, it seems possible to distinguish two broad classes of problems which shall be referred to as *time series modeling* (TSM) and *dynamical system modeling* (DSM). The former usually aims at problems such as forecasting, classification, geometrical, and statistical characterization of time series. The latter, on the other hand, aims at dynamical analysis, geometrical and topological characterization of dynamical systems, reconstruction of bifurcation diagrams, and the like. Although the intended application of the resulting models is different in each case, the algorithms used are often the same. Consequently, there is a gray area in between these two classes and a clearcut division is neither always possible, nor necessary. However, to recognize some of the subtle differences underlying such problems is usually convenient.

In this paper, various aspects of model building will be discussed in the context of the two problems mentioned in the previous paragraph. This will help appreciate the differences that exist among some alternatives. One of the key issues in this paper will be the use of prior knowledge in model building. It has been acknowledged that such an issue deserves greater attention in the literature [12]. In addressing this matter, it will be seen that to incorporate prior knowledge into models has received some attention in DSM. The types of information that have been used to build models will be briefly reviewed and a new type of prior information will be investigated, namely the presence of a periodic-point flip bifurcation. Therefore, a procedure will be proposed to build a model from dynamical data on a Poincaré section such that one of its periodic points will undergo a period-doubling (flip) bifurcation at a specified value of the parameter vector.

The remainder of the paper is organized as follows. Section II surveys some aspects of model building from data with and without prior knowledge. In Sec. III a new procedure for constrained model building is presented. The resulting model is constrained to have a period-doubling bifurcation besides approximating the dynamical data in some least-squares sense. The new procedure is illustrated and investigated in the light of numerical simulations in Sec. IV. The main conclusions of the paper are provided in Sec. V.

II. AN OVERVIEW OF SOME ASPECTS IN MODEL BUILDING

In what follows a rather general setting will be used to describe the main steps in model building from data produced by a nonlinear dynamical system. Then, assumptions will be made in order to turn the problems manageable. It is hoped that by addressing the model building problem in this way some of the potential pitfalls, due to the assumptions made, will become more clear.

It is assumed that data $\mathcal{Z}$ and possibly some auxiliary information $\mathcal{I}$ about a system $\mathcal{S}$ is available. The black-box model building problem consists of building a mathematical model $\mathcal{M}$ from the available data $\mathcal{Z}$. If, in addition to $\mathcal{Z}$, $\mathcal{I}$ is also used to determine $\mathcal{M}$, the procedure is often referred to

*FAX: +55 31 3409-4850. aguirre@cpdee.ufmg.br

as gray-box modeling. In either case, $\mathcal{M}$ should approximate the system $\mathcal{S}$ in some sense. In the gray-box case, apart from approximating the system $\mathcal{S}$, the model $\mathcal{M}$ should either exactly or approximately incorporate the auxiliary information used.

A. Model building as an optimization problem

The problem of building a model $\mathcal{M}$ that approximates the system $\mathcal{S}$ can be cast as an optimization problem. For the sake of argument, assume there is a cost function $J(\mathcal{S},\mathcal{M})$ which should be minimized with respect to features of the model $\mathcal{M}$. Therefore, the model that minimizes the aforementioned cost function would be equivalent, in the sense defined by J, to the system, that is, $\mathcal{M}\equiv_J\mathcal{S}$. The question is as follows: Does $\mathcal{M}\equiv_J\mathcal{S}$ guarantee $\mathcal{M}\equiv\mathcal{S}$? Another way of looking at the problem is to enquire the following: Which types of cost function J can be used in such a way that $\mathcal{M}\equiv_J\mathcal{S}$ implies $\mathcal{M}\equiv\mathcal{S}$ most of the time?

Of course, the cost function $J(\mathcal{S},\mathcal{M})$ is too general to be useful. To see this it suffices to consider that at this stage there is no representation of the system $\mathcal{S}$ and consequently such a cost function cannot be defined. Usually $\mathcal{S}$ is only known via the available data $\mathcal{Z}$ and by the auxiliary information $\mathcal{I}$, whenever the latter is available. Therefore, it would be natural to redefine the cost function as $J(\mathcal{Z},\mathcal{M})$, for black-box modeling, and $J(\mathcal{Z},\mathcal{I},\mathcal{M})$, for gray-box modeling. Each of these cases will be addressed in turn.

1. Black-box case

The last step taken in the previous paragraph (definition of the cost function) created an equivalence problem. Originally, it made sense to handle $\mathcal{S}$ and $\mathcal{M}$ in the same cost function because such entities are of the same type, that is, both are dynamical systems (although $\mathcal{S}$ is "abstract" and $\mathcal{M}$ is mathematical). Having seen that, it is reasonable to accept that replacing $\mathcal{S}$ with the data $\mathcal{Z}$ in the cost function will also require replacing $\mathcal{M}$ with some model data $Z_{\mathcal{M}}$.

In order to be more specific, assume that a subset $Z\in\mathbb{R}^{N\times r}$ of data is taken from $\mathcal{Z}$, $Z\subset\mathcal{Z}$. Z is assumed to be composed of at least one time series $y(k), k=1,\ldots,N$ called the output, and possibly other exogenous time series $u_1(k),\ldots,u_{r-1}(k), k=1,\ldots,N$ called inputs. If only one time series is available, it is interpreted as the output $y(k)$, and in such a case $Z=[y(1)\ldots y(N)]^T$.

Therefore, using Z and implementing the modification mentioned in the paragraph before the last, a practical cost function for black-box modeling would be $J(Z,Z_{\mathcal{M}})$, where

$$Z=\begin{bmatrix}\mathbf{z}_1^{\mathrm{T}}\\ \mathbf{z}_2^{\mathrm{T}}\\ \vdots\\ \mathbf{z}_N^{\mathrm{T}}\end{bmatrix}=[\mathbf{y}\ \mathbf{u}_1\ \ldots\ \mathbf{u}_{r-1}]=\begin{bmatrix}y(1) & u_1(1) & \ldots & u_{r-1}(1)\\ y(2) & u_1(2) & \ldots & u_{r-1}(2)\\ \vdots & \vdots & \ldots & \vdots\\ y(N) & u_1(N) & \ldots & u_{r-1}(N)\end{bmatrix}$$

$$Z_{\mathcal{M}}=\begin{bmatrix}\hat{\mathbf{z}}_1^{\mathrm{T}}\\ \hat{\mathbf{z}}_2^{\mathrm{T}}\\ \vdots\\ \hat{\mathbf{z}}_N^{\mathrm{T}}\end{bmatrix}=[\hat{\mathbf{y}}\ \mathbf{u}_1\ \ldots\ \mathbf{u}_{r-1}]=\begin{bmatrix}\hat{y}(1) & u_1(1) & \ldots & u_{r-1}(1)\\ \hat{y}(2) & u_1(2) & \ldots & u_{r-1}(2)\\ \vdots & \vdots & \ldots & \vdots\\ \hat{y}(N) & u_1(N) & \ldots & u_{r-1}(N)\end{bmatrix},$$

where $\hat{y}(k)=f(\mathbf{z}_{k-1})$ is the model one-step-ahead prediction. So, finally, many black-box model building techniques solve the following optimization problem

$$\min J(Z,Z_{\mathcal{M}})$$

$$\text{subject to } \boldsymbol{\theta}\in D, \tag{1}$$

where often $J(Z,Z_{\mathcal{M}})$ is chosen as the inner product $\langle\mathbf{y},\hat{\mathbf{y}}\rangle$, $\boldsymbol{\theta}$ is the parameter vector of $\mathcal{M}$, and D stands for the feasible set. When D is $\mathbb{R}^{\dim(\boldsymbol{\theta})}$, as it is usually the case in black-box modeling, the optimization problem (1) is said to be unconstrained. Hereafter in the unconstrained case the indication of the feasible set will be omitted.

2. Gray-box case

It should be recalled that, to attain Eq. (1), $\mathcal{Z}$ was replaced with Z and $\mathcal{M}$ with $Z_{\mathcal{M}}$ in the argument of $J(\mathcal{Z},\mathcal{M})$. This reveals that, in the gray-box model building procedure, $\mathcal{I}$ would also need to be replaced to render the problem manageable. In order to do this, let us assume that the particular type of information available, $I\subset\mathcal{I}$, can be mapped into the model. With some abuse of nomenclature this can be represented as $g^{-1}(I)=\mathcal{M}$ or equivalently as $g(\mathcal{M})=I$. By this it is meant that a certain type of information can be inserted into the model (this process is represented by g^{-1}) or can be extracted from the model (this would be represented by g). Therefore, $J(\mathcal{Z},\mathcal{I},\mathcal{M})$ becomes $J(Z,I,Z_{\mathcal{M}},g(\mathcal{M}))$ and the optimization problem to be solved is

$$\min J(Z,I,Z_{\mathcal{M}},g(\mathcal{M})). \tag{2}$$

A practical way of solving Eq. (2) is to decompose the cost function into two and solve the multiobjective problem

$$\min \lambda J(Z,Z_{\mathcal{M}})+(1-\lambda)J(I,g(\mathcal{M})),\lambda\in[0,\ 1]. \tag{3}$$

Another way of solving Eq. (2) is to use $g(\mathcal{M})=I$ to define constraints and to solve the constrained optimization problem

$$\min J(Z,Z_{\mathcal{M}})$$

$$\text{subject to } g(\mathcal{M})=I. \tag{4}$$

The interpretation of Eq. (4) is that not all "good" black-box models are acceptable, but now only a subset of models (composed by those that are consistent with the information I) is the feasible set. In other words, model $\mathcal{M}$ not only should minimize $J(Z,Z_{\mathcal{M}})$, but also should be consistent with

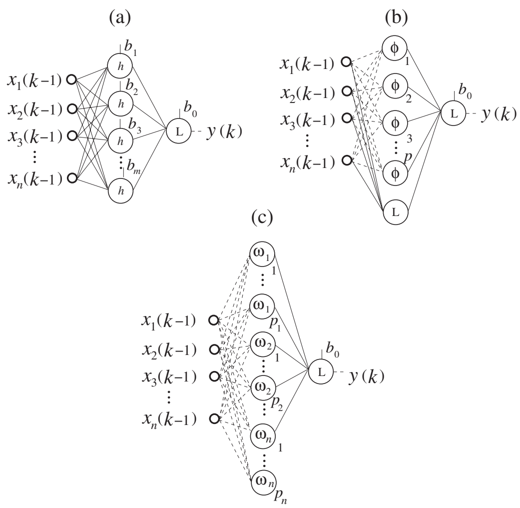

FIG. 1. Schematic representation of general model classes. (a) Typical perceptron model, (b) single-type basis function models, and (c) multitype basis function models. Solid lines correspond to parameters that must be estimated. Dashed lines indicate absence of parameters. Therefore (a) is nonlinear in the parameters while (b) and (c) are linear in the parameters. Usually ϕ depends on some parameters. When this is the case, such parameters are chosen beforehand.

the auxiliary information I, in the sense that $g(\mathcal{M})=I$ or, at least, $g(\mathcal{M})\approx I$.

More often that not, the numerical solutions of the optimization problems stated in Eqs. (3) and (4) are fairly standard. The great difficulty, however, is to find a certain type of auxiliary information I that relates to the model $\mathcal{M}$ by a well defined function g, such that $g(\mathcal{M})=I$. As it will be discussed below, the ease with which a certain type of information is mapped into a model will greatly depend on the type of information, on the model class, and on the particular model structure.

B. Model class

In this paper only single-output discrete-time model classes will be considered. Such model classes can be represented as $y(k)=f[\boldsymbol{x}(k-1)]$, where $y(k)$ is the scalar output at "time" k, f stands for the model, and $\boldsymbol{x}(k-1)$ is a vector of regressor (independent) variables taken at "time" $k-1$.

For the sake of discussion three broad classes of models will be mentioned. Such classes cover most of the examples of global discrete-time models to be found in the literature on nonlinear dynamics.

The *first* broad class is the perceptron model. Conventional simplifications in this class of models are (i) to take only one hidden layer of nodes, (ii) consider the output node linear, and (iii) consider all the activation functions h of the hidden layer nonlinear and the same. A perceptron model with such features is illustrated in Fig. 1(a) and can be described mathematically as

$$y(k)=b_o+\sum_{j=1}^{m} w_j^o h\left(b_j+\sum_{i=1}^{n} w_{ji}^h x_i(k-1)\right), \qquad (5)$$

where w_{ji}^h indicates a weight (to be estimated) of the hidden layer that connects the ith input to the jth neuron of the hidden layer. w_j^o is the weight (to be estimated) of the jth hidden neuron output, b's are constants, called bias parameters, and the neuron *activation function* is h. Finally, $n=\dim(\boldsymbol{x})$ and m is the number of neurons in the hidden layer. The function shown in the right-hand side of Eq. (5) is often called *feedforward* because there are no feedback loops *internal* to the network. It is important to notice that Eq. (5) is in the form $y(k)=f[\boldsymbol{x}(k-1)]$. Common choices for nonlinear activation functions are Gaussian, sigmoidal, and the hyperbolic tangent. The weights and the bias terms, on the other hand, are determined by optimization algorithms that search to minimize a cost function which usually depends on the difference between the given data and the network output. In a recent study, different approaches to network training were compared [13]. One of the first papers to build this kind of model for chaotic data seems to have been [14]. More recent and impressive results have been discussed in [15].

The *second* broad class is shown in Fig. 1(b). One important difference with respect to the first class is that now there are no weights associated to the connections between the inputs and the nodes in the hidden layer. The function ϕ is nonlinear and often depends on certain tuning parameters which are usually known when the weights associated to the connections indicated by solid lines [Fig. 1(b)] are to be estimated. As a consequence, the model is linear in the parameters. A common example of such a model class is obtained by choosing ϕ to be some radial function [7]

$$y(k)=b_0+\sum_{i}^{p} w_i\phi(\boldsymbol{x}(k-1))+\sum_{i=1}^{n} a_i x_i(k-i), \qquad (6)$$

where p is the number of radial basis functions. b_0, w_i, and a_i are the unknown parameters to be estimated. One of the first works to use this model class was [3], and in the hidden layer only nonlinear nodes were used. Work using this type of model has been recently surveyed in [16].

The *third* model class is illustrated in Fig. 1(c). The main difference with respect to the second class is that the basis functions are usually different, that is, $\omega_i\neq\omega_j$. Another important difference, not revealed in the figure, is that, whereas it is usually assumed in the second class that the input vector is uniform, in the third class such uniformity is not required (see Sec. II D). However, the main difference is that various basis functions are often used in Fig. 1(c) in such a way as to enable matching different data features. One possible choice of basis functions ω_i are monomials of different degrees [17]. This choice of basis functions constrains the resulting models to those cases in which the dynamics underlying the data can be approximated by a linear combination of nonlinear monomials. For systems that are more strongly nonlinear, other basis functions should be preferred. On the other hand, the choice of monomial basis functions enables building models which are more information-dependent than models for which all the basis functions are of the same type. This remark will be exploited further in the second part of the paper.

A class of models that lies somewhere in between the second and third classes is the wavelet network (wavenet). In such a model class the linear node in the hidden layer [see Fig. 1(b)] is not used, consequently the last summation in Eq. (6) is absent. The basis functions ϕ can be chosen as wavelet functions that depend on dilation and translation parameters which are taken *a priori* over a grid of values. In this respect the basis functions are of the same type, but have different scales. This will ensure that different time scales in the data are matched by some basis function. In [6] the "Mexican hat" wavelet function was used, whereas tensor product wavelet functions were used in [18], and *B*-splines in [19]. In the context of process control, relevant work has been accomplished choosing the basis functions to be Laguerre expansions of Volterra models [20].

C. Model structure selection

Having chosen a model class, the model structure selection problem is basically to determine which model topology is more adequate, given Z (black-box problem) or given Z and I (gray-box problem). One of the most fundamental aims in model structure selection is to find the *simplest model* that is consistent with Z (and eventually with I). Much work has been performed in this field and it would be impossible to survey even the main approaches in nonlinear dynamics [11,21–28]. In this section, however, a couple of relevant issues concerning structure selection will be pointed out.

First, it is important to notice that for the model classes illustrated in Figs. 1(a) and 1(b)—for which the activation and basis functions h and ϕ are usually chosen *a priori*—the model structure is basically determined simply by the *size* of the model. The inclusion of one more basis function or one more node in the hidden layer might improve (or not) performance but no new feature is gained with the new inclusion. In the context of black-box problems this is most welcome, because the model class is sufficiently general to cope, in principle, with most situations and the only concern is with the model size. One way of performing this task is using information criteria [22,27,29], although it has been shown that such criteria are unable to guarantee good overall *dynamical* performance [21,30,31].

On the other hand, models from the third class, see Fig. 1(c), require not only the decision of how large the model should be but especially which basis functions ω_i should compose the model. For models of this class, the simple use of information criteria is ineffective unless all possible combinations of basis functions are tried. So, first of all, it has to be decided *which sets* of basis functions are more adequate and then try to decide on *how many* basis functions should be used. This double concern, in fact, underlies the procedure developed in [32]. Such a procedure makes no assumptions on the type of basis functions, so much so that it was successfully applied to choose the centers in radial basis function models [33] and wavelet models [34,35]. However, noticing that the basis functions ω_i are found in groups (called term clusters Ω_i), a complementary procedure was proposed to aid in the selection of which class of basis functions are more adequate [36]. The formalism of term clusters has been found useful in diverse applications of model structure selection problems [26,37,38]. Detailed simulation studies have shown that the correct choice of *which* term clusters Ω_i should be considered to compose the model is generally more critical than *how many* functions are used [21,24]. Apart from this, it has been pointed out recently that the original algorithm put forward in [32] does not necessarily choose the best basis functions when the input is not sufficiently exciting [26]. A model-free procedure to choose the maximum lags in the embedding has been proposed in [39] and a tree-based search applied to the selection of model structure has been recently investigated in [40]. Several model classes have been compared in the modeling of presliding friction dynamics in [41,42].

In second place, it should be highlighted that in the context of black-box modeling, the double aspect of structure selection for the class of models illustrated in Fig. 1(c) is somewhat less convenient than that for the model classes shown in Fig. 1(a) and Fig. 1(b). On the other hand, because the sets of basis functions Ω_i code different type of information [43] such a feature could be used to advantage in the context of gray-box modeling *if the available information* could be mapped into the model, that is, if a function g can be defined such that $g^{-1}(I)=\mathcal{M}$ and $g(\mathcal{M})=I$ (see Sec. II A 2). Of course, the last assertion is valid for all the model classes illustrated in Fig. 1. However, because the model classes shown in Fig. 1(a) and Fig. 1(b) are more general, it is typically more difficult to establish g. Examples of g for different types of information and of model class will be provided in Sec. II F.

Before closing this section, it is argued that in choosing the model class, because of the parsimony principle, the most consistent approach would be to start off with the simplest model class (e.g., linear models) and then, having verified that the resulting model is inadequate, move on to a more general model representation.

D. Uniform and nonuniform embeddings

A key issue in modeling nonlinear dynamics is that of selecting an appropriate embedding space. In principle, this would include two stages: The choice of observables [44,45] and the choice of embedding parameters [46]. In many practical situations, although the observable is determined before data acquisition, the embedding parameters—basically the embedding dimension, d, and the time delay, τ—can be determined by the user *a posteriori*.

Using the nomenclature of Fig. 1, the input vector, at the left-hand side of the models, determines the embedding space. In particular, for time delay coordinates of autonomous systems we could have

$$\boldsymbol{x}(k-1)=[y(k-\tau)y(k-2\tau)\dots y(k-d\tau)]^T. \qquad (7)$$

It has been duly pointed out that the problem of choosing an embedding in the context of model building is a *bona fide* stage of the modeling procedure [47]. Also, the uniform embedding defined by taking the elements of $\boldsymbol{x}$ in Eq. (7) to be the coordinates is not necessarily optimal. Therefore, which elements (coordinates) should be chosen to compose $\boldsymbol{x}$ is also part of the modeling problem. An optimal solution to such a problem might require an embedding space in which the "temporal distances" from one coordinate to another are not necessarily the same. The authors of [47] call embedding spaces with this character irregular.

Despite this, it is commonplace to build discrete-time models using regular embeddings. To enable irregular embeddings (see Fig. 1) it suffices not to connect certain input nodes to the middle layer. Here it is pointed out that the choice of particular basis functions ω_i is in some cases equivalent to the choice of embedding coordinates which could turn out to be irregular. In fact, Eqs. (13) and (21) in [48] and Eq. (37) in [17] are some examples of models (automatically) built on irregular embeddings. In conclusion, it becomes clear that the modeling procedure followed in [17], for instance, includes the choice of embedding coordinates as a part of the modeling procedure, as pointed out in [47].

E. Model validation

In the context of *time series modeling* (TSM), it is usually assumed that there is a separate set of data Z_v, similar to Z, available for model selection.

For many models, the parameters are estimated by solving the problem outlined in Eq. (1) for $Z_{\mathcal{M}}$ taken as the one-step-ahead predictions. A key-point to realize is that the dynamical features of the model $\mathcal{M}$ are difficult to assess by analyzing $Z_{\mathcal{M}}$ [26,71,72] and a consequence of this is that solving Eq. (1) cannot possibly guarantee (although it is hoped that it will come close) that $\mathcal{M} \equiv \mathcal{S}$. As pointed out in Sec. II A, the bottom line is that even if the model approximates the system in terms of a particular choice of J, that is $\mathcal{M} \equiv_J \mathcal{S}$, that does *not* necessarily imply $\mathcal{M} \equiv \mathcal{S}$. In particular, that last remark is often true when J is the sum of squared residuals.

At this stage it will be useful to notice that, in theory, what can be guaranteed during model building is that a set of data produced by the model is consistent—in terms of J—to data measured from the system, that is, $Z_{\mathcal{M}} \equiv_J Z$. However, in practice, only a few (sometimes just one) realizations of the system are available for model validation (those data are indicated by Z_v) and all that can be verified is if $Z_{\mathcal{M}} \equiv_J Z_v$. Should either Z, Z_v or $Z_{\mathcal{M}}$ not clearly represent the dynamics of $\mathcal{S}$ or $\mathcal{M}$ adequately, very little can be said about the model quality. This nontrivial problem is at the center of model validation.

With the above discussion in mind, it can be appreciated that in order to increase robustness, from a dynamical point of view, it would be convenient that the optimization step, in the cost function $J(Z, Z_{\mathcal{M}})$ used a set of model data $Z_{\mathcal{M}}$ that should be *dynamically* more representative of the model than the one-step-ahead predictions. The ideal situation would be to build $Z_{\mathcal{M}}$ with free-run simulation data. This, unfortunately, would turn out to be computationally very demanding and probably would *not* apply to systems with positive Lyapunov exponents. Free-run predictions have been recently used to great advantage in the problem of structure selection [26] and network training [13]. Also, the multiple-shooting parameter estimator circumvents some of the problems that arise when $Z_{\mathcal{M}}$ is built with one-step-ahead predictions [49–51]. In a recent work the concept of synchronization of discrete models [52] was used in the problem of model selection [53].

F. Use of prior information

In general, the problem of building models from data *and* additional information (sometimes referred to as *a priori* or *auxiliary*) has been postponed. In the context of linear models or nonlinear process models, some results are available [54–58]. The general setting is to have the dynamical data $\mathcal{Z}$ *plus* some other source of information $\mathcal{I}$ and to use both in building a model, as discussed in Sec. II A 2.

As pointed out in [12] there seems to be less applications of gray-box modeling in the realm of nonlinear dynamics. One of the aims of this paper (see next section) is to put forward a procedure by which it is possible to impose a certain type of information I on a class of nonlinear discrete models built on a Poincaré section. In fact, as it will become clear later, the mention of a Poincaré section is only to emphasize that in the present paper the concern is with period-doubling of periodic-points in maps rather than the period-doubling of trajectories of flows. Before presenting the new procedure, it will be convenient to clarify under which circumstances gray-box modeling could be profitable and to mention some previous results in the field.

If all that is desired of the system is already available in the particular set of data used for model building, Z, then granted that the model class is sufficiently general and that the model structure is adequate, the resulting model should be a sufficiently accurate representation of the system. In this case there is no motivation to use an additional piece of information I. However, often in practice, for a number of different reasons—such as presence of noise, poor choice of observable, poor frequency content in the data, limited amplitude excursion, and the like—the available data Z either does not have all the desired information about the system or such information is difficult to obtain. In such cases it is conceivable that additional information, I, be available and it is natural to enquire if it is possible to use I to build the model. Thus, more often than not, if Z is of "good" quality and "sufficiently" complete, black-box modeling should be the practitioner's first choice.

In the realm of nonlinear dynamics, procedures have been put forward for building models using auxiliary information. The number of fixed points were used in [59]; the location of fixed points were used in [60] to solve problem (3). Information about the symmetry was recently used to constrain not only the topology but also the parameter estimates of network and radial-basis function models [61]. Lastly, topological features such as folding and tearing mechanisms [62,63] in addition to the location and local eigenstructure of fixed points has been used in [64].

In what follows a different type of information will be used in synthesizing nonlinear models from data. The resulting models must, by construction, be consistent with the additional information used. An important point to be noticed is that the type of information to be used in the next section can be comfortably handled by models of the type illustrated in Fig. 1(c) with monomial bases. Other model classes, while being better suited to approximate stronger nonlinearities, are less apt to have *hard* dynamical constraints imposed during model building. This tradeoff between approximation capability and handling of *a priori* information is one of the important points illustrated in the present work. Moreover, the previous discussion also suggests that the type of *a priori* information which is available and is desired to be used in the modeling procedure can provide a concrete aid in choosing which model class to select.

III. IMPOSING A PERIOD-DOUBLING BIFURCATION TO A DATA-DRIVEN MODEL

A. Statement of the problem

It is assumed that a set of data Z is available from a system $\mathcal{S}$. It is also assumed that the system undergoes a period-doubling bifurcation for a given parameter value μ_c. If the system is a flow for which a periodic orbit period-doubles, then it is further assumed that the data Z are taken at a Poincaré section and in this case the searched model will be of the Poincaré map and *not* of the original flow. It is desired to build a model $\mathcal{M}(\mu)$ from the data Z such that for $\mu=\mu_c$ a specified periodic point of the model undergoes a period-doubling (flip) bifurcation.

In this case I is the information about the desired bifurcation. In order to solve the gray-box problem (see Sec. II A 2) it is necessary to be able to map such information into the model, that is, it is necessary to find a function g such that $g^{-1}(I)=\mathcal{M}$. Here is where choosing the model class plays a role. If a certain type of information is available and should be used, only model classes for which g can be found are useful in this particular type of problems. In what follows it will be shown how to relate the information I with the model $\mathcal{M}$.

As stated, the present problem is not directly applicable because although it is realistic to get Z from a system that is known to undergo a period-doubling bifurcation, and even if a real parameter μ_S is measured, the parameter value that must be known is μ, which will depend on how the model $\mathcal{M}(\mu)$ is parametrized.

A more restrictive situation, but also more realistic would be to use a reference set of data Z_0 to build a nominal model $\mathcal{M}_0(\mu)$ in a black-box fashion (see Sec. II A 1). Assuming that the structure of $\mathcal{M}_0(\mu)$ is adequate, find the critical value μ_c for which a periodic point $\bar{y}$ of $\mathcal{M}_0(\mu)$ undergoes a period-doubling bifurcation. In this setting, the fact that the system undergoes a flip at $\mu=\mu_c$ is the additional information to be used, I.

As discussed in Sec. II F, the main motivation for using auxiliary information is to compensate for the loss or for the blurring of information in Z. Therefore, it is assumed that new data are gathered from the system, say Z_1, but such data are noisy. It is known that the effect of noise in dynamical data is to shift the bifurcations in the models built from such data (see [65], for instance). The problem then becomes to build a model $\mathcal{M}_1(\mu)$ from the noisy data set Z_1 using the fact that the system undergoes a flip, which for the nominal model happens at $\mu=\mu_c$. That additional information will serve as an anchor to grant robustness to the bifurcation structure of $\mathcal{M}_1(\mu)$. In practice Z_0 could be data collected under controlled circumstances and Z_1 collected in the field and therefore of lower quality.

B. Procedure

For the sake of presentation, the procedure will be described based on a quadratic map that undergoes a period-doubling bifurcation

$$y_k = 1 + \alpha y_{k-1} + \mu y_{k-1}^2, \tag{8}$$

where α and μ are parameters and the latter is taken as the bifurcation parameter. A similar procedure can be developed for normal forms.

Steady-state analysis of Eq. (8) results in $\bar{y}=1+\alpha\bar{y}+\mu\bar{y}^2$. One of the key points in the procedure that will follow is to realize that any discrete-time polynomial map composed with (any number of) terms taken from the term clusters [73] Ω_0 (constant), Ω_y and Ω_{y^2} will have a similar steady-state equation [59]. This last remark is very important because in many situations the system to be modeled will not be one-dimensional and consequently the map (8) will be insufficient. Nonetheless *the procedure to be developed will still apply*.

In order to simplify presentation, let us consider a three-dimensional map with all possible combinations of terms taken from Ω_0, Ω_y, and Ω_{y^2}:

$$\begin{aligned} y_k = f(\mathbf{y}_{k-1}) &= \theta_0 + \theta_1 y_{k-1} + \theta_2 y_{k-2} + \theta_3 y_{k-3} \\ &\quad + \theta_{11} y_{k-1}^2 + \theta_{12} y_{k-1} y_{k-2} + \theta_{13} y_{k-1} y_{k-3} \\ &\quad + \theta_{22} y_{k-2}^2 + \theta_{23} y_{k-2} y_{k-3} + \theta_{33} y_{k-3}^2 . \end{aligned} \tag{9}$$

Recalling that the state vector of this model is composed by $\mathbf{y}_{k-1}=[y_{k-3}y_{k-2}y_{k-1}]^T$, then $\mathbf{y}_k=F(\mathbf{y}_{k-1})$. The Jacobian matrix of Eq. (9) is given by

$$\mathrm{D}F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \dfrac{\partial f}{\partial y_{k-3}} & \dfrac{\partial f}{\partial y_{k-2}} & \dfrac{\partial f}{\partial y_{k-1}} \end{bmatrix}, \tag{10}$$

where

$$\frac{\partial f}{\partial y_{k-3}} = \theta_3 + \theta_{13} y_{k-1} + \theta_{23} y_{k-2} + 2\theta_{33} y_{k-3},$$

$$\frac{\partial f}{\partial y_{k-2}} = \theta_2 + \theta_{12} y_{k-1} + 2\theta_{22} y_{k-2} + \theta_{23} y_{k-3},$$

$$\frac{\partial f}{\partial y_{k-1}} = \theta_1 + 2\theta_{11} y_{k-1} + \theta_{12} y_{k-2} + \theta_{13} y_{k-3}.$$

The eigenvalues of Eq. (10) are given by the solutions of

$$\lambda^3 - \frac{\partial f}{\partial y_{k-1}}\lambda^2 - \frac{\partial f}{\partial y_{k-2}}\lambda - \frac{\partial f}{\partial y_{k-3}} = 0. \quad (11)$$

Because the main concern is to reproduce a flip bifurcation, Eq. (11) will be evaluated at the periodic-point (fixed-point) $\bar{y}$. This substitution yields

$$\lambda^3 - [\theta_1 + (2\theta_{11} + \theta_{12} + \theta_{13})\bar{y}]\lambda^2 - [\theta_2 + (\theta_{12} + 2\theta_{22} + \theta_{23})\bar{y}]\lambda - [\theta_3 + (\theta_{13} + \theta_{23} + 2\theta_{33})\bar{y}] = 0. \quad (12)$$

Steady-state analysis of Eq. (9) starts by substituting $\bar{y} = y_{k-1} = y_{k-2} = y_{k-3}$. This yields

$$\bar{y} = \Sigma_0 + \Sigma_y\bar{y} + \Sigma_{y^2}\bar{y}^2, \quad (13)$$

where the values of the cluster coefficients Σ_i are evident [59]. The similarity of Eqs. (13) and (8) is obvious and will be used in the following. The solutions of Eq. (13) yield the periodic points of the model, $\bar{y} = \bar{y}_i, i = 1, 2$. It should be pointed out that the steady-state analysis of any polynomial model of degree two, of any order [74], will yield Eq. (13), but of course for different values of the cluster coefficients. Closed solutions for models with three periodic-points (degree three) are given in [59].

The solution of Eq. (12) yields the eigenvalues of the Jacobian matrix DF at the periodic point, say $\bar{y} = \bar{y}_1$. If the model is to undergo a flip bifurcation then one of the eigenvalues must be $\lambda = -1$. Hence, taking $\bar{y} = \bar{y}_1$ and $\lambda = -1$ in Eq. (12) yields the following constraint:

$$\lambda^3 - [\theta_1 + (2\theta_{11} + \theta_{12} + \theta_{13})\bar{y}_1]\lambda^2 - [\theta_2 + (\theta_{12} + 2\theta_{22} + \theta_{23})\bar{y}_1]\lambda,$$

$$[\theta_3 + (\theta_{13} + \theta_{23} + 2\theta_{33})\bar{y}_1] = 0,$$

$$-\theta_1 + \theta_2 - \theta_3 - 2\bar{y}_1[\theta_{11} + \theta_{13} - \theta_{22} + \theta_{33}] = 1. \quad (14)$$

At this stage the problem can be stated as follows. Given the dynamical data Z, a polynomial model of *any order* is sought that will fit the data. For the sake of presentation only quadratic models will be considered, but the extension to models of higher degree is possible. The model so obtained, whichever the particular structure, will be characterized in steady state by Eq. (13). The parameters of the model must be estimated in such a way as to simultaneously satisfy the following.

(1) Yield a good dynamical fit to the dynamical data Z.

(2) Guarantee that $\bar{y}_1$ be a periodic point of the model.

(3) Guarantee that for $\Sigma_{y^2} = \mu_c$, $\bar{y}_1$ undergoes a flip, that is, one eigenvalue of the Jacobian matrix will be $\lambda = -1$ when evaluated at $\bar{y} = \bar{y}_1$.

The key issue now is to be able to write such requirements in terms of the model parameters, since $g(\mathcal{M}) = I$. In fact, that has already been done in Eqs. (13) and (14). Such constraints can be written in matrix form as

$$\mathbf{c} = S\boldsymbol{\theta}, \quad (15)$$

where

$$\mathbf{c}^T = [\bar{y}_1 \ \mu_c \ 1],$$

$$S = \begin{bmatrix} 1 & \bar{y}_1 & \bar{y}_1 & \bar{y}_1 & \bar{y}_1^2 & \bar{y}_1^2 & \bar{y}_1^2 & \bar{y}_1^2 & \bar{y}_1^2 & \bar{y}_1^2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & -1 & 1 & -1 & -2\bar{y}_1 & 0 & -2\bar{y}_1 & 2\bar{y}_1 & 0 & -2\bar{y}_1 \end{bmatrix},$$

$$\boldsymbol{\theta}^T = [\theta_0 \ \theta_1 \ \theta_2 \ \theta_3 \ \theta_{11} \ \theta_{12} \ \theta_{13} \ \theta_{22} \ \theta_{23} \ \theta_{33}].$$

The first constraint in Eq. (15) is actually Eq. (13); the second constraint imposes that the cluster coefficient Σ_{y^2}, which was chosen the bifurcation parameter, should have the specified value μ_c at the bifurcation point. Finally, the last constraint, which is Eq. (14), guarantees that whenever $\Sigma_{y^2} = \mu_c$, one eigenvalue of the Jacobian matrix will be $\lambda = -1$. All the remaining degrees of freedom will be used to fit the dynamical data Z. Clearly, the set of constraints (15)—which code the auxiliary information I—are in the form of Eq. (4), that is, Eq. (15) maps the information concerning the required bifurcation into the model by means of the model structure and parameters.

Finally, given the dynamical data, Z, the one-step-ahead prediction of the model (9), $Z_{\mathcal{M}}$, and the set of constraints (15), $g(\mathcal{M}) = I$, there is a well known solution to the problem (4), namely [66]

$$\hat{\boldsymbol{\theta}}_{\mathrm{CLS}} = (\Psi^T\Psi)^{-1}\Psi^T\mathbf{y} - (\Psi^T\Psi)^{-1}S^T(S(\Psi^T\Psi)^{-1}S^T)^{-1} \times (S\hat{\boldsymbol{\theta}}_{\mathrm{LS}} - \mathbf{c}), \quad (16)$$

where Ψ is the regressor matrix, $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{CLS}}$ are the least-squares unconstrained and constrained solutions, respectively, and S and $\mathbf{c}$ are defined in Eq. (15). The extension of this procedure to other polynomial models is possible.

IV. NUMERICAL RESULTS

The main points of the procedure presented in the previous section will be illustrated in this section by means of two simulated examples. In the first example, the aim is to illustrate the validity of the procedure described in Sec. III. In the second example, greater emphasis will be given to practical issues.

A. Hénon map

Data Z (N=500) was produced simulating the two-dimensional map [67]

$$y_k = 1 + 0.3y_{k-2} + \mu y_{k-1}^2, \quad (17)$$

with $\mu = -1.4$. The bifurcation diagram of Eq. (17) is shown in Fig. 2. The reconstruction of the bifurcation diagram of this map from different model structures was discussed in [59].

Eighty percent (80%) zero-mean Gaussian noise was added to the data. From such data the following model was obtained using the well-known least squares algorithm:

$$y_k = 0.5569 - 0.1077y_{k-1} + 0.1114y_{k-2} - 0.1490y_{k-3} - 0.3284y_{k-1}^2. \quad (18)$$

Model (18) was obtained in a purely black-box fashion and

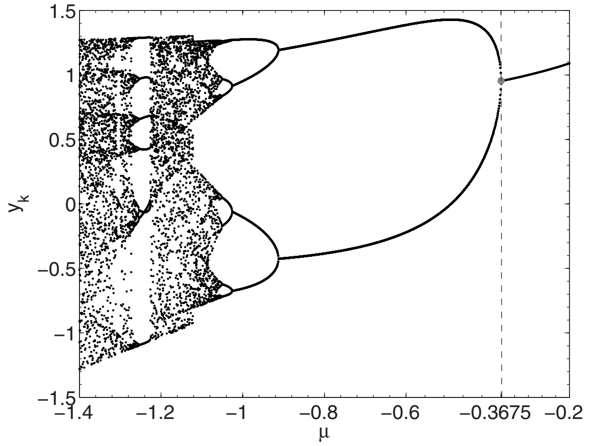

FIG. 2. Bifurcation diagram of the Hénon map (17).

its bifurcation diagram is shown in Fig. 3. From that figure it is seen that the first flip bifurcation has been shifted by the noise in the data. As a matter of fact, that shift increases with the intensity of the noise, as illustrated in Fig. 4.

It is now desired to use the procedure put forward in Sec. III to obtain another model that undergoes a flip at the original value of the bifurcation parameter $\mu \approx -0.37$. In this case the modeling step will use the *same* dynamical data (corrupted with 80% noise) plus the prior information concerning the flip bifurcation at $\mu = -0.3675$. The resulting model will be called a gray-box model.

Taking the data range into account, it becomes quite obvious that the periodic point of model (18) $\bar{y}_1 \approx 0.4$ (and not $\bar{y}_2 \approx -3.5$) should be used to build the constraints. Practical aspects of the estimation of fixed points from data have been considered in [68]. Having chosen which periodic point to use, a second step is to decide its location. This point needs further investigation as to its influence on the final model. For the sake of simplicity, in this example, it was decided to use the value that $\bar{y}_1$ would have if the parameter of the quadratic term (which is the bifurcation parameter) was $\theta_{11} = -0.3675$ (which is the critical value to have the period-

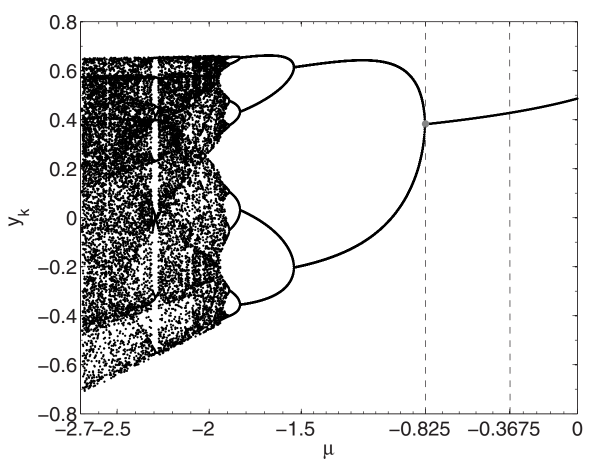

FIG. 3. Bifurcation diagram of black-box model (18). Notice how the first bifurcation point was shifted to $\mu \approx -0.825$, due to the noise in the data.

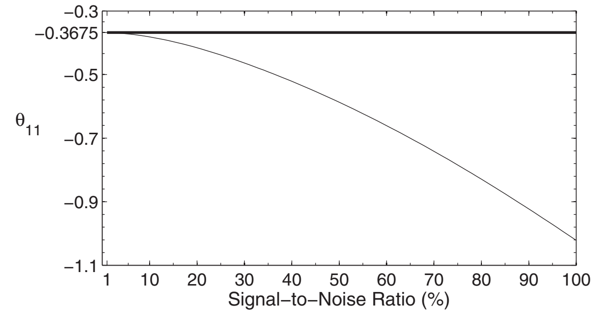

FIG. 4. Critical value of the bifurcation parameter $\mu = \theta_{11}$ (at which occurs the first flip bifurcation) as a function of the noise intensity for black-box models similar to Eq. (18). The line in boldface corresponds to the same, however for gray-box models, for which the first flip is imposed at a user-defined value. In this example, the value was $\mu = -0.3675$.

doubling). Therefore, for $\theta_{11} = -0.3675$, the periodic points of model (18) are $\bar{y}_1 = 0.4276$ and $\bar{y}_2 = -3.5439$.

As seen in Sec. III, the following restriction must hold

$$\begin{aligned} \bar{y}_1 &= \Sigma_0 + \Sigma_y \bar{y}_1 + \Sigma_{y^2} \bar{y}_1^2 \\ &= \theta_0 + [\theta_1 + \theta_2 + \theta_3]\bar{y}_1 + \theta_{11}\bar{y}_1^2, \end{aligned} \tag{19}$$

in order for the (new) model to have a periodic point at $\bar{y}_1$. The coefficient of the quadratic term is chosen as the bifurcation parameter that is $\mu = \theta_{11}$ to facilitate comparison with Eq. (17), but this is not a restriction to the method. Moreover, the auxiliary information to be used in this example is that the system undergoes a flip bifurcation at $\mu_c = -0.3675$. So, finally, the set of restrictions (15) can be written for this example as

$$\begin{bmatrix} 0.4276 \\ -0.3675 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & \bar{y}_1 & \bar{y}_1 & \bar{y}_1 & \bar{y}_1^2 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & -1 & -2\bar{y}_1 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_{11} \end{bmatrix},$$

$$\mathbf{c} = S\boldsymbol{\theta}, \tag{20}$$

where $\bar{y}_1 = 0.4276$ and $\mu = -0.3675$ were used. It should be clear that other values could be used here. For instance, by inspection of Fig. 2, a value closer to unity could be tried. In this example the *a priori* knowledge about the system is that the periodic point at $\bar{y}_1 = 0.4276$ undergoes a period-doubling bifurcation when $\mu \approx -0.36$, that is, $I = [0.4276 \ -0.3675 \ 1]^T$.

Therefore, using Eq. (16)—with $\mathbf{c}$ and S as defined in Eq. (20), and $\boldsymbol{\theta}$ being the parameters estimated by the LS, see model (18)—to reestimate the model parameters yields

$$\begin{aligned} y_k = 0.6062 - 0.2136 y_{k-1} + 0.2126 y_{k-2} \\ - 0.2595 y_{k-3} - 0.3675 y_{k-1}^2, \end{aligned} \tag{21}$$

which has periodic points at $\bar{y}_1 = 0.4276$ (imposed) and $\bar{y}_2 = -3.8578$ (free). The three eigenvalues of the Jacobian matrix evaluated at $\bar{y}_1$ are $\lambda_1 = -1$ (imposed), $\lambda_{2,3} = 0.2361 \pm 0.4515$ (free). Clearly, the fixed point $\bar{y}_1$ undergoes

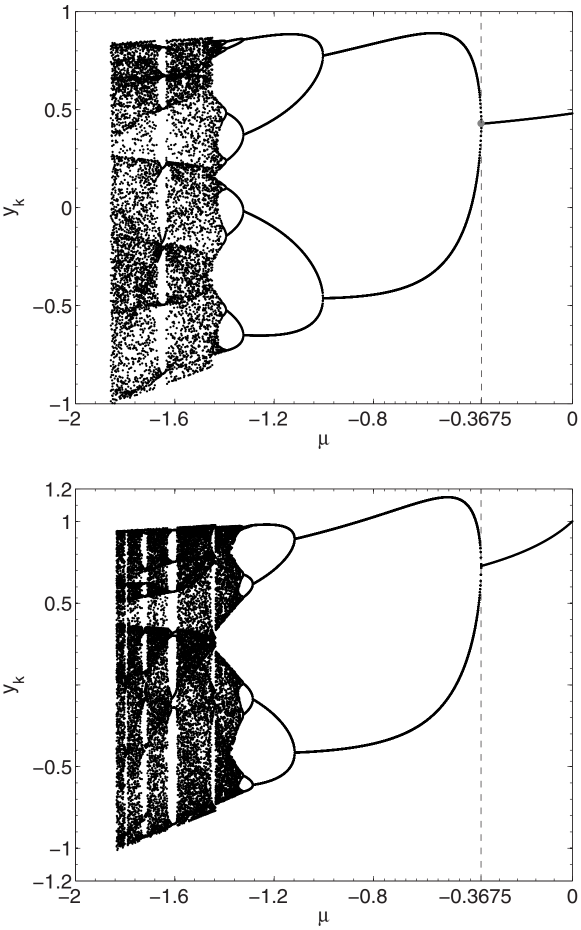

FIG. 5. Bifurcation diagrams of gray-box models. (a) Diagram of Eq. (21) with bifurcating periodic point imposed at $\bar{y}_1=0.4276$. (b) Diagram of another gray-box model obtained using $\bar{y}_1=0.7276$ instead. Notice how in both cases the first bifurcation point was successfully forced to $\mu \approx -0.3675$. Notice also how the location of the imposed periodic point is able to alter the periodicity pattern of the periodic windows.

a flip when the parameter of the quadratic term (which corresponds to μ) equals -0.3675, as desired. The bifurcation diagram of model (21) is shown in Fig. 5(a). As can be seen, the first bifurcation point was successfully forced to $\mu \approx -0.3675$, although the rest of the bifurcation diagram was shifted.

If the bifurcating periodic point is taken to be $\bar{y}_1 = 0.7276$ in Eq. (20), instead of $\bar{y}_1=0.4276$ the resulting model displays the bifurcation diagram shown in Fig. 5(b). As can be seen, the location of the bifurcating periodic point can be interpreted as an effective bifurcation parameter by which the pattern of periodic windows is changed. This dependence of the bifurcation pattern of a given model on a "generalized" parameter is an interesting topic that has been investigated recently [69].

Proceeding in this way, the location of the first flip bifurcation can be made insensitive to the noise intensity, as illustrated by the line in boldface in Fig. 4.

In order to further compare the dynamical performance of models (18) and (21), the procedure described in [53] was implemented and the results shown in Fig. 6. The interpretation of this figure is as follows. For any given value of the cost of synchronization ($J_{\rm rms}$), the gray-box model synchronizes better (smaller ϵ) than the black-box model. That is an indication that the gray-box model is closer to the underlying dynamics than the black-box model, at least for the considered bifurcation parameter value.

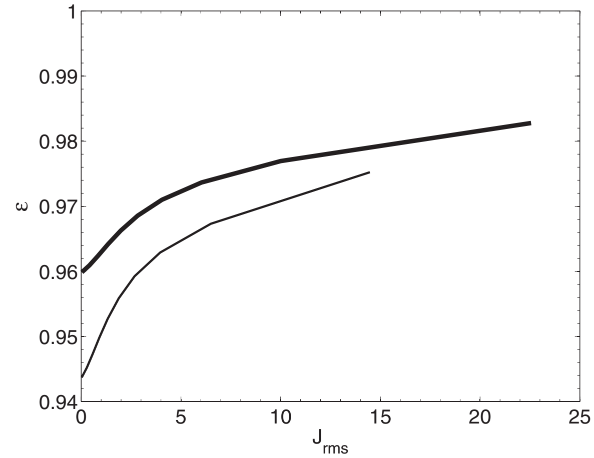

FIG. 6. Synchronization class vs cost of synchronization for identified models. Thick line corresponds to the black-box model (18) for $\mu=-2.68$ and the thin line to the gray-box model (21) for $\mu=-1.84$. The driving signal was produced with the original model (17) with $\mu=-1.4$.

In the present example, the system was a model with a mathematical structure (polynomial) very similar to that of the models. As a consequence, the value μ_c of the system was imposed directly on the model. In the next example a different system will be considered for which this will not be the case.

B. Sine map

This example will investigate the sine map

$$y_k = \mu \sin(\pi y_{k-1}), \tag{22}$$

which displays chaotic dynamics for $\mu=0.98$. The bifurcation diagram of map (22) is shown in Fig. 7. Throughout this example, the nomenclature established in Sec. III A will be used to facilitate comprehension.

It is assumed that the present modeling problem starts with a low-noise time series, Z_0, of the original system. In this example, such data were produced by iterating map (22) with $\mu=0.98$. This time series can be thought of as being obtained from a real system under controlled conditions, thus producing low-noise data. From 500 points of such high-quality data, the following model, $\mathcal{M}_0$, was estimated using least squares

$$\begin{aligned} y_k = & -0.0672 + 4.0810 y_{k-1} + 0.0065 y_{k-2} \\ & - 4.0528 y_{k-1}^2 - 0.0146 y_{k-1} y_{k-2}. \end{aligned} \tag{23}$$

Taking the parameter of the term $y_{k-1}y_{k-2}$ to be the bifurcation parameter μ, and varying μ within the range $-1 \leqslant \mu \leqslant 0$, results in the bifurcation diagram shown in Fig. 8. In

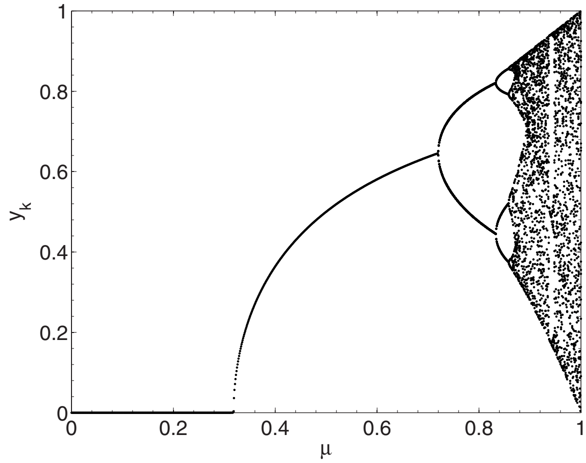

FIG. 7. Bifurcation diagram of the sine map (22).

this problem, the prior information comes from $\mathcal{M}_0$, or alternatively, from its bifurcation diagram shown in Fig. 8.

For $\mu=-0.71$ (first flip bifurcation) the model (23) has the following periodic points $\bar{y}_1=0.6257$ and $\bar{y}_2=0.0226$. No doubts, it is $\bar{y}_1$ that undergoes the first period-doubling bifurcation of the cascade.

It is interesting to notice that because Eq. (23) does not have the same mathematical structure as the original map (22), this example furnishes a somewhat more realistic scenario than in the previous example. Therefore we assume that *high quality* data Z_0 were obtained from a system and a model was built from such data. In this example, such a model is Eq. (23). From this model three important pieces of information will be used, namely (i) the structure, (ii) the critical bifurcation parameter value at the flip bifurcation point $\mu_c \approx -0.71$, and the location of the bifurcating periodic point at $\mu_c \approx -0.71$, which is $\bar{y}_1=0.6257$. In fact, such information is coded in the vector $I=[0.6257\ -0.71\ 1]^T$.

Now, consider that a new model $\mathcal{M}_1$ must be built from *new data* Z_1 that, for some reason, do not bring forth clearly

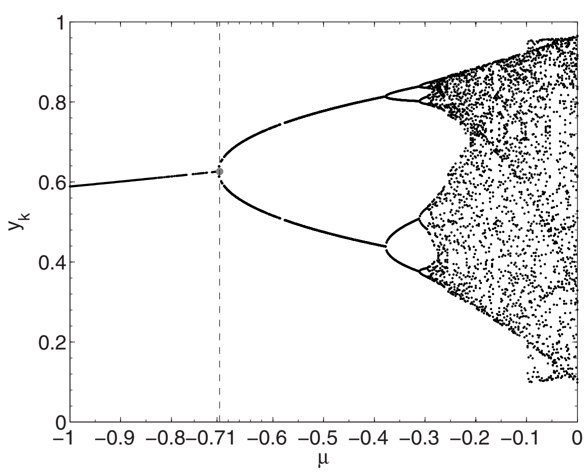

FIG. 8. Bifurcation diagram of the model (23) using the coefficient of the term $y_{k-1}y_{k-2}$ as the bifurcation parameter μ. Clearly, for $\mu=-0.0146$ the dynamical regime is chaotic and the first period-doubling occurs at $\mu_c \approx -0.71$.

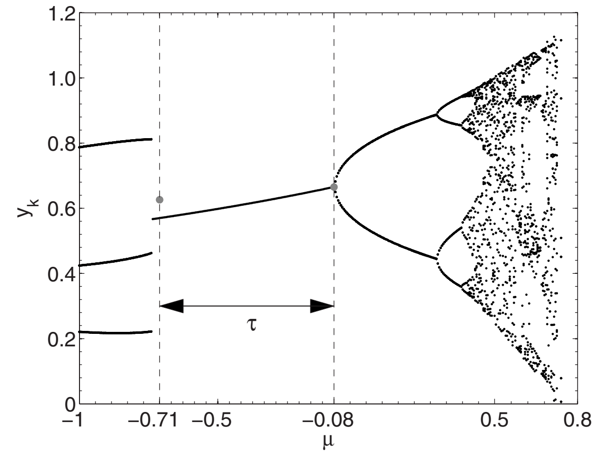

FIG. 9. Bifurcation diagram of the model (24) using the coefficient of the term $y_{k-1}y_{k-2}$ as the bifurcation parameter μ. Notice the noise-induced shift τ suffered by the first flip bifurcation.

bifurcation information. It is desired to use the new data Z and the *a priori* information I to build such a model.

In this example, the bifurcation information in the dynamical data set will be blurred by adding 20% zero-mean Gaussian noise. In other words, Z_1 was produced by adding 20% noise to Z_0. Proceeding to build an unconstrained (black-box) model yields the following:

$$y_k = 0.2444 + 2.8777y_{k-1} - 0.1698y_{k-2} - 3.0388y_{k-1}^2 + 0.1906y_{k-1}y_{k-2}, \tag{24}$$

for which the bifurcation diagram is shown in Fig. 9. Comparing the diagrams in Fig. 8 and Fig. 9, it is seen that the noise shifts the flip bifurcation to $\mu \approx -0.08$ *and* induces a spurious period-three solution, seen in the figure within the range $-1 < \mu < -0.72$.

In order to build a model $\mathcal{M}_1$ from the *noisy* set of data *and* the prior information I, the following set of constraints should be used

$$\begin{bmatrix} 0.6257 \\ -0.71 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & \bar{y}_1 & \bar{y}_1 & \bar{y}_1^2 & \bar{y}_1^2 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & -2\bar{y}_1 & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_{11} \\ \theta_{12} \end{bmatrix}, \tag{25}$$

$$\mathbf{c} = S\boldsymbol{\theta},$$

where $I=c$ and $g(\mathcal{M}_1)=S\boldsymbol{\theta}$.

A new (gray-box) model can be reestimated from the noisy data Z constrained by Eq. (25) using Eq. (16). Such a model, $\mathcal{M}_1$, is

$$y_k = 0.0988 + 2.6675y_{k-1} + 0.3017y_{k-2} - 2.6896y_{k-1}^2 - 0.71y_{k-1}y_{k-2}, \tag{26}$$

which has periodic points at $\bar{y}_1=0.6257$ (imposed), and $\bar{y}_2=-0.0464$ (free). The Jacobian matrix of model (26) evaluated at $\bar{y}_1$ has eigenvalues at $\lambda_1=-1$ (imposed) and $\lambda_2=-0.1425$ (free). From the bifurcation diagram of model

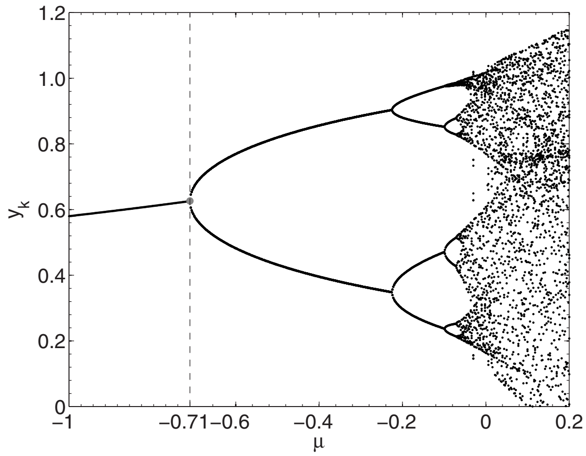

FIG. 10. Bifurcation diagram of the gray-box model (26). Notice how the first flip bifurcation has been successfully imposed at μ=−0.71.

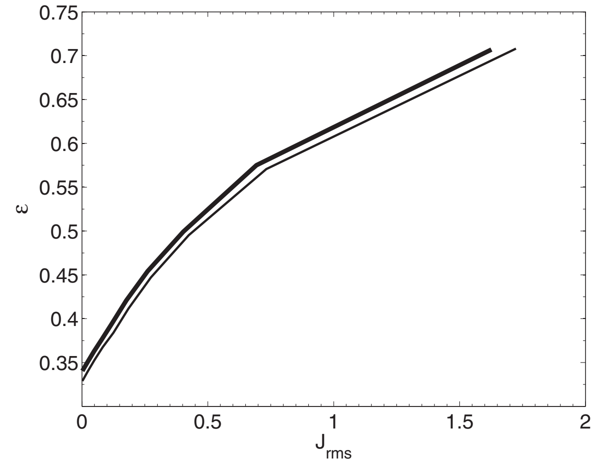

FIG. 11. Synchronization class vs cost of synchronization for identified models. Thick line corresponds to the black-box model (24) for μ=0.55 and the thin line to the gray-box model (26) for μ=0.10. The driving signal was obtained with the "original" model (23) for μ=−0.17.

(26), shown in Fig. 10, it is seen that not only the first bifurcation point has been brought back to the *a priori* (imposed) value μ_c=−0.71, but also the spurious period three regime has been eliminated, at least for the parameter range considered in the figure.

It should be noticed that the constraints used [see Eq. (25)] impose (1) the location of the periodic point at the first flip bifurcation, (2) one eigenvalue of the corresponding Jacobian matrix must be λ=−1, and (3) the specified critical value of the parameter μ. In this way, by imposing λ=−1, a period-doubling bifurcation is necessarily obtained at the chosen parameter value. All the rest of the diagram is shifted as a consequence of the parameter estimation algorithm using the noisy dynamical data.

As done for the previous example, the $\epsilon \times J_{\text{rms}}$ plot was computed for models (24) and (26). This is shown in Fig. 11. In this case, both models show very similar dynamical performances at this bifurcation parameter value. It should become clear at once that as we consider the bifurcation parameter closer to the value used in the constraint, i.e., μ =−0.71, the gray-box model clearly outperforms the black-box model.

V. DISCUSSION AND CONCLUSIONS

The field of modeling from data is mature. Many procedures are now available to build nonlinear models from data. It has been argued that in this field there are two realms of applications that, in practice, have different aims and, therefore, end up using somewhat different procedures, though the tools are basically the same. The two applications are *time series modeling* and *dynamical system modeling*. The former usually aims at forecasting and characterizing the data at hand, whereas the second usually aims at investigating aspects of the "global" dynamics of the system that produced the data.

Of course, such a classification is not unique nor can it be claimed that it precisely divides the wealth of available techniques into two distinct classes. Rather, there is a significant gray zone of methods that could be classified either way, depending on the focus. In fact, probably the best criterion to classify a given method of model building is to pay attention on how the obtained models are validated. This has been recently discussed at some length in [53].

Granted the aforementioned classification, the present paper is concerned with dynamic system modeling. Although much has been accomplished in this field since one of its pioneering papers [7], it seems that little attention has been given to the problem of incorporating additional information into the model building process, as pointed out in [12]. By "additional" it is meant any type of information apart from the time series from which models should be built.

More specifically, in this paper a procedure has been put forward by which it is possible to constrain the resulting model to have a periodic point at a certain location and to force it to undergo a *flip* bifurcation for a chosen value of a bifurcation parameter. Other constraints have been recently used in the context of parameter estimation [70]. In short, this paper has shown how to use *a priori* information (about a periodic point that undergoes a period-doubling bifurcation for a given bifurcation parameter value) in order to write down constraints that are useful in building polynomial models that, by design, will reproduce the used information.

It is interesting to point out that, for the sake of illustration, it was assumed that there is clean data (i.e., lab data) available from which a good model can be obtained. From such a model the *a priori* information is obtained, namely: the location of the periodic point at the first flip bifurcation, the model structure and the critical value of the bifurcation parameter at the flip. However, it should be clear that, in principle, any periodic-point location and any critical bifurcation parameter value could be used. If "unreasonable" values were used, probably the estimated models would be unstable under iteration.

Another interesting point is to notice that the bifurcation parameter could be taken, say, as the summation of the parameters of the quadratic terms of the model (this is known as the quadratic cluster coefficient [59]). For instance, in the last example this would yield the following set of constraints:

$$\begin{bmatrix} 0.6257 \\ -0.71 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & \bar{y}_1 & \bar{y}_1 & \bar{y}_1^2 & \bar{y}_1^2 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & -1 & 1 & -2\bar{y}_1 & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_{11} \\ \theta_{12} \end{bmatrix},$$

$$\mathbf{c} = S\boldsymbol{\theta}. \tag{27}$$

By using the constraints (27), the resulting model would have a periodic point at $\bar{y}_1=0.6257$ which would undergo a period-doubling bifurcation whenever $\theta_{11}+\theta_{12}=-0.71$. Therefore, there are many variations which could be tested following the procedure presented in this paper. For the sake of clarity, such a procedure was developed for a simple structure, but the extension to a more complex polynomial structure (i.e., with more terms, greater degree of nonlinearity, greater dimension, etc.) is possible. On the other hand, the application of these ideas to other representations like network models is not obvious at the moment and should be investigated.

The last remark leads us to point out that the type of *a priori* information that can be used is greatly determined by the model class. In the present paper such information was related to a periodic point and a parameter value at which it should undergo a period-doubling bifurcation. That type of information can be comfortably handled by nonlinear discrete-time models. Other model classes, while being better suited to approximate stronger nonlinearities, are less apt to have *hard* dynamical constraints imposed during model building. This tradeoff between approximation capability and handling of *a priori* information is one of the important points illustrated in the present work.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from CNPq and CAPES and Christophe Letellier for the critical reading of an early version of this paper.

[1] K. J. Åström and P. Eykhoff, Automatica **7**, 123 (1971).
[2] I. J. Leontaritis and S. A. Billings, Int. J. Control **41**, 303 (1985).
[3] D. S. Broomhead and D. Lowe, Complex Syst. **2**, 321 (1988).
[4] K. S. Narendra and K. Parthasarathy, IEEE Trans. Neural Netw. **1**, 4 (1990).
[5] T. Takagi and M. Sugeno, IEEE Trans. Syst. Man Cybern. **15**, 116 (1985).
[6] L. Cao, Y. Hong, H. Fang, and G. He, Physica D **85**, 225 (1995).
[7] M. Casdagli, Physica D **35**, 335 (1989).
[8] J. P. Crutchfield and B. S. McNamara, Complex Syst. **1**, 417 (1987).
[9] G. Gouesbet and J. Maquet, Physica D **58**, 202 (1992).
[10] J. B. Kadtke, J. Brush, and J. Holzfuss, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 607 (1993).
[11] A. I. Mees, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 669 (1993).
[12] P. E. Rapp, T. I. Schmah, and A. I. Mees, Physica D **132**, 133 (1999).
[13] P. Connally, K. Li, and G. W. Irwing, Neurocomputing **70**, 819 (2007).
[14] J. B. Elsner, J. Phys. A **25**, 843 (1992).
[15] R. Bakker, J. C. Schouten, C. L. Giles, F. Takens, and C. M. van den Bleek, Neural Comput. **12**, 2355 (2000).
[16] M. Small, *Applied Nonlinear Time Series Analysis* (World Scientific, Singapore, 2005).
[17] L. A. Aguirre and S. A. Billings, Int. J. Bifurcation Chaos Appl. Sci. Eng. **5**, 449 (1995).
[18] D. Allingham, M. West, and A. I. Mees, Int. J. Bifurcation Chaos Appl. Sci. Eng. **8**, 2191 (1998).
[19] S. A. Billings and D. Coca, Int. J. Bifurcation Chaos Appl. Sci. Eng. **9**, 1263 (1999).
[20] R. J. G. B. Campello, G. Favier, and W. C. do Amaral, Automatica **40**, 815 (2004).
[21] L. A. Aguirre and S. A. Billings, Physica D **80**, 26 (1995).
[22] K. Judd and A. Mees, Physica D **82**, 426 (1995).
[23] S. Lu, K. H. Ju, and K. H. Chon, IEEE Trans. Biomed. Eng. **48**, 1116 (2001).
[24] E. M. A. M. Mendes and S. A. Billings, Int. J. Bifurcation Chaos Appl. Sci. Eng. **8**, 535 (1998).
[25] E. M. A. M. Mendes and S. A. Billings, IEEE Trans. Syst. Man Cybern., Part A. Syst. Humans **36**, 597 (2001).
[26] L. Piroddi and W. Spinelli, Int. J. Control **76**, 1767 (2003).
[27] M. Small and C. K. Tse, Phys. Rev. E **66**, 066701 (2002).
[28] D. A. Smirnov, B. P. Bezruchko, and Ye. P. Seleznev, Phys. Rev. E **65**, 026205 (2002).
[29] T. Nakamura, D. Kilminster, and K. Judd, Int. J. Bifurcation Chaos Appl. Sci. Eng. **14**, 1129 (2004).
[30] R. Brown, N. F. Rulkov, and E. R. Tracy, Phys. Rev. E **49**, 3784 (1994).
[31] D. Xu and F. Lu, Chaos **16**, 043109 (2006).
[32] M. J. Korenberg, S. A. Billings, Y. P. Liu, and P. J. McIlroy, Int. J. Control **48**, 193 (1988).
[33] S. Chen, S. A. Billings, C. F. N. Cowan, and P. M. Grant, Int. J. Control **52**, 1327 (1990).
[34] S. A. Billings and H. L. Wei, IEEE Trans. Neural Netw. **16**, 862 (2005).
[35] H. L. Wei and S. A. Billings, Int. J. Syst. Sci. **35**, 511 (2004).
[36] L. A. Aguirre and S. A. Billings, Int. J. Control **62**, 569 (1995).
[37] T. A. Corrales and J. I. Auñón, IEEE Trans. Biomed. Eng. **47**, 472 (2000).
[38] K. Rodríguez-Vázquez and P. J. Fleming, Knowledge Inf. Syst. **8**, 235 (2005).
[39] B. Feil, J. Abonyi, and F. Szeifert, J. Process Control **14**, 593 (2004).

[40] J. Madár, J. Abonyi, and F. Szeifert, Ind. Eng. Chem. Res. **44**, 3178 (2005).
[41] U. Parlitz, A. Hornstein, D. Engster, F. Al-Bender, V. Lampaert, T. Tjahjowidodo, S. D. Fassois, D. Rizos, C. X. Wong, K. Worden, and G. Manson, Chaos **14**, 420 (2004).
[42] K. Worden, C. X. Wong, U. Parlitz, A. Hornstein, D. Engster, T. Tjahjowidodo, F. Al-Bender, D. D. Rizos, and S. D. Fassois, Mech. Syst. Signal Process. **21**, 514 (2007).
[43] L. A. Aguirre, M. V. Corrêa, and C. C. S. Cassini, IEE Proc.: Control Theory Appl. **149**, 343 (2002).
[44] C. Letellier and L. A. Aguirre, Chaos **12**, 549 (2002).
[45] C. Letellier, L. A. Aguirre, and J. Maquet, Commun. Nonlinear Sci. Numer. Simul. **11**, 555 (2005).
[46] C. J. Cellucci, A. M. Albano, and P. E. Rapp, Phys. Rev. E **67**, 066210 (2003).
[47] K. Judd and A. I. Mees, Physica D **120**, 273 (1998).
[48] L. A. Aguirre and S. A. Billings, Physica D **85**, 239 (1995).
[49] E. Baake, M. Baake, H. G. Bock, and K. M. Briggs, Phys. Rev. A **45**, 5524 (1992).
[50] J. Timmer, H. Rust, W. Horbelt, and H. U. Voss, Phys. Lett. A **274**, 123 (2000).
[51] H. U. Voss, J. Timmer, and J. Kurths, Int. J. Bifurcation Chaos Appl. Sci. Eng. **14**, 1905 (2004).
[52] L. A. B. Tôrres, Physica D **228**, 31 (2007).
[53] L. A. Aguirre, E. C. Furtado, and L. A. B. Tôrres, Phys. Rev. E **74**, 066203 (2006).
[54] L. A. Aguirre, M. F. S. Barroso, R. R. Saldanha, and E. M. A. M. Mendes, IEE Proc.: Control Theory Appl. **151**, 174 (2004).
[55] M. V. Corrêa, L. A. Aguirre, and R. R. Saldanha, IEEE Trans. Circuits Syst., I: Fundam. Theory Appl. **49**, 1376 (2002).
[56] E. Eskinat, S. H. Johnson, and W. L. Luyben, Ind. Eng. Chem. Res. **32**, 1981 (1993).
[57] T. A. Johansen, Automatica **32**, 337 (1996).
[58] H. J. A. F. Tulleken, Automatica **29**, 285 (1993).
[59] L. A. Aguirre and E. M. Mendes, Int. J. Bifurcation Chaos Appl. Sci. Eng. **6**, 279 (1996).
[60] E. G. Nepomuceno, R. H. C. Takahashi, G. F. V. Amaral, and L. A. Aguirre, Int. J. Bifurcation Chaos Appl. Sci. Eng. **13**, 1229 (2003).
[61] L. A. Aguirre, R. A. M. Lopes, G. F. V. Amaral, and C. Letellier, Phys. Rev. E **69**, 026701 (2004).
[62] C. Letellier, T. D. Tsankov, G. Byrne, and R. Gilmore, Phys. Rev. E **72**, 026212 (2005).
[63] G. Byrne, R. Gilmore, and C. Letellier, Phys. Rev. E **70**, 056214 (2004).
[64] G. F. V. Amaral, C. Letellier, and L. A. Aguirre, Chaos **16**, 013115 (2006).
[65] K. H. Chon, J. K. Kanters, R. J. Cohen, and N. H. Holstein-Rathlou, Physica D **99**, 471 (1997).
[66] N. R. Draper and H. Smith, *Applied Regression Analysis*, third ed. (John Wiley and Sons, New York, 1998).
[67] M. Hénon, Commun. Math. Phys. **50**, 69 (1976).
[68] L. A. Aguirre and A. V. P. Souza, Int. J. Bifurcation Chaos Appl. Sci. Eng. **8**, 2203 (1998).
[69] C. Letellier and E. M. A. M. Mendes, Chaos **15**, 013110 (2005).
[70] D. M. Walker, Int. J. Bifurcation Chaos Appl. Sci. Eng. **16**, 1067 (2006).
[71] L. A. Aguirre and S. A. Billings, Int. J. Bifurcation Chaos Appl. Sci. Eng. **4**, 109 (1994).
[72] K. Judd, in *Chaos and Its Reconstruction*, edited by G. Gouesbet, S. Meunier-Guttin-Cluzel, and O. Ménard (Nova Science, New York, 2003), pp. 179–214.
[73] The definitions of term clusters and cluster coefficients can be found in [36].
[74] The dynamical order of a discrete autonomous model is equivalent to the corresponding embedding dimension. The dynamical order should not be confused with the degree of a polynomial model.