



ELSEVIER

Physica D 158 (2001) 1–18

**PHYSICA** D

www.elsevier.com/locate/physd

# Structure-selection techniques applied to continuous-time nonlinear models

Luis A. Aguirre<sup>a,\*</sup>, Ubiratan S. Freitas<sup>a</sup>, Christophe Letellier<sup>b</sup>, Jean Maquet<sup>b</sup>

<sup>a</sup> *Laboratório de Modelagem, Análise e Controle de Sistemas Não Lineares, Departamento de Engenharia Eletrônica, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, Brazil*

<sup>b</sup> *Groupe d'Analyse Topologique et de Modélisation de Systèmes Dynamiques, CORIA UMR 6614, Université et INSA de Rouen, Place Emile Blondel, F-76821 Mont Saint-Aignan Cedex, France*

Received 6 September 2000; received in revised form 12 April 2001; accepted 28 June 2001

Communicated by C.K.R.T. Jones

---

## Abstract

This paper addresses the problem of choosing the multinomials that should compose a polynomial mathematical model starting from data. The mathematical representation used is a nonlinear differential equation of the polynomial type. Some approaches that have been used in the context of discrete-time models are adapted and applied to continuous-time models. Two examples are included to illustrate the main ideas. Models obtained with and without structure selection are compared using topological analysis. The main differences between structure-selected models and complete structure models are: (i) the former are more parsimonious than the latter, (ii) a predefined fixed-point configuration can be guaranteed for the former, and (iii) the former set of models produce attractors that are topologically closer to the original attractor than those produced by the complete structure models. © 2001 Elsevier Science B.V. All rights reserved.

*Keywords:* Structure selection; Topological analysis; Equations of motion from data

---

## 1. Introduction

It is an accepted fact that one of the great challenges in science is to obtain mathematical models that are able to explain and reproduce observed phenomena. A second challenge is to be able to construct such models from observed data. In order to do so, one must choose a mathematical representation to work with and subsequently define the model structure. Finally, assuming the representation is parametric, the model parameters are estimated in such a way as to optimally fit the data according to a given criterion. At present there are a number of *global* mathematical representations such as discrete autoregressive-type models [8], continuous-time models [6,12,19,22,23,30,39,41], radial basis function models [10,14] and even a composition of more than one type of basis functions [9,25]. Which to choose depends largely on the problem at hand, on the number and quality of the data and on a diversity of subjective reasons such as acquaintance with a particular representation and personal experience. Many of the aforementioned representations have been compared in [15].

---

\* Corresponding author. Tel.: +55-31-3499-4866; fax: +55-31-3499-4850.

E-mail address: [aguirre@cpdee.ufmg.br](mailto:aguirre@cpdee.ufmg.br) (L.A. Aguirre).

On the other hand, the choice of the model structure, i.e. the selection of the basis functions (regressors) that will compose the model is critical in most cases, regardless of the representation being used. The main motivation for estimating models with simplified structure is *not* computational, but rather dynamical, in the sense that models with incorrect structure typically display spurious dynamical regimes and may have a wrong configuration of fixed points (incorrect number and with different symmetry properties). Consequently, much attention has been devoted to the problem of how to adequately choose a parsimonious structure for nonlinear models [3,7,8,11,16,26,38]. Quite surprisingly, it seems that all such papers concern discrete-time models with the exception of the papers by Brown and colleagues and that by Bezruchko and Smirnov. Both papers deal with continuous-time models. In the former the authors delete certain subsets of terms from the entire candidate set in order to guarantee specific symmetry properties of the model function. In the latter, the authors consider the specific case of harmonically driven systems.

This paper discusses the application of structure-selection techniques to continuous-time polynomial models. Such a representation, to be described in Section 2.1, has been used with success to model real nonlinear systems from observed data. So far, no general structure-selection technique has been applied to such models [34,36]. This will be done and discussed in the present work. The procedure will be tested on simulated and real-data examples as discussed in Section 3. The results will be compared to those obtained without structure selection and the main advantages and difficulties of the procedure will be summarized in Section 4.

## 2. Modeling from data

### 2.1. The model representation

Consider a continuous-time dynamical system described by a set of ordinary differential equations,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; \boldsymbol{\mu}), \quad (1)$$

where  $\mathbf{x}(t) \in \mathbb{R}^n$  is a vector-valued function depending on a parameter  $t$  called the time and  $\mathbf{f}$ , the so-called vector field, is an  $n$ -component smooth function generating a flow  $\phi_t$ . Also,  $\boldsymbol{\mu} \in \mathbb{R}^p$  is the parameter vector with  $p$  components, assumed to be constant in this work. The system (1) is called the *original system* and for, say,  $n = 3$  can be written as

$$\dot{x} = f_1(x, y, z), \quad \dot{y} = f_2(x, y, z), \quad \dot{z} = f_3(x, y, z). \quad (2)$$

It is now assumed that the observer numerically records a single scalar time series. By convention, in this section, the observable is taken to be  $X_1 = x$ .

The aim is then to model a vector field equivalent to the original system using a basis consisting of the observable and its derivatives such as

$$\dot{X}_1 = X_2, \quad \dot{X}_2 = X_3, \quad \dots, \quad \dot{X}_{d_e} = F(X_1, X_2, \dots, X_{d_e}), \quad (3)$$

where  $d_e$  is the embedding dimension and  $F$  depends on the  $d_e$  variables which are  $x$  and the  $(d_e - 1)$  successive derivatives of  $x$ .  $F$  is called the model function and can be estimated by using a multivariate polynomial basis on nets [23] which may be built by means of a Gram–Schmidt orthogonalization procedure. The algorithm requires the definition of modeling parameters which are (i)  $d_e$ , the embedding dimension, (ii)  $N_c$ , the number of points from the data and their derivatives at which the function is evaluated. Such points will be referred to as centers in this paper, (iii)  $\Delta t$ , the time step between two successive centers. In general  $\Delta t$  is constant, but this is not a requirement,

(iv)  $N_p$ , the number of retained multinomials and (v)  $\tau_w$ , the window length on which the derivatives are computed by using polynomial interpolation over the window. Derivatives are then obtained by analytical differentiation. The estimated model function,  $\hat{F}$ , is

$$\hat{F}(X_1, X_2, \dots, X_{d_e}) = \sum_{p=1}^{N_p} \theta_p \psi^p, \quad (4)$$

where  $\theta_p$  are the parameters and  $\psi^p$  are multivariate monomials (or multinomials) of the form

$$\psi^p = X_1^{n_1}, X_2^{n_2}, \dots, X_{d_e}^{n_{d_e}}, \quad (5)$$

where the integers  $p$  are related to  $n_{d_e}$ -uplets  $(n_1, n_2, \dots, n_{d_e})$  by a bijective relationship discussed in [23]. The modeling parameters  $d_e$ ,  $N_c$ ,  $\Delta t$ ,  $N_p$  and  $\tau_w$  can be determined with the aid of an error function. Typically, at least 50 points per pseudo-period are required for a correct computation of the derivatives. When less than 50 samples per cycle are available, the data must be interpolated and/or smoothed in some way. In order to realize a spectrum-preserving interpolation, a Fourier method, based on zero padding of the direct Fourier Transform of the original time series may be used. This procedure is not sensitive to the number of bits used in digitization since a slight smoothing allows to recover a sufficient continuity between two successive points.

Writing (4) at  $N_c$  centers on the data yields a set of  $N_c$  equations of the form

$$\begin{bmatrix} \dot{X}_{d_e}(1) \\ \dot{X}_{d_e}(2) \\ \vdots \\ \dot{X}_{d_e}(N_c) \end{bmatrix} = \begin{bmatrix} \psi^{r_1}(1) & \dots & \psi^{m_p}(1) \\ \psi^{r_1}(2) & \dots & \psi^{m_p}(2) \\ \vdots & & \vdots \\ \psi^{r_1}(N_c) & \dots & \psi^{m_p}(N_c) \end{bmatrix} \begin{bmatrix} \theta_{r_1} \\ \vdots \\ \theta_{m_p} \end{bmatrix}, \quad (6)$$

$$\dot{\mathbf{x}} = \Psi \boldsymbol{\theta},$$

where  $r_i \neq r_j$  and  $1 \leq r_1, \dots, m_p \leq N_p$  and the numbers inside the parentheses indicate to which center the variable is related. It should be noted that at each center only one variable is measured and the successive derivatives are estimated. Once the model structure is determined (more on this in the next section) the parameter vector  $\hat{\boldsymbol{\theta}} = [\hat{\theta}_{r_1} \dots \hat{\theta}_{m_p}]^T$  can be estimated by standard least-squares techniques. The choice of which multinomials should compose the model, i.e. the choice of  $r_1, \dots, m_p$ , is fundamental in modeling problems and will be described in what follows.

## 2.2. Structure selection

The structure-selection problem in the continuous-time polynomial representation is to choose which multinomials  $\psi^p$ , out of a finite candidate set, should be included in the estimated model (4). The chosen multinomials (terms) should be the most relevant ones to the particular system under study. This section discusses different approaches help to select the most important terms of a model.

The first technique is to use information about the minimum number and possible symmetry of the original system fixed points. Using prior knowledge seems to be realistic even in the case where experimental data are investigated. For the continuous model (3) the fixed points are calculated taking

$$X_2 = 0, \quad X_3 = 0, \quad \dots, \quad \sum_{i=1}^{\ell} \theta_i (X_1)^i = 0, \quad (7)$$

where the  $\theta_i$  are the coefficients of the monomials which are independent of the coordinates  $X_2, X_3, \dots, X_{d_c}$  and  $\ell = n_1 + n_2 + \dots + n_{d_c}$  is the degree of nonlinearity. So, from (7), it is clear that only multinomials of the form  $X_1^i$  determine the values of the  $\ell$  fixed points. Moreover, if one knows how many fixed points are there in the original system and if  $\ell$  is set to such a value, the estimated model will have the same number of fixed points as the original system by construction. Also, information about the fixed-point symmetry is clearly related to the terms of the form  $(X_1)^i$  [2]. Information about the minimum number and possible symmetry of fixed points can be estimated directly from data [4]. A similar equivalent approach has been described in [11], but in that case the multinomials were deleted from the set of candidates based on symmetry properties of the functional form and not of the model fixed points.

Eq. (7) shows that for continuous-time models a relatively small number of terms is related to the model fixed points. Therefore, a major question still remains, namely which multinomials that do not affect the location and number of fixed points should be included in the model? In order to provide an answer to this question a second technique will be described which is based on a criterion known as the error reduction ratio (ERR) [8].

The ERR approach seems to be generally applicable since recently it was successfully applied to different type of model representations such as Volterra series [21] and forward neural networks [24]. Such a criterion is here used and applied to models of the form (3), apparently for the first time and are *not* restricted to systems with any particular type of symmetry. As a matter of fact such an approach can be applied in any problem in which a subset of regressors need to be chosen from a (usually much) larger set of candidates as for instance the choice of a limited number of centers of radial basis function models [17].

First, parameter estimation is performed for a linear-in-the-parameters model of the type

$$\dot{X}_{d_c}(t) = \sum_{i=1}^{N_p} g_i w_i(t) + \xi(t), \quad (8)$$

where  $\xi(t)$  is the error obtained when approximating  $\dot{X}_{d_c}(t)$  with the independent variables  $w_i(t)$ ,  $i = 1, \dots, N_p$ ;  $g_i$  are constant parameters and  $w_i(t)$  are constructed to be orthogonal over the data records, i.e.

$$\frac{1}{N_c} \sum_{t=1}^{N_c} w_i(t) w_j(t) = 0, \quad \overline{w_i(t) w_j(t)} = 0 \quad \forall i \neq j, \quad (9)$$

or simply  $\langle \mathbf{w}_i, \mathbf{w}_j \rangle = 0$ , where  $\mathbf{w}_i = [w_i(1) \dots w_i(N_c)]^T$ . In what follows, sometimes the argument  $t$  will be omitted but it should be borne in mind that for the continuous-time models, each restriction of the matrix equation (6) corresponds to a specific time (center). This is one of the greatest differences between the continuous and discrete-time model representations. The sum of square values of  $\dot{X}_{d_c}$  is  $\langle \dot{\mathbf{X}}_{d_c}, \dot{\mathbf{X}}_{d_c} \rangle$  or simply  $\dot{\mathbf{X}}_{d_c}^T \dot{\mathbf{X}}_{d_c}$ . Therefore, multiplying (8) by itself yields

$$\dot{X}_{d_c}^2 = \left( \sum_{i=1}^{N_p} \hat{g}_i w_i(t) + \xi(t) \right) \times \left( \sum_{j=1}^{N_p} \hat{g}_j w_j(t) + \xi(t) \right). \quad (10)$$

Taking the time average of (10) over the data records results in

$$\begin{aligned} \langle \dot{\mathbf{X}}_{d_c}, \dot{\mathbf{X}}_{d_c} \rangle &= \sum_{i=1}^{N_p} \hat{g}_i^2 \langle \mathbf{w}_i, \mathbf{w}_i \rangle + \sum_{i=1, j=1 \forall i \neq j}^{N_p} 2 \hat{g}_i \hat{g}_j \langle \mathbf{w}_i, \mathbf{w}_j \rangle + 2 \sum_{i=1}^{N_p} \hat{g}_i \langle \mathbf{w}_i, \boldsymbol{\xi} \rangle + \langle \boldsymbol{\xi}, \boldsymbol{\xi} \rangle \\ &= \sum_{i=1}^{N_p} \hat{g}_i^2 \langle \mathbf{w}_i, \mathbf{w}_i \rangle + \langle \boldsymbol{\xi}, \boldsymbol{\xi} \rangle. \end{aligned} \quad (11)$$

The last step in Eq. (11) was obtained remembering that, by construction, the regressors  $\mathbf{w}_i$  are orthogonal over the data and assuming that they are not correlated to the error vector  $\xi$ .

An interpretation of Eq. (11) is that the mean square value of  $\dot{X}_{de}(t)$  can be explained, using an orthogonal basis, as the summation of the square values of each orthogonal regressor multiplied by the respective squared parameter. The part that remains unexplained by the regressors equals the addition of the squared values of the error,  $\xi(t)$ .

Therefore, Eq. (11) enables one to quantify the importance of each independent variable, individually. To see this more clearly, consider the case in which the model is empty, i.e.  $N_p = 0$ . Obviously, in this case the “model” error is the entire observed signal, because the model does not explain the data at all. If the  $i$ th independent variable is added to the model, then the explained part is now  $\hat{g}_i^2 \langle \mathbf{w}_i, \mathbf{w}_i \rangle$ . Hence the ERR due to the inclusion of the  $i$ th independent variable can be expressed in terms of the signal mean square value  $\langle \dot{\mathbf{X}}_{de}, \dot{\mathbf{X}}_{de} \rangle$  as [8]

$$[\text{ERR}]_i = \frac{\hat{g}_i^2 \langle \mathbf{w}_i, \mathbf{w}_i \rangle}{\langle \dot{\mathbf{X}}_{de}, \dot{\mathbf{X}}_{de} \rangle}, \quad i = 1, 2, \dots, N_p. \quad (12)$$

Therefore the choice of the best independent variables among a large set of candidates can be accomplished selecting those with largest ERR. Also, for model (8), represented using an orthogonal basis, the parameters can be determined by

$$\hat{g}_i = \frac{\langle \mathbf{w}_i, \dot{\mathbf{X}}_{de} \rangle}{\langle \mathbf{w}_i, \mathbf{w}_i \rangle}, \quad i = 1, \dots, N_p. \quad (13)$$

In Appendix A, an algorithm is described that uses (12) and (13) to choose the most important regressors and estimate the respective parameters, simultaneously. Similar algorithms are the singular value decomposition (SVD) and principal component analysis (PCA) [20]. Some of these methods have been compared in the context of model selection in [18] and more recently in [24] where it has been reported that the ERR implementation runs typically much faster than the conventional SVD algorithm. Another interesting difference (see Appendix A) is that there is an available algorithm that estimates parameter and performs structure selection iteratively.

### 3. Results

#### 3.1. The Lorenz system

This section will consider the well-known Lorenz system [37]. To start with, 10 000 data points with 100 samples per pseudo-period were taken from the  $z$  variable in order to obtain a model. From such a time series 1125 centers were used. In this section, in addition to the structure-selection problem of choosing which terms should compose the model, the influence of center selection is also considered. It is pointed out that because the  $z$  variable is being observed, the symmetry plays no role in this problem because it is “invisible” from such an observable [32]. Therefore, this example illustrates the case in which the method is applied to systems with no symmetry.

The embedding dimension was estimated using the algorithm described in [13] which indicated that a 3D reconstructed phase space will suffice to describe the dynamics of the Lorenz system as observed from the  $z$  variable. One way of obtaining a model, without structure selection, is to predetermine an order of including each of the possible multinomials and gradually increase the number of such terms,  $N_p$ , until a valid model is obtained. However, if no good model is found, some of the modeling parameters can be altered and the procedure is repeated. An important feature of this approach is that the terms are always introduced into the model in the same predefined order [23]. Of course, testing all possible combinations would be so computationally demanding that it was not tried. The ERR criterion (12) will provide an alternative way to order all the possible multinomials.

If no structure selection is performed very few models are found that produce an attractor close to the original one. On the other hand, performing structure selection and choosing the centers equally spaced in time results in significant improvement. By centers here it is meant the points in state space used to write the set of restrictions. At these points the flow is approximated by a polynomial as seen in Eq. (4). In the context of radial basis functions, the centers have a different interpretation but still it seems that the problem of center selection in both representations is similar. The choice of centers in the context of radial basis functions has been investigated in [10,42]. Fig. 1 shows bi-dimensional projections of the original and model attractors. The original attractor is here the one induced by the  $z$  variable of the Lorenz system, i.e. embedded in the differential space spanned on the derivative coordinates  $(X, Y, Z) = (z, \dot{z}, \ddot{z})$  computed from the original dynamics. The use of differential embedding induced by the  $z$  variable, rather than the original phase portrait thus simplifies the comparison between the model attractors and the original system. In this case, no prior knowledge about fixed points is used but only the automatic ERR structure selection. Perhaps the main difference observed is that models with  $N_p = 24$ ,  $N_p = 46$  and  $N_p = 13$  produce attractors that are less developed than the original, i.e. the population of periodic orbits is not so large. On the other hand, the model with  $N_p = 50$  terms is only slightly different from the model with  $N_p = 13$ .

A glance at the attractors shown in Fig. 1 reveals that structure selection combined with center selection has indeed enabled to obtain a very competitive model that has nearly four times less terms than the best model obtained without structure selection. In this way, it was possible to obtain a 13-term model that reproduces the original dynamics very well. No doubt, simply pruning the 50-term model down to the first 13 terms and estimating parameters anew does not yield a good model at all.

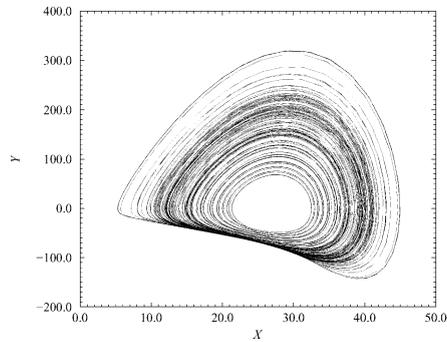
In this example, the choice of centers equally spaced in phase space proved helpful to improve model performance, in particular it helps to obtain models with attractors that have population of periodic orbits closer to the original. To understand the reason for this, consider Fig. 2 where the probability of the system visiting different regions of the Poincaré section  $P$  is shown. Clearly, the center of the Poincaré section is visited more often than the periphery. This means to say that if the centers are chosen to be equally spaced in time, as it usually happens, there is a high chance that very few or even no centers are chosen at the periphery of the attractor. Consequently, the modeling algorithm will have very little information of the dynamics in this part of the attractor. Conversely, most of the dynamical information will be related to the center of the chaotic band. This partially explains why models with  $N_p = 24$  and  $N_p = 46$  terms have approximately the right shape but fail to extend to the limits of the original attractor. When the centers are chosen equally spaced in the state space, the various parts of the attractor are equally well presented to the modeling algorithm. The main result of this was an obvious extension of the first-return map, as can be seen from Fig. 3d.

It is also important to notice that it becomes very difficult to decide from the bi-dimensional projections of Fig. 1 which is the best model between the two obtained using center selection. In order to do so, such models will be compared using topological analysis.

Although the original system is equivariant, the  $z$ -induced original attractor is not, since it is reconstructed from the variable that remains unchanged under the rotation symmetry [29,32]. In such a case, a Poincaré section is defined as,

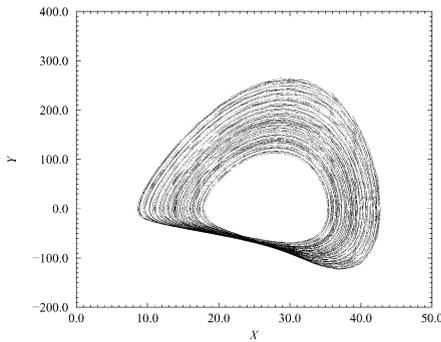
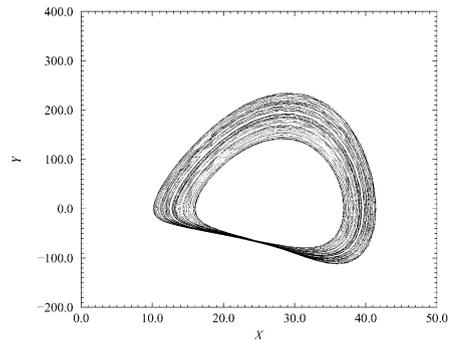
$$P \equiv \{(Y_n, Z_n) \in \mathbb{R}^2 | X_n = R - 1, \dot{X}_n > 0\}.$$

The first-return map to the Poincaré section  $P$  computed for the original dynamics (Fig. 3a) is equivalent to the Lorenz map obtained by Lorenz [37]. The other maps shown in Fig. 3 were computed from the models. Depending on the model used, the shape of the first-return map may differ significantly from the original one. It becomes apparent that the models with  $N_p = 24$  and  $N_p = 46$  terms are inadequate. On the other hand, models with  $N_p = 50$  and  $N_p = 13$  terms are quite similar. The two main differences seem to be the following. The attractor of the model with  $N_p = 13$  terms has a slightly smaller number of periodic orbits than for the 50-term model. In fact, only three



(a) Original dynamics

## Models without center selection

(b)  $N_p = 46$ (c)  $N_p = 24$ , with ERR

## Models with center selection

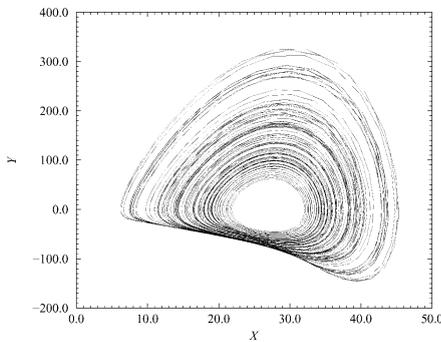
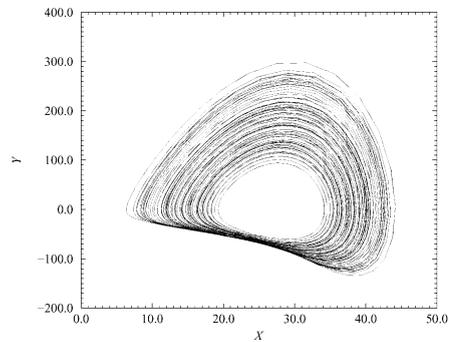
(d)  $N_p = 50$ (e)  $N_p = 13$ , with ERR

Fig. 1. Bi-dimensional projections of Lorenz attractor reconstructed (a) from the simulated  $z$  variable, and estimated models with (b) 46 terms, (c) 24 terms, (d) 50 terms and (e) 13 terms. For models in (b) and (c) the centers were chosen equally spaced in time, whereas for (d) and (e) the centers were chosen equally spaced in phase space. Moreover, the ERR criterion has been applied to obtain the models with attractors (c) and (e).

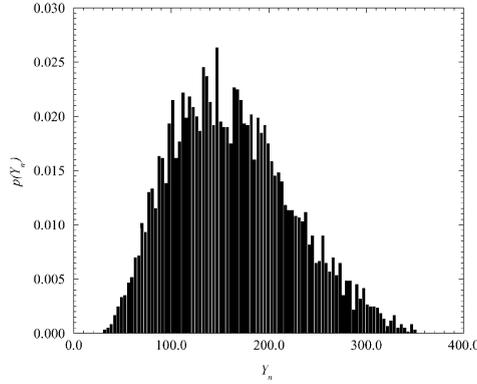


Fig. 2. Probability density function of visits to the Poincaré section. The middle part of the strip is more visited than the periphery.

out of 67 periodic orbits with period less than 9 are not found within the attractor generated by the former. On the other hand, the model with  $N_p = 50$  terms reveals a layered structure at the cusp. These ghost effects have been previously related to detrimental effects of over-parametrization in discrete-time models [1] and the same seems to be true for the present continuous-time models. This scenario has been avoided by means of the structure-selection procedure presented in Section 2.2. Note that the validation techniques here used are not restricted to 3D systems.

### 3.2. Circuit and data

The system under study is the well-known Chua's circuit, shown in Fig. 4. In this circuit, the nonlinearity comes from the so-called Chua's diode. The equations that describe the circuit are

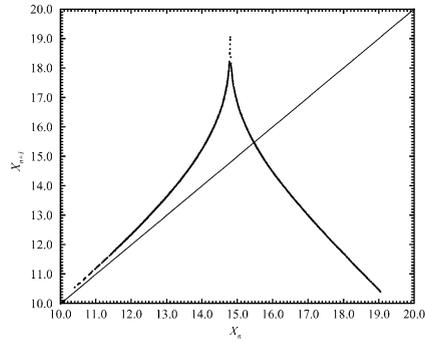
$$C_1 \frac{dv_1}{dt} = \frac{v_2 - v_1}{R} - i_d(v_1), \quad C_2 \frac{dv_2}{dt} = \frac{v_1 - v_2}{R} + i_L, \quad L \frac{di_L}{dt} = -v_2, \quad (14)$$

where  $v_i$  is the voltage across capacitor  $C_i$ ,  $i_L$  is the current through the inductor and the current through Chua's diode is given by

$$i_d(v_1) = \begin{cases} m_0 v_1 + B_p(m_0 - m_1), & v_1 < -B_p, \\ m_1 v_1, & |v_1| \leq B_p, \\ m_0 v_1 + B_p(m_1 - m_0), & v_1 > +B_p. \end{cases} \quad (15)$$

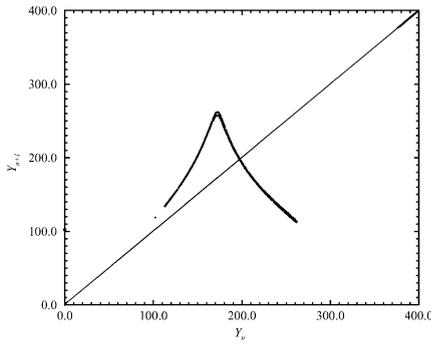
The following component values were used:  $C_1 = 11.0 \mu\text{F}$ ,  $C_2 = 45 \mu\text{F}$ ,  $L = 22 \text{ H}$  and  $R$  is a  $2.0 \text{ k}\Omega$  trimpot. Chua's diode was built using the two-operational-amplifier configuration suggested in [27] and the following parameters were measured:  $m_0 = -0.37 \pm 0.04 \text{ mS}$ ,  $m_1 = -0.68 \pm 0.04 \text{ mS}$  and  $B_p = 1.1 \pm 0.2 \text{ V}$ . This system has an inversion symmetry as discussed in [33]. It has three fixed points, one located at the origin of the phase space, and two, one being symmetric from the other, which have coordinates  $(v_1, v_2, i_L) = (B_p(m_0 \pm m_1)/(1/R - m_0), 0, -B_p(m_0 \pm m_1)/(1 - Rm_0))$ .

Varying the trimpot  $R$ , the dynamics of this circuit settles to several different regular and chaotic attractors. Of particular interest in this paper will be the double-scroll attractor attained at  $R \approx 1800 \Omega$ . In what follows, the observable was the voltage across the capacitor which is connected in parallel with Chua's diode. The data were collected from a real implementation of Chua's circuit. A 12 bit A/D acquisition board was used working at a sampling rate of 300 Hz. The whole data set is composed of  $10^5$  observations with approximately 100 samples per pseudo-period.

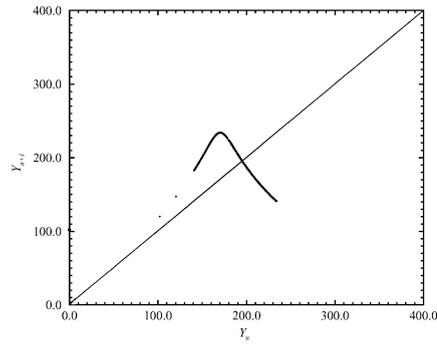


(a) Original system

Models without center selection

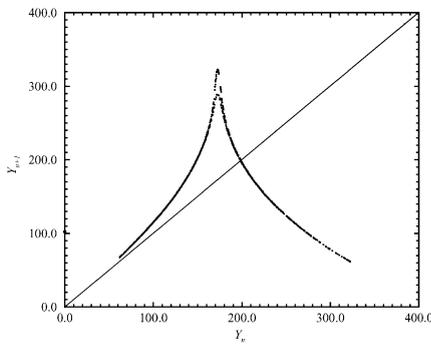


(b)  $N_p = 46$

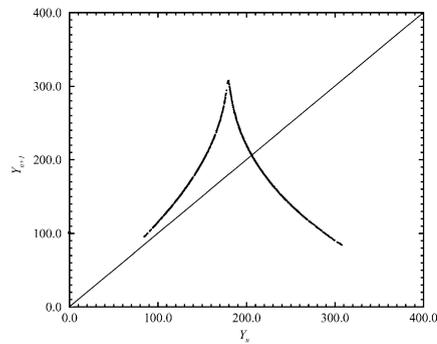


(c)  $N_p = 24$  with ERR

Models with center selection



(d)  $N_p = 50$



(e)  $N_p = 13$  with ERR

Fig. 3. First-return maps to the Poincaré section of Lorenz attractor obtained from (a) the simulated  $z$  variable, and estimated models with: (b) 46 terms, (c) 24 terms, (d) 50 terms and (e) 13 terms. For models in (b) and (c) the centers were chosen equally spaced in time, whereas for (d) and (e) the centers were chosen equally spaced in phase space. Moreover, the ERR criterion has been applied to obtain models with attractors (c) and (e).

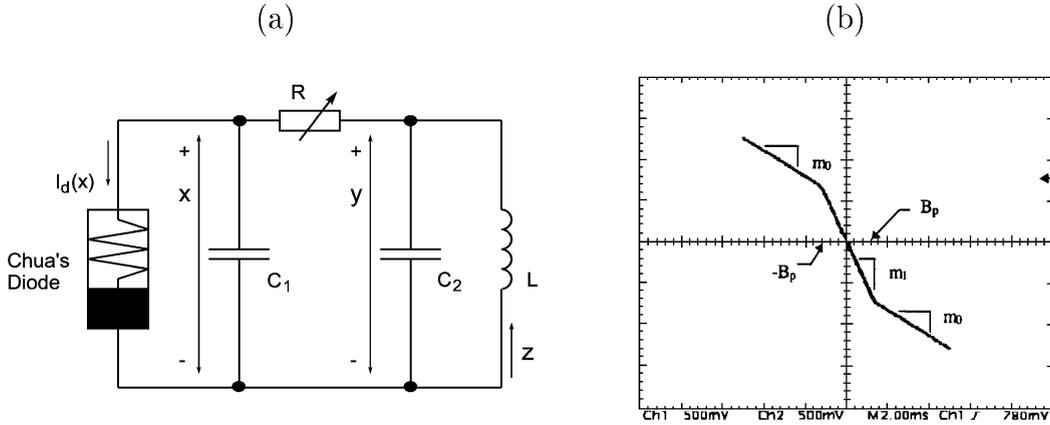


Fig. 4. (a) Chua's circuit, and (b) Chua's diode.

### 3.2.1. The parsimonious models

A 38-term model was obtained from the training time series without any structure selection, therefore, hereafter, such a model will be referred to as the *full model*. Conversely, the models obtained implementing the structure-selection procedure described in Section 2.2 will be referred to as *parsimonious models* in the sense that all such models will have less than 38 terms.

The parsimonious models were obtained by deleting the constant and  $(X_1)^2$  terms from the set of candidate multinomials. This is done to guarantee fixed-point symmetry in the estimated models, as discussed in the beginning of Section 2.2. The remaining terms in the model were automatically selected using (12). It should be noted that the order in which the terms are included into the model now is totally different from that of the full model. As will be seen below, this can result in parsimonious models with better dynamics and less terms.

### 3.2.2. Topological characterization

The chaotic attractor is embedded in a space spanned by the successive derivatives of the experimentally recorded time series,  $x$ . Such an embedding will permit the comparison of the state portrait reconstructed from the experimental data with the portrait generated by the model. The phase space reconstructed from the recorded time series has an embedding dimension equal to three as estimated by the algorithm described in [13], which coincides with the dimension of the original phase space. However, five derivative coordinates were used to obtain a global model. The two extra coordinates are required to obtain a good model. A similar feature has been observed when a global model is attempted from the  $z$  variable of the Rössler system. Only a 4D model has been obtained for successfully reproducing the dynamics [35]. In the present case, such an increase in the number of dynamical variables might result from the fact that the model function must be sufficiently complex to reproduce the discontinuous nonlinearity by a continuous polynomial of low degree.

Since the original dynamics can be embedded in a 3D space, only the first three coordinates will be used to characterize the topological structure of the reconstructed state space. Therefore, the first three coordinates are

$$\mathbf{x} = \begin{cases} X_1 = x(t), \\ X_2 = \dot{x}, \\ X_3 = \ddot{x}. \end{cases} \quad (16)$$

The chaotic attractor is displayed in Fig. 5. Since the system presents an order two inversion symmetry, every point in the time series may have a counterpart on the opposite side of the attractor. Consequently, the reconstructed

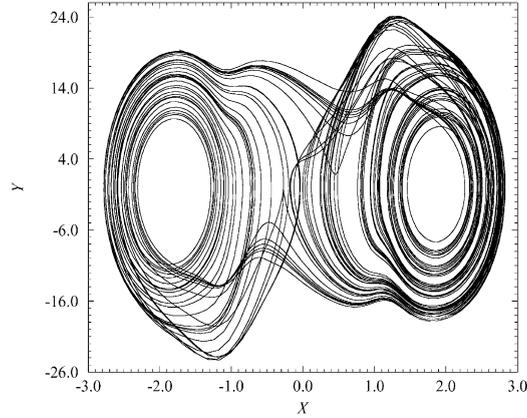


Fig. 5. Plane projection of the  $x$ -induced attractor generated by the experiments.

attractor should also present an inversion symmetry [32]. This information will be used at the end of this section and will prove to be useful in enhancing model quality. The phase portrait reconstructed from the  $x$  time series presents an inversion symmetry characterized by an operator  $\gamma$ , i.e.

$$\gamma \mathbf{F}(\mathbf{x}) = \mathbf{F}(\gamma \mathbf{x}), \quad (17)$$

where

$$\gamma \equiv \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (18)$$

and  $\mathbf{F}$  is the vector field associated with the model. This symmetry property is the one observed on the original phase space since the set of Eq. (14) is also equivariant with respect to the same operator  $\gamma$ . Since the double-scroll attractor is investigated, the original attractor is globally invariant under the action of the operator  $\gamma$ .

It has been shown that in the presence of symmetries, the dynamics could be advantageously analyzed in a fundamental domain modding out the symmetry properties [32] or more rigorously with the help of an image system [31]. The analysis then provides characteristic features associated with the underlying dynamics. It has also been illustrated that such a fundamental characterization could greatly simplify the analysis [33].

Again, the analysis of the dynamics starts with the computation of a first-return map obtained from a Poincaré section. In the case of a symmetric attractor, a fundamental map is used. This can be viewed as a first-return map to a Poincaré set defined as

$$\{(X_2, X_3) \in \mathbb{R}^2 | X_1 = 1.9, X_2 < 0\} \cup \{(X_2, X_3) \in \mathbb{R}^2 | X_1 = 1.9, X_2 > 0\} \quad (19)$$

built on the absolute value of one of the two variables ( $X_1, X_2$ ). The first-return map shown in Fig. 6 presents two critical points. The fact that this map does not exhibit any layered structure, confirms the hypothesis of symmetry.

### 3.2.3. Model validation

The first step in the validation was a simple visualization of plane projections of the attractors produced by each of the identified models. Many models were either unstable or simply settled to attractors very different from the original one. See Table 1 for a summary of these results for both the original and new procedures.

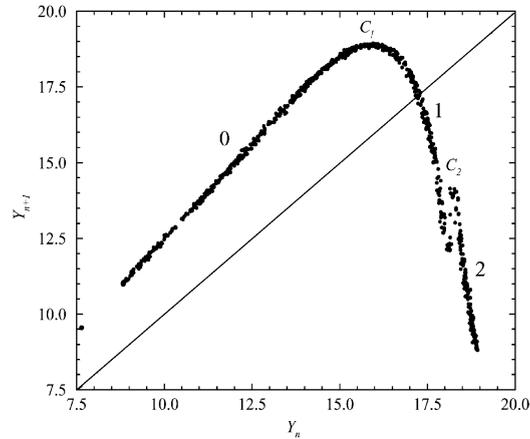


Fig. 6. Fundamental first-return map of the attractor reconstructed from the experimental data.

After this step 16 models were chosen as candidates. In a second step the largest Lyapunov exponents were estimated and this reduced the number of candidates down to three, namely the 29, 32 and 36-term models. In order to help comparisons, besides those three models the 18-term model was selected to perform the topological characterization. Fig. 7 shows the projections of the model attractors. It is pointed out that it would be quite difficult to choose any particular model solely based on these results.

The third step was to compare the fundamental first-return map computed for attractors generated by the models with the one associated with the experimental data, as shown in Fig. 8. According to this criterion, two other models are rejected. Indeed, models with 36 and 38 terms are characterized by fundamental first-return maps that have a layered structure which is not observed in the experimental data. In the case of the 36-term model, the only structure selection performed was based on fixed-point symmetry. The layered structure of the first-return maps of models with 36 and 38 terms could be an indication of over-parametrization [1]. As a result of this, the system becomes unsymmetrical [33].

On the other hand, the first-return maps of models with 29 and 32 terms are much closer to the experimental one and despite the differences, it seems fair to conclude that structure selection has been able to reduce the complexity of the full model resulting in parsimonious models with improved dynamics.

A possible reason for the modeling difficulties observed in this example is related to the fact that the flow is very fast in the vicinity of the trivial fixed point. In practice that means that very few centers are available in that region and therefore approximation of the flow around the trivial fixed point is generally poor. It is noteworthy that no such shortcoming is observed when discrete-time models are used [3]. Another important difference that could

Table 1  
Resulting dynamics of estimated models obtained without and with structure selection<sup>a</sup>

	Structure selection	
	Without ( $10 \leq N_p \leq 38$ )	With ( $8 \leq N_p \leq 36$ )
Unstable	18	4
Fixed-point or limit cycle	10	9
Spiral chaos	–	–
Double-scroll chaos	1	16

<sup>a</sup> The values inside the parenthesis represents the range of terms covered.

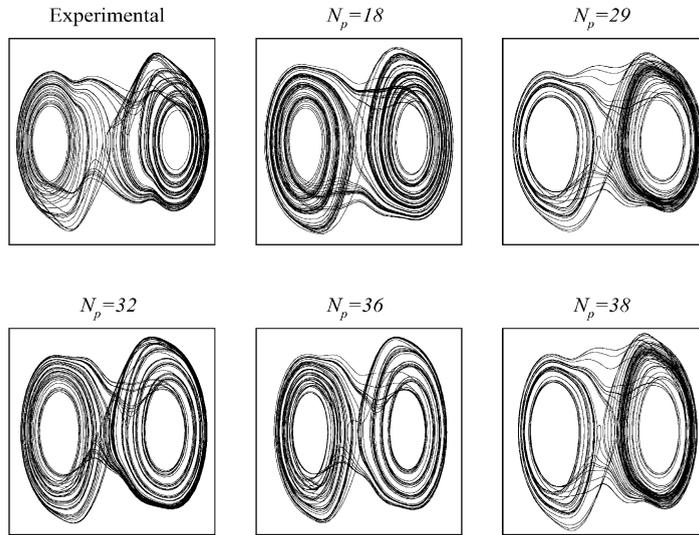


Fig. 7. Plane  $(x \times \dot{x}) \equiv (X_1 \times X_2)$  projections of the original attractor, the four parsimonious models and the full model ( $N_p = 38$ ).

be partially responsible for the different performances presented by these two representations is that whereas the parameters of a discrete model are usually estimated such as to minimize the one-step-ahead predictions, in the case of differential models the parameters minimize the flow at certain points (the centers) in state space.

The aforementioned difficulties and alternative solutions, such as deliberately choosing more centers close to the trivial fixed point, are currently being investigated and will be reported in a forthcoming paper. Nonetheless, the main point for now is that the techniques discussed in Section 2.2 can be used to obtain parsimonious continuous-time models.

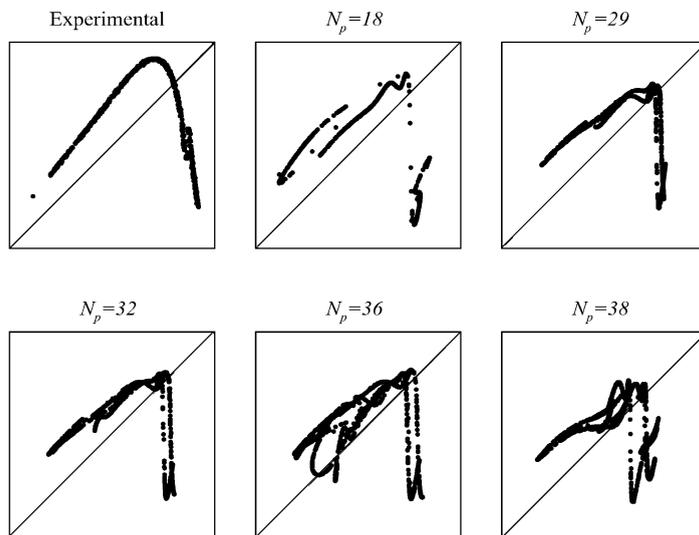


Fig. 8. First-return maps for the five different models.

The shortcomings of the identified models have been revealed mainly by a more detailed validation, using topological analysis. If such models had been validated using less exacting tools, as is usually the case in the literature, they would look very much alike and quite accurate.

As pointed out previously, because the system has an order two inversion symmetry, it is obviously desirable that the same should be true for the model. In order to help the algorithm account for this, a new model was obtained using a measure which respects the symmetry of the attractor, that is for each point in the differential embedding space, a symmetrical counterpart was included in the data used to estimate the model. For instance, if the original data set had the point  $\mathbf{x} = [X_1 X_2 \dots X_5]^T$ , then the mirror image of such a point,  $-\mathbf{x}$ , was inserted in the data set before choosing the centers. The modeling procedure outlined before was exactly the same in this case. For  $N_c = 2000$  and  $N_p = 20$ -term model was found. The first-return map of which is shown in Fig. 9. It is interesting to point out that among the candidate multinomials there are constant, linear, quadratic and cubic terms and that the ERR algorithm automatically excluded from this model the constant and quadratic terms which, if present, would destroy symmetry of the flow. Consequently symmetry of fixed points is still guaranteed because of the lack of the two terms: constant and  $X_1^2$ . It should be mentioned that this model possesses an important property of the experimental flow that is not shared by the other models since all the models mentioned in Fig. 8 have symmetric fixed points but the related flow is *not* symmetric due to the presence of quadratic terms of the form  $X_i X_j (\forall i, j \neq 1)$ .

Although improvement in some sense can be observed, the identified model still reveals difficulties in reproducing the original dynamics. Another two important differences are: the first-return map does not present the layered (ghostly) structure as before. This is a signature of perfect symmetry as pointed out in [33]. Secondly, the maximum of the first-return map shown in Fig. 9 does not coincide with the  $x = y$  line as observed in other models (Fig. 8), this has important consequences in the resulting population of periodic orbits. As a matter of fact, for the  $N_p = 20$  model the location of this maximum in relation to the  $x = y$  line is very similar to that of the experimental dynamics (Fig. 8). Moreover, the 20-term model map allows to define a clear partition of the attractor, unlike for the previous models. Nevertheless, because of the third branch which is increasing, rather than decreasing, as seen in the experimental map (Fig. 8), the topology of part of the attractor is different. Indeed, when linking numbers, i.e. the half sum of oriented crossings are computed, some differences are observed between the model and the dynamics underlying the experimental data. For instance, the pair of periodic orbits encoded by (200) and (10)

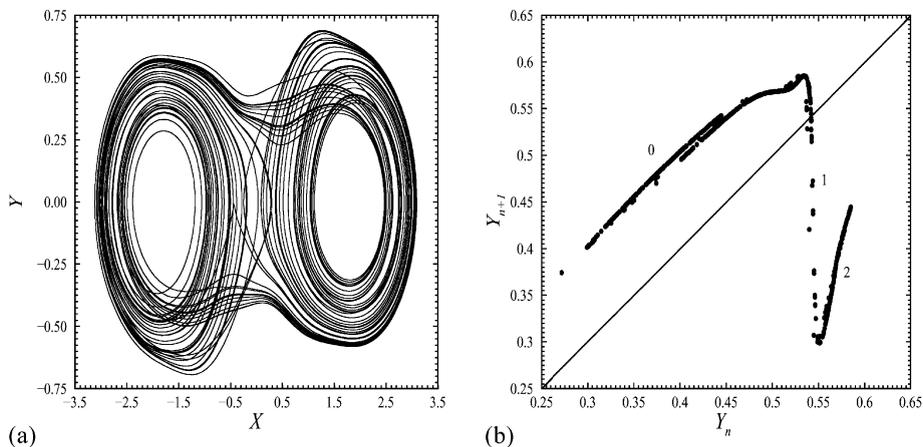


Fig. 9. Phase portrait (a) and the corresponding first-return map (b) generated by the symmetrical model. This model has 20 terms,  $N_p = 20$ .

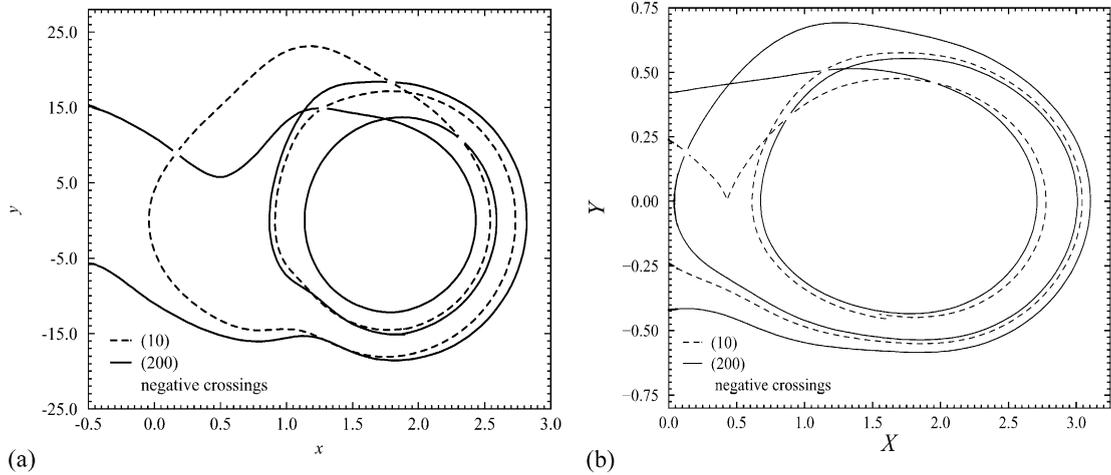


Fig. 10. (a) Experimental dynamics, and (b) model dynamics. The linking numbers between periodic orbits encoded by (200) and (10) is equal to  $-2$  for both the experimental dynamics and the 20-term model dynamics. Nevertheless, the period-2 orbit does not present the same symmetry properties. See text for a possible explanation.

are displayed for the experimental dynamics (Fig. 10a) and for the 20-term model (Fig. 10b). The linking number  $lk(200, 10)$  is equal to  $-2$  in both cases but it is noticed that the period-2 orbit of the model is a symmetric orbit, i.e. it visits the two scrolls in the same way, whereas it is asymmetric and confined in to a single scroll for the dynamics underlying the experimental data. Such an example confirms that the experimental and the 20-term model attractors are not topologically equivalent although both attractors are rather difficult to distinguish by simple visual inspection. A likely reason for such a difference is the third increasing branch in the first-return map of the model, as shown in Fig. 9b, which is not observed for the experimental dynamics. Interestingly, however, the first-return map of a Chua-type circuit with a smooth nonlinearity [28] also presents a similar increasing branch. At this stage, it seems correct to conjecture that the appearance of a third increasing branch might come as a limitation of the model representation used, since it will always yield a smooth nonlinearity rather than a piecewise-linear (see Fig. 4b) one of the actual system. Details on how such a topological analysis is carried out can be found in [33]. Note that computing linking numbers is only possible for 3D phase portraits.

#### 4. Conclusion

This paper has discussed the application of structure-selection techniques applied to a continuous-time model representation. Model structure is chosen based on an ERR criterion (12) that does *not* require any prior knowledge about the system. The number of fixed points may also be used for selecting the model structure. When the original system has symmetry properties, they can be used to constrain the model structure. The new procedure was tested using both simulated and real data from the Lorenz system and an implementation of an electronic oscillator, respectively. As a result, models with fewer terms and better dynamics were obtained. Model performance was assessed by visual inspection of the attractor geometry, largest Lyapunov exponent and, mainly, topological analysis. Whereas the first two criteria can only disqualify very bad models, topological characterization of attractors enables a much more detailed validation. In fact, the detrimental effects of (slight) over-parametrization can hardly be perceived using commonly used criteria for model validation.

Finally, it is mentioned that the identification procedure is sensitive to the number and location of centers (as for other representations) and to the quality of the estimated derivatives. Also, the application of the structure-selection procedure is sometimes sensitive to the set of candidate terms. Such problems are not unknown in the related literature and the search for appropriate solutions is currently an open field of research.

## Acknowledgements

Financial support from CNPq and CNRS is gratefully acknowledged.

## Appendix A. Simultaneous structure selection and parameter estimation

The starting point is to compose the following regressor matrix with all the candidate terms included. This should be accomplished before the first iteration. Therefore,

$$\tilde{\Psi}^{(0)} = [\Psi \quad \dot{\mathbf{X}}_{d_c}] = [\psi_1^{(0)} \dots \psi_{N_p}^{(0)} \quad \dot{\mathbf{X}}_{d_c}^{(0)}]. \quad (\text{A.1})$$

Matrix  $V_k \in \mathbb{R}^{k \times k}$  is a submatrix of  $V \in \mathbb{R}^{N_p \times N_p}$  that can be obtained by deleting rows and columns from  $k + 1$  to  $N_p$ . Matrix  $V$  is a top triangular matrix obtained by decomposition of the regressor matrix  $\Psi$  by means of an orthonormal matrix  $Q$ , i.e.

$$Q\Psi = \begin{bmatrix} V \\ 0 \end{bmatrix}. \quad (\text{A.2})$$

After performing  $k - 1$  Householder transformations,  $H^{(1)}, \dots, H^{(k-1)}$ , were performed to matrix  $\tilde{\Psi}^{(0)}$ , we have

$$\tilde{\Psi}^{(k-1)} = \begin{bmatrix} V_{k-1} & & & \\ & \psi_j^{(k-1)} & \dots & \psi_{N_p}^{(k-1)} & \dot{\mathbf{X}}_{d_c}^{(k-1)} \\ & \mathbf{0} & & & \end{bmatrix}, \quad (\text{A.3})$$

where  $\mathbf{0}$  is an  $(N - k + 1) \times (k - 1)$  matrix. A model that contains the first  $k - 1$  regressors and respective parameters will have the following residual sum of squares:

$$J_{MQ}^{*(k-1)} = \sum_{i=k}^{N_c} (\dot{X}_{d_c,i}^{(k-1)})^2, \quad (\text{A.4})$$

where  $\dot{X}_{d_c,i}^{(k-1)}$  denotes the variable  $\dot{X}_{d_c}$  at the  $i$ th center after  $(k - 1)$  transformations. Of course,  $\dot{\mathbf{X}}_{d_c}^{(k-1)} = [\dot{X}_{d_c,1}^{(k-1)} \dots \dot{X}_{d_c,N_c}^{(k-1)}]^T$ . When the  $k$ th transformation,  $H^{(k)}$ , is applied to  $\tilde{\Psi}^{(k-1)}$  one obtains  $\tilde{\Psi}^{(k)}$  and if the new regressor is then added to the model the residual sum of squares (A.4) is reduced to

$$J_{MQ}^{*(k)} = \sum_{i=k}^{N_c} (\dot{X}_{d_c,i}^{(k-1)})^2 - (\dot{X}_{d_c,j}^{(k)})^2 = \sum_{i=k+1}^{N_c} (\dot{X}_{d_c,i}^{(k)})^2.$$

Hence the objective is to choose among the columns of  $\psi_j^{(k-1)}, \dots, \psi_{N_p}^{(k-1)}$  that one for which  $(\dot{X}_{d_c,j}^{(k)})^2$  is largest. In other words, in the  $k$ th iteration the  $k$ th regressor should be chosen such that, when included in the model, it will

result in the greatest reduction of the residual sum of squares. In order to do so, the following variables should be determined:

$$a_j^{(k)} = \sum_{i=k}^{N_c} (\tilde{\psi}_{ij}^{(k-1)})^2, \quad j = k, \dots, N_p, \quad (\text{A.5})$$

$$b_j^{(k)} = \sum_{i=k}^{N_c} \tilde{\psi}_{ij}^{(k-1)} \dot{X}_{d_e,i}^{(k-1)}, \quad j = k, \dots, N_p, \quad (\text{A.6})$$

where  $\tilde{\psi}_{ij}^{(k-1)}$  is the  $(ij)$ th element of matrix  $\tilde{\Psi}^{(k-1)}$ . The benefit of including the  $j$ th regressor into the model can be quantified using the ERR of such a regressor  $(\dot{X}_{d_e,j}^{(k)})^2 / \langle \dot{X}_{d_e}, \dot{X}_{d_e} \rangle$  or expressing it in terms of  $a_j^{(k)}$  and  $b_j^{(k)}$  (see (12))

$$[\text{ERR}]_j^{(k)} = \frac{(\hat{g}_j^{(k)})^2 a_j^{(k)}}{\langle \dot{X}_{d_e}, \dot{X}_{d_e} \rangle} = \left( \frac{\langle \mathbf{w}_j, \dot{X}_{d_e} \rangle}{\langle \mathbf{w}_j, \mathbf{w}_j \rangle} \right)^2 \frac{a_j^{(k)}}{\langle \dot{X}_{d_e}, \dot{X}_{d_e} \rangle} = \left( \frac{b_j^{(k)}}{a_j^{(k)}} \right)^2 \frac{a_j^{(k)}}{\langle \dot{X}_{d_e}, \dot{X}_{d_e} \rangle} = \frac{(b_j^{(k)})^2}{a_j^{(k)} \langle \dot{X}_{d_e}, \dot{X}_{d_e} \rangle}. \quad (\text{A.7})$$

It is important to notice that the  $j$ th regressor ( $j = k$ ) will be included in the model at the  $k$ th iteration, after  $k - 1$  Householder transformations.  $[\text{ERR}]_j^{(k)}$  depends on  $a_j^{(k)}$  and  $b_j^{(k)}$ , which, on their turn, are determined after the  $(k - 1)$ th Householder transformation, as can be seen from (A.5) and (A.6). Finally, the structure-selection procedure at the  $k$ th iteration consists of determining  $[\text{ERR}]_j^{(k)}$  for all the candidate regressors that have not yet been included in the model, i.e.  $j = k, \dots, N_p$ . Denoting  $j_m$  the index for which the maximum occurs, the  $k$ th and  $j_m$ th columns of  $\tilde{\Phi}^{(k-1)}$  should be exchanged. After such permutation, the  $k$ th Householder transformation is performed.

The objective of using the ERR is to enable the user to arrange a set of candidate regressors in descending order of relevance. The best number of terms to compose the model, i.e. the cutting point at which the ERR will stop including terms in the model can be determined using other complimentary criteria such as Akaike's or Rissanen's information criteria [5,40].

The aforementioned algorithm uses the Householder transformation as a means to achieve orthogonalization. Other algorithms exist such as the classical and modified Gram–Schmidt and have been compared in [18].

## References

- [1] L.A. Aguirre, Recovering map static nonlinearities from chaotic data using dynamical models, *Physica D* 100 (1–2) (1997) 41–57.
- [2] L.A. Aguirre, E.M. Mendes, Nonlinear polynomial models: structure, term clusters and fixed points, *Int. J. Bifurc. Chaos* 6 (2) (1996) 279–294.
- [3] L.A. Aguirre, G.G. Rodrigues, E.M. Mendes, Nonlinear identification and cluster analysis of chaotic attractors from a real implementation of Chua's circuit, *Int. J. Bifurc. Chaos* 7 (6) (1997) 1411–1423.
- [4] L.A. Aguirre, A.V.P. Souza, An algorithm for estimating fixed points of dynamical systems from time series, *Int. J. Bifurc. Chaos* 8 (11) (1998) 2203–2213.
- [5] H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automat. Contr.* 19 (6) (1974) 716–723.
- [6] E. Baake, M. Baake, H.G. Bock, K.M. Briggs, Fitting ordinary differential equations to chaotic data, *Phys. Rev. A* 45 (8) (1992) 5524–5529.
- [7] B.P. Bezruchko, D.A. Smirnov, Constructing nonautonomous differential equations from experimental time series, *Phys. Rev. E* 63 (1) (2000) 1–7.
- [8] S.A. Billings, S. Chen, M.J. Korenberg, Identification of MIMO non-linear systems using a forward-regression orthogonal estimator, *Int. J. Contr.* 49 (6) (1989) 2157–2189.
- [9] J.L. Breen, N.P. Packard, A learning algorithm for optimal representation of experimental data, *Int. J. Bifurc. Chaos* 4 (2) (1994) 311–326.
- [10] D.S. Broomhead, D. Lowe, Multivariable functional interpolation and adaptive networks, *Complex Syst.* 2 (1988) 321–355.
- [11] R. Brown, V. In, E.R. Tracy, Parameter uncertainties in models of equivariant dynamical systems, *Physica D* 102 (3–4) (1997) 253–261.
- [12] R. Brown, N. Rulkov, N. Tuffillaro, Nonlinear prediction of chaotic time series, *Phys. Rev. E* 50 (6) (1994) 4488–4508.
- [13] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, *Physica D* 110 (1997) 43–52.

- [14] M. Casdagli, Nonlinear prediction of chaotic time series, *Physica D* 35 (1989) 335–356.
- [15] A. Çinar, Nonlinear time series models for multivariable dynamic processes, *Chemometrics Intelligent Laboratory Syst.* 30 (1995) 147–158.
- [16] P. Celka, J. Vesin, R. Vetter, R. Grueter, G. Thonet, E. Pruvot, H. Duplain, U. Scherrer, Parsimonious modeling of biomedical signals and systems: applications to the cardiovascular system, in: M. Akay (Ed.), *Nonlinear Biomedical Signal Processing. Part II*, IEEE Press, New York, 1999.
- [17] S. Chen, S.A. Billings, C.F.N. Cowan, P.M. Grant, Practical identification of NARMAX models using radial basis functions, *Int. J. Contr.* 52 (6) (1990) 1327–1350.
- [18] S. Chen, S.A. Billings, W. Luo, Orthogonal least squares methods and their application to nonlinear system identification, *Int. J. Contr.* 50 (5) (1989) 1873–1896.
- [19] J. Cremers, A. Hübler, Construction of differential equations from experimental data, *Z. Naturforsch* 42a (1987) 797–802.
- [20] G. Deco, D. Obradovic, *An Information Theoretic Approach to Neural Computing*, Springer, Berlin, 1996.
- [21] E. Floriani, T. Dudok de Wit, P. Le Gal, Nonlinear interactions in a rotating disk flow: from a Volterra model to the Ginzburg–Landau equation, *CHAOS* 10 (4) (2000) 834–847.
- [22] M. Giona, F. Lentini, V. Cimagalli, Functional reconstruction and local prediction of chaotic time series, *Phys. Rev. A* 44 (6) (1991) 3496–3502.
- [23] G. Gouesbet, C. Letellier, Global vector field reconstruction by using a multivariate polynomial  $l_2$  approximation on nets, *Phys. Rev. E* 49 (6) (1994) 4955–4972.
- [24] H.M. Henrique, E.L. Lima, D.E. Seborg, Model structure determination in neural network models, *Chem. Eng. Sci.* 55 (2000) 5457–5469.
- [25] A.D. Irving, T. Dewson, Determining mixed linear–nonlinear coupled differential equations from multivariate discrete time series sequences, *Physica D* 102 (1–2) (1997) 15–36.
- [26] J.B. Kadtko, J. Brush, J. Holzfuß, Global dynamical equations and Lyapunov exponents from noisy chaotic time series, *Int. J. Bifurc. Chaos* 3 (3) (1993) 607–616.
- [27] M.P. Kennedy, Robust OP Amp realization of Chua’s circuit, *Frequenz* 46 (3–4) (1992) 66–80.
- [28] A.I. Khibnik, D. Roose, L.O. Chua, On periodic orbits and homoclinic bifurcations in Chua’s circuit with a smooth nonlinearity, *Int. J. Bifurc. Chaos* 3 (2) (1993) 363–384.
- [29] G.P. King, I. Stewart, Phase space reconstructions for symmetric dynamical systems, *Physica D* 58 (1992) 212–228.
- [30] C.S.M. Lainscsek, F. Schürer, J. Kadtko, A general form for global dynamical data models for three-dimensional systems, *Int. J. Bifurc. Chaos* 8 (5) (1998) 899–914.
- [31] C. Letellier, R. Gilmore, Covering dynamical systems: two-fold covers, *Phys. Rev. E* 63 (2) (2001) 6206–6215.
- [32] C. Letellier, G. Gouesbet, Topological characterization of reconstructed attractors modding out symmetries, *J. Phys. II* 6 (1996) 1615–1638.
- [33] C. Letellier, G. Gouesbet, N. Rulkov, Topological analysis of chaos in equivariant electronic circuits, *Int. J. Bifurc. Chaos* 6 (12) (1996) 2531–2555.
- [34] C. Letellier, L. Le Sceller, P. Dutertre, G. Gouesbet, Z. Fei, J.L. Hudson, Topological characterization and global vector field reconstruction from an experimental electrochemical system, *J. Phys. Chem. A* 99 (1995) 7016–7027.
- [35] C. Letellier, J. Maquet, G. Gouesbet, L.A. Aguirre, On the non-equivalence of observables in phase-space reconstructions from recorded time series, *J. Phys. A* 31 (1998) 7913–7927.
- [36] C. Letellier, J. Maquet, H. Labro, L. Le Sceller, G. Gouesbet, F. Argoul, A. Arn’eo, Analyzing chaotic behaviour in a Belousov–Zhabotinskii reaction by using global vector field reconstruction, *J. Phys. Chem. A* 102 (50) (1998) 10265–10273.
- [37] E. Lorenz, Deterministic nonperiodic flow, *J. Atmos. Sci.* 20 (1963) 130–141.
- [38] E.M.A.M. Mendes, S.A. Billings, On over parametrization of nonlinear discrete systems, *Int. J. Bifurc. Chaos* 8 (3) (1998) 535–556.
- [39] P. Perona, A. Porporato, L. Ridolfi, On the trajectory method for the reconstruction of differential equations from time series, *Nonlinear Dyn.* 23 (1) (2000) 13–33.
- [40] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, Vol. 15, World Scientific, Singapore, 1989.
- [41] G. Rowlands, J.C. Sprott, Extraction of dynamical equations from chaotic data, *Physica D* 58 (1992) 251–259.
- [42] L.A. Smith, Identification and prediction of low dimensional dynamics, *Physica D* 58 (1992) 50–76.