Nonlinear Identification Using Prior Knowledge of Fixed Points: a Multiobjective Approach.

Article *in* International Journal of Bifurcation and Chaos · May 2003

DOI: 10.1142/S0218127403007187 · Source: dx.doi.org

4 authors:

Erivelton Nepomuceno
Maynooth University

Ricardo H. C. Takahashi
Federal University of Minas Gerais

Gleison Amaral
Federal University of São João del-Rei

Luis Antonio Aguirre
Federal University of Minas Gerais

International Journal of Bifurcation and Chaos, Vol. 13, No. 5 (2003) 1229–1246

NONLINEAR IDENTIFICATION USING PRIOR KNOWLEDGE OF FIXED POINTS: A MULTIOBJECTIVE APPROACH

ERIVELTON G. NEPOMUCENO*, RICARDO H. C. TAKAHASHI†, GLEISON F. V. AMARAL* and LUIS A. AGUIRRE*,‡
*Centro de Pesquisa e Desenvolvimento em Engenharia Elétrica,
†Departamento de Matemática,
Universidade Federal de Minas Gerais,
Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG, Brazil
†taka@mat.ufmg.br
‡aguirre@cpdee.ufmg.br

Received January 14, 2002; Revised March 23, 2002

This paper is devoted to the problem of model building from data produced by a nonlinear dynamical system. Unlike most published works that address the problem from a black-box perspective, in the present paper a procedure is developed that permits the use of prior knowledge about the location of fixed-points in addition to the data thus resulting in a gray-box approach. Numerical results using Chua's double-scroll attractor and the sine map are presented. As discussed, the suggested procedure is useful as a means to partially compensate for the loss of information due to noise and to improve dynamical performance in the presence of model structure mismatches. Preliminary results have indicated that the procedure outlined in this paper is a systematic way of searching for models in the vicinity of the black-box solution. This could have important consequences not only in model building but also in model validation.

Keywords: Gray-box identification; fixed points; multiobjective optimization.

1. Introduction

Since the papers by Packard and co-workers [Packard *et al.*, 1980] and that of Takens [1981], the field of nonlinear dynamics has witnessed great interest in what is called the state (or phase) space reconstruction problem [Casdagli *et al.*, 1991; Gibson *et al.*, 1992] (see also [Sauer *et al.*, 1991] for a comprehensive treatment of the subject). Briefly, the main goal is, usually starting from a single time series called the observable, to define a space spanned by variables that are functions of the observable in such a way that the reconstructed space shares dynamical and topological features of the original space. In this respect, there are a number of issues that are of utmost importance to enable a successful reconstruction in a nonideal environment. Among such issues we mention the choice of delay time [Rosenstein *et al.*, 1994] and data length [Kugiumtzis, 1996], and the investigation of the effects of basis variables [Casdagli *et al.*, 1991] and the choice of observable [Letellier *et al.*, 1998].

A different problem, although quite related, is that of functional approximation or simply modeling. Basically, this problem amounts to obtaining a mathematical model (for instance, one or more dynamical equations) directly from a set of available

‡Author for correspondence.

data (usually, but not necessarily, a single time series). If such a model is valid, it will display dynamics close to those underlying the data. A great variety of procedures have been suggested in the literature to obtain models for nonlinear dynamics from data. Examples include radial basis functions [Broomhead & Lowe, 1988], differential equations [Gouesbet & Letellier, 1994; Brown *et al.*, 1994; Lainscsek *et al.*, 1998], neural networks [Suykens & Vandewalle, 1995], difference equations [Aguirre *et al.*, 1997], mixed models [Irving & Dewson, 1997], wavelets [Billings & Coca, 1999] and Volterra models [Floriani *et al.*, 2000] to mention but a few.

A common feature to practically all the methods mentioned above is that in obtaining a model from data no prior knowledge is assumed. Moreover, in order to estimate the model parameters a single cost function is usually minimized. Techniques with such characteristics are usually referred to as black-box monovariable methods and, as can be seen from the aforementioned papers, are quite useful in a number of situations.

Assuming some information about the system is available prior to modeling, is it possible to effectively use such information during model building? How can this be done? What kind of information can be used? It is believed that such questions open new areas of research in the modeling of nonlinear dynamics and that the first answers will be specific to certain types of mathematical representations.

A few timid attempts to use prior information in modeling (usually referred to as gray-box modeling) nonlinear dynamics can be found in recent years. Brown and co-workers have modified the model structure in view of a symmetrical flow [Brown *et al.*, 1994]. Similarly, one of us has used symmetry of fixed points [Aguirre *et al.*, 1997] and the knowledge of steady-state relationships for the nonautonomous case [Aguirre *et al.*, 2000] also to constrain the space of viable model structures. In all such papers, however, the prior information is not directly used in the stage of parameter estimation in which the model parameters are estimated by the minimization of a cost function that takes into account a single objective, as for instance, the minimization of the sum of squared errors between data and model predictions.

This paper is believed to stand out from the aforementioned literature in at least two main points. Firstly, prior information (location of fixed points) will be used, in the context of nonlinear difference equations, to directly constrain parameter estimation and not only to help restrict some model structure classes, as in [Aguirre *et al.*, 2000]. Secondly, the parameter estimation stage will be accomplished by means of a multiobjetctive procedure, that is, the cost function takes into account more than one (in this paper, two) objective. The suggested procedure may be advantageous in many situations. It not only permits to use the information available about the location of fixed points but also enables the user to determine the weight with which such information should be used in the modeling procedure. The result is a continuum of models from which a particular candidate that best suits the user needs can be selected.

The remainder of the paper is organized as follows. Section 2 presents the motivation for using the multiobjective optimization approach. In Sec. 3 a brief background, containing a description of the NARMAX models and the concepts of term clusters and cluster coefficients is presented. Such concepts are used to calculate fixed points of polynomial NARMAX models. In Sec. 4 the use of prior knowledge is presented using both mono-objective and multiobjective approaches. Then, in Sec. 5 the multiobjective approach is illustrated in the parameter estimation for two chaotic systems: Chua's circuit and the sine map. In addition, a brief discussion on results is presented. Finally, Sec. 6 summarizes the main points of the paper.

2. Multiobjective Approach

Let Fig. 1 represent the set of *efficient solutions*, or the *Pareto set* of a biobjective optimization problem. This means that there does not exist any solution "below" that set, in the graphic of Fig. 1, and any method that minimizes f_1 and/or f_2, in the best case, will find solutions inside that set.

Black-box identification techniques (that rely on mono-objective unconstrained optimization algorithms) usually lead to solutions like p_1, that take into account only one objective having no control on other objectives (either explicitly defined or not). A gray-box identification procedure based on mono-objective optimization with aggregated objectives or with an objective taken as a constraint would yield a single solution that is likely to lead to a solution like p_b, since there is no reason for that procedure to find p_a.

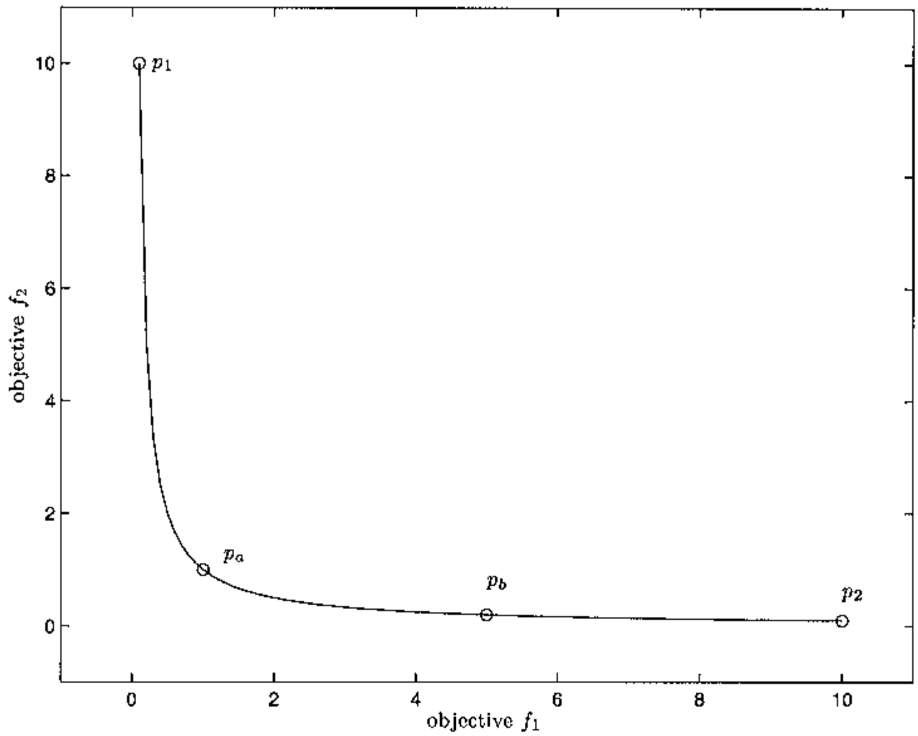

Fig. 1. The continuous curve linking point p_1 to point p_2 is a hypothetical Pareto set of a biobjective optimization problem. Solutions p_1 and p_2 denote the individual optima of the objective functions f_1 and f_2. Solutions p_a and p_b both belong to the Pareto set.

Multiobjective approaches can fall into two cases:

- If the criteria that are to be considered in the final choice of one solution are objectives f_1 and f_2 only, solution p_b is possibly not suitable, since there are better ones, that would represent a small degradation in one objective and a large enhancement in the other one. In this case, solution p_a would be possibly "the best one". This solution is near (in objective values) to both the individual optima of objectives f_1 and f_2. From another point of view, solution p_a, if modified, would lead to a rather small enhancement in one objective at a great cost in terms of the other one. The multiobjective optimization approach can lead to the choice of a solution like p_a.
- If, however, there are other criteria that cannot be expressed in terms of f_1 and f_2 but are also important in order to characterize meaningful solutions, then the whole set of *efficient solutions* should be examined in the search for the best solution. The solution, in this case, is found with some interaction with a "user". This is the case here, when qualitative properties of the model are important.

In both situations, when some kind of tradeoff analysis occurs, the multiobjective framework will probably be more helpful than mono-objective techniques.

3. Identification Algorithm

The mathematical representation used is the Nonlinear AutoRegressive Moving Average with eXogenous input (NARMAX) [Leontaritis & Billings, 1985a]. It is believed that the procedure suggested in this paper can be used in other representations as a systematic way of parameter estimation. Consider the NARMAX model described by the following equation:

$$y(k) = F\begin{bmatrix} y(k-1), \ldots, y(k-n_y) \\ u(k-d), \ldots, u(k-d-n_u+1) \\ e(k), \ldots, e(k-n_e) \end{bmatrix} \quad (1)$$

where n_y, n_u and n_e are the maximum lags considered for the process, input and noise terms, respectively, and d is the delay measured in sampling intervals, T_s. Moreover, $y(k)$ is a time series of the output while $u(k)$ is a time series of the input. $e(k)$

accounts for uncertainties, noise, unmodeled dynamics and the like. $F^\ell[\cdot]$ is some nonlinear function of $y(k)$, $u(k)$ and $e(k)$. In this paper $F^\ell[\cdot]$ is taken to be a nonlinear polynomial of degree $\ell \in \mathbb{Z}^+$. In order to estimate the parameters of such a polynomial, (1) is expressed as follows:

$$y(k) = \psi^{\mathrm{T}}(k-1)\hat{\boldsymbol{\theta}} + \xi(k) \tag{2}$$

where $\psi(k-1)$ is the vector of regressors (independent variables) that contains linear and nonlinear combinations of output, input and noise terms up to and including time $k-1$. The parameters corresponding to each term in such matrices are the elements of the vector $\hat{\boldsymbol{\theta}}$. Finally, $\xi(k)$ is the *residual* or *prediction errors* at time k which are defined as the difference between the measured data $y(k)$ and the one-step-ahead prediction $\psi^{\mathrm{T}}(k-1)\hat{\boldsymbol{\theta}}$.

A dynamical model as in (2) taken over a set of data, furnishes constraints which can be presented by a matrix equation as follows:

$$\mathbf{y} = \Psi\hat{\boldsymbol{\theta}} + \boldsymbol{\xi}\,. \tag{3}$$

The parameter vector $\hat{\boldsymbol{\theta}}$ that minimizes the inner product of the residual vector can be estimated by orthogonal least-squares techniques that minimize the cost function

$$J_{\mathrm{LS}}(\hat{\boldsymbol{\theta}}) = (\mathbf{y} - \Psi\hat{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{y} - \Psi\hat{\boldsymbol{\theta}})\,. \tag{4}$$

One of several advantages of such algorithms is that the error reduction ratio (ERR) can be easily obtained as a byproduct [Billings *et al.*, 1989]. This criterion provides an indication of which terms to include in the model by ordering all the candidate terms according to a hierarchy that depends on the relative importance of each term. After the terms have been ordered by the ERR, information criteria can be used to help decide a good cut-off point.

3.1. *Term clusters and cluster coefficients*

The deterministic part of a polynomial NARMAX model can be expanded as the summation of terms with degrees of nonlinearity in the range $1 \le m \le \ell$. Each mth order term is multiplied by a coefficient $c_{p,m-p}(n_1, \ldots, n_m)$ as follows

$$y(k) = \sum_{m=0}^{\ell}\sum_{p=0}^{m}\sum_{n_1,n_m}^{n_y,n_u} c_{p,m-p}(n_1, \ldots, n_m) \times \prod_{i=1}^{p} y(k-n_i) \prod_{i=p+1}^{m} u(k-n_i)\,, \tag{5}$$

where

$$\sum_{n_1,n_m}^{n_y,n_u} \equiv \sum_{n_1=1}^{n_y} \cdots \sum_{n_m=1}^{n_u}\,, \tag{6}$$

and the upper limit is n_y if the summation refers to factors in $y(k-n_i)$ or n_u for factors in $u(k-n_i)$.

Considering an asymptotically stable model in steady-state excited by a constant input, Eq. (5) can be written as

$$y(k) = \sum_{n_1,n_m}^{n_y,n_u} c_{p,m-p}(n_1, \ldots, n_m) \times \sum_{m=0}^{\ell} y(k-1)^p u(k-1)^{m-p}\,, \tag{7}$$

and the following definition can be presented.

Definition 3.1. $\sum_{n_1,n_m}^{n_y,n_u} c_{p,m-p}(n_1, \ldots, n_m)$ in Eq. (7) are the coefficients of the *term clusters* $\Omega_{y^p u^{m-p}}$, which contain terms of the form $y(k-i)^p u(k-j)^{m-p}$ for $m = 0, \ldots, \ell$ and $p = 0, \ldots, m$. Such coefficients are called *cluster coefficients* and are represented as $\sum_{y^p u^{m-p}}$.

A term cluster is a set of terms of the same type and the respective cluster coefficient is obtained by the summation of the coefficients for all the terms of the respective cluster which explain the type of nonlinearity.

All the possible clusters of an autonomous polynomial with degree of nonlinearity ℓ are Ω_0 = constant, $\Omega_y = \Omega_{y^2}, \ldots, \Omega_{y^\ell}$. Thus, the fixed point of a map with degree of nonlinearity ℓ, see Eqs. (5) and (6), are given by the roots of the following clustered polynomial [Aguirre & Mendes, 1996]

$$y(k) = c_{0,0} + y(k)\sum_{n_1=1}^{n_y} c_{1,0}(n_1) + y(k)^2 \sum_{n_1,n_2}^{n_y n_y} c_{2,0}(n_1, n_2) + y(k)^\ell \sum_{n_1,n_\ell}^{n_y n_y} c_{l,0}(n_1, \ldots, n_2)\,. \tag{8}$$

Finally, using the definition of cluster coefficients and dropping the argument k, the last equation can be written as follows

$$\Sigma_{y^\ell} y^\ell + \cdots + \Sigma_{y^2} y^2 + (\Sigma_y - 1)y + \Sigma_0 = 0\,, \tag{9}$$

where $\Sigma_0 = c_{0,0}$. From the last equation it becomes clear that an autonomous polynomial with degree of nonlinearity ℓ will have up to ℓ fixed-points if $\Sigma_{y^\ell} y^\ell \neq 0$.

In many practical situations $\Sigma_0 = c_{0,0} = 0$ and in this case the previous equation can be rewritten as

$$[\Sigma_{y^\ell} y^{\ell-1} + \cdots + \Sigma_{y^2} y + (\Sigma_y - 1)]y = 0\,. \qquad (10)$$

From Eq. (10) it becomes evident that the respective dynamical model has one trivial and $\ell - 1$ nontrivial fixed points.

4. Incorporating Prior Knowledge

In this section formulae will be developed that will enable taking into account prior information about fixed-point location. To this end, consider the vector-valued function of optimization objectives defined by:

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} J_1(\boldsymbol{\theta}) \\ \vdots \\ J_n(\boldsymbol{\theta}) \end{bmatrix}\,. \qquad (11)$$

If, say, function $J_1(\theta)$ stands for the square error function (4) and $J_2(\theta)$ stands for the sum of the squares of the fixed-points location error, the multiobjective problem becomes:

$$\begin{aligned} &\min\ \mathbf{J}(\boldsymbol{\theta}) \\ &\text{subject to } \{\boldsymbol{\theta} \in \mathbf{D} \end{aligned} \qquad (12)$$

in which $\mathbf{D}$ stands for the feasible set.

Generally, the minimization of functionals J_i have competing solutions. This is reasonable since, in most situations, the best solution will not necessarily coincide with any of the individual optimal solutions, being rather a compromise solution among the several performance criteria. The suitable conceptual framework for characterizing the solutions for this trade-off is furnished by the *multiobjective programming theory* [Chankong & Haimes, 1983]. The problem is to minimize, in some sense, the vector-valued objective $\mathbf{J}(\boldsymbol{\theta})$. The central concept in multiobjective programming is that of *efficient* or *Pareto optimal* solutions. The set Θ^* of efficient solutions $\boldsymbol{\theta}^*$ may be characterized by

$$\boldsymbol{\theta}^* \in \Theta^* \Leftrightarrow \{\nexists\ \boldsymbol{\theta} \in \mathbf{D} : \mathbf{J}(\boldsymbol{\theta}) \leq \mathbf{J}(\boldsymbol{\theta}^*) \text{ and } \mathbf{J}(\boldsymbol{\theta}) \neq \mathbf{J}(\boldsymbol{\theta}^*)\}\,. \qquad (13)$$

An efficient solution must belong to the feasible solution set. In Eq. (13), the less than or equal to and the different relations between vectors are defined as:

$$\begin{aligned} &\mathbf{V} \leq \mathbf{Z} \Leftrightarrow v_i \leq z_i,\ \forall\, i \in 1, \ldots, n \\ &\mathbf{V} \neq \mathbf{Z} \Leftrightarrow \exists\ i \in 1, \ldots, n | v_i \neq z_i \end{aligned} \qquad (14)$$

in which v_i and z_i are the components of the n-dimensional vectors $\mathbf{V}$ and $\mathbf{Z}$. A solution is *efficient* if and only if there does not exist any other solution that further minimizes any of the objective vector components without increasing at least one other component of the same vector.

In order to define computational procedures for determining the efficient solution set, a standard procedure is to characterize the efficient solution set in terms of solutions of appropriate scalar optimization problems. A common way for obtaining a scalar problem from a vector problem is *the weighting problem* P_w [Chankong & Haimes, 1983]. Let

$$W = \left\{ \mathbf{w} | \mathbf{w} \in \mathbb{R}^n,\ w_j \geq 0 \text{ and } \sum_{j=1}^{n} w_j = 1 \right\} \qquad (15)$$

be the set of non-negative weights. The weighting problem is defined for some $\mathbf{w} \in W$ as $P(\mathbf{w})$:

$$\min_{\boldsymbol{\theta} \in D} \sum_{i=1}^{n} w_i J_i(\boldsymbol{\theta})\,, \qquad (16)$$

where J_i are all objectives and constraints.

This method is very simple, but it leads to a correct solution only in the case of convex problems (all objective functions being convex functionals and the feasible set being a convex set) [Chankong & Haimes, 1983]. As the location of fixed-points can be stated as a quadratic function of the identification parameters [Aguirre & Mendes, 1996], it defines a convex functional, and the multiobjective problem becomes convex (note that the time series squared prediction error is also convex). Therefore, the weighting method can be applied in the case under analysis.

4.1. *Sum of squared error* versus *location of fixed points*

The use of (4) in parameter estimation problems is standard and is one of the most suitable approaches. Nonetheless, in some practical situations the use of prior knowledge in addition to (4) can be especially helpful in increasing the quality of the model [Johansen, 1996; Aguirre *et al.*, 2000].

The location of fixed-points can be obtained from the equations which describe the system, but it is also possible to estimate the fixed-points from some measured data [Aguirre & Souza, 1998]. Therefore, the use of the location of fixed-points as a prior knowledge seems to be suitable, as this knowledge is important and it is possible to estimate it *a priori*. Moreover, in NAR(MA)X models the fixed-points depend on the parameters, as described in Eq. (10). Thus, the parameter estimation problem with two objectives that should be achieved simultaneously, namely minimum sum of squared prediction errors and minimum error in fixed-point location, can be formulated. In particular, it is desirable to minimize the error between the location of fixed-points known *a priori* and the model fixed-points, see Eq. (9). Given ℓ fixed-points $[\alpha_1, \ldots, \alpha_2, \alpha_\ell]$, it is possible to calculate a set of cluster coefficients that define such fixed-points. That is, if a model has the following clustered form

$$\prod_{i=1}^{\ell}(\overline{y}-\alpha_i) = \Sigma_{y^\ell}\overline{y}^\ell + \cdots + \Sigma_{y^2}\overline{y}^2 + (\Sigma_y - 1)\overline{y} + \Sigma_0 = 0\,, \quad (17)$$

then such a model has fixed-points at $y = [\alpha_1, \ldots, \alpha_2, \alpha_\ell]$. Define the set Φ of these model coefficients:

$$\Phi = [\Sigma_{y^\ell}, \ldots, \Sigma_{y^2}, (\Sigma_y - 1), \Sigma_0]\,. \quad (18)$$

It is worth mentioning that (17) multiplied by any constant will furnish the same fixed points. Thus, in order to avoid a large variance of parameters, the following procedure can be carried out. The parameter set obtained by standard least-squares is used as reference; by means of Definition 3.1 a set of cluster coefficients are determined:

$$\Phi_{\mathrm{LS}} = [\Sigma_{\mathrm{LS}y^\ell}, \ldots, \Sigma_{\mathrm{LS}y^2}, (\Sigma_{\mathrm{LS}y} - 1), \Sigma_{\mathrm{LS}0}]\,. \quad (19)$$

Define the *Euclidean norm*, $\|\cdot\|$, of a parameter set as the square root of the sum of the squares of parameters. The Euclidean norm of the parameter set of any model associated to given fixed-points can be normalized, in order to become equal to the Euclidean norm of the set Φ_{LS}. Doing so, the normalized set Σ becomes:

$$\Sigma = \frac{\|\Phi_{\mathrm{LS}}\|}{\|\Phi\|}[\Sigma_{y^\ell}, \ldots, \Sigma_{y^2}, (\Sigma_y - 1), \Sigma_0]\,. \quad (20)$$

To minimize the error of location of fixed-points, the following function should be minimized

$$\begin{aligned} J_{\mathrm{FP}}(\hat{\boldsymbol{\theta}}) &= (\Sigma - \hat{\Sigma})^{\mathrm{T}}(\Sigma - \hat{\Sigma}) \\ &= (\Sigma - S\hat{\boldsymbol{\theta}})^{\mathrm{T}}(\Sigma - S\hat{\boldsymbol{\theta}}) \end{aligned} \quad (21)$$

where S is a linear map such that $\hat{\Sigma} = S\hat{\boldsymbol{\theta}}$.

It is easy to verify that this problem is convex. Thus *the weighting problem* P_w is a suitable multi-objective formulation. Using Eqs. (15) and (16), the parameter estimation problem with the use of prior knowledge (location of fixed points) is as follows

$$\min_{\hat{\boldsymbol{\theta}} \in D} J_{\mathrm{MO}}(\hat{\boldsymbol{\theta}}) = w_1 J_{\mathrm{LS}}(\hat{\boldsymbol{\theta}}) + w_2 J_{\mathrm{FP}}(\hat{\boldsymbol{\theta}})\,. \quad (22)$$

Considering $w = w_1 = 1 - w_2$ and replacing respectively J_{LS} and J_{FP} by Eqs. (4) and (21) yields

$$\begin{aligned} \min_{\hat{\boldsymbol{\theta}} \in D} J_{\mathrm{MO}}(\hat{\boldsymbol{\theta}}) &= w(\mathbf{y} - \Psi\hat{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{y} - \Psi\hat{\boldsymbol{\theta}}) \\ &\quad + (1-w)(\Sigma - S\hat{\boldsymbol{\theta}})^{\mathrm{T}}(\Sigma - S\hat{\boldsymbol{\theta}})\,. \end{aligned} \quad (23)$$

In order to minimize the cost function J_{MO} regarding to $\hat{\boldsymbol{\theta}}$, it is necessary to solve $(\partial J_{\mathrm{MO}}/\partial\hat{\boldsymbol{\theta}}) = 0$, that is

$$\begin{aligned} \frac{\partial J_{\mathrm{MO}}}{\partial\hat{\boldsymbol{\theta}}} &= w(\Psi^{\mathrm{T}}\Psi\hat{\boldsymbol{\theta}} - \Psi^{\mathrm{T}}\mathbf{y}) \\ &\quad + (1-w)(S^{\mathrm{T}}S\hat{\boldsymbol{\theta}} - S^{\mathrm{T}}\Sigma) = 0 \end{aligned} \quad (24)$$

and

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= [w\Psi^{\mathrm{T}}\Psi + (1-w)S^{\mathrm{T}}S]^{-1}[(w\Psi^{\mathrm{T}}\mathbf{y} \\ &\quad + (1-w)S^{\mathrm{T}}\Sigma]\,. \end{aligned} \quad (25)$$

It is a simple matter to show that

$$\frac{\partial^2 J}{\partial\hat{\boldsymbol{\theta}}^2} = w\Psi^{\mathrm{T}}\Psi + (1-w)S^{\mathrm{T}}S > 0\,. \quad (26)$$

Therefore, $\hat{\boldsymbol{\theta}}$ in Eq. (24) is the minimum of the cost function J. Furthermore, varying the weight w from 0 to 1 produces the *Pareto optimal* solutions. Finally, it is clear to see that when $w = 1$, Eq. (24) gives the solution of minimum sum of squared error and when $w = 0$, conversely, Eq. (24) yields the solution that takes into account exclusively the location of fixed-points.

5. Numerical Results

This section illustrates the application of the new methodology to two bench systems: a NAR model of Chua's Circuit [Chua & Hasler, 1993] and the sine-map with cubic-type nonlinearities.

5.1. *Chua's circuit*

This example uses a NAR model obtained from a real implementation of Chua's circuit [Aguirre *et al.*, 1997]

$$
\begin{aligned}
y(k) = &+3.523y(k-1) - 4.2897y(k-2)\\
&-0.2588y(k-4) - 1.7784y(k-1)^3\\
&+2.0652y(k-3) + 6.1761y(k-1)^2y(k-2)\\
&+0.1623y(k-1)y(k-2)y(k-4)\\
&-2.7381y(k-1)^2y(k-3)\\
&-5.5369y(k-1)y(k-2)^2 + 0.1031y(k-2)^3\\
&+0.4623y(k-4)^3 - 0.5247y(k-2)^2y(k-4)\\
&-1.8965y(k-1)y(k-3)^2\\
&+5.4255y(k-1)y(k-2)y(k-3)\\
&+0.7258y(k-2)y(k-4)^2\\
&-1.7684y(k-3)y(k-4)^2\\
&+1.18y(k-3)^2y(k-4) + \xi(k)
\end{aligned}
\tag{27}
$$

The system is autonomous and therefore has no eXogenous inputs, and the Moving Average part has been estimated to reduce bias but is not used for simulation. This example will use data generated by the model above in order to avoid structure mismatch problems. This constitutes the best possible situation for the traditional mono-objective (black-box) identification techniques. The multiobjective technique proposed in Sec. 4 will be employed in this case, for different additive white noise levels. In this way, the role of the new multiobjective technique can be assessed.

The model fixed points are approximately at $\overline{y} = 0, \pm 2.2417$ and the map has an estimated largest Lyapunov exponent (LLE) equal to $\lambda = 3.625 \pm 0.075$ bits/s. Figure 2 shows the model (double-scroll) attractor and Fig. 3 shows the first return map, both Figs. 2 and 3 were obtained by simulating model (27).

This example will be studied following two paths. First, using data obtained from the model (27) the system will be identified using the same structure. This is an unrealistically ideal case, where structure is known and the data is not corrupted by noise. Second, zero mean white noise with variance 0.0001 is added to the data yielding a signal noise ratio (SNR) of approximately 93.53 dB, compared to 72.30 dB of real data used in [Aguirre *et al.*, 1997]. In spite of the fact that in the present case the structure is assumed known, it will be seen that the conventional approach will not be able to accurately identify the system dynamics.

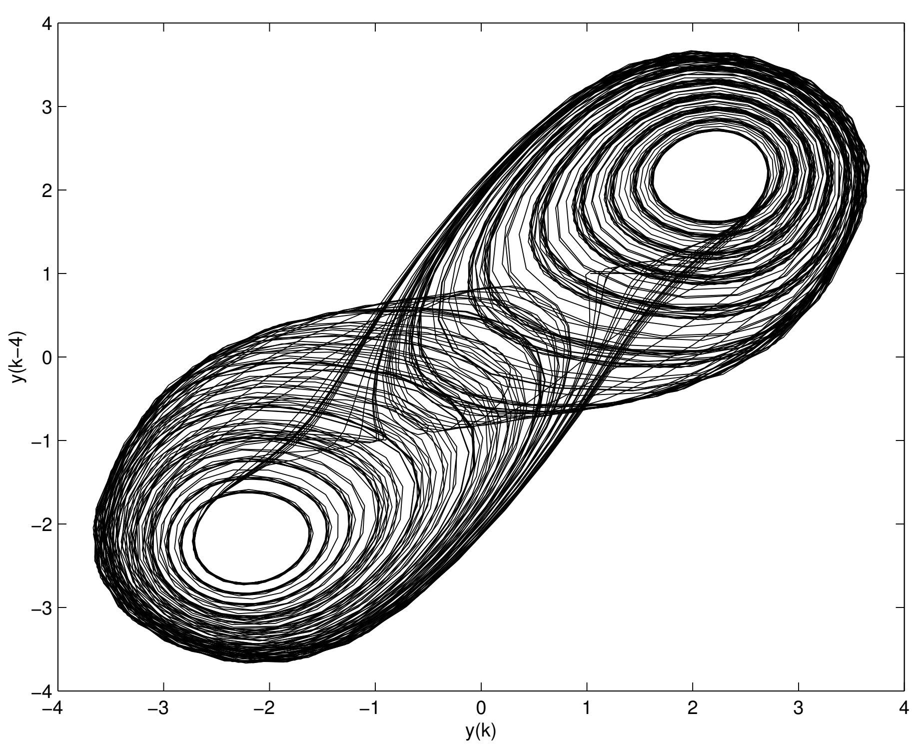

Fig. 2. Double-scroll attractor reconstructed from model (27).

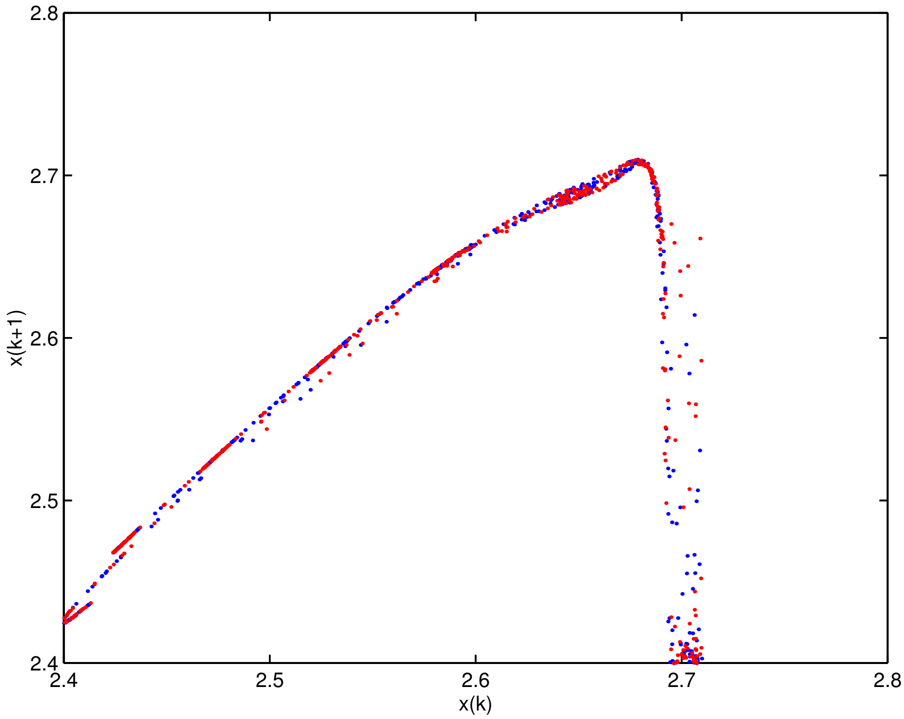

Fig. 3. First map return of model (27).

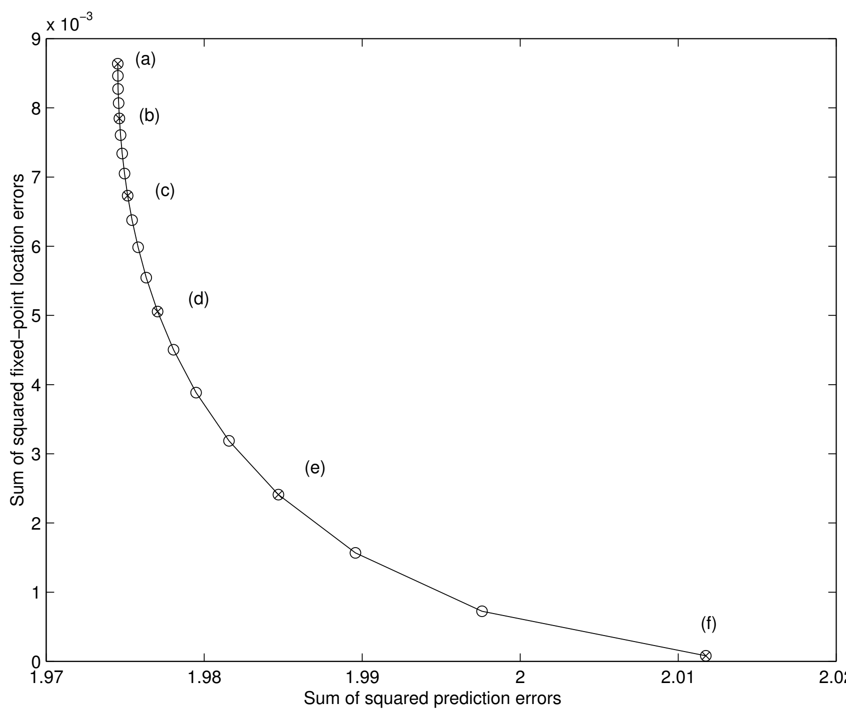

Fig. 4. *Pareto optimal* solutions. Some dynamic properties of models assigned by letters (a)–(f) are shown in Table 1.

5.1.1. *First case*

The knowledge of fixed points was included in the parameter estimation by means of (25). Twenty models have been identified. The *Pareto optimal* solutions are shown in Fig. 4, varying w from 0 to 1.

For the sake of simplicity six models (a)–(f), as shown in Fig. 4, will be analyzed further.

Table 1. Some static and dynamic properties of six models for the double-scroll model (27).

Model	Weight	SSPE	SSEFP	Fixed Point	LLE
(a)	0.001	1.9745	0.0086	0; $\pm$ 2.2466	2.8263 $\pm$ 0.0459
(b)	0.2111	1.9746	0.0078	0; $\pm$ 2.2473	2.6291 $\pm$ 0.1567
(c)	0.4212	1.9800	0.0037	0; $\pm$ 2.2606	2.0530 $\pm$ 0.0042
(d)	0.6313	1.9918	0.0013	0; $\pm$ 2.2141	–
(e)	0.8414	2.0016	0.0005	0; $\pm$ 2.2292	–
(f)	0.9990	2.0197	2.4251×10^{-7}	0; $\pm$ 2.2347	–

Table 1 shows the following parameters of the models (a)–(f): (i) Weight used in the P_w problem; (ii) sum of squared prediction error (SSPE); (iii) sum of squared error of fixed-points (SSEFP); (iv) the location of fixed-points; (v) and the LLE.

As expected, when the weight goes to 1 the sum of errors between the model fixed-points and the specified values goes to zero. Conversely, as the weight goes to zero, the solution gradually approaches the one obtained using the mono-objective

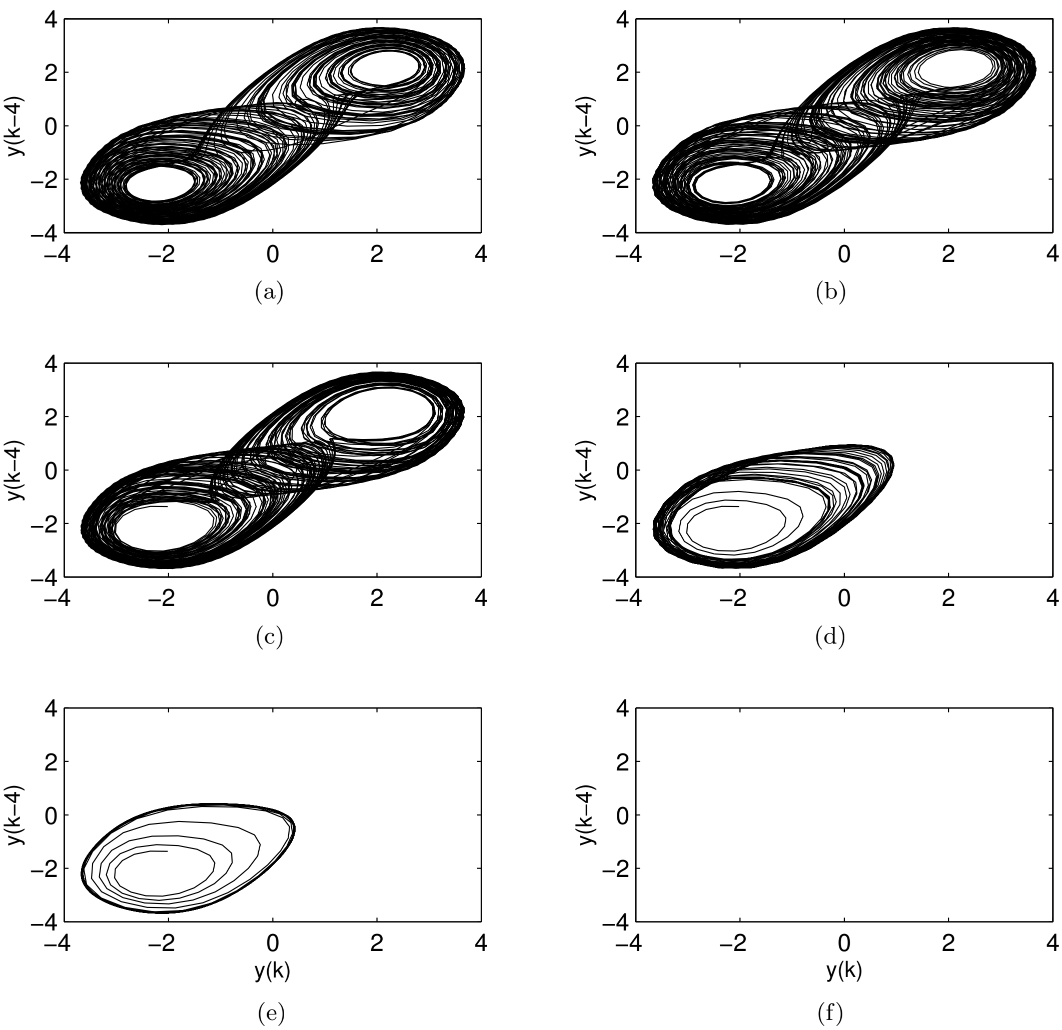

Fig. 5. Projections of double-scroll attractors corresponding to models (a)–(f), see Table 1 and Fig. 4. Noise-free case.

technique. Figure 5 shows the attractor bidimensional projections for models (a)–(f). It is interesting to notice that as the fixed-point is forced to its original value, the model dynamics deteriorate and eventually becomes unstable (see model (f)). The LLE algorithm did not converge for models (d) and (e) although such models were still stable.

The first return map is presented in Fig. 6. The shape of the map changes considerably from model (a) to model (f). Considering the whole set of model features (see Table 1), it seems that model (a) is the best one. This result was somewhat expected because, due to ideal case being simulated, there is a NAR model that simultaneously attains the minimum prediction error and the minimum fixed-point location error: the original model (27). In this case, the black-box identification indeed leads to a very good result. The variation of fixed-point error in Fig. 4 is almost negligible.

5.1.2. *Second case*

In the second case, the time series data was artificially corrupted with additive noise. In this case, one should expect that the minimum of the prediction error will not be close to the minimum of the fixed point error, as was for the noise free case. In other words, there should be greater difference between the models estimated by the minimization of the sum of squared prediction errors and by the minimization of sum of squared fixed-point errors. Therefore, a multiobjective solution would be highly desirable in this case.

Figure 7 shows the Pareto optimal solutions. The range of values spanned by the sum of squared fixed-point errors is much wider and the errors are ten times larger than in the noise-free case. Table 2 shows some static and dynamic properties of six models for this case. Figure 8 shows the projections

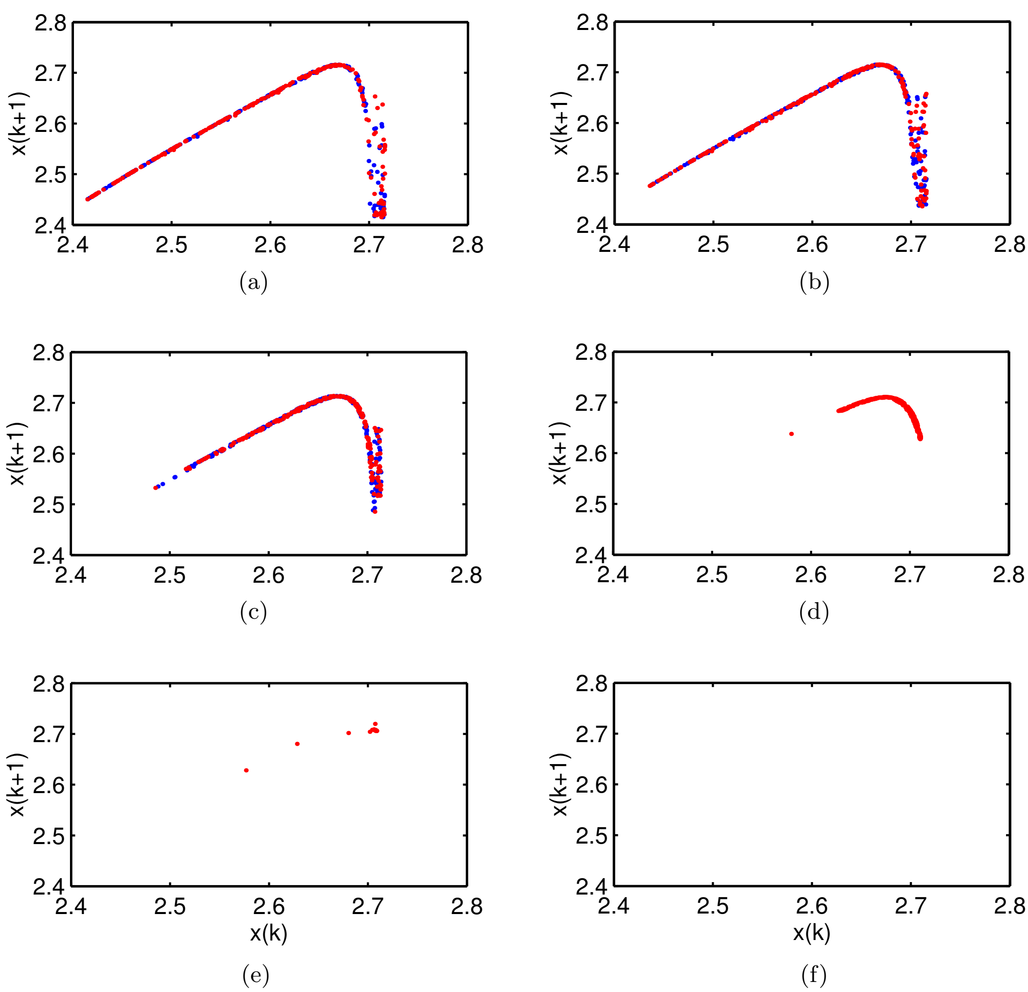

Fig. 6. First return maps of models (a)–(f), see Table 1 and Fig. 4. Noise-free case.

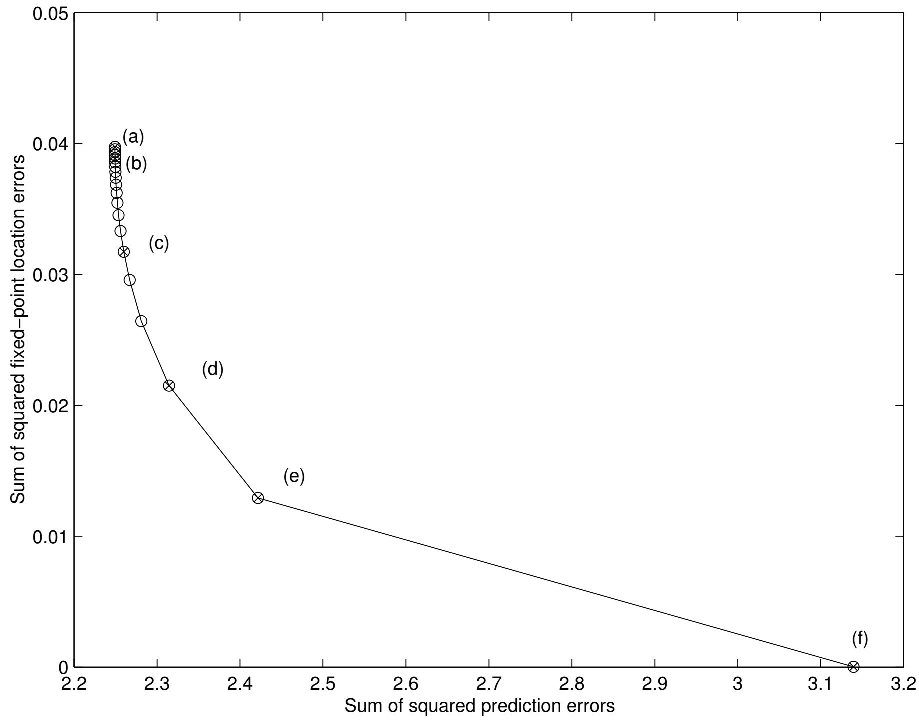

Fig. 7. *Pareto optimal* solutions. Some dynamic properties of models assigned by letters (a)–(f) are shown in Table 2.

of the double-scroll attractors for the models (a)–(f) and Fig. 9 shows the first return map for the mentioned models. In contrast to the noise-free case, the best model seems to be model (c) which is a multiobjective solution, that is, in estimating the parameters of this model two different objectives were simultaneously taken into account. This model keeps all dynamic properties (see Figs. 8 and 9) and has a sum of squared fixed-point error 25% better than for the model obtained with the black-box technique. In other words, with a negligible increase in the sum of squared prediction error, the multiobjective technique suggested in Sec. 4 enables the user to systematically obtain models with better performance in terms of fixed-point location. It should be realized that such models, in addition to having fixed-points that are better located, often have overall dynamics that are closer to the original system, as for instance model (c).

5.2. *The sine-map with cubic-type nonlinearities*

Consider the following map:

$$y(k) = \alpha \sin(y(k-1)) , \tag{28}$$

with $\alpha = 1.2\pi$. For initial condition $y(0) \in [-\pi, \pi]$. Equation (28) maps the interval $[-\pi, \pi]$ onto itself. The fixed points are approximately $\overline{y} = 0, \pm 2.4383$

Table 2. Static and dynamic properties of six models of the double-scroll model (27) for the noisy case.

Model	Weight	SSPE	SSEFP	Fixed Point	LLE
(a)	0.001	2.2494	0.0397	0; ± 2.2333	2.2122 ± 0.0330
(b)	0.2111	2.2495	0.0389	0; ± 2.2331	2.1475 ± 0.0512
(c)	0.4212	2.2599	0.03174	0; ± 2.2308	1.6107 ± 0.0051
(d)	0.6313	2.3145	0.02151	0; ± 2.2230	–
(e)	0.8414	2.4216	0.0129	0; ± 2.2279	–
(f)	0.9990	3.1395	2.0952×10^{-5}	0; ± 2.2520	–

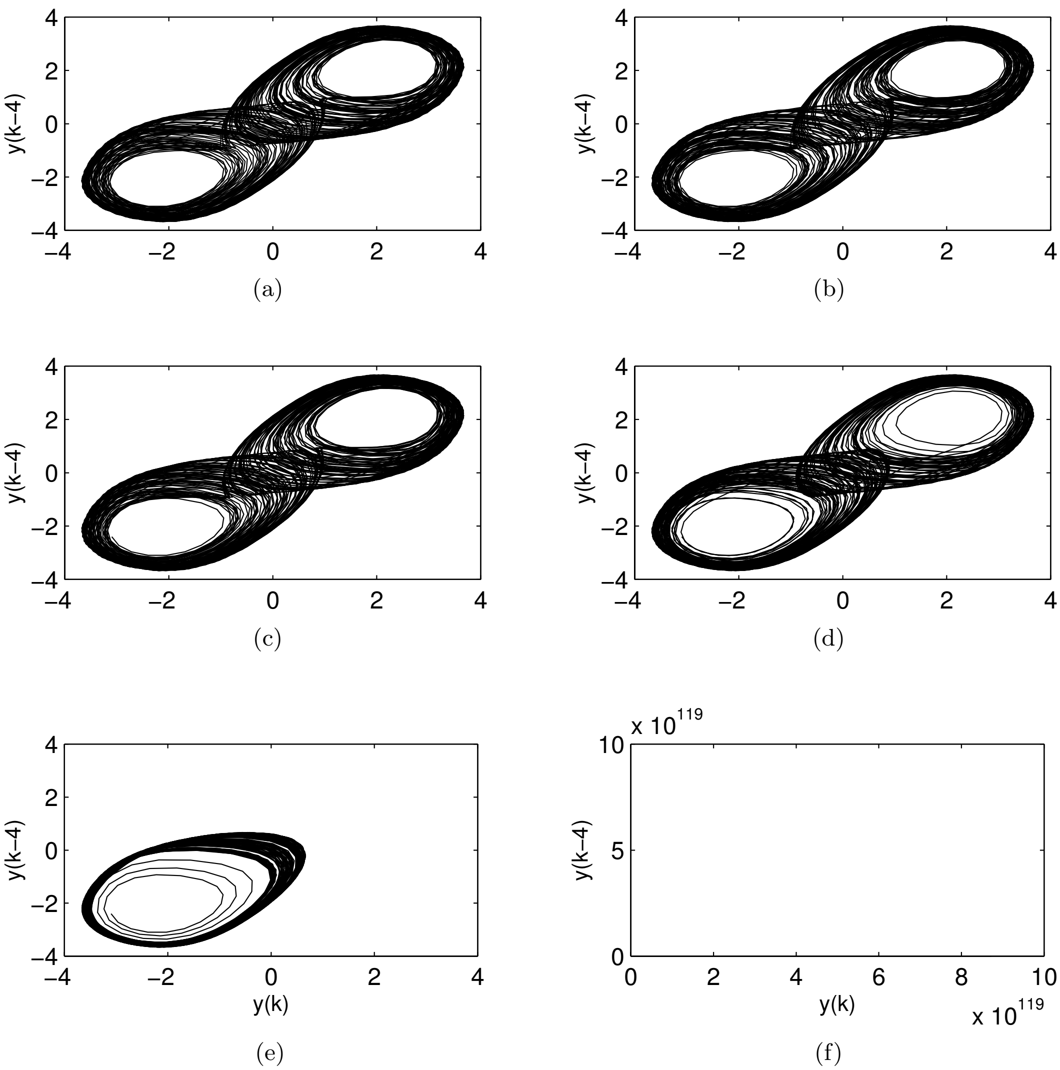

Fig. 8. Projections of double-scroll attractors corresponding to models (a)–(f), see Table 2 and Fig. 7. Noisy case.

and the map has an estimated LLE equal to $\lambda = 1.155 \pm 0.009$ bits/s. Figure 10 shows the first return map which is also a view of the attractor.

The following first-order model, using mono-objective techniques, was estimated from 1000 data points taken from the attractor shown in Fig. 10

$$y(k) = 2.6868y(k-1) - 0.2462y(k-1)^3\,, \quad (29)$$

with fixed-points at $\overline{y} = 0$, ± 2.6176. The sum of squared prediction errors is equal to 0.2319, and LLE is given by $\lambda = 1.1458 \pm 3.4545 \times 10^{-5}$ bits/s.

Incorporating the knowledge of fixed points by means of (25) during parameter estimation, twenty models have been identified. The *Pareto optimal* solutions are shown in Fig. 11. It is worth mentioning that it was necessary to vary the parameter w very close to 1 in order to produce a significant set of solutions. The same indices as before were used and are shown in Table 3.

Figure 12 shows the first return maps for the models (a)–(f). Observe that as the fixed-points tend to those of the system (indicated by the crossing of the $y = x$ line and the first return map of the sine map) the first return map of the models come closer to that of the original system, especially for values of $y(k)$ greater than 2 and less than -2. In this case, the benefit of using the information of fixed-points is quite obvious.

Model (b) displays interesting behavior. Although such a model has fixed-points that are not too far away from those of the original system, the model attractor reveals periodic dynamics rather than chaos. In order to further investigate this, Fig. 13 shows the histogram of the six models.

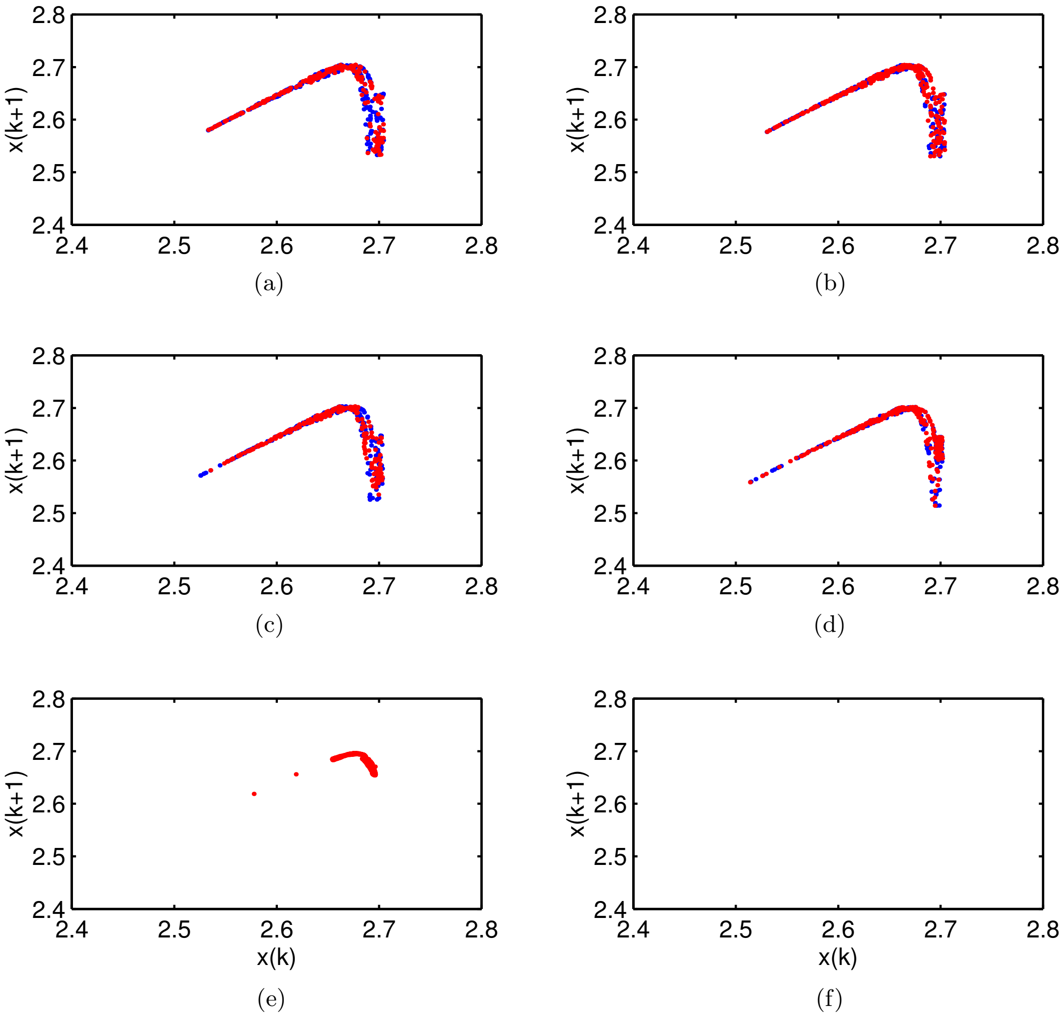

Fig. 9. First return maps of models (a)–(f), see Table 2 and Fig. 7. Noisy case.

Model (b) does not present a continuous distribution and, as expected, has a negative LLE. Thus it is possible to conjecture that a bifurcation has occurred along the Pareto as the weight, w is varied. In fact, a close view at the neighborhood of model (b) with the fixed-point error as the bifurcation parameter, shows the suspected bifurcation, as presented in Fig. 14. The behavior of the LLE is shown in Fig. 15. In fact, the bifurcation parameter is actually the weight, w, but there is a one-to-one relation between w and the sum of squared fixed-point errors, which is actually used to plot the bifurcation diagram.

Thus, another important consideration can be outlined: the multiobjective procedure defines a systematic way for searching solutions in the space of models. The black-box approach leads to only one point in such space for each model structure (a discrete set of solutions). The multiobjective approach allows a continuous search, starting in the black-box solutions, and visiting an infinite set of "reasonable" solutions. This can be important in the cases when standard MQ procedures does not produce models with desired properties at first but rather yields a solution which is in a vicinity of a good model [Aguirre *et al.*, 2002].

6. Final Remarks

This paper has discussed the use of prior knowledge in nonlinear system identification using a multiobjective approach. The difference between mono- and multiobjective approaches related to the incorporation of prior knowledge on fixed points has been outlined. The mono-objective approach leads directly to a single solution, while the multiobjective approach furnishes *efficient* or *Pareto optimal* solutions. The Pareto set constitutes an

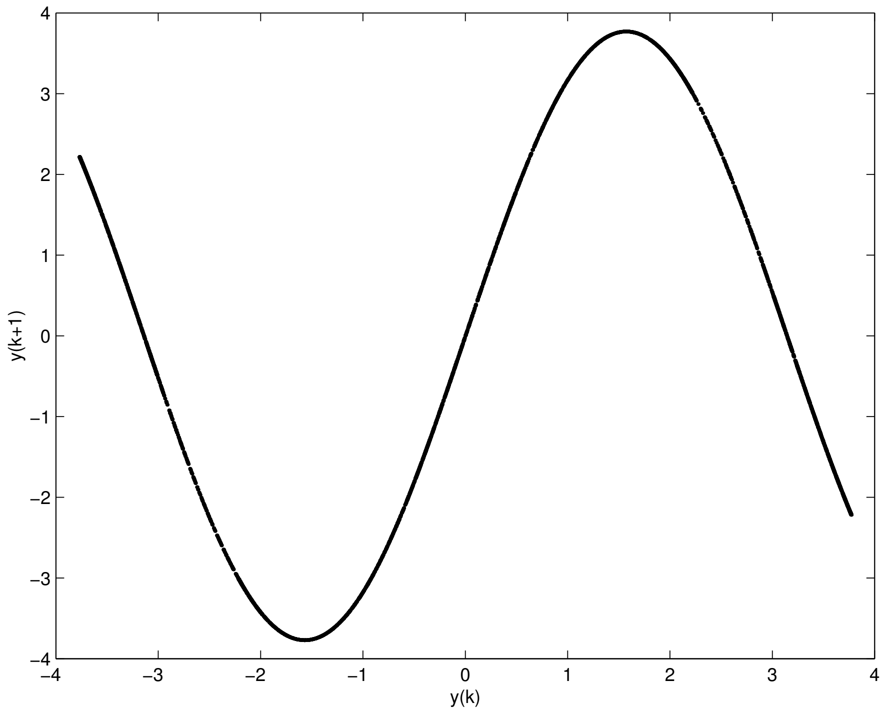

Fig. 10. First return map of the sine map, Eq. (28) with $\alpha = 1.2\pi$.

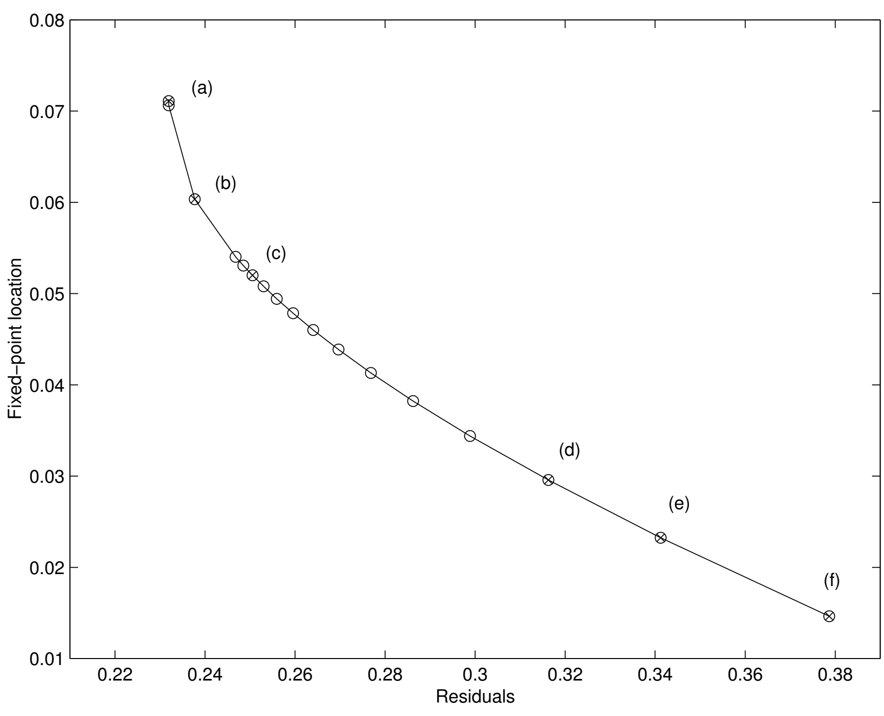

Fig. 11. *Pareto optimal* solutions for the sine map with $\alpha = 1.2\pi$.

object that guides the user interaction with the algorithm, allowing the user to decide what solution is more appropriate for his needs, through a *trade-off analysis*.

The presented methodology also allows to incorporate more than one kind of prior knowledge when this knowledge is related to parameter estimation directly. The use of multiobjective techniques

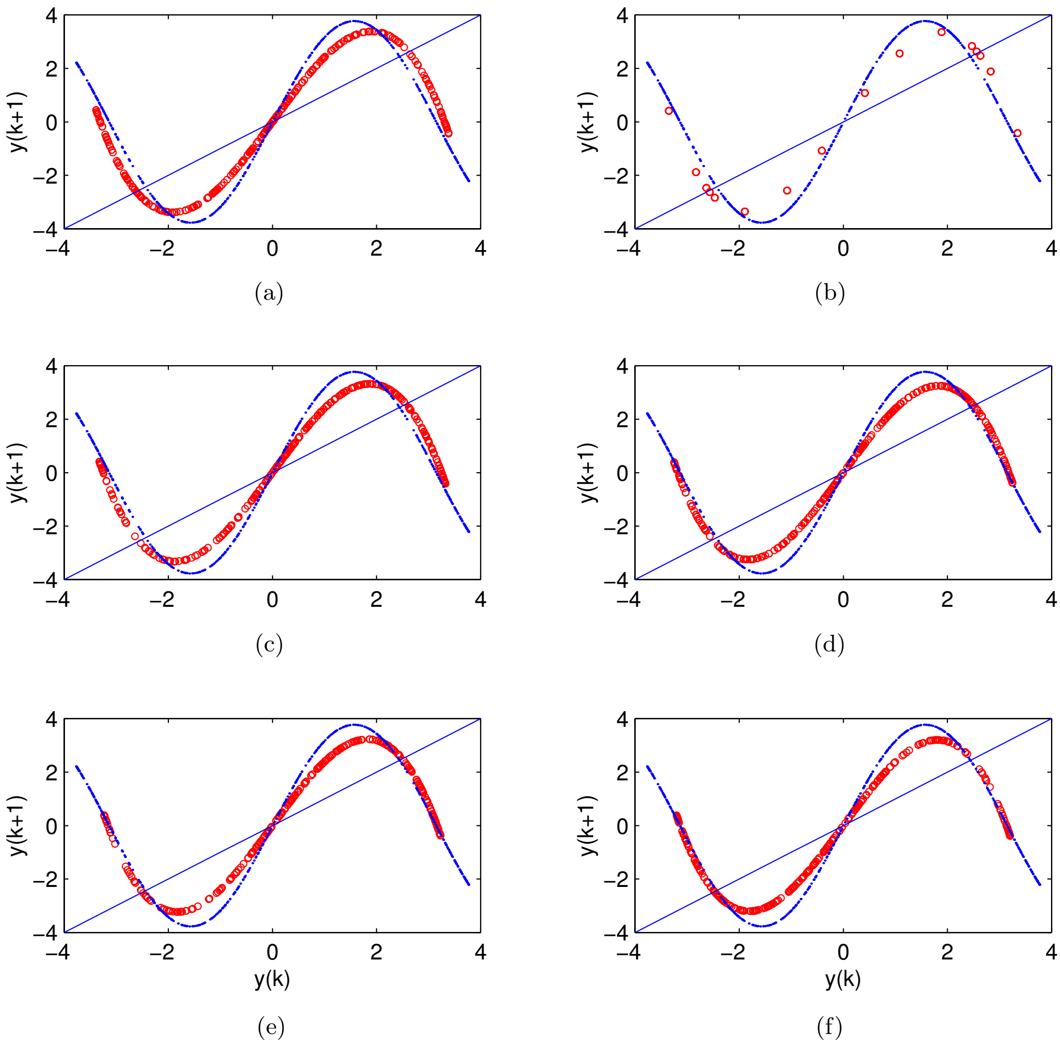

Fig. 12. First return maps of six models (a)–(f) for the sine map. $(\cdots)$ Original system; $(\circ)$ models, $(---)$ $y(k) = y(k+1)$. See Table 3. The fixed-points are located at the crossings of the respective first return maps and the $y = x$ line.

Table 3. Dynamic properties of sine-map.

Model	Weight	SSPE	SSEFP	Fixed Point	LLE
(a)	0	0.2319	0.0711	0; $\pm$ 2.6116	$1.0435 \pm 3.8190 \times 10^{-5}$
(b)	0.99998207821415	0.2377	0.0603	0; $\pm$ 2.5854	$-0.0249 \pm 6.1656 \times 10^{-5}$
(c)	0.99999141427971	0.2506	0.0520	0; $\pm$ 2.5652	$1.0225 \pm 1.0801 \times 10^{-5}$
(d)	0.99999777856618	0.3163	0.0296	0; $\pm$ 2.5103	$1.0264 \pm 9.1272 \times 10^{-6}$
(e)	0.99999848570829	0.3412	0.0232	0; $\pm$ 2.4949	$1.0223 \pm 2.5611 \times 10^{-5}$
(f)	0.99999919285389	0.3787	0.0146	0; $\pm$ 2.4739	$1.0096 \pm 1.4385 \times 10^{-5}$

is suitable for the purpose of dealing with two or more conflicting objectives simultaneously.

The main ideas have been applied to two chaotic bench systems: Chua's circuit and the sine map. In the first case, very little seems to be gained in the ideal case of clean data. However, as noise is added to the identification data, the use of prior knowledge results in better models. A possible explanation for this is that as the noise blurs information (dynamical and static) from the data, this

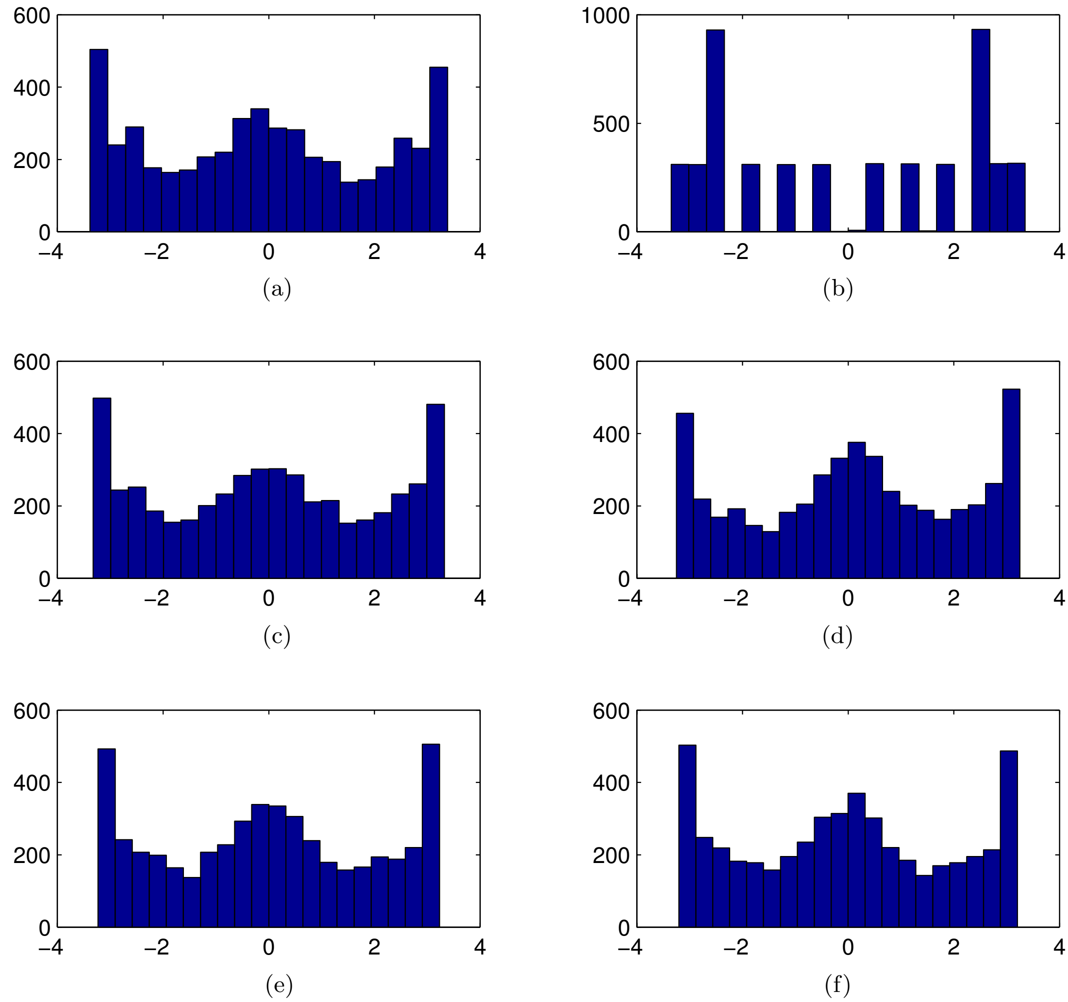

Fig. 13. Histograms that show the distribution in space of data produced by the six models (a)–(f) obtained for the sine map. See Table 3.

can be compensated for, at least to some extent, by means of using additional information. In a practical setting, the location of fixed-points would also have to be estimated from data. However, is has been shown that such estimation is far more robust to noise than the estimation of dynamical models [Aguirre & Souza, 1998].

In the case of the sine map, even in the total absence of noise, the identified models do not necessarily have a good set of fixed-points, mainly due to the fact that the model structures used had cubic nonlinearities while the original system had a sine-type nonlinearity. The use of fixed-point information in this example resulted in models with better dynamical properties in terms of first return map and slightly worse in terms of the largest Lyapunov exponent. In this example it also became clear that the weight used to produce the Pareto set could become an important bifurcation-like parameter that could be used to systematically search for good models in the vicinity of mono-objective black-box models thus overcoming shortcomings of other approaches. This has important consequences in model validation because it often happens that models which, at first sight, are not dynamically valid are in the vicinity of good models. The point is how can we systematically search for good models starting from a "seed model" [Aguirre *et al.*, 2002]? The procedure developed in Sec. 4 seems to be a good alternative.

It is believed that this paper has taken an important step in the modeling of nonlinear dynamics in suggesting a procedure to use prior knowledge in the model building process. Important as this might turn out to be, as pointed out in the introduction, there are a number of important questions that should be considered in order to move forward

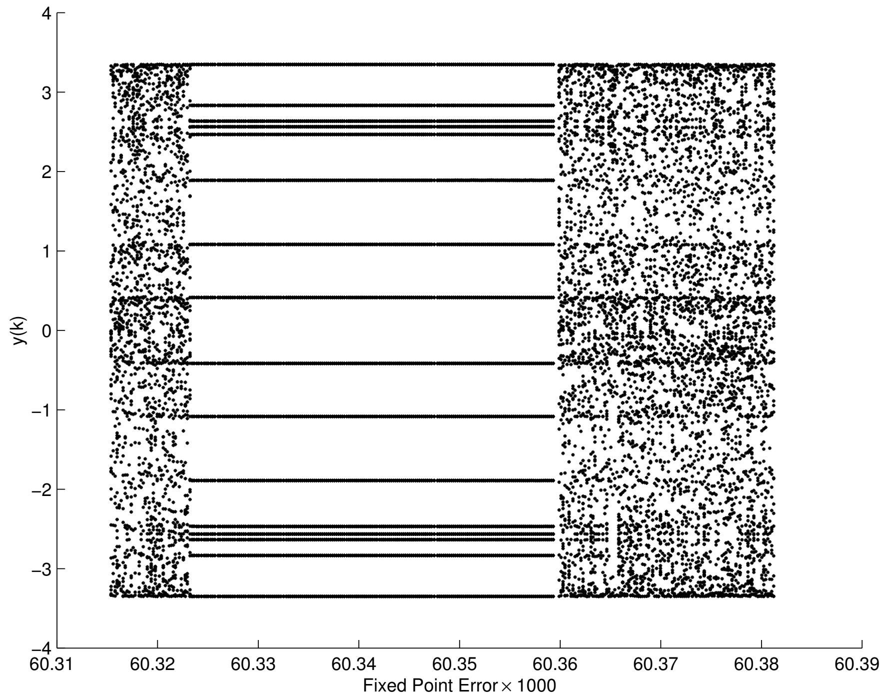

Fig. 14. Bifurcation diagram in the neighborhood of Fixed Point Error = 0.06035, used as the bifurcation parameter.

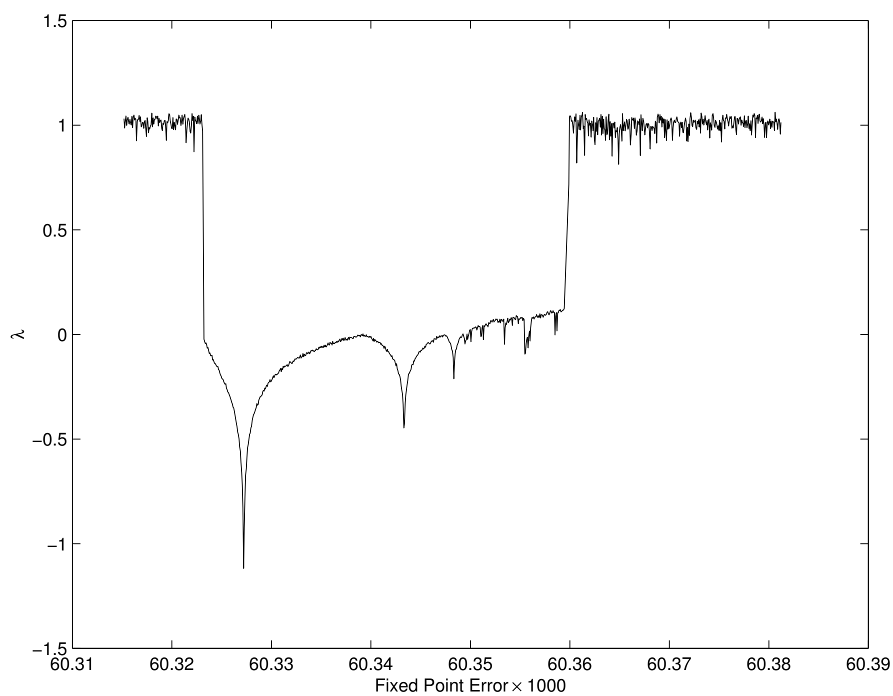

Fig. 15. Variation of the LLE in the neighborhood of Fixed Point Error = 0.06035.

such as: which types of information can be used? How can this be done using other representations? The authors believe that answers to these questions will be welcome contributions to the field.

Acknowledgments

The authors would like to thank CAPES, CNPq and FAPEMIG for supporting this work and

Dr. Eduardo Mendes for his help in computing the largest Lyapunov exponents.

References

Aguirre, L. A. & Mendes, E. M. [1996] "Nonlinear polynomial models: Structure, term clusters and fixed points," *Int. J. Bifurcation and Chaos* **6**, 279–294.

Aguirre, L. A., Rodrigues, G. & Mendes, E. [1997] "Nonlinear identification and cluster analysis of chaotic attractor from a real implementation fo Chua's circuit," *Int. J. Bifurcation and Chaos* **7**, 1411–1423.

Aguirre, L. & Souza, A. [1998] "An algorithm for estimating fixed points of dynamical systems from time series," *Int. J. Bifurcation and Chaos* **8**, 2203–2213.

Aguirre, L., Donoso-Garcia, P. & Santos-Filho, R. [2000] "Use of *a priori* information in the identification of global nonlinear model — a case study using a buck converter," *IEEE Trans. Circuits Syst.-I* **47**, 1081–1085.

Aguirre, L. A., Letellier, C. & Maquet, J. [2002] "Induced one-parameter bifurcations in identified nonlinear models," *Int. J. Bifurcation and Chaos* **12**, 135–145.

Billings, S. A., Chen, S. & Korenberg, M. J. [1989] "Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator," *Int. J. Contr.* **49**, 2157–2189.

Billings, S. A. & Coca, D. [1999] "Discrete wavelet models for identification and qualitative analysis of chaotic systems," *Int. J. Bifurcation and Chaos* **9**, 1263–1284.

Broomhead, D. S. & Lowe, D. [1988] "Multivariable functional interpolation and adaptive networks," *Compl. Syst.* **2**, 321–355.

Brown, R., Rulkov, N. & Tufillaro, N. [1994] "Nonlinear prediction of chaotic time series," *Phys. Rev.* **E50**, 4488–4508.

Casdagli, M., Eubank, S., Farmer, J. D. & Gibson, J. [1991] "State space reconstruction in the presence of noise," *Physica* **D51**, 52–98.

Chankong, V. & Haimes, Y. Y. [1983] *Multiobjective Decision Making: Theory and Methodology* (North-Holland–Elsevier, NY).

Chua, L. O. & Hasler, M. G. E. [1993] "Special issue on chaos in nonlinear electronic circuits," *IEEE Trans. Circuits Syst.-I* **40**, 10–11.

Floriani, E., Dudok de Wit, T. & Le Gal, P. [2000] "Nonlinear interactions in a rotating disk flow: From a Volterra model to the Ginzburg–Landau equation," *Chaos* **10**, 834–847.

Gibson, J. F., Farmer, J. D., Casdagli, M. & Eubank, S. [1992] "An analytic approach to practical state space reconstruction," *Physica* **D57**, p. 1.

Gouesbet, G. & Letellier, C. [1994] "Global vector field reconstruction by using a multivariate polynomial l_2 approximation on nets," *Phys. Rev.* **E49**, 4955–4972.

Irving, A. D. & Dewson, T. [1997] "Determining mixed linear-nonlinear coupled differential equations from multivariate discrete time series sequences," *Physica* **D102**, 15–36.

Johansen, T. [1996] "Identification of non-linear systems using empirical data and prior knowledge — an optimization approach," *Automatica* **32**, 337–356.

Kugiumtzis, D. [1996] "State space reconstruction parameters in the analysis of chaotic time series — the role of time series length," *Physica* **D95**, 13–28.

Lainscsek, C. S. M., Schürrer, F. & Kadtke, J. [1998] "A general form for global dynamical data models for three-dimensional systems," *Int. J. Bifurcation and Chaos* **8**, 899–914.

Leontaritis, I. J. & Billings, S. A. [1985] "Input–output parametric models for non-linear systems — Part II: Stochastic non-linear systems," *Int. J. Contr.* **41**, 329–344.

Letellier, C., Maquet, J., Le Sceller, L. & Aguirre, L. A. [1998] "On the non-equivalence of observables in phase-space reconstructions from recorded time series," *J. Phys.* **A31**, 7913–7927.

Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. [1980] "Geometry from a time series," *Phys. Rev. Lett.* **45**, 712–716.

Rosenstein, M. T., Collins, J. J. & De Luca, C. J. [1994] "Reconstruction expansion as a geometry-based framework for choosing proper delay times," *Physica* **D73**, 82–98.

Sauer, T., Yorke, J. A. & Casdagli, M. [1991] "Embedology," *J. Statist. Phys.* **65**, 579–616.

Suykens, J. A. K. & Vandewalle, J. [1995] "Learning a simple recurrent neural state space model to behave like Chua's double scroll," *IEEE Trans. Circuits Syst.-I* **42**, 499–502.

Takens, F. [1981] "Detecting strange attractors in turbulence," *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, Vol. 898, eds. Rand, D. A. & Young, L. S. (Springer-Verlag, Berlin), pp. 366–381.