Recovering map static nonlinearities from chaotic data using dynamical models

Luis Antonio Aguirre

Centro de Pesquisa e Desenvolvimento em Engenharia Elétrica. Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG, Brazil

Received 28 May 1996; accepted 4 July 1996
Communicated by Y. Kuramoto

Abstract

This paper is concerned with the estimation from chaotic data of maps with static nonlinearities. A number of issues concerning model construction such as structure selection, over-parametrization and model validation are discussed in the light of the shape of the static non-linearities reproduced by the estimated maps. A new interpretation of term clusters and cluster coefficients of polynomial models is provided based on this approach. The paper discusses model limitations and some useful principles to select the structure of non-linear maps. Some of the ideas have been tested using several non-linear systems including a boost voltage regulator map and a set of real data from a chaotic circuit.

1. Introduction

One of the most challenging steps in the construction of non-linear models from a set of data is the choice of the model structure. When the model is non-linear, there is a great diversity of possible structures and there is a danger of choosing a basis which is more complex than necessary. The aforementioned danger is recognized and accounted for by the parsimony principle which basically states that the model should be as simple as possible.

For a long time it seems that researchers always thought of model parsimony as a means of avoiding too large and statistically inefficient models. Also, because of computing limitations, parsimony was a necessary ingredient for any real application to be computationally feasible. As more flexible non-linear representations became popular and computer limitations seemed to fade, parsimony appears to have gained the status of simply a nice principle to bear in mind.

Another overall trend which can be observed is that more complex model structures have been applied. Brief descriptions and comparison of some model structure for non-linear systems can be found in [1–4]. This is desirable because such structures usually can adjust better to a window of data and to model with greater accuracy strong
nonlinearities found in the real world. On the other hand, however, there is an ever increasing danger to construct a model with an overcomplex structure.

In recent years, some attention has been devoted to the issue of model structure selection [5–9]. In particular, it has been argued that model overparametrization usually hampers the model of reproducing some dynamical invariants of the original system [10] and that simplified structures usually perform better [11]. In fact, some results reported in the literature seem to confirm the deleterious dynamical effects of overparametrization [12,13].

This paper is concerned with model construction of nonlinear dynamical systems. Special attention is given to systems with static nonlinearities which appear as nonlinear curves in a bidimensional embedding space. The shape of static nonlinearity in the data should have some bearing on the choice of model structure. Thus static nonlinearities with similar shapes can be modeled by similar structures, even if the original systems are completely different. This suggests that it would be helpful to have a library of static nonlinearities and the type of model structures required to model them.

The paper also points out to a link between NARMAX polynomial models and the respective static nonlinearity. In this respect, it is shown that the concepts of term clusters and cluster coefficients are useful. A number of issues concerning structure selection and model validation are discussed in the light of map static nonlinearities. In particular, it is shown that in some cases, the vanishing of a cluster coefficient is the result of fitting the static nonlinearity to the data. This sheds new light on the present understanding of some model building algorithms and in particular the concepts of cluster coefficients.

The paper is organized as follows: Section 2 provides the necessary background. Section 3 relates the term clusters present in a model and the static nonlinearity of such a model, even in the case when the model dynamical order is greater than one. Several structure selection issues are discussed in Section 4. The results discussed in Sections 3 and 4 are useful in detecting particular weaknesses of polynomial models. This is further discussed in Section 5. Two numerical examples are provided in Section 6 where a number of concepts and ideas presented throughout the paper are tested and illustrated. Finally, the main conclusions are summarized in Section 7.

2. Background

This section briefly overviews some basic concepts and definitions required to understand the paper.

2.1. Estimation of nonlinear autonomous models

Consider the nonlinear autoregressive moving average model with exogenous inputs (NARMAX) [14]

$$y(k) = f^l[y(k-1), \ldots, y(k-n_y), u(k-d), \ldots, u(k-n_u), e(k), \ldots, e(k-n_e)],$$  \hspace{1cm} (1)

where $n_y, n_u$ and $n_e$ are the maximum lags considered for the output, input and noise terms, respectively and $d$ is the delay measured in sampling intervals, $T_s$. Moreover, $u(k)$ and $y(k)$ are, respectively, input and output time series of length $N$ obtained by sampling the continuous data $u(t)$ and $y(t)$ at $T_s$. $e(k)$ accounts for uncertainties, possible noise, unmodeled dynamics, etc. and $f^l[\cdot]$ is some nonlinear function of $y(k)$, $u(k)$ and $e(k)$ with nonlinearity degree $l \in \mathbb{N}$. In this paper, the map $f^l[\cdot]$ is taken to be a polynomial of degree $l$.

The identification of a model consists roughly of two steps, namely structure selection and parameter estimation. In order to estimate the parameters of this map, Eq. (1) can be expressed in prediction error form as

$$y(k) = \psi^T(k-1)\hat{\Theta} + \xi(k),$$  \hspace{1cm} (2)

or in matrix form as
\[ y = \Psi \hat{\Theta} + \Xi, \]

where the hat indicates estimated values and \( \Xi = \{\xi(k)\}_{k=1}^{N} \) is the vector residuals or prediction errors which are defined as the difference between the measured data \( y(k) \) and the one-step-ahead prediction \( \psi^T(k-1)\hat{\Theta} \). The parameter vector \( \Theta \) can be estimated by minimizing the following cost function [15]:

\[ J_{LS}(\hat{\Theta}) = \| y - \Psi \hat{\Theta} \|, \]

where \( \| \cdot \| \) is the Euclidean norm. Moreover, least-squares minimization is performed using orthogonal techniques in order to effectively overcome two major difficulties in nonlinear model identification, namely (i) numerical ill-conditioning and (ii) structure selection.

The problem of structure selection is that of choosing the terms which should compose the model. This amounts to selecting the columns of the regressor matrix \( \Psi \). Clearly, if all possible terms were used to compose \( \Psi \), such a matrix would certainly be ill-conditioned and impractically large for most nonlinear systems. A criterion for structure selection which has proved helpful in many situations involving both real and simulated data is the error reduction ratio (ERR) [16] which is briefly described in what follows.

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

\[ y(t) = \sum_{i=1}^{n_p+n_{\xi}} g_i w_i(t) + \xi(t), \]

where \( n_p + n_{\xi} \) is the number of (process plus noise) terms in the model, \( \{g_i\}_{i=1}^{n_p+n_{\xi}} \) are constant parameters and the polynomials \( \{w_i(t)\}_{i=1}^{n_p+n_{\xi}} \) are constructed to be orthogonal over the data records. Finally, parameters of the model in Eq. (2) can be calculated from \( \{g_i\}_{i=1}^{n_p+n_{\xi}} \).

A criterion for selecting the most important terms in the model has been devised as a by-product of the orthogonal parameter estimation procedure. The maximum mean squared prediction error (MSPE) is achieved when no terms are included in the model, that is, when \( n_p + n_{\xi} = 0 \). In this case the MSPE equals \( y^2(t) \) where the over-bar indicates time averaging. The reduction in the MSPE due to the inclusion of the \( i \)th term, \( g_i w_i(t) \), in the auxiliary model of Eq. (5) is \( g_i^2 w_i^2(t) \). Expressing this reduction as a percentage of the total MSPE yields the error reduction ratio (ERR) [16]

\[ [\text{ERR}]_i = \frac{g_i^2 w_i^2(t)}{y^2(t)} \times 100, \quad i = 1, 2, \ldots, n_p + n_{\xi}. \]

Hence those terms with large values of ERR are selected to form the model.

It should be noted that for nonlinear systems structure selection includes order selection. For instance, \( y(k-1), y(k-2), y(k-1)y(k-2) \) and \( y(k-2)^3 \) could be candidate terms in a second-order model. But which of these terms do we really need? Structure selection aims to answer this question. The ERR criterion is used to sort all possible terms in decreasing order of importance according to the MSPE. Thus ERR provides a family of nested model structures and information criteria can then be used to choose a model within the family [9].

This paper is concerned with the estimation of autonomous maps of the form \( y(k) = f(y(k-1)) \), thus the exogenous variables will not be included in Eq. (1). Moreover, the moving average part will be discarded after parameter estimation.
2.2. Term clustering

The deterministic part of a NARMAX model, that is, a NARX model, can be expanded as the summation of terms with degrees of nonlinearity in the range $1 \leq m \leq l$. Each $m$th-order term can contain a $p$th-order factor in $y(k - n_i)$ and an $(m - p)$th-order factor in $u(k - n_i)$ and is multiplied by a coefficient $c_{p,m-p}(n_1, \ldots, n_m)$ as follows:

$$y(k) = \sum_{m=0}^{l} \sum_{p=0}^{m} \sum_{n_1, n_m} c_{p,m-p}(n_1, \ldots, n_m) \prod_{i=1}^{p} y(k - n_i) \prod_{i=p+1}^{m} u(k - n_i),$$

(7)

where

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

(8)

and the upper limit is $n_y$ if the summation refers to factors in $y(k - n_i)$ or $n_u$ for factors in $u(k - n_i)$. The representation in Eq. (7), also known as the Kolmogorov–Gabor polynomial approach, and similar variants have been considered by several authors [5,17].

In order to illustrate this representation, consider the following polynomial:

$$y(k) = c_{0,0} + \sum_{n_1=1}^{n_y} c_{1,0}(n_1) y(k - n_1) + \sum_{n_1=1}^{n_u} c_{0,1}(n_1) u(k - n_1)$$

$$+ \sum_{n_1=1}^{n_y} \sum_{n_2=1}^{n_y} c_{2,0}(n_1, n_2) y(k - n_1) y(k - n_2) + \sum_{n_1=1}^{n_y} \sum_{n_2=1}^{n_u} c_{1,1}(n_1, n_2) y(k - n_1) u(k - n_2)$$

$$+ \sum_{n_1=1}^{n_y} \sum_{n_2=1}^{n_u} c_{0,2}(n_1, n_2) u(k - n_1) u(k - n_2),$$

(9)

which is the expansion of Eq. (7) up to second order, that is, $l = 2$. It should be noted that the terms do not depend on the order of factors thus $y(k - n_1) y(k - n_2) = y(k - n_2) y(k - n_1)$.

Eq. (7) reveals that there are many possible terms in a polynomial model. However, it seems natural to consider groups of similar terms because in practice similar terms describe the same type of nonlinearity and dissimilar terms usually confer upon the model different dynamical characteristics. For instance, terms such as $y(k - 1)^2$, $y(k - 2) y(k - 3)$ and $y(k - 1) y(k - 3)$ would be members of the group which includes all the quadratic terms in the output variable. Of course, in a model each term has its own coefficient and not all possible term groups need be represented in a certain model.

The term groups mentioned in the last paragraph are called term clusters [18]. The set of terms represented by $\Omega_{y^{p}u^{m-p}}$ contains terms of the form $y(k - i)^p u(k - j)^{m-p}$ for $m = 0, \ldots, l$ and $p = 0, \ldots, m$. In other words, $y(k - i)^p u(k - j)^{m-p} \in \Omega_{y^{p}u^{m-p}}$ for $m = 0, \ldots, l$; $p = 0, \ldots, m$; $i = 1, \ldots, n_y$ and $j = d, \ldots, n_u$.

The summation of the coefficients of all the terms which pertain to a certain cluster is the cluster coefficient [18] and is denoted by $\Sigma_{y^{p}u^{m-p}}$. Generically, for Eq. (7) the cluster coefficients are $\sum_{n_1, n_m} c_{p,m-p}(n_1, \ldots, n_m)$.

For example, the term clusters represented in the model of Eq. (9) for $n_y = 2$ and $n_u = 1$ are $\Omega_0$ (the constant term), $\Omega_y$, $\Omega_u$, $\Omega_{y^2}$, $\Omega_{yu}$ and $\Omega_{u^2}$. The respective cluster coefficients are $\Sigma_0 = c_{0,0}$, $\Sigma_y = c_{1,0}(1) + c_{1,0}(2)$, $\Sigma_u = c_{0,1}(1)$, $\Sigma_{y^2} = c_{2,0}(1, 1) + c_{2,0}(1, 2) + c_{2,0}(2, 2)$, $\Sigma_{yu} = c_{1,1}(1, 1) + c_{1,1}(2, 1)$ and $\Sigma_{u^2} = c_{0,2}(1, 1)$. 
Summarizing, a cluster $\Omega_{y^m u^p}$ is a set of terms of the form $y(k - i)^p u(k - j)^m - p$ for $m = 0, \ldots, l$ and $p = 0, \ldots, m$, and the respective coefficient, $\Sigma_{y^m u^p}$, is the summation of the coefficients of all the model terms contained in such a cluster.

The set of candidate terms for a NARX model is the union of all possible clusters up to degree $l$. If a certain cluster is not required to compose a model, such a cluster is said to be spurious as opposed to the effective clusters needed to obtain a dynamically valid model.

It has been shown that if a certain term cluster is spurious, the respective coefficient will gradually become small or will oscillate around zero as the number of terms in the model is increased [18]. This procedure is simple, quite robust and can be used in structure selection problems. Moreover, it has been argued that it overcomes some of the drawbacks of more conventional methods [9] such as the zeroing-and-refitting [6].

In future sections it will be seen that terms pertaining to the same cluster have a similar effect on the static nonlinearity of the map. Another advantage of thinking in terms of clusters is that the clustered polynomial (see Eq. (11) below) is always in the form of a first-return map $y(k) = f(y(k - 1))$. Moreover, in Section 3.1, a geometrical interpretation will be provided for the fact that the cluster coefficient of spurious clusters tend to become small.

3. Map static nonlinearities

Consider an $n_y$th-order map $y(k) = f(y(k - 1), \ldots, y(k - n_y))$. As the map is iterated, past values become the arguments of $f(\cdot)$ in order to determine the next future value. When $f(\cdot)$ is nonlinear, it is well known that for some parameter values the sequence $y(k)$ can exhibit quite complicated dynamics even when $f(\cdot)$ has a simple form. It is often helpful to think of the sequence $y(k)$ as the result of iterating past values according to a certain nonlinear law which is precisely the static nonlinearity of the dynamical map.

In order to illustrate this, consider the well-known logistic map $y(k) = \alpha[1 - y(k - 1)]y(k - 1)$. $y(k)$ can be obtained iterating the last value of the sequence according to the parabola defined by $y = \alpha(x - x^2)$. This approach is widely used as it provides a visual way of undersanding the dynamics of the logistic map. Clearly, the parabola is the static nonlinearity of the logistic map.

The logistic map is of first order. What happens to the static nonlinearity if the map order is increased? Is the static nonlinearity still the same? In which sense is it the same? The answers to these questions are important in modeling problems where the order of the dynamics is not usually known a priori. In Section 3.1 and following sections, some answers to these questions will be suggested.

In the endeavour to attain a better fit between model and data, there is the danger of using a model of higher order than really required. Although a better fit is usually achieved (lower error variance), the dynamics of the resulting map might turn out to be very different from those of the system [10,19].

It should be noted that in many practical cases the static nonlinearity does appear to exist on the plane $y(k) \times y(k - 1)$ thus suggesting a first-order map. However, it might happen that a map of greater order is desired so that a better fit is attained. In such cases, in order to assess the static nonlinearity, an approximation is to take $x = y(k - 1) = \cdots = y(k - n_y)$. This yields $y = f(x)$, and $f(x)$ can be thought of as the projection of the static nonlinearity onto the $y \times x$ plane. Therefore, one can assess the effects of deliberately increasing the map order by comparing the projection $f(x)$ with the first-return map obtained from the data. This will be illustrated in Section 4.

3.1. Relation to term clusters

All the possible clusters of an autonomous polynomial with degree of nonlinearity $l$ are $\Omega_0 = \text{constant, } \Omega_y, \Omega_{y^2}, \ldots, \Omega_{y^l}$. Thus, the fixed points of a map with degree of nonlinearity $l$, see Eqs. (7) and (8), are given by
the roots of the following clustered polynomial:
\[
y(k) = c_{0,0} + y(k) \sum_{n_1=1}^{n_y} c_{1,0}(n_1) + y(k)^2 \sum_{n_1,n_2} c_{2,0}(n_1,n_2) \\
+ \cdots + y(k)^l \sum_{n_1,n_l} c_{l,0}(n_1, \ldots, n_l),
\]
(10)

where the definition of fixed points was used. Finally, using the definition of cluster coefficients and dropping the argument \(k\), Eq. (10) can be rewritten as follows:
\[
y = \Sigma y_1 y^l + \cdots + \Sigma y_2 y^2 + \Sigma y + \Sigma_0,
\]
(11)
where \(\Sigma_0 = c_{0,0}\) is a constant. From Eq. (11) it becomes clear that an autonomous polynomial with degree of nonlinearity \(l\) will have \(l\) fixed points if \(\Sigma y \neq 0\).

As mentioned above, given a nonlinear polynomial model of order \(n_y\), the static nonlinearity can be easily recovered by taking \(x = y(k - 1) = \cdots = y(k - n_y)\) which yields \(y = f(x)\). An important remark which should be made is that \(f(x)\) is the right hand side of the clustered polynomial in Eq. (11). That means to say that the monomials in \(f(x)\) correspond to the term-clusters present in the model and the coefficients of such monomials will be precisely the cluster coefficients.

Another remark is that because (11) is \textit{clustered}, it does not depend on the map dynamical order. Thus even if the order is greater than one, by using term clusters and cluster coefficients, it is possible to consider a projection of the static nonlinearity on the plane.

These observations help to interpret some previous results. In [18] it has been shown how the monitoring of cluster coefficients can be used to decide which term clusters are important in constructing a nonlinear dynamical polynomial model. The arguments in that reference were laid on a statistical basis. This section, however, suggests an alternative and simpler explanation. \textit{If there is an indication that a term cluster is spurious, that means that the estimation algorithm is adjusting the model in such a way as to adapt the algebraic relation} \(f(x)\) \textit{to the static nonlinearity of the original system}. For instance, in the case of the logistic equation the indication of cubic terms (terms in \(\Sigma y^3\)) as being spurious would come as no surprise because the static nonlinearity is a parabola which only needs linear and quadratic terms to be accurately represented. This example also illustrates that if cubic terms are included then, quite inevitably, the static nonlinearity of the resulting map will have an extra inflexion, thus characterizing a detrimental effect due to model overparametrization. Some of these issues will be pursued further and illustrated in Section 4.

4. Term clusters and structure selection issues

4.1. The sine-map with quadratic-type nonlinearities

In this section we follow up the observation made in Section 3. To begin with, consider the following map:
\[
y(k) = \alpha \sin(y(k - 1))
\]
(12)
with \(\alpha = \pi\). For an initial condition \(y(0) \in [0, \pi]\), Eq. (12) maps the interval \([0, \pi]\) onto itself. This map has fixed points approximately at \(y = \{0, 2.3137\}\) and an estimated Lyapunov exponent equal to \(\lambda = 0.997 \pm 0.002\) bits/s.

Taking 500 data points on this attractor the following two-term model was estimated:
\[
y(k) = 3.8901y(k - 1) - 1.2503y(k - 2)^2,
\]
(13)
which has fixed points at \( \bar{y} = \{0, 2.3115\} \) and Lyapunov exponent \( \lambda = 0.712 \pm 0.003 \) bits/s. The residual variance for this model is \( 4.4 \times 10^{-3} \). The choice of the structure of model (13) is easily justified by noticing that the static nonlinearity of the map which generated the data resembles the static nonlinearity of the logistic map within the range \([0, \pi]\). This suggests that the terms of the present model can be taken from the same clusters as the logistic map, that is, \( \Omega_y \) and \( \Omega_{y2} \).

Although the fixed points are accurately estimated, the Lyapunov exponent is almost 30% lower than the value estimated for the original map. In order to try to estimate a more accurate map, two things can be done as far as model structure is concerned, namely (i) increase the model order, and (ii) increase the degree of nonlinearity. Doing this and limiting the number of terms to five, the following model was estimated:

\[
y(k) = -0.1468 + 0.0055y(k - 1)^3 + 4.1309y(k - 1) - 1.3272y(k - 1)^2 - 0.0013y(k - 1)^2 y(k - 3),
\]

which has a positive Lyapunov exponent \( \lambda = 0.927 \pm 0.002 \) bits/s. The residual variance for this model is \( 3.6 \times 10^{-3} \). The terms in the model above were automatically chosen using the ERR criterion of Eq. (6). Clearly, the value of the Lyapunov exponent has been estimated with greater accuracy. The slight reduction in the residual variance is a consequence of the increase in the number of parameters. The price paid for this improvement is the inclusion of terms from the \( \Omega_0 \) and \( \Omega_{y3} \) clusters. Two direct consequences of this, irrespective of parameter values, are that the number of fixed points will now be three instead of two, and there will no longer be a trivial fixed point [20]. In fact, model (14) has fixed points at \( \bar{y} = \{0.0479, 2.3286, 315.8323\} \). Thus it is important to realize that whilst increasing model structure complexity might improve some invariants, other invariants may well deteriorate.

It is conjectured that the improvement in the estimate of the Lyapunov exponent comes as a consequence of fine-tuning the static nonlinearity of the map. This has been accomplished by the (automatic) inclusion of terms from other clusters. It is vital to realize that this happened because the original static nonlinearity cannot be perfectly described only in terms of clusters \( \Omega_y \) and \( \Omega_{y2} \). However, the inclusion of the constant term (cluster \( \Omega_0 \)) has made it impossible for the model to have a trivial fixed point. On the other hand, the inclusion of terms taken from cluster \( \Omega_{y3} \) introduces an extra fixed point.

In order to illustrate this matter, consider Fig. 1 which shows the static nonlinearities of the models (13) and (14). It becomes clear that the inclusion of \( \Omega_{y3} \) in model (14) has qualitatively altered the map static nonlinearity. The
appearance of an extra fixed point can be seen as a consequence of this. However, it is noteworthy that in the attractor region, marked in the figure, both static nonlinearities are quite similar. Thus, having included terms from \( \Omega_{y, s} \) in the model, it becomes apparent that the algorithm will estimate parameters in such a way as to ensure that in the region of the estimation data, the static nonlinearity resembles a parabola. This is done by reducing the relative importance of some coefficients in the clustered polynomial which determines the static nonlinearity. Finally, it is pointed out that the coefficients of the clustered polynomial are precisely the cluster coefficients. Therefore, the use of term clusters and cluster coefficients in structure selection problems can now be justified also in terms of static nonlinearity fitting. This is an important interpretation and complements previous results [18, 20].

4.2. The sine-map with cubic-type nonlinearities

In what follows the sine-map of Eq. (12) is considered anew with \( \alpha = 1.2\pi \). This equation maps the interval \([-\pi, \pi]\) onto itself. The fixed points are approximately \( \bar{y} = \{0, \pm 2.439\} \) and the map has an estimated Lyapunov exponent equal to \( \lambda = 1.155 \pm 0.009 \) bits/s. Fig. 2(a) shows the embedded attractor and its projections.

First, second- and third-order models were estimated from 500 data points taken from the attractor shown in Fig. 2. The first-order model is

\[
y(k) = 2.6893 y(k-1) - 0.2479 y(k-1)^3
\]

with residual variance of 0.31, fixed points at \( \bar{y} = \{0, \pm 2.610\} \) and Lyapunov exponent \( \lambda = 1.095 \pm 0.015 \) bits/s.

The second-order model is

\[
y(k) = 0.01925 y(k-1) y(k-2)^2 - 0.2402 y(k-1)^3
+ 2.5366 y(k-1) - 0.00278 y(k-1)^2 y(k-2)
\]

with residual variance of 0.29, fixed points at \( \bar{y} = \{0, \pm 2.621\} \) and Lyapunov exponent \( \lambda = 1.012 \pm 0.004 \) bits/s.

The third-order model is

\[
y(k) = -3.2782 \times 10^{-3} y(k-1) y(k-2)^2 - 0.2379 y(k-1)^3
+ 2.5933 y(k-1) - 3.1194 \times 10^{-2} y(k-1) y(k-2) y(k-3)
- 2.2703 \times 10^{-2} y(k-1)^2 y(k-3) + 9.2318 \times 10^{-3} y(k-3)^3
\]

with residual variance of 0.27, fixed points at \( \bar{y} = \{0, \pm 2.670\} \) and Lyapunov exponent \( \lambda = 1.129 \pm 0.007 \) bits/s.

The attractors of these models are shown in Figs. 2(b), (c) and (d) respectively. As can be seen, all the models settle to attractors which resemble the original one. Nonetheless, it is quite obvious that the model with greater structure complexity presents a fuzzier attractor, see Fig. 2(d). This seems to be an early indication of overparametrization. In fact, models with \( n_p = 7 \) and \( n_p > 8 \) are already unstable. It should be noted that the residual variance usually decreases as the number of fitted parameters increases. A consequence of this is the well-known fact that such an indicator is not usually a good model validator.

It is curious that in both cases, that is for \( \alpha = \pi \) and for \( \alpha = 1.2\pi \), the models with a slightly more complex structure yielded attractors with better largest Lyapunov exponents, namely models (14) and (17). This remark seems important for model validation because it brings to light that some dynamical invariants might, in some cases, fail to reveal model overparametrization.

An important difference noticed in the identification of models (13) and (14) compared to models (15) and (16) is that for the latter, the terms pertaining to cluster \( \Omega_{y, s} \) have significantly increased in importance whilst the terms in \( \Omega_{y, r} \) were not even chosen by the structure selection algorithm.
Fig. 2. Embedded attractors of (a) the map in Eq. (12) with $\alpha = 1.2\pi$, (b) the model in Eq. (15), (c) the model in Eq. (16), (d) the model in Eq. (17).

One way of quantifying this importance is by monitoring the values of the cluster coefficients for an increasing number of terms $n_p$. As discussed in [18] cluster coefficients of unimportant or spurious clusters either become very small or tend to oscillate around zero. This is shown in Fig. 3. This figure suggests that for the sine-map with $\alpha = \pi$ the effective clusters are $\Omega_y$ and $\Omega_{y2}$, and that for the case $\alpha = 1.2\pi$, the important clusters are $\Omega_y$ and $\Omega_{y3}$. This information is in accord with our expectations considering the type (shape) of static nonlinearity as well as the number of fixed points and the existence of a trivial fixed point. Concerning Fig. 3, it is pointed out that in every case, the estimated parameters used to compute the cluster coefficients were statistically significant with a confidence of 95%.
Fig. 3. Cluster coefficients for models with increasing number of process terms $n_p$: (a) $\Sigma_0$, (b) $\Sigma_2$, (c) $\Sigma_3$ and (d) $\Sigma_3$. Small or oscillatory values indicate clusters which are probably spurious [18]. The models were estimated from data generated by the map (12) for (-) $\alpha = 1.2\pi$ and (--) $\alpha = \pi$.

In the examples given in this section, overparametrization had basically two effects. First, the resulting attractors became fuzzier as seen in Fig. 2(d). Second, the largest Lyapunov exponent was better estimated. This suggests that in some cases slight overparametrization might be permitted to improve the estimation of some dynamic features of the resulting model at the expense of other dynamic invariants. But, of course, the ultimate effect of overparametrization is the utter loss of stability and should therefore be avoided.

4.3. Overparametrization effects

We now conclude this section with an example of how overparametrization affects the static non-linearity of the estimated models. In order to do so several models were estimated from the data on the attractor shown in Fig. 2(a). Both the order and the number of terms in the models were increased. The results are summarized in Fig. 4.

This figure brings to light the futility of trying to model the data with a global linear model since such a static linearity cannot possibly conform to the static nonlinearity of the original map. For the same reason a polynomial model with degree of nonlinearity equal to two would also fail to adjust to the static nonlinearity of the original system. On the other hand, a model with terms pertaining to $\Omega_{1,3}$ have a cubic-type static nonlinearity which comes much closer to the original attractor. If a model is overparametrized, the respective static nonlinearity is cubic-type but it might become out of phase with respect to the original attractor. This is indicated in Fig. 4 by the dashed lines which were sectioned in order to fit the figure scales. Note that this fact might well be seen as being responsible for the model instability.
Finally, it is pointed out that in Fig. 4, the worse nonlinear model (the overparametrized) has fixed points which are very close indeed to those of the original map. To see this, notice that the fixed points of the respective models are located at the intersections of the $y = x$ line (⋯ in the figure) with the respective static nonlinearity. However, although the location of the fixed points is much more accurate in this case, all the fixed points of this model are unstable. Therefore, here is another example in which poor models have some dynamic invariants (fixed-point locations) which are closer to the original system than those of some ‘good’ models. Clearly, this has some bearing on model validation.

5. Limitations of polynomial models

The real usefulness of polynomials to model dynamical data has always been an issue in the literature [2,3,21] although most researchers agree that such polynomials have a simple structure and therefore are amenable to analysis. However, a clear drawback which cannot be overlooked is that such models can become impractically large and unstable very easily [2,3]. Although this observation is actually true, it has been argued that most of these deleterious effects are consequences of model structure complexity [10]. If the structures of polynomial models are kept relatively simple, then such models can be used to model nonlinear dynamics in a number of situations [21]. This can be achieved by applying specialized algorithms developed for detecting the structure of nonlinear models [16,18].

However, it is known that polynomial models cannot be used to model any kind of nonlinear dynamics. Thus it seems that a question which still needs to be answered is: which type of dynamics can we expect to model adequately with nonlinear polynomial models? An exhaustive answer to this question is beyond our present understanding. It is believed, however, that the result discussed in Section 4 provide some insight into this issue and thus a further step can be taken in formulating an answer for the question above. For systems with static nonlinearities, the main
Fig. 5. (c) 500 data points on the tent attractor. Such data were used to identify nonlinear models. The static nonlinearities are shown for models with (−) three terms, (−−) four terms and (−−−) six terms. It is obvious that increasing the number of parameters of the model will hardly yield a good approximation.

conclusion can be summarized as follows: *A polynomial model with degree of nonlinearity $l$ will accurately describe a nonlinear dynamical map in so far as the static nonlinearity of the model (determined by the clustered polynomial) is able to conform to the static nonlinearity of the original map.*

Based on this conclusion, it becomes clear why a simple polynomial model with degree of nonlinearity $l = 3$ is quite able to resemble the dynamics of the sine-map which, in principle, has nothing of polynomial. On the other hand, it can be expected that a polynomial model would fail to provide a good description for the dynamics of, say, the *tent map* $y(k) = 1 - 2|y(k-1) - 0.5|$ which looks more polynomial than the sine-map. In fact, no polynomial model up to degree $l = 3$ has been found to reproduce the tent attractor, which now seems a predictable result.

Fig. 5 shows the tent attractor and the static nonlinearities of polynomial models identified from such data. The model terms were selected from linear, quadratic and cubic clusters. Moreover, models up to order three have been considered. Observing the figure, it becomes clear that simply increasing the number of terms or even increasing the model dynamical order will not significantly improve the approximation. The reason for this is that the type of static nonlinearity displayed by the system can hardly be approximated by a polynomial with a finite degree of nonlinearity.

It is interesting that the fixed points of the models whose static nonlinearities are shown in Fig. 5 are rather accurately estimated. In particular, the four-term model has fixed points located at 0.0691 and 0.6869 compared to 0.00 and 0.6667 of the original map. Thus, it seems that in the identification of a regressive model, good estimation of fixed points is usually achieved even if the static nonlinearity is not.

This last example illustrates a basic limitation of polynomial models in modeling nonlinear data in situations where the system nonlinearity can be modeled as a static nonlinearity. Although the model parameters are estimated in order to fit a *dynamical* model to a piece of data, it is most helpful to compare *static* nonlinearity of the model with that of the data to assess the feasibility to estimating a good model.
Therefore, in cases where a good polynomial cannot be obtained, this viewpoint will enable the user to search for other representations which will provide a better way of describing the static nonlinearity in the data. From the tent map static nonlinearity, it becomes obvious that a threshold model will do the trick. In fact, the tent map is precisely a threshold model.

It should be realised that the limitations discussed above have been furnished in the context of nonlinear maps. A first return map of a dynamic system might display a rather complicated shape thus suggesting that a low-order polynomial model would not be adequate. Nonetheless, a low-order polynomial could still be good enough to accurately model the system dynamics (not the first-return map) which are usually smooth functions of time. An example of this will be provided in Section 6.2.

6. Application to map estimation

The basic message of Sections 4 and 5 is that the static nonlinearity to which a nonlinear model can conform greatly depends on the model structure. If the model dynamical order is greater than one, then a type of projection of the static nonlinearity can be readily obtained from the clustered polynomial. This kind of information sheds some light on the question of model structure. In particular, in some cases it might become apparent that terms from a certain cluster should compose the model in order to be able to reproduce the static nonlinearities.

In what follows, two simple examples are provided. In both cases the main objective will be to obtain a map which should fit the first-return map reconstructed from data. It will be seen how the shape of the static nonlinearities in these cases suggests the model structure for the map.

6.1. A map for a feedback buck switching regulator

This example considers a buck switching regulator which operates in closed-loop in order to guarantee a certain voltage across the load [22]. A map was obtained from the circuit equations which gives one (sampled) value of the voltage across the load per cycle and has the general form [22]

\[ y(k) = \alpha y(k - 1) + \frac{h(d_n)2\beta E[E - y(k - 1)]}{y(k - 1)}. \] (18)

where \( \alpha, \beta \) and \( E \) are constants which only depend on the circuit components, \( d_n \) is the controller output (a voltage signal) and the saturation \( h(d_n) \) is given by

\[ h(d_n) = \begin{cases} 
0 & \text{if } d_n < 0, \\
1 & \text{if } d_n > 1, \\
d_n & \text{otherwise.} 
\end{cases} \] (19)

The attractor produced by the map (18) is shown in Fig. 6 where it was labeled ‘original attractor’. A glance at the shape of this attractor suggests the following: (i) the general appearance of the static nonlinearity is of a parabola, (ii) for ‘low’ values of \( y(k) \) the nonlinearity has a typically exponential shape, and (iii) for \( y(k) > 27 \) the static nonlinearity is basically a line. This information can be used to suggest the structure of a map as follows. The basic parabola is given by \( y = ax^2 + bx + c \). In order to guarantee that for \( x \to \infty \) the parabola becomes a line, it suffices to divide the parabola by \( x \). Finally, the exponential characteristic for \( x \to 0 \) is attained by adding the term \( d \exp(-x) \) which has negligible effect for large values of \( x \). A quick glance at the original attractor reveals that the exponential characteristic of the attractor for ‘small’ values of \( x \) does not take place in the vicinity of \( x = 0 \) but
rather at $x \approx 22$. Such attractor displacement can be achieved by including a constant value in the exponential term, that is $d \exp(22 - x)$. Thus the following dynamical map can be suggested:

$$y(k) = \frac{a y(k - 1)^2 + b y(k - 1) + c}{y(k - 1)} + d \exp[22 - y(k - 1)].$$

(20)

Conventional least-squares estimation of the parameters yields $a = 2.6204$, $b = -9.9875 \times 10$, $c = 1.4171 \times 10^3$ and $d = 4.6429 \times 10$. This map has a positive Lyapunov exponent $\lambda = 0.503 \pm 0.004$ bits/s. It is pointed out that the constant 22 inserted in the exponential term is not critical. Variations in this constant are, to some extent, compensated by the estimated parameter $d$. Moreover, the value $x \approx 22$ is suggested by the reconstructed attractor and no wild guess is required. In principle, estimation of this value is possible with nonlinear parameter estimation algorithms.

The static nonlinearity of map (20) and the respective attractor are shown in Fig. 6. One hundred iterates of the original and estimated maps are shown in Fig. 7. These figures show that a good agreement has been achieved although the respective maps are quite different. The explanation for this is that the map structures are capable of adjusting to similar static nonlinearities.
It should be emphasized that the objective of this example was not to propose a general procedure for map estimation. On the contrary, an ad hoc procedure has been followed. However, two important and rather general aspects in this example are: (i) based on the overall shape of the static nonlinearity, in some cases it is possible to suggest viable model structures, and (ii) good model structures can be apparently different provided they are able to adjust to similar nonlinearities.

6.2. Estimation of a first-return map from real data

This example uses a set of real data measured from an implementation of Chua’s circuit [23,24] evolving on the spiral attractor. The original data are composed of measurements of the voltage across Chua’s diode (the $x$ component) sampled at $T_s = 20\mu s$. The sequence of consecutive maxima in the time series was denoted $y(k)$. The first-return map of these data is shown in Fig. 8.

As seen from this figure, the static nonlinearity is quite smooth and considerably symmetric. Also, the data $y(k)$ strongly suggest that a low-dimensional polynomial map would suffice. First-order polynomial maps with degree of nonlinearity varying from 2 to 6 were identified directly from these data. The resulting first-return maps of the polynomials with even degrees of nonlinearity are shown in Fig. 8.

The least-squares estimated map with degree of nonlinearity equal to 6 is

$$
y(k) = -[2.0065 + 4.8134 y(k - 1) + 2.3744 \times 10 y(k - 1)^2 + 5.9455 \times 10 y(k - 1)^3
+ 6.1998 \times 10 y(k - 1)^4 + 2.8003 \times 10 y(k - 1)^5 + 4.5980 y(k - 1)^6],$$

which has a positive Lyapunov exponent $\lambda = 0.848 \pm 0.006 \text{ bits/s}$.

It is interesting to point out that the polynomials with odd degrees of nonlinearity did not reveal any significant improvement compared to the preceding even polynomial map. For instance, the map of degree 3 was not better than the map of degree 2. This should cause no surprise as the attractor in Fig. 8 displays even-type symmetry. From Eq. (11) it is seen that only certain clusters can be used to compose a model if such a model is supposed to retain some symmetry properties, especially concerning its fixed points. For details see [20].
Finally, it is pointed out that 5th-order polynomial models just with quadratic and cubic nonlinearities have been estimated [25] for the spiral attractor from the data which gave rise, via Poincaré sampling, to the data shown in Fig. 8. This illustrates that even if a polynomial model with small degree of nonlinearity cannot be estimated from first-return map data, such model representation can still be useful for modelling the dynamics underlying the time series.

7. Conclusions

This paper has discussed a number of aspects concerning the estimation of nonlinear dynamic maps with static nonlinearities. Several results described in the paper have consequences in many aspects of nonlinear system modeling. The main results, obtained using well-known maps and real data alike, are summarized and organized under specific headings in what follows.

Model validation. Which properties should be used as criteria to assess model quality? This is a nontrivial question in model validation. Because of several reasons, in the realm of nonlinear dynamics prediction errors are no longer considered good validators. Alternative criteria include dynamical invariants which usually are sensitive to parameter overparametrization, a desirable feature [26]. The results in this paper have shown that in some cases whilst slight overparametrization might deteriorate some invariants, others do improve, in particular the largest Lyapunov exponent. This calls for caution in choosing which dynamical invariants to use as criteria to assess model quality.

Overparametrization. In a previous paper, it was shown how overparametrization affects the dynamical invariants of mathematical models [10]. In this paper it has been shown that increasing the number of parameters does affect the shape of the model static nonlinearity. In the examples discussed in the paper, such effects were basically two, namely: (i) the first-return map becomes more fuzzy, and (ii) the static nonlinearity becomes out-of-phase with respect to the original one. The latter effect resulted in the loss of stability.

Structure selection. The shape of the first-return map is a picture of the type of nonlinearity which should be present in the model in order to reproduce the dynamics of the map. This remark has been used to formalize a limitation of polynomial models. Moreover, such a remark suggests that it would be helpful to have a library of static nonlinearities with the respective maps. In this way, whenever a first-return map of a system displayed a certain shape, one could resort to such a library and pick out the type of nonlinear terms required to reproduce the dynamics. This has been illustrated by means of an example which also served to highlight the fact that model structures which appear widely different can reproduce the same type of dynamics.

Term clustering interpretation. The use of term clusters and cluster coefficients in the structure selection of nonlinear systems was recently established [18]. In that paper, it was shown and argued that if a term cluster is unimportant, the respective coefficient will either become small or oscillate. The results in this paper have suggested a new interpretation for this fact, namely that the vanishing of a certain cluster coefficients is a way of adjusting the static nonlinearity of the map (which is defined by the map clustered polynomial) to the data. Thus a model which has an insignificant coefficient of the cubic cluster but a significant quadratic cluster, is actually adjusting the static nonlinearity to resemble a parabola in the neighborhood of the data.

Static nonlinearities of higher-order maps. It has been shown that if the order of a map is greater than one, a projection-type of the static nonlinearity can be obtained directly from the respective clustered polynomial, which is defined by the term clusters and the cluster coefficients. Thus, most of the results can be applied not only to first-return maps but also to higher-order models.

Finally, although the results discussed in this paper concern maps for which the nonlinearities can be modeled as static, it is believed that some of the ideas and concepts can be extended to other situations.
Acknowledgements

This work has been partially financed by CNPq (Brazil) under grant #351054/95-2. The Chua’s spiral attractor data have been kindly provided by Giovani G. Rodrigues.

References