Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 228 (2004) 421–430
Journal of Theoretical Biology
www.elsevier.com/locate/jtbi

Scalar modeling and analysis of a 3D biochemical reaction model

Jean Maquet[a], Christophe Letellier[a,*], Luis A. Aguirre[b]

[a] Université and INSA de Rouen, CORIA - UMR 6614, Avenue de, l'Université - BP 12, F-76801 Saint-Etienne du Rouvray cedex, France
[b] Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Belo Horizonte 31270-901, Brazil

Received 26 February 2003; received in revised form 1 September 2003; accepted 4 February 2004

Abstract

For many systems it is advantageous if analysis and modeling can be accomplished from a scalar time series because this greatly facilitates the experimental setup. Moreover, in real-life systems it is hardly true that all the state variables are available for analysis and modeling. Since the late 1980s, techniques have been put forward for building mathematical models from a scalar time series. One of the objectives of this paper is to verify if it is possible to obtain global non-linear models (non-linear differential equations) from scalar time series. Such data are obtained using a model of biochemical reaction with aperiodic (chaotic) oscillations as recently observed in the case of a glycolytic reaction (Nielsen, K., Sorensen, P.G., Hynne, F., 1997. Chaos in glycolysis. J Theor. Biol. 186, 303–306.). The main objective, however, is to investigate which state variable is more convenient for the task in practice. It is shown that observability indices seem to quantify quite well which variable should be preferred as the observable. The validity of the results are established performing rigorous topological analysis on the original system and the obtained models. The influence of noise, always present in experimental time series, on the dynamics underlying such a system is also investigated.

Keywords: Biochemical reaction; Global modeling; Observability

I. Introduction

When a biological, or more precisely a biochemical reaction is investigated, the complete knowledge of the states of the reaction may require the measurements of concentrations of all the interacting chemical species. To accurately measure the time evolution of a single species is already a great challenge and usually recording the entire set of state variables is not practically viable. Fortunately, to record a single species is often not a problem. One of the most interesting results in non-linear dynamical system theory is that the time evolution of a single species is sufficient to reconstruct a phase portrait equivalent to the original one provided the dimension of the reconstructed space is sufficiently high (Packard et al., 1980; Takens, 1981). Thus, it is possible to characterize precisely the nature of the dynamics and even to obtain global models directly from a single time series. See Gouesbet and Maquet (1992), Gouesbet and Letellier (1994) for examples using differential equations and Aguirre and Billings (1995) for examples using difference equations.

As exemplified by the Takens' theorem, it is usually believed that dynamical analyses or global modeling techniques may be used with the same ease, regardless of the observable chosen. In a number of practical situations, however, the choice of the observable, that is, the physical quantity which is recorded, does matter and does have a bearing on our ability to extract dynamical information from the reconstructed attractor (Letellier et al., 1998; Letellier and Aguirre, 2002). In the case of a biochemical reaction that involves a few species, the quality of the dynamical analysis may crucially depend on the species concentration chosen to be recorded.

To shed light on this relevant—though frequently overlooked issue—, this paper will discuss some results obtained during a numerical investigation of a three-variable biochemical prototype involving two enzymes with autocatalytic regulation proposed by Decroly and Goldbeter (1982).

In order to investigate this issue, observability indices will be computed using scalar time series obtained by

*Corresponding author. Universite et INSA de Rouen, CORIA UMR 6614, Place Emil Blondel, 76821 Mont Saint—Aignan Cedex, France. Tel.: +33-2-35-146557; fax: +33-2-35-708384.
E-mail address: christophe.letellier@coria.fr (C. Letellier).

0022-5193/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2004.02.004

observing the original model via one state variable at a time. This procedure will rank the state variables indicating which is the most suitable one to be used in modeling and analysis based on scalar time series. To validate such a ranking, a detailed study will be performed in which global models will be obtained from the individual scalar time series. The dynamical performance—quantified by a detailed topological analysis—of each model and the ease of modeling—quantified by the modeling parameters—are used to check on the consistency of the results obtained using the observability indices. Finally, the influence of multiplicative noise is investigated. It turns out that the dynamical variable for which the observability index is the greater, enables building very good dynamical global models and provides greater robustness against noise.

Although the modeling and topological analysis performed is rather involved, the calculation of the observability indices is quite straightforward. This means that such indices can be used in many other systems for which a state-space model is available to determine the best working variable before actually performing expensive experiments and going through time-consuming modeling and analysis procedures.

The remainder of the paper is organized as follows. In Section 2 some background material is provided. In particular, in Section 2.1, observability indices are reviewed and in Section 2.2, the problem of global modeling will be addressed. In Section 3.1, the biochemical model will be briefly introduced. The calculation of the specific observability indices for this model is provided in Section 3.2 and the modeling results are presented in Section 3.3 for the noise-free case and in Section 3.4 for the noisy case. Such results are discussed in Section 4 and, finally, Section 5 gives a conclusion.

2. Background

This section quickly covers background material. For the sake of space, all the details pertaining to the tools used in this investigation cannot be given in this section. The interested reader is referred to the cited literature.

2.1. Quantification of the observability

In spite of the general view, the different dynamical variables of a given system do not permit the same *practical*, information on the dynamics. As a consequence, the quality of the analysis may be affected by the choice of the variable used. This has been explicitly shown for the Rössler system (Letellier et al., 1998), a three-species food-chain model (Letellier et al., 2002) and other various cases (Letellier and Aguirre, 2002). When a dynamical variable is adequate for representing the dynamics the system is observable from such a variable. In this case, it is possible to build a phase portrait from its recorded values. This portrait, in turn, will allow an accurate characterization and a valid global model can be usually obtained from the data. The observability may be quantified using an index as introduced in Aguirre (1995) and Letellier et al. (1998).

The concept of observability in linear system theory is standard (Kailath, 1980); (Aguirre, 1995). Consider the system

$$\begin{cases} \dot{x} = Ax + Bu, \\ s = Cx, \end{cases} \tag{1}$$

where $x \in R^n$ is the state vector, $s \in R^r$ is the measurement vector, $u \in R^p$ is the input vector and $\{A, B, C\}$ are constant matrices. For a non-linear system, A is the Jacobian matrix of that system, B is the matrix defining the coupling between the system and external signals and C defines the measurement function. In all the cases here investigated, the systems are autonomous, i.e. $B=0$ or $u=0$. Thus system (1) is said to be state observable at time t_f if the initial state $x(0)$ can be uniquely determined from the knowledge of a finite time history of the output $x(\tau)$, $0 \leqslant \tau \leqslant t_f$, since the input $u(\tau) = 0$.

One way of testing whether system (1) is observable is to define the *observability matrix*,

$$Q = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-r} \end{bmatrix}. \tag{2}$$

System (1) is therefore state observable if matrix Q is full rank, that is if $\mathrm{rank}(Q) = n$. This definition is a "yes" or "no" measurement of observability, that is, the system is either observable or not. In practice, however, a system may gradually become unobservable as a parameter is varied or, for non-linear systems, it seems reasonable to suppose that there are regions in phase space that are less observable than others. We quantify the degree of observability with the observability index, defined as

$$\bar{\delta} = \frac{\frac{1}{T}\sum_{t=0}^{T} \lambda_{\min}[QQ^{\mathrm{T}}, x(t)]}{\frac{1}{T}\sum_{t=0}^{T} \lambda_{\max}[QQ^{\mathrm{T}}, x(t)]}, \tag{3}$$

where $\lambda_{\max}\lfloor QQ^{\mathrm{T}}, x(t)\rfloor$ indicates the maximum eigenvalue of matrix QQ^{T} estimated at point $x(t)$ (likewise for $\lambda_{\min}$) and $(\cdot)^{\mathrm{T}}$ indicates the transpose. T is the final time considered and, without loss of generality the initial time was set to be $t=0$. Then $0 \leqslant \delta(x) \leqslant 1$, and the lower bound is reached when the system is unobservable at point x. It should be noticed that index (3) is a type of condition number of the observability matrix. When a

single variable is measured, matrix C becomes a row vector and is directly responsible for any decrease in observability.

For instance, we thus obtain for the Rössler system (Letellier and Aguirre, 2002)

$$\begin{cases} \delta_x = 0.022 \pm 0.014, \\ \delta_x = 0.133 \pm 1.7 \times 10^{-14}, \\ \delta_x = 1.9 \times 10^{-4} \pm 0.024. \end{cases} \quad (4)$$

The variables can be ranked in descending degree of observability according to $y \rhd x \rhd z$. This means that the Rössler system is more observable when the y-variable is recorded than when the z-variable is recorded. In practice, a lot of global models were obtained from the x- or y-variable while they are almost impossible to obtain from the z-variable. This is not a hazard too if the x- and y-variables are most of the time used for a dynamical analysis.

2.2. Global modeling from a single time series

In the late 1980s, pioneering papers as for instance Crutchfield and McNamara (1987) and Cremers and Hübler (1987) have introduced the idea that it is possible to find a set of equations that might reproduce the dynamical behavior directly from a single recorded time series. This yields a so-called global model that hopefully correctly describes the dynamics over the whole phase space.

Let us consider an a priori unknown non-linear dynamical system defined by a set of ordinary differential equations

$$\dot{x} = f_\mu(x) \quad (5)$$

in which $x \in R^m$ and f_μ is the unknown set of functions defining the physical system. The solution vector $x_\mu(t)$ called the state vector describes a trajectory in the phase space. The quantity $u \in R^p$ is the parameter vector with p components, which, for a given time series, is assumed to be constant, that is the dynamics is assumed to be stationary. In such a case, the time t is not included explicitly in the vector field $f_\mu(x)$ and the system is said to be *autonomous*. In real situations, a good estimate of the dimension m is the embedding dimension d_E that can be computed using a false nearest-neighbors method (Cao, 1997).

Our purpose is to obtain a global model for the vector field f_μ from a single time series which is designated by x. The obtained model may then be built by using derivative coordinates based on vectors

$$X_n = \begin{vmatrix} X_1 = x(t_n), \\ X_2 = x^{(1)}(t_n), \\ \vdots \\ X_{d_E} = x^{(d_E-1)}(t_n), \end{vmatrix} \quad (6)$$

where $x^{(i)}(t_n)$ designates the ith derivative of the x-variable at time t_n. Working in such a differential space implies a model of the form

$$\begin{cases} \dot{X}_1 = X_2, \\ \dot{X}_2 = X_3, \\ \vdots \\ \dot{X}_{d_E} = F(X_1 X_2, \ldots, X_{d_E}) \end{cases} \quad (7)$$

in which F is a single unknown model function to estimate. When the model is good enough, a dynamical behavior equivalent to the one observed on the physical system studied is then generated by integrating numerically the obtained model. An important problem is to choose the basis on which the function F_x may be decomposed. Many kinds of functions types (representations) may be appropriate and have been used. For instance, Legendre polynomials are used in Cremers and Hübler (1987) or Gibson et al. (1992), rational functions have been used in Gouesbet and Maquet (1992). It seems to us that using a multivariate polynomials basis as introduced in Gouesbet and Letellier (1994) and Giona et al. (1991) is a more practical approach since a convergence theorem due to Weierstrass (Rice, 1964, 1969) exists to guarantee the existence of a good approximation of any analytical function, at least in the static case, and such a representation avoids some of the numerical problems encountered with rational functions Gouesbet and Maquet, 1992. Even after choosing the polynomial representation, it is still necessary to carefully select which subset of multinomials should be used to compose F, since this choice does strongly influence the model dynamical performance (Aguirre et al., 2001). Finally, the model parameters can be estimated using standard least-squares techniques.

The quality of the model depends on a certain set of parameters, called the *modeling* parameters. Three of them are particularly important. N_c is the number of centers retained for the estimation of the model function F. A center is a point of the measured time series and its $(d_E - 1)$ successive time derivatives. N_p is the number of centers retained per cycle. This is particularly sensitive parameters and, varying this modeling parameters may be sufficient to affect significantly the quality of the model (see Chapter 16 in Maquet et al. (2002)). N_k is the number of monomials used in the polynomial expansion. N_k is directly related to the degree of non-linearity used in the polynomial expansion. A balance between a sufficiently large number N_p of centers per cycle and N_c has to be found because when too many centers are retained, numerical errors prevent us to obtain accurate models. Depending on the complexity of the dynamics, N_p may vary from 10 to 30 with N_c limited to a few hundred, e.g. $N_c = 400$. $N_p$defines the precision with which the local structures are sampled and N_c defines

the statistics over the phase portrait. An important remark is in order. In the case of the present technique, the number N_c has no bearing on the model size, which is defined by the number N_k of parameters, unlike other model representations like radial basis function (RBF) models in which the number of centers and the number of parameters coincide.

3. Numerical results

3.1. The three-variable biochemical mode

The series coupling of two enzymes with autocatalytic regulation permits the construction of a three-variable biochemical prototype containing two instability-generating mechanisms (Decroly and Goldbeter, 1982). The substrate S is introduced at a constant rate into the system; this substrate is transformed by enzyme E_1 into product P_1, which serves as substrate for a second enzyme E_2 that transforms P_1 into P_2. Both allosteric enzymes are activated by their reaction product; P_1 and P_2 are thus positive effectors for enzymes E_1 and E_2, respectively. The system is considered as spatially homogeneous as in the case of experiments on glycolytic oscillations. The set of three ordinary differential equations thus reads as

$$\begin{cases} \dot{x} = V - \sigma_1\phi_1(x,y), \\ \dot{y} = q_1\sigma_1\phi_1(x,y) - \sigma_2\phi_2(x,y), \\ \dot{z} = q_2\sigma_2\phi_2(x,y) - K_s z, \end{cases} \tag{8}$$

where x, y, and z are (dimensionless) normalized concentrations of substrate S and of the reaction product P_1 and P_2, respectively. σ_1 and σ_2 are the normalized maximum rates of the enzyme E_1 and E_2. q_1 and q_2 quantify the ratios of the dissociation constants. When $q_1 > 1$ ($q_1 < 1$), the product P_1 varies faster (slower) than the substrate S. A similar feature between the dissociation constants of both reaction products is quantified by q_2. V denotes the substrate injection rate and K_s the apparent first-order rate constant for the removal of the final product in a reaction catalyzed by a Michaelian enzyme far from saturation by its substrate. The rate functions ϕ_1 and ϕ_2 of the allosteric enzymes E_1 and E_2 are given by

$$\phi_1(x,y) = \frac{x(1+x)(1+y)^2}{L_1 + (1+x)^2(1+y)^2} \tag{9}$$

and

$$\phi_2(x,y) = \frac{y(1+z)^2}{L_2 + (1+z)^2}. \tag{10}$$

For the sake of simplicity, the rate of enzyme E_2 depends in a linear manner on the concentration y of its substrate, that is the enzyme is never saturated by it. The model parameters are fixed to

$$\sigma_1 = 10.0\,\mathrm{s}^{-1}, \quad q_1 = 50.0, \quad L_1 = 5.0 \times 10^8,$$
$$\sigma_2 = 10.0\,\mathrm{s}^{-1}, \quad q_2 = 0.02, \quad L_2 = 100$$

with $K_s = 2.00$. It has been shown that this model goes through a period-doubling cascade when the parameter K_s is increased (Goldbeter, 1996). Beyond the accumulation point, the asymptotic behavior settles down onto a chaotic attractor (Fig. 1a). A topological analysis of such attractor was carried out by Letellier (2002).

When the control parameter is increased ($K_s = 2.024$), the dynamics develop and bursting is observed (Goldbeter, 1996). Such behaviors are characterized by large amplitude oscillations appearing on the x-variable time series (Fig. 2a) with a large period. These large amplitude oscillations present small period when they are investigated from the two other dynamical variables (Fig. 2b and c). These large amplitude bursts appear between two phases during which the oscillations typically correspond to the segment of trajectories visiting the population of unstable periodic orbits embedded within the attractor shown in Fig. 1a. This is a rather complicated case to investigate because two different time scales are present. Note that the global shape of the y- and z-time series are similar to those

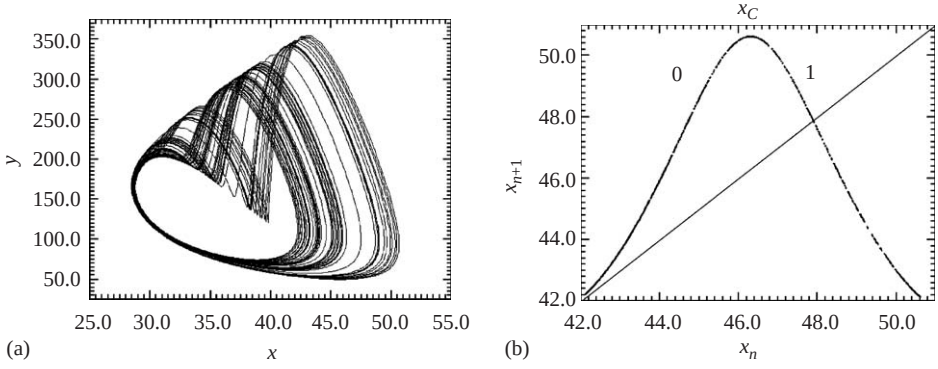

Fig. 1. Chaotic attractor solution of the three-variable biochemical model with $V = 0.45\,\mathrm{s}^{-1}$ and $K_s = 2.00$. The first-return map to the Poincaré section P is unimodal, that is constituted by two monotonic branches separated by a critical point. This critical point is located at a differentiable maximum (a) Chaotic attractor (b) First-return map.

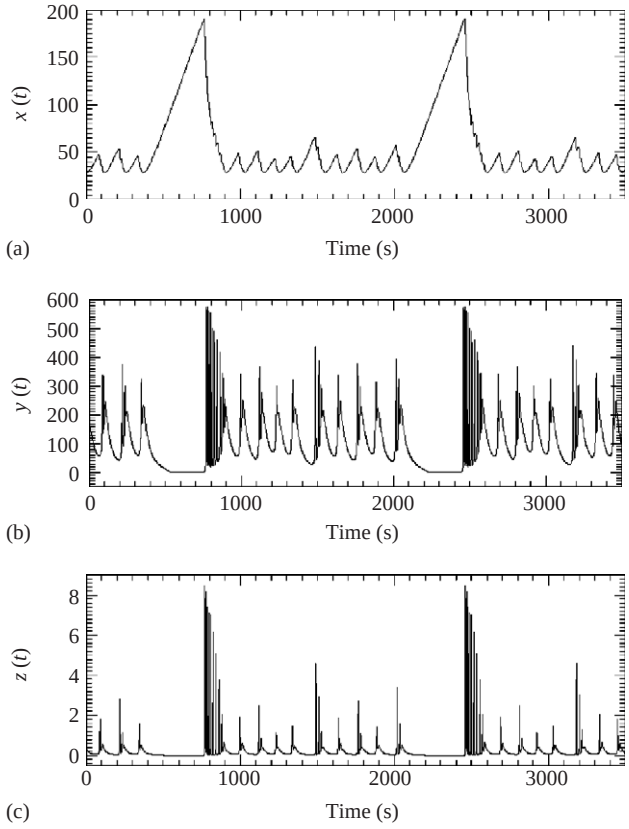

Fig. 2. Time series for the three variables of the biochemical model with $K_s = 2.024$. Large amplitude bursts are apparent (a) x-variable (b) y-variable and (c) z-variable.

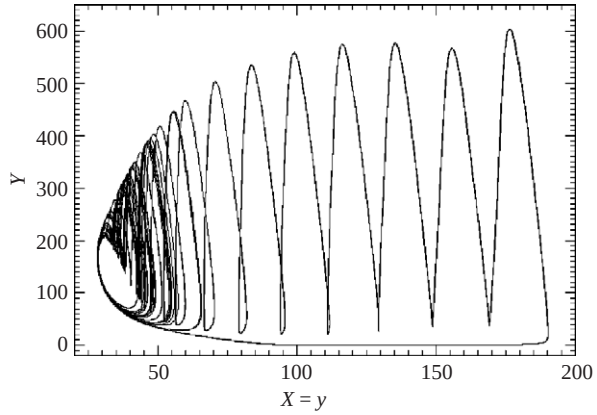

Fig. 3. Chaotic attractor with bursting generated by the biochemical model (8) integrated with $K_s = 2.024$.

generated by the neuron model proposed by Hindmarsh and Rose (1984). The chaotic attractor for which bursting is observed is shown in Fig. 3.

3.2. Observability of a biochemical model

The average observability indices are computed for the three variables of the biochemical model (8). They are equal to

$$\begin{cases} \delta_x = 8.2682 \times 10^{-11}, \\ \delta_y = 5.4648 \times 10^{-20}, \\ \delta_z = 4.2758 \times 10^{-15} \end{cases} \tag{11}$$

which means that the three variables may be ranked as

$$x \rhd z \rhd y \tag{12}$$

as far as observability is concerned. This order means that the dynamical variable x provides a better observability than the dynamical variable z or y. In other words, from these observability indices, it may be concluded that it should be better to measure the concentration of the substrate S rather than the concentration of products P_1 and P_2.

In what follows the observability of model (8) will be investigated in a more pragmatic way. Instead of computing indices, global models will be sought using, successively, the three state variables. Previous experience has shown that global modeling is sensitive to the choice of the observable and that using time series recorded from variables with relatively high observability usually lead to better models. Therefore, having computed the observability indices in this section, it is desired to find out how adequate are variables x, y and z for the purposes of modeling and to see if such a conclusion is consistent with the values of δ_x, δ_y and δ_z.

3.3. Modeling from noise-free time series

A global model is successively attempted from a time series of each dynamical variable. The modified algorithm developed by Cao (1997) for the false nearest-neighbor technique (Abarbanel and Kennel, 1993), yields an estimate for the embedding dimension equal to 3 irrespective of which dynamical variable is used (Letellier, 2002). Thus, in what follows only 3D models will be considered. The time series used for the estimation of the global models are generated by a numerical integration of the three-variable biochemical model (8) with a time step $\delta t = 0.02$ s. The best models were obtained for the modeling parameters as follows:

- $(N_c, N_p, N_k) = (396, 22, 56)$ for the x-variable,
- $(N_c, N_p, N_k) = (526, 60, 56)$ for the y-variable,
- $(N_c, N_p, N_k) = (400, 21, 56)$ for the z-variable.

These modeling parameters are almost the same with the exception of N_c and N_p for the y-variable. It is clearly more difficult to obtain a global model from the y-variable than from the two other variables. More centers are needed and, more significantly, much more centers per cycle must be retained for obtaining a model. Indeed, 60 centers per cycle, which is significantly more than the usual value, must be retained. Also, even if the models constructed from the x- and z-time series have very similar modeling parameters, a quick look at the

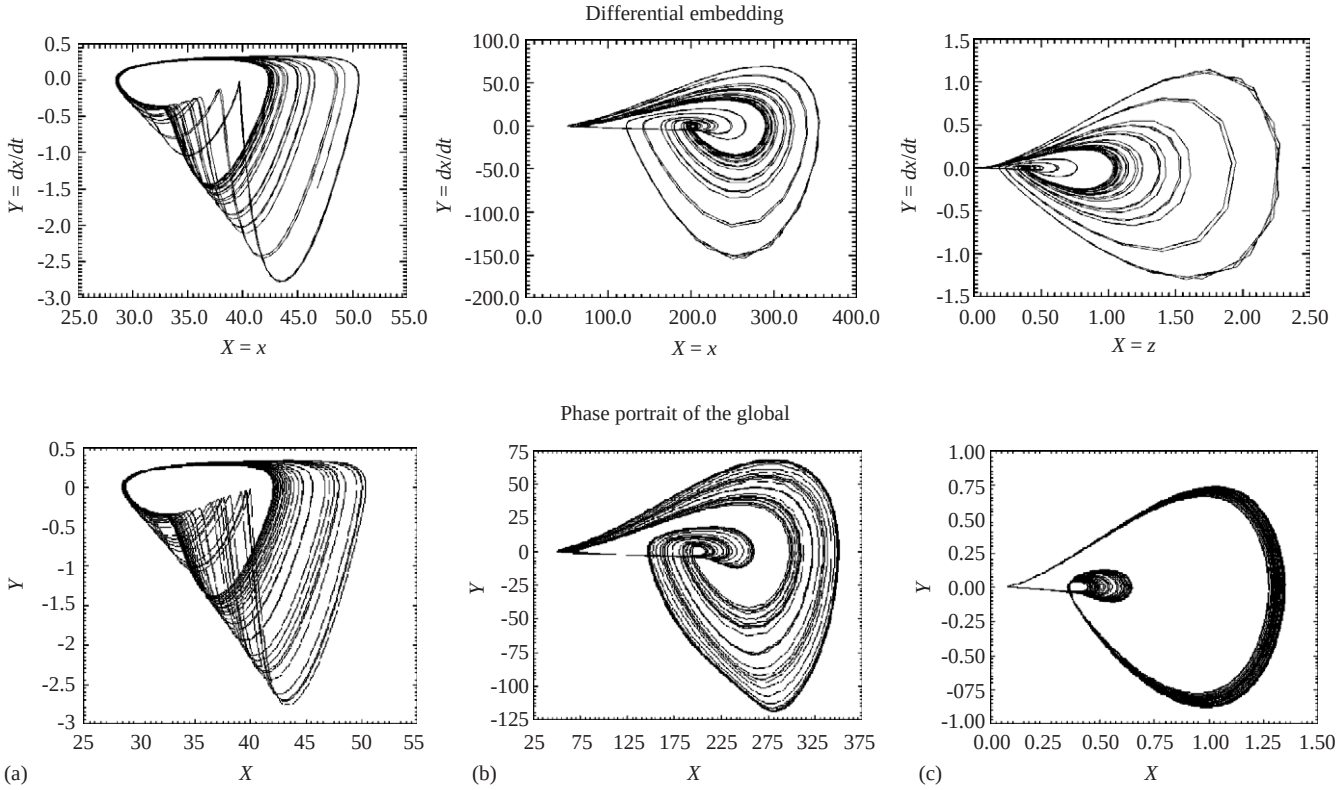

Fig. 4. Plane projections of the differential embeddings induced by the three variables of the biochemical model (8) and the portraits generated by the corresponding global models (a) x-variable (b) y-variable and (c) z-variable.

respective phase portraits shown in Fig. 4a and 4c clearly reveals the superiority of the model obtained from the x-time series, as it would be expected from the observability indices estimated in the previous section. There is therefore a connection between the observability indices and the ease with which a global model can be build from the respective observed time series.

The first-return maps to Poincaré sections for the original system and the global models are shown in Fig. 5. Such maps provide a means for detailed topological analysis by means of comparison of the population of periodic orbits, shown in Table 1.

3.4. Modeling from noisy time series

When real biochemical reactions are investigated, data are always contaminated by noise and one may expect that the non-equivalence between the dynamical variables will be further emphasized. In other words, it might be expected that, in general, the presence of noise will decrease the global observability of the dynamics and the correct choice of which variable to record becomes even more crucial. Moreover, it was observed that noise destabilizes limit cycles thus inducing chaos (Crutchfiled and Farmer, 1982) and that the parameter space may be changed under noise contamination (Ling and Lücke, 1986). The beginning of the bifurcation diagram (low-noise level) computed versus the noise standard deviation (Fig. 6) is quite similar to the one which is computed for the biochemical reaction (8) versus the control parameter. Increasing the noise level increases the population of periodic orbits embedded within the attractor (Fig. 3).

We would like to know what is the affect of the noise on the dynamics and its observability. To this end multiplicative noise was added in numerical simulations. This was accomplished by adding a random variable to each dynamical variable of the biochemical model (8). The standard deviations of these random variables are 0.001, 0.002 and 0.05, respectively. Even with such a small amount of noise, the dynamics is already significantly affected (Fig. 6). In particular, the bursting dynamics, observed for $K_s = 2.024$ on the noise free model (Fig. 3b), can be induced with noise ($\sigma = 0.002$) even for $K_s = 2.000$. A phase portrait (Fig. 7b) looks very similar to the attractor generated by the biochemical model (8) without any noise (Fig. 3).

With these noisy dynamics, no satisfactory global model from the y- and z-variables were found. Increasing the embedding dimension did not help. Such a lack of success in modeling results as a consequence of the strong blurring (by the noise) of the high-frequency

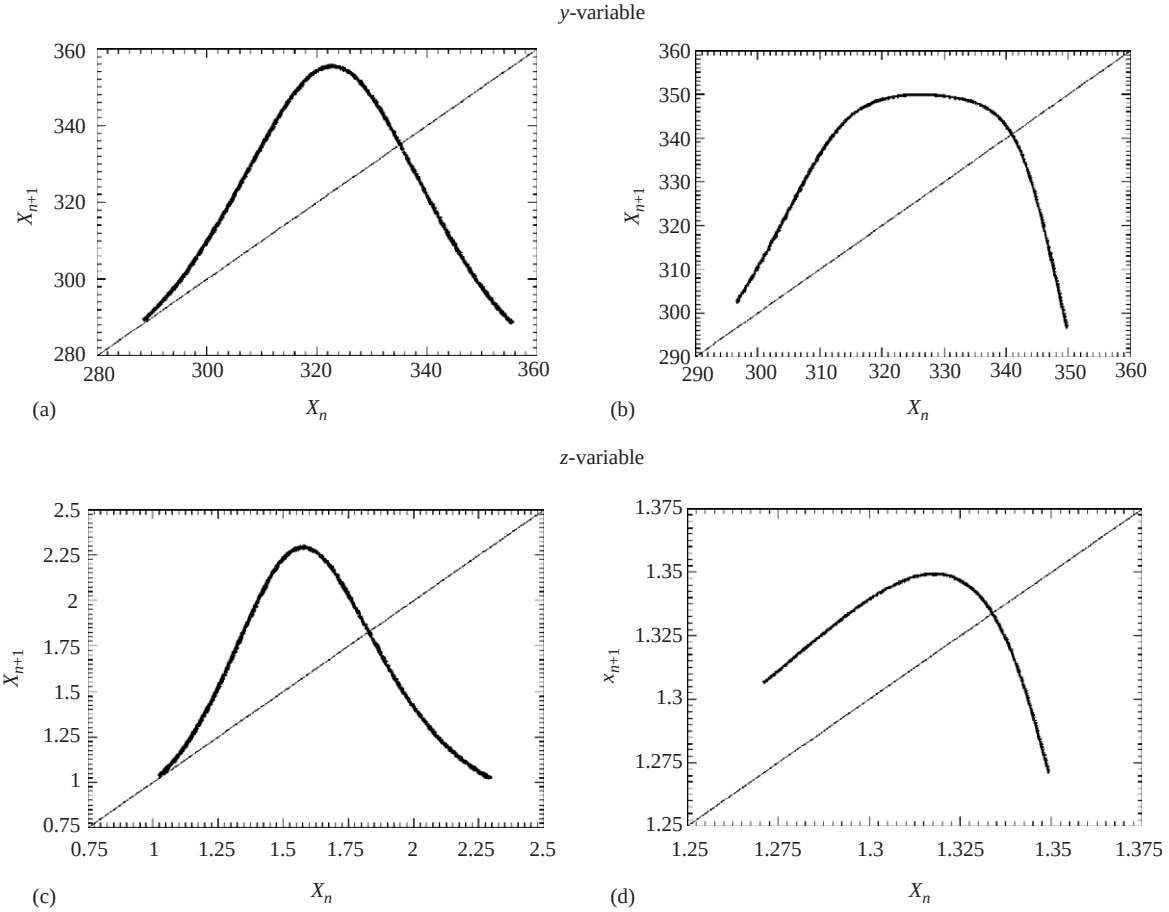

Fig. 5. First-return maps computed for global models obtained from the different variables of the biochemical model. Maps for the case of the x-variable (model and original) are not shown since no significant difference was observed (a) Differential embedding (b) Global model (c) Differential embedding (d) Global model.

Table 1
Population of periodic orbits extracted from the global models estimated from the different variable of the biochemical model

Sequence	x-variable	y-variable	z-variable
0	•	•	
1	•	•	•
10	•	•	•
1011	•	•	•
101110	•	•	•
101111	•	•	•
10111	•	•	•
10110	•	•	•
101	•	•	•
100	•	•	•
100101	•	•	•
10010	•	•	•
10011	•	•	•
100111	•	•	
100110	•	•	
1001	•	•	
1000	•	•	
1000010	•	•	
1000011	•	•	
10001	•	•	
10000	•	•	
100001	•		
100000	•		

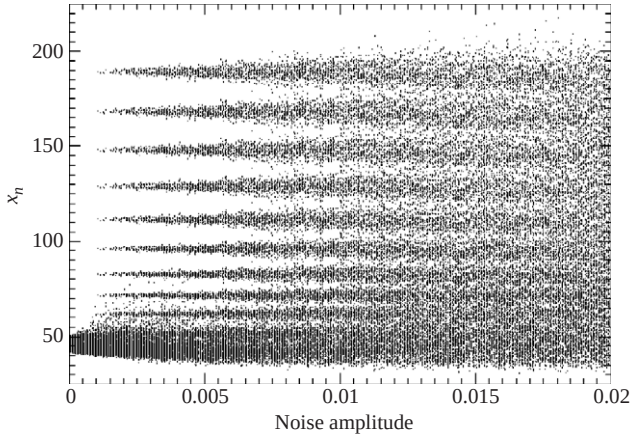

Fig. 6. Bifurcation diagram versus the noise standard deviation for the biochemical model with a multiplicative noise, that is a random variable is added to each of the three variables of the biochemical model at each time step δt ($K_s = 2.000$).

oscillations present in the time series of these variables (see Fig 2b and c).

Nevertheless, when the substrate concentration x is measured, it is possible to obtain a model, even with a

noise level as large as $\sigma = 0.05$. The best model was obtained for $(N_c, N_p, N_k) = (899, 200, 20)$. The number of centers per cycle has been increased significantly for being able to obtain a good global model. Moreover, the number of monomials retained in the polynomial expansion has been reduced to avoid the cumulation of numerical errors. The model from the biochemical reaction is therefore more simple from the noisy x-time series than from a noise free x-time series. Nevertheless, the quality of the modeled dynamics (Fig. 8b) is not as good as observed from noise-free data. Such a feature is quite expected and explains why it is much more difficult to obtain a satisfactory global model from experimental time series. The best model generates a phase portrait (Fig. 8b) which looks like the differential embedding built from the measurements of the x-variable (Fig. 8a). As usually observed, the corresponding dynamics is less developed than the dynamics of the noisy data. Such a feature may be easily explained by the fact that the global model is purely deterministic and, consequently, does not capture the stochastic component corresponding to the noise.

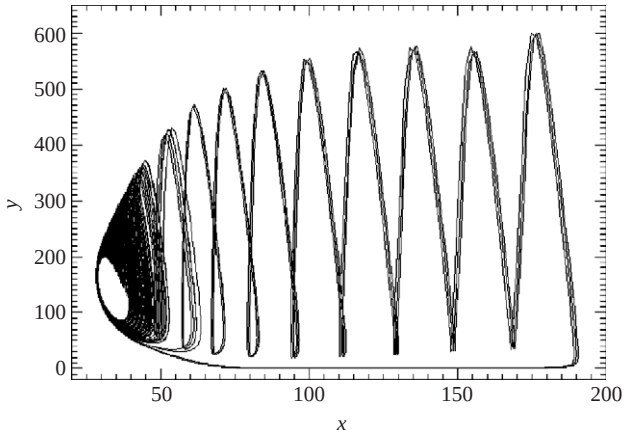

Fig. 7. Attractor generated by the biochemical model for a multiplicative noise with standard deviation $\sigma = 0.002$ and $K_s = 2.000$.

4. Discussion

4.1. Observability

It is important to notice that the computed indices do not convey any "absolute" measure of observability. On the contrary, they only provide relative information, that is, observability indices can be used for comparing variables involved in a given system. Also, this is the reason why symmetry—that could turn out to be quite harmful to the observability depending on the variable chosen—does not have any bearing on these indices (Letellier and Aguirre, 2002).

The values of the indices have very different orders of magnitude which suggest that the observability of the dynamics strongly depend on the choice of the observable. Such a feature could result from the length of the segment along which different cycles are quite difficult to distinguish on the differential embedding induced by the y- and z-variables. As it has been observed for the z-variable of the Rössler system, there is quite a significant time interval during which the amplitude of variables y and z are nearly constant while the amplitude of the x-variable evolves significantly (Fig. 2). Thus, when the y and z variables are recorded, the dynamics cannot be efficiently observed.

Recently, it was also shown that the observability of a dynamics from a given variable also depends on the nature of the coupling between the dynamical variables (Letellier and Aguirre, 2002). In particular, the more complex the analytical expression of the model function F, the less observable the dynamics. Consequently, since the y-time series (Fig. 2b) may even seem more favorable than the z-time series, the coupling between the dynamical variables is much more complicated when the dynamics is observed from the y-variable than from the z-variable.

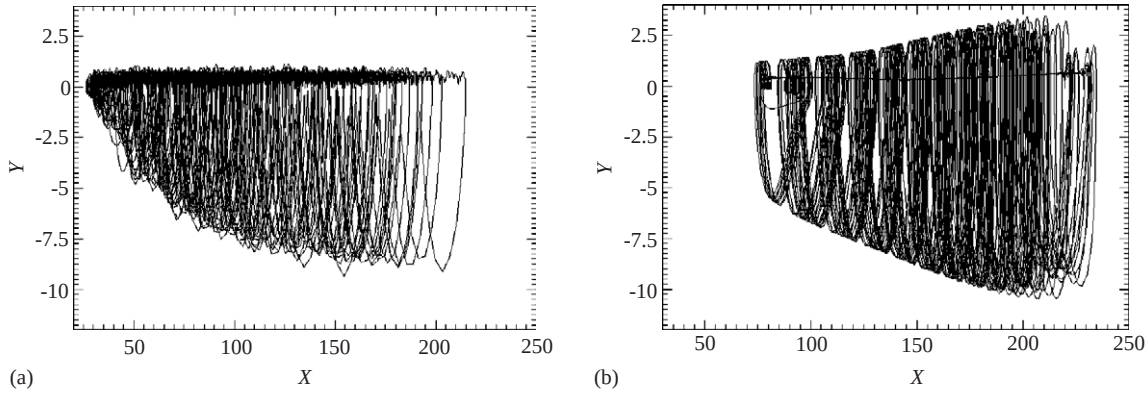

Fig. 8. Projections of the phase portraits reconstructed from the integration of the biochemical model with a multiplicative noise ($K_s = 2.000$ and $\sigma = 0.05$) and generated by the global model (a) Reconstructed attractor (b) Model attractor.

4.2. Modeling

The phase portraits generated by the three global models are shown in Fig. 4, where they can be compared to the differential embedding built from the three variables of the biochemical model (8). At first sight, the phase portrait generated by the global model estimated from the z-variable is the poorest since only a very small chaotic band is obtained (Fig. 4c). The phase portrait generated by the global model from the y-variable is just slightly less developed than the original dynamics and the one generated by the global model from the x-variable is qualitatively equivalent to the differential embedding induced by the same variable. These differences between the target dynamics and the models can also be appreciated comparing the first-return maps (Fig. 5). In particular, the shape of the first-return map computed for the global model estimated from the y-variable (Fig. 5b) is quite different from the one for the first-return map of the differential embedding (Fig. 5a). Nevertheless, the topological structure is preserved: one increasing branch and one decreasing branch are separated by a critical point located at a differential maximum. Moreover, the template corresponding to the phase portrait is the same as the one extracted from the phase portrait of the biochemical reaction (8) (Letellier and Aguirre, 2002). Small differences may be observed when the population of periodic orbits embedded within the attractors are compared. The population of the original attractor is reported in Table 1 with the populations for each global model. In the case of the global model estimated from the y-variable, only the last saddle-node bifurcation is missing. This results in the lack of the pair of period-6 orbits encoded (100001)–(100000). This is directly related to the fact that the increasing branch is slightly less developed.

For the z-variable, the increasing branch of the model first-return map (Fig. 5d) is significantly less developed than for the first-return map associated with the differential embedding (Fig. 5c). A clear signature of such a feature appears in the population of periodic orbits embedded within the attractor generated by the global model. No orbits appearing after the saddle-node bifurcation inducing the pair of period-5 orbits encoded by (10010)–(10011) are identified (Table 1), while the last saddle-node bifurcation involved in the biochemical reaction (8) for $K_s = 2.000$, induces the pair of period-6 orbits, respectively, encoded by (100001)–(100000). This means that, according to the unimodal forcing order (Letellier et al., 1995), one of the control parameter is not properly estimated in the global model. This happens quite often when a global model is attempted (see, for instance, Letellier et al. (1995)). Increasing the number of centers per cycles does not help to improve the quality of the model as it was required for the y-variable. The different shape of the first-return map for the model built from the z-variable is the responsible for the numbers of periodic orbits embedded within the model attractor and the original system (8) being so different. The dynamics is significantly less developed for this model.

The best model is obviously obtained from the x-variable. It has the same population of periodic orbits as the one of the differential embedding which is reported in Table 1. The dynamical performance of the two global models estimated from the y- and the z-variables is much poorer as a direct result of the poor observability of the dynamics from these variables when compared to the x-variable. Nevertheless, one has to note that the worst global model is the one estimated from the z-variable and not the model estimated from the y-variable as it was expected from the values of the observability indices. In this case, the fact that $\delta_y < \delta_z$ seems to be reflected by the fact that significantly larger values for N_c and N_p are required for the y-variable, most likely as a way of compensating for the poorer observability.

The results of this section, therefore, seem to support that the observability order (12) is consistent. Thus, when a biochemical reaction is to be investigated from a single time series, this study suggests that the best strategy would be to measure the concentration of the substrate instead of, as it could be believed, the last product of the reaction.

5. Conclusion

A three-variable biochemical model involving two enzymes with autocatalytic regulation was investigated based on scalar time-series. Such data were obtained by taking the substrate concentration and the two enzyme concentrations one at a time. It has been shown that the reaction modeling is sensitive to the choice of the observable, that is, to the choice of the concentration which is measured. Most likely, this sensitivity will also be observed in different attempts to analyse the system by other means. Such sensitivity is significantly increased in the more realistic case where the dynamics is contaminated by noise. From noisy data, global models were obtained only from substrate concentration measurements. From measurements of the two other concentrations, no global model was found. In other words, the deterministic nature of the underlying dynamics cannot be identified from the measurements. In any case, the dynamics generated by the global model obtained from the noisy x-time series corresponds to the dynamics of the biochemical model but with a slight change in one of the parameters. In other words, the model obtained corresponds to a close neighbor of the original dynamics in the space of models. This shift

due to noise has been recently observed and reported as being a common phenomenon in global modeling (Aguirre et al., 2002).

The observability indices employed in this paper can be easily computed for any system for which a rough state-space model is available. Such indices could then be used to determine the best working variable before actually performing expensive experiments and going through time-consuming modeling and analysis procedures to improve on the state-space model. In this way, it is believed that such indices provide a simple yet efficient and general tool to assist in the investigation of non-linear dynamics in general.

Acknowledgements

This work has been partially supported by CNPq (Brazil) and CNRS (France).

References

Abarbanel, H.D.I., Kennel, M.B., 1993. Local false nearest neighbors and dynamical dimensions from observed chaotic data. Phys. Rev. E 47 (5), 3057–3068.

Aguirre, L.A., 1995. Controllability and observability of linear systems: some non-invariant aspects. IEEE Trans. Edu. 38, 33–39.

Aguirre, L.A., Billings, S.A., 1995. Retrieving dynamical invariants from chaotic data using NARMAX models. Int. J. Bifur. Chaos 5 (2), 449–474.

Aguirre, L.A., Freitas, U.S., Letellier, C., Maquet, J., 2001. Structure selection techniques applied to continuous-time non-linear models. Physica D 158 (1–4), 1–18.

Aguirre, L.A., Maquet, J., Letellier, C., 2002. Induced one-parameter bifurcations. in identified models. Int. J. Bifur. Chaos 12 (1), 135–145.

Cao, L., 1997. Practical method for determining the minimum embedding dimension of a scalar time series. Physica D 110 (1 2), 43–52.

Cremers, J., Hübler, A., 1987. Construction of differential equations from experimental data. Z. Naturforsch. A 42, 797–802.

Crutchfiled, J.P., Farmer, J.D., 1982. Fluctuations and simple chaotic dynamics. Phys. Rep. 92 (2), 45–82.

Crutchfield, J.P., McNamara, B.S., 1987. Equations of motion from a data series. Complex Systems 1, 417–452.

Decroly, O., Goldbeter, A., 1982. Birhythmicity, chaos and other patterns of temporal self-organization in a multiply regulated biochemical system. Proc. Natl Acad. Sci. USA 79, 6917–6921.

Gibson, J.F., Farmer, J.D., Casdagli, M., Eubank, S., 1992. An analytic approach to practical state-space reconstruction. Physica D 57, 1–30.

Giona, M., Lentini, F., Cimagalli, V., 1991. Functional reconstruction and local prediction of chaotic time series. Phys. Rev. A 44 (6), 3496–3502.

Goldbeter, A., 1996. Biochemical Oscillations and Cellular Rhythms. Cambridge University Press, Cambridge.

Gouesbet, G., Maquet, J., 1992. Construction of phenomenological models from numerical scalar time series. Physica D 58, 202–215.

Gouesbet, G., Letellier, C., 1994. Global vector field reconstruction by using a multivariate polynomial L_2-approximation on nets. Phys. Rev. E 49 (6), 4955–4972.

Hindmarsh, J.L., Rose, R.M., 1984. A model for neuronal bursting using three coupled first-order differential equations. Proc. R. Soc. London B 221, 87–102.

Kailath, T., 1980. Linear Systems. Prentice-Hall, Englewood Cliffs, NJ.

Letellier, C., 2002. Topological analysis of chaos in a three-variable biochemical model. Acta Biotheor. 50 (1), 1–13.

Letellier, C., Aguirre, L., 2002. Investigating non-linear dynamics from time series: the influence of symmetries and the choice of observables. Chaos 12, 549–558.

Letellier, C., Dutertre, P., Maheu, B., 1995. Unstable periodic orbits and templates of the rössler system: toward a systematic topological characterization. Chaos 5 (1), 271–282.

Letellier, C., Le Sceller, L., Dutertre, P., Gouesbet, G., Fei, Z., Hudson, J.L., 1995. Topological characterization and global vector field reconstruction from an experimental electrochemical system. J. Phys. Chem. 99, 7016–7027.

Letellier, C., Maquet, J., Le Sceller, L., Gouesbet, G., Aguirre, L.A., 1998. On the non-equivalence of observables in phase space reconstructions from recorded time series. J. Phys. A 31, 7913–7927.

Letellier, C., Aguirre, L.A., Maquet, J., Aziz-Alaoui, A., 2002. Should all the species of a food chain be counted to investigate the global dynamics? Chaos, Solitons Fractals 13, 1099–1113.

Ling, S.J., Lücke, M., 1986. Additive and multiplicative noise on the bifurcation of the logistic map. Phys. Rev. A 33 (4), 2694–2703.

Maquet, J., Letellier, C., Gouesbet, G., 2002. Global modelling and differential embedding. In: Soofi, A., Cao, L. (Eds.), Modeling and Forecasting Financial Data: Techniques of Non-linear Dynamics. Kluwer Academic Publisher, New York.

Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S., 1980. Geometry from a time series. Phys. Rev. Lett. 45 (9), 712–716.

Rice, J.R., 1964,1969. The Approximation of Functions, Vol. 1; Vol. 2. Addison-Wesley, Reading, MA.

Takens, F., 1981. Detecting strange attractors in turbulence. Lecture Notes Math. 898, 366–381.