



**HAL**  
open science

# Polynomial search and global modeling: two algorithms for modeling chaos

Sylvain Mangiarotti, R. Coudret, Laurent Drapeau, Lionel Jarlan

## ► To cite this version:

Sylvain Mangiarotti, R. Coudret, Laurent Drapeau, Lionel Jarlan. Polynomial search and global modeling: two algorithms for modeling chaos. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2012, 86 (4), pp.046205. 10.1103/PhysRevE.86.046205. ird-01062696

**HAL Id: ird-01062696**

**<https://ird.hal.science/ird-01062696v1>**

Submitted on 10 Sep 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Polynomial search and global modeling – two new algorithms for modeling chaos

Mangiarotti<sup>a,\*</sup> S., Coudret<sup>a</sup> R., Drapeau<sup>a</sup> L. & Jarlan<sup>a</sup> L.

<sup>a</sup> *Centre d'Études Spatiales de la Biosphère, UPS-CNRS- CNES-IRD, Observatoire Midi-Pyrénées, 18 avenue Édouard Belin, 31401 Toulouse, France*

## Abstract:

Global modelling aims to build mathematical models of concise description. *Polynomial Model Search* (PoMoS) and *Global Modelling* (GloMo) are two complementary algorithms (freely downloadable at the following address: [http://www.cesbio.ups-tlse.fr/us/pomos\\_et\\_glomo.html](http://www.cesbio.ups-tlse.fr/us/pomos_et_glomo.html)) designed for the modelling of observed dynamical systems based on a small set of time series. Models considered in these algorithms are based on ordinary differential equations built on a polynomial formulation. More specifically, PoMoS aims at finding polynomial formulations from a given set of 1 to  $N$  time series, whereas GloMo is designed for single time series and aims to identify the parameters for a selected structure. GloMo also provides basic features to visualize integrated trajectories and to characterize their structure when it is simple enough: One allows for drawing the first return map for a chosen Poincaré section in the reconstructed space, another one computes the Lyapunov exponent along the trajectory.

In the present paper, global modelling from single time series is considered. A description of the algorithms is given and three examples are provided. The first example is based on the three variables of the Rössler attractor. The second one comes from an experimental analysis of the copper electrodisolution in phosphoric acid for which a less parsimonious global model was obtained in a previous study. The third example is an exploratory case and concerns the cycle of rainfed wheat under semi-arid climatic conditions as observed through a vegetation index derived from a spatial sensor.

---

\* To whom correspondence should be sent :  
sylvain.mangiarotti@ird.fr (S. Mangiarotti)

## 1 Introduction

A proper identification of chaos in nature – that is from real observational data – is a difficult task. Most of the methods that have been developed for its identification consist of checking some of the properties of chaotic dynamics. A chaotic dynamics is characterized by dynamical, geometrical and topological properties: unstable behaviors, fractal structure of the flow, and stretched and squeezed structure, respectively. These properties can be brought out using various concepts and tools. Instability can be estimated from the spectrum of the Lyapunov exponents [1]; geometrical structure can be investigated from the analysis of the correlation integral [2]; topological properties can be deduced from the structure of the embedded trajectory [3-5]. Practically, these properties can provide important arguments for chaos. However, algorithms associated with dynamical and geometrical properties are often very sensitive to noise, which strongly limits their validity when analyzing data stemming from real world or experimental conditions. Moreover, these analyses are not sufficient to reveal chaos. Actually, chaos requires two essential conditions: The first one is an underlying determinism, another one is the high sensitivity to initial condition. The dynamical and geometrical analysis can allow for validating the second condition only. Similarly, topological analysis can lead to a powerful way to characterize the type of chaos and to identify and understand the behavior in terms of trajectory, template, and knot. However, none of them can guarantee the determinism of the ongoing processes.

The question of the determinism is a hard question. Relatively few methodologies have been used to investigate the field. Determinism relates to the uniqueness and causal determination of states from time  $t$  to time  $t + dt$ . As a consequence, the ability to link dynamically the successive infinitesimal states of a system might be a reasonable proof of determinism. Methods based on the one-step prediction, such as the surrogate data analysis [6], have been suggested. This technique is known to exhibit a high robustness to noise. It also performs well at revealing underlying nonlinear processes but cannot be used to determine if a dynamics is deterministic or not [7]. Another approach based on noise titration was introduced in [8]. This method also showed a most efficient ability to investigate the strength of nonlinearities underlying dynamics. However, the approach is not always able to distinguish colored noise from deterministic chaos [9].

Another very interesting approach was introduced in the early 1990s [10]. Instead of investigating determinism from the one-step predictions, this approach consists of verifying that the tangent of the trajectories reconstructed in the phase space is a function of position in phase space. By construction, such an approach can guarantee local determinism in the phase space but not global since it does not allow for checking the coherency of the deterministic behaviors between all consecutive tangents. The global approach aims at finding an algebraic link between the current state of a variable and its current derivatives [11] or between the current state of a variable and its previous visited states [12]. In this sense, it also belongs to the latter approach [10]. However, by requiring the existence of a continuous correspondence between derivatives and the model state, global modeling goes further since guaranteeing the coherency of the dynamics inside the whole reconstruction. The global approach allows for a strong guarantee of the existence of a link between

infinitesimal states. At present, the global approach appears to be the most robust manner to probe and reveal the determinism of an underlying dynamics from experimental or observational time series. Therefore, associated with dynamical, geometrical or topological considerations, it appears as a powerful tool for the identification of chaos from real observations. The approach has been applied successfully to very different types of systems and data sets including, mostly, synthetic time series generated numerically from chaotic attractors [11-13] as well as real measurements gathered from controlled experiments [14-18]. Few examples could be obtained from real world measurements: One global model was obtained in astronomy for sunspot cycles [19]; another one in ecology for the Canadian lynx cycle [20].

Chaos has been one important source of research and development in the last decades. Many packages and libraries dedicated to the analysis of chaos from time series have been developed, among which is the TISEAN software [21]. This software provides a set of pre-existing methods organized into a standardized framework – and published as such – which does not incorporate tools for global modelling. The methodology of global modelling requires specific developments. To our knowledge, there is no software or libraries available at present, that include reliable tools for global modelling of chaos. One objective of this work is to introduce a couple of tools named Polynomial Model Search (PoMoS) and Global Modeling (GloMo) [22] to present their algorithms and their practical originality and to show their efficacy based on specific examples taken from various contexts. The present algorithms were developed as packages under R language [23-24]. The aim of these two algorithms is to investigate global models from time series. Both of them deal with ordinary differential equations of polynomial formulation. The first algorithm, PoMoS, is built to find the optimal terms in the polynomial formulation of the model. The second algorithm, GloMo, estimates the parameterization based on a Gram-Schmidt procedure [11] and also provides basic tools for visualization and analysis.

The paper is built as follows. First, the background of global modelling is briefly recalled in Section 2. A description of PoMoS and GloMo algorithms is given in Section 3. Three examples taken as benchmarks are presented in Section 4: The first example is based on the Rössler model whose variables offer very different levels of difficulty when attempting global modeling. A second example is based on an experimental data series resulting from the electrodisolution of copper in phosphoric acid. The last example relates to the dynamics of rainfed wheat as observed from satellite remote sensing in Morocco. For each system, the ability to retrieve the dynamics from single time series is tested. Results are discussed in Section 5, and compared to previous results when available (a simplified summary of the analytical solutions, and previously and presently obtained models is given in Table 1). Conclusions are presented in the last Section.

## 2 Theoretical background

One powerful result in the theory of nonlinear dynamical systems comes from the possibility to reconstruct a trajectory equivalent to the original set of variables from one single variable. This result has been expressed mathematically in a theorem by Takens [25]. The differential embedding is one of the methods enabling the reconstruction of such a trajectory from one single time series [26]. For a time series

$x(t)$ , this method consists of representing an original time series as a function of its successive derivatives as follows:

$$y = [x, \dot{x}, \ddot{x} \dots x^{(n)}] \quad (1)$$

where  $\dot{x}$ ,  $\ddot{x}$ , and  $x^{(n)}$  are the first, second, and  $n$ th derivatives of  $x(t)$ , respectively. When  $n$  equals to the dimension of the underlying dynamics, the reconstruction is generally complete. If so, the reconstruction is thus called an embedding and the associated dimension is called the embedding dimension  $d_e$ . One classical way to estimate  $d_e$  is the global false neighbors method [27]. However, the true embedding dimension is often difficult to estimate surely [28]. Therefore, the dimension of the global model may also be usefully investigated with a trial-and-error approach.

The differential formulation of a global model directly results from the differential embedding (Eq. 1). A quite general formulation of a dynamical system of  $n$  variables can be expressed as follows:

$$\begin{cases} \dot{x}_1 = f_1(x_1, x_2, \dots, x_n) \\ \dot{x}_2 = f_2(x_1, x_2, \dots, x_n) \\ \dots \\ \dot{x}_n = f_n(x_1, x_2, \dots, x_n) \end{cases} \quad (2)$$

Under invertible properties [29], if observed through  $X_1 = x_i$ , one of the original variable of the system, the system can be reformulated in the following canonical form:

$$\begin{cases} \dot{X}_1 = X_2 \\ \dot{X}_2 = X_3 \\ \dots \\ \dot{X}_n = F(X_1, X_2, \dots, X_n) \end{cases} \quad (3)$$

where function  $F(X_1, X_2, \dots, X_n)$  may be practically approximated by a polynomial  $P(X_1, X_2, \dots, X_n)$  of degree  $q$  since it was shown that a polynomial enables more stable and robust models than a rational [11]. However, it was shown that, in practice, the structure of the polynomial has to be chosen adequately since inappropriate combinations of terms may lead to spurious effects or to numerically unstable solutions [30]. Several classes of structure selection approach have been previously applied in this context. The first and simplest class of technique consists of removing one by one the terms of smaller contribution [31]. The second class of technique (initially developed for the NARMAX modeling) [32] consists of removing clusters of terms presenting compensative contributions [33]. The last class of techniques includes any technique based on a heuristic; one of these, based on a genetic algorithm, was applied to an ansatz model search [29].

The algorithm introduced in the present study for applying the structure selection belongs to the last class of heuristics and was specifically developed for global modelling. Compared to the previously evoked case [29], its originality is to be exclusively based on the binary parameterization of the model's structure related to

the presence/absence of each monomial. In other words, it aims at separating the identification of the model's structure from its precise parameterization. Another original aspect is that it keeps all the solutions visited during the search process in order to allow discriminating good from bad terms *a posteriori*. Finally, the approach was designed with the will to allow the user's dynamical interactions when desired.

### 3 Methodology

The aim of this section is to describe the overall functioning of the two algorithms PoMoS and GloMo which can be used either in a blind mode, or interactively through their associated interface.

#### 3.1 PoMoS algorithm

The PoMoS tool is based on a heuristic algorithm aiming at identifying the monomials of any polynomial ordinary differential equations. Technically, it can be applied to single or multiple time series. In the present case of global modeling considered here, PoMoS is used to identify the terms of one single polynomial  $P(X_1, X_2, \dots, X_n)$  for an optimal formulation of the dynamics as observed from variable  $X_1 = x_i$  (see Section 2). A schematic description of the algorithm is presented in Figure 1. Two baskets of models are used in the algorithm: Basket 1 collects models to be tested, basket 2 gathers the models already tested.

Initially, both baskets are empty: a set of  $N_0$  initial models is randomly generated and put into basket 1. An iterative loop is thus started:  $N_l$  models are picked up from basket 1, their suitability is estimated through a multivariate regression from which goodness of fit is quantified through a chosen criterion. Several criteria are provided, accounting for accuracy and size in a competitive way (their definitions are given in Appendix A). The best model among these  $N_l$  tested models is thus identified and is used to randomly generate  $N_2$  neighboring models. The distance between two models is defined as the number of polynomial terms added and removed. The best model is put back into basket 1 together with the neighboring models. Non optimal models are gathered into basket 2 for further (concurrent or posterior) analysis. The first loop is thus completed. Other loops are iteratively restarted until the number of models in basket 1 becomes lower than  $N_l$ . If so, the convergence of the heuristic algorithm is considered to be achieved. Last models available in basket 1 are thus tested and compared which leads to the 'best' model.

Practically, the algorithm requires three input parameters:  $N_0$  the number of model initially generated;  $N_l$  the number of models to be evaluated at each iterative loop; and  $N_2$  the number of models neighboring the current best model to be added into basket 1 at the end of each iterative loop.

#### 3.2 PoMoS interface

An interface is provided with PoMoS. This interface does not simply make it possible to start and follow the model search. It also provides interactive tools for

analyzing the current results and controlling the heuristic's parameters ( $N_1$  and  $N_2$ ) during the search process.

Two main analysis tools are available during the search process. Each of these aims at identifying current optimal terms. Therefore, these analyses are based on the relation between monomial attendance and the criterions of the models including these monomials. In order to take advantage of all the tests performed during the heuristic process, current analyses are based not only on the set of optimal models but on the whole set of models already evaluated. Considering the subensemble of models containing one chosen polynomial term, the usefulness of this term can be evaluated considering the distribution of criterion values associated to this subensemble. The first assistant tool provides this information by drawing a box plot, for each of the terms considered in the analysis. The second tool is based on a correlation analysis comparing the presence of one given term with the criterion value of the associated model's set. Note that low values of correlation are expected since correlation is applied to a heterogeneous set of data (one binary variable accounting for the presence or absence of one given monomial and one continuous variable accounting for the model fitting).

From our experience, these two tools have proven to be good indicators of what monomial should be kept or preferentially thrown out. The concurrent availability of the tools when running the algorithm is a powerful way for a human to control the heuristic parameters as well as for a more efficient convergence of the algorithm [34]. Indeed, the interface also allows for a dynamical control of the polynomial terms: Monomials can be added or removed; all terms of one generic polynomial defined by its maximal degree can be authorized, or specific selections of terms can be designed. Note that through the interface, this selection can be dynamically modified during the search process.

### 3.3 *GloMo algorithm and interface*

Contrarily to PoMoS, which can be applied also to multiple time series, GloMo is exclusively dedicated to global modeling from single time series. The object of this tool is the parameterization of a polynomial  $P(X_1, X_2, \dots, X_n)$  (see Section 2). The algorithm is built on a Gram-Schmidt procedure, following [11]. Technically, the polynomial formulation in the algorithm is fully generic and allows any model dimension with any polynomial degree.

Due to the huge number of models to explore in a global modeling approach, a search of model structures may advantageously be explored with PoMoS (or with another tool) as a first try. Of course, getting a model structure may not guarantee its reliability. Actually, because PoMoS was developed for selection among huge number of models, the algorithm was designed based on time saving tests. Therefore, practical tests for numerical integrability were not implemented in PoMoS. As a consequence, most often, a fully satisfying structure cannot be directly obtained from the models' structures provided by PoMoS. However, from our experience, when a reliable model is underlying the time series, such a model may be close to the 'best' model obtained with PoMoS. In other words, the best solution provided by PoMoS is often a good start.

In such a context, GloMo will provide an accurate estimate of the parameterization that will be used for testing the practical integrability of the model. In order to control the extension of the structure exploration, an interactive interface

was also designed. In particular, this interface allows adding or removing polynomial terms. After each structural change, a new identification can be performed and its reliability checked in term of numerical integrability. Information about the suitability of the monomials is also provided after parameter identification is run. This is done by removing each monomial one by one and comparing the fittings to the one obtained with all the terms. Note that such suitability can reflect the contribution of a term only in reference to the corresponding structure. Indeed, another structure may lead to different results. This reflects the local validity of the information provided by this test and justifies the prior use of the PoMoS algorithm for an overall (*i.e.* global in the sense used in optimization) view of the model ensemble to be explored.

### 3.4 *Experimental protocol*

No full systematic protocol can be applied when trying to get a global model with the PoMoS and GloMo algorithms from one given time series since interactions with the heuristic process are often very useful during a model search. Indeed, it can be noted retrospectively that, to get a model, the stages of the searching process are different for each case but also that no procedure can be directly exchanged from one case to another. However, although different for each case, some common hints can be identified.

When applying the heuristic, one usual difficulty is to know when to stop the research of the optimal model. This difficulty is closely linked to the impossibility for the criteria to discriminate properly an adequate model from a non robust or from a trivial model. From our experience, Akaike-like criteria generally allow for visiting the model space in a somehow transversal fashion, leading to visit alternatively large regions of non robustness (diverging solutions) and smaller regions of trivialness (fixed points or periodic orbits). Between these two types of region, some chaotic behaviors can sometimes be found. Indeed, the boundary between such clearly characterized regions generally appears fuzzy and possibly complex since behaviours observed there may often alternate discontinuously between nonrobustness, trivialness, nontrivial (*i.e.* multiple) periodic orbits and, potentially, chaos.

One consequence of this observation is that, when searching for a model, the analysis of the successive models visited with the heuristic – applied with human interaction or not – may provide interesting models that the final result might have missed, or lost. This observation can thus be turned into an informal rule when applying PoMoS and GloMo algorithms to time series. To reach a satisfying model, the successive suboptimal models progressively visited by the heuristic can be considered individually by using GloMo and checked one by one for their robustness and non trivialness.

When a global model is retrieved from one observed time series, it is well known that the original dynamics is not always precisely identified. Indeed, the dynamics identified is often shifted in the bifurcation diagram. Among various possible causes of this shift (sampling of the vector field in the phase space, parameters identification process, level of noise, quality of the model, etc.), one is the observability of the system from the measured variable ([35], see Section 4.1). Practically, once a global model is obtained, one simple way to overpass this difficulty consists of tuning one of the model's parameter. This was applied systematically here to the models obtained from the three systems taken as a benchmark in order to evaluate their ability to represent the original dynamics and,



when useful, to discriminate among the various models obtained. Note that the algebraic formulation of all the global models obtained in the present analysis is presented in [36].

### 3.5 Validation

Once a model is obtained, a validation should be performed. Without applying model validation, only the presence of a deterministic component in the underlying dynamics can be reliably claimed. To validate a chaotic model, one efficient approach consists in comparing the first return map obtained from the model (that is the pattern generated by plotting the coordinates  $R_n$  of the  $n$ th visiting points of a Poincaré section as a function of the coordinates  $R_{n+1}$  of the  $n + 1$ st returning visits) to the one reconstructed from the original signal. In practice, this requires to appropriately choose a Poincaré section which is not always an easy task when dealing with noisy data, jittery behaviours, complex structures, etc. For this reason, a precise description of each of the Poincaré sections should be always provided. The definition of all the sections and coordinate systems used in the present analysis is unambiguously provided in Table 2. Several levels of validation can be reached depending on the quality of the data and on the complexity of dynamics. (a) In the best case, when the dynamics is precisely retrieved, the shape of the first return maps (one obtained from the original data set, another one from the global model) should match. If so, an analysis of the branches' correspondence and of the orbits' recurrence may be performed to verify the attractor's structure. Such a procedure will be described hereafter. Note that a comparison of the bifurcation diagrams may also be applied since it is likely to be a very strong validation technique (see [30]). However, such a procedure may be applied only when the original system is known, which is generally not the case with global modeling. (b) If the structure of the global model's attractor is correct but the dynamical regime of the original behaviour poorly identified, the same type of pattern will be obtained from data and model (same number and type of branches) but with less developed dynamics (or eventually more, but from our experience this is more rare). (c) If the topology is partly lost (*e.g.*, less/more branches appear on the first return map, even after tuning the model's parameter), it is most probable that the system of equation retrieved is not equivalent to the original system, but similar only. (d) Finally, if no pattern can be clearly compared, only the presence of a deterministic component underlying the dynamics can be reliably conclusive.

In the best case mentioned above – case (a) corresponding to a quite good matching of the first return maps – a refined level of validation may be performed. Indeed, various degrees of reliability can be distinguished inside this former case. In a chaotic context, the precise validation of a model is considered to be a difficult task [5]. Indeed, due to the sensitivity to the initial conditions, predictions can be performed over relatively short horizons only. As a result, model validation over long time windows becomes more difficult. However, the existence of an attractor requires the dynamic to be ergodic and stationary; therefore, although the trajectory might be unpredictable, the attractor's structure should not vary in time and its statistical properties should remain identical. Validation can thus be based on the statistical properties of the attractor. And to go farther – with the aim to investigate global models' reliability in the long term – validation may be based on the statistical properties of the populations of periodic orbits. Since only an imperfect model can be expected in a real context, one way to quantify global model's reliability may thus

consist in estimating the number of periodic visit of the attractor that can be satisfyingly performed before losing the statistical properties of the original attractor. An iterative procedure is thus established in order to successively check model's reliability after 0, 1, 2, ..  $n$  periodic visits of the attractor. At the 0th iteration step, the comparison is static (it relates to the 1-symbol-length sequences), and thus basically allows for verifying that the modeled manifold has the same number of branch as the original data set. In practice, this step can be performed by comparing the distribution of the orbits. At the 1st iteration step, the comparison allows verifying that the same branch connexions are in presence within the original and the modelled manifolds. This can be checked by comparing the transition matrices (matrix providing the link between one symbol to another, thus relating to a 2-symbol length sequence). At larger iteration steps, the comparison will allow refining the (long-term) validation of the model. For the  $n$ th iteration step ( $n \geq 2$ ), validation will be performed by comparing the correspondence of the  $(n+1)$ -symbol lengths sequences. In practice, such validation procedure can be applied iteratively, increasing the sequence's length one by one until the correspondence between the original and the model's dynamics become incomplete (correspondence is considered to be incomplete as soon as one – or more – additional or missing sequence is found). The last iteration for which correspondence is fully satisfied corresponds to the horizon for which the structure validation can be statistically performed.

Practically, transition matrices can be estimated easily once a partition of the first return map can be performed. The transition matrix aims to characterize here the links existing inside the branched manifold underlying the attractor. It can be defined in two ways. The Markov transition matrix is defined here as the matrix describing the probability to move from one symbol to another, each symbol corresponding to one branch of the manifold, which can be determined from first return map's pattern. The binary transition matrix is a simplified view of the Markov transition matrix that basically provides the existence or the nonexistence of the same link. In practice, the binary transition matrix can be directly deduced from the Markov transition matrix by replacing nonzero values by one and keeping the zero values as so. Note that a more detailed description of the branched manifold may also be performed at this step, by depicting the rotations and the intertwining of the branches. This could be done through a complete topological analysis [3-4] and may allow for fully checking the attractor's structure. However, since based on a qualitative description of the flow's structure, a topological analysis will not enable checking the accuracy of the model over a time period longer than one periodic visit of the attractor. For long-term validation, an approach based on Markov transition matrix should be preferred.

The comparison of the populations of periodic orbits can be performed by comparing the symbols sequence. The following examples are based on an alphabet of four symbols (0, 1, 2, and 3). For the 1-symbol-length sequences, the comparison directly results from the presence – or the absence – of a branch existing in the original dynamics; as an example, the percentage of correspondence will be 100% if all the branches are retrieved, 75% if three branches are retrieved only, and so on. For 2-symbol-long sequences, this correspondence can be directly deduced from the binary transition matrix. An alphabet of four symbols leads to a  $4 \times 4$  terms matrix, if the binary transition matrix of the original's dynamics equals the one corresponding to the global model, then the correspondence is 100%; if one term differs between those matrices, correspondence is 93.75% and so on. For 3-symbol-length sequences and beyond this correspondence can be deduced from the number of common sequences retrieved in both original and modeled dynamics. The longer the symbol sequences

can be retrieved, the more accurate can be considered the model. These analyses were performed for each of the models obtained in this work (except when the partition of the first return map was not possible). The (Markov and binary) transition matrices are all provided in [36]. The boundary limits used to perform the symbols' partition are provided in Table 3 for each original data set and model. The ability of the global models' dynamics to represent the original one is also provided as a percentage of retrieved sequences for sequences of 1-symbol to 6-symbol lengths. In practice, an iteration is considered to be validated when correspondence strictly equals 100%.

## 4 Data

Three applicative cases are considered as benchmarks. The first case is purely theoretic. The second case is derived from measurements made under experimental conditions. The last case is taken from the observation of a real environmental system obtained from satellite remote sensing.

### 4.1 Rössler system

The Rössler system [37] is an interesting case to study since its three variables exhibit very different levels of observability of the underlying dynamics [35]. It is a low-dimensional system defined as follows:

$$\begin{cases} \dot{x}_1 = -x_2 - x_3 \\ \dot{x}_2 = x_1 + ax_2 \\ \dot{x}_3 = b + cx_3 + x_1x_3 \end{cases} \quad (4)$$

Synthetic time series of variables  $x_1$ ,  $x_2$ , and  $x_3$  were generated by integrating these equations with control parameters  $(a, b, c) = (0.520, 2., 4.)$  corresponding to the phase noncoherent regime characterized by fast oscillations around an external fixed point. Resulting time series are plotted in Figure 2. The reformulation of the system in a derivative form (Eq. 3) is obtained from the following transformation:

$$\phi_p = \begin{cases} X_1 = x_i \\ X_2 = f_i(x_1, x_2, x_3) \\ X_3 = \frac{\partial f_i}{\partial x_1} f_1 + \frac{\partial f_i}{\partial x_2} f_2 + \frac{\partial f_i}{\partial x_3} f_3 \end{cases} \quad (5)$$

When  $\phi_p$  is invertible, the exact model function  $F(X_1, X_2, \dots, X_n)$  of the canonical formulation (Eq. 3) can be obtained analytically [29]. For the Rössler system, solutions have been given in [11]. For the variable  $x_2$ , the function is a second degree polynomial, whereas for  $x_1$  and  $x_3$ , functions are rational:

$$F_2(X_1, X_2, X_3) = -b + (a+c)X_3 + cX_1 - (ac+1)X_2 - aX_1^2 - aX_2^2 - aX_1X_3 + X_2X_3 + (a^2+1)X_1X_2 \quad (6)$$

$$F_1(X_1, X_2, X_3) = ab + cX_1 + X_1^2 - aX_1X_2 + X_1X_3 - (ac+1)X_2 + (a+c)X_3 - \frac{bX_2 + X_1X_2 - aX_2^2 + X_2X_3}{a-c-X_1} \quad (7)$$

$$F_3(X_1, X_2, X_3) = b + (c-X_2)X_1 - X_2 + aX_3 + aX_1^2 + \frac{1}{X_1} \left[ (ab+3X_3)X_2 - bX_3 - aX_2^2 \right] + \frac{2X_2^2}{X_1^2} (b-X_2) \quad (8)$$

Level of observability can vary from one variable to another. Basically, a system is fully observable from one given variable if the full state of the system can be retrieved from it. However, in practice, it was shown that observability could vary by degree and an observability coefficient was introduced [35]. Practically, observability can be estimated from the Lie derivatives since it highly depends on the invertibility of  $\phi_p$ . The classification obtained for the Rössler system shows that the variables can be ranked according to the decreasing degree of observability as follows:  $x_2 \triangleright x_1 \triangleright x_3$ . From a modeling point of view, the observability coefficients bring important information about the potential difficulty to obtain a model.

However, this coefficient is only comparative inside one system but cannot be applied to compare the observability of variables between several systems [35]. Therefore, it cannot be practically used when only one single time series is available. Nonetheless, together with other factors such as dynamical perturbations and noisy measurements, level of observability may lead to easy, difficult, or impossible modeling. Therefore, observability is assuredly one important element to keep in mind as far as global modeling is considered. In order to better illustrate the ability of the present tools, the phase noncoherent regime was chosen to run the analyses. The phase noncoherent regime is characterized by a four-branch attractor that includes the presence of two timescales in the dynamics. This makes the conditions much more difficult to address when trying to capture the structure of the global mode and to reproduce the original dynamics satisfyingly.

## 4.2 *Electrodissolution of copper in phosphoric acid*

Many cases of chaotic behaviors have been identified or approached in chemical reactions and electrodisolutions. The case of copper electrodisolution in phosphoric acid is especially interesting here, since experiments were carefully conducted in the 1990s from which a global model was obtained and validated using a topological approach [14]. The time series used in the present work comes from the same experiment. This experiment was designed and studied in John Hudson's group at the University of Virginia [38]. The experimental setup consists in a copper cylinder (8.26 mm in diameter), inserted inside a tube of Teflon (20 mm in diameter) and rotating at a moderate speed of 4400 rotations per minute in an 85.7% (by mass) solution of phosphoric acid. A water bath was used to maintain the temperature at 20°C. The moderate value of rotation speed was justified by the necessity to facilitate the contact between the reactants on the one hand and to avoid turbulences on the other hand [39]. The time evolution of the dissolution current  $I(t)$  was measured

relatively to a reference electrode, which was separated from the solution by a capillary.

The sampling recording was characterized by large-amplitude oscillations of  $\pm 15$  mA (see Figure 3) exhibiting a chaotic phase portrait suggested by its first-return map. A 3-dimensional model was obtained from the data using the global approach [14]. The model function was a fifth-degree polynomial of 52 terms. The object of this revisited analysis is to check the ability of our tool to reach a similar or a more compact model.

### 4.3 *Vegetation index*

The Normalized Difference Vegetation Index (NDVI) product from the Global Inventory Modelling and Mapping Study (GIMMS) of the Global Land Cover Facility is used in this study [40], during the period 1982–2008. It is based on the Advanced Very High Resolution Radiometer sensor onboard National Oceanic Atmospheric Administration satellites. The NDVI is an efficient index for monitoring and modeling of grasslands and grain crops in semiarid regions [41-43]. The product used here has a 10-day sampling and an  $8 \times 8$ -km<sup>2</sup> resolution. It is corrected from sensor degradation over time and also accounts for atmospheric aerosols resulting from the eruptions of the El Chichon (in April 1982) and Mt Pinatubo (in June 1991).

The geographic area of the study is located in northern Morocco. It is defined by a window ranging from 6.2°W to 5.4°W in longitude and from 34.6°N to 35.4°N in latitude. Although this area includes one part of irrigated crops, rainfed production remains the dominant part of the area of study. A selection based on the land cover type has been applied to the data in order to obtain a more representative signal. The GLC2000 land cover map [44] was used for this purpose. Once the selection applied, the signal was spatially aggregated at the scale of the whole province by spatially averaging the data in order to reduce the effect of noise. A Savitzky-Golay method [45] was then applied with a 14-decade window to filter the resulting time series and to compute the successive derivatives with a 3-day resampling. The resulting series is shown in Figure 4.

## 5 Results and discussion

### 5.1 *Rössler system*

The object of this first application to the Rössler system is to show the efficacy of the tool when considering three variables  $x_2$ ,  $x_1$ , and  $x_3$  (see Eq. 4) enabling a decreasing level of observability of the original system. For this reason, the three variables are considered separately.

Variable  $x_2$  provides the best observability of the Rössler system which results from the linear relation between  $\dot{x}_2$  and variables  $(x_1, x_2, x_3)$  and can be deduced from Eqs. (4) and (5) (see [35]). Moreover, this variable also allows for a formally exact and even basic polynomial solution. For this reason, this first analysis can be considered as a potentially prohibitive test since it should absolutely be passed. The

best model obtained (lower Akaike Information Criterion) from  $x_2$  is obtained for maximal polynomial degree  $q = 2$  and corresponds exactly to the polynomial structure  $F_2(X_1, X_2, \dots, X_n)$  of the formal solution (see Eq. 6). Starting the structure search from  $q = 2$  or  $q = 3$ , it can be obtained by using PoMoS or by simply removing terms of smaller contribution one by one. Phase portraits reconstructed from the original data  $(x_2, \dot{x}_2, \ddot{x}_2)$  are shown in Figure 5a together with the simulation obtained from the global models Figures 5c and 5e. Global modeling often produces reliable model structures with imperfect parameterization [12]. In the first case, a 9-term model corresponding to the canonical solution (Eq. 6) was obtained. The development of its dynamical behavior appears fully retrieved as illustrated by the first return maps highlighting a quite good matching (Figures 5d). This is confirmed by the transition matrices (see Section 1.1 in [36]) and by the analysis of the populations of unstable periodic orbits for which symbolic sequences are fully retrieved (100% of the sequences match) until 3-symbol-length sequences (see Table 3). None significant improvement could be obtained by tuning the parameters by hand (not shown). Such improvement may be obtained by developing an optimization approach dedicated to this problem. Surprisingly, a 7-term model, thus smaller than the analytic solution (see Eq. 6), was also obtained (Figure 5e) by removing terms  $b$  and  $X_2$  from  $F_2(X_1, X_2, \dots, X_n)$ , suggesting a compensative contribution of these two terms. However, the phase portrait associated with this model of smaller size appears incomplete, and it should be noted that it was not found possible to adjust the dynamics to the original behavior by tuning the parameters. The transition matrix of this 7-term model is incomplete, leading to a short horizon of confidence corresponding to a 1-symbol length sequence only, which is quite poor. Although poor in quality, this smaller model is interesting in this analysis. Indeed, it shows that an approximate model can lead to a viable attractor (although incompletely reliable) of approximate dynamics characterized by a not fully developed (and not fully developable) dynamics. This does not prevent from obtaining a numerically robust (although approximate) model for the original dynamics. This model of smaller size (7 terms instead of the 9 analytic ones) is also interesting since it shows that, by oversimplifying the canonical system, some behaviors of the original system cannot be retrieved. Indeed, some behaviors provided by the two missing terms are obviously lost. This, also illustrates that the principle of parsimony can apply here, but should be applied with some caution. There is no reason to choose the smaller solution exclusively because it is smaller; but rather, when two solutions exhibit same behaviors, then the smaller one should be preferred since there is no reason to choose a more complicated one. In the present context, the 9-term model should assuredly be preferred. These results also show that a refined validation based on the comparison of the populations of periodic orbits can be a powerful approach.

Variable  $x_1$  provides a moderate observability of the Rössler system. The analytical formulation of  $F_1(X_1, X_2, \dots, X_n)$  also exists for this variable and is rational (see Eq. 6). The moderate level of observability results from the nonlinear relation existing between  $\dot{x}_1$  and variables  $(x_1, x_2, x_3)$  and can be obtained from Eqs. (4) and (5) (see [35]). Since only a polynomial formulation for the approximating function  $F$  is permitted in the present version of the algorithm, an exact formulation of the global model cannot be expected. However, despite the presence of rational terms, approximate functions of small size can be expected, and were obtained with degree 3 and 4 polynomials. Phase portraits reconstructed from the original data set and from

the global models are given in Figures 6a, 6c, 6f and 6h. None of these models allows for an accurate reconstruction of the phase portraits, as confirmed by the first return maps (see Figures 6b, 6d, 6g and 6i). At first sight, the best solution is obtained with the 13-term model (of maximum degree  $q = 4$ ) for which three of the four branches of the first return map are directly and clearly retrieved (see Figure 6d). A slightly less developed dynamics is obtained for the 10-term model (of maximum degree  $q = 3$ , Figure 6i), whereas the 9-term model (of maximum degree  $q = 4$ ) poorly exhibits three branches, the third one being extremely small (Figure 6g, respectively). The models' inaccuracy is also readable in the transition matrices (see Section 1.2 in [36]). By tuning one parameter of each of these three models, it was found that the 10-term model was obviously the best approximation (Figure 6j). Indeed, the full development of the 13-term ( $q = 4$ ) model's dynamic could only be partial and associated with discontinuities in the first return map (Figure 6e); and it was found not possible to tune the 9-term model (not shown). Contrarily, although not fully matching the original first return map, a gentle development of the dynamics was obtained with the 10-term model ( $q = 3$ ), including a small discontinuity only (Figure 6j). Moreover, it was found possible to fully retrieve 2-symbol length sequences with this latter (see Table 3), whereas sequences of only 1-symbol-length could be retrieved with the 13-term model and none with the 9-term model (see Section 1.2 also in [36]).

Finally, variable  $x_3$  provides the lower observability of the Rössler system. This low observability results from the nonlinear relations between  $\dot{x}_3$  and variables  $(x_1, x_2, x_3)$  and can be obtained from Eq. (4) and (5) (see [35]). The analytical formulation of  $F_3(X_1, X_2, \dots, X_n)$  also exists, it is rational (see Eq. 8), and includes degree 2 terms at the denominator. From the differential embedding (Fig. 7a), this variable exhibits a very spiky structure that makes global modelling especially difficult [46]. Several solutions were yet obtained in the past by using an over dimensioned 4D formulation [47], by using an ad hoc approach based on fixed points coordinates [48], or by using a very constraining structure selection with the Ansatz library [46]. However, none direct 3D global model could be obtained yet. This difficult context is further complicated here by the use of a phase noncoherent dynamics.

Despite the low observability and the spiky structure, a fifth degree polynomial approximation of 30 terms was obtained. Phase portraits and first return maps reconstructed from the initial data set and from the model are shown in Figures 7a & 7c. The chaotic behavior of the global model is confirmed by the first return map (Figure 7d). Three branches are obtained which means that a phase noncoherent regime is directly retrieved. However, the full development of the chaos is not obtained since only three branches appear, the third one being partly truncated. These characteristics are directly readable in the transition matrices (see Section 1.3 in [36]). A rather gentle development of the chaos was obtained by tuning one of the parameters (Figure 7e) giving rise to the full development of the four branches of the original data set. As a result, transition matrices could become reliable and even, a full agreement of the symbols sequences is found until 2-symbol length sequences (see Table 3). Some imperfections can be noted, however. A fuzzy behavior is visible in the first return map, on the left side of the pattern, corresponding to the faster oscillations, more difficult to detect. These imperfections are likely to result from the especially acid situation characterized by a spiky structure in a context of low observability and further complicated by the two timescales of the phase noncoherent regime. It is therefore a very convincing result since, at present, no 3-dimensional

model could be obtained yet from this difficult variable of the Rössler system without using strong structure selection.

## 5.2 *Electrodissolution of copper in Phosphoric acid*

The analysis of the time evolution of current intensity measured from the experiment introduced in Section 4.2 has led to a 21-term model of maximum polynomial degree  $q = 4$ . A phase portrait obtained from the measured signal (Figure 8a) can be compared to the portrait obtained from the model Figure 8c. The retrieved model exhibits the same structure. The chaotic behaviour is confirmed in both cases from the double-branch pattern of the first return maps (Figures 8b & 8d). A larger central hole with a slight shift can however be noted in the model, leading to a poor development of the first branch of the first return map, and it was found not possible to significantly develop the dynamics by manually tuning one of the parameter (not shown). The direct consequence of the poor development of the dynamics is easily readable in the transition matrices (see Section 2 in [36]).

Therefore, another solution (21-p.<sup>\*</sup>) obtained by considering a shifted time series of longer length (see Figure 3) was also considered. For this reason, the two models have the same structure and only their parameterization differs. The second model exhibits a limit cycle (see Figure 8e) very close to a period-5 cycle as depicted in the first return map (see Figure 8f). Although less chaotic with its initial parameterization, this other model could be efficiently tuned up, leading to a developed chaos close to the original dynamics (Figure 8g). The transition matrix of the tuned version of this second model (21-p.<sup>\*opt</sup>) is in agreement with the one obtained from the original signal. Further, symbol sequences resulting from this latter model are in agreement with the original sequences until 6-symbol-length sequences.

When only one single variable is available, the observability coefficients of the system cannot be estimated. However, it can be noted that, unlike Rössler- $x_2$  or Rössler- $x_1$  and similar to Rössler- $x_3$ , a model was hard to find for this application, and only a high degree polynomial solution could be obtained ( $q = 4$ ). Moreover, as found with Rössler- $x_3$ , a high sensitivity to the time series' window used to identify the model was noted. Such sensitivity to the time series' size is a common feature of global modeling. This sensitivity results from the identification process which is a difficult task in a context of chaotic behavior. Indeed, when more data is available to identify the parameterization, a more precise parameterization may be expected. Unfortunately, more data will also lead to more complex shapes of the objective function. In the case of a chaotic behavior, the complexity of this shape will rapidly tend to fractal structures that simple identification approaches may not be able to overtake optimally. The difficulties in obtaining a satisfying model are likely to arise from the complex shape of the objective function complicated here by the intermediate or poor level of observability of the system as considered from the measurement of current intensity. This assumption may be investigated interestingly by reconsidering the dynamics in another experimental configuration which would involve several electrodes simultaneously, for comparison; or by considering other variables of the system.



### 5.3 Dynamics of rainfed wheat

The time evolution of the vegetation index (Figure 4) introduced in Section 4.3 has been analyzed, leading to one 15-term model. The phase portraits obtained from the original data set and from the retrieved model are shown in Figures 9a and 9b. Some differences are noticeable that include the filtering of short timescale events observed in the original portrait that may result from transitory effects associated with meteorological perturbations or to spatially differentiated behaviors of the region of study. Another emphasized difference comes from the higher density of trajectories observed in the center hole of the simulated trajectory, whereas the original phase portrait clearly exhibits an empty hole. Such differences may result from a suboptimal parameterization. Indeed, it is well known that global modeling may lead to a good structure (*i.e.* equivalent to the original one) but to an inaccurate identification of the regime: The attractor is generally not located exactly at the same place in a bifurcation diagram of the original system [12]. Despite the differences noticed, obtaining such a model is quite an important point since it clearly shows the robustness of the algorithm to get a satisfactory model from a rather jittery time series. It is also a quite important result from a thematic point of view since it brings a strong evidence of a deterministic component underlying the signal of rainfed wheat observed from space. More generally, to our knowledge, there was previously no such strong element of evidence of deterministic behavior underlying the dynamics of vegetation.

The resulting attractor exhibits a complex structure, for which no simple first return map is accessible. Its unstable behavior is confirmed by a clearly positive first Lyapunov exponent  $\lambda_1 = +1.90 \pm 0.02$  [1]. These analyses provide a strong argument for a chaotic behavior in which instability is guaranteed by the first Lyapunov exponents and where determinism is ensured by obtaining the (deterministic) global model. The more in-depth analysis of the attractor will require a dedicated study.

## 6 Conclusions

PoMoS and GloMo, a couple of algorithms developed under R language and dedicated to global modeling, is presented and tested in this work. The tools are applied to three different systems in order to show their capacity in providing satisfying models of small size from single time series. Models of smaller size are preferred for two reasons. One reason is to avoid redundant terms that may bias the correspondence with the original dynamics. Indeed, it is well known that the addition of one single term in a dynamical system may completely change the system's behavior. Another reason is to remove spurious terms that may foster dynamical or numerical instabilities and potentially lead to a short- or long-term divergence. Obtaining a model of small size is thus a powerful sign of strength of the model. However, oversimplification should be avoided also. Nonetheless, from our experience, obtaining oversimplified models from global modeling appears rarer and may be considered with less apprehension since it may lead to simplified but more robust dynamics rather than spurious or numerically unstable ones. In other words, but depending on the context of application, oversimplification may lead to an interesting solution by default, especially if no better solution can be found.

Various levels of validation are distinguished in the paper, and a validation method accounting for long-term dynamical behaviors is introduced. These various levels of validation were tested onto the global models obtained with the two

algorithms with the aim to have a significant overview of the results. The capacity of the two algorithms was thus evaluated not only based on their ability to reach viable models, but also on the reliability of the models obtained, reliability being evaluated with a refined validation technique, when possible. The quality of a model was estimated in terms of phase portrait, first return map, transition matrix, and population of unstable periodic orbits. The model's size was also considered as an interesting criterion for comparison to the analytic or previously published solutions. The ability of the tool was first tested on the Rössler system, which variables exhibit various levels of difficulty ranging from very low to high when attempting a global modeling. The algorithms were also tested on two sets of real observations. One set relates to the electrodisolution of copper in phosphoric acid, which data was obtained under experimental conditions. Another set concerns the dynamics of rainfed wheat under semiarid climatic conditions observed by satellite remote sensing over northern Morocco.

It was possible to get a global model for each of the cases considered in the study. Except for the dynamics of rainfed wheat, all the levels of validation could be achieved (often after applying an *a posteriori* parameter tuning): (a) Determinism was shown by obtaining a global model; (b) the first return map's patterns could be satisfyingly compared to the one reconstructed from the original data set (the same number of branches could be obtained); (c) the same development of the dynamics could be checked qualitatively (verifying the branches' lengths); (d) the transition matrices could be satisfyingly compared; and (e) the symbols sequences could be checked (until 2-symbol-length sequences, at least). These models are more parsimonious compared to the previously published results. Surprisingly, although not fully reliable, a 7-term model was obtained for the first variable of the Rössler system while a formal solution of nine monomials is ideally known. The model obtained from the time evolution of a current measured in an experiment of electrodisolution also has a smaller size (21 parameters) than the previously published model of 52 parameters. The efficiency of the PoMoS and GloMo algorithms is powerfully exhibited by obtaining a 3-dimensional global model from the last variable of the Rössler system (noted  $x_3$  in the present work), which is characterized by a low observability. This result is especially significant since no 3D model could be directly obtained before, without using a very constraining selection technique or *a priori* knowledge.

Validation could not be performed onto the model of rainfed wheat due to the jittery structure of the phase portrait resulting from the original signal which probably results from the noisy conditions. Nonetheless, these new tools have permitted the extraction of a deterministic component from the signal of vegetation index observed from space. This is an important result since, to our knowledge, the deterministic behavior has never been argued with such a strong element of evidence for any type of vegetation. This is also important since determinism is one of the two essential conditions for chaos. Since the dynamics captured from the vegetation index exhibits a relatively high level of complexity which is characterized by a chaotic regime, it offers a strong evidence for a chaotic behavior of rainfed wheat in northern Morocco.

Finally, thanks to the two algorithms PoMoS and GloMo, it is shown that dynamics of quite complex behaviors can be modeled by global models of canonical form and of quite small size: 10 terms only for the Rössler- $x_1$  variable, 30 terms only for the Rössler- $x_3$  variable, 21 terms for the electrodisolution of copper in phosphoric acid; and 15 terms for the dynamics of rainfed wheat in northern Morocco. This latter model also suggests that relatively complex dynamical structures can be modeled with canonical formulation of rather small size. Developments of global modeling were

initiated in the early 1990s; the present results contribute to provide quite encouraging new illustrations of the potential of the approach.

## 7 Appendix A

Two main optimization criteria are provided in PoMoS. Each is expressed as the sum of two terms, one term evaluating the maximum likelihood, another term corresponding to the model size defined here as the model's number of parameter. These two terms are in competition since maximum likelihood invariably leads to maximize the models' size. The Akaike Information Criterion  $C^{\text{AIC}}$  [49] is defined as follows:

$$C^{\text{AIC}} = -2\ln(L) + 2k \quad (9)$$

where  $L$  is the likelihood and  $k$  is the number of parameters independently adjusted within the model. In practice,  $L$  is estimated from the residuals of the least-squares fitting. The Bayesian Information Criterion  $C^{\text{BIC}}$  [50] allows for a stronger penalization of the model's size through the following expression:

$$C^{\text{BIC}} = -2\ln(L) + k \ln(n_0) \quad (10)$$

where  $n_0$  is the sample size (corresponding to the size of the time series).

The theoretical backgrounds of the two criteria are different since  $C^{\text{AIC}}$  aims at finding a satisfying trade-off between likelihood and model size, whereas  $C^{\text{BIC}}$  aims at identifying the true or quasitru model. Their efficacy can vary drastically depending on factors including the size of the sample, the presence of the true model in the model ensemble, and the size of this model ensemble. In practice, due to the stronger penalization of the model size in the  $C^{\text{BIC}}$  criterion,  $C^{\text{BIC}}$  may lead to sub-dimensioned models, whereas  $C^{\text{AIC}}$  will generally lead to over-dimensioned ones.

### Acknowledgments:

This work was supported by the Institut de Recherche pour le Développement (IRD), the Centre National d'Etudes Spatiales (CNES), and Les Enveloppes Fluides et l'Environnement (LEFE) program of the Institut National des Sciences de l'Univers (INSU). Interactions with the research group Dynamique et Contrôle des Ensembles Complexes (DYCOEC) was especially rich and encouraging for this project. We would like to thank Professor John Hudson warmly for providing us with his data of current from the electrodisolution of copper in phosphoric acid made available through the Atomosyd Web site (<http://www.atomosyd.net/>). We would like also to thank the two anonymous reviewers for their constructive review and, especially, one reviewer for his detailed comments and his very interesting suggestion to develop the validation approach by comparing the populations of unstable orbits; among others, this suggestion made it possible to improve the paper significantly.

## References

- [1] Wolf A., Swift J.B., Swinney H.L. & Vastano J.A., 1985. Determining Lyapunov exponents from a time series. *Physica D*, 16, 285–317.
- [2] Grassberger P. and Procaccia I., 1983. Characterization of strange attractors, *Phys. Rev. Lett.*, 50, 346–349.
- [3] Tufillaro N.B., Abbott T. & Reilly J., 1992. An experimental approach to nonlinear dynamics and chaos. Addison-Wesley: New York.
- [4] Gilmore R., 1998. Topological analysis of chaotic dynamical systems. *Review of Modern Physics*, 70, 1455–1530.
- [5] Gilmore R. & Lefranc M., 2002. *The Topology of Chaos, Alice in Stretch and Squeezeland*, Wiley, 495pp.
- [6] Theiler J., Eubank S., Longtin A., Galdrikian B. & Farmer J.D., 1992. Testing for nonlinearity in time series: the method of surrogate data, *Physica D*, 58, 77–94.
- [7] Gomes M.E.D., Souza A.V.P., Guimarães H.N. & Aguirre L.A., 2000. Investigation of determinism in heart rate variability, *Chaos*, 10 (2), 398–410.
- [8] Poon C.-S. & Barahona M., 2001. Titration of Chaos with Added Noise, *Proc. Natl. Acad. Sci. of USA*, 98 (13), 7107–7112.
- [9] Freitas U., Letellier C. & Aguirre L., 2009. Failure in distinguishing colored noise from chaos using the "noise titration" technique, *Physical Review E*, 79, 035201.
- [10] Kaplan D.T. & Glass L., 1992. Direct test for determinism, *Physical Review Letters*, 68 (4), 427–430.
- [11] Gouesbet G. & Letellier C., 1994. Global vector-field reconstruction by using a multivariate polynomial  $L_2$  approximation on nets, *Physical Review E*, 49 (6), 4955–4972.
- [12] Letellier C., Aguirre L.A. & Freitas U.S., 2009. Frequently asked questions about global modeling. *Chaos* 19(2), 023103.
- [13] Judd K. & Mees I., 1995. On selecting models for nonlinear time series. *Physica D*, 82, 426–444.
- [14] Letellier C., Le Sceller L., Maréchal E., Dutertre P., Maheu B., Gouesbet G., Fei Z. & Hudson J.L., 1995. Global vector field reconstruction from a chaotic experimental signal in copper electrodisolution, *Physical Review E*, vol. 51, 5, pp. 4262–4266.
- [15] Letellier C., Le Sceller L., Gouesbet G., Lusseyran F., Kemoun A., & Izrar B., 1997. Recovering deterministic behavior from experimental time series in mixing reactor, *AIChE Journal*, 43, 9, pp. 2194–2202.
- [16] Letellier C., Maquet J., Labro H., Le Sceller L., Gouesbet G., Argoul F. & Arnéodo A., 1998. Analysing chaotic behavior in a Belousov–Zhabotinsky reaction by using a global vector field reconstruction. *J. Phys. Chem. A*, 102, 10265–10273.
- [17] Letellier C., Ménard O., Klinger Th., Piel A. & Bonhomme G., 2001. Dynamical analysis and map modeling of a thermionic diode plasma experiment. *Physica D*, 156, 169–178.
- [18] Corrêa M.V., Mendes E.M.A.M. & Aguirre L.A., 2000. Modeling chaotic dynamics with discrete nonlinear rational models. *Int. J. Bifurcation Chaos*, 10(5), 1019–1032.
- [19] Letellier C., Aguirre L.A., Maquet J. & Gilmore R., 2006. Evidence for low dimensional chaos in the sunspot cycles. *Astronomy & Astrophysics*, 449, 1, 379–387.

- [20] Maquet J., Letellier C. & L.A. Aguirre, 2007. Global models from the Canadian lynx cycles as a direct evidence for chaos in real ecosystems. *J. of Mathematical Biology*, 55(1), 21-39.
- [21] Hegger R., Kantz H. & Schreiber T., 1999. Practical implementation of nonlinear time series methods: The TISEAN package, *Chaos*, 9, 413.
- [22] PoMoS & GloMo can be freely downloaded at the following address: [http://www.cesbio.ups-tlse.fr/us/pomos\\_et\\_glomo.html](http://www.cesbio.ups-tlse.fr/us/pomos_et_glomo.html).
- [23] Ihaka R. & Gentleman R., 1996. R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics*, 5, 299–314. See also the Comprehensive R Archive Network's Web site at <http://cran.r-project.org/>.
- [24] Chambers J.M., 2008. Software for data analysis: programming with R, Springer, New York, ISBN 978-0-387-75935-7. See also the Comprehensive R Archive Network's website at <http://cran.r-project.org/>.
- [25] Takens F., 1981. Detecting strange attractor in turbulence, in *Lecture Notes in Mathematics*, edited Rand D. and L.S. Young Vol.898, Springer Verlag, Berlin, 366–381.
- [26] Packard N. H., Crutchfield J. P., Farmer J. D., Shaw R.S., 1980. Geometry from a time series, *Physical Review Letters*, 45 (9), 712–716, doi:10.1103/PhysRevLett.45.712.
- [27] Kennel M.B. & Abarbanel H.D.I., 2002. False neighbors and false strands: A reliable minimum embedding dimension algorithm. *Physical Review E*, 66, 1–18, doi:10.1103/PhysRevE.66.026209.
- [28] Letellier C., Moroz I.M. & Gilmore R., 2008. Comparison of tests for embeddings. *Phys. Rev. E*, 78, 026203, doi:10.1103/PhysRevE.78.026203.
- [29] Lainscsek C., Letellier C., Schürer F., 2001. Ansatz library for global modeling with a structure selection. *Physical Review E*, 64, 016206.
- [30] Aguirre L.A., Billings S.A., 1995. Dynamical Effects of Overparametrization in Nonlinear Models, *Physica D*, **80** (1, 2), 26–40.
- [31] Billings S.A., Chen S. & Korenberg M.J., 1989. Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator, *Int. J. Control*, **49** (6), 2157–2189.
- [32] Aguirre L.A., Jácome C.R.F., 1998. Cluster Analysis of NARMAX Models for Signal-Dependent Systems, *IEE Proceedings, Part D -- Control Theory and Applications*, **145**(4), 409–414.
- [33] Aguirre L.A., Freitas U.S., Letellier C., Maquet J., 2001. Structure selection techniques applied to continuous-time nonlinear models, *Physica D*, **158** (1-4), 1–18.
- [34] Note that using the interface in a preliminary run may be a useful step before applying a systematic search in a blind mode since allowing for an efficient tuning of the heuristic.
- [35] Letellier C., Aguirre L.A. & Maquet J., 2005. Relation between observability and differential embeddings for nonlinear dynamics. *Physical Review E*, **71**, 066213.
- [36] See Supplemental Material at [URL will be inserted by publisher] for the explicit formulation of the global models and associated transition matrices obtained (1) for the Rössler system from the three variables ( $x_1, x_2, x_3$ ); (2) for the electrodisolution of copper in phosphoric acid; and (3) for the cycles of rainfed wheat observed from satellite data in northern Morocco.
- [37] Rössler O., 1976. An Equation for Continuous Chaos, *Physics Letters*, 57A (5), p. 397–398.

- [38] Bassett M. R. & Hudson J. L., 1989. Quasi-Periodicity and chaos during an electrochemical reaction, *The Journal of Physical Chemistry*, 93, 2731–2737.
- [39] Fei Z., Hudson J. L. & Kelly R. G., 1994. A low-noise contact for rotating electrodes, *Journal of the Electrochemical Society*, 141 (9), L123–L124.
- [40] Tucker C.J., Pinzon J.E., Brown M.E., Slayback D.A., Pak E.W., Mahoney R., Vermote E.F. & Saleous N.E., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, 26:20, 4485–4498, doi:10.1080/01431160500168686.
- [41] Tucker C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127–150.
- [42] Jarlan L., Tourre Y., Mougin E, Philippon N., Mazzega P., 2005. Dominant patterns of AVHRR NDVI interannual variability over the Sahel and linkages to key climate signals (1982–2003). *Geophysical Research Letters*, 32, L04701, doi:10.1029/2004GL021841.
- [43] Mangiarotti S., Mazzega P., Jarlan L., Mougin E., Baup F. & Demarty J., 2008. Evolutionary bi-objective optimization of a semi-arid vegetation dynamics model with NDVI and  $\sigma_0$  satellite data. *Remote Sensing of the Environment*, vol. 112, no. 4, pp. 1365–1380.
- [44] Mayaux P., Bartholomé E., Fritz S., Belward A., 2004. A new land-cover map of Africa for the year 2000. *J. Biogeogr.*, 31, 861–877.
- [45] Savitzky A., Golay M. J. E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Analytical Chemistry*, 36 (8), 1627–1639.
- [46] Lainscsek C., Letellier C. & Gorodnitsky I., 2003. Global modeling of the Rössler system from the  $z$ -variable, *Physics Letters A*, 314 (5-6), 409–427.
- [47] Letellier C., Maquet J., Le Sceller L., Gouesbet G., Aguirre L., 1998. On the non-equivalence of observables in phase space reconstructions from recorded time series. *Journal of Physics A*, **31**, 7913-7927.
- [48] Le Sceller L., Letellier C. & Gouesbet G., 1999. Structure selection for global vector field reconstruction by using the identification of fixed points, *Physical Review E*, 60 (2), 1600–1606.
- [49] Akaike H., 1974. A new look at the statistical model identification, *IEEE Transactions on Automatic Control*. 19 (6), 716–723, doi: 10.1109/TAC.1974.1100705
- [50] Schwarz G., 1978. Estimating the dimension of a model, *The Annals of Statistics*, 6 (2), 461–464.

## Tables

	Analytic solution	Polynomial solution	Rational solution		Previous 3D-models	PoMoS & GloMo	
	existence	# terms	Nominator	Denominator	# terms	# terms	degree
			degree	degree			
Rössler- $x_2$	yes	9	2	0	9	7	2
Rössler- $x_I$	yes	-	2	1	33	9	2
						10	3
						9	4
Rössler- $x_3$	yes	-	2	2	none*	13	4
$I(t)$	no	-	-	-	52	21	4
$v(t)$	no	-	-	-	none	15	3

Table 1: Synthetic presentation of the previously published and presently obtained global models in terms of number of monomial in the analytic solution, nominator's and denominator's maximum polynomial degree.

\* None direct 3D global model could be obtained yet. However, solutions could be obtained by over dimensioning (4D) the model formulation [47], by using an ad hoc approach based on fixed points coordinates [48], or by using the very constraining Ansatz library [46].

Variable	Model	Poincaré section					1 <sup>st</sup> return
		reference	normal	abscissa	ordinate	selected	
		point	vector $\mathbf{n}$	vector $\mathbf{v}_a$	vector $\mathbf{v}_b$	zone	
Rössler- $x_2$	origin.	[-0.6; 0; 0]	[-0.447; -0.894; 0.]	[0.003; -0.001; 1.]	[0.894; -0.447; -0.003]	$R_b > -0.3$	$\mathbf{v}_b$
	9-p.	"	"	"	"	"	"
	7-p.	"	"	"	"	$R_b > +0.1$	"
Rössler- $x_7$	origin.	[0.; 0.; 0.]	[-0.050; 0.999; 0.]	[0.998; 0.050; 0.]	[-0.019; -0.001; 1.]	$R_a < +0.3$	$\mathbf{v}_b$
	13-p.	"	"	"	"	"	"
	13-p.opt.	"	"	"	"	$R_a < +0.2$	"
	9-p.	"	"	"	"	$R_a < +0.3$	"
	origin.	[0.; 0.; 0.]	[1.; 0.; 0.]	[0.; 0.031; 1.]	[0.; 1.; -0.031]	$R_b < 120$	"
	10-p.	"	"	"	"	"	"
10-p.opt.	"	"	"	"	$R_b < 100$	"	
Rössler- $x_3$	origin.	[0.; 0.; 0.]	[0.; 1; 0.]	[-1.; 0.; 0.002]	[-0.002; 0.; -1.]	$R_a > -0.6$	$\mathbf{v}_a$
	30-p.	"	"	"	"	$R_a > -0.49$	"
	30-p.opt.	"	"	"	"	"	"
$I(t)$	origin.	[44.2; 4; 0]	[1; 0.020; 0]	[-0.003; 0.128; 0.992]	[-0.02; 0.991; -0.128]	$R_a > 0.$	$\mathbf{v}_a$
	21-p.	"	"	"	"	-	"
	21-p.*	"	"	"	"	-	"
	21-p.*opt	"	"	"	"	-	"

Table 2: Definition of the Poincaré sections used to construct the first return maps of the original and global models presented in Figures 5, 6, 7 and 8. The Cartesian coordinates system  $(O, \mathbf{n}, \mathbf{v}_a, \mathbf{v}_b)$  used for each Poincaré section are provided, where  $\mathbf{n}$  is the normal to the Poincaré section. The direction ( $\mathbf{v}_a$  or  $\mathbf{v}_b$ ) chosen to construct the first return map is specified. Restrictions applied to the resulting plane are also provided, coordinates  $R_a$  and  $R_b$  referring to directions  $\mathbf{v}_a$  and  $\mathbf{v}_b$ , respectively.



		Boundaries (along $R_a$ or $b$ )			% of sequence retrieved						Symbols	Maximum
Model		between symbols			sequence's length						present /	length
		0 & 1	1 & 2	2 & 3	1	2	3	4	5	6	retrieved	validated sequence
Rössler- $x_2$	origin.	1.17	2.42	2.99							{0; 1; 2; 3}	
	9-p.	1.19	2.41	3.01	100.	100.	100.	76.2	55.3	47.8	{0; 1; 2; 3}	3
	7-p.	1.51	2.66	3.09	100.	93.8	37.5	-	-	-	{0; 1; 2; 3}	1
Rössler- $x_7$	origin.	66.0	118.5	142.5							{0; 1; 2; 3}	
	13-p.	63.0	117.7	-	75	50	-	-	-	-	{0; 1; 2}	0
	13-p.opt.	64.0	119.0	144.0	100.	93.7	84.3	53.9	-	-	{0; 1; 2; 3}	1
	9-p.	-	127.5	152.8	75	50	-	-	-	-	{1; 2; (3)}	0
	origin.	12.4	23.3	28.3							{0; 1; 2; 3}	
	10-p.	12.1	21.3	-	75	50	-	-	-	-	{0; 1; 2}	0
	10-p.opt.	12.0	22.8	26.8	100.	100.	98.4	90.6	53.5	-	{0; 1; 2; 3}	2
Rössler- $x_3$	origin.	-0.36	-0.28	-0.26							{0; 1; 2; 3}	
	30-p.	-0.36	-0.20	-	75	50	-	-	-	-	{0; 1}	0
	30-p.opt.	-0.36	-0.19	-0.16	100.	100.	96.9	90.6	56.25	-	{0; 1; 2; 3}	2
$l(t)$	origin.	330									{0; 1}	
	21-p.	590			100.	75.	-	-	-	-	{0; 1}	1
	21-p.*	529			100.	75.	-	-	-	-	{0; 1}	1
	21-p.*opt	490			100.	100.	100.	100.	100.	100.	{0; 1}	6

Table 3: Boundary limits used when partitioning the first return maps, % of sequences retrieved by the model for symbol sequence of 1– to 6–symbol length sequences, symbols retrieved and length of the larger symbol sequences that could be validated.

## Figure Caption (color online only)

Figure 1: Operating diagram of PoMoS algorithm: (1) a set of  $N_0$  initial models is randomly generated and put in basket 1; (2) if the number of model is large enough ( $N > N_l$ ), models are tested based on one chosen criterion (see the Appendix); (3) best models are selected, whereas bad ones are rejected into basket 2; (4) the best model is used to generate neighboring models (a model distance is defined for this purpose) which are added into basket 1; (5) stages (2) to (4) are repeated until the number of model in basket 1 becomes too low ( $N \leq N_l$ ); (6) the model research is thus stopped.

Figure 2: Original signal from the Rössler system for variables  $x_1$  (top),  $x_2$  (middle) and  $x_3$  (bottom).

Figure 3: Selected windows of an experimental signal of current intensity  $I(t)$  measured in an experiment of electrodisolution of copper in phosphoric acid obtained in John Hudson's group [38]. The complete data set is available on the atomosyd Web site (<http://www.atomosyd.net/>).

Figure 4: Vegetation index signal  $v(t)$  of rainfed wheat. The time series was derived from the NDVI product provided by the GIMMS [40] at  $8 \times 8$ -km<sup>2</sup> resolution and 10-day sampling. It was obtained by averaging the signal spatially over a window located in northern Morocco and by applying a Savitzky-Golay filtering (details are given in the text).

Figure 5: Differential embeddings (left panels) and first return maps (right panels) reconstructed from the original variable  $x_2$  of the Rössler system (top panels) and from two global models obtained from it: one 9-term model (middle panels) and one 7-term model (bottom panels).

Figure 6: Differential embeddings (left panels) and first return maps (middle column and right column panels) reconstructed from the original variable  $x_1$  of the Rössler system (top panels) and from three global models obtained from it: one fourth degree 13-term model (second line panels), one fourth degree 9-term model (third line panels), and one third degree 10-term model (bottom line panels). The first return maps reconstructed from the models directly obtained with GloMo algorithm are plotted in the middle column panels. The first return maps obtained after tuning one of the parameters are plotted (when available) in the right column panels.

Figure 7: Differential embeddings (left panels) and first return maps (middle column and right column panels) reconstructed from the original variable  $x_3$  of the Rössler system (top panels) and from the fifth degree 30-term global model obtained with GloMo (bottom panels). In the bottom panels, the middle panel corresponds to the first return map of the 30-term model directly obtained from GloMo algorithm, whereas the right panel corresponds to the modified version of the same model obtained by tuning one parameter.

Figure 8: Differential embeddings (left panels) and first return maps (middle and right panels) reconstructed from the original current intensity  $I(t)$  measured from an

experience of copper electrodisolution (top panel) and from the two global models obtained from it. Both have the same structure: fourth degree 21-term models. The parameterization of the first one (middle panel) was obtained from the shorter time series presented in Figure 3; and from the longer time series for the second one (bottom panel). First return maps directly obtained by applying GloMo algorithm are plotted in the middle column panels. First return maps obtained after tuning one parameter of the second model is plotted in the right column panel.

Figure 9: Differential embeddings reconstructed from the original vegetation index  $v(t)$  (left panel) and from the third degree 15-term global model(right panel).

# Figures

Figure 1:

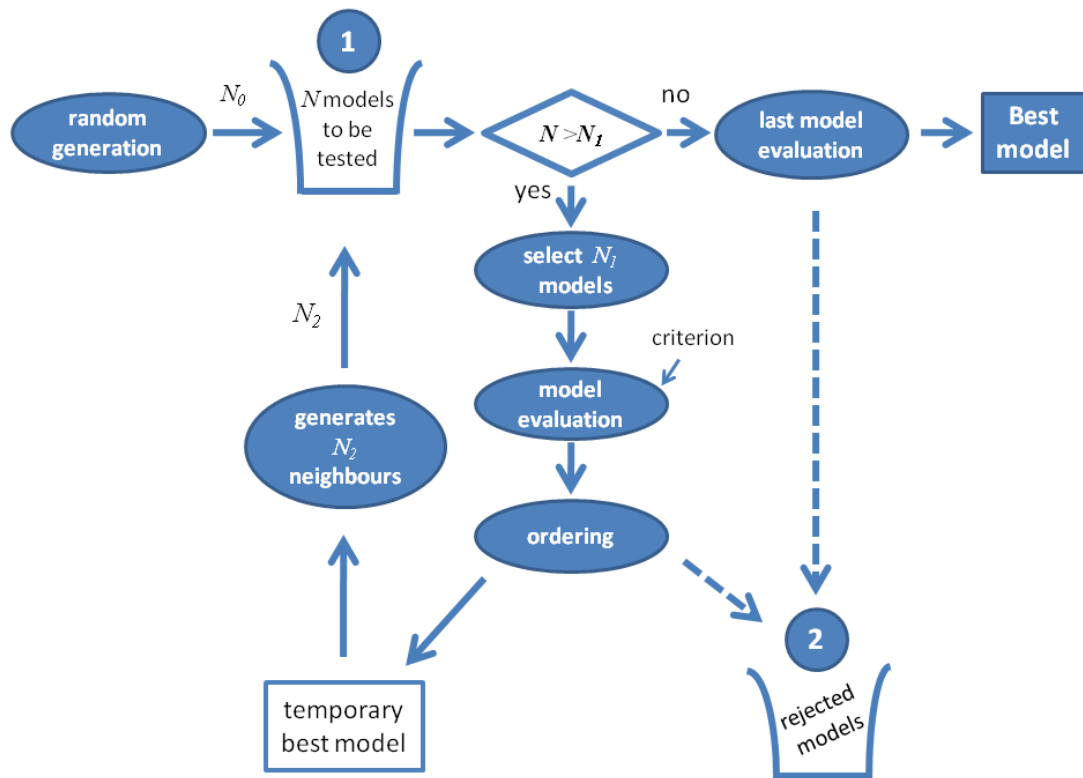


Figure 2:

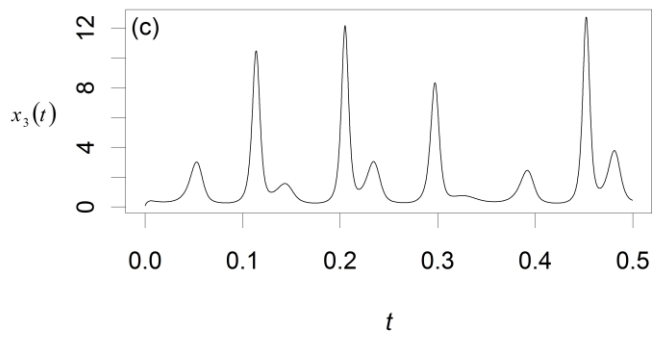
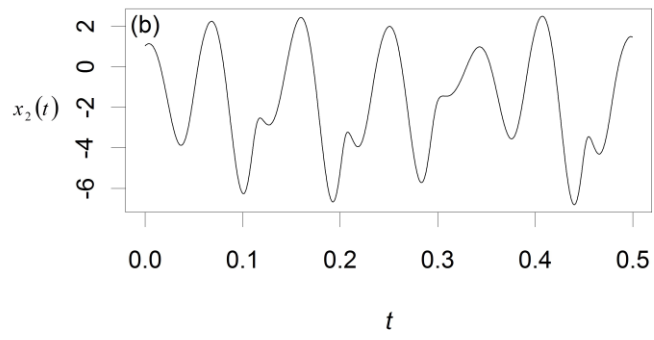
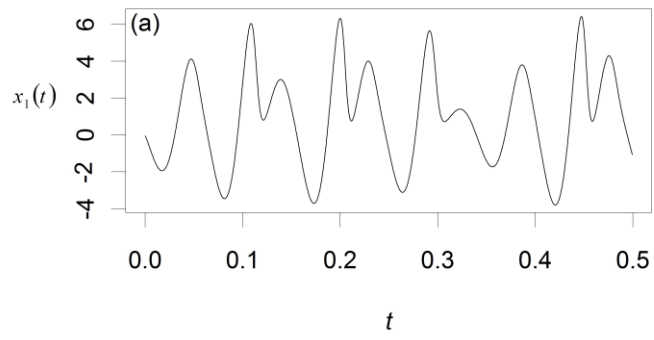


Figure 3:

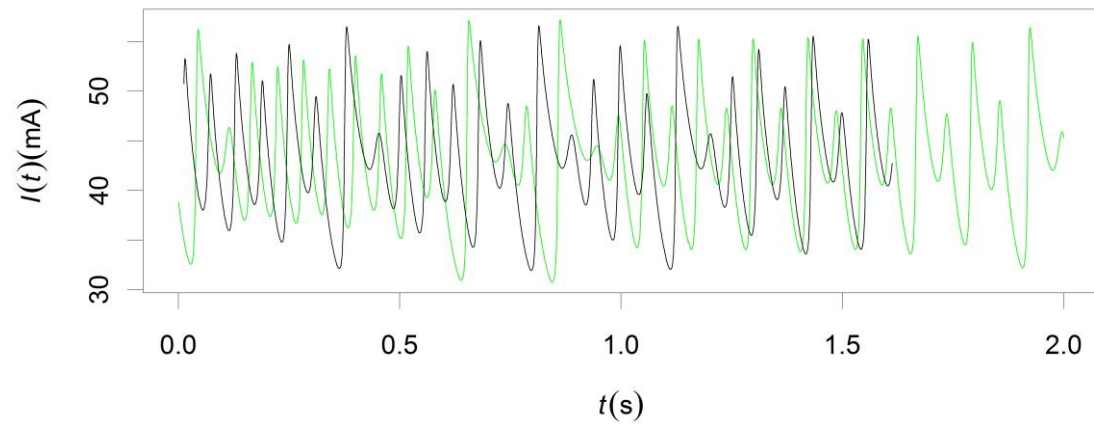


Figure 4:

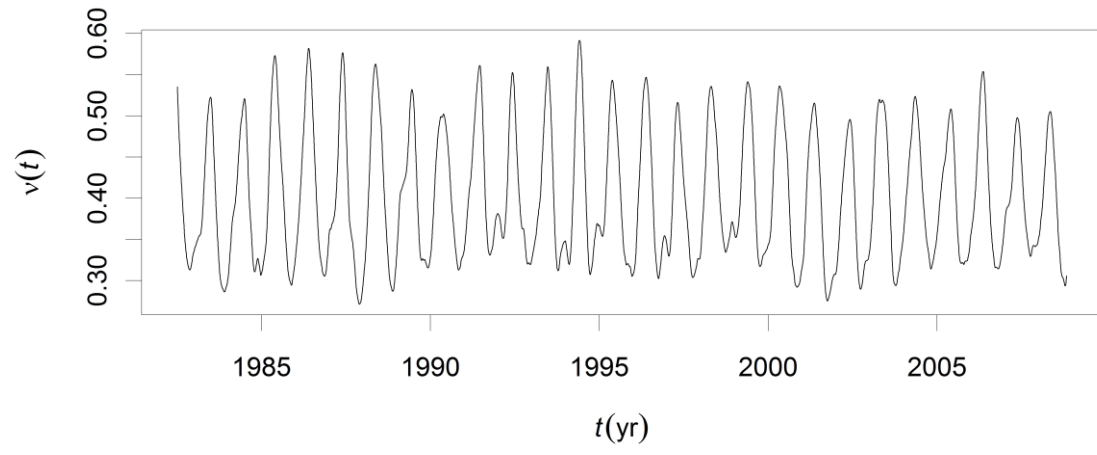


Figure 5:

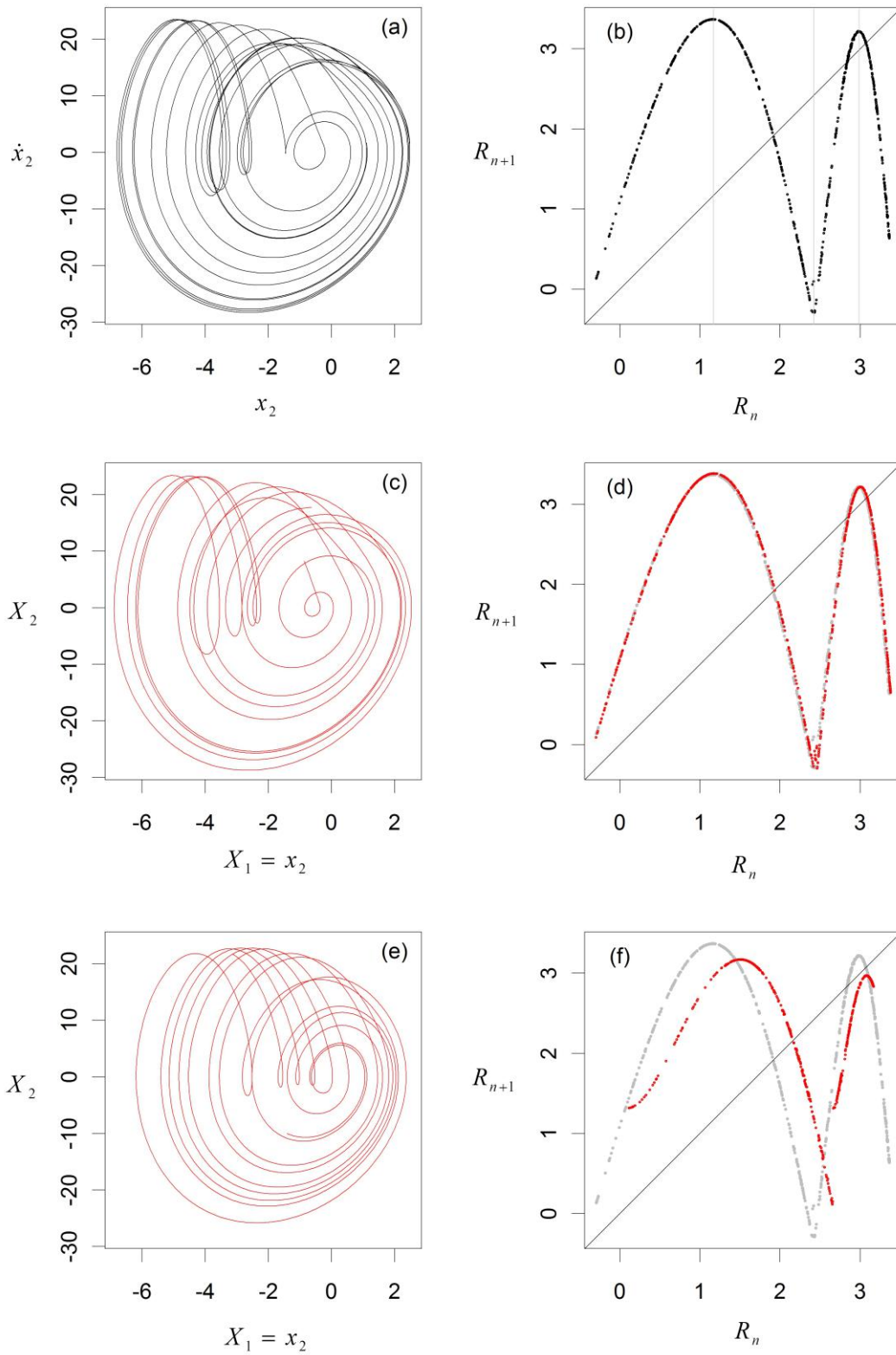




Figure 6:

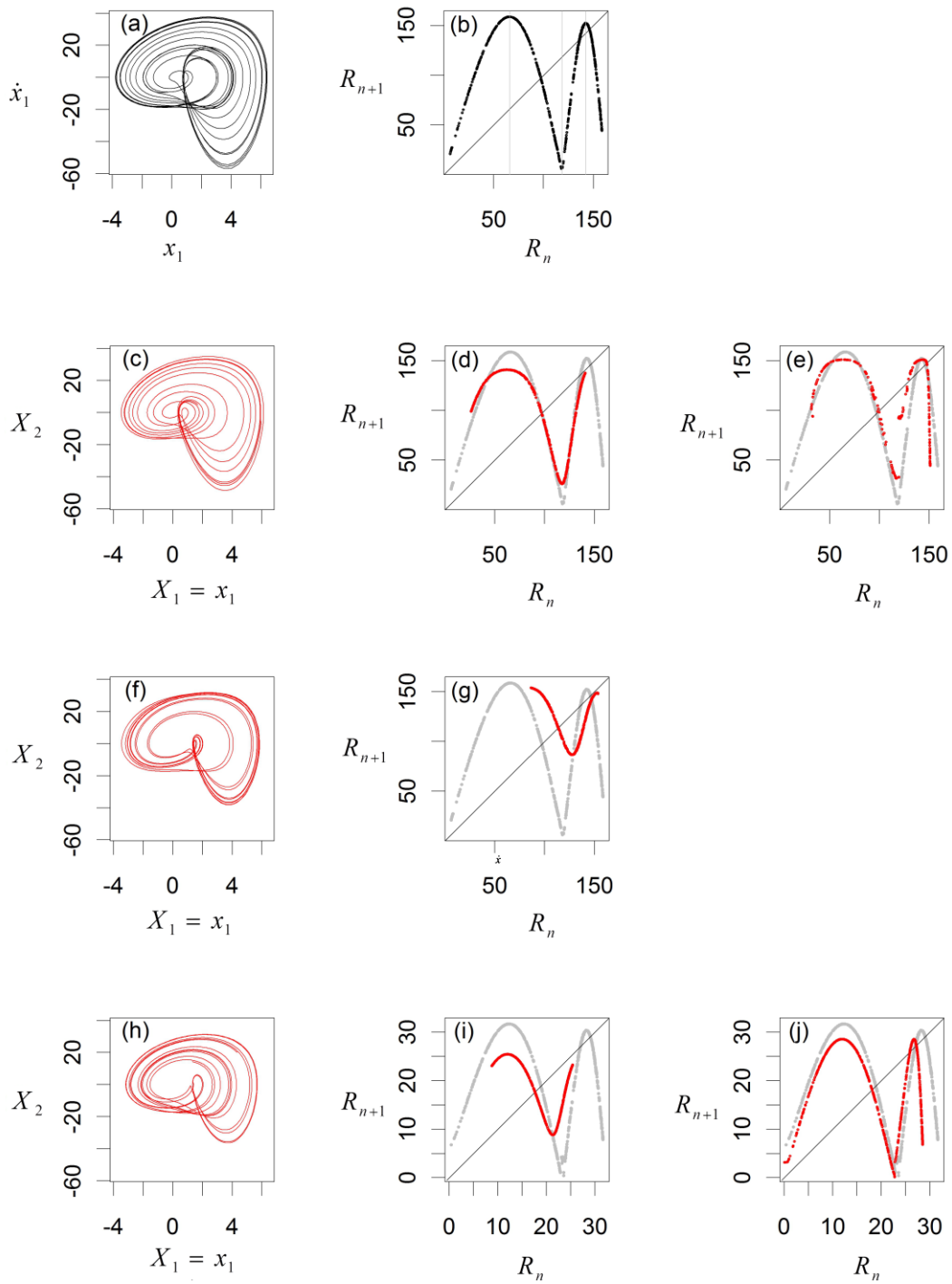


Figure 7:

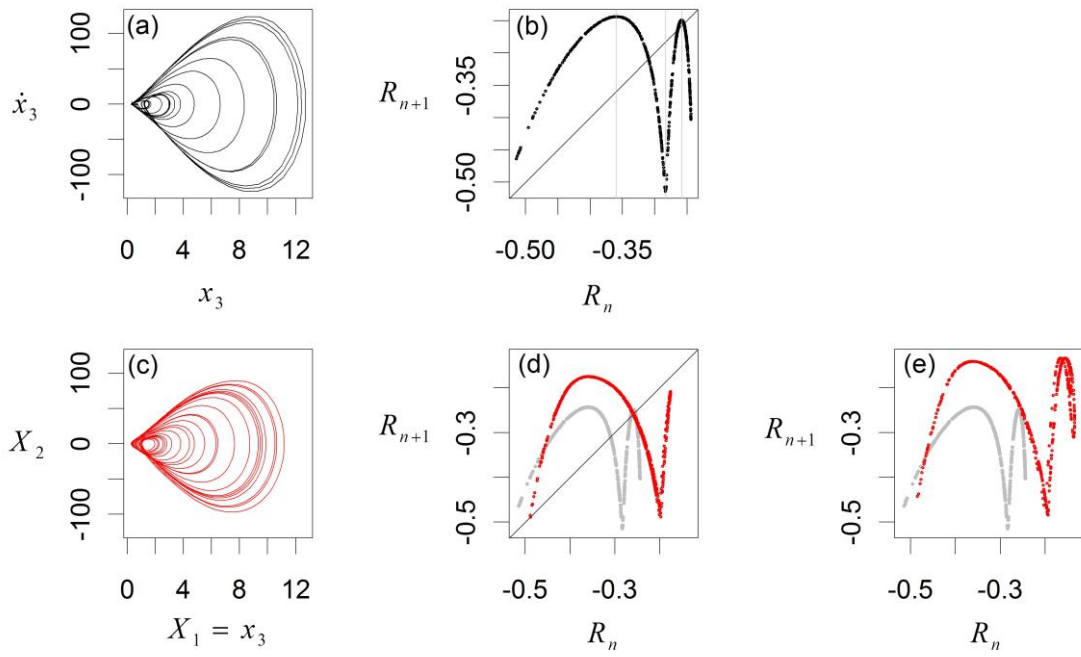


Figure 8:

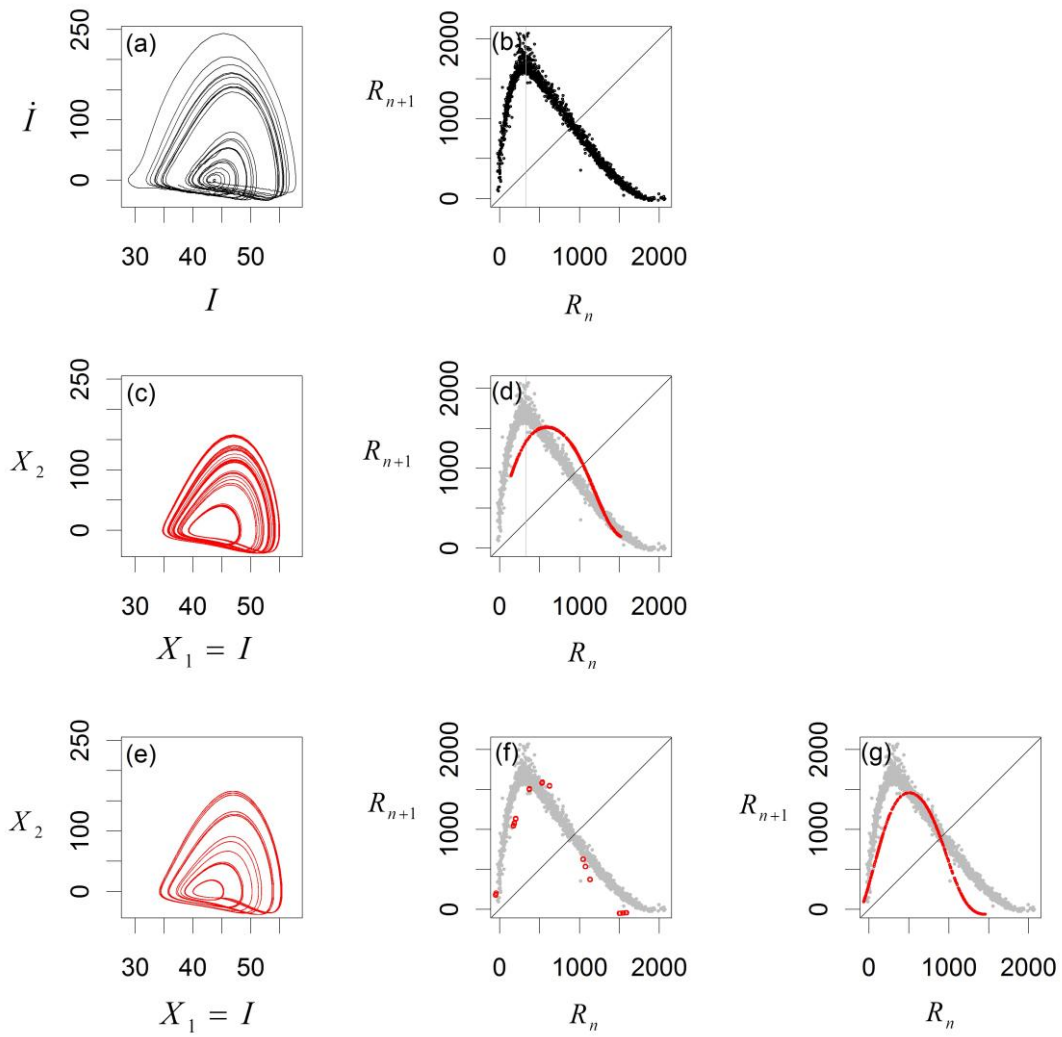


Figure 9:

