Can the original equations of a dynamical system be retrieved from observational time series?

Cite as: Chaos **29**, 023133 (2019); https://doi.org/10.1063/1.5081448
Submitted: 14 November 2018 . Accepted: 01 February 2019 . Published Online: 25 February 2019

Sylvain Mangiarotti, and Mireille Huc

Can the original equations of a dynamical system be retrieved from observational time series?

Cite as: Chaos **29**, 023133 (2019); doi: 10.1063/1.5081448
Submitted: 14 November 2018 · Accepted: 1 February 2019 ·
Published Online: 25 February 2019

Sylvain Mangiarotti and Mireille Huc

AFFILIATIONS

Centre d'Études Spatiales de la Biosphère, UPS-CNRS-CNES-INRA-IRD, Observatoire Midi-Pyrénées, 18 avenue Édouard Belin, 31401 Toulouse, France

ABSTRACT

The aim of the present work is to investigate the possibility to retrieve the original sets of dynamical equations directly from observational time series when all the system variables are observed. Time series are generated from chosen dynamical systems, and the global modeling technique is applied to obtain optimal models of parsimonious structure from these time series. The obtained models are then compared to the original equations to investigate if the original equations can be retrieved. Twenty-seven systems are considered in the study. The Rössler system is first used to illustrate the procedure and then to test the robustness of the approach under various conditions, varying the initial conditions, time series length, dynamical regimes, subsampling (and resampling), measurement noise, and dynamical perturbations. The other 26 systems (four rational ones included) of various algebraic structures, sizes, and dimensions are then considered to investigate the generality of the approach.

In many domains in science, one important problem is to unveil the equations underlying the observed dynamics. Equations' reconstruction is generally performed throughout a long historical process in a movement back and forth between observations and modeling. Though, since equations do arise from observations, one could expect that these equations could be directly derived from observational data. It is shown in the present paper that, in numerous nontrivial cases, the algebraic structure of original dynamical systems can be retrieved from observational time series when the original structure is polynomial or close to polynomial. One interest of such an approach is to obtain interpretable equations when these are unknown which is the case for numerous physical, ecological, and biological systems.

I. INTRODUCTION

It is only recently that the idea to obtain dynamical equations directly from observational time series has emerged. Historically, model-building from data originates from the fields of Electric Engineering and Statistics[1] and was almost exclusively focused on linear problems.[2,3] Nonlinear problems started to be investigated later, mainly based on discrete modeling approaches.[4–6] Modeling in Ordinary Differential Equations (ODEs) started to be investigated in the 1990s but exclusively based on single time series.[7,8] The theory of nonlinear dynamical systems offered a fertile ground for such an objective for various reasons. First, it enables to generate complex deterministic behaviors from low-dimensional systems which is one important difficulty to tackle with before considering higher-dimensional systems. Second, its theoretical background is well adapted for modeling deterministic behaviors presenting a high sensitivity to the initial conditions (chaotic behaviors) which is the case for numerous real world behaviors (e.g., climatic, hydrological, ecological, and epidemiological dynamics). This theory is *a fortiori* also adapted for modeling linear and weakly nonlinear dynamics. Third, the global modeling technique derived from this theory has proven to be a powerful approach to obtain models from observational time series resulting from synthetic data,[7,9–13] experimental data,[8,12,14] and real world observations.[12,15–18]

It is much more recently that this technique could be used to obtain continuous equations from multivariate time series[19,20] (discrete equations could be obtained earlier[21]) and to detect nonlinear couplings.[22]

Even much fewer studies have explored the question of equation re-constructibility of which difficulties have been pointed out associated with non-uniqueness of the solution

either relating to time scaling or resulting from identical jerk forms or to common statistical properties.[23–25]

In a single case, for a model directly obtained from observational data, an interpretation could even be proposed for each term of the reconstructed model.[19] Specific questions arise from this result: Can equations' structures obtained from the time series be interpreted? And thus, just before: Can original equations of the dynamics be unveiled from observational time series?

To investigate these questions, simple low-dimensional systems should be considered first in order to have prohibitive examples of the potential of the approach: if it is possible to retrieve the equations entire structure from a set of observational time series, then it should be possible to do it—at least—for equations of the simplest formulation. In that respect, the dynamical systems considered in this study will be nontrivial—low-dimensional ($d = 3$ to 5)—systems. Twenty-seven systems of various algebraic structure and dynamics will be considered here.

The paper is organized as follows. The theoretical background is introduced in Sec. II, where the global modeling technique is presented, the problem is stated, and the algorithms and modeling procedure are described. The synthetic data generated for the analysis are explained in Sec. III. The results of the experiments are then presented in Sec. IV and separated in five subsections. In Sec. IV A, the Rössler system is first considered under ideal conditions to provide a careful description of the procedure (this procedure will be applied to all the experiments). The robustness of the approach is then tested under various restrictive conditions (still on the Rössler system) in Sec. IV B. Analyses are then tested on other three-dimensional quadratic systems (in Sec. IV C), cubic and higher-dimensional systems (in Sec. IV D), and non-polynomial systems (in Sec. IV E). Conclusions are drawn in Sec. V.

II. THEORETICAL BACKGROUND

A. Global modeling

The aim of the global modeling technique is to obtain sets of discrete, continuous, or delayed equations directly from observational time series.[7,11,26] In its principle, the approach is quite general and can also be applied to splitted time series, that is, either presenting gaps in time or observed at different locations (concomitantly or not).[13] It has also been proven, both analytically and numerically that, under conditions of phase synchronicity, it is possible to apply this technique to spatially aggregated time series.[13] In the present work, we will focus on ODEs of general form

$$\begin{cases} \dot{x}_1 = F_1(x_1, x_2, \ldots, x_n), \\ \quad \vdots \\ \dot{x}_n = F_n(x_1, x_2, \ldots, x_n) \end{cases} \tag{1}$$

that will be modeled by polynomial structures, with the aim to retrieve the original equations' structure when it is polynomial itself, or to get good polynomial approximation for it when it is not. The objective of the present work being to obtain the original algebraic formulation of the original system, it will be assumed that all the system variables $(x_1, x_2, \ldots, x_n)$ are observed. The initial data set, for each system, will then consist of n time series $\{x_i^{obs}(t_k)\}_{i=1\ldots n, k=1\ldots N_{ech}}$ where N_{ech} is the number of data point for each observed time series x_i^{obs} at time t_k.

B. Statement of problem

The objective of the present study is to investigate under which conditions the original formulation of a dynamical system can be retrieved from observations. The problem is very general since our objective is not just to reproduce the observations (that is, a specific trajectory provided by observed time series): our aim is retromodelling, that is, to retrieve an algebraically interpretable set of equations and ideally, *the* original set of equations. This implies that the methodology should be general enough not to reproduce strictly the particular temporal pattern of the original system but rather to unveil all the solutions of the original system. The phase space is an oriented space well designed for this purpose, since enabling to represent all the system states. For a deterministic system, all the solutions are included in this space. To retrieve the original equations, the phase space generated by the model should be equivalent to the phase space reconstructed from the observed time series.

To investigate the possibility to retrieve the original equations from observational time series, the following methodology is used: (a) Time series are generated from known dynamical systems. To investigate the robustness of the approach, the original system may be perturbed and the time series degraded by various factors. (b) The global modeling technique is applied to get a dynamical model of optimally concise formulation. For this purpose, a selection technique is designed based on the rejection of the models of which phase space is of bad agreement with the original phase space (see Secs. II C and II D). (c) Then, the algebraic structure of the best model is compared to the original system. Ideally, the retrieved structure should be identical to the original system structure used to generate the observational time series.

In practice, it is thus required to have an algorithm enabling to obtain optimal models (structure and parameterization) based on appropriate criteria. Moreover, this algorithm should run in a reasonably short time. In the present context, it is assumed that all the variables of the system are observed. To retrieve the complete set of equations, it will be required to retrieve one equation of polynomial form

$$\dot{X}_l = \sum \mu_i \Theta_i \tag{2}$$

for each variable X_l, where $\Theta_i = \prod_{j,k_j} X_j^{k_j}$ are the polynomial terms (such as $\sum k_j \leq q$) and μ_i are their corresponding coefficients.

To run the algorithm, the following parameters are required: (1) the dimension d of the reconstruction, that is, the number of variables used as input if derivatives are not

used in the reconstruction (since it is assumed here that all the system variables are available, we have $d = n$, where n is the number of observed variables); (2) the maximum polynomial degree q used for the description (this degree is assumed to be *a priori* unknown). The number p of possible terms for a single equation directly relies on d and q, such as $p = \frac{(n+q)!}{q!\,n!}$. (3) The maximum number N_p^{max} of model parameters which cannot exceed $n \times p$. To reduce the computing time, it may be limited intentionnally and increased progressively.

The algorithm is organized in four stages. Without any knowledge nor information, the number of possible models for each equation will directly depend on the number of variables considered in the formulation and on the maximum polynomial degree used in the formulation. The resulting number of possible model structure is absolutely huge: $2^{n\times p}$. For instance, for $d = 3$ and $q = 2$, a single equation can have $p = 10$ polynomial terms (X_1, X_2, X_3, X_1^2, X_2^2, X_3^2, the cross terms X_1X_2, X_1X_3, X_2X_3 plus the constant) leading to 2^{10} potential structures for one equation. Since three equations have to be retrieved ($n = d = 3$), the number of possible models will be $2^{3\times10}$. Our algorithm has to find the original model structure among more than one billion models. For $d = 6$ and $q = 2$ with six equations to retrieve, the number of possible models will be $2^{6\times28}$ ($\approx 3.7 \cdot 10^{50}$), etc. To be able to tackle with such tremendous numbers of potential model structures, a drastic (but also very accurate) selection technique is required.

C. Structure selection technique

In the present context, one equation has to be obtained for each variable X_l, that is, $\dot{X}_l = P_l(X_1, X_2, \ldots, X_n)$ for $l = 1 \ldots n$. In the first stage, the equation for each variable X_l is considered individually, for which an ensemble of possible structure is preselected. The selection is performed as follows. At the beginning, all the p terms of the polynomial are included. A Gram-Schmidt technique is used to compute the model coefficients μ_i. The first possible equation structure $E_{1,p}$ is then obtained for Eq. (1) (corresponding to variable X_l). The variance σ of the residual signal $\epsilon(t)$ defined as

$$\epsilon = \dot{X}_l^{obs} - \sum \mu_i \Theta_i^{obs}, \tag{3}$$

is computed for it. To estimate the contribution of each term in this first equation, a leave-one-out method is used: each term Θ_i is set aside one by one by setting its parameter μ_i to zero. For each corresponding sub-model i (such as $\mu_i = 0$), the variance σ_i^2 of the residual signal is computed. The relative contribution of each monomial is deduced from the ratio

$$K_i = \sigma_i^2/\sigma^2. \tag{4}$$

The smaller the K-ratio, the smaller the usefulness of the corresponding ith monomial. The term of smallest contribution is thus removed, and the Gram-Schmidt technique is used again to estimate the parameters of this new formulation in order to have a precise estimate of this potential equation structure corresponding now to $E_{1,p-1}$. Indeed, in a context of strongly nonlinear relations between the variables, even removing a single term may strongly modify the coefficients of the remaining variables, and therefore, the parameters should be re-estimated for each substructure. The contribution of each remaining $p-1$ monomial is thus re-estimated from the $E_{1,p-1}$ equation using again the leave-one-out technique. The monomial of least importance is then definitively removed and a precise computation of the parameters is performed with the Gram-Schmidt technique from which equation structure $E_{1,p-2}$ is deduced. The same algorithm is repeated until no monomial remains (corresponding to $E_{1,0}$). The $p+1$ possible equations $E_{1,p}$ to $E_{1,0}$ constitute the ensemble of potential equation structures for the dynamic of variable X_1. We have thus reduced our ensemble from 2^p to $p+1$ possible structures for variable X_1, among which the original formulation should be included if our selection technique is appropriate and accurate enough.

The same selection technique is used for all the variables X_l leading to an ensemble of $(p+1)$ potential equations for each variable. Their combinations represent $(p+1)^n$ potential models $\{E_{l,k}\}_{l\in[1\ldots n],k\in[0\ldots p]}$, which remains considerable but which appears manageable for numerical integration tests if the number of variables n and the maximal polynomial degree q are sufficiently small (e.g., 11^3 models for $n = 3$ and $q = 2$). In order to accelerate the model search, only models of size $N_p \le N_p^{max}$ may be considered where N_p may be increased progressively as long as no valid model can be obtained.

D. General model selection procedure

The following model selection procedure is used: (1) The automatic selection technique presented in Sec. II C is applied to the time series to obtain an ensemble of automatically preselected models. (2) This ensemble of preselected models is then tested in terms of numerical integrability (diverging models are rejected). (3) The remaining models are then selected by default, based on their performances to reproduce the original phase space: fixed points and period-1 cycles are automatically rejected (except if the original phase portrait is a period-1 cycle); remaining models that are obviously not able to reproduce the original phase portrait are rejected by a visual inspection (illustration of such a rejection will be given in Sec. II A). A rejection based on more refined characterization may be required when the models' performances are visually not obvious. If several models remain after this rejection process, the model of more parsimonious formulation is preferred. When several models of same size remain, then the structure common to these suboptimal models is considered as best. (4) This optimal structure is then compared to the original system.

III. DATA

To test the ability of the global modeling technique to unveil the original form of the dynamical equations from observational time series, it is necessary to generate synthetic data sets that will be used for the various experiments considered in this study. The first ensemble of experiments aims to

characterize the sensitivity and the robustness of the global modeling technique to external factors: sensitivity to (1) initial conditions, (2) data sampling time, (3) time series length, (4) dynamical regime, (5) measurement noise, and (6) dynamical perturbations. All these analyses will take the Rössler system (R76)[27]

$$\dot{x} = -y - z; \quad \dot{y} = x + ay; \quad \dot{z} = b + z(x - c) \tag{5}$$

as a case study. This system is paradigmatic of chaos and has been extensively studied. It has seven terms and only a single nonlinearity (xz in the third equation). For $(a, b, c) = (0.432, 2, 4)$, it produces a phase-coherent chaos which is a common situation in chaos that should thus be considered in first.

The other parts of the analysis aim to test the generality of the tools by investigating a wide diversity of systems. The description of the other systems considered in the analysis will be given throughout the analysis, in order to facilitate the comparison between the original equations and the models retrieved by global modeling.

For all the models presented in the study, the time series were generated using a fourth-order Runge-Kutta method. For all the analyses, the time series of the first derivatives are required that will be computed using a Savitzky-Golay filter. This filter will also have a slight smoothing effect on the time series. Various time series preprocessing may also be applied depending on the type of degradation. When necessary, resampling will be performed using cubic splines and smoothing using Butterworth filter. The main characteristics of all the time series considered in the study, as well as the detailed equations of all the dynamical systems used in the analysis and the models obtained for them are provided as the supplemental material.

IV. APPLICATION AND RESULTS

A. An example: The Rössler system

A first analysis is given under ideal conditions to exemplify the process. The three time series $x(t)$, $y(t)$, and $z(t)$ presented in Fig. 1 and obtained by numerical integration of the Rössler system are used in this example. A phase space projection reconstructed from these time series is presented in Fig. 2 (black curves). These time series are of moderate length (six oscillations), free of noise, and have a sampling time that provides a very good description of the time evolutions for the three variables ($N_{ech} = 2000$ data point for each time series).

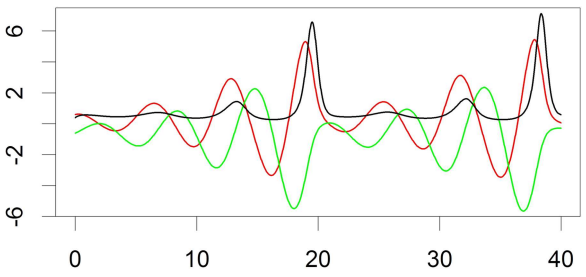


FIG. 1. Original time series (no unit) $x(t)$ (in red), $y(t)$ (in green), and $z(t)$ (in black) generated with the Rössler system [Eq. (5)] with $(a, b, c) = (0.432, 2, 4)$ and initial conditions $(x_0, y_0, z_0) = (0.6, -0.6, 0.4)$.

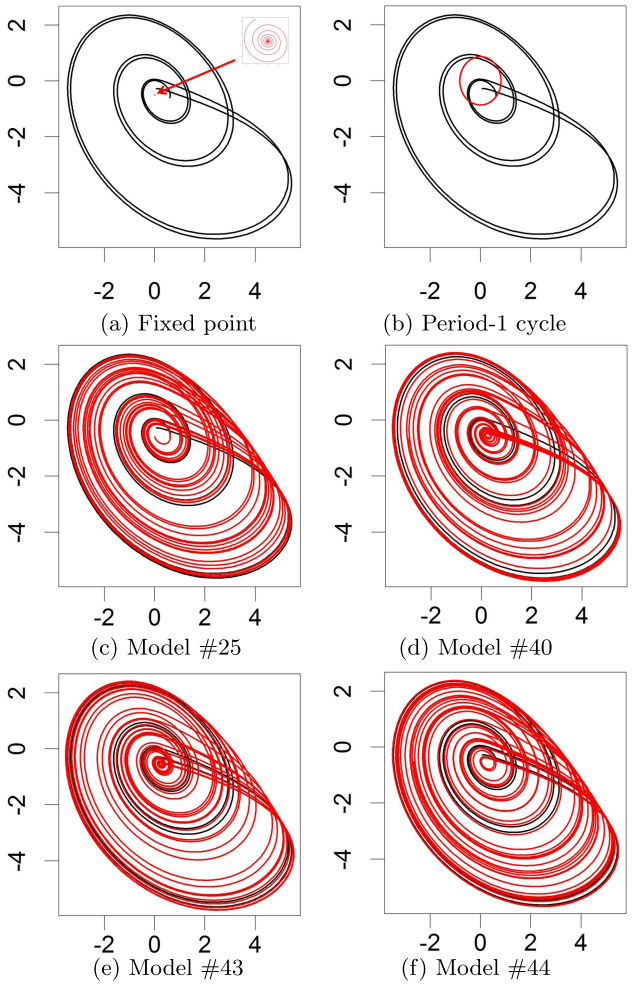


FIG. 2. Global models (in red) superimposed to the original (black) phase portraits in (x,y) projection.

The global modeling technique is applied to this set of three time series using the following input parameters: model dimension $d = 3$ (one dimension for each measured variable), maximum polynomial degree $q = 2$, and model maximum size $N_p^{max} = 8$ (with the idea to restart the modeling process with alternative algorithm parameters if no satisfying model can be obtained). Fifty-six models are preselected by the automatic structure selection procedure defined in Sec. IV C, and their numerical integrability is then tested. Thirty-one models are automatically rejected (not shown), either (1) as diverging if

Not a Number are produced by numerical integration or (2) if the estimated trajectory is going too far from the original data set (by default, a four sigma test is applied to each variable). Models that remain are rejected manually if the models' dynamic is obviously too simple compared to the original data, that is, either (3) when a model converges to a fixed point [it was the case for six models, one example is given Fig. 2(a)] or (4) to period-1 cycles, which phase portraits are poorly compatible with the original phase portrait [as it is the case for 15 other models, e.g., Fig. 2(b)].

Once these models rejected, only four models remained [see Figs. 2(c)–2(f)]: one 7-term model (#25) and three 8-term models (#40, #43, and #44). The four ones can reproduce the complexity of the original dynamic. Their phase portraits are highly consistent with the original one. Therefore, it is not necessary to investigate models of larger size ($N_p^{max} > 8$) or of larger polynomial degree ($q > 2$). Note that it should not be expected to retrieve strictly the original trajectory. Indeed, for chaotic behavior, to obtain exactly the same original trajectory, it would be required to retrieve—both the initial conditions and the model parameters—with an infinite precision which is impossible in practice, and which is not our aim here since our method aims to apply under various degraded conditions (subsampled and noisy time series, perturbed systems). Based on the phase portraits, there is *a priori* no reason to prefer one model or another among these four. To choose among them under such conditions, the model of more concise formulation should be preferred, that is, here, model #25

$$\begin{cases} \dot{x} = -0.9979\,y - 0.9987\,z, \\ \dot{y} = 0.9990\,x + 0.4316\,y, \\ \dot{z} = 1.9874 - 3.9747\,z + 0.9937\,xz \end{cases} \tag{6}$$

(the number of digits was truncated to facilitate the readability, see the supplementary material for more details). The algebraic structure of this model #25 effectively corresponds to the original set of equations [Eq. (5)]. The original system structure is thus effectively retrieved from the original time series and the error on the estimated parameters does not exceed 3%. The largest error is found in the second equation [it is associated to parameter a, Eq. (5)], it does not exceed 1% elsewhere.

Interestingly, the other three models have the same algebraic structure plus one extra monomial, such as

$$\begin{cases} \dot{x} = -\alpha_1\,y - \alpha_2\,z \quad (+\alpha_3 z^2), \\ \dot{y} = \beta_1\,x + \beta_2\,y \quad (+\beta_3 xz), \\ \dot{z} = \gamma_1 - \gamma_2\,z + \gamma_3\,xz \quad (+\gamma_4 x^2), \end{cases} \tag{7}$$

with $(\alpha_3, \beta_3, \gamma_4) = (0, 0, 0.001238)$ for model #40, $(0, -0.000\,1493, 0)$ for #43, and $(0.0008870, 0, 0)$ for #44. The four models thus all have the same common structure. This structural consistency, as well as the small amplitudes of the additional terms contribute to strengthen the conclusion that, at least under ideal conditions, the equations' original structure of a dynamical system of polynomial form can be retrieved from observational time series. The robustness and the generality of the approach will be investigated in Secs. IV B– IV E.

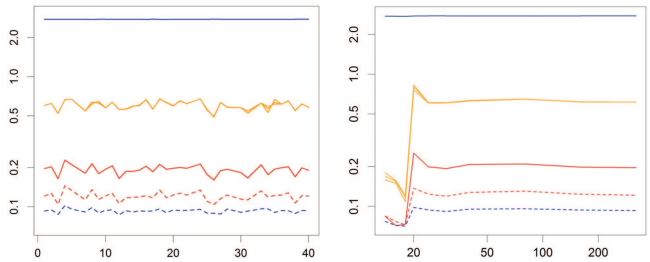
B. Sensitivity analyses

1. Initial conditions

To illustrate the low sensitiveness of the approach to the initial conditions, an ensemble of 40 non-overlapping time series was generated, each starting from different initial conditions on the attractor. The procedure illustrated in Sec. II A was used in each case. The original model formulation could be retrieved for 36 simulations. For the four remaining simulations of this ensemble, no model was obtained (all were rejected, that is, no detection but also no erroneous detection). This shows that it may be useful to restart a simulation with another window analysis when no model is obtained. The error associated with the seven model parameters are plotted in Fig. 3(a) that shows that the parameter precision is weakly dependent on the window, and that both high and low heterogeneity may take place between the various model coefficients, even in the same equation. Note that considering transients (with initial conditions located far from the attractor) led to similar results (not shown).

2. Dynamical regime

In this experiment, four alternative dynamical regimes of the Rössler system are considered, three of lower complexity (periodic cycles of periods one, two, and four) and one of higher complexity (a phase non-coherent chaotic regime). Their original phase portraits are plotted in Fig. 4 (in light gray). The global modeling technique was applied to these four situations following the procedure presented in Sec. IV A. For all the cases, most of the models were rejected automatically, and the remaining models could be separated in two classes. In the first class, the correspondence with the original dynamic was quite poor [as previously observed for the reference case in Figs. 2(a)–2(b)] and could be rejected straightforward; in the



(a) Initial conditions (b) Length T

FIG. 3. Models parameter error (in %) estimated (a) from an ensemble of 40 non-overlapping time series and (b) from time series of increasing length $T = N_{ech} \times \delta t$ ($N_{ech} = 700$ to 16 000). Parameters from the first, second, and third equations are, respectively, plotted in red, blue, and orange. For the third equation, it is difficult to distinguish the three lines (in orange) since the error is almost identical for each. Vertical lines correspond to $T = 24$ and $T = 100$.

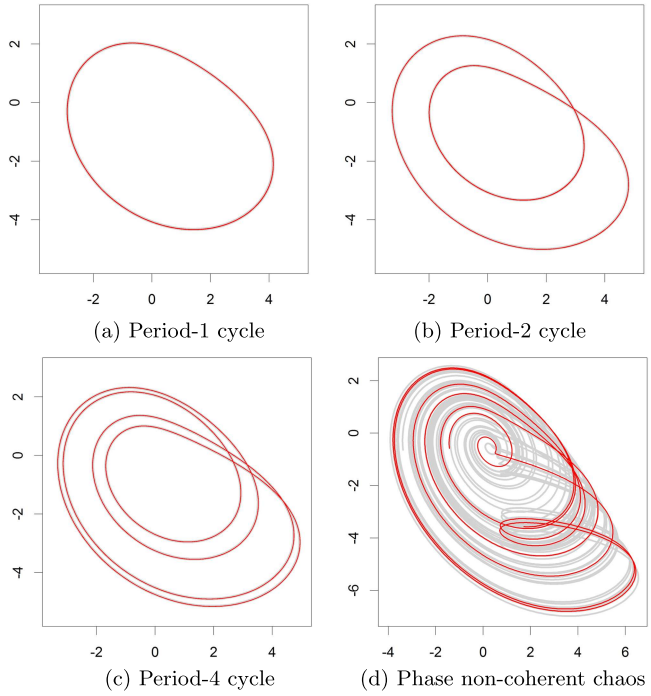


FIG. 4. Phase portraits of the Rössler-*1976* system for periodic regimes: period-1 cycle for $a = 0.33$ (a), period-2 for $a = 0.37$ (b), period-4 for $a = 0.38$ (c), phase non-coherent chaos for $a = 0.52$ (d). Other parameters: $(b = 2, c = 4)$. Original phase portrait are plotted in gray, global models in red. For periodic regimes, original and model phase portraits are indistinguishable because almost identical.

second class, models were very similar to the original portraits [see Fig. 4(d)] or even almost identical to the original phase portrait presented in Figs. 4(a)–4(c) for the periodic cases. In all the cases, the models obtained in the second class had exactly the same algebraic structure as the models #25, #40, #43, and #44 obtained for the reference case [Eqs. (6) and (7)]. The models corresponding to #25 were preferred in all the cases for their higher concision. The equations' original structure was thus retrieved in all cases, the parameters error progressively increasing from 0.36% for period-1 cycle to 1.08% for the phase non-coherent chaotic regime.

3. Time series length

To test the sensitiveness of the modeling technique to the time series length, experiments with time series of lengths varying from $N_{ech} = 400$ to 16 000 data points were tested, the smaller size corresponding to a single oscillation along the chaotic attractor. No model could be obtained for the smallest time series which probably comes from a lack of data leading to a low representativeness of the original phase space. Models could be obtained for all the time series of lengths equal to or longer than approximately two complete cycles ($N_{ech} \geq 700$). For these models, the parameter errors vary from 0.07% to 2.7%. The parameter errors were found to converge for a time series of duration $T = 24$ time units (1200 data points), corresponding to approximately four oscillations [see Fig. 3(b)]. Smaller errors were obtained for the shortest time series ($N_{ech} < 1000$ data points), which results from the fluctuations of the observability in the phase space.

4. Sampling time

The computation of the derivatives being necessary to apply the global modeling technique, a sufficiently good sampling time is thus required. Unfortunately, under experimental conditions, it is not always possible to choose the sampling time. Under subsampling conditions, it may be necessary to resample the observational time series at a higher resolution. Various levels of subsampling were considered in this experiment to investigate the conditions for retrieving the original equations' structure. Since it was shown in the previous paragraph that the error associated with time series length converges for a time series length of four oscillations ($T = 24$ units of time), this duration was used for the present experimentation. The analyzed time series were thus resampled at various sampling times from a time series of four oscillations to test the sensitiveness to the sampling time, decimating the number of points for each time series by a factor $N = 2$ to 80, corresponding to sampling times of $\Delta t = N \times \delta t$ with an initial sampling time $\delta t = 1/50$. Still following the same procedure, it was first attempted to get models from this set of subsampled time series. Results are presented in Table I. The algebraic structure of the original system could be retrieved for moderate subsamplings, with a parameter error quickly increasing from 2.66% (for $\Delta t = 0.04$) to 14.9% (for $\Delta t = 0.1$). A model could also be obtained for $\Delta t = 0.2$ but this included one extra term ($+0.055766x^2$) in the third equation and with error levels reaching 54.4% for the properly detected coefficients. All models were rejected for more degraded subsampling conditions (no erroneous detection).

The set of subsampled time series was then resampled using cubic splines, and the approach was applied again to the resampled time series (results also presented in Table I). It was found possible to retrieve the original system structure for $\Delta t = 0.04$ to 0.6. For more degraded conditions, results were found more erratic: models were all rejected for $\Delta t = 0.8$, 1.2, and 1.4, but a model could be obtained for $\Delta t = 1.0$ and 1.6. The formulations of the two latter models did not exactly correspond to the original formulation: six terms over seven were properly retrieved in both cases (i.e., one term undetected) but one extra term for $\Delta t = 1.4$, and two for $\Delta t = 1.6$, were erroneously detected. In terms of parameter error level, it is found that, for moderated original subsampling ($\Delta t = 0.02$ to 0.4), the parameter error remains extremely low ($< 1.5\%$). Obviously, time series resampling can improve highly the model structure detection and the precision of the model parameters. For more degraded sampling conditions $\Delta t \geq 50 \times \delta t$, which corresponds to seven or less data points per oscillations, higher error levels can be reached (up to 50%-70% here). Moreover, the chances to get a model becoming more erratic (higher probability to have all the model rejected) and

TABLE I. Structure and parameter maximum error (in %) of the models obtained from the subsampled and resampled time series as a function of the original sampling time Δt (with $T = 24$ for all the experiments).

N	1	2	5	10	20	30	40	50	60	70	80
Δt	0.02	0.04	0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
					Subsampled time series						
Proper	7	7	7	7	…	…	…	…	…	…	…
Missing	0	0	0	0	…	…	…	…	…	…	…
Extra	0	0	0	1	…	…	…	…	…	…	…
Err. (%)	0.67	2.66	14.9	54.[a]	…	…	…	…	…	…	…
					Resampled time series						
Proper	7	7	7	7	7	7	…	6	…	…	6
Missing	0	0	0	0	0	0	…	1	…	…	1
Extra	0	0	0	0	0	0	…	1	…	…	2
Err. (%)	0.67	0.67	0.68	0.72	1.33	5.6	…	61.[a]	…	…	53.[a]

[a]Estimated for the parameters detected by the model.

less robust (some erroneous detections) under such conditions, it may be useful to try several time series windows when attempting to get a model. Eight points per oscillations can be considered as good conditions for the global modeling technique, although models can be expected down to four. To check the robustness of the analysis, the same experiments were performed again considering a longer time period $T = 100$. Very similar results were obtained (see the supplementary material).

5. Measurement noise

Measurement noise is a common limitation when considering observational time series. It refers to perturbations that alter the measured signal without perturbing the observed system. Such noisy time series $s^{obs}(t)$ are easily generated by adding a noise to the original (free of noise) signal such as

$$s^{obs}(t) = s(t) + \epsilon_s(t), \tag{8}$$

where $\epsilon_s(t)$ is the measurement noise applied to the original time series $s(t)$ and $\epsilon(t)$ is assumed here to be independent and identically distributed (iid) Gaussian noise. The level of measurement noise is defined as

$$l_m = \sigma_m^2/\sigma_s^2 \times 100 \tag{9}$$

(in percent) with σ_m^2 the variance of the measurement noise $\epsilon(t)$ and σ_s^2 is the variance of the original signal $s(t)$; here, $(\sigma_x^2, \sigma_y^2, \sigma_z^2) \approx (1.854^2, 1.758^2, 1.122^2)$. Levels of noise ranging from 0.01% to 50% were applied. Following the procedure defined in Sec. IV A, the global modeling technique was applied to this set of noisy time series. However, the results in terms of model structure detection were found partly erratic, the original structure being recurrently partly retrieved, only. To investigate the probability to unveil the original system structure, an ensemble of 25 simulations was launched for each measurement noise level, and the modeling procedure was applied to each time series. Results are summarized in Table II. For each ensemble, results are separated in four classes: (a) Original system retrieved straightforward; (b) Original model approximately retrieved (i.e., with either one term undetected or one erroneous term); (c) Original model not retrieved (with more than one erroneous or missing detection); and (d) No model obtained (all the models are rejected). Results show that, even for low levels of noise ($l_m \leq 5\%$), there is a non-negligible probability (15.5%) to have an imperfect detection of the original system structure. For high level of noise ($l_m \geq 20\%$), the risk to obtain an erroneous model structure [class (c)] is also high (23.0%), although the probability not to obtain any model becomes much higher (74.0%). These results may be improved using a refined analysis. Indeed, when imperfect models are obtained, it is common to have several models of same size (number of parameters identical). In such conditions, there is no reason to prefer one model rather than another. In such a situation, the algebraic structure common to the valid models can be considered as more robust. The optimal structure is then obtained indirectly (e). The results after taking this refined analysis into account can thus be separated in four new classes: (f) Original system properly retrieved (directly or indirectly); (g) Original system almost perfectly retrieved (with either one undetected or erroneous term); (h) Original system erroneously retrieved (with more than one term erroneous or missing); the class (d) with no model detection remaining unchanged. This additional procedure enables to improve the detection. However, the risk to have an imperfect detection remains high (6% on average, against 15.5% before) even for low levels of noise, and the risk to have erroneous detections remains almost unchanged for higher noise levels. A smoothing was then applied to each time series and the global modeling applied again (results also reported in Table II). Smoothing appears very efficient to improve the model structure detection: the number of direct detection is globally increased, in particular, when higher levels of noise are concerned. No situation of class (c) and (h) is met anymore. Most of the detections with a single erroneous/undetected term (b) now give rise to a proper indirect detection (e). Situations for which no model can be

TABLE II. Structure and parameter maximum error (in %) of the models obtained under various levels of measurement noise.

l_m	0.01	0.2	0.5	1	2	5	10	20	30	40	50
					Noisy time series						
(a) Direct	23	16	21	22	22	21	13	3	0	0	0
(b) ±1	2	9	4	2	2	4	6	1	2	0	0
(c) $\pm n$	0	0	0	0	0	0	0	7	3	7	3
(d) No model	0	0	0	1	1	0	6	14	20	18	22
(e) Indirect	…	5	3	2	2	4	2	1	0	0	0
(f) Proper (a)+(e)	23	21	24	24	24	25	15	4	0	0	0
(g) Almost perfect	2	4	1	0	0	0	4	1	2	0	0
(h) Erronous	0	0	0	0	0	0	0	6	3	6	3
					Smoothed time series						
(a) Direct	22	22	23	22	24	24	23	20	21	20	15
(b) ±1	3	3	2	3	1	1	1	4	1	2	6
(c) $\pm n$	0	0	0	0	0	0	0	0	0	0	0
(d) No model	0	0	0	0	0	0	1	1	3	3	4
(e) Indirect	3	3	2	3	1	1	1	4	0	1	3
(f) Proper (a)+(e)	25	25	25	25	25	25	24	24	21	21	18
(g) Almost perfect	0	0	0	0	0	0	0	0	1	1	3
(h) Erronous	0	0	0	0	0	0	0	0	0	0	0

obtained (d) becomes rare and only for the highest levels of noise. In terms of parameter error [see Fig. 5(a)], it is found that the effect of noise becomes more effective for noise levels such as $l_m \geq 0.5\%$. Error increases regularly with rising noise for all the parameters but one in the second equation which error remains stable even for higher levels of noise. In several cases, parameter smoothing [see Fig. 5(b)] will tend to increase the parameter error [see the error parameters for the third equation in Eq. (5)], although this is not systematic. The main interest of the smoothing is to make the global modeling technique applicable and much more robust in terms of model structure detection, especially when no model would be obtained otherwise.

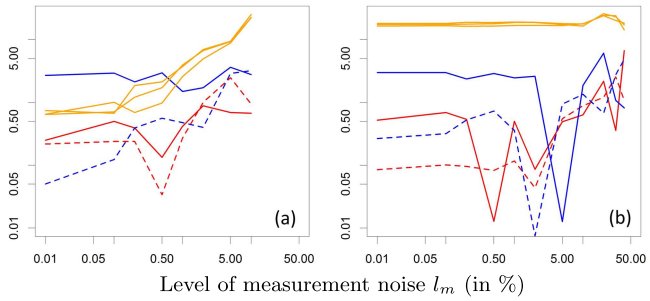


FIG. 5. Sensitivity to measurement noise before (a) and after (b) smoothing. Models' parameter error (in %) is plotted as a function of the percentage of added noise l_m (respectively, up to 10% and 50% for non-smoothed and smoothed signals). Same color codes as Fig. 3.

6. Dynamical perturbation

Dynamical perturbation is defined here as a perturbation of the original dynamical system that will alter the current state of the studied system. Such a perturbation may thus change the system long term evolution. It is therefore fundamentally different from the measurement noise which will affect only the observational time series not the system itself. To investigate the influence of dynamical noise, observational time series must be generated by perturbing the dynamical system during the integration process. The Rössler system can thus be rewritten as stochastic differential equations

$$\begin{cases} \dot{x} = -y - z + \mu_x(t), \\ \dot{y} = \ x + ay \ + \mu_y(t), \\ \dot{z} = b + z(x - c) + \mu_z(t), \end{cases} \tag{10}$$

where μ_x, μ_y, and μ_z are assumed to be iid Gaussian noise of amplitude level

$$l_d = \sigma_d^2/\dot{\sigma}_s^2 \times 100 \tag{11}$$

(in percent) with σ_d^2 is the variance of the dynamical noise $\mu_.(t)$ applied to variables x, y, or z at each integration time step, and $\dot{\sigma}_s^2$ is the variance of the original signal derivative $\dot{s}(t)$; in the present case (with $\delta t = 1/50$), we have $(\dot{\sigma}_x^2, \dot{\sigma}_y^2, \dot{\sigma}_z^2) \approx (0.0395^2, 0.0339^2, 0.0348^2)$. In practice, $\mu_x(t)$, $\mu_y(t)$, and $\mu_z(t)$ are chosen in order to have the same level of dynamical noise l_d for each variable. Levels of dynamical noise ranging from $l_d = 1\%$ to 800% were applied (note that, compared to the variance of the original variables, this level would appear three orders of magnitudes lower since $1040 > \sigma_s^2/\dot{\sigma}_s^2 > 2700$). The procedure presented in Sec. IV A was then applied to this set of noisy time series. For high dynamical noise levels ($l_d \geq 700\%$), models were found to be all systematically rejected in

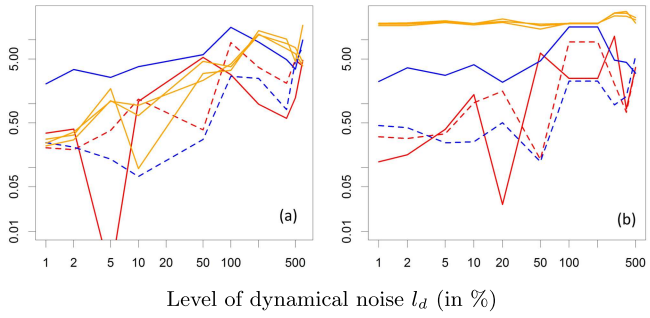


FIG. 6. Sensitivity to dynamical noise before (a) and after (b) smoothing. Models' parameter error (in %) is plotted as a function of the percentage of added noise l_m (up to 600% for both non-smoothed and smoothed signals). Same color codes as Fig. 3.

both noisy and smoothed cases. The analysis will thus focus on noise levels such as $l_d \leq 600\%$. Starting with the non-smoothed time series, the original formulation of the Rössler system could be retrieved in almost all cases with an error on the parameters increasing with the level of noise (see Fig. 6). The effect of the dynamical noise on the parameter error level appears qualitatively very similar to the effect of measurement noise (Fig. 5). Punctually, for $l_d = 20\%$ and 300%, all the models were rejected (no detection). Starting from the same time series after a smoothing was applied to them, the original Rössler system was retrieved in all the cases except for the largest level of noise ($l_d = 600\%$) for which a model was obtained which algebraic structure of the last equation did not entirely match the original system (the constant term b was missing and two extra terms were erroneously detected; note, however, that the magnitude of the spurious terms x^2 and z^2 was comparatively low). The approach can thus be considered as robust to dynamical noise although some erroneous detection may happen when dynamical perturbation level becomes too high. A smoothing preprocessing appears superfluous to deal with dynamical noise. The Savitzky-Golay filter applied to compute the derivatives that will also have a slight smoothing effect on the original time series (see Sec. III) is therefore sufficient to tackle with low to moderate levels of dynamical perturbation.

C. Other quadratic three-dimensional systems

In Sec. IV B, all the experiments were based on the Rössler system in order not to mix the influence of the system structure with the one of the tested factor. The object of Sec. IVC is now to test other algebraic structures. Nineteen other dynamical systems are now considered, all three-dimensional and quadratic. In each case, the procedure introduced in Sec. IV A is used; additional details are punctually given when further analyses are required. Results are synthesized in Table III, and additional information are provided as the supplemental material.

The Nosé-Hoover (NH84) system[28,29]

$$\dot{x} = ay; \quad \dot{y} = yz - x; \quad \dot{z} = b - y^2 \tag{12}$$

is a very interesting case of study because its three variables—although fully deterministically coupled—are almost completely uncorrelated. For the parameterization $(a, b) = (0.2, 3)$, its dynamic is chaotic and the correlation between the three variables does not exceed 0.15 ($C_{x,z} \approx 0.142$, $C_{x,y} \approx 0.025$, and $C_{y,z} \approx 0.016$). It has only five terms among which two nonlinear ones (yz in the second equation, y^2 in the third one). The model

$$\begin{cases} \dot{x} = 0.196y, \\ \dot{y} = 0.854yz - 1.08x, \\ \dot{z} = 2.744 - 0.916y^2 \end{cases} \tag{13}$$

was obtained when applying the global modeling technique. Despite very low correlation levels, the structure of the original system is directly retrieved with a precision on the parameters ranging from 2% to 15%.

The same procedure was applied to the Genesio-Tesi (GT92),[30] and the Sprott-F, G, H, K, M, O, P, Q, and S systems,[31] that all have six terms with a single nonlinear one included. Their formulation is thus very concise. In all these cases, still following the same procedure, the original system structure was retrieved directly from observational time series, the parameter error varying from 0.1% to 2.1%.

Systems of higher algebraic complexity were then considered. The Lorenz-1963 (L63) system[32]

$$\dot{x} = ay - ax; \quad \dot{y} = -y + bx - xz; \quad \dot{z} = -cz + xy \tag{14}$$

has seven terms including two nonlinearities (xz and xy in the second and third equations, respectively). For the parameterization $(a, b, c) = (10, 28, 8/3)$, it produces the famous two-scroll chaotic attractor. The following model was obtained:

$$\begin{cases} \dot{x} = 9.748y - 9.748x, \\ \dot{y} = 24.482x - 0.913xz, \\ \dot{z} = -2.581z + 0.968xy, \end{cases} \tag{15}$$

in which one term ($-y$ in the second equation) is missing. The seven-term original formulation was actually included in the preselected models but it could not be retained because of its less parsimonious formulation. The obtained model corresponds to one of the three smallest subsystems of the general Lorenz-like (sLL1-3) systems discovered by Lainscsek[24] and characterized by an identical differential formulation. This result highlights here the ability of the approach to retrieve minimal algebraic formulation of the equations rather than the original equations. Note that the sLL1-3 systems are algebraically equivalent exclusively when reformulating the dynamics from variable x. But their original algebraic formulations being different, it is thus possible to retrieve their original structure when the three variables (x, y, z) are observed, as confirmed here by applying the global modeling technique. This result clearly confirms the ability of the approach to distinguish equations of minimal formulation one from another.

TABLE III. Systems detected and undetected structure, and sensitivity to noise: system name, dimension n, polynomial degree q, number of monomial (nonlinear ones included), parameter error range, number of properly detected terms (nonlinear ones included), of undetected ones (nonlinear ones included), and of erroneous detection, maximum measurement noise l_m^{max} for which the result could be confirmed.

System	n	q	N_P	err	Proper	Miss.	Erroneous	l_m^{max}
R76	3	2	7(1)	[0.09; 0.6]	7(1)	0	0	50
NH84	3	2	5(2)	[2.0; 14.7]	5(2)	0	0	50
GT92	3	2	6(1)	[0.6; 2.1]	6(1)	0	0	10
SprF	3	2	6(1)	[0.6; 2.0]	6(1)	0	0	20
SprG	3	2	6(1)	[0.5; 1.8]	6(1)	0	0	40
SprH	3	2	6(1)	[0.9; 1.8]	6(1)	0	0	20
SprK	3	2	6(1)	[0.6; 1.9]	6(1)	0	0	5
SprM	3	2	6(1)	[0.1; 0.3]	6(1)	0	0	40
SprO	3	2	6(1)	[0.1; 0.6]	6(1)	0	0	2
SprP	3	2	6(1)	[0.1; 0.4]	6(1)	0	0	10
SprQ	3	2	6(1)	[0.2; 0.4]	6(1)	0	0	2
SprS	3	2	6(1)	[0.3; 0.5]	6(1)	0	0	5
BS81	3	2	6(2)	[2.3; 5.4]	5(2)	1(0)	0	30
L63	3	2	7(2)	[2.5; 12.6]	6(2)	1(0)	0	5
sLL1	3	2	6(2)	[2.9; 5.2]	6(2)	0	0	5
sLL2	3	2	6(2)	[0.7; 1.4]	6(2)	0	0	20
sLL3	3	2	6(2)	[2.5; 4.9]	6(2)	0	0	5
Li12	3	2	9(4)	[3.5; 12.0]	9(4)	0	0	2
cord12	3	2	11(4)	[8.4; 30.1]	6(3)	5(1)	0	10
L84	3	2	11(6)	[3.4; 66.2]	6(4)	5(2)	0	10
M02	3	3	5(1)	[0.01; 0.1]	5(1)	0	0	0.2
RFOsc	4	3	9(3)	[1.1; 23.7]	8(3)	1(0)	0	0.5
dyn15	5	2	15(3)	[0.05; 43.8]	10(3)	5	0	0.2

The same procedure was also applied to four other systems: the Lorenz-like Burke-Shaw (BS81) system,[33] the Li (Li12) system[34] that produces toroidal chaos, the Lorenz-1984 (L84)[35] that produces a weakly dissipative chaos, and the cord system (cord12).[36] The completed algebraic structure was retrieved directly only for the Li12 system. Simplified algebraic structures (i.e., one or more undetected terms) were obtained for BS81, L84, and cord12 systems, without erroneous detection (see Table III).

A sensitivity test to measurement noise was applied to each system. This showed that the approach is robust to noise although sensitivity can highly vary from one system to another (from $l_m^{max} = 2\%$ for SprO, SprQ, and Li12 up to 50% for R76 and NH84, see Table III for details).

D. Non-quadratic and higher-dimensional systems

The approach was then tested on three other systems, higher-dimensional and/or cubic. The five-dimensional (dyn15) system[37]

$$\begin{cases} \dot{x}_1 = -ax_1 + x_2x_3 - px_4 + qx_5, \\ \dot{x}_2 = -bx_1 - ax_2 + x_1x_3 - px_4 + qx_5, \\ \dot{x}_3 = 1 - x_1x_2, \\ \dot{x}_4 = cx_2, \\ \dot{x}_5 = q(x_1 + x_2 + x_4) \end{cases} \tag{16}$$

derived from the Rikitake dynamo system[38] has 15 terms, among which three nonlinearities, and produces a hyperchaotic behavior for $a = 1, b = 1, c = 0.7, p = 1.1$, and $q = 0.1$. The 10-term model

$$\begin{cases} \dot{x}_1 = -1.185x_1 + 1.089x_2x_3, \\ \dot{x}_2 = -2.095x_1 - 0.706x_2 - 0.76x_4 + 0.95x_1x_3, \\ \dot{x}_3 = 0.998 - 0.998x_1x_2, \\ \dot{x}_4 = 0.700x_2, \\ \dot{x}_5 = 0.144x_1 \end{cases} \tag{17}$$

was the smaller model able to reproduce the two scrolls observed in the original (x_1, x_2) projection. Compared to the original system (16), five terms are missing, without any erroneous detection.

Compared to quadratic systems, cubic systems are relatively rare. The simplest cubic system (M02)

$$\dot{x} = y; \quad \dot{y} = z; \quad \dot{z} = -az - x + xy^2 \tag{18}$$

was introduced by Malasoma.[39] No quadratic model could be obtained for this case, and the original system structure could be entirely retrieved with its cubic term, with parameter error lower than 0.1%.

The Reciprocally Forced Oscillators (RFOsc)

$$\begin{cases} \dot{x} = y & \dot{y} = aw - by - cx^3, \\ \dot{z} = w & \dot{w} = dw - ez^2w - z^3 + fy \end{cases} \tag{19}$$

is introduced here in order to have a study case of mixed complexity ($n = 4, q = 3$). For $(a, b, c, d, e, f) = (0.7, 0.01, 5, 0.2, 2.206, 0.25)$, a chaotic regime is produced. No valid quadratic model could be obtained for it. The best model able to reproduce the complexity of the original phase space is

$$\begin{cases} \dot{x} = 0.968y, & \dot{y} = 0.534w - 4.56x^3, \\ \dot{z} = 0.985w, & \dot{w} = 0.179w - 1.99z^2w - 0.99z^3 + 0.24y \end{cases} \tag{20}$$

in which one term is undetected ($-by$ in the second equation) and that does not include any erroneous detection. The sensitivity test to measurement noise showed that systems MO2, RFOsc, and dyn15 are more sensitive to noise than the 3D quadratic systems.

The results obtained in Secs. IV C and IV D clearly illustrate the fact that the algebraic structure of dynamical systems can be retrieved, and often completely, when the system is polynomial and sufficiently concise. This proves that statistical criteria (such as the ones used in the structure selection process, see Sec. IV C) can be very efficient for model preselection. Nonetheless, since such statistical properties depend on the model complexity (number of parameters, polynomial degree, etc.), the same properties may be obtained with different model structures (as shown by Ref. 25). By construction, classical statistical properties are not adapted to characterize unambiguously the deterministic properties of a dynamical system, in particular, its phase space. As a consequence, such properties may not be sufficient to identify the best model and may not enable the retrieval of the the equations original structure. Note that criteria including a penalty increasing with model size should *a fortiori* be avoided in the structure selection process because such criteria will add a strong constraint on the model size (and thus on its structure) generally without any objective knowledge. From this respect, a qualitative comparison between the original and the model phase portraits appears to be a much more powerful criterion to reject spurious models among potential models preselected on statistical properties. At the mean time, it was also found that the less parsimonious the systems, the more difficult the complete retrieval of the original equations. Several factors can contribute to this limitation: (1) More terms in an equation will lead to increase parameters error (and the residual signal) during the identification process. (2) Terms of low amplitude will be more affected by such a situation and therefore more difficult to detect unambiguously. (3) Polynomial terms may also be more difficult to distinguish one from another when variables are close to synchronization. Finally (4), although polynomial enables to define a non-redundant basis, non-uniqueness cannot be guaranteed systematically under numerical conditions due to finite precision and finite data length. For these reasons, models of more concise form should be systematically preferred if the validation cannot justify the contribution of additional terms. The principle of parsimony (Ockham's razor) appears to be a powerful element of selection for equations structure identification but it should advantageously take the dynamical properties into account rather than common statistical properties.

E. Non-polynomial systems

To test the approach under non-polynomial conditions, four other systems with rational terms were also investigated: the host-immune-tumor system (HIT01),[40] the phyplankton-zooplakton-nutrient system (PZN14),[41] as well as one equivariant double cover with Reflection symmetry (EDCR) and one double cover with $R_Z(\pi)$ symmetry (DCπ) of the R76 system.[42]

The HIT01 system

$$\begin{cases} \dot{x} = \rho_1 x(1 - x) - \alpha_{13} xz, \\ \dot{y} = \dfrac{\rho_2 yz}{1 + z} - \alpha_{23} yz - \delta_2 y, \\ \dot{z} = \rho_3 z(1 - z) - \alpha_{31} xz - \alpha_{32} yz \end{cases} \tag{21}$$

produces a chaotic behavior for $\rho_1 = 0.518, \alpha_{13} = 1.5, \rho_2 = 4.5, \alpha_{23} = 0.2, \delta_2 = 0.5, \rho_3 = 1, \alpha_{31} = 1$, and $\alpha_{32} = 2.5$.[43] The model

$$\begin{cases} \dot{x} = 0.5171x - 1.498x^2 - 0.5171xz, \\ \dot{y} = 2.7697yz - 0.4029y, \\ \dot{z} = 0.9823z - 0.9836z^2 - 2.464xz - 0.9773yz \end{cases} \tag{22}$$

was obtained for it. All the terms are properly detected (i.e., neither erroneous nor missing detection) except the rational one which was approximated by a Taylor series to zeroth order (see the supplementary material). A similar result was obtained for the EDCR system, except that the rational term was approximated by a Taylor series to second order. For the other two systems (PZN14 and DCπ), the models were all rejected (no spurious detection). The equations of all the systems and models are provided as the supplementary material.

These examples show that the obtained polynomial models can be the result of an approximation. These experiments also show that the probability to fail to detect a deterministic behavior is high when the original system is non-polynomial. Though, it was proven that polynomial can be often obtained to approximate rational functions in the case of differential models reconstructed from univariate time series, for instance, for modeling the Rössler-76 dynamics from variables x or z.[7,12] Such a situation is, however, much less probable when all the system variables are observed since an approximation of a rational can be a big source of perturbation. Multivariate global models are much more sensitive than univariate ones because errors will propagate from all the equations in the multivariate case whereas it will mostly originate from a single equation in the univariate one. It is probably another reason why multivariate models in ODEs could not be obtained earlier from real world data.

V. CONCLUSIONS

Several experiments have been carried out in this study to investigate under which conditions, the original equations of dynamical systems can be retrieved, in part or entirely, from

observational time series, when all the system variables are observed. It is shown that, for polynomial systems of concise formulation, the original structure can be exactly and entirely retrieved. For polynomial systems with more terms, most of the terms of the original equations can generally be efficiently retrieved and erroneous terms are found to be rare and comparatively of much smaller amplitudes.

The sensitivity tests showed that the approach is independent to the initial conditions, and that, up to a certain extent, it is robust to measurement noise and dynamical perturbations (respectively, up to $l_m = 50\%$ of the original signal variance and $l_d = 500\%$ of the incremental signal variance for the Rössler-76 system). However, sensitivity can highly vary when considering other systems. It is also shown that a signal of two oscillations and a sampling time of four to six data points per oscillation may be sufficient to retrieve the original system equation structure. Experiments show that smoothing and resampling preprocessing can improve considerably the detection of the original equations under measurement noise and subsampling conditions, respectively.

It is also proven that the approach can be potentially applied to dynamical systems of various dimensions and not exclusively to systems of quadratic form and to very diversified dynamical regimes (periodic, chaotic, hyperchaotic, toroidal or not, phase coherent and phase non-coherent chaos, weakly and strongly dissipative chaos). It is noted that, using the algorithms and procedures introduced here, obtaining models with erroneous terms is relatively rare but that terms may be undetected when the original structure is poorly concise. It is observed that, using the present technique, it is much more common not to detect any model rather than to detect a model with a single or more erroneous terms. One consequence of this latter observation is that alternative analysis windows can be tried when no model, or non-valid models, are obtained.

It is also proven that the approach can be applied to detect deterministic couplings even when the variables of the system are almost fully uncorrelated ($C < 0.15$).

To investigate the generality of the results, the approach was also tested on four non-polynomial systems. For the two more complex systems, all the models were rejected (no spurious detection). For the two other systems, the polynomial part of the systems could be detected and the rational part approximated by a Taylor series.

The present study strongly supports the possibility to retrieve the exact original system structure when the original system formulation is polynomial and sufficiently concise and to obtain a good polynomial approximation of the original system otherwise. These overall conclusions therefore suggest that process-based interpretations can be reasonably proposed for models directly derived from observational time series.

SUPPLEMENTARY MATERIAL

The supplementary material includes information about the time series used for the analyses. The results of an alternative test to subsampling. The equations of the dynamical systems and of the models obtained for them under free of noise conditions.

ACKNOWLEDGMENTS

Part of this work was supported by the French programs LEFE (INSU) and Défi InFiNiTy (CNRS).

REFERENCES

[1]L. A. Aguirre and C. Letellier, "Modeling nonlinear dynamics and chaos: A review," Math. Probl. Eng. **2009**, 238960 (2009).
[2]P. C. Young, "Parameter estimation for continuous-time models—A survey," Automatica **17**(1), 23–39 (1981).
[3]D. C. Saha and G. P. Rao, *Identification of Continuous Dynamical Systems—The Poisson Moment Functionals, (PMF) Approach* (Springer-Verlag, Berlin, 1983).
[4]S. A. Billings, "Structure detection and model validity tests in the identification of nonlinear systems," ACSE Rep. **196**, 22 (1982).
[5]S. A. Billings and I. J. Leontaritis, "Parameter estimation techniques for nonlinear systems," IFAC Proceedings **15**(4), 505–510 (1982).
[6]J. P. Crutchfield and B. S. McNamara, "Equations of motion from a data series," Compl. Syst. **1**, 417 (1987).
[7]G. Gouesbet and C. Letellier, "Global vector field reconstruction by using a multivariate polynomial L_2-approximation on nets," Phys. Rev. E **49**(6), 4955–4972 (1994).
[8]C. Letellier, L. Le Sceller, E. Maréchal, P. Dutertre, B. Maheu, and G. Gouesbet, "Global vector field reconstruction from a chaotic experimental signal in copper electrodissolution," Phys. Rev. E **51**(5), 4262–4266 (1995).
[9]C. Letellier, L. A. Aguirre, and U. S. Freitas, "Frequently asked questions about global modeling," Chaos **19**, 023103 (2009).
[10]C. Lainscsek, C. Letellier, and C. Schürrer, "Ansatz library for global modeling with a structure selection," Phys. Rev. E **64**, 016206 (2001).
[11]C. Lainscsek, C. Letellier, and I. Gorodnitsky, "Global modeling of the Rössler system from the z-variable," Phys. Lett. A **314**(5–6), 409 (2003).
[12]S. Mangiarotti, R. Coudret, L. Drapeau, and L. Jarlan, "Polynomial search and global modeling: Two algorithms for modeling chaos," Phys. Rev. E **86**, 046205 (2012).
[13]S. Mangiarotti, F. Le Jean, M. Huc, and C. Letellier, "Global modeling of aggregated and associated chaotic dynamics," Chaos Solitons Fractals **83**, 82–96 (2016).
[14]C. Letellier, S. Mangiarotti, I. Sendiña-Nadal, and O. E. Rössler, "Topological characterization versus synchronization for assessing (or not) dynamical equivalence," Chaos **28**(4), 045107 (2018).
[15]S. Mangiarotti, L. Drapeau, and C. Letellier, "Two chaotic global models for cereal crops cycles observed from satellite in northern Morocco," Chaos **24**(2), 023130 (2014).
[16]S. Mangiarotti, "Modélisation globale et caractérisation topologique de dynamiques environnementales," Theses and habilitation to direct researches (Université de Toulouse III, 2014).
[17]S. Mangiarotti, A. K. Sharma, S. Corgne, L. Hubert-Moy, L. Ruiz, M. Sekhar, and Y. Kerr, "Can the global modelling technique be used for crop classification?," Chaos Solitons Fractals **106**, 363–378 (2018).
[18]J. Maquet, C. Letellier, and L. A. Aguirre, "Global models from the Canadian lynx cycles as a direct evidence for chaos in real ecosystems," J. Math. Biol. **55**(1), 21–39 (2007).
[19]S. Mangiarotti, "Low dimensional chaotic models for the plague epidemic in Bombay (1896-1911)," Chaos Solitons Fractals **81**, 184–196 (2015).
[20]S. Mangiarotti, M. Peyre, and M. Huc, "A chaotic model for the epidemic of Ebola virus disease in West Africa (2013-2016)," Chaos **26**, 113112 (2016).
[21]L. A. Aguirre and S. A. Billings, "Retrieving dynamical invariants from chaotic data using NARMAX models," Int. J. Bifurcat. Chaos **9**(2), 449–474 (1995).

[22]S. Mangiarotti, I. Sendiña-Nadal, and C. Letellier, "Using global modelling to unveil hidden couplings in small network motifs," Chaos **28**, 123110 (2018).
[23]C. Lainscsek, "Nonuniqueness of global modeling and time scaling," Phys. Rev. E **84**, 046205 (2011).
[24]C. Lainscsek, "A class of Lorenz-like systems," Chaos **22**, 013126 (2012).
[25]A. M. Barbosa, R. H. C. Takahashi, and L. A. Aguirre, "Equivalence of nonlinear model structures based on Pareto uncertainty," IET Control Theory Appl. **9**(16), 2423–2429 (2015).
[26]L. A. Aguirre and S. A. Billings, "Dynamical effects of overparametrization in nonlinear models," Physica D **80**(1–2), 26–40 (1995).
[27]O. E. Rössler, "An equation for continuous chaos," Phys. Lett. A **57**(5), 397–398 (1976).
[28]S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," Mol. Phys. **52**(2), 255–268 (1984).
[29]W. G. Hoover, "Nonlinear conductivity and entropy in the two-body Boltzmann gas," J. Stat. Phys. **42**(3–4), 587–600 (1986).
[30]R. Genesio and A. Tesi, "Harmonic balance methods for the analysis of chaotic dynamics in nonlinear systems," Automatica **28**(3), 531–548 (1992).
[31]J. C. Sprott, "Some simple chaotic flows," Phys. Rev. E **50**(2), 647–650 (1994).
[32]E. N. Lorenz, "Deterministic non-periodic flow," J. Atmos. Sci. **20**(2), 130–141 (1963).
[33]R. Shaw, "Strange attractor, chaotic behavior and information flow," Z. Naturforschung A **36**, 80–112 (1981).
[34]D. Li, "A three-scroll chaotic attractor," Phys. Lett. A **372**, 387–393 (2007).
[35]E. N. Lorenz, "Irregularity: A fundamental property of the atmosphere," Tellus **36A**, 98–110 (1984).
[36]C. Letellier and L. A. Aguirre, "Required criteria for recognizing new types of chaos: Application to the cord attractor," Phys. Rev. E **85**, 036204 (2012).
[37]S. Vaidyanathan, V.-T. Pham, and C. K. Volos, "A 5-D hyperchaotic Rikitake dynamo system with hidden attractors," Eur. Phys. J. **224**, 1575–1592 (2015).
[38]T. Rikitake, "Oscillations of a system of disk dynamos," Proc. Cambridge Philosoph. Soc. **54**, 89–105 (1958).
[39]J.-M. Malasoma, Chaos Solitons Fractals **13**, 1835–1842 (2002).
[40]L. G. De Pillis and A. Radunskaya, "A mathematical tumor model with immune resistance and drug therapy: An optimal control approach," J. Theor. Med. **3**, 79–100 (2001).
[41]A. Sharma, A. K. Sharma, and K. Agnihotri, "The dynamic of plankton-nutrient interaction with delay," Appl. Math. Comput. **231**, 503–515 (2014).
[42]R. Gilmore and C. Letellier, *The Symmetry of Chaos* (Oxford University Press, New York, 2007).
[43]M. Itik and S. P. Banks, "Chaos in a three-dimensional cancer model," Int. J. Bifurcat. Chaos **20**(1), 71–79 (2010).