Modeling dynamics from only output data

Evan Monroig,[1,][*] Kazuyuki Aihara,[2] and Yozo Fujino[1]

[1]*Department of Civil Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[2]*Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*

We examine the problem of reconstructing input-output systems from time series data. Although the method of delays has already been used in the case where both input and output are measured, in some cases, the inputs cannot be measured, and hence, the method of delays cannot be used. On the basis of ideas derived from existing embedding theorems, we propose to build models by using delays of multivariate observations of output data. Assuming that the inputs are few, we use several observations for obtaining information about the inputs, and the remaining observations for obtaining information about the state of the system. Numerical examples on a discrete map and a continuous-time system show that input-output systems can indeed be identified by using multivariate observations of output data only. We also discuss the application of this method to the analysis of coupled systems or complex networks, by partitioning such large systems and analyzing each subsystem separately. The models used in this paper are nonpredictive models; thus, they cannot be used to predict the future behavior of the system. However, since they model the dynamics of the system, they have other possible applications such as change detection and noise reduction.

I. INTRODUCTION

In the analysis of experimental time series originating from low-dimensional nonlinear dynamical systems, the *method of delays* [1] is widely used for reconstructing a state space in which the dynamics can be viewed. This method is theoretically justified by Takens's embedding theorem [2,3]. Thus, for an autonomous deterministic dynamical system, a scalar time series can contain sufficient information to reconstruct a state space that is equivalent to the state space of the original dynamical system. One can then use the reconstruction, for example, to estimate dynamical invariants of the system, to predict future values of the time series, to remove noise, or to control the system (see [4,5]).

Since Takens's theorem applies only to autonomous deterministic systems, strictly speaking, it cannot be directly applied to a measured time series because of the presence of noise or an external input. First, signals obtained from real systems are affected by *observational noise*, which modifies the measured variable. A second type of noise is *dynamical noise*, which acts on the internal state of the systems; such systems are known as *stochastic systems*. Finally, in many fields, *input-output systems* are considered instead of autonomous systems.

Many studies have been carried out in order to overcome these difficulties. The effect of observational noise on the reconstruction of a state space has been well studied [6]. The applicability of such reconstruction to input-output systems was conjectured by Casdagli [7], and indeed, several embedding theorems were later proved for deterministically forced systems [8] and for stochastic systems and input-output systems [9]. See [10] for a review of these theorems.

Another question that recurs in the literature, but for which embedding theorems have not yet been stated or

proved [11], is the use of *multivariate observations* in the reconstruction of a state space [12–16]. If the dynamics of a dynamical system can be reconstructed from a scalar variable, then in certain cases, the reconstruction will be intuitively improved by the use of several variables. Muldoon *et al.* [17] showed that by performing many simultaneous observations such that an embedding of the original state space becomes automatic, even the noise can be reconstructed.

In this paper, on the basis of ideas derived from the theorem given by Muldoon *et al.* [17] and from Takens's theorem for stochastic systems [9], we propose to build models for low-dimensional deterministic input-output systems with unobserved inputs. We assume that the inputs are few, and we separate multivariate observations into primary observations and secondary observations. The secondary observations are used for obtaining information about the input, while delays of the primary observations are used for obtaining information about the state of the system.

In Sec. II, we review some of the existing embedding theorems. In Sec. III, we compare our approach to the existing embedding theorems and propose some theoretical arguments for the existence of a functional relation between delayed multivariate observations. In Sec. IV, we verify the functional relations of Sec. III by analyzing three numerically simulated systems: a discrete map with an external input, a continuous-time system with an external input, and finally two coupled continuous-time systems. We evaluate the quality of the reconstruction on the basis of predictability. In Sec. V, we summarize and discuss this research and its possible applications.

II. PHASE SPACE RECONSTRUCTION

A. Takens's embedding theorem

Consider a deterministic finite-dimensional dynamical system whose state at time t is represented by a vector x on

---

*evan.monroig@m4x.org

a compact m-dimensional manifold $\mathcal{M}$. The time evolution of the system is given by a map $f^t:\mathcal{M}\rightarrow\mathcal{M}$, so that if at time t_0 the system is in state x_0, then at time $t+t_0$ its state is $x(t+t_0)=f^t(x_0)$. Such a map is usually given by the solution of a differential equation on $\mathcal{M}$.

In most applications, the state x is observed not directly but through a scalar observable y that is related to x by an observation function $\varphi:\mathcal{M}\rightarrow\mathbb{R}$. The system is observed at discrete times $t_k=t_0+k\tau$ $(k=0,1,\dots)$ with a sampling interval τ in order to obtain a scalar time series $\{y_k\}$,

$$y_k = \varphi(x_k), \tag{1}$$

where $x_k=f^{k\tau}(x_0)$. The states at discrete times $t_0,t_1,\dots$ are related by the equation

$$x_{k+1} = f(x_k), \tag{2}$$

where the notations were simplified by rescaling the time so that $\tau=1$ and $f=f^\tau$.

On the basis of the observed scalar time series $\{y_k\}$, it is possible to reconstruct a multidimensional state space that is equivalent to the state space of the original system; from such a state space, one can then build predictive models. One way of doing so is to use the method of delays [1]: use successive delays of the time series $\{y_k\}$ to build a new d-dimensional variable $\mathbf{v}_k$,

$$\mathbf{v}_k = (y_k,y_{k+1},\dots,y_{k+d-1}).$$

The integer d is called *embedding dimension*, and $\mathbf{v}_k$ is called *reconstruction vector* or *delay vector*.

From Eqs. (1) and (2), it follows that $\mathbf{v}_k$ and x_k are related by the equation

$$\mathbf{v}_k = \Phi(x_k), \tag{3}$$

where $\Phi:\mathcal{M}\rightarrow\mathbb{R}^d$ is called *delay map* or *reconstruction map* and is defined as follows:

$$\Phi(x) = \Phi_{f,\varphi}(x) = [\varphi(x),\varphi(f(x)),\dots,\varphi(f^{d-1}(x))].$$

Takens's embedding theorem [2,3,10] then shows that for $d\geq 2m+1$ and for *typical* f and φ of class C^2, the delay map Φ is a smooth coordinate change between the original state space, where x_k lies, and the reconstructed state space, where $\mathbf{v}_k$ lies.

To be precise, the delay map is then an *embedding*, i.e., it maps $\mathcal{M}$ diffeomorphically onto its image: it is one-to-one and continuously differentiable, and its inverse defined on $f(\mathcal{M})$ is also continuously differentiable. For a precise statement of Takens's theorem, refer to [10].

Since Φ is an embedding, a predictive relation can be obtained in the following manner. The coordinate change Φ can be used to transport the map f, which becomes $F=\Phi\circ f\circ\Phi^{-1}$ in the reconstructed coordinates, where the symbol "$\circ$" denotes function composition. Then, from Eq. (3), we know that $\mathbf{v}_k=\Phi(x_k)$ is in the image of Φ. Thus, F can be applied to it;

$$F(\mathbf{v}_k) = \Phi\circ f\circ\Phi^{-1}(\mathbf{v}_k),$$

$$= \Phi\circ f\circ\Phi^{-1}\circ\Phi(x_k) \quad\text{by Eq. (3)},$$

$$= \Phi\circ f(x_k),$$

$$= \Phi(x_{k+1}) \quad\text{by Eq. (2)},$$

$$= \mathbf{v}_{k+1} \quad\text{by Eq. (3)}.$$

By rewriting the above equation in terms of the observed time series $\{y_k\}$, we can express the equations of dynamics of $\mathbf{v}_k$ as follows:

$$(y_{k+1},y_{k+2},\dots,y_{k+d}) = F(y_k,y_{k+1},\dots,y_{k+d-1}).$$

Here, the first $d-1$ components of F are a trivial shift of coordinates, and hence, if we write the last component of F as $G:\Phi(\mathcal{M})\rightarrow\mathbb{R}$, we obtain the following predictive relation:

$$y_{k+d} = G(y_k,y_{k+1},\dots,y_{k+d-1}).$$

The function G can be estimated from observed data; thus, the future of the time series $\{y_k\}$ can be predicted.

B. Generalizations to forced systems

We now consider a discrete-time system that is excited by an external force, namely, a forced system. The state at discrete time k is represented by a vector x_k on a compact m-dimensional manifold $\mathcal{M}$, and a multidimensional external input u_k, which lies on a compact n-dimensional manifold $\mathcal{N}$, is added.

The input u is considered to be a stochastic input or the state of a dynamical system that is coupled to the considered system. If the input is the state of a low-dimensional deterministic system, such that, for example, $u_{k+1}=g(u_k)$, then by using the methods of Stark [8], it is possible to simultaneously reconstruct both systems and obtain a predictive relation; this case is not considered here. Also, a nonstationary input is not considered.

The evolution of the state of the system is given by a map $f:\mathcal{M}\times\mathcal{N}\rightarrow\mathcal{M}$ so that the state at discrete time $k+1$ is related to the state and the input at discrete time k as follows:

$$x_{k+1} = f(x_k,u_k). \tag{4}$$

As before, we observe a scalar value y_k, which is related to the state x_k by an observation function $\varphi:\mathcal{M}\rightarrow\mathbb{R}$.

The question we now ask is whether we can reconstruct a state space equivalent to the original one, given only the input and output time series $\{u_k\}$ and $\{y_k\}$. The purpose of this reconstruction is to obtain an input-output model of the system.

Stark *et al.* [9] proved an embedding theorem that is applicable to a wide class of stochastic dynamical systems as well as to input-output systems. They called this theorem Takens's theorem for stochastic systems. For $d\geq 2m+1$, for *typical* f and φ, and for *typical* input sequence u, the corresponding delay map is an embedding. Further, we have the

following predictive relation, which was conjectured by Casdagli [7]:

$$y_{k+d} = G(y_k, y_{k+1}, \ldots, y_{k+d-1}, u_k, u_{k+1}, \ldots, u_{k+d-1}). \quad (5)$$

However, G is only defined for almost every $u_k, u_{k+1}, \ldots, u_{k+d-1}$. Hence, it is not known whether G is regular with respect to the input terms [10].

Muldoon *et al.* [17] have a different approach; they use multivariate observations. Assume that the observation function $\varphi : \mathcal{M} \to \mathbb{R}^p$ gives p independent observables, with $p \geq 2m+1$. Then, on the basis of Whitney's embedding theorem [18], φ is generically an embedding. Thus, from one multivariate observation y_k, information about the state x_k can be obtained. The idea developed by Muldoon *et al.* [17] is to combine this observation with its successor y_{k+1}, in order to obtain information about the input u_k. More explicitly, the reconstruction map $\Phi_{f,\varphi} : \mathcal{M} \times \mathcal{N} \to R^{2p}$ defined by

$$\Phi_{f,\varphi}(x,u) = [\varphi(x), \varphi(f(x,u))]$$

is an embedding [26]. In other words, knowledge of the state x_k (through y_k) and its successor x_{k+1} (through y_{k+1}) gives us information about the input u_k. However, this theorem cannot be directly used to derive a predictive relation for future observations. In Sec. III, we use the main idea of this theorem in order to build a functional relation that does not rely on knowing the inputs.

III. FORCED SYSTEMS WITH UNKNOWN INPUTS

In Sec. II, we reviewed two theorems that are related to Takens's embedding theorem and are applicable to forced systems. The theorem by Stark *et al.* [9] allows the reconstruction of an input-output system, and, by assuming the knowledge of the input, we may use this theorem to obtain the predictive relation given in Eq. (5). The theorem by Muldoon *et al.* [17] allows the reconstruction of an input-output system by using multivariate observations; however, this theorem does not lead to a predictive relation.

In the case of unmeasured input, we now argue that it is possible to conciliate the two approaches and derive a functional relation that uses multivariate observations but does not require any knowledge of the input. Our approach is to combine multivariate observations with the method of delays.

We separate the problem into two cases. In the simpler case, in which the entire state can be observed, we derive a functional relation to estimate a part of the state. Then, in the more general case for which the state is observed through observation functions, we argue that a similar functional relation, but one that involves delays, can be obtained.

Both functional relations are nonpredictive relations; thus, they cannot be used to predict the future behavior of the system. However, since they model the dynamics of the system, they have other possible applications such as change detection and noise reduction (see Sec. V).

A. Observing the entire state

We first assume that the entire state is directly observed and that the manifold $\mathcal{M}$, where the state x lies, is itself a Cartesian product, $\mathcal{M} = \mathcal{M}_a \times \mathcal{M}_b$, where $\mathcal{M}_a$ and $\mathcal{M}_b$ are two compact manifolds of respective dimensions m_a and m_b. Therefore, x may be decomposed as $x = \{x_a, x_b\}^T$, and Eq. (4) can be expressed as follows:

$$x_{a,k+1} = f_a(x_k, u_k), \quad (6a)$$

$$x_{b,k+1} = f_b(x_k, u_k), \quad (6b)$$

where f_a and f_b are the corresponding components of the map f. We call x_a the *primary* state variables, and x_b the *secondary* state variables. Here, the word *primary* is used to imply that these are the state variables of primary interest. The secondary state variables are used to uniquely determine the input, as explained below.

An important hypothesis of the theorem by Muldoon *et al.* [17] is that the restriction $f_x \equiv f(x, \cdot) : \mathcal{N} \to \mathcal{M}$ is an embedding for each $x \in \mathcal{M}$. This hypothesis allows the unique determination of the input u_k from one state x_k and the next state x_{k+1}.

We consider a more restrictive hypothesis that states that the restriction $f_{b,x} \equiv f_b(x, \cdot) : \mathcal{N} \to \mathcal{M}_b$ is an embedding for each $x \in \mathcal{M}$. By using Eq. (6b), we have $x_{b,k+1} = f_{b,x_k}(u_k)$, and therefore, u_k can be uniquely determined as follows:

$$u_k = f_{b,x_k}^{-1}(x_{b,k+1}).$$

Inserting the value for u_k into Eq. (6a) leads to a functional relation on the basis of which the primary state variables, $x_{a,k+1}$, can be inferred,

$$x_{a,k+1} = f_a(x_k, u_k) = f_a(x_k, f_{b,x_k}^{-1}(x_{b,k+1})). \quad (7)$$

Equation (7) can be rewritten as follows:

$$x_{a,k+1} = G(x_{a,k}, x_{b,k}, x_{b,k+1}). \quad (8)$$

The function G can then be estimated from time series data, and the next values of the primary state variables can be estimated. To this end, G should be at least continuous, which is the case, as the following lemma shows.

Lemma 1. Let $\mathcal{M}_a$, $\mathcal{M}_b$, and $\mathcal{N}$ be compact manifolds of respective dimensions m_a, m_b, and n, and let $\mathcal{M} = \mathcal{M}_a \times \mathcal{M}_b$. Let $f_a : \mathcal{M} \times \mathcal{N} \to \mathcal{M}_a$ and $f_b : \mathcal{M} \times \mathcal{N} \to \mathcal{M}_b$ be C^1 maps such that $f_{b,x} \equiv f_b(x, \cdot) : \mathcal{N} \to \mathcal{M}_b$ is an embedding for each $x \in \mathcal{M}$. Then, the function G defined for $x \in \mathcal{M}$ and $y \in f_{b,x}(\mathcal{N})$ by $G(x,y) = f_a(x, f_{b,x}^{-1}(y))$ is well defined and of class C^1.

Proof. Let $\Psi : \mathcal{M} \times \mathcal{N} \to \mathcal{M} \times \mathcal{M}_b$ be the function defined by $\Psi(x,u) = (x, f_b(x,u))$. By assumption f_b is of class C^1 and for $x \in \mathcal{M}$, $f_{b,x}$ is an embedding; therefore, it is easy to show that Ψ is of class C^1, is injective, and is an immersion. Thus, since $\mathcal{M}$ and $\mathcal{N}$ are compact, Ψ is an embedding, and its inverse Ψ^{-1} is defined on $\Psi(\mathcal{M} \times \mathcal{N})$ and is of class C^1. Then, $G = f_a \circ \Psi^{-1}$, which is of class C^1.

B. Using observation functions

We now suppose that the state is indirectly observed through two observation functions $\varphi_a : \mathcal{M} \to \mathbb{R}^{p_a}$ and $\varphi_b : \mathcal{M} \to \mathbb{R}^{p_b}$. Thus, we have two time series $\{y_a\}$ and $\{y_b\}$ given at discrete time k by

$$y_{a,k} = \varphi_a(x_k), \quad y_{b,k} = \varphi_b(x_k).$$

Here also, we call y_a the *primary observations*, and y_b the *secondary observations*. Paralleling the reasoning of Sec. III A, we assume that the restriction $\varphi_b \circ f_x \equiv \varphi_b \circ f(x, \cdot) : \mathcal{N} \rightarrow \mathbb{R}^{p_b}$ is an embedding for each $x \in \mathcal{M}$. Thus, the knowledge of x_k and $y_{b,k+1}$ uniquely determines the input u_k. This can be seen by replacing the expression of $y_{b,k+1}$ using the equation of dynamics [Eq. (4)],

$$y_{b,k+1} = \varphi_b(x_{k+1}) = \varphi_b(f(x_k, u_k)) = \varphi_b \circ f_{x_k}(u_k).$$

The previous equation shows that $y_{b,k+1}$ is in the image of $\varphi_b \circ f_{x_k}$. Hence, by using our hypothesis, we can apply its inverse $(\varphi_b \circ f_{x_k})^{-1}$ and obtain

$$u_k = (\varphi_b \circ f_{x_k})^{-1}(y_{b,k+1}).$$

Inserting this expression of u_k in the equation of dynamics [Eq. (4)] then yields

$$x_{k+1} = f(x_k, u_k) = f(x_k, (\varphi_b \circ f_{x_k})^{-1}(y_{b,k+1})),$$

which may be rewritten as

$$x_{k+1} = \widetilde{f}(x_k, y_{b,k+1}). \tag{9}$$

Equation (9) can be viewed as the equation of dynamics of a hypothetical input-output system whose state at discrete time k is x_k and whose input at discrete time k is $y_{b,k+1}$. The state of this input-output system is observed through the observation function φ_a. If Takens's theorem for stochastic systems [9] could be applied to this system, then similarly to Eq. (5), the functional relation

$$y_{a,k} = G(y_{a,k-1}, y_{a,k-2}, \ldots, y_{a,k-d}, y_{b,k}, y_{b,k-1}, \ldots, y_{b,k-(d-1)})$$

would hold for sufficiently large d.

However, Takens's theorem for stochastic systems cannot be applied to this case for the following reason. The theorem requires that for fixed y, the evolution operator of the system, i.e., the function $\widetilde{f}_y : \mathcal{M} \rightarrow \mathcal{M}$ defined by

$$\widetilde{f}_y(x) = \widetilde{f}(x, y) = f(x, (\varphi_b \circ f_x)^{-1}(y)),$$

be a diffeomorphism of $\mathcal{M}$. From the expression of $\widetilde{f}_y$, one can see that $\varphi_b(\widetilde{f}_y(x)) = y$. Thus, $\widetilde{f}_y(\mathcal{M}) \subset \varphi_b^{-1}(\{y\})$. If $\widetilde{f}_y$ is a diffeomorphism, the equality $\varphi_b^{-1}(\{y\}) = \mathcal{M}$ will hold. This implies that φ_b will be a constant function on $\mathcal{M}$, which is contrary to our assumption that $\varphi_b \circ f_x$ is an embedding for each $x \in \mathcal{M}$. The problem is that we are trying to use φ_b to determine the input, and then φ_a to obtain information about the entire state of the system, including the part that we already know from φ_b.

Instead, assume as in Sec. III A that $\mathcal{M}$ is a product of two manifolds $\mathcal{M} = \mathcal{M}_a \times \mathcal{M}_b$ and that the function $f_b(x, \cdot) : \mathcal{N} \rightarrow \mathcal{M}_b$ is an embedding. Further, assume that φ_a depends only on x_a, that φ_b depends only on x_b, and that φ_b is an embedding of $\mathcal{M}_b$. Then, Eq. (8) holds, and since φ_b is an embedding, $x_{b,k}$ is uniquely determined by $y_{b,k} = \varphi_b(x_{b,k})$, and $x_{b,k+1}$ is uniquely determined by $y_{b,k+1} = \varphi_b(x_{b,k+1})$. Therefore, Eq. (8) can be rewritten as follows:

$$x_{a,k+1} = \widetilde{f}_a(x_{a,k}, y_{b,k}, y_{b,k+1}). \tag{10}$$

With these assumptions, the aforementioned problem does not occur. If Takens's theorem for stochastic systems is applicable to the hypothetical dynamical system of Eq. (10), the following functional relation holds:

$$y_{a,k} = G(y_{a,k-1}, y_{a,k-2}, \ldots, y_{a,k-d}, y_{b,k}, y_{b,k-1}, \ldots, y_{b,k-d}). \tag{11}$$

IV. NUMERICAL SIMULATIONS

In order to verify the validity of the functional relations derived in Sec. III A [Eq. (8)] and Sec. III B [Eq. (11)], in Secs. IV B–IV D we introduce and analyze three systems: a discrete map that is a discretized version of the Rössler system by finite differences, the continuous-time Rössler system, and two coupled continuous-time Rössler systems.

A. Methodology

For each system to be analyzed, we build models based on multivariate observations and without any knowledge of the input excitation. Further, we evaluate their estimation errors on separate test data. In this section, we present the models used and the methodology used for verifying the validity of the functional relations of Eqs. (8) and (11).

In both cases, the data that we analyze come in the form of a multivariate time series $\{y_k\}_{1 \le k \le N}$ consisting of p individual channels $y_{i,k}$, $i = 1, \ldots, p$. Each channel represents an independent observation of the state of the system. Following the notations of Sec. III B, some of these observations will be primary observations, and others will be secondary observations. From the primary observations, we select one channel as the target variable z_k. The target z_k is the value that we wish to estimate. For example, z_k may be the first channel of the observed time series, $y_{1,k}$. The multivariate time series $\{y_k\}_{1 \le k \le N}$ and associated target time series $\{z_k\}_{1 \le k \le N}$ form a training set $\mathcal{T}$. From the training data in $\mathcal{T}$, we wish to identify the functional relation

$$z_k = G(\mathbf{v}_k), \tag{12}$$

where $\mathbf{v}_k$ is a vector formed from delayed values of the observed time series. We write the delay vector $\mathbf{v}_k$ in a generic form, using nonuniform delays [19],

$$\mathbf{v}_k = (y_{i_1, k-\tau_1}, y_{i_2, k-\tau_2}, \ldots, y_{i_d, k-\tau_d}).$$

The reconstruction vector $\mathbf{v}_k$ has d components. For the jth component, we independently choose the channel i_j and the delay τ_j. Usually, we restrain $\tau_j > 0$ so that the target z_k depends only on past delays; Eq. (12) is then a predictive relation. However, in the case of an input-output system without any knowledge of the input, as the relation from Eq. (11) shows, one of the delays for the secondary outputs is 0. Therefore, when modeling this relation, we allow $\tau_j = 0$ for the channels chosen as secondary outputs. In this paper, we take $\mathbf{v}_k$ as the right-hand side of Eqs. (8) or (11).

Once the reconstruction vector $\mathbf{v}_k$ is chosen, we use the measured time series for estimating the function G of Eq.

(12). To this purpose, we use a local linear approximation (refer to [5]) defined as follows. For a test point $\tilde{\mathbf{v}}$ with associated target $\tilde{z}$, define $\mathcal{U}_\eta(\tilde{\mathbf{v}})$ as the set of η points among the reconstruction vectors $\mathbf{v}_k$ of the training set $\mathcal{T}$ that are closest to the test point $\tilde{\mathbf{v}}$ in terms of the Euclidean distance. The number of neighbors $\eta = |\mathcal{U}_\eta(\tilde{\mathbf{v}})|$ is chosen beforehand. Then, the local linear approximation of the target $\tilde{z}$ is

$$z_{pred} = a \cdot \tilde{\mathbf{v}} + b,$$

where the coefficients a and b are obtained by least-squares fitting over the set of neighbors $\mathcal{U}_\eta(\tilde{\mathbf{v}})$.

In this paper, we perform all the local linear approximations by using neighborhoods containing $\eta = 20$ neighbors. The optimal choice of η depends on the nonlinearity of the equations, on the level of noise, and on the dimension of the reconstruction vector $\mathbf{v}_k$. Thus, a fixed number of neighbors is not optimal for all the systems that are studied in Secs. IV B–IV D, and not optimal even for all the relations of a given system. Several values were tried, and the choice $\eta = 20$ is close to the optimal for the systems that are studied here; for $\eta = 10$ or $\eta = 50$, the results were similar to the results shown in this paper, with the exception of cases for which the dimension of the reconstruction vector is close to η.

In order to estimate the accuracy of the model, we compute the root mean square of the estimation error of the model on a test time series, which we normalize by the standard deviation of the time series to be estimated,

$$E = \frac{1}{\sigma(\tilde{z})} \sqrt{\frac{1}{N_{test}} \sum_{k=1}^{N_{test}} [\tilde{z}_k - z_{pred,k}]^2}. \qquad (13)$$

In the rest of the paper, we refer to E as normalized estimation error, or simply as estimation error.

We now examine the source of the estimation errors. The functional relation that we are modeling is in the form of Eq. (12). If it holds exactly, then as the number of points N in the training time series increases, better and better approximations of the function G can be made, and the normalized estimation error E will become closer and closer to zero. For the local linear approximation, the scaling law is as follows:

$$E = \mathcal{O}(N^{-2/D}), \qquad (14)$$

where N is the length of the training data, and, in the case of autonomous systems, D is the information dimension of the attractor [20,21]. For input-output systems, the same scaling law is expected to hold, with D being the dimension of the space in which the delay vector $\mathbf{v}_k$ is constrained to lie [7]. With a random input, the dimension is $D = m + n_i$, where m is the state-space dimension and n_i the number of different inputs that appear in the delay vector.

Alternatively, if Eq. (12) does not hold exactly, then the relation between the delay vector $\mathbf{v}_k$ and the target variable z_k is of the form

$$z_k = G(\mathbf{v}_k) + \varepsilon_k, \qquad (15)$$

where ε_k is the noise term, or innovation, that does not depend on the variables included in $\mathbf{v}_k$. In this case, in the limit of large N, the normalized estimation error E is limited by the signal-to-noise ratio of the time series. For large N, the normalized estimation error E is then approximately

$$E \approx \frac{\sigma(\varepsilon)}{\sigma(z)}. \qquad (16)$$

In the present study, there may be several causes for not being able to estimate the target z_k: first, when the reconstruction vector $\mathbf{v}_k$ does not have a sufficient number of delay terms, a functional relation does not exist between $\mathbf{v}_k$ and z_k; second, when the function G is not continuously differentiable, the local linear approximation fails; third, when the input u_k cannot be uniquely determined from the variables included in the reconstruction vector $\mathbf{v}_k$, the proposed relations do not hold. It is the third case that we are looking for. In this case, the innovation ε_k of Eq. (15) mainly consists of a term depending on u_k. Therefore, we can estimate the standard deviation of the innovation ε_k, and thus the normalized estimation error E in the limit of large N, by estimating the contribution to z_k of the terms depending on u_k in the equations of dynamics of the system.

In Secs. IV B–IV D, we use this reasoning for verifying the validity of the functional relations. If, as the training data length N is increased, the estimation error does not decrease to a value below the threshold given by Eq. (16), the input u_k cannot be uniquely determined as was supposed in Sec. III. Conversely, an estimation error following the scaling law of Eq. (14) indicates a successful model based on only the output data.

In order to better discriminate between the two cases, we need to make a further adjustment to the procedure. As Casdagli [21] noted and as we also observed, very infrequent bad estimations seem to affect the normalized estimation error E. Hence, for large N, Eq. (14) does not hold. Casdagli [21] suggests discarding 10% of the worst estimations, or alternatively, replacing the arithmetic mean in Eq. (13) by a geometric mean. In this paper, we removed 10% of the worst estimations when computing the normalized estimation error E. The use of a geometric mean in Eq. (13) also gave similar results, which we have not discussed here.

Finally, we would like to mention that no observational noise was added to the simulated time series, and the time series were normalized in the following manner so as to give an equal weight to every channel when computing Euclidean distances for the local linear approximations. Each channel of the training time series was first normalized by subtracting the sample mean and dividing by the sample standard deviation so that each channel of the resulting time series has zero mean and unit variance. Each channel of the test time series was then normalized by the same factors (so that the channels of the test data may or may not have zero mean and unit variance). The reason for using the same factors is that the behavior of the system, and hence the values estimated by the models, depends on the amplitude of the data, i.e., on the amplitude of the current states and of the inputs. By normalizing the training data and the test data by the same factors, we can retain this dependency.
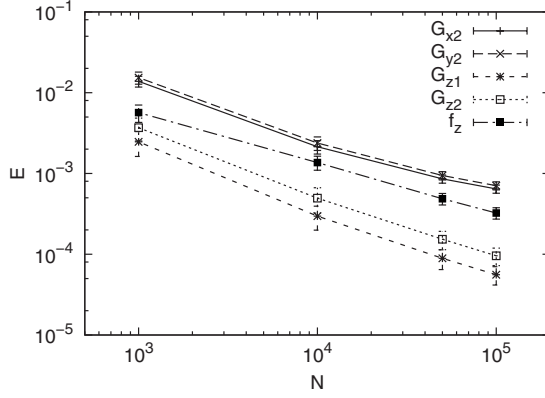
FIG. 1. Normalized estimation error for a local linear approximation of the Rössler map as a function of the training data length N. The functions f_x, f_y, G_{x1}, and G_{y1} were not included since their estimation error was of the order of the round-off error of floating-point computations. For each value of N, the mean and the standard deviation for ten sets of training and test data are shown.

B. Rössler map

1. Model and data

As the first example, we use a discretized version of the Rössler system. We call this system Rössler map. The original Rössler system [22] is given by

$$\dot{x} = -\omega y - z, \tag{17a}$$

$$\dot{y} = \omega x + ay, \tag{17b}$$

$$\dot{z} = b + z(x - c). \tag{17c}$$

We transform it into a forced system by adding an input force $u(t)$ on the first or second equations,

$$\dot{x} = -\omega y - z + \alpha_x u(t), \tag{18a}$$

$$\dot{y} = \omega x + ay + \alpha_y u(t), \tag{18b}$$

$$\dot{z} = b + z(x - c). \tag{18c}$$

We use the third finite difference model presented by Letellier *et al.* [23], and to it, we add the above input force. Thus, the equations for the Rössler map that we use are as follows:

$$x_{k+1} = x_k - \varphi_1(\omega y_k + z_k - \alpha_x u_k), \tag{19a}$$

$$y_{k+1} = (1 + a\varphi_2)y_k + \varphi_2(\omega x_{k+1} + \alpha_y u_k), \tag{19b}$$

$$z_{k+1} = b\varphi_3 + [1 + \varphi_3(x_{k+1} - c)]z_k. \tag{19c}$$

This is a nonstandard finite difference model; φ_1, φ_2, and φ_3 are coefficients that depend on the integration time step h and on the time scales of the system. Following Letellier *et al.* [23], we use the values

$$\varphi_1 = h, \quad \varphi_2 = \frac{1 - e^{-ah}}{a}, \quad \varphi_3 = \frac{1 - e^{-x_c h}}{x_c},$$

where $x_c = (-c + \sqrt{c^2 - 4ab})/2$. We use the parameter values $\omega = 1$, $a = 0.432$, $b = 2$, $c = 4$, an integration step $h = 0.1$, and input coefficients $\alpha_x = \alpha_y = 1$.

Several training time series of the Rössler map with N data points were generated by evaluating Eq. (19) for $N + 1000$ time steps and discarding the first 1000 values, for increasing values of N. Each time series had different randomly chosen initial conditions and different external input $\{u_k\}$, which was unfiltered random Gaussian white noise with zero mean and unit variance. In order to compute estimation errors, a test time series with 1000 points was also generated for each training time series.

2. Full state observation

We now examine Eq. (19) in more detail in order to understand which functional relations can be expected to hold. First, observe that the terms x_k, y_k, z_k, and u_k appear in all three equations, through the term x_{k+1} for Eqs. (19b) and (19c). Thus, the next states can be expressed as functions of the immediately preceding input and states,

$$x_{k+1} = f_x(x_k, y_k, z_k, u_k), \tag{20a}$$

$$y_{k+1} = f_y(x_k, y_k, z_k, u_k), \tag{20b}$$

$$z_{k+1} = f_z(x_k, y_k, z_k, u_k). \tag{20c}$$

The functions f_x and f_y are linear, while f_z is nonlinear in that it involves second-order polynomial terms because of the presence of the term $x_{k+1}z_k$ in Eq. (19c). Therefore, local linear approximations of f_x and f_y should have very low estimation errors, while local linear approximations of f_z should follow the scaling law of Eq. (14).

Consider the case of a full state observation. We assume that the hypothesis of Sec. III A holds so that if we have the full knowledge of the state (x_k, y_k, z_k), only one of the states x_{k+1}, y_{k+1}, or z_{k+1} is necessary to uniquely determine the input u_k. Considering successively x, y, or z as the secondary variable, we thus define six functional relations on the basis of Eq. (8),

$$y_{k+1} = G_{y1}(x_{k+1}, x_k, y_k, z_k), \tag{21a}$$

$$z_{k+1} = G_{z1}(x_{k+1}, x_k, y_k, z_k), \tag{21b}$$

$$x_{k+1} = G_{x1}(y_{k+1}, x_k, y_k, z_k), \tag{21c}$$

$$z_{k+1} = G_{z2}(y_{k+1}, x_k, y_k, z_k), \tag{21d}$$

$$x_{k+1} = G_{x2}(z_{k+1}, x_k, y_k, z_k), \tag{21e}$$

$$y_{k+1} = G_{y2}(z_{k+1}, x_k, y_k, z_k). \tag{21f}$$

By using Eq. (19), we can derive analytical expressions for the above functions as follows: G_{y1} and G_{x1} are completely linear; G_{z1} and G_{z2} are polynomial; and G_{x2} and G_{y2} are

TABLE I. Estimation error thresholds for the Rössler map. The values given are the mean for ten data sets of 1000 points, and the standard deviation.

Variable	Threshold
x	$4.12 \times 10^{-2} \pm 2.66 \times 10^{-3}$
y	$4.49 \times 10^{-2} \pm 4.76 \times 10^{-3}$
z	$1.19 \times 10^{-2} \pm 1.41 \times 10^{-3}$

rational functions with a pole at $z_k = 0$. (The time series that were generated for the Rössler map do not include state-space regions with $z_k = 0$.)

For each of the functions of Eq. (21) to be approximated, Fig. 1 shows the normalized estimation error E as a function of the training data length N. As the estimation errors of local linear approximations of the functions f_x, f_y, G_{x1}, and G_{y1} were as expected on the order of the round-off error of floating-point computations, their values were not included. For the sake of comparison, the estimation errors of local linear approximations of f_z are also shown in Fig. 1. In Fig. 1, since both axes are shown on a logarithmic scale, it is apparent that all estimation errors decrease as a negative power of the data length N, following the scaling law of Eq. (14).

We further validate the estimation errors by comparing them to thresholds as defined in Sec. IV A. For each state variable x_k, y_k, and z_k, the corresponding threshold was computed in the following manner. Assume that the noise ε_k mainly consists of an additive term depending on the input u_k, and approximate it by an estimate of the contribution of the terms depending on u_k in the equations of dynamics. For example, when estimating x_{k+1}, the noise term $\varepsilon_{x,k}$ is approximated by

$$\varepsilon_{x,k} = f_x(x_k, y_k, z_k, u_k) - f_x(x_k, y_k, z_k, \bar{u}_k),$$

where the bar represents the sample average. The threshold is then obtained by dividing the standard deviation of ε_x by that of x as per Eq. (16). The resulting thresholds are shown in Table I. All estimation errors of Fig. 1 are well below the thresholds for training time series with more than 10 000 points.

3. Observation functions

We now consider the case of Sec. III B, for which the state is observed through two observation functions φ_a and φ_b. Since the entire state is not observed, the functional relation of Sec. III A cannot be used. For several choices of observation functions φ_a and φ_b and for increasing d, we evaluate the quality of local linear approximations of the functional relation of Eq. (11).

Table II contains the definition of the pairs (φ_a, φ_b) of observation functions that were considered as well as the thresholds computed in the same manner as that described in Sec. IV B 2, and the normalized estimation errors obtained for local linear approximations of Eq. (11) using training time series with $N = 100\,000$ points.

For $d \geq 2$, the estimation error in all cases is at least close to one order smaller than the corresponding threshold, which indicates that for $d \geq 2$, the functional relations hold in all the cases listed in Table II.

Figures 2 and 3 show the evolution of the estimation error as a function of the training data length N, for case 1 and case 4 of Table II. Case 1 is one of the cases with the smallest estimation error; case 4 is one of the cases with the largest estimation error. The figures show that the estimation error for $d = 1$ stagnates, while for $d \geq 2$, the estimation error follows the scaling law of Eq. (14).

TABLE II. Thresholds and estimation errors for local linear approximations of the Rössler map using various primary (φ_a) and secondary (φ_b) observation functions. The threshold for each primary observation function φ_a was computed from data sets of 1000 points as described in Sec. IV B 2. The functional relation that is approximated is that of Eq. (11), with increasing d. The estimation errors were computed using 100 000 training data points and 1000 test data points. All values shown are the mean for ten sets of data, and the standard deviation as a percentage of the mean.

| Case	φ_a	φ_b	Threshold	$E(d=1)$	$E(d=2)$
1	x	y	$4.12 \times 10^{-2} \pm 6.5\%$	$1.55 \times 10^{-2} \pm 12.8\%$	$2.99 \times 10^{-5} \pm 22.6\%$
2	y	x	$4.49 \times 10^{-2} \pm 10.6\%$	$6.81 \times 10^{-3} \pm 13.3\%$	$2.17 \times 10^{-5} \pm 20.8\%$
3	$x+y+z$	xy	$5.91 \times 10^{-2} \pm 9.0\%$	$2.24 \times 10^{-2} \pm 8.6\%$	$7.03 \times 10^{-3} \pm 14.1\%$
4	xy	$x+y+z$	$3.69 \times 10^{-2} \pm 15.0\%$	$1.29 \times 10^{-2} \pm 10.5\%$	$1.09 \times 10^{-3} \pm 14.1\%$
5	z	xy	$1.19 \times 10^{-2} \pm 11.8\%$	$6.96 \times 10^{-3} \pm 16.8\%$	$8.44 \times 10^{-4} \pm 17.0\%$
6	z	$x-y$	$1.19 \times 10^{-2} \pm 11.8\%$	$2.89 \times 10^{-3} \pm 12.6\%$	$2.95 \times 10^{-4} \pm 20.0\%$
7	z	$x+y$	$1.19 \times 10^{-2} \pm 11.8\%$	$2.18 \times 10^{-2} \pm 21.8\%$	$5.89 \times 10^{-4} \pm 17.8\%$
8	x	yz	$4.12 \times 10^{-2} \pm 6.5\%$	$2.00 \times 10^{-2} \pm 11.6\%$	$1.97 \times 10^{-3} \pm 14.3\%$
9	x	$y+z$	$4.12 \times 10^{-2} \pm 6.5\%$	$5.95 \times 10^{-3} \pm 10.9\%$	$1.62 \times 10^{-4} \pm 19.6\%$
10	x	$y-z$	$4.12 \times 10^{-2} \pm 6.5\%$	$1.88 \times 10^{-2} \pm 11.3\%$	$3.80 \times 10^{-4} \pm 28.6\%$
11	y	xz	$4.49 \times 10^{-2} \pm 10.6\%$	$9.97 \times 10^{-3} \pm 10.5\%$	$1.55 \times 10^{-3} \pm 8.5\%$
12	y	$x+z$	$4.49 \times 10^{-2} \pm 10.6\%$	$7.50 \times 10^{-3} \pm 11.6\%$	$2.22 \times 10^{-4} \pm 19.6\%$
13	y	$x-z$	$4.49 \times 10^{-2} \pm 10.6\%$	$1.47 \times 10^{-2} \pm 8.8\%$	$1.14 \times 10^{-3} \pm 16.7\%$
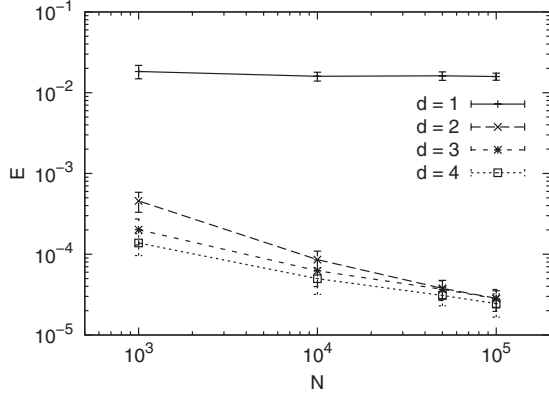
FIG. 2. Normalized estimation error for a local linear approximation of the Rössler map with observation functions $\varphi_a(x, y, z) = x$ and $\varphi_b(x, y, z) = y$, as a function of the training data length N. For each value of N, the mean and the standard deviation for ten sets of training and test data are shown.

C. Rössler system

1. Model and data

The second example is that of the Rössler system given by Eq. (18), with parameter values $a = 0.398$, $b = 2$, $c = 4$, and $\omega = 1$. We define three cases of input coefficients: (a) $\alpha_x = 1$ and $\alpha_y = 0$, (b) $\alpha_x = 0$ and $\alpha_y = 1$, and (c) $\alpha_x = 1$ and $\alpha_y = 1$. By studying these three cases, it is possible to confirm, as the following results show, that if the inputs do not affect the chosen secondary variables, or if the number of secondary variables is insufficient, the relations do not hold: the estimation error stops to decrease after reaching a certain level. For each of these cases, several training and test time series were generated with the same procedure as described in Sec. IV B 1, by simulating the Rössler system of Eq. (18) with the fourth-order Runge-Kutta method (see for example [24]) with an integration time step $h = 0.1$.

For an integration step, in the fourth-order Runge-Kutta equations, the input $u(t)$ is evaluated at the beginning of the step $[u(t_k) = u_k]$, at the middle of the step $[u(t_k + h/2)]$, and at
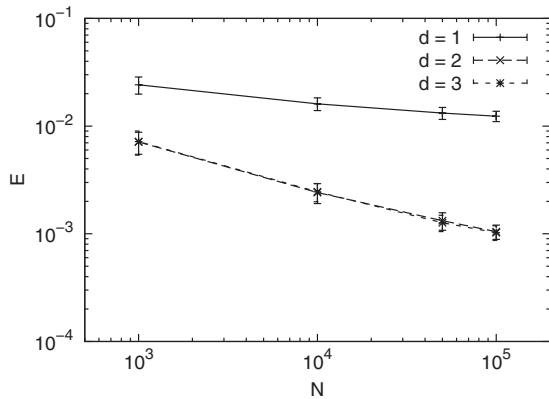
FIG. 3. Normalized estimation error for a local linear approximation of the Rössler map with observation functions $\varphi_a(x, y, z) = xy$ and $\varphi_b(x, y, z) = x + y + z$, as a function of the training data length N. For each value of N, the mean and the standard deviation for ten sets of training and test data are shown.

the end of the step $[u(t_k + h) = u_{k+1}]$. We assume that the input $u(t)$ is constant during one time step, so that $u(t_k + h/2) = u_k$. Because of this integration method, in the discretized equations that evolve the state (x_k, y_k, z_k) to $(x_{k+1}, y_{k+1}, z_{k+1})$, there are input terms depending on u_k in the three equations, and input terms depending on u_{k+1} in the first equation (when $\alpha_x = 1$) and in the second equation (when $\alpha_y = 1$). Thus, the generic form of the equations of dynamics of the discretized system is as follows:

$$x_{k+1} = f_x(x_k, y_k, z_k, u_k, u_{k+1}), \tag{22a}$$

$$y_{k+1} = f_y(x_k, y_k, z_k, u_k, u_{k+1}), \tag{22b}$$

$$z_{k+1} = f_z(x_k, y_k, z_k, u_k). \tag{22c}$$

2. Full state observation

We first consider the case of a full state observation. Since the discretized equation, Eq. (22) involves two inputs u_k and u_{k+1}, we need at least two state variables to uniquely determine the inputs u_k and u_{k+1} as described in Sec. III A. We have at our disposition only three state variables. We choose two of these as the secondary variables, and the remaining one as the primary variable. Thus, we consider the following functional relations based on Eq. (8):

$$x_{k+1} = G_x(y_{k+1}, z_{k+1}, x_k, y_k, z_k), \tag{23a}$$

$$y_{k+1} = G_y(x_{k+1}, z_{k+1}, x_k, y_k, z_k), \tag{23b}$$

$$z_{k+1} = G_z(x_{k+1}, y_{k+1}, x_k, y_k, z_k). \tag{23c}$$

Here, all the functions of Eqs. (22) and (23) are polynomial or rational functions of their arguments. We have noted in Sec. IV C 1 that when $\alpha_x = 1$ and $\alpha_y = 0$, f_y and f_z do not depend on the input u_{k+1}. Therefore, the input u_{k+1} cannot be uniquely determined unless x is chosen as one of the secondary variables; hence, Eq. (23a) does not hold in case (a). Similarly, when $\alpha_x = 0$ and $\alpha_y = 1$, f_x and f_z do not depend on the input u_{k+1}. Therefore, the input u_{k+1} cannot be uniquely determined unless y is chosen as one of the secondary variables; hence, Eq. (23b) does not hold in case (b).

Figure 4 shows the normalized estimation error of local linear approximations of G_x, G_y, and G_z as a function of the training data length N, while Table III gives the thresholds of Eq. (16), computed in the same way as for the Rösser map in Sec. IV B 2.

From Fig. 4(a), we see that for case (a), for which $\alpha_x = 1$ and $\alpha_y = 0$, the estimation error of approximations of G_y and G_z continues to decrease below the corresponding threshold as N is increased, which indicates successful models using only output data. However, the estimation error of an approximation of G_x does not decrease below the threshold as N is increased. This was expected, since as mentioned above, in case (a) using the variable x_{k+1} is indispensable in order to uniquely determine the second output u_{k+1}.

Similarly, Fig. 4(b) shows that for case (b), for which $\alpha_x = 0$ and $\alpha_y = 1$, the estimation error of an approximation of G_z continues to decrease below the threshold, while G_y is
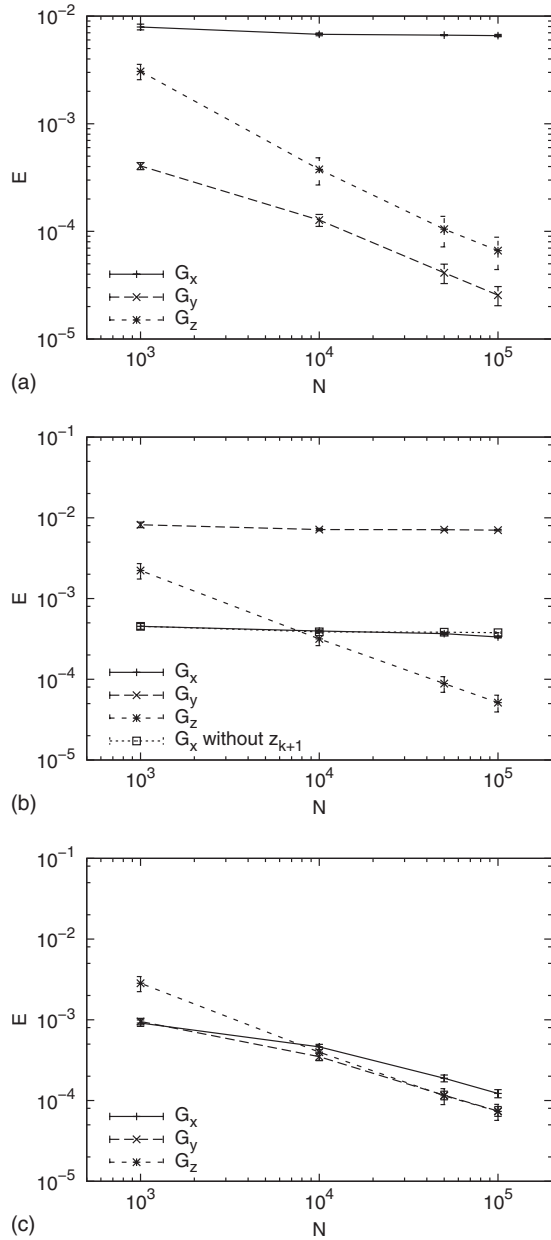
FIG. 4. Normalized estimation error for a local linear approximation of the Rössler system as a function of the training data length N, with parameters (a) $\alpha_x=1$ and $\alpha_y=0$, (b) $\alpha_x=0$ and $\alpha_y=1$, and (c) $\alpha_x=1$ and $\alpha_y=1$. For each value of N, the mean and the standard deviation for ten sets of training and test data are shown.

limited by the inability to uniquely determine the input u_{k+1} from the variables y_{k+1}, z_{k+1}, x_k, y_k, and z_k. Nevertheless, note that the estimation error for G_x is below the threshold but does not follow the scaling law. This can be interpreted as follows. In Fig. 4(b), the estimation error for an approximation of G_x without using the variable z_{k+1} is also included. The estimation error is similar to that of an approximation of G_x. Thus, the information contained in z_{k+1} was not useful to obtain information about the input u_k; only the information contained in y_{k+1} was useful. However, since y_{k+1} depends not only on u_k but also on u_{k+1}, the input u_k can be only partially determined. Thus, the estimation of x_{k+1} is not accurate due to the partial knowledge of u_k.

TABLE III. Estimation error thresholds for the Rössler system. The values given are the mean for ten data sets of 1000 points, and the standard deviation.

Variable	Input coefficients	Threshold
x	$\alpha_x=1,\ \alpha_y=0$	$3.71\times10^{-2}\pm1.88\times10^{-3}$
y	$\alpha_x=1,\ \alpha_y=0$	$2.33\times10^{-3}\pm9.44\times10^{-5}$
z	$\alpha_x=1,\ \alpha_y=0$	$6.18\times10^{-3}\pm3.67\times10^{-4}$
x	$\alpha_x=0,\ \alpha_y=1$	$2.32\times10^{-3}\pm1.60\times10^{-4}$
y	$\alpha_x=0,\ \alpha_y=1$	$4.19\times10^{-2}\pm2.79\times10^{-3}$
z	$\alpha_x=0,\ \alpha_y=1$	$2.13\times10^{-4}\pm9.94\times10^{-6}$
x	$\alpha_x=1,\ \alpha_y=1$	$3.29\times10^{-2}\pm2.49\times10^{-3}$
y	$\alpha_x=1,\ \alpha_y=1$	$3.83\times10^{-2}\pm3.62\times10^{-3}$
z	$\alpha_x=1,\ \alpha_y=1$	$5.90\times10^{-3}\pm4.60\times10^{-4}$

Finally, Fig. 4(c) shows that for case (c), for which $\alpha_x=1$ and $\alpha_y=1$, the estimation error of approximations of G_x, G_y, and G_z continues to decrease below the threshold as N is increased. This indicates that the functional relations hold, and that any two of the state variables x_{k+1}, y_{k+1}, and z_{k+1} can be used to uniquely determine the two inputs u_k and u_{k+1}, assuming also the knowledge of the previous state (x_k, y_k, z_k).

3. Observation functions

We now consider the case for which the state is observed through two observation functions φ_a and φ_b. When choosing the observation functions, it is important to ensure that the main assumption of Sec. III B holds, i.e., when the knowledge of the state (x_k, y_k, z_k) is assumed, the secondary observation $\varphi_b(x_{k+1}, y_{k+1}, z_{k+1})$ uniquely determines the inputs, which here are u_k and u_{k+1}. Ignoring the fact that u_k and u_{k+1} are the same input at two different times, we need to have at least two secondary observations. Here again, for several choices of φ_a and φ_b, and for increasing d, we evaluate the normalized estimation errors of local linear approximations of the functional relation of Eq. (11).

We present here the results only for case (b) with $\alpha_x=0$ and $\alpha_y=1$; for cases (a) and (c), the results were similar.

Table IV contains the definition of the pairs (φ_a, φ_b) of observation functions that were considered as well as the thresholds computed in the same manner as that described in Sec. IV B 2, and the normalized estimation errors obtained for local linear approximations of Eq. (11) using training time series with $N=100\,000$ points. For cases 2 to 6, when $d\geq1$ the estimation errors are at least one order below the threshold; for cases 1 and 7, when $d\geq2$ the estimation errors are one order below the threshold. As a representative example, Fig. 5 shows the estimation error for case 5 as a function of the training data length N. One can see that the estimation error follows the scaling law of Eq. (14).

D. Two coupled Rössler systems

Finally, we analyze a system that consists of two coupled Rössler systems, the first one having additional input forces:

TABLE IV. Thresholds and estimation errors for local linear approximations of the Rössler system using various primary (φ_a) and secondary (φ_b) observation functions, for input coefficients $\alpha_x=0$ and $\alpha_y=1$. The threshold for each primary observation function φ_a was computed from data sets of 1000 points as described in Sec. IV B 2. The functional relation that is approximated is that of Eq. (11), with increasing d. The estimation errors were computed using 100 000 training data points and 1000 test data points. All values shown are the mean for ten sets of data, and the standard deviation as a percentage of the mean.

Case	φ_a	φ_b	Threshold	$E(d=1)$	$E(d=2)$
1	z	$(xy,x-y)$	$2.13\times10^{-4}\pm4.7\%$	$2.09\times10^{-4}\pm19.4\%$	$2.12\times10^{-5}\pm15.0\%$
2	$x+y+z$	$(xy,x-y)$	$2.55\times10^{-2}\pm7.4\%$	$4.61\times10^{-4}\pm13.0\%$	$4.02\times10^{-4}\pm5.5\%$
3	$x+y+z$	$(xy+z,yz+x)$	$2.55\times10^{-2}\pm7.4\%$	$4.03\times10^{-4}\pm6.2\%$	$3.69\times10^{-4}\pm6.1\%$
4	$xy+z$	$(yz+x^2,(x+y)^2)$	$3.58\times10^{-2}\pm4.9\%$	$1.29\times10^{-3}\pm12.2\%$	$9.12\times10^{-4}\pm7.2\%$
5	$x+y+z$	$(yz+x^2,(x+y)^2)$	$2.55\times10^{-2}\pm7.4\%$	$7.67\times10^{-4}\pm11.3\%$	$5.31\times10^{-4}\pm7.1\%$
6	$xz+y$	$(xy,x-y)$	$1.51\times10^{-2}\pm10.8\%$	$3.35\times10^{-4}\pm11.0\%$	$2.39\times10^{-4}\pm4.3\%$
7	z	$(yz+x^2,(x+y)^2)$	$2.13\times10^{-4}\pm4.7\%$	$2.99\times10^{-4}\pm19.5\%$	$2.53\times10^{-5}\pm19.3\%$

$$\dot{x}_1 = -\omega_1 y_1 - z_1 + u(t), \tag{24a}$$

$$\dot{y}_1 = \omega_1 x_1 + a_1 y_1, + \varepsilon(y_2 - y_1), \tag{24b}$$

$$\dot{z}_1 = b_1 + z_1(x_1 - c_1), \tag{24c}$$

$$\dot{x}_2 = -\omega_2 y_2 - z_2, \tag{24d}$$

$$\dot{y}_2 = \omega_2 x_2 + a_2 y_2 + \varepsilon(y_1 - y_2), \tag{24e}$$

$$\dot{z}_2 = b_2 + z_2(x_2 - c_2), \tag{24f}$$

with parameter values $a_1=a_2=0.22$, $b_1=b_2=0.1$, $c_1=c_2=8.5$, $\omega_1=0.98$, $\omega_2=1.03$, and a coupling parameter $\varepsilon=0.1$. With the same procedure as in Sec. IV B 1, time series for the two coupled Rössler systems were generated by using the fourth-order Runge-Kutta method with an integration step $h=0.1$.

We consider the two coupled systems as not one system but two separate systems. The state variables of the first Rössler system are x_1, y_1, and z_1, and the external inputs are

FIG. 5. Normalized estimation error for a local linear approximation of the Rössler system with observation functions $\varphi_a(x,y,z)=x+y+z$ and $\varphi_b(x,y,z)=(yz+x^2,(x+y)^2)$, as a function of the training data length N. The input parameters were $\alpha_x=0$ and $\alpha_y=1$. For each value of N, the mean and the standard deviation for ten sets of training and test data are shown.

u and y_2. The state variables of the second Rössler system are x_2, y_2, and z_2, and the external input is y_1.

However, for discretized versions of these systems, the external inputs are different, as shown below. Consider a discretization using the fourth-order Runge-Kutta method. As in Sec. IV C 1, we assume that the input is constant during one time step. Equation (24) can be inserted into the Runge-Kutta equations in order to obtain discretized equations of the two coupled systems. The discretized equations are fourth-order approximations. Reproducing the complete set of equations here would take considerable space; however, for the second Rössler system, we succinctly write the general form of these equations:

$$x_{2,k+1} = f_{x_2}(x_{2,k}^0, y_{2,k}^1, z_{2,k}^1, y_{1,k}^2, x_{1,k}^3, z_{1,k}^4, u_k^4), \tag{25a}$$

$$y_{2,k+1} = f_{y_2}(y_{2,k}^0, x_{2,k}^1, y_{1,k}^1, z_{2,k}^2, x_{1,k}^2, z_{1,k}^3, u_k^3), \tag{25b}$$

$$z_{2,k+1} = f_{z_2}(z_{2,k}^0, x_{2,k}^1, y_{2,k}^2, y_{1,k}^3, x_{1,k}^4), \tag{25c}$$

where superscripts have been added as follows. Note that the functions f_{x_2}, f_{y_2}, and f_{z_2} are polynomials. When the terms of the polynomials are ordered according to the increasing power of h, the superscript of each variable in the right-hand side of Eq. (25) is the order of the first term in which this variable appears. For example, $y_{1,k}$ first appears in a term multiplied by h^2 in the function f_{x_2}, in a term multiplied by h in the function f_{y_2}, and in a term multiplied by h^3 in the function f_{z_2}.

Equation (25) shows that the inputs of the discretized second Rössler system are $y_{1,k}$, $x_{1,k}$, $z_{1,k}$, and u_k. We cannot in general uniquely determine four inputs from the next values of the three state variables x_2, y_2, and z_2, or from observables that depend on them, since this requires at the very least four state variables. However, if we consider a second-order approximation in h, Eq. (25) shows that only $y_{1,k}$ and $x_{1,k}$ are inputs of the second Rössler system.

In order to verify this approach, we consider output-only models that rely on the assumption of having only two inputs $y_{1,k}$ and $x_{1,k}$, and we identify, from the data, local linear approximations of the following three functional relations:
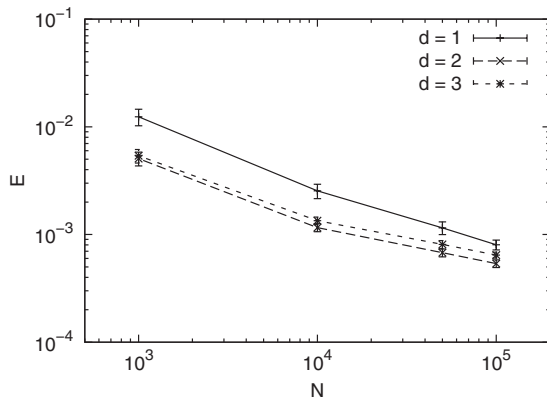
$$x_{2,k+1} = G_{x_2}(y_{2,k+1}, z_{2,k+1}, x_{2,k}, y_{2,k}, z_{2,k}),$$

$$y_{2,k+1} = G_{y_2}(x_{2,k+1}, z_{2,k+1}, x_{2,k}, y_{2,k}, z_{2,k}),$$

$$z_{2,k+1} = G_{z_2}(x_{2,k+1}, y_{2,k+1}, x_{2,k}, y_{2,k}, z_{2,k}).$$

These functional relations are the same as those for the Rössler system in Eq. (23). For comparison, we also identify local linear approximations for the functions f_{x_2}, f_{y_2}, and f_{z_2}, in which we include variables with increasing order in h. For example, $f_{x_2}^1$ is a first-order approximation of f_{x_2} depending only on the variables $x_{2,k}$, $y_{2,k}$, and $z_{2,k}$.

The normalized estimation error as a function of the training data length N is shown in Fig. 6. Figure 6(b) shows that the local linear approximation using only output data to estimate y_2 has an estimation error that is smaller than that of a first-order approximation of f_{y_2}, but larger than that of a second-order approximation of f_{y_2}. This was expected since the second-order approximation depends on both $y_{1,k}$ and $x_{1,k}$, which cannot be uniquely determined from the two secondary state variables, $x_{2,k+1}$ and $z_{2,k+1}$.

In Figs. 6(a) and 6(c) we observe that, for an estimation of x_2 or z_2, local linear approximations using only output data outperform local linear approximations of the functions f_{x_2} and f_{z_2}. This is contrary to our expectation. A possible explanation is that the local linear approximation did not only model the relations between primary and secondary variables using the equations for the second Rössler system, but also reconstructed part of the dynamics of the first Rössler system.

V. SUMMARY AND DISCUSSIONS

We approached the problem of building models of input-output systems without any knowledge of the input. After examining existing embedding theorems, we proposed some theoretical arguments for the existence of a functional relation between delays of multivariate observations. The main assumptions are that the number of inputs is small and that the inputs can be uniquely determined by the current state and some of the observations at the next discrete time. Thus, multivariate observations are split into primary observations and secondary observations. The secondary observations are used for obtaining information about the input, while delays of the primary observations are used for obtaining information about the state of the system.

We first studied two examples, namely, a discrete map that is a discretized version of the Rössler system by using finite differences, and the Rössler system simulated by the Runge-Kutta method. We showed that, when compared to what they should be if the inputs could not be taken account of in the model, the estimation errors obtained for local linear models using only observed data are smaller by at least close to one order. Furthermore, the estimation errors follow the classical scaling law $E = \mathcal{O}(N^{-2/D})$. This indicates that successful models can be built by using only output data. A condition for this is that the secondary observations are functions of the states that are affected by the input.
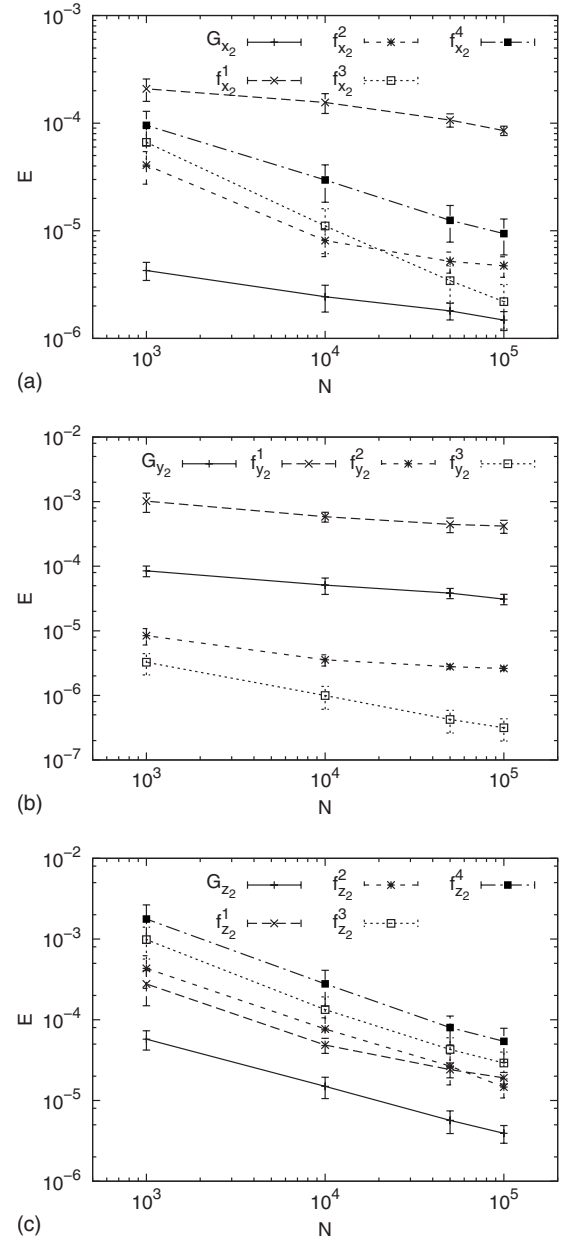

FIG. 6. Normalized estimation error for a local linear approximation of the two coupled Rössler systems as a function of the training data length N. The estimated variables are (a) x_2, (b) y_2, and (c) z_2. For each value of N, the mean and the standard deviation for ten sets of training and test data are shown.

An important application of this method is the analysis of coupled systems or complex networks. We can partition such a large system into several subsystems and study them separately. In this way, models based on the deterministic dynamics of each subsystem can be built, which would not be possible for the whole system because of its high dimensionality.

In order to verify the feasibility of this approach for the case of continuous-time systems, we studied the example of two coupled Rössler systems. The results indicate that in the case of a noiseless time series originating from these two continuous-time coupled systems, we cannot fully decouple the two subsystems to obtain a model of one of the two subsystems only. Although the two subsystems are coupled

only by one variable when viewed as continuous-time systems, the discrete equations controlling the evolution of a time series originating from these subsystems have more coupling terms. In the case of a discretization with the Runge-Kutta method, it was shown that because of the continuous-time nature of the system, in the discretized equations the inputs of the considered subsystem include all the state variables of the other subsystem. This dependence originates from the fact that information flows through each variable in the continuous-time equations. Despite this, the effect of some of the inputs on the equations of the considered subsystem is negligible; therefore, the proposed relations hold approximately.

Only noiseless time series were used in this paper. The effect of observational noise is reported elsewhere [25]. For noisy observations, the dependence on the additional variables will be hidden by the noise since these variables appear in the discretized equations in terms of higher order in the integration time h. In such a case, we can expect to separate the dynamics of the two subsystems since the noise hides the flow of information from the variables of one subsystem to the other.

It should be noted that the two proposed functional relations are nonpredictive relations; thus, they cannot be used to predict the future behavior of the system. However, since they model the dynamics of the system, they have other applications such as change detection and noise reduction. The application to change detection is reported elsewhere [25]. Also, preliminary results indicate that for noisy data, the estimates of the primary observations are less noisy than the original time series; thus, noise in the primary observations can be reduced.

ACKNOWLEDGMENTS

[1] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).

[2] F. Takens, Lect. Notes Math. **898**, 366 (1981).

[3] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).

[4] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, Berlin, 1996).

[5] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1999).

[6] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, Physica D **51**, 52 (1991).

[7] M. Casdagli, in *Nonlinear Modeling and Forecasting*, edited by M. Casdagli and S. Eubank (Addison-Wesley, Reading, 1992), Vol. XII.

[8] J. Stark, J. Nonlinear Sci. **9**, 255 (1999).

[9] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke, J. Nonlinear Sci. **13**, 519 (2003).

[10] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke, Nonlinear Anal. Theory, Methods Appl. **30**, 5303 (1997).

[11] J. Stark, Philos. Trans. R. Soc. London, Ser. A **358**, 41 (2000).

[12] L. Cao, A. Mees, and K. Judd, Physica D **121**, 75 (1998).

[13] S. P. Garcia and J. S. Almeida, Phys. Rev. E **72**, 027205 (2005).

[14] Y. Hirata, H. Suzuki, and K. Aihara, Phys. Rev. E **74**, 026202 (2006).

[15] L. A. Aguirre and C. Letellier, J. Phys. A **38**, 6311 (2005).

[16] S. Boccaletti, D. L. Valladares, L. M. Pecora, H. P. Geffert, and T. Carroll, Phys. Rev. E **65**, 035204(R) (2002).

[17] M. R. Muldoon, D. S. Broomhead, J. P. Huke, and R. Hegger, Dyn. Stab. Syst. **13**, 175 (1998).

[18] H. Whitney, Ann. Math. **37**, 645 (1936).

[19] K. Judd and A. Mees, Physica D **120**, 273 (1998).

[20] J. D. Farmer and J. J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).

[21] M. Casdagli, Physica D **35**, 335 (1989).

[22] O. E. Rössler, Phys. Lett. **57A**, 397 (1976).

[23] C. Letellier, S. Elaydi, L. A. Aguirre, and A. Alaoui, Physica D **195**, 29 (2004).

[24] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 2002).

[25] E. Monroig, Ph.D. thesis, University of Tokyo, 2009.

[26] It also preserves the product structure of $\mathcal{M} \times \mathcal{N}$, but we do not use this in the present discussion.