Estimability and dependency analysis of model parameters based on delay coordinates

J. Schumann-Bischoff,* S. Luther,† and U. Parlitz‡

Biomedical Physics Group, Max Planck Institute for Dynamics and Self-Organization, Am Faßberg 17, 37077 Göttingen, Germany
and Institute for Nonlinear Dynamics, Georg-August-Universität Göttingen, Am Faßberg 17, 37077 Göttingen, Germany

(Received 29 February 2016; revised manuscript received 9 August 2016; published 28 September 2016)

In data-driven system identification, values of parameters and not observed variables of a given model of a dynamical system are estimated from measured time series. We address the question of estimability and redundancy of parameters and variables, that is, whether unique results can be expected for the estimates or whether, for example, different combinations of parameter values would provide the same measured output. This question is answered by analyzing the null space of the linearized delay coordinates map. Examples with zero-dimensional, one-dimensional, and two-dimensional null spaces are presented employing the Hindmarsh-Rose model, the Colpitts oscillator, and the Rössler system.

DOI: 10.1103/PhysRevE.94.032221

I. INTRODUCTION

Computer simulations based on mathematical models are an important method for analyzing and applying dynamical systems in physics and many other scientific fields. In many cases, models are given by a set of ordinary or partial differential equations (ODEs or PDEs) including parameters which have to be specified by suitable measurements or using estimation methods based on time series measured from the process the model aims at describing.

The latter data-driven state and parameter estimation methods include synchronization- and observer-based methods [1–4], different types of Kalman filters [5–7], or particle filters [8,9]. With optimization-based methods [10,11], the unknown states and parameters are estimated by minimizing a cost function (for example, by using numerical optimization or by solving the Euler-Lagrange equations [7,12]). Its Bayesian probabilistic background is described, for example, in Refs. [7,13,14]. In the geosciences these data-assimilation methods are known as 4D-Var [15,16].

All the above-mentioned methods provide estimates for the model variables and parameters. However, it is also important to know how accurate and unique these estimates are. If there are different solutions for certain model variables and parameters which describe the measured data with a comparable accuracy, then this is a hint that the data do not contain enough information for achieving (almost) unique estimates or that some parameters are redundant for specifying the dynamics. In many cases, it is therefore desirable to identify those quantities which cannot uniquely be estimated from the available time series, a topic which is closely related to the concept of observability [17–26]. While this task is completely solved for linear systems, it remains a challenge for nonlinear models [21,22]. In general, in control theory one considers models given by nonlinear ODEs

$$\dot{\mathbf{x}}(t) = \mathbf{F}[\mathbf{x}(t), \mathbf{p}, \mathbf{u}(t)], \tag{1}$$

with the model state $\mathbf{x} = (x_1, x_2, \ldots, x_D)^{\mathrm{tr}}$, the input (control variable) $\mathbf{u}(t)$, and the parameter vector $\mathbf{p} = (p_1, p_2, \ldots, p_{N_\mathrm{p}})^{\mathrm{tr}}$. Furthermore, a scalar measurement function is defined by

$$y(t) = h[\mathbf{x}(t)], \tag{2}$$

whose output signal $y(t)$ represents measured data.

Often in this context one is interested in the *identifiability* of the model Eq. (1) for some specific parameter values. For a definition of identifiability we follow Ref. [27] and consider two solutions $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$, $t \geqslant t_0$, of Eq. (1) generated using the same initial state $\mathbf{x}(t_0)$, the same input $\mathbf{u}(t)$, and the parameter vectors $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively. If an initial state $\mathbf{x}(t_0)$ and an input $\mathbf{u}(t)$ exists such that the trajectories coincide $[\mathbf{x}_1(t) = \mathbf{x}_2(t)\, \forall\, t \geqslant t_0]$ only if $\mathbf{p}_1 = \mathbf{p}_2$, then the model is *identifiable at* $\mathbf{p}_1$. Note that the identifiability is a property independent of the measurement function.

An important issue for identifiability is the existence of a suitable input $\mathbf{u}(t)$. It has to apply an excitation to the system that produces different responses for different parameter values for some suitable initial value of the system; i.e., the input has to be *informative*.

In contrast to the identifiability, the *observability* is a property of the system consisting of the model Eq. (1) and the measurement function Eq. (2). Let us consider two solutions $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$, $t \geqslant t_0$, of Eq. (1), each generated with the same input $\mathbf{u}(t)$ and having the same output $y(t)$. If this implies that the initial states of both solutions have to be equal, $\mathbf{x}_1(t_0) = \mathbf{x}_2(t_0)$, then the system is observable at $\mathbf{x}_1$. This means that a system is observable at $\mathbf{x}(t_0)$ if and only if $\mathbf{x}(t_0)$ can be uniquely reconstructed from the output $y(t)$ and the input $\mathbf{u}(t)$, $t \geqslant t_0$. If two different outputs $y_1(t)$ and $y_2(t)$ are available (resulting from two different measuring functions) for a solution $\mathbf{x}(t)$ with initial value $\mathbf{x}(t_0)$ and we are able to determine $\mathbf{x}(t_0)$ from both sets of finite data $[\mathbf{u}(t), y_1(t)]$ and $[\mathbf{u}(t), y_2(t)]$, then the system is observable from both chosen outputs. In general, one can interpret model parameters $\mathbf{p}$ as (additional) state variables with trivial dynamics $\dot{\mathbf{p}} = 0$. In this sense not only the state variables but also the parameters of an observable (extended) system can be uniquely estimated.

Often nonlinear dynamical models are formulated without an input, for example when investigating systems that cannot be (arbitrarily) driven or where only a measurement is available without external input $\mathbf{u}(t)$. Therefore, in this article we

*jan.schumann-bischoff@ds.mpg.de
†stefan.luther@ds.mpg.de
‡ulrich.parlitz@ds.mpg.de

consider this special case and focus on the input-free model

$$\dot{\mathbf{x}}(t) = \mathbf{F}[\mathbf{x}(t),\mathbf{p}] \tag{3}$$

instead of Eq. (1). Furthermore, we consider the common situation that only the output $y(t)$ [Eq. (2)] instead of the full state vector $\mathbf{x}(t)$ is experimentally accessible.

For a given model and a given measurement function it is possible that some state variables or parameters can be uniquely estimated from the given time series and others not. So the full state vector cannot be determined and the system is nonobservable. In this article we address this case and propose a method for identifying *estimable* quantities which can, in principle, be determined uniquely from the available time series and groups of *redundant* quantities (parameters) whose values cannot uniquely be determined from the given data.

To achieve this goal we make use of time-delay coordinates [28–32] of $y(t)$ and consider the K-dimensional forward delay coordinates map G,

$$\mathbf{G} : \mathbb{R}^D \times \mathbb{R}^{N_\mathrm{p}} \to \mathbb{R}^K, \quad (\mathbf{x}(t),\mathbf{p}) \mapsto \mathbf{G}[\mathbf{x}(t),\mathbf{p}], \tag{4}$$

with

$$\begin{aligned}\mathbf{g} &= \mathbf{G}[\mathbf{x}(t),\mathbf{p}] \\ &= [y(t), y(t+\tau), \ldots, y[t+(K-1)\tau]]^{\mathrm{tr}},\end{aligned} \tag{5}$$

where τ is the delay time. In general, the delay reconstruction vector $\mathbf{G}[\mathbf{x}(t),\mathbf{p}]$ consists of a time series of K samples. Here $\mathbf{G}[\mathbf{x}(t),\mathbf{p}]$ is computed based on the measurement function Eq. (2), the model Eq. (3), and an initial state. For that, the output $y(t+i\tau), i = 0,\ldots,K-1$, of the system [see Eq. (2)] is computed from the full state $\mathbf{x}(t+i\tau)$, which, in turn, is computed by integrating the model Eq. (3) from t to $t+i\tau$ using the initial state $\mathbf{x}(t)$.

According to [33] it is sufficient to choose $K \geqslant 2(D+N_\mathrm{p})+1$ to make statements about the estimability of model parameters and variables. For smaller K the correctness of the statements is, in general, not guaranteed.

If the map $\mathbf{G}$ is locally invertible at given $\mathbf{x}$ and $\mathbf{p}$, then $\mathbf{x}$ and $\mathbf{p}$ can be uniquely reconstructed given $y(t)$ at present and delayed times. That is, $\mathbf{x}$ and $\mathbf{p}$ are *locally estimable*. This is the case if the $K \times (D+N_\mathrm{p})$ Jacobian matrix $D\mathbf{G} = D\mathbf{G}(\mathbf{x}(t),\mathbf{p})$, computed with respect to $\mathbf{x}(t)$ and $\mathbf{p}$, is locally invertible at $(\mathbf{x}(t),\mathbf{p})$, i.e., has full rank.

To compute $D\mathbf{G}$ we consider the flow

$$\phi^{\tau'} : \mathbb{R}^D \otimes \mathbb{R}^{N_\mathrm{p}} \to \mathbb{R}^D, \quad (\mathbf{x}(t),\mathbf{p}) \mapsto \mathbf{x}(t+\tau') \tag{6}$$

generated by Eq. (3), such that $y(t+\tau') = h[\phi^{\tau'}(\mathbf{x}(t),\mathbf{p})]$. The components of $D\mathbf{G}$ consist of derivatives of $y(t+\tau')$, $\tau' = 0,\tau,2\tau,\ldots$, computed with respect to $x_j(t)$ and p_k which, in turn, contain derivatives of the form $\partial\phi_i^{\tau'}(\mathbf{x}(t),\mathbf{p})/\partial x_j(t)$ and $\partial\phi_i^{\tau'}(\mathbf{x}(t),\mathbf{p})/\partial p_k$ ($\phi_i^{\tau'}$ is the ith component of $\phi^{\tau'}$) and are solutions of the linearized model equations (see, e.g., Ref. [34] for details).

Instead of considering the observability of the full system, here we focus on identifying the (local) estimability of individual variables and parameters. If some of these unknown quantities (parameters and variables) are not locally estimable, we investigate their relationships and address the question of which parameters may be fixed to obtain local estimability for all (remaining) quantities. The approach discussed in detail in Sec. II A is based on investigating the null space of the Jacobian matrix $D\mathbf{G}$ of the map $\mathbf{G}$. Not estimable quantities and their relationships are identified by a suitable choice of the basis of the null space using methods adopted from Ref. [35].

In Refs. [34,36] it is investigated how small perturbations of the delay reconstruction vector $\mathbf{g}$ are mapped to small perturbations in the state and parameter space. This approach is extended in Sec. II B to find out how the unknown quantities are locally correlated.

In Sec. III the optimization-based (weak-4D-Var type) state and parameter estimation algorithm from Refs. [37,38] is revisited. This estimation method is then used in the subsequent examples to evaluate the results obtained by the application of the previously suggested analyses. In particular, we make use of the concept of a profile likelihood [39,40], where one model parameter is manually tuned and all others are estimated (beside the model variables) by minimizing a cost function.

Three examples, the Colpitts oscillator [41], the Rössler model [42–44], and the Hindmarsh-Rose neuron model [45], are discussed in Sec. IV. In Refs. [43,44], based on derivative coordinates, an algebraic method was presented which can be used to find functional relationships between model parameters of polynomial models [the right-hand side of Eq. (3) is a polynomial vector field] so that the measured variable (and its higher-order time derivatives) remains unchanged. We demonstrate for the Rössler model that our (more general) approach for identifying not estimable variables and parameters and their relations provides the same results as the method presented in Refs. [43,44].

II. THEORY

We investigate the local estimability of model variables and parameters of a model, Eq. (3), with a scalar measurement function, Eq. (2), by means of a delay reconstruction map, Eq. (5).

Following the approach presented in Refs. [34,36] we consider how small perturbations of the reconstructed state $\mathbf{g}$ are related to variations of the state vector $\mathbf{x}$ and the parameters $\mathbf{p}$. This approach can be easily extended to multivariate measurement functions [46]. To simplify the discussion we introduce a vector of all $N_\mathrm{w} = D + N_\mathrm{p}$ unknowns $\mathbf{w} = (w_1, w_2, \ldots, w_{N_\mathrm{w}})^{tr} = (\mathbf{x},\mathbf{p})^{\mathrm{tr}} \in \mathbb{R}^{N_\mathrm{w}}$. If perturbations $\Delta\mathbf{g} = \tilde{\mathbf{g}} - \mathbf{g}$ are (infinitesimally) small, then the unknowns $\mathbf{w} = \mathbf{G}^{-1}(\mathbf{g})$ and the perturbed unknowns $\tilde{\mathbf{w}} = \mathbf{G}^{-1}(\tilde{\mathbf{g}})$ can be used in the linearization

$$\mathbf{G}^{-1}(\tilde{\mathbf{g}}) = \mathbf{G}^{-1}(\mathbf{g}) + D\mathbf{G}^{-1}(\tilde{\mathbf{g}} - \mathbf{g}) \tag{7}$$

$$\Rightarrow \tilde{\mathbf{w}} - \mathbf{w} = D\mathbf{G}^{-1}(\tilde{\mathbf{g}} - \mathbf{g}) \tag{8}$$

to compute the (resulting) perturbation of unknown quantities,

$$\Delta\mathbf{w} = \tilde{\mathbf{w}} - \mathbf{w} = D\mathbf{G}^{-1}\Delta\mathbf{g}, \tag{9}$$

with $\Delta\mathbf{w} = (\Delta w_1, \Delta w_2, \ldots, \Delta w_{N_\mathrm{w}})$. The Jacobian matrix $D\mathbf{G}^{-1}$ of $\mathbf{G}^{-1}$ is the (pseudo) inverse of the Jacobian $D\mathbf{G}(\mathbf{w}) = D\mathbf{G}(\mathbf{x},\mathbf{p})$ and can be computed by inversion of its singular value decomposition [35,47] (SVD),

$$D\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{tr}}, \tag{10}$$

where the $K \times N_w$ matrix $\mathbf{U}$ is column orthonormal and the $N_w \times N_w$ matrix $\mathbf{V}$ is orthonormal; i.e., $\mathbf{V}^{tr} = \mathbf{V}^{-1}$. The elements of the $N_w \times N_w$ diagonal matrix

$$\mathbf{S} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{N_w}) \tag{11}$$

are the singular values $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_{N_w} \geqslant 0$ and the pseudoinverse of $D\mathbf{G}$ is then given by

$$D\mathbf{G}^{-1} = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^{tr}. \tag{12}$$

A. Dependency analysis of variables and parameters

The Jacobian matrix $D\mathbf{G}$ contains information about the local estimability of variables and parameters. To obtain this information, we rewrite Eq. (9) as

$$D\mathbf{G}\Delta\mathbf{w} = \Delta\mathbf{g}. \tag{13}$$

If there exists a $\Delta\mathbf{w} \neq 0$ for which $\Delta\mathbf{g} = 0$, then we know that there exist perturbations of $\mathbf{w}$ which do not lead to perturbations of $\mathbf{g}$; that is, they do not affect the output signal $y(t)$. Since they have no impact on the measured signal, the values of all quantities w_i which are involved in such perturbations cannot be uniquely estimated from a y time series. To identify nonestimable quantities, we therefore want to find out whether any $\Delta\mathbf{w} \neq 0$ exists, which fulfills

$$D\mathbf{G}\Delta\mathbf{w} = 0; \tag{14}$$

that is, we want to compute the null space (or kernel) null($D\mathbf{G}$) of the matrix $D\mathbf{G}$. If the dimension of null($D\mathbf{G}$), the nullity, is $D_N = 0$, then the null space contains only the null vector and we know that any small variation of $\mathbf{w}$ leads to a perturbation of $\mathbf{g}$ and, hence, all quantities in $\mathbf{w}$ are locally estimable. If $D_N > 0$, then certain elements w_i of $\mathbf{w}$ can be varied by $\Delta w_i \neq 0$ *without* a corresponding perturbation of $\mathbf{g}$ (i.e., with $\Delta\mathbf{g} = 0$). Therefore, these quantities w_i are not locally estimable from measurements given by h, and in the following they are also called *locally redundant*. For all i for which it is $\Delta w_i \neq 0$ the associated columns $D\mathbf{G}^{(i)}$ of $D\mathbf{G}$ are linearly dependent, as one can see when expanding Eq. (14) to

$$D\mathbf{G}^{(1)}\Delta w_1 + \cdots + D\mathbf{G}^{(N_w)}\Delta w_{N_w} = 0. \tag{15}$$

To investigate the null space, we exploit the SVD [35,47] of $D\mathbf{G}$, Eq. (10). The nullity D_N is given by the number of vanishing singular values ($\sigma_i = 0$). In cases where $D_N > 0$ the null space of $D\mathbf{G}$ can be spanned by D_N basis vectors. For ordered singular values $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_{N_w - D_N + 1} = \cdots = \sigma_{N_w} = 0$ a set of D_N orthonormal basis vectors $\mathbf{v}_B^{(i)}$, $i = 1, \ldots, D_N$, spanning null($D\mathbf{G}$) is given by the last D_N columns of $\mathbf{V}$. Then every perturbation $\Delta\mathbf{w} \in$ null($D\mathbf{G}$) can be expressed as a linear combination of these basis vectors,

$$\Delta\mathbf{w} = a_1\mathbf{v}_B^{(1)} + \cdots + a_{D_N}\mathbf{v}_B^{(D_N)}, \tag{16}$$

with the real coefficients $\mathbf{a} = (a_1, \ldots, a_{D_N})^{tr}$. If the ith components of all basis vectors are zero, then a perturbation $\Delta w_i \neq 0$ is *not* possible within null($D\mathbf{G}$). This means that any (small) perturbation of w_i leads to a perturbation $\Delta g \neq 0$ of the measured time series represented by $\mathbf{g}$ and, therefore, w_i can be uniquely reconstructed from $\mathbf{g}$.

To summarize the basis vectors, we introduce a $N_w \times D_N$ basis matrix $\mathbf{V}_B = [\mathbf{v}_B^{(1)}, \ldots, \mathbf{v}_B^{(D_N)}]$ whose columns are the last D_N columns of $\mathbf{V}$. Then, Eq. (16) can be rewritten as

$$\Delta\mathbf{w} = \mathbf{V}_B\mathbf{a}. \tag{17}$$

If the ith row of $\mathbf{V}_B$ contains zeros only, then this implies $\Delta w_i = 0$ and a perturbation of w_i within null($D\mathbf{G}$) is not possible. As mentioned before, this means that the variable or parameter w_i can be uniquely reconstructed from $\mathbf{g}$.

Since the SVD is typically computed numerically (as in the examples in Sec. IV), the singular values may be very small but not exactly zero. Therefore, in the following a singular value σ_i is considered as "zero" (or vanishing) if the ratio of σ_i and the maximal singular value σ_1 is of the order of magnitude of the machine precision, $\sigma_i/\sigma_1 \approx 10^{-15}$.

A situation which may occur (but is not considered in this article) is that the spectrum of singular values can show a smooth transition to very small singular values (of orders of magnitude 10^{-15}). In this case, one may not clearly distinguish between vanishing and nonvanishing singular values and an additional criterion for specifying "vanishing" singular values would be required. In the examples we consider in Sec. IV, however, that this difficulty does not occur and vanishing singular values can clearly be identified because nonvanishing singular values are orders of magnitude larger than vanishing ones.

1. *Investigating the null space*

In the case when locally not estimable variables and parameters are recognized, the question arises as to which of them may be removed from the analysis or the estimation problem so that, as a result, all remaining quantities are locally estimable. We address this issue by considering the situation for a one- and a multidimensional null space separately.

If the null space of $D\mathbf{G}$ is $D_N = 1$ dimensional only and, hence, can be spanned by the basis vector $\mathbf{v}_B$ (last column of $\mathbf{V}$), then it contains any variation

$$\Delta\mathbf{w} = a \cdot \mathbf{v}_B, \tag{18}$$

where $a \in \mathbb{R}$.

The *pattern of zero and nonzero elements* of $\mathbf{v}_B$ determines which quantities w_i are locally estimable and which are not. This is illustrated in Fig. 1. In Fig. 1(a) only the first component of $\mathbf{v}_B$ is different from zero and any perturbation $\Delta w_i \neq 0$ of w_2 or w_3 would leave the null space and have an impact of the reconstructed state ($\Delta g \neq 0$). Therefore, in this case w_2 and w_3 are estimable from the given time series. Figure 1(b) illustrates a case where the basis vector $\mathbf{v}_B$ contains only a single vanishing component and w_3 is estimable. If all components of $\mathbf{v}_B$ are different from zero [Fig. 1(c)], then all quantities w_i are not locally estimable. In this case one of the quantities has to be fixed to a specific value such that no perturbations of this quantity are possible any longer. For the three-dimensional example shown in Fig. 1 this means that perturbations are restricted to a two-dimensional plane which intersects the line representing the null space spanned by $\mathbf{v}_B$ in a point. The additional constraint (here $\Delta w_2 = 0$) thus reduces the dimension of the null space from one to zero such that both remaining quantities (here w_1 and w_3) become locally estimable.

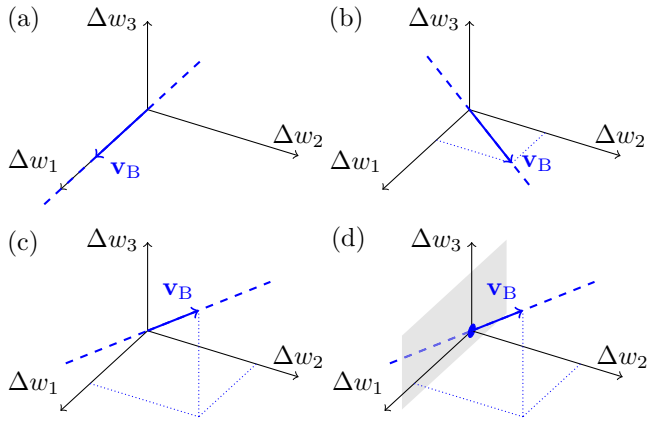

FIG. 1. Illustration of different orientations of a one-dimensional null space.

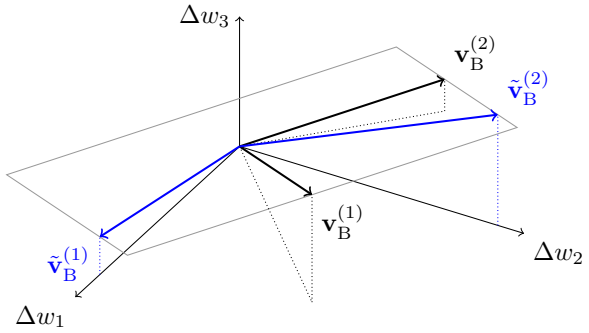

FIG. 2. Illustration of different bases of a two-dimensional null space. $\mathbf{v}_\mathrm{B}^{(1)}$ and $\mathbf{v}_\mathrm{B}^{(2)}$ denote the orthonormal basis vectors obtained with the SVD, while $\tilde{\mathbf{v}}_\mathrm{B}^{(1)}$ and $\tilde{\mathbf{v}}_\mathrm{B}^{(2)}$ are the nonorthogonal basis vectors with vanishing components computed using Eq. (22).

Fixing w_i by prohibiting its variation via $\Delta w_i = 0$ by setting w_i to a fixed value instead of estimating it in an estimation problem immediately sets $a = 0$ in Eq. (18) if $v_{\mathrm{B},i} \neq 0$. Therefore, variations of all other quantities are also prohibited because of $\Delta\mathbf{w} = 0$. In other words, setting any variable w_i whose corresponding component of the basis vector $\mathbf{v}_\mathrm{B}$ does not vanish ($v_{\mathrm{B},i} \neq 0$) to a fixed value makes all other quantities locally estimable.

Next we consider the situation where the null space of $D\mathbf{G}$ is $D_\mathrm{N} > 1$ dimensional. That means that at least D_N quantities have to be fixed by prohibiting their variation to make all quantities locally estimable. The variation within the null space can be expressed as a linear combination of the basis vectors $\mathbf{v}_\mathrm{B}^{(i)}$ constituting the basis matrix $\mathbf{V}_\mathrm{B}$ (obtained via SVD; see Sec. II A) and the real coefficients summarized in $\mathbf{a}$; see Eqs. (16) and (17).

The question is now which quantities are to be fixed so that all remaining quantities become locally estimable, i.e., $\Delta\mathbf{w} = 0$. Component Δw_i vanishes if the corresponding ith row of the basis matrix $\mathbf{V}_\mathrm{B}$ contains only vanishing elements. If this is not the case, the corresponding prefactor a_i has to be set to zero by fixing another quantity w_j. This depends on the pattern of zero elements of the basis matrix $\mathbf{V}_\mathrm{B}$. If, for example, for three unknowns and $D_\mathrm{N} = 2$ the basis matrix has the form

$$\mathbf{V}_\mathrm{B} = \left[\mathbf{v}_\mathrm{B}^{(1)}, \mathbf{v}_\mathrm{B}^{(2)}\right] = \begin{pmatrix} v_{\mathrm{B},1}^{(1)} & 0 \\ v_{\mathrm{B},2}^{(1)} & v_{\mathrm{B},2}^{(2)} \\ v_{\mathrm{B},3}^{(1)} & v_{\mathrm{B},3}^{(2)} \end{pmatrix}, \tag{19}$$

then Eq. (17) implies that fixing (w_1, w_2) or (w_1, w_3) makes all quantities locally estimable [$(\Delta w_1, \Delta w_2) = (0,0)$ or $(\Delta w_1, \Delta w_3) = (0,0)$ imply $\Delta\mathbf{w} = 0$], while fixing (w_2, w_3) does not imply $a_1 = 0$ and $\Delta\mathbf{w} = 0$.

The occurrence of zero elements of the basis matrix $\mathbf{V}_\mathrm{B}$ may have two possible reasons. (i) If quantity w_i is locally estimable, then the ith row of $\mathbf{V}_\mathrm{B}$ is occupied by zero elements (and thus $\Delta w_i = 0$). This pattern of zeros occurs whenever a SVD of a Jacobian matrix is computed at a state vector where the local estimability condition is fulfilled and since this is typically the case for (almost) all states, it is also characteristic for the distribution of the corresponding component computed with a representative set of states on the attractor. In this sense, (locally) estimable quantities w_i are immediately visible in the basis matrix $\mathbf{V}_\mathrm{B}$ as rows with zero elements. (ii) The second type of vanishing element of the basis matrix $\mathbf{V}_\mathrm{B}$ is less obvious, because it depends on the (proper) orientation of the basis vectors of the null space. This issue is illustrated in Fig. 2 for the case $D_\mathrm{N} = 2$. The numerical SVD provides a pair of orthonormal vectors $\mathbf{v}_\mathrm{B}^{(1)}$ and $\mathbf{v}_\mathrm{B}^{(2)}$ which span the null space but contain no vanishing elements. On the other hand, there exist (in general, nonorthogonal) basis vectors which are partly aligned with the axes of the quantities to be estimated and thus possess vanishing elements as shown for a particular example ($\tilde{\mathbf{v}}_\mathrm{B}^{(1)}$ and $\tilde{\mathbf{v}}_\mathrm{B}^{(2)}$) in Fig. 2. In the following we introduce a systematic change and choice of the basis of the null-space that leads to additional vanishing components of the basis vectors. An important feature of the resulting patterns of vanishing components in the basis matrix is that they are the same at different locations in state space (if the local estimability does not change on the attractor). To compute this basis transformation, we follow the approach from Ref. [35] for finding linear dependent columns of a matrix. For any basis of the null space the following holds:

$$D\mathbf{G} \cdot \mathbf{V}_\mathrm{B} = 0. \tag{20}$$

First, we choose D_N quantities and define the square $D_\mathrm{N} \times D_\mathrm{N}$ matrix $\mathbf{V}_{\mathrm{B},1}$ to contain the D_N rows of $\mathbf{V}_\mathrm{B}$ which are associated with these quantities, with the restriction that $\mathbf{V}_{\mathrm{B},1}$ is nonsingular. This implies that the nullity of $\mathbf{V}_{\mathrm{B},1}$ is zero. Prohibiting the variations of the chosen quantities w_i via $\Delta w_i = 0$ immediately sets all coefficients $a_j = 0$, which makes every quantity locally estimable and, therefore, is already an answer to the question of which quantities to fix. Nevertheless, the new basis to construct also allows to read off which other quantities may be fixed instead to obtain local estimability.

One possible strategy to choose D_N quantities is to try different sets of chosen quantities, compute the singular values of $\mathbf{V}_{\mathrm{B},1}$, and choose the combination with the largest ratio $\sigma_{\min}/\sigma_{\max}$. $\mathbf{V}_{\mathrm{B},2}$ contains all rows of $\mathbf{V}_\mathrm{B}$ not contained in $\mathbf{V}_{\mathrm{B},1}$. Similarly, we define $D\mathbf{G}_1$ which contains the D_N columns of $D\mathbf{G}$ which are associated with the chosen quantities. $D\mathbf{G}_2$

contains all remaining columns of $D\mathbf{G}$. Next, as suggested in Ref. [35], we rewrite Eq. (20) to

$$[\mathrm{D}\mathbf{G}_1 \quad \mathrm{D}\mathbf{G}_2]\begin{bmatrix}\mathbf{V}_{\mathrm{B},1}\\ \mathbf{V}_{\mathrm{B},2}\end{bmatrix} = \mathrm{D}\mathbf{G}_1 \cdot \mathbf{V}_{\mathrm{B},1} + \mathrm{D}\mathbf{G}_2 \cdot \mathbf{V}_{\mathrm{B},2} = 0, \tag{21}$$

multiply with $\mathbf{V}_{\mathrm{B},1}^{-1}$ from the right, and obtain

$$D\mathbf{G}_1 \cdot \mathbb{1}(D_\mathrm{N}) + D\mathbf{G}_2 \cdot \mathbf{V}_{\mathrm{B},2}\mathbf{V}_{\mathrm{B},1}^{-1} = [D\mathbf{G}_1 \quad D\mathbf{G}_2]\underbrace{\begin{bmatrix}\mathbb{1}(D_\mathrm{N})\\ \mathbf{V}_{\mathrm{B},2}\mathbf{V}_{\mathrm{B},1}^{-1}\end{bmatrix}}_{\widetilde{\mathbf{V}}'_\mathrm{B}} = 0, \tag{22}$$

where $\mathbb{1}(D_\mathrm{N})$ is the $D_\mathrm{N} \times D_\mathrm{N}$ identity matrix. Now, we reorder the columns of $[D\mathbf{G}_1 \quad D\mathbf{G}_2]$ to obtain $D\mathbf{G}$ again. If we apply the same reordering to the rows of $\widetilde{\mathbf{V}}'_\mathrm{B}$, then we obtain the $N_\mathrm{w} \times D_\mathrm{N}$ matrix $\widetilde{\mathbf{V}}_\mathrm{B}$ and therefore the newly ordered version of Eq. (22) is

$$D\mathbf{G} \cdot \widetilde{\mathbf{V}}_\mathrm{B} = 0. \tag{23}$$

Due to the identity matrix in $\widetilde{\mathbf{V}}'_\mathrm{B}$ the columns of $\widetilde{\mathbf{V}}'_\mathrm{B}$, and hence the columns of $\widetilde{\mathbf{V}}_\mathrm{B}$, are linearly independent. Since the number of columns of $\widetilde{\mathbf{V}}_\mathrm{B}$, D_N, is equal to the dimension of the null space of $D\mathbf{G}$, the columns of $\widetilde{\mathbf{V}}_\mathrm{B} = [\widetilde{\mathbf{v}}_\mathrm{B}^{(1)}, \ldots, \widetilde{\mathbf{v}}_\mathrm{B}^{(D_\mathrm{N})}]$ form another basis (beside the orthogonal basis $\mathbf{V}_\mathrm{B}$) of the null space of $D\mathbf{G}$. Depending on the selection of D_N rows of $\mathbf{V}_\mathrm{B}$, this new basis is unique and there are no (rotational) degrees of freedom left for the basis vectors. This feature is crucial for comparing results at different states and for computing histograms of relevant quantities as it will be done for the examples presented in Sec. IV.

Furthermore, all basis vectors have a mutually different pattern of nonzero elements. Every variation of $\mathbf{w}$ within the null space of $D\mathbf{G}$ can then be expressed in terms of the new basis vectors with the real coefficients $\widetilde{a}_i$,

$$\Delta\mathbf{w} = \widetilde{a}_1\widetilde{\mathbf{v}}_\mathrm{B}^{(1)} + \cdots + \widetilde{a}_{D_\mathrm{N}}\widetilde{\mathbf{v}}_\mathrm{B}^{(D_\mathrm{N})}. \tag{24}$$

How the information about possible perturbations along the new basis vectors can be used to find local relationships between variables and parameters is discussed in the next section.

2. *Local relationships*

To reveal some local relationships between quantities, we consider variations $\Delta\mathbf{w}^{(i)}$ along each basis vector $\widetilde{\mathbf{v}}_\mathrm{B}^{(i)} = [\widetilde{v}_{\mathrm{B},1}^{(i)}, \ldots, \widetilde{v}_{\mathrm{B},N_\mathrm{w}}^{(i)}]^\mathrm{tr}$, which are given by

$$\Delta\mathbf{w}^{(i)} = \widetilde{a}_i\widetilde{\mathbf{v}}_\mathrm{B}^{(i)} \tag{25}$$

and can be obtained by setting $\widetilde{a}_j = 0$ for all $j \neq i$ in Eq. (24). In the case of a one-dimensional null space it is $\widetilde{\mathbf{v}}_\mathrm{B}^{(i)} = \mathbf{v}_\mathrm{B}$; see Eq. (18). This illustrates that simultaneous variations of all quantities w_k, where the corresponding components $v_{\mathrm{B},k}^{(i)}$ are nonzero, are possible without affecting the output of the system at present and delayed times, $\Delta\mathbf{g} = 0$. All other quantities are kept unchanged. That indicates a local dependency between these quantities. A *set of locally dependent quantities* exists for every basis vector and contains all quantities where the corresponding component of the basis vector is nonzero. Note that such a set is not unique because the patterns of nonzero elements of $\widetilde{\mathbf{v}}_\mathrm{B}^{(i)}$ depend of the choice of quantities to construct $\mathbf{V}_{\mathrm{B},1}$.

Furthermore, to investigate how quantities w_j and w_k locally depend on each other (still assuming variations along $\widetilde{\mathbf{v}}_\mathrm{B}^{(i)}$ only), dividing the jth row of Eq. (25) by the kth row, we consider

$$\Delta w_j^{(i)} = \frac{\widetilde{v}_{\mathrm{B},j}^{(i)}}{\widetilde{v}_{\mathrm{B},k}^{(i)}}\Delta w_k^{(i)}. \tag{26}$$

Whether an increase of w_k leads to an increase or a decrease of w_j depends on the signs of the components of the basis vectors. This point is illustrated in greater detail in the examples discussed in Sec. IV. It should be noted that computing Eq. (26) based on another basis vector than $\widetilde{\mathbf{v}}_\mathrm{B}^{(i)}$ may lead to a different dependency, but that would not be a contradiction (see the example with the Rössler system, Sec. IV B).

B. Correlation analysis

In Sec. II A we described a way to detect locally dependent quantities (variables and parameters) by investigating the null space of $D\mathbf{G}$ for cases where the null space is one- or higher dimensional. There, no assumptions were made about the distribution of perturbations $\Delta\mathbf{g}$ in Eq. (13). Here we follow Refs. [34,36] and make the assumption that the perturbations $\Delta\mathbf{g}$ are multivariate normal distributed with a covariance matrix Σ_g and a mean of zero,

$$\Delta\mathbf{g} \sim \mathcal{N}(0, \Sigma_\mathrm{g}). \tag{27}$$

Since $\Delta\mathbf{g}$ is locally mapped to $\Delta\mathbf{w}$ via the linear function Eq. (9), $\Delta\mathbf{w}$ is also multivariate normal distributed [13, Theorem 2.11] with a mean of zero and a covariance matrix Σ_w,

$$\Delta\mathbf{w} \sim \mathcal{N}(0, \Sigma_\mathrm{w}), \tag{28}$$

where the covariance matrix is given by [13, Theorem 2.11]:

$$\Sigma_\mathrm{w} = \mathrm{D}\mathbf{G}^{-1}\Sigma_\mathrm{g}[\mathrm{D}\mathbf{G}^{-1}]^\mathrm{tr}. \tag{29}$$

Furthermore, the perturbed state $\tilde{\mathbf{w}}$ [see Eq. (9)] is also normally distributed [13, Theorem 2.11],

$$\tilde{\mathbf{w}} \sim \mathcal{N}(\mathbf{w}, \Sigma_\mathrm{w}), \tag{30}$$

and has the same covariance matrix.

In Refs. [34,36] it was assumed that the perturbations of the delay reconstruction vector $\mathbf{g}$ are uncorrelated and have the same variance ζ^2 (a typical assumption for measurement noise). That is, its covariance matrix is diagonal, $\Sigma_\mathrm{g} = \zeta^2 \cdot \mathbb{1}$, where $\mathbb{1}$ denotes the identity matrix. Using Eq. (12) the standard deviation of single parameters or variables is given by the diagonal elements of

$$\sqrt{\Sigma_\mathrm{w}} = \zeta\sqrt{D\mathbf{G}^{-1}[D\mathbf{G}^{-1}]^\mathrm{tr}} \tag{31}$$

$$= \zeta\sqrt{[\mathbf{V}\mathbf{S}^{-2}\mathbf{V}^\mathrm{tr}]}, \tag{32}$$

where the square roots are meant to be computed componentwise. Since ζ is a factor only, we set $\zeta = 1$ and

have defined the *measure of uncertainty* [34,36] for w_k:

$$\nu_k = \nu(w_k) = \sqrt{[\mathbf{V}\mathbf{S}^{-2}\mathbf{V}^{\mathrm{tr}}]_{kk}}. \tag{33}$$

The larger ν_k is, the larger is the uncertainty when estimating w_k and the worse w_k can be estimated.

In contrast to Refs. [34,36], we now consider the nondiagonal elements to investigate the correlation between two different quantities. This is done by considering the $N_{\mathrm{w}} \times N_{\mathrm{w}}$ *linear Pearson's correlation matrix* $\boldsymbol{\rho}$ consisting of the correlation coefficients

$$\rho_{ij} = \rho(w_i, w_j) = \frac{\Sigma_{\mathrm{w},ij}}{\sqrt{\mathrm{var}(w_i)}\sqrt{\mathrm{var}(w_j)}} \in [-1,1] \tag{34}$$

and containing the variances $\mathrm{var}(w_k) = \Sigma_{\mathrm{w},kk}$. Since Σ_{w} is the covariance matrix of both $\Delta\mathbf{w}$ and $\tilde{\mathbf{w}}$, their correlation matrix $\boldsymbol{\rho}$ is also equal. For the diagonal elements, the following holds: $\rho_{kk} = 1$. Now let us consider the following cases.

$\rho_{ij} = \pm 1$. There is a perfect positive ($\Delta w_i > 0 \Rightarrow \Delta w_j > 0$) or negative ($\Delta w_i > 0 \Rightarrow \Delta w_j < 0$) correlation between both perturbations. In this case there is a locally linear relationship between both quantities.

$\rho_{ij} = 0$. Since $\Delta\mathbf{w}$ and $\tilde{\mathbf{w}}$ are multivariate normal distributed, this implies that Δw_i and Δw_j, as well as $\tilde{w}_i$ and $\tilde{w}_j$, are statistically independent (see [13, Theorem 2.10]).

$0 < |\rho_{ij}| < 1$. The larger $|\rho_{ij}|$ is, the stronger is the correlation between Δw_i and Δw_j.

To analyze the correlations between quantities, one can compute $\boldsymbol{\rho}$ for different states and generate histograms showing the distributions of the correlation coefficients ρ_{ij}. If, for example, $\rho_{ij} \approx -1$ for most of the analyzed states, then this is a hint that it is difficult to estimate quantities w_i and w_j simultaneously when performing state and parameter estimation based on experimentally measured data from a dynamical process. One can expect that the anticorrelation between w_i and w_j reflects in the estimated values of w_i and w_j. If w_i is estimated to a larger value than the (usually unknown) "true"' value in the dynamical process, then it is likely that w_j is estimated to a smaller value than the corresponding "true" value.

It is known (see Ref. [35]) that a large correlation coefficient $|\rho_{ij}| \approx 1$ implies an (almost) linear relationship between Δw_i and Δw_j, but a linear dependency between more than two perturbations does not necessarily lead to a correlation coefficient $|\rho_{ij}| \approx 1$. Therefore, we suggest to apply the dependency analysis from Sec. II A 2 first to detect locally dependent quantities and quantities which are, in principle, locally estimable. If $D\mathbf{G}$ is nonsingular ($D_{\mathrm{N}} = 0$), or if one fixes quantities in the analysis until every variation $\Delta\mathbf{w} \neq 0$ leads to $\Delta\mathbf{g} \neq 0$, then we suggest to apply the correlation analysis.

However, even in the case of a nonsingular $D\mathbf{G}$ a weak dependency might not be revealed by the correlation analysis [35]. Another way to quantify weak dependencies is the collinearity diagnostics suggested in Ref. [35].

III. STATE AND PARAMETER ESTIMATION ALGORITHM

The task of state and parameter estimation is to find a trajectory for the model variables $\{\mathbf{x}(n)\}$, $\mathbf{x}(n) = \mathbf{x}(t_n) \in \mathbb{R}^D$, with $t_n = n\Delta t$ and $n = 0, \ldots, N-1$ and a vector of model parameters in such a way that the trajectory fits an (here univariate) experimental data time series $\{\eta(n)\}$, $\eta(n) = \eta(t_n) \in \mathbb{R}$, with $t_n = n\Delta t$ and $n = 0, \ldots, N-1$ [via a measurement function Eq. (2)], on the one hand and fulfills the model equations on the other hand. The approach chosen here [37,38] is similar to weak constraint 4D-VAR [7,10,13,48,49] and is based on minimizing a cost function. Here models in the form of ODE Eq. (3) are used. Approximating the time derivative by a finite difference $\Delta\mathbf{x}(n)/\Delta t$ and introducing an error term $\mathbf{u}(n) = \mathbf{u}(t_n)$ in the model, the discretized equations read

$$\frac{\Delta\mathbf{x}(n)}{\Delta t} = F[\mathbf{x}(n), \mathbf{p}] + \mathbf{u}(n). \tag{35}$$

To estimate a trajectory $\{\mathbf{x}(n)\}$ and model parameters $\mathbf{p}$ the cost function

$$C(\{\mathbf{x}(n)\}, \mathbf{p}) = \sum_{n=0}^{N} \left\{ \frac{\alpha}{N}[\eta(n) - h(\mathbf{x}(n))]^2 + \frac{\alpha - 1}{N D}\mathbf{u}(n)^{\mathrm{tr}}\mathbf{B}\mathbf{u}(n) \right\} + C_3 + C_4 \tag{36}$$

has to be minimized with respect to the entire trajectory $\{\mathbf{x}(n)\}$ and the parameters $\mathbf{p}$. The trajectory and parameters which minimize $C(\{\mathbf{x}(n)\}, \mathbf{p})$ are then considered as the solution of the estimation problem. To estimate a smooth trajectory, a Hermite interpolation is performed by the additional sum C_3 and to force variables and parameters to stay in predefined boundaries the sum C_4 is required. Both terms are described in more detail in Ref. [37,38]. Furthermore, a homotopy parameter $\alpha \in (0,1)$ is used to cope with local minima in the cost function [12]. If $\alpha \approx 1$, the estimated trajectory is very close to the data [large modeling errors $\mathbf{u}(n)$ are allowed], and if $\alpha \approx 0$, the estimated trajectory fulfills the model equations very well, but might not match the data. Additionally, a matrix $\mathbf{B}$ is introduced for an individual weighting of the error $\mathbf{u}(n)$. In the weak constraint 4D-Var formulation $\mathbf{B}$ is the inverse covariance matrix of the modeling error $\mathbf{u}(n)$. In the examples we consider in Sec. IV, however, $\mathbf{B}$ is chosen to be a diagonal $D \times D$ matrix. The diagonal elements b_{ii} are set heuristically so that the estimated solutions for the model variables are smooth. We observed that too-small b_{ii} may lead to relatively large errors $u_i(n)$ and, therefore, to a nonsmooth estimated trajectory of x_i (at least in the case of noisy data). The exact forms of $\mathbf{B}$ used in the examples are provided in Sec. IV.

For optimization the Levenberg-Marquard [50–52] algorithm is used, where derivatives and sparsity structures are computed by means of the automatic differentiation software ADOL-C [53–55].

IV. EXAMPLES

The aim of the examples is to demonstrate how the dependency and correlation analysis of model states and parameters discussed in Secs. II A and II B can be applied.

In case of the Colpitts oscillator (Sec. IV A) the null space of $D\mathbf{G}$, Eq. (13), is one dimensional and both the dependency and correlation analysis are applied. The results are then compared to results obtained by applying the state and parameter estimation algorithm from Sec. III.

In Sec. IV B the dependency analysis from Sec. II A is applied to the Rössler model. A two-dimensional null space of $D\mathbf{G}$ was found. Furthermore, it is demonstrated how sets of dependent variables and parameters can be found. These results are then compared to results presented in Ref. [44], where relationships between model parameters were identified in polynomial vector fields.

In the third example the Hindmarsh-Rose model is studied. It was found that the null space of $D\mathbf{G}$ is zero dimensional. In theory, this means that all variables and parameters should be estimable. Applying the correlation analysis from Sec. II B, however, shows that there is still a strong correlation between certain model parameters. This correlation was also observed when the state and parameter estimation algorithm from Sec. III is applied.

A. Colpitts oscillator

In this section we use the model of the chaotic Colpitts oscillator [41] to investigate local dependencies of model parameters and variables. The system is given by $D = 3$ model equations

$$\begin{aligned} \dot{x}_1 &= p_1 x_2, \\ \dot{x}_2 &= -p_2(x_1 + x_3) - p_3 x_2, \\ \dot{x}_3 &= p_4(x_2 + 1 - e^{-x_1}), \end{aligned} \tag{37}$$

with the $N_\mathrm{p} = 4$ parameters $\mathbf{p} = (5,0.08,0.7,6.3)$ (equations and parameter values are taken from Ref. [56]) and the measurement function

$$h(\mathbf{x}) = x_1. \tag{38}$$

To generate states for the analysis the model Eq. (37) is integrated 10^6 steps with a step size of 0.1 (no transient included) using the parameters $\mathbf{p} = (5,0.08,0.7,6.3)$. Then, every 100th point from the trajectory is used for the analysis resulting in 10^4 states. The dimension of the delay coordinate vector, Eq. (5), is set to $K = 20$ so that with $K > 2(D + N_\mathrm{p}) + 1 = 15$ the sufficient condition from Ref. [33] is fulfilled and wrong statements regarding the estimability due to too-small K are not to be expected. Then the delay reconstruction map Eq. (5) used for the following analysis is given by

$$\begin{aligned} &\mathbf{G}[\mathbf{x}(t),\mathbf{p}] \\ &\quad = [x_1(t),x_1(t+\tau),\ldots,x_1[t+(K-1)\tau]]^{\mathrm{tr}}. \end{aligned} \tag{39}$$

Next, its $K \times D + N_\mathrm{p}$ Jacobian matrix $D\mathbf{G}$ is evaluated for the 10^4 states. How $D\mathbf{G}$ can be evaluated for a specific state and parameter vector is discussed in Sec. I.

First, we compute histograms for the singular values σ_i, $i = 1,\ldots,7$ of $D\mathbf{G}$, Eq. (11), including all four model parameters in the analysis; see Fig. 3. The histograms are computed in the following way. (i) The delay time τ is set to the smallest considered value $\tau = 0.1$. (ii) Then all normalized singular values σ_i/σ_1 ($i = 1,\ldots,7$) are computed for each of the 10^4 different states on the attractor using the fixed τ. (iii) A histogram for each σ_i/σ_1 is computed and plotted vertically (color coded) in the corresponding subplot. (iv) This process is repeated with a slightly increased τ. We can see that for all investigated states and independent on τ only the smallest

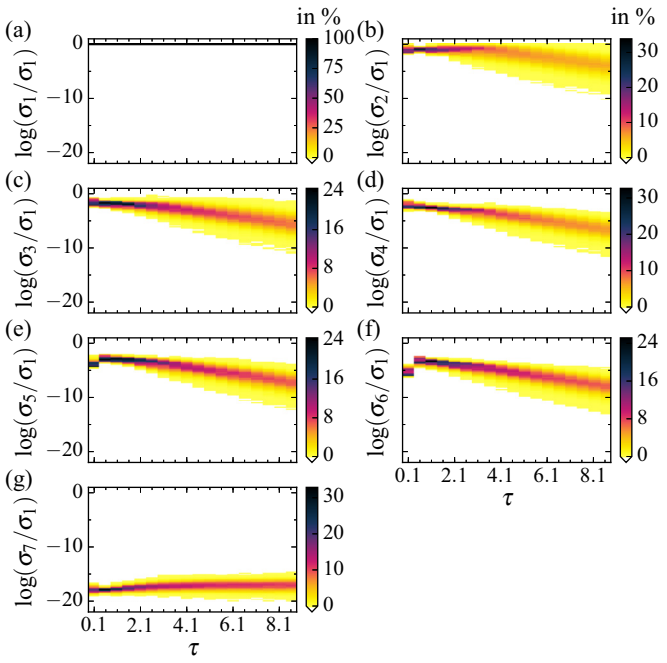

FIG. 3. Histograms (color coded) of the normalized singular values σ_i/σ_1, Eq. (11) (computed with delay reconstruction dimension $K = 20$; log denotes the logarithm with base 10), of $D\mathbf{G}$ computed with respect to all variables ($D = 3$) and parameters ($N_\mathrm{p} = 4$) of the Colpitts oscillator, Eq. (37), where x_1 is assumed to be measured, Eq. (38). For each τ the singular values σ_i are computed for 10^4 points on the attractor using $\mathbf{p} = (5,0.08,0.7,6.3)$. The smallest normalized singular values σ_7/σ_1 is of magnitude smaller than 10^{-15} for (almost) all states and τ and, hence, is numerically very close to zero, indicating a one-dimensional null space of $D\mathbf{G}$, Eq. (13).

normalized singular values σ_7/σ_1 are very close to zero. This indicates a one-dimensional null space ($D_\mathrm{N} = 1$) of $D\mathbf{G}$ and means that there exist $\Delta\mathbf{w} \neq 0$ with $\Delta\mathbf{g} = 0$; see Eqs. (13) and (14). Therefore, some of the model variables or parameters are locally not estimable because its variation does not lead to a perturbation $\Delta\mathbf{g} \neq 0$ of the measured signal $y(t)$ at present time t and delayed times.

Since $D_\mathrm{N} = 1$, the last column of $\mathbf{V}$ [see Eq. (10)], $\mathbf{v}_\mathrm{B} = \mathbf{v}^{(7)}$, is a basis vector of the null space of $D\mathbf{G}$ (see Sec. II A). To simplify the association of the components of $\mathbf{v}_\mathrm{B}$, the associated variable or parameter is mentioned in brackets. Since both $\mathbf{v}_\mathrm{B}$ and $-\mathbf{v}_\mathrm{B}$ are valid basis vectors (and both variants may be obtained by the SVD of $D\mathbf{G}$), $\mathbf{v}_\mathrm{B}$ is normalized by the sign of an arbitrary nonvanishing component $v_\mathrm{B}(p_i)$ associated with the parameter p_i. We have chosen $v_\mathrm{B}(p_4)$ and normalize $\mathbf{v}_\mathrm{B}$ via $\mathbf{v}_\mathrm{B} \to \mathrm{sgn}(v_\mathrm{B}(p_4)) \cdot \mathbf{v}_\mathrm{B}$.

To check whether some variables or parameters are locally estimable, we consider histograms of the components of $\mathbf{v}_\mathrm{B}$, $v_\mathrm{B}(w_i)$, for the same 10^4 states and delay times τ used previously; see Fig. 4. Only $|v_\mathrm{B}(x_1)| < 10^{-10}$ and $|v_\mathrm{B}(p_3)| < 10^{-10}$ are close to zero for all 10^4 states and all considered τ. This means that it is (almost) impossible to vary x_1 and p_3 without changing $\mathbf{g}$. Hence, x_1 and p_3 are as locally estimable. All other quantities are locally redundant.

Local variations of the redundant quantities keeping $\mathbf{g}$ unchanged are only simultaneously possible; see Eq. (18). To

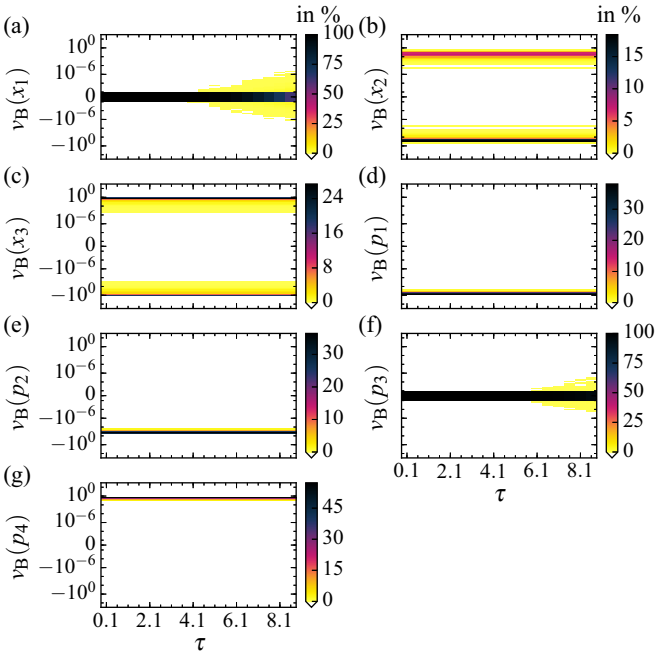

FIG. 4. Histograms (color coded) of the components $v_B(w_i)$ of a basis vector spanning the one-dimensional (see Fig. 3) null space of $D\mathbf{G}$ (computed with delay reconstruction dimension $K = 20$) computed with respect to all variables ($D = 3$) and parameters ($N_p = 4$) of the Colpitts oscillator, Eq. (37), where x_1 is assumed to be measured, Eq. (38). For each τ the components $v_B(w_i)$ are computed for 10^4 points from the attractor using $\mathbf{p} = (5, 0.08, 0.7, 6.3)$. Only $v_B(x_1)$ and $v_B(p_3)$ are within the interval $[-10^{-10}, 10^{-10}]$ for (almost) all states and delay times τ and numerically very close to zero, indicating that only x_1 and p_3 are locally estimable.

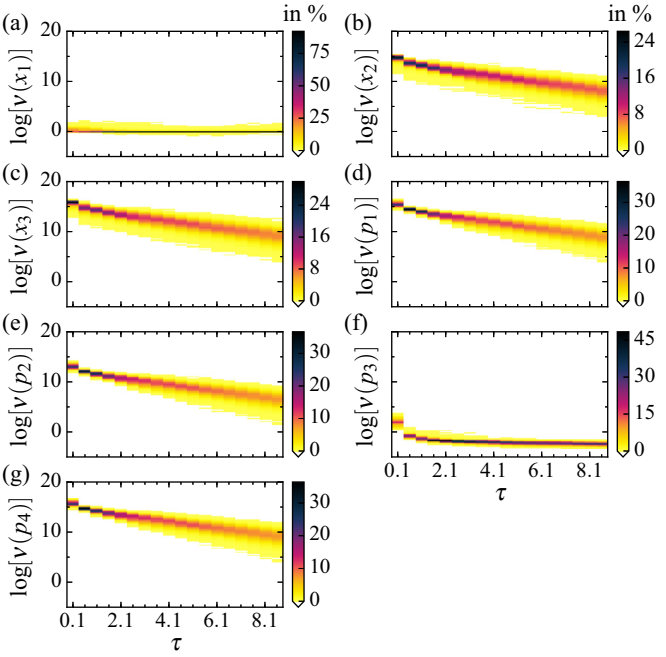

FIG. 5. Histograms (color coded) of the measure of uncertainty $v(w_i)$, Eq. (33) (computed with delay reconstruction dimension $K = 20$; log denotes the logarithm with base 10), of all variables ($D = 3$) and parameters $N_p = 4$ of the Colpitts oscillator where x_1 is assumed to be measured, Eq. (38). For each delay time τ the measure of uncertainty $v(w_i)$ [see Eq. (33)] is computed for 10^4 points on the attractor using $\mathbf{p} = (5, 0.08, 0.7, 6.3)$. On average, $v(w_i)$ is quite large for x_2, x_3, p_1, p_2, and p_4, and much smaller for x_1 and p_3, indicating a good local estimability only for x_1 and p_3.

find out how these quantities pairwise depend on each other [see Eq. (26)], we consider the ratio of the signs of $v_B(w_i)$ read off from Fig. 4. The histograms associated with the model parameters show that there the signs of $v_B(w_i)$ are independent on the state and the delay time τ. According to Eq. (26) we see the following. (i) A positive variation of p_4, $\Delta p_4 > 0$, leads to a negative variation of p_1 and p_2, $\Delta p_1, \Delta p_2 < 0$, because of $v_B(p_4)/v_B(p_1) < 0$ and $v_B(p_4)/v_B(p_2) < 0$ and vice versa. (ii) A variation $\Delta p_4 \neq 0$ leads to no change of p_3.

To confirm that only x_1 and p_3 are locally estimable, the measure of uncertainty $v(w_i)$, Eq. (33), was computed for the same 10^4 points from the attractor and delay times τ as previously with $K = 20$. First, $v(w_i)$ was computed, including all four model parameters in the analysis. Histograms for all $v(w_i)$ are shown in Fig. 5.

One can see that, on average, $v(w_i)$ is very large (magnitude 10^{10} to 10^{15}) for all i, except $v(x_1)$ and $v(p_3)$. This indicates a high uncertainty for x_2, x_3, p_1, p_2, and p_4. However, estimating x_1 and p_3 should give accurate results due to much smaller $v(w_i)$ (on average). This coincides with the results from investigating the null space of $D\mathbf{G}$, which also indicated a good local estimability of x_1 and p_3 since variations of these quantities lead to variations of $\mathbf{g}$.

To check this result a twin experiment is performed where all model parameters are estimated beside the model variables from a x_1 time series using the estimation method from Sec. III. To obtain the data time series, the model Eq. (37) was integrated and the true solution $\mathbf{z}(t)$ is used to generate the time series $\{\eta(t_n)\}$ ($t_n = n \cdot 0.01$, $n = 0, \ldots, 6000$) with

$$\eta(t_n) = z_1(t_n). \tag{40}$$

The corresponding measurement function, Eq. (38), is then used for the estimation of all model variables and all parameters by minimizing the cost function, Eq. (36). The weighting matrix $\mathbf{B}$ is chosen to be a 3×3 identity matrix. Figure 6 shows the estimated solution. In panel (a) one can see that the output of $h(\mathbf{x})$ (blue line) perfectly matches the data (light green line). In panels (b), (d), and (e) one can see that the estimated solution of the corresponding model variable x_i (blue line) matches the true solution z_i (red dashed line) only for the first variable x_1. Furthermore, one can see in panel (c), a magnified version of panel (b), that the magnitude of the estimated solution of x_2 is much smaller than the true solution z_i. The model parameters are estimated to $\mathbf{p} = (5573.7,\ 1.62 \times 10^{-4},\ 0.700,\ 2.79)$. Only p_3 was estimated to the correct value used to generate the data. These results were correctly predicted by the previously computed measure of uncertainty [$v(p_3)$ is much smaller compared to the uncertainties of the other parameters] (Fig. 5) and the components $v_B(w_i)$ of a basis vector of the null space of $D\mathbf{G}$ (Fig. 4). The component $v_B(p_3)$ is very close to zero, indicating that a variation of p_3 is (almost) impossible without a change of $\mathbf{g}$.

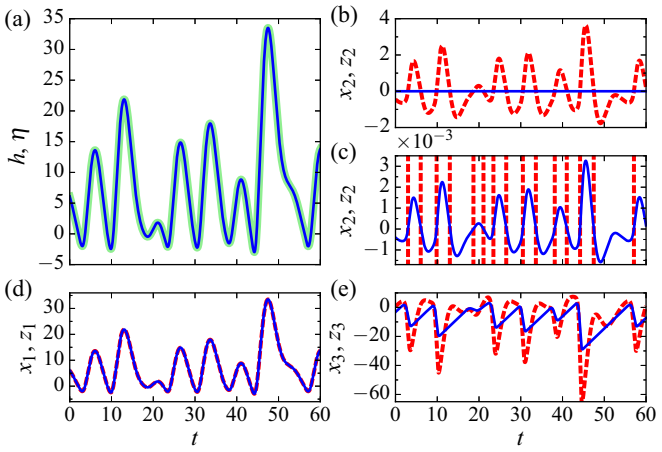

FIG. 6. State and parameter estimation of the Colpitts oscillator Eq. (37) via measurement function $h(\mathbf{x}(t))$, Eq. (38), from the data time series $\{\eta(t_n)\}$, Eq. (40). (a) Output of $h(\mathbf{x}(t))$ (blue line) matches the data $\{\eta(t_n)\}$ (light green line). (b),(d),(e) Estimated model variables x_i (blue line) and the true solutions z_i, $i = 1,2,3$, (red dashed line) used to generate $\{\eta(t_n)\}$. (c) Zoomed version of (b). Only x_1 matches its true solution z_1. Parameters are estimated to $\mathbf{p} = (5573.7, 1.62 \cdot 10^{-4}, 0.700, 2.79)$. That is, only p_3 was estimated to the value used to generate the data.

There exists one set of locally dependent quantities (since $D_N = 1$) consisting of x_2, x_3, p_1, p_2, and p_4. Prohibiting variations of one of these quantities via $\Delta w_i = 0$ in Eq. (18), for example by setting w_i to a fixed value instead of estimating it in an estimation problem, sets $a = 0$. As a result, variations of all other quantities are not possible within the null space of $D\mathbf{G}$, which means that all other quantities became locally estimable. For example, fixing p_3 does not set a to zero in Eq. (18) because, according to Fig. 4, $v_B(p_3)$ is (very close to) zero. Therefore, fixing p_3 would not make all other quantities locally estimable.

We decided to remove p_4 from the analysis by setting it to the true value $p_4 = 6.3$. One can verify that the null space of $D\mathbf{G}$, with the column associated with p_4 removed, is zero dimensional by considering the singular values of $D\mathbf{G}$ (not shown here). Then we considered the correlation coefficients $\rho(w_i, w_j)$, Eq. (34), which describes local correlations between quantities. The correlation analysis from Sec. II B was performed for all model variables x_1, x_2, and x_3 and the model parameters p_1, p_2, and p_3. The computation of the correlation coefficients is based on the same 10^4 states from the Colpitts attractor and the measurement function Eq. (38), as previously. Figure 7 shows the histograms of all correlation coefficients (without the diagonal elements of $\boldsymbol{\rho}$, because they are 1 anyway) for different τ and a reconstruction dimension of $K = 20$. The plots (a) to (o) show histograms of $\rho(w_i, w_j)$ for different quantities w_i and w_j using different delay times τ. In all plots the correlation coefficients are, independent of τ, distributed over the whole interval $[-1, 1]$. One can see that no perfect correlations occur because there is no histogram with $\rho(w_i, w_j) \approx \pm 1$ for all investigated states and delay times τ. Therefore, state and parameter estimation should give accurate results.

In a twin experiment in Ref. [37] the parameters p_1, p_2, and p_3 of the Colpitts oscillator were estimated from a noisy time series of the first model variable x_1. All estimated parameter values coincide with the values used to generate the data and are hence estimated correctly. This also coincides with results obtained when computing the measure of uncertainty only with the first three, instead of all four, model parameters. To obtain this result the previous computation of $v(w_i)$ was repeated including only x_1, x_2, x_3, p_1, p_2, and p_3 in the analysis. Histograms of $v(w_i)$ are shown in Fig. 8. We can see that the uncertainties are of magnitude 1 (for larger τ) and, hence, much smaller than in Fig. 5 (where all four parameters are included in the analysis).

To confirm that the set of locally dependent quantities is correctly predicted by the dependency analysis, the twin experiment from the beginning of this section is repeated. There, the model of the Colpitts oscillator, Eq. (37), was adapted to a x_1 time series (no measurement noise) and all four parameters were estimated. Instead of estimating all four parameters simultaneously, we adapt the concept of a

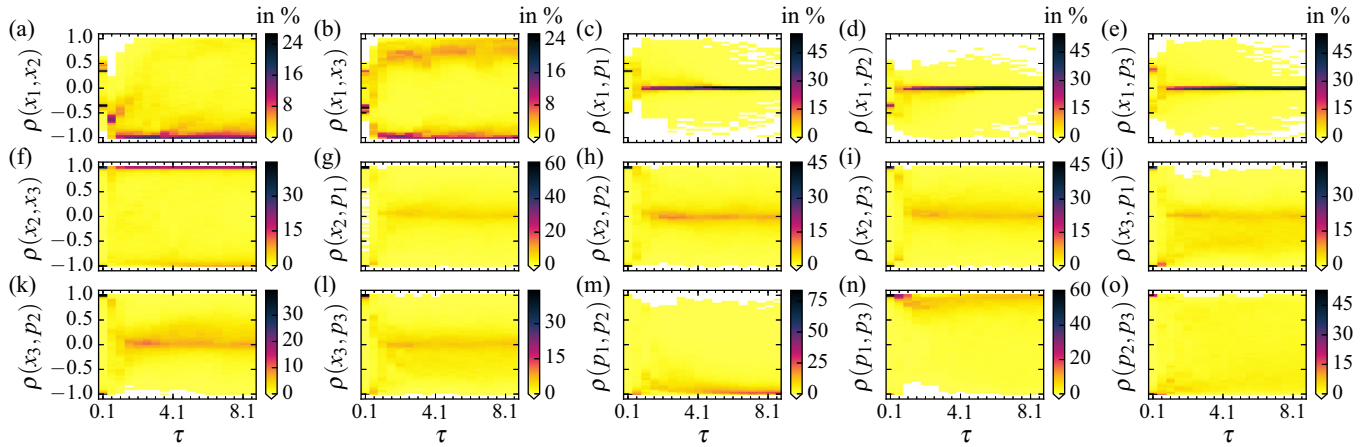

FIG. 7. Histograms (color coded) of the correlation coefficients $\rho(w_i, w_j)$, Eq. (34) (computed with reconstruction dimension of $K = 20$), of all variables and the parameters p_1, p_2, and p_3 of the Colpitts oscillator, Eq. (37), where x_1 is assumed to be measured, Eq. (38). For each delay time τ the coefficients $\rho(w_i, w_j)$ are computed for 10^4 points on the attractor. In all plots it is $-1 < \rho(w_i, w_j) < 1$ for many states and τ indicating not very strong local dependencies between the quantities.

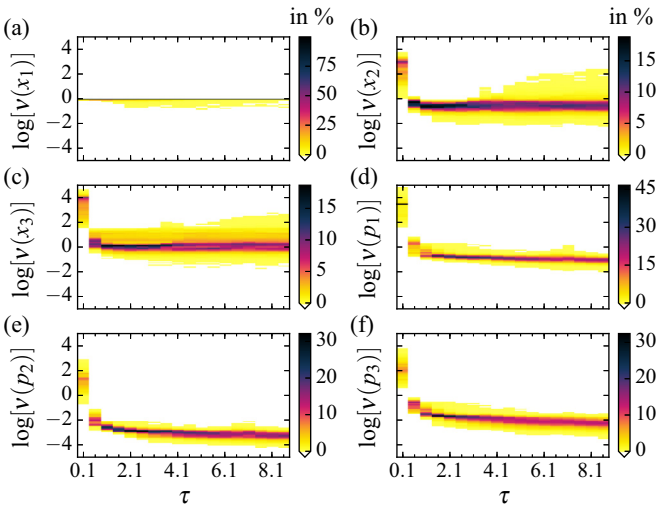

FIG. 8. Similar to Fig. 5, but the measure of uncertainty, $\nu(w_i)$, was only computed for x_1, x_2, x_3, p_1, p_2, and p_3 (without p_4). For all quantities $\nu(w_i)$ is very small (on average), indicating that states and parameters can be estimated correctly. log denotes the logarithm with base 10.

profile likelihood [39,40]. Only the first three parameters are estimated and p_4 is set to different values. The dependency of the estimated values of p_1, p_2, and p_3 on the parameter p_4 is shown in Fig. 9. In more detail, the simulation was performed in the following steps. (i) p_4 is fixed to $p_4 = 4.0$. (ii) Beside the model variables, the parameters p_1, p_2, and p_3 are estimated, as shown (blue dots) in plots (a), (b), and (c). The corresponding value of the cost function C, Eq. (36), for the estimated solution is shown in panel (d). (iii) p_4 is slightly increased and the

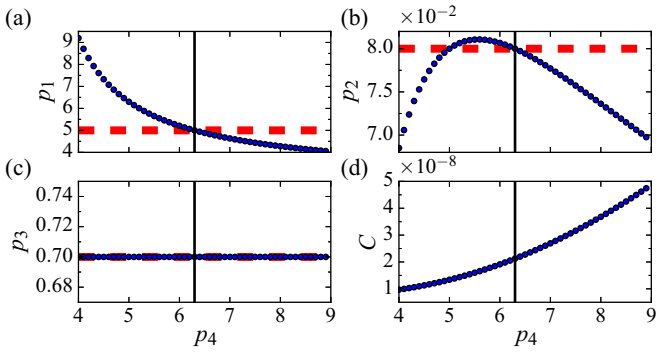

FIG. 9. State and parameter estimation (using the algorithm from Sec. III) of the Colpitts oscillator, Eq. (37), from a clean x_1 time series, Eq. (38). (a)–(c) To generate the data, the parameter values p_1, p_2, p_3, indicated by the red dashed lines, and $p_4 = 6.3$ [black vertical line in (a)–(d)] were used. For each fixed p_4 the other parameters (blue dots) and the model variables were estimated (concept of profile likelihood [39,40]). (a),(b) The local dependency of p_1 and p_2 on p_4 around $p_4 = 6.3$ is consistent with the dependency analysis (see Fig. 4) since $\Delta p_4 > 0 \Rightarrow \Delta p_1, \Delta p_2 < 0$ and vice versa. (c) p_3 is estimated correctly independent of p_4 (consistent with the dependency analysis). (d) Cost function C, Eq. (36), is very small at the estimated solutions, indicating a relatively flat valley in the minimum.

estimation is repeated. The red dashed lines in panels (a)–(c) show the value of the corresponding parameter used to generate the data. In plots (a)–(d) the vertical black line at $p_4 = 6.3$ shows the value of p_4 used to generate the data.

First of all, one can see that for all values of p_4 the cost at the estimated solution is very small (order of magnitude of 10^{-8}) so that it is likely that the estimated solutions are in a global minimum of the cost function. A possible reason why C is not exactly constant for different p_4 might be truncation errors. Further, for $p_4 = 6.3$ the estimated values of all other parameters coincide with the values used to generate the data. If p_4 is fixed to a value slightly smaller than 6.3 (for example $p_4 = 5.5$), then the estimated values for p_1 and p_2 are larger than the values used to generate the data (therefore $\Delta p_1, \Delta p_2 > 0$), as shown in Figs. 9(a) and 9(b), respectively. If now p_4 is fixed to a value slightly larger than 6.3 (for example, $p_4 = 7.0$), then p_1 and p_2 are estimated to too-small values. In Fig. 9(b) one can see that the dependencies are only locally correct (around $p_4 = 6.3$) because for $p_4 < 5$ the parameter p_2 is estimated to too-small values. Furthermore, p_3 is independent of p_4.

These results coincide with the local dependencies found in the dependency analysis when considering the ratio of the components of the basis vectors (see Fig. 4).

B. Rössler model

In this section we use the $D = 3$ dimensional Rössler model [42] in a form presented in Refs. [43,44]. The model equations read

$$\begin{aligned} \dot{x}_1 &= p_1 x_2 + p_2 x_3, \\ \dot{x}_2 &= p_3 x_1 + p_4 x_2, \\ \dot{x}_3 &= p_5 + p_6 x_3 + p_7 x_1 x_3, \end{aligned} \tag{41}$$

where in the following the $N_p = 7$ dimensional parameter vector $\mathbf{p} = (-1, -1, 1, 0.1, 0.1, -14, 1)$ is used. To compare our results with Ref. [44], we also assume that x_2 is measured,

$$h(\mathbf{x}) = x_2. \tag{42}$$

To investigate the estimability, we first integrate the Rössler model, Eq. (41), for 10^4 time steps with a step size of 0.1. Then, every 10th state (1000 different states in total) is used for the following analysis. For each state and for different delay times τ the Jacobian $D\mathbf{G}$, Eq. (10), and its SVD of the $K = 50$ dimensional delay reconstruction map Eq. (5) were computed with respect to all three model states and all seven model parameters. Therefore, $D\mathbf{G}$ is a 50×10 matrix. Note, that $K > 2(D + N_p) + 1 = 21$ and therefore the condition of Ref. [33] is fulfilled, as discussed in Sec. I.

Histograms of the normalized singular values σ_i/σ_1, $i = 1, \ldots, 10$, Eq. (11), of $D\mathbf{G}$ are shown in Fig. 10. We can see that the smallest normalized singular values, σ_9/σ_1 and σ_{10}/σ_1 are, for almost all states and τ, smaller than 10^{-15} and, hence, are very close to zero. This means that the null space of $D\mathbf{G}$ is $D_N = 2$ dimensional for all investigated states and all τ, and it can be spanned by the two basis vectors $\mathbf{v}_B^{(1)} = \mathbf{v}^{(9)}$ and $\mathbf{v}_B^{(2)} = \mathbf{v}^{(10)}$ (two last columns of $\mathbf{V}$); see Sec. II A. Both vectors are orthogonal and constitute the 10×2 basis matrix $\mathbf{V}_B = [\mathbf{v}_B^{(1)}, \mathbf{v}_B^{(2)}]$; see Eq. (20). If the ith row of $\mathbf{V}_B$ contains

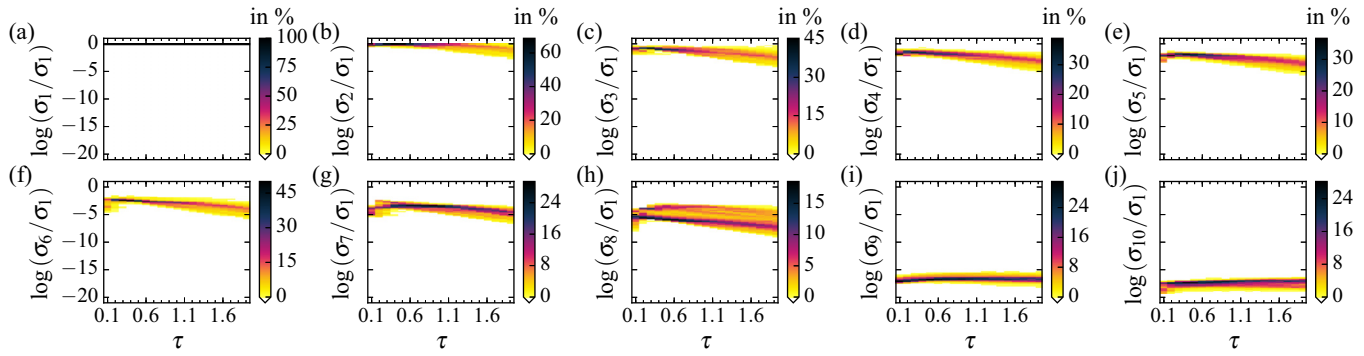

FIG. 10. Histograms (color coded) of the normalized singular values σ_i/σ_1, Eq. (11) (computed with delay reconstruction dimension $K = 50$; log denotes the logarithm with base 10), using all variables ($D = 3$) and parameters ($N_p = 7$) of the Rössler-model, Eq. (41), where x_2 is assumed to be measured, Eq. (42). For each delay time τ the normalized singular values σ_i/σ_1 of $D\mathbf{G}$ are computed for 1000 states from the attractor. The two smallest normalized singular values σ_9/σ_1 and σ_{10}/σ_1 are of magnitude smaller than 10^{-15} for (almost) all states and τ and, hence, are numerically very close to zero. This indicates a two-dimensional null space of $D\mathbf{G}$ and, therefore, the existence of redundant variables or parameters.

zeros only, then the quantity w_i is locally estimable because a perturbation of this quantity is not possible inside the null space of $D\mathbf{G}$ and, therefore, would lead to a perturbation $\Delta\mathbf{g} \neq 0$; see Eq. (13).

In this example, all quantities are ordered by $\mathbf{w} = [x_1, x_2, x_3, p_1, p_2, p_3, p_4, p_5, p_6, p_7]$. Hence, the quantity w_i is associated with the components $v_{\mathrm{B},i}^{(1)}$ and $v_{\mathrm{B},i}^{(2)}$ of the basis vectors. For example, $v_{\mathrm{B},5}^{(1)}$ and $v_{\mathrm{B},5}^{(2)}$ are associated with p_2. To simplify the notation in the following discussion, the associated quantities will be provided in brackets, e.g., $v_{\mathrm{B},5}^{(1)} = v_{\mathrm{B}}^{(1)}(p_2)$ and $v_{\mathrm{B},5}^{(2)} = v_{\mathrm{B}}^{(2)}(p_2)$. Then, the basis matrix is

$$\mathbf{V}_{\mathrm{B}} = \begin{pmatrix} v_{\mathrm{B}}^{(1)}(x_1) & v_{\mathrm{B}}^{(2)}(x_1) \\ v_{\mathrm{B}}^{(1)}(x_2) & v_{\mathrm{B}}^{(2)}(x_2) \\ v_{\mathrm{B}}^{(1)}(x_3) & v_{\mathrm{B}}^{(2)}(x_3) \\ v_{\mathrm{B}}^{(1)}(p_1) & v_{\mathrm{B}}^{(2)}(p_1) \\ v_{\mathrm{B}}^{(1)}(p_2) & v_{\mathrm{B}}^{(2)}(p_2) \\ v_{\mathrm{B}}^{(1)}(p_3) & v_{\mathrm{B}}^{(2)}(p_3) \\ v_{\mathrm{B}}^{(1)}(p_4) & v_{\mathrm{B}}^{(2)}(p_4) \\ v_{\mathrm{B}}^{(1)}(p_5) & v_{\mathrm{B}}^{(2)}(p_5) \\ v_{\mathrm{B}}^{(1)}(p_6) & v_{\mathrm{B}}^{(2)}(p_6) \\ v_{\mathrm{B}}^{(1)}(p_7) & v_{\mathrm{B}}^{(2)}(p_7) \end{pmatrix}. \tag{43}$$

To identify quantities where prohibiting its variation, for example, fixing them in an estimation problem, makes all quantities locally estimable, we apply the dependency analysis suggested in Sec. II A. In general, for multidimensional null spaces ($D_N > 1$) basis vectors obtained via SVD contain zero elements only for estimable quantities, but not the maximal number of vanishing components that could be achieved with a suitable basis as illustrated in Fig. 2. Furthermore, due to their (rotational) degree of freedom, these basis vectors are not unique such that results obtained at different states are difficult to compare. Therefore, we construct a new basis of the null space of $D\mathbf{G}$ (as introduced in Sec. II A 1) where all basis vectors will have mutually different patterns of nonzero elements. In the first step we have to choose two arbitrary quantities (two, because here we have a two-dimensional null space of $D\mathbf{G}$) so that $\mathbf{V}_{\mathrm{B},1}$, Eq. (44), will be nonsingular and its inverse exists so that Eq. (22) can be evaluated. In this example we have chosen the two quantities p_5 and p_7 and split $\mathbf{V}_{\mathrm{B}}$ into

$$\mathbf{V}_{\mathrm{B},1} = \begin{pmatrix} v_{\mathrm{B}}^{(1)}(p_5) & v_{\mathrm{B}}^{(2)}(p_5) \\ v_{\mathrm{B}}^{(1)}(p_7) & v_{\mathrm{B}}^{(2)}(p_7) \end{pmatrix} \tag{44}$$

and

$$\mathbf{V}_{\mathrm{B},2} = \begin{pmatrix} v_{\mathrm{B}}^{(1)}(x_1) & v_{\mathrm{B}}^{(2)}(x_1) \\ v_{\mathrm{B}}^{(1)}(x_2) & v_{\mathrm{B}}^{(2)}(x_2) \\ v_{\mathrm{B}}^{(1)}(x_3) & v_{\mathrm{B}}^{(2)}(x_3) \\ v_{\mathrm{B}}^{(1)}(p_1) & v_{\mathrm{B}}^{(2)}(p_1) \\ v_{\mathrm{B}}^{(1)}(p_2) & v_{\mathrm{B}}^{(2)}(p_2) \\ v_{\mathrm{B}}^{(1)}(p_3) & v_{\mathrm{B}}^{(2)}(p_3) \\ v_{\mathrm{B}}^{(1)}(p_4) & v_{\mathrm{B}}^{(2)}(p_4) \\ v_{\mathrm{B}}^{(1)}(p_6) & v_{\mathrm{B}}^{(2)}(p_6) \end{pmatrix}. \tag{45}$$

Next we split $D\mathbf{G}$ into the 50×2 matrix $D\mathbf{G}_1 = [D\mathbf{G}^{(8)}, D\mathbf{G}^{(10)}]$ containing the eighth and the tenth column of $D\mathbf{G}$ (the first column is associated with p_5 and the second column is associated with p_7) and the 50×8 matrix $\mathrm{D}\mathbf{G}_2$ containing the remaining columns of $D\mathbf{G}$.

Histograms of both normalized singular values σ_i/σ_1 of $\mathbf{V}_{\mathrm{B},1}$, computed based on the same states and delay times τ as previously, are shown in Fig. 11. Since the smallest singular value is relatively large (larger than 10^{-2}) for all states and τ, $\mathbf{V}_{\mathrm{b},1}^{-1}$ exists and is unique.

Therefore, we can compute $\widetilde{\mathbf{V}}_{\mathrm{B}}'$ in Eq. (22) for each considered state and τ. In the next step we reorder the columns of $[D\mathbf{G}_1, D\mathbf{G}_2]$ in a way such that we obtain $D\mathbf{G}$ again.

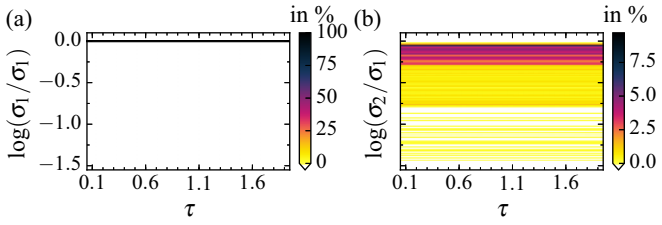

FIG. 11. Histograms (color coded) of the normalized singular values σ_i/σ_1 of the square matrix $\mathbf{V}_{\mathrm{B},1}$, Eq. (44), using all variables ($D = 3$) and parameters ($N_\mathrm{p} = 7$) of the Rössler model, Eq. (41) (based on the same states, delay times τ, etc., as in Fig. 10; log denotes the logarithm with base 10). Since for all states and τ the smallest normalized singular values σ_2/σ_1 are relatively large, $\mathbf{V}_{\mathrm{B},1}$ is not singular and has, therefore, a unique inverse.

Performing the same reordering on the rows of $\widetilde{\mathbf{V}}'_\mathrm{B}$ yields

$$\widetilde{\mathbf{V}}_\mathrm{B} = \begin{pmatrix} \widetilde{v}_\mathrm{B}^{(1)}(x_1) & \widetilde{v}_\mathrm{B}^{(2)}(x_1) \\ \widetilde{v}_\mathrm{B}^{(1)}(x_2) & \widetilde{v}_\mathrm{B}^{(2)}(x_2) \\ \widetilde{v}_\mathrm{B}^{(1)}(x_3) & \widetilde{v}_\mathrm{B}^{(2)}(x_3) \\ \widetilde{v}_\mathrm{B}^{(1)}(p_1) & \widetilde{v}_\mathrm{B}^{(2)}(p_1) \\ \widetilde{v}_\mathrm{B}^{(1)}(p_2) & \widetilde{v}_\mathrm{B}^{(2)}(p_2) \\ \widetilde{v}_\mathrm{B}^{(1)}(p_3) & \widetilde{v}_\mathrm{B}^{(2)}(p_3) \\ \widetilde{v}_\mathrm{B}^{(1)}(p_4) & \widetilde{v}_\mathrm{B}^{(2)}(p_4) \\ 1 & 0 \\ \widetilde{v}_\mathrm{B}^{(1)}(p_6) & \widetilde{v}_\mathrm{B}^{(2)}(p_6) \\ 0 & 1 \end{pmatrix} = \begin{bmatrix} \widetilde{\mathbf{v}}_\mathrm{B}^{(1)} & \widetilde{\mathbf{v}}_\mathrm{B}^{(2)} \end{bmatrix} \tag{46}$$

and fulfills Eq. (23). The vectors $\widetilde{\mathbf{v}}_\mathrm{B}^{(1)}$ and $\widetilde{\mathbf{v}}_\mathrm{B}^{(2)}$ contain the components of the first and the second column of $\widetilde{\mathbf{V}}_\mathrm{B}$, respectively. Due to the ones and zeros in $\widetilde{\mathbf{V}}_\mathrm{B}$ both vectors are linear independent and, since the null space of $D\mathbf{G}$ is $D_\mathrm{N} = 2$, form a basis of $D\mathbf{G}$ with different patterns of nonzero elements. In this basis it is $\widetilde{v}_\mathrm{B}^{(1)}(p_5) = \widetilde{v}_\mathrm{B}^{(2)}(p_7) = 1$ and $\widetilde{v}_\mathrm{B}^{(2)}(p_5) = \widetilde{v}_\mathrm{B}^{(1)}(p_7) = 0$.

To further investigate the structure of the null space of $D\mathbf{G}$, we consider histograms of all components of $\widetilde{\mathbf{v}}_\mathrm{B}^{(1)}$ (Fig. 12) and $\widetilde{\mathbf{v}}_\mathrm{B}^{(2)}$ (Fig. 13) computed using the same states and delay times τ as previously. Every variation $\Delta\mathbf{w}$ within the null space of $D\mathbf{G}$ can be expressed as

$$\Delta\mathbf{w} = \widetilde{a}_1\widetilde{\mathbf{v}}_\mathrm{B}^{(1)} + \widetilde{a}_2\widetilde{\mathbf{v}}_\mathrm{B}^{(2)}; \tag{47}$$

see Eq. (24). $\Delta\mathbf{w}$ is thus controlled by two scalars $\widetilde{a}_1$ and $\widetilde{a}_2$. By reading off the patterns of zero and nonzero elements of $\widetilde{\mathbf{v}}_\mathrm{B}^{(1)}$ and $\widetilde{\mathbf{v}}_\mathrm{B}^{(2)}$ from Figs. 12 and 13, the structure of Eq. (47) is

$$\Delta\mathbf{w} = \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \\ \Delta p_1 \\ \Delta p_2 \\ \Delta p_3 \\ \Delta p_4 \\ \Delta p_5 \\ \Delta p_6 \\ \Delta p_7 \end{pmatrix} = \widetilde{a}_1 \begin{pmatrix} 0 \\ 0 \\ * \\ 0 \\ * \\ 0 \\ 0 \\ * \\ 0 \\ 0 \end{pmatrix} + \widetilde{a}_2 \begin{pmatrix} * \\ 0 \\ 0 \\ * \\ * \\ * \\ 0 \\ 0 \\ 0 \\ * \end{pmatrix}. \tag{48}$$

A "0" means that this component is zero and "$*$" means that this component is nonzero for (almost) all investigated states and τ.

One can read off from Eq. (48) that variations $\Delta x_2, \Delta p_4, \Delta p_6 \neq 0$ are not possible without leaving the null space of $D\mathbf{G}$. Therefore, variations of x_2, p_4, and p_6 lead to a perturbation $\Delta\mathbf{g} \neq 0$ and, thereby, to a perturbation of $\mathbf{g}$ in Eq. (5), that is, of the output of the system at present and delayed times. This means that these quantities are locally estimable.

One can also read off from Eq. (48) how all other unknown quantities can be made locally estimable by prohibiting the variation of only two quantities. In a state and parameter estimation problem these two quantities would then be set to fixed values and not be estimated. The goal is to prohibit variations of quantities until $\widetilde{a}_1 = \widetilde{a}_2 = 0$. To accomplish that, in this example there exist several ways. For examples, fixing p_1 and p_5 by $\Delta p_1 = \Delta p_5 = 0$ implies $\widetilde{a}_1 = \widetilde{a}_2 = 0$ because $\widetilde{v}_\mathrm{B}^{(1)}(p_5)$ and $\widetilde{v}_\mathrm{B}^{(2)}(p_1)$ are nonzero. Or by setting $\Delta p_3 = 0$, which sets $\widetilde{a}_2 = 0$, and $\Delta p_2 = 0$, which then sets $\widetilde{a}_1 = 0$.

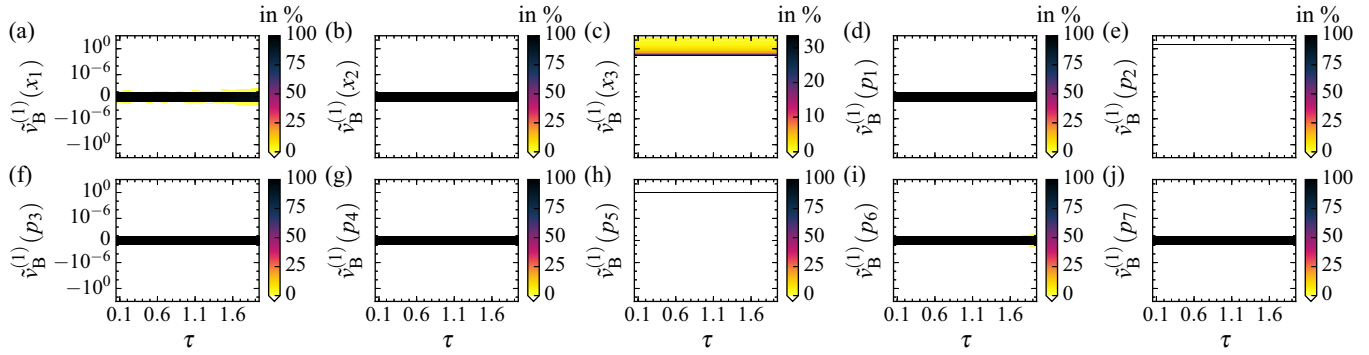

FIG. 12. Histograms (color coded) of the components of the first basis vector $\widetilde{\mathbf{v}}_\mathrm{B}^{(1)}$ (logarithm with base 10), Eq. (46), of the two-dimensional null space of $D\mathbf{G}$ using all variables ($D = 3$) and parameters ($N_\mathrm{p} = 7$) of the Rössler model, Eq. (41) (based on the same states, delay times τ, etc., as in Fig. 10). (c),(e),(h) Only the components associated with x_3, p_2, and p_5 are nonzero for all states and τ. (a),(b),(d),(f),(g),(i),(j) All other components are within the interval $[-10^{-10}, 10^{-10}]$ for (almost) all states and delay times τ and, hence, numerically very close to zero.

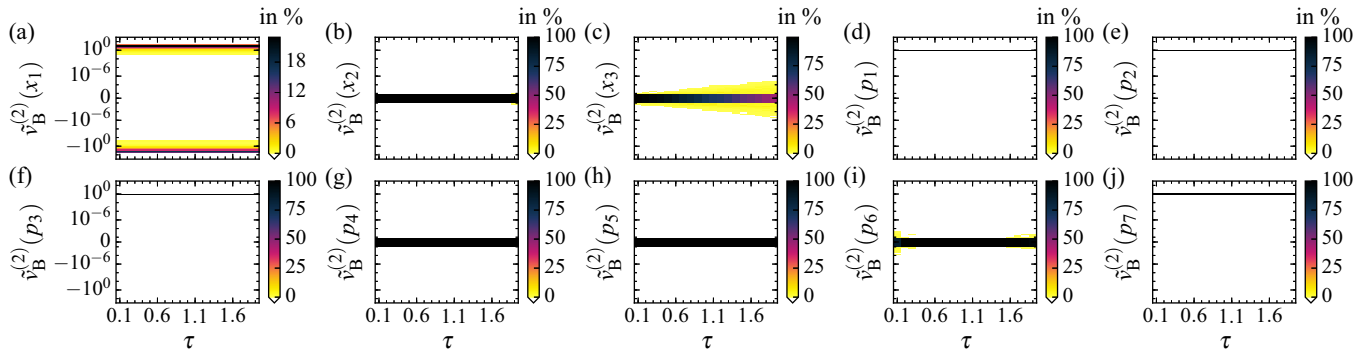

FIG. 13. As Fig 12, but histograms of the components of $\widetilde{\mathbf{v}}_B^{(2)}$, Eq. (46), are shown. (a),(c),(d),(e),(f),(j) Only the components associated with x_1, x_3, p_1, p_2, p_3, and p_7 are nonzero for all states and τ. (b),(c),(g),(h),(i) All other components are within the interval $[-10^{-10},10^{-10}]$ and, hence, numerically very close to zero for (almost) all states and delay times τ.

This shows the advantage of considering the basis vectors $\widetilde{\mathbf{v}}_B^{(1)}$ and $\widetilde{\mathbf{v}}_B^{(1)}$ instead of $\mathbf{v}_B^{(1)}$ and $\mathbf{v}_B^{(2)}$ (directly obtained via SVD). The basis vectors $\widetilde{\mathbf{v}}_B^{(1)}$ and $\widetilde{\mathbf{v}}_B^{(1)}$ contain a maximal number of zero elements and their pattern of zeros is unique, which makes it possible to consider histograms of particular elements obtained from different states of an attractor.

In Refs. [43,44] the authors also assumed that x_2 is measured and they used an analytical calculation based on derivative coordinates to show that Eq. (41) can be rewritten as

$$\begin{aligned}\dot{x}_1 &= \frac{1}{b}p_1x_2 + \frac{1}{ab}p_2x_3,\\ \dot{x}_2 &= bp_3x_1 + p_4x_2,\\ \dot{x}_3 &= ap_5 + p_6x_3 + bp_7x_1x_3,\end{aligned} \tag{49}$$

with the (introduced) scaling parameters $a,b \in \mathbb{R}$. This model can produce the same dynamics of the second model variable for arbitrary a and b. Therefore, a and b have a role comparable to that of $\widetilde{a}_1$ and $\widetilde{a}_2$ in Eqs. (47) and (48). It is only possible to uniquely recover the parameters p_4 and p_6 from a x_2 time series (only these parameters have no scaling factors). This coincides with our results from the dependency analysis that the only locally estimable parameters are p_4 and p_6.

With the scaling parameter a in Eq. (49) the same model parameters can be varied as with $\widetilde{a}_1$ in Eq. (48). The same holds for b in Eq. (49) and $\widetilde{a}_2$ in Eq. (48).

Furthermore, the way that parameters pairwise locally depend on each other is predicted correctly by the dependency analysis when using the same parameter values also used in the model Eq. (41) for the previous analysis, $\mathbf{p} = (-1, \quad -1, \quad 1, \quad 0.1, \quad 0.1, \quad -14, \quad 1)$. Let us start with $a = b = 1$ and slightly change $a \mapsto a > 1$. This leads to the following variations: $p_5 \mapsto ap_5 > p_5$ because $p_5 > 0$, and $p_2 \mapsto p_2/a > p_2$ because $p_2 < 0$. Hence, an increase of p_5 goes along with an increase of p_2. Since, according to Fig. 12, the ratio of the associated components of the basis vector $\widetilde{\mathbf{v}}_B^{(1)}$ is $\widetilde{v}_B^{(1)}(p_2)/\widetilde{v}_B^{(1)}(p_5) > 0$ for all states and delay times τ, a small variation $\Delta p_5 >$ leads to a variation $\Delta p_2 > 0$; see Eq. (26). In the same way one can verify, by considering a small variation of b and the components of $\widetilde{\mathbf{v}}_B^{(2)}$ in Fig. 13, that the pairwise local dependency between the other model parameters is equally predicted by Eq. (49) and by the dependency analysis.

It should be noted that choosing p_5 and p_7 is not the only possible combination to construct $\mathbf{V}_{B,1}$, Eq. (44), and with that a different $\widetilde{\mathbf{V}}_B$. We repeated the dependency analysis by choosing p_2 and p_7, checking that $\mathbf{V}_{B,1}$ is nonsingular for the investigated states and delay times τ, and were able to verify that the results of the analysis are consistent with a transformation of the model Eq. (49). For that, we introduced new scaling parameters e and f, where setting $a = e/f$ and $b = f$ in Eq. (49) yields

$$\begin{aligned}\dot{x}_1 &= \frac{1}{f}p_1x_2 + \frac{1}{e}p_2x_3,\\ \dot{x}_2 &= fp_3x_1 + p_4x_2,\\ \dot{x}_3 &= \frac{e}{f}p_5 + p_6x_3 + fp_7x_1x_3.\end{aligned} \tag{50}$$

Equation (50) was then used for the verification.

C. Hindmarsh-Rose model

In this example we use the Hindmarsh-Rose (HR) neuron model [45], which generates typical neuronal activity such as spiking and bursting governed by dynamics on separated time scales. The system consists of the model equations

$$\begin{aligned}\dot{x}_1 &= -x_1^3 + p_1x_1^2 + x_2 - x_3,\\ \dot{x}_2 &= 1 - p_2x_1^2 - x_2,\\ \dot{x}_3 &= p_3[x_1 + p_5(p_4 - x_3)],\end{aligned} \tag{51}$$

and a measurement function

$$h(\mathbf{x}) = x_1, \tag{52}$$

where x_1 denotes the membrane potential. x_2 and x_3 describe slow and fast ion current rates, respectively, and the values of the model parameters are $\mathbf{p} = (3,5,0.004,3.19,0.25)$.

To obtain representative states of the system, we first integrate the model Eq. (51) for 10^6 steps with a step size of 0.1. Then, every 100th state was used for the analysis resulting in 10^4 states from the attractor. For each state the Jacobian matrix $D\mathbf{G}$ of the delay reconstruction map Eq. (5) was computed with respect to all $D = 3$ model variables and all $N_p = 5$ model parameters for different delay times τ using a reconstruction dimension of $K = 20$. Since

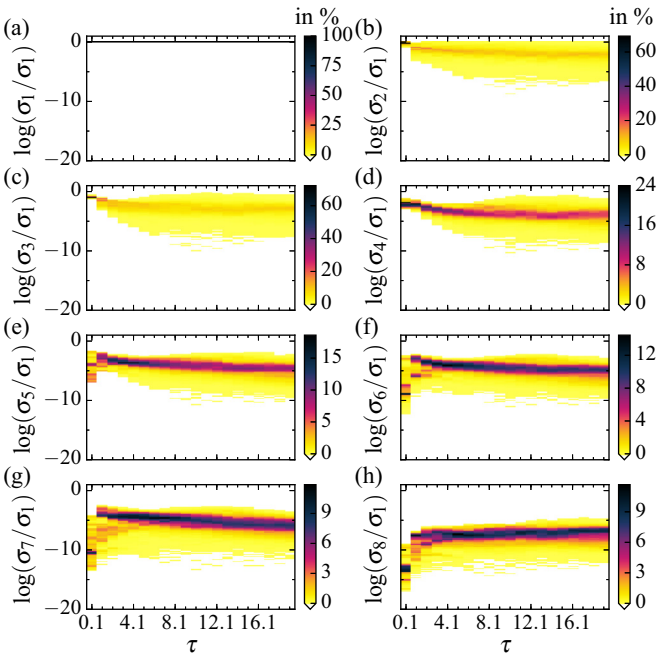

FIG. 14. Histograms (color coded) of the normalized singular values σ_i/σ_1, Eq. (11) (computed with delay reconstruction dimension $K = 20$; log denotes the logarithm with base 10), using all variables ($D = 3$) and parameters $N_p = 5$ of the HR model, Eq. (51), where x_1 is assumed to be measured, Eq. (52). For each τ the singular values σ_i are computed for 10^4 points on the attractor using $\mathbf{p} = (3,5,0.004,3.19,0.25)$. The smallest normalized singular value σ_8/σ_1 is of magnitude $\sigma_8/\sigma_1 \approx 10^{-7}$ and, hence, numerically greater than zero.

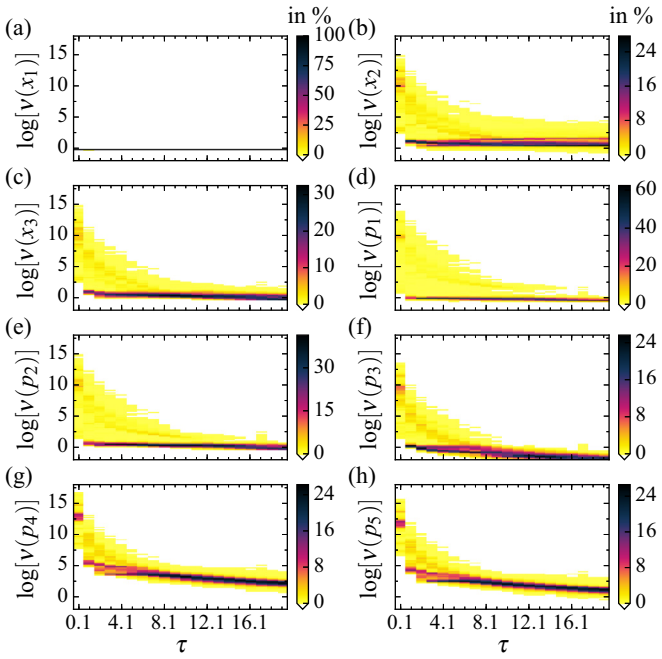

FIG. 15. Histograms (color coded) of the measure of uncertainty $\nu(w_i)$, Eq. (33) (computed with delay reconstruction dimension $K = 20$; log denotes the logarithm with base 10), of all $D = 3$ variables and all $N_p = 5$ parameters for the HR-model, Eq. (51), where x_1 is assumed to be measured, Eq. (52). For each delay time τ, $\nu(w_i)$ are computed for 10^4 points on the attractor using $\mathbf{p} = (3,5,0.004,3.19,0.25)$. (a)–(f) Uncertainties of x_1, x_2, x_3, p_1, p_2, and p_3 converge (on average) to comparatively small values for larger τ. (g),(h) Uncertainties of p_4 and p_5 converge (on average) to comparatively larger values, indicating a worse local estimability compared to the other quantities.

$K > 2(D + N_p) = 17$, the sufficient condition from Ref. [33] is fulfilled.

Histograms of the normalized singular values σ_i/σ_1 of $D\mathbf{G}$, Eq. (11), computed based on all states and for different τ, are shown in Fig. 14. We can see that even the smallest singular value σ_8/σ_1 is of magnitude $\sigma_8/\sigma_1 \approx 10^{-7} > 0$ for increasing τ. Because the example using the Colpitts oscillator shows that in this context singular values can numerically also be much closer to zero (σ_7/σ_1 in Fig. 3 converges to values of magnitude 10^{-16}), we consider all singular values of this result to be nonzero. Therefore, the dimension of the null space of $D\mathbf{G}$, see Eqs. (13) and (14), is $D_N = 0$. This means that for all investigated states and delay times τ there exist no variation of model parameters or variables $\Delta\mathbf{w} \neq 0$ such that the output of the system at present and delayed times is kept unchanged, $\Delta\mathbf{g} = 0$. Therefore, in principle, all quantities are locally estimable.

Next we consider histograms of the uncertainties, Eq. (33), for all three model variables and all five parameters based on the same 10^4 states, K, measurement function Eq. (52) and τ as previously; see Fig. 15. For large τ on average $\nu(p_4)$ and $\nu(p_5)$ are relatively large. This is a hint that the estimation of p_4 and p_5 from a x_1 time series is difficult. In Ref. [36] comparable histograms of $\nu(w_i)$ were considered for a fixed delay time and different reconstruction dimensions K (up to $K \approx 2000$) giving the same result that $\nu(p_4)$ and $\nu(p_5)$ converge to large values.

In the following we consider histograms of the correlation coefficients $\rho(w_i, w_j)$, Eq. (34), computed for all $D = 3$ model variables and all $N_p = 5$ model parameters based on the same 10^4 states, K, measurement function Eq. (52), and τ as previously; see Fig. 16. In Figs. 16(a)–16(g) the distributions of correlation coefficients between all quantities and x_1 are shown. One can see that for most states and delay times τ the correlation coefficients are numerically close to zero, indicating only a weak correlation between x_1 and the other quantities. Taking into account that the uncertainty of x_1 is comparably small (see Fig. 15), it should be relatively easy to estimate x_1 from a x_1 time series.

In Figs. 16(h)–16(a1) the distributions of the correlation coefficients are spread over the whole interval $\rho(w_i, w_j) \in [-1,1]$, almost independent of τ. Since for many states and τ it is $1 < \rho(w_i, w_j) < 1$, correlations between the quantities exist, but are not all too strong (on average).

Special attention should be payed to the distributions of $\rho(p_4, p_5)$ shown in Fig. 16(b1). $\rho(p_4, p_5)$ is (close to) -1 for almost all of the analyzed states and τ, indicating a relatively strong negative correlation between p_4 and p_5. This is a hint of an almost linear relation between both quantities. It is very likely that an increase of p_4 leads to a decrease of p_5 by keeping the delay reconstruction vector $\mathbf{g}$ in Eq. (5), which is the output of the system at present and delayed times, almost unchanged.

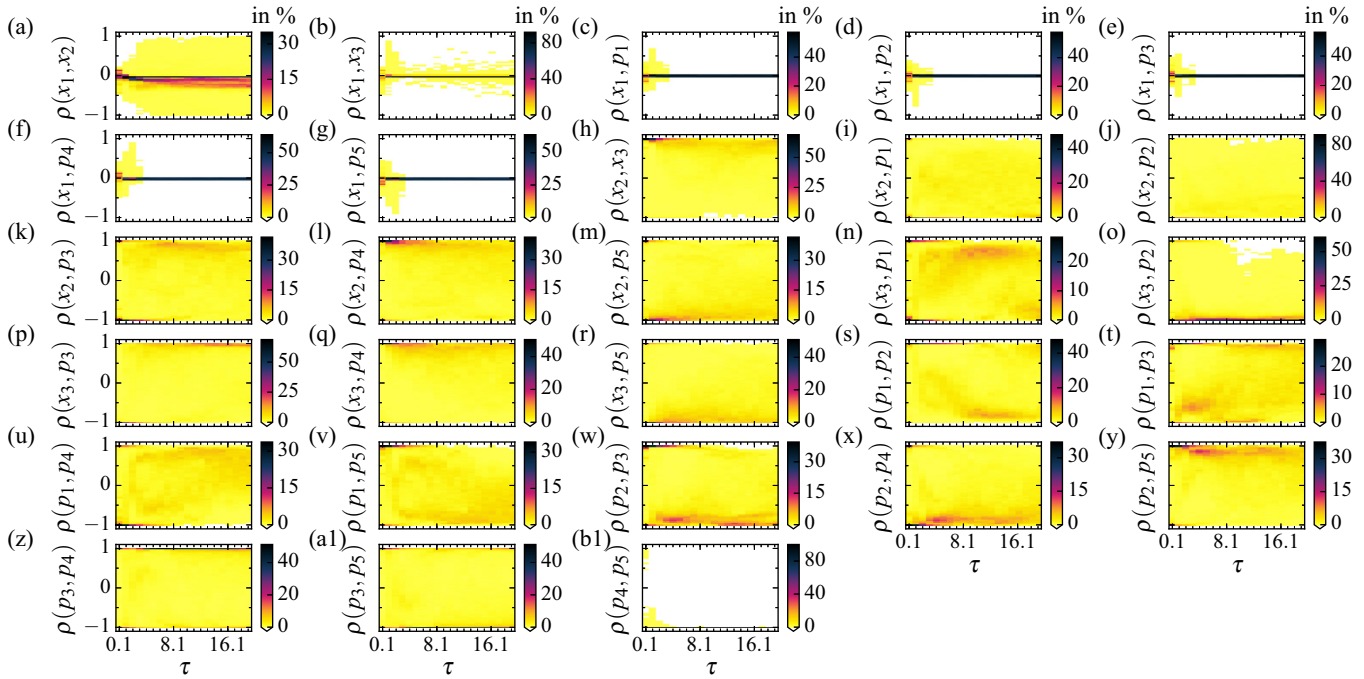

FIG. 16. Histograms (color coded) of the correlation coefficients ρ_{ij} (computed with $K = 20$) of all variables and parameters of the HR model, Eq. (51), where x_1 is assumed to be measured, Eq. (52). For each τ the correlation matrix $\boldsymbol{\rho}$, Eq. (34) is computed for 10^4 points on the attractor using $\mathbf{p} = (3,5,0.004,3.19,0.25)$. (a)–(a1) The quantities are not very correlated, indicating that they should be estimable from a x_1 time series. (b1) There is a strong, but not perfect, negative correlation between p_4 and p_5.

However, there are still analyzed states where $\rho(p_4,p_5)$ is not (close to) -1 (especially for smaller τ), indicating that the negative correlation is not perfect. Hence, in principle, it should be possible to estimate p_4 and p_5 from a x_1 time series, although the estimation might be difficult and error prone.

Thus, one can expect that the simultaneous estimation of all variables and all parameters should give good results, but it is likely that the estimated values of p_4 and p_5 are sensitive with respect to small numerical (truncation) errors or small variations in the data (for example, the length of the time series, its underlying dynamics, or a slightly different noise level).

To illustrate this effect, a twin experiment is performed where all variables and parameters of the HR model Eq. (51) are estimated from noisy artificial data of the first model variable x_1 using the estimation method described in Sec. III. The diagonal weighting matrix in the cost function Eq. (36) to minimize is chosen to be $\mathbf{B} = \mathrm{diag}(1,1,10)$.

To generate the data, the model was integrated using the parameters $\mathbf{p} = (3,5,0.004,3.19,0.25)$. The solution $\mathbf{z}(t)$ (denoted as the "true solution") is then used to create the noisy data time series $\{\eta(t_n)\}$, $t_n = 0,0.5,\ldots 1500$, with

$$\eta(t_n) = z_1(t_n) + \xi(t_n), \quad \xi(t_n) \sim \mathcal{N}(0,0.004) \tag{53}$$

(SNR = 18.3 dB). All model variables have to be estimated at the times $t_m = 0,0.05,\ldots,1500$ using the measurement function Eq. (52) together with all five model parameters. Hence, data are available only at every tenth time step. Figures 17(b), 17(c), and 17(d) show that estimated solutions of x_i (blue lines) for the model variables and the corresponding true solution z_i (red dashed lines) match quite well. In Fig. 17(a) one can see that the output of the measurement function $h(\mathbf{x}(t))$ (blue line) coincides with the noisy data $\{\eta(t_n)\}$ (green circles). As predicted by the correlation analysis, the estimated values of the first three parameters ($p_1 = 3.00$, $p_2 = 4.99$, and $p_3 = 0.004\,02$) are very close to the values used to generate the data. However, p_4 and p_5 are estimated to $p_4 = 4.19$ and $p_5 = 0.191$, which is a much larger deviation to the values used to generate the data. This is in coincidence

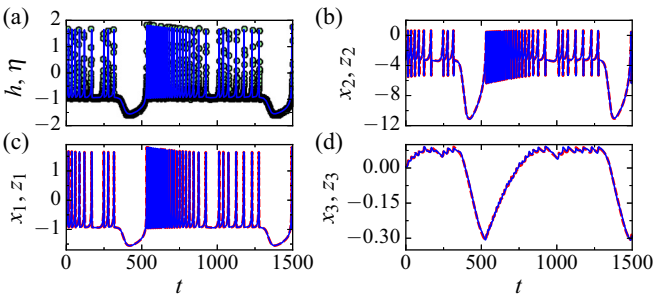

FIG. 17. State and parameter estimation (using the method from Sec. III) of all $D = 3$ variables and all $N_p = 5$ parameters of the HR-model, Eq. (51), from a noisy x_1 time series $\{\eta(t_n)\}$, Eq. (53), using the measurement function $h(\mathbf{x})$, Eq. (52). (a) Output $h(\mathbf{x})$ (blue line) matches the data $\{\eta(t_n)\}$ (green circles). (b),(c),(d) Estimated model variables x_i (blue lines) match the true solutions z_i (red dashed lines; used to generate $\{\eta(t_n)\}$). Parameters are estimated to $\mathbf{p} = (3.00,4.99,0.00402,4.19,0.191)$. That is, p_1, p_2, and p_3 are estimated to values similar to those used to generate the data. p_4 and p_5 are estimated to much less accurate values.

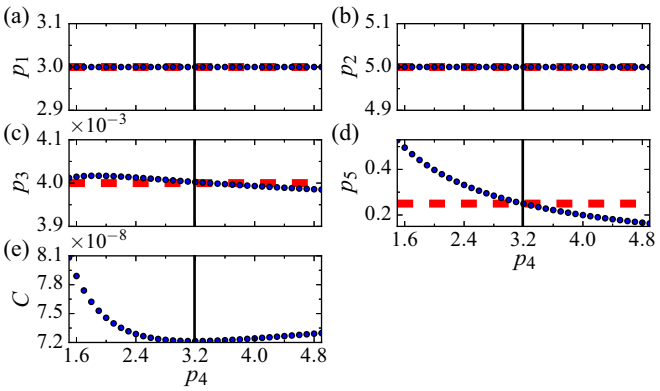

FIG. 18. State and parameter estimation (using algorithm in Sec. III) of the HR model, Eq. (51), from a clean x_1 time series. (a)–(d) The parameter values shown by the red dashed lines were used, beside $p_4 = 3.19$ [black vertical line in (a)–(e)], to generate the data. For each fixed p_4 all other parameters were estimated (blue dots) beside the model variables (profile likelihood approach [39,40]). That p_5 is estimated to a too-small value, if p_4 is too large, is consistent with the negative correlation between both parameters; see Fig. 16(b1). Nevertheless, p_1, p_2, and p_3 are estimated (almost) correctly to values used to generate the data, (almost) independent of the value of p_4. (e) Although cost function C, Eq. (36), is very small for all p_4, C exhibits a minimum around $p_4 \approx 3.19$ (used to generate the data).

with that the uncertainties of p_4 and p_5 are larger compared to uncertainties of the other parameters (on average); see Fig. 15.

The parameter p_4 is estimated to a too-large value and p_5 to a too-small value. This relation coincides with the negative correlation found in the correlation analysis; see Fig. (16).

Although p_4 and p_5 are not estimated close to the correct values, its estimates are not completely wrong. Another twin experiment showed that it is possible to estimate all five parameters close to the values used to generate the data if one removes the measurement noise in Eq. (53) and provides data at every time step the variables will be estimated. This confirms the hypothesis that the simultaneous estimation of p_4 and p_5 should, in principle, be possible, but it will not be very robust.

To confirm the predicted negative correlation of p_4 and p_5, the twin experiment is repeated without measurement noise in the data. Furthermore, data are available at every time step where the model variables are estimated. p_4 is fixed to different values and p_1, p_2, p_3, and p_5 are estimated beside the model variables. The dependency of the estimated parameter values on p_4 (profile likelihood approach [39,40]) is shown in Fig. 18.

The horizontal red dashed lines in Figs. 18(a) to 18(d) show the parameter values p_1, p_2, p_3, and p_5, respectively, used to generate the data. The value of p_4 used for generating the data is shown in Figs. 18(a) to 18(e) by the vertical black line at $p_4 = 3.19$. In Figs. 18(a) to 18(c) one can see that the estimated values of p_1 and p_2 coincide with true values (red dashed lines) used to generate the data sufficiently well, independently of the value of p_4. Only a very weak dependency of p_3 on p_4 was observed. Figure 18(d) shows that an increase of $p_4 > 3.19$ goes along with estimated values of $p_5 < 0.25$ which are smaller than the value used to generate the data. This dependency between both parameters was predicted correctly by the negative correlation $\rho(p_4, p_5) \approx -1$ for most states and τ; see Fig. 16(b1).

Figure 18(e) shows that the value of the cost function, Eq. (36), at the estimated solution is relatively small for all considered values of p_4. Nevertheless, the cost function exhibits a minimum around $p_4 = 3.19$. The fact that it is not very steep makes the estimation process not very robust. Errors in the estimation problem (measurement noise, truncation errors, ...) may easily shift the minimum to a different value of p_4 and, hence, lead to a wrong estimation of p_4 and p_5. This is likely to be the reason why the state and parameter estimation from a noisy time series with fewer data points (see Fig. 17) gives worse results for p_4 and p_5.

V. CONCLUSION

When estimating parameters and trajectories of variables of a dynamical model from time series, often the used estimation method does not provide information about the accuracy and uniqueness of the estimates.

In this article we address this problem and exploit features of the null space of the Jacobian matrix of the delay coordinates map to identify parameters and variables of a dynamical model which cannot be uniquely estimated from a measured time series (generated by the same model of a process the model aims at describing). The proposed method is applicable to state and parameter vectors obtained, for example, as estimated solutions from a state- and parameter-estimation algorithm. This analysis not only identifies those unknown quantities which are not estimable (in a strict sense), but also information about their relations. Using this information, one can specify the impact of setting one (or more) of the nonestimable quantities to fixed values. If the dimension of the null space of the Jacobian matrix of the delay coordinates map is D_N, then setting a suitable set of D_N non-estimable unknowns to fixed values makes all remaining unknowns estimable. Criteria for selecting these D_N unknowns are obtained from the orientation of the null space and the pattern of vanishing elements of its basis matrix. These aspects were illustrated with (time series from) the Colpitts oscillator and the Rössler system. Even if all parameters and not measured variables are locally estimable, there may still be unknowns which are strongly correlated. This case can be identified by a suitable correlation analysis, as shown with the Hindmarsh-Rose model. All concepts presented and discussed can be generalized to multivariate time series and spatially extended systems or dynamical networks. Furthermore, they may also serve to answer the question as to whether two identical systems can be synchronized by a given unidirectional coupling signal.

ACKNOWLEDGMENTS

S.L. and U.P. acknowledge support from the BMBF (Grant No. FKZ031A147, GO-Bio), the DFG (Grant No. SFB 1002, project A18), and the German Center for Cardiovascular Research (DZHK e.V.).

[1] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990).
[2] U. Parlitz, L. Junge, and L. Kocarev, Phys. Rev. E **54**, 6253 (1996).
[3] D. Rey, M. Eldridge, M. Kostuk, H. D. I. Abarbanel, J. Schumann-Bischoff, and U. Parlitz, Phys. Lett. A **378**, 869 (2014).
[4] D. Rey, M. Eldridge, U. Morone, H. D. I. Abarbanel, U. Parlitz, and J. Schumann-Bischoff, Phys. Rev. E **90**, 062916 (2014).
[5] R. E. Kalman, J. Basic. Eng. **82**, 35 (1960).
[6] G. Evensen, J. Geophys. Res. **99**, 10143 (1994).
[7] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, 2nd ed. (Springer, Dordrecht, New York, 2009).
[8] P. J. van Leeuwen, Mon. Weather Rev. **137**, 4089 (2009).
[9] P. J. van Leeuwen, Q. J. R. Meteorol. Soc. **136**, 1991 (2010).
[10] Y. Sasaki, Mon. Weather Rev. **98**, 875 (1970).
[11] H. D. I. Abarbanel, D. R. Creveling, R. Farsian, and M. Kostuk, SIAM J. Appl. Dyn. Syst. **8**, 1341 (2009).
[12] J. Bröcker, Q. J. R. Meteorol. Soc. **136**, 1906 (2010).
[13] A. H. Jazwinski, *Stochastic Processes and Filtering Theory* (Academic Press, New York, 1970).
[14] H. Abarbanel, *Predicting the Future: Completing Models of Observed Complex Systems* (Springer, New York, 2013).
[15] O. Talagrand and P. Courtier, Q. J. R. Meteorol. Soc. **113**, 1311 (1987).
[16] P. Courtier and O. Talagrand, Q. J. R. Meteorol. Soc. **113**, 1329 (1987).
[17] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory* (Wiley & Sons, New York, 1967).
[18] R. Hermann and A. J. Krener, IEEE Trans. Autom. Control **22**, 728 (1977).
[19] E. D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, 2nd ed. (Springer, New York, 1998).
[20] H. Nijmeijer, Int. J. Control **36**, 867 (1982).
[21] C. Letellier, L. A. Aguirre, and J. Maquet, Phys. Rev. E **71**, 066213 (2005).
[22] C. Letellier, L. Aguirre, and J. Maquet, Commun. Nonlinear. Sci. Numer. Simul. **11**, 555 (2006).
[23] H. Nijmeijer and A. v. d. Schaft, *Nonlinear Dynamical Control Systems* (Springer Science & Business Media, New York, 1990).
[24] L. Aguirre, IEEE Trans. Educ. **38**, 33 (1995).
[25] R. E. Kalman, in *Proceedings of the First IFAC Congress* (Butterworth, London, 1961), Vol. 1, p. 481.
[26] R. E. Kalman, J. Soc. Ind. Appl. Math. **1**, 152 (1962).
[27] M. Gevers, A. S. Bazanella, D. F. Coutinho, and S. Dasgupta, Automatica **63**, 38 (2016).
[28] D. Aeyels, SIAM J. Control Optim. **19**, 595 (1981).
[29] F. Takens, in *Dynamical Systems and Turbulence, Warwick 1980*, edited by D. Rand and L.-S. Young, Lecture Notes in Mathematics Vol. 898 (Springer, Berlin, Heidelberg, 1981), p. 366.
[30] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[31] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge Nonlinear Science Series Vol. 7 (Cambridge University Press, Cambridge, U.K., 1997).
[32] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, 2nd ed. (Springer Verlag, Berlin, 1997).
[33] E. D. Sontag, J. Nonlinear Sci. **12**, 553 (2003).
[34] U. Parlitz, J. Schumann-Bischoff, and S. Luther, Chaos **24**, 024411 (2014).
[35] D. A. Belsley, *Conditioning Diagnostics: Collinearity and Weak Data in Regression*, 1st ed. (Wiley-Interscience, New York, 1991).
[36] U. Parlitz, J. Schumann-Bischoff, and S. Luther, Phys. Rev. E **89**, 050902 (2014).
[37] J. Schumann-Bischoff and U. Parlitz, Phys. Rev. E **84**, 056214 (2011).
[38] J. Schumann-Bischoff, S. Luther, and U. Parlitz, Commun. Nonlinear. Sci. Numer. Simul. **18**, 2733 (2013).
[39] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer, Bioinformatics **25**, 1923 (2009).
[40] A. Raue, V. Becker, U. Klingmüller, and J. Timmer, Chaos **20**, 045105 (2010).
[41] M. Kennedy, IEEE Trans. Circuits Syst. I **41**, 771 (1994).
[42] O. E. Rössler, Phys. Lett. A **57**, 397 (1976).
[43] C. Lainscsek, Phys. Rev. E **84**, 046205 (2011).
[44] C. Lainscsek, J. Weyhenmeyer, T. J. Sejnowski, and C. Letellier, Chaos Solitons Fractals **76**, 182 (2015).
[45] J. L. Hindmarsh and R. M. Rose, Proc. R. Soc. London, Ser. B **221**, 87 (1984).
[46] J. Schumann-Bischoff, U. Parlitz, H. D. I. Abarbanel, M. Kostuk, D. Rey, M. Eldridge, and S. Luther, Chaos **25**, 053108 (2015).
[47] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, Cambridge, U.K., New York, 2007).
[48] P. J. van Leeuwen and G. Evensen, Mon. Weather Rev. **124**, 2898 (1996).
[49] A. C. Lorenc, Q. J. R. Meteorol. Soc. **112**, 1177 (1986).
[50] K. Levenberg, Q. Appl. Math. **2**, 164 (1944).
[51] D. W. Marquardt, J. Soc. Ind. Appl. Math. **11**, 431 (1963).
[52] M. I. A. Lourakis, *Computer Vision – ECCV 2010*, edited by K. Daniilidis, P. Maragos, and N. Paragios, Lecture Notes in Computer Science Vol. 6312 (Springer, Berlin, Heidelberg, 2010), pp. 43–56.
[53] A. Griewank, D. Juedes, and J. Utke, ACM Trans. Math. Softw. **22**, 131 (1996).
[54] S. F. Walter, Pyadolc (2015), https://github.com/b45ch1/pyadolc.
[55] A. Walther and A. Griewank, *Combinatorial Scientific Computing*, CRC Computational Science Series (Chapman & Hall, Boca Raton, FL, 2012), Chap. Getting Started with ADOL-C, p. 181.
[56] D. R. Creveling, P. E. Gill, and H. D. I. Abarbanel, Phys. Lett. A **372**, 2640 (2008).