Using waveform information in nonlinear data assimilation

Daniel Rey,[*] Michael Eldridge, Uriel Morone, and Henry D. I. Abarbanel[†]
Department of Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0374, USA

Ulrich Parlitz and Jan Schumann-Bischoff
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 8 September 2014; published 22 December 2014)

Information in measurements of a nonlinear dynamical system can be transferred to a quantitative model of the observed system to establish its fixed parameters and unobserved state variables. After this learning period is complete, one may predict the model response to new forces and, when successful, these predictions will match additional observations. This adjustment process encounters problems when the model is nonlinear and chaotic because dynamical instability impedes the transfer of information from the data to the model when the number of measurements at each observation time is insufficient. We discuss the use of information in the waveform of the data, realized through a time delayed collection of measurements, to provide additional stability and accuracy to this search procedure. Several examples are explored, including a few familiar nonlinear dynamical systems and small networks of Colpitts oscillators.

DOI: 10.1103/PhysRevE.90.062916 PACS number(s): 05.45.Pq, 02.30.Zz

I. INTRODUCTION

In constructing models of complex systems, the dynamical states and fixed parameters of the model are typically unknown and must therefore be inferred through data generated by observing the system. To test or validate a model requires an accurate estimate of its fixed parameters and its unobserved state variables, which then must be used to predict the outcome of new measurements when the same system is subjected to forces different from those that were used to construct the estimate. This enterprise of incorporating information from measured data into the properties of a predictive model is known as data assimilation in geophysical sciences and is practiced in a wide spectrum of scientific inquiries, including numerical weather prediction [1], systems biology [2,3], biomedical engineering [4], chemical engineering [5], biochemistry [6], coastal and estuarine modeling [7,8], cardiac dynamics [9], and nervous system networks [10,11], among many others.

We wish to emphasize throughout this paper that estimation alone is not enough when seen through the measured state variables only. One can, and often does, estimate the observations well, but this sheds little or no light on our knowledge of the unobserved states and unknown parameters, both of which must also be known in order to predict beyond the observation window. Prediction then is the metric one must adopt to assess the quality of a model's consistency with given data.

Previous work has shown that when the system under consideration yields chaotic trajectories the dynamical instability associated with sensitivity to initial conditions impedes the successful identification of the initial state and parameters of the system [12,13]. In particular, it has been observed that many data assimilation techniques require a minimum number of measurements to succeed, even when the noise levels are low [13–16].

This paper expands on a method introduced in Ref. [17], which can function successfully even when the available measurements are fewer than what was previously shown to be necessary. This is possible because, rather than comparing the estimate to the observations at individual points in time, we instead compare the waveforms of the data and model output over some period. The idea is that the waveform contains additional information, which can be used to improve the accuracy of the estimate for the unmeasured states and parameters.

The paper is organized as follows: Section II details the structure of a data assimilation problem and introduces some established approaches. We describe the problem in terms of the probability distribution of possible trajectories conditioned on the observations. We then illustrate the challenges posed by chaotic instability and present our solution to controlling these instabilities by using the information in the waveform of the observed data, via time delayed measurements. Section III describes the central numerical obstacle in our technique, inverting an ill-conditioned matrix, and discusses heuristics to improve its stability. Section IV presents examples on a variety of models from nonlinear dynamics with different dimensions and degrees of instability. Section V discusses a method for directly estimating the critical number of measurements, and Section VI investigates the network properties of coupled oscillators. We summarize our results in Sec. VII.

II. ASSIMILATING DATA INTO MODELS OF OBSERVED PROCESSES

During an assimilation or measurement window $[0,T]$, data from an observed system are presented to a system model. Various methods are employed [12,13] to estimate the fixed parameters and full state of the model (both observed and unobserved state variables) at the end of the assimilation window $t = T$. To validate the model and the estimates,

[*]drey@physics.ucsd.edu
[†]Marine Physical Laboratory (Scripps Institution of Oceanography), California, USA.

predictions are compared with further observations in such a way that information about these subsequent observations is not utilized to further modify the estimates of the fixed model parameters.

The model is stated in terms of differential equations for fields $\phi_k(\mathbf{r},t)$ or point objects $q_\alpha(t)$, so one must estimate all of the $\phi_k(\mathbf{r},T)$ or $q_\alpha(T)$ in order to predict the dynamical behavior for $t > T$. We reduce the continuous set of independent variables $(\mathbf{r},t)$ to a finite grid in space and time, arriving at a set of state variables $x_a(t_n) = x_a(n)$ where $a = \{1,2,\ldots,D\}$ and $n = \{0,1,\ldots,N\}$. The resulting state $\mathbf{x}(n) = \{x_1(n),x_2(n),\ldots,x_D(n)\}$ follows the rule

$$x_a(n+1) = f_a(\mathbf{x}(n)), \tag{1}$$

constituting our model. In this discrete time formulation we have treated the N_P fixed parameters as state variables satisfying $x_j(n+1) = x_j(n)$ for $j = \{1,2,\ldots,N_P\}$. For purposes of our discussion, we will often use the continuous time version of this discrete time map in the form

$$\frac{dx_a(t)}{dt} = F_a(\mathbf{x}(t)), \tag{2}$$

although all calculations are actually performed with Eq. (1).

The information we wish to transfer to this model resides in the L measurements $\mathbf{y}(n) = \{y_1(n),y_2(n),\ldots,y_L(n)\}$ made at each time t_n within an observation window $\{t_0,t_1,\ldots,t_n,\ldots,t_m = T\}$. To connect the measurements $y_l(n)$ with the solution of the dynamical equations described by the model we must specify a "measurement function," which realizes the data in terms of the model output $\mathbf{x}(n)$ as $y_l(n) = h_l(\mathbf{x}(n))$. When $y_l(n) \approx h_l(\mathbf{x}(n))$ within the estimation window, the model is said to be consistent with the data, while validation of the model requires prediction of the observables $h_l(\mathbf{x}(t))$ for $t > T$.

The method of estimating the states and parameters $x_a(n)$ relies on systematic adjustment from some initial state $x_a^{(0)}(n)$ through an iterative process that produces a sequence of estimates

$$x_a^{(0)}(n) \to x_a^{(1)}(n) \to x_a^{(2)}(n) \to \cdots \to x_a^{(J)}(n)$$

to a final estimate $x_a^{(J)}(n)$ using some numerical method deemed to converge as $J \to \infty$ to a "correct" answer: $h_l(\mathbf{x}^J(n)) \approx y_l(n)$. The adjustments to the $\mathbf{x}^{(j)}(n)$ are perturbations to the states and parameters to improve the relationship $y_l(n) \approx h_l(\mathbf{x}^{(j)}(n))$, taking the dynamical rules of the model into account.

Our discussion will primarily focus on the case where the model developed for understanding observational data is perfect. That is, the data has no model errors and provides a deterministic constraint on how estimations and predictions are carried out. In this limit where the model dynamics are known, a simple technique for the direct transfer of information from observations to the dynamical model involves adding a nonphysical control term associated with each measurement to perturb the state of the model system $\mathbf{x}(t)$ toward the observations $\mathbf{y}(t)$ as the model evolves in time. The equations for these "coupled dynamics" are given by

$$\frac{dx_l(t)}{dt} = F_l(\mathbf{x}(t)) + \sum_{l'=1}^{L} g_{l,l'}(t)[y_{l'}(t) - h_{l'}(\mathbf{x}(t))], \tag{3}$$

for the $l = \{1,2,\ldots,L\}$ measured states, and

$$\frac{dx_k(t)}{dt} = F_k(\mathbf{x}(t)),$$

for the $k = \{L+1,L+2,\ldots,D\}$ unmeasured states. The control term $\mathbf{g}(t)$ is positive definite and has a narrow peak centered at $t = t_n$, so that it impacts the model trajectory only at times when an observation is made.

This construction has been implemented in the meteorological literature for many years, where it is called "nudging," Newtonian relaxation, or 4DDA, and is rooted in the theory of controls and dynamical systems [18,19]. From a dynamical systems perspective, the control term $\mathbf{g}(t)$ transfers information from the measured data to the model state by coupling the estimated (model) system to the true (physical) system to promote the synchronization of the model with the data [20].

This process is essentially a dynamical inverse, wherein the model state and parameters are deduced from the measured data. The model dynamics act as a filter that supplies additional information about the unobserved states of the model, which are required to construct an accurate estimate of the state of the true system. This idea of using the model as a filter is well established and is the core idea behind algorithms like the Kalman–Bucy filter [21], as well as its various extensions. In those algorithms, the coupling term $\mathbf{g}(t)$ is dynamical and chosen to minimize an estimate of the error covariance [12].

For our purposes however, we focus on the simple case where $\mathbf{g}(t)$ is constant and diagonal. Its value must be chosen judiciously, to synchronize the model output with the measured data without destabilizing the model. When this is accomplished, accurate prediction follows.

A. The action $A_0(\mathbf{X}) = -\ln[P(\mathbf{X}|\mathbf{Y})]$

In practice, the model is almost never perfectly accurate. When model errors are present or when the dynamics of the model are stochastic, the iterative process taking $\mathbf{x}^{(j)}(n) \to \mathbf{x}^{(j+1)}(n)$ may be formulated as [13]

(i) A numerical optimization procedure to estimate an "optimal" path of the states

$$\mathbf{X} = \{\mathbf{x}(0),\mathbf{x}(1),\ldots,\mathbf{x}(m)\}.$$

(ii) A Monte Carlo algorithm seeking to make an accurate estimate of a conditional probability density function $P(\mathbf{X}|\mathbf{Y})$ for all states in the observation window, conditioned on the collection of observations

$$\mathbf{Y} = \{\mathbf{y}(0),\mathbf{y}(1),\ldots,\mathbf{y}(m)\}.$$

Prediction beyond the measurement window, $t > T$, requires all components of the state $\mathbf{x}(T)$ and either the deterministic dynamical rule (1) or a dynamical rule that includes stochastic model errors.

Since the data are noisy and the model inevitably has errors, most applications require us to estimate $P(\mathbf{X}|\mathbf{Y})$. This distribution contains all information relevant to the data assimilation problem. It allows one to decide whether the best estimate for the state $\mathbf{X}$ is the mean, mode (i.e., the maximum *a posteriori* estimate), or some other measure of the distribution. Moreover, it allows us to quantify the uncertainty in our estimate by computing statistical quantities as functions

$G(\mathbf{X})$ of the path $\mathbf{X}$ with the form

$$E[G(\mathbf{X})|\mathbf{Y}] = \frac{\int d\mathbf{X} P(\mathbf{X}|\mathbf{Y}) G(\mathbf{X})}{\int d\mathbf{X} P(\mathbf{X}|\mathbf{Y})} = \frac{\int d\mathbf{X} \exp[-A_0(\mathbf{X})] G(\mathbf{X})}{\int d\mathbf{X} \exp[-A_0(\mathbf{X})]}. \quad (4)$$

The action $A_0(\mathbf{X})$ in Eq. (4) is composed of

(i) Terms moving the model state from time t_n to time t_{n+1} through the observation window.

(ii) Terms associated with the modification of the conditional probability distribution at times when measurements are made.

The general formulation, which incorporates noisy measurements and model errors, is given by

$$A_0(\mathbf{X}) = -\sum_{n=0}^{m} CMI(\mathbf{x}(n),\mathbf{y}(n)|\mathbf{Y}(n-1)) - \sum_{n=0}^{m-1} \ln[P(\mathbf{x}(n+1)|\mathbf{x}(n))] - \ln[P(\mathbf{x}(0))].$$

The term $P(\mathbf{x}(0))$ is the initial distribution of the states at the beginning of the assimilation window t_0. If no prior information is available, this distribution is taken to be uniform and can be ignored as an additive constant. The term $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ is the transition probability for the state $\mathbf{x}(n) \to \mathbf{x}(n+1)$. For deterministic models, this term is a delta function. The conditional mutual information term is

$$CMI(\mathbf{x}(n),\mathbf{y}(n)|\mathbf{Y}(n-1)) = \ln\left[\frac{P(\mathbf{x}(n),\mathbf{y}(n)|\mathbf{Y}(n-1))}{P(\mathbf{x}(n)|\mathbf{Y}(n-1))P(\mathbf{y}(n)|\mathbf{Y}(n-1))}\right], \quad (5)$$

where $\mathbf{Y}(n) = \{\mathbf{y}(n),\mathbf{y}(n-1),\ldots,\mathbf{y}(0)\}$ is the collection of measurements up to time t_n. This term contains the additional information transferred from the current measurement $\mathbf{y}(n)$ to the model $\mathbf{x}(n)$, conditioned on the past measurements in $\mathbf{Y}(n-1)$.

If the measurement noise and model errors are Gaussian distributed with respective inverse covariance matrices $\mathbf{R}^m$ and $\mathbf{R}^f$, the action becomes [13]

$$A_0(\mathbf{X}) = \sum_{n=0}^{m}\sum_{l,l'=1}^{L}\left[\delta m_l(n)\frac{R^m_{l,l'}(n)}{2}\delta m_{l'}(n)\right] + \sum_{n=0}^{m-1}\sum_{a,a'=1}^{D}\left[\delta f_a(n)\frac{R^f_{a,a'}}{2}\delta f_{a'}(n)\right] - \ln[P(\mathbf{x}(0))], \quad (6)$$

where

$$\delta\mathbf{f}(n) := \mathbf{x}(n+1) - \mathbf{x}(n) - \int_{t_n}^{t_{n+1}} \mathbf{f}(\mathbf{x}(t'))dt',$$

$$\delta\mathbf{m}(n) := \mathbf{y}(n) - \mathbf{h}(\mathbf{x}(n))$$

are deviations from the model and measurements, with $\mathbf{R}^f$ and $\mathbf{R}^m$ as their respective inverse covariances.

There is much discussion in the data assimilation literature [12] focused on the development of numerical methods for evaluating the path integral in Eq. (4). Since these integrals tend to be high dimensional, the methods can generally be divided into two categories:

(i) stationary path methods, which seek the paths where $\partial A_0(\mathbf{X})/\partial\mathbf{X} = 0$ and assumes they are the dominant contribution to the integral;

(ii) Monte Carlo methods, which directly sample the distribution $\exp[-A_0(\mathbf{X})]$.

The connection between the two approaches is given by the fact that $P(\mathbf{X}) = \exp[-A_0(\mathbf{X})]$ is the limiting distribution for a distribution $P(\mathbf{X},s)$ of orbits $\mathbf{X}(s)$ satisfying the Langevin equation

$$\frac{dX_\sigma(s)}{ds} = -\frac{\partial A_0(\mathbf{X}(s))}{\partial X_\sigma(s)} + \sqrt{2}\eta_\sigma(s),$$

where the parameter s denotes "algorithmic time." Here, σ is the collection of indices of the path $\mathbf{X}$, and $\eta_\sigma(s)$ is a Gaussian distributed random variable with mean zero, variance unity, and independent at each algorithmic time s. The distribution $P(\mathbf{X},s)$ satisfies a Fokker–Planck equation whose distribution as $s \to \infty$ is $P(\mathbf{X}) = \exp[-A_0(\mathbf{X})]$. The Langevin equation shows the connection between the minima of $A_0(\mathbf{X})$ where $\partial A_0(\mathbf{X})/\partial X_\sigma = 0$ and the distribution of fluctuations $P(\mathbf{X})$ about those minima induced by $\eta_\sigma(s)$.

Our attention in this paper is on the ability to succeed with these methods when multiple stationary paths or multiple local minima of the action are present. These local minima are due to chaotic instability in the dynamics and impede the identification of the optimal path or the Monte Carlo calculation of $P(\mathbf{X}|\mathbf{Y})$ [14]. Thus, even in an ideal situation where the model is known exactly and the data have no noise, estimating the unobserved states and parameters of the model may still be difficult when the dynamics are chaotic.

B. Chaotic instability as an impediment to success and the "critical" number of measurements L_c

When the system under consideration yields chaotic time series $y_l(n)$ and $x_a(n)$, there arises a serious impediment to many of the iterative processes used across multiple scientific fields to search for the set of states and parameters that most closely matches the observed data [14]. This impediment is common to all of the approaches discussed thus far; namely,

(i) the dynamical synchronization (nudging) approach;

(ii) variational or optimization methods, which seek a minimum of $\mathbf{A_0}(\mathbf{X},\mathbf{Y}) = -\ln[P(\mathbf{X}|\mathbf{Y})]$;

(iii) the Monte Carlo estimation framework, which directly samples $P(\mathbf{X}|\mathbf{Y})$.

In both the variational and Monte Carlo frameworks, the problem is manifested as multiple minima in the action $A_0(\mathbf{X},\mathbf{Y}) = -\ln[P(\mathbf{X}|\mathbf{Y})]$ caused by the instability associated with the sensitivity to initial conditions characteristic of chaotic motion. Since small perturbations in the initial values of the path yield large deviations of the action, incoherence of chaotic flows for slightly differing initial states or parameters causes the search surface $A_0(\mathbf{X})$ to be riddled with local minima. The presence of these local minima significantly impedes the algorithmic search for the minimizers.

In the synchronization approach, the impediment arises from instabilities on the L-dimensional synchronization

manifold, where $y_l(n) = h_l(\mathbf{x}(n))$ in the D-dimensional state space. Such behavior may be characterized quantitatively by the conditional Lyapunov exponents (CLEs) for motion on the D-L dimensional submanifold governed by Eq. (1). A necessary condition to achieve synchronization of chaotic systems requires all Lyapunov exponents to be negative [22]. When any of the D-L CLEs are positive, the synchronization manifold is unstable and we observe multiple minima in $\mathbf{X}$ for $A_0(\mathbf{X})$.

While there does not yet exist a rigorous mathematical framework establishing these observations, there exists substantial numerical evidence. For instance, the book by Evensen [12] shows (in Fig. 6.1) multiple minima generated by the Lorenz 1963 model [23] in the graph of the cost function associated with the strong variational method (strong 4D-Var), in which the dynamical equations are used as nonlinear equality constraints, and only the initial conditions appear as control variables in the optimization. Although he does not connect this with the instability on the synchronization manifold, this connection is made in Ref. [24].

Similarly, multiple local minima are also observed in the weak version of 4D-Var, in which model errors are incorporated into the cost function [15]. As the weak 4D-Var method is directly related to the Monte Carlo method through the Langevin equation (7), dynamical instability impacts these techniques as well [15].

For each approach, the impediments to the search for states (and parameters) are removed by increasing the number of measurements L to a value $L_c \leqslant D$. This value L_c we call the critical number of measurements, above which the search surfaces become smooth in $\mathbf{X}$. This smoothing of the action is analogous to a phase transition in the number of measurements L [14]. For instance, given a perfect dynamical model and perfect measured data (no noise), either $L < L_c$ and the search space is riddled with numerous local minima, or $L \geqslant L_c$ and the space is smooth with a single, unique (global) minimum. In the latter case, we observe that predictions made by using one of the aforementioned approaches succeed with high-probability, regardless of the choice of initial condition. Whereas when $L < L_c$, the process is likely to be unsuccessful unless additional knowledge about the initial state of the system is available.

This transition is most evident in the context of synchronization. When $L \geqslant L_c$ the model output synchronizes with the data, otherwise it does not. To understand this quantitatively, recall how the coupling matrix $\mathbf{g}(t)$ modifies the Jacobian of the dynamics (2),

$$\frac{\partial \mathbf{F}(\mathbf{x}(t_n))}{\partial \mathbf{x}} \to \frac{\partial \mathbf{F}(\mathbf{x}(t_n))}{\partial \mathbf{x}} - \mathbf{g}(t_n).$$

With enough measurements $L \geqslant L_c$, a judicious choice of coupling $\mathbf{g}(t)$ is capable of making all the positive CLEs negative, thereby establishing the conditions necessary to synchronize the model with the data. Since only the rows of the Jacobian that correspond to measured state components are altered, the number of measurements is crucial to the success of this process.

Consider a singular value decomposition (SVD) of the Jacobian $\partial \mathbf{F}(\mathbf{x}(t))/\partial \mathbf{x}$ and denote the unstable subspace as the space spanned by the singular vectors whose associated singular values are greater than one [25]. In this unstable subspace, perturbations from the true solution grow exponentially regardless of how close the model estimate is to the truth. The coupling term in Eq. (3) uses information from the measurements to control this unstable subspace. Consequently, one needs enough measurements to span the unstable subspace so that a proper choice of coupling $\mathbf{g}(t)$ may remove the dynamical instability and thereby establish the conditions required to achieve synchronization.

We speak a bit loosely in this paper about the "number of required measurements" L_c. The precise statement must address: (i) the number of measurements, (ii) which states are measured, (iii) the measurement "function" $\mathbf{h}(\mathbf{x}(t))$, as well as (iv) the temporal resolution of the time series. For simplicity, we make the assumption that the measurements are projections $h_l(\mathbf{x}(t)) = x_l(t)$ and the time series is dense or near continuous, such that a measurement is available at every time step Δt of the numerical integration. Since not all measurements carry the same amount of information, we focus on finding a minimal subset L_c that provides enough information to stabilize the instabilities in the model. We shall see that focusing on L_c in this way provides a good sense of how many measurements are required to achieve reliable predictions.

C. Using time delayed measurements to further stabilize the transfer of information

This has been a somewhat general introduction to the problem facing many scientists when seeking to create quantitative models of complex systems. The main issue addressed in this paper arises in the typical situation where the set of measurements L remains smaller than L_c. One must estimate D state variables $\mathbf{x}(T)$ at the end of the measurement window in order to predict. When $L < L_c$ the estimation process is seriously hindered and predictions for $t > T$ will be unreliable.

We are concerned in this paper with removing these impediments in a manner that places the smallest burden on the experiment. That is, we wish to develop techniques to reduce L as much as possible while maintaining the same successful prediction rate obtained when $L > L_c$ and no prior knowledge of the state is known. Our goal is to extract as much information as possible from a given set of measurements, since in most applications the number of measurements is tightly constrained, perhaps by cost, time, or other technological considerations.

Our suggestion is to use information stored in the waveform of the measurements in addition to the values of these quantities at the measurement times to augment the number of observations and, more precisely, to pass more information about the observed system to the model. This idea was previously discussed in Ref. [17], but here we give a more detailed explanation of the method and provide additional numerical results.

In particular, we use the measurements $\mathbf{y}(t_n)$ as well as a collection of the time delayed versions of those measurements as our observations. For this task, we collect all of the measurements at time t_n along with $D_M - 1$ time delayed versions of $\mathbf{y}(t_n)$ into a LD_M-dimensional vector, which we call

$$\mathbf{Y}(t_n) := \{\mathbf{y}(t_n), \mathbf{y}(t_n + \tau), \ldots, \mathbf{y}(t_n + (D_M - 1)\tau)\}. \quad (7)$$

In component form, it may be written as

$$Y_{k:l}(t_n) = y_l(t_n + (k-1)\tau),$$

where $l = \{1,2,\ldots,L\}$ and $k = \{1,2,\ldots,D_M\}$.

The use of time delays of observed data to provide a setting for representing information in nonlinear systems is quite mature and very well tested in the analysis of chaotic behavior [26–32]. In phase space reconstruction, they provide a proxy state space for analyzing properties of the source of chaotic motions. Here, the number of required delays is dictated by geometric considerations, provided the time delay τ yields components for the equivalent of $\mathbf{y}(t_n)$ that are independent in some, usually heuristic, sense.

The usual practice is to use each measurement $\mathbf{y}(t_n)$ independently of measurements at different times. Of course, these measurements are not totally independent of each other, as they come from a dynamical system that describes the physical processes underlying the system's time evolution. The dependence comes from the idea that proceeding from an observation $y(t)$ (using $L = 1$ for illustration) to a time delayed observation $y(t+\tau)$ utilizes some dynamical rule involving all of the degrees of freedom of the observed system, not just those that are observed. So if τ is long enough for the unobserved states of the system to have acted in sufficient magnitude to influence $y(t+\tau)$, then $y(t+\tau)$ possesses information about the overall dynamics not available in $y(t)$ alone. The utility and mathematical value of the time delay construction we develop rests precisely on the information residing in those connections.

Suppose we are able for some physical reason to observe only $L = 1$ variable, $z_0(t)$, in a $K+1$ dimensional dynamical system with other variables $z_k(t)$; $k = \{1,2,\ldots,K\}$ satisfying the differential equations

$$\frac{dz_0(t)}{dt} = G_0(z_0(t),\mathbf{z}(t)),$$
$$\frac{dz_k(t)}{dt} = G_k(z_0(t),\mathbf{z}(t)),$$

then

$$z_0(t+\tau) = z_0(t) + \int_t^{t+\tau} dt' G_0(z_0(t'),\mathbf{z}(t')),$$

and additional information about the time course of the other variables $\mathbf{z}(t)$ for $[t,t+\tau]$ resides in $z_0(t+\tau)$ while it is absent in $z_0(t)$.

If τ is too short relative to the natural times of the $\mathbf{z}(t)$, effectively nothing new will be usable in $z_0(t+\tau)$ about the $\mathbf{z}(t)$. Similarly, if τ is too long compared to the timescale of chaotic behavior, the values of $z_0(t)$ and $z_0(t+\tau)$ will be incoherent with respect to each other. So a balanced choice of τ, perhaps as given by the first minimum of the average mutual information between them, is appropriate [30,31,33]. This line of reasoning regarding the selection of time delays applies here for the purposes of extracting additional information from our measurements.

However, it is important to recognize that our use of time delays is quite distinct from its role in nonlinear dynamics, in which one seeks independent coordinates that construct a proxy phase space to the underlying physical space using the measured variables. By contrast, our goal here is to use the information in the time delayed observations to inform a model about the state of the physical system representing the processes yielding the observations.

The argument regarding the number of components D_M is different as well. For phase-space reconstruction, the sufficient number of time delays needed to reconstruct the entire phase space can be determined geometrically. By contrast, in our application the time delays are used to control the unstable subspace of the dynamics, so the number of required time delays is a dynamical quantity, which should be less than or equal to the number of delays required to reconstruct the entire phase space.

Furthermore, our numerical examples will show that the number of required time delays is approximately equivalent to the dimension of the unstable subspace, averaged over a long trajectory. Next, however, we propose an extension of the synchronization or nudging technique described in Eq. (3) that incorporates information from time delayed measurements.

D. Synchronization using information from time delayed measurements

Following our definition of $\mathbf{y}(t_n)$ in Eq. (7), we construct the corresponding time delayed model state:

$$\mathbf{S}(\mathbf{x}(t)) := \{\mathbf{h}(\mathbf{x}(t)),\mathbf{h}(\mathbf{x}(t+\tau)),\ldots,\mathbf{h}(\mathbf{x}(t+(D_M-1)\tau))\}.$$

Its components may be written as

$$S_{k:l}(\mathbf{x}(t)) = h_l(\mathbf{x}(t+(k-1)\tau)),$$

where $l = \{1,2,\ldots,L\}$ and $k = \{1,2,\ldots,D_M\}$. In the framework we have described, we want the model output $\mathbf{S}(\mathbf{x}(t))$ to be equal to the data vector $\mathbf{Y}(t)$ as an indicator of synchronization between the data and the model output.

The time delay vector $\mathbf{S}(\mathbf{x}(t))$ is constructed from a map $\mathbf{x}(t) \to \mathbf{S}(\mathbf{x}(t))$, and thus satisfies the dynamical equation,

$$\frac{dS_{k:l}(\mathbf{x}(t))}{dt} = \sum_{a=1}^{D} \frac{\partial S_{k:l}(\mathbf{x}(t))}{\partial x_a(t)} F_a(\mathbf{x}(t)).$$

Setting aside for now worries about the details of the inverse map $\mathbf{S}(\mathbf{x}(t)) \to \mathbf{x}(t)$, this expression gives us a dynamical equation in $\mathbf{S}$ space:

$$\frac{dS_{k:l}(\mathbf{x}(t))}{dt} = \mathcal{F}_{k:l}(\mathbf{S}(\mathbf{x}(t))). \tag{8}$$

Following the idea in Eq. (3), we introduce a control term in $\mathbf{S}$ space whose role is to stabilize the chaotic motion using information expressed in $\mathbf{S}$ space:

$$\frac{dS_{k:l}(\mathbf{x}(t))}{dt} = \mathcal{F}_{k:l}(\mathbf{S}(\mathbf{x}(t))) + \sum_{l'=1}^{L}\sum_{k'=1}^{D_M} g'_{k:l,k':l'}(t)[Y_{k':l'}(t) - S_{k':l'}(\mathbf{x}(t))], \tag{9}$$

where $\mathbf{g}'(t)$ is a coupling gain matrix defined in $\mathbf{S}$ space. Mapping back to the physical space $\mathbf{x}(t)$ we arrive at

$$\frac{dx_a(t)}{dt} = F_a(\mathbf{x}(t)) + \sum_{a'=1}^{D} g_{a,a'}(t)\delta x_{a'}(t), \tag{10}$$

where, in matrix notation,

$$\delta\mathbf{x}(t) := \frac{\partial\mathbf{x}(t)}{\partial\mathbf{S}(\mathbf{x}(t))}\cdot\mathbf{g}'(t)\cdot[\mathbf{Y}(t)-\mathbf{S}(\mathbf{x}(t))],$$

and $\mathbf{g}(t)$ is an additional coupling matrix, defined in $\mathbf{x}$ space. As before, these coupling terms $\mathbf{g}(t)$ and $\mathbf{g}'(t)$ are localized pulses so their contribution is only active at times when measurements occur.

This equation displays the manner in which information from $\mathbf{Y}(t)$ is transferred to the model $\mathbf{x}(t)$ via the dynamical equations. This form of the dynamics is utilized throughout the measurement window to estimate the model output states and parameters $\mathbf{x}(t)$ required to match the data $\mathbf{Y}(t)$. When measurements are completed, we set the coupling matrices $\mathbf{g}(t),\mathbf{g}'(t)=0$ to predict for $t>T$ using the uncoupled dynamics (1).

The term $\partial\mathbf{x}/\partial\mathbf{S}(\mathbf{x})$ is a generalized inverse of the Jacobian $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ of the forward map to time delay space $\mathbf{x}(t)\rightarrow\mathbf{S}(\mathbf{x}(t))$. This Jacobian is constructed by integrating the variational equation [30] for the uncoupled dynamics (1):

$$\frac{d\Phi_{ab}(t',t_n)}{dt} = \sum_{c=1}^{D}\frac{\partial F_a(\mathbf{x}(t'))}{\partial x_c(t')}\Phi_{cb}(t',t_n),$$

$$\Phi_{ab}(t',t_n) := \frac{\partial x_a(t')}{\partial x_b(t_n)},\quad \Phi_{ab}(t_n,t_n)=\delta_{ab},$$

in the interval $[t_n,t_n+(D_M-1)\tau]$. This allows us to construct the Jacobian of the time delay model vector,

$$\frac{\partial S_{k:l}(\mathbf{x}(t_n))}{\partial x_a(t_n)} = \frac{\partial x_l(t_n+(k-1)\tau)}{\partial x_a(t_n)} = \Phi_{la}(t_n+(k-1)\tau,t_n).$$

Since $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ has dimensions $LD_M\times D$, it is not uniquely invertible and a generalized inverse must be used. For this paper, we use the pseudoinverse of this matrix; the details of its calculation will be given in the next section.

Also note that, for notational simplicity, here we have assumed that the measurements are projections of the state, $y_l(t_n)=h_l(\mathbf{x}(t_n))=x_l(t_n)$. To derive the more general expression, one would simply have to include the Jacobian of the measurement function $\mathbf{h}(\mathbf{x})$ in the definition of $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$.

ALGORITHEM 1. TIME DELAY SYNCHRONIZATION

for $n=\{0,1,\ldots,N\}$ **do**

1. COMPUTE $\mathbf{S}(\mathbf{x}(t_n))$ AND $\partial\mathbf{S}(\mathbf{x}(t_n))/\partial\mathbf{x}(t_n)$ VIA THE UNCOUPLED DYNAMICS.

2. COMPUTE THE COUPLING PERTURBATION $\delta\mathbf{x}(t_n)$,

$$\partial\mathbf{S}(\mathbf{x}(t_n))/\partial\mathbf{x}(t_n)\cdot\delta\mathbf{x}(t_n)=\mathbf{g}'(t_n)\cdot[\mathbf{Y}(t_n)-\mathbf{S}(t_n)].$$

3. TAKE A SMALL STEP VIA THE COUPLED DYNAMICS,

$$\mathbf{x}(t_{n+1})\leftarrow\mathbf{x}(t_n)+\Delta t[\mathbf{F}(\mathbf{x}(t_n))+\mathbf{g}(t_n)\cdot\delta\mathbf{x}(t_n)].$$

end for

The algorithm for determining the state $\mathbf{x}(t)$ within the observation window $0\leqslant t\leqslant T$ is outlined in Alg. (I). At each time step t_n, we compute the model time delay vector $\mathbf{S}(\mathbf{x}(t_n))$ and the Jacobian $\partial\mathbf{S}(\mathbf{x}(t_n))/\partial\mathbf{x}(t_n)$. The results are then used to evaluate the coupling perturbation $\delta\mathbf{x}(t_n)$. The process is repeated in this way, mapping back and forth between the physical and time delay spaces until the end of the observation window is reached.

Note that the integration time step Δt can be chosen much smaller than τ or the typical size of $t_{n+1}-t_n$ over the assimilation window. This may be desirable to achieve stability of the numerical scheme used for advancing the dynamics from a measurement time to the next measurement time.

It is also worth noting that, in the limit $D_M=1$, the time delay formulation (10) reduces to the standard nudging control (3). Several important differences, however, are realized when $D_M>1$:

(i) Information from the time delays of the observations is presented to the physical model equations.

(ii) The framework is easily extended to incorporate nonlinear measurement functions $\mathbf{h}(\mathbf{x})$.

(iii) The impact of the coupling terms is not limited to measurement times when $t=t_n$. All measurements within the current time delay window $[t,t+(D_M-1)\tau]$ can be incorporated, regardless of the current time step.

(iv) All components of the model state $\mathbf{x}(t)$ are influenced by the control term, not just the observed components. Consequently, the fixed parameters of the model may be estimated as a natural result of the synchronization process by including them as additional state variables, satisfying $F_a(\mathbf{x}(t))=0$.

(v) The time delay technique allows one to extract additional information from *existing measurements*.

The latter point is extremely important, as in many applications additional measurements may be prohibitively expensive, time consuming, or not technologically feasible. The benefits of using time delays will be displayed in further detail in the context of the numerical examples presented later in the paper. For the moment however, we divert our attention to a technical matter that is of crucial importance. Namely, the calculation of control term $\partial\mathbf{x}/\partial\mathbf{S}(\mathbf{x})$ as a regularized local inverse.

III. COMPUTING THE PSEUDOINVERSE OF $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$

We now discuss some of the details regarding the computation of the pseudoinverse $\partial\mathbf{x}/\partial\mathbf{S}(\mathbf{x}):=[\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}]^+$. We wish to solve the linear system of equations for $\delta\mathbf{x}$:

$$\frac{\partial\mathbf{S}(\mathbf{x})}{\partial\mathbf{x}}\cdot\delta\mathbf{x}=\delta\mathbf{S}(\mathbf{x}):=\mathbf{g}'\cdot[\mathbf{Y}-\mathbf{S}(\mathbf{x})],\tag{11}$$

where the explicit time dependence has been suppressed. We wish to determine the perturbation in physical space $\delta\mathbf{x}$ that produces the perturbation $\delta\mathbf{S}(\mathbf{x})$ in time delay space. This task may be formulated as an optimization problem that seeks to minimize a least squared objective function:

$$\left[\frac{\partial\mathbf{S}(\mathbf{x})}{\partial\mathbf{x}}\cdot\delta\mathbf{x}-\delta\mathbf{S}(\mathbf{x})\right]^2.$$

In general, $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ is a $LD_M\times D$ rectangular matrix and therefore its inverse is not unique; the system may be underdetermined or overdetermined depending on the choice of D_M.

The common solution for such ill-posed problems such as this is to include a *regularization* term in the objective function [34,35],

$$\left[\frac{\partial\mathbf{S}(\mathbf{x})}{\partial\mathbf{x}}\cdot\delta\mathbf{x}-\delta\mathbf{S}(\mathbf{x})\right]^2+[\Gamma\cdot\delta\mathbf{x}]^2.\tag{12}$$

This process, known as Tikhonov regularization, allows us to choose Γ to give preference for particular solutions with desirable properties. Here we choose $\Gamma = \alpha \mathbf{I}$ where $\mathbf{I}$ is a $D \times D$ dimensional identity matrix, which in the limit $\alpha \to 0$ recovers the expression for the Moore–Penrose pseudoinverse. In addition to being arguably the simplest choice for Γ, this form selects for solutions to Eq. (11) that minimizes the least squares norm of $\delta\mathbf{x}$. The regularization in Eq. (12) leads to the expression for $\delta\mathbf{x}$ of

$$\delta\mathbf{x} = \left[\Gamma + \frac{\partial \mathbf{S}(\mathbf{x})}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{S}(\mathbf{x})}{\partial \mathbf{x}}\right]^{-1} \cdot \frac{\partial \mathbf{S}(\mathbf{x})}{\partial \mathbf{x}} \cdot \delta\mathbf{S}(\mathbf{x}),$$

where only a square $D \times D$ matrix needs to be inverted. This choice agrees intuitively with the interpretation of $\delta\mathbf{x}$ as a perturbation control.

We do not imply that this choice is optimal. Indeed, optimality must depend on the specific problem and, more specifically, on the form of noise in the measurement vector $\delta\mathbf{S}(\mathbf{x})$. For instance, it is known that certain choices of $\Gamma(t)$ can implement low-pass filter properties which can be used to enforce smoothness of the solution. However, for the purposes of this paper and the numerical experiments herein, we focus on one approach: the pseudoinverse.

A. Computing the pseudoinverse with singular value decomposition

There are many numerical approaches available for constructing the pseudoinverse of an $m \times n$ matrix $\mathbf{M}$. The simplest choice involves the direct inversion of the matrix product,

$$\mathbf{M}^{+} = (\mathbf{M}^{T} \cdot \mathbf{M})^{-1}\mathbf{M}^{T}. \tag{13}$$

This technique is known to incur numerical stability problems, which become especially problematic when $\mathbf{M}$ is ill conditioned. The reason is that if $\mathbf{M}$ has condition number κ then the product $\mathbf{M}^{T} \cdot \mathbf{M}$ has condition number κ^2, and will be considerably more ill conditioned than $\mathbf{M}$.

An alternative approach that does not suffer from such instability involves a SVD of the matrix $\mathbf{M}$ [36]. A generalization of eigenvalue decomposition from square to nonsquare matrices, the SVD decomposes an $n \times m$ matrix $\mathbf{M}$ into a product of three matrices,

$$\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^{\dagger} \tag{14}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices of size $n \times n$ and $m \times m$, respectively, Σ is an $m \times n$ rectangular diagonal matrix of singular values σ_i, and $\mathbf{V}^{\dagger}$ denotes the conjugate transpose of the matrix $\mathbf{V}$. The SVD is unique up to permutations and sign exchanges of the singular values. Most algorithms choose the singular values to be positive and ordered such that $\sigma_1 > \sigma_2 \cdots > \sigma_{r_{\max}}$ where $r_{\max} = \min(m,n)$.

Once the SVD is known, the pseudoinverse can be constructed as

$$\mathbf{M}^{+} = \mathbf{V}\Sigma^{+}\mathbf{U}^{\dagger}. \tag{15}$$

where Σ^{+} is defined by taking the reciprocal of each nonzero element along the diagonal, leaving the zeros in place. In practice however, only elements larger than some small tolerance are taken to be nonzero, while the others are replaced by zeros. This choice of tolerance determines the rank of the inverse, which we will show plays a crucial role in the numerical stability of the algorithm and governs its overall performance. To this end, we now discuss methods for choosing the rank of the inverse.

B. Rank considerations

The default tolerance used in most linear algebra routines to compute the pseudoinverse, which is on the order of the machine precision, has proven to be insufficient for our purposes, as evidenced by our numerical experiments. Choosing such a small tolerance will lead to the inversion of very small singular values, which in turn produces excessively large control perturbations $\delta\mathbf{x}$, and these will quickly push the model system into an unstable regime, resulting in numerical overflow.

By significantly raising this tolerance [e.g., from $O(10^{-16})$ to $O(10^{-3})$] the calculations can be stabilized but its performance is markedly degraded, presumably because information about the unobserved states is being discarded. In practice, a smaller rank corresponds directly to a smaller control $\delta\mathbf{x}$. The balance, therefore, is between a large enough $\delta\mathbf{x}$ to synchronize the model states with the data and a small enough $\delta\mathbf{x}$ to keep the numerical methods stable.

We now explore some ideas for choosing the rank of the inverse. One option is to choose the inverse to have constant rank throughout the entire estimation process. This has several attractive features:

(i) If the rank is chosen conservatively small, the calculations are numerically stable.

(ii) It provides insight into its role in stabilizing the synchronization manifold. The rank of the inverse appears to be roughly equivalent to the number of measurements needed to achieve synchronization in twin experiments.

(iii) The pseudoinverse, which in general is a discontinuous operation, can be made continuous by specifying a constant rank. This is important because it allows the derivative of the inverse to be properly defined. This is necessary, for instance, to calculate the Lyapunov exponents of the error propagation, which are often used to prove convergence of optimal control techniques [37].

The main drawback with this choice is that it must be made conservatively enough to avoid numerical instability along the entire trajectory. While this global choice is not an issue in many circumstances, nonetheless, it discards useful information in areas of state space where numerical instability is less of a concern.

Through numerical experiments we have observed that the $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ matrix is more well conditioned in regions with higher local Lyapunov exponents. This makes some intuitive sense, as the degeneracy of $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ is due to the lack of independence among the various components of $\mathbf{S}(\mathbf{x})$, which in turn is related to the rate of information flow among the various state variables $\mathbf{x}(t)$. Larger local Lyapunov exponents indicate increased dynamical mixing among the physical states as well as improved conditioning of the $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ matrix. In other words, the optimal rank of the inverse fluctuates along the trajectory and the SVD method actually appears to perform better in regions where the dynamics are more locally chaotic.

We have considered algorithms for adaptively choosing the rank of the Jacobian to maximize the amount of information transferred by the control coupling, without causing numerical instabilities. One idea that has proved effective for this task imposes a continuity constraint on the solution $\mathbf{x}(t)$ by ensuring that some measure of magnitude of the control coupling $\delta\mathbf{x}$ is not too large relative to the corresponding magnitude of the unperturbed vector field $\mathbf{F}(\mathbf{x})$. This can be implemented in several ways depending on the choice of norm. For instance, selecting the L^2 norm and choosing a tolerance ϵ, we have

$$||\delta\mathbf{x}||_2 \leqslant \epsilon||\mathbf{F}(\mathbf{x})||_2 = \epsilon\left(\sum_{a=1}^{D} F_a(\mathbf{x})^2\right)^{1/2}.$$

Given positive singular values of $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ ordered as $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_{r_{\max}}$, where $r_{\max} = \min(D_M L, D)$, the choice of rank r can be expressed via the inequality

$$\begin{aligned} ||\delta\mathbf{x}||_2 &= \left|\left|\left(\frac{\partial\mathbf{S}(\mathbf{x})}{\partial\mathbf{x}}\right)^{-1}\delta\mathbf{S}\right|\right|_2 \\ &\leqslant \left|\left|\left(\frac{\partial\mathbf{S}(\mathbf{x})}{\partial\mathbf{x}}\right)^{-1}\right|\right|_2 ||\delta\mathbf{S}||_2 \\ &\leqslant \frac{||\delta\mathbf{S}||_2}{\sigma_r}. \end{aligned}$$

In this case, select the largest r such that

$$\frac{||\delta\mathbf{S}||_2}{\sigma_r} \leqslant \epsilon||\mathbf{F}(\mathbf{x})||_2,$$

to guarantee that $||\delta\mathbf{x}||_2$ does not grow too large with respect to the magnitude of the vector field $||\mathbf{F}(\mathbf{x}(t))||_2$.

Another useful choice involves the L^∞ norm,

$$\left|\left|\frac{\delta\mathbf{x}}{\mathbf{F}(\mathbf{x})}\right|\right|_\infty := \max_{1\leqslant a\leqslant D}\left|\frac{\delta x_a}{F_a(\mathbf{x})}\right| \leqslant \epsilon, \tag{16}$$

where the vector division is performed by component. This can be implemented by explicit calculation of the inverse and the corresponding control coupling. Starting with the rank $r = 1$, construct the control coupling using only the largest singular value σ_1, and check whether the expression in Eq. (16) holds. If this condition is true, increase the rank by one and perform the check again by using the inverse constructed from the two largest singular values. The process is then iterated until full rank is reached or the condition fails. In the latter case, the result from the previous iteration is used. Thus, the choice for r can be written compactly as follows:

$$r = \underset{1\leqslant r\leqslant r_{\max}}{\operatorname{argmax}}\left[\left|\left|\frac{\delta\mathbf{x}_r}{\mathbf{F}(\mathbf{x})}\right|\right|_\infty \leqslant \epsilon\right],$$

where $\delta\mathbf{x}_r$ is the control coupling constructed from the inverse of $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ containing the r largest singular values.

There are several advantages for choosing r based on the size of the perturbation relative to the dynamics. For instance, assuming the dynamics inherently stable it is reasonable to think that maintaining the modified derivatives on the same scale will keep the trajectory in a stable regime. Moreover, the L^∞ approach normalizes the effective threshold to account for the different state variables. This is important because the choice of rank should not depend on the units in which the dynamical equations are expressed.

We reiterate that these techniques are heuristic choices that, in the following numerical examples, have demonstrated improved performance over the constant rank approach. For these experiments, selecting $\epsilon \approx 10$ appeared to consistently stabilize the calculations, while selecting a high rank in regions of phase space where the time delay construction is better conditioned and its inverse is less unstable. We make no claims to the optimality of these suggestions.

Certainly, other good choices are available. For instance, selecting a low-pass operator (e.g., a difference operator or a weighted Fourier operator) for the Tikhonov matrix in Eq. (12) is known to enforce smoothness and may help combat the effects of measurement noise [35]. Another idea is to use L^1 norm for the regularization term in Eq. (12), so that

$$\left(\frac{\partial\mathbf{S}(\mathbf{x})}{\partial\mathbf{x}}\cdot\delta\mathbf{x} - \delta\mathbf{S}\right)^2 + ||\Gamma\cdot\delta\mathbf{x}||_1.$$

This formulation may be useful when the optimal control perturbation is sparse, as this choice of norm optimizes for sparsity and is related to recent developments in the theory of compressed sensing [38]. We have also yet to investigate using a nonuniform time delay. For instance, it may be possible to choose the delays adaptively to generate vectors via $\mathbf{x} \to \mathbf{S}(\mathbf{x})$ that are in some way "optimally" well conditioned.

Although interesting, these considerations are beyond the scope of this paper, which seeks to give a general introduction to the use of time delayed measurements in data assimilation. Thus, we turn now to some concrete numerical examples that illustrate the capability of the time delay synchronization technique.

IV. USING TIME DELAY INFORMATION IN EXAMPLES

We now illustrate these ideas and developments with examples that address the applicability of the time delay technique for state and parameter estimation of chaotic dynamical systems. Along the way we discover strengths and weaknesses, and we try to point out both.

We examine four model dynamical systems as test beds for our ideas. Three of these systems are small, well-investigated dynamical models: the Lorenz 1963, the Rössler 1979, and the Lorenz 1996 models [23,39,40]. In addition, we extend the analysis to network models with chaotic Colpitts oscillators at the nodes [41]. For each example, we demonstrate that the time delay control scheme extracts enough information from a single measured variable (i.e., a scalar time series) to achieve accurate estimates and predictions for the unobserved states and parameters of the system. This is a significant improvement over the standard $D_M = 1$ coupling procedure in Eq. (3), for which it will be shown that a single measured state component is in fact insufficient for most of the examples presented here.

To evaluate our technique we perform "twin" experiments, in which the data $\mathbf{x}^{\text{data}}(t)$ are generated from the same model used to perform the state and parameter estimation. This allows us to directly compare our estimates and predictions for *all* state components, not just those that are observed. In this case, we are able to calculate the physical or $\mathbf{x}$-space

synchronization error,

$$\mathrm{SE}_x^2(t) := \frac{1}{D}\sum_{a=1}^{D}\left[x_a^{\mathrm{model}}(t) - x_a^{\mathrm{data}}(t)\right]^2, \tag{17}$$

as a metric of the error between the model and data trajectories. Since our models are deterministic, as $\mathrm{SE}_x(t) \to 0$, the model will exactly reproduce the unobserved as well as the observed data.

In real experiments however, the unobserved states are unknown. In this situation, we instead use the synchronization error in **S** space:

$$\mathrm{SE}_s^2(t) := \frac{1}{D_M L}\sum_{l=1}^{L}\sum_{k=1}^{D_M}[Y_{k:l}(t) - S_{k:l}(\mathbf{x}(t))]^2, \tag{18}$$

and we argue that—for large enough D_M—it serves as a suitable indicator of convergence.

To illustrate the general applicability of our technique we present these examples as a series of "real" experiments by performing the assimilation as though the data had been collected from a partially observable system. No information from the unobserved variables was used to produce the state estimates. We only supplement the predictions, when needed for comparison, with data from the unobserved variables. In this way, we hope to convey the capability of our method in the context of actual experiments.

Before reporting the results of our numerical investigations, recall that we are solving the controlled or regularized model differential equations given in Eq. (10). Parameters are estimated by treating them as state variables with trivial dynamics $F_a(\mathbf{x}(t)) = 0$. All numerical integration was performed by using an explicit fourth-order Runge–Kutta algorithm. During an the assimilation window, measurements are available at every time step Δt. Unless otherwise specified, the coupling matrices $\mathbf{g}(t)$ and $\mathbf{g}'(t)$ are taken to be identity matrices when $0 \leqslant t \leqslant T$. For $t > T$, we predict by removing the control or coupling terms, so $\mathbf{g}(t), \mathbf{g}'(t) \to \mathbf{0}$ and no additional information is utilized from the measurements.

Since we are working with deterministic models, without model error, we do not require any of the probabilistic machinery discussed earlier. Although one may place the time delay method fully within the general path integral formulation [13], we do not do so here. Instead, we simply modify the dynamical equations with the control terms Eq. (10). After long enough time evolution, the states in the model will match the states of system and we take these as our initial conditions for prediction.

A. Lorenz 1963 model

We begin with the Lorenz 1963 [23] model whose equations of motion are given by

$$\begin{aligned}
\frac{dx_1(t)}{dt} &= p_1[x_2(t) - x_1(t)],\\
\frac{dx_2(t)}{dt} &= x_1(t)[p_2 - x_3(t)] - x_2(t),\\
\frac{dx_3(t)}{dt} &= x_1(t)x_2(t) - p_3 x_3(t),
\end{aligned} \tag{19}$$

where the parameters are chosen to be $p_1 = 10, p_2 = 60,$ and $p_3 = 8/3$.

To produce the data, we integrate these equations with a time step $\Delta t = 0.01$ for $t = [0,T]$ where $T = 10 = 1000\Delta t$ and select a measurement function $y(t) = h(\mathbf{x}(t)) = x_1(t)$; so $L = 1$. The initial conditions for both the physical system $\mathbf{x}^{\mathrm{data}}(0)$ and the model system $\mathbf{x}^{\mathrm{model}}(0)$ are chosen at random from a uniform distribution that roughly spans the size of the attractor. However, the initial condition for the observed component $x_1(0)$ was chosen to match the data.

1. *Estimating states only*

Fixing the parameters p_1, p_2, p_3 at the values used to generate the data, we perform our calculations using a constant, uniform coupling $g = 10$ so that $g\Delta t = 0.1$. The matrix $\mathbf{g}'(t)$ is taken to be unity and the pseudoinverse is constructed using full rank $r = r_{\max} = \min(D_M, D)$. The time delay is chosen to be $\tau = 0.1 = 10\Delta t$, which is consistent with the average mutual information criterion [30,31,33].

The estimation proceeds by numerically integrating the coupled equations Eq. (10) throughout the assimilation window $t = [0,10]$. Then, setting $\mathbf{g}(t), \mathbf{g}'(t) \to \mathbf{0}$, we continue the integration to predict for $t = [10,20]$.

The trajectory of the experimental synchronization error $\mathrm{SE}_s(t)$ throughout the assimilation window is shown in the top panel of Fig. 1 for $D_M = \{1,2,3\}$. Note in particular how the $D_M = 1$ coupling is insufficient to achieve synchronization. This result however, is not at odds with the work of Pecora and Carroll, whose synchronization scheme replaces $x_1(t)$ by $y(t)$ in the dynamical equations, corresponding to the limit $g \to \infty$ [22]. By increasing the coupling to $g = 100$ we provide enough control strength to synchronize the systems with no time delays, $D_M = 1$.

This result does, however, suggest that the addition of time delays provides stronger coupling with lower values of g. This can be seen by noting how choosing $D_M > 1$ generates rapid convergence of the experimental synchronization error, and that the system converges to a synchronized state considerably faster with $D_M = 3$ compared with $D_M = 2$. Furthermore, we have checked that selecting $D_M > 3$ does not further improve the convergence rate, which we suspect is due to the fact that choosing $D_M > 3$ does not increase the rank of the inverse. In this case, the system is observable enough so that $D_M = 3$ provides a basis of measurements that spans the entire $D = 3$ state space. These results provide a simple demonstration of how effectively the time delays transfer additional information from the unobserved states to stabilize the synchronization manifold.

The true test however, of any data assimilation scheme is its ability to predict the behavior beyond the assimilation window. To this end, the bottom panel of Fig. 1 shows the estimates and predictions for the observed state component $x_1(t)$ for each $D_M = \{1,2,3\}$. As expected, the predictions for $D_M = 1$ are poor whereas for $D_M = \{2,3\}$ they are exceptionally accurate throughout the entire prediction window. The fact that $D_M = \{2,3\}$ produced excellent predictions but $D_M = 1$ did not supports use of the **S**-space synchronization error (18) as an experimentally viable indicator of convergence. Since this is a twin experiment, we actually know all the "unobserved" data

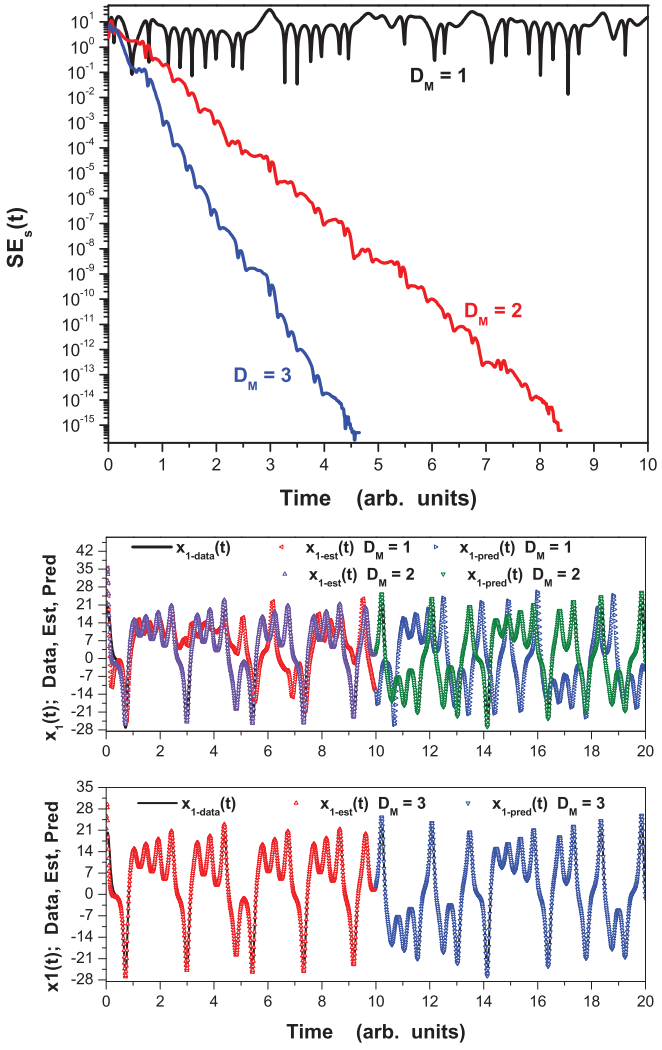

FIG. 1. (Color online) (top) Synchronization error $\mathrm{SE}_s(t)$ for state estimation in the Lorenz 1963 system with $D_M = \{1,2,3\}$ and parameters fixed to their true values. One time delay does not synchronize the systems because the coupling is too small. However, selecting $D_M = 2$ or 3 generates rapid convergence to synchronization of the model output $x_1(t)$ to $y(t)$. (bottom) Data, estimates, and predictions for the observed $x_1(t)$ component of the Lorenz 1963 model with $D_M = \{1,2\}$ (top subpanel) and $D_M = 3$ (bottom subpanel). In agreement with the top panel the estimates and predictions made with $D_M = 1$ are poor whereas with $D_M = \{2,3\}$ they are accurate.

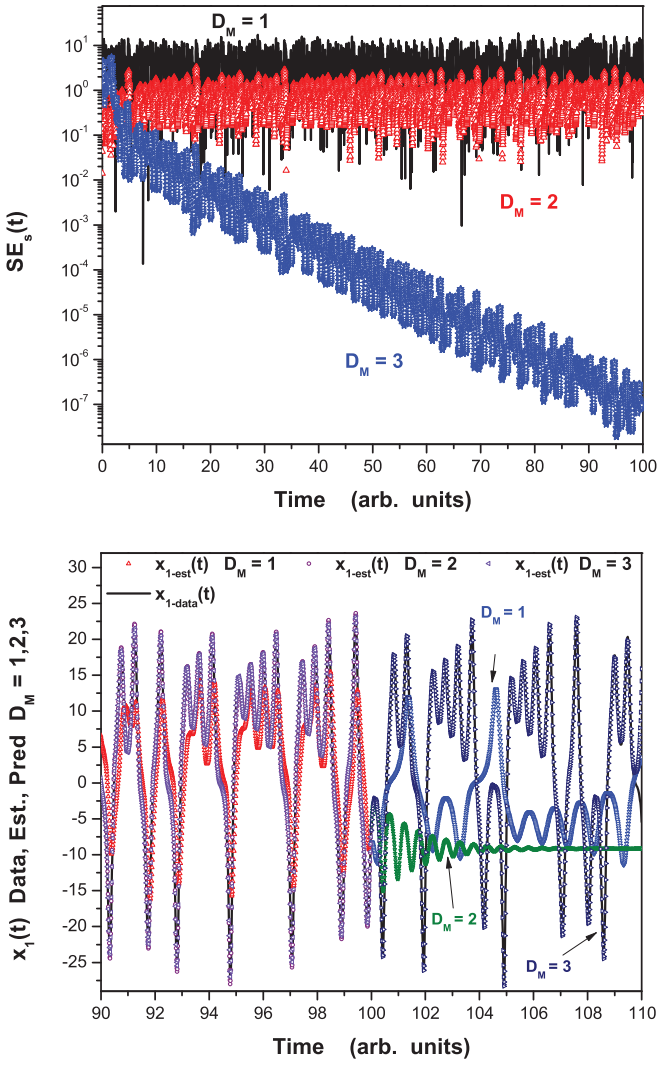

FIG. 2. (Color online) (top) $\mathrm{SE}_s(t)$ with $D_M = \{1,2,3\}$ for the extended Lorenz 1963 system where the parameters are treated as additional states. Three time delays are needed to synchronize. Including parameters can increase L_c. (bottom) Data, estimates, and predictions of the observed $x_1(t)$ for $D_M = \{1,2,3\}$ when parameters are included as state variables. Predictions made with $D_M = \{1,2\}$ are poor, but accurate with $D_M = 3$. Estimates for $D_M = 2$ match the data well but the predictions are not accurate, indicating the importance of using predictions (rather than "data fits") to validate the model.

time series, so we may verify the predictions of the unobserved state components directly. We have done this, and the results (not shown) confirm our comments.

2. *Estimating states and parameters*

Next, we estimate the parameters for this system by extending Eq. (19) to include the parameters as state variables. We now have six dynamical equations $\mathbf{p}(t) = \{p_1(t), p_2(t), p_3(t)\} = \{x_4(t), x_5(t), x_6(t)\}$ with $d\mathbf{p}(t)/dt = 0$. Only the time delay control appears in the vector field of the p_k. The initial values of the parameters are chosen to be 50% of their known values, and the coupling matrix is selected as $\mathbf{g}(t) = \mathrm{diag}(\{10,10,10,100,100,100\})$. That is, the parameters are subject to ten-fold larger coupling than the states. The assimilation proceeds as before, except over an extended observation window $T = 100 = 10^4 \Delta t$. The coupling is then turned off to predict for [100,110].

Trajectories of $\mathrm{SE}_s(t)$ are shown in the top panel of Fig. 2 for $D_M = \{1,2,3\}$. Synchronization proceeds more slowly than in the previous example where parameters are fixed. Notably however, for the extended system, $D_M = 2$ is no longer sufficient. This can be further established by examining the parameters at the end of the assimilation window (see Table I). As expected, for $D_M = 1$ the parameters have not changed from their initial values, because $D_M = 1$ coupling only perturbs the measured state components [here $x_1(t)$]

TABLE I. Estimated parameter values for the (extended, parameters treated as state variables) Lorenz 1963 model. The true values are $\mathbf{p} = \{10.0, 60.0, 2.667\}$.

D_M	Estimated p_1	Estimated p_2	Estimated p_3
1	5.0000	30.0000	1.3333
2	33.8313	25.2357	3.4764
3	10.0000	59.9999	2.6667

and is therefore unable to perform parameter estimation. For $D_M = 2$, the parameter estimates are poor and for $D_M = 3$ they are very accurate. In the latter case, the relative errors $\epsilon_i^{\text{rel}} := (p_i^{\text{model}} - p_i^{\text{data}})/p_i^{\text{data}}$ are all $O(10^{-6})$ or smaller.

Forecasts for the observed state variable $x_1(t)$ are shown in the bottom panel of Fig. 2. The estimates and predictions for $D_M = 1$ are not acceptable. Selecting $D_M = 2$ on the other hand, generates a very reasonable "fit" to the data during the assimilation window, but it results in poor prediction. This raises two important points:

(i) It illustrates our statement that the merit of any data assimilation scheme must be judged by its capability to predict, not just fit, the data.

(ii) For the extended system (with parameters included), selecting $D_M = 2$ is no longer sufficient to achieve synchronization. This suggests that promoting parameters into states with trivial dynamics can increase L_c.

Thus, we have demonstrated the capability of our method to successfully estimate the state and parameters of a simple Lorenz 1963 system. These results notwithstanding, this system is not so interesting from the standpoint of demonstrating the true power of this technique, since we know one measured state component is sufficient to synchronize the systems by using the $D_M = 1$ coupling method, provided the coupling gain is chosen high enough. The rest of our examples do not share this property and are thus more suitable for investigating the problem of assimilating data with an insufficient number of measurements.

B. Rössler hyperchaos

We now investigate the four-dimensional (4D) Rössler system described by [39]

$$\begin{aligned}
\frac{dx_1(t)}{dt} &= -x_2(t) - x_3(t),\\
\frac{dx_2(t)}{dt} &= x_1(t) + p_1 x_2(t) + x_4(t),\\
\frac{dx_3(t)}{dt} &= p_2 + x_1(t)x_3(t),\\
\frac{dx_4(t)}{dt} &= p_3 x_3(t) + p_4 x_4(t).
\end{aligned}$$

We generate a time series $\mathbf{x}^{\text{data}}(t)$ by using a time step of $\Delta t = 0.025$ starting from the initial condition $\mathbf{x}^{\text{data}}(0) = \{-20, 0, 0, 15\}$ with a parameter set $\mathbf{p}^{\text{data}} = \{0.25, 3.0, -0.5, 0.05\}$, for an observation window $T = 20 = 800\Delta t$. As in the previous example, we choose a measurement function $y(t) = h(\mathbf{x}(t)) = x_1(t)$, so $L = 1$.

To initiate our time delay algorithm, the three *unobserved* initial model conditions are selected randomly from a uniform distribution that spans the attractor, so that $\mathbf{x}^{\text{model}}(0) = \{-20, -18.6, 25.7, 122.4\}$.

Parameters are estimated by treating them as four additional state variables $\mathbf{p}(t) = \{x_5(t), x_6(t), x_7(t), x_8(t)\}$ with $d\mathbf{p}(t)/dt = 0$. The initial parameter estimates are selected to be $\mathbf{p}^{\text{model}}(0) = \{0.125, 1.5, -0.25, 0.025\}$; namely, 50% of the known values.

We encountered some initial problems with numerical stability, which we attribute to the fact that the 4D Rössler attractor is rather inhomogeneous. That is, the $x_3(t)$ state spends most of its time near zero but is punctuated by short excursions to relatively large values. To increase the stability of the computations we used the L^2 rank selection procedure described above, with $\epsilon = 10$, and imposed constraints on all parameters to keep them within the window $[-10, 10]$.

The calculations are carried out by using $\tau = 4\Delta t$ and a uniform coupling $g = 10$ so $g\Delta t = 0.25$. As before, $\mathbf{g}'(t)$ is taken to be unity. At the end of the observation window, the model parameters are fixed at their estimated values, and we then predict for a subsequent $200 = 8000\Delta t$ time units.

In the top panel of Fig. 3 we plot $\text{SE}_s(t)$ for $D_M = \{6, 8, 13\}$. For $D_M = 6$ synchronization does not occur whereas for $D_M = \{8, 13\}$ it does. The middle panel of Fig. 3 displays $\text{SE}_s(t)$ for $D_M = 8$ beyond the observation window. After the coupling is switched off the error grows at a rate that is roughly consistent with the maximum Lyapunov exponent of the system. The bottom panel displays the estimate (red) and prediction (blue) of the observed $x_1(t)$ along with the known data. Excellent predictions indicate good estimates of the unobserved states and parameters. The eventual deviation of the predictions from the known data is due to the chaotic behavior of the system.

Since this is a twin experiment, we may directly investigate the behavior of the unobserved states of the system. In the top panel of Fig. 4 we display the unobserved state $x_4(t)$. As expected, the estimates and predictions are quite good. A similar comparison for the parameter estimates is shown in the bottom panel. While the estimates may vary initially, they soon settle on the correct values. Numerical results for the parameter estimates are compiled in Table II. The values reported are the relative errors at the end of the observation window.

C. Lorenz 1996 model

We now turn to the example of the Lorenz 1996 model [40], which is studied widely in the geophysical literature [42]. The model describes a ring of $D > 3$ coupled oscillators, which obey the differential equations

$$\frac{dx_a(t)}{dt} = x_{a-1}(t)[x_{a+1}(t) - x_{a-2}(t)] - x_a(t) + p_1, \qquad (20)$$

where $a = \{1, 2, \ldots, D\}$ and the indices are permuted cyclically i.e., $x_0(t) = x_D(t), x_{D+1}(t) = x_1(t)$ and $x_{-1}(t) = x_{D-1}(t)$.

When the forcing parameter p_1 is large enough, this model exhibits extensive chaos so that the number of positive Lyapunov exponents scales with the number of spatial dimensions D [43]. Similarly, the number of measurements required

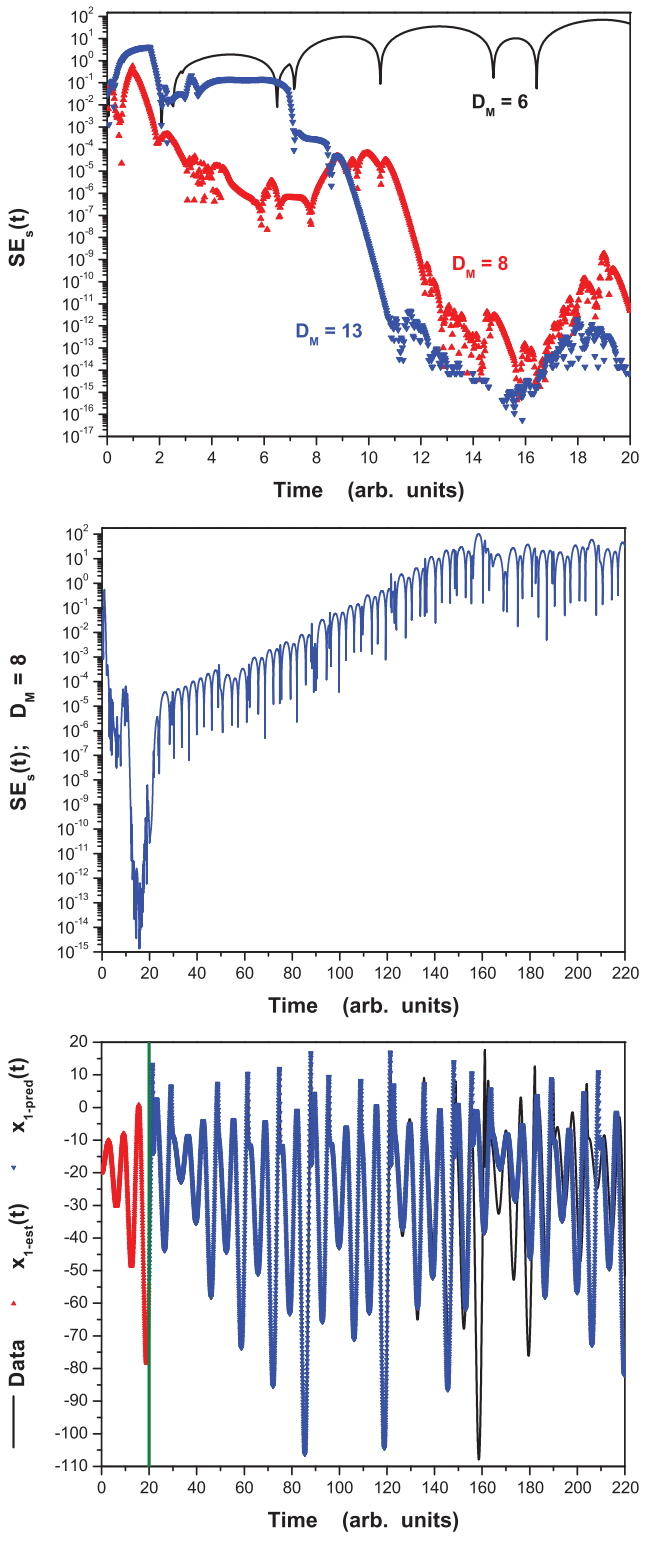

FIG. 3. (Color online) (top) $\mathrm{SE}_s(t)$ for $D_M = \{6,8,13\}$ for the Rössler system including parameters as state variables. (middle) Long $\mathrm{SE}_s(t)$ trajectory for $D_M = 8$. Trajectories begin to diverge immediately after the coupling is removed at $T = 20$. The rate of error growth is consistent with the largest Lyapunov exponent of the system. (bottom) Known (black), estimated (red), and predicted (blue) trajectories of the observed component $x_1(t)$. The prediction deviates from the data around $t \approx 160$ in agreement with the $\mathrm{SE}_s(t)$ results.

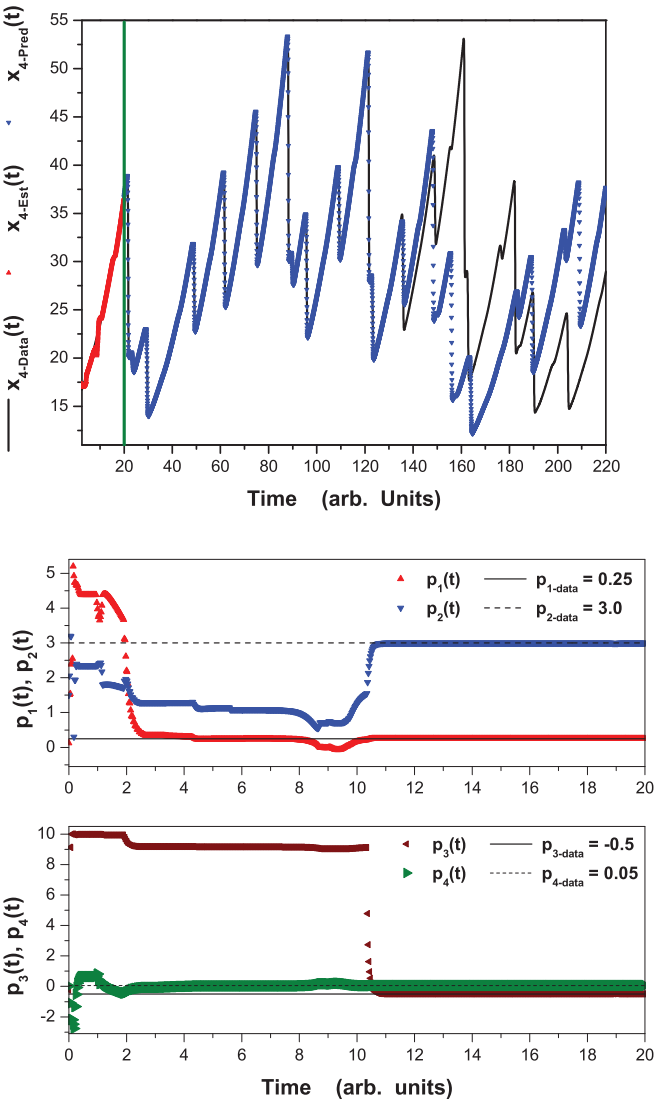

FIG. 4. (Color online) (top) Known (black), estimated (red), and predicted (blue) time series for an *unobserved* state variable $x_4(t)$ of the Rössler system with $D_M = 8$. The prediction fails near $t = 140$ due to the chaotic behavior of this system. Only in a twin experiment are we able to compare an unobserved state variable with known data. The initial condition for $x_4(0)$ in the model, as noted in the text, was 122.5. This was reduced to about 17 after about 100 time steps of $\Delta t = 0.025$. In the figure we started the time axis at $t = 2.5$ so the display was not compressed by the need to display the very large initial guess of x_4. (bottom) Estimates of the four unknown parameters of the Rössler system within the observation window. All parameters are bounded with a window $[-10,10]$ to improve numerical stability. All parameters converge to their known values.

to stabilize the synchronization manifold is also proportional to D. Previous work [14,24] has shown that, with a global forcing parameter $p_1 = 8.17$, the standard coupling scheme (3) involving one control term in the differential equations of each measured state requires approximately $L_c \approx 0.4D$ to achieve synchronization. Since the dimension D may be chosen freely, this makes the Lorenz 1996 system an excellent testing ground for investigating the behavior of data assimilation techniques in the context insufficient measurements. For our purposes, it

TABLE II. Relative error of the four parameter estimates for the Rössler system.

D_M	ϵ_1^{rel}	ϵ_2^{rel}	ϵ_3^{rel}	ϵ_4^{rel}
6	29.7088	0.4368	1.1004	46.0390
8	1.8877×10^{-11}	4.1588×10^{-9}	4.1174×10^{-8}	4.7842×10^{-10}
13	1.3742×10^{-12}	4.9737×10^{-12}	3.8792×10^{-10}	9.8734×10^{-12}

will further demonstrate how the time delay dimensions serve as additional measurements.

1. D = 20 with a single global forcing parameter

First, we look at a system of size $D = 20$ and extend it to include the single global parameter p_1 as a 21st state variable $x_{21}(t)$ with dynamics $dx_{21}/dt = 0$. We observe only the first state component $h(\mathbf{x}(t)) = x_1(t)$, so $L = 1$. Data is generated by using a time-step of $\Delta t = 0.01$. We select a constant coupling $g = 10$, so that $g\Delta t = 0.1$ and time delay $\tau = 0.1 = 10\Delta t$. The inverse, $\partial\mathbf{x}/\partial\mathbf{S}(\mathbf{x})$ is taken to have full rank $r = D_M$ and the parameters are not subject to any constraints.

This example also includes additive white noise in the measurement $y(t) \to y_1(t) + \eta(t)$. The noise is generated by choosing $\eta(t)$ from a uniform distribution centered around zero $U(-\alpha,\alpha)$. The amplitudes $\alpha = \{0.0, 6.34 \times 10^{-5}, 0.0011, 0.020\}$ (arbitrary units) are chosen so that the signal-to-noise ratios (SNRs) are, respectively, SNR $= \{\infty, 100, 75, 50\}$ dB, where for a uniform distribution,

$$\text{SNR} := 10\log_{10}\left(\frac{\langle y(t)^2\rangle - \langle y(t)\rangle^2}{|\alpha|^2/3}\right),$$

$$\langle x\rangle := \frac{1}{T}\sum_{n=1}^{T} x(t_n).$$

The estimate is performed by using the same data trajectory $y(t)$ for each of the noise amplitudes. Trajectories of the experimental synchronization error $\text{SE}_s(t)$ are shown in the top and bottom panels of Fig. 5 for SNR $= \infty$ and SNR $=$ 100 dB, respectively. Each plot includes traces for $D_M = \{1,8,9,10,12,14\}$.

When no noise is present, a clear transition to synchronization is evident between $D_M = 8$ and $D_M = 9$ for the extended system with 21 degrees of freedom. This allows us to identify $L_c \approx 9$, which is in agreement with previous work [24]. This rule also holds when the SNR $= 100$ dB. In this case, our results show the synchronization error quickly converges down to the approximate level of the noise. However, as the SNR is further decreased, this transition becomes less apparent. For SNR $= \{75, 50\}$ (not shown), the fluctuations of the synchronization error are roughly the order of magnitude of the noise.

Estimates and predictions for the observed variable $x_1(t)$ are shown in Fig. 6 for no added noise and in Fig. 7 for SNR $= 100$ dB. Here again, we see a clear distinction between the accuracy of the predictions between $D_M = 8$ and $D_M = 9$ when the noise levels are low, SNR $\geqslant 100$ dB. However, for higher noise levels SNR $\leqslant 75$ dB (not shown), the estimates are good but the predictions are poor regardless of D_M, indicating poor parameter estimates.

Since this is a twin experiment, we may check the parameter estimates directly. These results are shown in Table III. As expected the estimates for SNR $\geqslant 100$ are accurate when $D_M \geqslant 9$. However, as noise levels are further increased, the accuracy of the estimates deteriorates markedly. In this regime, increasing D_M seems detrimental to the parameter estimates.

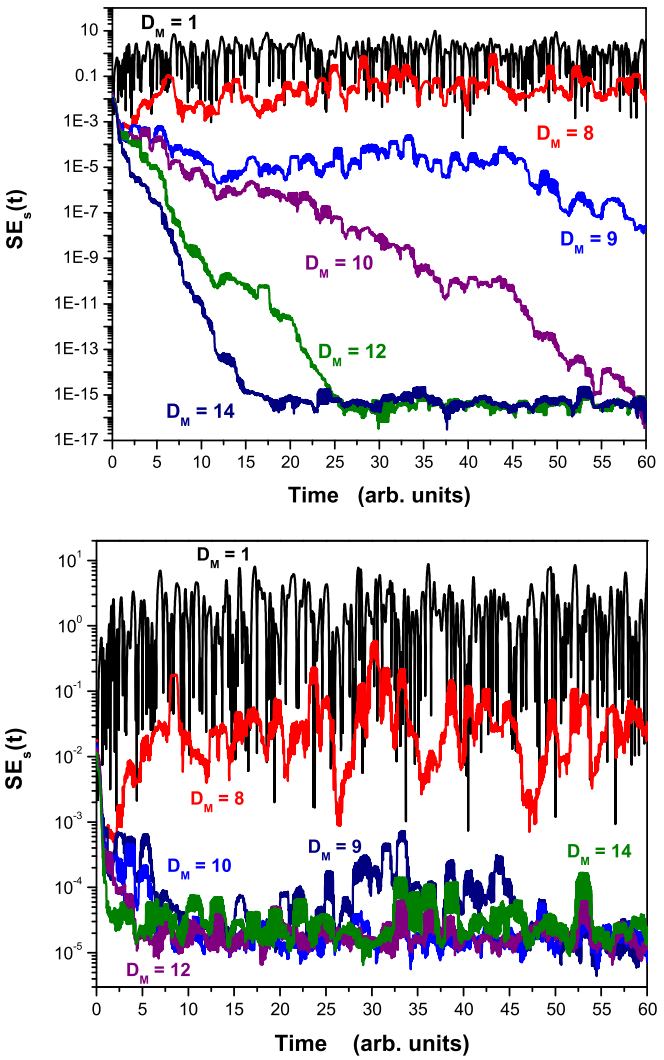

FIG. 5. (Color online) (top) $\text{SE}_s(t)$ for the Lorenz 1996 system with $D = 20$, and augmented with a single forcing parameter. No noise is added to the measured state $x_1(t)$ so that SNR $= \infty$. Traces are shown for various $D_M = \{1,8,9,10,12,14\}$. Synchronization is achieved with $D_M > 8$ allowing us to identify $L_c = 9$. (bottom) $\text{SE}_s(t)$ for the Lorenz 1996 system with $D = 20$ and augmented with a single forcing parameter. Uniformly distributed white noise is added to the measured state $x_1(t)$ so that SNR $= 100$ dB. Selecting $D_M > 8$ allows the systems to synchronize to within the level of the noise.

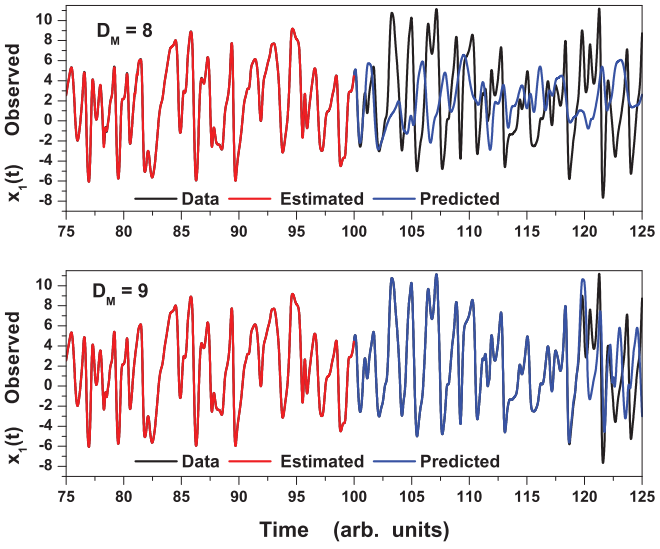

FIG. 6. (Color online) Estimates and predictions for the observed $x_1(t)$ for the Lorenz 1996 model with no additive noise (SNR $= \infty$). Traces are shown for (top) $D_M = 8$ and (bottom) $D_M = 9$. Note that, while $D_M = 8$ generates excellent estimates, the predictions are poor, indicating that unobserved states are not correctly determined. Selecting $D_M = 9$ however, produces quality predictions implying that the states and parameters are well estimated.

This sensitivity may indicate instability in the pseudoinverse. Indeed, we have checked that further decreasing the SNR causes the calculations to become unstable with $D_M > 10$.

Reducing the rank of the inverse stabilizes the calculations but does not improve the estimates in this case. However, we have seen evidence that results may be improved by choosing a larger D_M while fixing the inverse rank at a lower value to ensure stable calculations (e.g., $r = 10$). Increasing the time

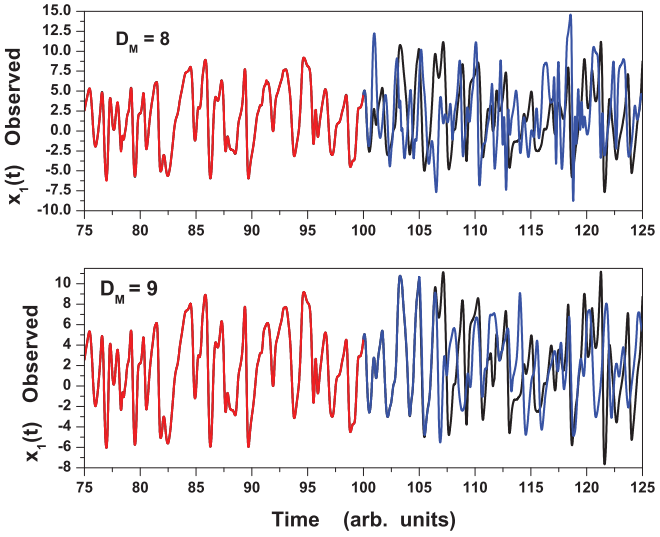

FIG. 7. (Color online) Estimates and predictions for the observed $x_1(t)$ for the Lorenz 1996 model with additive noise (SNR $= 100$ dB). Traces are shown for (top) $D_M = 8$ and (bottom) $D_M = 9$. As expected, selecting $D_M = 9$ produces good predictions, although not as good as the case with no noise.

TABLE III. Parameter estimates for the Lorenz 1996 model with $D = 20$ and a single, global parameter for various SNRs. The actual parameter value is $p_1 = 8.17$.

D_M	SNR $= \infty$	SNR $= 100$ dB	SNR $= 75$ dB	SNR $= 50$ dB
1	8.9259	8.9259	8.9259	8.9259
8	4.6297	10.4429	9.4413	10.4346
9	8.1700	8.1702	2.1007	−8.7913
10	8.1700	8.1718	2.3666	5.4988
12	8.1700	8.1707	9.6669	2.2544
14	8.1700	8.1701	12.3476	684.1818

delay τ has also been observed to improve robustness against noise as the addition of time delay coordinates tends to act as a low-pass filter. However, there is a tradeoff with this tactic. As the length of the time delay vector gets long with respect to the Lyapunov time, the inverse of the largest Lyapunov exponent, the $\partial \mathbf{S}(\mathbf{x})/\partial \mathbf{x}$ matrix becomes more ill conditioned and small errors in the data are amplified. Consequently, a good method for choosing the rank of the matrix is especially crucial when noise is involved and when the maximum time delay time $D_M \tau$ is long. Furthermore, we expect the noise robustness to be further improved by adapting the coupling terms $\mathbf{g}(t)$, $\mathbf{g}'(t)$ in some optimal manner that incorporates estimates for the error covariance, such as what is done for the Kalman–Bucy filter.

2. *D = 10 with different forcing for each oscillator*

Our next example uses the Lorenz 1996 model Eq. (20) with $D = 10$ and *different* values for the forcing parameters for each dimension ($p_1 \to p_i$ for $i = \{1, \ldots, D\}$). The values of these parameters are given in Table IV and are selected in this way to break the symmetry of the original model. Proceeding as usual, we construct the extended system consisting of $D + D = 20$ states and parameters to perform the estimates. All other parameters remain the same as the previous example. Also, no additional measurement noise was included in this simulation so SNR $= \infty$.

Figure 8 shows the temporal evolution of the synchronization error $\mathrm{SE}_s(t)$ for different delay dimensions D_M. While $D_M \leqslant 5$ is not sufficient for achieving synchronization,

TABLE IV. Estimated and known values for ten forcing parameters f_a in the Lorenz 1996 model with D $= 10$ and $D_M = 1, 5, 6, 10$.

Actual value f_a	Estimated $D_M = 1$	Estimated $D_M = 5$	Estimated $D_M = 6$	Estimated $D_M = 10$
5.7	6.198	5.349	5.700	5.699
7.1	8.059	7.100	7.100	7.100
9.6	9.940	3.879	9.597	9.599
6.2	6.785	−2.439	6.204	6.200
7.5	7.723	4.569	7.495	7.499
8.4	9.151	13.463	8.403	8.400
5.3	5.555	−0.003	5.295	5.300
9.7	10.205	−0.261	9.702	9.699
8.5	9.199	−12.887	8.499	8.500
6.3	7.190	8.955	6.299	6.300

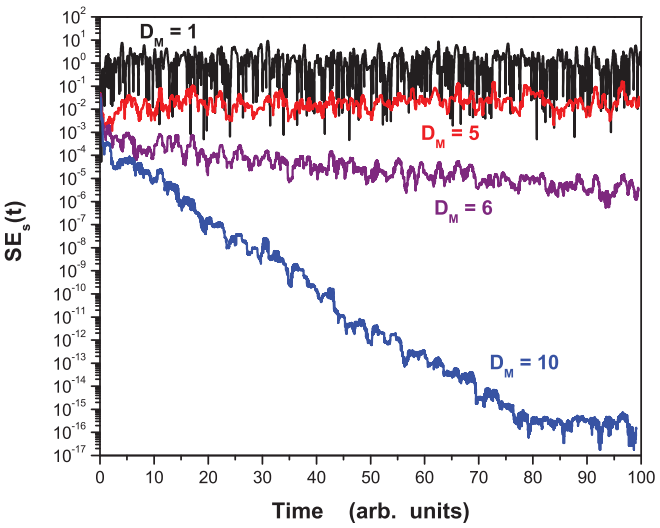

FIG. 8. (Color online) Experimental synchronization error $\mathrm{SE}_s(t)$ for the Lorenz 1996 model with $D = 10$ and different forcing p_a in each component for $D_M = \{1,5,6,10\}$. This shows that, in addition to the state variables, ten parameters may be estimated when $D_M = 10$.

the simulation with $D_M = 6$ shows a slow convergence to zero and $D_M = 10$ exhibits a clear and fast transition to synchronization. This allows us to identify $L_c \approx 6$ for the extended system. This is confirmed in Fig. 9 where in the top panel the predictions fail for $D_M = 1$ and $D_M = 5$, but succeed for $D_M = 10$ as shown in the bottom panel.

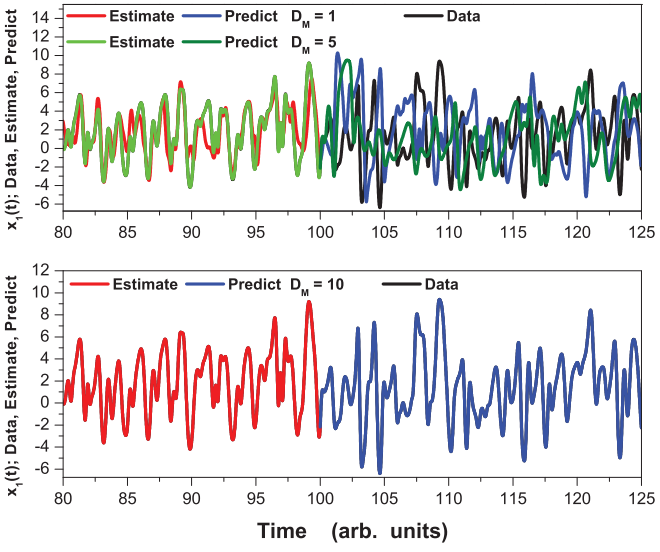

FIG. 9. (Color online) Estimate ($t < 100$) and prediction ($t > 100$) for $x_1(t)$ of the Lorenz 1996 model with $D = 10$ and different forcing p_a in each component during the synchronization shown in Fig. (8). (top) For $D_M = 1$ and $D_M = 5$ the estimation and the prediction is not good nor is the model output synchronized to the data. (bottom) For $D_M = 10$, we have excellent estimation and prediction.

V. DIRECT ESTIMATION OF L_c

We have now examined several examples of chaotic oscillators in which the use of additional information from the waveform of the data permits estimation of parameters and states when only $L = 1$ measurement is made at each observation time. In particular, we have seen that the time delays act in some sense as additional measurements and are able to reduce the number of measurements L required to achieve accurate estimates and reliable predictions. For instance, previous work with the Lorenz 1996 system showed that success in this endeavor requires $L \geqslant L_c \approx 0.4D$ measurements without time delays [14,24]. However, the results here show that success can be achieved using only $L = 1$ measurement as long as roughly $D_M \geqslant L_c$ time delays are used.

The fact that the critical number of time delays is approximately the same as the L_c is no accident. As we mentioned above, L_c is related to the number of unstable dimensions of the dynamics. We now give a technique for directly estimating this critical value.

Consider a long trajectory $\mathbf{x}(t_n)$ generated by Eq. (2) and sampled at discrete times $n = \{0,1,\ldots,N\}$. At each point t_n, evaluate $\partial\mathbf{F}(\mathbf{x}(t_n))/\partial\mathbf{x}$ and construct its SVD,

$$\frac{\partial\mathbf{F}(\mathbf{x}(t_n))}{\partial\mathbf{x}} = \mathbf{U}(t_n)\mathbf{S}(t_n)\mathbf{V}^{\dagger}(t_n).$$

Let $\{\sigma_1(t_n),\sigma_2(t_n),\ldots,\sigma_D(t_n)\}$ be the collection of singular values along the path. The local dimension of the unstable subspace is given by counting the number of singular values whose value is greater than unity. Consequently, a direct estimate for L_c can be obtained by averaging these values over the entire path. Specifically, the estimate is given by

$$L_c \approx \frac{1}{m+1}\sum_{n=0}^{m}\sum_{a=1}^{D}\Theta[\ln(\sigma_a(t_n))], \tag{21}$$

where $\Theta[\cdot]$ is the usual Heaviside theta function.

When this numerical technique is applied to the noiseless Lorenz 1996 system with a fixed, global parameter $p_1 = 8.17$ the $L_c \approx 0.4D$ scaling rule is reproduced. Applying this technique to the Lorenz 1996 system with $D = 10$ and 10 distinct parameters yields an estimate of $L_c \approx D$. This estimate, while not at odds with the above results, is a bit high, as we have observed synchronization with as low as $D_M = 6$. The transition with $D_M = 6$, however, takes much longer, as can be seen in Fig. 8.

The reason for this, we argue, is related to the fact that incorporating parameters into the model modifies the spectrum of the Jacobian $\partial\mathbf{F}(\mathbf{x}(t))/\partial\mathbf{x}$ to have singular values that are close to zero. These "slightly" unstable dimensions tend to get "averaged out" so to speak, when the assimilation window is long, allowing synchronization to occur with fewer than D measurements. Similar behavior was observed for the Lorenz 1963 and Rössler systems.

These results further strengthen our argument that L_c is closely related to the number of locally unstable directions in phase space, or more precisely, the ergodic average of this quantity. Also, the fact that the critical number of time delays is approximately equal to L_c supports the idea that to successfully synchronize the model with the observed data, one requires the

set of measurements (either physical or time delayed) to span the unstable subspace of the dynamics.

This idea of incorporating information from time delayed measurements to regularize the search for the correct model states and parameters is not new by any means. In particular, we have recently discovered that the method discussed here and in Ref. [17] (also proposed earlier in Ref. [25]) is fundamentally equivalent to a control theoretic construct known as a Newton observer, which was first introduced by Moraal and Grizzle in Refs. [44,45]. The idea is that by using time delays in this way, the perturbation $\delta\mathbf{x}(t)$ is essentially the Newton step associated with the observability equation. We elaborate this point in more detail below.

We begin with the standard definition of the nonlinear observability matrix $\partial\Phi(\mathbf{x})/\partial\mathbf{x}$, in which

$$\Phi(\mathbf{x}) := \begin{bmatrix} h(\mathbf{x}) \\ \mathcal{L}_F h(\mathbf{x}) \\ \vdots \\ \mathcal{L}_F^{D-1} h(\mathbf{x}) \end{bmatrix} \tag{22}$$

is the collection of repeated Lie derivatives $\mathcal{L}_F h(\mathbf{x}) = \mathbf{F}(\mathbf{x}) \cdot \nabla h(\mathbf{x})$ of the measurement function $\mathbf{h}(\mathbf{x})$ with respect to the vector field $\mathbf{F}$ [37,46]. Note that the use of Φ here differs from its use in section 2D to represent the variational matrix, though the two are related. The system is said to be *locally observable* at a point $\mathbf{x}_0$ if and only if

$$\mathbf{rank}\left[\frac{\partial\Phi(\mathbf{x}_0)}{\partial\mathbf{x}_0}\right] = D.$$

When the system is locally observable at a point $\mathbf{x}_0$, there exists a neighborhood Ω such that for every $\mathbf{z} \in \Omega$ the point $\mathbf{z} \neq \mathbf{x}_0$ is distinguishable from $\mathbf{x}_0$, in the sense that $\mathbf{h}(\mathbf{z}) \neq \mathbf{h}(\mathbf{x}_0)$. Intuitively, this means that, at the point $\mathbf{x}_0$, there is enough information from the measurement and the dynamics to infer the entire state of the true, physical system.

In principle one can perform this inversion locally, without having to use a dynamical process, by solving the following nonlinear system of equations for $\mathbf{x}$:

$$\mathcal{Y} := \begin{pmatrix} y \\ y^1 \\ \vdots \\ y^{D-1} \end{pmatrix} = \Phi(\mathbf{x}), \tag{23}$$

where $y^i := d^i y/dt^i$ are higher-order time derivatives of the measured data. This can be done for instance, with a Newton's method approach, which involves a series of iterates $\mathbf{x}^i$:

$$\mathbf{x}^{i+1} - \mathbf{x}^i = \left(\frac{\partial\Phi(\mathbf{x}^i)}{\partial\mathbf{x}^i}\right)^{-1} \cdot [\mathcal{Y} - \Phi(\mathbf{x}^i)].$$

For this process to succeed, the system must be locally observable so that the Jacobian $\partial\Phi(\mathbf{x})/\partial\mathbf{x}$ has full rank [37,46].

The vector-valued functions $\Phi(\mathbf{x})$ and $\mathbf{S}(\mathbf{x})$ are similar in that they both contain information about the time evolution of the states. In particular, $\mathbf{S}(\mathbf{x})$ can be considered a time delayed version of $\Phi(\mathbf{x})$. While $\Phi(\mathbf{x})$ is easier to work with analytically, performing the inversion of Eq. (23) is rarely useful in practice, because it requires one to measure high-order derivatives of the data or approximate them with finite differences. The latter approach is numerically unstable when measurement noise is present, as the finite difference approximation acts as a high-pass filter [46].

The time delay formulation on the other hand, does not have this problem since the derivatives on the left-hand side of Eq. (23) are replaced with time delayed values of the measurements. As Takens noted [28], time delays carry the same information as the derivatives but are far less sensitive to measurement noise. The same Newton's method approach can be performed using time delays,

$$\mathbf{x}^{i+1} - \mathbf{x}^i = \left(\frac{\partial\mathbf{S}(\mathbf{x}^i)}{\partial\mathbf{x}^i}\right)^{-1} \cdot [\mathbf{Y} - \mathbf{S}(\mathbf{x}^i)]. \tag{24}$$

Note that this process is *static*. That is, it is carried out at a single time t. Compare this with the *dynamic* process in Eq. (10), for which the control perturbation is essentially the right-hand side of Eq. (24). The immediate connection between the "observation space" $\Phi(\mathbf{x})$ and the time delay space $\mathbf{S}(\mathbf{x})$ suggests that the static process (24) can only converge to the correct solution when $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ has full rank.

In terms of the dynamical process Eq. (10), the observability criterion ensures that one can modify *all* of the eigenvalues of the error system

$$\mathbf{e}(t) := \mathbf{x}^{\text{model}}(t) - \mathbf{x}^{\text{data}}(t)$$

to converge at a desired rate [37,46]. In our numerical experiments, we observe precipitous drops in the synchronization error in regions where $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ is well-conditioned enough to construct the full-rank inverse. We consider this empirical evidence for the correspondence between our time delay approach and observability.

In addition, the connection with observability provides a different perspective on the time delay approach. Namely, at each time step we are solving a time delayed version of the observability (23) to estimate the error between the model and the data, which is then fed back into the model system after being modified by an appropriately chosen coupling (gain) $\mathbf{g}(t)$. When $D_M = 1$ the estimate uses only information available at the current time and when the inverse $\partial\mathbf{x}/\partial\mathbf{S}(\mathbf{x})$ is full rank the estimate provides full state feedback.

When the observability condition is not satisfied the static process fails. There is however, a weaker condition known as "detectability," which requires all of the unobservable modes of the system to decay asymptotically [37,46]. If this condition holds, the dynamical process will still succeed as we are able to control all of the locally unstable directions associated with error growth.

This is essentially what we mean by the suggestion that the set of measurements must span the unstable dynamical subspace. In nonlinear systems, however, the analysis is more difficult as this subspace changes dynamically in time, so that we may not always have a spanning set of measurements at each point along the trajectory. For our purposes, we are interested in an ergodic or "infinite horizon" estimation process where, although we may not be able to control all of the instabilities at every point, we nonetheless have enough measurements to initiate the transition to synchronization given a "long enough" time series of measurements, i.e., $T \to \infty$.

The purpose of this discussion has been to introduce a direct estimate Eq. (21) for L_c in terms of the average number of unstable directions in the dynamics and to acknowledge the apparent connection with observability. These ideas have had some mention in the data assimilation literature. For instance, the unstable dynamical subspace has been used for selecting ensemble members in ensemble forecasts and for identifying sensitive regions to target for further observation [42]. Also, optimization-based approaches such as moving horizon estimation seek to incorporate a moving time window of observations [47,48]. However, the true value of the Newton observer (time delay synchronization) technique lies in its ability to deal with poorly observable system in a systematic way [45]. It was with such systems in mind that we independently rediscovered the work of Moraal and Grizzle some two decades later, because these systems are altogether common in applications where the number of degrees of freedom in the model far exceeds the number of observations. With that said, we now turn to our final example, which involves the estimation of a small network of chaotic oscillators.

VI. NETWORKS OF CHAOTIC OSCILLATORS

One particular goal for our time delay method is to provide a means to analyze networks of oscillators, such as those found in nervous systems. As in practical geophysical dynamics (for example, numerical weather prediction) sparse measurements of the network behavior under selected forcing is to be expected. One strategy [11] for understanding the underlying physical properties of such problems is to analyze carefully the properties of the nodes; namely, the specific oscillators such as the ones we have covered here, and then use the same approach to analyze the nature and strengths of the couplings among the oscillators at the nodes to complete a model for the network as a whole.

In the case of nervous system networks, we have many neurons (nodes) connected by a variety of links (e.g., synaptic and gap junction). In practice, we cannot measure the detailed intracellular properties of more than one or a few of the nodes. If we, however, have determined the biophysics of each node from the analysis of isolated neurons, we require a tool to allow the estimate of the connectivity so the functional behavior of the network can be quantified.

A. A small network of chaotic Colpitts oscillators

Pursuing this goal, we examine a small network of well-studied chaotic oscillators. Each of the $M = 3$ nodes shown in Fig. 10 is a Colpitts oscillator that is forced by a voltage across a known circuit. A chaotic regime of behavior is reached from a fixed point for each oscillator through a bifurcation sequence including a limit cycle.

In particular, we investigate a ring of oscillators with connected with unidirectional coupling. The state of each oscillator is given by $x_a^i(t)$, where $i = \{1,2,3\}$ is the node index and $a = \{1,2,3\}$ denotes three internal state variables

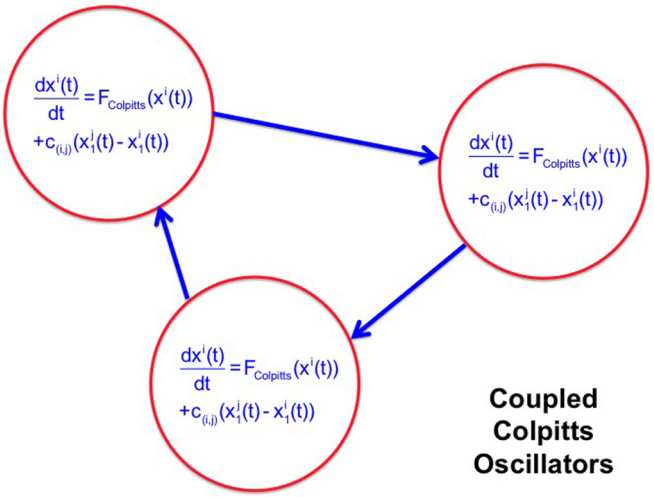

FIG. 10. (Color online) Diagram of a unidirectionally coupled network of three Colpitts oscillators.

for each node. The dynamical equations are given by

$$\begin{aligned}
\frac{dx_1^{(i)}(t)}{dt} &= p_1^{(i)} x_2^{(i)}(t) + c_{(i+1,i)}\big(x_1^{(i+1)}(t) - x_1^{(i)}(t)\big),\\
\frac{dx_2^{(i)}(t)}{dt} &= -p_2^{(i)}\big(x_1^{(i)}(t) + x_3^{(i)}(t)\big) - p_3^{(i)} x_2^{(i)}(t),\\
\frac{dx_3^{(i)}(t)}{dt} &= p_4^{(i)}\big(x_2^{(i)}(t) + 1 - \exp\big[-x_1^{(i)}(t)\big]\big),
\end{aligned} \tag{25}$$

where the indices are permuted cyclically so that $\mathbf{x}^{(M+1)}(t) = \mathbf{x}^{(1)}(t)$. The parameters $c_{(i+1,i)} \geqslant 0$ are constant coupling constants that serve as connections among the individual oscillators.

The Colpitts oscillator is comprised of standard R, L, C components together with a single bipolar transistor. The only nonlinearity is the exponential function $\exp[-x_1^{(i)}(t)]$ coming from the transistor dynamics. These equations are a rescaled representation of the physical equations of state. The derivation of these dynamical equations from Kirchoff's laws is given in Refs. [14,49]. The states x_1, x_2, and x_3, respectively correspond to the voltage between the transistor emitter and its base, the current through the inductor and the voltage at the transistor collector and its base.

When $p_1 \geqslant 3.5$ or so, the oscillator expresses chaotic behavior. Following Ref. [14], we select $p_1^{(i)} = 5.0$, $p_2^{(i)} = 0.0797$, and $p_4^{(i)} = 0.6898$ for all three oscillators. To break the ring symmetry, we select $p_3^{(1)} = 3$, $p_3^{(2)} = 3.5$, $p_3^{(3)} = 4$ as well as $c_{(2,1)} = 0.8$, $c_{(3,2)} = 0.9$, $c_{(1,3)} = 1.0$. Direct integration of Eq. (25) confirms that the individual oscillators do not synchronize with each other. This is important, because a synchronized network may require fewer measurements than an unsynchronized network. Indeed, synchronization of oscillators in a network may allow population behaviors by effectively reduce the degrees of freedom of the network in a functional manner.

B. Estimating the states of the network

To begin we fix all parameters and the couplings among oscillators to their known values, and use the time delay method to estimate the state of the network system given only the scalar time-series $y(t) = h(\mathbf{x}(t)) = x_1^{(1)}(t)$, so $L = 1$. A constant time delay $\tau = 0.2 = 20\Delta t$ and coupling gain $g\Delta t = 0.1$ were selected. To improve numerical stability during the transient period, the L_2 adaptive rank algorithm was used with a tolerance $\epsilon = 10$. Initial conditions for the model were chosen at random from an arbitrary trajectory on the attractor.

Results for the state estimation procedure are shown in the top panel of Fig. 11. The experimental synchronization error is plotted as a function of time for $D_M = \{1,3,5,9\}$. $D_M = \{1,2\}$ is insufficient; one needs $D_M = 3$ to achieve synchronization. Our analysis of estimates and predictions for individual states verified that $D_M = 3$ indeed produces excellent predictions, whereas $D_M = \{1,2\}$ does not. This result gives an estimate of $L_c \approx 3$ for the case under consideration, where only the states are to be determined.

Furthermore, the rate of convergence does not increase monotonically with the number of measurements. That is, $D_M = \{4,5\}$ have a slower convergence rate than $D_M = 3$, and $D_M = 6$ does not appear to converge at all. This illustrates the importance of the proper choice of D_M, as there is a tradeoff between the rate of convergence and the stability of the procedure. Note that the adaptive rank algorithm did not impact this result because, apart from about 100 time steps at the beginning of the assimilation window, full rank was used, i.e., $r = \min(D, D_M)$.

In addition, note that the fastest rate of convergence is achieved with $D_M = 9$. This choice is a special case where $D_M = D$ and so the $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ matrix is square and may be inverted exactly. While, theoretically, such an embedding allows the entire state to be reconstructed instantaneously at a single time t, in practice the matrix is often too ill conditioned for this technique to be of use. The adaptive rank algorithm counteracts this numerical instability by selecting the largest rank r that produces a stable perturbation. In this case however, we observe that, as the estimated state approaches the true value, the adaptive rank algorithm selects a full-rank inverse $r = D_M = D$. This indicates that the time delay construction is well-conditioned enough so that the exact inverse can be used to generate a perturbation $\delta\mathbf{x}$ that is small relative to the dynamics. When this happens, the estimate converges remarkably quickly to the true result, as evidenced by the steep dive for the $D_M = 9$ trace in the top panel of Fig. 11.

We argue that this accelerated convergence brought about by the full-rank inverse of $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ is intrinsically related to the observability condition familiar from control theory. As we suggested earlier, the $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ can be considered a time delayed version of the observability matrix. When a well-conditioned, full-rank inverse exists, the error between the true and estimated states is well approximated by $\delta\mathbf{x}$ and the estimate converges quickly to its true value. Thus, while $D_M = 3$ time delays appears to be necessary to stabilize the chaotic subspace of the dynamics to provide asymptotically stable convergence, selecting $D_M = 9$ provides rapid convergence that is less numerically stable.

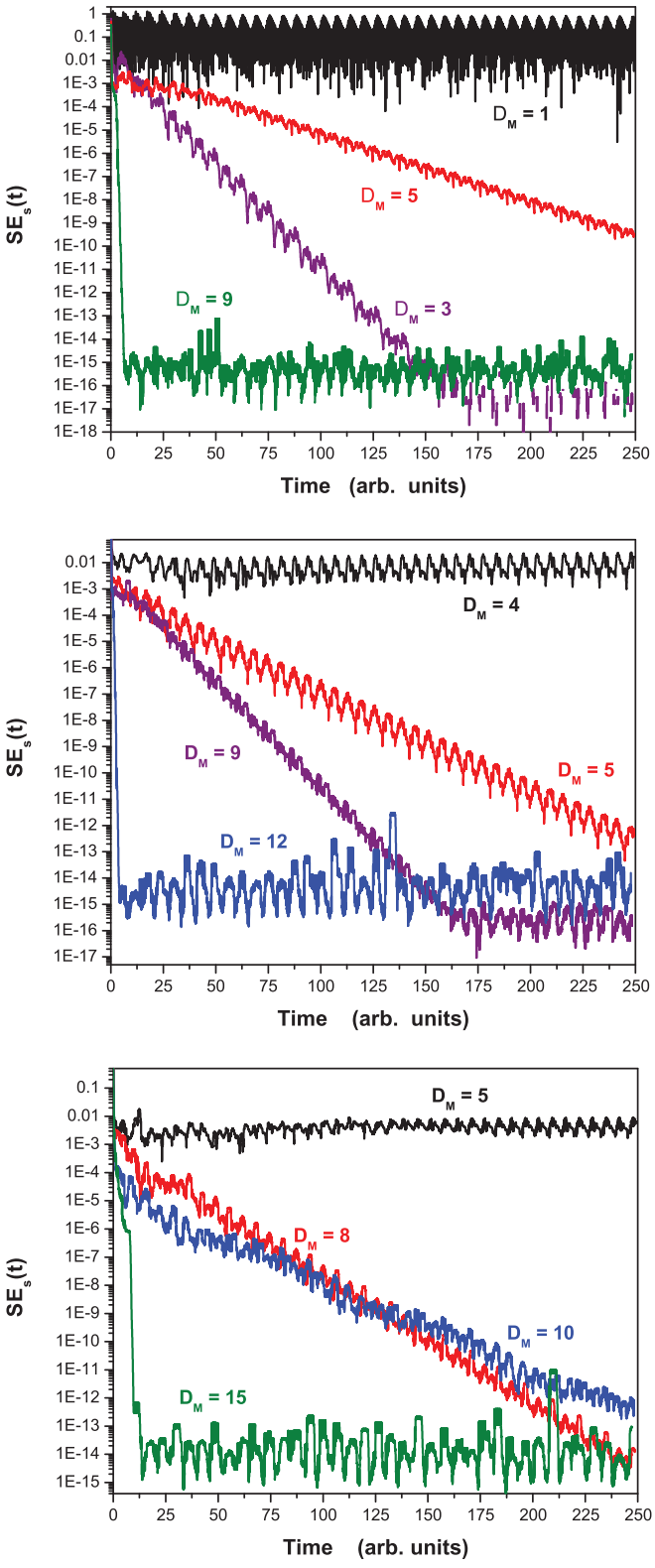

FIG. 11. (Color online) (top) $\mathrm{SE}_s(t)$ for state estimates of a network of three Colpitts oscillators. All model and coupling parameters are fixed to their true values and the network topology is known. (middle) $\mathrm{SE}_s(t)$ for state and ring coupling parameter estimates. Model parameters are not estimated and the network topology is known. (bottom) $\mathrm{SE}_s(t)$ for state and ring coupling parameter estimates. The model has been expanded to include backwards couplings, so the network topology is estimated as well.

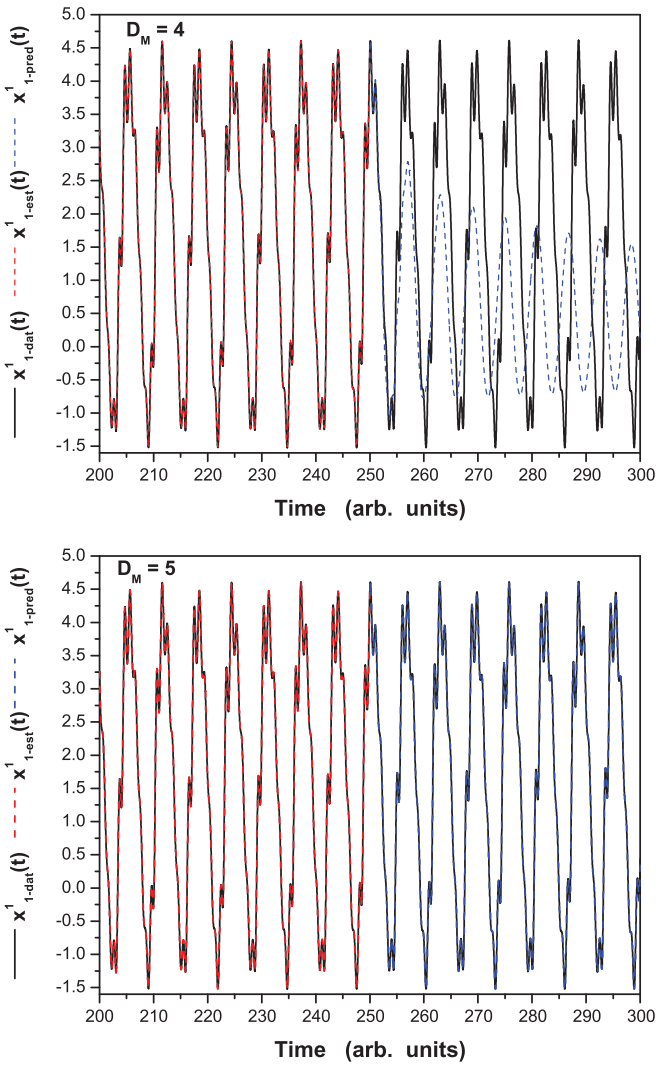

FIG. 12. (Color online) Ring of three Colpitts oscillators. (top) Known (black), estimated (red), and predicted (blue) trajectories for the observed $x_1^{(1)}(t)$ state component with $D_M = 4$. Although the estimate is quite good, poor predictions confirm that $D_M = 4$ time delays do not provide enough information to successfully estimate the state of the system as well as the ring coupling parameters. (bottom) Estimated and predicted trajectory for the observed $x_1^{(1)}(t)$ state component with $D_M = 5$. As anticipated from the synchronization error results, the estimates and predictions are quite accurate, indicating that the estimation procedure was successful.

Furthermore, depending on the system being studied it appears that it is not always possible to construct such a well-conditioned full-rank time delay space. In particular when parameters are being estimated, different parameters only may be observable within disparate regions of phase space and thus our localized time delay vector will not be able to capture the behavior of all parameters at a given point on the attractor.

C. Estimating the states and the couplings between nodes

Next, we fix the parameters $\mathbf{p}^{(i)}$ to their known values and include the internode couplings $c_{(i+1,i)}$ in the estimation procedure. This is directly relevant to analyses of neuron networks where we may have some knowledge of the cells individually, but we wish to explore the connectivity which underlies the function of the network.

We proceed as before, using the same time delay $\tau = 0.2 = 20\Delta t$ and coupling gain $g\Delta t = 0.1$ as well as the adaptive rank selection with $\epsilon = 10$. The initial conditions for the ring coupling parameters are chosen to be one half of their correct values, i.e., $c_{(2,1)} = 0.4$, $c_{(3,2)} = 0.45$, $c_{(1,3)} = 0.5$.

In the middle panel of Fig. 11, the experimental synchronization error is plotted as a function of time for $D_M = \{4,5,9,12\}$. Results show that $D_M = 5$ time delays are required to achieve synchronization. The $D_M = 1$ case was not computed here, because the ring coupling parameter estimates are guaranteed to be incorrect without the use of time delays. Trajectories with $D_M = 3$ and $D_M = 6$ were also computed, but not shown as the results proved to be numerically unstable. Also, here again we see that a full-rank $r = D_M = D = 12$ inverse is available, because the $D_M = 12$ trajectory synchronizes very rapidly.

As in previous sections, we validate the results as if this were an actual experiment. In the top and bottom panels of Fig. 12, we plot the estimate and predicted trajectories of the observed $x_1^{(1)}(t)$ state component for $D_M = 4$ and $D_M = 5$, respectively. As expected, the prediction for $D_M = 4$ is poor despite that its estimate looks quite accurate. This once again demonstrates the necessity of using predictions to validate the quality of a model's consistency with experimental results. On the other hand, the $D_M = 5$ estimate produces accurate predictions that do not diverge for a considerable time after the end of the assimilation window [largest Lyapunov exponent is $O(10^{-1})$]. Similar results are obtained for $D_M = 9$ and $D_M = 12$, although these trajectories are not shown.

Estimated values for the ring coupling parameters are shown in Table V. As expected, the estimates are accurate only when $D_M \geqslant 5$. This result demonstrates the potential of the time delay procedure for performing state and parameter estimation on a network of coupled chaotic oscillators. Specifically, it shows that the waveform of a scalar signal from a state of a single oscillator carries enough information to determine both the states of the neighboring nodes in the network as well as the coupling parameters that determine the interaction between the neighbors. This of course assumes that the network topology is known. In the next section, we dispense with this assumption and investigate whether the algorithm is capable of determining the functional connectivity of this simple network.

TABLE V. Estimated ring coupling parameters for a network of three Colpitts oscillators with known topology. True parameter values are $c_{(2,1)} = 0.8$, $c_{(3,2)} = 0.9$, $c_{(1,3)} = 1.0$.

D_M	Estimated $c_{(2,1)}$	Estimated $c_{(3,2)}$	Estimated $c_{(1,3)}$
4	4.7370	2.5639	1.4645
5	0.8000	0.9000	1.0000
9	0.8000	0.9000	1.0000
12	0.8000	0.9000	1.0000

D. Estimating the functional connectivity of the network

Until this point, our model has been constructed by assuming that the connectivity of the network is known, but the strength of the connections is not. In many practical applications however, this information is not available. For instance, when modeling small neurobiological networks, one often has some idea of the number of nodes in the network, and perhaps even some notion of their physical connectivity. The functional connectivity of the network however (i.e., the relative synaptic strengths) is generally unknown and therefore must be determined from experimental data. We now investigate this prospect of network topology estimation, within the context of our simple Colpitts network.

To this end, we expand our network model (25) to include coupling in both directions so that the dynamical equations for the $x_1^{(i)}$ states are now, with $c_{(j,i)} \geqslant 0$,

$$\frac{dx_1^{(i)}(t)}{dt} = p_1^{(i)} x_2^{(i)}(t) + \sum_{j=1}^{3} c_{(j,i)}\left(x_1^{(j)}(t) - x_1^{(i)}(t)\right).$$

The twin experiment data is generated as before, so that the true values of the additional coupling parameters are $c_{(1,2)} = c_{(2,3)} = c_{(3,1)} = 0$. The initial conditions for these couplings are chosen to be symmetric, so that $c_{(2,1)} = c_{(1,2)} = 0.4$, $c_{(3,2)} = c_{(2,3)} = 0.45$, $c_{(1,3)} = c_{(3,1)} = 0.5$. All other parameters are the same as before.

Also, note that all self-couplings $c_{(i,i)}$ are implicitly zero. This however, is just a consequence of how the network coupling model was defined; the procedure may be easily generalized to estimate self-coupling parameters as well.

In the bottom panel of Fig. 11, we display experimental synchronization error trajectories for $D_M = \{5,6,8,10,12,15\}$. Results are similar to those shown in the previous section. Synchronization requires $D_M \geqslant 8$ time delays and the full-rank $r = D_M = D = 15$ inverse synchronizes rapidly. Results with $D_M = 7$ were numerically unstable and are not shown. Known, estimated, and predicted trajectories for the observed $x_1^{(1)}(t)$ are shown in the top and bottom panels of Fig. 13 for $D_M = 6$ and $D_M = 8$, respectively. As anticipated from the synchronization error results, the prediction for $D_M = 8$ is quite accurate whereas for $D_M = 6$ it is not. The estimated coupling parameters shown in Table VI further confirm this result. Only the estimates made with $D_M \geqslant 8$ time delays are accurate, allowing us to identify $L_c \approx 8$ for this extended model, in which the connectivity of the network is unknown.

The main point of this calculation was to demonstrate that the time delay method is capable of determining the function connectivity of a network of chaotic oscillators, within the assumption that the model is known: both for the internal dynamics within a node and the coupling between nodes. In particular, we have shown that knowing $x_1^{(1)}(t)$ alone is enough to determine the functional connectivity of this small network of three Colpitts oscillators. That is, we are able to correctly estimate the values of both the forward and backward couplings, the latter of which are zero. Furthermore, we have learned that expanding the model in this way (i.e., to include coupling in both directions) increases the number of required time delays from $D_M = 5$ to $D_M = 8$.

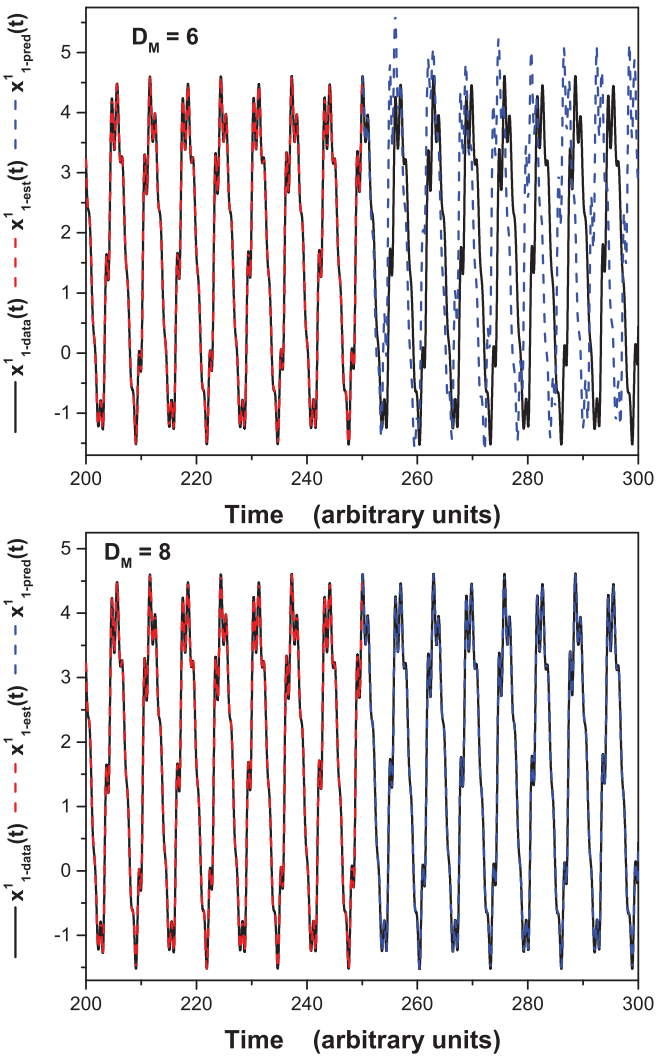

FIG. 13. (Color online) (top) Estimated and predicted trajectory for the observed $x_1^{(1)}(t)$ state component with $D_M = 6$ using the expanded network model that includes backwards coupling (unknown network topology). Although the estimate is quite good, poor predictions confirm that $D_M = 6$ time delays do not provide enough information to successfully estimate the state of the system as well as the ring coupling parameters. (bottom) Estimated and predicted trajectory for the observed $x_1^{(1)}(t)$ state component with $D_M = 8$. As anticipated from the synchronization error results, the estimates and predictions are quite accurate, indicating that the estimation procedure was successful.

We remark however, that this case of $M = 3$ is exceedingly simple, requiring estimation of only three additional parameters (the backwards couplings). For a general network, the number of coupling terms to be estimated grows as M^2 and thus we expect that, at some M, a single, scalar measurement will not be enough.

Nonetheless, the twin experiment framework presented here, together with the time delay algorithm, supplies crucial information about the observability of the system being studied. It provides, for instance, a lower bound estimate on the number of required measurements and also offers a way to determine which nodes should be targeted for observation:

TABLE VI. Estimated ring coupling parameters for a network of three Colpitts oscillators in which the network topology is unknown. True values are $c_{(2,1)} = 0.8$, $c_{(3,2)} = 0.9$, $c_{(1,3)} = 1.0$, and $c_{(1,2)} = c_{(2,3)} = c_{(3,1)} = 0$. Results show that with $D_M \geqslant 8$, the connectivity of the network is successfully predicted because the backwards couplings are estimated to be zero.

D_M	Estimated $c_{(2,1)}$	Estimated $c_{(3,2)}$	Estimated $c_{(1,3)}$
5	7.3325	0.5190	2.3770
6	0.8968	3.0695	1.3499
8	0.8000	0.9000	1.0000
10	0.8000	0.9000	1.0000
12	0.8000	0.8999	1.0000
15	0.8000	0.9000	1.0000
D_M	$c_{(1,2)}$	$c_{(2,3)}$	$c_{(3,1)}$
5	2.3385	-2.8119×10^{-1}	3.6094
6	-1.9911	5.2784×10^{-1}	3.9568×10^{-1}
8	3.8706×10^{-11}	-3.3491×10^{-11}	2.1787×10^{-12}
10	-8.3628×10^{-10}	1.0565×10^{-9}	2.0178×10^{-10}
12	1.0308×10^{-5}	-9.8415×10^{-6}	-1.4026×10^{-7}
15	-3.5215×10^{-10}	3.0340×10^{-10}	2.9756×10^{-12}

since some nodes may provide more data than others. Such information would be highly beneficial for experimental design purposes, as it allows one to directly investigate the constraints imposed by one's limited measurement capability. In other words, our framework allows one to determine, in principle, whether enough measurements are available to successfully determine the connectivity of the network, and predict its subsequent behavior.

VII. DISCUSSION AND SUMMARY

The idea of using the waveform of measurements—that is, the measurement at time t and its time delays—has been investigated in the context of synchronization-based state and parameter estimation for chaotic dynamical systems as a means to transfer additional information from observed data to a model. An algorithm has been presented that uses this additional information to generate dynamical coupling between the data and model systems and its capability has been demonstrated by using the Lorenz 1963 and 1996 models, the four-dimensional "hyperchaotic" Rössler model, as well as recurrent networks of chaotic oscillators.

These examples demonstrate that when only one state variable is observed, utilizing $D_M \geqslant L_c$ time delays stabilizes the synchronization manifold enough to enable accurate estimation of unknown states and parameters, and permit accurate predictions beyond the observation window. In this way, the time delays are capable of significantly reducing the number of measurements required to achieve accurate estimates and reliable predictions.

In practice, the number of available measurements is often tightly constrained (e.g., by cost or technological considerations) and are typically sparse compared with the number of degrees of freedom of the model. For instance, in the analysis of a shallow water model of geophysical flow, it was shown that $L_c \approx 0.7D$ [16], while in operational weather prediction systems (such as the European Centre for Medium Range Weather Forecasts) only about 10^7 measurements are typical for models with 10^8 or 10^9 degrees of freedom [50].

When additional measurements are unavailable, time delays offer another means to further stabilize the search space. Regarding the shallow water model for instance, recent work has shown that, by using time delays in this way, enough information is extracted from the height field alone to permit synchronization between the data and the model [51]. These results demonstrate a proof of concept that time delays may be used to effectively reduce the total number of measurements required to achieve this goal.

The form of the time delayed coupling has some desirable properties as well. For instance, in the case where $D_M = 1$ it reduces to the classical form (3). Also, when $D_M > 1$, it generates control perturbations on *all* state components, and our results have shown that, by including the parameters as state variables augmented with trivial dynamics $d\mathbf{p}/dt = 0$, parameter estimation occurs as a natural result of the synchronization process. This is an improvement over the classical (i.e., $D_M = 1$) form, which typically requires nonlinear optimization techniques to estimate the parameters. Additionally, one could use this method in conjunction with other nonlinear estimation procedures as a means to improve the estimate when $L < L_c$.

There also appears to be a direct correspondence between the sufficient number of measurements L_c and the number of time delays required to stabilize the synchronization manifold. This is interesting for a number of reasons. For one, although it is reasonable that in each case there should exist such a sufficient condition, there is no reason to assume *a priori* that they should be the same. The fact that they appear to be (roughly) equal indicates that this condition may be an invariant property of the dynamics. Indeed, we have observed the same phenomenon by using other approaches (e.g., variational optimization and Markov chain Monte Carlo [49,52]), which suggests these other methods may also benefit from the inclusion of time delays.

This result also highlights clearly the distinction between the use of time delays here, for the purpose of state and parameter estimation, and its familiar application in nonlinear dynamics for reconstructing the phase space of a partially observable dynamical system. For instance, the Kaplan–Yorke dimension [53] for the $D = 20$ Lorenz 1996 system is $D_A \approx 12$, so the sufficient dimension for phase space reconstruction is [29] $2D_A \approx 24$, whereas the required number of time delays is $D_M \approx 9$. The time delays stabilize the synchronization manifold using a fraction of the sufficient number of delays needed for full attractor reconstruction.

We also wish to note that, in practice, there is a finite limit to the amount of information available from the time delays of a single scalar time series. For instance, with the Lorenz 1996 system we observed that, regardless of the chosen dimension D, a threshold occurs around $D_M \approx 12$. Continuing to increase D_M beyond this threshold causes the $\partial \mathbf{S}(\mathbf{x})/\partial \mathbf{x}$ matrix to become highly ill conditioned, and therefore requires a lower choice of rank to maintain stable computations. We suspect that restricting the rank in this way effectively limits the number of stable dimensions transferred from the control coupling. In other words, we have seen evidence that there exists a

correspondence between the required number of measured states L_c, the number of time delays D_M, and the rank r of the inverse. Given the threshold $D_M \approx 12$ and the empirical scaling rule $L_c \approx 0.4D$, this suggests that we should not be able to synchronize a system with $D > 25$ by using only a single measurement, which is indeed the limit observed in our numerical experiments (although these results are not shown here).

We further suggest that this threshold is due to a limited amount of information available in a time-series that is locally bounded by the characteristic timescale of the chaos. That is, holding τ fixed and increasing D_M to extend the time delay vector far beyond the Lyapunov time should not provide any additional information, as the later points are too far de-correlated to be of any use. Likewise, increasing D_M by decreasing τ and holding the total length fixed should in principle provide enough information. However, in this case we are restricted by the noise level of the system (or if no noise is present, by finite numerical precision). Both cases are indicated by ill conditioning of $\partial \mathbf{S}(\mathbf{x})/\partial \mathbf{x}$, and the threshold on D_M likely a manifestation of both effects.

Thus, it is crucial to the success of the algorithm that the parameters D_M and τ are chosen appropriately. They must be large enough to provide additional information about the unobserved states, yet not so large as to induce numerical instability in the calculation of the inverse. There exist many techniques for choosing τ that have been developed for the purposes of attractor reconstruction, such as the first minimum of average mutual information between measurements. These methods are likely to be applicable here as well, although for the examples shown here changing τ by a few dt did not noticeably impact the results.

Also, whereas here we have only considered forward time delays, recent work [54,55] has shown that a mixture of forward and backward delays can further improve the conditioning of $\partial \mathbf{S}(\mathbf{x})/\partial \mathbf{x}$. Whether or not mixed delays provides superior results for synchronization is currently under investigation.

Moreover, although these examples have been limited to the case where the number of measurements $L = 1$, our formulation generalizes easily to the case where $L > 1$. In particular, given D_{M_i} time delays available in each of $i = \{1, \ldots, L\}$ measurements, the number of time delays required to stabilize the estimation should satisfy $\sum_{i=1}^{L} D_{M_i} > L_c$. Note that this is only a rough approximation because it is quite clear that the amount of information contained in each state component is different in general, and not additive, in the sense that measuring two mutually dependent variables may not provide as much information as each variable contributes individually. We remark however, that the twin experiment framework is a useful tool for determining the relative value of a given measurement. Such information is essential for analyzing the costs and benefits of obtaining further measurements.

The inclusion of time delays comes of course with an additional computational cost, mainly associated with the integration steps required to construct the time delay vectors and its Jacobian, as well as solving for the perturbation itself. The baseline for comparison is the simple nudging algorithm (3), which is recovered in the limit $D_M = 1$. Certainly, clever algorithmic improvements are required in order to reduce this overhead as much as possible. For instance, one idea is to reduce the resolution of the model, initialize it with existing measurements, run the assimilation, and then interpolate to recapture the desired resolution for forward prediction. It may also be possible to update $\mathbf{S}(\mathbf{x})$ directly with the perturbation, so that it does not need to be recomputed in its entirety at each time step. Other such ideas will surely emerge as well, while the technique is scaled up to larger problems.

Finally, extension of this method to more complex models, or high-dimensional models representing numerical approximations to partial differential equations appears possible. In particular, applications of this approach for numerical weather prediction or the analysis of biological neural networks are currently under investigation. These applications typically permit too few measurements than are required to stabilize the estimation procedure and the results presented here suggest that the incorporation of time delays will allow us to extract more information from *existing measurements* to improve our state and parameter estimates and generate more accurate predictions.

ACKNOWLEDGMENTS

The research leading to the results has received funding from the European Community's Seventh Framework Program FP7/2007-2013 under Grant Agreement No. HEALTH-F2-2009-241526, EUTrigTreat. This work was funded in part under a grant from National Science Foundation (PHY-0961153). Partial support from the Department of Energy CSGF program (DE-FG02-97ER25308) for D. Rey is appreciated. Partial support has come from the ONR MURI Program (N00014-13-1-0205). We thank S. Luther for discussions on state and parameter estimation and for his continuing support.

[1] A. C. Lorenc and T. Payne, Q. J. R. Meteorol. Soc. **133**, 607 (2007).

[2] Maria Rodriguez-Fernandez, Jose A. Egea, and Julio R. Banga, BMC Bioinf. **7**, 483 (2006).

[3] A. Pokhilko *et al.*, Mol. Sys. Biol. **6**, 416 (2010).

[4] A. Horvath and D. Manini, in *International Conference on BioMedical Engineering and Informatics* (IEEE Computer Society, Los Alamitos, CA, 2008), Vol. 1, p. 713.

[5] R. Xiong, P. J. Wissmann, and M. A. Gallivan, Comput. Chem. Eng. **30**, 1657 (2006).

[6] D. Dochain, J. Process Control **13**, 801 (2003).

[7] Z. Yang and J. M. Hamrick, Estuarine, Coastal Shelf Sci. **62**, 13 (2005).

[8] N. Margvelashvili *et al.*, Environ. Modell. Software **40**, 191 (2013).

[9] R. H. Clayton *et al.*, Prog. Biophys. Mol. Biol. **104**, 22 (2011).

[10] Q. J. M. Huys and L. Paninski, PLoS Comput. Biol. **5**, e1000379 (2009).

[11] D. Meliza, M. Kostuk, H. Huang, A. Nogaret, H. D. I. Abarbanel, and D. Margoliash (unpublished).

[12] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, 2nd ed. (Springer, Berlin, Heidelberg, 2008).

[13] H. D. I. Abarbanel, *Predicting the Future: Completing Models for Observed Complex Systems* (Springer-Verlag, New York, 2013).

[14] H. D. I. Abarbanel, D. R. Creveling, R. Farsian, and M. Kostuk, J. Appl. Dyn. Syst. **8**, 1341 (2009).

[15] J. C. Quinn, Ph.D. thesis, University of California, San Diego, 2010 (unpublished).

[16] W. Whartenby, J. Quinn, and H. D. I. Abarbanel, Mon. Weather Rev. **141**, 2502 (2013).

[17] D. Rey, M. Kostuk, M. Eldridge, H. D. I. Abarbanel, J. Schumann-Bischoff, and U. Parlitz, Phys. Lett. A **378**, 869 (2014).

[18] J. E. Hoke and R. A. Anthes, Mon. Weather Rev. **104**, 1551 (1976).

[19] S. Lakshmivarahan and John M. Lewis, *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, edited by Seon Ki Park and Liang Xu (Springer, Berlin, Heidelberg, 2013), Vol. II, Chap. 2.

[20] U. Parlitz, L. Junge, and L. Kocarev, Phys. Rev. E **54**, 6253 (1996).

[21] J. Amezcua, K. Ide, E. Kalnay, and S. Reiche, Q. J. R. Meteorol. Soc. **140**, 995 (2014).

[22] L. M Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990).

[23] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).

[24] Mark Kostuk, Ph.D. thesis, University of California, San Diego, 2012 (unpublished).

[25] L. Junge and U. Parlitz, Phys. Rev. E **64**, 055204(R) (2001).

[26] D. Aeyels, SIAM J. Control Optim. **19**, 595 (1981).

[27] D. Aeyels, Syst. Contr. Lett. **1**, 92 (1981).

[28] F. Takens, in *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, edited by D. A. Rand and L.-S. Young (Springer-Verlag, Berlin, Heidelberg, 1981), Vol. 898, p. 366.

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

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

[31] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. (Cambridge University Press, Cambridge, 2004).

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

[33] A. Fraser and H. Swinney, Phys. Rev. A **33**, 1134 (1986).

[34] A. Tikhonov, A. Goncharsky, V. Stepanov, and A. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems* (Kluwer Academic Publishers, Dordrecht, 1995).

[35] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, New York, 2007), Sec. 19.5.

[36] G. H. Golub and C. Reinch, Numer. Math. **14**, 403 (1970).

[37] H. Nijmeijer and A. V. der Schaft, *Nonlinear Dynamical Control Systems* (Springer, New York, 1990).

[38] E. Candés, J. Romberg, and T. Tao, Commun. Pure Appl. Math. **59**, 8 (2006).

[39] O. E. Rössler, Phys. Lett. A **71**, 155 (1979).

[40] E. N. Lorenz, in *Proceedings of the Seminar on Predictability* (ECMWF, Reading, 1996), Vol. 1.

[41] E. H. Colpitts, US Patent No. 1,624,537 (1918).

[42] T. Palmer and R. Hagedorn, *Predictability of Weather and Climate* (Cambridge University Press, Cambridge, 2006).

[43] A. Karimi and M. R. Paul, Chaos **20**, 043105 (2010).

[44] P. E. Moraal, Ph.D. thesis, University of Michigan, 1994 (unpublished).

[45] P. E. Moraal and J. W. Grizzle, in *Proceedings of the 34th Conference on Decision and Control, New Orleans* (IEEE, 1995), p. 109.

[46] C. Chi-Tsong, *Linear System Theory and Design* (Oxford University Press, New York, 1995).

[47] H. Michalska and D. Mayne, IEEE Trans. Autom. Control **40**, 995 (1995).

[48] W. Kang, IEEE Trans. Autom. Control **51**, 344 (2006).

[49] J. C. Quinn, P. H. Bryant, D. R. Creveling, S. R. Klein, and H. D. I. Abarbanel, Phys. Rev. E **80**, 016201 (2009).

[50] L. Isaksen and S. English, lecture delivered at the European Centre for Medium-Range Weather Forecasts Training Course, June, 2013 (unpublished).

[51] Z. An, D. Rey, and H. D. I. Abarbanel (unpublished).

[52] J. Quinn and H. D. I. Abarbanel. Q. J. R. Meteorol. Soc. **136**, 1855 (2010).

[53] J. L. Kaplan and J. A. Yorke, in *Functional Differential Equations and Approximations of Fixed Points*, Proceedings, Bonn, July 1978 (Springer-Verlag, Berlin, 1979), p. 204.

[54] U. Parlitz, J. Schumann-Bischoff, and S. Luther, Chaos **24**, 024411 (2014).

[55] U. Parlitz, J. Schumann-Bischoff, and S. Luther, Phys. Rev. E **89**, 050902(R) (2014).