



CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Using waveform information in nonlinear data assimilation

Daniel Rey, Michael Eldridge, Uriel Morone, Henry D. I. Abarbanel, Ulrich Parlitz, and Jan Schumann-Bischoff

Phys. Rev. E **90**, 062916 — Published 22 December 2014

DOI: [10.1103/PhysRevE.90.062916](https://doi.org/10.1103/PhysRevE.90.062916)

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, CA 92093-0374

Ulrich Parlitz and Jan Schumann-Bischoff

Max Planck Institute for Dynamics and Self-Organization

Institute for Nonlinear Dynamics

Georg-August-Universität Göttingen

Am Fassberg 17, 37077

Göttingen, Germany

(Dated: December 2, 2014)

Abstract

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.

^{*} drey@physics.ucsd.edu

[†] Marine Physical Laboratory (Scripps Institution of Oceanography)

14 **CONTENTS**

15	I. Introduction	3
16	II. Assimilating data into models of observed processes	4
17	A. The action $A_0(\mathbf{X}) = -\log[P(\mathbf{X} \mathbf{Y})]$	6
18	B. Chaotic instability as an impediment to success and the ‘critical’ number of	
19	measurements L_c	9
20	C. Using time delayed measurements to further stabilize the transfer of information	11
21	D. Synchronization using information from time delayed measurements	14
22	III. Computing the pseudoinverse of $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$	17
23	A. Computing the pseudoinverse with singular value decomposition	18
24	B. Rank considerations	19
25	IV. Using time delay information in examples	23
26	A. Lorenz 1963 model	24
27	1. Estimating states only	25
28	2. Estimating states and parameters	26
29	B. Rössler hyperchaos	27
30	C. Lorenz 1996 model	28
31	1. $D = 20$ with a single global forcing parameter	29
32	2. $D = 10$ with different forcing for each oscillator	31
33	V. Direct estimation of L_c	31
34	VI. Networks of Chaotic Oscillators	35
35	A. A small network of chaotic Colpitts oscillators	36
36	B. Estimating the states of the network	37
37	C. Estimating the states and the couplings between nodes	38
38	D. Estimating the functional connectivity of the network	40
39	VII. Discussion and Summary	41
40	Acknowledgements	45

42 **I. INTRODUCTION**

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

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

60 Previous work has shown that when the system under consideration yields chaotic tra-
61 jectories the dynamical instability associated with sensitivity to initial conditions impedes
62 the successful identification of the initial state and parameters of the system [12, 13]. In
63 particular, it has been observed that many data assimilation techniques require a minimum
64 number of measurements to succeed, even when the noise levels are low [12, 14–16].

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

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

82 II. ASSIMILATING DATA INTO MODELS OF OBSERVED PROCESSES

83 During an assimilation or measurement window $[0, T]$ data from an observed system
84 are presented to a system model. Various methods are employed [12, 13] to estimate the
85 fixed parameters and full state of the model (both observed and unobserved state variables)
86 at the end of the assimilation window $t = T$. To validate the model and the estimates,
87 predictions are compared with further observations in such a way that information about
88 these subsequent observations is not utilized to further modify the estimates of the fixed
89 model parameters.

90 The model is stated in terms of differential equations for fields $\phi_k(\mathbf{r}, t)$ or point objects
91 $q_\alpha(t)$, so one must estimate all of the $\phi_k(\mathbf{r}, T)$ or $q_\alpha(T)$ in order to predict the dynamical
92 behavior for $t > T$. We reduce the continuous set of independent variables (\mathbf{r}, t) to a
93 finite grid in space and time arriving at a set of state variables $x_a(t_n) = x_a(n)$ where $a =$
94 $\{1, 2, \dots, D\}$ and $n = \{0, 1, \dots, N\}$. The resulting state $\mathbf{x}(n) = \{x_1(n), x_2(n), \dots, x_D(n)\}$
95 follows the rule

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

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

99 time map in the form

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

100 though all calculations are actually performed with Eqn. (1).

101 The information we wish to transfer to this model resides in the L measurements
 102 $\mathbf{y}(n) = \{y_1(n), y_2(n), \dots, y_L(n)\}$ made at each time t_n within an observation window
 103 $\{t_0, t_1, \dots, t_n, \dots, t_m = T\}$. To connect the measurements $y_l(n)$ with the solution of the
 104 dynamical equations described by the model we must specify a ‘measurement function’,
 105 which realizes the data in terms of the model output $\mathbf{x}(n)$ as $y_l(n) = h_l(\mathbf{x}(n))$. When
 106 $y_l(n) \approx h_l(\mathbf{x}(n))$ within the estimation window, the model is said to be consistent with the
 107 data, while validation of the model requires prediction of the observables $h_l(\mathbf{x}(t))$ for $t > T$.

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

$$x_a^{(0)}(n) \rightarrow x_a^{(1)}(n) \rightarrow x_a^{(2)}(n) \rightarrow \dots \rightarrow x_a^{(J)}(n)$$

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

115 Our discussion will primarily focus on the case where the model developed for under-
 116 standing observational data is perfect. That is, the data has no model errors and provides
 117 a deterministic constraint on how estimations and predictions are carried out. In this limit
 118 where the model dynamics are known, a simple technique for the direct transfer of informa-
 119 tion from observations to the dynamical model involves adding a nonphysical control term
 120 associated with each measurement to perturb the state of the model system $\mathbf{x}(t)$ toward the
 121 observations $\mathbf{y}(t)$ as the model evolves in time. The equations for these ‘coupled dynamics’
 122 are given by

$$\frac{dx_l(t)}{dt} = F_l(\mathbf{x}(t)) + \sum_{\nu=1}^L g_{l,\nu}(t) (y_\nu(t) - h_\nu(\mathbf{x}(t))), \quad (3)$$

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

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

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

127 This construction has been implemented in the meteorological literature for many years,
 128 where it is called ‘nudging’, Newtonian relaxation, or 4DDA, and is rooted in the theory of
 129 controls and dynamical systems [18, 19]. From a dynamical systems perspective, the control
 130 term $\mathbf{g}(t)$ transfers information from the measured data to the model state by coupling the
 131 estimated (model) system to the true (physical) system to promote the synchronization of
 132 the model with the data [20].

133 This process is essentially a dynamical inverse, wherein the model state and parameters
 134 are deduced from the measured data. The model dynamics act as a filter that supplies
 135 additional information about the unobserved states of the model, which are required to
 136 construct an accurate estimate of the state of the true system. This idea of using the model
 137 as a filter is well-established and is the core idea behind algorithms like the Kalman-Bucy
 138 filter [21], as well as its various extensions. In those algorithms, the coupling term $\mathbf{g}(t)$ is
 139 dynamical and chosen to minimize an estimate of the error covariance [13].

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

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

144 In practice, the model is almost never perfectly accurate. When model errors are present
 145 or when the dynamics of the model are stochastic, the iterative process taking $\mathbf{x}^{(j)}(n) \rightarrow$
 146 $\mathbf{x}^{(j+1)}(n)$ may be formulated as [12]:

- 147 • A numerical optimization procedure to estimate an ‘optimal’ path of the states

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

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

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

151 Prediction beyond the measurement window, $t > T$, requires all components of the state
 152 $\mathbf{x}(T)$ and either the deterministic dynamical rule Eqn. (1) or a stochastic version of the
 153 model errors is present.

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

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

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

- 161 • Terms moving the model state from time t_n to time t_{n+1} through the observation
 162 window.
- 163 • Terms associated with the modification of the conditional probability distribution at
 164 times when measurements are made.

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

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

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

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

172 where $\mathbf{Y}(n) = \{\mathbf{y}(n), \mathbf{y}(n-1), \dots, \mathbf{y}(0)\}$ is the collection of measurements up to time t_n .
 173 This term contains the additional information transferred from the current measurement
 174 $\mathbf{y}(n)$ to the model $\mathbf{x}(n)$, conditioned on the past measurements in $\mathbf{Y}(n-1)$.

175 If the measurement noise and model errors are Gaussian distributed with respective
 176 inverse covariance matrices \mathbf{R}^m and \mathbf{R}^f , the action becomes [12]

$$\begin{aligned}
 A_0(\mathbf{X}) = & \sum_{n=0}^m \sum_{l,l'=1}^L \left[\delta m_l(n) \frac{R_{l,l'}^m(n)}{2} \delta m_{l'}(n) \right] \\
 & + \sum_{n=0}^{m-1} \sum_{a,a'=1}^D \left[\delta f_a(n) \frac{R_{a,a'}^f}{2} \delta f_{a'}(n) \right] - \log[P(\mathbf{x}(0))],
 \end{aligned} \tag{6}$$

177 where

$$\begin{aligned}
 \delta \mathbf{f}(n) & := \mathbf{x}(n+1) - \mathbf{x}(n) - \int_{t_n}^{t_{n+1}} \mathbf{f}(\mathbf{x}(t')) dt' \\
 \delta \mathbf{m}(n) & := \mathbf{y}(n) - \mathbf{h}(\mathbf{x}(n))
 \end{aligned}$$

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

180 There is much discussion in the data assimilation literature [13] focused on the develop-
 181 ment of numerical methods for evaluating the path integral in Eqn. (4). Since these integrals
 182 tend to be high dimensional, the methods can generally be divided into two categories:

- 183 • Stationary path methods, which seek the paths where $\partial A_0(\mathbf{X})/\partial \mathbf{X} = 0$ and assumes
 184 they are the dominant contribution to the integral
- 185 • Monte Carlo methods, which directly sample the distribution $\exp[-A_0(\mathbf{X})]$.

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

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

189 where the parameter s denotes ‘algorithmic time’. Here, σ is the collection of indices of the
 190 path \mathbf{X} , and $\eta_\sigma(s)$ is a Gaussian distributed random variable with mean zero, variance unity
 191 and independent at each ‘algorithmic time’ s . The distribution $P(\mathbf{X}, s)$ satisfies a Fokker-
 192 Planck equation whose distribution as $s \rightarrow \infty$ is $P(\mathbf{X}) = \exp[-A_0(\mathbf{X})]$. The Langevin

193 equation shows the connection between the minima of $A_0(\mathbf{X})$ where $\partial A_0(\mathbf{X})/\partial X_\sigma = 0$ and
 194 the distribution of fluctuations $P(\mathbf{X})$ about those minima induced by $\eta_\sigma(s)$.

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

201 **B. Chaotic instability as an impediment to success and the ‘critical’ number of** 202 **measurements L_c**

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

- 207 • The dynamical synchronization (nudging) approach
- 208 • Variational or optimization methods, which seek a minimum of $\mathbf{A}_0(\mathbf{X}, \mathbf{Y}) = -\log[P(\mathbf{X}|\mathbf{Y})]$
- 209 • The Monte Carlo estimation framework, which directly samples $P(\mathbf{X}|\mathbf{Y})$.

210 In both the variational and Monte Carlo frameworks, the problem is manifested as mul-
 211 tiple minima in the action $A_0(\mathbf{X}, \mathbf{Y}) = -\log[P(\mathbf{X}|\mathbf{Y})]$ caused by the instability associated
 212 with the sensitivity to initial conditions characteristic of chaotic motion. Since small per-
 213 turbations in the initial values of the path yield large deviations of the action, incoherence
 214 of chaotic flows for slightly differing initial states or parameters causes the search surface
 215 $A_0(\mathbf{X})$ to be riddled with local minima. The presence of these local minima significantly
 216 impedes the algorithmic search for the minimizers.

217 In the synchronization approach, the impediment arises from instabilities on the L -
 218 dimensional synchronization manifold, where $y_l(n) = h_l(\mathbf{x}(n))$ in the D -dimensional state
 219 space. Such behavior may be characterized quantitatively by the conditional Lyapunov ex-
 220 ponents (CLEs) for motion on the $D - L$ submanifold governed by Eqn. (1). A necessary

221 condition to achieve synchronization of chaotic systems requires all Lyapunov exponents to
 222 be negative [22]. When any of the $D - L$ CLEs are positive, the synchronization manifold
 223 is unstable and we observe multiple minima in \mathbf{X} for $A_0(\mathbf{X})$.

224 While there does not yet exist a rigorous mathematical framework establishing these
 225 observations, there exists substantial numerical evidence. For instance, the book by
 226 Evensen [13] shows (in Fig. 6.1) multiple minima generated by the Lorenz 1963 model [23]
 227 in the graph of the cost function associated with the strong variational method (strong
 228 4D-Var), in which the dynamical equations are used as nonlinear equality constraints, and
 229 only the initial conditions appear as control variables in the optimization. Though he does
 230 not connect this with the instability on the synchronization manifold, this connection is
 231 made in [24].

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

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

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

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

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

255 Consider a singular value decomposition (SVD) of the Jacobian $\partial\mathbf{F}(\mathbf{x}(t))/\partial\mathbf{x}$, and denote
256 the unstable subspace as the space spanned by the singular vectors whose associated singular
257 values are greater than one [43]. In this unstable subspace, perturbations from the true
258 solution grow exponentially regardless of how close the model estimate is to the truth. The
259 coupling term in Eqn. (3) uses information from the measurements to control this unstable
260 subspace. Consequently, one needs enough measurements to span the unstable subspace
261 so that a proper choice of coupling $\mathbf{g}(t)$ may remove the dynamical instability and thereby
262 establish the conditions required to achieve synchronization.

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

273 **C. Using time delayed measurements to further stabilize the transfer of informa-** 274 **tion**

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

280 for $t > T$ will be unreliable.

281 We are concerned in this paper with removing these impediments in a manner that places
282 the smallest burden on the experiment. That is, we wish to develop techniques to reduce
283 L as much as possible while maintaining the same successful prediction rate obtained when
284 $L > L_c$ and no prior knowledge of the state is known. Our goal is to extract as much
285 information as possible from a given set of measurements, since in most applications the
286 number of measurements is tightly constrained, perhaps by cost, time or other technological
287 considerations.

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

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

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

297 In component form, it may be written as

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

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

299 The use of time delays of observed data to provide a setting for representing information in
300 nonlinear systems is quite mature and very well-tested in the analysis of chaotic behavior [25–
301 31]. In phase space reconstruction, they provide a proxy state space for analyzing properties
302 of the source of chaotic motions. Here the number of required delays is dictated by geometric
303 considerations, provided the time delay τ yields components for the equivalent of $\mathbf{Y}(t_n)$ that
304 are independent in some, usually heuristic, sense.

305 The usual practice is to use each measurement $\mathbf{y}(t_n)$ independently of measurements at
306 different times. Of course, these measurements are not totally independent of each other,
307 as they come from a dynamical system that describes the physical processes underlying the

308 system's time evolution. The dependence comes from the idea that proceeding from an
 309 observation $y(t)$ (using $L = 1$ for illustration) to a time delayed observation $y(t + \tau)$ utilizes
 310 some dynamical rule involving all of the degrees of freedom of the observed system, not just
 311 those that are observed. So if τ is long enough for the unobserved states of the system to
 312 have acted in sufficient magnitude to influence $y(t + \tau)$, then $y(t + \tau)$ possesses information
 313 about the overall dynamics not available in $y(t)$ alone. The utility and mathematical value
 314 of the time delay construction we develop rests precisely on the information residing in those
 315 connections.

316 Suppose we are able for some physical reason to observe only $L = 1$ variable, $z_0(t)$, in a
 317 $K + 1$ dimensional dynamical system with other variables $z_k(t)$; $k = \{1, 2, \dots, K\}$ satisfying
 318 the differential equations

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

319 then

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

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

322 If τ is too short relative to the natural times of the $\mathbf{z}(t)$, effectively nothing new will
 323 be usable in $z_0(t + \tau)$ about the $\mathbf{z}(t)$. Similarly, if τ is too long compared to the time
 324 scale of chaotic behavior, the values of $z_0(t)$ and $z_0(t + \tau)$ will be incoherent with respect
 325 to each other. So a balanced choice of τ , perhaps as given by the first minimum of the
 326 average mutual information between them, is appropriate [29, 30, 32]. This line of reasoning
 327 regarding the selection of time delays applies here for the purposes of extracting additional
 328 information from our measurements.

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

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

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

345 **D. Synchronization using information from time delayed measurements**

346 Following our definition of $\mathbf{Y}(t_n)$ in Eqn. (7), we construct the corresponding time delayed
 347 model state

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

348 Its components may be written as

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

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

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

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

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

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

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

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

358 where $\mathbf{g}'(t)$ is a coupling gain matrix defined in \mathbf{S} space. Mapping back to the physical space
 359 $\mathbf{x}(t)$ we arrive at

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

360 where, in matrix notation,

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

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

364 This equation displays the manner in which information from $\mathbf{Y}(t)$ is transferred to the
 365 model $\mathbf{x}(t)$ via the dynamical equations. This form of the dynamics is utilized throughout
 366 the measurement window to estimate the model output states and parameters $\mathbf{x}(t)$ required
 367 to match the data $\mathbf{Y}(t)$. When measurements are completed, we set the coupling matrices
 368 $\mathbf{g}(t), \mathbf{g}'(t) = 0$ to predict for $t > T$ using the uncoupled dynamics Eqn. (1).

369 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
 370 map to time delay space $\mathbf{x}(t) \rightarrow \mathbf{S}(\mathbf{x}(t))$. This Jacobian is constructed by integrating the
 371 variational equation [29] for the uncoupled dynamics Eqn. (1)

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

372

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

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

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

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

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

ALGORITHM 1. TIME DELAY SYNCHRONIZATION

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

382 The algorithm for determining the state $\mathbf{x}(t)$ within the observation window $0 \leq t \leq T$ is
 383 outlined in Alg. (1). At each time step t_n , we compute the model time delay vector $\mathbf{S}(\mathbf{x}(t_n))$
 384 and the Jacobian $\partial\mathbf{S}(\mathbf{x}(t_n))/\partial\mathbf{x}(t_n)$. The results are then used to evaluate the coupling
 385 perturbation $\delta\mathbf{x}(t_n)$. The process is repeated in this way, mapping back and forth between
 386 the physical and time delay spaces until the end of the observation window is reached.

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

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

- 394 • Information from the time delays of the observations is presented to the physical model
 395 equations.

- 396 • The framework is easily extended to incorporate nonlinear measurement functions
397 $\mathbf{h}(\mathbf{x})$.
- 398 • The impact of the coupling terms is not limited to measurement times when $t = t_n$.
399 All measurements within the current time delay window $[t, t + (D_M - 1)\tau]$ can be
400 incorporated, regardless of the current time step.
- 401 • All components of the model state $\mathbf{x}(t)$ are influenced by the control term, not just
402 the observed components. Consequently, the fixed parameters of the model may be
403 estimated as a natural result of the synchronization process by including them as
404 additional state variables, satisfying $F_a(\mathbf{x}(t)) = 0$.
- 405 • The time delay technique allows one to extract additional information from *existing*
406 *measurements*.

407 The latter point is extremely important, as in many applications additional measurements
408 may be prohibitively expensive, time-consuming, or not technologically feasible. The benefits
409 of using time delays will be displayed in further detail in the context of the numerical
410 examples presented later in the paper. For the moment however, we divert our attention
411 to a technical matter that is of crucial importance. Namely, the calculation of control term
412 $\partial\mathbf{x}/\partial\mathbf{S}(\mathbf{x})$ as a regularized local inverse.

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

414 We now discuss some of the details regarding the computation of the pseudoinverse
415 $\partial\mathbf{x}/\partial\mathbf{S}(\mathbf{x}) := (\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x})^+$. We wish to solve the linear system of equations for $\delta\mathbf{x}$

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

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

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

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

423 The common solution for such ill-posed problems such as this is to include a *regularization*
 424 term in the objective function [33, 34],

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

425 This process, known as Tikhonov regularization, allows us to choose $\mathbf{\Gamma}$ to give preference
 426 for particular solutions with desirable properties. Here we choose $\mathbf{\Gamma} = \alpha\mathbf{I}$ where \mathbf{I} is a
 427 $D \times D$ dimensional identity matrix, which in the limit $\alpha \rightarrow 0$ recovers the expression for the
 428 Moore-Penrose pseudoinverse. In addition to being arguably the simplest choice for $\mathbf{\Gamma}$, this
 429 form selects for solutions to Eqn. (11) that minimizes the least squares norm of $\delta\mathbf{x}$. The
 430 regularization in Eqn. (12) leads to the expression for $\delta\mathbf{x}$

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

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

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

438 **A. Computing the pseudoinverse with singular value decomposition**

439 There are many numerical approaches available for constructing the pseudoinverse of an
 440 $m \times n$ matrix \mathbf{M} . The simplest choice involves the direct inversion of the matrix product,

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

441 This technique is known to incur numerical stability problems, which become especially
 442 problematic when \mathbf{M} is ill-conditioned. The reason is that if \mathbf{M} has condition number κ then

443 the product $\mathbf{M}^T \cdot \mathbf{M}$ has condition number κ^2 , and will be considerably more ill-conditioned
 444 than \mathbf{M} .

445 An alternative approach that does not suffer from such instability involves an SVD of
 446 the matrix \mathbf{M} [35]. A generalization of eigenvalue decomposition from square to non-square
 447 matrices, the SVD decomposes an $n \times m$ matrix \mathbf{M} into a product of three matrices,

$$\mathbf{M} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^\dagger \quad (14)$$

448 where \mathbf{U} and \mathbf{V} are unitary matrices of size $n \times n$ and $m \times m$ respectively, $\mathbf{\Sigma}$ is an $m \times n$
 449 rectangular diagonal matrix of singular values σ_i , and \mathbf{V}^\dagger denotes the conjugate transpose
 450 of the matrix \mathbf{V} . The SVD is unique up to permutations and sign exchanges of the singular
 451 values. Most algorithms choose the singular values to be positive and ordered such that
 452 $\sigma_1 > \sigma_2 \dots > \sigma_{r_{max}}$ where $r_{max} = \min(m, n)$.

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

$$\mathbf{M}^+ = \mathbf{V} \cdot \mathbf{\Sigma}^+ \cdot \mathbf{U}^\dagger. \quad (15)$$

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

460 B. Rank considerations

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

467 By significantly raising this tolerance (e.g. from $\mathcal{O}(10^{-16})$ to $\mathcal{O}(10^{-3})$) the calculations
 468 can be stabilized but its performance is markedly degraded, presumably because information

469 about the unobserved states is being discarded. In practice, a smaller rank corresponds
470 directly to a smaller control $\delta\mathbf{x}$. The balance, therefore, is between a large enough $\delta\mathbf{x}$ to
471 synchronize the model states with the data and a small enough $\delta\mathbf{x}$ to keep the numerical
472 methods stable.

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

- 476 • If the rank is chosen conservatively small, the calculations are numerically stable.
- 477 • It provides insight into its role in stabilizing the synchronization manifold. The rank
478 of the inverse appears to be roughly equivalent to the number of measurements needed
479 to achieve synchronization in twin experiments.
- 480 • The pseudoinverse, which in general is a discontinuous operation, can be made contin-
481 uous by specifying a constant rank. This is important because it allows the derivative
482 of the inverse to be properly defined. This is necessary, for instance, to calculate the
483 Lyapunov exponents of the error propagation, which are often used to prove conver-
484 gence of optimal control techniques [36].

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

489 Through numerical experiments we have observed that the $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ matrix is more well-
490 conditioned in regions with higher local Lyapunov exponents. This makes some intuitive
491 sense, as the degeneracy of $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ is due to the lack of independence among the various
492 components of $\mathbf{S}(\mathbf{x})$, which in turn is related to the rate of information flow among the
493 various state variables $\mathbf{x}(t)$. Larger local Lyapunov exponents indicate increased dynamical
494 mixing among the physical states as well as improved conditioning of the $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ matrix.
495 In other words, the ‘optimal’ rank of the inverse fluctuates along the trajectory and the SVD
496 method actually appears to perform better in regions where the dynamics are more locally
497 chaotic.

We have considered algorithms for adaptively choosing the rank of the Jacobian to max-
imize the amount of information transferred by the control coupling, without causing nu-

merical 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,

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

498 Given positive singular values of $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ ordered as $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{r_{max}}$, where $r_{max} =$
 499 $\min(D_M L, D)$, the choice of rank r can be expressed via the inequality,

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

In this case, select the largest r such that

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

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

502 Another useful choice involves the L^∞ norm,

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

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 Eqn. (16) holds. If this condition is true, increase the rank by one and perform the check again 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,

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

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

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

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

517 Certainly, other good choices are available. For instance, selecting a low-pass operator
 518 (e.g., a difference operator or a weighted Fourier operator) for the Tikhonov matrix in
 519 Eqn. (12) is known to enforce smoothness and may help combat the effects of measurement
 520 noise [34]. Another idea is to use L^1 norm for the regularization term in Eqn. (12) so that,

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

521 This formulation may be useful when the ‘optimal’ control perturbation is sparse, as this
 522 choice of norm optimizes for sparsity and is related to recent developments in the theory of
 523 compressed sensing [37]. We have also yet to investigate using a non-uniform time delay.
 524 For instance, it may be possible to choose the delays adaptively to generate vectors via
 525 $\mathbf{x} \rightarrow \mathbf{S}(\mathbf{x})$ that are in some way ‘optimally’ well-conditioned.

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

530 **IV. USING TIME DELAY INFORMATION IN EXAMPLES**

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

535 We will examine four model dynamical systems as testbeds for our ideas. Three of these
 536 systems are small, well-investigated dynamical models: the Lorenz 1963, the Rössler 1979
 537 and the Lorenz 1996 models [23, 38, 39]. In addition, we extend the analysis to network
 538 models with chaotic Colpitts oscillators at the nodes [40]. For each example, we demonstrate
 539 that the time delay control scheme extracts enough information from a single measured
 540 variable (i.e., a scalar time series) to achieve accurate estimates and predictions for the
 541 unobserved states and parameters of the system. This is a significant improvement over the
 542 standard $D_M = 1$ coupling procedure in Eqn. (3), for which it will be shown that a single
 543 measured state component is in fact insufficient for most of the examples presented here.

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

$$SE_x^2(t) := \frac{1}{D} \sum_{a=1}^D (x_a^{model}(t) - x_a^{data}(t))^2 \quad (17)$$

549 as a metric of the error between the model and data trajectories. Since our models are
 550 deterministic, as $SE_x(t) \rightarrow 0$, the model will exactly reproduce the unobserved as well as
 551 the observed data.

552 In real experiments however, the unobserved states are unknown. In this situation, we
 553 instead use the synchronization error in \mathbf{S} -space

$$SE_s^2(t) := \frac{1}{D_M L} \sum_{l=1}^L \sum_{k=1}^{D_M} (Y_{k:l}(t) - S_{k:l}(\mathbf{x}(t)))^2, \quad (18)$$

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

555 To illustrate the general applicability of our technique we present these examples as a
 556 series of ‘real’ experiments by performing the assimilation as though the data had been

557 collected from a partially observable system. No information from the unobserved variables
 558 was used to produce the state estimates. We only supplement the predictions, when needed
 559 for comparison, with data from the unobserved variables. In this way, we hope to convey
 560 the capability of our method in the context of actual experiments.

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

569 Since we are working with deterministic models, without model error, we do not require
 570 any of the probabilistic machinery discussed earlier. Although one may place the time
 571 delay method fully within the general path integral formulation [12], we do not do so here.
 572 Instead, we simply modify the dynamical equations with the control terms Eqn. (10). After
 573 long enough time evolution, the states in the model will match the states of system and we
 574 take these as our initial conditions for prediction.

575 A. Lorenz 1963 model

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

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

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

578 To produce the data, we integrate these equations with a time step $\Delta t = 0.01$ for $t = [0, T]$
 579 where $T = 10 = 1000 \Delta t$ and select a measurement function $y(t) = h(\mathbf{x}(t)) = x_1(t)$; so
 580 $L = 1$. The initial conditions for both the physical system $\mathbf{x}^{data}(0)$ and the model system
 581 $\mathbf{x}^{model}(0)$ are chosen at random from a uniform distribution that roughly spans the size of

582 the attractor. However, the initial condition for the observed component $x_1(0)$ was chosen
583 to match the data.

584 1. *Estimating states only*

585 Fixing the parameters p_1, p_2, p_3 at the values used to generate the data, we perform our
586 calculations using a constant, uniform coupling $g = 10$ so that $g \Delta t = 0.1$. The matrix
587 $\mathbf{g}'(t)$ is taken to be unity and the pseudoinverse is constructed using full rank $r = r_{max} =$
588 $\min(D_M, D)$. The time delay is chosen to be $\tau = 0.1 = 10 \Delta t$, which is consistent with the
589 average mutual information criterion [29, 30, 32].

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

593 The trajectory of the experimental synchronization error $SE_s(t)$ throughout the assimila-
594 tion window is shown in the **Top** panel of Fig. 1 for $D_M = \{1, 2, 3\}$. Note in particular how
595 the $D_M = 1$ coupling is insufficient to achieve synchronization. This result however, is not
596 at odds with the work of Pecora and Carroll, whose synchronization scheme replaces $x_1(t)$
597 by $y(t)$ in the dynamical equations, corresponding to the limit $g \rightarrow \infty$ [22]. By increasing
598 the coupling to $g = 100$ we provide enough control strength to synchronize the systems with
599 no time delays, $D_M = 1$.

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

610 The true test however, of any data assimilation scheme is its ability to predict the be-
611 havior beyond the assimilation window. To this end, the **Bottom** panel of Fig. 1 shows the

612 estimates and predictions for the observed state component $x_1(t)$ for each $D_M = \{1,2,3\}$.
 613 As expected, the predictions for $D_M = 1$ are poor whereas for $D_M = \{2,3\}$ they are ex-
 614 ceptionally accurate throughout the entire prediction window. The fact that $D_M = \{2,3\}$
 615 produced excellent predictions but $D_M = 1$ did not, supports use of the \mathbf{S} -space synchro-
 616 nization error Eqn. (18) as an experimentally viable indicator of convergence. Since this is
 617 a twin experiment, we actually know all the ‘unobserved’ data time series, so we may verify
 618 the predictions of the unobserved state components directly. We have done this, and the
 619 results (not shown) confirm our comments.

620 2. Estimating states and parameters

621 Next, we estimate the parameters for this system by extending Eqn. (19) to include the pa-
 622 rameters as state variables. We now have six dynamical equations $\mathbf{p}(t) = \{p_1(t), p_2(t), p_3(t)\} =$
 623 $\{x_4(t), x_5(t), x_6(t)\}$ with $d\mathbf{p}(t)/dt = 0$. Only the time delay control appears in the vector
 624 field of the p_k . The initial values of the parameters are chosen to be 50% of their known
 625 values, and the coupling matrix is selected as $\mathbf{g}(t) = \text{diag}(\{10, 10, 10, 100, 100, 100\})$. That
 626 is, the parameters are subject to ten-fold larger coupling than the states. The assimilation
 627 proceeds as before, except over an extended observation window $T = 100 = 10^4 \Delta t$. The
 628 coupling is then turned off to predict for [100, 110].

629 Trajectories of $\text{SE}_s(t)$ are shown in the **Top** panel of Fig. 2 for $D_M = \{1, 2, 3\}$. Syn-
 630 chronization proceeds more slowly than in the previous example where parameters are fixed.
 631 Notably however, for the extended system $D_M = 2$ is no longer sufficient. This can be
 632 further established by examining the parameters at the end of the assimilation window (see
 633 Table I). As expected, for $D_M = 1$ the parameters have not changed from their initial
 634 values, as $D_M = 1$ coupling only perturbs the measured state components (here $x_1(t)$) and
 635 is therefore unable to perform parameter estimation. For $D_M = 2$, the parameter estimates
 636 are poor and for $D_M = 3$ they are very accurate. In the latter case, the relative errors
 637 $\epsilon_i^{\text{rel}} := (p_i^{\text{model}} - p_i^{\text{data}})/p_i^{\text{data}}$, are all $\mathcal{O}(10^{-6})$ or smaller.

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

- 642 • It illustrates our statement that the merit of any data assimilation scheme must be
643 judged by its capability to predict, not just fit the data.
- 644 • For the extended system (with parameters included), selecting $D_M = 2$ is no longer
645 sufficient to achieve synchronization. This suggests that promoting parameters into
646 states with trivial dynamics can increase L_c .

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

654 B. Rössler hyperchaos

655 We now investigate the four dimensional Rössler system described by [38]

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

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

660 To initiate our time delay algorithm the three *unobserved* initial model conditions
661 are selected randomly from a uniform distribution that spans the attractor, so that
662 $\mathbf{x}^{model}(0) = \{-20, -18.6, 25.7, 122.4\}$.

663 Parameters are estimated by treating them as four additional state variables $\mathbf{p}(t) =$
664 $\{x_5(t), x_6(t), x_7(t), x_8(t)\}$ with $d\mathbf{p}(t)/dt = 0$. The initial parameter estimates are selected to
665 be $\mathbf{p}^{model}(0) = \{0.125, 1.5, -0.25, 0.025\}$, namely 50% of the known values.

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

672 The calculations are carried out using $\tau = 4 \Delta t$ and a uniform coupling $g = 10$ so
673 $g \Delta t = 0.25$. As before, $\mathbf{g}'(t)$ is taken to be unity. At the end of the observation window, the
674 model parameters are fixed at their estimated values and, we then predict for a subsequent
675 $200 = 8000 \Delta t$ time units.

676 In the **Top** panel of Fig. 3 we plot $SE_s(t)$ for $D_M = \{6, 8, 13\}$. For $D_M = 6$ synchroniza-
677 tion does not occur whereas for $D_M = \{8, 13\}$ it does. The **Middle** panel displays $SE_s(t)$ for
678 $D_M = 8$ beyond the observation window. After the coupling is switched off the error grows
679 at a rate that is roughly consistent with the maximum Lyapunov exponent of the system.
680 The **Bottom** panel displays the estimate (red) and prediction (blue) of the observed $x_1(t)$
681 along with the known data. Excellent predictions indicate good estimates of the unobserved
682 states and parameters. The eventual deviation of the predictions from the known data is
683 due to the chaotic behavior of the system.

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

691 C. Lorenz 1996 model

692 We now turn to the example of the Lorenz 1996 model [39], which is studied widely in
693 the geophysical literature [41]. The model describes a ring of $D > 3$ coupled oscillators,
694 which obey the differential equations

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

695 where $a = \{1, 2, \dots, D\}$ and the indices are permuted cyclically i.e. $x_0(t) = x_D(t), x_{D+1}(t) =$
 696 $x_1(t)$ and $x_{-1}(t) = x_{D-1}(t)$.

697 When the forcing parameter p_1 is large enough this model exhibits extensive chaos so that
 698 the number of positive Lyapunov exponents scales with the number of spatial dimensions
 699 D [42]. Similarly, the number of measurements required to stabilize the synchronization
 700 manifold is also proportional to D . Previous work [14, 24] has shown that with a global
 701 forcing parameter $p_1 = 8.17$, the standard coupling scheme Eqn. (3) involving one control
 702 term in the differential equations of each measured state requires approximately $L_c \approx 0.4 D$
 703 to achieve synchronization. Since the dimension D may be chosen freely, this makes the
 704 Lorenz 1996 system an excellent testing ground for investigating the behavior of data assim-
 705 ilation techniques in the context insufficient measurements. For our purposes, it will further
 706 demonstrate how the time delay dimensions serve as additional measurements.

707 1. $D = 20$ with a single global forcing parameter

708 First, we look at a system of size $D = 20$ and extend it to include the single global
 709 parameter p_1 as a 21st state variable $x_{21}(t)$ with dynamics $dx_{21}/dt = 0$. We observe only
 710 the first state component $h(\mathbf{x}(t)) = x_1(t)$, so $L = 1$. Data is generated using a time-step
 711 of $\Delta t = 0.01$. We select a constant coupling $g = 10$, so that $g \Delta t = 0.1$ and time delay
 712 $\tau = 0.1 = 10 \Delta t$. The inverse, $\partial \mathbf{x} / \partial \mathbf{S}(\mathbf{x})$ is taken to have full rank $r = D_M$ and the
 713 parameters are not subject to any constraints.

714 This example also includes additive white noise in the measurement $y(t) \rightarrow y_1(t) + \eta(t)$.
 715 The noise is generated by choosing $\eta(t)$ from a uniform distribution centered around zero
 716 $U(-\alpha, \alpha)$. The amplitudes $\alpha = \{0.0, 6.34 \cdot 10^{-5}, 0.0011, 0.020\}$ (arbitrary units) are chosen
 717 so that the signal to noise ratios are respectively $\text{SNR} = \{\infty, 100, 75, 50\}$ dB, where for a
 718 uniform distribution

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

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

719 The estimation is performed using the same data trajectory $y(t)$ for each of the noise
 720 amplitudes. Trajectories of the experimental synchronization error $\text{SE}_s(t)$ are shown in the
 721 **Top** and **Bottom** panels of Fig. 5 for $\text{SNR} = \infty$ and $\text{SNR} = 100$ dB respectively. Each plot

722 includes traces for $D_M = \{1, 8, 9, 10, 12, 14\}$.

723 When no noise is present, a clear transition to synchronization is evident between $D_M = 8$
724 and $D_M = 9$ for the extended system with 21 degrees-of-freedom. This allows us to identify
725 $L_c \approx 9$, which is in agreement with previous work [24]. This rule also holds when the
726 SNR = 100 dB. In this case, our results show the synchronization error quickly converges
727 down to the approximate level of the noise. However, as the SNR is further decreased, this
728 transition becomes less apparent. For SNR = {75, 50} (not shown), the fluctuations of the
729 synchronization error are roughly the order of magnitude of the noise.

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

735 Since this is a twin experiment, we may check the parameter estimates directly. These
736 results are shown in Table III. As expected the estimates for SNR \geq 100 are accurate
737 when $D_M \geq 9$. However, as noise levels are further increased however, the accuracy of the
738 estimates deteriorates markedly. In this regime, increasing D_M seems detrimental to the
739 parameter estimates. This sensitivity may indicate instability in the pseudoinverse. Indeed,
740 we have checked that further decreasing the SNR causes the calculations to become unstable
741 with $D_M > 10$.

742 Reducing the rank of the inverse stabilizes the calculations, but does not improve the
743 estimates in this case. However, we have seen evidence that results may be improved by
744 choosing a larger D_M while fixing the inverse rank at a lower value to ensure stable calcula-
745 tions (e.g. $r = 10$). Increasing the time delay τ has also been observed to improve robustness
746 to noise as the addition of time delay coordinates tends to act as a low pass filter. However,
747 there is a trade-off with this tactic. As the length of the time delay vector gets long with
748 respect to the Lyapunov time, the inverse of the largest Lyapunov exponent, the $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$
749 matrix becomes more ill-conditioned and small errors in the data are amplified. Conse-
750 quently, a good method for choosing the rank of the matrix is especially crucial when noise
751 is involved and when the maximum time delay time $D_M \tau$ is long. Furthermore, we expect
752 the noise robustness to be further improved by adapting the coupling terms $\mathbf{g}(t)$, $\mathbf{g}'(t)$ in
753 some ‘optimal’ manner that incorporates estimates for the error covariance, such as what is

754 done for the Kalman-Bucy filter.

755 2. $D = 10$ with different forcing for each oscillator

756 Our next example uses the Lorenz 1996 model Eqn. (20) with $D = 10$ and *different* values
757 for the forcing parameters for each dimension ($p_1 \rightarrow p_i$ for $i = \{1, \dots, D\}$). The values of
758 these parameters are given in Table IV and are selected in this way to break the symmetry
759 of the original model. Proceeding as usual, we construct the extended system consisting of
760 $D + D = 20$ states and parameters to perform the estimates. All other parameters remain
761 the same as the previous example. Also, no additional measurement noise was included in
762 this simulation so $\text{SNR} = \infty$.

763 Fig. 8 shows the temporal evolution of the synchronization error $\text{SE}_s(t)$ for different
764 delay dimensions D_M . While $D_M \leq 5$ is not sufficient for achieving synchronization, the
765 simulation with $D_M = 6$ shows a slow convergence to zero and $D_M = 10$ exhibits a clear
766 and fast transition to synchronization. This allows us to identify $L_c \approx 6$ for the extended
767 system. This is confirmed in Fig. 9 where in the top panel the predictions fail for $D_M = 1$
768 and $D_M = 5$, but succeed for $D_M = 10$ as shown in the bottom panel.

769 **V. DIRECT ESTIMATION OF L_c**

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

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

782 Consider a long trajectory $\mathbf{x}(t_n)$ generated by Eqn. (2) and sampled at discrete times
 783 $n = \{0, 1, \dots, N\}$. At each point t_n , evaluate $\partial\mathbf{F}(\mathbf{x}(t_n))/\partial\mathbf{x}$ and construct its SVD,

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

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

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

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

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

795 The reason for this, we argue, is related to the fact that incorporating parameters into
 796 the model modifies the spectrum of the Jacobian $\partial\mathbf{F}(\mathbf{x}(t))/\partial\mathbf{x}$ to have singular values that
 797 are close to zero. These ‘slightly’ unstable dimensions tend to get ‘averaged out’ so to speak,
 798 when the assimilation window is long, allowing synchronization to occur with fewer than D
 799 measurements. Similar behavior was observed for the Lorenz 1963 and Rössler systems.

800 These results further strengthen our argument that L_c closely related to the number of
 801 locally unstable directions in phase space, or more precisely, the ergodic average of this
 802 quantity. Also, the fact that the critical number of time delays is approximately equal to
 803 L_c supports the idea that to successfully synchronize the model with the observed data,
 804 one requires the set of measurements (either physical or time delayed) to span the unstable
 805 subspace of the dynamics.

806 This idea of incorporating information from time delayed measurements to regularize the
 807 search for the correct model states and parameters is not new by any means. In particular, we
 808 have recently discovered that the method discussed here and in [17] (also proposed earlier
 809 in [43]) is fundamentally equivalent to a control theoretic construct known as a Newton

810 observer, which was first introduced by Moraal and Grizzle in [44, 45]. The idea is that by
 811 using time-delays in this way, the perturbation $\delta\mathbf{x}(t)$ is essentially the Newton step associated
 812 with the observability equation. We elaborate this point in more detail below.

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

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

815 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
 816 function $\mathbf{h}(\mathbf{x})$ with respect to the vector field \mathbf{F} [36, 46]. The system is said to be *locally*
 817 *observable* at a point \mathbf{x}_0 if and only if

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

818 When the system is locally observable at a point \mathbf{x}_0 , there exists a neighborhood Ω such
 819 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 $\mathbf{h}(\mathbf{z}) \neq$
 820 $\mathbf{h}(\mathbf{x}_0)$. Intuitively, this means that at the point \mathbf{x}_0 there is enough information from the
 821 measurement and the dynamics to infer the entire state of the true, physical system.

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

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

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

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

826 For this process to succeed, the system must be locally observable so that the Jacobian
 827 $\partial\Phi(\mathbf{x})/\partial\mathbf{x}$ has full rank [36, 46].

828 The vector-valued functions $\Phi(\mathbf{x})$ and $\mathbf{S}(\mathbf{x})$ are similar in that they both contain infor-
 829 mation about the time-evolution of the states. In particular, $\mathbf{S}(\mathbf{x})$ can be considered a time

830 delayed version of $\Phi(\mathbf{x})$. While $\Phi(\mathbf{x})$ is easier to work with analytically, performing the
 831 inversion of Eqn. (23) is rarely useful in practice, as it requires one to measure high-order
 832 derivatives of the data or approximate them with finite differences. The latter approach is
 833 numerically unstable when measurement noise is present, as the finite difference approxima-
 834 tion acts as a high-pass filter [46].

835 The time delay formulation on the other hand, does not have this problem since the
 836 derivatives on the left hand side of Eqn. (23) are replaced with time delayed values of
 837 the measurements. As Takens noted [27], time delays carry the same information as the
 838 derivatives but are far less sensitive to measurement noise. The same Newton’s method
 839 approach can be performed using time delays,

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

840 Note that this process is *static*. That is, it is carried out at a single time t . Compare this
 841 with the *dynamic* process in Eqn. (10), for which the control perturbation is essentially the
 842 right hand side of Eqn. (24). The immediate connection between the ‘observation space’
 843 $\Phi(\mathbf{x})$ and the time delay space $\mathbf{S}(\mathbf{x})$ suggests that the static process Eqn. (24) can only
 844 converge to the correct solution when $\partial \mathbf{S}(\mathbf{x})/\partial \mathbf{x}$ has full rank.

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

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

847 to converge at a desired rate [36, 46]. In our numerical experiments, we observe precipitous
 848 drops in the synchronization error in regions where $\partial \mathbf{S}(\mathbf{x})/\partial \mathbf{x}$ is well-conditioned enough to
 849 construct the full rank inverse. We consider this empirical evidence for the correspondence
 850 between our time delay approach and observability.

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

857 When the observability condition is not satisfied the static process fails. There is however,
 858 a weaker condition known as ‘detectability’, which requires all of the unobservable modes of

859 the system to decay asymptotically [36, 46]. If this condition holds, the dynamical process
860 will still succeed as we are able to control all of the locally unstable directions associated
861 with error growth.

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

870 The purpose of this discussion has been to introduce a direct estimate Eqn. (21) for L_c
871 in terms of the average number of unstable directions in the dynamics and to acknowledge
872 the apparent connection with observability. These ideas have had some mention in the data
873 assimilation literature. For instance, the unstable dynamical subspace has been used for
874 selecting ensemble members in ensemble forecasts and for identifying sensitive regions to
875 targeted for further observation [41]. Also, optimization-based approaches such as moving
876 horizon estimation seek to incorporate a moving time window of observations [47, 48]. How-
877 ever, the true value of the Newton observer (time-delay synchronization) technique lies in
878 its ability to deal with poorly observable system in a systematic way [45]. It was with such
879 systems in mind that we independently rediscovered the work of Moraal and Grizzle some
880 two decades later, as these systems are altogether common in applications where the num-
881 ber of degrees of freedom in the model far exceeds the number of observations. With that
882 said, we now turn to our final example, which involves the estimation of a small network of
883 chaotic oscillators.

884 VI. NETWORKS OF CHAOTIC OSCILLATORS

885 One particular goal for our time delay method is to provide a means to analyze networks
886 of oscillators, such as those found in nervous systems. As in practical geophysical dynamics
887 (for example, numerical weather prediction) sparse measurements of the network behavior
888 under selected forcing is to be expected. One strategy [11] for understanding the underlying

889 physical properties of such problems is to analyze carefully the properties of the nodes,
 890 namely the specific oscillators such as the ones we have covered here, and then use the same
 891 approach to analyze the nature and strengths of the couplings among the oscillators at the
 892 nodes to complete a model for the network as a whole.

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

899 A. A small network of chaotic Colpitts oscillators

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

904 In particular, we investigate a ring of oscillators with connected with unidirectional cou-
 905 pling. The state of each oscillator is given by $x_a^i(t)$, where $i = \{1, 2, 3\}$ is the node index and
 906 $a = \{1, 2, 3\}$ denotes three internal state variables for each node. The dynamical equations
 907 are given by

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

908 where the indices are permuted cyclically so that $\mathbf{x}^{(M+1)}(t) = \mathbf{x}^{(1)}(t)$. The parameters
 909 $c_{(i+1,i)} \geq 0$ are constant coupling constants that serve as connections among the individual
 910 oscillators.

911 The Colpitts oscillator is comprised of standard R, L, C components together with a
 912 single bipolar transistor. The only nonlinearity is the exponential function $\exp(-x_1^{(i)}(t))$
 913 coming from the transistor dynamics. These equations are a rescaled representation of the
 914 physical equations of state. The derivation of these dynamical equations from Kirchoff's laws

915 is given in [14, 49]. The states x_1, x_2 , and x_3 respectively correspond to the voltage between
 916 the transistor emitter and its base, the current through the inductor and the voltage at the
 917 transistor collector and its base.

918 When $p_1 \geq 3.5$ or so, the oscillator expresses chaotic behavior. Following [14], we select
 919 $p_1^{(i)} = 5.0$, $p_2^{(i)} = 0.0797$, and $p_4^{(i)} = 0.6898$ for all three oscillators. To break the ring sym-
 920 metry, we select $p_3^{(1)} = 3$, $p_3^{(2)} = 3.5$, $p_3^{(3)} = 4$ as well as $c_{(2,1)} = 0.8$, $c_{(3,2)} = 0.9$, $c_{(1,3)} = 1.0$.
 921 Direct integration of Eqn. (25) confirms that the individual oscillators do not synchronize
 922 with each other. This is important, as a synchronized network may require fewer measure-
 923 ments than an unsynchronized network. Indeed, synchronization of oscillators in a network
 924 may allow population behaviors by effectively reduce the degrees of freedom of the network
 925 in a functional manner.

926 **B. Estimating the states of the network**

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

933 Results for the state estimation procedure are shown in the **Top** panel of Fig. 11. The
 934 experimental synchronization error is plotted as a function of time for $D_M = \{1, 3, 5, 9\}$.
 935 $D_M = \{1, 2\}$ is insufficient; one needs $D_M = 3$ to achieve synchronization. Our analysis
 936 of estimates and predictions for individual states verified that $D_M = 3$ indeed produces
 937 excellent predictions, whereas $D_M = \{1, 2\}$ does not. This result gives an estimate of $L_c \approx 3$
 938 for the case under consideration, where only the states are to be determined.

939 Furthermore, the rate of convergence does not increase monotonically with the number
 940 of measurements. That is, $D_M = \{4, 5\}$ have a slower convergence rate than $D_M = 3$, and
 941 $D_M = 6$ does not appear to converge at all. This illustrates the importance of the proper
 942 choice of D_M , as there is a trade-off between the rate of convergence and the stability of the
 943 procedure. Note that the adaptive rank algorithm did not impact this result because apart
 944 from about 100 time steps at the beginning of the assimilation window, full rank was used

945 i.e., $r = \min(D, D_M)$.

946 In addition, note that the fastest rate of convergence is achieved with $D_M = 9$. This
947 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
948 be inverted exactly. While theoretically, such an embedding allows the entire state to be
949 reconstructed instantaneously at a single time t , in practice the matrix is often too ill-
950 conditioned for this technique to be of use. The adaptive rank algorithm counteracts this
951 numerical instability, by selecting the largest rank r that produces a stable perturbation.
952 In this case however, we observe that as the estimated state approaches the true value, the
953 adaptive rank algorithm selects a full rank inverse $r = D_M = D$. This indicates that the
954 time delay construction is well-conditioned enough so that the exact inverse can be used to
955 generate a perturbation $\delta\mathbf{x}$ that is small relative to the dynamics. When this happens, the
956 estimate converges remarkably quickly to the true result, as evidenced by the steep dive for
957 the $D_M = 9$ trace in the **Top** panel of Fig. 11.

958 We argue that this accelerated convergence brought about by the full-rank inverse of
959 $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ is intrinsically related to the observability condition familiar from control theory.
960 As we suggested earlier, the $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$ can be considered a time delayed version of the
961 observability matrix. When a well-conditioned, full-rank inverse exists, the error between
962 the true and estimated states is well-approximated by $\delta\mathbf{x}$ and the estimate converges quickly
963 to its true value. Thus, while $D_M = 3$ time delays appears to be necessary to stabilize the
964 chaotic subspace of the dynamics to provide asymptotically stable convergence, selecting
965 $D_M = 9$ provides rapid convergence that is less numerically stable.

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

971 **C. Estimating the states and the couplings between nodes**

972 Next, we fix the parameters $\mathbf{p}^{(i)}$ to their known values and include the internode couplings
973 $c_{(i+1,i)}$ in the estimation procedure. This is directly relevant to analyses of neuron networks
974 where we may have some knowledge of the cells individually, but we wish to explore the

975 connectivity which underlies the function of the network.

976 We proceed as before, using the same time delay $\tau = 0.2 = 20 \Delta t$ and coupling gain
977 $g \Delta t = 0.1$ as well as the adaptive rank selection with $\epsilon = 10$. The initial conditions for
978 the ring coupling parameters are chosen to be one half of their correct values i.e. $c_{(2,1)} =$
979 0.4 , $c_{(3,2)} = 0.45$, $c_{(1,3)} = 0.5$.

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

987 As in previous sections, we validate the results as if this were an actual experiment. In
988 the **Top** and **Bottom** panels of Fig. 12 we plot the estimate and predicted trajectories of
989 the observed $x_1^{(1)}(t)$ state component for $D_M = 4$ and $D_M = 5$ respectively. As expected,
990 the prediction for $D_M = 4$ is poor despite that its estimate looks quite accurate. This once
991 again demonstrates the necessity of using predictions to validate the quality of a model's
992 consistency with experimental results. On the other hand, the $D_M = 5$ estimate produces
993 accurate predictions that do not diverge for a considerable time after the end of the assimila-
994 tion window (largest LE is $\mathcal{O}(10^{-1})$). Similar results are obtained for $D_M = 9$ and $D_M = 12$
995 although these trajectories are not shown.

996 Estimated values for the ring coupling parameters are shown in Table V. As expected,
997 the estimates are accurate only when $D_M \geq 5$. This result demonstrates the potential of
998 the time delay procedure for performing state and parameter estimation on a network of
999 coupled chaotic oscillators. Specifically, it shows that the waveform of a scalar signal from
1000 a state of a single oscillator carries enough information to determine both the states of the
1001 neighboring nodes in the network as well as the coupling parameters that determine the
1002 interaction between the neighbors. This of course assumes that the network topology is
1003 known. In the next subsection, we dispense with this assumption and investigate whether
1004 the algorithm is capable of determining the functional connectivity of this simple network.

1005 **D. Estimating the functional connectivity of the network**

1006 Until this point, our model has been constructed assuming that the connectivity of the
 1007 network is known, but the strength of the connections is not. In many practical applications
 1008 however, this information is not available. For instance, when modeling small neurobiological
 1009 networks, one often has some idea of the number of nodes in the network, and perhaps
 1010 even some notion of their physical connectivity. The functional connectivity of the network
 1011 however (i.e., the relative synaptic strengths) is generally unknown and therefore must be
 1012 determined from experimental data. We now investigate this prospect of network topology
 1013 estimation, within the context of our simple Colpitts network.

1014 To this end, we expand our network model Eqn. (25) to include coupling in both directions
 1015 so that the dynamical equations for the $x_1^{(i)}$ states are now, with $c_{(j,i)} \geq 0$,

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

1016 The twin experiment data is generated as before, so that the true values of the additional
 1017 coupling parameters are $c_{(1,2)} = c_{(2,3)} = c_{(3,1)} = 0$. The initial conditions for these couplings
 1018 are chosen to be symmetric, so that $c_{(2,1)} = c_{(1,2)} = 0.4$, $c_{(3,2)} = c_{(2,3)} = 0.45$, $c_{(1,3)} = c_{(3,1)} =$
 1019 0.5 . All other parameters are the same as before.

1020 Also, note that all self-couplings $c_{(i,i)}$ are implicitly zero. This however, is just a con-
 1021 sequence of how the network coupling model was defined; the procedure may be easily
 1022 generalized to estimate self-coupling parameters as well.

1023 In the **Bottom** panel of Fig. 11, we display experimental synchronization error trajec-
 1024 tories for $D_M = \{5, 6, 8, 10, 12, 15\}$. Results are similar to those shown in the previous sub-
 1025 section. Synchronization requires $D_M \geq 8$ time delays and the full rank $r = D_M = D = 15$
 1026 inverse synchronizes rapidly. Results with $D_M = 7$ were numerically unstable and are not
 1027 shown. Known, estimated and predicted trajectories for the observed $x_1^{(1)}(t)$ are shown in
 1028 the **Top** and **Bottom** panels of Fig. 13 for $D_M = 6$ and $D_M = 8$ respectively. As anticipated
 1029 from the synchronization error results, the prediction for $D_M = 8$ is quite accurate whereas
 1030 for $D_M = 6$ it is not. The estimated coupling parameters shown in Table VI further confirm
 1031 this result. Only the estimates made with $D_M \geq 8$ time delays are accurate, allowing us
 1032 to identify $L_c \approx 8$ for this extended model, in which the connectivity of the network is
 1033 unknown.

1034 The main point of this calculation was to demonstrate that the time delay method is
1035 capable of determining the function connectivity of a network of chaotic oscillators, within
1036 the assumption that the model is known: both for the internal dynamics within a node
1037 and the coupling between nodes. In particular, we have shown that knowing $x_1^{(1)}(t)$ alone
1038 is enough to determine the functional connectivity of this small network of three Colpitts
1039 oscillators. That is, we are able to correctly estimate the values of both the forward and
1040 backward couplings, the latter of which are zero. Furthermore, we have learned that ex-
1041 panding the model in this way (i.e. to include coupling in both directions) increases the
1042 number of required time delays from $D_M = 5$ to $D_M = 8$.

1043 We remark however, that this case of $M = 3$ is exceedingly simple, requiring estimation
1044 of only three additional parameters (the backwards couplings). For a general network, the
1045 number of coupling terms to be estimated grows as M^2 and thus we expect that at some M
1046 a single, scalar measurement will not be enough.

1047 Nonetheless, the twin experiment framework presented here, together with the time delay
1048 algorithm, supplies crucial information about the observability of the system being studied.
1049 It provides for instance, a lower bound estimate on the number of required measurements and
1050 also offers a way to determine which nodes should be targeted for observation: since some
1051 nodes may provide more data than others. Such information would be highly beneficial
1052 for experimental design purposes, as it allows one to directly investigate the constraints
1053 imposed by one's limited measurement capability. In other words, our framework allows
1054 one to determine, in principle, whether enough measurements are available to successfully
1055 determine the connectivity of the network, and predict its subsequent behavior.

1056 VII. DISCUSSION AND SUMMARY

1057 The idea of using the waveform of measurements—that is, the measurement at time t
1058 and its time delays—has been investigated in the context of synchronization-based state
1059 and parameter estimation for chaotic dynamical systems as a means to transfer additional
1060 information from observed data to a model. An algorithm has been presented that uses
1061 this additional information to generate dynamical coupling between the data and model
1062 systems and its capability has been demonstrated using the Lorenz 1963 and 1996 models,
1063 the 4-dimensional ‘hyperchaotic’ Rössler model, as well as recurrent networks of chaotic

1064 oscillators.

1065 These examples demonstrate that when only one state variable is observed, utilizing
1066 $D_M \geq L_c$ time delays stabilizes the synchronization manifold enough to enable accurate
1067 estimation of unknown states and parameters, and permit accurate predictions beyond the
1068 observation window. In this way, the time delays are capable of significantly reducing the
1069 number of measurements required to achieve accurate estimates and reliable predictions.

1070 In practice, the number of available measurements is often tightly constrained (e.g., by
1071 cost or technological considerations) and are typically sparse compared with the number of
1072 degrees of freedom of the model. For instance, in the analysis of a shallow water model of
1073 geophysical flow, it was shown that $L_c \approx 0.7 D$ [16], while in operational weather prediction
1074 systems (such as the European Centre for Medium Range Weather Forecasts) only about
1075 10^7 measurements are typical for models with 10^8 or 10^9 degrees of freedom [50].

1076 When additional measurements are unavailable, time delays offer another means to fur-
1077 ther stabilize the search space. Regarding the shallow water model for instance, recent work
1078 has shown that by using time delays in this way, enough information is extracted from the
1079 height field alone to permit synchronization between the data and the model [51]. These
1080 results demonstrate a proof-of-concept that time delays may be used to effectively reduce
1081 the total number of measurements required to achieve this goal.

1082 The form of the time delayed coupling has some desirable properties as well. For instance,
1083 in the case where $D_M = 1$ it reduces to the classical form Eqn. (3). Also, when $D_M > 1$, it
1084 generates control perturbations on *all* state components and our results have shown that,
1085 by including the parameters as state variables augmented with trivial dynamics $d\mathbf{p}/dt = 0$,
1086 parameter estimation occurs as a natural result of the synchronization process. This is
1087 an improvement over the classical (i.e. $D_M = 1$) form, which typically requires nonlinear
1088 optimization techniques to estimate the parameters. Additionally, one could use this method
1089 in conjunction with other nonlinear estimation procedures as means to improve the estimate
1090 when $L < L_c$.

1091 There also appears to be a direct correspondence between the sufficient number of mea-
1092 surements L_c and the number of time delays required to stabilize the synchronization man-
1093 ifold. This is interesting for a number of reasons. For one, although it is reasonable that in
1094 each case there should exist such a sufficient condition, there is no reason to assume *a priori*
1095 that they should be the same. The fact that they appear to be (roughly) equal indicates

1096 that this condition may be an invariant property of the dynamics. Indeed, we have observed
 1097 the same phenomenon using other approaches (e.g. variational optimization and Markov
 1098 Chain Monte Carlo [49, 52]), which suggests these other methods may also benefit from the
 1099 inclusion of time delays.

1100 This result also highlights clearly the distinction between the use of time delays here,
 1101 for the purpose of state and parameter estimation, and its familiar application in nonlinear
 1102 dynamics for reconstructing the phase space of a partially observable dynamical system. For
 1103 instance, the Kaplan-Yorke dimension [53] for the $D = 20$ Lorenz 1996 system is $D_A \approx 12$, so
 1104 the sufficient dimension for phase space reconstruction is [28] $2D_A \approx 24$, whereas the required
 1105 number of time delays is $D_M \approx 9$. The time delays stabilize the synchronization manifold
 1106 using a fraction of the sufficient number of delays needed for full attractor reconstruction.

1107 We also wish to note that, in practice, there is a finite limit to the amount of information
 1108 available from the time delays of a single scalar time series. For instance, with the Lorenz
 1109 1996 system we observed that, regardless of the chosen dimension D , a threshold occurs
 1110 around $D_M \approx 12$. Continuing to increase D_M beyond this threshold causes the $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$
 1111 matrix to become highly ill-conditioned, and therefore requires a lower choice of rank to
 1112 maintain stable computations. We suspect that restricting the rank in this way effectively
 1113 limits the number of stable dimensions transferred from the control coupling. In other words,
 1114 we have seen evidence that there exists a correspondence between the required number of
 1115 measured states L_c , the number of time delays D_M and the rank r of the inverse. Given the
 1116 threshold $D_M \approx 12$ and the empirical scaling rule $L_c \approx 0.4 D$, this suggests that we should
 1117 not be able to synchronize a system with $D > 25$ using only a single measurement, which
 1118 is indeed the limit observed in our numerical experiments (although these results are not
 1119 shown here).

1120 We further suggest that this threshold is due to a limited amount of information avail-
 1121 able in a time-series that is locally bounded by the characteristic time-scale of the chaos.
 1122 That is, holding τ fixed and increasing D_M to extend the time delay vector far beyond the
 1123 Lyapunov time should not provide any additional information, as the later points are too
 1124 far decorrelated to be of any use. Likewise, increasing D_M by decreasing τ and holding the
 1125 total length fixed should in principle provide enough information. However, in this case we
 1126 are restricted by the noise level of the system (or if no noise is present, by finite numerical
 1127 precision). Both cases are indicated by ill-conditioning of $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$, and the threshold on

1128 D_M likely a manifestation of both effects.

1129 Thus, it is crucial to the success of the algorithm that the parameters D_M and τ are
1130 chosen appropriately. They must be large enough to provide additional information about
1131 the unobserved states, yet not so large as to induce numerical instability in the calculation
1132 of the inverse. There exist many techniques for choosing τ that have been developed for
1133 the purposes of attractor reconstruction, such as the first minimum of average mutual in-
1134 formation between measurements. These methods are likely to be applicable here as well,
1135 although for the examples shown here changing τ by a few dt did not noticeably impact the
1136 results.

1137 Also, whereas here we have only considered forward time delays, recent work [54, 55] has
1138 shown that a mixture of forward and backward delays can further improve the conditioning
1139 of $\partial\mathbf{S}(\mathbf{x})/\partial\mathbf{x}$. Whether or not mixed delays provides superior results for synchronization is
1140 currently under investigation.

1141 Moreover, although these examples have been limited to the case where the number of
1142 measurements $L = 1$, our formulation generalizes easily to the case where $L > 1$. In partic-
1143 ular, given D_{M_i} time delays available in each of $i = \{1, \dots, L\}$ measurements, the number
1144 of time delays required to stabilize the estimation should satisfy $\sum_{i=1}^L D_{M_i} > L_c$. Note that
1145 this is only a rough approximation because it is quite clear that the amount of information
1146 contained in each state component is different in general, and not additive, in the sense that
1147 measuring two mutually dependent variables may not provide as much information as each
1148 variable contributes individually. We remark however, that the twin experiment framework
1149 is a useful tool for determining the relative value of a given measurement. Such information
1150 is essential for analyzing the costs and benefits of obtaining further measurements.

1151 The inclusion of time-delays comes of course with an additional computational cost,
1152 mainly associated with the integration steps required to construct the time delay vectors
1153 and its Jacobian, as well as solving for the perturbation itself. The baseline for comparison
1154 is the simple nudging algorithm Eqn. (3), which is recovered in the limit $D_M = 1$. Certainly,
1155 clever algorithmic improvements are required in order to reduce this overhead as much as
1156 possible. For instance, one idea is to reduce the resolution of the model, initialize it with
1157 existing measurements, run the assimilation and then interpolate, to recapture the desired
1158 resolution for forward prediction. It may also be possible to update $\mathbf{S}(\mathbf{x})$ directly with
1159 the perturbation, so that it does not need to be recomputed in its entirety at each time

1160 step. Other such ideas will surely emerge as well, while the technique is scaled up to larger
1161 problems.

1162 Finally, extension of this method to more complex models, or high-dimensional models
1163 representing numerical approximations to partial differential equations appears possible. In
1164 particular, applications of this approach for numerical weather prediction or the analysis of
1165 biological neural networks are currently under investigation. These applications typically
1166 permit too few measurements than are required to stabilize the estimation procedure and
1167 the results presented here suggest that the incorporation of time delays will allow us to
1168 extract more information from *existing measurements* to improve our state and parameter
1169 estimates and generate more accurate predictions.

1170 ACKNOWLEDGEMENTS

1171 The research leading to the results has received funding from the European Community's
1172 Seventh Framework Program FP7/2007-2013 under grant agreement no. HEALTH-F2-2009-
1173 241526, EUTrigTreat. This work was funded in part under a grant from National Science
1174 Foundation (PHY-0961153). Partial support from the Department of Energy CSGF pro-
1175 gram (DE-FG02-97ER25308) for D. Rey is appreciated. Partial support has come from the
1176 ONR MURI Program (N00014-13-1-0205). We thank S. Luther for discussions on state and
1177 parameter estimation and for his continuing support.

-
- 1178 [1] A. C. Lorenc and T. Payne, "4D-Var and the butterfly effect: Statistical four-dimensional
1179 data assimilation for a wide range of scales," *Q. J. R. Meteorol. Soc.* **133**, 607 (2007).
- 1180 [2] Maria Rodriguez-Fernandez, Jose A Egea and Julio R Banga, "Novel metaheuristic for pa-
1181 rameter estimation in nonlinear dynamic biological systems," *BMC Bioinformatics* **7** (2006)
1182 483.
- 1183 [3] A. Pokhilko *et al.*, "Data assimilation constrains new connections and components in a com-
1184 plex, eukaryotic circadian clock model," *Mol. Sys. Bio.* **6**, 416 (2010).
- 1185 [4] A. Horva'th and D. Manini, "parameter estimation of Kinetic Rates in Stochastic Reaction
1186 Networks by the EM Method," *International Conference on BioMedical Engineering and In-*

- 1187 *formatics*, 2008.
- 1188 [5] R. Xiong, P.J. Wissmann, and M.A. Gallivan, “An extended Kalman filter for in situ sensing
1189 of yttria-stabilized zirconia in chemical vapor deposition.” *Computers & Chemical Engineering*
1190 **30**, 1657 (2006).
- 1191 [6] D. Dochain, “State and parameter estimation in chemical and biochemical processes: a tuto-
1192 rial,” *Journal of Process Control* **13**, 801 (2003).
- 1193 [7] Z. Yang and J.M. Hamrick, “Optimal control of salinity boundary condition in a tidal model
1194 using a variational inverse method,” *Estuarine, Coastal and Shelf Science* **62**, 13 (2005).
- 1195 [8] N. Margvelashvili *et. al.*, “Satellite data assimilation and estimation of a 3D coastal sediment
1196 transport model using error-subspace emulators,” *Environmental Modelling & Software* **40**,
1197 191 (2013).
- 1198 [9] R. H. Clayton *et. al.*, “Models of cardiac tissue electrophysiology: Progress, challenges and
1199 open questions,” *Prog. Biophys. Mol. Biol.* **104**, 22 (2011).
- 1200 [10] Q. J. Huys and L. Paninski, “Smoothing of, and parameter estimation from, Noisy Biophysical
1201 Recordings,” *PLoS Comput Biol.* **5**, (2009).
- 1202 [11] D. Meliza, M. Kostuk, H. Huang, A. Nogaret, H. D. I. Abarbanel, and D. Margoliash, unpub-
1203 lished (2013).
- 1204 [12] H. D. I. Abarbanel, *Predicting the Future: Completing Models for Observed Complex Systems*,
1205 Springer-Verlag, New York, 2013.
- 1206 [13] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, (2nd ed.), Springer, (2008).
- 1207 [14] H.D.I Abarbanel, D. R. Creveling, R. Farsian, and M. Kostuk, “Dynamical State and Param-
1208 eter Estimation,” *SIAM J. Appl. Dyn. Syst.* **8**, 1341 (2009).
- 1209 [15] J. C. Quinn, *A Path Integral Approach to Data Assimilation in Stochastic Nonlinear Systems*,
1210 UCSD, Physics PhD, June, 2010.
- 1211 [16] W. Whartenby, J. Quinn, and H. D. I. Abarbanel, “The Number of Required Observations
1212 in Data Assimilation for a Shallow Water Flow,” *Monthly Weather Review* **141**, 2502-2518
1213 (2013).
- 1214 [17] D. Rey, M. Kostuk, M. Eldridge, H. D. I. Abarbanel, J. Schumann-Bischoff, and U. Parlitz,
1215 “Accurate State and Parameter Estimation in Nonlinear Systems with Sparse Observations,”
1216 *Physics Letters A.* **378**, 869-873 (2014).

- 1217 [18] J. E. Hoke and R. A. Anthes, “The Initialization of Numerical Models by a Dynamic-
1218 Initialization Technique”, *Monthly Weather Review* **104**, 1551-1156 (1976).
- 1219 [19] S. Lakshmivarahan and John M. Lewis, “Nudging Methods: A Critical Overview,” Chapter
1220 2 in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, (Vol. II), Seon
1221 Ki Park and Liang Xu, Editors, Springer (2013).
- 1222 [20] U. Parlitz, L. Junge and L. Kocarev, “Synchronization based parameter estimation from time
1223 series,” *Phys. Rev. E* **54**, 6253-6529 (1996).
- 1224 [21] J. Amezcua, K. Ide, E. Kalnay and S. Reiche, “Ensemble transform Kalman Bucy filters”, *Q.*
1225 *J. R. Meteorol. Soc.* **140**, 9951004 (2014).
- 1226 [22] L. Pecora and T. Carroll. *Phys. Rev. Lett.* 64 821 (1990).
- 1227 [23] E. N. Lorenz, “Deterministic nonperiodic flow,” *Journal of the atmospheric sciences* **20.2**,
1228 130-141 (1963).
- 1229 [24] Mark Kostuk, *Synchronization and statistical methods for the data assimilation of HVC neuron*
1230 *models*. UCSD, Physics PhD, June, 2012. <http://escholarship.org/uc/item/2fh4d086>
- 1231 [25] D. Aeyels, “Generic observability of differentiable systems,” *SIAM J. Control Optim.* **19**, 595
1232 (1981).
- 1233 [26] D. Aeyels, “On the number of samples necessary to achieve observability,” *Syst. Control Lett.*
1234 **1**, 92 (1981).
- 1235 [27] F. Takens, “Detecting strange attractors in turbulence,” In D. A. Rand and L.-S. Young.
1236 “Dynamical Systems and Turbulence,” Lecture Notes in Mathematics, vol. 898. Springer-
1237 Verlag, 366 (1981).
- 1238 [28] T. Sauer, J. A. Yorke, and M. Casdagli, “Embedology.” *J. Stat. Phys.* **65**, 579 (1991).
- 1239 [29] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer-Verlag, New York, 1996.
- 1240 [30] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, (2nd ed.), Cambridge University
1241 Press, Cambridge, UK, 2004.
- 1242 [31] Y. Hirata, H. Suzuki and K. Aihara, “Reconstructing state spaces from multivariate data
1243 using variable time delays,” *Phys. Rev. E* **74**, (2006).
- 1244 [32] A. Fraser and H. Swinney, *Phys. Rev. A* **33**, 2 (1986).
- 1245 [33] A. Tikhonov, A. Goncharsky, V. Stepanov, A. Yagola. *Numerical Methods for the Solution of*
1246 *Ill-Posed Problems*, Kluwer Academic Publishers (1995).

- 1247 [34] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, “Section 19.5. Linear Regularization Meth-
1248 ods,” *Numerical Recipes: The Art of Scientific Computing*, (3rd ed.), New York: Cambridge
1249 University Press (2007).
- 1250 [35] G. Golub and R. Christian, “Singular value decomposition and least squares solutions,” *Nu-
1251 merische Mathematik* **14.5**, 403-420 (1970).
- 1252 [36] H. Nijmeijer and A.V. der Schaft, *Nonlinear dynamical control systems*, Springer, 1990.
- 1253 [37] E. Candés, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate
1254 measurements,” *Communications on Pure and Applied Mathematics* **59**, 8 (2006).
- 1255 [38] O. E. Röessler, “An equation for hyperchaos,” *Physics Letters A* **71.2**, (1979).
- 1256 [39] E. N. Lorenz, “Predictability A problem partly solved,” in *Proceedings of the Seminar on
1257 Predictability*, Volume 1, ECMWF: Reading, UK, (1996) 1.
- 1258 [40] US 1624537, E. H. Colpitts, “Oscillation generator,” published 1 February 1918, issued 12
1259 April 1927
- 1260 [41] T. Palmer and R. Hagedorn, *Predictability of weather and climate*, Cambridge, 2006.
- 1261 [42] A. Karimi and M. R. Paul, “Extensive chaos in the Lorenz-96 model,” *Chaos: An Interdisci-
1262 plinary Journal of Nonlinear Science* 20.4 (2010).
- 1263 [43] L. Junge U. Parlitz, “Synchronization using dynamic coupling,” *Physical Review E* **64.5**,
1264 055204 (2001).
- 1265 [44] P. E. Moraal, “Nonlinear observer design: Theory and applications to automotive control,”
1266 Ph.D. dissertation, University of Michigan (1994).
- 1267 [45] P. E. Moraal and J. W. Grizzle. “Asymptotic observers for detectable and poorly observable
1268 systems.” In *IEEE Conf. Decision and Control*, New Orleans, 109-114 (1995).
- 1269 [46] C. Chi-Tsong, *Linear system theory and design*, Oxford University Press, 1995.
- 1270 [47] H. Michalska and D. Mayne, “Moving horizon observers and observer-based control,” *Auto-
1271 matic Control*, IEEE Transactions on 40.6, 995-1006 (1995).
- 1272 [48] W. Kang, “Moving horizon numerical observers of nonlinear control systems,” *Automatic
1273 Control*, IEEE Transactions on 51.2, 344-350 (2006).
- 1274 [49] J. C. Quinn, P. H. Bryant, D. R. Creveling, S. R. Klein, and H. D. I. Abarbanel, “State and
1275 Parameter and State Estimation of Experimental Chaotic Systems Using Synchronization,”
1276 *Physical Review E* **80**, (2009).

- 1277 [50] Lectures of L. Isaksen and S. English at the European Centre for Medium-Range Weather
1278 Forecasts Training Course, June, 2013.
- 1279 [51] Z. An, D. Rey, and H. D. I. Abarbanel, “Estimating the State of a Geophysical System with
1280 Sparse Observations.” *Monthly Weather Review* (2014 under review).
- 1281 [52] J. Quinn, and H. D. I. Abarbanel. “State and parameter estimation using Monte Carlo eval-
1282 uation of path integrals,” *Quarterly Journal of the Royal Meteorological Society* **136**, 652
1283 (2010).
- 1284 [53] J. L. Kaplan and J. A. Yorke, “In Functional Differential Equations and Approximations of
1285 Fixed Points,” Proceedings, Bonn, July 1978. Berlin: Springer-Verlag, 204, 1979.
- 1286 [54] U. Parlitz, J. Schumann-Bischoff, and S. Luther, “Local observability of state variables and
1287 parameters in nonlinear modeling quantified by delay reconstruction,” *Chaos: An Interdisci-
1288 plinary Journal of Nonlinear Science* **24.2**, (2014).
- 1289 [55] U. Parlitz, J. Schumann-Bischoff, and S. Luther, “Quantifying uncertainty in state and pa-
1290 rameter estimation.” *Physical Review E* **89.5** (2014).

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

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

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

D_M	SNR = ∞	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

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 $p_1 = 8.17$.

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

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

	Estimated	Estimated	Estimated
D_M	$c_{(2,1)}$	$c_{(3,2)}$	$c_{(1,3)}$
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

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

	Estimated	Estimated	Estimated
D_M	$c_{(2,1)}$	$c_{(3,2)}$	$c_{(1,3)}$
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

	Estimated	Estimated	Estimated
D_M	$c_{(1,2)}$	$c_{(2,3)}$	$c_{(3,1)}$
5	2.3385	-2.8119e-1	3.6094
6	-1.9911	5.2784e-1	3.9568e-1
8	3.8706e-11	-3.3491e-11	2.1787e-12
10	-8.3628e-10	1.0565e-09	2.0178e-10
12	1.0308e-05	-9.8415e-06	-1.4026e-07
15	-3.5215e-10	3.0340e-10	2.9756e-12

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 $D_M \geq 8$, the connectivity of the network is successfully predicted because the backwards couplings are estimated to be zero.

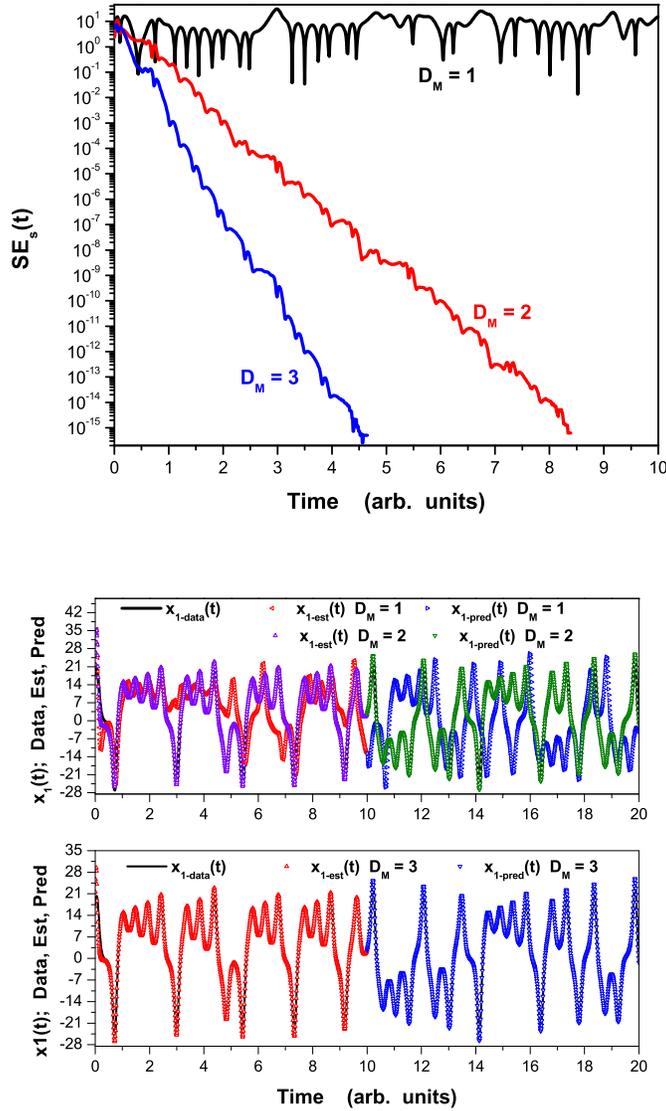


FIG. 1. (Color online) **Top:** Synchronization error $SE_s(t)$ for state estimation in the Lorenz 1963 system with $D_M = \{1, 2, 3\}$ and parameters fixed to their true values. One time delay does not synchronize the systems because the coupling is too small. However, selecting $D_M = 2$ or 3 generates 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 sub-panel) and $D_M = 3$ (bottom sub-panel). In agreement with the top panel the estimates/predictions made with $D_M = 1$ are poor whereas with $D_M = \{2, 3\}$ they are accurate.

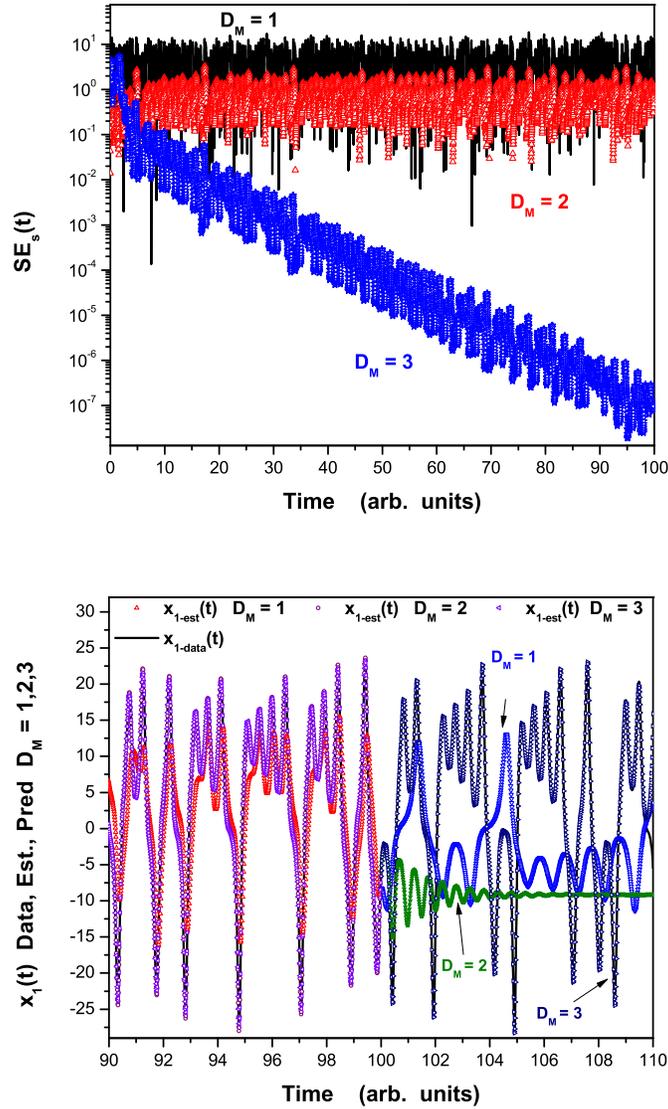


FIG. 2. (Color online) **Top** $SE_s(t)$ with $D_M = \{1, 2, 3\}$ for the extended Lorenz 1963 system where the parameters are treated as additional states. Three time delays are needed to synchronize. Including parameters can increase L_c . **Bottom** Data, estimates and predictions of the observed $x_1(t)$ 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.

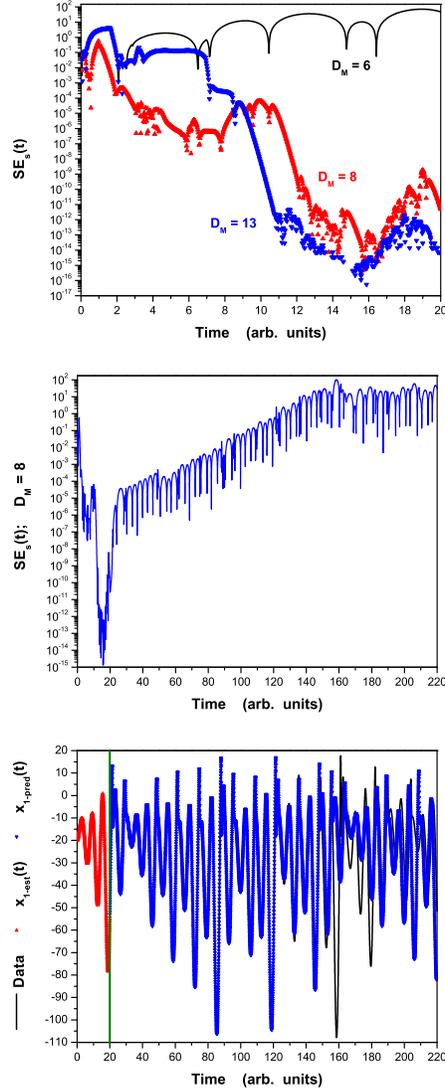


FIG. 3. (Color online) **Top** $SE_s(t)$ for $D_M = \{6, 8, 13\}$ for the Rössler system including parameters as state variables. **Middle** Long $SE_s(t)$ trajectory for $D_M = 8$. Trajectories begin to diverge immediately after the coupling is removed at $T = 20$. The rate of error growth is consistent 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 $SE_s(t)$ results.

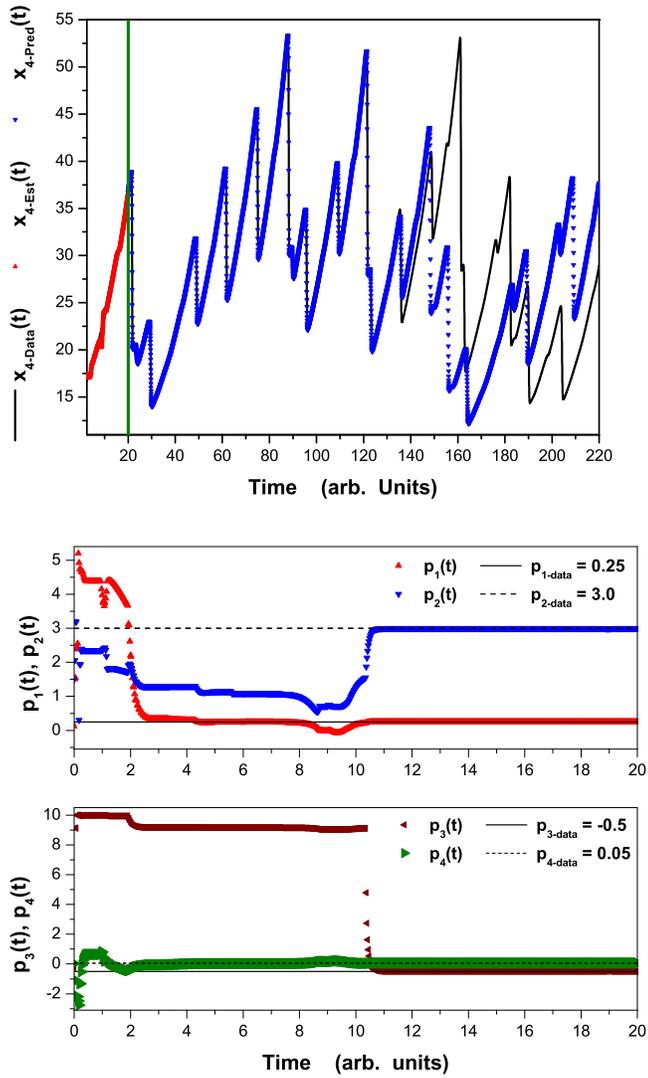


FIG. 4. (Color online) **Top** Known (black), estimated (red), and predicted (blue) time-series for an *unobserved* state variable $x_4(t)$ of the Rössler system with $D_M = 8$. The prediction fails near $t = 140$ due to the chaotic behavior of this system. Only in a twin experiment are we able to compare 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.

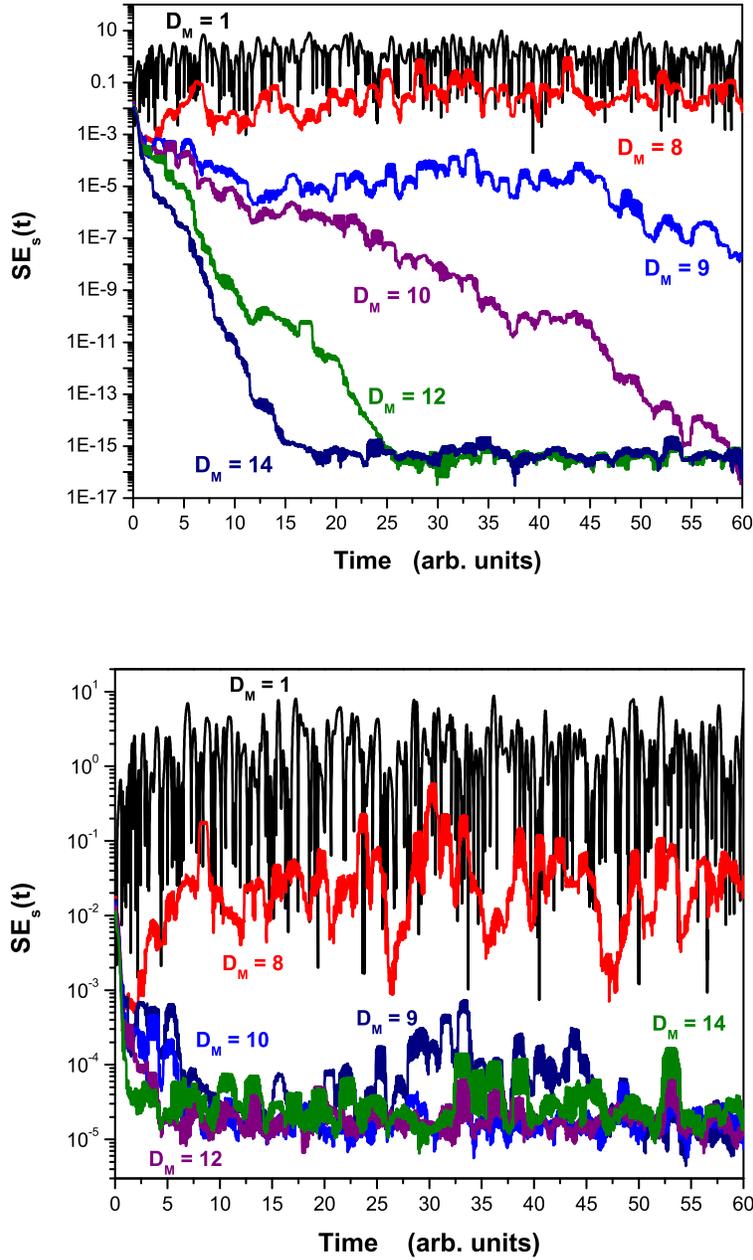


FIG. 5. (Color online) **Top** $SE_s(t)$ for the Lorenz 1996 system with $D = 20$, and augmented with a single forcing parameter. No noise is added to the measured state $x_1(t)$ so that $\text{SNR} = \infty$. Traces are shown for various $D_M = \{1, 8, 9, 10, 12, 14\}$. Synchronization is achieved with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

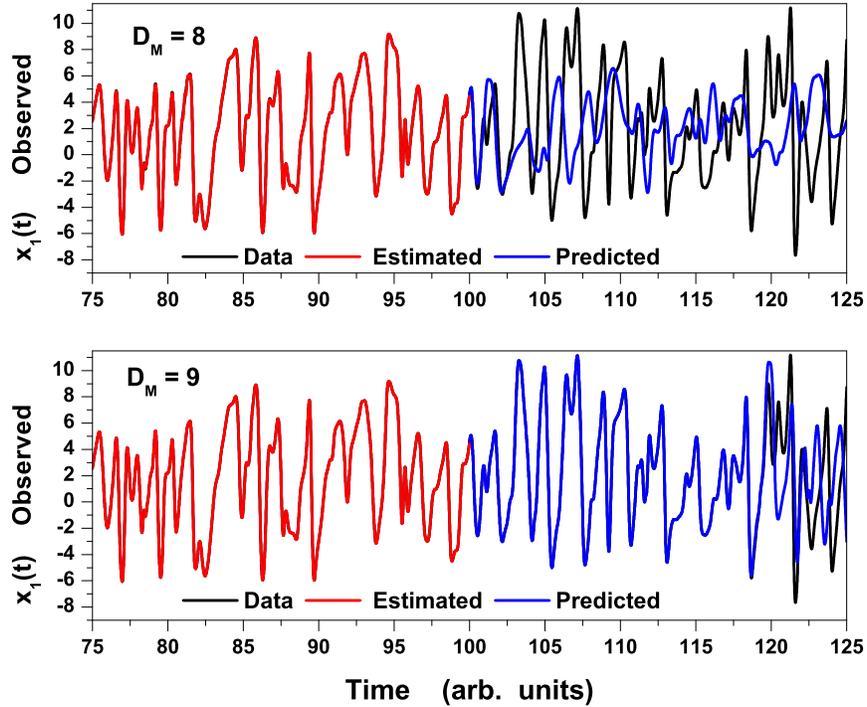


FIG. 6. (Color online) Estimates and predictions for the observed $x_1(t)$ for the Lorenz 1996 model with no additive noise ($\text{SNR} = \infty$). Traces are shown for $D_M = 8$ (**Top**) and $D_M = 9$ (**Bottom**). Note that while $D_M = 8$ generates excellent estimates, the predictions are poor, 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.

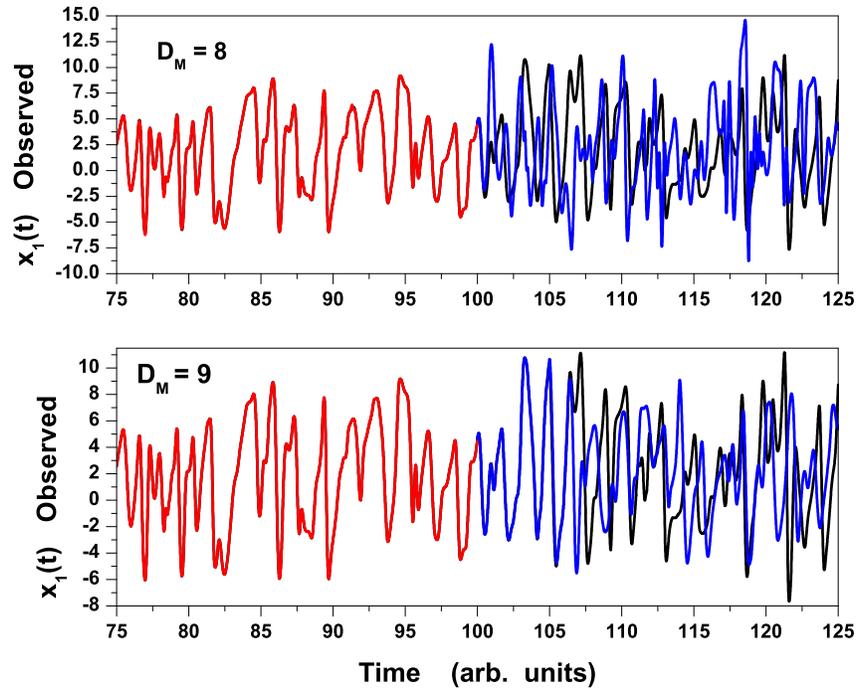


FIG. 7. (Color online) Estimates and predictions for the observed $x_1(t)$ for the Lorenz 1996 model with additive noise (SNR = 100 dB). Traces are shown for $D_M = 8$ (**Top**) and $D_M = 9$ (**Bottom**). As expected, selecting $D_M = 9$ produces good predictions. Although not as good as the case with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

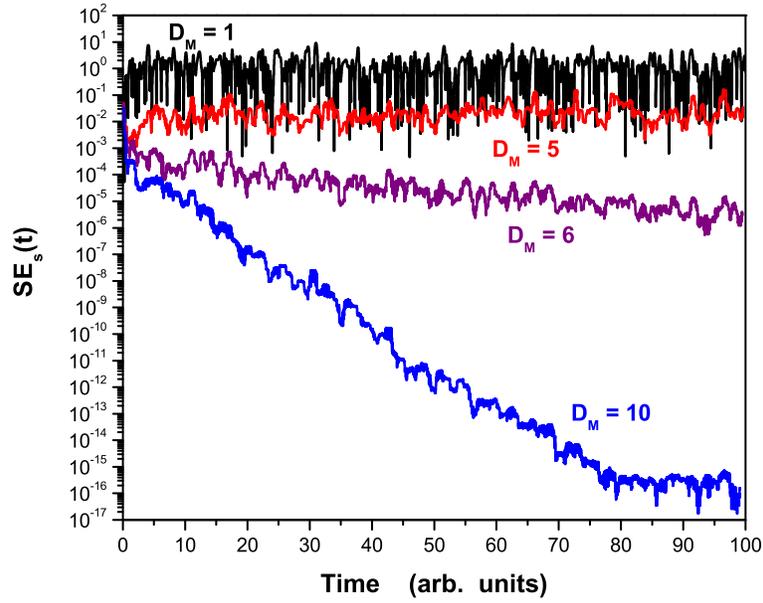


FIG. 8. (Color online) Experimental synchronization error $SE_s(t)$ for the Lorenz 1996 model with $D = 10$ and different forcing p_a in each component, for $D_M = \{1, 5, 6, 10\}$. This shows that, in addition to the state variables, ten parameters may be estimated when $D_M = 10$.

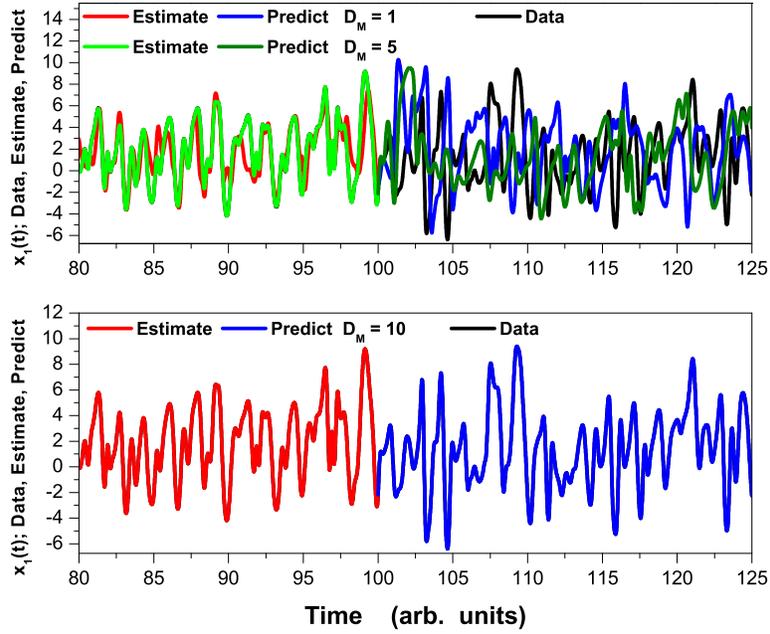


FIG. 9. (Color online) Estimate ($t < 100$) and prediction ($t > 100$) for $x_1(t)$ of the Lorenz 1996 model with $D = 10$ and different forcing p_a in each component during the synchronization shown in Fig. (8). For $D_M = 1$ and $D_M = 5$ (**Top**) the estimation and the prediction is not good nor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

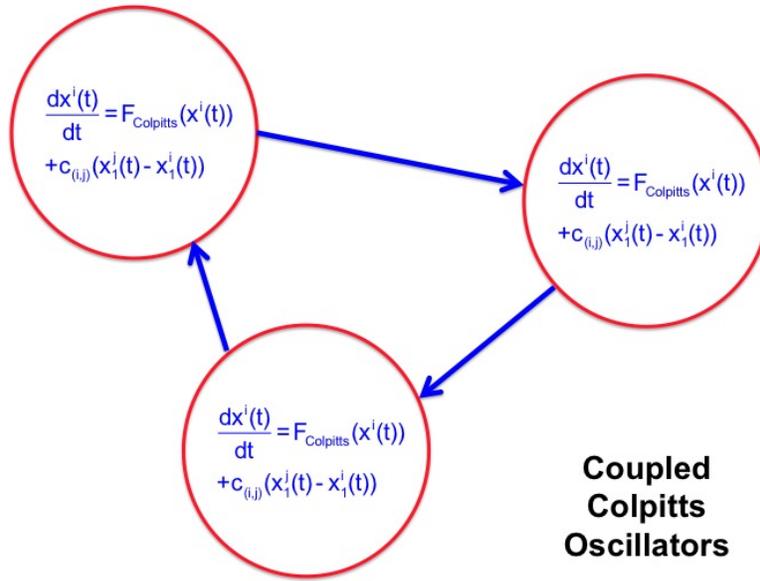


FIG. 10. Diagram of a unidirectionally coupled network of three Colpitts oscillators.

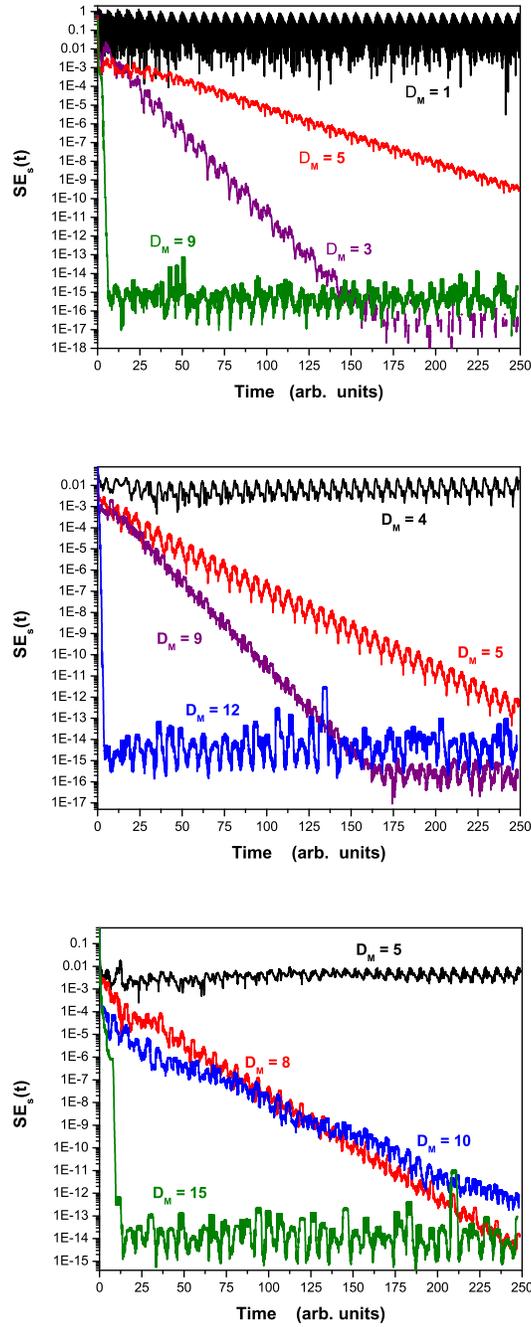


FIG. 11. (Color online) **Top** $SE_s(t)$ for state estimates of a network of three Colpitts oscillators. All model and coupling parameters are fixed to their true values and the network topology is known. **Middle** $SE_s(t)$ for state and ring coupling parameter estimates. Model parameters are not estimated and the network topology is known. **Bottom** $SE_s(t)$ for state and ring coupling parameter estimates. The model has been expanded to include backwards couplings, so the network topology is estimated as well.

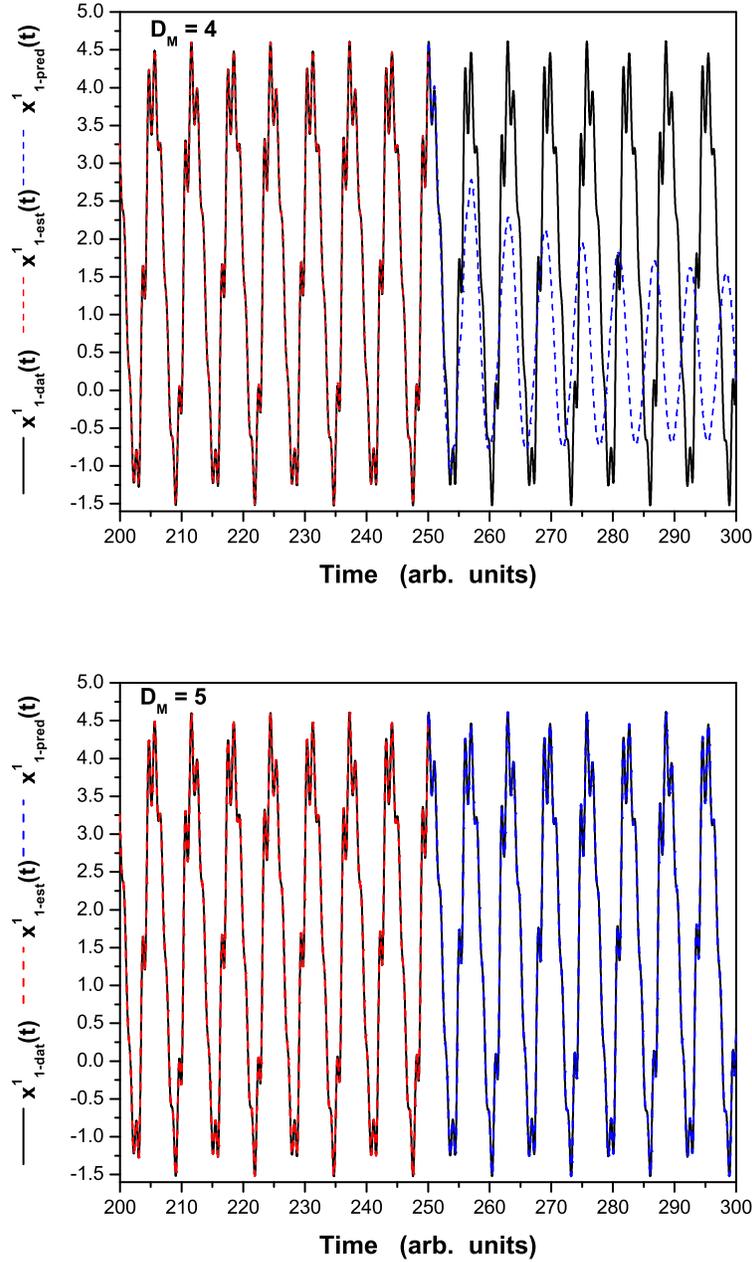


FIG. 12. (Color online) Ring of 3 Colpitts Oscillators **Top** Known (black), estimated (red) and predicted (blue) trajectories for the observed $x_1^{(1)}(t)$ state component with $D_M = 4$. Although the estimate is quite good, poor predictions confirm that $D_M = 4$ time delays do not provide enough 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, indication that the estimation procedure was successful.

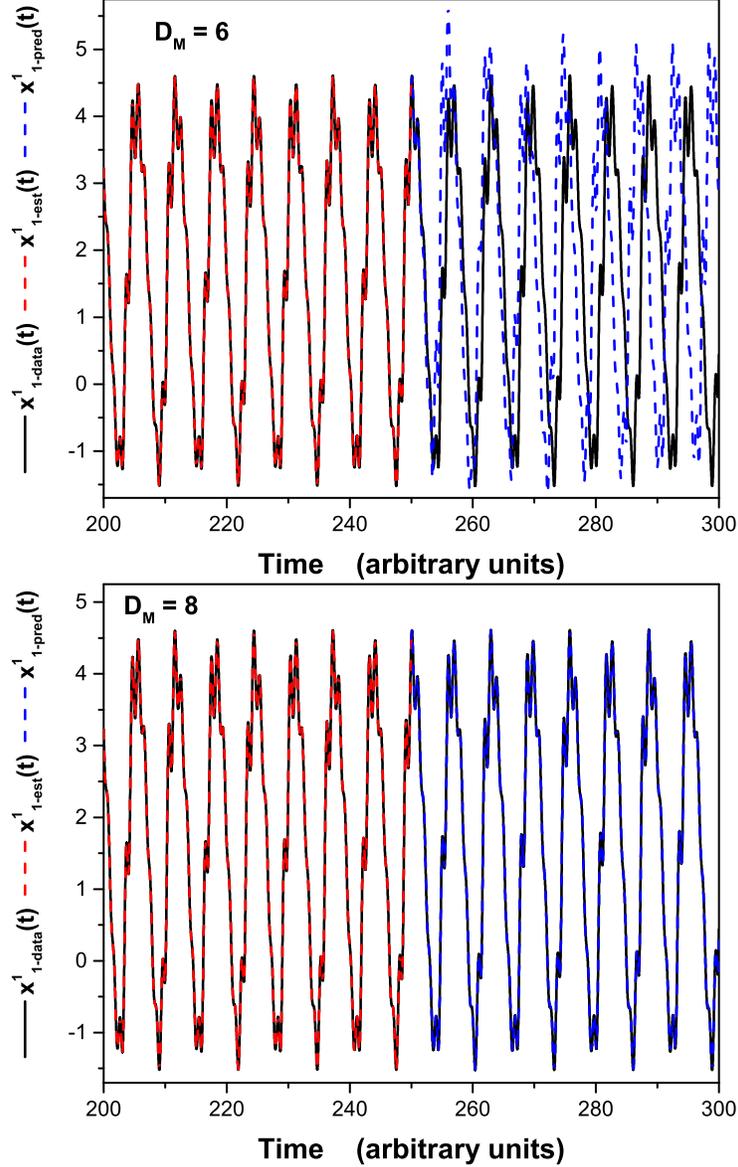


FIG. 13. (Color online) **Top** Estimated and predicted trajectory for the observed $x_1^{(1)}(t)$ state component with $D_M = 6$ using the expanded network model that includes backwards coupling (unknown network topology). Although the estimate is quite good, poor predictions confirm that $D_M = 6$ 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, indication that the estimation procedure was successful.