Data assimilation with regularized nonlinear instabilities

Henry D. I. Abarbanel[a], Mark Kostuk[b] and William Whartenby[b*]
[a]*Department of Physics and Marine Physical Laboratory (Scripps Institution of Oceanography), La Jolla, California, USA*
[b]*Department of Physics, University of California San Diego, La Jolla, California, USA*
*Correspondence to: W. Whartenby, Institute for Nonlinear Science, University of California San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0402, USA. E-mail: wwhartenby@physics.ucsd.edu

In variational formulations of data assimilation, the estimation of parameters or initial state values by a search for a minimum of a cost function can be hindered by the numerous local minima in the dependence of the cost function on those quantities. We argue that this is a result of instability on the synchronization manifold where the observations are required to match the model outputs in the situation where the data and the model are chaotic. The solution to this impediment to estimation is given as controls moving the positive conditional Lyapunov exponents on the synchronization manifold to negative values and adding to the cost function a penalty that drives those controls to zero as a result of the optimization process implementing the assimilation. This is seen as the solution to the proper size of 'nudging' terms: they are zero once the estimation has been completed, leaving only the physics of the problem to govern forecasts after the assimilation window.

We show how this procedure, called Dynamical State and Parameter Estimation (DSPE), works in the case of the Lorenz96 model with nine dynamical variables. Using DSPE, we are able to accurately estimate the fixed parameter of this model and all of the state variables, observed and unobserved, over an assimilation time interval $[0, T]$. Using the state variables at T and the estimated fixed parameter, we are able to accurately forecast the state of the model for $t > T$ to those times where the chaotic behaviour of the system interferes with forecast accuracy.

Key Words: parameter estimation; DSPE; unstable synchronization manifold; conditional Lyapunov exponent

Received 12 May 2009; Revised 27 January 2010; Accepted 8 February 2010; Published online in Wiley InterScience 16 April 2010

Citation: Abarbanel HDI, Kostuk M, Whartenby W. 2010. Data assimilation with regularized nonlinear instabilities. *Q. J. R. Meteorol. Soc.* **136**: 769–783. DOI:10.1002/qj.600

1. Introduction

One of the difficulties in assimilating data for nonlinear problems, such as those arising in forecasting atmospheric and oceanic behaviour, is associated with the unstable or chaotic dynamics of the models and the underlying geophysical systems. As clearly explained by Evensen (2007, Chapter 6), this leads to irregular surfaces in the parameter or initial condition dependence of a penalty function or cost function with which one wishes to estimate these quantities. A detailed discussion of how this irregular surface appears in the Lorenz model is given (Evensen, 2007), and this issue has been known in the nonlinear dynamics community for some years (K. Brueckner, 1983, personal communication; Roulston, 1999; Voss *et al.*, 2004).

The notion of a synchronization manifold arises when one considers a unidirectional flow of information from a transmitter to a receiver. Here the 'transmitter' is the source of measurements or observations and the 'receiver' is the model system. We identify the state of the 'transmitter' in this article as $\mathbf{x}(t)$ and the 'receiver' as $\mathbf{y}(t)$. If there are L measured quantities, known functions $z_l(t) = h_l\{\mathbf{x}(t)\}$; $l = 1, 2, \ldots L$, of the state variables of the observed system, then the criterion of model quality requiring $z_l(t) \approx y_l(t)$ is a measure of the synchronization $\mathbf{x}(t) \approx \mathbf{y}(t)$ of the transmitter and the receiver. Generally there are more state variables than the L measured items, so the subspace (or manifold) on which synchronization occurs is a subspace of the overall transmitter plus receiver space. This subspace, here L-dimensional, is called the synchronization manifold. It is stability in this subspace that is of interest. The stability indices for chaotic systems are called Lyapunov exponents. When restricted to this submanifold and conditioned on the data signals $z_l(t)$, these stability indices are called conditional Lyapunov exponents of the model system (Abarbanel, 1996; Kantz and Schreiber, 2004).

By considering the data assimilation problem as a question of synchronizing (Abarbanel, 1996; Yang *et al.*, 2006) an observed data stream with a model of the geophysical dynamics, we argue that the irregularities in the state and parameter estimation problem are associated with instabilities on the synchronization manifold. These instabilities arise when there are positive conditional Lyapunov exponents (CLEs) (Pecora and Carroll, 1990; Abarbanel, 1996; Kantz and Schreiber, 2004; Abarbanel *et al.*, 2008) expressed by motions on this synchronization manifold. Varying parameters and initial state values as one searches for an optimum set of values perturbs the motion on the synchronization manifold and causes irregular excursions from it, thus impeding the systematic search for accurate values of these quantities. We show how one can regularize the search for fixed parameter and state variables in data assimilation problems by adding terms to the equations of motion that reduce the CLEs to negative values, and adding terms to the cost or penalty function that set the regularizing terms to zero at the termination of the estimation effort.

We present a general argument for the need for the method we develop, and we give a detailed example using the Lorenz96 model (Lorentz, 1996; Lorentz and Emanuel, 1998). We show how one can use time series data to regularize the estimation of unobserved state variables, including their initial values at the beginning of the assimilation period, to provide accurate values of these state variables at the end of an assimilation period. Along with the observed state variables and the estimation of any fixed parameters, this estimate of the state at the end of the assimilation period provides the information required to forecast the behaviour of the model. We show explicitly how this works in the Lorenz96 model, and demonstrate that the accuracy allows prediction or forecasting within the time horizon dictated by the largest positive Lyapunov exponent of the model (Abarbanel, 1996; Kantz and Schreiber, 2004).

To the extent that irregularities in parameter and state variable surfaces impede forecasting, this regularization procedure eliminates that barrier and allow an environment where the focus can be on improvements in the model physics.

Our procedure requires small changes to the familiar four-dimensional variational principles, 4D-Var, discussed widely in the literature, and when implemented with the 'direct method' of nonlinear numerical optimization (Barclay *et al.*, 1998; Gill *et al.*, 1998, 2002a,b, 2005), does not require linearization at any stage of the formulation.

The focus of this paper is the presentation of the Dynamical State and Parameter Estimation (DSPE) method; summarized here are some highlights of this method that will be further discussed in what follows:

- The use of synchronization allows for the simultaneous determination of parameters and unobserved state variables.
- The use of control terms, and their values after the numerical optimization is completed, allows for the relative comparison of multiple models using a dataset. If the control lead to synchronization, the model is not consistent with the data.
- Enforcing negative CLEs during the optimization process leads to a reduction of observational noise intensity in the resultant state variable estimates.
- The method offers a means for determining how many measurements are needed to permit its success.
- The procedure may be scaled to problems of any size (dimensional and temporal). The practical application to numerical weather prediction problems is likely to be dependent on the capabilities of the nonlinear optimizer.
- The nonlinear problem does not require linearization.

1.1. A general view of variational principles for state and parameter estimation

The use of information in time series observations to inform the estimation of unobserved state variables and fixed parameters in the model has three components; it may be useful to keep these clearly in mind as they often are mixed up together in various approaches.

- The metric to use for comparison of the observations and the model output must have sufficient smoothness in the parameters and values of the state variables at the beginning of the observation period to allow a search procedure to avoid the impediment of local minima. In our approach this is achieved through a discussion of the stability of the subspace where the data and the model output are compared. If this surface is unstable, small perturbations in the parameters or state variables during the search will lead to large excursions away from the desired equality of the model output and the data.
- One must seek an efficient and reliable numerical method to perform the search through parameter and initial state variable space. We have selected the method called SNOPT (Barclay *et al.*, 1998; Gill *et al.*, 1998, 2002a,b, 2005) and have used it successfully as shown later in this paper. However, there are alternative, well-tested methods available in the public domain and commercially. We have not attempted to select among these methods in the work reported here.
- The data usually provide fewer observations than the combined number of state variables and parameters, so one should have some notion of how many observations are required to formulate a smooth comparison metric and in what manner they should be assimilated into the dynamical model.

2. Formulation of the nonlinear data assimilation problem

2.1. One unstable direction

In this section we will formulate the data assimilation problem presented in the literature (Evensen, 2007; Blum *et al.*, 2008) in nonlinear dynamical language. As a result, this section has some repetition of previous presentations of this problem.

The task at hand is to take observations of a dynamical system with state variables $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots]$ and use these to estimate the fixed parameters $\mathbf{p} = [p_1, p_2, \ldots, p_K]$ as well as the D initial conditions $\mathbf{y}(0) = [y_1(t_0), y_2(t_0), \ldots, y_D(t_0)]$ of a model of the system. We will write the model either as ordinary differential equations or as a set of discrete time equations associated with the $N+1$ observation times $t_m : \{t_0, t_1, \ldots, t_N = T\}$ for the data.

We imagine that a subset of the state variables $x_l(t_m) = x_l(m)$; $l = 1, 2, \ldots, L$, are observed over a time interval $[t_0, t_N = T]$. We associate these observations with state variables of the model with $y_l(m)$ being the equivalent of the observed $x_l(m)$. The remaining $D - L$ model variables we call $\mathbf{y}_R(m) = [y_{L+1}(m), y_{L+2}(m), \ldots, y_D(m)]$.

We could, with some additional notation, formulate the problem in terms of observed variables $z_l(m) = h_l\{\mathbf{x}(m)\}$, where we have L nonlinear observation functions $h_l(\mathbf{x})$. This would mean we need to compare the model output $\mathbf{y}(m)$ using this nonlinear measurement function, and we would look for equality of $h_l\{\mathbf{y}(m)\}$ with the observations $z_l(m)$. As this adds little to the procedure we introduce (except some notation), we proceed with $h_l\{\mathbf{x}(m)\} = x_l(m)$. The measurement is thus a projection of the full state $\mathbf{x}(m)$ to the $l = 1, 2, \ldots, L$ axes.

One could make a nonlinear change of variables with the $h_l(\mathbf{y})$ as L of the new coordinates, and then the identification of the model coordinates with the observed signals would be natural in that coordinate system.

The dynamics of the model system are given as an iterated map moving the dynamical description from some time t_m to time t_{m+1}:

$$\left.\begin{aligned} y_l(t_{m+1}) &\equiv y_l(m+1) \\ &= f_l\{\mathbf{y}(t_m), \mathbf{p}\} = f_l\{\mathbf{y}(m), \mathbf{p}\}, \\ &\qquad l = 1, 2, \ldots, L; \\ \mathbf{y}_R(t_{m+1}) &\equiv \mathbf{y}_R(m+1) \\ &= \mathbf{f}_R\{\mathbf{y}(t_m), \mathbf{p}\} = \mathbf{f}_R\{\mathbf{y}(m), \mathbf{p}\}, \end{aligned}\right\} \tag{1}$$

where $\mathbf{y}(m) = [y_1(m), y_2(m), \ldots, y_L(m), \mathbf{y}_R(m)]$. A formulation of the dynamics as differential equations is equivalent to this, and the discrete time equations may just be the differential equation solver one has selected.

To estimate the fixed parameters $\mathbf{p}$ and the initial values of $\mathbf{y}(0)$, we introduce the familiar minimal variance cost function

$$C\{\mathbf{y}(0), \mathbf{p}\} = \frac{1}{2(N+1)} \times \sum_{m=0}^{N} \left[\sum_{l=1}^{L} \left\{x_l(m) - y_l(m)\right\} \frac{1}{\sigma_l} \left\{x_l(m) - y_l(m)\right\} \right], \tag{2}$$

where the error $x_l(m) - y_l(m)$ is taken as having variance σ_l. The minimization of this cost function is assumed to lead to good estimates of the $\mathbf{p}$ and the $\mathbf{y}_R(0)$. (As it does not change the focus of our discussion, we set $\sigma_l = 1$.)

As noted in the introduction, the surfaces where one searches for values of the $\mathbf{p}$ and the $\mathbf{y}_R(0)$ are irregular when the dynamics of the model system are chaotic. To see how this arises, consider the estimation of $y_a(0)$. For this we must seek a zero of

$$\frac{\partial C\{\mathbf{y}(0), \mathbf{p}\}}{\partial y_a(0)} = \frac{1}{(N+1)} \sum_{m=0}^{N} \sum_{l=1}^{L} \frac{\partial y_l(m)}{\partial y_a(0)} \{y_l(m) - x_l(m)\}. \tag{3}$$

Now we note that, along with $\partial \mathbf{y}_R(m)/\partial y_a(0)$, the quantity $\partial y_l(m)/\partial y_a(0)$ satisfies

$$\frac{\partial y_j(m+1)}{\partial y_a(0)} = \sum_{b=1}^{D} \frac{\partial f_j\{\mathbf{y}(m)\}}{\partial y_b(m)} \frac{\partial y_b(m)}{\partial y_a(0)}, \quad j = 1, 2, \ldots, D, \tag{4}$$

where the Jacobian matrix

$$Df_{jb}(\mathbf{y}) = \frac{\partial f_j(\mathbf{y})}{\partial y_b} \tag{5}$$

of the dynamics now appears. Equation (4) is an exact result of the dynamics. The usual discussion of the role of the Jacobian matrix is in terms of a linearization of the equations of motion about a given orbit (Abarbanel *et al.*, 2008); this result makes the role of the perturbation in $\mathbf{y}(0)$ explicit. The solution of the equation for $\partial y_j(N+1)/\partial y_a(0)$ is the product of N Jacobian matrices evaluated along the orbit $\mathbf{y}(t_m)$:

$$\mathbf{Df}^N(\mathbf{y}) = \mathbf{Df}\{\mathbf{y}(N)\} \cdot \mathbf{Df}\{\mathbf{y}(N-1)\} \cdots \mathbf{Df}\{\mathbf{y}(1)\}, \tag{6}$$

using $\partial y_j(0)/\partial y_a(0) = \delta_{ja}$. The Lyapunov exponents of the model system are evaluated using the Oseledec multiplicative ergodic theorem (Oseledec, 1968; Abarbanel *et al.*, 2008) which tells us that the eigenvalues of the matrix (where T represents transpose)

$$\mathbf{OSL}\left\{\mathbf{y}(0), N\right\} = \left[\mathbf{Df}^N(\mathbf{y})^{\mathrm{T}} \cdot \mathbf{Df}^N(\mathbf{y})\right]^{(1/2N)}, \tag{7}$$

as $N \to \infty$,

- exist,
- are independent of $\mathbf{y}(0)$ within a basin of attraction of the dynamics, and
- are independent of the coordinate system under a smooth change of coordinates.

These eigenvalues are the Lyapunov numbers $\mathbf{e}^{\lambda_i}, i = 1, 2, \ldots, D$ which yield the Lyapunov exponents λ_i. Chaos results when any exponent is positive (or equivalently any eigenvalue of the Oseledec matrix lies outside the unit circle). If the dynamics of the model arise from autonomous differential equations, one of the eigenvalues is unity (or $\lambda_i = 0$ for some exponent).

It is the eigenvalues outside the unit circle that cause the irregularities in the $\{\mathbf{y}(0), \mathbf{p}\}$ surfaces of the cost function. To address this, let us start by assuming just one eigenvalue is outside the unit circle. Ordering the exponents, we will refer to the largest Lyapunov exponent as λ_1, so $\lambda_1 > 0$. Further, for the moment we consider the data to consist of one set of observations $x_1(m)$. We introduce regularization of the unwanted behaviour on the synchronization manifold $\mathbf{x}(m) = \mathbf{y}(m)$ with the addition to the dynamical equations of a term

$$\left.\begin{aligned} y_1(m+1) =& f_1\{\mathbf{y}(m), \mathbf{p}\} \\ &+ u(m)\{x_1(m) - y_1(m)\}, \\ \mathbf{y}_R(m+1) =& \mathbf{f}_R\{\mathbf{y}(m), \mathbf{p}\}, \end{aligned}\right\} \tag{8}$$

where the control or gain $u(m)$ is positive and is allowed to vary along a system orbit, or equivalently along the attractor, as the instabilities associated with $\mathbf{x}(m) = \mathbf{y}(m)$ are not generally uniform in the system state space. This change in the dynamics modifies the Jacobian, Eq. (5), to

$$Df_{ij}\{\mathbf{y}(m)\} \to Df_{ij}\{\mathbf{y}(m)\} - \delta_{11} u(m). \tag{9}$$

This is enough information to locally alter the largest Lyapunov exponent from a positive value to a negative value. If $u(m)$ were constant over the orbit, so $u(m) = k$, then we expect λ_1 to behave as $-k$ for large k.

Taking $u(m) = k$ for a moment, we see from the dynamical equations

$$\left.\begin{aligned} y_1(m+1) =& f_1\{\mathbf{y}(m), \mathbf{p}\} + k\{x_1(m) - y_1(m)\}, \\ \mathbf{y}_R(m+1) =& \mathbf{f}_R\{\mathbf{y}(m), \mathbf{p}\}, \end{aligned}\right\} \tag{10}$$

that for $k \to \infty$ the data $x_1(m)$ and the model $y_1(m)$ synchronize for any dynamics $\mathbf{f}(\mathbf{y}, \mathbf{p})$. Equivalently, for large k,

$$C\{\mathbf{y}(0), \mathbf{p}, k\} = \frac{1}{2(N+1)} \sum_{m=0}^{N} \{x_1(m) - y_1(m)\}^2 \tag{11}$$

behaves as k^{-2} and the derivatives with respect to $\mathbf{y}(0)$ or $\mathbf{p}$ become numerically so flat that these quantities cannot be accurately estimated. This means we need to balance the magnitude of the control term: k cannot be too small or we have irregular surfaces in $\{\mathbf{y}(0), \mathbf{p}\}$ space, nor can it be too large or the estimation becomes essentially infeasible. To achieve such a balance, we need a penalty for large k (or large $u(m)$), and we introduce this into the cost function as

$$C\{\mathbf{y}(0), \mathbf{p}, u\} = \frac{1}{2(N+1)} \sum_{m=0}^{N} \left[\{x_1(m) - y_1(m)\}^2 + u(m)^2\right]. \tag{12}$$

We could have used any function of $u(m)^2$ that vanishes as $u(m) \to 0$, but a simple quadratic works well.

We have now phrased this variational optimization problem as a least-squares problem comprising the minimization of the cost function (Eq. (12)) subject to the equality constraints (Eq. (10)).

We call this variational formulation with a regularized parameter and initial condition surface Dynamical State and Parameter Estimation (DSPE). We will show shortly that this achieves the regularization of the surfaces in $\{\mathbf{y}(0), \mathbf{p}\}$ space, allowing estimates of the fixed parameters $\mathbf{p}$ and state variable quantities, and has the important additional benefit that, when the estimation is completed, we will have achieved synchronization $x_1(m) \approx y_1(m)$, *and* we will have driven $u(m) \to 0$, so the dynamics of the problem are just those equations dictated by the physics with the regulation removed.

This method of regularizing the behaviour on the synchronization manifold has made its appearance in the literature of various disciplines over many decades. In control theory it is known as a Luenberger observer (Luenberger, 1964, 1966, 1971, 1979; Kailath, 1980; Ciccarella *et al.*, 1993, 1995; Moral and Grizzle, 1995; Nijmeijer and Mareels, 1997; Huijberts, 1999; Huijberts *et al.*, 2001; Roset and Nijmeijer, 2004), in atmospheric sciences it is known as 'nudging' (Auroux and Blum, 2008; Telford *et al.*, 2008), and in nonlinear dynamics it has been explored as a method for parameter estimation for some years (Parlitz, 1996; Parlitz *et al.*, 1996; Dedieu and Ogorzalek, 1997; Maybhate and Amritkar, 1999; Sakaguchi, 2002; Tokuda *et al.*, 2002; Konnur, 2003; Huang, 2004; Voss *et al.*, 2004; Yu *et al.*, 2006; Creveling *et al.*, 2007, 2008; Parlitz and Yu, 2008; Sorrentino and Ott, 2009). With one exception that we have found (So *et al.*, 1994), it has not been associated with stability of the synchronization manifold. The development of a variational formulation that regularized motion on the synchronization manifold and allowed state and parameter estimation where the controls ($u(m)$) were sent to zero at the end of the procedure was first explored in Abarbanel *et al.* (2008).

The goal of the 'observer' as discussed by Luenberger and others is to estimate the unobserved state variables of a model system when provided with a subset of the state variables of the observed system being represented by the model. In its original formulation in the 1960s and 1970s, it coupled information in the form of time series information into the dynamics of the model system. This coupling usually took the form of a constant times the difference of the measured input and the model output added to the dynamics of the model variables. In the earliest work, one was restricted to linear dynamics and the coupling was a constant whose value was unimportant as it was the complete system state, observed and unobserved variables, one wished to estimate. The observer was required to match the model output to the observed variables at each time after a transient, and this was usually stated as a match as time became large. Nonlinear versions of such investigations (Nijmeijer and Mareels, 1997; Huijberts, 1999; Huijberts *et al.*, 2001) proceeded in the same spirit. Nudging (Auroux and Blum, 2008; Telford *et al.*, 2008) is a variant of these considerations.

DSPE is clearly a descendant of these ideas. It adds two requirements: (1) it recognizes that the goal is stability (So *et al.*, 1994) on the synchronization manifold, and as that stability varies on the attractor explored by the orbits of the model system, the constant coupling becomes a function of time, and (2) it recognizes that the goal is the use of the state and parameter estimates, at the end of the time window when observations are made, to predict forward in time using the original dynamical equations. In this case, it is important that the unphysical 'nudging' or control of the model orbit to the observations be absent for the predictions, and DSPE adds a penalty to the variational principle used for estimation of parameters and state variables representing a cost for the use of the control term. Our examples show that this has the effect of reducing the controls to zero when the estimation procedure is completed, leaving one with the original physics introduced into the model.

While there are many examples to be found of the use of control theory ideas in the assimilation of data information into model equations, we comment on two of recent vintage for comparison. Chen and Lu (2002) add a term depending on the state variables, the control, and time-dependent variables which are given dynamical equations driving them to a fixed point where they become equal to the model parameters. The observations are coupled to the model equations much as in this paper, with the specifics of the coupling associated with a Lyapunov function for the system. No cost function is identified, and in the examples in Chen and Lu (2002) knowledge of all state variables is required, so only parameters are estimated, and that is done pointwise in time. Although that paper discussed several quite interesting low-dimensional examples, and work has been done on computing Laypunov functions for certain kinds of problems (Marióssn, 2002; Geisl, 2007), we were unable to find any method which could practically construct the Lyapunov function in general.

Another method that has been quite productive comes from the work of Bewley and colleagues (Bewley *et al.*, 2001; Cessna *et al.*, 2007). In the latter paper an innovative method for minimizing the standard least-squares cost function is discussed that uses the linearized adjoint equations to the model dynamics to search backward from the final time of observations. This approach, while numerically efficient, does not address the many local minima in the cost function, but the authors do recognize it. Their suggestion is to utilize small temporal windows of observations to avoid these, and this is accord with suggestions in Pires *et al.* (1996) and Evensen (2007).

2.2. Many unstable directions

We have dealt with a single unstable direction on the synchronization manifold. This will not be the usual situation for a complex dynamical system such as atmospheric dynamics. Suppose we have many unstable directions associated with the synchronization manifold. How can we determine the number of those unstable directions?

If the dynamics $\mathbf{y}(m+1) = \mathbf{f}\left\{\mathbf{y}(m+1), \mathbf{p}\right\}$ have U positive CLEs, we need to introduce U independent pieces of information to provide regulation in the U directions. Probably this can be accomplished in many ways, but we mention two:
(1) use U independent measurements or observations $x_l(m)$ to provide information about U directions in state space. If $U \le L$, then the formulation with L observations can be used.
(2) If $U > L$, one will need to use some version of the state space embedding theorem (Aeyels, 1981a,b; Takens, 1981; Sauer *et al.*, 1991; Garcia and Almeida, 2005a,b; Hirata *et al.*, 2006) to extract the needed information. We have formulated the solution to the latter problem in a recent paper (Abarbanel *et al.*, 2009).
In this paper we explore (1) in the context of the Lorenz96 model.

The method of state space embedding (Abarbanel,1996; Kantz, 2004) uses the information in the waveform of an observed time series to identify the number of dynamical degrees of freedom contained in the measurements. It looks at the measurements at a time t_n as well as measurements at earlier times $\{t_{n-1}, t_{n-2}, \ldots\}$. The number of such time delays determines the number of independent pieces of information contained in the waveform, and permits one to extract more information from the observations to present to the model system to assist in estimating state variables and parameters.

To determine the value of U, we have two approaches. One uses the Oseledec multiplicative ergodic theorem for finite, though large, N and utilizes a recursive QR method (Abarbanel, 1996) to determine the positive Lyapunov exponents. The other notes that the synchronization error (our cost function) will vanish as the equivalent of k (or $u(m)$) above becomes large, and it couples into the dynamics one, two, three, ... pieces of information at large k until that error goes to zero. We will use the second approach in this paper, and we will explore the first in a later publication (Kostuk *et al.*, 2010).

To be more specific, we go to a formulation of DSPE in continuous time, so we are considering D ordinary differential equations rather than D iterated maps:

$$\left.\begin{aligned}\frac{\mathrm{d}y_a(t)}{\mathrm{d}t} &= F_a\{\mathbf{y}(t), \mathbf{p}\} + \sum_{l=1}^{U} u_{al}(t)\{x_l(t) - y_l(t)\},\\ &\qquad a = 1, 2, \ldots, L,\\ \frac{\mathrm{d}\mathbf{y}_R(t)}{\mathrm{d}t} &= \mathbf{F}_R\{\mathbf{y}(t), \mathbf{p}\}.\end{aligned}\right\} \tag{13}$$

Taking the controls $u_{al}(t) \equiv u_{al}$ (or 'nudges') to be constants, we evaluate

$$C\{\mathbf{y}(0), \mathbf{p}, \mu\} = \frac{1}{2T}\int_0^T \left[\sum_{a=1}^{U} \{x_a(t) - y_a(t)\}^2\right] \mathrm{d}t \tag{14}$$

as a function of some $y_r(0)$ or some p_k for various fixed values of u_{al}. We do this for $U = 0, 1, 2, \ldots$, until the cost function surface in the $y_r(0)$ or p_k becomes smooth. This tells us the minimum number of controls we require.

Knowing U, we then introduce U independent pieces of information into the dynamics, as in Eq. (13), and use these equations as constraints on the minimization of

$$C\{\mathbf{y}(0), \mathbf{p}, \mu\} = \frac{1}{2T}\int_0^T \left\{\sum_{a=1}^{U}\left[\{x_a(t) - y_a(t)\}^2 + \sum_{b=1}^{U} u_{ab}(t)\, u_{ab}(t)\right]\right\} \mathrm{d}t. \tag{15}$$

As we use the 'direct method' for the numerical solution of this optimization problem (Barclay *et al.*, 1998; Gill *et al.*, 1998, 2002a,b, 2005), at the conclusion of the estimation procedure we will have estimated values of the following quantities:

1. the parameters $\mathbf{p}$,
2. the state variables $\mathbf{y}(t_m)$ at the observation times $t_m = \{t_0, t_1, t_2, \ldots, t_N\}$ where $t_0 = 0$ and $t_N = T$, and
3. the controls $u_{al}(t_m)$ at the observations times.

The latter are zero, within some numerical error, when the model is consistent (section 3).

The full set of estimated parameters $\mathbf{p}$ and the full state of the model at time $t_N = T$, namely $\mathbf{y}(t_N = T)$, now permits us to use the dynamics with controls removed $d\mathbf{y}(t)/dt = \mathbf{F}\{\mathbf{y}(t), \mathbf{p}\}$ to predict or forecast the model behaviour for $t > T$. Because of the chaotic behaviour of the physical system, we will need to re-estimate the state variables and any 'fixed' parameters (which in fact have some variation) at some later time in order to continue an accurate forecast. That time horizon is approximately a time ${\lambda_1}^{-1}$ after T.

All of these statements hold equally well for the dynamics expressed as discrete time maps.

3. The role of the controls

In our DSPE method, we introduce the control ('nudging') terms to regularize the motion on the synchronization manifold $x_l(t) = y_l(t)$; $l = 1, 2, \ldots L$; however, we wish to have them essentially zero when the estimation of any parameters and the model state variables is completed. These control variables have dimensions of inverse time, and it is more sensible to provide a dimensionless comparison of the importance of the control terms

$$\sum_{l=1}^{L} u_{al}(t)\left\{x_l(t) - y_l(t)\right\}$$

relative to the model vector fields $F_a\{\mathbf{y}(t), \mathbf{p}\}$ for $a = 1, 2, \ldots, L$ where they are introduced. We suggest evaluating the L dimensionless ratios

$$R_a(t) = \frac{F_a\{\mathbf{y}(t), \mathbf{p}\}^2}{F_a\{\mathbf{y}(t), \mathbf{p}\}^2 + \left[\sum_{l=1}^{L} u_{al}(t)\{x_l(t) - y_l(t)\}\right]^2}. \quad (16)$$

These ratios lie between zero and unity. When they are close to 1, this indicates that the synchronization has succeeded because the model is consistent with the data and the control terms

$$\sum_{l=1}^{L} u_{al}(t)\{x_l(t) - y_l(t)\}$$

are negligible with respect to the dynamics $F_a\{\mathbf{y}(t), \mathbf{p}\}$. When the $R_a(t)$ are near zero, this indicates that, although the synchronization has succeeded, the model is *not* consistent with the data because large values of the controls have forced the model dynamical variables $y_l(t)$ to the values from the data $x_l(t)$.

In addition, we may use the $u_{al}(t)$ to test the consistency of the model. One can easily check that estimating the parameters and states $\mathbf{y}(t)$ in a model using incompatible data will result in the values of $R_a(t)$ over the estimation time $[0, T]$ deviating significantly from unity, with many values near zero.

4. Numerical optimization method: SNOPT

To solve the nonlinear numerical optimization problem set by DSPE, Eqs. (13) and (15), we have used the 'direct method' embodied in the software SNOPT (Barclay *et al.*, 1998; Gill *et al.*, 1998, 2002a,b, 2005). This method introduces a time 'grid' $t_0, t_1, t_2, \ldots, t_N$ covering the observation interval $[0, T]$ and uses as variables in the optimization algorithm the values of the state variables $y_a(t_m)$, the control variables $u_{al}(t_m)$, and the unknown fixed parameters $p_1, p_2, \ldots, p_K$. The numerical optimization is an iterative method with equality constraints given by the equations of motion. If one has weak constraints associated with an estimate of model errors (Evensen, 2007), then the optimization is unconstrained.

In this large space, an initial estimate $\{y_a^{(0)}(t_m), u_{al}^{(0)}(t_m), p_k^{(0)}\}$ is given along with tolerances on the amount the current local solution can be improved. SNOPT produces a sequence of estimates

$$\left\{y_a^{(i)}(t_m), u_{al}^{(i)}(t_m), p_k^{(i)}\right\} \rightarrow \left\{y_a^{(i+1)}(t_m), u_{al}^{(i+1)}(t_m), p_k^{(i+1)}\right\}$$

until the tolerance is met or the number of iterations reaches a preset maximum allowed number.

$$\left\{y_a^{(\text{final})}(t_m), u_{al}^{(\text{final})}(t_m), p_k^{(\text{final})}\right\}$$

are then reported.

5. Lorenz96 model; $K = 9$

To illustrate the use of DSPE, we report on a calculation using the Lorenz96 model (Lorenz, 1996) with $K = 9$ variables. In this section we treat the problem with no observational noise or model error, so we expect our model variables $y_a(t)$; $a = 1, 2, \ldots, K = 9$ to be equal to the data presented from an independent evaluation of the model giving state variables $\mathbf{x}(t)$. We solve the Lorenz96 equations

$$\frac{dx_a(t)}{dt} = x_{a-1}(t)\{x_{a+1}(t) - x_{a-2}(t)\} - x_a(t) + f; \quad a = 1, 2, \ldots, K, \quad (17)$$

with $x_{-1}(t) = x_{K-1}(t), x_0(t) = x_K(t)$, and $x_{K+1}(t) = x_1(t)$, with $K = 9$ and some choice of $x_a(0)$. We selected the single parameter, representing the forcing of the model, as $f = 5.0117$ so that the solutions of these equations were chaotic and not fixed points at $x_a = f$ or periodic orbits. These fixed points and limit cycles are unstable for this value of f.

We passed $U = 0, 1, 2, \ldots, 9$ time series $x_l(t)$; $l = 1, 2, \ldots, U$ to the model equations

$$\frac{dy_a(t)}{dt} = y_{a-1}(t)\{y_{a+1}(t) - y_{a-2}(t)\} - y_a(t) + f, \quad (18)$$
$$a = 1, 2, \ldots, K,$$

with $y_{-1}(t) = y_{K-1}(t)$, $y_0(t) = y_K(t)$, and $y_{K+1}(t) = y_1(t)$, with $K = 9$, and introducing

$$\frac{dy_a(t)}{dt} = y_{a-1}(t)\{y_{a+1}(t) - y_{a-2}(t)\} - y_a(t) + f + k_a\{x_a(t) - y_a(t)\}, \quad (19)$$

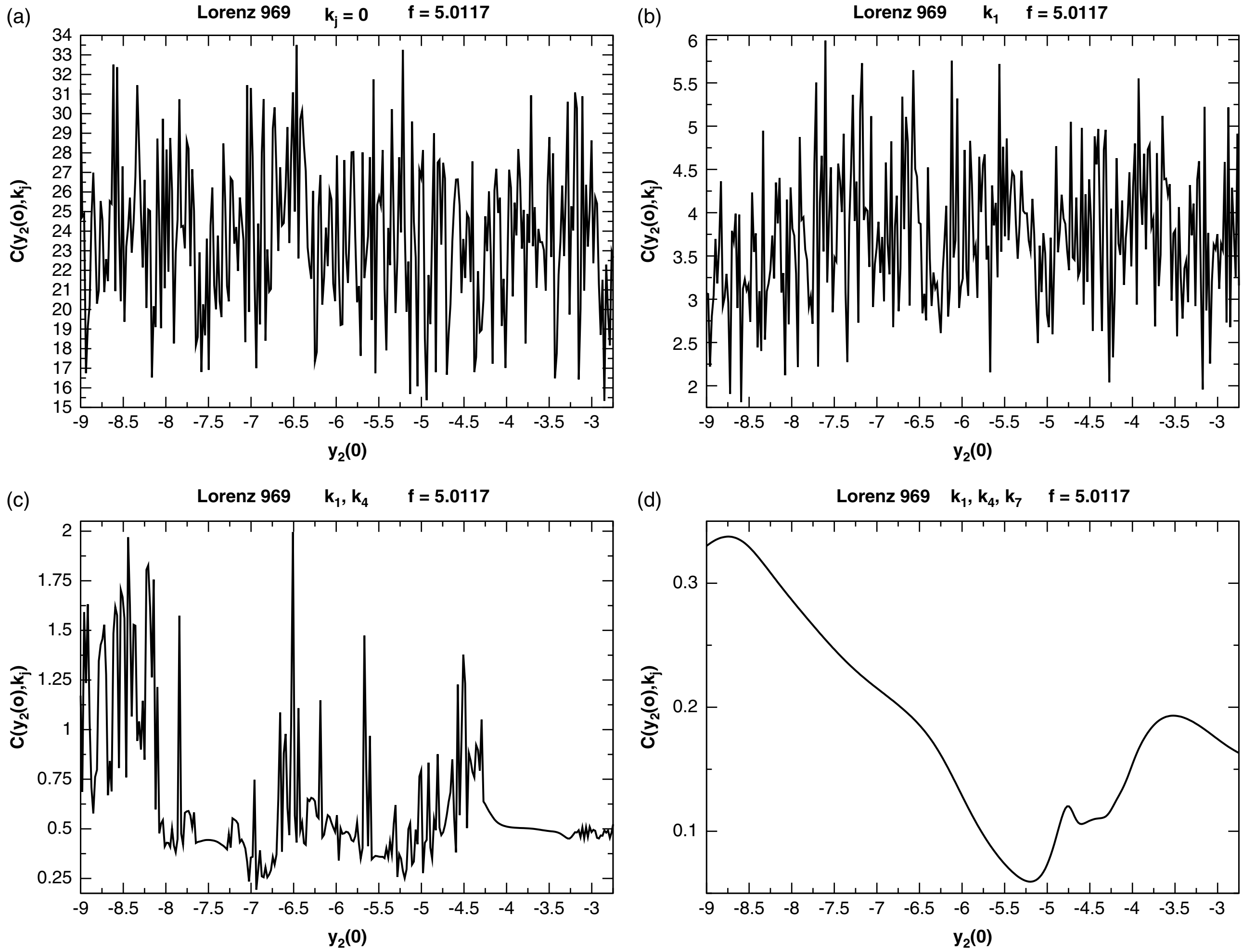

Figure 1. The cost function or synchronization error $C\{\mathbf{y}(0), \mathbf{p}, k_1, k_2, \ldots, k_9\}$ from Eq. (20) as a function of the initial value $y_2(0)$. (a) All $k = 0$, (b) $k_1 \neq 0$; $k_1 = 9.3$, (c) $k_1, k_4 \neq 0$; $k_1 = k_4 = 9.3$, (d) $k_1, k_4, k_7 \neq 0$; $k_1 = k_4 = k_7 = 9.3$.

we evaluated the synchronization error

$$C\{\mathbf{y}(0), \mathbf{p},\ k_1, k_2, \ldots, k_9\} = \frac{1}{2T}\int_0^T \left\{ \sum_{a=1}^{U} \{x_a(t) - y_a(t)\}^2 \right\} \mathrm{d}t \tag{20}$$

as a function of $y_2(0)$ for different choices of U and values of $k_a;\ a = 1, 2, .., 9$.

First we selected all $k_a = 0$, and show in Figure 1(a) the cost function (14) as it varies in $y_2(0)$ with no couplings. The ragged, irregular nature of the variation in $y_2(0)$ reflects the instability of the synchronization manifold. In particular the contributions of the data $x_a(t)$ and the model output $y_a(t)$ are essentially random with respect to each other. These irregularities mask the minimum at the correct value of $y_2(0) \approx -5.2$, and the numerous local minima will present any search routine with substantial impediments to finding a good estimate. This situation applies to all initial conditions $\mathbf{y}(0)$ as well as any parameters, here only f.

In Figure 1(b) we now couple the observed variable $x_1(t)$ into the equation for $y_1(t)$ with $k_1\{x_1(t) - y_1(t)\}$ selecting the (somewhat arbitrary but large enough for eventual synchronization) value $k_1 = 9.3$. The irregularity in the variation of $y_2(0)$ persists with this single coupling. Next, in Figure 1(c) we also couple the observation $x_4(t)$ into the equation for $y_4(t)$ as $k_4\{x_4(t) - y_4(t)\}$, selecting $k_4 = k_1 = 9.3$, and we see that the irregularity in the variation of $y_2(0)$, though reduced, persists. Finally, in Figure 1(d)) additionally adding in the information contained in the observation $x_7(t)$ coupled to the $y_7(t)$ equation as $k_7\{x_7(t) - y_7(t)\}$, selecting $k_1 = k_4 = k_7 = 9.3$, we see that the variation of the cost function in $y_2(0)$ has become smooth and now permits the identification of a good estimate for the value $y_2(0) \approx -5.2$. The cost function $C\{\mathbf{y}(0), \mathbf{p}, k_1, k_2, \ldots, k_9\}$ for these values of k_a has become nearly zero, indicative of synchronization.

This pattern held for any one, two or three observations $x_a(t)$ we coupled into the model equations, for any initial condition $y_r(0)$, and for the estimation of the single parameter f in the Lorenz96 model. This calculation strongly suggests that the Lorenz96 model with $K = 9$ at the selected parameter value $f = 5.0117$ has three positive CLEs. It also indicates that, to use DPSE to estimate the *unobserved* state variables of the model–$y_2(t), y_3(t), y_5(t), y_6(t), y_8(t)$, and $y_9(t)$ in this example, we need to pass three pieces of information to the model. Once we have estimated those unobserved state variables over the observation period $t_0 \leq t \leq T$, we can use that and the value of f to forecast the model behaviour for $t > T$. This we do next.

We now solve the Lorenz96 system as in Eq. (17) with $f = 5.0117$ for the 'observed' data $x_a(t)$. We sampled the model output at $\Delta t = 0.01$ corresponding to an equivalent time of 6/5 h (Lorenz, 1996; Lorenz and Emanuel, 1998). Using the guidance from the smoothness of the cost function in initial condition and parameter space, we passed to the model the three observations $x_1(t_m), x_3(t_m), x_7(t_m)$, Eq. (19).

We then asked our numerical optimization code, SNOPT, to minimize the cost function

$$C(\mathbf{y}, \mu) = \frac{1}{2T}\int_0^T \Big[\{x_1(t) - y_1(t)\}^2 + \{x_3(t) - y_3(t)\}^2 + \{x_7(t) - y_7(t)\}^2 + u_1(t)^2 + u_2(t)^2 + u_3(t)^2\Big]\mathrm{d}t, \tag{21}$$

subject to the equations of motion

$$\left.\begin{aligned} \frac{\mathrm{d}y_1}{\mathrm{d}t} &= (y_2 - y_8)\,y_9 - y_1 + f + u_1(x_1 - y_1), \\ \frac{\mathrm{d}y_3}{\mathrm{d}t} &= (y_4 - y_1)\,y_2 - y_3 + f + u_2(x_3 - y_3), \\ \frac{\mathrm{d}y_7}{\mathrm{d}t} &= (y_8 - y_5)\,y_6 - y_7 + f + u_3(x_7 - y_7), \\ \frac{\mathrm{d}y_j}{\mathrm{d}t} &= (y_{j+1} - y_{j-2})\,y_{j-1} - y_j + f; \\ &\quad j \neq 1,\ 3,\ 7. \end{aligned}\right\} \tag{22}$$

We used $N = 1045$ data points where observations were made. In the time units described by Lorenz and Emanuel (1998), this corresponds to about 7.5 weeks of data. We could have used a smaller dataset, but we wanted to make sure the irregularities associated with chaotic behaviour manifested themselves in the 'data' so that, when applying this method to realistic datasets, we could be certain that the presence of chaotic oscillations had been addressed.

In Figure 2(a) we show the known, $x_4(t), x_8(t)$, and the estimated, $y_4(t)$ and $y_8(t)$, at the 1045 locations where 'data' $x_a(t)$ were available. Similarly, in Figure 2(b), we show the known, $x_1(t), x_2(t)$, and the estimated, $y_1(t)$ and $y_2(t)$, at the 1045 locations where 'data' $x_a(t)$ were available. The RMS error in the state estimation was about 1.15% across the observations. From these estimated state variables, we selected the values at t_{1043} from which to predict forward in time or forecast the model behaviour. The error in the values of the $y_a(t_{1043})$ were about 0.3% compared to the known $x_a(t_{1043})$. We chose to forecast from a slightly earlier time than T, just to have a small overlap with the training data.

In Figure 3 we now show the predictions or forecasts of the model behaviour for $t > T = t_{1045}$ using our estimates of the state of the model at T and the value of f. Here we compare the data $x_2(t > T)$ and $x_8(t > T)$ with the predictions from the model for $y_2(t > T)$ and $y_8(t > T)$ and the initial states determined by our estimation procedure. Predictions for all other state variables of the model were equally accurate. The model solution remains close to the data for about 1000 time steps. This translates into a largest Lyapunov exponent $\lambda_1 \approx 1.0$ in units of 6/5 h.

To evaluate the role of the three controls $u_1(t_m), u_2(t_m)$, and $u_3(t_m)$, we calculated the three ratios $R_1(t_m), R_3(t_m)$, and $R_7(t_m)$, (Eq. (16)), over $[0, T]$ using our model equations, our estimated values for $\mathbf{y}(t_m)$, our known observations $x_1(t_m), x_3(t_m)$, and $x_7(t_m)$, and the estimated values of $u_1(t_m), u_2(t_m)$ and $u_3(t_m)$. We also estimated the one fixed parameter f in the Lorenz96 equations, and as that was within 1% of the known value, we do not discuss it further.

In Figure 4 we show the values for $R_1(t_m), R_3(t_m)$, and $R_7(t_m)$ over the observation period. It is clear that the synchronization of the data $x_a(t_m)$ and the model $y(t_m)$ is associated with the accuracy of the model, not the presence of large $u_a(t_m)$. Examination of the $u_a(t_m)$ themselves (not shown) verifies this with the caveat that they are dimensional while our ratios are dimensionless.

5.1. *Observational noise*

Our final numerical results concern the case where the observations $x_l(t_m)$ are contaminated by random noise. This means we present $x_l(t_m) + \eta_l(t_m)$ to our model where the $\eta_l(t_m)$ are zero-mean random errors in the measurements. While there are many choices for the statistics of the $\eta_l(t_m)$, we report the results of using noise that is independent at each time t_m and uncorrelated between measurements, $< \eta_l(t_m)\, \eta'_l(t_n) > \propto \delta_{ll'}\delta_{mn}$. Each noise term was taken to be uniformly distributed in the interval $[-A, A]$, making the RMS value of the noise $A/\sqrt{3}$. This allows us to define a signal-to-noise ratio (in dB) for each observed, noisy dataset as

$$(S/N)_a = 20\log_{10}\left(\frac{RMS(x_a)\sqrt{3}}{A}\right), \tag{23}$$

with $RMS(x_a)$ the RMS value of the observed dataset $x_a(t_m)$.

Using these definitions, we presented noisy observed data $x_l(t_m) + \eta_l(t_m)$ with $S/N = 20$ dB and $S/N = 30$ dB. Our 'clean' data above have essentially $S/N = \infty$ dB. In these calculations, we added noise to each of the three signals $x_1(t_m)$, $x_3(t_m)$, and $x_7(t_m)$ used previously in providing information to the model system. Noise was also added to the other, unobserved, data $x_2(t_m), x_4(t_m), x_5(t_m), x_6(t_m), x_8(t_m), x_9(t_m)$.

In Figure 5(a) we show the three quantities $x_5(t_m)$ without noise added, with noise $x_5(t_m) + \eta_5(t_m)$, and the estimated model output $y_5(t_m)$. In Figure5(b) we remove the noisy signal $x_5(t_m) + \eta(t_m)$ to provide an easier comparison of the model output $y_5(t_m)$ and the noise-free data signal $x_5(t_m)$. Note that y_5 is an unobserved signal. One of the observed signals, $x_7(t_m)$, along with its noisy version $x_7(t_m) + \eta_7(t_m)$ as presented to the model, and the model output $y_7(t_m)$ are shown in Figure 6(a), and in Figure 6(b) we have again removed the noisy signal. Clearly the accuracy of the estimated signals, observed or unobserved, has degraded with the addition of noise (here we have a $S/N = 30$ dB), but the accuracy at the end of the assimilation period ($T \approx 1000\Delta t$) is still high.

In Figure 7 we present the same information as displayed in Figure 5, now using a signal-to-noise ratio of 20 dB. Now we see significant degradation of the estimated signal $y_5(t_m)$ relative to either the noisy or clean input until the later part of the assimilation period. In Figure 7(b) we see that the estimate is rather accurately tracking the clean input signal $x_5(t_m)$.

To look at the quality of the estimated signals in the presence of observational noise, we evaluated the consistency ratios $R_1(t_m), R_3(t_m)$, and $R_7(t_m)$ for each case, $S/N = 30$ dB and $S/N = 20$ dB. These are shown in Figure 8 where we

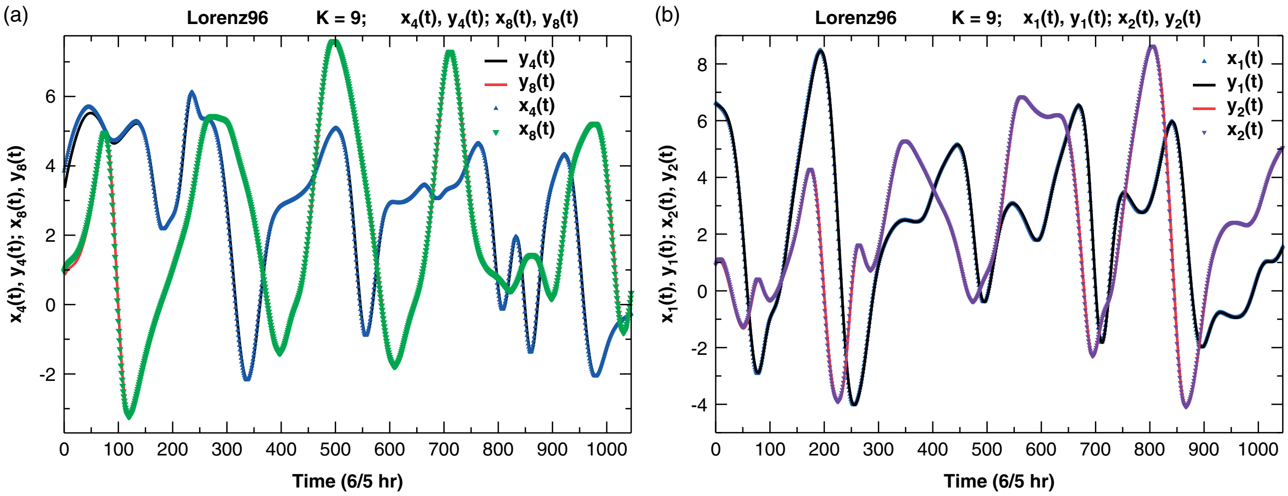

Figure 2. Information from Eq. (22) using DSPE. (a) shows $x_4(t), y_4(t), x_8(t), y_8(t)$. (b) shows the known $x_1(t), x_2(t)$ and the estimated $y_1(t), y_2(t)$.

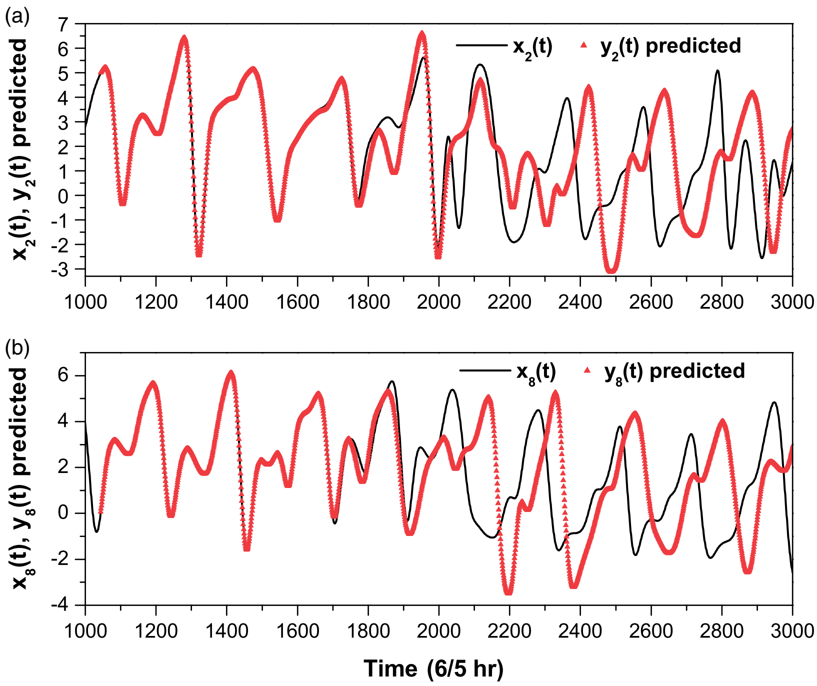

Figure 3. Using the estimated values of the model variables at the end of the estimation or assimilation period ($t = T$), the model equations were used to evaluate $y_a(t > T)$. (a) shows $y_2(t)$ compared to the known $x_2(t)$, while (b) shows the known $x_8(t)$ and the estimated model variable $y_8(t)$. The deviation of the model predictions comes from the chaotic behaviour of this dynamical system.

see that, in the regions of the assimilation time where the estimates $y_a(t_m)$ are quite different from either the noisy or clean data, the controls are working very hard to move the estimate toward the clean signal $x_a(t_m)$. Where the estimates are accurate, we see that the ratios are near unity, indicating consistency of the model with the data. This is another benefit of using the direct method of solving the nonlinear numerical optimization problem posed by DSPE.

Our last two figures (Figures 9 and 10) take the estimated values of two of the model outputs $y_2(T)$ and $y_8(T)$ at the end of the assimilation period and use the model differential equations with $u_a(t_m) = 0$ to predict their behaviour for $t > T$. These are then compared with the data for $x_2(t > T)$ and $x_8(t > T)$. The accuracy of these forecasts breaks down much sooner than in the clean case shown earlier (Figure 3). This is attributable to the loss of accuracy in the estimates $y_2(T)$ and $y_8(T)$ in the presence of noise. The forecasts for the other model outputs are consistent with the two we have displayed here.

It is interesting to note that, in each of these cases, the model $y_a(t_m)$ estimation of the system state is closer to the noise-free $x_a(t_m)$ than to the noisy data presented to the model. This is not particular to this Lorenz96 example, but occurs in other examples as well (Abarbanel *et al.*, 2010). The reason for this reduction in noise is that the introduction of the control or regularization terms $u_{al}(t_m)$ locally reduces positive conditional Lyapunov exponents to negative values. Negative CLEs cause perturbations from the synchronization manifold to return exponentially rapidly to the synchronization manifold, and these perturbations by the noise in the observations are drawn back to the unperturbed orbit. The efficacy of this depends on the local Lyapunov exponents, and we do not have a method for evaluating this outside of the numerical results reported

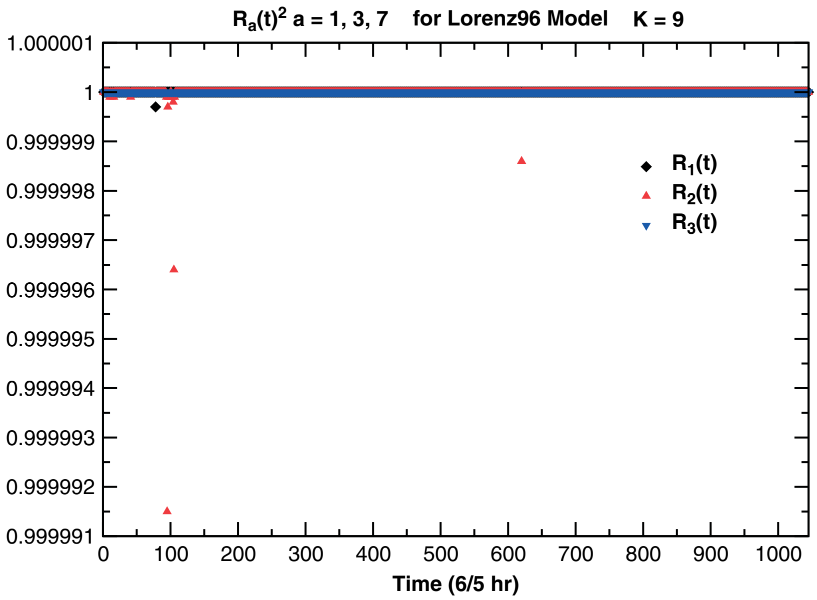

Figure 4. Equation (16) was used to evaluate the role of the control ('nudging') terms. If the ratio $R_a(t)$ is order unity, as here, the model is consistent with the data, and the role of the control terms is not important.

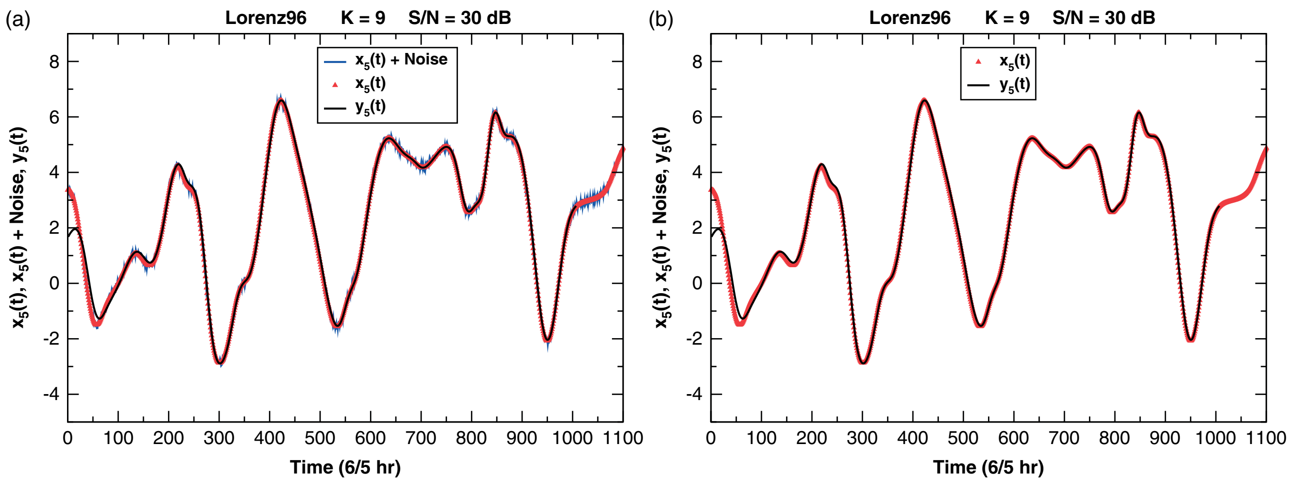

Figure 5. (a) The clean $x_5(t_m)$, noisy $x_5(t_m) + \eta_5(t_m)$, and estimated model output $y_5(t_m)$ when noise with a signal-to-noise ratio of 30 dB has been added to the signal presented to the model. (b) is as (a) but with the noisy signal removed, so that comparison between the estimated model output and the clean data signal can more easily be made. Note that the output $y_5(t_m)$ has smoothed out the noisy signal presented to the model. The improvement here is a factor of ≈ 2.15.

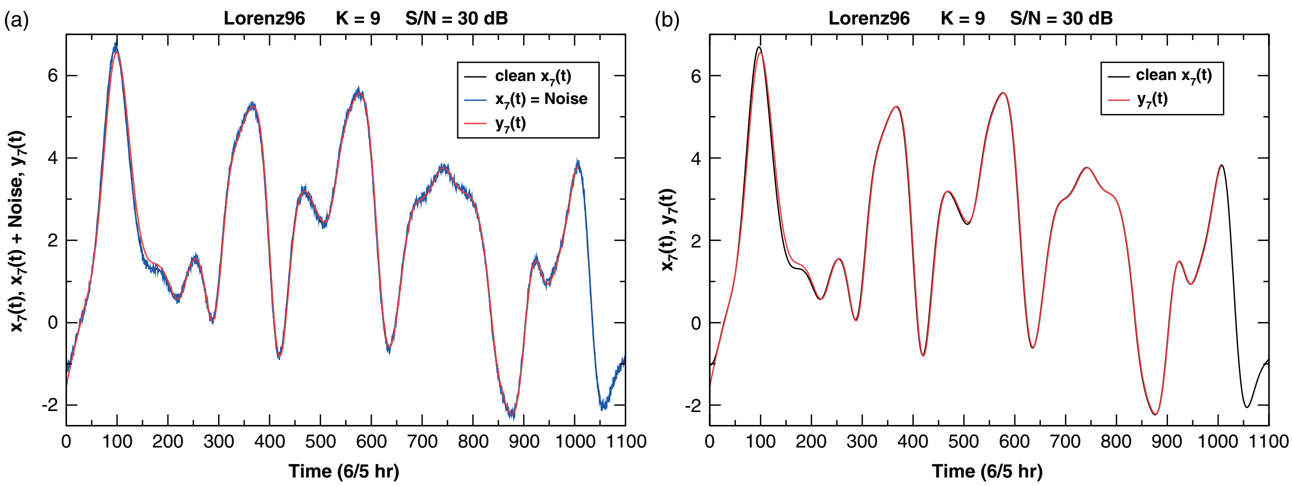

Figure 6. (a) The clean $x_7(t_m)$, noisy $x_7(t_m) + \eta_7(t_m)$, and estimated model output $y_7(t_m)$ when noise with a signal-to-noise ratio of 30 dB has been added to the signal presented to the model. (b) is as (a) but with the noisy signal removed so that comparison between the estimated model output and the clean data signal can more easily be made. Note that the output $y_7(t_m)$ has smoothed out the noisy signal presented to the model.

here. In the case of our observations contaminated by noise with a S/N of 30 dB, we have evaluated the noise reduction relative to that in the measurement as 6.7 dB or, translated into the RMS signal, this is a reduction in the influence of noise by a factor of 2.2.

The ability of nonlinear systems to use *negative* CLEs to reduce noise in a signal via a *single* pass of the noisy signal through the dynamical system was known to Soviet radio engineers in the 1950s and 1960s (M. Rabinovich, 2008, personal communication), and for curious reasons was never published explicitly. Some results are suggested by Kuznetsov *et al.* (1965).

In our procedure we have the mixed situation where, during the estimation procedure, the control ('nudge') terms are non-zero and used to regularize motions near the synchronization manifold, and in that role also reduce the noise perturbations. At the end of the estimation task, those controls are essentially zero yet, along the way, they provided a useful result. They are unable to provide as much noise reduction along system trajectories where the Lyapunov exponents are zero, but transverse to these trajectories, either in transients or on the system attractor, they work quite well in noise reduction.

6. Discussion

In this paper we have connected the irregularities in the minimal variance estimation of model initial conditions and fixed parameters with the instabilities of the subspace of the combined data plus model state space, the synchronization manifold, where observed data are required to equal model output. These instabilities are associated with positive conditional Lyapunov exponents for motions on that manifold, and we have shown how to regularize those instabilities and smooth out the surfaces on which one searches for estimates of parameters and initial conditions. The particular method for reducing CLEs to negative values used here is known in the atmospheric/oceanic literature as 'nudging,' in the control theory literature as a Luenberger observer, and it has been widely explored in nonlinear dynamics as a tool for parameter estimation of nonlinear models from data. There are other methods for synchronizing data and model systems (Abarbanel, 1996), but we have not explored them here.

We showed that adding control terms of the form $\mathbf{u}(t)\{\mathbf{x}(t) - \mathbf{y}(t)\}$ to the dynamical equations for the model state variables $\mathbf{y}(t)$ couples information from the observations $\mathbf{x}(t)$ in a manner that allows us to use the controls along an orbit $\mathbf{y}(t)$ to reduce CLEs through the Oseledec multiplicative ergodic theorem to negative values. Augmenting a variational principle for synchronization through minimization of

$$\frac{1}{2T}\int_0^T \{\mathbf{x}(t) - \mathbf{y}(t)\}^2 \mathrm{d}t$$

by adding a cost for the control variables, so one minimizes

$$\frac{1}{2T}\int_0^T \left[\{\mathbf{x}(t) - \mathbf{y}(t)\}^2 + \mu(t)^2\right] \mathrm{d}t. \tag{24}$$

This produces the added attractive result that, when the estimation procedure is completed, $\mathbf{u}(t) \approx 0$ and the dynamical equations are those dictated by the physics of the problem.

In estimating the fixed parameter and state variables of a dynamical system using observations, one is addressing three problems, more or less all together, so we wish to separate them as challenges remaining in each aspect of the overall problem:

- one must determine a cost or objective function that permits regularization of the instabilities of nonlinear systems related to chaotic behaviour of the system. Without this regularization, the incoherent oscillations of the observed data and the model output associated with sensitivity to initial conditions or parameter variations will impede any accurate search for these quantities.
- when an objective function has been established and equations of motion consistent with the desire to regularize motions on the synchronization manifold are implemented, then one must have an accurate numerical method for solving the implied variational problem, either in its strong (as here) or weak constraint format.
- one must establish validation criteria to examine the model produced by the outcome of the numerical procedure. Most tests of models are able to invalidate them. Here we proposed a consistency check associated with the regularization method for nonlinear systems.

The main goal of this paper has been a discussion of a regularized cost function and associated equations of motion with controls that permit estimation of fixed parameter and state variables. Our formulation was stated as a variational principle we called Dynamical State and Parameter Estimation (DSPE) consisting of the minimization of

$$C(\mathbf{y}, \mathbf{u}, \mathbf{p}) = \frac{1}{2(N+1)} \times \sum_{m=0}^{N} \left\{ \sum_{l=1}^{L} \{x_l(t_m) - y_l(t_m)\}^2 + \sum_{a,l=1}^{L} u_{al}(t_m)^2 \right\}, \tag{25}$$

subject to the equations of motion

$$\begin{aligned} y_a(t_{m+1}) &= f_a\{\mathbf{y}(t_m), \mathbf{p}\} \\ &\quad + \sum_{l=1}^{L} u_{al}(t_m)\{x_l(t_m) - y_l(t_m)\}, \\ &\quad a = 1, 2, \ldots, D, \end{aligned} \tag{26}$$

where D is the dimension of the model state space. Observations are made, in this formulation, of $x_l(t_m)$ for $l = 1, 2, \ldots, L$.

This is a small departure from standard settings for 4D-Var widely used in data assimilation. It is a form of strong 4D-Var as the dynamical equations are imposed as equality constraints on the numerical minimization of the cost function.

If one observes functions of the $\mathbf{x}(t)$, $z_l(t_m) = h_l\{\mathbf{x}(t_m)\}$ then the cost function and the modified equations of motion

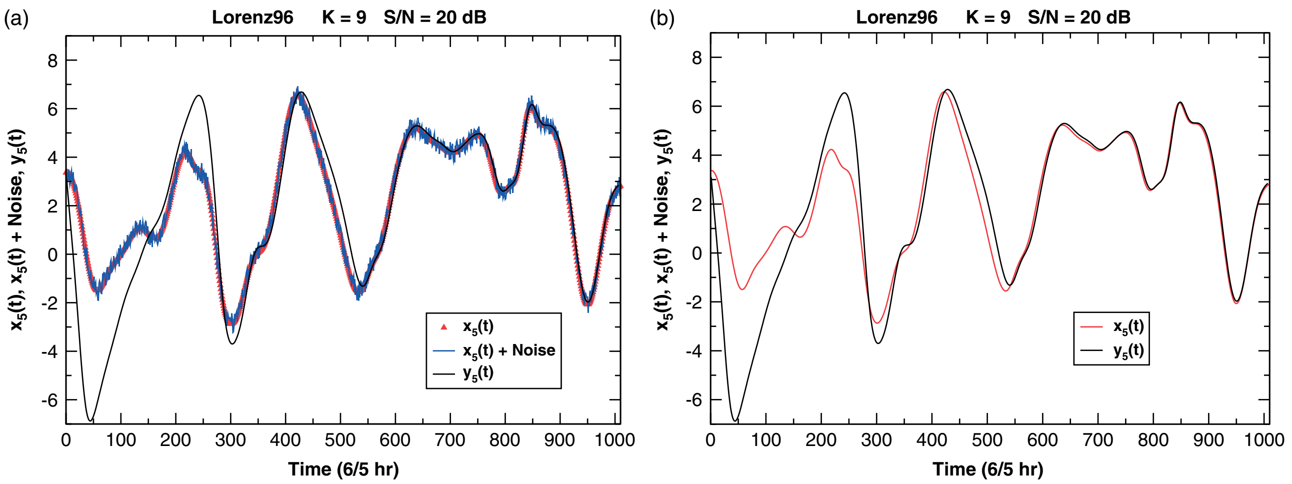

Figure 7. (a) The clean $x_5(t_m)$, noisy $x_5(t_m) + \eta_5(t_m)$, and estimated model output $y_5(t_m)$ when noise with a signal-to-noise ratio of 20 dB has been added to the signal presented to the model. (b) is as (a) but with the noisy signal removed so that comparison between the estimated model output and the clean data signal can more easily be made. Note that the output $y_5(t_m)$ has smoothed out the noisy signal presented to the model.

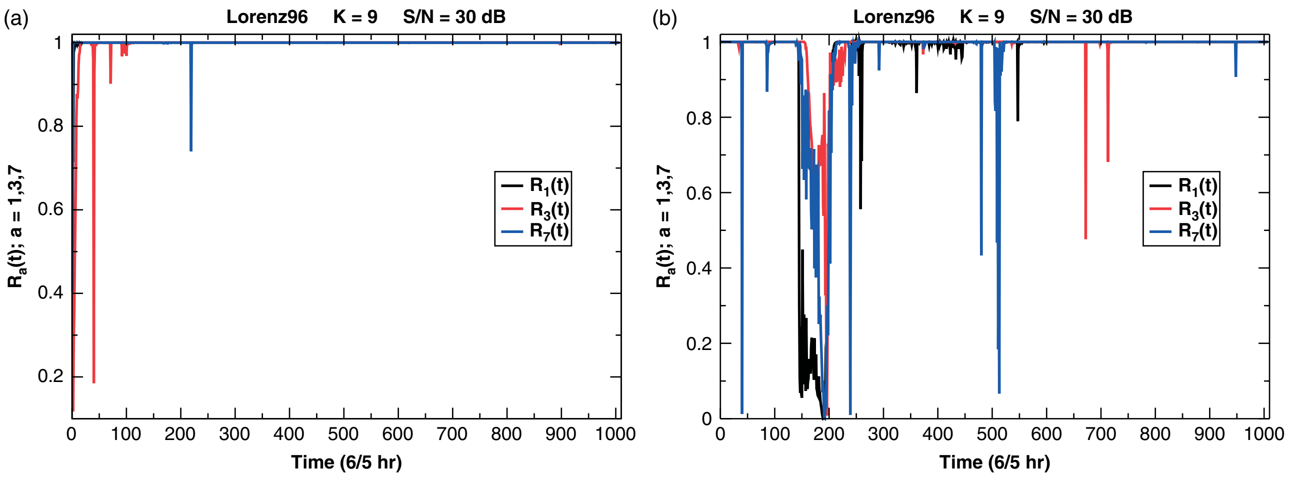

Figure 8. The consistency ratios $R_1(t_m), R_3(t_m), R_7(t_m)$ for data with observational noise of signal-to-noise ratios of (a) 30 dB and (b) 20 dB. The consistency test tells us that the model does not do well in representing the data for the early part of the whole assimilation period $[0, T \approx 1000\Delta t]$, and that the controls are working hard to make the model output closer to the data, but not succeeding well. In the latter part of the assimilation period, the model output is achieving an accurate representation of the clean data and is doing so without assistance from the controls. This suggests that the estimates of the state variables $y_a(T)$ at the end of the assimilation period may be quite accurate and provide the basis for an accurate forecast. Figures 9 and 10 show these forecasts.

should reflect this as the requirement to minimize

$$C(\mathbf{y}, \mathbf{u}, \mathbf{p}) = \frac{1}{2(N+1)} \times \sum_{m=0}^{N} \left\{ \sum_{l=1}^{L} [z_l(t_m) - h_l\{\mathbf{y}(t_m)\}]^2 + \sum_{a,l=1}^{L} u_{al}(t_m)^2 \right\}, \tag{27}$$

subject to the equations of motion

$$y_a(t_{m+1}) = f_a\{\mathbf{y}(t_m), \mathbf{p}\} + \sum_{l=1}^{L} u_{al}(t_m)[z_l(t_m) - h_l\{\mathbf{y}(t_m)\}]; \tag{28}$$
$$a = 1, 2, \ldots, D.$$

In our minimization of variance formulation, we have set all correlation matrices to unity, and correlations beyond this are rather easy to add to our formulation. Finally, we have used the strong-constraint formulation implying no model error either from lack of knowledge about underlying physical processes or from environmental influences driving the dynamical equations. This too can be incorporated into our formulation.

We have not attempted to explore the spectrum of numerical methods for solving the variational problem posed, but we have used the 'direct method' which treats all state variables at each observation time t_m, all the fixed parameters, and all the control or regularization variables $u_a(t_m)$ on an equal footing. Because of the quadratic penalty term in the cost function for the regularization variables $u_{ab}(t_m)$, once the optimization has achieved synchronization $x_l(t_m) \approx y_l(t_m)$, these regularization variables are driven to zero. This resolves the dilemma remaining after 'nudging' methods have been employed (Auroux and Blum, 2008; Telford *et al.*, 2008): the nudging variables are gone. What remains is the system physics: $y_a(t_{m+1}) = f_a\{\mathbf{y}(t_m), \mathbf{p}\}$; $a = 1, 2, \ldots, D$ with estimated parameters.

One useful result of the particular numerical approach we have adopted is that the estimated state variables available at the completion of the estimation procedure include the

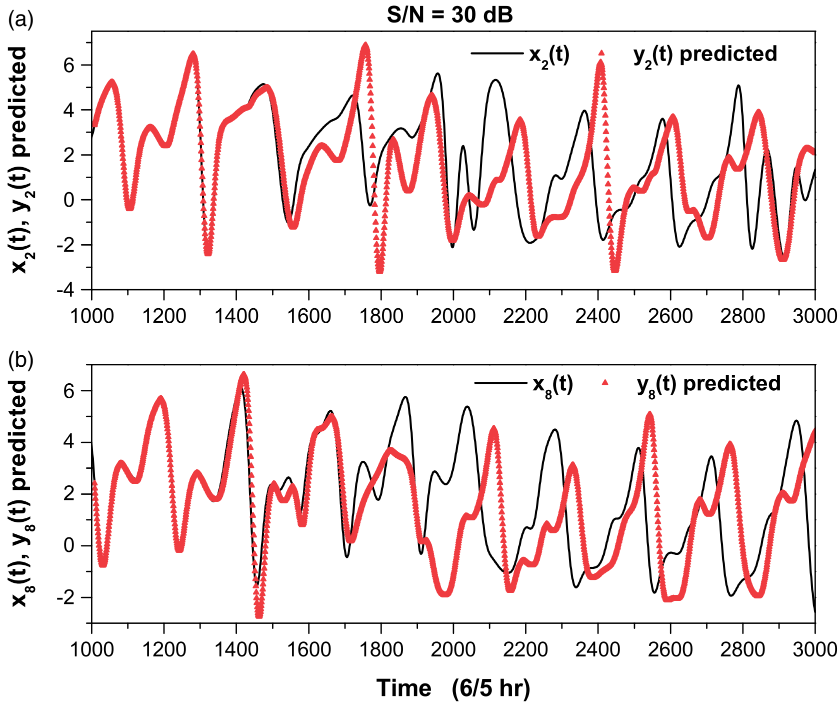

Figure 9. Using the estimated values of the model variables at the end of the estimation or assimilation period ($t = T$), the model equations with signal-to-noise ratio 30 dB were used to evaluate $y_a(t > T)$. (a) shows $y_2(t)$ compared to the known $x_2(t)$, while (b) shows the known $x_8(t)$ as well as the estimated model variable $y_8(t)$. The deviation of the model predictions comes from the chaotic behaviour of this dynamical system as well as the errors in the estimates of $\mathbf{y}(T)$ due to the noise.

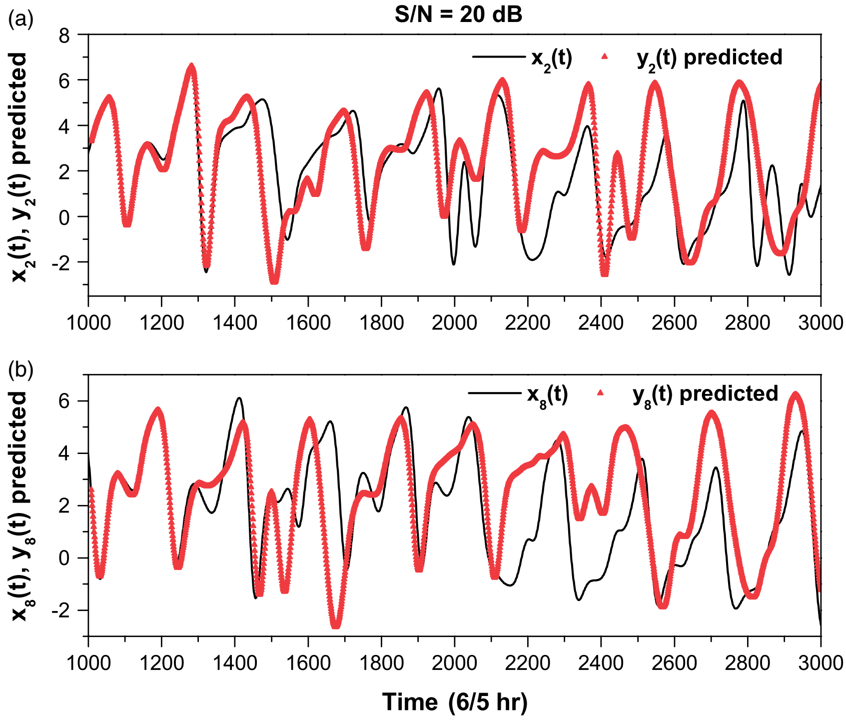

Figure 10. As Figure 9, but with a signal-to-noise ratio of 20 dB.

value of the full state of the system $\mathbf{y}(T)$ at the end of the observation interval $t_N = T$. With that state in hand, along with our estimation of the fixed parameters, we can use $y_a(t_{m+1}) = f_a\{\mathbf{y}(t_m), \mathbf{p}\}$; $a = 1, 2, \ldots, D$ to forecast $\mathbf{y}(t > T)$ with some confidence.

If we had estimated only the state of the model $\mathbf{y}(t_0)$ at the beginning of the observation period, we would have been required to use the nonlinear equations to evaluate $\mathbf{y}(t_N = T)$. That process carries with it all the uncertainties of chaotic orbits of nonlinear systems: any uncertainty in $\mathbf{y}(t_0)$ and round-off error inherent in the numerical methods to go from $\mathbf{y}(t_0)$ to $\mathbf{y}(T)$ would be amplified by the intrinsic instabilities of the nonlinear system.

We note again that, having eliminated the many local (false) minima in the dependence of the cost function on parameters or state variable values by utilizing a regularized dynamics, we may focus our attention on the development of the physics of the model system and numerical methods for extracting information from the data.

One more item of note is the possibility of identifying how many measurements are actually required to determine the full state of the model. This is the number of positive CLEs of the model conditioned on the data.

Finally there are a number of important items addressed in the extensive literature on data assimilation in geophysical problems that we have not addressed in this paper. While it is our plan to come back to these issues both in simplified models such as the 1996 Lorenz model and geophysically more interesting models, we have not done so yet. The issues include:

- model errors. These are often considered as random noise added to the dynamical equations (Evensen, 2007). The formulation of these error terms is to incorporate them into the cost function making the dynamical model part of the cost function itself.
- the relation between the formulation of DPSE and various forms of Kalman filtering (Kalnay, 2003, Chapters 5 and 6). While we have not addressed this comparison here, we recognize the use of Kalman filtering as a local linear approximation to nonlinear systems and as a useful and numerically helpful approach. By using the DSPE equations in a 'leapfrog' mode where the nonlinear integration works from step n to $n+1$, then the control terms work from $n+1$ to $n+2$, one can replicate, in format, the familiar background, analysis cycle (Kalnay, 2003; Evensen, 2007).
- in our reported work, we have assumed that observations are simple, indeed identity, functions of the system state variables, $h_l\{\mathbf{x}(t_m)\} = x_l(t_m)$. One can extend this to include nonlinear functions of the states $h_l\{\mathbf{x}(t_m)\}$ with a bit more notation, but we have not evaluated the implications for our algorithms. In principle, any nonlinear function relating the system state and the observations, if known explicitly, is not a theoretical problem, and, if not known explicitly, could be represented as a transfer function between states $x_l(t_m)$ and observations with additional fixed parameters to be estimated.
- if the number of unstable directions on the synchronization manifold exceeds the number of observations available, $U > L$, in the notation we have used in this paper, then one must employ some like time delay embedding to provide enough information to the model to assure that motions on the synchronization manifold are regularized. We have addressed this elsewhere (Abarbanel *et al.*, 2009), and will do so in the future in a geophysical context.
- we have not really addressed at all how one might cast the various theoretical observations about regularizing the synchronization manifold into practical, efficient numerical algorithms.
- we have not addressed here the interesting and important issue of ensemble methods and how our observations would be used in that context. It might be useful to add to the 'state', whose probability distribution is being estimated, the control terms $u_{al}(t_m)$, and one might wish to recognize the role of the invariant measure as a probability distribution instead of assuming that approximate Gaussian distributions will be enough.

However one chooses to implement DSPE, the recognition that stabilizing motion on the synchronization manifold is essential leads one to consider how many observations are actually required to determine parameters and states (at the end of the assimilation period). If one requires U as identified here, and if there are more than U observations, then one is not only in good shape but may use the 'extra' observations as a check on the forecasts. If there are fewer observations than required to stabilize the synchronization manifold, then additional information, perhaps in the form of constructing an embedding space, would be required.

Acknowledgements

We are very appreciative to M. I. Rabinovich for pointing out to us results of Stratonovich and colleagues apparently known to all workers in nonlinear dynamics in the former Soviet Union, and somehow not communicated effectively elsewhere. Our colleague Philip Gill, one of the creators of SNOPT, has provided invaluable advice on the use of SNOPT for this class of optimization problems. This work was partially funded by a grant from the National Science Foundation, NSF PHY0097134, and by a grant from the National Institutes of Health, NIH R01 NS40110-01A2. HDIA is partially supported by the NSF-sponsored Center for Theoretical Biological Physics at the University of California San Diego. The research on numerical optimization was partially funded by NSF grant NSF DMS-0511766. HDIA appreciates the hospitality and stimulating environment of the Institute of Neuroinformatics, University/ETH Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; in particular, many thanks to R. Hahnloser, K. Hepp, and R. Douglas.

References

Abarbanel HDI. 1996. *The Analysis of Observed Chaotic Data*. Springer: Berlin, Heidelberg, New York.

Abarbanel HDI, Creveling D, Gill PE. 2008. State and parameter estimation in nonlinear systems as an optimal tracking problem. *Phys. Lett. A* **372**: 2640–2644.

Abarbanel HDI, Creveling DR, Farsian R, Kostuk M. 2009. Dynamical state and parameter estimation. *SIAM J. Appl. Dyn. Syst.* **8**: 1341–1381.

Abarbanel HDI, Bryant P, Gill PE, Kostuk M, Rofe J, Singer Z, Toth B, Wong E. 2010. Dynamical parameter and state estimation in neuron models. In *Neuronal Variability and its Functional Significance*. Glanzman D, Ding M. (eds.) Oxford University Press: Oxford, UK. *in press*.

Aeyels, D. 1981a. Generic observability of differentiable systems. *SIAM J. Control Optim.* **19**: 595–603.

Aeyels D. 1981b. On the number of samples necessary to achieve observability. *Syst. Control Lett.* **1**: 92–94.

Auroux D, Blum J. 2008. A nudging-based data assimilation method for oceanographic problems: The back and forth nudging (BFN) algorithm. *Nonlin. Proc. Geophys.* **15**: 305–319.

Barclay A, Gill PE, Rosen JB. 1998. SQP methods in optimal control. In *Variational Calculus, Optimal Control and Applications*. Bulirsch R, Bittner L, Schmidt WH, Heier K. (eds.) Volume 124 *of International Series on Numerical Mathematics*. Birkhäuser: Basel, Boston, Berlin. 207–222.

Bewley TR, Moin P, Temam R. 2001. DNS-based predictive control of turbulence: an optimal benchmark for feedback algorithms. *J. Fluid Mech.* **447**: 179–225.

Blum J, Le Dimet F-X, Navon IM. 2008. Data assimilation of geophysical fluids. In *Computational Methods for the Atmosphere and the Oceans. Handbook of Numerical Analysis*, vol. XIV, Temam RM, Tribbia JJ. (eds.) 377–433.

Cessna J, Colburn C, Bewley TR. 2007. Multiscale retrograde estimation and forecasting of chaotic nonlinear systems. In *Proc. 46th IEEE Conference on Decision and Control*, New Orleans, La, USA, 12–14 December 2007. 2205–2210.

Chen S, Lu J. 2002. Parameters identification and synchronization of chaotic systems based on adaptive control. *Phys. Lett. A* **299**: 353–358.

Ciccarella G, Dalla Morta M, Germani A. 1993. Observers for discrete time nonlinear systems. *Syst. Control Lett.* **20**: 373–382.

Ciccarella G, Dalla Morta M, Germani A. 1995. A robust observer for discrete time nonlinear systems. *Syst. Control Lett.* **24**: 291–300.

Creveling D, Jeanne J, Abarbanel HDI. 2007. Parameter estimation using balanced synchronization. *Phys. Lett. A* **372**: 2043–2047.

Creveling D, Jeanne J, Abarbanel HDI. 2008. Estimation of parameters in nonlinear systems using balanced synchronization. *Phys. Rev. E* **77**: 016208.

Dedieu H, Ogorzalek MJ. 1997. Identifiability and identification of chaotic systems based on adaptive synchronization. *IEEE Trans. Circuits Syst. I* **44**: 948–962.

Evensen G. 2007. *Data Assimilation: The Ensemble Kalman Filter*. Springer: Berlin, Heidelberg, New York.

Garcia SP, Almeida JS. 2005. Nearest neighbor embedding with different time delays. *Phys. Rev. E* **71**: 037204.

Garcia SP, Almeida JS. 2005. Multivariate phase space reconstruction by nearest neighbor embedding with different time delays. *Phys. Rev. E* **72**: 027205.

Geisl P. 2007. *Construction of Global Lyapunov Functions using Radial Basis Functions*. Springer: Berlin, Heidelberg, New York.

Gill P, Murray W, Saunders MA. 1998. *User's Guide for SNOPT 5.3: A FORTRAN package for large-scale nonlinear programming*. Systems Optimization Laboratory: Stanford University, Calif, USA.

Gill PE, Murray W, Saunders MA. 2002a. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM J. Optimization* **12**: 979–1006.

Gill P, Murray W, Saunders MA. 2002b. *User's Guide for SNOPT version 6: A FORTRAN package for large-scale nonlinear programming*. Systems Optimization Laboratory: Stanford University, Calif, USA. www.cam.ucsd.edu/∼peg/papers/sndoc6.pdf.

Gill P, Murray W, Saunders MA. 2005. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review* **47**: 99–131.

Hirata Y, Suzuki H, Aihara K. 2006. Reconstructing state spaces from multivariate data using variable time delays. *Phys. Rev. E* **74**: 026202.

Huang D. 2004. Synchronization-based estimation of all parameters of chaotic systems from time series. *Phys. Rev. E* **69**: 067201.

Huijberts HJC. 1999. On existence of extended observer forms for nonlinear discrete-time systems. In *New Directions in Nonlinear Observer Design*. Nijmeijer H, Fossen TI. (eds.) Springer: Berlin, Heidelberg, New York. 77–92.

Huijberts HJC, Lilge T, Nijmeijer H. 2001. Nonlinear discrete-time synchronization via extended observers. *Int. J. Bifurc. Chaos* **11**: 1997–2006.

Kailath T. 1980. *Linear Systems*. Prentice-Hall: Englewood Cliffs, NJ, USA.

Kalnay E. 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press: Cambridge, UK.

Kantz H, Schreiber T. 2004. *Nonlinear Time Series Analysis*. (2nd edition). Cambridge University Press: Cambridge, UK.

Konnur R. 2003. Synchronization-based approach for estimating all model parameters of chaotic systems. *Phys. Rev. E* **67**: 027204.

Kostuk M, Whartenby W, Quinn J, Bryant P, Abarbanel HDI. 2010. Determining the number of measurements required for nonlinear data assimilation and prediction. *Phys. Rev. E submitted*.

Kuznetsov PI, Stratonovich RL, Tikhonov VI. 1965. Various papers in *Nonlinear Transformations of Stochastic Processes*. Wise J, Cooper DC. (transl.) Pergamon Press: London.

Lorenz EN. Predictability – A problem partly solved. 1996. In *Proceedings of the Seminar on Predictability*, volume 1. ECMWF: Reading, UK. 1–18.

Lorenz EN, Emanuel KA. 1998. Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci* **55**: 399–414.

Luenberger DG. 1964. Observing the state of a linear system, *IEEE Trans. Mil. Electron.* **3MIL**: 74–80.

Luenberger DG. 1966. Observers for multivariable systems. *IEEE Trans. Automat. Contr.* **ACll**: 190–197.

Luenberger DG. 1971. An introduction to observers. *IEEE Trans. Automat. Contr.* **AC16**: 396–602.

Luenberger DG. 1979. *Introduction to Dynamic Systems, Theory, Models and Applications.* Wiley: New York.

Marióson SF. 2002. Lyapunov function construction for ordinary differential equations with linear programming. *Dyn. Syst. Internat. J.* **17**: 137–150.

Maybhate A, Amritkar RE. 1999. Use of synchronization and adaptive control in parameter estimation from a time series. *Phys. Rev. E* **59**: 284.

Moral PE, Grizzle JW. 1995. Observer design for nonlinear systems with discrete-time measurements. *IEEE Trans. Autom. Control* **40**: 395–404.

Nijmeijer H, Mareels IMY. 1997. An observer looks at synchronization. *IEEE Trans. Circuits Syst. I* **44**: 882–890.

Oseledec VI. 1968. A multiplicative ergodic theorem: Ljapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.* **19**: 197–231.

Parlitz U. 1996. Estimating model parameters from time series by auto-synchronization. *Phys. Rev. Lett.* **76**: 1232–1235.

Parlitz U, Yu D. 2008. 'Synchronization and control-based parameter identification'. Preprint, University of Göttingen, Germany.

Parlitz U, Junge L, Kocarev L. 1996. Synchronization-based parameter estimation from time series. *Phys. Rev. E* **54**: 6253–6529.

Pecora L, Carroll T. 1990. Synchronization in chaotic systems. *Phys. Rev. Lett.* **64**: 821–824.

Pires C, Vautard R, Talagrand O. 1996. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus A* **48**: 96–121.

Roset B, Nijmeijer H. 2004. Observer-based model predictive control. *Int. J. Control* **77**: 1452–1462.

Roulston MS. 1999. 'Data assimilation in chaotic systems'. Woods Hole Oceanographic Institution: Mass, USA. http://www.whoi.edu/cms/files/rep_roulston_21485.pdf.

Sakaguchi H. 2002. Parameter evaluation from time sequences using chaos synchronization, *Phys. Rev. E* **65**: 02720.

Sauer T, Yorke JA, Casdagli M. 1991. Embedology. *J. Statist. Phys.* **65**: 579–616.

So P, Ott E, Dayawansa WP. 1994. Observing chaos: Deducing and tracking the state of a chaotic system from limited observation. *Phys. Rev. E* **49**: 2650–2660.

Sorrentino F, Ott E. 2009. Using synchronization of chaos to identify the dynamics of unknown systems. *Chaos* **19**: 033108; doi:10.1063/1.3186458.

Takens F. 1981. *Detecting strange attractors in turbulence*. *Lecture Notes in Mathematics* 898: 366–381. Springer: Berlin, Heidelberg.

Telford PJ, Braesicke P, Morgenstern O, Pyle JA. 2008. Technical Note: Description and assessment of a nudged version of the new dynamics Unified Model. *Atmos. Chem. Phys.* **8**: 1701–1712.

Tokuda I, Parlitz U, Illing L, Kennel MB, Abarbanel HDI. 2002. Parameter estimation for neuron models. *Proceedings of 7th Experimental Chaos Conference*, San Diego, Calif, USA.

Voss H, Timmer J, Kurths J. 2004. Nonlinear system identification from uncertain and indirect measurements. *Int. J. Bifurc. Chaos* **14**: 1905–1933.

Yang S-C, Baker D, Li H, Cordes K, Huff M, Nagpal G, Okereke E, Villafañe J, Kalnay E, Duane GS. 2006. Data assimilation as synchronization of truth and model: Experiments with the three-variable Lorenz system. *J. Atmos. Sci.* **63**: 2340–2354.

Yu D, Righero M, Kocarev L. 2006. Estimating topology of networks. *Phys. Rev. Lett.* **97**: 188701.

