Nonlinear Processes in Geophysics

Synchronicity in predictive modelling: a new view of data assimilation

G. S. Duane[1]**, J. J. Tribbia**[1]**, and J. B. Weiss**[2]

[1]National Center for Atmospheric Research, PO Box 3000, Boulder, CO 80307, USA
[2]Department of Atmospheric and Oceanic Sciences, UCB 311, University of Colorado, Boulder, CO 80309, USA

Abstract. The problem of data assimilation can be viewed as one of synchronizing two dynamical systems, one representing "truth" and the other representing "model", with a unidirectional flow of information between the two. Synchronization of truth and model defines a general view of data assimilation, as machine perception, that is reminiscent of the Jung-Pauli notion of synchronicity between matter and mind. The dynamical systems paradigm of the synchronization of a pair of loosely coupled chaotic systems is expected to be useful because quasi-2D geophysical fluid models have been shown to synchronize when only medium-scale modes are coupled. The synchronization approach is equivalent to standard approaches based on least-squares optimization, including Kalman filtering, except in highly non-linear regions of state space where observational noise links regimes with qualitatively different dynamics. The synchronization approach is used to calculate covariance inflation factors from parameters describing the bimodality of a one-dimensional system. The factors agree in overall magnitude with those used in operational practice on an ad hoc basis. The calculation is robust against the introduction of stochastic model error arising from unresolved scales.

1 Introduction

A computational model of a physical process that provides a stream of new data to the model as it runs must include a scheme to combine the new data with the model's prediction of the current state of the process. The goal of any such scheme is the optimal prediction of the future behavior of the physical process. While the relevance of the data assimilation problem is thus quite broad, techniques have been investigated most extensively for weather modeling, because

of the high dimensionality of the fluid dynamical state space, and the frequency of potentially useful new observational input. Existing data assimilation techniques (3DVar, 4DVar, Kalman Filtering, and Ensemble Kalman Filtering) combine observed data with the most recent forecast of the current state to form a best estimate of the true state of the atmosphere, each approach making different assumptions about the nature of the errors in the model and the observations.

An alternative view of the data assimilation problem is suggested here. The objective of the process is not to "now-cast" the current state of reality, but to make the model converge to reality in the future. Recognizing also that a predictive model, especially a large one, is a semi-autonomous dynamical system in its own right, influenced but not determined by observational input from a co-existing reality, it is seen that the guiding principle that is needed is one of synchronism. That is, we seek to introduce a one-way coupling between reality and model, such that the two tend to be in the same state, or in states that in some way correspond, at each instant of time. The problem of data assimilation thus reduces to the problem of synchronization of a pair of dynamical systems, unidirectionally coupled through a noisy channel that passes a limited number of "observed" variables.

While the synchronization of loosely coupled regular oscillators with limit cycle attractors is ubiquitous in nature (Strogatz, 2003), synchronization of chaotic oscillators has only been explored recently, in a wave of research spurred by the seminal work of Pecora and Carroll (1990). Chaos synchronization can be surprising because it implies that two systems, each effectively unpredictable, connected by a signal that can be virtually indistinguishable from noise, nevertheless exhibit a predictable relationship. Chaos synchronization has indeed been used to predict new kinds of weak teleconnection patterns relating different sectors of the global climate system (Duane, 1997; Duane et al., 1999; Duane and Tribbia, 2004).

Correspondence to: G. S. Duane
(gduane@ucar.edu)

It is now clear that chaos synchronization is surprisingly easy to arrange, in both ODE and PDE systems (Kocarev et al., 1997; Duane and Tribbia, 2001, 2004). A pair of spatially extended chaotic systems such as two quasi-2D fluid models, if coupled at only a discrete set of points and intermittently in time, can be made to synchronize completely. The application of chaos synchronization to the tracking of one dynamical system by another was proposed by So et al. (1994), so the synchronization of the fluid models suggests a natural extension to meteorological data assimilation that has not heretofore been recognized.

Since the problem of data assimilation arises in any situation requiring a computational model of a parallel physical process to track that process as accurately as possible based on limited input, it is suggested here that the broadest view of data assimilation is that of machine perception by an artificially intelligent system. Indeed, the new field of Dynamic Data Driven Application Systems (DDDAS) is defined as the real-time modelling of evolving physical systems based on select observations[1]. Like a data assimilation system, the human mind forms a model of reality that functions well, despite limited sensory input, and one would like to impart such an ability to the computational model. In the artificial intelligence view of data assimilation, the additional issue of model error can be approached naturally as a problem of machine learning, as discussed in the concluding section.

In this more general context, the role of synchronism is reminiscent of the psychologist Carl Jung's notion of synchronicity in his view of the relationship between mind and the material world. Jung had noted uncanny coincidences or "synchronicities" between mental and physical phenomena. In collaboration with Wolfgang Pauli (Jung and Pauli, 1955), he took such relationships to reflect a new kind of order connecting the two realms. (The new order was taken to explain relationships between seemingly unconnected phenomena in the objective world as well.) It was important to Jung and Pauli that synchronicities themselves were distinct, isolated events, but as described in Sect. 2.1, such phenomena can emerge naturally as a degraded form of chaos synchronization.

A principal question that is addressed in this paper is whether the synchronization view of data assimilation is merely an appealing reformulation of standard treatments, or is different in substance. The first point to be made is that all standard data assimilation approaches, if successful, do achieve synchronization, so that synchronization defines a more general family of algorithms that includes the standard ones. It remains to determine whether there are synchronization schemes that lead to faster convergence than the standard data assimilation algorithms. It is shown here analytically that optimal synchronization is equivalent to Kalman filtering when the dynamics change slowly in phase space, so that the same linear approximation is valid at each point

in time for the real dynamical system and its model. When the dynamics change rapidly, as in the vicinity of a regime transition, one must consider the full nonlinear equations and there are better synchronization strategies than the one given by Kalman filtering or ensemble Kalman filtering. The deficiencies of the standard methods, which are well known in such situations, are usually remedied by ad hoc corrections, such as "covariance inflation" (Anderson, 2001). In the synchronization view, such corrections can be derived from first principles.

This paper takes a broad view of data assimilation by a model system, defined as a set of differential equations, that is coupled to noisy data obtained from a "true system", defined by the same set of differential equations, with the possible addition of a stochastic term to represent model error. We begin by reviewing the phenomenology of chaos synchronization generally in Sect. 2.1, and an application to geophysical fluid systems in Sect. 2.2. A brief review of standard data assimilation is provided in Sect. 2.3. In Sect. 3 the synchronization approach is compared to standard approaches. The optimal synchronization problem for a coupled pair of stochastic differential equations is framed as a problem of finding the coupling that gives the tightest synchronization in a linear approximation with observational noise. The optimal coupling thus derived can be compared to forms used in standard data assimilation. The difference becomes large in regions of state-space where nonlinearities are important. In Sect. 4, a comparison of the two approaches for the full nonlinear case is used to estimate covariance inflation factors that would be needed to adjust the Kalman filter scheme to give optimal synchronization, for both perfect models and models including stochastic error from unresolved scales. Section 5 concludes by expanding on the view of data assimilation as machine perception and discussing automatic model adaptation in the synchronization framework.

2 Background: synchronized chaos and data assimilation

2.1 Chaos synchronization

The phenomenon of chaos synchronization was first brought to light by Fujisaka and Yamada (1983) and independently by Afraimovich et al. (1986). Extensive research on the synchronization of chaotic systems in the '90s was spurred by the work of Pecora and Carroll (1991), who found that two Lorenz (1963) systems would synchronize when the X or Y variable of one was slaved to the respective X or Y variable of the other, despite sensitive dependence on initial values of the other variables. (Synchronization does not occur if the Z variables are analogously coupled.)

In this paper we consider a weaker *diffusive* form of coupling, as illustrated by the following pair of bidirectionally

---

[1]http://www.dddas.org

coupled Lorenz systems:

$$\dot{X} = \sigma(Y - Z) + \alpha(X_1 - X)$$
$$\dot{Y} = \rho X - Y - XZ$$
$$\dot{Z} = -\beta Z + XY$$

$$ \tag{1} $$

$$\dot{X}_1 = \sigma(Y_1 - Z_1) + \alpha(X - X_1)$$
$$\dot{Y}_1 = \rho X_1 - Y_1 - X_1 Z_1$$
$$\dot{Z}_1 = -\beta Z_1 + X_1 Y_1$$

where α parameterizes the coupling strength. The two Lorenz systems synchronize rapidly for appropriate values of α, and also do so for unidirectional coupling, defined by removing the term in α from the first equation.

For a pair of coupled systems that are not identical, as with an imperfect model of a physical system, synchronization may still occur, but the correspondence between the states of the two systems, that defines the *synchronization manifold* in state space is different from the identity. In this situation, known as *generalized synchronization*, we have two different dynamical systems $\dot{x}=F(x)$ and $\dot{y}=G(y)$, with $x, y \in R^n$, modified in some manner so as to define two coupled systems $\dot{x}=\hat{F}(x, y)$ and $\dot{y}=\hat{G}(y, x)$. The systems are said to be generally synchronized iff there is some locally invertible function $\Phi : R^n \to R^n$ such that $||\Phi(x)-y|| \to 0$ as $t \to \infty$ (Rulkov et al., 1995). Generalized synchronization can be shown to occur even for very different systems, as with a Rossler system coupled to a Lorenz system, but with a correspondence function Φ that is nowhere smooth (Pecora et al., 1997).

It is commonly not the existence, but the stability of the *synchronization manifold* in state space that distinguishes coupled systems exhibiting synchronization from those that do not (such as Eq. 1 for different values of α). As the coupling is weakened, bursts of desynchronization (a special case of on-off intermittency) interrupt the synchronized behavior. On-off synchronization, that can also arise from noise in the communication channel between the two systems, is a second way that identical synchronization is found to degrade (Ashwin et al., 1994). In the data assimilation application, it corresponds to "catastrophes" of large model drift that can arise from observational noise (Baek et al., 2004).

2.2 Synchronization between geophysical fluid systems

Pairs of 1D PDE systems of various types, coupled diffusively at discrete points in space and time, were shown to synchronize by Kocarev et al. (1997). Synchronization in geophysical fluid models was demonstrated by Duane and Tribbia (2001), originally with a view toward predicting and explaining new families of long-range teleconnections (Duane and Tribbia, 2004).

The uncoupled single-system model, derived from one described by Vautard et al. (1988), is given by the quasi-
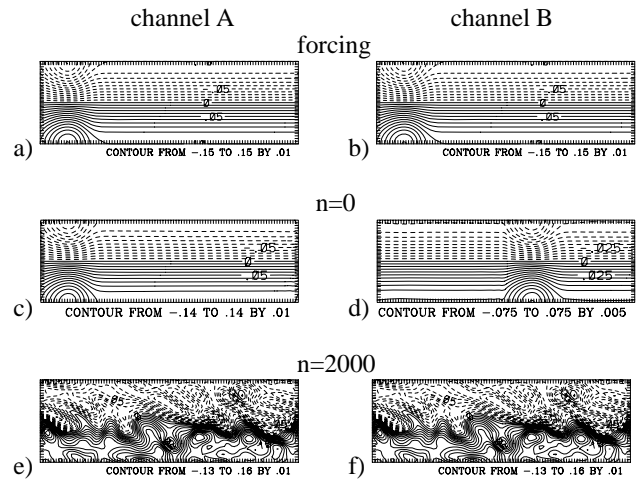

Fig. 1. Streamfunction (in units of 1.48×10^9 m^2 s^{-1}) describing the forcing ψ^* (**a, b**), and the evolving flow ψ (**c–f**), in a parallel channel model with bidirectional coupling of medium scale modes for which $|k_x|>k_{x0}=3$ or $|k_y|>k_{y0}=2$, and $|k| \leq 15$, for the indicated numbers n of time steps in a numerical integration. Parameters are as in Duane and Tribbia (2004). An average streamfunction for the two vertical layers $i=1, 2$ is shown. Synchronization occurs by the last time shown (**e, f**), despite differing initial conditions.

geostrophic equation for potential vorticity q in a two-layer reentrant channel on a β-plane:

$$\frac{Dq_i}{Dt} \equiv \frac{\partial q_i}{\partial t} + J(\psi_i, q_i) = F_i + D_i \tag{2}$$

where the layer $i=1, 2$, ψ is streamfunction, and the Jacobian $J(\psi, \cdot) = \frac{\partial \psi}{\partial x}\frac{\partial \cdot}{\partial y} - \frac{\partial \psi}{\partial y}\frac{\partial \cdot}{\partial x}$ gives the advective contribution to the Lagrangian derivative D/Dt. Equation (2) states that potential vorticity is conserved on a moving parcel, except for forcing F_i and dissipation D_i. The discretized potential vorticity is

$$q_i = f_0 + \beta y + \nabla^2 \psi_i + R_i^{-2}(\psi_1 - \psi_2)(-1)^i \tag{3}$$

where $f(x, y)$ is the vorticity due to the Earth's rotation at each point (x, y), f_0 is the average f in the channel, β is the constant df/dy and R_i is the Rossby radius of deformation in each layer. The forcing F is a relaxation term designed to induce a jet-like flow near the beginning of the channel: $F_i = \mu_0(q_i^* - q_i)$ for q_i^* corresponding to the choice of ψ^* shown in Fig. 1a. The dissipation terms D, boundary conditions, and other parameter values are given in Duane and Tribbia (2004).

Two models of the form (2), $Dq^A/Dt=F^A+D^A$ and $Dq^B/Dt=F^B+D^B$ were coupled diffusively in one direction by modifying one of the forcing terms:

$$F_k^B = \mu_k^c[q_k^A - q_k^B] + \mu_k^{\text{ext}}[q_k^* - q_k^B] \tag{4}$$

where the flow has been decomposed spectrally and the subscript k on each quantity indicates the wave number k spectral component. (The layer index i has been suppressed.)

channel A channel B

n=40:

n=280:

CONTOUR FROM −10 TO 10 BY 1   CONTOUR FROM −8 TO 7 BY 1

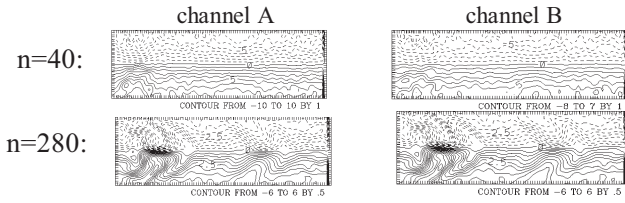CONTOUR FROM −6 TO 6 BY .5   CONTOUR FROM −6 TO 6 BY .5

Fig. 2. Flow results as in Fig. 1, but with the inter-channel coupling restricted to the 10-local-bred-vector subspace as in Eq. (5). Synchronization is apparent by the time step shown.

The two sets of coefficients $\mu_{\boldsymbol{k}}^c$ and $\mu_{\boldsymbol{k}}^{\text{ext}}$ were chosen to couple the two channels in some medium range of wavenumbers and to force each channel only with the low wavenumber components of the background flow.

It was found that the two channels rapidly synchronize if only the medium scale modes are coupled (Fig. 1), starting from initial flow patterns that are arbitarily set equal to the forcing in one channel, and to a rather different pattern in the other channel. (Results are shown for bidirectional coupling defined by adding an equation for $F_{\boldsymbol{k}}^A$ analogous to Eq. 4. The synchronization behavior for coupling in just one direction is very similar.) With unidirectional coupling, the synchronization effects data assimilation from the A channel into the B channel.

One question about data assimilation that may be addressed in the synchronization context concerns the definition and minimum number of variables that must be assimilated to give adequate predictive skill. It has been argued, for instance, that atmospheric dynamics is locally low-dimensional, and that a small number of locally selected bred vectors spans the effective state space in each local region (Patil et al., 2001). Bred vectors are commonly used to specify likely directions of forecast error (Toth and Kalnay, 1993, 1997).

If Patil et al. (2001)'s argument about low "BV-dimension" is correct, then it should only be necessary to couple two channels in the subspace defined by the properly chosen bred vectors, in each local region, to synchronize the two channels. That is, it should be possible to replace Eq. (4) by

$$F^B(x, y) = \mu^{BV} \sum_i [(q^A - q^B) \cdot b_i]_k b_i(x, y) + F^{B\text{ext}}(x, y) \quad (5)$$

for $(x, y) \in r^k$, where the b_i for $i = 1 \dots 10$ are an orthonormal set of vectors formed from ten bred vectors B_i by Gram-Schmidt orthogonalization in each local region separately. That is, the bred vectors B_i are first computed globally for the channel as a whole, as in Toth and Kalnay (1997). Then, the channel is divided rectangularly into a 20×16 patchwork of local regions r^k $k = 1 \dots 320$. The set of vectors b_i is formed from B_i by Gram-Schmidt orthogonalization of the set $B_i(x, y):(x, y) \in r^k$ for each r^k and concatenating the re-

sulting vectors over all regions. The dot product in brackets in Eq. (5) is computed separately for each local region:

$$[v \cdot w]_k \equiv \sum_{(x, y) \in r^k} v(x, y) w(x, y) \quad (6)$$

so that $320 \times 10 = 3200$ independent coefficients $[(q^A - q^B) \cdot b_i]_k$ are computed at each instant of time. Thus $[b_i \cdot b_j]_k = \delta_{ij}$ for all k. The overall coupling strength is given by μ^{BV}, and the external forcing by the jet is defined by $F_{\boldsymbol{k}}^{B\text{ext}} \equiv \mu_{\boldsymbol{k}}^{\text{ext}} [q_{\boldsymbol{k}}^* - q_{\boldsymbol{k}}^B]$ as before.

It is found that two channels coupled in a truncated bred vector basis according to Eq. (5) do synchronize, as illustrated in Fig. 2, and do not synchronize with fewer independent regions or if fewer bred vectors are used to define the coupling subspace. The total number of independent coefficients, however, is comparable to or larger than the total number of Fourier components in the mid-range of scales that was seen to be effective for synchronization in the coupling scheme (4).

The lesson is that the synchronization phenomenon does not appear to be very sensitive to the detailed choice of coupling subspace. A similar conclusion was reached by Yang et al. (2004) for synchronizing Lorenz systems. Those authors obtained only small improvement by using bred vectors or singular vectors instead of single-variable coupling. In the present case of spatially extended models, it seems that any basis that captures the essential physical phenomena in each local region, phenomena that can be described in terms of a middle range of scales, is adequate for synchronization and hence for data assimilation.

2.3 Data assimilation

Standard data assimilation, unlike synchronization, estimates the *current* state $\boldsymbol{x}_T \in R^n$ of one system, "truth", from the state of a model system $\boldsymbol{x}_B \in R^n$, combined with noisy observations of truth. The best estimate of truth is the "analysis" $\boldsymbol{x}_A$, which is the state that minimizes error as compared to all possible linear combinations of observations and model. That is

$$\boldsymbol{x}_A \equiv \boldsymbol{x}_B + \boldsymbol{\Lambda}(\boldsymbol{x}_{\text{obs}} - \boldsymbol{x}_B) \quad (7)$$

minimizes the analysis error $<(\boldsymbol{x}_A - \boldsymbol{x}_T)^2>$ for a stochastic distribution given by $\boldsymbol{x}_{\text{obs}} = \boldsymbol{x}_T + \boldsymbol{\xi}$ where $\boldsymbol{\xi}$ is observational noise, for properly chosen $n \times n$ gain matrix $\boldsymbol{\Lambda}$. The standard methods to be considered in this paper correspond to specific forms for the generally time-dependent matrix $\boldsymbol{\Lambda}$.

The simple method known as 3dVar uses a time-independent $\boldsymbol{\Lambda}$ that is based on the time-averaged statistical properties of the observational noise and the resulting forecast error. Let the matrix

$$\mathbf{R} \equiv <\boldsymbol{\xi}\boldsymbol{\xi}^T> = <(\boldsymbol{x}_{\text{obs}} - \boldsymbol{x}_T)(\boldsymbol{x}_{\text{obs}} - \boldsymbol{x}_T)^T> \quad (8)$$

be the observation error covariance, and the matrix

$$\mathbf{B} \equiv <(\boldsymbol{x}_B - \boldsymbol{x}_T)(\boldsymbol{x}_B - \boldsymbol{x}_T)^T> \quad (9)$$

be the "background" error covariance, describing the deviation of the model state from the true state. If both covariance matrices are assumed to be constant in time, then the optimal linear combination of background and observations is:

$$x_A = \mathbf{R}(\mathbf{R} + \mathbf{B})^{-1}x_B + \mathbf{B}(\mathbf{R} + \mathbf{B})^{-1}x_{\mathrm{obs}} \qquad (10)$$

The formula (10), which simply states that observations are weighted more heavily when background error is greater and conversely, defines the 3dVar method in practical data assimilation, based on empirical estimates of $\mathbf{R}$ and $\mathbf{B}$. The 4dVar method, which will not be considered here, generalizes Eq. (10) to estimate a short history of true states from a corresponding short history of observations.

The Kalman filtering method, that is popular for a variety of tracking problems, uses the dynamics of the model to update the background error covariance $\mathbf{B}$ sequentially. The analysis at each assimilation cycle i is:

$$x_A^i = \mathbf{R}(\mathbf{R} + \mathbf{B}^i)^{-1}x_B^i + \mathbf{B}^i(\mathbf{R} + \mathbf{B}^i)^{-1}x_{\mathrm{obs}}^i \qquad (11)$$

where the background x_B^i is formed from the previous analysis x_A^{i-1} simply by running the model $M : R^n \to R^n$

$$x_B^i = M_{i-1 \to i}(x_A^{i-1}) \qquad (12)$$

as is done in 3dVar. But now the background error is updated according to

$$\mathbf{B}^i = \mathbf{M}_{i-1 \to i}\mathbf{A}^{i-1}\mathbf{M}_{i-1 \to i}^T + \mathbf{Q} \qquad (13)$$

where $\mathbf{A}$ is the analysis error covariance $\mathbf{A} \equiv <(x_A - x_T)(x_A - x_T)^T>$, given conveniently by $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{R}^{-1}$. The matrix $\mathbf{M}$ is the *tangent linear model* given by

$$\mathbf{M}_{ab} \equiv \left.\frac{\partial M_b}{\partial x_a}\right|_{x=x_A} \qquad (14)$$

The update formula (13) gives the minimum analysis error $<(x_A - x_T)^2> = Tr\mathbf{A}$ at each cycle. The term $\mathbf{Q}$ is the covariance of the error in the model itself, as discussed in Sect. 4.

3 Comparison of synchronization with standard methods of data assimilation

3.1 Optimal coupling for synchronization of stochastic differential equations

To compare synchronization to standard data assimilation, we inquire as to the coupling that is optimal for synchronization, so that this coupling can be compared to the gain matrix used in the standard 3dVar and Kalman filtering schemes. The general form of coupling of truth to model that we consider in this section is given by a system of stochastic differential equations:

$$\dot{x}_T = f(x_T)$$
$$\dot{x}_B = f(x_B) + \mathbf{C}(x_T - x_B + \xi) \qquad (15)$$

where true state $x_T \in R^n$ and the model state $x_B \in R^n$ evolve according to the same dynamics, given by f, and where the noise ξ in the coupling (observation) channel is the only source of stochasticity. The form (15) is meant to include dynamics f described by partial differential equations, as in the last section. The system is assumed to reach an equilibrium probability distribution, centered on the synchronization manifold $x_B = x_T$. The goal is to choose a time-dependent matrix $\mathbf{C}$ so as to minimize the spread of the distribution.

Note that if $\mathbf{C}$ is a projection matrix, or a multiple of the identity, then Eq. (15) effects a form of nudging. But for arbitrary $\mathbf{C}$, the scheme is much more general. Indeed, continuous-time generalizations of 3DVar and Kalman filtering can be put in the form (15).

Let us assume that the dynamics vary slowly in state space, so that the Jacobian $\mathbf{F} \equiv Df$, at a given instant, is the same for the two systems

$$Df(x_B) = Df(x_T) \qquad (16)$$

where terms of $O(x_B - x_T)$ are ignored. Then the difference between the two Eqs. (15), in a linearized approximation, is

$$\dot{e} = \mathbf{F}e - \mathbf{C}e + \mathbf{C}\xi \qquad (17)$$

where $e \equiv x_B - x_T$ is the synchronization error.

The stochastic differential equation (17) implies a deterministic partial differential equation, the Fokker-Planck equation, for the probability distribution $\rho(e)$:

$$\frac{\partial \rho}{\partial t} + \nabla_e \cdot [\rho(\mathbf{F} - \mathbf{C})e] = \frac{1}{2}\delta\nabla_e \cdot (\mathbf{C}\mathbf{R}\mathbf{C}^T\nabla_e\rho) \qquad (18)$$

where $\mathbf{R} = <\xi\xi^T>$ is the observation error covariance matrix, and δ is a time-scale characteristic of the noise, analogous to the discrete time between molecular kicks in a Brownian motion process that is represented as a continuous process in Einstein's well known treatment. Equation (18) states that the local change in ρ is given by the divergence of a probability current $\rho(\mathbf{F} - \mathbf{C})e$ except for random "kicks" due to the stochastic term.

The PDF can be taken to have the Gaussian form $\rho = N\exp(-e^T\mathbf{K}e)$, where the matrix $\mathbf{K}$ is the inverse spread, and N is a normalization factor, chosen so that $\int \rho d^n e = 1$. For background error covariance $\mathbf{B}$, $\mathbf{K} = (2\mathbf{B})^{-1}$. In the one-dimensional case, $n=1$, where C and K are scalars, substitution of the Gaussian form in Eq. (18), for the stationary case where $\partial\rho/\partial t = 0$ yields:

$$2B(C - F) = \delta RC^2 \qquad (19)$$

Solving $dB/dC = 0$, it is readily seen that B is minimized (K is maximized) when $C = 2F = (1/\delta)B/R$.

In the multidimensional case, $n > 1$, the relation (19) generalizes to the *fluctuation-dissipation relation*

$$\mathbf{B}(\mathbf{C} - \mathbf{F})^T + (\mathbf{C} - \mathbf{F})\mathbf{B} = \delta\mathbf{C}\mathbf{R}\mathbf{C}^T \qquad (20)$$
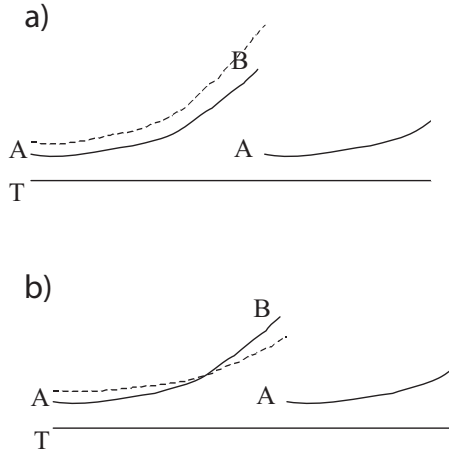
a)

b)

Fig. 3. An analysis cycle, with trajectories shown for the true state "T", the model evolving from the initial analysis "A" to the next forecast, or background "B", and an alternative model run (dotted line) starting from an inferior "analysis" that is further from the initial truth. In the case **(a)** where $Df(\boldsymbol{x}_T)=Df(\boldsymbol{x}_A)$, then a worse analyis will always produce a worse forecast, but in the general case **(b)** where $Df(\boldsymbol{x}_T)\neq Df(\boldsymbol{x}_A)$, nonlinearities may allow a worse analysis to evolve to a better forecast (the trajectories do not actually cross).

that can be obtained directly from the stochastic differential equation (17) by a standard proof that is reproduced in Appendix A. **B** can then be minimized element-wise. Differentiating the matrix equation (20) with respect to the elements of **C**, we find

$$d\mathbf{B}(\mathbf{C} - \mathbf{F})^T + \mathbf{B}(d\mathbf{C})^T + (d\mathbf{C})\mathbf{B} + (\mathbf{C} - \mathbf{F})d\mathbf{B}$$
$$= \delta[(d\mathbf{C})\mathbf{R}\mathbf{C}^T + \mathbf{C}\mathbf{R}(d\mathbf{C})^T] \qquad (21)$$

where the matrix $d\mathbf{C}$ represents a set of arbitrary increments in the elements of **C**, and the matrix $d\mathbf{B}$ represents the resulting increments in the elements of **B**. Setting $d\mathbf{B}=0$, we have

$$[\mathbf{B} - \delta\mathbf{C}\mathbf{R}](d\mathbf{C})^T + (d\mathbf{C})[\mathbf{B} - \delta\mathbf{R}\mathbf{C}^T] = 0 \qquad (22)$$

Since the matrices **B** and **R** are each symmetric, the two terms in Eq. (22) are transposes of one another. It is easily shown that the vanishing of their sum, for arbitrary $d\mathbf{C}$, implies the vanishing of the factors in brackets in Eq. (22). Therefore $\mathbf{C}=(1/\delta)\mathbf{B}\mathbf{R}^{-1}$, as in the 1D case.

3.2 Optimal synchronization vs. least-squares data assimilation

Turning now to the standard methods, so that a comparison can be made, it is recalled that the analysis $\boldsymbol{x}_A$ after each cycle is given by:

$$\boldsymbol{x}_A = \mathbf{R}(\mathbf{R} + \mathbf{B})^{-1}\boldsymbol{x}_B + \mathbf{B}(\mathbf{R} + \mathbf{B})^{-1}\boldsymbol{x}_{\text{obs}}$$
$$= \boldsymbol{x}_B + \mathbf{B}(\mathbf{R} + \mathbf{B})^{-1}(\boldsymbol{x}_{\text{obs}} - \boldsymbol{x}_B) \qquad (23)$$

In 3dVar, the background error covariance matrix **B** is fixed; in Kalman filtering it is updated after each cycle using the linearized dynamics. The background for the next cycle is computed from the previous analysis by integrating the dynamical equations:

$$\boldsymbol{x}_B^{n+1} = \boldsymbol{x}_A^n + \tau f(\boldsymbol{x}_A^n) \qquad (24)$$

where τ is the time interval between successive analyses. Thus the forecasts satisfy a difference equation:

$$\boldsymbol{x}_B^{n+1} = \boldsymbol{x}_B^n + \mathbf{B}(\mathbf{R} + \mathbf{B})^{-1}(\boldsymbol{x}_{\text{obs}}^n - \boldsymbol{x}_B^n) + \tau f(\boldsymbol{x}_A^n) \qquad (25)$$

We model the discrete process as a continuous process in which analysis and forecast are the same:

$$\dot{\boldsymbol{x}}_B = f(\boldsymbol{x}_B) + 1/\tau \mathbf{B}(\mathbf{B} + \mathbf{R})^{-1}(\boldsymbol{x}_T - \boldsymbol{x}_B + \boldsymbol{\xi})$$
$$+ O[(\mathbf{B}(\mathbf{B} + \mathbf{R})^{-1})^2] \qquad (26)$$

using the white noise $\boldsymbol{\xi}$ to represent the difference between observation $\boldsymbol{x}_{\text{obs}}$ and truth $\boldsymbol{x}_T$. The continuous approximation is valid so long as f varies slowly on the time-scale τ.

It is seen that when background error is small compared to observation error, the higher order terms $O[(\mathbf{B}(\mathbf{B}+\mathbf{R})^{-1})^2]$ can be neglected and the optimal coupling $\mathbf{C}=1/\delta\mathbf{B}\mathbf{R}^{-1}$ is just the form that appears in the continuous data assimilation equation (26), for $\delta=\tau$. Thus under the linear assumption that $Df(\boldsymbol{x}_B)=Df(\boldsymbol{x}_T)$, the synchronization approach is equivalent to 3dVar in the case of constant background error, and to Kalman filtering if background error is dynamically updated over time. The equivalence can also be shown for an exact description of the discrete analysis cycle, by comparing it to a coupled pair of synchronized maps. See Appendix B.

The equivalence between synchronization and standard methods in the linear case actually follows easily from a comparison of the optimization principles that define the two approaches. In the standard approaches, the form (23) minimizes the expected value of $(\boldsymbol{x}_A-\boldsymbol{x}_T)^2$, as compared to all other linear combinations of $\boldsymbol{x}_{\text{obs}}$ and $\boldsymbol{x}_B$. But if double linearization (16) holds then the minimization of $<(\boldsymbol{x}_A-\boldsymbol{x}_T)^2>$ implies the minimization of $<(\boldsymbol{x}_B-\boldsymbol{x}_T)^2>$ at any future time.

To see this, first consider the background error after a period of time τ, just before the next analysis, as in Fig. 3a. If double linearization (16) holds, then this projected background error is related to the initial analysis error by

$$\boldsymbol{e}(t) = T\left[\exp\int_{t_o}^t \mathbf{F}(t')dt'\right]\boldsymbol{e}(t_o) \equiv \mathbf{M}\boldsymbol{e}(t_o) \qquad (27)$$

where the notation T before the expression in brackets denotes time-ordering. The expectations are related by

$$\mathbf{B}=<\boldsymbol{e}(t)\boldsymbol{e}^T(t)>=\mathbf{M}<\boldsymbol{e}(t_o)\boldsymbol{e}^T(t_o)>\mathbf{M}^T=\mathbf{M}\mathbf{A}\mathbf{M}^T \qquad (28)$$

If we consider a more general "analysis" $\boldsymbol{x}_A^{\Lambda}$ formed from a general linear combination of forecast and observations

$$\boldsymbol{x}_A^{\Lambda} \equiv \boldsymbol{x}_B + \mathbf{\Lambda}(\boldsymbol{x}_{\text{obs}} - \boldsymbol{x}_B) \qquad (29)$$

and compute a general "analysis error covariance" $\mathbf{A}^{\mathbf{\Lambda}} \equiv <(\boldsymbol{x}_A^{\mathbf{\Lambda}} - \boldsymbol{x}_T)(\boldsymbol{x}_A^{\mathbf{\Lambda}} - \boldsymbol{x}_T)^T>$ accordingly, we seek to minimize the trace $Tr\mathbf{A}^{\mathbf{\Lambda}} = <(\boldsymbol{x}_A^{\mathbf{\Lambda}} - \boldsymbol{x}_T)^2>$. But it is readily shown that a solution $\mathbf{\Lambda}$ of $d_{\mathbf{\Lambda}} Tr\mathbf{A}^{\mathbf{\Lambda}} = 0$ is also a solution of $d_{\mathbf{\Lambda}} Tr\mathbf{M}(t)\mathbf{A}^{\mathbf{\Lambda}}\mathbf{M}^T(t) = 0$ for $0 < t \leq \tau$. (The operator "$d_{\mathbf{\Lambda}}$" produces a *matrix* with elements that are the derivatives of the argument with respect to the corresponding elements of $\mathbf{\Lambda}$.) Thus the best analysis gives the best forecast just prior to the next analysis. It is then readily seen that such a forecast will also give the best subsequent analysis and so on.

In the fully nonlinear case $Df(\boldsymbol{x}_B) \neq Df(\boldsymbol{x}_T)$, the best analysis may not give the best forecast, as illustrated schematically in Fig. 3b. In this more general situation, the optimal coupling scheme for synchronization may differ from that used in standard data assimilation methods. As a thought experiment, imagine a dynamical system that suddenly switches between two different dynamical regimes, given by different sets of equations, e.g. a 3-variable system that switches between Lorenz and Rossler dynamics, at regular intervals. If the period between switches is long enough, a Kalman filter will take a little time to adjust the coupling to the new dynamics after each switch. This is a sub-optimal arrangement. The optimal arrangement would switch the coupling scheme completely each time the dynamics switches.

A version of this thought experiment, realized numerically, is depicted in Fig. 4, using a non-autonomous three-variable system that switches, at periodic intervals Δ, between Lorenz dynamics and the same dynamics with the roles of X and Z reversed:

$$
\left.
\begin{aligned}
\dot{X} &= \sigma(Y - X) \\
\dot{Y} &= \rho X - Y - XZ \\
\dot{Z} &= -\beta Z + XY
\end{aligned}
\right\}
$$
if $n\Delta \leq t < (n + 1/2)\Delta$ $\qquad (n \in Z^+)$

$$(30)$$

$$
\left.
\begin{aligned}
\dot{X} &= -\beta X + ZY \\
\dot{Y} &= \rho Z - Y - ZX \\
\dot{Z} &= \sigma(Y - Z)
\end{aligned}
\right\}
$$
if $(n + 1/2)\Delta \leq t < (n + 1)\Delta$

Synchronization of a master-slave pair of such systems was effected using coupling in the X and Y variables, and alternately in the Z and Y variables, combinations known to be effective for the Lorenz system and the "reversed Lorenz" system, respectively:

$$
\left.
\begin{aligned}
\dot{X}_1 &= \sigma(Y_1 - X_1) + k(X - X_1) \\
\dot{Y}_1 &= \rho X_1 - Y_1 - X_1 Z_1 + k(Y - Y_1) \\
\dot{Z}_1 &= -\beta Z_1 + X_1 Y_1
\end{aligned}
\right\}
$$
if $n\Delta \leq t < (n + 1/2)\Delta$

$$(31)$$

$$
\left.
\begin{aligned}
\dot{X}_1 &= -\beta X_1 + Z_1 Y_1 \\
\dot{Y}_1 &= \rho Z_1 - Y_1 - Z_1 X_1 + k(Y - Y_1) \\
\dot{Z}_1 &= \sigma(Y_1 - Z_1) + k(Z - Z_1)
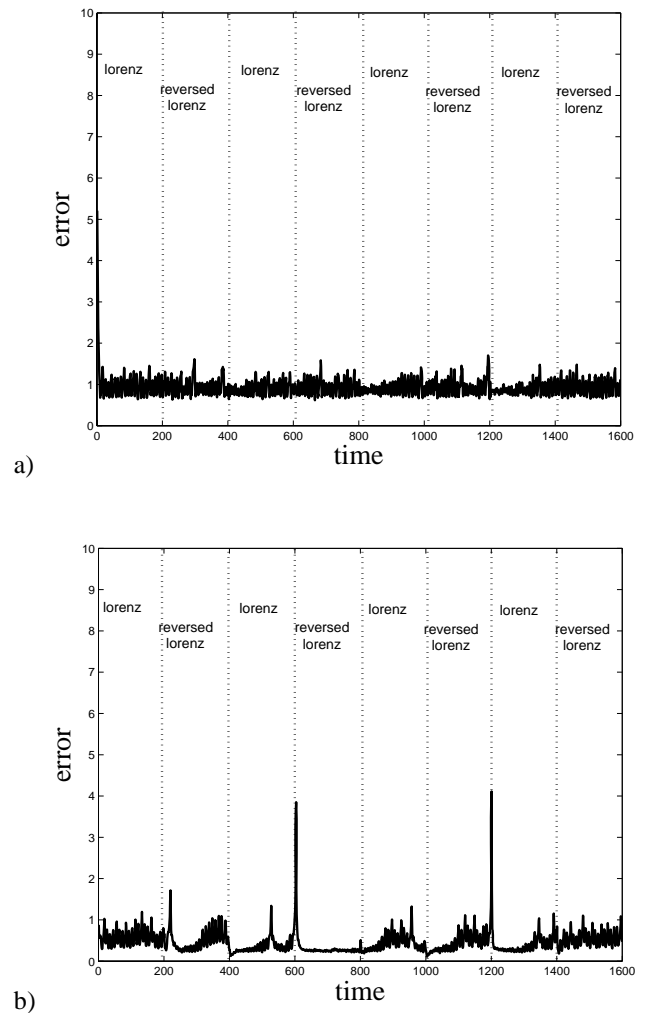\end{aligned}
\right\}
$$

a)

b)

Fig. 4. Synchronization error between two copies of the alternating Lorenz system (30), with standard Lorenz parameters and period $\Delta = 400$, coupled in two ways: **(a)** by simply alternating xy-coupling and zy-coupling (31), with coupling strength $k=140$ and **(b)** by Kalman filtering with algorithm and parameters as in (Yang et al., 2004), except with low values of the covariance inflation factor $1+\delta=1.05$ and of the standard deviation $\mu=0.025$ of the random perturbations added to the analysis error covariance. The average error for 100 realizations of the stochastic process, all with the same initial conditions, is plotted in each panel. Runge-Kutta numerical integration was performed with stepsize $\Delta t=0.01$. Time is shown in units equal to .08 in the nondimensional units of Eq. (31). Observations were assimilated every eight time steps in the Kalman filter algorithm; the coupling term in Eq. (31) was analogously turned on only every eight time steps. The Kalman filtering algorithm gives lower error on the average, but there are some desynchronization bursts at times of regime transitions (dotted lines) between the Lorenz and "reversed Lorenz" phases.

if $(n + 1/2)\Delta \leq t < (n + 1)\Delta$

The alternate coupling method (31) is compared with the Kalman filter algorithm that directs the slave system, defined

by X_1, Y_1, Z_1, based on "observations" of the X, Y, Z system. It is seen that the Kalman filter approach (Fig. 4b) gives bursts of desynchronization just after the transitions, as expected, unlike the coupling (31) (Fig. 4a), although the Kalman filter performance is better on the average. (The plots in Fig. 4 are averages over a large number of realizations of the stochastic process. Each realization is based on exactly the same master system trajectory. The desynchronization phenomenon occurs when the Lorenz/reversed Lorenz transition takes place at certain points on the master system attractor, but not when it takes place at other points.)

There are several lessons to be learned from this extreme and somewhat artificial example. First, the optimal synchronization scheme, with time-varying coupling, would reduce both the average error and the bursting phenomenon. That would be the ideal way to do data assimilation. But second, the Kalman filtering approach is almost always better than any of the coupling algorithms described in the synchronization literature. Specifically, the bursts (corresponding to "catastrophes" in the data assimilation literature) are short-lived even in this extreme example, lasting for only one or two time steps.

In the geophysical realm, effective non-autonomy is common, as with phenomena influenced by the diurnal or annual cycles for example. More generally, the highly nonlinear regions of phase space where the assumption (16) fails, and the optimality of the Kalman filter is expected to break down correspond to regime transitions. It is in such regions, that typically occupy small volumes of phase space, where the synchronization approach is expected to improve on standard methods to a small degree.

4 Synchronization vs. data assimilation for strongly nonlinear dynamics

In a region of state space where nonlinearities are strong and Eq. (16) fails, the prognostic equation for error (17) must be extended to incorporate nonlinearities. Additionally, model error due to processes on small scales that escape the digital representation should be considered. While errors in the parameters or the equations for the explicit degrees of freedom require deterministic corrections, the unresolved scales, assumed dynamically independent, can only be represented stochastically. The physical system is governed by:

$$\dot{x}_T = f(x_T) - \xi_M \tag{32}$$

in place of Eq. (15a), where ξ_M is model error, with covariance $Q \equiv <\xi_M \xi_M^T>$. The error equation (17) becomes

$$\dot{e} = (F - C)e + Ge^2 + He^3 + C\xi + \xi_M \tag{33}$$

where we have included terms up to cubic order in e, with $H < 0$ to prevent divergent error growth for large $||e||$. In the multi-dimensional case, Eq. (33) is shorthand for a tensor equation in which G and H are tensors of rank three and

rank four (and the restrictions on H are more complex). In the one-dimensional case, which we shall analyze here, G and H are scalars.

The Fokker-Planck equation is now:

$$\frac{\partial \rho}{\partial t} + \nabla_e \cdot \{\rho[(F - C)e + Ge^2 + He^3]\}$$
$$= \frac{1}{2}\delta \nabla_e \cdot [(CRC^T + Q)\nabla_e \rho] \tag{34}$$

Using the ansatz for the PDF ρ:

$$\rho(e) = N \exp(-Ke^2 - Le^3 - Me^4) \tag{35}$$

with a normalization factor $N = [\int_{-\infty}^{\infty} de \ \exp(-Ke^2 - Le^3 - Me^4)]^{-1}$, we obtain from Eq. (34) the following relations between the dynamical parameters and the PDF parameters:

$$F - C = \frac{1}{2}\tau(C^2R + Q)(-2K)$$
$$G = \frac{1}{2}\tau(C^2R + Q)(-3L) \tag{36}$$
$$H = \frac{1}{2}\tau(C^2R + Q)(-4M)$$

The goal is to minimize the background error:

$$B(K, L, M) = \frac{\int_{-\infty}^{\infty} de \ e^2 \exp(-Ke^2 - Le^3 - Me^4)}{\int_{-\infty}^{\infty} de \ \exp(-Ke^2 - Le^3 - Me^4)}. \tag{37}$$

Using Eq. (36) to express the arguments of B in terms of the dynamical parameters, we find $B(K, L, M) = B(K(C), L(C), M(C)) \equiv B(C)$, and can seek the value of C that minimizes B, for fixed dynamical parameters F, G, H.

For grounding in choosing appropriate parameter values, one might consider the nonlinearities of typical systems in geophysical fluid dynamics. The parameters G and H can be taken from a Taylor expansion of the model system dynamics about the true state. That is, the expansion $f(y) = f(x) + ef'(x) + \frac{1}{2}e^2 f''(x) + \frac{1}{6}e^3 f'''(x)$, for $x = x_T$, $y = x_B$, implies $G = \frac{1}{2}f''(x)$ and $H = \frac{1}{6}f'''(x)$, from which (33) follows, noting (15). But then the prognostic equation for the true system can be similarly expanded about $x = x_o$:

$$\dot{x} = f(x_o) + (x - x_o)F + (x - x_o)^2 G + (x - x_o)^3 H \tag{38}$$

Suppose the true system (38) describes motion in a double-well potential with the central fixed point at $x = x_o$, so that $f(x_o) = 0$. It is at such points, e.g. the central fixed point in the Lorenz system, that nonlinearities are usually greatest. The Lorenz (1984) system, intended to represent large-scale atmospheric circulation using just a few variables has similar structure: there are two regimes with central unstable fixed points which are about equidistant from the unstable fixed point at the center of the whole attractor that marks the boundary between the two regimes.
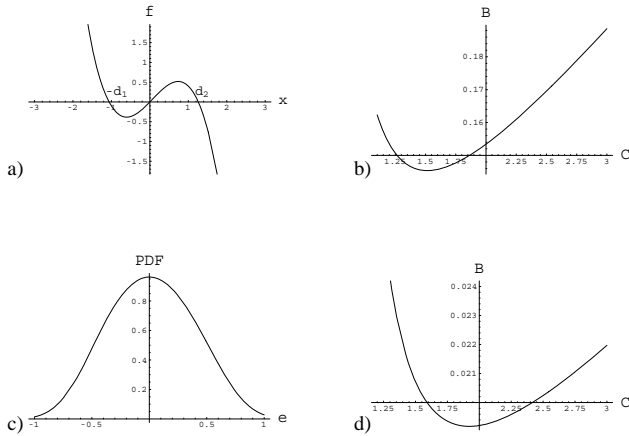
Fig. 5. For a true system $\dot{x}=f(x)$, with f defined by (38) describing a two-well potential about $x=x_o=0$, **(a)**, background error B is plotted against coupling strength C for bimodality parameters $d_1=1.06$, $d_2=1.26$, observation error $R=1$ and inter-observation time $\tau=0.1$ **(b)**. Although the distribution of true states is bimodal, the error distribution $\rho(e)$ is unimodal **(c)**. B vs. C for a shorter inter-observation time $\tau=0.01$ **(d)** gives a covariance inflation factor near unity (see text).

Table 1. Covariance inflation factor vs. bimodality parameters d_1, d_2, for 50% model error in the resolved tendency.

				d_1			
		.75	1.	1.25	1.5	1.75	2.
	.75	1.26	1.26	1.28	1.30	1.32	1.34
	1.	1.26	1.23	1.23	1.25	1.27	1.29
d_2	1.25	1.28	1.23	1.22	1.23	1.24	1.25
	1.5	1.30	1.25	1.23	1.22	1.23	1.24
	1.75	1.32	1.27	1.24	1.23	1.23	1.23
	2.	1.34	1.29	1.25	1.24	1.23	1.23

Matching the gross structure of the two-well potential to the gross structure of the Lorenz '63 or Lorenz '84 systems, one can find values of G and H, that will give the correct distances between the central fixed point and the two outlying fixed points, respectively. For the Lorenz '84 system, these distances are $d_1=1.06$ and $d_2=1.26$ respectively. The prognostic equation for error (33) has fixed points in the desired positions iff $G=.15$ and $H=-.75$, in a re-scaled time coordinate for which $F=1$.

For the perfect model case, $Q=0$, the function $B(C)$ is then found numerically to have the form plotted in Fig. 5b, with a distinct optimum at B=0.15, C=1.5, where the inter-observation time $\tau=0.1$ (6 h in dimensional units) is chosen to be an order of magnitude smaller than the dynamical time scale F^{-1} and we assume a large observation error $R=1$, of the same order as the typical displacement of x about the assumed unstable fixed point. Note that although the true state of the system is distributed bimodally, the distribution (35) of synchronization error $\rho(e)$ shown in Fig. 5c is unimodal, because of smearing by observational noise.

The coupling that gives optimal synchronization can again be compared with the coupling used in standard data assimilation, as for the linear case. In particular, one can ask whether the "covariance inflation" scheme that is used as an ad hoc adjustment in Kalman filtering (Anderson 2001) can reproduce the C values found to be optimal for synchronization. The form $C=\tau^{-1}B(B+R)^{-1}$ is replaced by the adjusted form

$$C = \frac{1}{\tau}\frac{\mathcal{F}B}{\mathcal{F}B+R} \qquad (39)$$

where $\mathcal{F}$ is the covariance inflation factor. For the example depicted in Fig. 5b, the optimal value $C=1.5$ would be generated by an inflation factor $\mathcal{F}=1.2$.

The optimization problem was solved numerically with results as plotted in Table 1 for a range of values of the bimodality parameters d_1 and d_2, giving dynamical parameters $G=1/d_2-1/d_1$ and $H=-1/(d_1d_2)$. Results are displayed for the case where the amplitude of model error in Eq. (32) is about 50% of the resolved tendency $\dot{x}_T$, with the resulting model error covariance $Q=0.04$ approximately one-fourth of the background error covariance B. The covariance inflation factors are remarkably constant over a wide range of parameters and agree with typical values used in operational practice.

For the range of bimodality parameters considered, the large stochastic model error makes little difference in the estimated covariance inflation factors, typically changing $\mathcal{F}$ only by about ±0.001. Indeed, for the linear case ($d_1=d_2=\infty$), the optimal coupling is still $1/\delta BR^{-1}$, as for a perfect model. The stochastic model error results in an extra constant term δQ on the RHS of the Fluctuation-Dissipation Relation (20) that does not affect the derivatives in the subsequent optimization procedure. (However, Q does enter the prognostic equation (13) for B in the linear case, just as in Kalman filtering.)

As the inter-observation time τ becomes smaller, B decreases at the minimum point, and the form (39) implies a decreasing covariance inflation factor. For the case $\tau=0.01$ shown in Fig. 5d, the required factor is near unity. (Precisely, $\mathcal{F}=1.02$.) That is, the coupling required for continuous-time Kalman filtering approaches the optimal coupling for synchronization. That the advantage conferred by the synchronization approach obtains only for sizable inter-observation times suggests some commonality with non-linear generalizations of Kalman filtering (e.g. Miller and Ghil, 1994). The difference between optimal synchronization and nonlinear Kalman filtering, while likely to be small, merits further investigation, since the two approaches are defined by different goals. Similarly, the use of ensembles to represent non-Gaussian PDF's in an empirical way (Anderson, 2003)

would seem to address some of the same issues treated in this section, and a comparison is warranted.

In the synchronization view, the values typically used for the covariance inflation factor in the presence of model nonlinearity and stochastic model error are readily explained. Sampling error, as it affects estimates of background covariance B, is not treated here, but might be expected to enter in a similar way and not to substantially alter the conclusions of the optimal coupling analysis.

5 Concluding remarks: data assimilation as machine perception

That ad hoc covariance inflation factors used in operational practice can be explained naturally in the synchronization view of data assimilation suggests a deep relevance for that viewpoint. Nothing in the foregoing discussion of synchronization and data assimilation is limited to meteorological processes. In any situation in which a computational predictive model of a physical process receives a stream of new data as it is running, the synchronization of the physical process and the model is the true goal of any assimilation scheme. As suggested in the introduction, the philosophical basis for the proposed use of synchronization appears to be the idea of synchronicity as espoused by Jung and Pauli, originally in a psychological context.

While a weather-prediction model is not usually viewed as artificially intelligent software, it forms an internal representation of the external world as complex as that of any robot, albeit without the motor component. One could envision augmenting the statistical data assimilation algorithms with rules to discriminate between good and bad observations, and with rules to transform the observed data in complex ways based on the current model state. Neurobiological paradigms may guide the design of such artificially intelligent predictive systems. It is known that interneuronal synchronization across wide distances in the brain plays a role in the grouping of percepts that is a prerequisite to higher processing, and may even underlie consciousness (Strogatz, 2003; Von Der Malsburg and Schneider, 1986). These findings lend credibility to the suggestion that a synchronization principle is also fundamental to the relationship between the brain and the external world, and that synchronization should be a cornerstone in the design of a neuromorphic, artificially intelligent predictive model.

Machine learning might also be realized in the synchronization context, so as to correct for deterministic model error in the resolved degrees of freedom. By allowing model parameters to vary slowly, generalized synchronization would be transformed to more nearly identical synchronization. Indeed, parameter adaptation laws can be added to a synchronously coupled pair of systems so as to synchronize the parameters as well as the states. Parlitz (1996) showed for example that two unidirectionally coupled Lorenz systems

with different parameters:

$$
\begin{aligned}
\dot{X} &= \sigma(Y - X) \\
\dot{Y} &= \rho X - Y - XZ \\
\dot{Z} &= -\beta Z + XY
\end{aligned}
$$

$$(40)$$

$$
\begin{aligned}
\dot{X}_1 &= \sigma(Y - X_1) \\
\dot{Y}_1 &= \rho_1 X_1 - \nu Y_1 - X_1 Z_1 + \mu \\
\dot{Z}_1 &= -\beta Z_1 + X_1 Y_1
\end{aligned}
$$

could be augmented with parameter adaptation rules:

$$
\begin{aligned}
\dot{\rho}_1 &= (Y - Y_1)X_1 \\
\dot{\nu} &= (Y_1 - Y)Y_1 \\
\dot{\mu} &= Y - Y_1
\end{aligned}
$$

$$(41)$$

so that the Lorenz systems would synchronize, and additionally $\rho_1 \to \rho$, $\nu \to 1$, and $\mu \to 0$. Note that parameters cease to adapt when the systems are perfectly synchronized with $Y - Y_1 = 0$. Generalizations to PDEs would allow model parameters to automatically adapt. In complex cases, a stochastic component (in the parameters) might be necessary to allow parameters to jump among multiple basins of attraction, most of which are sub-optimal. The stochastic approach could perhaps be extended to a genetic algorithm that would make random qualitative changes in the model as well, until synchronization is achieved.

The main competing approach to the tracking of reality by a predictive model is Kalman filtering, or generalizations thereof that use Bayesian reasoning to estimate the current state. Further development of the optimal synchronization approaches to provide more refined modifications of the Kalman filter in select regions of state space will be of interest in any situation where a Kalman filter is used to track a highly nonlinear process.

Conversely, unidirectional synchronization can always be viewed as a data assimilation problem, by regarding the slave system as a "model" of the master. The synchronization properties of a bidirectionally coupled system can often be inferred from the study of a corresponding unidirectional configuration. The optimally modified Kalman filter that is needed for data assimilation will therefore also be useful for optimizing the synchronization of dynamical systems generally.

Appendix A

Derivation of the fluctuation-dissipation relation for synchronously coupled differential equations with noise in the coupling channel

Consider the stochastic differential equation for synchronization error (17), rewritten as:

$$
\frac{de}{dt} = (\mathbf{F} - \mathbf{C})e + \mathbf{C}\boldsymbol{\xi}
$$

$$(A1)$$

where e is the synchronization error vector, $\mathbf{F}$ is a matrix representing the linearized dynamics, $\mathbf{C}$ is the coupling matrix, and $\boldsymbol{\xi}$ is a time-dependent vector of white noise process satisfying $<\boldsymbol{\xi}(t)>=0$ and

$$< \boldsymbol{\xi}(t)\boldsymbol{\xi}^T(t') > = \delta\mathbf{R}\delta(t-t') \tag{A2}$$

where $\mathbf{R}$ is the observation error covariance matrix and δ is the time over which the physical noise decorrelates. The solution to Eq. (A1) is:

$$\boldsymbol{e}(t) = \mathrm{e}^{(\mathbf{F}-\mathbf{C})t}\boldsymbol{e}(0) + \int_0^t dt' \mathrm{e}^{(\mathbf{F}-\mathbf{C})(t-t')}\mathbf{C}\boldsymbol{\xi}(t') \tag{A3}$$

Thus the mean synchronization error is

$$< \boldsymbol{e}(t) > = \mathrm{e}^{(\mathbf{F}-\mathbf{C})t}\boldsymbol{e}(0) \tag{A4}$$

and the synchronization error variance is

$$< [\boldsymbol{e}(t) - \mathrm{e}^{(\mathbf{F}-\mathbf{C})t}\boldsymbol{e}(0)][\boldsymbol{e}(t) - \mathrm{e}^{(\mathbf{F}-\mathbf{C})t}\boldsymbol{e}(0)]^T > =$$
$$\int_0^t dt' \int_0^t dt'' \mathrm{e}^{(\mathbf{F}-\mathbf{C})(t-t')}\mathbf{C}<\boldsymbol{\xi}(t')\boldsymbol{\xi}^T(t'')>\mathbf{C}^T \mathrm{e}^{(\mathbf{F}-\mathbf{C})^T(t-t'')} \tag{A5}$$

or

$$< \boldsymbol{e}(t)\boldsymbol{e}^T(t) > = \mathrm{e}^{(\mathbf{F}-\mathbf{C})t}\boldsymbol{e}(0)\boldsymbol{e}^T(0)\mathrm{e}^{(\mathbf{F}-\mathbf{C})^Tt}$$
$$+ \int_0^t dt' \mathrm{e}^{(\mathbf{F}-\mathbf{C})(t-t')}\delta\mathbf{C}\mathbf{R}\mathbf{C}^T \mathrm{e}^{(\mathbf{F}-\mathbf{C})^T(t-t')} \tag{A6}$$

If $\mathbf{C}$ is chosen so that the system synchronizes, in the absence of noise, as $t\to\infty$, then the first term on the right hand sided of Eq. (A6) vanishes in this limit. The system with noise approaches a stationary state with $<\boldsymbol{e}>=0$ and

$$< \boldsymbol{e}\boldsymbol{e}^T > \equiv \mathbf{B} = \int_0^\infty dt \mathrm{e}^{(\mathbf{F}-\mathbf{C})t}\delta\mathbf{C}\mathbf{R}\mathbf{C}^T \mathrm{e}^{(\mathbf{F}-\mathbf{C})^Tt} \tag{A7}$$

Differentiating the integrand in Eq. (A7) with respect to t and and using Eq. (A7) to simplify the resulting expression, we find

$$(\mathbf{F}-\mathbf{C})\mathbf{B} + \mathbf{B}(\mathbf{F}-\mathbf{C})^T$$
$$= \int_0^\infty dt \frac{d}{dt}[\mathrm{e}^{(\mathbf{F}-\mathbf{C})t}\delta\mathbf{C}\mathbf{R}\mathbf{C}^T \mathrm{e}^{(\mathbf{F}-\mathbf{C})^Tt}]$$
$$= [\mathrm{e}^{(\mathbf{F}-\mathbf{C})t}\delta\mathbf{C}\mathbf{R}\mathbf{C}^T \mathrm{e}^{(\mathbf{F}-\mathbf{C})^Tt}]\Big|_0^\infty \tag{A8}$$

The last expression in brackets vanishes at the upper limit for the case of stable synchronization, so we have

$$(\mathbf{F}-\mathbf{C})\mathbf{B} + \mathbf{B}(\mathbf{F}-\mathbf{C})^T = -\delta\mathbf{C}\mathbf{R}\mathbf{C}^T \tag{A9}$$

which is the fluctuation-dissipation relation (20).

Appendix B

Optimal coupling for synchronization of discrete-time maps

In Sects. 3.1 and 3.2, standard data assimilation was compared to optimal synchronization of differential equations by considering the continuous-time limit of the discrete analysis cycle. Instead, one can leave the analysis cycle intact and compare it to a discrete-time version of optimal synchronization, i.e. to optimally synchronized maps.

We begin by deriving a fluctuation-dissipation relation (FDR) for stochastic difference equations. Consider the stochastic difference equation with additive noise,

$$\boldsymbol{x}(n+1) = \mathbf{F}\,\boldsymbol{x}(n) + \boldsymbol{\xi}(n) \quad < \boldsymbol{\xi}(n)\boldsymbol{\xi}(m)^T >= \mathbf{R}\,\delta_{n,m}, \tag{B1}$$

where $\boldsymbol{x}, \boldsymbol{\xi} \in R^n$, $\mathbf{F}, \mathbf{R}$ are $n\times n$ matrices, $\mathbf{F}$ is assumed to be stable, and $\boldsymbol{\xi}$ is Gaussian white noise. One can prove by induction that the solution to this equation, with initial condition $\boldsymbol{x}(p)$, is

$$\boldsymbol{x}(n+1) = \mathbf{F}^{n+1-p}\boldsymbol{x}(p) + \sum_{m=p}^n \mathbf{F}^{m-p}\boldsymbol{\xi}(n+p-m) \tag{B2}$$

We first wish to find the equilibrium covariance matrix $\boldsymbol{\Gamma}=<\boldsymbol{x}\boldsymbol{x}^T>$. If the initial condition is in the infinite past then the equilibrium covariance is the covariance at any finite iteration and it is convenient to choose iteration one. Since $\mathbf{F}$ is stable the initial condition is forgotten and we obtain:

$$\boldsymbol{x}(1) = \sum_{m=0}^\infty \mathbf{F}^m \boldsymbol{\xi}(-m), \tag{B3}$$

and then

$$\boldsymbol{\Gamma} = \sum_{m=0}^\infty \mathbf{F}^m \mathbf{R}\mathbf{F}^{m^T}. \tag{B4}$$

One can then show that $\boldsymbol{\Gamma}$ satisfies the matrix FDR

$$\mathbf{F}\,\boldsymbol{\Gamma}\,\mathbf{F}^T - \boldsymbol{\Gamma} + \mathbf{R} = 0. \tag{B5}$$

Now consider a model that takes the analysis at step n to a new background at step $n+1$, given by a linear matrix $\mathbf{M}$. That is, $\boldsymbol{x}_B(n+1)=\mathbf{M}\boldsymbol{x}_A(n)$. Also, $\boldsymbol{x}_T(n+1)=\mathbf{M}\boldsymbol{x}_T(n)$. Since $\boldsymbol{x}_A(n)=\boldsymbol{x}_B(n)+\mathbf{B}(\mathbf{B}+\mathbf{R})^{-1}(\boldsymbol{x}_{\mathrm{obs}}(n)-\boldsymbol{x}_B(n))$, where $\boldsymbol{x}_{\mathrm{obs}}=\boldsymbol{x}_T+\boldsymbol{\xi}$, we derive a difference equation for $\boldsymbol{e}\equiv\boldsymbol{x}_B-\boldsymbol{x}_T$:

$$\boldsymbol{e}(n+1)=\mathbf{M}(\mathbf{I}-\mathbf{B}(\mathbf{B}+\mathbf{R})^{-1})\boldsymbol{e}(n)+\mathbf{M}\mathbf{B}(\mathbf{B}+\mathbf{R})^{-1}\boldsymbol{\xi}. \tag{B6}$$

For synchronously coupled maps, on the other hand, we have

$$\boldsymbol{e}(n+1) = (\mathbf{M}-\mathbf{C})\boldsymbol{e}(n+1) + \mathbf{C}\boldsymbol{\xi}, \tag{B7}$$

and with the FDR as derived above:

$$(\mathbf{M}-\mathbf{C})\mathbf{B}(\mathbf{M}-\mathbf{C})^T - \mathbf{B} + \mathbf{C}\mathbf{R}\mathbf{C}^T = 0 \tag{B8}$$

Differentiating the matrix equation (B8) with respect to the elements of $\mathbf{C}$, as in the continuous-time analysis, we find

$$0 = (\mathbf{M} - \mathbf{C})d\mathbf{B}(\mathbf{M} - \mathbf{C})^T + (-d\mathbf{C})\mathbf{B}(\mathbf{M} - \mathbf{C})^T$$
$$+ (\mathbf{M} - \mathbf{C})\mathbf{B}(-d\mathbf{C})^T - d\mathbf{B} + d\mathbf{C}\mathbf{R}\mathbf{C}^T + \mathbf{C}\mathbf{R}d\mathbf{C}^T . \quad \text{(B9)}$$

We seek a matrix $\mathbf{C}$ for which $d\mathbf{B}{=}0$ for arbitrary $d\mathbf{C}$, and thus

$$(-d\mathbf{C})[\mathbf{B}(\mathbf{M} - \mathbf{C})^T - \mathbf{R}\mathbf{C}^T]$$
$$+ [(\mathbf{M} - \mathbf{C})\mathbf{B} - \mathbf{C}\mathbf{R}](-d\mathbf{C})^T = 0 \quad \text{(B10)}$$

for arbitrary $d\mathbf{C}$. The two terms are transposes of one another, and it is easily shown, as in the continuous-time case, that the quantities in brackets must vanish. This gives the optimal matrix

$$\mathbf{C} = \mathbf{M}\mathbf{B}(\mathbf{B} + \mathbf{R})^{-1} \quad \text{(B11)}$$

which upon substitution in Eq. (B7) reproduces the standard data assimilation form (B6), confirming the equivalence.

References

Afraimovich, V. S., Verichev, N. N., and Rabinovich, M. I.: Stochastic synchronization of oscillation in dissipative systems, Radiophys. Quantum Electron., 29, 795–803, 1986.

Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, Mon. Wea. Rev., 129, 2884–2903, 2001.

Anderson, J. L.: A local least-squares framework for ensemble filtering, Mon. Wea. Rev., 131, 634–642, 2003.

Corazza, M., Kalnay, E., Patil, D. J., Yang, S.-C., Morss, R., Cai, M., Szunyogh, I., Hunt, B. R., and Yorke, J. A.: Use of the breeding technique to estimate the structure of the analysis "errors of the day", Nonlin. Processes Geophys., 10, 233–243, 2003, http://www.nonlin-processes-geophys.net/10/233/2003/.

Ashwin, P., Buescu, J., and Stewart, I.: Bubbling of attractors and synchronisation of chaotic oscillators, Phys. Lett. A, 193, 126–139, 1994.

Baek, S. J., Hunt, B. R., Szunyogh, I., Zimin, A., and Ott, E.: Localized error bursts in estimating the state of spatiotemporal chaos, Chaos, 14, 1042–1049, 2004.

Corazza, M., Kalnay, E., Patil, D. J., Yang, S.-C., Morss, R., Cai, M., Szunyogh, I., Hunt, B. R., and Yorke, J. A.: Use of the breeding technique to estimate the structure of the analysis "errors of the day", Nonlin. Processes Geophys., 10, 233–243, 2003, http://www.nonlin-processes-geophys.net/10/233/2003/.

Duane, G. S.: Synchronized chaos in extended systems and meteorological teleconnections, Phys. Rev. E, 56, 6475–6493, 1997.

Duane, G. S.: Synchronized chaos in climate dynamics, in: Proc. 7th Experimental Chaos Conference, edited by: In, V., Kocarev, L., Carroll, T. L., et al., AIP Conference Proceedings 676, Melville, New York, 115–126, 2003.

Duane, G. S. and Tribbia, J. J.: Synchronized chaos in geophysical fluid dynamics, Phys. Rev. Lett., 86, 4298–4301, 2001.

Duane, G. S. and Tribbia, J. J.: Weak Atlantic-Pacific teleconnections as synchronized chaos, J. Atmos. Sci., 61, 2149–2168, 2004.

Evensen, G.: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10 143–10 162, 1994.

Fujisaka, H. and Yamada, T.: Stability theory of synchronized motion in coupled-oscillator systems, Prog. Theor. Phys., 69, 32–47, 1983.

Jung, C. G. and Pauli, W.: The interpretation of nature and the psyche, Pantheon, New York, 1955.

Kocarev, L., Tasev, Z., and Parlitz, U.: Synchronizing spatiotemporal chaos of partial differential equations, Phys. Rev. Lett., 79, 51–54, 1997.

Lorenz, E. N.: Deterministic nonperiodic flows, J. Atmos. Sci., 20, 130–141, 1963.

Lorenz, E. N.: Irregularity – a fundamental property of the atmosphere, Tellus A, 36, 98–110, 1984.

Miller, R. N. and Ghil, M.: Advanced data assimilation in strongly nonlinear dynamical systems, J. Atmos. Sci., 51, 1037–1056, 1994.

Parlitz, U.: Estimating model parameters from time series by autosynchronization, Phys. Rev. Lett., 76, 1232–1235, 1996.

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A., and Ott, E.: Local low dimensionality of atmospheric dynamics, Phys. Rev. Lett., 86, 5878–5881, 2001.

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

Pecora, L. M., Carroll, T. L., Johnson, G. A., Mar, D. J., and Heagy, J. F.: Fundamentals of synchronization in chaotic systems, concepts, and applications, Chaos, 7, 520–543, 1997.

Rulkov, N. F., Sushchik, M. M., and Tsimring, L. S.: Generalized synchronization of chaos in directionally coupled chaotic systems, Phys. Rev. E, 51, 980–994, 1995.

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

Strogatz, S. H.: Sync: The Emerging Science of Spontaneous Order, Theia, New York, 338 pp., 2003.

Toth, Z. and Kalnay, E.: Ensemble forecasting at NMC – the generation of perturbations, Bull. Amer. Meteor. Soc., 74, 2317–2330, 1993.

Toth, Z. and Kalnay, E.: Ensemble forecasting at NCEP and the breeding method, Mon. Wea. Rev., 125, 3297–3319, 1997.

Vautard, R., Legras, B., and Déqué, M.: On the source of midlatitude low-frequency variability. Part I: A statistical approach to persistence, J. Atmos. Sci., 45, 2811–2843, 1988.

Von Der Malsburg, C. and Schneider, W.: A neural cocktail-party processor, Biol. Cybernetics, 54, 29–40, 1986.

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