Automatic Parameter Estimation in a Mesoscale Model Without Ensembles

Gregory S. Duane and Joshua P. Hacker

National Center for Atmospheric Research, Boulder, CO gduane,hacker@ucar.edu

In numerical forecasting, unknown model parameters have been estimated from a time series of observations by regarding them as extra state variables, and applying standard data assimilation methods that use ensembles to represent background error. In many situations, however, the use of ensembles is prohibitively expensive and/or impracticable because of the inability to properly account for model error in the initialization scheme. If one is seeking to estimate model parameters as data is assimilated, it is possible to take advantage of the assumed relative constancy of such parameters over large regions of time and space to derive an estimate from a single realization. The approach follows from a general result on synchronously coupled dynamical systems, where one system here represents "truth" and the other "model": If two such systems can be made to synchronize when their corresponding parameters are identical, for any coupling scheme (such as might be used in conventional data assimilation) a parameter estimation law can generally be added that will dynamically reduce a total cost (Lyapunov) function including parameter mismatch terms as well as state mismatch terms.

The approach is used to estimate a parameter that quantifies the effect of soil moisture in a single-column version of the Weather Research and Forecasting (WRF) model. The scheme can be extended to infer a 2D map of soil parameter values for a 3D model, using the fact that the parameter is slowly varying almost everywhere. Discontinuities are represented as additional degrees of freedom, and the Lyapunov function is augmented so as to penalize for horizontal variations in the soil parameter value except at locations of such discontinuities. The constrained optimization approach that is proposed should be useful for a variety of parameter estimation problems in numerical weather prediction (NWP), and will extend the power of ensemble methods.

1 Introduction

Any scheme for meteorological data assimilation has the goal of synchronizing a computational model with the real climate system, based on a limited set of noisy observations. Where the purpose is prediction of the state of the real system in the not-too-distant future, this synchronization view contrasts *a priori* with the usual view, in which observations are combined with the current model state to form the best possible estimate of the *current* state of the real system.

In this paper, we extend previous work on synchronization-based data assimilation, summarized in the next two sections, to show that parameters can be readily synchronized, as well as states. The synchronization view lends itself particularly well to the estimation of slowly varying parameters, a point made with a simplified, single-column version of an actual weather prediction model in Section 4. It is argued that the synchronization view also lends itself to the estimation of local parameter values that are slowly varying in space almost everywhere. In the concluding section, it is argued that the approach to parameter estimation can be extended to a more general scheme for machine learning.

2 Background: Data Assimilation and Synchronized Chaos

The phenomenon of chaos synchronization [1] was first brought to light by Fujisaka and Yamada [2] and independently by Afraimovich et al. [3], but extensive research on the subject in the '90s was spurred by the seminal work of Pecora and Carroll [4], who considered two chaotic systems in a master-slave relationship defined by a shared subsystem. Pecora and Carroll considered configurations such as the following combination of Lorenz systems:

$$\begin{aligned} \dot{X} &= \sigma(Y - X) & & \\ \dot{Y} &= \rho X - Y - XZ & \dot{Y}_1 &= \rho X - Y_1 - XZ_1 \\ \dot{Z} &= -\beta Z + XY & \dot{Z}_1 &= -\beta Z_1 + XY_1 \end{aligned} \tag{1}$$

which synchronizes rapidly, slaving the Y_1, Z_1-subsystem to the master X, Y, Z-subsystem, as seen in Fig. 1, despites differing initial conditions and despite sensitive dependence on initial conditions.

If we imagine that the first Lorenz system represents the world, and that the second Lorenz system is a predictive model, then synchronization effects data assimilation of observed variables into the running model. The only observed variable in the foregoing example is X, but that is sufficient to cause the desired convergence of model to truth. Synchronization is known to be tolerant of reasonable levels of noise, as might arise in the observation channel, and occurs with partial coupling schemes that do not completely replace a model variable with a variable of the observed system.

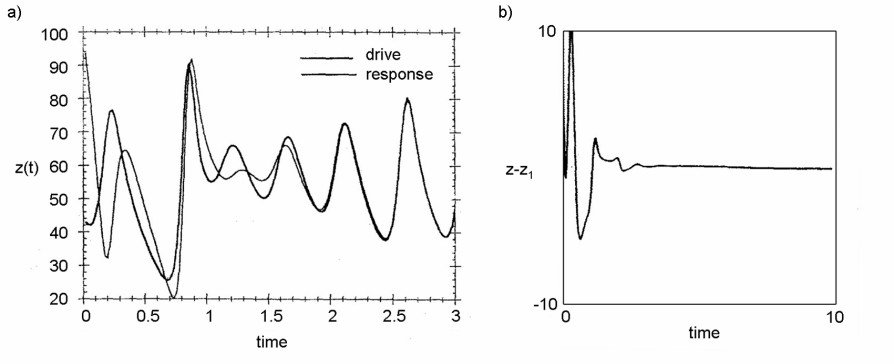

Fig. 1. The trajectories of the synchronously coupled Lorenz systems in the Pecora-Carroll complete replacement scheme (1) rapidly converge (a). Differences between corresponding variables approach zero (b).

Specifically, systems can also synchronize when coupled *diffusively*, as with a pair of directionally coupled Rossler systems:

$$\begin{aligned} \dot{X} &= -Y - Z + \alpha(X_1 - X) & \dot{X}_1 &= -Y_1 - Z_1 + \alpha(X - X_1) \\ \dot{Y} &= X + aY & \dot{Y}_1 &= X_1 + aY_1 \\ \dot{Z} &= b + Z(X - c) & \dot{Z}_1 &= b + Z_1(X_1 - c) \end{aligned} \tag{2}$$

where α parametrizes the coupling strength. The diffusive coupling scheme can be seen to resemble the "nudging" approach to data assimilation [5]. For judicious choice of nudging coefficient, it can be seen to resemble 3DVar, and for time-varying coefficient, Kalman filtering (as defined in e.g. [6]).

It is commonly not the existence, but the stability of the synchronization manifold that distinguishes coupled systems exhibiting synchronization from those that do not (such as (2) for different values of α). N Lyapunov exponents can be defined for perturbations in the N-dimensional space that is transverse to the synchronization manifold $\mathcal{M}$. If the largest of these, $h^{\perp}_{max}$, is negative, then motion in the synchronization manifold is stable against transverse perturbations. In that case, the coupled systems will synchronize for some range of differing initial conditions. However, since $h^{\perp}_{max}$ only determines *local* stability properties, the size of the basin of attraction for the synchronized regime remains unknown. As $h^{\perp}_{max}$ is increased through zero, the system undergoes a *blowout bifurcation*. For small positive values of $h^{\perp}_{max}$, on-off synchronization occurs (a special case of on-off intermittency), as illustrated in Fig. 2b, where degradation results from a time lag in the coupling. The other panels of Fig.2 show an increasing rate of bursting as the time-lag increases. Vestiges of synchronization are discernible even far from the blowout bifurcation point (Fig.2c), a phenomenon that was used to predict new teleconnection patterns [7].

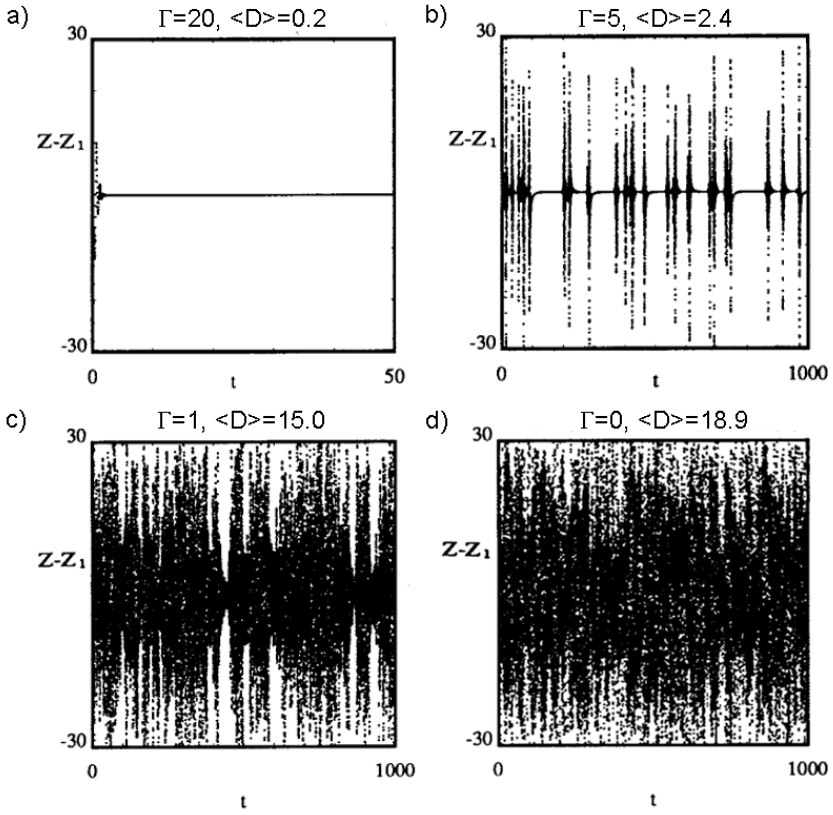

Fig. 2. The difference between the simultaneous states of two Lorenz systems with time-lagged coupling, represented by $Z(t) - Z_1(t)$ vs. t for various values of the inverse time-lag Γ illustrating complete synchronization (a), intermittent or "on-off" synchronization (b), partial synchronization (c), and de-coupled systems (d). Average euclidean distance $\langle D \rangle$ between the states of the two systems in X, Y, Z-space is also shown. The trajectories are generated by adaptive Runge-Kutta numerical integrations with $\sigma = 10.$, $\rho = 28.$, and $\beta = 8/3$.

The early work on synchronized chaos was spurred by an intended application to secure communications, since the signal connecting the two synchronized systems can be difficult to distinguish from noise. Practical applications to cryptography have not emerged, largely because system parameters can be extracted from the coupling signal, with some effort, for low-dimensional systems. But the essence of the phenomenon is that two systems that are effectively unpredictable, connected by a signal that may be almost indecipherable, can still exhibit significant correlations. It is argued that this phenomenon makes weather prediction possible, and will be more generally useful for real-time computational modeling of natural systems.

3 Background: Synchronization in Geophysical Fluid Dynamics

Synchronization in geophysical fluid models was demonstrated by Duane and Tribbia [8], originally with a view toward predicting and explaining new families of long-range teleconnections [9]. The uncoupled single-system model in this work was derived from one described by Vautard *et al.* [10].

The model is given by the quasigeostrophic equation for potential vorticity q in a two-layer reentrant channel on a β-plane:

$$\frac{Dq_i}{Dt} \equiv \frac{\partial q_i}{\partial t} + J(\psi_i, q_i) = F_i + D_i \tag{3}$$

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. Eq. (3) states that potential vorticity is conserved on a moving parcel, except for forcing F_i and dissipation D_i. The discretized potential vorticity is

$$q_i = \mathrm{f}_0 + \beta \mathrm{y} + \nabla^2 \psi_\mathrm{i} + \mathrm{R}_\mathrm{i}^{-2}(\psi_1 - \psi_2)(-1)^\mathrm{i} \tag{4}$$

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 $d\mathrm{f}/\mathrm{dy}$ 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: $F_i = f_0(q_i^* - q_i)$ for q_i^* corresponding to the choice of ψ^* shown in Figure 3a. The dissipation terms D_i, boundary conditions, and other parameter values are given in Ref. [9].

Two models of the form (3), $Dq^A/Dt = F^A + D^A$ and $Dq^B/Dt = F^B + D^B$ were coupled diffusively in one direction by modifying one of the forcing terms:

$$F_\mathbf{k}^B = f^B[a_\mathbf{k}(q_\mathbf{k}^* - q_\mathbf{k}^B) + b_\mathbf{k}(q_\mathbf{k}^A - q_\mathbf{k}^B)] \tag{5}$$

where the flow has been decomposed spectrally and the subscript $\mathbf{k}$ on each quantity indicates the wave number $\mathbf{k}$ spectral component. (The layer index i has been suppressed.) The two sets of coefficients $a_\mathbf{k}$ and $b_\mathbf{k}$ 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:

$$a_\mathbf{k} = \begin{cases} 0 & \text{if } |k_x| \le k_{x0} \text{ and } |k_y| \le k_{y0} \\ (k_n/|\mathbf{k}|)^4 & \text{if } |\mathbf{k}| > k_n \\ 1 - (k_0/|\mathbf{k}|)^4 & \text{otherwise} \end{cases}$$

$$b_\mathbf{k} = \begin{cases} 1 - a_\mathbf{k} & \text{if } |\mathbf{k}| \le k_n \\ 0 & \text{if } |\mathbf{k}| > k_n \end{cases}$$

as in Ref. [9], where the constants k_0, k_{x0}, k_{y0} and k_n are defined.

It was found that the two channels thus coupled rapidly synchronize (Fig. 3), starting from initial flow patterns that are arbitarily set equal to the forcing in one channel, and to a different pattern in the other channel. (Results

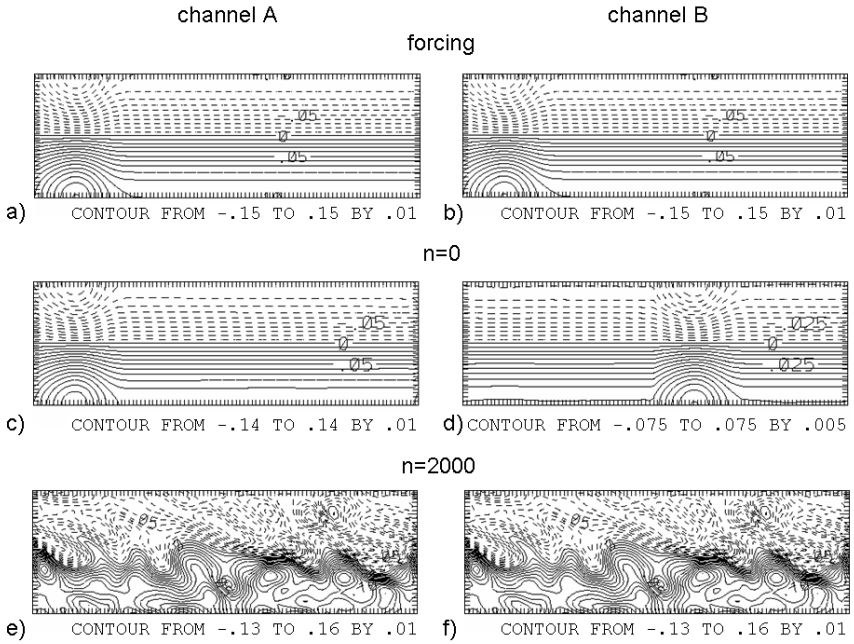

Fig. 3. Streamfunction (in units of $1.48 \times 10^9 m^2 s^{-1}$) describing the forcing ψ^* (a,b), and the evolving flow ψ (c-f), in a parallel channel model with bidirectional coupling of medium scale modes for which $|k_x| > k_{x0} = 3$ or $|k_y| > k_{y0} = 2$, and $|k| \leq 15$, for the indicated numbers n of time steps in a numerical integration. Parameters are as in Ref. [9]. 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.

are shown for bidirectional coupling defined by adding an equation for $F^A_{\mathbf{k}}$ analogous to (5). 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.

4 Parameter Adaptation in a Mesoscale 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 that is defined by a complex non-identical correspondence between variables in the two models 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

[11] showed for example that two unidirectionally coupled Lorenz systems with different parameters:

$$\begin{array}{ll} \dot{X} = \sigma(Y - X) & \dot{X}_1 = \sigma(Y - X_1) \\ \dot{Y} = \rho X - Y - XZ & \dot{Y}_1 = \rho_1 X_1 - \nu Y_1 - X_1 Z_1 \\ \dot{Z} = -\beta Z + XY & \dot{Z}_1 = -\beta Z_1 + X_1 Y_1 \end{array} \tag{6}$$

could be augmented with parameter adaptation rules:

$$\begin{aligned} \dot{\rho}_1 &= (Y - Y_1)X_1 \\ \dot{\nu} &= (Y_1 - Y)Y_1 \\ \dot{\mu} &= Y - Y_1 \end{aligned} \tag{7}$$

so that the Lorenz systems would synchronize, and additionally $\rho_1 \to \rho$, $\nu \to 1$, and $\mu \to 0$.

Equations for a synchronously coupled pair of systems can in fact always be augmented to allow parameter adaptation as well, provided that relevant dynamical variables are observed, as shown by Duane, Yu, and Kocarev [12]. Consider, for example, two quasigeostrophic channel models of the form (3) coupled according to (5), which are known to synchronize, as discussed in the background section. The model parameter to be estimated, f^B, is an overall coefficient in the forcing term

$$F^B_{\mathbf{k}} = f^B a_{\mathbf{k}}(q^*_{\mathbf{k}} - q^B_{\mathbf{k}}) + f^B b_{\mathbf{k}}(q^A_{\mathbf{k}} - q^B_{\mathbf{k}}) \tag{8}$$

where the layer index l is suppressed and the coefficients $a_{\mathbf{k}}, b_{\mathbf{k}}$ are slightly smoothed step functions of $\mathbf{k}$, as before, so that each spectral component is either coupled to the corresponding component in the A system or to the background flow q^* or neither. The coefficients are chosen as before so as to couple only the medium-scale components. (The forcing for the A system is correspondingly: $F^A_{\mathbf{k}} = f^A a_{\mathbf{k}}(q^*_{\mathbf{k}} - q^A_{\mathbf{k}})$.)

The parameter estimation rule in spectral space:

$$\dot{f}^B = \sum_{\mathbf{k} \in S} (q^A_{\mathbf{k}} - q^B_{\mathbf{k}})[a_{\mathbf{k}}(q^*_{\mathbf{k}} - q^B_{\mathbf{k}}) + b_{\mathbf{k}}(q^A_{\mathbf{k}} - q^B_{\mathbf{k}})] \tag{9}$$

for a restricted range of wavenumbers in S, as in the figure, causes f^B to converge to f^A, as shown in Fig. 4 .

The derivation of the rule (9) is instructive. One chooses the parameter adaptation rule so that a Lyapunov function that quantifies both state error and parameter error is monotonically decreasing, using the fact that a Lyapunov function for the identical-parameter situation is known to be monotonically decreasing, since the identical systems synchronize. The latter, "core" Lyapunov function, $L_o(q^A, q^B)|_{f^A = f^B} \equiv \int d^2x \ (q^A - q^B)^2$, is known to be decreasing at any point (q^A, q^B) in a large region of the coupled-system state space. Consider the more general Lyapunov function for the case of parameter mismatch:

$$L(q^A, q^B, f^A, f^B) \equiv (f^A - f^B)^2 + \int d^2x\ (q^A - q^B)^2 \tag{10}$$

and compute the time derivative

$$\dot{L} = -2\dot{f}^B(f^A - f^B) + \frac{d}{dt}\int d^2x\ (q^A - q^B)^2 \tag{11}$$

The second term on the right-hand side can be expanded using the dynamical equation (3), with the forcing term as defined in (8), to compute the extra contribution to $\dot{q}^B$ from the time-varying coefficient f^B

$$\begin{aligned}
\frac{d}{dt}\int d^2x\ (q^A - q^B)^2 &= 2\int d^2x\ (q^A - q^B)(\dot{q}^A - \dot{q}^B) \\
&= \dot{L}_o|_{f^B=f^A} + 2(f^A - f^B)\times \\
&\qquad \times \int d^2x\ (q^A - q^B)\sum_{\mathbf{k}} \left[a_{\mathbf{k}}(q^*_{\mathbf{k}} - q^B_{\mathbf{k}}) + b_{\mathbf{k}}(q^A_{\mathbf{k}} - q^B_{\mathbf{k}})\right] e^{i\mathbf{k}\cdot\mathbf{x}} \\
&= \dot{L}_o|_{f^B=f^A} + 2(f^A - f^B)\sum_{\mathbf{k}}(q^A_{\mathbf{k}} - q^B_{\mathbf{k}})\left[a_{\mathbf{k}}(q^*_{\mathbf{k}} - q^B_{\mathbf{k}}) + b_{\mathbf{k}}(q^A_{\mathbf{k}} - q^B_{\mathbf{k}})\right]
\end{aligned} \tag{12}$$

where the sum on the second line, multiplied by $f^A - f^B$, is the extra contribution to F in (3). From (11) and (12), it is seen that if we choose the adaptation law (9), with S universal, we will have

$$\dot{L} = \dot{L}_o|_{f^B=f^A} \leq 0 \tag{13}$$

as desired. The adaptation law for restricted S can be derived by using a different, correspondingly truncated Lyapunov function.

It is also instructive to consider the effect of a simpler parameter adaptation law. If one ignores the occurence of the parameter f^B in the coupling of the two channels and only retains the first term in the forcing (8), then the parameter adaptation law that guarantees (13) is:

$$\dot{f}^B = \int d^2x\ (q^* - q^B)(q^A - q^B) \tag{14}$$

Under the truncated adaptation rule (14) the monotonic convergence of f^B to the correct value f^A (Fig. 5a) is replaced by oscillatory convergence, as plotted in Fig. 5b. The robustness of the general approach to parameter estimation is apparent in this example.

The general approach that we have illustrated was formalized by Duane et al. [12]. Consider a "real system" given by ODE's:

$$\dot{\mathbf{x}} = \boldsymbol{f}(\mathbf{x}, \mathbf{p}), \tag{15}$$

$$\dot{\mathbf{p}} = 0, \tag{16}$$

where $\mathbf{x} \in I\!R^N$, $\boldsymbol{f} : I\!R^N \rightarrow I\!R^N$, and $\mathbf{p} \in I\!R^m$ is the vector of (unknown, constant) parameters of the system. We further assume that $\mathbf{s} = \mathbf{s}(\mathbf{x})$, where $\mathbf{s} : I\!R^N \rightarrow I\!R^n$, $n \leq N$, is an n dimensional vector representing the experimental measurement output of the system. A "computational model" of the system (15) is given by:

$$\dot{\mathbf{y}} = \boldsymbol{f}(\mathbf{y}, \mathbf{q}) + \upsilon(\mathbf{y}, \mathbf{s}), \tag{17}$$

$$\dot{\mathbf{q}} = \boldsymbol{N}(\mathbf{y}, \mathbf{x} - \mathbf{y}), \tag{18}$$

where $\boldsymbol{N}(\mathbf{y}, 0) = 0$, and υ is the control signal. Let $\mathbf{e} \equiv \mathbf{y} - \mathbf{x}$ and $\mathbf{r} \equiv \mathbf{q} - \mathbf{p}$. Choose a positive definite Lyapunov function $L_o(\mathbf{e})|_{\mathbf{q}=\mathbf{p}}$. Assume that the control signal υ is designed such that there is some time t_0 for which $\dot{L}_o(\mathbf{e}(t))|_{\mathbf{q}=\mathbf{p}} < 0$ when $\mathbf{e}(t) \neq 0$ and $\dot{L}_o(\mathbf{e}(t))|_{\mathbf{q}=\mathbf{p}} = 0$ when $\mathbf{e}(t) = 0$, for all $t > t_0$. That is, after time t_0, the system proceeds monotonically toward synchronization. Let $\boldsymbol{h} \equiv \boldsymbol{f}(\mathbf{y}, \mathbf{r}+\mathbf{p}) - \boldsymbol{f}(\mathbf{y}-\mathbf{e}, \mathbf{p})$. Duane *et al.* [12] established the following theorem:

Theorem 1. *Assume that (i) the control law υ in (17) is designed such that the synchronization manifold $\mathbf{x} = \mathbf{y}$ is globally asymptotically stable, (ii) $\boldsymbol{f}$ is linear in the parameters $\mathbf{p}$, and (iii) the parameter estimation law (18) is designed such that*

$$N_j = -\delta_j \sum_i \left(\frac{\partial L_o}{\partial e_i} \right) \left(\frac{\partial h_i}{\partial r_j} \right),$$

where δ_j are positive constants. Then the synchronization manifold $\mathbf{y} = \mathbf{x}$, $\mathbf{p} = \mathbf{q}$ is globally asymptotically stable.

The theorem ensures the stability of the synchronization manifold $\mathbf{y} = \mathbf{x}, \mathbf{p} = \mathbf{q}$. It says that if the two systems synchronize for the case of identical parameters, then the parameters of the "real system" can be estimated when they are not known *a priori*, provided that each partial derivative $\partial L_o / \partial e_i$ is known for which the vector $\partial h_i / \partial r_j$ $(j = 1, ...)$ is not zero. For the usual form $L_o \equiv \sum_i (e_i)^2$, the requirement is that x_i be known if the equation for $\dot{y}_i$ contains parameters that one seeks to estimate. By considering a more general Lyapunov function that is defined in terms of some subset S of the state variables, or their indices, $L_o \equiv \sum_{i \in S} c_i (e_i)^2$ for positive coefficients c_i, one obtains the looser requirement for each desired parameter, that x_i be known for at least some i for which the $\dot{y}_i$ equation contains that parameter. (Convergence may be slower if fewer x_i are known.)

As a more realistic example than the quasigeostrophic channel model, the Weather Research and Forecasting (WRF) model was considered, as adapted for weather prediction over military test ranges for the Army Test and Evaluation Command (ATEC). The ATEC application is based on observations that are so frequent that they can be assumed to occur at every numerical time step, so that a continuously coupled differential equation system can be taken to reflect the actual data assimilation scenario.

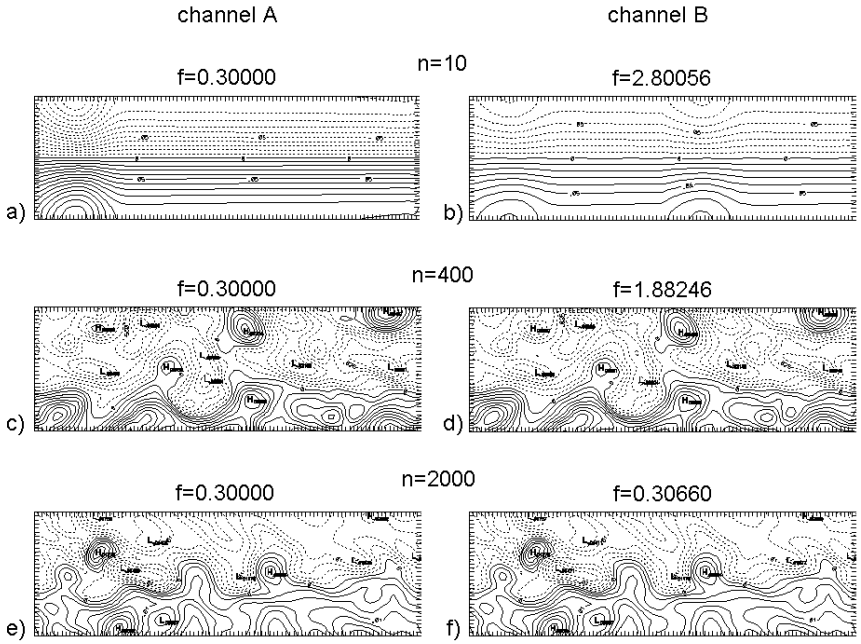

Fig. 4. The evolving flow ψ (a-f) for two quasigeostrophic channel models that are synchronously coupled as in (Duane and Tribbia 2001; Duane and Tribbia 2004) (but in one direction only), and with the forcing parameter f^B for the second channel (denoted μ_o in the reference) allowed to vary according to the truncated parameter adaptation rule (9) with $S = \{\mathbf{k} : \mathbf{k}_x, \mathbf{k}_y \leq 12\}$ (in waves per channel-length). Starting from the initial value $f^B = 3.0$ at time step $n = 0$ (not shown), f^B converges to the value of the corresponding parameter $f^A = 0.3$ in the first channel, as the flows synchronize. (An average of the two layers $l = 1, 2$ is shown.)

At a relevant level of model detail, the prognostic equation for humidity (water vapor mixing ratio) q is:

$$\frac{\partial q}{\partial t} = \frac{\partial}{\partial z}\left\{K\left(\frac{\partial q}{\partial z} - Mf(u_0, T_0, ...)\right)\right\} \tag{19}$$

where K is a moisture diffusivity, and $M = M(x, y)$ quantifies the impact of soil moisture at each location (x, y), which is a function f of state variables such as the zonal wind u_0, temperature T_0, etc. at the surface. To study the estimation of M using the synchronization method, attention is restricted to a single vertical column $(x, y) = (x_0, y_0)$ and a model is introduced that is diffusively coupled to ("nudged" by) the true state. The model humidity q_m, for instance, is governed by:

$$\frac{\partial q_m}{\partial t} = \frac{\partial}{\partial z}\left\{K\left(\frac{\partial q_m}{\partial z} - Mf(u_{m0}, T_{m0}, ...)\right)\right\} + c(q_{obs} - q_m) \tag{20}$$

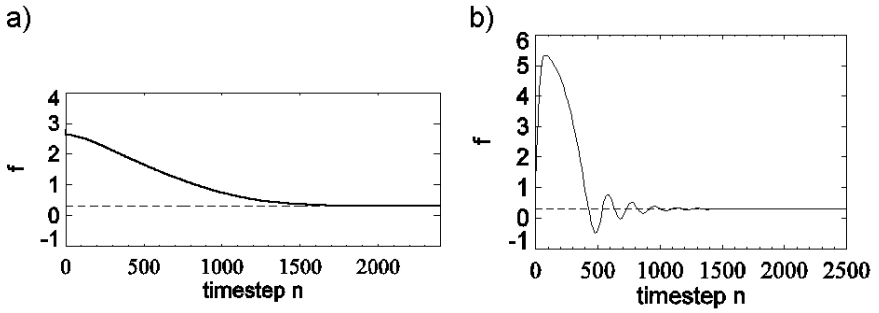

Fig. 5. a) Convergence of f^B to $f^A = 0.3$ (dashed line) for the synchronously coupled quasigeostrophic channels, displayed in Fig. 4, and b) convergence for the simplified parameter adaptation rule (14). Monotonic convergence is replaced by oscillatory convergence.

where q_{obs} is the observed humidity (at any level z where an observation is taken) that is the sum of the true q and observational noise. c is a coupling ("nudging") coefficient. Similar equations govern the evolution of temperature T, wind speed u, and other model variables, but the parameter M is thought to enter only the humidity equation (20).

In accordance with the theorem, extended to PDEs in a straightforward way, M for the model was made to vary with observational input as:

$$\dot{M} \sim -\frac{\partial K}{\partial z} f(u_{m0}, T_{m0}, ...)(q_{obs}(z) - q_m(z)) \tag{21}$$

for the case of observations taken at just one level z. For observations at multiple levels, (21) is simply averaged over several values of z.

Repeated convergence of M to its true value, each time followed by a burst away from synchronization, is seen in Fig. 6. The behavior differs from the smooth convergence in the channel model example because the state variables do not converge in the time interval under consideration (Fig. 6e) - in contrast to the nearly complete synchronization of the two channel models. Importantly, in contrast to the example of the channel model, the prognostic equation (19) is an idealization of the behavior of the adapted WRF model, as implemented in software, the details of which were not completely known to the authors.

In a 3-dimensional model, soil moisture availability $M(x, y)$ is a slowly varying function of ground position (x, y) almost everywhere. It is not slowly varying only for a set of positions of measure zero, at which land cover abruptly changes. It is suggested that the Lyapunov function approach can be readily extended so as to estimate such a parameter field that is slowly varying almost everywhere.

One simply introduces a Lyapunov function of the form:

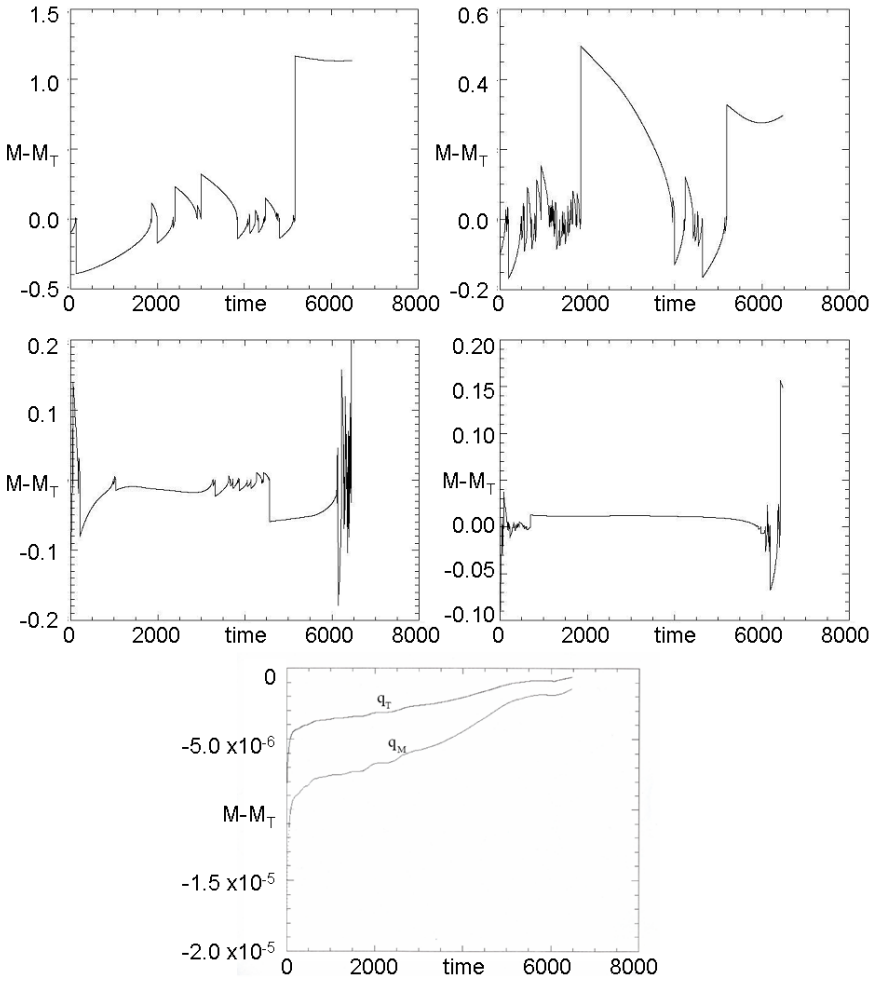

Fig. 6. The variable model parameter M converges to the true value M_T with repeated "bursting", in the dynamical parameter adaptation scheme that only requires a single realization $M - M_T$ is plotted for several cases: a).observations at 7 points in the column, but nudging only at surface level, with nudging coefficient $c = 0.01$ in (20), b) observations and nudging as in a), but with $c = 0.15$, c) observations and nudging at 7 points, with $c = 0.0025$, d) observations and nudging at 4 points, with $c = 0.015$. Results are unstable, but for all cases the correct value of the parameter can be identified. State variables, plotted in e) for observations and nudging at all levels, also do not converge completely over the time interval shown.

$$L = L_{\text{single-column}} + \sum_{x,y} \sum_{(x',y') \in N(x,y)} \big[A(M_{x,y} - M_{x',y'})^2 (1 - l_{x,y,x',y'})^2 + B(1 - l_{x,y,x',y'})^2 (l_{x,y,x',y'})^2 \big] + C\Phi(l) \tag{22}$$

and derives dynamical equations such that L is monotonically decreasing. The term in (22) with coefficient A tends to force smooth spatial variation in M except at locations where the new field l has a value near unity. The variable $l_{x,y,x',y'}$ is conceptually located at a position between (x, y) and (x', y'), and is intended to represent a linear discontinuity in land cover. $N(x, y)$ denotes the local neighborhood of point (x, y). The term with coefficent B tends to binarize the values of l, so that either $l \approx 0$ or $l \approx 1$. The expression $\Phi(l)$ (multiplied by an arbitrary coefficient C) denotes a collection of terms that tends to make the discontinuites along which $l \approx 1$ one-dimensional in $(x, y)-$ space, by inhibiting neighboring parallel "edges", and favoring neighboring contiguous edges.

While the suggested extension of the single-column approach has not yet been tested, it can be expected to succeed on theoretical and empirical grounds. The use of multiple neighboring columns to estimate a local soil parameter promises improvement in principle. The treatment of discontinuities resembles methods that have been effectively applied to image segmentation [13].

5 Concluding Remarks: From Parameter Estimation to Model Learning

The extension of the parameter estimation method to 3D, suggests a further extension to qualitative model learning. For problems of qualitative model optimization, as for the estimation of a 2D parameter field, the requisite Lyapunov function has multiple local optima, as does (22). The optimization problem contrasts with those described by a quadratic Lyapunov function that possess a single basin of attraction. Unlike the quadratic case, a stochastic component in the adaptation procedure might play an essential role. The stochastic component would allow jumps among the basins of attraction of the different local optima defined by the deterministic scheme, as in Fig. 7. The resulting approach would resemble that of a genetic algorithm, with a "mutation rate" proportional to synchronization error.

The synchronization approach to data assimilation and model learning stands in contrast to the use of ensembles to effectively estimate background error [14]. In an anthropomorphic view of machine learning, the synchronization approach appears more natural - one learns "on the fly", rather than forming multiple copies of oneself to test alternative possibilities. In the case of estimating slowly varying parameters, one is effectively using ergodicity to replace an ensemble average by a time average, computed dynamically, as in (9) and (21). For the full three-dimensional WRF model, the single-column version of which was discussed in the last section, such a replacement is essential, since the dimensionality of mesoscale models precludes the use of a large enough ensemble, with currently available computational resources.

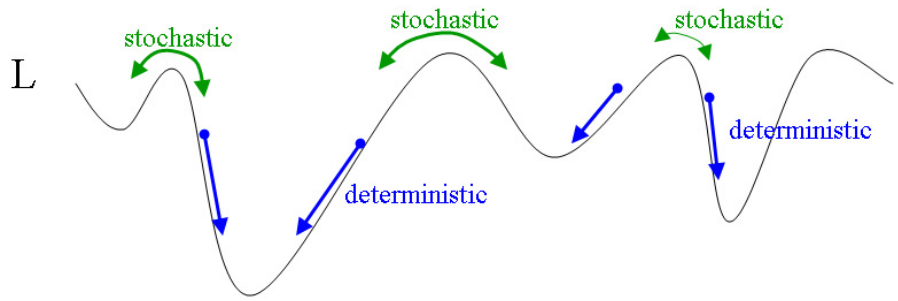

Fig. 7. Deterministic parameter estimation rules cause parameters to reach local optima of the Lyapunov function L. Stochasticity (e.g. in a simulated annealing algorithm) allows jumps among different basins of attraction.

Acknowledgements. The author thanks Joe Tribbia, Jeff Weiss, Dongchuan Yu, and Ljupco Kocarev for useful discussions. This work was partially supported under NSF Grant 0327929. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

References

1. S. Boccaletti, J. Kurths, G. Osipov, D.L. Valladares, C.S. Zhou, The synchronization of chaotic systems, *Physics Reports,* **366,** 1–101 (2002)
2. H. Fujisaka, T. Yamada, Stability Theory of Synchronized Motion in Coupled-Oscillator Systems, *Prog. Theor. Phys,* **69,** 32–47 (1983)
3. V.S. Afraimovich, N.N. Verichev, M.I. Rabinovich, Stochastic synchronization of oscillation in dissipative systems, *Radiophys. Quantum Electron.,* **29,** 795–803 (1986)
4. L.M. Pecora, T.L. Carroll, Synchronization in chaotic systems, *Phys. Rev. Lett.,* **64,** 821–824 (1991)
5. R.A. Anthes, Data Assimilation and Initialization of Hurricane Prediction Models, *J. Atmos. Sci.,* **31,** 702–719 (1974)
6. E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability,* Cambridge University Press, Cambridge (2003)
7. G.S. Duane, Synchronized chaos in extended systems and meteorological teleconnections, *Phys. Rev. E* **56,** 6475–6493 (1997)
8. G.S. Duane, J.J. Tribbia, Synchronized Chaos in Geophysical Fluid Dynamics, *Phys. Rev. Lett.* **86,** 4298–4301 (2001)
9. G.S. Duane, J.J. Tribbia, Weak Atlantic-Pacific Teleconnections as Synchronized Chaos, *J. Atmos. Sci.* **61,** 2149–2168 (2004)
10. R. Vautard, B. Legras, M. Déqué, On the Source of Midlatitude Low-Frequency Variability. Part I: A Statistical Approach to Persistence, *J. Atmos. Sci.* **45,** 2811–2844 (1988)
11. U. Parlitz, Estimating Model Parameters from Time Series by Autosynchronization, *Phys. Rev. Lett.* **76,** 1232–1235 (1996)

12. G.S. Duane, D. Yu, L. Kocarev, Identical synchronization, with translation invariance, implies parameter estimation, *Phys. Lett. A,* **371,** 416–420 (2007)
13. C. Koch, J. Marroquin, A.Yuille, Analog "Neuronal" Networks in Early Vision, *Proc. Nat. Acad. Sci.,* **83,** 4263–4267 (1986)
14. J.L. Anderson, An Ensemble Adjustment Kalman Filter for Data Assimilation, *Mon. Wea. Rev.* **129,** 2884–2903 (2001)