Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Estimating the State of a Geophysical System with Sparse Observations : Time Delay Methods to Achieve Accurate Initial States for Prediction

Zhe An[1], Daniel Rey[1], JingXin Ye[1], and Henry D. I. Abarbanel[1, 2]

[1]Department of Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0374, USA
[2]Marine Physical Laboratory (Scripps Institution of Oceanography) University of California, San Diego 9500 Gilman Drive La Jolla, CA 92093-0374, USA

Correspondence to: Zhe An (z1an@ucsd.edu)

Abstract.

The data assimilation process, in which observational data is used to estimate the states and parameters of a dynamical model, becomes seriously impeded when the model expresses chaotic behavior and the number of measurements is below a critical threshold, L_s. Since this problem of insufficient measurements is typical across many fields, including numerical weather prediction, we analyze a method introduced in Rey et al (2014a, b) to remedy this matter, in the context of the nonlinear shallow water equations on a β-plane. This approach generalizes standard nudging methods by utilizing time delayed measurements to augment the transfer of information from the data to the model. We will show it provides a sizable reduction in the number of observations required to construct accurate estimates and high-quality predictions. For instance, in Whartenby et al (2013) we found that to achieve this goal, standard nudging requires observing approximately 70% of the full set of state variables. Using time delays, this number can be reduced to about 33%, and even further if Lagrangian drifter information is also incorporated.

1 Introduction

The ability to forecast the complex behavior of the earth's coupled ocean, atmosphere system lies at the core of modern numerical weather prediction (NWP) efforts. To successfully predict such behavior requires both a good model of the underlying physical processes as well as an accurate estimate of the state of the model when the observations are completed and the predictions begin. The latter is indispensable, even if one has a perfect model, as the accuracy of the prediction is crucially determined by the quality of the estimated initial state values: if the state of the model at the end of observation window is inaccurate, the forecasts will be undependable.

Suppose one is given a D-dimensional dynamical model and L measurements are made at each observation time $t_n = t_0 + n\Delta t$; $n = 0, 1, ..., N$ throughout an observation window $[t_0 \leq t \leq t_N = T]$. As discussed in Cardinali (2013) there are now about 3-4 $\times 10^7$ daily observations available to NWP models with order 10^8 (or more) degrees of freedom. In our earlier work Whartenby et al (2013) we showed that, using familiar nudging methods, a nonlinear shallow water flow on a β-plane driven by Ekman pumping required about 70% of the $3N^2$ dynamical variables on an $N \times N$ grid to be measured to result

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Discussions
Open Access

in accurate predictions. Using the time delay methods introduced in Rey et al (2014a, b) we show here that number can be reduced to about 33% of the model state variables, and using information from 20 drifters as well, can be further reduced to about 27% of the model state variables that must be observed. Increasing the number of drifters to 64, we have results showing that only 16% of the total degrees of freedom of the model need be observed. These outcomes now bring the data assimilation methods using time delays into the practical range where their use for improving the accuracy of forecasts such as those at ECMWF Cardinali (2013) may be feasible.

While we do not discuss it here in detail, the same results for a 4DVar assimilation method Evensen (2008); Bennett (1992) should be expected. The key dynamical achievement in increasing the number of measurements at each observation time is to control the growth of unstable directions in the local dynamics of the underlying dynamical equations. One can achieve this control through additional measurements introduced into the cost function or action of a variational principle Abarbanel (2013) or via a nudging approach directly implemented in the dynamical equations, as here. The results appear to be equivalent.

A modified version of time delay nudging has been introduced by Pazo (2015), and, although their approach has not yet been applied to geophysical flows, we have no doubt that they will also find the substantial improvement in reducing the number of observations at each measurement time that we report here.

When the number L of measured state components is sparse compared with the total number of degrees of freedom D, $L \ll D$, the main issue is what to do when $L < L_s$. That is, when the number of available measurements is 'insufficient' to produce accurate estimates and dependable predictions. This paper examines this situation and, in particular, explains how to augment observational data taken from a complex system with additional information residing in the time delays of these measurements. In particular, we show how using time delays in this way effectively reduces the number of physical measurements required to solve the initialization problem. This in turn provides more reliable predictions, when the model is accurate, and also facilitates the identification of errors in the model. These benefits (among others) will be demonstrated here in calculations on a core geophysical model: the nonlinear shallow water equations.

Although this discussion will focus solely on a specific geophysical system, the methods we describe here have broad applicability across the quantitative study of the underlying physical or biological properties appearing in many complex systems. The notable feature of high dimensional dynamics and sparse ($L \ll D$) measurements is typical in the process of examining the consistency of observed data and quantitative models of complex nonlinear systems: from functional nervous systems to genetic transcription dynamics to complex earth systems models, among many other examples Abarbanel (2013).

2 Transferring Information from the Waveform of Observations to a Model

While the details of the discussion in this section may be found in Rey et al (2014a, b) we briefly repeat the ideas of time delay nudging here. We emphasize that we introduce few additional tools in the time delay methods in this paper. Our goal is to demonstrate within the context of a nonlinear geophysical flow how the ideas can be used toward the practical goal of reducing the number of physical measurements required to achieve an accurate initialization for predictions.

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Discussions
Open Access

In data assimilation we seek to use the information in observations to determine properties of a model that describes the dynamics producing those observations. These properties include unknown parameters as well as the time dependence of unobserved state variables. The model acts as a nonlinear filter coupling the observed states to the parameters and unobserved states.

During a temporal observation or assimilation window $[0 \le t \le T]$, we perform L observations $\mathbf{y}(t_n)$ at each time $t_n = \{t_0, t_1, t_2, \ldots, t_N = T\}$. Between observations, the system moves its D-dimensional state $\mathbf{x}(t)$ ahead in time via a deterministic set of ordinary differential equations

$$\frac{dx_a(t)}{dt} = F_a(\mathbf{x}(t)). \tag{1}$$

If the dynamics of the system is described by partial differential equations (such as with the fluids in an earth systems model) the ordinary differential equations may be realized by discretizing the partial differential equations on a grid, in which the grid label as well as the vectorial nature of the state variables are collected into the index $a = \{1, \ldots, D\}$. In addition, we assume here for simplicity that the measurements $\mathbf{y}(t)$ are simply projections of the overall state of the system $\mathbf{x}(t)$. With the usual noisy measurements we would have $y_\ell(t) = x_\ell(t) + \text{noise}$; $\ell = 1, 2, ..., L$. This is not imperative however, as what follows may be generalized to handle an arbitrary observation operator where $\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t)) + \text{noise}$.

The main objective is to estimate the model state at the end of the assimilation window $\mathbf{x}(T)$ using information from the sparse observations $L \ll D$, and then use this estimate to predict the system's subsequent behavior for $t > T$ using Eq. (1). The accuracy of these predictions, when compared with additional measured data in the prediction window $t > T$, serves as a metric to validate both the model and the assimilation method, through which the unobserved states of the system are determined. This is crucial, because it establishes a necessary condition on L that is required to synchronize the model output with the data and thereby obtain accurate estimates for the unobserved states of the system, which are also required to make good predictions.

In the limit that the model is known precisely, a familiar strategy for transferring information from the measurements to the model involves the addition of a coupling or control or nudging term to Eq. (1),

$$\frac{dx_a(t)}{dt} = F_a(\mathbf{x}(t)) + \sum_{\ell=1}^{L} g_{a\ell}(t)\big[y_\ell(t) - x_\ell(t)\big]. \tag{2}$$

In this expression, $\mathbf{g}(t)$ is a $D \times L$ matrix, whose elements are nonzero, only when $t = t_n$, in rows that correspond to measured states. In this way, the coupling term perturbs the measured states of the system to drive the observed model states towards the data.

With enough observations L, a sufficiently strong coupling will alter the Jacobian of the dynamical system Eq. (2) so that all its (conditional) Lyapunov exponents are negative. See Pecora & Carroll (1990); Abarbanel (1996); Kantz & Schreiber (2004). This is important, as it establishes a necessary condition on L required to synchronize the model output with the data and thereby obtain an accurate estimate for the unobserved states of the system which are also needed to make good predictions.

This long-standing procedure, known as nudging in the geophysics and meteorology literature, has been shown Abarbanel et al (2009) to fail when the number of measurements at a given time is smaller than a critical value L_s. An explicit example of this

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Open Access
Discussions

will be shown later in our discussion of geophysical shallow water flow. Since in practice however, additional measurements are often scarce, we must find another means to overcome this deficit in L.

One way to proceed Rey et al (2014a, b); Pazo (2015) involves the recognition that additional information resides in the temporal derivatives of the observations. In practice, however, this derivative information cannot be measured directly, so it must be approximated via finite differences. For instance, one may approximate $d\mathbf{y}(t_n)/dt$ with $[\mathbf{y}(t_n+\tau)-\mathbf{y}(t_n)]/\tau$ where τ is some multiple of the time differences between measurements. The drawback here is that the derivative operation acts as a high-pass filter, and is thus quite susceptible to measurement noise. Although convenient from an analytical perspective, the derivative is not directly suitable for the purposes considered here.

Rather, we proposed a technique to extract additional information from the waveform of observations by establishing an extended state space, created from an $L \times D_M$ dimensional vector of the measurements and its time delays Rey et al (2014a, b). The components of this extended space measurement vector are denoted by

$$\mathbf{Y}_{\ell,k}(t) = y_\ell(t+(k-1)\tau);\;\; \ell = 1,2,...,L;\; k = 1,2,...,D_M. \tag{3}$$

This is a collection of D_M-dimensional time delays (indexed by k) for each observed $y_\ell(t_n)$.

This idea stems from a well-known technique in the analysis of nonlinear dynamical systems, where this structure is employed as a means of reconstructing unambiguous orbits of a partially observable system Aeyels (1981a, b); Mañé (1981); Sauer et al (1991); Takens (1981); Abarbanel (1996); Kantz & Schreiber (2004). In that nonlinear dynamical context, the proxy space of time delayed observations serves as a way to invert the projection associated with measuring $L < D$ components of the underlying dynamics. The main idea is that the new information beyond $\mathbf{y}(t_n)$ lies in $\mathbf{y}(t_n+\tau)$ itself; the derivative operation is just another way of accessing this information.

In the present context, however, our use of time delay coordinates is quite distinct. Rather than reconstructing the topology of the attractor, we instead use the time delay construction to control the local instabilities in the dynamics. That is, for our purposes, D_M need only be large enough to effectively increase the amount of information transferred from the L measurements to a value above the critical threshold, L_s.

For this task, we construct the corresponding time delay model vectors, whose components are given by

$$\mathbf{S}_{\ell,k}(\mathbf{x}(t)) = x_\ell(t+(k-1)\tau);\;\; \ell = 1,2,...,L;\; k = 1,2,...,D_M. \tag{4}$$

The dynamical rule for the vectors $S_{\ell,k}(\mathbf{x}(t))$ is given by,

$$\frac{dS_{\ell,k}(\mathbf{x}(t))}{dt} = \sum_{a=1}^{D} \frac{\partial S_{\ell,k}(\mathbf{x}(t))}{\partial x_a(t)} F_a(\mathbf{x}(t)) = \sum_{a=1}^{D} \frac{\partial S_{\ell,k}(\mathbf{x}(t))}{\partial x_a(t)} \frac{dx_a}{dt}. \tag{5}$$

Following the idea expressed in Eq. (2), we introduce a control term in time delay space,

$$\frac{dS_{\ell,k}(\mathbf{x}(t))}{dt} = \sum_{a=1}^{D} \frac{\partial S_{\ell,k}(\mathbf{x}(t))}{\partial x_a(t)} F_a(\mathbf{x}(t)) + \sum_{k'=1}^{D_M} \sum_{\ell'=1}^{L} \mathbf{G}_{\ell,k,\ell',k'}(t) \big[Y_{\ell',k'}(t) - S_{\ell',k'}(\mathbf{x}(t))\big]. \tag{6}$$

Transforming this rule back into the physical space of the $\mathbf{x}(t)$, we arrive at our required result,

$$\frac{dx_a(t)}{dt} = F_a(\mathbf{x}(t)) + \mathbf{g}_{ab}(t) \frac{\partial x_b(t)}{\partial S_{\ell,k}}(\mathbf{x}(t))\, \mathbf{G}_{\ell,k,\ell',k'}(t) \left[Y_{\ell',k'}(t) - S_{\ell',k'}(\mathbf{x}(t))\right] \tag{7}$$

where repeated indices are summed over. There are two control terms: $\mathbf{g}(t)$ and $\mathbf{G}(\mathbf{t})$, which act respectively in physical space and in time delay space. The result is at each step of the integration of the controlled (nudged) dynamical equations Eq. (2), the added control term moves the full state vector in time delay space toward the full time delay measurement vector.

The matrix $\partial\mathbf{x}(t)/\partial\mathbf{S}(\mathbf{x}(t))$ determines the directions in phase space along which the controls are applied, and it is to be understood as the generalized inverse of the $L\,D_M \times D$ rectangular matrix $\partial\mathbf{S}(\mathbf{x}(t))/\partial\mathbf{x}(t)$. The latter is constructed from elements of the variational matrix $\Phi_{ab}(t,t_n) = \partial x_a(t)/\partial x_b(t_n)$ for $t \geq t_n$ starting at each time step t_n by integrating the variational equation

$$\frac{d\Phi_{ab}(t,t_n)}{dt} = \sum_{c=1}^{D} DF(\mathbf{x}(t))_{ac}\,\Phi_{cb}(t,t_n) \qquad \Phi_{ab}(t_n,t_n) = \delta_{ab}, \tag{8}$$

from $t = t_n$ to $t = t_n + \tau(D_M - 1)$. $DF(\mathbf{x}(t))_{ab} = \partial F_a(\mathbf{x}(t))/\partial x_b(t)$ is the Jacobian of the vector field and δ_{ab} is the Kronecker delta. Both the map $\mathbf{x}(t) \to \mathbf{S}(\mathbf{x}(t))$ and the variational equation Eq. (8) are computed using the uncoupled dynamics Eq. (1). For further details on this calculation, see Rey et al (2014a, b).

It is worth noting that in the limit $D_M = 1$ the time delay formulation Eq. (7) reduces to the standard nudging control Eq. (2). Two important differences however are realized when $D_M > 1$: (1) information from the time delays of the observations is presented to the physical model equations, and (2) all components of the model state $\mathbf{x}(t)$ are influenced by the control term, not just the observed components. Consequently, the fixed parameters of the model may be estimated as a natural result of the synchronization process by including them as additional state variables, satisfying $dp_r(t)/dt = F_r(\mathbf{x}(t)) = 0$.

The main achievement of this technique is that it extracts additional information from the time-series of *existing* measurements. The value of this statement will become more clear in the context of our simple geophysical example.

3 Twin Experiments

We test our time delay nudging procedure through a series of numerical simulations called 'twin experiments' Durandet al (2002); Blum et al (2009); Blum (2010). After solving the original dynamical equations Eq. (1) forward from some initial condition $\mathbf{x}(0)$, the observed data is taken as the projection down to the L observed components. Gaussian noise $N(0,\sigma)$ is added to each component to simulate observation error, but the model itself is assumed known perfectly. So this framework tests the estimation procedure, not the model. Removing the issue of model error allows us to assess the weaknesses and strengths of the estimation algorithm and explore in detail the manner in which the unobserved variables are determined. When successful, it provides confidence that the method may be applied to real data. When it fails, it helps us figure out why.

In a twin experiment, the full state of the system is known at all times. However, in a *true* experiment we have only the observed quantities to compare. In this case, we can monitor our process by calculating the observable synchronization error between the model and the data,

$$SE(t)^2 = \frac{1}{L}\sum_{\ell=1}^{L}\left[x_\ell^s(t) - y_\ell^s(t)\right]^2, \tag{9}$$

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

where we introduce the scaled variables $x^s_\ell(t) = [x_\ell(t) - x^{min}_\ell(t)]/[x^{max}_\ell(t) - x^{min}_\ell(t)]$ and $x^{min/max}_\ell(t)$ are the minimum or maximum values of $x_\ell(t)$, over the entire assimilation window. The same definition holds for $y^s_\ell(t)$. This rescales all data and observed model states to lie in the interval $[0,1]$, so that each state component's contribution to the synchronization error is weighted approximately equally.

The synchronization error Eq. (9) involves only observable state variables, which makes it a more realistic way to monitor the success of the estimation, since the 'true' error is only known in the context of a twin experiment. Furthermore, when the estimation is complete (at time $t = T$), the coupling terms $\mathbf{g}(t)$ and $\mathbf{G}(t)$ are set to zero and the uncoupled dynamics Eq. (1) are integrated forward from the estimated $\mathbf{x}(T)$ to construct a forecast for $t > T$, which may then be compared with additional observations $\mathbf{y}(t > T)$. It is crucial to compare $SE(t)$ for *both* estimates *and* predictions, as the former is just a 'fit' involving measured quantities, while the latter relies on accurate determination of the unmeasured variables as well. No additional information is passed from the data to the model during the prediction phase.

Thus, in a true experiment the prediction error should be used exclusively. Accurate estimates alone are not sufficient to validate the model or indicate the success of the estimation procedure, as they do not any information about the unobserved states.

We have previously shown that when the synchronization error Eq. (9) decreases in time to very small values, the full state $\mathbf{x}(T)$ has been accurately estimated and the prediction is quite good Whartenby et al (2013) — assuming the model is known precisely. On the other hand, when the synchronization error does not decrease to very small values, the full state $\mathbf{x}(T)$ is not well estimated and the prediction is unreliable. Moreover, the synchronization error appears to decrease only when the number of time delayed observations $L \times D_M$, and the magnitude of the elements of the coupling matrices $\mathbf{g}$ and $\mathbf{G}$ are 'large enough'. The precise meaning of this statement will become apparent shortly, in the context of our simple geophysical example.

4 Nonlinear Shallow Water Equations

To illustrate how time delays utilize information latent in the waveform of a time series of observations, we now describe in detail its application to a model of shallow water flow on a mid-latitude β-plane. This geophysical fluid dynamical model, previously examined by Pedlosky (1987) and Whartenby et al (2013), among many others, is at the core of earth system flows used in NWP. Of course, production NWP models contain much more detail than this example, and those models also describe the dynamics over a whole sphere. We argue that the results presented here, for this simplified model, will be applicable for establishing the initial state of those models and predicting their subsequent behavior. We do not underestimate the numerical challenges implicit in this extrapolation.

As the depth of the atmosphere/ocean fluid layer (order $10-15$ km) is markedly less than the earth's radius (6400 km), the shallow water equations for two dimensional flow are an excellent approximation to the fluid dynamics in such a geometry. Three fields on a mid-latitude plane describe the fluid flow $\{u(\mathbf{r},t), v(\mathbf{r},t), h(\mathbf{r},t)\}$: the north-south velocity $v(\mathbf{r},t)$, the east-west velocity $u(\mathbf{r},t)$, and the height of the fluid $h(\mathbf{r},t)$, with $\mathbf{r} = \{x,y\}$. The fluid is taken as a single, constant density layer and is driven by wind stress $\tau(\mathbf{r},t)$ at the surface $z = h(\mathbf{r},t)$ through an Ekman layer. These physical processes satisfy the

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Open Access
Discussions

Parameter	Physical Quantity	Value in Twin Experiments
Δt	Time Step	36 s
ΔX	East-West Grid Spacing	50 km
ΔY	North-South Grid Spacing	50 km
H_0	Equilibrium Depth	5.1 km
f_0	Central value of the Coriolis parameter	5×10^{-5} s^{-1}
β	Meridional derivative of the Coriolis parameter	2.0×10^{-11}m^{-1}s^{-1}
F/ρ	Wind Stress	0.2 m^{2}s^{-3}
A	Effective Viscosity	10^{-4} m^{2}s^{-1}
ϵ	Rayleigh Friction	2×10^{-8} s^{-1}

Table 1. Parameters used in the generation of the shallow water 'data' for the twin experiment. All fields as well as $\{x,y,t\}$ were scaled by the values in the table, so all calculations were done with dimensionless variables.

following dynamical equations with $\mathbf{u}(\mathbf{r},t) = \{u(\mathbf{r},t), v(\mathbf{r},t)\}$,

$$\begin{aligned}\frac{\partial \mathbf{u}(\mathbf{r},t)}{\partial t} &= -\mathbf{u}(\mathbf{r},t)\cdot\nabla\mathbf{u}(\mathbf{r},t) - g\,\nabla h(\mathbf{r},t) + \mathbf{u}(r,t)\times f(y)\,\hat{z} + A\,\nabla^2\mathbf{u}(\mathbf{r},t) - \epsilon\,\mathbf{u}(\mathbf{r},t)\\ \frac{\partial h(\mathbf{r},t)}{\partial t} &= -\nabla\cdot(h(\mathbf{r},t)\,u(\mathbf{r},t)) - \hat{z}\cdot\mathrm{curl}\left[\frac{\tau(\mathbf{r},t)}{f(y)}\right].\end{aligned} \tag{10}$$

The Coriolis force is linearized about the equator $f(y) = f_0 + \beta y$ and the wind-stress profile is selected to be $\tau(\mathbf{r},t) = \{F/\rho\cos(2\pi y), 0\}$. The parameter A represents the viscosity in the shallow water layer, ϵ is Rayleigh friction and $\hat{z}$ is the unit vector in the z-direction. The values we have used for the model parameters are given in Table 1. With these fixed parameters the shallow water flow is chaotic, and the largest Lyapunov exponent for this flow is $\lambda_{max} = 0.0325/h \approx 1/31\,h$.

We have analyzed this flow using the enstrophy conserving discretization scheme from Sadourny (1975) on a periodic grid of size N_Δ^2 for $N_\Delta = \{16, 32, 64\}$ with periodic boundary conditions. Using the twin-experiment framework with simple nudging coupling controls given in Eq. (2), we estimated that approximately 70% of the $D = 3N_\Delta^2$ degrees of freedom must be observed in order to synchronize the model output with the data. These results on the required number of observations agree with previous calculations, presented in Whartenby et al (2013). The actual control in Eq. (2) was applied only in the observed components and only along the diagonal.

As the results are consistent across the various grid sizes that were investigated, we restrict our discussion here to the case where $N_\Delta = 16$, so that the total number of degrees of freedom $D = 3N_\Delta^2 = 768$ and $L_s \approx 524 = 0.68\,D$ has been explicitly calculated for this case. We are confident that despite the numerical challenges associated with scaling the algorithm up to larger D, the results presented here for $N_\Delta = 16$ will also remain valid for higher grid resolution.

In the discussion above, which included reference to the lectures of Cardinali Cardinali (2013), we see that the requirement of having to observed 70% of the model dynamical variables exceeds the measurements now available by at least a factor of two; more if the NWP model is larger yet.

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Open Access
Discussions

5 Results with Time Delay Nudging for the Shallow Water Equations

We now demonstrate the capability of the time delay method to reduce the number of measurements required to generate accurate predictions. In particular, no information about the tangential velocity fields is used to estimate the state. This strategy was shown to fail in Whartenby et al (2013) with static $D_M = 1$ nudging. With the addition of time delays however, it is possible to estimate the full state (heights and velocities) when only height information is available.

We assume height measurements alone are made at each grid point (i,j) for $i,j = \{1,2,\ldots,16 = N_\Delta\}$, so $L = 256 < 524 \approx L_s$. The initial state $\mathbf{x}(t_0)$ for the model and the data are taken to have the form,

$$h^{(i,j)}(t_0) = \left(\frac{\pi A_0}{N_\Delta \Delta Y}\right)^2 \left[\cos(\omega_\phi \phi(\mathbf{r}(i,j)) + \delta_\phi) + \cos(\omega_\theta \theta(\mathbf{r}(i,j)) + \delta_\theta)\right] + H_0$$

$$u^{(i,j)}(t_0) = -A_0 \frac{\partial \psi(\mathbf{r}(i,j))}{\partial x} \qquad v^{(i,j)}(t_0) = A_0 \frac{\partial \psi(\mathbf{r}(i,j))}{\partial y}$$

where the parameters $H_0 = 5100,\ A_0 = 10^6$ and

$$\psi(\mathbf{r}) = \cos(\omega'_\phi \phi(\mathbf{r}) + \delta'_\phi) \sin(\omega'_\theta \theta(\mathbf{r}) + \delta'_\theta).$$

The functions ϕ and θ respectively evaluate the latitude and longitude at the point $\mathbf{r}^{(i,j)}$ on the grid. The parameters $\omega_\phi, \omega_\theta, \omega'_\phi, \omega'_\theta$ and $\delta_\phi, \delta_\theta, \delta'_\phi, \delta'_\theta$ are chosen arbitrarily to distinguish the initial conditions between the model and data. All fields as well as $\{x, y, t\}$ were scaled by the values in Table (1), so all calculations were done with dimensionless variables. Although the time delay method is capable of estimating the model parameters, here they are treated as known.

The coupling matrix $\mathbf{G}(t)$ is taken to be diagonal with different weights for the heights and for the velocities. In particular, $G_{u,v}\,\Delta t = 0.5$ and $G_h\,\Delta t = 1.5$ with $\Delta t = 0.01\,h = 36\,s$. These are chosen because the height values are several orders of magnitude larger than the flow velocities. For instance, the average heights are around $5000 \pm 30\,m$ whereas average velocities are approximately $0 \pm 5\,m/s$. The time delay space coupling $\mathbf{g}(t)$ is taken to be the identity matrix, as all the height measurements are assumed to be known with equal precision.

The time delay was selected to be $\tau = 10\,\Delta t = 0.1\,h$, in order to maintain a balance between numerical stability and the common criterion of independence between the components of $\mathbf{S}(\mathbf{x}(t))$. To further justify this choice, we also calculated the first minimum of the average mutual information to be $\tau \approx 30\,\Delta t$. Hence, our selected value is smaller, though still consistent with the standard convention. Furthermore, the results were reasonably stable to changing its value within a few Δt.

To estimate the state, we integrated the modified differential equations Eq. (7) within a data assimilation window $[0,T]$; $T = 5\,h = 500\,\Delta t$ with various $D_M = \{1,6,8,10\}$. The coupling terms were switched off at $t = T$, and the estimate $\mathbf{x}(T)$ was used as initial conditions to generate predictions until $t = 500\,h$.

Short and long term synchronization error Eq. (9) trajectories $SE(t)$ are plotted in Figure 1 for various D_M. Choosing $D_M = \{1,6\}$, yields a synchronization error that remains around its initial value of 0.005 until the end of the five hour observation window. After the coupling is switched off, the error rises very rapidly until stabilizing around 0.1 for the remainder of the prediction window. By contrast, for $D_M = \{8,10\}$ the synchronization error falls steeply to order 10^{-6} within the observation

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics Discussions

window. It then subsequently rises as $e^{\lambda_{max}(t-T)}$, where λ_{max} agrees with the largest Lyapunov exponent calculated for this flow. This exponential rate of growth is particularly evident in the long trajectory displayed in the Right Panel of Figure 1.

Since $D_M \geq 8$ produces error values several orders of magnitude smaller than those obtained with $D_M \leq 6$, we expect the state estimates $\mathbf{x}(T)$ obtained with $D_M \geq 8$ to be quite accurate when compared with the estimates for $D_M \leq 6$. These estimates are now evaluated as they would be in a true experiment: no information about the unobserved states is used. We compare subsequent predictions on the observed heights with additional measured data, but in the prediction window $t > 5$ no information about the new measurements is passed back to the model. In Figure 2 the known (black), estimated (red), and predicted (blue) height trajectories are shown for an arbitrarily selected grid point $h^{(6,4)}(t)$. Short and long term prediction trajectories computed with $D_M = 6$ are displayed in Figure 2 upper panels respectively. Corresponding results for $D_M = 8$ are shown in the lower panels. As anticipated, the predictions for $D_M = 8$ are clearly superior to those obtained with $D_M = 6$. Just a reminder note here, we used L = 256 = 33% of the total 768 dynamical variables as observed, then used time delay information on the waveform of the measurements to provide the required additional information.

The failure of predictions obtained with $D_M = 6$ is a result of poor estimates of the unobserved states (i.e. fluid velocities) at $t = T$. Although in an actual experiment we would not be able to verify this statement directly, we may do so here in the context of a twin experiment because the full state information is available. Velocity profiles $u^{(6,4)}(t)$ displaying short and long time comparisons between the known (black), estimated (red) and predicted (blue) values are given in Figure 3 for $D_M = 6$ in the upper panels, and for $D_M = 8$ in the lower panels. We find the situation is indeed as anticipated; the estimates and predictions are quite unacceptable for $D_M = 6$, whereas for $D_M = 8$ they are highly accurate. The same striking improvement in predictive accuracy was obtained for the other velocity component $v^{(6,4)}(t)$. These results are plotted in Fig. 4.

As this point is the key theme of this paper, we take the liberty of repeating that the number of *physical* measurements is just 33% of the overall dynamical variables. Using standard nudging (D_M = 1) we found the same flow required $L = 524$ or $\approx 70\%$ of the dynamical variables to be observed. The required additional information to make accurate estimates of all states at t = 5 h comes from the waveform information in the time delays.

Predictions were also calculated for $D_M = 1$ and $D_M = 10$, but these results are not shown. They agree with the synchronization error calculations in Figure 1, in that the predictions generated with $D_M = 10$ are just as accurate as those for $D_M = 8$. Likewise, predictions with $D_M = 1$ (i.e. simple nudging) are very poor, in accordance with our previous results in Whartenby et al (2013).

The success of this procedure when $D_M \geq 8$ is attributed to information transferred from the additional time delays. In particular, our modification Eq. (7) to the dynamical equations alters the Jacobian of the dynamics $\mathbf{DF}(\mathbf{x}(t))$, adjusting the conditional Lyapunov spectrum of the modified system in a manner dependent upon the coupling strengths $\mathbf{G}(t)$, $\mathbf{g}(t)$ and the presented data. When enough information is available, and the coupling is strong enough, these conditional Lyapunov exponents will all be negative, allowing the coupled systems of data and model output to synchronize.

This outcome suggests, in turn, that reducing the coupling strength will have a detrimental effect on the quality of the estimation procedure and the resulting prediction. We investigate this now, by performing the same calculations as above with $D_M = 10$ but reducing the coupling on the height so we have $G_h\,\Delta t = g_u\,\Delta t = g_v\,\Delta t = 0.5$. The synchronization error $SE(t)$,

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Open Access
Discussions

shown in Figure 4, Upper Left Panel, stabilizes to a level three orders of magnitude higher than was achieved with $G_h\,\Delta t = 1.5$, suggesting that the assimilation procedure has failed. This is confirmed in the remainder of Figure 4, which displays the known (black), estimated (red) and predicted (blue) values for $h^{(6,4)}(t)$, $u^{(6,4)}(t)$, and $v^{(6,4)}(t)$, respectively. Although the height estimate is rather good and the corresponding prediction is not terrible, at least for the first 15-20 hours after the end of the assimilation window, the unobserved states are clearly not well estimated at any time $t \leq T$. This result demonstrates that proper choice of coupling is crucial to the success of the procedure. The fact that the height estimates appear to be rather accurate also emphasizes the point that, in a true experiment, the success of the assimilation procedure must be evaluated against the predictions—not the estimates.

In addition, until now we have conveniently chosen to observe the height field at all $L = N^2$ grid locations. We now attempt to reduce L even further, so as to relieve some of the demand on the data collection process. This could simulate for instance, finite resolution in satellite measurements. In particular, keeping all other parameters fixed, we executed our estimation procedure using $L = 252$ and $L = 248$ height measurements chosen at arbitrary grid points. Note that these L values are rather close to $L = 256 = N_\Delta^2$ considered above. From the synchronization errors $SE(t)$, displayed in the Upper Left Panel of Figure 6, it is evident that for $L = 252$ rapid and accurate synchronization is still achieved, while for $L = 248$ it is not. In addition, the known (black), estimated (red), and predicted values (blue) for $h^{(6,4)}(t)$ are shown in the other panels of Figure 6 for $L = 248$ and $L = 252$ respectively. Results for the unobserved velocity fields agree as well, though these results are not shown. Thus, in accordance with our previous results, the connection between the synchronization and accurate predictions remains intact.

Moreover, even with the additional time delays, it appears that it may not be possible to significantly reduce the number of required height measurements. We remark, however, that the overall space of parameters appearing in our formalism has not been thoroughly explored and that by further adjusting these parameters ($\mathbf{G}(t)$, $\mathbf{g}(t)$, D_M, and τ), it may be possible to produce good predictions with even fewer observations. It may also be interesting to investigate this question as the grid resolution is increased. One would expect, that at some point the resolution should be high enough to not necessitate further measurements.

5.1 Measurements with Gaussian Noise

In operational data assimilation in meteorology, one challenge is that the measurement contains observation error. An effective assimilation algorithm must be robust under such noise contamination. We investigated the sensitivity to the initial condition noise for the $D_M = 8$ and $D_M = 10$ time delay data assimilation with L = 252 as described above in detail.

The shallow water model and the data generation were taken to be the same as above, but to the initial condition for the data we added iid Gaussian white noise $N(0,1)$ in the form

$$\phi_{data}(\mathbf{r},t_0) = \psi_{model}(\mathbf{r},t_0) + C_{data}N(0,1)$$
$$h_{data}^{(i,j)}(t_0) = C_{height}N(0,1),$$

and we selected $C_{data} = 10^6$ and $C_{height} = 1652$.

Figure 7 compares the synchronization error with and without noise in the observation. The synchronization error still falls rapidly within the observation window and, after that, rises in an exponential manner as expected. Even with noise in the

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Discussions

initial conditions propagating through the dynamical variables via the dynamical equations, choosing a sufficient time delay dimension, here $D_M = 8$ and $D_M = 10$ still results in small synchronization error when observations are completed. The time delay nudging method remains robust under imperfect observations.

6 Using Drifter Data with Time Delays

Another greatly expanding source of observations about ocean flows come now through measurements of the position $\mathbf{r}(t)$ of Lagrangian drifters Mariano et al (2000). Sparse observation is a bottleneck in modern data assimilation in oceanography. Several papers have shown that Lagrangian observations can be a good supplement to the traditional observations which are made on a fixed grid Kuznetsov et al (2003). The Lagrangian observations are used to estimate an Eulerian velocity field for most assimilation schemes Molcard et al (2003); Piterbarg et al (2008); Salman et al (2006).

In this section, we combine the time delay method used in the previous twin experiment simulation with a data set from drifter measurements and assimilate the information they contain into estimates for the state variables $\{h(\mathbf{r},t), u(\mathbf{r},t), v(\mathbf{r},t)\}$. Since the data type of the drifter measurements differs substantially from the velocity and height information, our target is to show the time delay nudging method can be used in the data assimilation with drifters as well. Without a complicated setup and fine-tuning the coupling coefficients, the assimilation method still provides a solid estimation of the state variables.

In this section, 20 drifter positions are measured within the estimation window, the initial deployment positions are chosen randomly across the N_Δ^2 grid locations. After the initial deployment, the drifters move between grid points providing information not available from grid point measurements alone.

The dynamics of drifters is described as two-dimensional fluid parcel motion on the surface of the water layer. Since the positions of drifters are continuous values, the velocities of the drifters are estimated by a smooth linear interpolation Press et al (2015); Thompson and Emery (2014) of the discrete velocity field of the water layer.

The initial conditions of the data $\psi_{data}(\mathbf{r},t_0)$ and $h^{(i,j)}_{data}(\mathbf{r},t_0)$ and of the model $\psi_{model}(\mathbf{r},t_0)$ and $h^{(i,j)}_{model}(\mathbf{r},t_0)$ are related by $\psi_{data}(\mathbf{r},t_0) = C_0\psi_{model}(\mathbf{r},t_0)$ and $h^{(i,j)}_{data}(\mathbf{r},t_0) = C_0 h^{(i,j)}_{model}(\mathbf{r},t_0)$. We chose $C_0 = 1.0 + 0.1\times\eta$ with η selected from a uniform distribution in the interval [-1,1]. The velocity fields are found as above, using $\psi(\mathbf{r},t_0)$ as a stream function.

The time delayed nudging method with hybrid measurements can be written as a control added to the dynamical equations in a combined state variable plus drifter time delay space. The observations are written as $\mathbf{Y}_{drifter}(t) = \{\mathbf{Y}_{\ell,k}(t), \mathbf{R}_{drifter}(t)\}$ and the total state variable is written as $\mathbf{S}_{drifter}(t) = \{\mathbf{S}_{\ell,k}(t), \mathbf{S}_{drifter}(t)\}$. The time delayed observations of drifter $n = 1,2,...,N_D$ at location $\mathbf{r}^{(n)}(t)$ are incorporated in

$$\mathbf{R}_{drifter}(t)\} = \{\mathbf{r}^{(n)}_{obs}(t), \mathbf{r}^{(n)}_{obs}(t+\tau), \mathbf{r}^{(n)}_{obs}(t+2\tau), ..., \mathbf{r}^{(n)}_{obs}(t+(D_M-1)\tau)\}$$

A similar expression for the model drifter coordinates enters $\mathbf{S}_{drifter}(t)$, which are determined by the Lagrangian equations

$$\frac{d\mathbf{r}^{(n)}(t)}{dt} = \mathbf{u}(\mathbf{r}^{(n)}_{obs}(t),t).$$

In Figure 8 we present results for the synchronization error of observed quantities when we select $D_M = 8$ and all other parameters as in the previous calculations with no drifters. We present (in red) the synchronization error when we have L = 208

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Open Access
Discussions

observations of heights and $N_D = 20$ drifter location observations, and we present (in blue) the same synchronization error when L = 208 and $N_D = 0$ drifters are deployed. We see that with L = 208, namely 27% of the heights observed and 20 drifters observed, we move into a regime where small synchronization error results within our five hour observation window. With no drifters, L = 208 does not result in small synchronization errors, and this is consistent with the results reported just above.

We have further investigated how the geographic distribution of the drifters influences the size of the synchronization error, although these results are not displayed here. One striking result was that when L = 128 heights $h^{(i,j)}(t)$ were observed over 6h without drifters, the synchronization error remained large. The homogeneous addition of $N_D = 64$ drifters in a 30 minute observation window, reduced the synchronization error nearly to zero. This underlines the utility of adding time delay coordinates to the observed drifter locations to extract further information useful to determining the state of the model for an initial state for predictions Molcard et al (2003); Piterbarg et al (2008); Salman et al (2006).

We have not yet explored how to make a balance between the number of drifters tracked and the number of height (or other) measurements employed. It is clear, however, that drifter data can be of substantial use in estimating the state of a geophysical flow using time delay coordinates to enhance the value of each existing drifter measurement.

7 Conclusions

In an earlier paper Whartenby et al (2013) we showed that using standard nudging, in which a control term such as Eq. (2) uses observations directly, requires 70% of the dynamical variables $\{u(\mathbf{r},t), v(\mathbf{r},t), h(\mathbf{r},t)\}$ at every measurement time to establish synchronization between model output and observations. In that paper, we utilized additional information from the waveform of the time series to reduce the number of required measurements to about 30% of all dynamical variables. Further, using information from drifters, which probe between grid points, we were able to show further reduction in the required number of observations to achieve excellent predictions.

Although we have done this all within our model of shallow water flow on a β-plane driven by surface winds, each step of our work can be used to analyze increasingly realistic and complex models of the ocean, atmosphere system. When observations are sparse, the time delay method may be the critical step in allowing accurate estimations of the full state of the model for use in predictions. This in turn allows one to achieve accurate state estimates and predictions in cases where simple nudging (i.e. $D_M = 1$) fails, due to insufficient measurements $L < L_s$.

These same issues regarding chaotic instability have also been observed in 4DVar variational calculations Evensen (2008), which are now standard practice in data assimilation. For instance, given our assumption that the model is known precisely, the strategy behind the strongly constrained 4DVar framework involves minimizing the least squares distance, or cost function, between the observed states and the measurements,

$$\sum_{n=0}^{N} \sum_{\ell=1}^{L} [y_\ell(t_n) - x_\ell(t_n)]^2, \tag{11}$$

subject to dynamical constraints imposed by the model Eq. (1).

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics Discussions

In our discussion so far we have taken the model to be precise, without model errors. This assumption can be relaxed however, by incorporating the natural errors in the observations and the uncertainty in the initial model state into a statistical physics path integral. In this formulation, one specifies an 'action' $A_0(\mathbf{X})$ on the path $\mathbf{X}$, which is discretized in time through the observation window $[0,T]$ such that its components are $X_{a:n} = x_a(t_n)$ for $a = \{1,2,\ldots,D\}$ and $n = \{1,\ldots,N\}$. The conditional probability density on the path $P(\mathbf{X}) = \exp[-A_0(\mathbf{X})]$ determines the expected value of all functions of $\mathbf{X}$. Knowing the form of the action allows one to compute conditional expected values of any function of $\mathbf{X}$, including the path of maximum likelihood and its statistical moments, which are used for uncertainty quantification. Instead of the precisely known data and predictions shown here, one has expected state values along with error bars on the estimates and predictions. Our previous work Whartenby et al (2013) is an example of this statistical calculation in the context of shallow water flow.

In practice the action is not known explicitly, and an approximation is constructed from conditional Markov transition probabilities $P(\mathbf{x}(n+1)|\mathbf{x}(n))$ for the state $\mathbf{x}(t_n) \to \mathbf{x}(t_{n+1})$ using a discrete time version of Eq. (1) as well as a rule that quantifies the modification of the probability distribution associated with the transfer of information from noisy measurements Abarbanel (2013). Under standard assumptions, the information transfer term in the action is essentially Eq. (11). The model error term, assuming additive uncorrelated Gaussian errors is proportional to

$$\sum_{n=0}^{N}\sum_{a=1}^{D}\left[\frac{dx_a(t_n)}{dt} - F_a(\mathbf{x}(t_n))\right]\frac{R_f(a)}{2}\left[\frac{dx_a(t_n)}{dt} - F_a(\mathbf{x}(t_n))\right], \tag{12}$$

all expressed in discrete time. The inverse covariance matrix $\mathbf{R}_f$ accounts for the uncertainty associated with the reduced resolution in state space from the deterministic formulation. Using the term Eq. (12) along with the objective function Eq. (11) relaxes the dynamical constraints associated with the equations of motion Eq. (1) to produce a form that is functionally equivalent to weak constrained 4DVar—assuming no information about the initial condition or prior is known. Further details on this derivation and the associated calculations can be obtained from Abarbanel (2013).

Minimizing the action to determine the path is equivalent to the minimization of the synchronization error with model errors. This equivalence can be made more precise by considering the continuous time limit, and writing the action as

$$\begin{aligned} A_0(\mathbf{x}(t),\dot{\mathbf{x}}(t),t) &= \int_{t_0}^{t_f} dt\left\{\frac{R_m(t,\ell)}{2}\sum_{\ell=1}^{L}(\mathbf{x}_\ell(t) - y_\ell(t))^2 + \frac{R_f(a)}{2}\sum_{a=1}^{D}\left(\frac{dx_a(t)}{dt} - F_a(\mathbf{x}(t))\right)^2\right\}, \\ &= \int_{t_0}^{t_f} dt\, L(\mathbf{x}(t),\dot{\mathbf{x}}(t),t). \end{aligned}$$

The extrema of this action are given by the Euler-Lagrange equations

$$R_f(a)\left\{\frac{d^2x_a(t)}{dt^2} + (DF - DF^T)_{ab}\frac{dx_b(t)}{dt}\right\} - \frac{1}{2}\frac{\partial F_c(\mathbf{x}(t))F_c(\mathbf{x}(t))}{\partial x_a(t)} = R_m(\ell,t)\delta_{a,\ell}(x_\ell(t) - y_\ell(t)),$$

where $DF_{ab} = \partial F_a(\mathbf{x})/\partial x_b$. This equation, subject to the boundary conditions on $p_a(t)\delta x_a(t)$ at $t = t_0, t_f$ with

$$p_a(t) = \frac{\partial L(\mathbf{x}(t),\dot{\mathbf{x}}(t),t)}{\partial \dot{x}_a(t)},$$

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

shows that 'nudging' comes in a natural manner in weak 4DVar as a balance between measurement error terms and model error terms.

When $L < L_s$, this process is also plagued by multiple local minima, just as in the deterministic case considered here. The main point of this discussion is that the problem of insufficient measurements appears to be quite general, affecting a broad class of data assimilation techniques: from synchronization and nudging, to strong/weak optimization as well as stochastic methods like particle filters and Markov-Chain Monte Carlo. Whether or not the addition of time delays will benefit these other methods is left for future work.

The framework presented here allows one to directly estimate the minimum number of observations at each measurement time required for accurate predictions, L_s, under the assumption that the model is perfectly accurate. This is important because in practice the actual available number of observations is naturally constrained, and our method provides a way to determine whether the available measurements provide enough information to adequately solve the initialization problem. If this cannot be done, then any attempts at prediction on real data will almost surely fail. On the other hand, when the process succeeds, it increases confidence that predictive failures associated with the assimilation of real data arise from inadequacies in the model. When the model is wrong, as it typically will be in practice, this framework allows us to determine whether the model is at fault, or whether we require more observations. This in turn allows us to focus more on improving the model, without concurrent concerns regarding the estimation procedure. In other words, when predictions fail, our strategy provides a useful diagnostic framework to help determine where to concentrate our efforts: improving the model or collecting more data.

The inclusion of time delays comes of course with an additional computational cost, mainly associated with the integration steps required to construct the time delay vectors and its Jacobian, as well as solving for the perturbation itself. The baseline for comparison is the simple nudging algorithm Eq. (2), which is recovered in the limit $D_M = 1$. Certainly, algorithmic improvements are required in order to reduce this overhead as much as possible. One idea is to reduce the resolution of the model, initialize it with existing measurements, run the assimilation and then interpolate, to recapture the desired resolution for forward prediction.

In summary, the transfer of information from measurements on a chaotic complex system to a quantitative model of the processes in that system, is impeded when the number of measurements at each measurement time is too small. A sufficient number of required measurements L_s can be established by an examination of the model in a twin experiment, but the number of possible measurements L may be such that $L < L_s$. This paper suggests, and explores in a key geophysical model of earth system flow—the nonlinear shallow water equations—how one may extract further information from the time delays of the measurements to overcome the impediments inherited from the dynamical instability.

To this end, we presented an algorithm that assumes no model error, and utilizes control terms to enable the synchronization of data with the model output and thereby estimate the full model state $\mathbf{x}(T)$. This estimate was then used to generate predictions for $t > T$ and the quality of these predictions was used to evaluate the accuracy of this estimate. By adjusting various parameters of the algorithm (e.g. number of time delays, the number of measurements, and the strength of the couplings) it was shown how one may accurately initialize the model state prior to predictions using a sparse subset of observations.

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Discussions
Open Access

In addition, we have demonstrated the capability of our algorithm in application to a model of geophysical shallow water flow, implemented with $D = 3N_{\Delta}^2 = 768$ degrees of freedom. We expect that this formalism will generalize to systems substantially larger than the one presented here, although we do not underestimate the numerical challenges involved in its extension to say, the scale of existing NWP models with $D \approx 10^8 - 10^9$ such as those presented in Cardinali (2013); ECMWF (2013). A crucial issue here regarding memory scalability is the sparseness of the Jacobian matrix $\partial\mathbf{S}(\mathbf{x}(t))/\partial\mathbf{x}(t)$, as storage of an $O(D^2)$ matrix is clearly impractical for very large models. Initial numerical results however, suggest that this matrix does indeed have sparse structure, at least for the simple partial differential equations systems we have examined thus far. It also appears that a similar trick using adjoint equations as what is done in incremental 4DVar may be applicable here. This strategy will be reported elsewhere.

While we do not contend that the time delay construction will be beneficial for all data assimilation techniques, we recognize the issue of insufficient measurements as a critical bottleneck in our current ability to predict the behavior of complex, chaotic systems. Such systems are quite typical in practice, and thus this issue warrants further investigation. Harking back to the introduction, we note that in the report by Cardinali Cardinali (2013) she indicates that 30-40 million daily observations are now available, and that many NWP models comprise 10^8 degrees of freedom. If we may extend the qualitative trends seen here for shallow water flow, in which time delays provides successful predictions with order of only 30% of state variables observed, then there may be a role already now for time delay operations in existing production NWP computations.

Acknowledgements. This work was funded in part under a grant from the US National Science Foundation (PHY-0961153). Partial support from the Department of Energy CSGF program (DE-FG02-97ER25308) for D. Rey is appreciated. Partial support from the MURI Program (N00014-13-1-0205) sponsored by the Office of Naval Research is also acknowledged.

References

Abarbanel, Henry D. I., *Analysis of Observed Chaotic Data*, Springer-New York (1996).

Abarbanel, Henry D. I., D. R. Creveling, R. Farsian, and M. Kostuk, "Dynamical State and Parameter Estimation," *SIAM J. Appl. Dyn. Syst.* **8**, 1341-1381 (2009).

Abarbanel, H. D. I., M. Kostuk, W. G. Whartenby, "Data Assimilation with Regularized Nonlinear Instabilities," *Quarterly Journal of the Royal Meteorological Society* **136**, 769-783 (2010).

Abarbanel, Henry D. I., *Predicting the Future: Completing Models of Observed Complex Systems*, Springer-Verlag, New York (2013).

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

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

Apte, A., M. Hairer, A.M. Stuart and J. Voss, "Sampling The Posterior: An Approach to Non-Gaussian Data Assimilation", *Physica D: Nonlinear Phenomena* **230**, 50-64, (2007). DOI:10.1016/j.physd.2006.06.009.

Bennett, A. F., "Inverse Methods in Physical Oceanography", Cambridge University Press (1992).

Blum, Jacques, Francois-Xavier Le Dimet, and I. Michael Navon. Chapter in *Computational Methods for the Atmosphere and the Oceans*, Volume **14**: Special Volume of Handbook of Numerical Analysis. R. Temam and J. Tribbia, eds. Elsevier Science Ltd, New York. Hardbound, 784 pages, publication date: NOV-2008 ISBN-13: 978-0-444-51893-4 (Philippe G. Ciarlet, Editor). (2009)

Blum, Jacques "Data assimilation for geophysical problems: variational and sequential techniques," Online Presentation from University of Nice Sophia Antipolis, France. jblum@unice.fr (2010).

Cardinali, C., "Data Assimilation: Observation Impact on the Short Range Forecast," ECMWF Lecture Notes, 2013. http://www.ecmwf.int/publications/

Cox, H., "On the Estimation of State Variables and Parameters for Noisy Dynamic Systems," *IEEE Trans. Automatic Control* **9**, 5-12 (1964).

Durand, Fabien, Lionel Gourdeau, Thierry Delcroix and Jacques Verron, "Assimilation of sea surface salinity in a tropical Oceanic General Circulation Model (OGCM): A twin experiment approach, " *Journal of Geophysical Research* **107**, 8004 (2002).

Evensen, Geir, *Data Assimilation, The Ensemble Kalman Filter*, 2^{nd} ed., Springer, (2008). ISBN = 978-3-540-3830.

Friedland, B. and I. Bernstein, "Estimation of the State of a Nonlinear Process in the Presence of Nongaussian Noise and Disturbances," *J. Franklin Inst.* **281**, 455-480 (1966).

Lectures of Lars Isaksen and Stephen English at the European Centre for Medium-Range Weather Forecasts Training Course, June, 2013.

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

Kuznetsov, L., K. Ide, and C. K. R. T. Jones,"A Method for Assimilation of Lagrangian Data", *Mon. Wea. Rev.* **131**, 2247?2260 (2003).

Mañé, R. "On the dimension of the compact invariant sets of certain nonlinear maps". In D. A. Rand and L.-S. Young. Dynamical Systems and Turbulence, Lecture Notes in Mathematics, **898**. Springer-Verlag. 230?242 (1981).

Mariano, A. J., A. Griffa, T. M. Zgkmen, and E. Zambianchi, "Lagrangian Analysis and Predictability of Coastal and Ocean Dynamics", *J. Atmos. Oceanic Technol.* **19**, 1114?1126 (2002).

Molcard, A., L. I. Piterbarg, A. Griffa, T. M. Özgökmen, and A. J. Mariano, "Assimilation of drifter observations for the reconstruction of the Eulerian circulation field", *J. Geophys. Res. Oceans* **593** 2156-2202 (2003).

Pazo, D., A. Carrassi and J. M. Lopez, "Data-assimilation by delay-coordinate nudging," *Quarterly Journal of the Royal Meterological Society* DOI: 10.1002/qj.2732; Accepted Online 30 December 2015

Pecora, L. and T. Carroll, "Synchronization in chaotic systems," *Physical Review Letters* **64**, 821-824 (1990).

Pedlosky, J., *Geophysical Fluid Dynamics*, 2^{nd} Edition, Springer-Verlag, (1987). ISBN = 0-387-90368-2

Piterbarg, L. I., "Optimal estimation of Eulerian velocity field given Lagrangian observations", *Applied Mathematical Modelling* **32**, 2133?2148 (2008).

Press, W. H, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes, The Art of Scientific Computing, 3rd Edition*, Cambridge University Press, (2007). See pp. 136-139

Rey, Daniel, Michael Eldridge, Mark Kostuk, Henry D. I. Abarbanel, Jan Schumann-Bischoff, and Ulrich Parlitz, "Accurate State and Parameter Estimation in Nonlinear Systems with Sparse Observations", *Physics Letters A* **378**, 869-873.(2014).

Rey, Daniel, Michael Eldridge, Uriel Morone, Henry D. I. Abarbanel, Ulrich Parlitz, and Jan Schumann-Bischoff, " Using waveform information in nonlinear data assimilation," Phys. Rev. E 90, 062916 (2014).

Sadourny, R., "The Dynamics of Finite Difference Models of the Shallow Water Equations", *Journal of Atmospheric Sciences* **32**, 680-689 (1975).

Salman, H., L. Kuznetsov, C. K. R. T. Jones, and K. Ide, "A Method for Assimilating Lagrangian Data into a Shallow-Water-Equation Ocean Model," *Mon. Wea. Rev.* **134**, 1081?1101 (2006).

Sauer, T., J. A. Yorke, and M.Casdagli, "Embedology,? *J. Statistical Physics* **65**, 579-616 (1991).

Takens, F., "Detecting strange attractors in turbulence". In D. A. Rand and L.-S. Young. *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, **898**. Springer-Verlag. 366?381 (1981).

Thomson, Richard E. and William J. Emery, *Data Analysis Methods in Physical Oceanography (Third Edition)*, Elsevier B. V. (2014). Chapter 3, "Statistical Methods and Error Handling"

Whartenby, W., J. Quinn, and H. D. I. Abarbanel, "The Number of Required Observations in Data Assimilation for a Shallow Water Flow," *Monthly Weather Review* **141**, 2502-2518, (2013).

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Open Access
Discussions

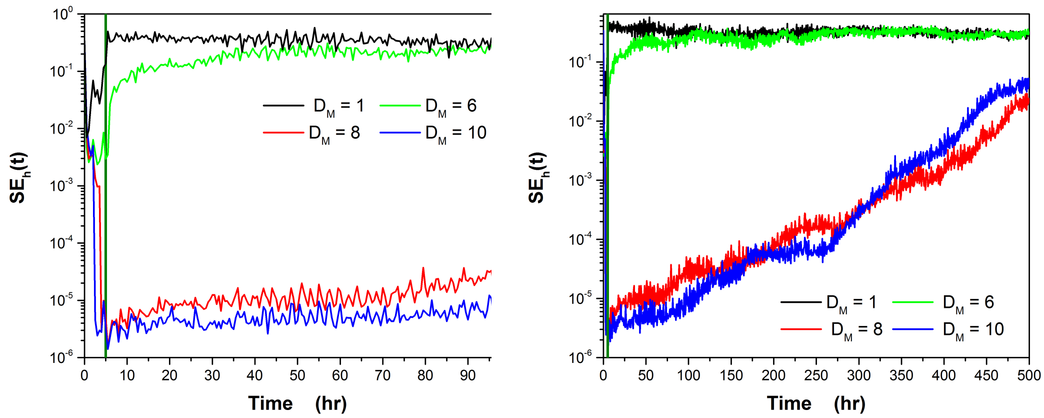

Figure 1. Synchronization error as a function of time, Eq. (9), computed with $D_M = \{1,6,8,10\}$, $g_h\,\Delta t = 1.5$, $g_u\,\Delta t = g_v\,\Delta t = 0.5$ and $\tau = 10\,\Delta t = 0.1\,h$. Assimilation is performed for $t \leq 5$ h. **Left Panel**The couplings are then switched off and trajectories are evolved with the original dynamical equations Eq. (10) until $t = 100\,h$. Note that in the $t \geq 5$ window, the error in the trajectories grow roughly with the largest Lyapunov exponent of the system $\lambda_{max} \approx 1/31h$. Synchronization is evident when $D_M = \{8,10\}$ and not for $D_M = \{1,6\}$. This suggests that accurate predictions will be obtained $D_M = \{8,10\}$. **Right Panel** Same calculation extended to t = 500 h.

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Open Access
Discussions

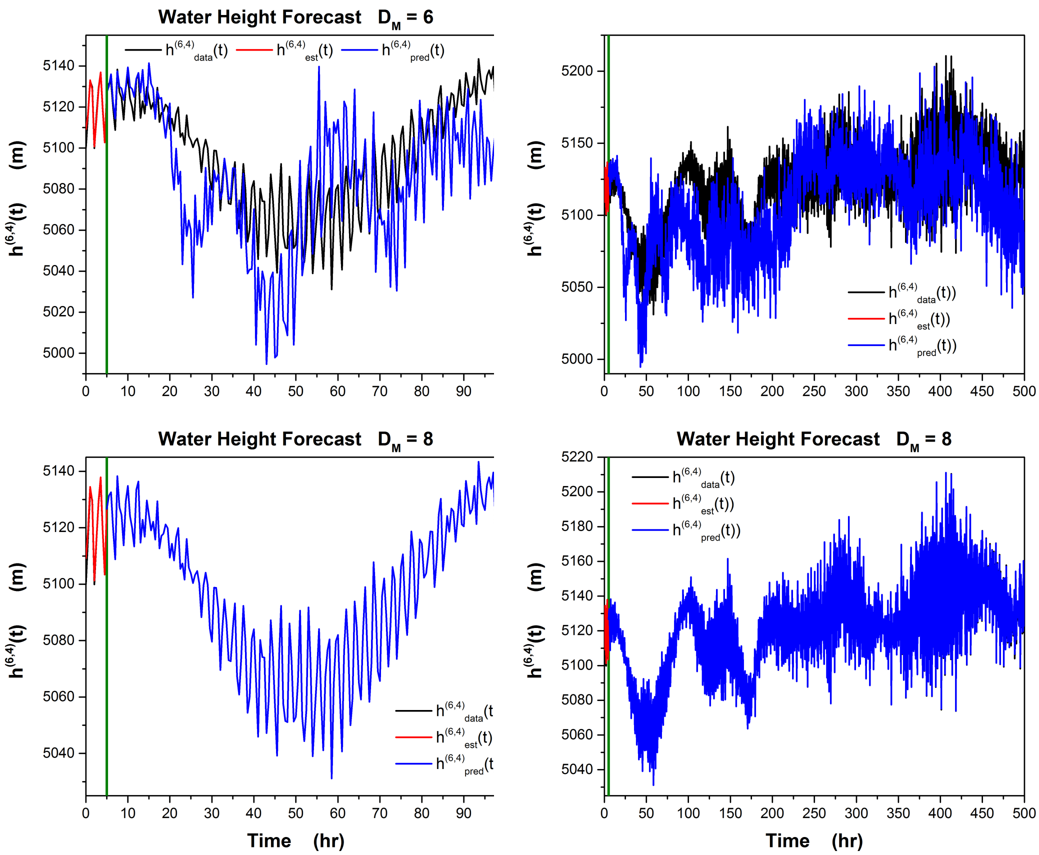

Figure 2. Upper Left Panel Known (black), estimated (red) and predicted (blue) for the observed height values $h^{(6,4)}(t)$ at grid point $(6,4)$ for $D_M = 6$. Observations are for $0 \leq t \leq 5$ h. Predictions are for $5 \leq t \leq 100$ h. In a true experiment, one may evaluate the validity of the model and the estimation procedure in this way, by comparing observed data with additional measurements. In accordance with the synchronization error results. **Upper Right Panel** Same calculation for $D_M = 6$ for a prediction window $5 \leq t \leq 500$ h. **Lower Left Panel** Same calculation except $D_M = 8$. Prediction window is $5 \leq t \leq 100$ h. **Lower Right Panel** Same calculation except $D_M = 8$. Prediction window is $5 \leq t \leq 500$ h.

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

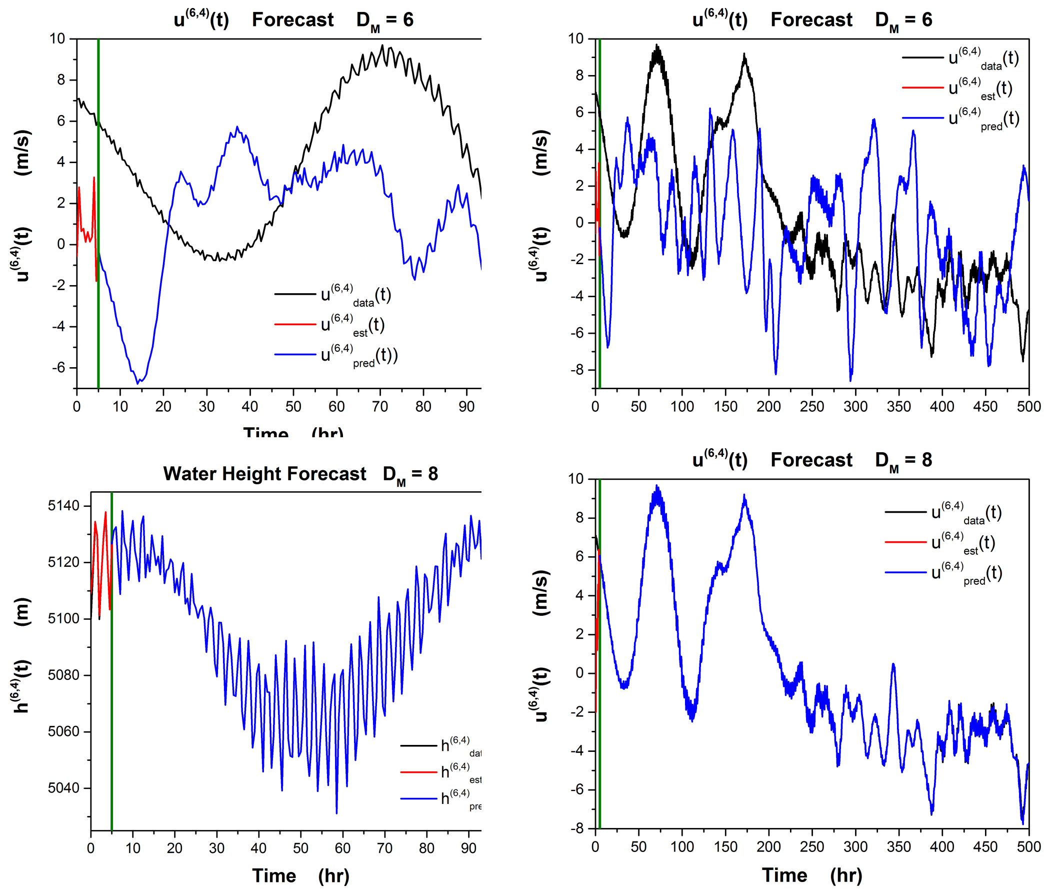

Figure 3. Upper Left Panel Known (black), estimated (red) and predicted (blue) for the observed x-velocity values $u^{(6,4)}(t)$ at grid point $(6,4)$ for $D_M = 6$. Observations are for $0 \le t \le 5$ h. Predictions are for $5 \le t \le 100$ h. In a true experiment, one may evaluate the validity of the model and the estimation procedure in this way, by comparing observed data with additional measurements. In accordance with the synchronization error results. **Upper Right Panel** Same calculation for $D_M = 6$ for a prediction window $5 \le t \le 500$ h. **Lower Left Panel** Same calculation except $D_M = 8$. Prediction window is $5 \le t \le 100$ h. **Lower Right Panel** Same calculation except $D_M = 8$. Prediction window is $5 \le t \le 500$ h.

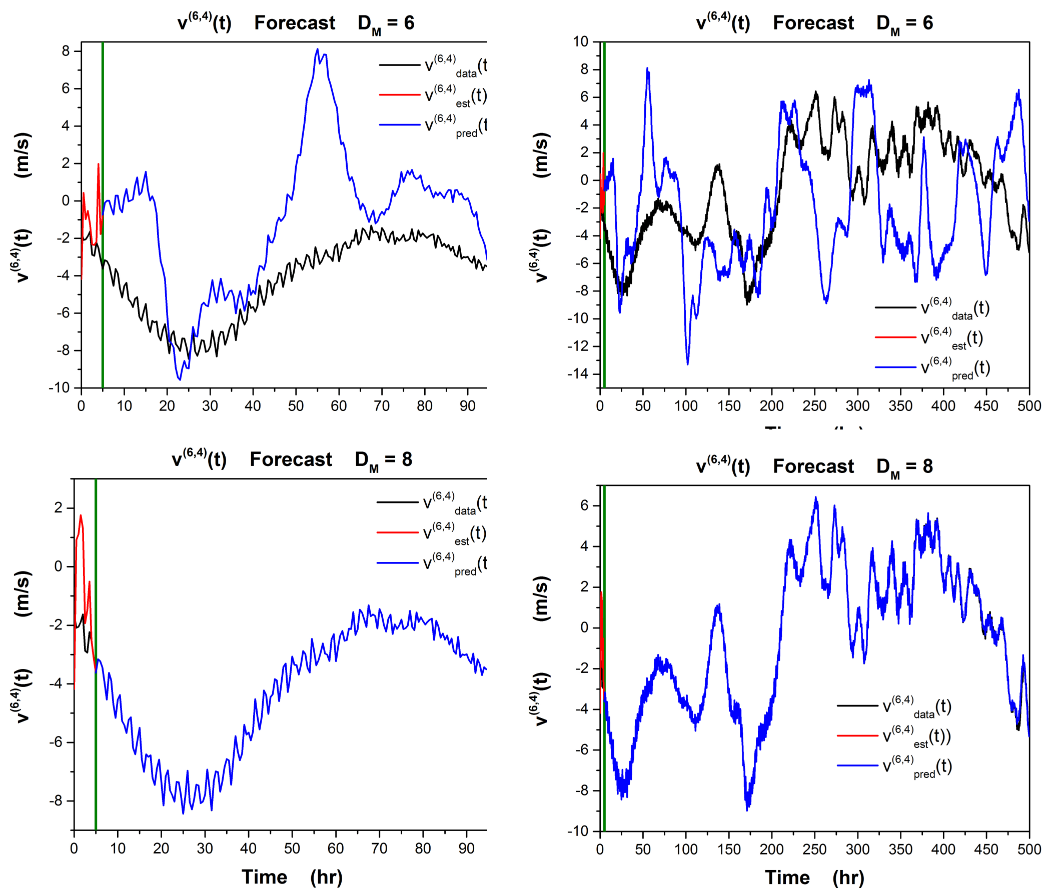

Figure 4. Upper Left Panel Known (black), estimated (red) and predicted (blue) for the observed y-velocity values $v^{(6,4)}(t)$ at grid point $(6,4)$ for $D_M = 6$. Observations are for $0 \le t \le 5$ h. Predictions are for $5 \le t \le 100$ h. In a true experiment, one may evaluate the validity of the model and the estimation procedure in this way, by comparing observed data with additional measurements. In accordance with the synchronization error results. **Upper Right Panel** Same calculation for $D_M = 6$ for a prediction window $5 \le t \le 500$ h. **Lower Left Panel** Same calculation except $D_M = 8$. Prediction window is $5 \le t \le 100$ h. **Lower Right Panel** Same calculation except $D_M = 8$. Prediction window is $5 \le t \le 500$ h.

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Open Access
Discussions

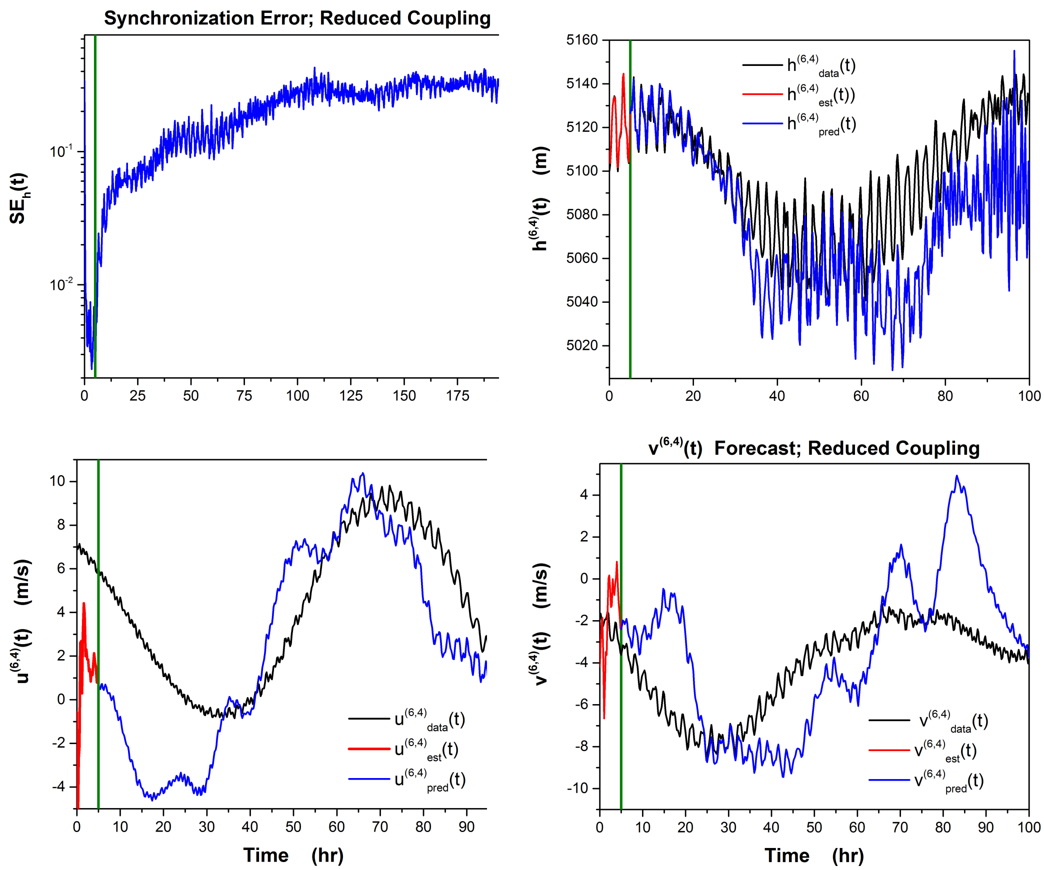

Figure 5. Data assimilation results with $D_M = 10$ and reduced coupling on the height component $h^{(6,4)}(t)$ at location (6,4), $g_h\,\Delta t = g_u\,\Delta t = g_v\,\Delta t = 0.5$. All other parameters are the same. **Upper Left Panel** $SE_h(t)$ for $0 \le t \le 200$ h. **Upper Right Panel** Known (black), estimated (red) and predicted (blue) for the observed height values $h^{(6,4)}(t)$ at grid point $(6,4)$ for $D_M = 10$. Observations are for $0 \le t \le 5$ h. Predictions are for $5 \le t \le 100$ h. **Lower Left Panel** Known (black), estimated (red) and predicted (blue) for the observed x-velocity values $u^{(6,4)}(t)$ at grid point $(6,4)$ for $D_M = 6$. Observations are for $0 \le t \le 5$ h. Predictions are for $5 \le t \le 100$ h. **Lower Right Panel** Known (black), estimated (red) and predicted (blue) for the observed y-velocity values $v^{(6,4)}(t)$ at grid point $(6,4)$ for $D_M = 6$. Observations are for $0 \le t \le 5$ h. Predictions are for $5 \le t \le 100$ h.

Nonlin. Processes Geophys. Discuss., doi:10.5194/npg-2016-22, 2016
Manuscript under review for journal Nonlin. Processes Geophys.
Published: 22 March 2016

Nonlinear Processes in Geophysics
Open Access
Discussions

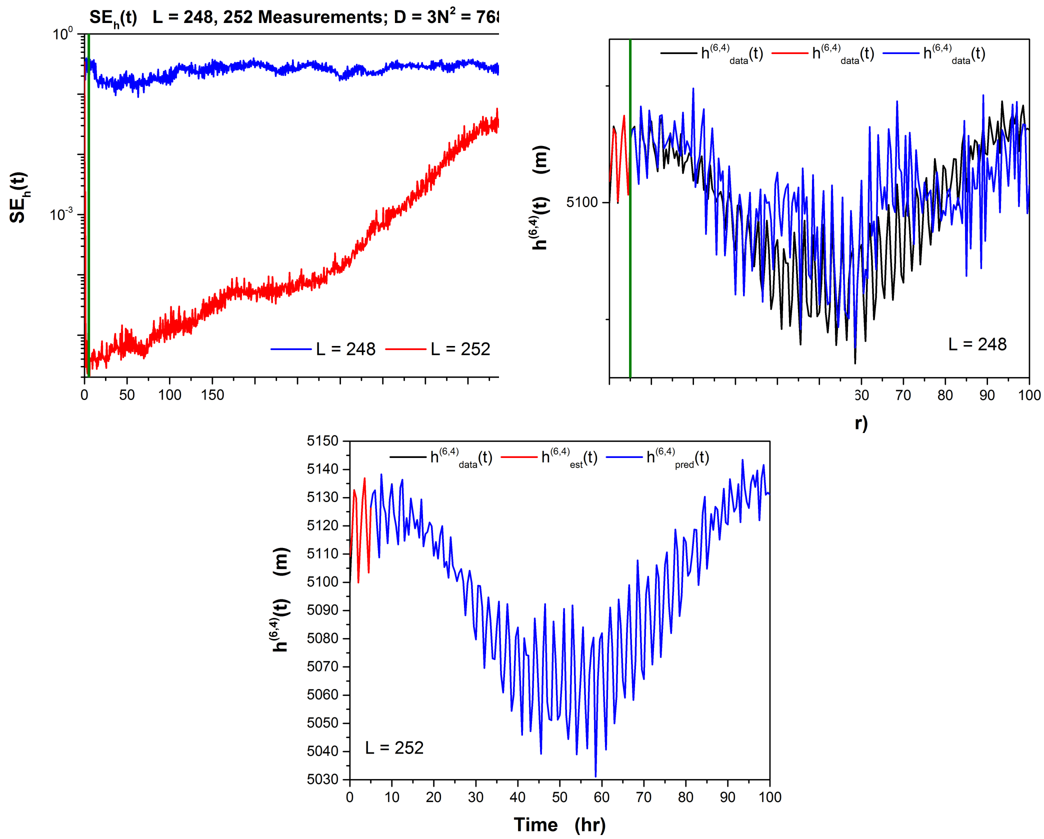

Figure 6. Synchronization error and known, estimated, and predicted height values for L = 248 height measurements at each observation time and for L = 252 height measurements at each observation time. **Upper Left Panel** $SE_h(t)$ for L = 248 and L = 252 over $0 \leq t \leq 5$ h in the observation window, and $5 \leq t \leq 500$ h after the couplings are removed. **Upper Right Panel** Known (black), estimated (red), and predicted (blue) values of the height $h^{(6,4)}(t)$ at gridpoint (6,4) for $0 \leq t \leq 100$ h for L = 248. **Lower Panel** Known (black), estimated (red), and predicted (blue) values of the height $h^{(6,4)}(t)$ at gridpoint (6,4) for $0 \leq t \leq 100$ h for L = 252. This shows the rather sharp transition between bad predictions (L = 248) and good predictions (L = 252).

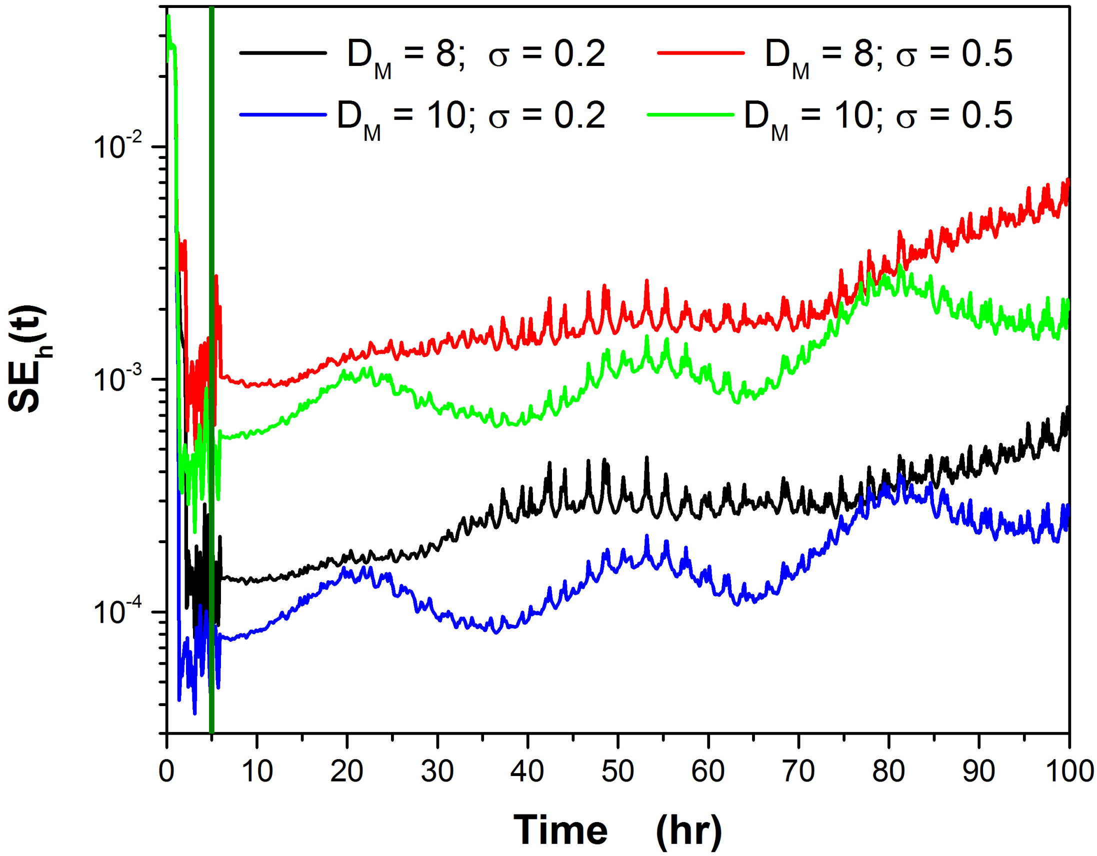

Figure 7. The effect of noise levels in the initial condition for the solution of the model equations Eq. (10) on $SE_h(t)$. We show the results for $D_M = \{8, 10\}$ for added Gaussian noise $N(0, \sigma)$ with $\sigma = \{0.2, 0.5\}$. For this range of noise levels added to the initial condition for generating the data in our twin experiments, we see that the detailed values of $SE_h(t)$ change. In the case of both $D_M = 8$ and $D_M = 10$, $SE_h(t)$ still becomes quite small in the observation window $0 \leq t \leq 5$ h, suggesting that predictions for $t \geq 5$ will remain robustly accurate.

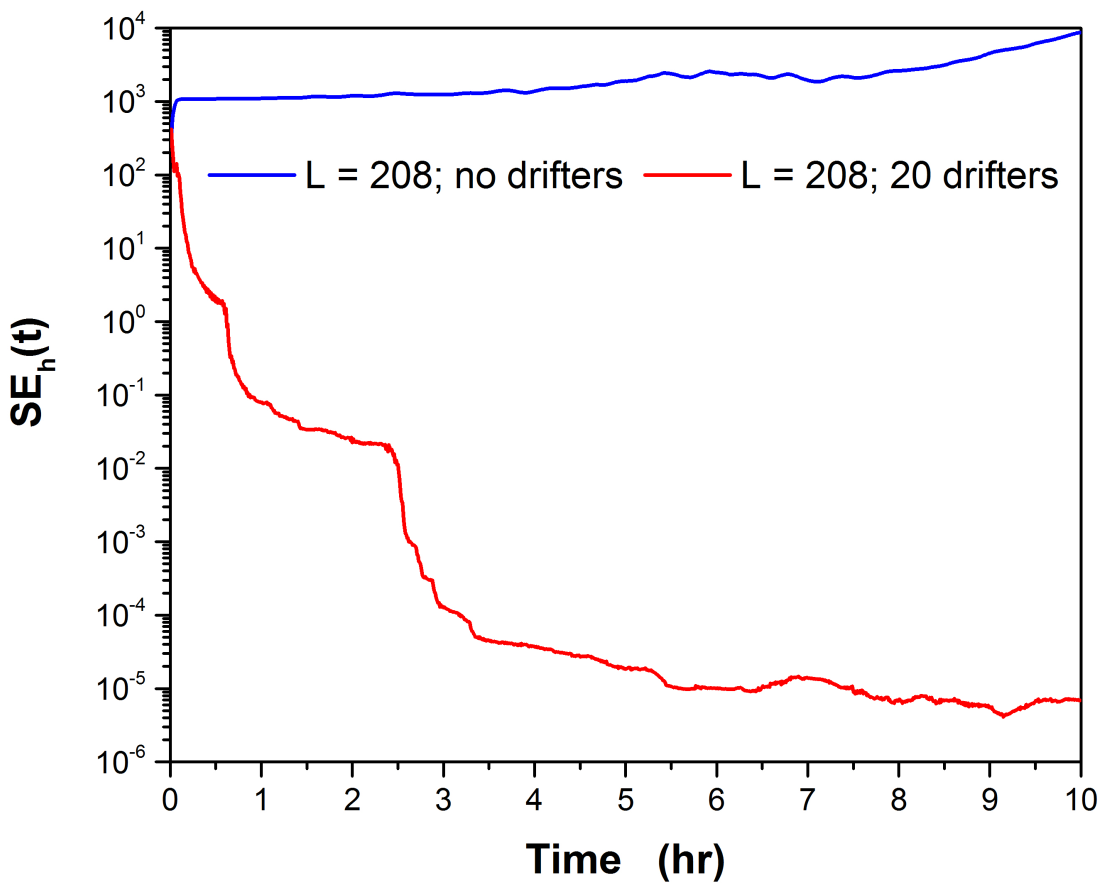

Figure 8. $SE_h(t)$ for our standard twin experiment described in detail earlier when we utilize drifter information, and when we do not utilize drifter information. When the number of observations of height is L = 208, we see that without drifter information (blue line) there is no synchronization and correspondingly inaccurate predictions (not shown). When information from 20 Lagrangian drifters is added during data assimilation using time delay nudging, $SE_h(t)$ decreases very rapidly (red line) indicating predictions will be very accurate (also not shown). The efficacy of small numbers of drifters is clear in this example.