Authors' Response to Referee, Dr. S.G. Penny

First, may we thank the referee for the thorough reading of our paper and for the detailed suggestions of changes and improvements. This document is our first response to your comments, and when the discussion period for the paper is completed, we will incorporate our comments here and responses to all commenters into a revised version of the paper. Again, these remarks have been most helpful.

The Authors: An, Rey, Ye and Abarbanel
=================================

1. Techniques for dealing with a sparse observational networks are critically important, particularly for ocean and climate reanalyses that attempt to reconstruct the past state of the Earth system (e.g. Compo et al., 2011; http://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2.html). The experiment scenarios described here by the authors are perhaps most applicable to the estimation of the global ocean state after the introduction of satellite altimeters, e.g. TOPEX/Poseidon in late 1992 (https://sealevel.jpl.nasa.gov/missions/topex/), with their final set of experiments having a potential application to leverage data from the Global Drifter Program (http://www.aoml.noaa.gov/phod/dac/index.php). Thus from a practical point of view, the time-delay method has potential merit for operational scale data assimilation (DA) and reanalysis.

==============================

We have examined the links you provide, and it does indeed look like fruitful directions for the use of the time delay method. We think it fair to evaluate ourselves critically and recognize that we may not be prepared to tackle, with present personnel levels and computational resources, something of this magnitude. However, we agree about the importance of this problem and appreciate your encouraging comments and suggested applications. There is no doubt in our mind that these items are in our future, and we look forward to pursuing them.

==

2. Because of such potential, the authors should give a bit more explicit description about how these ideas compare to common methods like 4DVar or the 4D Ensemble Kalman Filter (EnKF), both of which utilize observations over an extended time window. The authors could give a more thorough depiction of how their ideas could be incorporated in these existing systems in order to facilitate a higher likelihood that an operational center might adopt the approach.

=====================================

Again, we agree in toto with the referee's comments. We have taken a more cautious path to these comparisons with ExtKF, EnKF, and traditional 4DVar methods [but see our paper in NPG (Improved variational methods in statistical data assimilation J. Ye, N. Kadakia, P. J. Rozdeba, H. D. I. **Abarbanel** and J. C. Quinn Nonlin. Processes Geophys., 22, 205-213, doi:10.5194/npg-22-205-2015, 2015))].

The comparisons require, in our opinion, another paper dedicated to those, and, if we want to be fair about the comparisons, we feel we need to do them in cooperation with colleagues who have experience with those methods. We do have such colleagues at Scripps Institution of Oceanography, and we will be working with them on just these matters.

We considered going into more detail in this paper and decided it might take away from its main point, which was to demonstrate the benefit of using time delays in a simple geophysical model, and its application to drifter measurements.
====================

3. The sea surface height is closely connected to the near surface currents via the geostrophic balance, particularly in midlatitudes. Thus it is expected that unobserved currents would be well constrained by proper estimation of the surface height. For example, sea surface heights and sea surface winds are used to construct an estimate of ocean surface currents for the OSCAR product (http://www.oscar.noaa.gov/index.html). However, the examples given by the authors could perhaps be described as a supplement for the tropical region where this relationship breaks down. For future work, a natural extension would be to address a slightly more sophisticated example consisting of multiple vertical layers and the modeling of the temperature and salinity components of the density. This experiment would give a better test of estimating unobserved variables. For example, observing only temperature while estimating salinity is a challenging problem for ocean reanalyses before the Argo era.

   ==============

Thank you for these suggestions. We will look closely at how the geostrophic balance plays into directing the dynamical outcome of our use of time-delays. However, as the geostrophic wind is related to the gradient of the height variable, it may be that this provides a different general constraint on the solutions to the shallow water equations.
================================

4. A brief statement could be made about the applicability of the time-delay approach, for example, to the tropical observing system of moored buoys (TAO/TRITON). These are stationary sensors generating data about once every 10 minutes, but the majority of this data is not used in DA because most global scale ocean assimilation systems use analysis cycles that span multiple days. Even a coupled ocean/atmosphere DA system cycling every 6 hours could benefit from better use of this data. I suggest investigating the TPOS-2020 (Tropical Pacific Observing System) effort for the potential to inform the future development of this and other observing systems (http://tpos2020.org). A weakness in the chosen experiments scenarios that should be acknowledged is that the approach has not been tested on time-delay observations with errors that are correlated in the time dimension. This is particularly important in ocean DA because errors of representativeness often dominate (versus instrument errors).

================================

This is an excellent suggestion. It would appear to provide information from an unused (by us, and it appears many others) data source for useful information about ocean models.

==============================

5. I suggest an experiment, perhaps for future work, in which you run 2 model resolutions. The high resolution run is treated as 'truth', from which observations are drawn. The low resolution model is what you are synchronizing via DA. Set up appropriately, this should give you 'natural' errors of representativeness in the observations that may be correlated in time with the errors of future or past observations. Does the time-delay method still work effectively in this experiment scenario?

   ==============================

This is another good idea. We actually considered including such an experiment, to investigate the impact of finite resolution and model errors arising from subgrid scale processes. Ultimately, we decided to leave these considerations for a future paper, and focus here on the perfect model scenario. We see no impediment to the use of time delays in this scenario; indeed, it may provide information from "spatial delays" (also used in the past for nonlinear dynamical descriptions of waves propagating in nonlinear materials) presently not incorporated in our own work.

====================================

6. The time-delay method is described in comparison to nudging as a baseline. I would like to see the authors compare a simple 4DVar to the time-delay method as well (via experiment) to give context into how their method compares to a more state-of-the-art DA.

A thorough comparison is planned for a future paper, where we discuss in detail the connection between our method and 4DVar. The revised paper we will prepare notes this as future work. The simple answer at this time is that we know how to introduce time delays into what we call the action, often called the 4DVar cost function, and we have not <u>yet</u> used this augmented cost function (and our method of 4DVar as noted above) on this problem.

7. It seems that the time-delay information for the observations and model state applied with what is essentially a diagonal coupling term emulates a similar effect as the cross-covariances that would in effect apply a non-diagonal coupling term to the innovations computed at different times throughout the window. The authors should discuss how the off-diagonal coupling used in most operational DA relates to the diagonal coupling with time-delay observations used here.

   ==============================

We agree with your statements here about the cross-correlations. The off-diagonal terms here arise from the generalized inverse of the time delayed innovations. The diagonal coupling term in time delay space could for instance damp the effect of measurements further in the future, which have more uncertainty due to dynamical instability.

A similar effect could be achieved from 4DVar with a uniform prior and a time distributed observation error matrix, but we would rather discuss this in a future paper that more thoroughly explores the connection between time delayed nudging and 4DVar.

====================================

8. The impact of observation error on synchronization via the nudging approach is not addressed very thoroughly. I'd like to see some evaluation of the sensitivity to observation error in the assessment of the method. The authors should describe how their method is impacted by outliers in the observed data. Is the method sensitive to such outliers? I'd like to see an example.
   =========================

When observation error is present, the model will synchronize to within the noise ball of the `true' solution, when the model is known perfectly and enough observations are present. We recognize however that for many DA methods the goal is to reduce the RMSE below the noise level, but this was not the case here as we chose to consider the sparsity of observations as the dominant effect, rather than observational noise. As a result, we elected to only include a brief investigation, to show that our method is not significantly impacted by very small observational errors.

To be clear though, you are right that enough noise will 'break' this method, or at least severely impede its chances of success. The degree of regularization needed for the generalized inverse of dS/dx is commensurate with the observational errors of the system.

In addition to these remarks, we use the synchronization error as our "monitor" of the reduction of the model output error to indicate when we have sufficient observations at each measurement time. These errors are limited by the noise in the observations.

General Technical Corrections:

We do not comment on these, really valuable—to us—comments. We have addressed each of them in our rewrite of the submitted paper, and on revision after the end of the NPG discussion period, we will note each change we have made built upon these detailed, and appreciated, comments. Thank you.

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 problem of forecasting the behavior of a complex dynamical system through analysis of observational time-series data becomes difficult when the system expresses chaotic behavior and the measurements are sparse, in both space and/or time. Despite the fact that this situation is quite typical across many fields, including numerical weather prediction, the issue of whether the available observations are 'sufficient' for generating successful forecasts is still not well-understood. An analysis by Whartenby et al (2013) found that in the context of the nonlinear shallow water equations on a β-plane, standard nudging techniques require observing approximately 70% of the full set of state variables. Here we examine the same system using a method introduced by Rey et al (2014a), which generalizes standard nudging methods to utilize time delayed measurements. We show that in certain circumstances, it provides a sizable reduction in the number of observations required to construct accurate estimates and high-quality predictions. In particular, we find that this estimate of 70% can be reduced to about 33% using time delays, and even further if Lagrangian drifter locations are also used as measurements..

1 Introduction

The ability to forecast the complex behavior of global circulation in the coupled Earth 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 at the end of the analysis or observation window. When the model is chaotic, even if it is known precisely, the accuracy of the prediction depends on the accuracy of the initial state estimate. This is due to sensitive dependence to the initial conditions, which was first identified by Lorenz (1963).

Here we consider an idealized situation where a perfect dynamical model describes the deterministic time evolution of a set of D state variables. We also assume a set of L measurements are taken at each observation time from the physical system at a uniform sampling interval Δt, which is assumed to be small relative to the time scale of the dynamics.

Our main concern here is the case where the measurements are sparse in state space, so $(L \ll D)$. Although this discussion will focus solely on a specific geophysical system (the shallow water equations), the methods we describe here have broad applicability across the quantitative study of the underlying physical or biological properties appearing in many complex systems. In particular, the situation of high dimensional dynamics and sparse 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, among many other examples (Abarbanel (2013)).

As discussed by Cardinali (2013), operational NWP models at the European Centre for Medium-Range Weather Forecasting (ECMWF) now contain upwards of 10^8 degrees of freedom. These models are analyzed using $3-4\times10^7$ daily observations, a large portion of which are often discarded. Given the scale of these calculations, the question of whether the remaining observations are in fact 'sufficient' for producing accurate analyses and forecasts is of the utmost importance.

To clarify the term 'sufficient' we refer to the analysis by Whartenby et al (2013), which showed that familiar nudging methods, when applied to a chaotic, shallow water flow on a β-plane driven by Ekman pumping, require observation of roughly 70% of the $3N_\Delta^2$ dynamical variables. That is, to achieve accurate forecasts, these methods required measuring the height variable h and at least one of the two velocity variable u, v at each of the $N_\Delta\times N_\Delta$ grid points. Additionally, the prediction accuracy was shown to drop precipitously when the number of observations L drops below a critical threshold L_s, which was identified as the number of 'required' observations.

The existence of this critical threshold, despite the otherwise ideal circumstances, raises a number of questions. Most notably this one: what can be done if the number of observations L is constrained to be less than L_s? In examining this question, it is worth considering that the value L_s depends on a number of factors, including the chosen data assimilation algorithm. This fact suggests that one might be able to effectively reduce this threshold by modifying the algorithm in a way that more efficiently utilizes the information in the available observations.

This paper will investigate this idea using the method introduced by Rey et al (2014a, b). This modifies a standard nudging technique to include additional information in the *time delays* of the observations. In particular, we will show that by using time delays the estimate of 70% given by Whartenby et al (2013) can be reduced to roughly 33%, and can be even further reduced if positional observations from Lagrangian drifters are also used. These outcomes suggest that time delays may be useful for reducing the number of required observations to meet the practical constraints of operational NWP.

We now briefly describe the concept of time delayed nudging. Further details can be found in Rey et al (2014a, b).

2 Time delayed nudging

The system is assumed to be described by a mathematical model, whose state is given by a D-dimensional vector $\mathbf{x}(t)$. The model defines a dynamical rule for evolving the $\mathbf{x}(t)$ in time, which we assume can be represented as a set of ordinary differential equations (ODEs)

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(\mathbf{x}(t), t). \tag{1}$$

If the dynamics of the system are described by partial differential equations (PDEs), such as with fluids in an earth systems model, these ODEs may be realized by discretizing the PDEs on a grid.

Measurements of the physical system are recorded during an observation window $0 \le t \le T = N\,\Delta t$, where L observations $\mathbf{y}(t_n)$ are taken at each time $t_n = n\,\Delta t$ for $n = 0, 1, \ldots, N$. The measurements $\mathbf{y}(t)$ are related to the state vector $\mathbf{x}(t)$ through a measurement operator, which for simplicity we take here to be a constant $L \times D$ projection matrix $\mathbf{H}$, so that $\mathbf{y}(t) = \mathbf{H} \cdot \mathbf{x}(t) + \text{noise}$. This assumption is not imperative however.

The overall objective is to estimate the full model state $\mathbf{x}(T)$ at the end of the assimilation window using information from observations, 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 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.

When 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),

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(\mathbf{x}(t),t) + \mathbf{H}^{\dagger} \cdot \mathbf{G}(t) \cdot \big(\mathbf{y}(t) - \mathbf{H} \cdot \mathbf{x}(t)\big). \tag{2}$$

where $\mathbf{H}^{\dagger}$ denotes the transpose, and $\mathbf{G}(t)$ is an $L \times L$ matrix that is nonzero only at times t_n where measurements occur. For simplicity, when $\mathbf{G}(t)$ is non-zero, it is assumed to be constant and diagonal, so the coupling terms only affect the measured states.

This long-standing procedure, known as 'nudging' in the geophysics and meteorology literature, has been shown to fail when the number of measurements at a given time is smaller than a critical value L_s (Abarbanel et al (2009)). This can be understood by noting that the coupling term perturbs the observed model states, driving them towards the data. With enough observations L, and a sufficiently strong coupling $\mathbf{G}(t)$, this control term alters the Jacobian of the dynamical system Eq. (2) so that all its (conditional) Lyapunov exponents are negative — see e.g., Pecora & Carroll (1990); Abarbanel (1996); Kantz & Schreiber (2004). That is, the log of the maximum eigenvalue of the matrix $[\mathbf{\Phi}(T,t_0)^{\dagger} \cdot \mathbf{\Phi}(T,t_0)]^{1/2T}$ is negative, where $\mathbf{\Phi}(t,t')$ is the solution to the variational equation

$$\frac{d\mathbf{\Phi}(t,t')}{dt} = \mathbf{D\tilde{F}}(\mathbf{x}(t),t) \cdot \mathbf{\Phi}(t,t') \qquad \Phi_{ab}(t,t') = \delta_{ab}, \tag{3}$$

along the trajectory given by Eq. (2) and

$$\mathbf{D\tilde{F}} = \mathbf{DF}(\mathbf{x}(t),t) - \mathbf{H}^{\dagger} \cdot \mathbf{G}(t) \cdot \mathbf{H}$$

is its Jacobian. That is, $DF_{ab}(\mathbf{x}(t),t) = \partial F_a(\mathbf{x}(t),t)/\partial x_b(t)$ and δ_{ab} is the Kronecker delta, so $\Phi(t',t')$ is an identity matrix. This establishes a necessary condition on L required to synchronize the model with the data. Numerical experiments have shown that when this condition is not met, estimates are not accurate and predictions are unreliable (Abarbanel et al (2009); Abarbanel (2013)). An example of this will be given later in our discussion of geophysical shallow water flow.

It is therefore important to understand for a given problem, whether $L > L_s$. If this condition is not satisfied and additional measurements cannot be made, then we must find another means to overcome this deficit in L.

One way to proceed involves the recognition that additional information resides in the temporal derivatives of the observations. In practice, however, this derivative information cannot be measured directly, although it can be approximated via finite differences, for instance by approximating $d\mathbf{y}(t_n)/dt$ 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 noise in the measurements.

Alternatively, it has been known for some time in the nonlinear dynamics literature that this additional information in the derivative is also available in the *time delay* of the measurements, $\mathbf{y}(t_n+\tau)$. This process can be repeated as many times as needed to form a D_M dimensional vector of time delays, which we call $\mathbf{S}(t)$.

This idea provides the basis for the well-established 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 (see e.g., Aeyels (1981a, b); Mañé (1981); Sauer et al (1991); Takens (1981); Kantz & Schreiber (2004); Abarbanel (1996)). By mapping to a proxy space of time delayed observations, one is able invert the projection associated with measuring $L<D$ components of the underlying dynamics, by using fact that new information beyond $\mathbf{y}(t_n)$ lies in $\mathbf{y}(t_n+\tau)$. The derivative operation is just another (albeit less numerically robust) way of accessing this information.

Note that the *time delay* τ and the *embedding dimension* D_M are parameters that need to be chosen appropriately for the system, although a number of useful heuristics are available (Abarbanel (1996)). Moreover, Takens (1981) proved that that taking $D_M > 2\,D_A$, where D_A is the fractal dimension of the attractor, is sufficient to unambiguously reconstruct the topology of the attractor. It is worth noting however that this condition is only sufficient, and the procedure often succeeds with considerably a smaller value of D_M.

In the estimation context, the time delays are used in a slightly different way. Instead of reconstructing the topology of the attractor, they are used to control local instabilities in the dynamics, which cause errors in the analysis to grow. In other words, D_M does not need to embed the entire space. Rather, it only needs to be large enough to effectively increase the amount of information transferred from the L measurements to a value above the critical threshold, L_s.

Using this idea Rey et al (2014a, b) proposed a technique to extract additional information from time delayed observations by constructing an extended state space $\mathbf{S}(t)$, created from an $L\cdot D_M$ dimensional vector of the measurements and its time delays. The components of this time delayed observation vector are denoted by

$$\mathbf{Y}^{\dagger}(t_n) = \{\mathbf{y}^{\dagger}(t_n), \mathbf{y}^{\dagger}(t_n+\tau), \ldots, \mathbf{y}^{\dagger}(t_n+\tau\,(D_m-1))\}, \tag{4}$$

where D_M is the dimension of the time delayed vector $\mathbf{Y}(t_n)$, and τ is the delay, which here is assumed to be an integer multiple of Δt. The corresponding time delay model vectors $\mathbf{S}(\mathbf{x}(t))$ are given by

$$\mathbf{S}^{\dagger}(\mathbf{x}(t)) = \{[\mathbf{H}\cdot\mathbf{x}(t)]^{\dagger}, [\mathbf{H}\cdot\mathbf{x}(t+\tau)]^{\dagger}, \ldots, [\mathbf{H}\cdot\mathbf{x}(t+\tau\,(D_m-1))]^{\dagger}\}, \tag{5}$$

where the values $\mathbf{x}(t'>t)$ are constructed by integrating the *uncoupled* dynamics, Eq. (1), forward in time. The time evolution for $\mathbf{S}(\mathbf{x}(t))$ is given by the chain rule,

$$\frac{d\mathbf{S}(\mathbf{x}(t))}{dt} = \mathbf{DS}(\mathbf{x}(t))\cdot\mathbf{F}(\mathbf{x}(t),t), \tag{6}$$

where the Jacobian $\mathbf{DS}(\mathbf{x}(t)) = \partial\mathbf{S}(\mathbf{x}(t))/\partial\mathbf{x}(t)$ with respect to $\mathbf{x}(t)$ can be computed using the variational Eq. (3), by substituting the Jacobian of the uncoupled model $\mathbf{D}\tilde{\mathbf{F}} \to \mathbf{DF}$. Furthermore, in analogy with Eq. (2), we introduce a control term

$\mathbf{g}(t)$ in time delay space

$$\frac{d\mathbf{S}(\mathbf{x}(t))}{dt} = \mathbf{DS}(\mathbf{x}(t)) \cdot \mathbf{F}(\mathbf{x}(t),t) + \mathbf{g}(t) \cdot \big(\mathbf{Y}(t) - \mathbf{S}(\mathbf{x}(t))\big). \tag{7}$$

We then transform back to physical space, by multiplying both sides of this equation by $[\mathbf{DS}(\mathbf{x}(t))]^{-1}$, to get

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(\mathbf{x}(t),t) + \mathbf{G}(t) \cdot [\mathbf{DS}(\mathbf{x}(t))]^{-1} \cdot \mathbf{g}(t) \cdot \big(\mathbf{Y}(t) - \mathbf{S}(\mathbf{x}(t))\big), \tag{8}$$

Note there are now two control terms, $\mathbf{G}(t)$ and $\mathbf{g}(t)$, which act in physical and in time delay space respectively. Also, since $\mathbf{DS}(\mathbf{x}(t))$ is a $(L \cdot D_m) \times D$ matrix, it is generally not square so its pseudoinverse $[\mathbf{DS}(\mathbf{x}(t))]^{+}$ is used.

At each step of the integration of the controlled (nudged) dynamical equations Eq. (8), the control term perturbs the full state vector in time delay space $\mathbf{S}(\mathbf{x}(t))$ toward the time delay measurement vector $\mathbf{Y}(t)$, allowing it to extract additional information from the waveform of the *existing measurements*. The value of this statement will become more clear later on.

Furthermore, in the limit $D_M = 1$ the time delay formulation Eq. (8) reduces to the standard nudging control Eq. (2). Two important differences however are realized when $D_M > 1$. First, information from the time delays of the observations is presented to the physical model equations. And second, all components of the model state $\mathbf{x}(t)$ are influenced by the control term, not just the observed components. This, for example, allows fixed parameters $\mathbf{p}$ of the model may be estimated as a natural result of the synchronization process by including them as additional state variables, satisfying $d\mathbf{p}(t)/dt = 0$.

Also worth mentioning is that here we are not using time 'delays' but rather a *time advanced* formulation, which looks forward in time. The reason for this is related to the goal of controlling the propagation of errors on the unstable manifold as the system is integrated forward in time, which are locally described by Eq. (3) so the time advanced construction is a natural choice, although both formulations are acceptable. This also brings up a concern regarding what to do at the end of the assimilation window. One option is to switch to a time delayed formulation, or perhaps a mixed formulation that uses delays both forwards and backwards in time. Though comparing the performance of various choices would likely be interesting, we do not consider this issue further. Rather, our numerical experiments use only a time advanced formulation, by choosing the end of the observation window so that the last observation $\mathbf{y}(T + \tau(D_M - 1))$ is always available.

There appear to be considerable similarities between this method and those of strong constraint 4DVar (Lewis et al, (1985); Talagrand et al (1987)), which are now standard practice in data assimilation Rabier et al (2000). Although in this form, time delay nudging certainly cannot handle a system of the size used in operational NWP, as the variational equation requires manipulating $D \times D$ matrices, it may be possible to avoid this issue, for instance by using adjoints, similar to what is done in practice (Courtier et al (1994)). This formulation will be given in a subsequent paper.

Pazo (2016) gave a simplified version of time delay nudging that requires considerably less computation. Although their approach has not yet been applied to geophysical flows, it is worth investigating whether it is also capable of achieving the same reduction in L_s.

Furthermore, while it known that chaotic behavior in the model can cause serious issues for strong constraint 4DVar (Pires et al (1996)), perhaps less well-known is that similar observability thresholds have also been observed in both nudging and 4DVar (Abarbanel et al (2009); Abarbanel (2013)), even with weak constraints (Quinn & Abarbanel (2010)). In fact, the value of L_s appears remarkably consistent across a variety of formulations, which it may be a rather fundamental quantity. It is also evident that Kalman related methods can do better if the observation operator is allowed to adapt to the unstable subspace (Law et al (2014)). This is the related to the fact that the Riccati equation for the error covariance propagation targets the unstable subspace (Trevisan & Palatella (2011); Gurumoorthy et al (2015)). Related ideas have appeared in the literature on assimilation in the unstable subspace (Trevisan et al (2010); Palatella et al (2013)).

We are currently working on unifying the motivating ideas behind time delay nudging with the variational action principle of weak constraint 4DVar. This and other related connections to 4DVar will be given in a subsequent paper that will also compare time delay method with a few other common data assimilation techniques.

For the moment however, no additional theory will be introduced. Instead, we focus its application to a core geophysical model. Namely, the shallow water equations.

3 Twin Experiments

We test our time delay nudging procedure through a series *twin experiments* (Durand et al (2002); Blum et al (2009); Blum (2010)). After solving the original dynamical equations Eq. (1) forward from a preselected 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.

To simulate the conditions of a *true experiment* we monitor our progress by calculating the observable synchronization error, namely the root mean square deviation between the data and the observed model states,

$$SE(t_n) = \sqrt{\frac{1}{L}\left|\mathbf{H}\cdot\mathbf{x}^s(t_n) - \mathbf{y}^s(t_n)\right|^2}, \tag{9}$$

where scaled variables have been introduced such that $x^s_\ell(t) = [x_\ell(t) - x^{min}_\ell(t)]/[x^{max}_\ell(t) - x^{min}_\ell(t)]$ 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 roughly equal. While this could make the result sensitive to outliers in the data, it did not appear to be an issue here.

When the estimation is completed 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 $\mathbf{y}(t > T)$. Comparing against the forecast provides confidence that the unobserved state variables are also accurately estimated.

It was previously shown by Whartenby et al (2013) that when the synchronization error Eq. (9) decreases to very small values, the full state $\mathbf{x}(T)$ is accurately estimated and the forecast is quite good. Conversely, when this fails to occur, the full state $\mathbf{x}(T)$ is not well estimated and the prediction is unreliable.

In Rey et al (2014a, b), this contraction of the synchronization error was only observed when the number of time delayed observations $L \times D_M$, and the magnitude of the coupling matrices $\mathbf{g}(t)$, $\mathbf{G}(t)$ were 'large enough'. The precise meaning of this statement will become apparent shortly.

4 Nonlinear Shallow Water Equations

We now describe the application of time delay nudging to a nonlinear 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, operational models contain considerably more detail than this example, and those models also describe the dynamics over a sphere. However, we suspect the results presented here for this simplified model will be applicable to more complex models as well, though we do not underestimate the numerical challenges in this extrapolation.

As the depth of the coupled atmosphere ocean fluid layer ($10-15$ km) is markedly less than the earth's radius (6400 km), the shallow water equations for two dimensional flow provide a good approximation to the fluid dynamics of the ocean. 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 following dynamical equations with $\mathbf{u}(\mathbf{r},t)=\{u(\mathbf{r},t), v(\mathbf{r},t)\}$,

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

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 estimated to be $\lambda_{max}=0.0325/h\approx 1/31\,h$, measuring the average growth rate of random perturbations.

We have analyzed this flow using the enstrophy conserving discretization scheme given by Sadourny (1975) on a grid of size N_Δ^2 for $N_\Delta=\{16,32,64\}$ with periodic boundary conditions. Using the twin-experiment framework, with simple nudging given in Eq. (2) and a static observation operator, approximately 70% of the $D=3N_\Delta^2$ degrees of freedom must be observed in order to synchronize the model output with the data (Whartenby et al (2013)). 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 $D=3N_\Delta^2=768$. For this case, Whartenby et al (2013) estimated $L_s\approx 524=0.68\,D$ using a uniform grid of observations. In other words, the height field and one of the velocity fields at each grid point needed to be observed to achieve reliable results.

Based on the discussion above and the lectures by Cardinali (2013), we see that the requirement of having to observe 70% of the model dynamical variables exceeds the measurements now available by at least a factor of two. This requirement is expected to be higher in practice, when the model and observations contain substantial errors. Furthermore, it is worth investigating whether the number of required observations eventually stabilizes to some finite value as the model resolution increases, but this is left for a future investigation.

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.

5 Results with Time Delay Nudging for the Shallow Water Equations

We now demonstrate that the time delay method is capable of reducing L_s, by showing that it can construct successful estimates and predictions without directly observing the horizontal velocity fields. This strategy was shown to fail by Whartenby et al (2013) with static ($D_M = 1$) nudging. Thus, we assume height measurements alone are made at each grid point (i,j) for $i,j = \{1,2,\ldots,16 = N_\Delta\}$, so $L = 256 < 524 \approx L_s$, as estimated by Whartenby et al (2013).

The initial state $\mathbf{x}(t_0)$ for the model and the data are taken to have the form,

$$h^{(i,j)}(t_0) = \left(\frac{\pi A_0}{N_\Delta \Delta Y}\right)^2 \left[\cos(\omega_\phi \phi(\mathbf{r}^{(i,j)}) + \delta_\phi) + \cos(\omega_\theta \theta(\mathbf{r}^{(i,j)}) + \delta_\theta)\right] + H_0$$
$$u^{(i,j)}(t_0) = A_0 \frac{\partial \psi(\mathbf{r}^{(i,j)})}{\partial y} \qquad v^{(i,j)}(t_0) = -A_0 \frac{\partial \psi(\mathbf{r}^{(i,j)})}{\partial x} \tag{11}$$

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

$$\psi(\mathbf{r}) = \cos(\omega'_\phi \phi(\mathbf{r}) + \delta'_\phi) \sin(\omega'_\theta \theta(\mathbf{r}) + \delta'_\theta). \tag{12}$$

The functions ϕ and θ respectively evaluate the latitude and longitude at the point $\mathbf{r}^{(i,j)}$ on the grid. All fields as well as the variables $\{x,y,t\}$ were scaled by the values in Table (1), to make them dimensionless. The parameters $\omega_\phi, \omega_\theta, \omega'_\phi, \omega'_\theta$ and $\delta_\phi, \delta_\theta, \delta'_\phi, \delta'_\theta$ are chosen arbitrarily, so that the phase and period of the initial condition are different for truth and the estimate. Also, although the method is capable of estimating the static model parameters, here they are considered 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 $\Delta t = 0.01\,h = 36\,s$. The values of G_h are larger than G_u, G_v, since the average height is $5000 \pm 30\,m$, three orders of magnitude higher than the average velocity $0 \pm 5\,m/s$. 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 temporal precision throughout the observation window.

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 $\mathbf{S}(\mathbf{x}(t))$. The first minimum of the average mutual information was also calculated to be $\tau \approx 30\,\Delta t$, which is reasonably close to our choice, and the results did not change if its value was shifted by a few Δt.

5.1 Choosing D_M

The state was estimated by integrating the coupled differential equations Eq. (8) from $t = 0$ to $T = 5\,h = 500\,\Delta t$ with various $D_M = \{1, 6, 8, 10\}$. The coupling terms were then switched off at $t = T$ 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 window. It then subsequently rises as $e^{\lambda_{max}(t-T)}$, where $\lambda_{max} \approx 1/31h$ 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, by comparing predictions on the observed heights with additional data. 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 $D_M = 6$.

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. 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.

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 Whartenby et al (2013).

5.2 Reducing the coupling strength

In the previous discussion it was suggested 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 $SE(t)$, shown in Figure 5, Upper Left Panel, stabilizes to a level three orders of magnitude higher than was achieved with $G_h\,\Delta t = 1.5$, suggesting failure. 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 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 required, although we have not developed a systematic way of choosing these values. 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 forecasts—not the analyses.

5.3 Further reducing the number of measurements

In addition, until now we have conveniently chosen to observe the height field at all $L = N^2 = 256$ grid locations. We now attempt to reduce L even further, by repeating the analysis with $L = 252$ and $L = 248$ height measurements, chosen at arbitrary grid points. From the results 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, even with time delays, 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 study has not been thoroughly explored. Additional refinement of the parameters $\mathbf{G}(t),\ \mathbf{g}(t),\ D_M,$ and τ may further reduce this constraint, for instance by allowing $\mathbf{G}(t)$ to be non-diagonal.

5.4 Noise in the observations

We now repeat the above calculations for $L = 252$ with Gaussian noise $N(0, \sigma)$ added to the height observations. A comparison is shown in Figure 7 for $\sigma = \{0.2, 0.5\}$ and $D_M = \{8, 10\}$. The synchronization error still falls rapidly within the observation window, although not to $O(10^{-5})$, as in the noiseless case. In the prediction window, it rises in an exponential manner as expected. These results were included to show that the method appears to be relatively robust to small errors in the observations. A more thorough examination of the impact of imperfect observations will be given elsewhere.

5.5 Using drifter data

Another quite important source of observations about ocean flows is being provided by position measurements $\mathbf{r}(t)$ of Lagrangian drifters (Mariano et al (2002)). Such observations have been shown to be a good supplement to the traditional observations made on a fixed grid (Kuznetsov et al (2003)) and they can also be used to estimate an Eulerian velocity field (Molcard et al (2003); Piterbarg et al (2008); Salman et al (2006)). In this section, we combine the time delay method with a data set from drifter measurements to show that they can provide accurate estimates for the grid state variables $\{h(\mathbf{r}^{(i,j)},t), u(\mathbf{r}^{(i,j)},t), v(\mathbf{r}^{(i,j)},t)\}$, without much additional effort.

We monitor the positions of N_D drifters deployed at randomly chosen grid locations and afterwards allowed to move freely to provide spatially continuous measurements between grid points. The dynamics of drifters are described as two-dimensional fluid parcel motion on the surface of the water layer, which are determined by the Lagrangian equations

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

where $\mathbf{r}^{(n)}(t)$ is the position of the n^{th} drifter and this equation was simulated by linear interpolation of the discrete velocity fields (Press et al (2015); Thompson and Emery (2014)). Hybrid measurements are incorporated into the time delay nudging method by combining the grid variables and the collective drifter positions

$$\mathbf{R}^{\dagger}(t) = \{[\mathbf{r}^{(1)}(t)]^{\dagger}, [\mathbf{r}^{(2)}(t)]^{\dagger}, \ldots, [\mathbf{r}^{(N_D)}(t)]^{\dagger}\}$$

into a single hybrid state vector. The corresponding time delayed vectors are $\mathbf{Y}_{drifter}(t) = \{\mathbf{Y}_{grid}(t), \mathbf{Y}_{drifter}(t)\}$ and $\mathbf{S}_{drifter}(t) = \{\mathbf{S}_{grid}(t), \mathbf{S}_{drifter}(t)\}$ respectively, where

$$\mathbf{Y}^{\dagger}_{drifter}(t) = \{\mathbf{R}^{\dagger}_{data}(t), \mathbf{R}^{\dagger}_{data}(t+\tau), \ldots, \mathbf{R}^{\dagger}_{data}(t+\tau(D_M-1))\}$$
$$\mathbf{S}^{\dagger}_{drifter}(t) = \{\mathbf{R}^{\dagger}_{model}(t), \mathbf{R}^{\dagger}_{model}(t+\tau), \ldots, \mathbf{R}^{\dagger}_{model}(t+\tau(D_M-1))\}.$$

In contrast to the previous results, where the initial conditions for the grid variables were taken to differ in both phase and frequency between the true solution and the estimate, here the initial conditions only vary in amplitude. That is, the initial conditions of the data $\psi_{data}(\mathbf{r}^{(i,j)},t_0)$ and $h_{data}(\mathbf{r}^{(i,j)},t_0)$ and of the model $\psi_{model}(\mathbf{r}^{(i,j)},t_0)$ and $h_{model}(\mathbf{r}^{(i,j)},t_0)$ are related by $\psi_{data}(\mathbf{r}^{(i,j)},t_0) = C_0\,\psi_{model}(\mathbf{r}^{(i,j)},t_0)$ and $h_{data}(\mathbf{r}^{(i,j)},t_0) = C_0\,h_{model}(\mathbf{r}^{(i,j)},t_0)$. We choose $C_0 = 1.0 + 0.1\eta$, 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. This was done in order to improve the results, as we found that the drifter results were more sensitive to the choice of initial condition than the results from the previous section, without drifters. Plots showing the initial positions of drifters for the two cases considered below ($N_D = 20$ and $N_D = 64$) are shown in Figure 8. They were also deactivated when they reached the boundary of the grid, so the number of operational drifters decreases throughout the estimation window.

In Figure 9, we show the synchronization error of observed quantities when for $D_M = 8$, keeping all other parameters the same as in the previous calculations. We present (in red) the synchronization error for L = 208 height observations and $N_D = 20$ drifter observations, and we show (in blue) the same synchronization error when $L = 208$ and $N_D = 0$ drifters are deployed. With $L = 208$, namely, observing 27% of the heights and 20 drifters, the synchronization error converges to a small value within the five hour observation window. Without drifters, the estimation fails.

Furthermore, by increasing the number of drifters to $N_D = 64$ within a 30 minute observation window, synchronization can be achieved with $L = 128$ height observations. Snapshots of the fields at different times throughout the estimation and prediction window are shown in Figure 10 for comparison.

Thus, although we have not yet explored how to balance between the number of drifters tracked and the number of height (or other) measurements employed, it is clear from these preliminary results that drifter data can be useful for improving the observability of the system, and that the time delay method provides a way to incorporate this information into the analysis.

6 Discussion and Summary

The transfer of information from measurements of a chaotic dynamical system to a quantitative model of the system is impeded when the number of measurements at each measurement time is below an approximate threshold L_s, which can be established in a twin experiment. Whartenby et al (2013) previously showed that for a nonlinear model for shallow water flow, a standard nudging technique given by Eq. (2) requires direct observation of roughly 70% of the dynamical variables $\{h(\mathbf{r},t), u(\mathbf{r},t), v(\mathbf{r},t)\}$ at each measurement time to synchronize the model output with the observations.

Here we have demonstrated how information in the time delays of the observations may be used to reduce this requirement to about 30%, in which only the height fields need be observed. Moreover, it appears L_s can be even further reduced by adding positional information from drifters, which interpolate the height field at locations between grid points.

Although all this has been done on a simplified model of shallow water flow, implemented with only $D = 3N_\Delta^2 = 768$ degrees of freedom, the process can be used to analyze increasingly realistic and complex models of coupled earth systems. Since the successful analysis of simulated data is typically a prerequisite for success with real data, when the model is wrong (as it generally will be in practice) this methodology provides some idea as to whether the model is at fault, or whether more observations are needed.

Furthermore, 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 operational NWP models. We also suspect this issue of insufficient measurements to be a critical limitation in our current ability to predict the behavior of complex, chaotic systems. Since such systems are quite typical in practice, these issues need to be examined with more realistic models.

Harking back to the introduction, we note that in the report by Cardinali (2013) indicates that 30-40 million daily observations are now available at the ECMWF, and that many NWP models comprise upwards of 10^8 degrees of freedom. If the qualitative trends shown here, in which time delays provide successful predictions with only 30% of the state variables observed, can be extended to substantially larger systems, then this method may indeed be useful for improving the forecasts of existing operational NWP models.

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. We would also like to thank the reviewers for their thorough reading and thoughtful suggestions that have helped significantly improve this manuscript.

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* **230**, 50-64, (2007).

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 (2009).

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

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

Courtier, Philippe, J. N. Thépaut, and A. Hollingsworth, "A strategy for operational implementation of 4D-Var, using an incremental approach," *Quarterly Journal of the Royal Meteorlogical Society* **120**, 1367-1387 (1994).

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

Fabien Durand, L. Gourdeau, T. Delcroix, and J. 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).

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.

Karthik S. Gurumoorthy, C. Grudzien, A. Apte, A. Carrassi, and C.K.R.T. Jones, "Rank deficiency of Kalman error covariance matrices in linear perfect model" *arXiv preprint arXiv:1503.05029* (2015).

Kantz, H. and T. Schreiber, *Nonlinear Time Series Analysis*, 2^{nd} ed., 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).

Kody J.H. Law, D. Sanz-Alonso, A. Shukla, and A. M. Stuart, "Filter accuracy for chaotic dynamical systems: fixed versus adaptive observation operators," *arXiv preprint arXiv:1411.3113* (2014).

Lewis, J. M. and Derber, J. C., "The use of adjoint equations to solve a variational adjustment problem with advective constraints," *Tellus A*, 37A 309–322 (1985).

Edward N. Lorenz "Deterministic Nonperiodic Flow," *Journal of the Atmospheric Sciences* **20** 130-141 (1963).

Pazo, D., A. Carrassi and J. M. Lopez, "Data-assimilation by delay-coordinate nudging," *Quarterly Journal of the Royal Meteorological Society* **142** 1290-1299 (2016)

Palatella L., A. Carrassi, and A. Trevisan, "Lyapunov vectors and assimilation in the unstable subspace: theory and applications," *Journal of Physics A: Mathematical and Theoretical* **46** (25) (2013).

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

Pires, C., R. Vautard, and O. Talagrand. "On extending the limits of variational assimilation in nonlinear chaotic systems," *Tellus A* **48.1** 96-121 (1996).

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

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

Quinn, J. C., H. D. Abarbanel. "State and parameter estimation using Monte Carlo evaluation of path integrals," *Quarterly Journal of the Royal Meteorological Society* **136 (652)** 1855-1867 (2010).

Rabier, Florence, H. Järvinen, E. Klinker, J.F. Mahfouf, and A. Simmons, "The ECMWF operational implementation of four-dimensional variational assimilation I: Experimental results with simplified physics," *Quarterly Journal of the Royal Meteorological Society* **126**, 1143-1170 (1994).

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," *Monthly Weather Review* **134**, 1081-1101 (2006).

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

Talagrand, O., and P. Courtier, "Variational Assimilation of Meteorological Observations With the Adjoint Vorticity Equation. I: Theory," *Quarterly Journal of the Royal Meterology Society* **113**, 1311–1328 (1987).

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”

Anna Trevisan, M. D’Isidoro, and O. Talagrand, “Four dimensional variational assimilation in the unstable subspace and the optimal subspace dimension,” *Quarterly Journal of the Royal Meteorological Society* **136**, no. 647, 487-496 (2010).

Trevisan, A., and L. Palatella, “On the Kalman Filter error covariance collapse into the unstable subspace,” *Nonlinear Processes in Geophysics* **18**, no. 2, 243-250, (2011).

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).

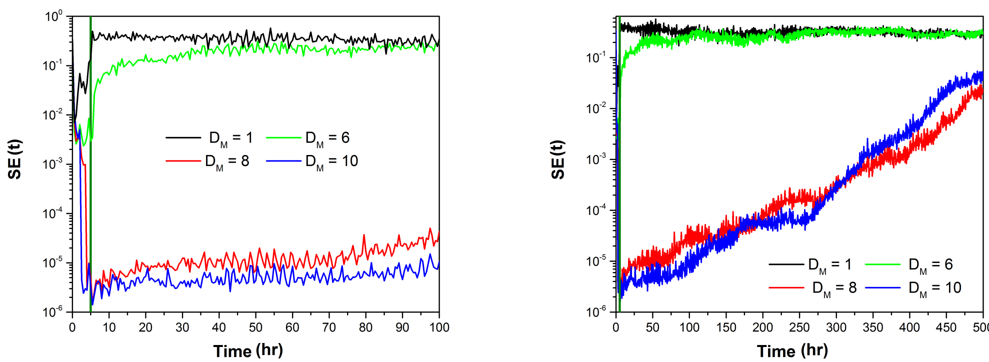

Figure 1. Synchronization error $SE(t)$, defined in Eq. (9), computed with $D_M = \{1, 6, 8, 10\}$, $G_h\,\Delta t = 1.5$, $G_u\,\Delta t = g_v\,\Delta t = 0.5$ and $\tau = 10\,\Delta t = 0.1\,h$. Assimilation is performed for $t \leq 5$ hr. **Left Panel** The couplings are then switched off and predictions are generated using the original dynamical equations Eq. (10) until $t = 100\,h$. In the prediction window ($t \geq 5$), 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 $D_M = \{1, 6\}$, suggesting that accurate predictions will be obtained $D_M = \{8, 10\}$. **Right Panel** The same calculation, but extended to t = 500 hr.

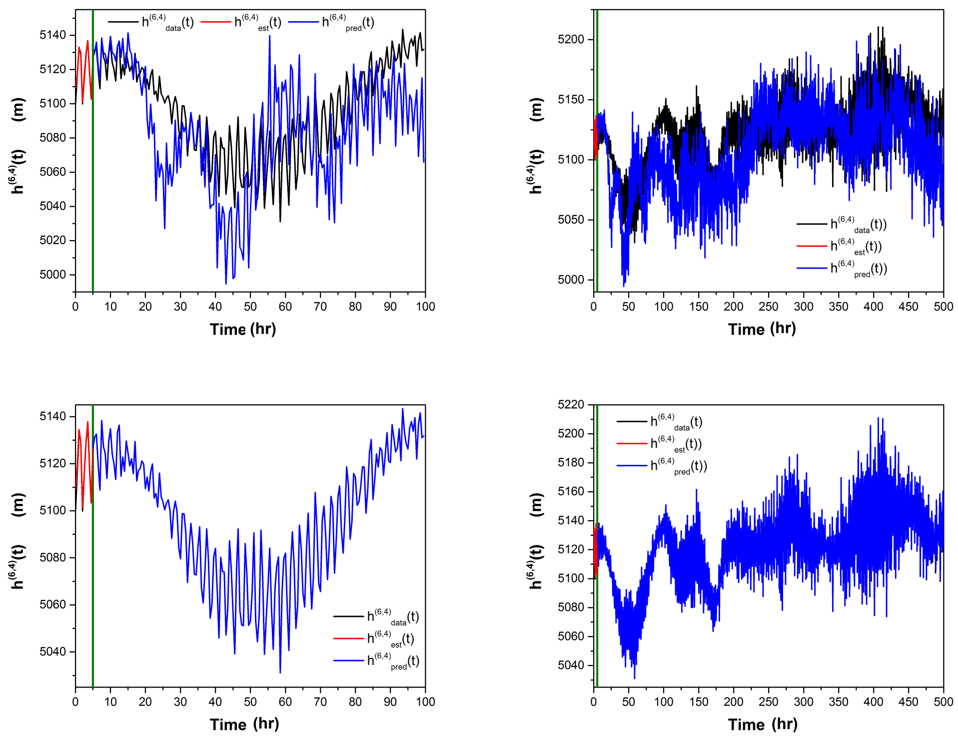

Figure 2. Upper Left Panel Known (black), estimated (red) and predicted (blue) for the observed height values $h^{(6,4)}(t)$ at grid point $(6,4)$ for $D_M = 6$. Observations are for $0 \leq t \leq 5$ hr. Predictions are for $5 \leq t \leq 100$ hr. **Upper Right Panel** The same calculation for $D_M = 6$ for a longer prediction window $5 \leq t \leq 500$ hr. **Lower Left Panel** The same calculation except $D_M = 8$. Prediction window is $5 \leq t \leq 100$ hr. **Lower Right Panel** The same calculation except $D_M = 8$. Prediction window is $5 \leq t \leq 500$ hr.

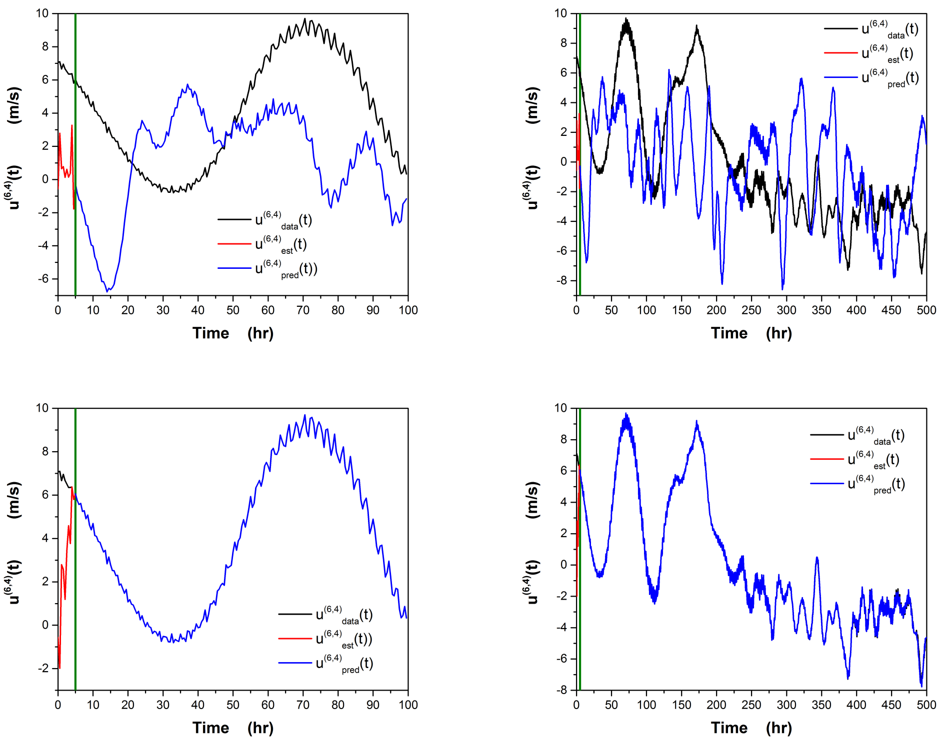

Figure 3. Upper Left Panel Known (black), estimated (red) and predicted (blue) for the observed x-velocity values $u^{(6,4)}(t)$ at grid point $(6,4)$ for $D_M = 6$. Observations are for $0 \leq t \leq 5$ hr. Predictions are for $5 \leq t \leq 100$ hr. **Upper Right Panel** The same calculation for $D_M = 6$ for a longer prediction window $5 \leq t \leq 500$ hr. **Lower Left Panel** The same calculation except $D_M = 8$. Prediction window is $5 \leq t \leq 100$ hr. **Lower Right Panel** The same calculation except $D_M = 8$. Prediction window is $5 \leq t \leq 500$ hr.

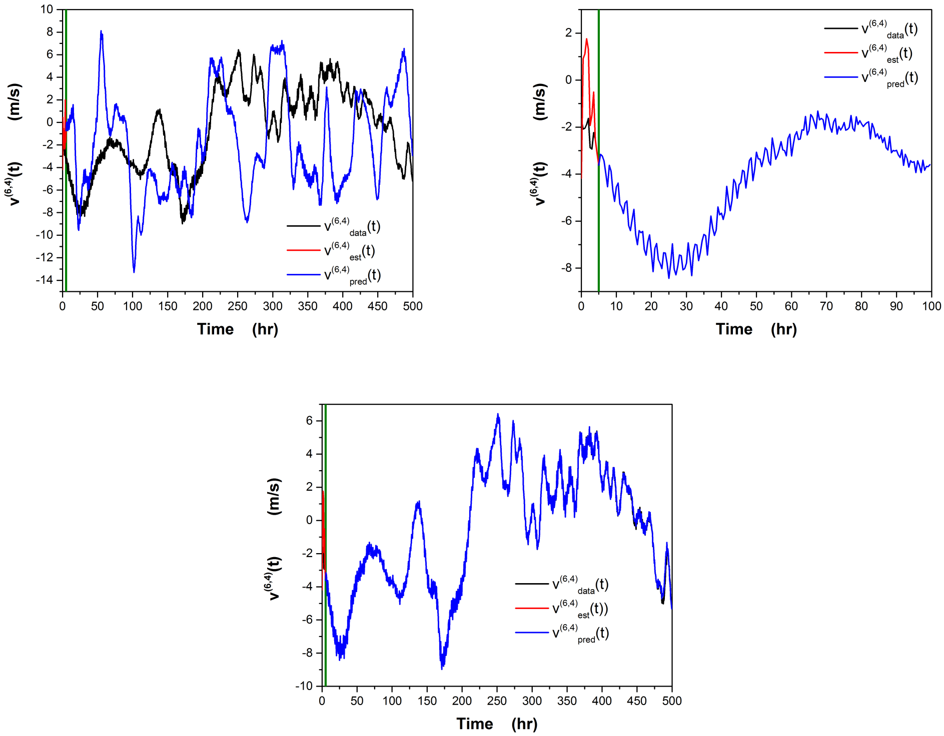

Figure 4. Upper Left Panel Known (black), estimated (red) and predicted (blue) for the observed y-velocity values $v^{(6,4)}(t)$ at grid point $(6,4)$ for $D_M = 6$. Observations are for $0 \leq t \leq 5$ hr. Predictions are for $5 \leq t \leq 100$ hr. **Upper Right Panel** The same calculation for $D_M = 6$ for a longer prediction window $5 \leq t \leq 500$ hr. **Lower Left Panel** The same calculation except $D_M = 8$. Prediction window is $5 \leq t \leq 100$ hr. **Lower Right Panel** The same calculation except $D_M = 8$. Prediction window is $5 \leq t \leq 500$ hr.

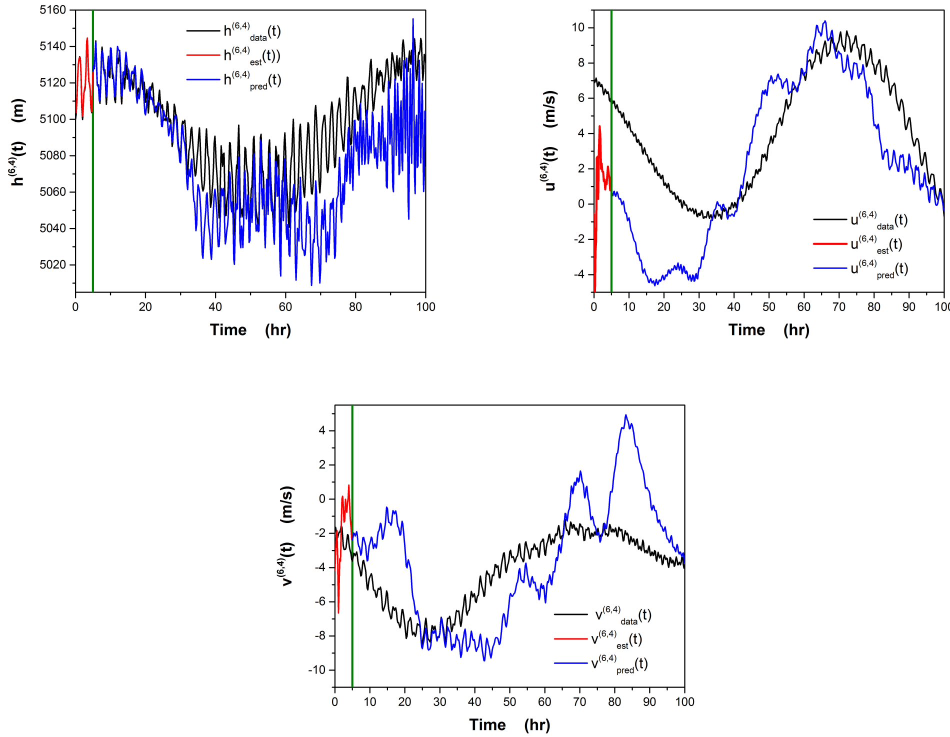

Figure 5. Data assimilation results with $D_M = 10$ and reduced coupling on the height component $h^{(6,4)}(t)$ at location (6,4), $g_h\,\Delta t = g_u\,\Delta t = g_v\,\Delta t = 0.5$. All other parameters are the same. **Upper Left Panel** $SE(t)$ for $0 \leq t \leq 200$ hr. **Upper Right Panel** Known (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 $0 \leq t \leq 5$ hr. Predictions are for $5 \leq t \leq 100$ hr. **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 $0 \leq t \leq 5$ hr. Predictions are for $5 \leq t \leq 100$ hr. **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 $0 \leq t \leq 5$ hr. Predictions are for $5 \leq t \leq 100$ hr.

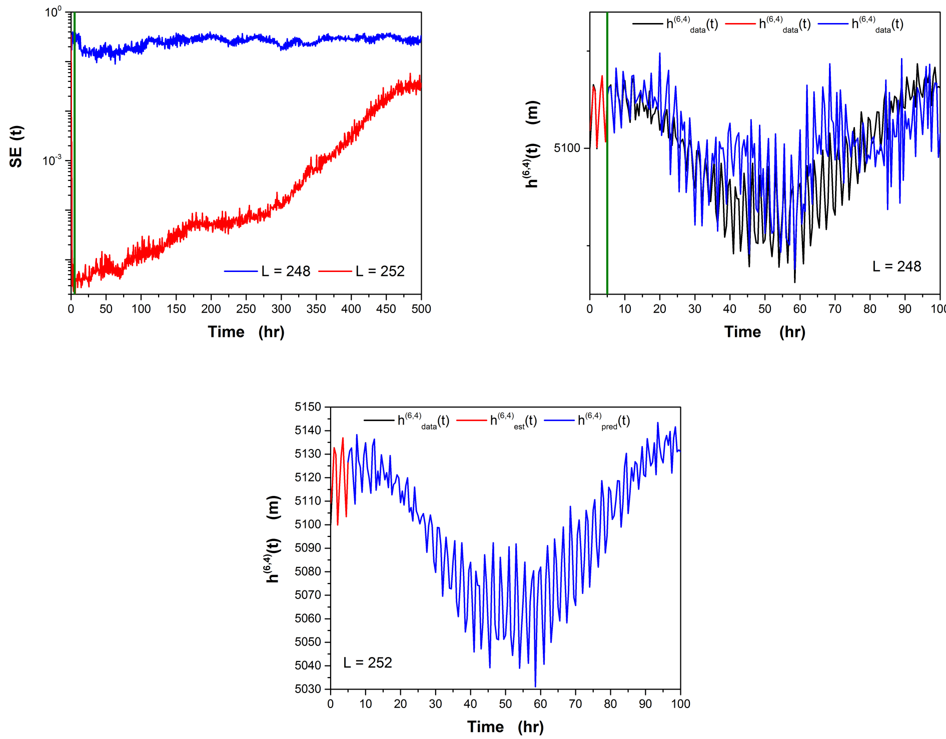

Figure 6. Synchronization error and known, estimated, and predicted height values for L = 248 height measurements at each observation time and for L = 252 height measurements at each observation time. **Upper Left Panel** $SE(t)$ for L = 248 and L = 252 over $0 \leq t \leq 5$ h in the observation window, 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).

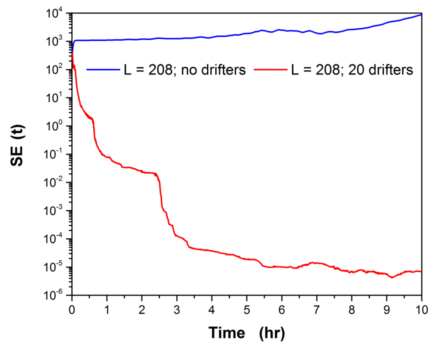

Figure 7. The effect of noise levels in the initial condition for the solution of the model equations Eq. (10) on $SE(t)$. We show the results for $D_M = 8$ and 10 for added Gaussian noise $N(0, \sigma)$ with $\sigma = 0.2$ and 0.5. For this range of noise levels added to the initial condition for generating the data in our twin experiments, we see that the detailed values of $SE(t)$ change. In the case of both $D_M = 8$ and $D_M = 10$, $SE(t)$ 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.

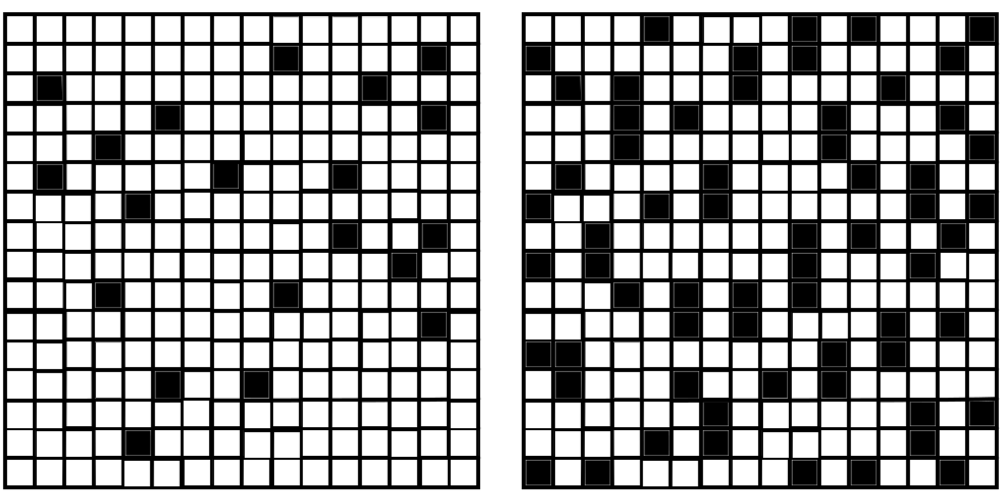

Figure 8. Initial positions for **Left Panel** $N_D = 20$ drifters and **Right Panel** $N_D = 64$ drifters.

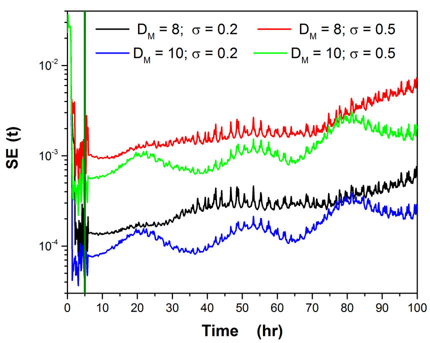

Figure 9. $SE(t)$ for our standard twin experiment described in detail earlier when we utilize drifter information, and when we do not utilize drifter information. When the number of observations of height is L = 208, we see that without drifter information (blue line) there is no synchronization and correspondingly inaccurate predictions (not shown). When information from 20 Lagrangian drifters is added during data assimilation using time delay nudging, $SE(t)$ 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.

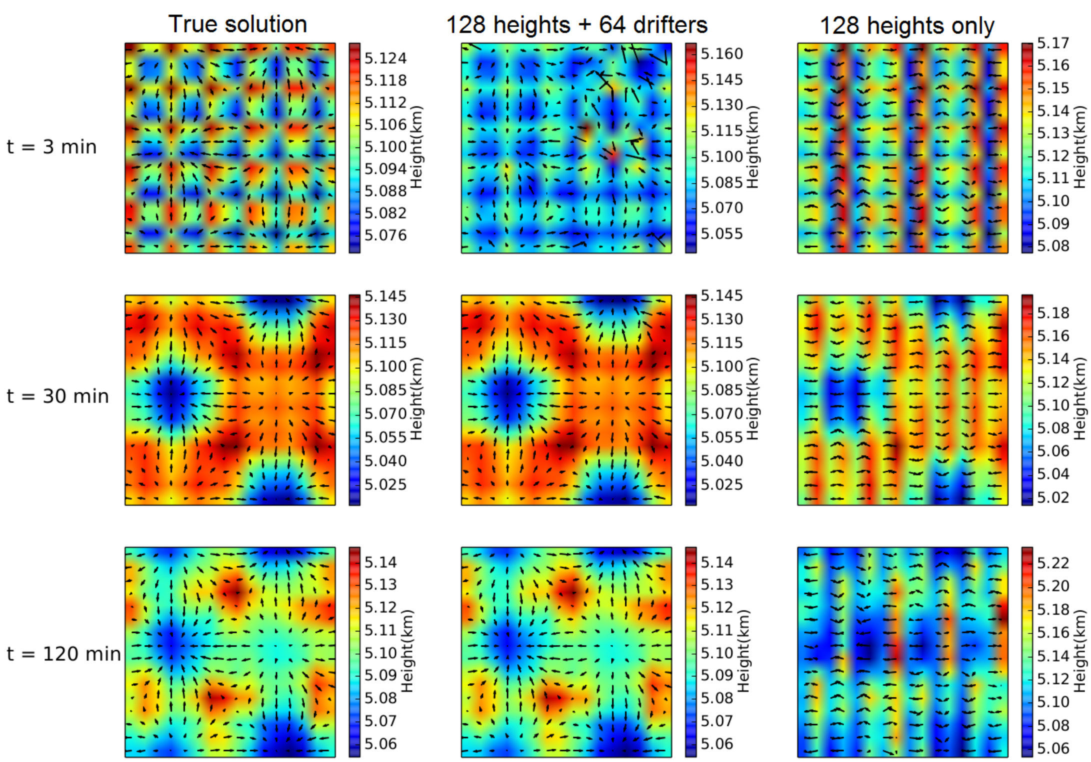

Figure 10. Comparison of the estimated and predicted fields $\{\mathbf{h}(t), \mathbf{u}(t), \mathbf{v}(t)\}$ between the truth (**Left Column**) and analyses, run with observations of 128 height variables, both with (**Center Column**) and without drifters (**Right Column**). Snapshots are taken 3 min (**Upper Row**) into the assimilation window, at 30 min the end of the assimilation window (**Center Row**), and 90 min into the prediction window (**Bottom Row**).