Probabilistic Noise Reduction

James A. Hansen*

Space Science Department, Rutherford Appleton Laboratory

Chilton, Didcot, OX11 0QX

United Kingdom &

Department of Mathematics, University of Oxford

&

Leonard A. Smith

Department of Mathematics, University of Oxford

Oxford, OX1 3LB

United Kingdom

*Corresponding author

Email: jhansen@mit.edu

Abstract.

State estimation is an important factor in the production of accurate forecasts. Great effort is expended in reducing the noise inherent in observations, to produce a 'best' estimate of the true system state. But noisy observations necessitate a probabilistic, not a deterministic, approach to state estimation. A state's probabilistic description is rarely Gaussian, and requires information beyond variance magnitude; the correct distribution is provided by the underlying structure of the system attractor. The concepts of finite-time stable and unstable sets are introduced and data assimilation-based methods for their estimation are developed. Four-dimensional variational assimilation proves adept at finding the finite-time stable set valid at the beginning of assimilation windows while the ensemble Kalman filter is capable of approximating the finite-time unstable set at any time that an observation is available. Combining the results of the two schemes produces a probabilistic estimate of the system state that is superior to either in isolation.

1. Introduction

Improving initial conditions is widely viewed as the primary means for improving forecasts of a deterministic system. Initial conditions are typically improved by combining model information with observations to produce a single improved estimate of the system state (Ghil *et al.*, 1997). Alternatively, one may take a probabilistic approach to state estimation (Evensen, 1994). We show that improved initial conditions can be obtained through the combination of hitherto distinct probabilistic state estimation techniques. The systematic tendencies of four-dimensional variational assimilation (4D-Var) and the ensemble Kalman filter (EnKF) are exploited to improve probabilistic estimates of the true system state.

Given a perfect model and imperfect observations, to improve forecasts one must improve initial conditions. If it were possible to produce an initial condition corresponding to the true system state, it would be possible to produce perfect forecasts. But uncertainties inherent in observations mean that even given a perfect model it is impossible to produce the "true" initial conditions (Judd, 2000). For this reason, a probabilistic approach is well suited to both the initial condition estimation problem and the forecasting problem. If one can determine the correct distribution of possible initial conditions, a sample from that distribution (i.e. and ensemble) can be propagated forward in time to produce an accurate description of possible forecast states. In this paper techniques are employed to estimate the probability density function (PDF) of initial conditions.

In nonlinear dynamics terms, improving estimates of system states is typically denoted "noise reduction"; noise reduction techniques aim to combine all available information about the state of a system in an effort to produce the best estimate possible of the underlying true states. Typical noise reduction techniques include the geometric approaches described in Grassberger *et al* (1993) and the variational approach of Davies (1994).

The aim in geophysics is different. Rather than reducing noise in the historical record of observations, the emphasis is placed on improving the approximation of the true system state, which is then used an an initial condition for forecasting. Producing improved initial conditions for geophysical systems is known as "data assimilation". In the context of meteorology and oceanography, observations and model equations are combined; both variational methods and direct methods are employed. A variational approach such as four-dimensional variational assimilation (described in section 3) shifts state estimates onto a locally stable set of points (defined below), which when evolved *forward* in time collapse towards a single system state. A direct approach such as the ensemble Kalman filter (described in section 4.1) shifts state estimates onto a locally unstable set of points which, when evolved *backwards*

in time, collapse towards a single system state. By combining results from the two techniques we lesson the shortcomings of each. Contrary to Pires *et al* (1996) we demonstrate the ability to exploit local uncertainty dynamics to produce a reduction in variance along the stable set; this remains possible even when using only variational data assimilation.

As we are interested in the PDF of noise reduced states, an ensemble approach to noise reduction is employed: noise is reduced in an ensemble of states consistent with a single observation. A perfect model scenario is used throughout this work, allowing us to assess the ability of the various noise reduction techniques without the added complication of model error. It is appreciated that some noise reduction techniques are better suited to scenarios where system dynamics are not well known; both of the noise reduction techniques highlighted in this work are critically dependent on an accurate dynamical representation of the system.

Stable and unstable manifolds are introduced in section 2, along with generalisations to finite-time dynamics. It is shown that for the finite-time case it is possible to produce a greater degree of noise reduction than is suggested by a simple interpretation of the infinite-time theory. Four-dimensional variational assimilation is introduced as a method for finding the finite-time stable set in section 3, while methods for finding the finite-time unstable set, the ensemble Kalman filter and nonlinear noise reduction, are discussed in section 4. Combining the estimates of the finite-time stable and unstable set leads to improved initial conditions and forecasts, as shown in section 5. Section 6 provides a brief statement of conclusions.

2. Forecasting Nonlinear Systems

In a chaotic system, error in initial conditions will often (but need not ever) grow exponentially with forecast time, strongly motivating attempts to reduce initial uncertainty. Observations are often used in efforts to improve estimates of initial conditions (analyses), but measurement error will frustrate all attempts to obtain exact initial conditions. While measurement errors are generally assumed to be drawn from a Gaussian distributions, no such distribution can be assumed for analysis errors. A system's true state will lie on the system's attractor; the distribution of possible initial conditions should be consistent not only with the magnitude of the expected analysis uncertainty, but also with the attractor's local structure in state space, leading to highly non-Gaussian distributions.

Insofar as attractor structure is related to the evolution of the system, one approach for obtaining initial conditions that reflect local attractor structure is to utilise time series of observations called an "assimilation window", such an approach is used in section 3. By defining a model trajectory which

minimises the misfit between observations and model states, it is generally felt that the resulting analysis will evolve towards the manifold that contains the system attractor. This approach aims for analyses which are consistent with attractor structure at the final time, the final position in the 'assimilation window'. The states at the beginning of the window can suffer from problems with unconstrained regions in cost function space; regions in which significantly different states yield nearly identical cost function values. Further, the states at the beginning of the window do not reflect the local attractor structure, but rather the structure of the system's relevant local finite-time stable set, a concept discussed below.

In a hyperbolic system [1] (Hirsch and Pugh, 1970) a set of states exists whose trajectories converge to the trajectory of a given "true" state, $\mathbf{x}^t(t)$, as $t \to \infty$. This set of points is the stable manifold, W_s, formally defined as the set of points satisfying

$$\hat{\mathbf{x}} \in W_s(\mathbf{x}^t) \Leftrightarrow \lim_{t\to\infty} \|\hat{\mathbf{x}}(t) - \mathbf{x}^t(t)\| = 0, \tag{1}$$

where $\hat{\mathbf{x}} = \hat{\mathbf{x}}(0)$ and $\mathbf{x}^t = \mathbf{x}^t(0)$. Note that the superscript t denotes 'truth' while the parenthetical t denotes time. Similarly, there exists a set of states, W_u, whose trajectories converge to that of any true state, $\mathbf{x}^t(t)$, on the attractor as $t \to -\infty$. This set is called the unstable manifold,

$$\hat{\mathbf{x}} \in W_u(\mathbf{x}^t) \Leftrightarrow \lim_{t\to-\infty} \|\hat{\mathbf{x}}(t) - \mathbf{x}^t(t)\| = 0. \tag{2}$$

It is unlikely that any system (or model) of geophysical interest would be hyperbolic, and further, the relevance of stable and unstable manifolds is extremely limited when dealing with finite-lengthed trajectories. While *sets* of states will exist that collapse towards specified trajectories, these sets may not define manifolds [2]. Further, the well-known properties of stable and unstable manifolds unfold over an infinitely long trajectory; we will, of course, be using finite-lengthed trajectories in this paper. Over finite time, for example, all uncertainties may decrease, even in a chaotic system (Smith *et al.*, 1999). For these reasons we now introduce the concept of finite-time stable and unstable sets.

Finite-time stable and unstable sets can be defined in a number of ways, and several are discussed in Appendix A. Fortunately, this is not a liability; the property that is of interest is independent of the definition employed. We are interested in the set of points that, when integrated forward/backward to $\pm\tau$ collapse toward the true state at that time, $\mathbf{x}^t(\pm\tau)$; these sets are then denoted the finite-time

[1] A system in which stable and unstable manifolds always have transverse intersections (Arrowsmith and Place, 1990).

[2] Manifolds require, amongst other things, the properties of continuity and differentiability.

stable and unstable sets, respectively. The definitions employed are;

$$\tilde{\mathbf{x}} \in \widetilde{W}_s(\mathbf{x}, \tau, \epsilon) \Leftrightarrow \|\tilde{\mathbf{x}}(\tau) - \mathbf{x}^t(\tau)\| < \epsilon \tag{3}$$

($\epsilon > 0$ and $\tau > 0$) for the finite-time stable set, and

$$\tilde{\mathbf{x}} \in \widetilde{W}_u(\mathbf{x}, \tau, \epsilon) \Leftrightarrow \|\tilde{\mathbf{x}}(-\tau) - \mathbf{x}^t(-\tau)\| < \epsilon \tag{4}$$

for the finite-time unstable set. In general both $\widetilde{W}_s$ and $\widetilde{W}_u$ will have the dimension of the state space, while W_s and W_u will have a lower dimension. Alternative definitions and a convergence property are given in Appendix A.

Methods for finding points in $\widetilde{W}_s(\mathbf{x}^t)$ and $\widetilde{W_u}(\mathbf{x}^t)$ are demonstrated using the Ikeda system (Ikeda, 1979), a non-hyperbolic system described by the equations:

$$x_{i+1} = 1 + \mu(x_i \cos(t) - y_i \sin(t)) \tag{5}$$

$$y_{i+1} = \mu(x_i \sin(t) + y_i \cos(t)) \tag{6}$$

$$t = a - \frac{b}{x_i^2 + y_i^2 + 1} \tag{7}$$

where typical values of a, b and μ are 0.4, 0.6 and 9.0, respectively. The global structure of the Ikeda system is shown as the black points in figure 1 (the curved line in the figure is discussed in section 3).

Given a time series of observations, to find $\widetilde{W}_s(\mathbf{x}^t)$ one can search for initial conditions which yield trajectories consistent with the observations and their associated uncertainty over some specified window in time. "Consistency" is a subjective term; if one defines consistent to mean 'within four standard deviations', then one searches for initial conditions that produce trajectories that are as long as possible given the constraint that the model state is never farther than 4σ from each observation. Such a trajectory is said to ι-shadow the observations (Gilmour and Smith, 1997).

Initial conditions that ι-shadow (are elements of $\widetilde{W}_s(\mathbf{x}^t)$) are shown as a function of trajectory length, τ in figure 2a. A circular uncertainty isopleth with radius $r = 0.055$ is placed around the point of interest and each of its 10 subsequent images. The area bounded by the isopleth is filled uniformly with 2^{10} points. Each of these points is integrated forward, and those that are inconsistent with future observations (those that fall outside of a future point's expected uncertainty) are sequentially eliminated. The traditional view (Pires *et al.*, 1996) is that as the number of observations is increased, the consistent points at initial time will collapse onto the local stable manifold; this view is at odds with figure 2a. Such a view implies that if one is considering only trajectories forward in time it is not possible for initial points to be noise reduced along the stable manifold; once the stable manifold has been reached, there is no information in the observations which will shift those points along the

manifold towards the true system state. As a result, the noise reduced points will still span the original uncertainty magnitude but will do so only along the local directions of the stable manifold. Note that the definitions of stable and unstable manifolds do not explicitly consider finite-lengthed trajectories of observations, they require that points on the stable/unstable manifold collapse onto the same point as $t \to \pm\infty$.

When finite-lengthed time series of observations are considered, it is possible to achieve contraction along the finite-time stable/unstable set. The ι-shadowing results of figure 2a demonstrate this contraction (the other panel of figure 2 is discussed elsewhere). The outer-most grey points in figure 2a are those which were consistent at initial time, but inconsistent with the uncertainty associated with the system state after one step. The outer set of magenta points are those which became inconsistent after two steps. Notice that the points which are consistent after two steps do not simply result in a greater constriction along the major axis defined by the points which are consistent after one step; their major axis has a different orientation. This direction change, coupled with the inclusion of the one-step information, produces a set of consistent initial conditions which have effectively experienced contraction along the stable manifold. Points made inconsistent by steps 3, 4, 7, 8, 9, 10 are shown in green, yellow, red, dark green, purple and dark blue respectively. The inner-most light blue points are those which are still consistent after ten steps.

An explanation for the time-dependent orientation of the $\widetilde{W_s}(\mathbf{x}^t)$ seen in figure 2a can be found through an examination of the system's linearised dynamics. Initial singular vectors (see, for example, Strang (1988)) define directions in which infinitesimal perturbations will experience maximum growth after a specified time-scale. Figure 4 plots the initial singular vectors constructed over one step, two steps and ten steps for the same initial condition used to construct figure 2. The $\mathbf{sv_1}$ (dashed) represent growing directions, while the $\mathbf{sv_2}$ (dash-dotted) represent shrinking directions, the number in parentheses specifies the optimisation time used in the singular vector construction. The second initial singular vector (the shrinking direction) for the one step optimisation time, $\mathbf{sv_2}(1)$, is oriented in the same direction as the major axis of the set of points in figure 2a which are consistent at one step. Notice the large rotation between the leading one step singular vector, $\mathbf{sv_1}(1)$, and the leading two step singular vector, $\mathbf{sv_1}(2)$. Again, the second initial singular vector from the two step optimisation time, $\mathbf{sv_2}(2)$, is oriented in the same direction as the major axis of the set of points in figure 2a which are consistent after two steps. The local dynamics of the attractor are such that the short time-scale uncertainty behavior is independent of the long time-scale uncertainty behavior. The details of such behavior is, of course, both system and state dependent, but the general picture is not unusual (e.g. see

Smith *et al* (1999) and McSharry (1999)).

Ultimately, we wish to capture the underlying structure of the system's attractor; information that is contained in the $W_u(\mathbf{x}^t)$. The $\widetilde{W_s}(\mathbf{x}^t)$ does not directly provide information on the attractor structure, but any $\widetilde{W_s}(\mathbf{x}^t)$ will contain elements of $W_s(\mathbf{x}^t)$ and $W_u(\mathbf{x}^t)$. The elements of $W_u(\mathbf{x}^t)$ will eventually be stretched along the $W_u(\mathbf{x}^t)$. Thus attractor structure information is provided at the end of the trajectory instead of the beginning.

3. Noise reduction onto the stable set: Four-dimensional variational assimilation

First we consider a noise reduction method that approximates the $\widetilde{W_s}(\mathbf{x}^t)$. Four-dimensional variational assimilation (4d-Var) is a familiar and operationally feasible method of noise reduction in geophysics (see Ghil *et al* (1997) and references therein). It utilises information in both space and time to blend observations and model dynamics in an effort to produce a superior estimate of a system's state. 4d-Var takes trajectories of observations and their associated uncertainty information into account in its effort to produce initial conditions. This is accomplished through the minimisation of a cost function which measures the misfit between the model state and observations. Initial conditions produced by 4d-Var need not ι-shadow. Consistency is not a constraint in the 4d-Var minimisation; trajectories which are inconsistent with observed states are not overly penalised (although consistency checks can be performed): the variational approach *assumes* an ι-shadow (i.e. a consistent trajectory) exists. Variational approaches to noise reduction appear in many independent fields; a discussion of the standard method for variational noise reduction in nonlinear systems can be found in Davies (1994), where all available observational information is included in the minimisation process. The application of 4d-Var differs from such an approach in that minimisation occurs over short assimilation windows rather than across all available data. This is often justified on computational grounds, but becomes an even greater issue when one accepts the inevitability of model imperfections; no model exists that can shadow our complete record of weather observations, for example.

Using the notation of Ide *et al* (1997), the 4d-Var cost function is defined as

$$\begin{aligned} J[\mathbf{x}(t_0)] &= \frac{1}{2}(\mathbf{x}(t_0)-\mathbf{x}^b(t_0))^T\mathbf{B}_0^{-1}(\mathbf{x}(t_0)-\mathbf{x}^b(t_0)) \\ &\quad +\frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i[\mathbf{x}(t_i)]-y_i^o)^T\mathbf{R}_i^{-1}(\mathbf{H}_i[\mathbf{x}(t_i)]-\mathbf{y}_i^o), \end{aligned} \tag{8}$$

where $\mathbf{x}(t_0)$ are the model initial conditions, $\mathbf{x}^b(t_0)$ is the first guess, or background state produced by the model (typically as a short-term forecast), and $\mathbf{B}_0^{-1}$ is the inverse of the background error covariance

matrix. The second term calculates the misfit between observations, $\mathbf{y}_i^o$, distributed over the time interval $[t_0, t_n]$ and the model state expressed in observation space, $\mathbf{H}_i[\mathbf{x}(t_i)]$, weighted by the inverse of the observational error covariance matrix, $\mathbf{R}_i^{-1}$. The $\mathbf{H}_i$ operator is a map between model space and observation space, and in this work is the identity matrix. By locating a minimum of the cost function, one finds initial conditions which produce a model trajectory through the available observations which minimises the distances between model states and observed states. Note that in using this cost function one is assuming the model is perfect.

The application of 4d-Var is now illustrated using the same initial condition and ten successive images of the Ikeda map used in section 2. A Gaussian distribution of 2^{10} points is constructed around each point such that $4\sigma = 0.055$. This distribution is treated as a probability density function (PDF) of likely observations. 4d-Var is applied to each of the 2^{10} initial conditions, each using a time series of observations with its own realisation of the observational noise[3]. This produces a new, noise-reduced PDF at the beginning of the assimilation window. In short, the analyses are obtained by minimising

$$\begin{aligned} J[\mathbf{x}_j(t_0)] &= \frac{1}{2}(\mathbf{x}_j(t_0) - \mathbf{x}_j^b(t_0))^T \mathbf{B}_0^{-1}(\mathbf{x}_j(t_0) - \mathbf{x}_j^b(t_0)) \\ &\quad + \frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i[\mathbf{x}_j(t_i)] - y_{ij}^o)^T \mathbf{R}_i^{-1}(\mathbf{H}_i[\mathbf{x}_j(t_i)] - \mathbf{y}_{ij}^o),\ j = 1, 2^{10}, \end{aligned} \tag{9}$$

where the $\mathbf{x}_j^b(t_0)$ are the first guess initial conditions and the $\mathbf{y}_{ij}^o$ are re-noised observations at time t_i consistent with the uncertainty in $\mathbf{y}_{ij}^o$ as given by $\mathbf{R}_i$. A conjugate-gradient minimisation routine is used (Press *et al.*, 1986). As 4d-Var is dependent on the number of observations considered during an assimilation, initial PDFs are calculated for each assimilation window length between two points and ten points, inclusive, that is $N_a = 2, 3, \ldots, 10$. In this particular application, the background state, $\mathbf{x}^b(t_0)$ has the same uncertainty characteristics as the observations, $\mathbf{B}_0 = \mathbf{R}_i = \mathbf{I}$. Setting $\mathbf{B}_0 = \mathbf{I}$ allows the model dynamics to completely dictate the eventual analysis value; any other choice of $\mathbf{B}_0$ will preferentially weight specific regions of state space. While the method presented here is not restricted to $\mathbf{B}_0 = \mathbf{I}$, this choice allows for an unambiguous interpretation of results.

Figure 2b shows the noise reduced distribution of points at the *beginning*[4] of the assimilation window for each N_a. The magenta points show the noise reduced states for $N_a = 1$, when only the initial condition and its first image are considered. Notice that the major axis of the distribution of

---

[3]In order to satisfy the assumption that the observations are random variables, noise is added to each observation before assimilation is performed. See Burgers *et al* (1998) for theoretical justification.

[4]Predictions, of course, are launched from the *end* of the assimilation window.

points is oriented in the same direction as points consistent after one step in figure 2a, the direction of the one-step contracting singular vector, $\mathbf{sv_2}(2)$. The noise reduced states for $N_a = 2$ are shown in green. Again, the major axis is rotated into the direction of the two-step second initial singular vector, $\mathbf{sv_2}(2)$. As N_a is increases from 3 to 4, 7, and 8 (the same step-lengths shown in figure 2a), the post assimilation distribution of points go through yellow, red, dark green, and purple respectively, each additional point resulting in a further contraction onto the $\widetilde{W_s}(\mathbf{x}^t)$. The dark blue points are the analysis distribution for $N_a = 9$. While contraction towards the $\widetilde{W_s}(\mathbf{x}^t)$ continues to be a feature, notice the scattering of points (centred near (0.3,-0.3)) below the main distribution. This secondary grouping remains pronounced in the noise reduced distribution of light blue points valid for $N_a = 10$. The main grouping of points about the true initial condition agree well with the consistent region defined by ι-shadowing. This is not true for the secondary grouping, highlighting one of 4d-Var's limitations; the danger of locating localised minima which need not be consistent with future observations (an excellent discussion of this effect is given in Pires *et al* (1996)). None of the points in the secondary grouping will ι-shadow beyond four steps.

The localised minimum arises from the dynamics of the $W_s(\mathbf{x}^t)$. The solid line in figure 2b lies on the $W_s(\mathbf{x}^t)$. Moving to the right from the true initial condition, the stable manifold curves down and away before reversing direction and passing back through the points grouped in the local minimum. So the "local minimum" isn't a local minimum in the traditional sense, but rather an expression of the states that are unconstrained by the available time series of observations as reflected by the $\widetilde{W_s}(\mathbf{x}^t)$; the secondary grouping is only disconnected from the primary set of noise reduced points because of the magnitude of observational uncertainty. For any given realisation of the time series of observations, the cost function may have a value that is lower in the region of the secondary grouping than in the primary region; the secondary grouping only becomes a "local minimum" when the cost function is defined with respect to truth. Figure 1 shows a portion of the initial condition's $W_s(\mathbf{x}^t)$ in the context of the entire Ikeda attractor. It is not always the case that minima are formed by the $W_s(\mathbf{x}^t)$ folding back through the area of interest so directly; it is clear from figure 1 that the structure of the $W_s(\mathbf{x}^t)$ is such that it may take several large excursions before passing back through the region of the initial condition. Increasing the length of the assimilation window is likely to increase the number of minima as the structure of the $W_s(\mathbf{x}^t)$ becomes more complex.

4. Noise reduction onto the unstable manifold

Both the ensemble Kalman filter (EnKF) and nonlinear noise reduction (NNR) produce noise reduced states that are consistent with the $\widetilde{W_u}(\mathbf{x}^t)$. The EnKF, a variant of the extended Kalman filter (see Jazwinski (1969)), exploits the tendency of model states to collapse onto model attractors. In the familiar (to nonlinear dynamicists) technique of NNR, the local subspace containing the system manifold is estimated using a database of past observations and noise is then "reduced" by moving an observation towards this manifold.

4.1. Ensemble Kalman Filter

The EnKF was first introduced by Evensen (Evensen (1994)). It is a method that is finding increased application in both oceanography (Evensen and van Leeuwen, 1996) and meteorology (Houtekamer and Mitchell, 1998). The EnKF is intrinsically probabilistic, making it well suited to providing analysis PDF results. Instead of performing a separate assimilation operation on each initial estimate of the system state (as is done for the 4d-Var approach in section 3), the EnKF produces the uncertainty estimates directly from ensemble integrations. The EnKF is a sequential technique; model states are integrated forward until an observation is encountered, at which point the model states and observed state are combined to produce new, improved estimates of the true state of the system.

State estimates are produced through the combination of short term model forecasts, observations and the associated uncertainty of each. A full description of the method is given in Appendix B. The remainder of this paragraph describes the EnKF equations in a method analogous to the extended Kalman Filter. The equations are:

$$\mathbf{x}_i^f(t) = \mathbf{F}(\mathbf{x}_i^a(t-1)), \text{ for } i = 1, n \tag{10}$$

$$\mathbf{P}^f(t) = \frac{1}{n-1}(\mathbf{A}^f(t) - \mathbf{M}^f(t))(\mathbf{A}^f(t) - \mathbf{M}^f(t))^T \tag{11}$$

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)^T(\mathbf{H}(t)\mathbf{P}^f(t)\mathbf{H}(t)^T + \mathbf{R}(t))^{-1} \tag{12}$$

$$\mathbf{x}_i^a(t) = \mathbf{x}_i^f(t) + \mathbf{K}(t)[\mathbf{y}_i^o(t) - \mathbf{H}(t)\mathbf{x}_i^f(t)], \text{ for } i = 1, n \tag{13}$$

$$\mathbf{P}^a(t) = \frac{1}{n-1}(\mathbf{A}^a(t) - \mathbf{M}^a(t))(\mathbf{A}^a(t) - \mathbf{M}^a(t))^T. \tag{14}$$

In equation 10, each analysis value from the previous time-step, $\mathbf{x}_i^a(t-1)$, is propagated forward using the forecast model, $\mathbf{F}$, to produce a forecast ensemble member, $\mathbf{x}_i^f(t)$ (n is the ensemble size). The forecast error covariance, $\mathbf{P}^f(t)$, is defined directly from the forecast ensemble in equation 11. $\mathbf{A}^f(t)$ is a matrix made up of each forecast ensemble member, and $\mathbf{M}^f(t)$ is a matrix constructed from the

vector mean of $\mathbf{A}^f(t)$. The gain matrix, $\mathbf{K}(t)$, is calculated in equation 12 using the forecast error covariance and the observational error covariance, $\mathbf{R}(t)$. Here $\mathbf{H}(t)$ is simply an operator that maps one from model space to observation space. Once the gain matrix is in hand, it is possible to produce an analysis ensemble by "correcting" each member of the forecast ensemble using equation 13. A correction term is defined by weighting the difference between a forecast ensemble member and an observational realisation, $\mathbf{y}_i^o(t)$, with the gain matrix. Finally, the analysis error covariance, $\mathbf{P}^a(t)$, is computed in equation 14 directly from the analysis ensemble. $\mathbf{A}^a(t)$ is a matrix made up of each analysis ensemble member, and $\mathbf{M}^a(t)$ is a matrix constructed from the vector mean of $\mathbf{A}^a(t)$. Because we are considering the perfect model scenario, no model error covariances appear in equations 10-14. There are issues of independence associated with correcting ensemble members used in the construction of the gain matrix (see Houtekamer and Mitchell (1998)), but they are ignored in this work.

If desired, a single analysis value can be produced at each time by calculating the mean of the analysis ensemble:

$$\mathbf{x}^a(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^a(t). \tag{15}$$

Although $\mathbf{x}^a(t)$ provides a useful deterministic analysis value for verification, only in the case where the analysis has a Gaussian distribution (unlikely in nonlinear models) will it provide a maximum likelihood estimate of the true state. Indeed, the mean of the analysis ensemble can easily occupy a region of state space where the probability of the analysis PDF is zero. An alternative deterministic analysis is provided by the most probable member of the ensemble, but again, the trajectory produced by this member need not be "good". The most sensible approach is to remain in a probabilistic framework.

The EnKF system begins with an initially Gaussian analysis ensemble, but integrating the ensemble forward in time allows states that lie off the model attractor to collapse towards it. This blending of model trajectories and observations produces noise reduced distributions that are far from Gaussian, as can be seen in figure 3a, where EnKF noise reduction is performed using the same initial conditions, observations, and realisations of observations used in the 4d-Var experiments of section 3. The noise reduced points (black dots) appear to cluster onto the $\widetilde{W_u}(\mathbf{x}^t)$, an estimate of the manifold containing the system attractor. Some of the noise reduced points lie far from the manifold containing the attractor; this is due to a combination of observations pulling analyses away from the attractor, and to the time-scale of collapse onto the attractor being larger than a single integration step.

4.2. Nonlinear noise reduction

The aim of "traditional" nonlinear noise reduction (NNR) techniques is to produce improved estimates of the true state of a system (see e.g. Kostelich and Yorke (1988)). Unlike 4d-Var and the EnKF, NNR accomplishes this using only a database of noisy observations; the system equations are not required. This is of substantial benefit should a good representation of the system's dynamics be unavailable, but when the dataset is long compared to the recurrence time of the system, NNR provides an attractive alternative, a geometric method for producing $\widetilde{W_u}(\mathbf{x}^t)$.

If a system evolves in a space with a dimension greater than the dimension of the manifold on which the dynamics lie, then it is possible to move an observed point closer to the manifold by eliminating the observation's projection into dimensions not populated by the manifold. To accomplish this, a local estimate of the subspace spanned by the manifold is determined using past observations that are near the observation of interest. A subspace is defined by the covariance eigenvectors that account for the majority of the variance of mass in the distribution of points[5]. Ideally, this gives the subspace occupied by the attractor, while the remaining eigenvectors define a subspace occupied only due to noise. An observation is noise reduced by removing its projection into the noise subspace. If an ensemble of observations are constructed around the point one wishes to noise reduce and each is noise reduced in turn, an analysis PDF can be constructed which is expected to lie closer to the locally unstable set. Probabilistic NNR is applied to the Ikeda system by Hansen (1998).

The geometric NNR approach is similar in concept to a technique called 'normal mode initialisation' (Daley, 1991) used in atmospheric data assimilation. Instead of locating eigenmodes geometrically they are determined dynamically, and instead of minimising an analysis' projection into the noise subspace one minimises an analysis' projection into fast time-scale subspaces. The aim is to produce an analysis that won't suffer from dynamic 'shocks' when propagated forward, but it can also be discussed in terms of attempting to determine the local attractor structure.

5. Combined probability

The distributions of analyses produced by 4d-Var at the *beginning* of the assimilation window tend to lie along the $\widetilde{W_s}(\mathbf{x}^t)$, while the distributions produced by the EnKF at any observation time tend to lie along the $\widetilde{W_u}(\mathbf{x}^t)$. In general, their combination produces a state estimate that leads to an initial condition estimate superior to either in isolation. It is worth mentioning that in the linear, deterministic

[5]Deciding on the number of eigenvectors to use is typically a subjective exercise.

context, 4d-Var and the Kalman filtering approach are equivalent (Lorenc, 1986). In nonlinear systems they are not equivalent. Further, the combined approach aims for probabilistic state estimation; by combining the distribution obtained at the *beginning* of the 4d-Var assimilation window and the EnKF distribution we achieve a distribution that outperforms either in isolation.

The finite-time stable and unstable sets are not co-incident in the region of the Ikeda system shown in figures 2 and 3, and useful information can be obtained by combining the results from 4d-Var and the EnKF. Figure 3b shows a re-sampling of points from the $N_a = 10$ 4d-Var analysis (figure 2b) retaining those points which are most likely given the EnKF distribution (figure 3a). This sample was determined by weighting each point in the 4d-Var analysis ensemble with a probability given by the EnKF analysis; members with weights greater than zero were retained. The most striking feature of the combined set is the removal of the local minimum found in the 4d-Var analysis. Further, the combined set exhibits significant contraction along $\widetilde{W_s}(\mathbf{x}^t)$; the set spans only the range of the local branch of the attractor.

Figure 3b shows only one application of our technique. To demonstrate the robustness of this approach, experiments were carried out on 1024 initial conditions spread across the attractor. For each initial condition, a 4d-Var analysis PDF, EnKF analysis PDF and a combined 4d-Var/EnKF PDF was produced. The quality of the analyses were assessed by determining the area contained within the 90% contour, the area in which 90% of the analyses reside. A binning procedure was used to estimate the PDFs' area, with a resulting bias towards an overestimate of the total uncertainty, but an underestimate of the impact of local minima, diluting the performance of the combined method. Histograms of the results normalised by the area of the 90% contour produced by observational uncertainty are plotted in figure 5; 4d-Var results in panel a), and EnKF results in panel b). A histogram of areas for the combined approach is not shown, but members would occupy only the first bin in the histograms shown in figure 5. The 4d-Var analysis areas are, on average narrower than the EnKF analysis areas, magnitude. In every instance considered, the combined PDF contained truth.

The combination of 4d-Var and EnKF noise reduction gives better analysis distributions than either of the methods in isolation, but what about its ability to forecast? Recall that the analyses being produced are constructed around the state at the *beginning* of the 4d-Var assimilation window. If one is assimilating data in real time, before a forecast can be run the analysis PDF must be propagated forward to the end of the assimilation window. When the $N_a = 10$ 4d-Var distribution and the combined distribution are evolved to the end of the assimilation window (the time at which forecasts are launched), both of the evolved ensembles project strongly onto the $\widetilde{W_u}(\mathbf{x}^t)$. The 4d-Var ensemble's variance along $\widetilde{W_u}(\mathbf{x}^t)$ is larger than that of the combined set; further, the combined set is better centred

on the true state of the system. A cumulative distribution constructed from the squared distance from truth of the points in the 4d-Var and combined ensembles (not shown) conveys the same information: for this initial condition, the combined ensemble has a larger fraction of points near the true state and a narrower spread than the 4d-Var ensemble. If cumulative distributions for the EnKF ensemble and observational distribution are included, the three noise reduction methods significantly outperform the observational distribution, while the 4d-Var and combined ensembles outperform the EnKF ensemble. Quantitatively, for this initial condition the combined ensemble unambiguously outperforms the 4d-Var ensemble beyond the 70th percentile, while 80% of the combined ensemble states fall within a radius that is 33% smaller than that needed to contain 80% of the 4d-Var states.

Figure 6 plots mean errors for 4d-Var, EnKF and combined ensembles as a function of time from the beginning of the 4d-Var assimilation window. The values are normalised by the error resulting from an ensemble constructed around an observation at the beginning of the assimilation window and integrated forward. The negative numbers along the x-axis represent the fact that the combined analysis is valid at the beginning of the 4d-Var data assimilation window; the 4d-Var and combined analysis PDF must be propagated to the end of the window before forecasts can be launched. Only forecast errors are plotted for the EnKF as it uses no 'window'.

The combined distribution outperforms the 4d-Var distribution both over the 10 step assimilation window and over the 10 forecast steps considered. This stems primarily from the reduction of local minima through the inclusion of the EnKF PDF. Values shown are the most likely (the mean value), which do not necessarily coincide with the most probable (the bin with the highest probability value). The 4d-Var mean analysis error is 0.0144, compared with a most probable value of 0.0042; the PDFs are clearly far from Gaussian. For comparison, the combined distribution's mean analysis error is 0.0011 with a nearly identical most probable value; the structure of the $\widetilde{W_s}(\mathbf{x}^t)$ has significant curvature in many regions of the Ikeda system. The inclusion of the EnKF information removes curvature that is not consistent with the local attractor structure and improves the relevance of ensemble mean statistics. It is important to note that this feature is not unique to the EnKF. Similar information could be obtained through three-dimensional variational assimilation, or from the states at the final location of the previous 4d-Var window. The EnKF was chosen for this work because the large number of $W_s(\mathbf{x})$-based localised minima encountered during 4d-Var minimisation resulted in relatively poor $\widetilde{W_u}(\mathbf{x})$. It is important to note that these localised minima problems may be system and/or minimisation routine specific.

6. Conclusions

This paper has introduced an improved method for producing analysis ensembles that relies on the combination of four-dimensional variational assimilation and the ensemble Kalman filter. The method presented here takes a probabilistic approach to the problem, producing PDFs of noise reduced states rather than a single 'best' estimate. Given a nonlinear system, a PDF is to be preferred over any 'best' estimate . All experiments were performed in the perfect model context using observations distributed about truth. While results are illustrated for the Ikeda map, the method is not constrained to low-dimensional, chaotic systems. Insofar as a dynamical system of interest possesses non-coincident finite-time stable and unstable sets, the techniques described in this paper are expected to provide improved analysis ensembles. The limitations of implementation are practical (available computing power) rather than theoretical.

The distribution of noise reduced points produced by 4d-Var at the *beginning* of the assimilation window tend to fall onto the $\widetilde{W_s}(\mathbf{x}^t)$, while those produced by the EnKF (which has no window) tend to fall onto the $\widetilde{W_u}(\mathbf{x}^t)$. By combining distributions it is possible to produce a noise reduced PDF that exploits the benefits of each. As discussed in section 2, the finite-time dynamics of the system render it possible to obtain contraction along the $\widetilde{W_s}(\mathbf{x}^t)$ using only 4d-Var, contrary to the results of Pires (Pires *et al.*, 1996). Additional improvement is produced by incorporating the EnKF.

The ability to determine the local structure of a model's attractor has implications beyond noise reduction. When an analysis ensemble is consistent with the underlying system attractor, then both the ensemble members and the true system state are drawn from the same distribution. The same is true for the ensemble forecasts and associated true states. This results in consistent ensemble forecasts, that is, events that are forecast to occur 10% of the time will actually occur 1 time in 10. Further, consistent analysis ensembles provide accurate analysis (and forecast) error covariance information. Such information can be useful for further noise reduction efforts and for sensitivity (singular vector) studies.

7. Acknowledgments

The authors wish to thank Kevin Judd for his significant contribution to the development of the concepts of finite-time stable and unstable sets, Pat McSharry for his detailed knowledge of the Ikeda system, Myles Allen for useful conversations regarding implications in geophysical systems, and two anonymous reviewers.

Appendix A: Finite-time Stable and Unstable Sets

Broadly, a finite-time stable/unstable set is one which tends to collapse onto a single state as one propagates the set forward/backward for a time τ. Such sets do not have a unique definition: three are given below followed by a necessary convergence property.

ϵ formulation

For $\epsilon > 0$,

$$\tilde{\mathbf{x}} \in \widetilde{W}_s(\mathbf{x}^t, \tau, \epsilon) \Leftrightarrow \|\tilde{\mathbf{x}}(\tau) - \mathbf{x}^t(\tau)\| < \epsilon \tag{A1}$$

for the finite-time stable set, and

$$\tilde{\mathbf{x}} \in \widetilde{W}_u(\mathbf{x}^t, \tau, \epsilon) \Leftrightarrow \|\tilde{\mathbf{x}}(-\tau) - \mathbf{x}^t(-\tau)\| < \epsilon \tag{A2}$$

for the finite-time unstable set. This is the definition used in this work. It requires that all points in the $\widetilde{W}_s(\mathbf{x}^t)/\widetilde{W_u}(\mathbf{x}^t)$ must lie within a distance ϵ of the true state at $\pm\tau$.

δ formulation

For $\delta > 0$,

$$\tilde{\mathbf{x}} \in \widetilde{W}_s(\mathbf{x}^t, \tau, \delta) \Leftrightarrow \sum_{i=0}^{\tau} \|\tilde{\mathbf{x}}(i) - \mathbf{x}^t(i)\| < \delta \tag{A3}$$

for the finite-time stable set, and

$$\tilde{\mathbf{x}} \in \widetilde{W}_u(\mathbf{x}^t, \tau, \delta) \Leftrightarrow \sum_{i=0}^{-\tau} \|\tilde{\mathbf{x}}(i) - \mathbf{x}^t(i)\| < \delta \tag{A4}$$

for the finite-time unstable set. Instead of requiring points to be within some ϵ of the true state at $\pm\tau$, the δ formulations specifies that the sum of the distances between the trajectories between 0 and $\pm\tau$ must be less than some δ.

λ formulation

For $c > 0$ and $\lambda > 0$,

$$\tilde{\mathbf{x}} \in \widetilde{W}_s(\mathbf{x}^t, \tau, c, \lambda) \Leftrightarrow \|\tilde{\mathbf{x}}(\tau) - \mathbf{x}^t(\tau)\| < c\|\tilde{\mathbf{x}}(0) - \mathbf{x}^t(0)\|e^{-\lambda\tau} \tag{A5}$$

for the finite-time stable set, and

$$\tilde{\mathbf{x}} \in \widetilde{W}_u(\mathbf{x}^t, \tau, c, \lambda) \Leftrightarrow \|\tilde{\mathbf{x}}(-\tau) - \mathbf{x}^t(-\tau)\| < c\|\tilde{\mathbf{x}}(0) - \mathbf{x}^t(0)\|e^{-\lambda\tau} \tag{A6}$$

for the finite-time unstable set. This definition specified that points in the $\widetilde{W_s}/\widetilde{W_u}$ must collapse onto the true trajectory exponentially fast.

Convergence property

Whichever definition of finite-time stable and unstable sets is employed, it is critical that it satisfy the following criterion: for all $\mathbf{x}$ there exits some τ_0 such that

$$\mathbf{x} \in \widetilde{W_s},\ \forall\ \tau \geq \tau_0\ \Leftrightarrow\ x \in W_s \tag{A7}$$

and

$$\mathbf{x} \in \widetilde{W_u},\ \forall\ \tau \leq \tau_0\ \Leftrightarrow\ x \in W_u. \tag{A8}$$

Equations A7 and A8 state that there exists some trajectory length beyond which the points in the $\widetilde{W_s}/\widetilde{W_u}$ are also in the infinite-time stable/unstable set.

Appendix B: The Ensemble Kalman Filter

The Kalman filter is used to obtain an optimal estimate (in the least-squares sense) of a system state given a model state and observations. Developed in 1960 (Kalman, 1960), it is a method that is capable of extracting the maximum amount of information available in an observation. This is achieved by combining model states and system observations and utilising information about the uncertainty in each.

A more detailed derivation of the Kalman filter can be found in Jazwinski (1969); only an outline is presented here. The Kalman filter's derivation is based on maximum likelihood. As such, both observational error and model error are assumed to be Gaussian, white noise processes.

A Kalman Filter variant called the ensemble Kalman Filter (Evensen and van Leeuwen, 1996) is utilised in this work. The EnKF is identical in application to the extended Kalman Filter introduced below. It differs only in its method for producing estimates of the forecast error covariance matrix and its handling of model error. Before discussing the EnKF directly, the extended Kalman filter is discussed.

Consider a model given by

$$\mathbf{x}^f(t) = \mathbf{F}(\mathbf{x}^a(t-1)) \tag{B1}$$

and a system

$$\mathbf{x}^t(t) = \mathbf{G}(\mathbf{x}^t(t-1)). \tag{B2}$$

The model need not be a perfect representation of the system, but it is assumed that any discrepancy between the model and the system can be expressed as

$$\mathbf{G}(\mathbf{x}^t(t-1)) = \mathbf{F}(\mathbf{x}^t(t-1)) + \boldsymbol{\eta}(t) \tag{B3}$$

where $\boldsymbol{\eta}$ is a Gaussian white noise process that is intended to account for all model error. The expected value of $\boldsymbol{\eta}$ is taken as zero, $E(\boldsymbol{\eta}) = q = 0$, with a model error variance of $E(\boldsymbol{\eta}\boldsymbol{\eta}^T) = \mathbf{Q}$.

Observations take the form

$$\mathbf{y}^o(t) = \mathbf{H}(t)\mathbf{x}^t(t) + \boldsymbol{\epsilon}_o \tag{B4}$$

where $E(\boldsymbol{\epsilon}_o) = 0$ and $E(\boldsymbol{\epsilon}_o\boldsymbol{\epsilon}_o^T) = \mathbf{R}$. The operator $\mathbf{H}(\mathbf{t})$ provides a transformation from model space to observation space. It accounts for such things as observation grids which do not correspond to the model grid and performing the transformation model space to observation space.

The Kalman filter produces an estimate of the system state, $\mathbf{x}^a(t)$, given a forecast, $\mathbf{x}^f(t)$, observations, $\mathbf{y}^o(t)$ and the associated model, forecast and observational uncertainties. The equation for the analysis takes the form

$$\mathbf{x}^a(t) = \mathbf{x}^f(t) + \mathbf{K}(t)[\mathbf{y}^o(t) - \mathbf{H}(t)\mathbf{x}^f(t)]. \tag{B5}$$

where $\mathbf{K}(\mathbf{t})$ is a 'gain' term specifying the relative weight given to the forecast and the observations.

Assuming the analysis error and the model error are independent, that is, $E(\boldsymbol{\eta}\boldsymbol{\epsilon}_o) = 0$, expressions can be written for the forecast error covariance, $\mathbf{P}^f(t)$ and the analysis error covariance, $\mathbf{P}^a(t)$

$$\mathbf{P}^f(t) = \mathbf{F}'\mathbf{P}^a(t-1)\mathbf{F}'^T + \mathbf{Q} \tag{B6}$$

and

$$\mathbf{P}^a(t) = \mathbf{P}^f(t) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}^f(t) \tag{B7}$$

where $\mathbf{F}'$ is the linearisation of $\mathbf{F}$.

The application of the extended Kalman filter is then as follows

$$\mathbf{x}^f(t+1) = \mathbf{F}(\mathbf{x}^a(t)) \tag{B8}$$

$$\mathbf{P}^f(t) = \mathbf{F}'\mathbf{P}^a(t-1)\mathbf{F}'^T + \mathbf{Q} \tag{B9}$$

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)^T(\mathbf{H}(t)\mathbf{P}^f(t)\mathbf{H}(t)^T + \mathbf{R}(t))^{-1} \tag{B10}$$

$$\mathbf{x}^a(t) = \mathbf{x}^f(t) + \mathbf{K}(t)[\mathbf{y}^o(t) - \mathbf{H}(t)\mathbf{x}^f(t)] \tag{B11}$$

$$\mathbf{P}^a(t) = \mathbf{P}^f(t) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}^f(t), \tag{B12}$$

A forecast of the system state is made from the current analysis (equation B8), as is a forecast of the analysis error covariance (equation B9). Using the forecast of the analysis error covariance and the current observational uncertainty, the gain matrix is calculated (equation B10), and from it and the predicted model state, the new analysis is determined (equation B11). Finally, using the gain term and the forecast of the analysis error covariance, the best estimate of the current analysis error covariance is computed (equation B12).

The ensemble Kalman filter was introduced by Evensen (1994) as a method for avoiding the expensive calculation of the forecast error covariance matrix (equation B9) necessary for the extended Kalman filter derived above. Instead of mapping an estimate of the analysis error covariance forward under a model's linear uncertainty propagator, the EnKF uses the full model equations to integrate forward a number of initial conditions from which the error covariances can be estimated. Similarly, instead of estimating the analysis error covariance with equation B12, its value can be determined from the analysed states of the ensemble members after an application of equation B11 to each.

In the extended Kalman filter, the model error term, $\mathbf{Q}$, is added to the estimate of the propagated forecast error covariance. Such a simple treatment of model error is not possible for the EnKF. Model error must be accounted for in the integration of the ensemble of initial conditions. Evensen argues that this can be accomplished by forcing the model with "pseudo random fields with specified variance and covariance," and provides an appendix outlining the construction of such fields. In the current work, a perfect model is employed and no model error term is necessary.

References

Arrowsmith, D. K., and Place, C. M. 1990. *An introduction to dynamic systems.* Cambridge: Cambridge University Press

Burgers, G., van Leeuwen, P. J., and Evensen, G. 1998. Analysis Scheme in the Ensemble Kalman Filter. Monthly Weather Review, 126,1719–1724

Daley, R. 1991. *Atmospheric Data Analyis.* Cambridge University Press

Davies, M. 1994. Noise reduction schemes for chaotic time series. Physica D, 79,174–192

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophyical Research, 99(C5),10,143–10,162

Evensen, G., and van Leeuwen, P. J. 1996. Assimilation of Geosat Data for the Agulhas Current Using the Ensemble Kalman Filter with a Quasigeostrophic Model. Monthly Weather Review, 124,85–96

Ghil, M., K. Ide, A. Bennett, Courtier, P., Kimoto, M., Nagata, M., Saiki, M., and Sato, N. (eds) 1997. *Data Assimilation in Meteorology and Oceanography: Theory and Practice.* Vol. 75

Gilmour, I., and Smith, L.A. 1997. Enlightenment in Shadows. *In:* Kadtke, J.B, and Bulsara, A. (eds), *Applied Nonlinear Dynamics and Stochastic Systems Near the Millenium*

Grassberger, P., Hegger, R., Kantz, H., Schaffrath, C., and Schreiber, Th. 1993. On noise reduction methods for chaotic data. Chaos, 3,127–140

Hansen, J. A. 1998. *Adaptive Observations in Spatially-extended, Nonlinear Dynamical Systems.* Ph.D. Thesis, Oxford University

Hirsch, W. M., and Pugh, C. C. 1970. Stable manifolds and hyperbolic sets. Proc. Symp. Pure Math., 14,133–163

Houtekamer, P. L., and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. Monthly Weather Review, 126(3),796–811

Ide, K., Courtier, P., Ghil, M., and Lorenc, A.C. 1997. Unified notation for data assimilation: Operational, sequential and variational. *Pages 181–189 of:* Ghil, M., Ide, K., Bennett, A., Courtier, P., Kimoto, M., Nagata, M., Saiki, M., and Sato, N. (eds), *Data Assimilation in Meteorology and Oceanography: Theory and Practice.* A Collection of Papers Presented at the WMO Second

International Symposium on Assimilation of Observations in Meteorology and Oceanography, 13-17 March 1995, Tokyo, Japan

Ikeda, K. 1979. Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system. Opt. Commun., 30,257

Jazwinski, A. H. 1969. Adaptive Filtering. Automatica, 5,475–485

Judd, K. 2000. *University of Western Australia, Department of Mathematics and Statistics.* Personal communication

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. J. Basic Eng., 82,35–45

Kostelich, E. J., and Yorke, J. A. 1988. Noise reduction in dynamical systems. Phys. Rev. A, 38,1649–1652

Lorenc, A. C. 1986. Analysis method for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society, 112,1177–1194

McSharry, P. E. 1999. *Innovations in Consistent Nonlinear Deterministic Prediction.* Ph.D. thesis, University of Oxford

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T, and Flannery, B. P. 1986. *Numerical Recipes in FORTRAN.* Cambridge University Press

Smith, L. A., Ziehmann, C., and Fraedrich, K. 1999. Uncertainty dynamics and predictability in nonlinear systems. Quarterly Journal of the Royal Meteorological Society, 125,2855–2888

Strang, G. 1988. *Linear Algebra and its Applications.* Saunders

*

List of Figures

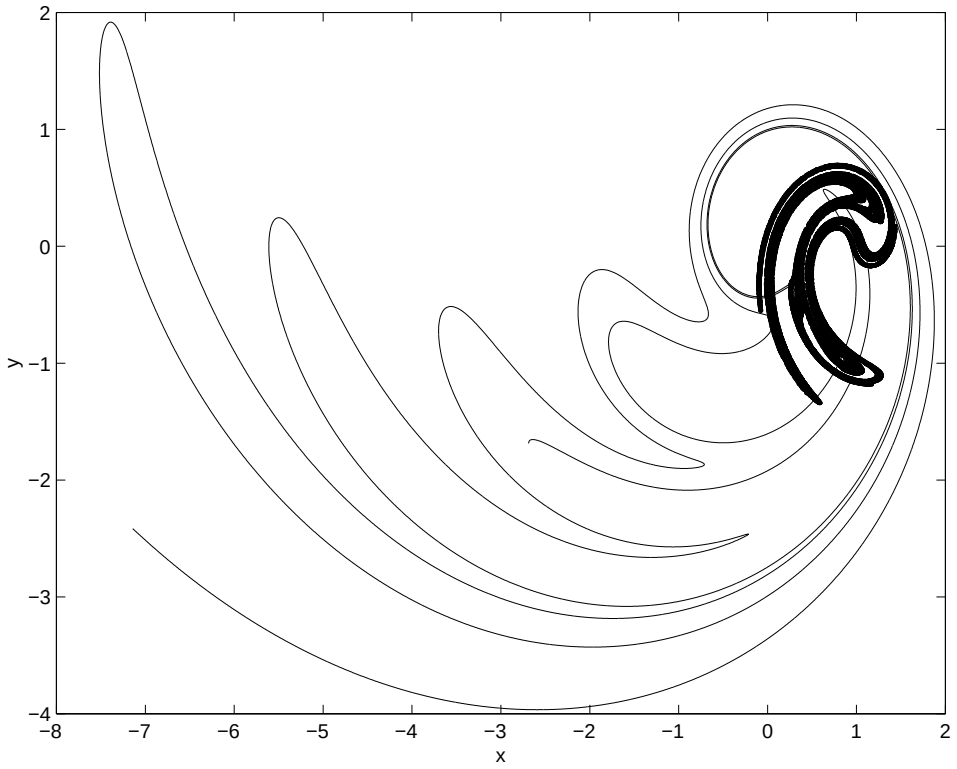
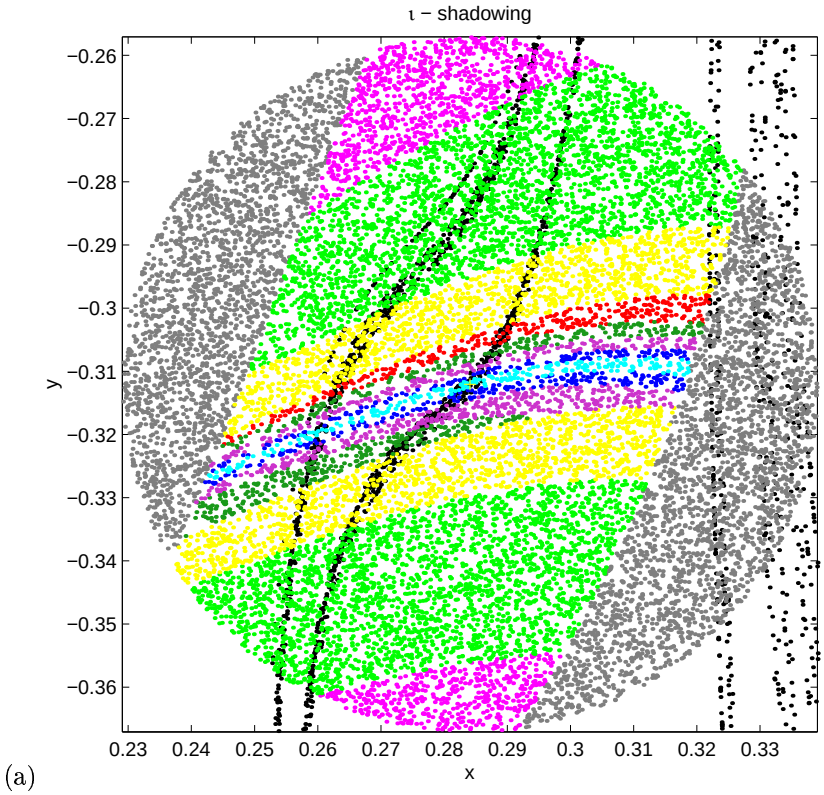
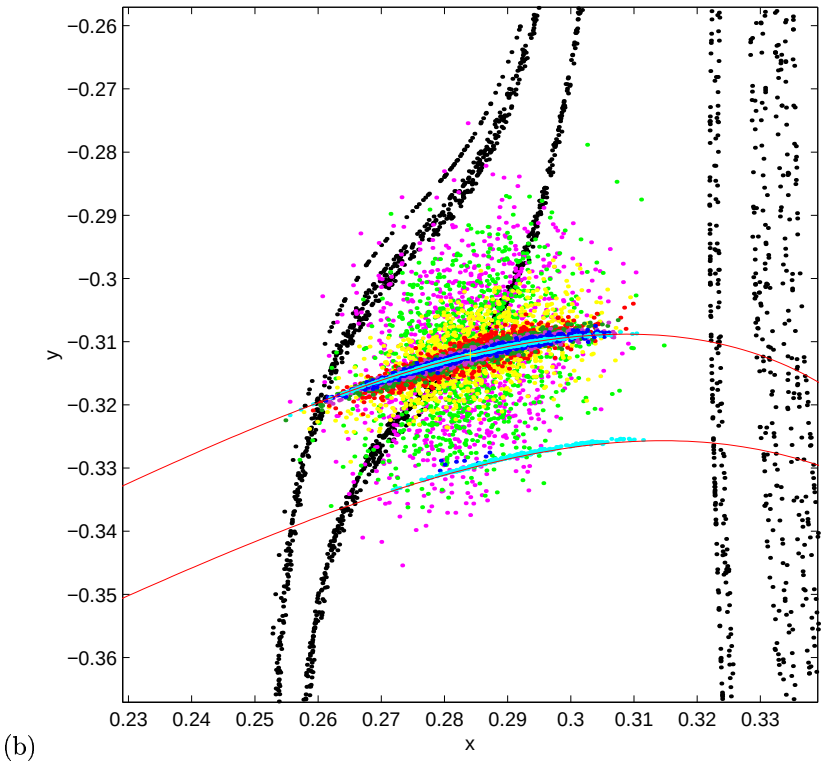


Figure 1. The Ikeda attractor and the structure of the $\widetilde{W_s}(\mathbf{x}^t)$ for the point in figures 2 and 3. The Ikeda attractor is seen as the dots on the right side of the picture while the $\widetilde{W_s}(\mathbf{x}^t)$ is shown as the looping line. The $\widetilde{W_s}(\mathbf{x}^t)$ has a rich structure; each time it loops back through the region surrounding the point about which it is constructed it will cause a valley in cost function space.

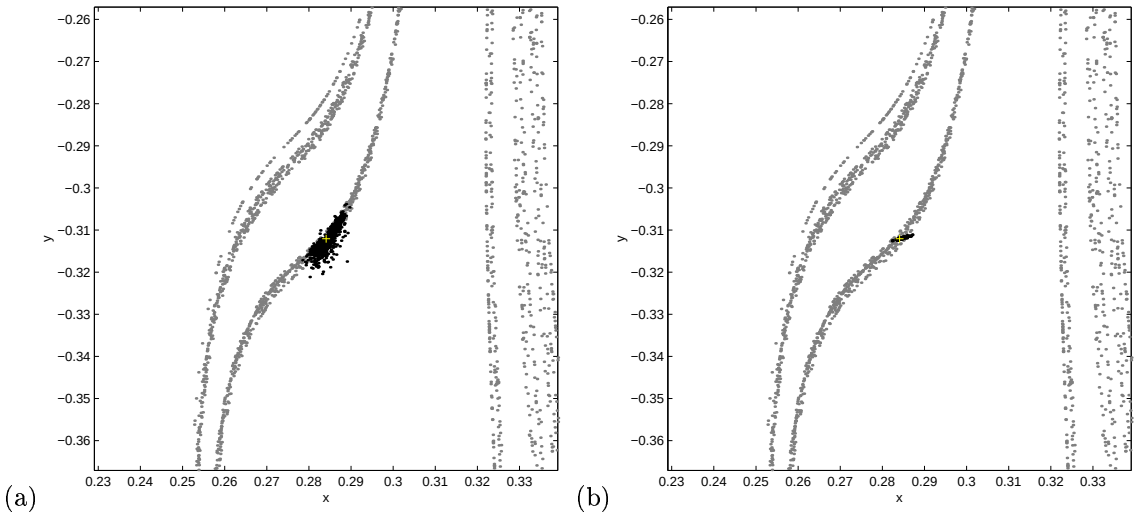


(a)



(b)

Figure 2. Noise reduced distributions produced by a) ι-shadowing, and b) 4d-Var. See text for details.



(a) (b)

Figure 3. Noise reduced distributions produced by a) the EnKF and b) the combination of the $N_a = 10$ 4d-Var PDF (figure 2b) and the EnKF PDF. In both panels, the grey dots are points on the Ikeda attractor.

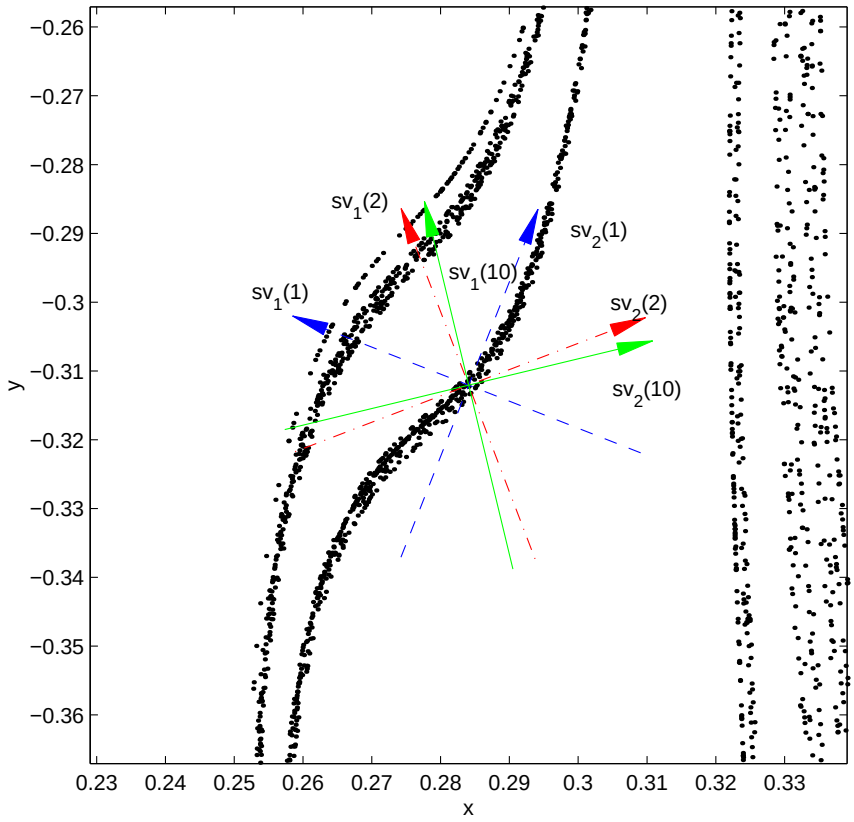


Figure 4. Orientation of the one, two and ten step initial singular vectors. The orientation of the second one step singular vector $\mathbf{sv_2(1)}$, the shrinking direction, is the same as the orientation of the points which ι-shadow after one step (figure 2a) and the same as the primary orientation of the 4d-Var ensemble for $N_a = 1$ (figure 2b). There is a large degree of rotation between the one step and the two step singular vectors. This accounts for the contraction along the $\widetilde{W}_s(\mathbf{x}^t)$ seen in both the ι-shadowing results and the 4d-Var results of figure 2.

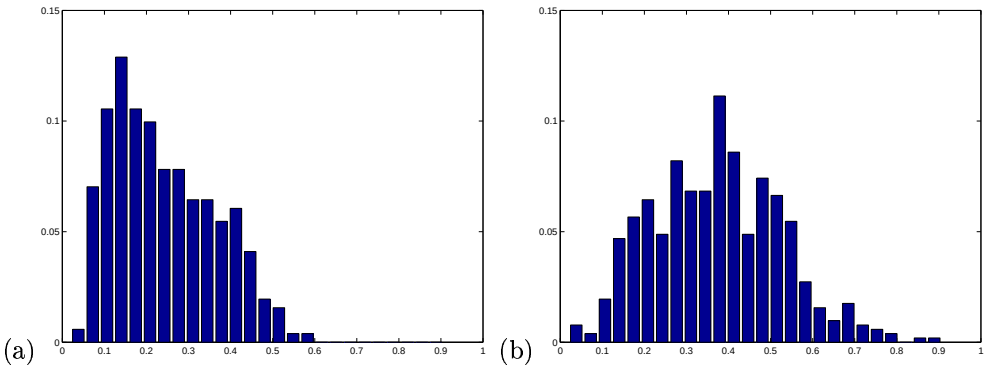


Figure 5. Distribution of the area containing 90% of the probability in noise reduced analysis PDFs for a) 4d-Var, and b) the EnKF over 1024 different initial conditions. The areas are normalised by the area containing 90% of the observational probability. The areas produced by 4d-Var are smaller than those produced by the EnKF, consistent with the 4d-Var ensembles lying along a line given by the $\widetilde{W}_s(\mathbf{x}^t)$ while the EnKF ensembles reflect the underlying attractor structure. The combination of the 4d-Var and EnKF PDFs results results in an improvement of nearly two orders of magnitude. Although the histogram is not shown, all areas of the combined results would fall into the first bin of panels a) and b).

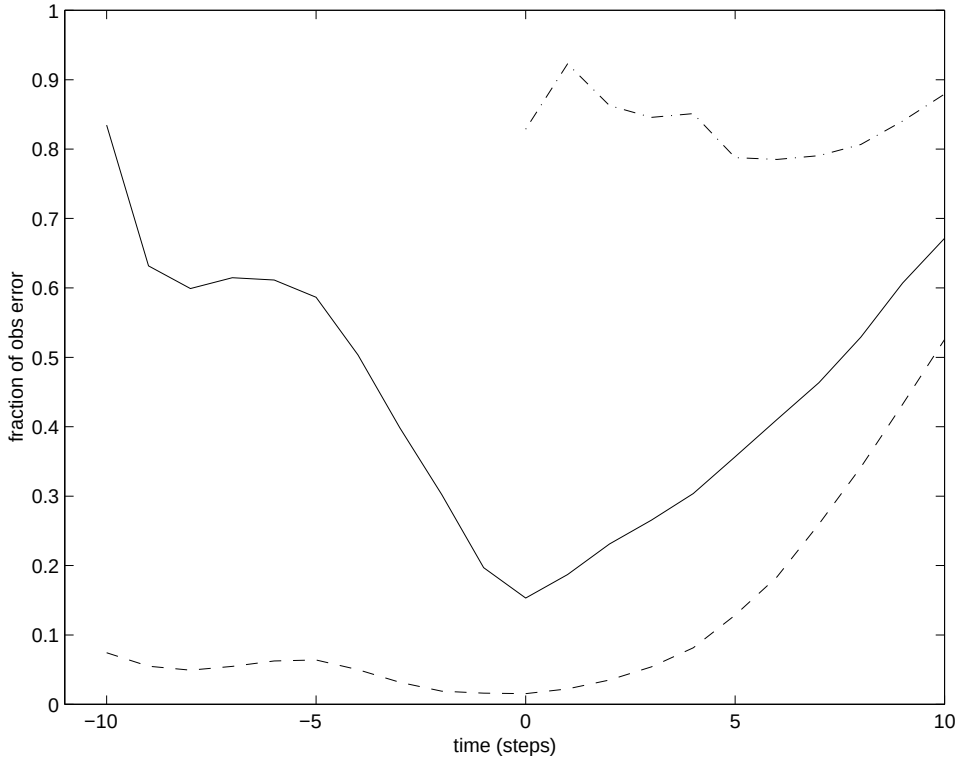


Figure 6. Mean distance between the ensemble mean and truth for $N_a = 10$ 4d-Var ensembles (solid), the EnKF Ensembles (dot-dashed), and the combined 4d-Var/EnKF ensembles (dashed) over 1024 different initial conditions. Analysis ensembles for 4d-Var and combined 4d-Var/EnKF were constructed and integrated forward 20 steps. The first 10 steps covered the assimilation window, with the next 10 steps being proper predictions. As the EnKF ensembles do not utilise a 'window' they were integrated over only the 10 forecast steps. All errors are normalised by the error that results from performing no noise reduction and simply constructing an ensemble about an observation. It is seen that the combined ensemble outperforms the 4d-Var ensemble at every step.