Efficient fully nonlinear data assimilation for geophysical fluid dynamics

Article

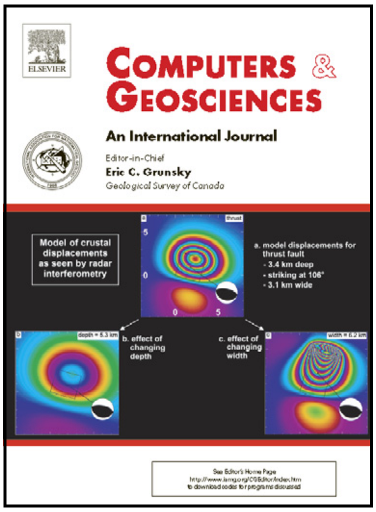
Accepted Version

Van Leeuwen, P. J. and Ades, M. (2013) Efficient fully nonlinear data assimilation for geophysical fluid dynamics. Computers & Geosciences, 55. pp. 16-27. ISSN 0098-3004 doi: 10.1016/j.cageo.2012.04.015 Available at https://centaur.reading.ac.uk/35769/

To link to this article DOI: http://dx.doi.org/10.1016/j.cageo.2012.04.015

Publisher: Elsevier

Author's Accepted Manuscript



www.elsevier.com/locate/cageo

Efficient fully nonlinear data assimilation for geophysical fluid dynamics

Peter Jan van Leeuwen, Melanie Ades

PII: S0098-3004(12)00138-0
DOI: http://dx.doi.org/10.1016/j.cageo.2012.04.015
Reference: CAGEO2889

To appear in: *Computers & Geosciences*

Received date: 6 January 2012
Revised date: 16 April 2012
Accepted date: 19 April 2012

Cite this article as: Peter Jan van Leeuwen and Melanie Ades, Efficient fully nonlinear data assimilation for geophysical fluid dynamics, *Computers & Geosciences,* http://dx.doi.org/10.1016/j.cageo.2012.04.015

Efficient fully nonlinear data assimilation for geophysical fluid dynamics

Peter Jan van Leeuwen and Melanie Ades

Data Assimilation Research Centre, Department of Meteorology, University of Reading, PO Box 243, Reading, RG6 6BB, United Kingdom

Abstract

A potential problem with Ensemble Kalman Filter is the implicit Gaussian assumption at analysis times. Here we explore the performance of a recently proposed fully nonlinear particle filter on a high-dimensional but simplified ocean model, in which the Gaussian assumption is not made. The model simulates the evolution of the vorticity field in time, described by the barotropic vorticity equation, in a highly nonlinear flow regime. While common knowledge is that particle filters are inefficient and need large numbers of model runs to avoid degeneracy, the newly developed particle filter needs only of the order of 10-100 particles on large scale problems. The crucial new ingredient is that the proposal density cannot only be used to ensure all particles end up in high-probability regions of state space as defined by the observations, but also to ensure that most of the particles have similar weights. Using identical twin experiments we found that the ensemble mean follows the truth reliably, and the difference from the truth is captured by the ensemble spread. A rank histogram is used to show that the truth run is indistinguishable from any of the particles, showing statistical consistency of the method.

Key words: Data Assimilation, Inverse Modeling, Particle Filter, Ensemble Kalman Filter

1. Introduction

Numerical models for simulation and prediction of the evolution of systems in the geosciences are becoming ever more complex. While relatively simple linear balances tend to dominate the systems at large scales, with increasing resolution more and more nonlinear processes are involved. Furthermore, with the coupling of many physical, chemical and biological systems extremely complex behaviour with highly nonlinear feebacks has to be simulated.

To the extent that these flows are initial value problems our incomplete knowledge of the exact initial conditions leads to incomplete knowledge of the evolution of the system. This forces us to think in terms of uncertainty, which can be described in probabilistic terms. The evolution equations for the related probability densities have been known for decades (see e.g. Jazwinski, 1970). If the system is Markovian, our present knowledge of the system in the form of a probability density function evolves according to the Kolmogorov or Fokker-Plank equation. This theory can be applied for small dimensional systems, but the systems we study in the geosciences are not so.

When observations of the system are available, their information on the system can be incorporated using Bayes Theorem, in which the prior probability density function (pdf from now on), representing our prior knowledge, is multiplied by the likelihood, i.e. the probability density of the observations given a specific model state. This then leads to the so-called posterior pdf, that describes our updated knowledge of the system. This process of updating the prior pdf with observations is called *data assimilation*, and its goal is to determine properties of this posterior pdf. It should be realised that this posterior pdf is unlikely to be ever at our disposal in full because the size of the state space is huge, typically 100 million for numerical weather prediction. We can only infer statistical moments like mean, covariance, percentiles, and modes.

It is stressed here that the data-assimilation problem as specified above is a multiplication problem and not an inverse problem: Bayes Theorem (see equation (1)) shows that one has to *multiply* the prior pdf with the likelihood to obtain the posterior pdf. There is no inversion needed to obtain the posterior. Also parameter estimation falls in this framework: the prior pdf of the parameters is updated through multiplication with the likelihood to obtain the posterior pdf of the parameters. Obviously, one needs the relation between the parameters and the observations in the likelihood, and that typically involves integrating a full numerical model, but that doesn't make the problem an inverse problem. The emphasis of this paper is on estimation of the pdf of the model variables represented by a state vector, and not on that of parameters.

When the posterior pdf is unimodal or the majority of the posterior probability mass is concentrated around a mode of the posterior pdf it makes sense to concentrate on the mode of the posterior pdf. The problem of finding the mode is usually formulated as an inverse problem, i.e. a problem in which a matrix has to be inverted, although there is no necessity to do so. Examples are variational algorithms that try to find the mode by exploring the gradient of the log of the posterior pdf. In the geosciences these methods are known as e.g. gradient methods, 3DVar, 4DVar (Talagrand and Courtier, 1987), representer method (Bennett, 1992), PSAS (Courtier, 1997), depending on details of the solution method. The Ensemble Kalman filter (Evensen, 1994, Burgers et al., 1998) is slightly different in that it tries to find the posterior mean (the least-squares estimate, which is the mean by definition), but because of the linearity assumptions in the Kalman filterit is assumed implicitly that the mean is close to the mode. This has led to confusion that data assimilation is all about finding this mode in the geophysical and the so-called inverse-problem communities, and in some cases hampered progress to more nonlinear multimodal problems.

In this paper we propose solutions to highly nonlinear high-dimensional data-assimilation problems. Our stating point is the particle filter (e.g. Gordon et al., 1993), in which an ensemble of model runs is performed, representing our prior knowledge of the system. Each ensemble member, or particle, is weighted with its distance to observations when these become available. The distance norm is determined by the value of the pdf of the observations given this particle, so the likelihood of the observations given this particle. The weights are the relative probabilistic weights of the particles, so e.g. the mean of the ensemble now becomes a weighted mean in the posterior pdf.

It is well known that in systems with moderate dimensions, say of order 10 and higher, particle filters tend to be degenerate, meaning that the weights vary too much. Typically after one or a few updates with observations the relative weight of one particle is close to one, while that of all others is very close to zero. This means that e.g. a weighted mean is in fact based on only one particle, so all statistical information in the ensemble is lost. To prevent this from happening several methods have been proposed, starting from resampling (Gordon et al., 1993) to more complicated or approximating solutions (see e.g. Doucet et al., 2001, and Van Leeuwen, 2009, for a review of applications in the geosciences). None of the proposed methods is applicable to systems with dimensions larger than say of order 100, without having to need millions of particles, so millions of model integrations. As mentioned, our goal is perhaps 100 million dimensional systems, and this number keeps on increasing with the size and speed of supercomputers.

In this paper we discuss a new particle filter methodology that is applicable to systems of much higher dimension, and which up to the dimensions we tested it on has perfect scaling, i.e. the number of particles is independent of the dimension of the state vector. The secret is a proper use of the proposal density, that allows much more freedom than perhaps anticipated in earlier work. Typically, the proposal density has been used to steer the particles to high-probability areas as defined by the observations in state space, but when the number of independent observations is large, the relative weights of the particles will vary enormously, leading to degeneracy. Here we exploit the fact that the proposal density can in addition be used to obtain similar relative weights for the particles, thus avoiding degeneracy. The method is introduced in Van Leeuwen (2010), and Van Leeuwen (2011) discussed applications to systems of up to 1000 dimensions using only about 20 particles. In this paper, the method is outlined and its performance on a geophysical system with about 65,000 dimensions is demonstrated.

The paper is organised as follows. The next section discusses Particle Filtering in general, followed by a section on the new method. It is highlighted why other particle filter formulations fail, and how the new method can be successful. Then the numerical model simulating the barotropic vorticity equation is described, followed by initial results when applying the new particle filter to that system. A concluding chapter closes the paper.

2. Particle filtering

The probability density function (pdf) of the state vector is represented, and approximated, by a discrete set of delta functions centred around a set of model states, called the particles. Using this representation of the prior pdf of the model in Bayes theorem

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)\, dx} \tag{1}$$

where x is the state vector, and y is the observation vector, one finds:

$$p(x|y) = \sum_{i=1}^{N} w_i \delta(x - x_i) \tag{2}$$

in which the weights w_i are related to how close each particle is to the observations:

$$w_i = \frac{p(y|x_i)}{\sum_{j=1}^{N} p(y|x_j)} \tag{3}$$

The density $p(y|x_i)$ is the likelihood, i.e. the probability density of the observations given the model state x_i. It is related to the fact that we cannot make perfect observations, any observation comes with a measurement error, and hence this density is the pdf of the errors in the observations due to the measurement process. In the data-assimilation problem it is given, and often assumed to be Gaussian:

$$p(y|x_i) = A \exp\left[-\frac{1}{2}(y - H(x_i))^T R^{-1}(y - H(x_i))\right] \tag{4}$$

in which $H(x_i)$ is the measurement operator, which projects the model state onto the observation space, R is the observation error covariance matrix, and A is a normalisation constant.

Unfortunately, weights vary wildly, and even when resampling is applied, only a few particles will have relatively high weight, so will have any statistical significance. This as called *filter degeneracy* and is a very serious problem in standard particle filtering (Snyder et al, 2008). Several methods have been proposed to solve this problem (see review for the geosciences by Van Leeuwen, 2009), but none of these is directly applicable to large-dimensional geophysical problems.

To understand why this is the case, consider the following. Since the particles x_i are not the evolution of the true system, the distance in observation space between the observation y_j and the particle equivalent $H_j(x_i)$ will on average be similar to or larger than a typical observation error (from the Cauchy-Swartz inequality), so it can be expected that $(y_j - H_j(x_i))^T R_{jj}^{-1}(y_j - H_j(x_i))$ will typically be similar to or larger than 1 for each observation y_j. Assuming M independent observations, $(y - H(x_i))^T R^{-1}(y - H(x_i))$ is expected to be of order M or larger. So, to start with, the likelihood for each particle will be fairly small.

However, the particle filters works with relative weights, so we need to address the variation of the likelihood with the particles. Let us assume that the particles are drawn from a Gaussian with covariance B centred around the true state. Clearly, the larger B, the larger the variation in the weights of the particles. Let us now assume, to illustrate the argument, that HBH^T is of similar magnitude as R. In that case, the argument of the exponent in the likelihood is a χ-squared variable with M degrees of freedom. Such a variable has mean M, and standard deviation $\sqrt{2M}$. This means that the relative weights of the particles differ by a factor $\exp\sqrt{2M}$. Assuming a moderate 50 independent observations, the weights will vary by a factor $\exp(50) \approx 5.0\ 10^{21}$, so the particle filter will be degenerate when the number of independent observations grows, and serious improvement is needed.

3. The new method

The new method that will be explored consists of two ingredients. The first ingredient is that the particles are steered towards the future observations by choosing a specific form of model forcing that tends to pull the model towards the observations. This is an old idea in particle filtering, and has been explored in the Lorenz 1963 and 1996 models in Van Leeuwen (2010, 2011). Assume the model equation to be written as

$$x^n = f(x^{n-1}) + \beta^n \tag{5}$$

in which $f(..)$ denotes the deterministic part of the model and β^n is the stochastic part, and n is the time index. Instead of using this, the model equation is modified to:

$$x^n = f(x^{n-1}) + \hat{\beta}^n + K(y^m - H(x^{n-1})) \tag{6}$$

in which $\hat{\beta}^n$ is random forcing which might have different characteristics from the original random forcing, and y^m denotes future observations at time $m > n$. The main difference with the original model equation is the relaxation term that tends to pull the particle to the future observations y^m with a strength given by matrix K. This relaxation matrix will depend on the application, and an example is given below. This looks like cheating in the sense that the model forcing is not chosen from the probability density of the model error, but as something that we like better. Also, the different particles will have different strength of the 'pulling' term dependent on how far they are from the future observations, so we seem to loose control over the statistical meaning of each particle. However, this different forcing can be compensated for exactly by changing the relative weights of the particles.

In particle filter jargon, we have implemented a proposal transition density instead of using the original transition density. The original transition density is denoted as $p(x^n|x^{n-1})$ specifying how probable state x^n is given state x^{n-1} at the previous time step. For the original model equation (5) this density is given by the pdf of β^n. If the β^n are Gaussian distributed as $N(0, Q)$, we find:

$$p(x^n|x^{n-1}) = N(f(x^{n-1}), Q) \tag{7}$$

The proposal transition density can be written as, assuming a Gaussian distribution for the $\hat{\beta}^n$ with covariance $\hat{Q}$:

$$q(x^n|x^{n-1}, y^m) = N(f(x^{n-1}) + K(y^m - H(x^{n-1})), \hat{Q}) \tag{8}$$

Also $\hat{Q}$ can be problem dependent. In the example discussed below we choose it equal to Q. Note that the proposal transition density does depend on the future observations. Furthermore, the relaxation term is part of the deterministic proposal model, since the observations are given.

The question now is how the weights are affected when we arrive at the observations at time m. To this end, let us write the prior pdf at time m as:

$$\begin{aligned} p(x^m) &= \int p(x^m, x^{m-1}, ..., x^0)\, dx^{m-1}...dx^0 \\ &= \int p(x^m|x^{m-1})...p(x^1|x^0)p(x^0)\, dx^{0:m-1} \end{aligned} \tag{9}$$

in which we exploited the Markovian property of the model, and introduced the shorthand notation $dx^{n-1}...dx^0 = dx^{0:n-1}$. Furthermore, the previous set of observations was present at time 0 in this notation. The integrand can be multiplied and divided by the proposal transition densities to find:

$$p(x^m) = \int \frac{p(x^m|x^{m-1})}{q(x^m|x^{m-1}, y^m)} ... \frac{p(x^1|x^0)}{q(x^1|x^0, y^m)} q(x^m|x^{m-1}, y^m)...q(x^1|x^0, y^m)p(x^0)\, dx^{0:n-1} \tag{10}$$

In the original model we draw random samples from $p(x^0)$ and from each of the $p(x^i|x^{i-1})$ as indicated above. Using the proposed model we draw samples from $p(x^0)$ and from the proposal transition densities $q(x^i|x^{i-1}, y^m)$. Doing the latter, realising that this creates delta functions for times 0 to $n-1$, we can perform the integrations and find for the prior at time m:

$$p(x^m) = \sum_{i=1}^{N} \hat{w}_i \delta(x - x_i) \tag{11}$$

in which the weights are given as:

$$\hat{w}_i = \frac{p(x_i^m|x_i^{m-1})}{q(x_i^m|x_i^{m-1}, y^m)} \cdots \frac{p(x_i^1|x_i^0)}{q(x_i^1|x_i^0, y^m)} \tag{12}$$

So where we had equally weighted particles in the standard particle filter for the prior, we now have weighted particles. These weights are related to the fact that we changed the model equations. They specify how probable the move from x^n to x^{n-1} is in the original model, normalised by that probability in the modified model.

Finally, to find the full posterior weights we use Bayes theorem to include the likelihood, leading to

$$w_i \propto p(y^m|x_i^m) \frac{p(x_i^m|x_i^{m-1})}{q(x_i^m|x_i^{m-1}, y^m)} \cdots \frac{p(x_i^1|x_i^0)}{q(x_i^1|x_i^0, y^m)} \tag{13}$$

Making sure that all particles end up relatively close to the observations still does not avoid wildly varying weights in large-dimensional systems. Clearly, ending up close to the observations reduces the variance in the likelihood weights, but the variance in the weights related to the proposal density are nonzero, and can be substantial.

The second new ingredient is that we ensure that all posterior weights are of equivalent size. This is achieved in two stages: first, use the scheme mentioned above for all time steps up to time $n-1$ and perform a deterministic time step with each particle that ensures that most of the particles have equal weight; and secondly, add a very small random perturbation to ensure that Bayes theorem is satisfied. There are many ways to accomplish both stages.

Let us assume that the observation errors and the errors in the model equations are Gaussian distributed. The weights can be written as:

$$w_i \propto p(y^m|x^m) \frac{p(x^m|x_i^{m-1})}{q(x^m|x_i^{m-1}, y^m)} w_i^{rest} \tag{14}$$

leaving the last time step open. w_i^{rest} contains the weights from all time steps up to time $n-1$, which are now given (we have done all these steps). Ignoring the proposal transition density part for the moment, making the weight of each particle equal to $\exp(-C)$, say, leads to the following quadratic equation for particle x_i at time m:

$$\frac{1}{2}(x^m - f(x_i^m))^T Q^{-1}(x^m - f(x_i^m)) + \\ \frac{1}{2}(y - Hx_i^m)^T R^{-1}(y - Hx_i^m) - \log(w_i^{rest}) = C \quad (15)$$

Now any quadratic form has a minimum, and depending on the value for C this equations has two, one, or zero real roots for a one dimensional system. Zero roots means that the particle is unable to reach this specified weight $\exp(-C)$; the w_i^{rest} factor for such a particle is too low. Clearly, we don't want the weight of each particle to be the same as the worst particle. We have chosen here a weight C such that 80% can reach it, and the other 20% will be ignored for now. They will re-enter the ensemble via the resampling step later on.

Once C is chosen, an infinite number of solutions exist if the dimension of the system is larger than 1. A simple choice is to enforce

$$x_i^m = f(x_i^{m-1}) + \alpha_i M(y^m - H(f(x_i^{m-1}))) \quad (16)$$

in which $M = QH^T(HQH^T + R)^{-1}$, Q is the error covariance of the model errors, and R is the error covariance of the observations. α_i is a scalar that is determined from equation (15), and we obtain for each α_i, (see Van Leeuwen, 2010, 2011)]:

$$\alpha_i = 1 - \sqrt{1 - b_i/a_i} \quad (17)$$

in which $a_i = 0.5x_i^T R^{-1} HKz$ and $b_i = 0.5x_i^T R^{-1} x_i - C - \log w_i^{rest}$. Here $z = y^m - H(f(x_i^{m-1}))$, C is the chosen weight level, and w_i^{rest} denotes the relative weights of each particle i up to this time step, related to the proposal density explained above.

Note that the last time step so far is a purely deterministic step: we have chosen C, and directly calculated x_i^m. Of course, this last step towards the observations cannot be fully deterministic, as can be seen from Eq. (13). A deterministic proposal would mean that the proposal transition density q can be zero while the target transition density p is non zero, leading to division by zero: a deterministic move the transition density is a delta function. In the example presented below the proposal transition density was chosen to be a Gaussian. Since the weights have q in the denominator a draw from the tail of a Gaussian could lead to a very high weight for a particle that is perturbed by a relatively large amount, resulting in the opposite of the intended outcome. We didn't encounter this problem in this experiment.

To avoid this potential problem q could be chosen in the last step before the observations as a mixture density

$$q(x^m|x') = (1-\gamma)U(-a,a) + \gamma N(0,a^2) \tag{18}$$

in which x' is the particle after the deterministic step outlined above. A draw from this density would be performed as follows. First, we determine from which density U, or N, we will draw the stochastic perturbation, e.g. by drawing u from a uniform density $U[0,1]$ and if $u < \gamma$ we draw from the normal density $N(0,a^2)$, and we draw from the uniform density $U(-a,a)$ otherwise. By choosing γ very small we most likely draw from the uniform density $U(-a,a)$. For small a we can completely control the size of the stochastic perturbation to the state vector. If by chance we have to choose from $N(0,a^2)$ we most likely draw from near the peak of this Gaussian. It is very unlikely to draw from the Gaussian and at the same time draw from the tail of that Gaussian. It is mentioned that γ can be made dependent on the number of particles to control the number of times we actually draw from the Gaussian, and keep that number small.

4. The barotropic vorticity equation and statistical set up

The barotropic vorticity equation describes how the vorticity field ζ changes with time through advection of the vorticity field by the velocity field:

$$\frac{\partial q}{\partial t} + u\frac{\partial q}{\partial x} + v\frac{\partial q}{\partial y} = \beta \tag{19}$$

in which u the eastward and v the northward velocity, and in which we included a random forcing β. The vorticity field is related to the velocity field as

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 0 \tag{20}$$

Because the divergence of the horizontal velocity field is zero:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{21}$$

a streamfunction can be defined as

$$u = -\frac{\partial \psi}{\partial y} \qquad v = \frac{\partial \psi}{\partial x} \tag{22}$$

Combining this with the evolution equation for the vorticity field leads to the following set of equations that have to be solved at every time step:

$$\begin{aligned} \frac{\partial q}{\partial t} - \frac{\partial \psi}{\partial y}\frac{\partial q}{\partial x} + \frac{\partial \psi}{\partial x}\frac{\partial q}{\partial y} &= \beta \\ q = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} & \end{aligned} \tag{23}$$

This set of equations is solved on a double periodic domain of 256 by 256 grid points and grid spacing $\Delta x = \Delta y = 1/256$, leading to a state dimension of close to 65,000. At each time step the vorticity field is updated using a semi-Lagrangian scheme with time step $\Delta t = 0.04$, followed by an update of the streamfunction via an inversion of the second equation using FFT's.

The stochastic term is chosen from a multivariate Gaussian with mean zero, variance 0.01, and a Gaussian spatial correlation with decorrelation lengthscale 4 gridpoints. It is integrated using a simple Euler scheme. Because we are not interested in the specific stochastic evolution, but in the overall properties of the stochastic equation, the accuracy is $O(\Delta t)$. The initial condition was a random vorticiy field with nondimensional amplitude 1 and spatial decorrelation length of 10 grid points. This initial condition results in highly nonlinear turbulent flow structures. Without the random forcing, the evolution of the system would follow that of 2-D turbulence, cascading energy to the largest scales. However, the random forcing keeps on injecting energy at smaller scales, so the flow remains fully turbulent throughout the whole data assimilation experiment. The whole experiment lasted 600 time steps.

The vorticity field was observed every 50 time steps on every gridpoint, giving about 65,000 observations every time step. The observations were obtained from a truth run and independent random measurement noise with standard deviation 0.05 was added to each observation. This should be compared to the typical nondimensional vorticity values of about 1. We determined the decorrelation timescale τ of the system by averaging the correlation time series at several points in the field, and taking τ as the time scale at which the correlation was $1/e$. We found a decorrelation time scale of this system of about 26 time steps. Since we observe the system every 50 time steps this is an extremely hard nonlinear data assimilation problem.

Only 24 particles were used to track the posterior pdf. Because we observe the full state vector we chose K in the relaxation term as a scalar with maximum value $K = 0.1$. (In general, when only part of the state vector is observed, or when H is a nonlinear function, K will be a matrix.) The random forcing covariance $\hat{Q}$ was the same as in the original model Q. This value for K is equal to the standard deviation of the model errors, chosen such that the relaxation term will be of that order of magnitude. Furthermore, K was chosen to vary linearly from zero to its maximum value between observation times. This allows the ensemble to spread out due to the random forcing initially, and pulling harder and harder towards the new observation the closer the system comes to the new observation time. No tuning has been applied in this example of the new particle filter, the reasonable values used for the parameters in the scheme applied to the Lorenz 1993 and 1996 systems (see Van Leeuwen, 2010,2011) have been implemented directly. More detailed experiments in which the sensitivity of the results to these specific choices for K will be described in another paper in preparation.

5. Results

Here a few initial results using the new particle filter with equivalent weights are shown. Figure 1 shows the vorticity field at time 50, and figure 2 the mean of the particles at that time. The two field are almost identical to the eye, showing that the new method is able to track the truth in this highly nonlinear regime.

Figure 3 shows the vorticity field at time 600, and its particle filter counterpart is shown in figure 4. Again the close tracking is very encouraging.

Figure 5 shows the absolute value of the difference between the ensemble mean and the truth run at time 50. This can be compared to the standard deviation in the ensemble in figure 6. Figures 7 and 8 show the same, but now after 600 time steps. Although the spread around the truth is underestimated at several locations, it is over estimated elsewhere, and the averages over the fields are almost equal. Given the statistical nature of these estimates, this is satisfactory.

A check on the workings of the equivalent weights scheme is to visualise the weights before resampling. Figure 9 shows that the weights are distributed as they should: they display small variance around the equal weight value 1/20 for the 80% of the 24 particles. Note that the particles with zero weight had too small weight to be included in the equivalent weight scheme, and will be resampled from the rest. Because the weights vary so little the weights can be used back in time, generating a smoother solution for this high-dimensional problem with only 24 particles. The results presented here refer to the filter solution only.

One of the questions one could ask is if these results could have been obtained with one of the standard scheme's used in meteorology or oceanography, like 4DVar or variants of the EnKF. When concentrating on the mean this might be so, but clearly the structure of the full pdf cannot be reconstructed with these methods. An example is depicted in figure 10, which shows the posterior pdf for the vorticity value at a certain point after 600 time steps. The non-Gaussian structure, hinting at bimodality, cannot be captured by any of these traditional methods.

Variational methods like 4DVar typically provide no error estimate because that is too expensive for large-dimensional problems like encountered in e.g. numerical weather prediction. From a scientific point of view this is not satisfactory. Furthermore, in a situation like depicted in figure 10 the usefulness of just the modal value would be limited, and also an error estimate based on the

Hessian, so the local curvature, has limited significance. Finally, given that the observation times are about two decorrelation time scales apart, 4DVar might struggle to convergence, but that is not tested here.

Ensemble Kalman filter methods do assume Gaussian prior and posterior densities. It might be possible by tuning inflation factors and localisation functions to obtain a proper evolution of the ensemble mean, comparable with the true evolution of the system. The ensemble generated with an EnKF might show bimodal structures, but it is unclear if these are real since the update does assume Gaussian pdf's, leading to a posterior pdf that does have the correct posterior mean and covariance (under the Gaussian assumption). The actual positions of the ensemble members differ in the different variants of the ensemble Kalman filter and have no direct statistical meaning. Furthermore, it is unclear if the ensemble covariance would have any real scientific meaning for these highly non-Gaussian pdf's. The ensemble spread is perhaps more used for tuning the system to ensure the correct evolution of the mean via inflation and localisation, than representing actual covariances in these highly nonlinear systems.

It is stressed here that no tuning has been applied in this example of the new particle filter, the reasonable values used for the parameters in the scheme applied to the Lorenz 1993 and 1996 systems (see Van Leeuwen, 2010, 2011) have been implemented directly.

An important issue is the quality of the scheme to infer the full posterior pdf. We have seen that the mean is close to the truth, but that could be due to e.g. extreme relaxation, so that all particles are very close to the observations, and so close to the truth in this high dimensional system. To investigate the quality of the ensemble we calculated a rank histogram using the ensemble values at every 50th time step and at every 4th grid point in each row and column of the field, assuming they were close to independent. For each time instance we rank the value of the truth in each gridpoint in the ensemble values at that gridpoint. This is done through ranking the values for the ensemble members from low to high, and determining where the truth lies in this ranking. The rank histogram is constructed by adding a value of 1 to that bin in which the truth falls, e.g. bin 4 is increased by 1 if the truth ranks between ensemble member 3 and 4. This is repeated for each gridpoint as mentioned above, and for each time instance, generating one rank histogram. The result is depicted in figure 11.

The second way to generate a rank histogram is to rank the observations in the measured ensemble members perturbed by the normal measurement error. This is the method of choice when the truth is not available, as in any real situation. Figure 11 shows the ranking of the truth, but both methods give similar results.

In the ideal case any of the particles could act as the truth, resulting in a uniform histogram. A low bias of the particles would yield a histogram in which the truth is biased to the higher rankings, so a histogram with higher bins to the right, and vice versa. An under dispersive ensemble will give rise to a truth value that is either lower or higher than the typical ensemble, resulting in a U-shaped histogram. Finally, an over dispersive ensemble leads to a histogram with a hump in the middle. Although the present histogram in figure 11 is not uniform, it is close to it given the statistical noise. One could do a proper confidence interval test, but that is not attempted here. It is remarkable to see how flat the histogram is, realising the high dimension of the system, the long interval between observations, and the fact that we only use 24 particles. The hump in the middle of the histogram might indicate an over dispersive ensemble, but the peaks at the end of the interval tend to show the opposite, an under dispersive ensemble. Or perhaps we see an over dispersive ensemble with biases in both directions. This little discussion shows that weakness of the histogram, it can be nearly flat for several reasons, not all of them positive. But, not withstanding that, the results are encouraging.

The results so far are encouraging and a much more detailed analysis of the present results, looking e.g. more closely at the posterior pdf's, the sensitivity to the observation uncertainty, and the spatial and temporal frequency of the observations. This will be reported on in a future paper.

6. Conclusions and discussion

The effectiveness of a new particle filter that exploits the proposal density and allows small ensemble sizes has been demonstrated on the highly nonlinear 65,000 dimensional barotropic vorticity equation that simulates ocean eddy processes. It was shown using identical twin experiments that the ensemble mean closely follows the truth, and that the ensemble spread is a good measure of the difference between the two. The nonlinear character of the problem is highlighted by studying the posterior pdf's, which often tend to show bimodal behaviour. Finally, a rank histogram for the whole experiment was shown to be close to uniform, indicating that the statistics of the ensemble is sound.

The advantage of this method is the enormous freedom in the two steps that make up the new method. The first adds terms to the model equations that force the model towards the future observations. The simple additive terms allow easy implementation in any simulation code for atmosphere or ocean, or more generally any computer code that simulates a Markov process. But also more sophisticated proposals can be used, like methods that optimise paths on each particle, e.g. a weak-constraint 4DVar solution on each particle. Note that the 4DVar would be special in the sense that the initial condition of the 4DVar is fixed, the particle position at time zero, but a model error term has to be included. Furthermore, since a 4DVar is a deterministic solution a random perturbation has to be added to each time step after the full 4DVar solution has been obtained.

The second crucial step allows the weights to be almost equal. Without this step the particle filter would still be degenerate with a large number of independent observations in the present settings. Also here much freedom exists in how this term is implemented. We replaced the search for the intersection of a hyperplane and the pdf in the 65,000 dimensional space by a simple line search, but many other possibilities can be explored. There is an interesting connection with new developments in rare event simulation using Monte-Carlo methods. Also there good proposal densities are essential, and advances have been made that allow simulation with minimal Monte-Carlo statistical errors (see e.g. Vanden Eijnden and Weare, 2012). These links will be pursued in future work.

One of the main questions is why this particle filter works in this high-dimensional system with only 24 particles. The reason is not entirely clear yet, but is most likely related to the following. First, one has to realise that there is no inherent problem related to the size of the space spanned by a small number of the particles with a high number of independent observations. (This would be the case for an ensemble Kalman filter.) The clearest examples are variational methods like 4DVar that are able to absorb all observations in a single model run. In the present implementation of the particle filter we do not run a complete 4DVar on each particle but a very crude approximation to that through the relaxation term. (One could run a 4DVar on each particle, as mentioned above, which is what the implicit particle filter of Chorin and Tu (2009) does, but that would be much more expensive, although probably better.) This, however is not enough to avoid filter divergence of the particle filter, i.e. the fact that the likelihood and proposal weights vary too much, with one particle getting a weight close to one, and the others all weights close to zero, when the number of independent observations is large. For that a scheme like the equivalent-weights step is needed to allow for the majority of the particles to have very similar weights, thus avoiding degeneracy. The actual dimension of the manifold on which the dynamics happens will be (much) smaller than the 65,000. The barotropic vorticity dynamics exhibits spatial and temporal coherency in which the smaller-scale motions tend to be slaved to the larger scales. (However, it should be realised that the small-scale random forcing does destroy this coupling to some extent.) It should be realised that of interest is the dimension of the dynamics given the observations, which will be different from that of the dynamical manifold of a free run. Exploring this fact is a very exciting research direction in which the data assimilation community and the dynamical systems community will have to work closely together. It will be clear, however that the dimension of this manifold will be much higher than 24.

Some variant of the EnKF could be used as proposal density for the particle filter, allowing e.g. for localisation, which is not straightforward in particle filtering (see e.g. Van Leeuwen, 2009, Papadakis et al, 2010, Prakash et al, 2011). This might allow us to ignore the relaxation scheme at each time step. The localised EnKF scheme could then be followed by the equivalent weights scheme. This is one direction of further research.

One of the main advantages of this particle filter scheme is that no reference is made to the covariance of the model state. It is well known that 4DVar stands or falls with the quality of the covariance of the initial state, the so-called B matrix. An enormous research effort has been spent, and is still spent on improving this B matrix. Also Ensemble Kalman Filters rely on the accuracy of the ensemble covariance matrix. This is why so much effort has gone into, and is still going into better inflation and localisation schemes. All these issues play no role in particle filtering.

It is well realised in the geoscienes community that errors in the model equations have to be included in the data-assimilation schemes. However, a proper statistical description of these errors is hard to come by. Even if it is assumed that the errors are Gaussian distributed, the mean, related to a model bias, and its covariance need to be specified, which is not easy. But that doesn't mean we should not go forward, especially when we realise that this will be the proper way to model improvement. As soon as an estimate of the statistical properties of the model errors is obtained, implementation in ensemble data assimilation methods like EnKf and particle filters is relatively easy because random realisations for these error estimates can be added directly to each ensemble member (and similar for multiplicative errors). Much more research is needed to come up with efficient implementations in variational methods. So, particle filters like the one explored here force us to consider where we are weakest: the errors in the model equations, and these particle filters are not distracted by problems in covariance structures in the model states themselves.

Although the results presented here might be promising, much more research is needed before questions on suitability for e.g. numerical weather forecasting can be answered. For example, we observed the full state vector at observation times, which is never the case for any real application from the geosciences. We are working hard on partially observed systems now. On the other hand, we observed the system at twice the decorrelation time scale, which makes the problem extremely hard since the information from previous observations is lost to a very large extent, showing the robustness of the method.

Another critical issue is to what extent the method pulls the system out of quasi balanced states. In numerical weather prediction tremendous progress was made when models were forced to stay close to balanced states, greatly suppressing artificial gravity waves that ruined the forecasts. It should be realised that as soon as we accept a statistical description of model errors model balances will be perturbed. So the question is if and how the proposal density will perturb the model balances more than just the random forcing. By keeping the stochastic part of the proposal density of similar magnitude as the original transition density that part should not add extra perturbations. The deterministic relaxation term can grow quite large, but if that becomes problematic we can restrict its size to some maximum value without problem. Another option is to project the relaxation terms to some sort of slow manifold, as is done in high-resolution numerical weather prediction ensemble Kalman filter applications. However, when the dynamics is strongly nonlinear it is unclear what the actual balances are. Finally, the essential equivalent weights step can be large too. Also here we could limit the size of the deterministic move, but this might destroy the possibility for majority of the weights to be equivalent. Also, projection on a slow manifold might help here too, with the same caveat as above. More research into these aspects are needed, and will no doubt be problem dependent.

Acknowledgements

We thank the National Centre for Earth Observation (NCEO) and the National Environmental Research Council (NERC) for support via grants. Also two anonymous reviewers are thanked for critical comments that improved the paper substantially.

REFERENCES

Bennett, A., 1992, Inverse Methods in Physical Oceanography, Cambridge University Press, Cambridge.

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

Chorin, A.J., and X. Tu, 2009, Implicit sampling for particle filters, PNAS, 106, 17249-17254.

Courtier, P., 1997, Dual formulation of four-dimensional variational assimilation, Quarterly Journal of the Royal Meteorological Society, 123(B), 2449-2461.

Doucet, A., De Freitas,N., and Gordon, 2001, N., Sequential Monte-Carlo methods in practice, Springer, Berlin.

Evensen, G., 1994, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, Journal of Geophysical Research, 99, 10143-10162.

Gordon, N.J., D.J. Salmond, and A.F.M. Smith., 1993, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEEE Proceedings-F, 140, 107-113.

Jazwinki, A.H., 1970, Stochastic processes and filtering theory, Academic Press, 376p.

Papadakis, N., E.Memin, A. Cuzol, and N.Gengembre, 2010, Data assimilation with the weighted ensemble Kalman filter, Tellus, 62, 673-697.

Prakask J., S.C.Patwardhan, and S.L.Shah, 2011, On the choice of importance distributions for unconstrained and constrained state estimation using particle filters, J. Process Control, 21, 3-16.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008, Obstacles to high-dimensional particle filtering, Monthly Weather Review, 136, 4629-4640.

Talagrand, O., and P. Courtier, 1987, Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, Quarterly Journal of the Royal Meteorological Society, 113, 1311-1328.

Vanden Eijnden E. and J. Weare, 2012, Rare event simulation with vanishing error for small noise diffusions, Communications Pure Applied Math., in press.

Van Leeuwen, P.J., 2009, Particle Filtering in Geosciences, Monthly Weather Review, 137, 4089-4114.

Van Leeuwen, P.J., 2010, Nonlinear Data Assimilation in geosciences: an extremely efficient particle filter, Quarterly Journal of the Royal Meteorological Society., 136, 1991-1996.

Van Leeuwen, P.J. 2011, Efficient non-linear Data Assimilation in Geophysical Fluid Dynamics, Computers and Fluids, doi:10,1016/j.compfluid.2010.11.011.

List of Figures

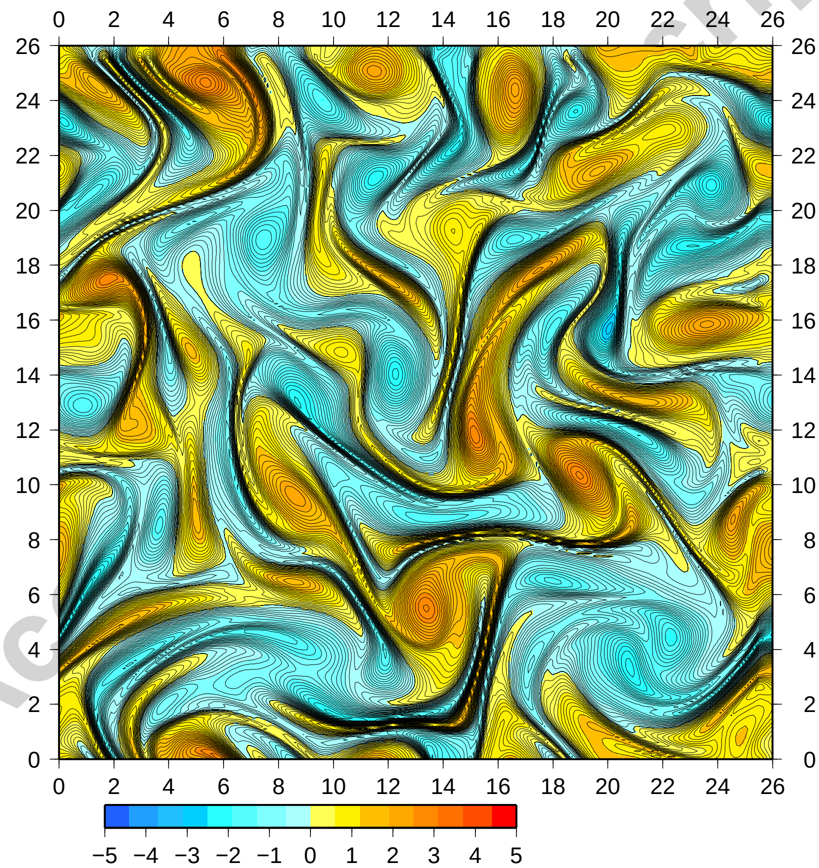


Figure 1: Snap shot of the vorticity field of the truth at time 50. Note the highly chaotic state of the field.

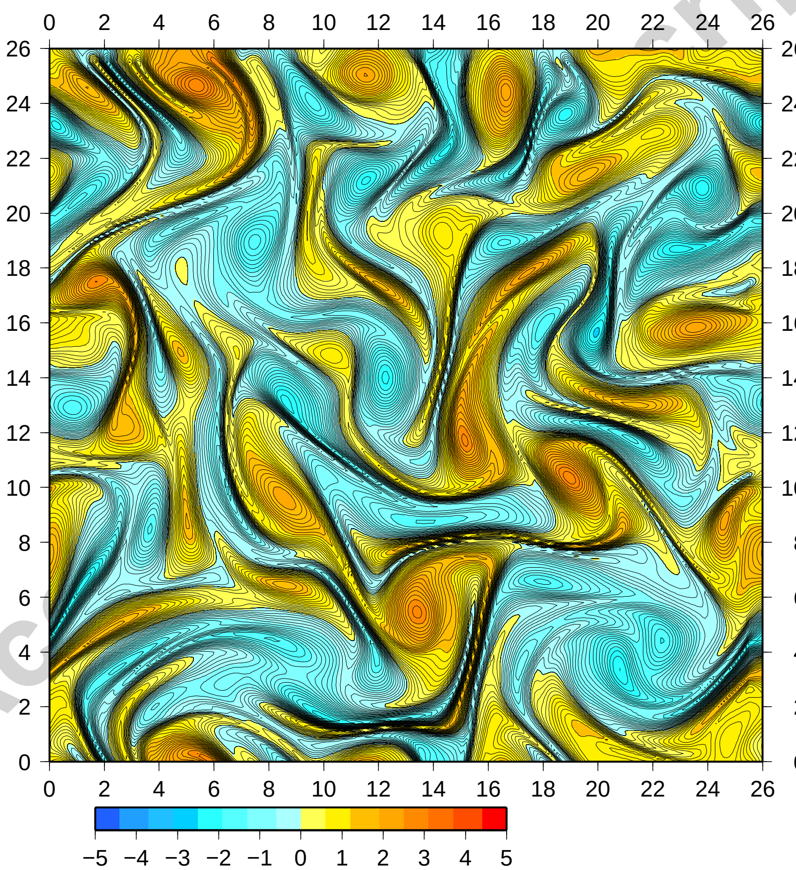


Figure 2: Snap shot of the vorticity field of the mean of the particle filter mean at time 50. Compare with figure 1 and note the close to perfect tracking.

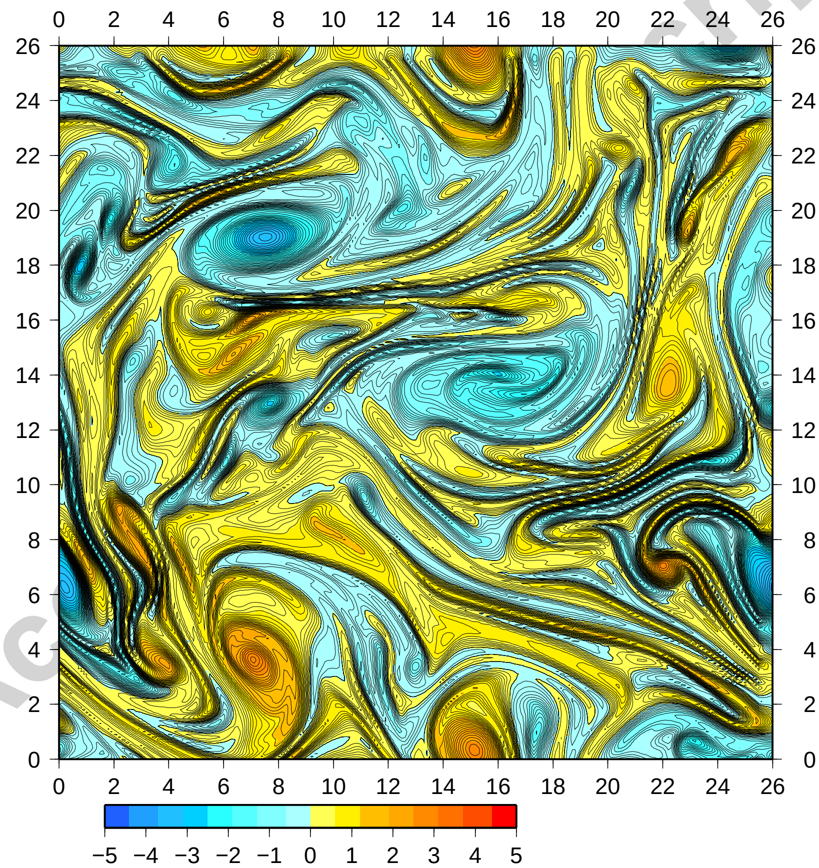


Figure 3: Snap shot of the vorticity field of the truth at time 600. Note again the highly chaotic state of the field.

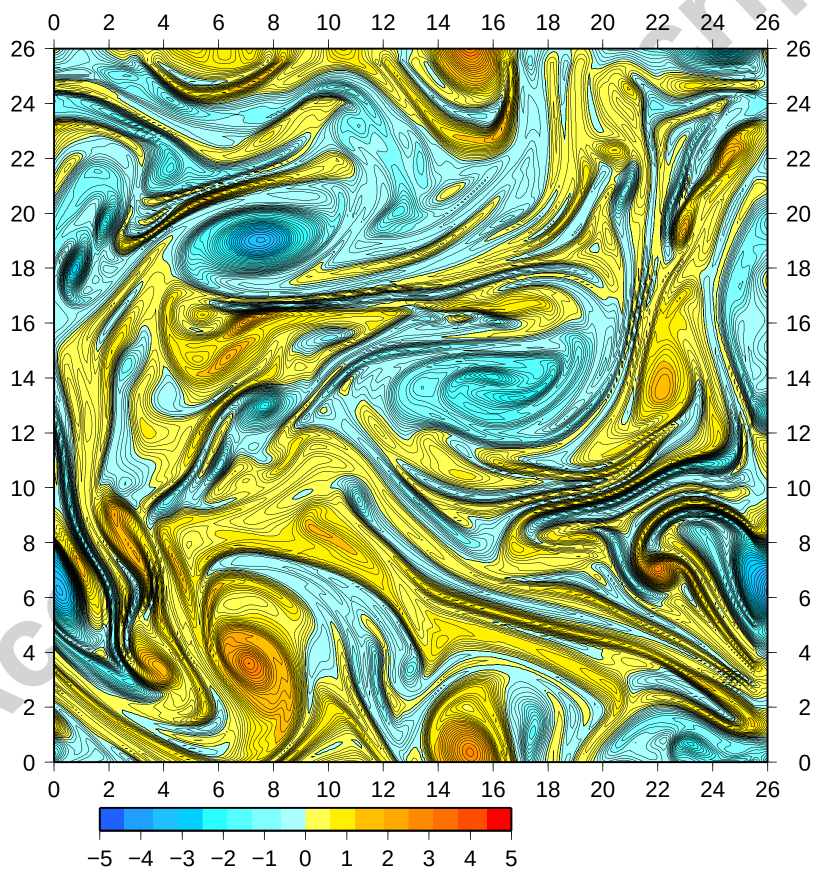


Figure 4: Snap shot of the vorticity field of the mean of the particle filter mean at time 600. Compare with figure 3 and note the close to perfect tracking.

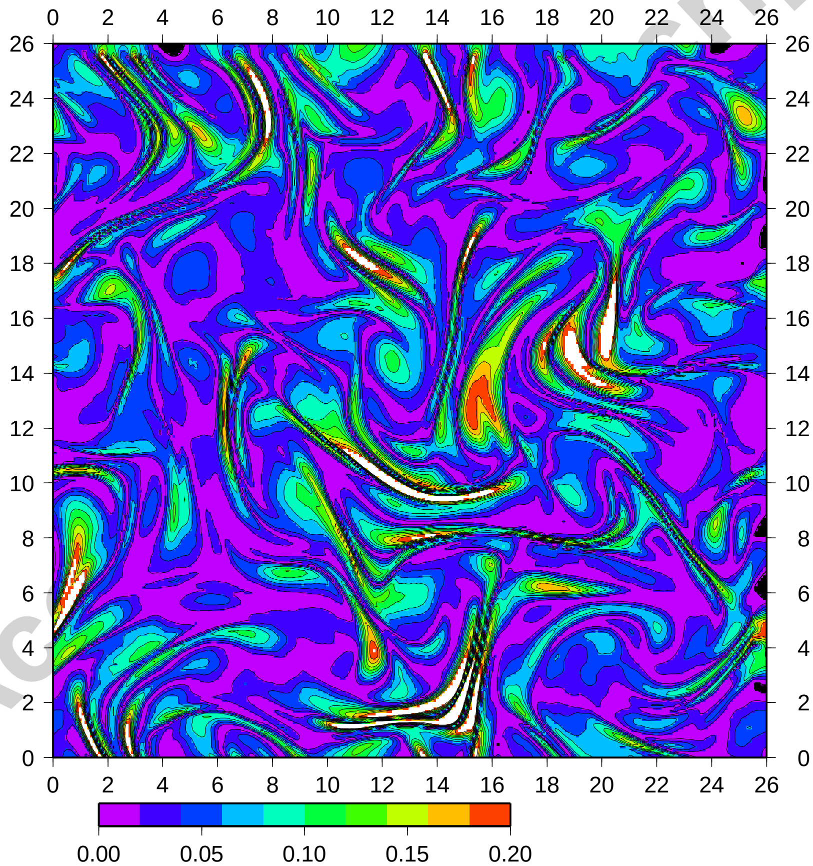


Figure 5: Snap shot of the absolute value of the mean-truth misfit at time 50. Note the highly irregularity of the field, reflecting the statistical nature of the estimate.

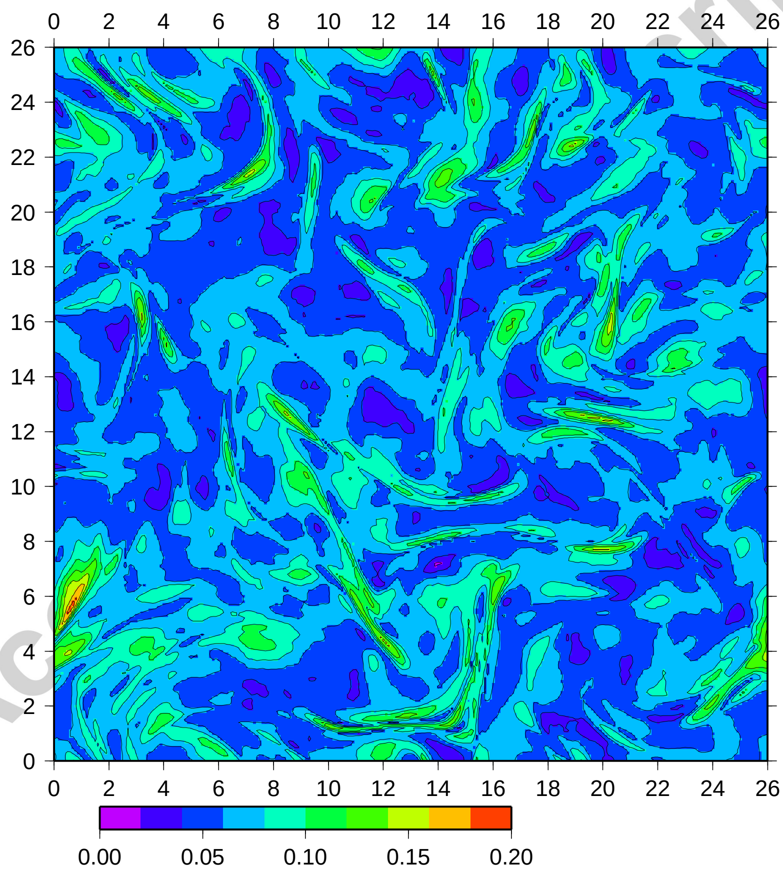


Figure 6: Snap shot of the absolute value of the standard deviation in the ensemble at time 50 for comparison with figure 5. The ensemble underestimates the spread at several locations, but averaged over the field it is slightly higher, 0.074 versus 0.056.

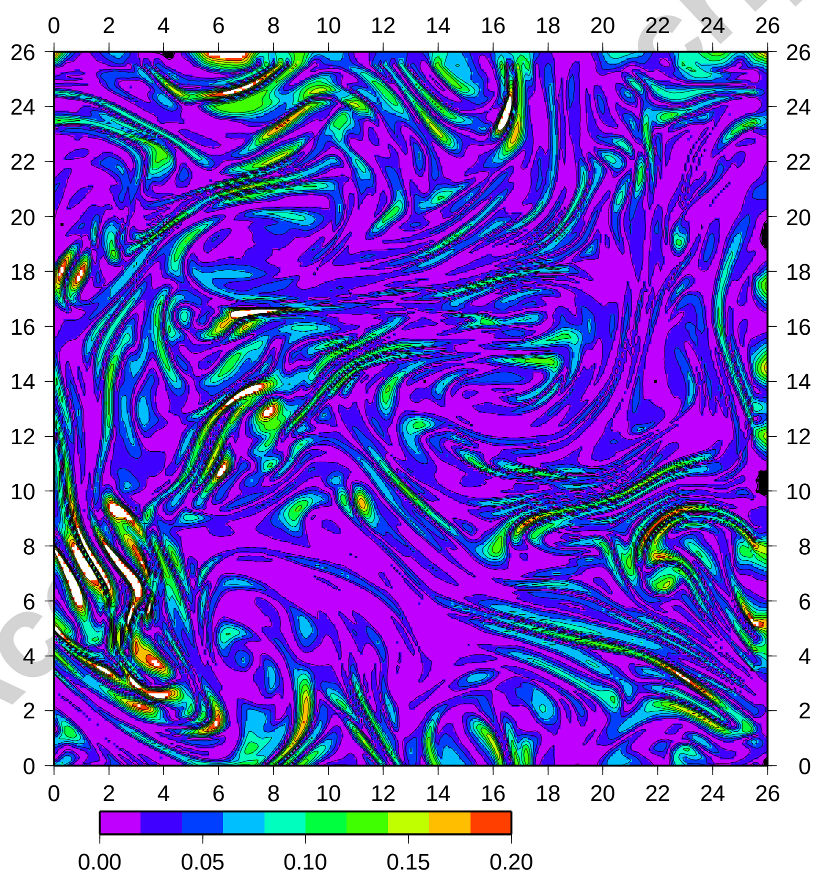


Figure 7: Snap shot of the absolute value of the mean-truth misfit at time 600.

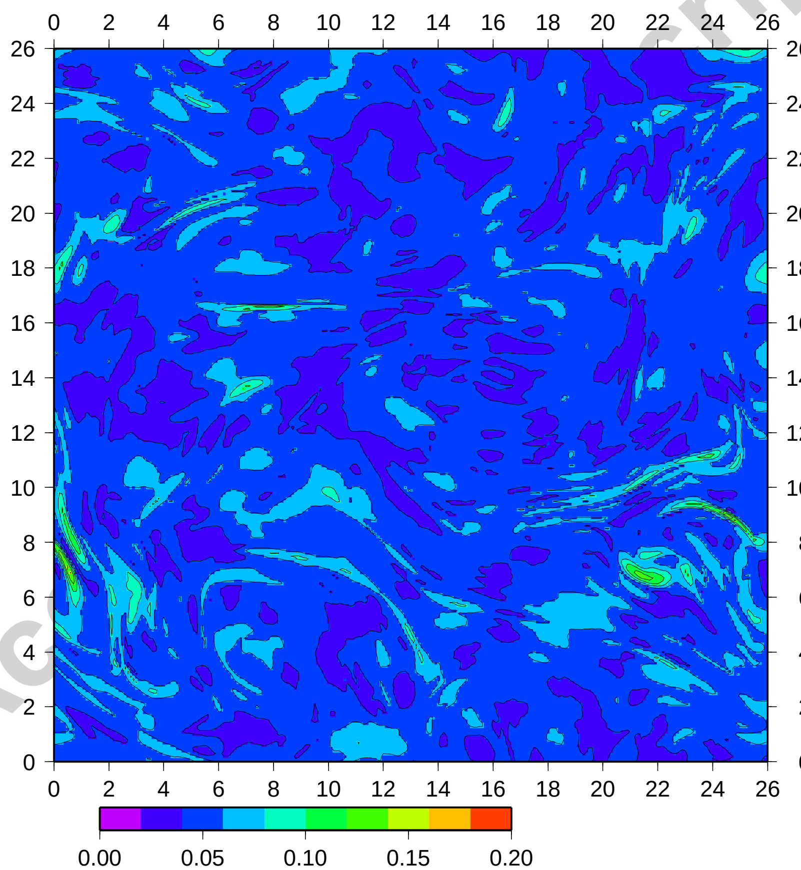


Figure 8: Snap shot of the absolute value of the standard deviation in the ensemble at time 600 for comparison with figure 7

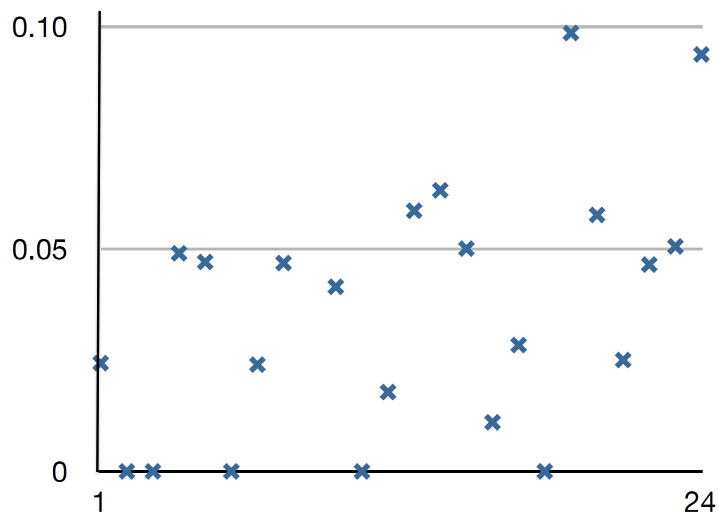


Figure 9: Weights distribution of the particles before resampling. All weights cluster around 0.05, which is close to 1/24 for uniform weights (using 24 particles). The 5 particles with weights zero will be resampled. Note that the other particles form the smoother estimate.

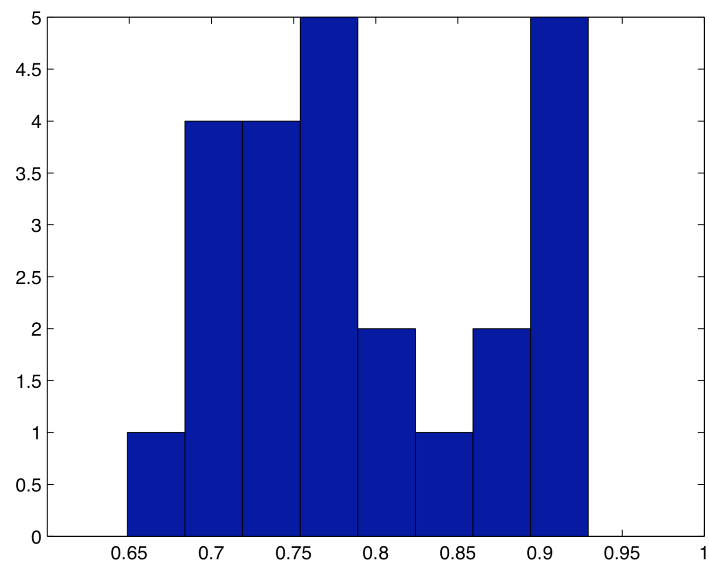


Figure 10: Estimate of the posterior pdf of the vorticity value at point (200,200) after 600 time steps. The bimodal structure shows that variational methods that look for the mode of this pdf have little meaning, and also methods based on the EnKF will not be able to represent this structure accurately.

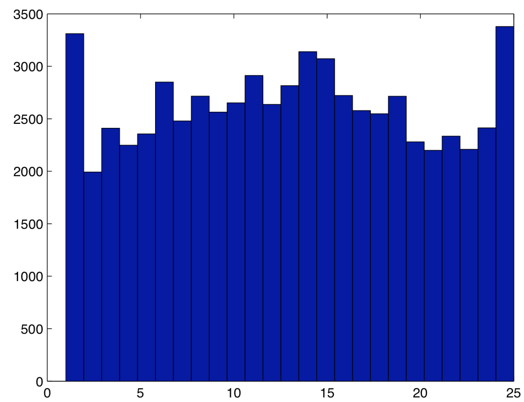


Figure 11: Rank histogram of how the truth ranks in the ensemble. The nearly uniform distribution shows that each particle could act as the truth, suggesting the ensemble is of good quality. (But see discussion in text.)

University of Reading
Department of Meteorology
University of Reading
Earley Gate, PO Box 243
Reading RG6 6BB

phone +44 (0)118 378 5578
fax +44 (0)118 378 8905
email p.j.vanleeuwen@reading.ac.uk

5 January 2012

Highlights:

1. First application of fully nonlinear data assimilation method to complex geophysical fluid dynamical system.
2. First application of full particle filter to 65,000 dimensional system
3. Detailed discussion of advantages and disadvantages of the new method