EFFICIENT NONLINEAR DATA ASSIMILATION FOR OCEANIC MODELS OF INTERMEDIATE COMPLEXITY

Peter Jan van Leeuwen

p.j.vanleeuwen@reading.ac.uk
Data Assimilation Research Centre
Department of Meteorology, University of Reading
PO Box 243, Reading, RG6 6BB, United Kingdom

ABSTRACT

A fully nonlinear particle filter is used on a simplified ocean model, consisting of the barotropic vorticity equation. While common knowledge is that particle filters are inefficient and need large numbers of model runs to avoid degeneracy, the newly developed particle filters need only of the order of 10-100 particles on large scale problems. Also, we show that the scaling is perfect in that increasing the dimension of the system does not need more particles. This opens the possibility for fully nonlinear filtering/smoothing in very high dimensional state spaces, e.g. for numerical weather forecasting.

Index Terms— Data Assimilation; Particle filtering; high dimensional; Bayes theorem; nonlinear filtering

1. INTRODUCTION

Numerical model for simulation and prediction of atmospheric and oceanic flows are becoming ever more complex. While relatively simple linear balances dominate the flows at large scales, with increasing resolution 2 and 3 D turbulence have to be resolved, leading to highly nonlinear flow structures. Another related area were nonlinearity is crucial is in climate models were many physical, chemical and biological systems are coupled leading to extremely complex behaviour. To the extend 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, which can be described in probabilistic terms. If the system is Markov, our present knowledge of the system in the form of a probability density function evolves according to the Kolmogorov or Fokker-Plank equation.

When observations of the system are present, 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 updates 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. Also parameter estimation falls in this framework: the prior pdf of the parameters is updated through the likelihood to the posterior pdf of the parameters. When the system at hand and the relation between observations and state vector are close to linear 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 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. 3DVar, 4DVar [8], representer method[1], PSAS [3], depending on details of the solution method. The Ensemble Kalman filter [4],[2] 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 filter the mean is equal to the mode. This has led to confusion in at least the geophysical, but also the so-called inverse-problem, community in what one is actually trying to solve, and and in some cases hampered progress to more nonlinear problems.

In this paper we propose solutions to highly nonlinear high-dimensional data-assimilation problems. Our stating point is the particle filter [6], 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

This research was sponsored by the National Environmental Research Council via the National Centre for Earth Observation, and projects NE/H008853/1, and NE/I005196/1.

distance norm is determined by the value of the pdf 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.

It is well known that in systems with moderate dimensions, say 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 [6] to more complicated or approximating solutions (see e.g.[4], and [9] for a review of applications in the geosciences). None of the proposed method is applicable to systems with larger than say 100, without having to need millions of particles, so millions of model integrations. As mentioned, our final goal is 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. The method is introduced in [10]. In this paper, the method is outlined and its performance on systems with up to 70,000 dimensions is demonstrated.

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 the particles. Using this representation of the prior pdf of the model in Bayes theorem one finds:

$$p(\psi|d) = \sum_{i=1}^{N} w_i \delta(\psi - \psi_i) \tag{1}$$

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

$$w_i = \frac{p(d|\psi_i)}{\sum_{j=1}^{N} p(d|\psi_j)} \tag{2}$$

The density $p(d|\psi_i)$ is the likelihood, i.e. the probability density of the observations given the model state ψ_i, which is given in the data-assimilation problem, and often taken as a Gaussian:

$$p(d|\psi_i) = A \exp\left[-\frac{(d - H(\psi_i))^2}{2\sigma^2}\right] \tag{3}$$

in which $H(\psi_i)$ is the measurement operator, which projects the model state on the observation d, and σ is the standard deviation of the observation error. When more observations are available, which might have correlated errors, the above is replaced by the joint pdf of all these measurements.

Unfortunately, weights vary wildly even when resampling is applied, and again 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 [7]. Several methods have been proposed to solve this problem (see review for the geosciences by [9], but none of these is directly applicable to large-dimensional geophysical problems.

3. THE NEW METHOD

The new method that will be explored in this research proposal consists of two new ingredients. The first new 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, but has not been explored in the geosciences. Assume the model equation to be written as

$$\psi^n = f(\psi^{n-1}) + \beta^n \tag{4}$$

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:

$$\psi^n = f(\psi^{n-1}) + \hat{\beta}^n + K(d^{n+m} - H(\psi^{n-1})) \tag{5}$$

in which $\hat{\beta}$ is random forcing which might have different characteristics from the original random forcing, and d^{n+m} denotes future observations at time $n + m$. The main difference with the original model equation is the 'nudging' term that tends to pull the particle to the observations. 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. The weights are modified as (see e.g. [4],[9]):

$$w_i \propto p(d^n|\psi_i^n) \frac{p(\psi_i^n|\psi_i^{n-1})}{q(\psi_i^n|\psi_i^{n-1} d^{n+m})} \tag{6}$$

The extra factor in the numerator is the probability that the actual state of particle i moves from ψ_i^{n-1} to state ψ_i^n, which can be calculated from the pdf of the random forcing β^n. The extra factor in the denominator follows from the pdf of random forcing $\hat{\beta}$, with mean $K(d^{n+m} - H(\psi^{n-1}))$.

But there is more. Making sure that all particles end up relatively close to the observations still does not mean that the weights will not vary wildly in large-dimensional systems. The second new ingredient is that we ensure that all posterior weights are almost equal. This consists of two stages: first perform a deterministic time step with each particle that ensures that most of the particles have equal weight, and then add a very small random step to ensure that Bayes theorem is satisfied. There are infinitely many ways to do this.

A simple choice for the first stage is enforcing

$$\psi_i^{n+m} = f(\psi_i^{n+m-1}) + \alpha_i K(d^{n+m} - H(f(\psi_i^{n+m-1}))) \quad (7)$$

in which $K = QH^T(HQH^T + R)^{-1}$, 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 to be determined such that the weights are equal. Exploiting the explicit expressions for the weights we obtain for each α_i, see [10]:

$$\alpha = 1 - \sqrt{1 - b_i/a_i} \quad (8)$$

in which $a_i = 0.5x_i^T R^{-1} HKx$ and $b_i = 0.5x_i^T R^{-1} x_i - C - \log w_i^{rest}$. Here $x = d^{n+m} - H(f(\psi_i^{n+m-1}))$, 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.

Of course, this last step towards the observations cannot be fully deterministic, as can be seen from Eq. (6). 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, because for a deterministic move the transition density is a delta function. The proposal transition density could be chosen a Gaussian, but since the weights have q in the denominator a draw from the tail of a Gaussian would lead to a very high weight for a particle that is perturbed by a relatively large amount. To avoid this q is chosen in the last step before the observations as a mixture density

$$q(\psi^n|\psi') = (1-\alpha)U(-a, a) + \alpha N(0, a^2) \quad (9)$$

in which ψ' the particle before the last random step. By choosing α small the change of having to choose from $N(0, a^2)$ can be made as small as desired. For instance, it can be made dependent on the number of particles N.

4. RESULTS

Here a few results using the new particle filter with almost equal weights are shown. Figure 1 shows the application of the method to the highly chaotic barotropic vorticity equation, governed by:

$$\begin{aligned} \frac{\partial q}{\partial t} - \frac{\partial \psi}{\partial y}\frac{\partial q}{\partial x} + \frac{\partial \psi}{\partial x}\frac{\partial q}{\partial y} &= \beta \\ q = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} & \end{aligned} \quad (10)$$

in which q is the vorticity field, ψ is the streamfunction, and β is a random noise term representing errors in the model equations. It was chosen from a multivariate Gaussian with mean zero, variance 0.01, and decorrelation lengthscale 4 gridpoints. The equations are implemented on a 256 X 256 grid, using a semi-Lagrangian scheme with time step $\Delta t = 0.08$, grid spacing $\Delta x = \Delta y = 1/256$, leading to a state dimension of close to 70,000. The vorticity field was observed every 25 time steps on every 4th gridpoint, giving about 4,000 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.

Only 24(!) particles were used to track the posterior pdf. In the application of the new particle filter we chose $K = 0.1$ in the nudging term (except for the last time step before the new observations, where the 'almost equal weight' scheme was used, as explained above), multiplied by a linear function that is zero half way the two updates and growing to one at the new observation time. The random forcing was the same as in the original model. This allows the ensemble to spread out due to the random forcing, and pulling harder and harder towards the new observation the closer to the new update time.

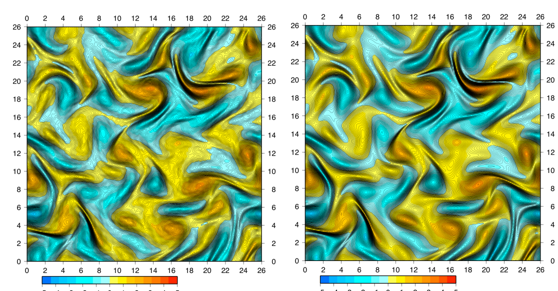

Fig. 1. Snap shot of the vorticity field of the truth (right) and the particle filter mean (left) at time 25. Note the highly chaotic state of the fields, and the close to perfect tracking.

Figure 1 shows the difference between the mean and the truth after 25 time steps, and figure 2 the ensemble standard deviation compared to the absolute value of the mean-truth misfit. Clearly, the truth is well represented by the mean of the ensemble. Figure 2 shows that although the spread around the truth is underestimated at several locations, it is over estimated elsewhere,

Finally, figure 3 shows that the weights are distributed as they should: they display small variance around the equal weight value 1/24 for the 24 particles. Note that the parti-

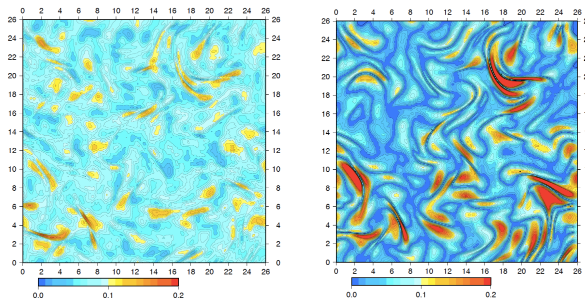

Fig. 2. Snap shot of the absolute value of the mean-truth misfit and the standard deviation in the ensemble. The ensemble underestimates the spread at several locations, but averaged over the field it is slightly higher, 0.074 versus 0.056.

cles with zero weight had too small weight to be included in the almost equal 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.

5. CONCLUSIONS AND DISCUSSION

A new particle filter has been introduced that exploits the proposal density and allows small ensemble sizes on very large dimensional problems. It was demonstrated here on the highly nonlinear 70,000 dimensional barotropic vorticity equation that simulates ocean eddy processes.

The big 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 of ocean, or more general any computer code that simulates a Markov process. But also more sophisticated proposals can be used, like e.g. a weak-constraint 4DVar solution on each particle, or an Ensemble Kalman filter. 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 a large freedom exists in how this term is implemented. We replaced the search for the intersection of a hyperplane and the pdf in the 70,000 dimensional space by a simple line search, but many other possibilities can be explored.

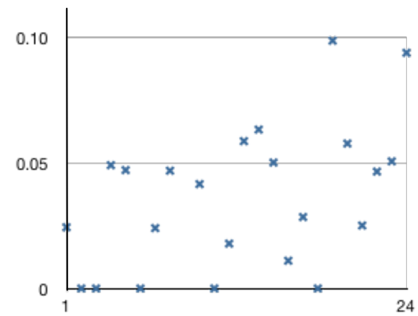

Fig. 3. Weights distribution of the particles before resampling. All weights cluster around 0.05, which is close to 1/24 for uniform weights (using 24 particles). The 5 particles with weights zero will be resampled. Note that the other particles form the smoother estimate.

6. REFERENCES

[1] A. Bennett, A. *Inverse Methods in Physical Oceanography*, (Cambridge University Press, Cambridge, 1992).

[2] G. Burgers, Van Leeuwen, P.J., and Evensen, G. ”Analysis Scheme in the Ensemble Kalman Filter”, *Monthly Weather Rev.*, 126, 1719-1724, 1998.

[3] P. Courtier, ” Dual formulation of four-dimensional variational assimilation”, *Q.J.Royal Met. Soc.*, 123(B), 2449-2461, 1997.

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

[5] G. Evensen, ”Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics”, *J. Geophys. Res.*, 99, 10143-10162, 1994.

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

[7] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson ”Obstacles to high-dimensional particle filtering”, *Mon. Wea. Rev.*, 136, 4629-4640, 2008.

[8] O. Talagrand and P. Courtier, ”Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory” *Quart. J. Roy. Meteor. Soc.*, 113, 1311-1328, 1987.

[9] P.J. Van Leeuwen, ”Particle Filtering in Geosciences”, *Monthly Weather Rev.*, 137, 4089-4114, 2009.

[10] P.J. Van Leeuwen, ”Nonlinear Data Assimilation in geosciences: an extremely efficient particle filter”, *Quart. J. Roy. Meteor. Soc..*, 136, 1991-1996, 2010.