

Ensemble Kalman filters, Sequential Importance Resampling and beyond

Peter Jan van Leeuwen

*Institute for Marine and Atmospheric research Utrecht (IMAU)
Utrecht University, P.O.Box 80005, 3508 TA Utrecht, The Netherlands
leeuwen@phys.uu.nl*

ABSTRACT

Data assimilation in high-resolution atmosphere or ocean models is complicated because of the nonlinearity of the problem. Several methods to solve the problem have been presented, all having their own advantages and disadvantages. In this paper so-called particle methods are discussed, with emphasis on Sequential Importance Resampling (SIR) and a new variant of that method. Reference is made to related methods, in particular to the Ensemble Kalman filter (EnKF). A detailed comparison between the EnKF and the SIR is made using the nonlinear KdV equation. It is shown that SIR produces good results in a highly nonlinear multi-layer quasi-geostrophic ocean model. Since the method needs at least 500 ensemble members or particles with the present-day observing system, new variants have to be studied to reduce that number in order to make the method feasible for real applications, like (seasonal) weather forecasting. In the new variant discussed here the number of members can be reduced by at least a factor 10 by guiding the ensemble to future observations. In this way the method starts to resemble a smoother. It is shown that the new method gives promising results, and the potentials and drawbacks of the new method are discussed.

1 Introduction

The highly nonlinear nature of atmospheric and oceanic flows together with the relative sparse observation network (especially in oceanography) make the data-assimilation problem nonlinear too. By inspecting a probability density function (pdf) of a multi-layer quasi-geostrophic model of the ocean we immediately see that the nonlinearity is substantial: when starting from a Gaussian the pdf becomes multimodal in a matter of days (see figure 1). The pdf was constructed by running an ensemble of models each with a slightly perturbed initial condition and adding model noise at each time step (see section 4).

The message from such a figure is that data-assimilation methods that contain linearizations of the true dynamics are expected to work not too well. Indeed, methods like the Extended kalman filter (EKF) will not work in problems like these, because they assume that the errors in the optimal estimate evolve with linear dynamics. It is expected that when the resolution of the numerical models increases this problem becomes even more stressing because the dynamics tend to become more nonlinear.

A possible way to circumvent this problem is by letting the errors evolve with the nonlinear model equations by performing an ensemble of model runs. This approach leads to the well-known Ensemble Kalman filter (EnKF), as presented by Evensen (1994, see also Burgers et al., 1998). A practical advantage of this method above the original EKF is the fact that only a limited number of ensemble members (100-500) is needed, compared to the equivalent of N ensemble runs (with linear dynamics) in the EKF, in which N is the size of the state space, which can easily be 1 million.

One problem remains, however, and that is the analysis step, when the model is confronted with observations. In all Kalman-filter-like methods it is assumed at analysis time that either the model is linear or that the pdf of

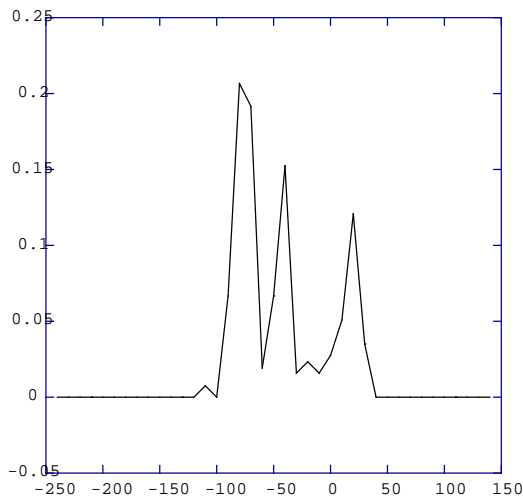


Figure 1: Example of a probability density function in a certain model point, obtained from a data-assimilation run with a multi-layer quasi-geostrophic model.

the model is Gaussian. For observations the assumption is that the pdf is Gaussian, and that the measurement operator (that produces the model equivalent of an observation) is linear. In methods like the EnKF, the Kalman-filter update equations are used even when the above mentioned assumptions are not fulfilled. What is done is that the variance of the ensemble is used in the kalman-filter equation. So, it is assumed that this equation does hold, but the rest of the method is 'as nonlinear as possible'. With the latter I mean that due to the ensemble nature of the method it is possible to use nonlinear measurement operators, and each ensemble member is updated individually. One can show that, under the assumptions of the original kalman-filter, the updated ensemble has the correct covariance. (Note that the covariance is never explicitly calculated in ensemble methods.) However, since these assumptions are relaxed to some extent it is in fact not entirely clear what is done statistically. For instance, the prior ensemble will in general have a third-order moment (skewness) that is unequal to zero. So, extra information on the true pdf is present, which is a good thing. However, the update is done independent of this skewness. So, the updated ensemble will have a skewness that is deformed by the update process, and it is unclear what its meaning is. It might be completely wrong, deviating the estimated pdf from the true one.

So-called adjoint methods take the nonlinearity of the dynamics into account directly. A serious problem is, however, that the method does not provide an error estimate in a highly nonlinear context. Note that the inverse of the Hessian, which is sometimes claimed to provide this error estimate, only does so in a linear context. Another problem is the assumption that the model dynamics and so-called physics (that is, the diabatic processes) are free of errors. Unfortunately, we are a long way from that assumption. Finally, the error-free model dynamics make the final estimate extremely dependent on initial conditions in the (nearly) chaotic system that the true atmosphere or ocean is. This latter problem can be avoided by avoiding long assimilation time intervals, as is done in incremental 4DVAR.

An advantage of the latter methods is the fact that the model evolution is always balanced. The Kalman-filter-like methods can produce unbalanced states at analysis times because each updated ensemble member is just a linear combination of the prior ensemble. For instance, negative concentrations or unbalanced dynamics can occur. This is due to the Gaussian assumption on the pdf's, while e.g. a concentration does not have a Gaussian pdf.

Another interesting fact of all methods presented above is that they have to perform matrix inversions. The size of the matrix depends on the number of observations, and can become rather large. So, it seems wise to

try to keep the number of observations low for optimal numerical performance, while, from a statistical point of view, we want to use as many observations as possible.

Is it possible to create a data-assimilation method that can handle non-Gaussian pdf's, nonlinear measurement functionals, provide error estimates and has no problems with millions of observations? To answer this question one has to return to the most general description of the data-assimilation problem. Bayes has shown that the probability density of a model ψ given a new set of observations d is given by:

$$f_m(\psi|d) = \frac{f_d(d|\psi)f_m(\psi)}{\int f_d(d|\psi)f_m(\psi)d\psi}. \tag{1}$$

In words, the pdf of the model given the observations is given by the product of the prior pdf of the model and that of the observations given the model. The denominator is a normalization constant. Note that, because we need the pdf of the observations given the model, we are not troubled with the true value of the observed variable, only with its measurement and the model estimate.

A close look at the equation shows that the posterior pdf can be obtained by just multiplying the densities of model and observations. So, data assimilation is that simple. In its purest form it has nothing to do with matrix inversions, it is not an inverse problem in that sense. At the same time we see how ensemble methods can be used. The ensemble members, or particles, are a certain representation of the prior model pdf, and the posterior pdf is represented by a reweighting of these ensemble members. This weighting is dependent on the value of the observations given an ensemble member. Sequential Importance sampling is just doing that: it creates an ensemble of models, runs that ensemble forward until observations become available (so far the method is identical to the other ensemble methods like the EnKF), weight each ensemble member with these observations, and continue the integration. This procedure is depicted in figure 2, and the actual calculation is presented in the next section.

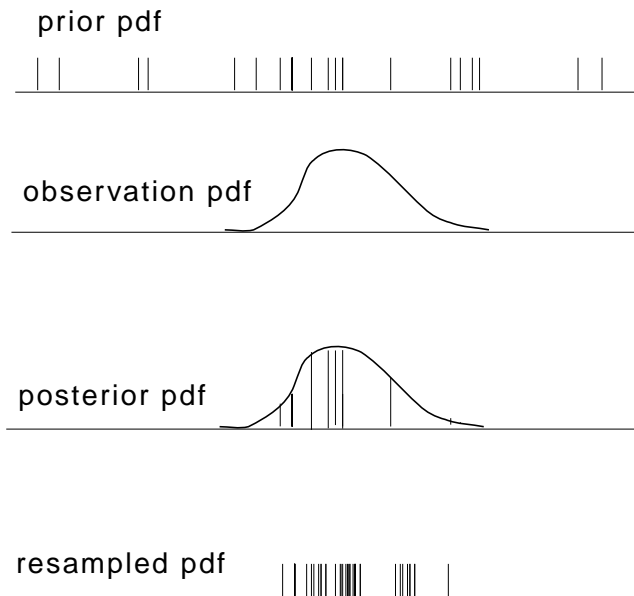


Figure 2: Sequential Importance Resampling: the prior pdf, as represented by an ensemble is multiplied with the observation pdf (not necessarily a Gaussian) to obtain the posterior pdf as represented by the ensemble. This posterior pdf is resampled to give each member equal weight again. The horizontal axis denoted one model variable.

A practical problem is evident in figure 2, namely that some ensemble members have nothing to do with the observations: they are just way off. These members get a very low weight compared to the others. The ratio

of these weights can easily be a million or more. Clearly, these members have no influence on the first few moments of the pdf, they contain no real information anymore. This has let people to so-called resampling (Gordon et al, 1993). In the original formulation one just draws the complete ensemble randomly from the weighted posterior ensemble. Clearly, posterior ensemble members with very low weight have a very low probability to be drawn, while members with large weights can be drawn more than once. This is nothing more than abandoning those members that contain no information, and stress those that have. In the present implementation members with large weights are copied, and the number of copies is related to their weight. This copying and abandoning is done such that the total number of members remains the same as before the update. This presents, in words, Sequential Importance Resampling; we elaborate on this in the next section.

In Van Leeuwen (2003) was shown that this method works in a highly nonlinear corner of the World Ocean, where the size of the state-space was about $2 \cdot 10^5$. Unfortunately, the size of the ensemble needed was rather high; good results were obtained with a size of 495. This is considered too large for operational purposes. The idea to reduce this number further arose from communications with Kevin Judd. He suggests to use a so-called shadowing trajectory (or rather pseudo orbit) to first detect where the observations are before an ensemble is created to follow that trajectory. He proposes to use 4DVAR to generate the shadowing trajectory, but the drawback is that that is rather expensive (of the order of 100 ensemble integrations), so that the gain is lost in generating this first trajectory. Approximate ways to construct the trajectory are under construction, as are extensions to pseudo-orbits for models including dynamical errors. However, instead of generating such a trajectory one could let the ensemble know before hand where the observations are. This results in a SIR algorithm in which the SIR is applied with the same set of observations before the actual observation time. This leads to the Guided Sequential Importance Resampling. It can be shown that this is a valid procedure as long as the weighting is done properly. In this paper it is shown that this method can reduce the number of ensemble members needed to about 30 for the problem under consideration.

In the present paper the methods are presented in the next section. In section 3 a comparison of the performance of the EnKF and the SIR is presented to highlight the differences in a strongly nonlinear environment (clearly not to judge on which method is best...). Section 4 discusses the SIR and especially the GSIR for a real-sized problem with real observations. The paper closes with a discussion of the relative merits.

2 Sequential Importance Resampling

In this section we discuss the algorithms for the SIR and the GSIR. What both methods have in common is the reweighting of ensemble members as soon as observations are available. Before the observations are taken into account, the pdf of the model is described by a swarm of particles or ensemble members. Each of them contains the same amount of information of this pdf, each has weight $1/N$ in which N the size of the ensemble. When observations are available these members are reweighted. Bayes tells us (see (1)) that this new weight should be:

$$w_i = \frac{f_d(d|\psi_i)}{\sum_{i=1}^N f_d(d|\psi_i)}. \quad (2)$$

in which f_d the pdf of the observations. When the observations are Gaussian distributed this leads to:

$$w_i = \frac{1}{A} \exp \left[-\frac{(d - H\psi_i)^2}{2\sigma^2} \right] \quad (3)$$

in which H is the measurement operator (that can be nonlinear), and in which the normalization A is of no importance because we know that the total weights of all members should equal 1. The next two sections describe the two SIR-filters in detail.

2.1 The SIR algorithm

The algorithm is explained rather easily. For details I refer to Van Leeuwen (2003). The procedure that I followed in this paper closely resembles importance resampling, as mentioned above. In the original SIR new members are drawn randomly from the weighted posterior ensemble. Here, instead of choosing randomly from the distribution determined by the weights, members with large weights are chosen directly from the distribution in the following way. First, this density is multiplied by the total amount of ensemble members. For each so-obtained weight that is larger than 1 the integer part of that weight determines the number of identical copies of the corresponding ensemble member. So, if the original weight of member i was $w_i = 0.115$, with an ensemble size of 100, the new weight $\tilde{w}_i = 100 * 0.115 = 11.5$. This results in 11 copies of the ensemble member, while $\tilde{w}_i = 0.5$ remains. This procedure is followed for all weights. Finally, all remaining parts \tilde{w}_i form a new density from which the rest of the ensemble is drawn, according to the rules of stochastic importance resampling described above. The reason for the deviation from the basic importance resampling is described in Van Leeuwen (2003) and is of little consequence for what follows.

2.2 The GSIR algorithm

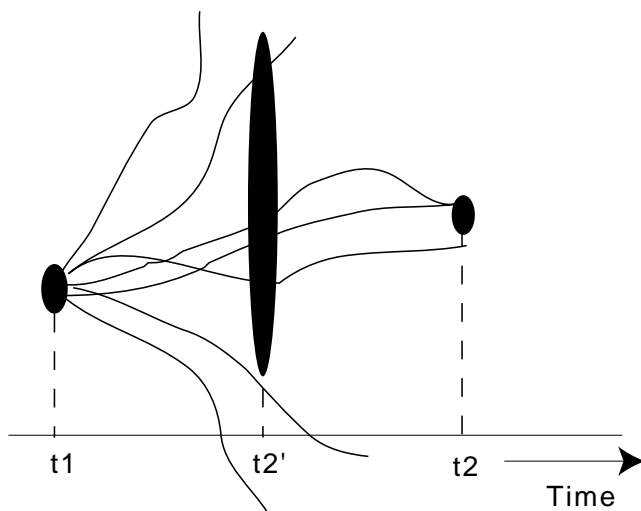


Figure 3: Guided Sequential Importance Resampling: a SIR is performed at $t2'$ before the actual measurement time $t2$ to guide the ensemble towards the observations at $t2$. The ellipses denote the observations with their standard deviation.

Now that the SIR algorithm is given, the GSIR is easily explained. Figure 3 shows the method. When the SIR is applied at the previous observation time, we wait until the next observations come in. Then we integrate the ensemble forward in time for a few time steps. We now assume that the new observations are performed at this time and perform a SIR-step. Obviously, we make an error here, because the observations are not to be used yet. However, we want to know already at this stage which members are going in the right direction, and which members are shooting off to remote areas in state space not supported by the observations at all. To avoid being too strict we increase the measurement error by a large factor, 100 say. The system is so nonlinear (or the model is so bad), that a number of ensemble members is unable to get a reasonable weight at this stage. By using the SIR we just abandon these already.

After performing the SIR at this intermediate time step before the actual observations, we continue integrating the ensemble, again a few time steps. Then the procedure described above is repeated, but the error in the

observations is increased less than in the previous step, let's say a factor 10. Again the SIR is performed at this intermediate step. This is repeated a few times if necessary.

Then we integrate the ensemble up to the true measurement time. Again we perform a SIR step, but because the ensemble members are already relatively close to the observations (guided to them as it were), the weights will not differ too great. This allows one to greatly reduce the number of members. In the application described in section 4 the reduction is more than a factor 30! This brings the number of members needed to about 30, which is sensible for real applications, given the ensemble forecasts and number of iterations performed in the 4-DVAR at this moment. It is even cheaper...

3 Nonlinear filtering: A simple example with the KdV equation

In this section the SIR method is applied to the Korteweg-deVries equation to study its behavior in nonlinear systems. For comparison the ensemble Kalman filter is applied to the same problem. In this way the difference between the conventional Kalman update and a variance-minimizing solution can be investigated. Details of this comparison are presented in Van Leeuwen (2003).

The Korteweg-deVries (KdV) equation describes the nonlinear evolution of a field u subject to advection and dispersion. It reads:

$$u_t + 6uu_x + u_{xxx} = 0 \tag{4}$$

Shape conserving solutions called solitons are allowed by a balance between the steepening of the wave form due to nonlinear advection and the dispersion from the third spatial derivative. We start by a field of the form:

$$u(x, 0) = \frac{0.5a}{(\cosh(\sqrt{a}(x - x_0)))^2} \tag{5}$$

in which x_0 is the position of the maximum of the wave form, and $0.5a$ its amplitude (see Fig. 4a). The KdV equation will move the wave form towards positive x values with a speed a , while conserving its shape. Important for the following is that the soliton is stable to small perturbations.

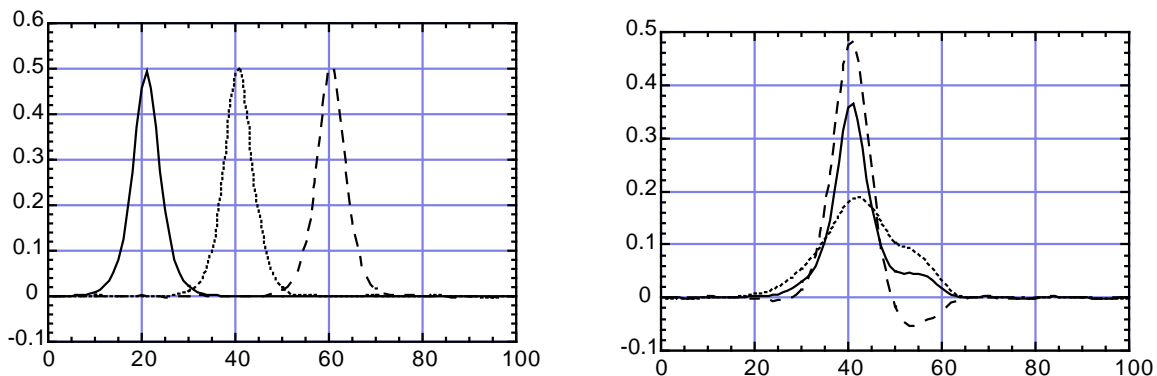


Figure 4: Left: True solution of the KdV equation at time 0, 10 and 20. Right: Prior (dotted), EnKF (dashed) and SIR (solid) ensemble mean solutions on $t=10$.

Several experiments have been performed with the new filter and the ensemble Kalman filter. We present one experiment in detail here that highlights the differences between the two methods. We first form a true solution by integrating this form with $a = 1$ over a domain of length 50, with periodic boundary conditions, $x_0 = 20$ and $\Delta x = 0.5$. The generation of the ensemble and the integration details are described in Van Leeuwen (2003).

This solutions is measured six times, on $t = 10$, at $x = 37$, $x = 40$, and at $x = 43$, and on $t = 20$, at $x = 57$, $x = 60$, and at $x = 63$. To these pseudo observations random Gaussian noise with zero mean and standard deviation 0.05 was added. Note that the observations are taken around the peak values of the wave form.

In figure 4b the mean of the ensemble before the update at $t = 10$ is given. The decrease of the amplitude can be attributed to the spread in amplitudes a , leading to an ensemble of soliton-like waves with different propagation speed. In the same figure the mean of the ensemble after analysis at $t = 10$ is given for the Ensemble Kalman filter and for the sequential importance resampling filter. The measurements are indicated by the crosses. The first thing that strikes the eye is that the EnKF solution comes much closer to the observations than the SIRF. The EnKF solution is too close to the observations because the variance in the prior ensemble is very large, so the EnKF assumes there is little information in the prior ensemble. In Fig. 5a the prior, posterior and observational density are given at the peak of the true soliton, on $t = 10$, at $x = 40$. The densities are created using the frequency interpretation on pre-specified intervals. Varying the intervals within reasonable bands showed that the features visible are robust. Clearly, the prior is non Gaussian because the solitons are always nonnegative. Several ensemble members moved too slow or too fast to have a significant value for the solution at $x = 40$. The SIRF gives the exact solution because the ensemble has converged.

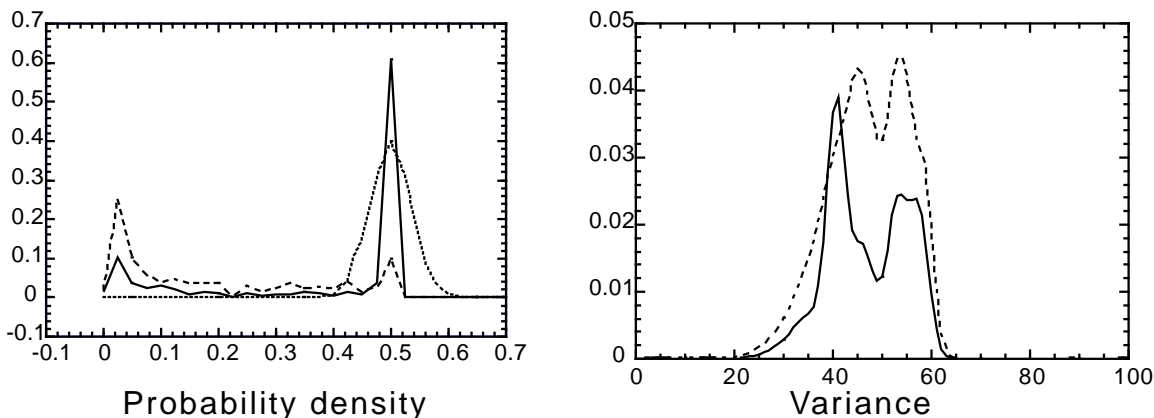


Figure 5: Left: Prior (dashed), posterior (solid) and observational (dotted) pdf on $t=10$ at $x=40$. Right: prior (dashed) and posterior (solid) variance on $t=10$.

Variance estimates for a truly variance minimizing solution like the SIR also show unfamiliar behavior. Figure 5b shows the prior and posterior variance estimates for the SIRF at $t = 10$. Interestingly, the posterior variance is higher than the prior variance at the measurement point $x = 40$. So, in a conventional way of thinking the uncertainty in the estimate at that point is increased due to the measurement. A more accurate inspection of the full prior and posterior densities shows that the uncertainty, defined here as the entropy (see Van Leeuwen, 2003), has decreased, but the second moment of the density does not show it. One can easily show that the entropy of a Gaussian distributed variable is proportional to its variance. For probability densities other than the Gaussian this is not necessarily true. For instance, for bimodal densities the variance loses its meaning. In that case the entropy is a more sensible estimate, as Shannon (1948) showed. In our case we find that the entropy of the prior density is $H = 1.1587$, and that of the posterior density is $H = 0.7286$. So, the entropy has decreased.

The analysis at $t = 20$ (not shown here) gives a SIR estimate very close to the observations. This has to do with the correct update of the ensemble at $t = 10$, leading to a relatively large part of the ensemble rather close to the truth. The EnKF analysis is behaving very wild. This is due to the fact that the relatively strong update at $t = 10$ generated negative values, and the KdV solution is very sensitive to such values, leading to large waves that travel in the wrong direction. This shows another feature of nonlinear filtering: truly nonlinear filters are

'property conserving', that is, they take into account the conservation laws that are in the governing equations because ensemble members are not mixed, they are merely weighted.

It should be mentioned that modifications to the standard EnKF can be used to overcome the problems presented here, but the point we want to make is that nonlinear filtering gives unexpected results sometimes, and needs no adaptations for specific problems.

4 The SIR and its Guided variant

In this section the SIR and the GSIR are applied to a large-scale problem to study its behavior in such a setting. We study the ocean area around South Africa. The Agulhas Current runs along the east coast of South Africa southward and retroflects just after leaving the continent at its most southern tip, back into the Indian Ocean. At the retroflection point large Agulhas rings are shed, moving into the South Atlantic. The area is an ideal test bed for data assimilation methods due to the highly nonlinear dynamics and the availability of high-quality satellite altimeter measurements of the height of the sea surface. Since this height is directly related to the pressure it is an important dynamical constraint on the flow.

The area is modeled by a 5-layer quasi-geostrophic ocean model with a horizontal resolution of 10 km, with 251×140 gridpoints. The first baroclinic Rossby deformation radius is about 35 km in this area. The layer depths are 300, 300, 400, 100 and 3000 m, respectively, with densities of 1026, 1026.9, 1027.25, 1027.65 and 1027.75 kg/m^3 . The time stepping was done with leap-frog, with an Euler step every 67th step to suppress the computational mode. A time step of 1 hour was chosen, close to the CFL-limit, for optimal accuracy. Small-scale noise was reduced by a Shapiro filter of order 8. Boundary conditions are such that features are leaving the domain with twice the speed of the fastest 5 wave modes. Only the inflow of the Agulhas Current at the eastern side of South Africa was prescribed. Because this inflow is super critical for all baroclinic (Kelvin) waves this last condition is well posed for those waves. Problems with barotropic waves did not arise. (Note that the values at the boundaries are part of the data assimilation problem.) The model is able to produce realistic ring-shedding events and general meso-scale motion in the area, as compared to satellite observations and in situ ship measurements.

4.1 Statistics

The initial streamfunction error (uncertainty) was taken space independent, with values of 4000, 3000, 2000, 1000 and $1000 \text{ m}^2/\text{s}$ for the layer models. Every day a random error of 0.05 times these values was added to describe the model error. The spatial correlation of the errors was Gaussian with a decorrelation length of twice the Rossby radius of deformation. The state space, consisting of the 5 streamfunction fields for the 5 layers, has a dimension of about $2 \cdot 10^5$. The ensemble size was 1024, where the SIR seems to have converged (see Van Leeuwen, 2003), and 32 for the GSIR.

4.2 Observations

The observations were satellite altimeter height data from the TOPEX/Poseidon and the ERS-2 satellites. These two satellites cover the model area with tracks that are about 150 (T/P) and about 70 (ERS) km apart. T/P has a repeat orbit of 10 days, ERS has a repeat orbit of 35 days. The along-track resolution is 7 km, of which we used every fifth point. The observations that are used in the data-assimilation experiment are collected over 1 day and the resulting batches are offered to the model. So, each batch has only a partial coverage of the

domain, a few tracks, that differs from day to day. The observational error was specified as $5000 \text{ m}^2/\text{s}$, which corresponds to about 4 cm in sea-surface height.

The shape of the probability density of the altimeter observations is a difficult matter. Due to the weighting procedure the SIRF is very sensitive to the tails of that density. In general, we know little of those tails. It has been suggested, however, that the tails of a Gaussian are too small: the square in the exponent cuts off large deviations from the observations very drastically. So, outliers, meaning bad measurements here, may have a tremendous effect on the behavior of the filter. The Lorentz density is used in this paper, but better alternatives can probably be found quite easily. Advantage of this density is that it has a shape very similar to a Gaussian near the peak (symmetric and quadratic), but it is much broader away from the peak. The observational error is taken equal to σ , half the full-width at half maximum, so equal to a Gaussian in this respect.

The SIR and GSIR filters were implemented on a Origin 3800 parallel computer using up to 256 processors, with MPI. The speedup was close to 100 %, while the f90 code is extremely simple.

4.3 Results

We concentrate here on the comparison between the SIR and the GSIR. The prior density shown in figure 1 is from the SIR experiment, showing that Kalman-filter-like methods might not do a good job.

Figure 6a shows the streamfunction evolution of the upper layer for days 2, 10 and 20, as determined from the SIR with 1024 members. Figure 6b shows the same for the GSIR with 32 members. Comparing these two figures shows that the two methods produce nearly identical results. The differences are within the ensemble standard deviations for the SIR and the GSIR (not shown). Also these second moments show a close resemblance. The conclusion from these results is that the GSIR works in this highly nonlinear environment with only 32 ensemble members.

5 Summary and discussion

A new data assimilation method, a variant of sequential importance resampling, has been presented. It is truly variance minimizing for nonlinear dynamics, unlike schemes based on the (ensemble) Kalman filter. It is argued that due to its very nature data assimilation is not an inverse problem, or a minimizing problem, although it is usually cast in that form. The method is based on importance resampling, a well-known method in control theory (see e.g. Doucet et al, 2001). The SIR filter has been shown to work well in large-scale problems in Van Leeuwen (2003), but needed at least 500 ensemble members. New in this paper is a new variant, the Guided SIR, in which the number of ensemble members can be decreased at least a factor 30 by using observations backward in time.

First the special effects in nonlinear filtering were highlighted in an example with the KvD equation, where the EnKF and the SIR were compared. We found that the Gaussian assumption made in the EnKF at analysis time can lead to incorrect updating. Furthermore, it is possible that the variance of the posterior density is larger than that of the prior density at some locations as the new filter showed. This is impossible when the Gaussian assumption is made. The entropy of the pdf, a measure of the uncertainty, did decrease however. Finally, it was shown that the new filter avoids unbalanced states.

So, to sum up, the SIR is truly variance minimizing, no matrix inversions are needed, the observations can be distributed non Gaussian and the measurement functionals can be nonlinear without any problem. Furthermore, it preserves prior model constraints like positive definiteness unlike Kalman filter-like methods that mix states at analysis time, and provides error estimates unlike 4DVAR-like methods. Finally, it is easy to implement and

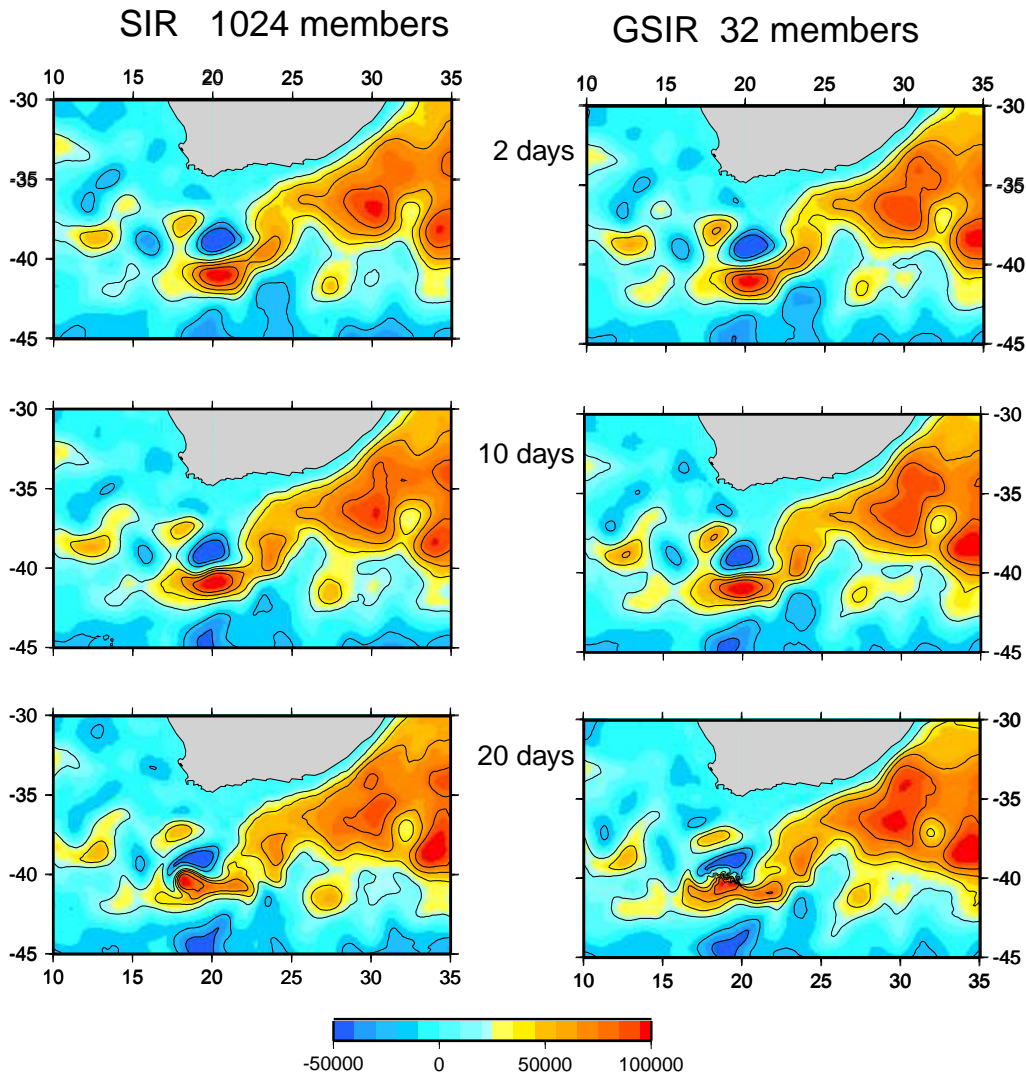


Figure 6: Upper-layer streamfunction for SIR with 1024 members (left) and GSIR with 32 members on day 2, 10 and 20.

parallel, by its very nature.

Secondly the GSIR was compared to the SIR with a large-scale problem. Because of the high dimension of the state space in real-size applications, care has to be taken when formulating the probability density of the observations. It was argued that because of outliers a Gaussian density is too narrow, i.e., its tails are too low. One measurement can degrade the solution if this is not taken into account. A Lorentz profile was used in this paper, leading to a more robust filter that seemed to work quite well.

It was found that a SIR with 1024 members produces the same ensemble mean as a GSIR with 32 members, that is, within the error bound of the SIR. Also the standard deviations of both methods are of comparable size. This leads us to the conclusion that the GSIR is a promising method for real operational problems because it combines all advantages of the SIR with a relatively low ensemble size.

One disadvantage over the SIR is that two new parameters have to be chosen, the frequency of intermittent SIR steps and the variance inflation of the observations at these steps. Preliminary experiments show that the outcome is not too sensitive to these choices, as long as they are in a range around the values presented in this paper. However, more research, probably with trial and error, is needed to make definite statements about these

matters.

References

- [1] Burgers, G., P. J. van Leeuwen, and G. Evensen, On the analysis scheme of the Ensemble Kalman Filter, *Monthly Weather Rev.*, , 1719–1724, 1998.
- [2] Doucet, A., N. de Freitas, N.J. Gordon (Eds.) *Sequential Monte Carlo methods in practice* Springer, New York, 581p., 2001.
- [3] Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, **99**(C5), 10,143–10,162, 1994.
- [4] Gordon, N.J., D.J. Salmond, and A.F.M. Smith Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEEE Proceedings-F*, **140**, 107-113, 1993.
- [5] Shannon, C.E. A mathematical theory of communication, *Bell Syst. Tech. J.*, **27**, 379-423, 1948.
- [6] Van Leeuwen, P.J. An variance-minimizing filter for large-scale applications, *Mon. Weather Rev.*, *resubmitted* 2003. Available from <http://www.phys.uu.nl/leeuwen>.