Assimilation of Geosat Altimeter Data for the Agulhas Current Using the Ensemble Kalman Filter with a Quasigeostrophic Model

GEIR EVENSEN
Nansen Environmental and Remote Sensing Center, Bergen, Norway

PETER JAN VAN LEEUWEN
IMA, University of Utrecht, Utrecht, The Netherlands
(Manuscript received 14 December 1994, in final form 16 May 1995)

ABSTRACT

The ring-shedding process in the Agulhas Current is studied using the ensemble Kalman filter to assimilate Geosat altimeter data into a two-layer quasigeostrophic ocean model. The properties of the ensemble Kalman filter are further explored with focus on the analysis scheme and the use of gridded data. The Geosat data consist of 10 fields of gridded sea surface height anomalies separated 10 days apart that are added to a climatic mean field. This corresponds to a huge number of data values, and a data reduction scheme must be applied to increase the efficiency of the analysis procedure. Further, it is illustrated how one can resolve the rank problem occurring when a too large dataset or a small ensemble is used.

1. Introduction

The Agulhas Current is a western boundary current flowing along the east coast of South Africa. Its water originates from the Mozambique channel (see, e.g., Sætre and da Silva 1984) and from east of Madagascar (e.g., Lutjeharms et al. 1981) as part of the subtropical gyre in the Indian Ocean. After the Agulhas reaches the tip of the continent, it makes a large anticyclonic turn and flows back east into the southern Indian Ocean as the Agulhas Return Current. At the turning point, the so-called retroflexion area, large anticyclonic eddies are formed that travel into the South Atlantic Ocean (Lutjeharms and Gordon 1987; Lutjeharms and van Ballegooijen 1988; Gordon and Haxby 1990; Wukker et al. 1990; Feron et al. 1992; and Van Ballegooijen et al. 1994). These eddies are the largest in the whole ocean and are believed to play a significant role in the general thermohaline circulation (De Ruijter 1982; Gordon 1986; Gordon et al. 1992). Several studies have estimated the magnitude of the transport and the number of eddies that are shed. Using satellite thermal infrared imagery, Lutjeharms and van Ballegooijen (1988) found that six to nine large eddies containing warm water were shed annually from the Agulhas Retroflexion, while Feron et al. (1992) observed stronger variability with four to eight eddies a year from an altimetric analysis.

The mechanism of ring shedding is still under investigation. The anticyclonic turn of the Agulhas can be explained theoretically by the conservation of potential vorticity, including the planetary vorticity (Ou and De Ruijter 1986; De Ruijter and Boudra 1985). In Lutjeharms and van Ballegooijen (1984) a barotropic-free inertial jet model was used to study the area. They found that a relatively low volume transport of the Agulhas resulted in a more westward penetration of the current, and vice versa. However, it turned out that their calculated path depended strongly on the bottom velocity, which seems to be in contrast with the more persistent flow patterns observed.

In De Ruijter and Boudra (1985) and Boudra and De Ruijter (1986), experiments were conducted with a barotropic and a multilayer isopycnic model, respectively, to investigate the Atlantic–Indian Ocean circulation. The studies concentrated on the influence of external parameters on the retroflexion mechanism and the interbasin circulation. They showed that the position of the zero of the wind stress curl is important for the amount of leakage from the Indian to the Atlantic Ocean. They found only regular ring formation in the multilayer model when the Agulhas was unrealistically weak. Boudra and Chassignet (1988) and Chassignet and Boudra (1988) also used a multilayer isopycnic model and found regular ring formation with a more realistic Agulhas strength and a more realistic shape of South Africa. The rings were shed along the east coast of the continent, while the observations indicated a
shedding south of the continent. The same ring-shedding position was found with the Fine Resolution Antarctic Model (FRAM).

A reduced gravity model was used by Ou and De Ruijter (1986). The separation of an inertial boundary current from a curved coastline was studied. Aside from the volume transport of the current, the curvature and the angle of separation associated with it (measured clockwise from the south) were found to be critical parameters. A larger volume transport resulted in a separation at a less negative latitude, which increased the generation of anticyclonic vorticity and a sharper eastward turn at a higher latitude. A low volume transport can even cause the current to round the continent. The jet can intersect itself if the angle of separation is large enough, which corresponds to a separation at a more negative latitude. The model is able to predict intersections and thus ring formations for parameter values corresponding to the Agulhas. Their study shows that the Agulhas is strong enough to make the anticyclonic turn and flow back in the Indian Ocean, but at the same time it is weak enough to allow for intersection on itself, thus producing large anticyclonic rings.

To study the details of the ring-shedding process, data assimilation can be used. Holland et al. (1991) used a five-layer quasigeostrophic model in which they assimilated Geosat altimeter data. They concentrated on the characteristics of two different assimilation methods, nudging and reinitialization, and not on the physics of the system. In this paper the ring-shedding process is studied using the ensemble Kalman filter introduced by Evensen (1994b) (see also the examples in Evensen 1994a) to assimilate gridded Geosat altimeter data from the Agulhas Current into a two-layer quasigeostrophic ocean model.

The main difference between the ensemble Kalman filter and the extended Kalman filter is the way the forecast error statistics is calculated. The extended Kalman filter integrates an approximate equation for the error covariance matrix where all higher-order statistical moments are neglected. This linearization has proven to be inconsistent for strongly nonlinear dynamics, and it will in many cases be a too simplified closure approximation (see Evensen 1992; Miller et al. 1994). The ensemble Kalman filter integrates an ensemble of model states from which the error covariances can be calculated. This is equivalent to a Monte Carlo method for integrating Kolmogorov's equation, which is the fundamental equation for evolution of error statistics. No closure approximations or linearizations have been applied, and the error evolution is exact to numerical truncation errors in the limit of an infinite size of the ensemble, given that the initial error distribution is exact and model errors are Gaussian.

The quasigeostrophic model (see the appendix) will probably not give a very accurate description of the dynamics in the Agulhas retroflection region. However, the model includes the most dominant physics such as barotropic and baroclinic instabilities, the topographic influence, and the planetary β effect, which are processes that play a key role for the eddy shedding and meandering growth in the Agulhas current. Further, the model is nonlinear and therefore makes the advanced data assimilation problem very challenging. This work presents a first attempt of performing an advanced data assimilation experiment where the ensemble Kalman filter is used to assimilate real data.

An overview of the Geosat dataset is given in the next section; then in section 3, the fundamentals of the ensemble Kalman filter including its analysis procedure and a data reduction scheme is discussed. Results from a pure model and ensemble forecast are given in sections 4 and 5. Finally the data assimilation experiments are presented in section 6, followed by discussions in section 7.

2. Geosat data

The Geosat data consist of 10 gridded fields of sea surface height anomalies separated 10 days apart. The data have been subject to standard corrections; for example, the sea level changes attributable to tides and static response to atmospheric pressure are removed, the geoid signal is eliminated, and corrections for orbit errors are applied [see Holland et al. (1991) for a detailed discussion of the data]. The sea surface height anomalies are added to climatic mean fields from the Southern Ocean Atlas of Gordon (1982) and the resulting fields are interpolated to the numerical grid where \( \Delta x = \Delta y = 22 \) km. Errors in the sea surface height are estimated to be 10 cm compared to the strong signal that is of the order 100 cm in the Agulhas region. The strong signal in the altimeter data for this particular region makes it an appealing test site for altimeter data assimilation experiments. The altimetric data field at \( T = 0 \) is used as an initial condition for the model and is given in Fig. 1, while another six data fields are plotted in Fig. 2. The data contain a well-identified Agulhas current flowing along the South African coast before it separates from the coast, retroflects, and leaves the region at the eastern boundary. Part of the circumpolar current can also be seen in the data. Note the eddy located at 39.5°S, 18.0°E at day 30, which 10 days later has been shed from the Agulhas Current and has left the model domain. This will be further discussed in connection with the model integrations and data assimilation experiments.

3. Ensemble Kalman filter

The ensemble Kalman filter (EnKF) is based on the theory of stochastic dynamic prediction that describes the evolution of error statistics. This theory was first introduced by Epstein (1969), and several publications have later extended it (see, e.g., Gleeson 1970; Fleming
where $\psi$ is the state vector, $g$ is a nonlinear vector function, and $dq \in \mathbb{R}^N$ is assumed to be a vector of random noise that is white in time and has a multivariate Gaussian distribution with zero mean in the spatial coordinates. The evolution of the probability density for this equation is described by the Kolmogorov's equation which for the QG model simplifies to

$$\frac{\partial \phi}{\partial t} + \sum_{i=1}^{n} g_i \frac{\partial \phi}{\partial \psi_i} = \sum_{i,j=1}^{n} \frac{Q_{ij}}{2} \frac{\partial^2 \phi}{\partial \psi_i \partial \psi_j},$$

(3)

where $Q = Q^T$ is the covariance matrix for the model errors. A derivation of this equation, which is the fundamental equation for evolution of error statistics, can be found in Jazwinski (1970). The probability density function represents the density of an infinite ensemble of possible ocean states, each having an associated probability number. The width of the probability density function corresponds to the variance of the ensemble and represents the errors in the predicted solution. Note that the stochastic forcing introduces a diffusion term that tends to flatten the probability density function (spreading the ensemble) during the integration, that is, the probability decreases and the errors increase.

If this equation could be solved for the probability density function, it would be possible to calculate statistical moments like the mean state and the error covariances at different time levels. In the extended Kal-

![Fig. 1. Initial upper and lower streamfunction fields with contour interval 0.35. The upper-layer field is the altimetric data field at $T = 0$, and in the lower layer this field is scaled down with a factor 0.4. The bathymetry is given with a contour interval of 540 m, and the initial error variance field has a contour interval of 0.0025.](image1)

![Fig. 2. Geosat altimeter data for the first 60 days with contour interval of 0.35.](image2)

1971a,b; Epstein and Pitcher 1972; Leith 1971, 1974; Pitcher 1977; Salmon et al. 1976; and Seidman 1981). These publications discuss both the use of Monte Carlo methods, which form the basis for the error prediction in the EnKF, and the use of approximate stochastic dynamic prediction on which the extended Kalman filter is based.

**a. Error prediction**

It is worthwhile to briefly review the basics of predictability theory since advanced data assimilation methods are built directly on these concepts. The state vector at a specified time, $\psi$, can be represented by a single point in an $n$-dimensional phase space $\mathbb{R}^n$. Thus, time evolution of the state vector $\psi$ is described by continuous motion of the point along a trajectory in phase space. The uncertainty in the state vector can be represented by a large ensemble of possible states, each assigned an individual probability number. Suppose there are $N$ points altogether, where $N$ is a very large number, and $dN$ is their density (points per volume increment) at any location. As the number of such phase points approaches infinity, one can define a probability density function

$$\phi(\psi) = \frac{dN}{N},$$

(1)

which can vary throughout the space. In Evensen (1994b) it was shown that the QG model could be written in discrete form as an Itô stochastic differential equation describing a Markov process

$$d\psi = g(\psi, t) dt + dq,$$

(2)
man filter, equations for the moments of $\phi$ can be derived from Kolmogorov’s Eq. (3). For linear dynamics these equations are uncoupled, and if the initial probability density is Gaussian, that is, given by the mean and the covariance, it will remain so for all times. This is the reason why the Kalman filter for linear models is the optimal sequential data assimilation method; that is, an exact equation is used for the evolution of the error covariance matrix.

For a nonlinear model, the mean and the covariance matrix will not in general characterize $\phi(\Psi, t)$, because the hierarchy of equations for the different statistical moments are coupled, and even if the initial distribution is Gaussian, it will not remain so. However, the mean and the covariance matrix do determine the mean path and the dispersion about that path. In the extended Kalman filter the coupling to the higher-order moments is removed by the crude closure assumption that their contribution can be neglected.

An alternative method for predicting error statistics is to solve Kolmogorov’s Eq. (3) using Monte Carlo methods. A large cloud of ocean states, that is, points in phase space, can be used to represent a specific probability density function. By integrating such an ensemble of states forward in time it is easy to calculate approximate estimates for moments of the probability density function at different time levels. In this context the Monte Carlo method might be considered a particle method in phase space. When the size $N$ of the ensemble increases, the errors in the solution for the probability density will approach zero at a rate proportional to $N^{-1/2}$. For practical ensemble sizes, say $O(100)$, the errors will be dominated by statistical noise, not by closure problems or unbounded error variance growth as have been observed in the extended Kalman filter (see Evensen 1992, 1994b).

When the Monte Carlo method is applied one first calculates a best-guess initial condition based on available information from data and statistics. The model solution calculated from this initial state is denoted the central forecast. The uncertainty in the best-guess initial condition is represented by the initial variance. An ensemble of initial states is then generated where the mean equals the best-guess initial condition, and the variance is specified based on knowledge of the uncertainty in the first-guess initial state. The covariance or smoothness of the ensemble members should reflect the true scales of the system; for example, the internal Rossby radius is the physical scale for a quasigeostrophic model. A procedure for generating such pseudorandom fields with a specified variance and covariance was outlined in Evensen (1994b).

The effect of external error growth must be included to give reliable estimates for the evolution of errors. If these errors were known, their effect could be included by integrating each ensemble member as a stochastic differential equation where the stochastic forcing fields are drawn from an ensemble having the correct model error probability distribution.

b. Analysis procedure

In Evensen (1994b) an analysis scheme was proposed where the traditional update equation used in the Kalman filter is applied, except that the gain is calculated from the error covariances provided by the ensemble. It was also illustrated that a new ensemble with error statistics representing the analyzed state could be generated by updating each ensemble member individually using the same analysis equation.

Traditionally the analysis in the Kalman filter is calculated from

$$\Psi_k^f = \Psi_{k+1}^f + K_k(d_k - H_k\Psi_k^f),$$  \hspace{1cm} (4)

where the measurement equation is defined as

$$d_k = H_k\Psi_k^f + \epsilon_k$$  \hspace{1cm} (5)

and relates the data linearly to the true state through the measurement matrix $H_k$. The errors in the data $\epsilon_k$ are assumed white in time and with an error covariance matrix $W_k = \epsilon_k \epsilon_k^T$. The optimal variance-minimizing weights are given by the Kalman gain

$$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + W_k)^{-1}.$$  \hspace{1cm} (6)

This analysis procedure can be characterized as the optimal variance-minimizing method. An inherent assumption is that the error statistics are Gaussian with vanishing higher-order statistical moments. This is in general not true for nonlinear dynamics where the probability density function may be far from Gaussian and higher-order moments may contribute significantly (see Miller 1994 for an example). An optimal-analysis procedure would be to calculate the probability density function from the ensemble and then calculate the most probable state using Bayes theorem. However, the computational load is at present extreme for higher-dimensional problems and such methods have been applied only with good results for low-dimensional problems (E. F. Carter and R. N. Miller 1994, personal communication).

An ensemble of ocean states can be stored in the matrix $A_k^{n \times N}$, where $n$ is the number of state variables for each ensemble member and $N$ is the number of ensemble members. Given the ensemble forecast $A_k^f$, one can calculate the error covariance matrix for the ensemble from

$$P_k^e = \frac{(A_k^e - M_k^e)(A_k^e - M_k^e)^T}{N - 1},$$  \hspace{1cm} (7)

where $M_k^e$ contains the predicted ensemble mean in each column. This means that the rank of the error covariance matrix $P_k^e$ will be less than or equal to the number of members in the ensemble. The rank of the
so-called representer matrix $R_i = H_i P_i H_i^T$ will be the least of the rank of $P_i$ and the rank of $H_i$.

This has a serious implication for applications where the number of data are greater than the ensemble size, which makes the representer matrix $R_i$ singular. Adding the measurement error covariance matrix increases the rank for the matrix to be inverted in (6), but there is still no guarantee that the system becomes well conditioned.

However, there are ways of resolving this rank problem. Note that the covariance functions calculated from the ensemble will still give a good estimate of the actual covariance functions, and the addition of even further ensemble members will only help reducing the statistical noise. It is convenient to rewrite the analysis (4) as

$$\psi_i = \psi'_i + B_i^T b_i,$$

with $B_i = H_i P_i^T$, where the rows in $B_i$ are influence functions or so-called representatives for each of the measurements. The vector $b_i$ contains amplitudes for each of the influence functions and is found by solving the system

$$(H_i P_i^T H_i^T + W_i) b_i = d_i - H_i \psi'_i.$$

Note that it is not necessary to calculate the full error covariance matrix to compute the analyzed estimate from (8) and (9). The influence functions are found from

$$B_i = \frac{H_i (A_i - M_i) (A_i - M_i)^T}{N - 1},$$

where the $m \times N$ matrix $S = H_i (A_i - M_i) (A_i - M_i)^T(N - 1)^{-1}$. The representer matrix $R_i$ can be calculated from

$$R_i = H_i P_i H_i^T = H_i B_i^T S = S S^T(N - 1)^{-1}.$$ The coefficients $b_i$ can now be solved for in (9) if the matrix $R_i + W_i$ is nonsingular.

A data reduction scheme is often required for practical applications when gridded data are used. If the gridded data are applied directly that requires the inversion of a huge matrix with order equal to the number of grid points in the gridded data fields. The conditioning of this matrix will also be poor due to the correlation between neighboring grid points. In the cases below data values from every third grid point are used. This reduces the dimension of the system (9) with a factor of 1/9, and a total of 297 data points are used in each analysis. The distance between the data points becomes about 66 km. Subsampling the data also improves the conditioning of (9), due to the lower correlation between the applied neighboring data points. It is also expected that such subsampling yields only minor differences in the results since the unused data contain only additional information about small-scale structures in the data fields.

By using an eigenvalue decomposition when solving (9) and discarding the contribution to $b_i$ from the noisy eigenvectors corresponding to the least significant eigenvalues, it is possible to eliminate the noise resulting from the poor conditioning. Here 130 to 140 eigenvalues, accounting for about 95% of the variance, were used in the analyses. This corresponds to finding the Moore–Penrose inverse solution of (9) after a formal reduction of the rank of the system (Bennett 1992).

This approach does in fact resolve the conditioning problem caused by dependent measurements, and it is also possible to have a larger number of measurements than ensemble members, which results in a rank problem. The analysis will be well posed even with a very small ensemble when only the contribution from non-zero eigenvalues is used. However, the accuracy of the analysis is still determined by the size of the ensemble. For an extensive discussion of this procedure, see Bennett (1992, chapter 6), where it was used to resolve the poor conditioning caused by linearly dependent observations.

An ensemble with the correct analyzed error statistics must be calculated during each analysis. In Evensen (1994b) it was shown that if the same gain is used to update each individual member of the ensemble, the resulting ensemble will have the correct error statistics. In the alternative analysis procedure given above, this corresponds to solving (9) for each ensemble member, that is, each ensemble member results in a new right-hand side, and thus a new $b_i$, and the analysis is then calculated for each individual ensemble member using (8).

4. Model forecast

First a pure model forecast will be generated and compared with the data. This will provide some information about the properties of the model, and it is also important to verify if the model is at all valid for this region. The same initial streamfunction is used for all experiments, where the upper-layer streamfunction is given by the first of the 10 gridded data fields and the lower-layer streamfunction is the same as the upper layer but scaled down by a factor 0.4. These initial fields are shown in Fig. 1 together with the initial variance of the ensemble and the bathymetry.

The upper- and lower-layer streamfunction from the pure model run with no assimilation of data are shown in Figs. 3 and 4. During the first part of the integration the model solution compares well with the gridded data; that is, the forecast generated by the model looks reasonable compared to the typical structures seen in the gridded data (compare Figs. 2 and 3).

However, after about two months of integration a strong instability develops in the Agulhas Current along the continent. This instability can be understood as follows. The quasigeostrophic model conserves potential vorticity. As the water flows along the east coast...
of South Africa, it gains planetary vorticity and loses relative vorticity. This relative vorticity is in a real situation dissipated in the boundary layer along the coast. However, the model has a free-slip boundary, so the velocity gradients are smaller and this dissipation is less strong. Consequently, when this water reaches the detachment point it has accumulated a large negative relative vorticity, which results in the formation of a circulation cell centered at 36.5°S, 20.5°E. Note that the reason for the retroflection is the gain in relative vorticity after detachment (see the introduction). In the model run the current gains extra relative vorticity due to the free-slip condition before detachment. This results in a faster turn immediately after detachment in the pure model run. Water entrained on the eastward side of the main current also has a negative relative vorticity when it reaches the turn in the coastline at 34.5°S, 25.0°E. Indeed, a strong eddy forms, which eventually distorts the main current. At day 60 this eddy even attracts water from the Agulhas Return Current, thus enhancing and disturbing the Agulhas Current even more.

Note also that the detachment of the ring located at 40.0°S, 18.0°E at day 20 is slower in the model forecast than in the data (compare Figs. 2 and 3). The ring is shed and escapes the region between day 30 and day 40 in the data, while in the model forecast it escapes the region between day 40 and day 50. This feature is consistent with comparisons of quasigeostrophic and primitive equation models for instabilities of eastward-flowing jets. In a quasigeostrophic model the instabilities of the jet grow faster, but the final wave steepening and ring shedding is slower compared to a more realistic primitive equation model. This is because the horizontal scales become so small during the wave steepening that ageostrophic effects are important. Another interesting feature is the barotropization of the Agulhas ring just after its detachment (see Fig. 4). This is in agreement with observations of Agulhas Rings (e.g., Gordon and Haxby 1990).

The model-predicted fields contain stronger variability than the data, which may be a result of a too strong smoothing of the altimeter data during the interpolation to the numerical grid. It is hoped that the data will be able to constrain the major flow pattern, while the model will regenerate some smaller-scale structures in the data assimilation experiments to be discussed below.

5. Ensemble forecast

Before the data assimilation experiments are performed it is instructive to examine the statistics produced by a pure ensemble integration. The error statistics provided by an ensemble integration will, if the
model is realistic enough, be valid for very long integration times compared to the error covariance evolution in the extended Kalman filter. The dynamical instabilities are saturated in the ensemble integration, while those in the integration of the error covariance equation are allowed to have unbounded growth due to the statistical linearization or closure applied (Evensen 1994b). If realistic system noise is included, the variance of the ensemble provides an estimate of the errors in the model forecast.

Here an ensemble of 500 members is used, and the initial error variance field for the upper and lower layer is given in Fig. 1. Further, the system noise is neglected to examine the sole properties of the internal dynamical instabilities. The time evolution of the mean square errors is given in Fig. 5. The mean square errors grow nearly exponentially during this 90-day integration. This is a too short integration time for the error growth to saturate; in fact, it may not saturate at all, because it has not been ensured that the approximate open boundary scheme conserves energy (Evensen 1993).

The model forecast, the mean, and the error variance are given in Fig. 6 at the final day. The mean solution is smoother than the pure model forecast but it still contains eddies and meanders of similar structures and at approximately the same locations as in the pure model forecast. The strongest error growth is located in the region close to the South African land boundary. In this area nonlinear instabilities can extract energy from the strong Agulhas current, and this is also the most eddy active region in the model solutions. If a measurement antenna is to be designed for this region, it is obvious that the largest number of measurements should be taken from the areas with largest error variance, at least if one searches for a low-variance estimate of the ocean state.

6. Data assimilation experiments

The error-covariance statistics for the data is specified using a correlation model

$$\text{cov}(x_1, x_2) = e(x_1) e(x_2) \exp \left( - \frac{||x_1 - x_2||^2}{r^2} \right).$$

(11)

where the horizontal decorrelation length is $r_x = 3.0$ and the error variance is $e^2(x) = 0.02$ in the interior of the model domain, but with errors decreasing to zero approaching the coastal boundary (see Fig. 1). Since the initial data field is used to initialize the model, the same parameters are used for the initial error statistics. A vertical decorrelation equal to $\exp(-0.3)$ is imposed between layer 1 and 2 for the initial error statistics.

We will not give an elaborate discussion on how to specify the correct model errors, but rather argue for a crude inclusion of the model errors in this experiment. Even if the model errors are unknown, they should be represented in some way, and here normal distributed model errors with zero mean and a specified covariance are assumed. The correlation model (11) was used with the same decorrelation length $r_x = 3.0$ but with a much lower variance $e^2(x) = 0.0005$. This low variance ensures that the main contribution to the error covariance

![Figure 5. Time evolution of mean-square errors in the pure ensemble integration case.](image1)

![Figure 6. Results from the pure ensemble integration at day 90. The contour interval for the variance plots (bottom) is 0.1, while it is 0.0025 for all other variance plots in this paper.](image2)
evolution is still from the dynamical evolution. As a cheaper alternative to integrating each ensemble member properly as a stochastic differential equation, we have for now chosen to add pseudo random fields drawn from a distribution having the prescribed error statistics to each ensemble member every eighth time step, which corresponds to the characteristic time $\Delta t = 1.0$ for the dynamics. In doing this a component of random walk in phase space is included, which will increase the variance of the ensemble.

Now an experiment using the full ensemble Kalman filter is performed with an ensemble size of 500. A time series of the analyzed estimates for the upper and lower layer is given in Figs. 7 and 8. From these plots it is first of all evident that the streamfunction estimates are similar to the data and contain the major structures observed by the data. The Agulhas Current is clearly less unstable than what was found in the pure model forecast. Note the eddy formed just below the southern tip of South Africa in Figs. 8 and 9. It is not visible in the data, so it is an essential feature of the ensemble of model runs; it was also observed in the mean of the ensemble. However, maybe the formulation of the model poses the problem. Because of the overestimation of depth changes in quasigeostrophic models, a depth-influence reduction of 90% was applied. In reality the depth of the lower layer is zero at this location, which maybe would prevent this eddy from being formed.

Let us now concentrate on the shedding of the Agulhas Rings. If we compare the EnKF results with the pure
model forecast and the data, it can be observed that the ring is formed faster in the EnKF experiment than in the model prediction, and in more agreement with the data. The too slow final wave steepening and ring shedding in the quasigeostrophic model, which is caused by the lack of ageostrophic effects in the quasigeostrophic model, is here accounted for by the assimilation of the data, which of course contain these ageostrophic effects. It is interesting to see that the lower layer does not show this difference in ring-formation speed.

Another difference between the model forecast and the EnKF experiment is the fact that the Agulhas ring formed in the former is stronger than in the EnKF experiment. This can be seen in both layers. In the lower layer, also the horizontal extension of the ring during formation is larger in the model forecast.

Finally, a difference in the position of the neck region, the region where the ring is still connected with the Agulhas Current, can be observed. This effect is more visible in the lower layers. From Fig. 4, the model forecast, the neck region is centered at about 39.0°S, 20.0°E. In Fig. 8, the EnKF experiment, the neck region is centered at 40.5°S, 19.0°E. The difference is about 180 km. Especially the northern rim of the neck region seems to be active in the ring-shedding event. This can be expected because a westward-flowing jet is more unstable than an eastward-flowing jet. The model forecast seems to attribute significance for ring shedding to both northern and southern rims of the neck region. Maybe part of this can be explained by the stronger ring in the model forecast, which entrains water northward in the neck region. Another reason can be that Rossby wave growth is faster in quasigeostrophic models than in primitive equation models; ageostrophic mo-

7. Discussion

Here fields of gridded Geosat data, separated 10 days apart, were used in a data assimilation experiment to
study the ring-shedding process in the Agulhas current. The new ensemble Kalman filter (EnKF) proposed by Evensen (1994b) was used to assimilate the data into a two-layer quasigeostrophic (QG) model. We found that the method produced results consistent with the data and that the assimilation of data provided a means to correct for deficiencies or neglected physics in the QG model. The method limited the too fast meander growth observed in QG models compared to more advanced models, and the eddy shedding, which is to slow in QG model due to the lack of ageostrophic effects, was enhanced by the data. These results suggest that a data assimilation system can indeed be useful for pure physical process studies and can be used to account for ageostrophic effects contained in the data that are missing in the QG model. A conclusion of this work is that by allowing the model to contain errors it is possible to introduce to the results physics contained in the data that were neglected in the model formulation. In a strong constraint variational formulation this would not be possible. The model predictions at each analysis time were compared with the data, and it was found that except for the effects mentioned above, the QG model produced relatively realistic results on the time and spatial scales used here.

The ensemble Kalman filter is another data assimilation method where model predicted error statistics are used to determine the influence functions and weights between the data and the model forecast. Model-predicted error statistics are also used in other advanced methods like the Kalman filter, the Kalman smoother or sweep algorithm (Bennett and Budgell 1989), and the representor method (Bennett 1992).

The disadvantage of suboptimal schemes like optimal interpolation is that the analysis is based on ad hoc assumptions made for the forecast covariances. It is very hard to verify that the assumed statistics is consistent, especially for nonlinear systems with unstable dynamics, where the forecast variance and covariance may vary significantly over the domain.

On the other hand, if the model-predicted error statistics used in advanced methods are wrong or biased, then the data assimilation procedure may result in poor estimates. Note that this is not a problem with the data assimilation method, but one should probably not use that particular model at all since it is inconsistent with the physics studied.

The use of a finite ensemble size will introduce some errors in the estimates of the statistical moments used in the assimilation algorithm. Statistical noise in the covariance fields will result in incorrect relative weights between dynamics and data, and noise in the covariance functions will exist all through the domain and contribute to unsystematic errors in the estimate. Note that the amplitude of these errors can be made negligible by increasing the ensemble size, however, for practical applications, especially using primitive equation models in the ocean or operational forecast models for the atmosphere, a very large ensemble is not affordable.

To avoid the influence of noise in the covariance functions when using a small ensemble, it is possible to apply a correlation model like (11) but still calculate the estimated standard deviation of the error field, $\epsilon(x)$, from the ensemble. Alternative methods can be constructed based on the analysis (9) and (10). Even if an analytical covariance model replaces the influence functions contained in $B$, in (10), one can still calculate the representor matrix directly from the ensemble. Further, one could imagine an approach where one first calculates the influence functions from the ensemble and next fit analytical covariance functions to the predicted ones. This would result in a class of ensemble optimal interpolation methods, which at least take advantage of predicted error variances. However, the main advantage of the ensemble Kalman filter, namely, the possibility of using consistent influence functions based on the current state of the ensemble, is lost.

The analysis scheme becomes extremely expensive to compute when gridded data are used due to the large matrix that must be inverted. Here, the gridded data were subsampled and only data from every third grid point were used. This seems to be a consistent approach since additional data values will add a number of small eigenvalues to the inversion and therefore contribute only to low-amplitude and small-scale noise in the estimated fields. The strong correlation between neighboring data values is reduced when the data are subsampled, and this also improves the conditioning of the analysis.

If an eigenvalue decomposition is used for solving the system and only the significant eigenvalues are used, one can resolve the problem with poor conditioning caused by correlated measurements (see Bennett 1992). Here it was shown that this method also resolves the problem with the singular matrix introduced when a larger number of measurements than ensemble members is used. In addition, this approach corresponds to introducing new influence functions $B[Z$ where the $m \times m$ matrix $Z$ contains the eigenvectors of the matrix in (9). Those of these new influence functions that correspond to the most significant eigenvalues will be smooth and contain only low-wavenumber information, while the influence functions corresponding to the small and less significant eigenvalues will contain all the high wavenumber information. Thus, by including only the contribution from the significant eigenvalues, one resolves the problem with poor conditioning, and the influence of the noise in the error covariance statistics is reduced.

It is not clear whether it would be better to use the original but corrected data collected along the satellite tracks, as opposed to interpolating them onto a grid first. The analysis scheme used in sequential algorithms like the Kalman filter has the nice property of performing the actual data interpolation directly, using the
model forecast as background field and the predicted error statistics for determining the weights. It may be simpler to work with gridded data than the original raw data collected along the satellite tracks. Problems with using the raw data directly are connected to the interpolation needed because of the model time step and the specific ground-track pattern. In addition, the response of a nonlinear system to isolated updates along a satellite track may also lead to spurious effects. As an example, a single datum measuring one location of some dynamical structure may generate a “false” Rossby wave in the analysis. Unless new data are available soon enough to correct this, such spurious effects may corrupt the results. However, if gridded data are used, then one should use a consistent interpolation algorithm like objective analysis, which provides error estimates for the gridded data field, and this error estimate should again be used in the assimilation procedure. It is also expected that structures in the gridded fields may be lost during the gridding process. However, in the Agulhas region the Geosat signal is very strong (of order 1 m and with errors of about 10 cm), and it is believed that the gridded data are reasonably accurate too.

Acknowledgments. The authors would like to thank W. P. Budgell and P. Houtekamer for interesting discussions during part of this work. Thanks also to W. R. Holland for making the gridded data available. C. Evensen was supported by the Research Council of Norway, and P. J. van Leeuwen was sponsored by the Space Research Organization Netherlands (SRON) under Grant EO-002. This work has also received support from the Norwegian Super Computing Committee (TRU) through a grant of computing time.

APPENDIX

Ocean Model

The ocean model is the multilayered and nonlinear quasi-geostrophic model on a β plane (Pedlosky 1987). It describes conservation of potential vorticity ζ in each layer on an f plane. The mean-layer thicknesses are Dl, and the density in each layer is ρl, where l denotes layer number: l = 1 in the upper layer. Here Ψl is the streamfunction in layer l. The horizontal length scale is Rl, the internal Rossby radius of deformation of the upper layer, given by $R_l = [(\rho_2 - \rho_1)/gD_l](\rho_0f^2)^{-1}$, where g is the gravitational acceleration, $\rho_0$ is averaged density, and f is the Coriolis parameter. The characteristic horizontal velocity is denoted $U_l$, yielding a timescale $T = R_l/U_l$. The pressure scale is $\rho_0fU_R$, and the streamfunction scale is $U_R R_l$. The nondimensional quasi-geostrophic equations are

$$ \left( \frac{\partial}{\partial t} + u_l \frac{\partial}{\partial x} + v_l \frac{\partial}{\partial y} \right) \zeta = 0, \quad l = 1, n, \quad (A1) $$

where $n_l$ is the number of layers and the velocities are the geostrophic approximations

$$ u_l = -\frac{\partial \Psi_l}{\partial y}, \quad v_l = \frac{\partial \Psi_l}{\partial x}. \quad (A2) $$

The vorticity in each layer is given by

$$ \zeta_l = \nabla^2 \Psi_l + Fr_{l,2}(\Psi_2 - \Psi_1) + \beta y, \quad (A3a) $$

$$ \zeta_l = \nabla^2 \Psi_l - Fr_{l,1}(\Psi_l - \Psi_{l-1}) + Fr_{l,2}(\Psi_{l+1} - \Psi_l), + \beta y \quad (A3b) $$

$$ \zeta_{l,1} = \nabla^2 \Psi_{l,n} - Fr_{l,n}(\Psi_{n} - \Psi_{n-1}) + \beta y + \eta, \quad (A3c) $$

where $l = 2, n, l - 1$. When $\zeta$ is known, this is a set of coupled Helmholtz equations for the streamfunction $\Psi$. The Laplacian is $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$, and the constants $Fr_{l,1}$ and $Fr_{l,2}$ are “nondimensional Froude numbers”

$$ Fr_{l,1} = \frac{D_l \rho_2 - \rho_l}{D_l \rho_l - \rho_{l-1}}, \quad Fr_{l,2} = \frac{D_l \rho_2 - \rho_l}{D_l \rho_l - \rho_{l+1}}. \quad (A4) $$

The bottom topography term is

$$ \eta = \frac{1}{\epsilon} \frac{h_b}{D_n}, \quad (A5) $$

with $\epsilon$ as the Rossby number and $h$ as bottom topography.

For the experiments in this paper a two-layer version with a density jump of 1 kg m$^{-3}$ is used. The equations are solved on a 51 × 65 grid where $\Delta x = \Delta y = 22$ km compared to the internal Rossby radius, which is about 33 km in this region. The upper and lower layers are, respectively, 900 and 4500 m thick. The time-stepping routine was discussed in Evensen (1992). A fourth-order Shapiro filter is used every sixteenth time step to dissipate $2\Delta x$ noise. The open boundary conditions are described in detail by Evensen (1993).

REFERENCES


