Application of ensemble-based techniques in fish stock assessment

Rune Grønnevik & Geir Evensen

SARSIA

Grønnevik R, Evensen G. 2001. Application of ensemble-based techniques in fish stock assessment. *Sarsia* 86:517-526.

A basic fish-stock assessment system requires an integrated use of a model for the time evolution of a fish stock and information about the fish stock from catch data. Typical for common fish stock assessment systems has been the use of fairly simplistic data assimilation methodologies for the integration of observations with the dynamical models. On the other hand, there has been a fast development of assimilation techniques which can be used with highly nonlinear and complex dynamical models. In this paper some of these methods, which are based on Monte Carlo formulations, will be examined with a simple fish stock model. The methods used are the ensemble Kalman filter, the ensemble Kalman smoother and the simpler ensemble smoother.

Geir Evensen & Rune Grønnevik, Nansen Environmental and Remote Sensing Center, Edvard Griegsvei 3A, N-5059 Bergen, Norway. – Geir Evensen also at Department of Mathematics, University of Bergen, Norway. E-mail: geir@nersc.no

Keywords: Fish stock estimation; data assimilation; ensemble Kalman Filter; ensemble Kalman Smoother.

1 INTRODUCTION

Gudmundsson (1987, 1994) introduced the extended Kalman filter (EKF) as a filtering technique for use in fish stock assessment by implementing and applying it with a prediction model for Icelandic cod. The EKF is a filtering method similar to the original Kalman filter where the error covariance statistics are evolved forward in time using a linearized version of the dynamical model. It has been shown in a number of studies that this linearization may be too severe when the model dynamics become too nonlinear, e.g., Evensen (1992) and Miller & al. (1994). This has initiated the search for new filtering techniques which are designed to handle strongly nonlinear dynamical models in a consistent manner.

One such method, the ensemble Kalman filter (EnKF), was proposed by Evensen (1994). It was derived as a Monte Carlo method for solving the Kolmogorov's equation for the time evolution of the probability density function for the model state. Together with an efficient algorithm for the computation of the analysis step, the EnKF provides a consistent and efficient formalism which makes it possible to apply filtering techniques with highly nonlinear models in huge state spaces. The EnKF has been used and examined in several applications, e.g., Evensen & van Leeuwen (1996), Evensen (1997a), and Houtekamer & Mitchell (1998).

Later developments include the ensemble smoother (ES) from van Leeuwen & Evensen (1996) which allows for the information from measurements to be propagated backward in time to improve the estimates at prior times. A recent extension of this method is the ensemble Kalman smoother (EnKS), which has proven to work significantly better with nonlinear dynamics and is formulated as a direct extension of the EnKF. Comparison of results from the ES and the EnKS provides an indication of the importance of nonlinear effects in the dynamical model.

The ensemble-based methods comprise an approach to data assimilation which attempts to efficiently solve the fully nonlinear assimilation problem rather than simplifying the already approximate EKF, see e.g. Todling & Sivakumaran (1998) for a summary of different suboptimal Kalman smoother implementations. The ensemble techniques have been shown to resolve major problems experienced with the EKF in Evensen (1994) and Miller & al. (1999). First of all it is ensured that the time evolution of error statistics does not suffer from linearizations or other approximations in the assimilation schemes. Further, there are no restrictions imposed if the basic model is extended to a more complex and nonlinear one at a later stage. This is one of the major strengths of the ensemble techniques. The methods are model independent and it is possible to improve, extend or even replace the model while keeping the present assimilation system unchanged. Thus, it is possible to develop a generic and model independent implementation of the ensemble methods.

The main purpose of this manuscript is to introduce the ensemble-based assimilation techniques to the field of fish stock assessment. Here the EnKF, ES and EnKS have been implemented with the fish stock model proposed by Gudmundsson (1987, 1994), and his experiments have been revisited using the new methods.

2 THE FISH STOCK PREDICTION MODEL

The model by Gudmundsson (1987, 1994) starts with the youngest ages and flows towards the oldest, and numbers at age are converted to catch at age. Thus we need a recruitment model to estimate the number of fish entering the fish stock each year. In the absence of such a model we have used a constant recruitment with high variance. Gudmundsson (1994) has given several alternative models for describing the fishing mortality, but the main features are the same for all of these. The major assumption is that the fishing mortality during one year will be somehow related to the fishing mortality in previous years.

The model by Gudmundsson (1994) is based on the following equations, where we have used the common notation where a is age, C is catch, N is abundance, F is fishing mortality, M is the natural mortality which is chosen to be constant and $Z = F + M$ is total mortality, i.e.,

$$C(a,t) = N(a,t)e^{-Z(a,t)} F(a,t) / Z(a,t) \quad \text{for } 4 \le a \le A, \qquad (1)$$

$$N(a,t) = N(a-1,t-1)e^{-Z(a-1,t-1)} \quad \text{for } 5 \le a \le A, \qquad (2)$$

$$\log F(a,t) = U(a,t) + V(t) + \mu_1(a,t) \quad \text{for } 4 \le a \le A, \qquad (3)$$

where fish is assumed to enter the fishery at age $a = 4$ and the oldest fish caught is $A = 10$. The abundance of four year old fish in equation (2) is determined by a recruitment model (assumed constant in our model). In equation (3), $U(a, t)$ represents the selectivity. It may differ between different fishing fleets as it is determined by the technology employed in fishing (e.g. mesh size). Further $V(t)$ represents the overall change in fishing mortality due to fishing gear. The selectivity is given as

$$U(a,t) = \begin{cases} U(a,t-1) + \mu_2(a,t), & \text{for } 4 \le a < a_m, \\ U(a_m,t-1) + \mu_2(a_m,t), & \text{for } a_m \le a \le A, \end{cases} \qquad (4)$$

where a_m is set to 9. Thus, fish older than a_m experience no difference in selectivity. To avoid an underdetermined system the following conditions are prescribed to model $V(t)$,

$$V(t) = Y(t) + \mu_3(t), \qquad (5)$$

$$Y(t) = Y(t-1) + \alpha + \mu_4(t). \qquad (6)$$

Here α is a trend term, i.e. a constant improvement of fishing gear or constantly increasing fishing effort. We have modeled the fishing mortality as a multivariate random walk process through the random variables, μ_i, which are mutually independent, normally distributed with mean equal to zero, has a prescribed variance (discussed below) and are serially uncorrelated. Considering actual causes of variations in fishing mortality rates, the weather and irregular variations in conditions in the sea are mainly transitory, but technological development and changes in fleet size induce more permanent variations.

We have used approximately the same model parameters and initial and boundary conditions as in Gudmundsson (1994) to allow for comparison of results. This includes the use of a constant recruitment rate with high initial variance or uncertainty and a constant natural mortality, but unlike Gudmundsson (1994) we have neglected effects from immigration and emigration. The model is examined using catch at age data on Icelandic cod from 1977 to 1990.

When the model is integrated forward in time the model variables are updated in the sequence $Y(t)$, $V(t)$, $U(a, t)$, $\log F(a, t)$ and abundance $N(a, t)$. A practical implementation of the system of equation is obtained by modeling the stochastic errors as follows:

$$Y(t) = (Y(t-1) + \alpha)(1 + \sigma_0\mu_4(t)), \qquad (7)$$

$$V(t) = Y(t)(1 + \sigma_0\mu_3(t)), \qquad (8)$$

$$U(a,t) = \begin{cases} U(a,t-1)(1+\sigma_0\mu_2(a,t)), & \text{for } 4 \le a < a_m, \\ U(a_m,t-1)(1+\sigma_0\mu_2(a_m,t), & \text{for } a_m \le a \le A, \end{cases} \qquad (9)$$

$$\log F(a,t) = (U(a,t) + V(t))(1 + \sigma_0\mu_1(a,t)) \quad \text{for } 4 \le a < A, \qquad (10)$$

$$N(a,t) = N(a-1,t-1)e^{-Z(a-1,t-1)} \quad \text{for } 5 \le a < A \qquad (11)$$

$$C(a,t) = N(a,t)e^{-Z(a,t)} F(a,t) / Z(a,t) \quad \text{for } 4 \le a < A. \qquad (12)$$

Thus, with the variance of $\mu_i = 1$, the stochastic forcing in the model have a magnitude determined by the standard deviation $\sigma_0 = 0.04$ multiplied with the actual values of the model variables. This results in a stochastic forcing which is proportional to 4 % of the value of the model variables. Note that the catch is computed as a diagnostic variable after the abundance is updated.

3 ASSIMILATION METHODS

Gudmundsson (1987, 1994) used the so-called extended Kalman filter (EKF) to merge the information from observations with the dynamical fish stock model. The EKF is a filter method, which means that the model is integrated forward in time and every time there are measurements available these are used to reinitialize the model before the integration continues. Neglecting the time index and denoting a model forecast and analysis as ψ^f and ψ^a respectively, with measurements contained in $\boldsymbol{d}$, and the respective covariances for model forecast, analysis and measurements as $\boldsymbol{P}^f$, $\boldsymbol{P}^a$ and $\boldsymbol{R}$, the analysis equation becomes

$$\psi^a = \psi^f + \boldsymbol{P}^f \boldsymbol{H}^T (\boldsymbol{H} \boldsymbol{P}^f \boldsymbol{H}^T + \boldsymbol{R})^{-1} (\mathrm{d} - \boldsymbol{H}^f), \quad (13)$$

with the reduced error covariances given as

$$\boldsymbol{P}^a = \boldsymbol{P}^f - \boldsymbol{P}^f \boldsymbol{H}^T(\boldsymbol{H} \boldsymbol{P}^f \boldsymbol{H}^T + \boldsymbol{R})^{-1} \boldsymbol{H} \boldsymbol{P}^f. \quad (14)$$

This reinitialization is determined as a weighted linear combination of the model prediction and the measurements. The weights are the inverses of the error covariances for the model prediction and the measurements, and the optimal linear-combination becomes the Best Linear Unbiased Estimator (BLUE).

The error covariances for the measurements, $\boldsymbol{R}$, need to be prescribed based on our best knowledge about their accuracy and the methodologies used to collect them. The error covariances for the model prediction are computed by solving an equation for the time evolution of the error covariance matrix of the model state. Given a linear dynamical model written on discrete form as

$$\psi_{k+1} = \boldsymbol{F}\psi_k, \quad (15)$$

the error covariance equation becomes

$$\boldsymbol{P}_{k+1} = \boldsymbol{F} \boldsymbol{P}_k \boldsymbol{F}^T + \boldsymbol{Q}. \quad (16)$$

The matrix $\boldsymbol{Q}$ is the error covariance matrix for the model errors. The model is assumed to contain errors, e.g. due to neglected physics and numerical approximations.

If the model is nonlinear, e.g. on the form

$$\psi_{k+1} = \boldsymbol{f}(\psi_k), \quad (17)$$

the error covariance equation (16) is still used but with $\boldsymbol{F}$ the tangent linear operator, or Jacobian of $\boldsymbol{f}(\psi)$. Thus, a linearized and approximate equation is used for the prediction of error statistics in the EKF. A comprehensive discussion of the properties of the EKF can be found in the literature, but for a convenient summary which intercompares the EKF with EnKF to be discussed next, see the review by Evensen (1997b).

3.1 Ensemble Kalman filter

The EnKF was first proposed by Evensen (1994) to resolve a major problem related to the use of the EKF with nonlinear dynamics in large state spaces. The linearization used in the error covariance equation has been shown to be invalid in a number of applications, e.g. by Evensen (1992) and Miller & al. (1994). In fact, the equation (16) is no longer the fundamental equation for the error evolution when the dynamical model is nonlinear. In this case, by using (16) one neglects contributions from higher order statistical moments, as a statistical closure approximation. For a nonlinear model where we appreciate that the model is not perfect and therefore contains model errors we can write it as a stochastic differential equation (on continuous form) as

$$d\psi = \boldsymbol{f}(\psi)dt + d\boldsymbol{q}. \quad (18)$$

This equation states that an increment in time will yield an increment in ψ and in addition there will be a random contribution to the increment from $d\boldsymbol{q}$ which is the stochastic forcing representing the model errors. From this equation one can derive the Fokker-Planck or Kolmogorov's equation which describes the time evolution of the probability density $\phi(\psi)$ of the model state,

$$\frac{\partial \phi}{\partial t} + \sum_{i=1}^{n} \frac{\partial f_i \phi}{\partial \psi_i} = \sum_{i,j=1}^{n} \frac{Q_{ij}}{2} \frac{\partial^2 \phi}{\partial \psi_i \partial \psi_j} \quad (19)$$

where $\boldsymbol{Q} = \overline{d\boldsymbol{q}d\boldsymbol{q}^T}$ is the covariance matrix for the model errors and n is the dimension of the model state. This equation does not apply any important approximations and can be considered as the fundamental equation for the time evolution of error statistics. A detailed derivation is given in Jazwinski (1970). The error covariance equation (16) can be derived by taking the first statistical moment of the Fokker Planck equation. The equation describes the change of probability density in a local "volume", which is dependent on the divergence term describing a probability flux into the local "volume" representing the impact of the dynamical equation, and the diffusion term which tends to flatten the probability density due to the effect of stochastic model errors.

The EnKF applies a Monte Carlo method to solve this equation. The probability density can be represented using a large ensemble of model states and evolved in time by integrating these model states forward in time according to the model dynamics described by the stochastic differential equation (18). This ensemble prediction is equivalent to solving the Fokker Planck equation using a Monte Carlo method, a procedure which forms the backbone for the EnKF.

An ensemble representation of the analysis equations (13-14) is used for the computation of the analysis and is further discussed in Section 3.4. See also Burgers & al. (1998) for a detailed discussion of the analysis scheme and the tutorial review of the EnKF and the EKF given by Evensen (1997b).

To summarize, ensemble integrations are used to predict error statistics which are used to update or reinitialize the model solution with a "subspace" BLUE estimate, whenever observations are available. It becomes a subspace BLUE because normally the number of members in the ensemble is less than the independent degrees of freedom in the model state.

3.2 Ensemble smoother

The ensemble smoother (ES) was first proposed by van Leeuwen & Evensen (1996) as a method for solving for the variance minimizing estimate defined by the Bayes theorem,

$$\phi(\psi \mid \boldsymbol{d}) = \frac{\phi(\psi)\phi(\boldsymbol{d} \mid \psi)}{\phi(\boldsymbol{d})}. \tag{20}$$

This equation states that the probability density for the model state, given a vector of measurements, $\boldsymbol{d}$, is equal to the density of the model state without any information from measurements, $\phi(\psi)$, times the density of the measurements, $\phi(\boldsymbol{d}|\psi)$. The denominator, $\phi(\boldsymbol{d})$ is the integral of the numerator and ensures that the integral of the posterior density becomes equal to one, (the probability of finding the model state, somewhere, is equal to one).

The ensemble smoother (ES) attempts to find the solution which minimizes the posterior error variance. Realizing that the density $\phi(\psi)$ can be found from an ensemble integration which solves the Fokker Planck equation, and that the density for the measurements often can be prescribed as a Gaussian, a variance minimizing analysis scheme similar to the one used in the EnKF can be applied. By neglecting non-Gaussian contributions in $\phi(\psi)$ one can compute the mean and error covariance of the model prediction as a function of space and time. This information can be used together with the measurement vector and the error covariance of the measurements to compute a variance minimizing (BLUE) analysis in space and time.

3.3 Ensemble Kalman smoother

Evensen & van Leeuwen (2000) derived a smoother formulation where the measurements are processed sequentially in time. The method was named ensemble Kalman smoother. It is based on an assumption that the measurements are independent in time which is most often the case, and possesses several important properties. The EnKF solution is the first guess solution for the EnKS, and the smoother solution can be found by computing a BLUE analysis computing the impact of measurements backward in time at each an analysis step in the EnKF. If the model is linear this method will give identical results to those from the ES, since the two methods are solving the same general formulation. If the model is nonlinear, this approach will have an advantage over the ES since observations are assimilated sequentially during the forward ensemble integration. This keeps the ensemble "on track" and it will be more consistent with the measurements before the smoother analysis is computed. This is in contrast to the ES which uses a pure ensemble integration, where non-Gaussian contributions can freely develop in the first guess for the analysis. In addition to introducing the EnKS, Evensen & van Leeuwen (2000) explain the connections between the different filtering methods and the impact of non-Gaussian contributions to the analysis.

3.4 The ensemble-based analysis scheme

Here we give a brief introduction to the analysis scheme used by the ensemble methods. The forecasted model states are stored in a matrix $\boldsymbol{A}$ where each column contains one member of the ensemble. An ensemble approximation of the model error covariance matrix is then given by

$$\boldsymbol{P}_e^f = \frac{1}{n_{ens}-1}(\boldsymbol{A}^f - \overline{\boldsymbol{A}}^f)(\boldsymbol{A}^f - \overline{\boldsymbol{A}}^f)^T, \tag{21}$$

where each column of the matrix $\overline{\boldsymbol{A}}^f$ contains the mean of the ensemble,

$$\overline{\psi} = \frac{1}{n_{ens}} \sum_{j=1}^{n_{ens}} \psi_j \tag{22}$$

The overline denotes an expectation value and the ensemble mean, $\overline{\psi}$, is considered to be the best guess estimate, and the spreading of the ensemble around the mean gives the error variance in the ensemble.

In order to have a variance minimizing analysis scheme, Burgers & al. (1998), we also need to create an ensemble of observations, $\boldsymbol{D}$, where each column, $\boldsymbol{d}_j$, is a perturbed measurement vector created by adding a noise vector to the observations

$$\boldsymbol{d}_j = \boldsymbol{d} + \varepsilon_j, j = 1,\ldots, n_{ens}. \tag{23}$$

Here ε_j is a vector of observation noise picked randomly from a Gaussian distribution with mean equal to zero and the standard deviation equal to the measurement standard deviation. Thus the measurement error covariance matrix is given by

$$\boldsymbol{R} \approx \boldsymbol{R}_e = \overline{\varepsilon\varepsilon^T}. \tag{24}$$

The analysis can now be computed from

$$\boldsymbol{A}^a = \boldsymbol{A}^f + \boldsymbol{P}_e^f \boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}_e^f \boldsymbol{H}^T + \boldsymbol{R})^{-1}(\boldsymbol{D} - \boldsymbol{H}\boldsymbol{A}^f). \tag{25}$$

This equation illustrates some of the properties of the ensemble methods used here. The new analyzed ensemble is generated by updating every member of the forecasted ensemble using the equations for the standard BLUE with covariances computed from the ensemble. Note that an ensemble of measurements is also created to ensure that the analyzed ensemble will have the correct covariance. Thus, using these methods, there is no need for additional resampling to create the new ensemble. This is a great computational advantage over nonlinear filters such as those discussed by, e.g., Anderson & Anderson (1999) which need additional resampling at every analysis step.

In the EnKF, the equation (25) describes the analysis if $\boldsymbol{P}_e^f$ is the model error covariance at the particular time t_k when the analysis is computed. On the other hand, if $\boldsymbol{P}_e^f$ is the error covariance for the whole model state in space *and time*, then this equation will correspond to the analysis scheme in the ES. For the EnKS the equation is used to process the observations sequentially in time and for every analysis the error covariances will contain correlations from time t_k and backward in time to the initial state. Note that these correlations are computed based on the ensemble resulting from all previously computed analyses. See Evensen & van Leeuwen (2000) for a more detailed discussion of the analysis scheme for the filter and smoothers.

4 EXAMPLE

In the following discussion we have examined the different data assimilation methods in an application with catch at age data for Icelandic cod in the period from 1977 to 1990. Effects of emigration have been neglected. However, there was a significant immigration of 8 years old cod in 1981 and 6 years old cod in 1990, from the east Greenland cod stock which is treated as an error source for the model. In Gudmundsson (1987, 1994), the immigration needed to be known in advance and incorporated into the model equations.

4.1 Model configuration and initialization

We have tried to choose the model parameters close to those used by Gudmundsson. The natural mortality is set to $M = 0.2$, and the trend parameter α, signifying an improvement of fishing gear and increased effort in time is chosen to be $\alpha = 0.025$. The fish enter the fishery at age 4, fish older than 9 years will experience no difference in selectivity and the oldest fish caught are 10 years old. Thus for each of the variables the age, a, varies from 4 to 10, except for $U(a, t)$ where a varies from 4 to 8 (with $a_m = 9$) and the variables Y, V and $U(a_m, t)$ which are independent of a.

The model state vector, ψ, contains the variables $N(a, t)$, $\log F(a, t)$, $U(a, t)$, $Y(t)$, $V(t)$, $U(a_m, t)$ and $C(a, t)$. The catch $C(a, t)$ has been included as a diagnostic variable for practical reasons which allows for the use of a linear measurement functional operating directly on the catch variable, rather than a nonlinear functional operating on the model state. Every time we have a measurement of the catch, which in this particular experiment is at every time step (i.e. once a year), the predicted catch is calculated and the data assimilation is performed. The updates of the rest of the model state from catch observations is performed through the use of multivariate cross correlations contained in the ensemble of model states.

As with the model parameters, the initial and boundary conditions are chosen close to the values used by Gudmundsson (1994). The initial conditions of $N(a, t)$, $\log F(a, t)$, $U(a, t)$, $Y(t)$, $V(t)$ and $U(a_m, t)$ are given in Table 1, and the initial catch is calculated from these values. The boundary condition $N(4, t)$ has in the lack of a recruitment model been given the same value as $N(4, 1)$.

4.2 Specification of error statistics

All statistical information is represented by an ensemble. We have used 500 ensemble members to reduce statistical fluctuations in the results although 100-150 would probably be sufficient based on experiences from previous applications. The ensemble members are generated from the formula

$$\psi_j = \overline{\psi}(1 + v_j\sqrt{\sigma^2}), \tag{26}$$

where v_j is a random number taken from a Gaussian distribution with mean equal to zero and standard deviation equal to one.

Table 1. Initial values for the experiment with Icelandic cod stocks. The abundance N is given in million fish. The other variables have no units.

$N(4, 1) = 133.82$	$\log F(4, 1) = -1.7487$	$U(4, 1) = -1.64$
$N(5, 1) = 63.75$	$\log F(5, 1) = -1.0527$	$U(5, 1) = -0.95$
$N(6, 1) = 47.92$	$\log F(6, 1) = -0.7744$	$U(6, 1) = -0.67$
$N(7, 1) = 15.69$	$\log F(7, 1) = -0.3011$	$U(7, 1) = -0.20$
$N(8, 1) = 8.93$	$\log F(8, 1) = -0.2169$	$U(8, 1) = -0.11$
$N(9, 1) = 3.35$	$\log F(9, 1) = -0.1508$	$U(a_m, 1) = -0.05$
$N(10,1) = 1.00$	$\log F(10, 1) = -0.1800$	$V(1)\&Y(1) = -0.1$

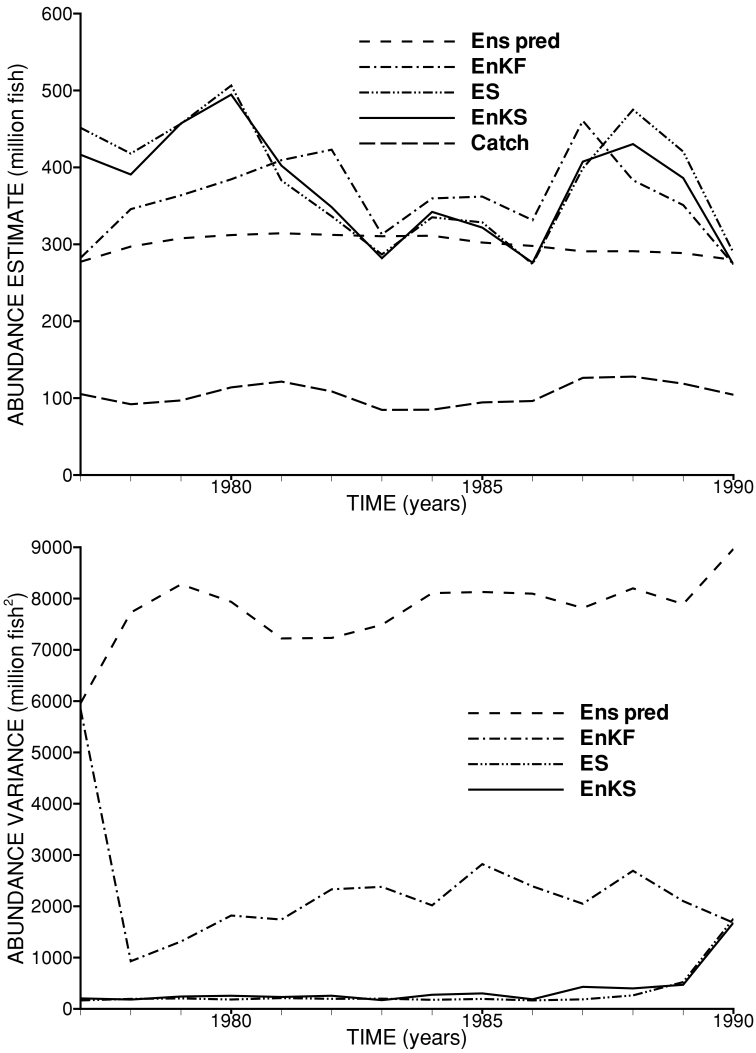

Fig. 1. The upper plot shows estimated abundance in million fish for 4-10 years old fish. The long dashed line is observed catch. The lower plot shows the corresponding variances of the abundance estimates for 4-10 years old fish.

For the initial values of N the variance is given the fairly large value, $\sigma^2 = 0.25$, to account for our lack of knowledge about the initial size of the fish stock. The initial values of *U*, *V* , *Y* and log *F* have been given a variance, $\sigma^2 = 0.025$.

The dynamical variance is incorporated into the model in a similar way and is set to σ^2 = 0:01 (equivalent to 10 % standard deviation in a model prediction from one year to the next). It applies to the entire vector of state every time this is integrated forward in time. This variance could have been set higher for the years when immigrational effects are important since the model solution is less reliable in these years.

The ensemble of measurements is generated in the same way and the measurement variance is set to $\sigma^2 = 0.01$. Thus the measurements and a one-year model forecast from a perfect initial condition could be expected to have errors of the same order. However, since the initial conditions and the boundary conditions are poorly known, and at the same time rather dominant, the analyzed solution is expected to be closer to the observations than to the model forecast.

4.3 Discussion of results

The following discussion considers the results from an experiment where the three assimilation methods, i.e., the EnKF, the ES and the EnKS are used with identical setups. Further, an additional pure ensemble integration (Ens pred) with no assimilation of data is performed. This allows us to examine the impact of assimilation of

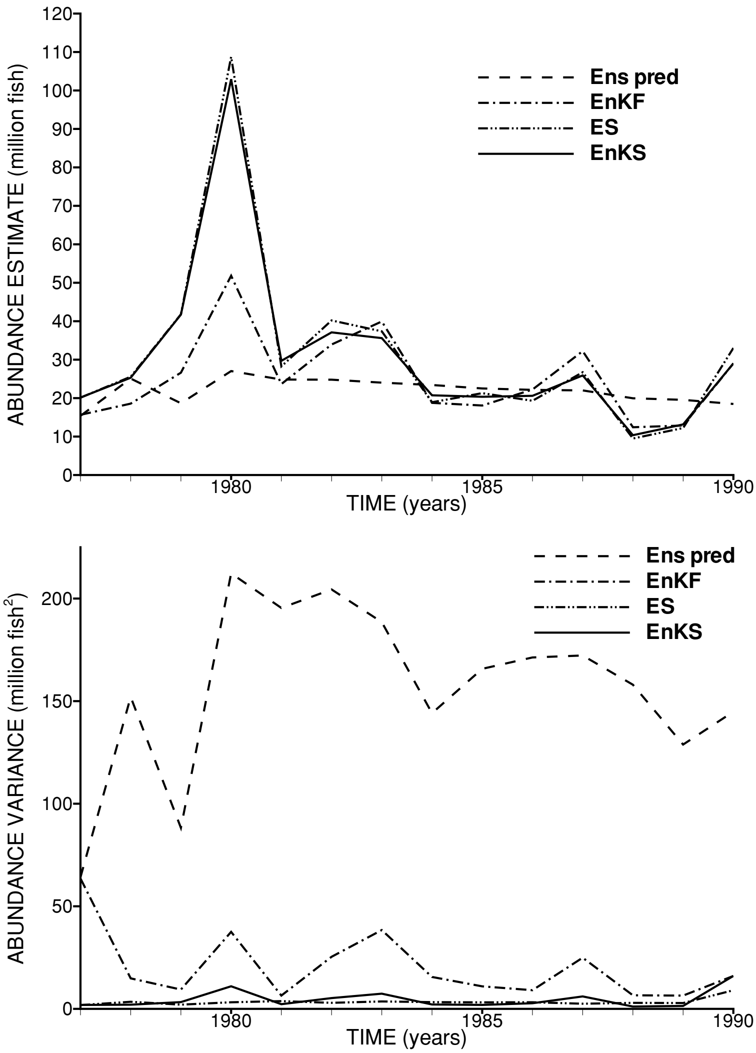

Fig. 2. The upper plot shows estimated abundance in million fish for 7 years old fish. The lower plot shows the corresponding variances of the abundance estimates for 7 years old fish.

data both for the estimated solution and the error variances.

The upper plot in Fig. 1 shows the sum of the abundance of 4 to 10 years old cod, i.e., the total stock as described by the model, for the different assimilation methods and the pure ensemble prediction. In addition the observed catch data are shown. The lower plot shows the corresponding error variances. The ensemble prediction results in a fairly steady total stock of around 300 million fish. The variance of the ensemble prediction indicates a large uncertainty corresponding to a standard deviation of about 90 million fish, i.e. almost one third of the total stock.

The impact of the catch data is obvious when examining the assimilation estimates. All the methods predict an increased size of the stock around 1986 to 1989. The smoothers also update the fish stock at the initial time by bringing information from data backward in time. The error variances are also reduced dramatically by the introduction of catch data. The EnKF predicts an error variance indicating a standard deviation of around 44 million fish, while the smoothers further reduce the standard deviation to around 14 million fish.

As the theory predicts, the EnKF and the EnKS ends up with the same estimate and error variance in the final year, and the EnKS gives an improved solution for all previous years. The EnKS and the ES solutions are fairly close, indicating that the model is only weakly nonlinear, resulting in a nearly Gaussian density for the model prediction. Trends in the estimates are otherwise consistent

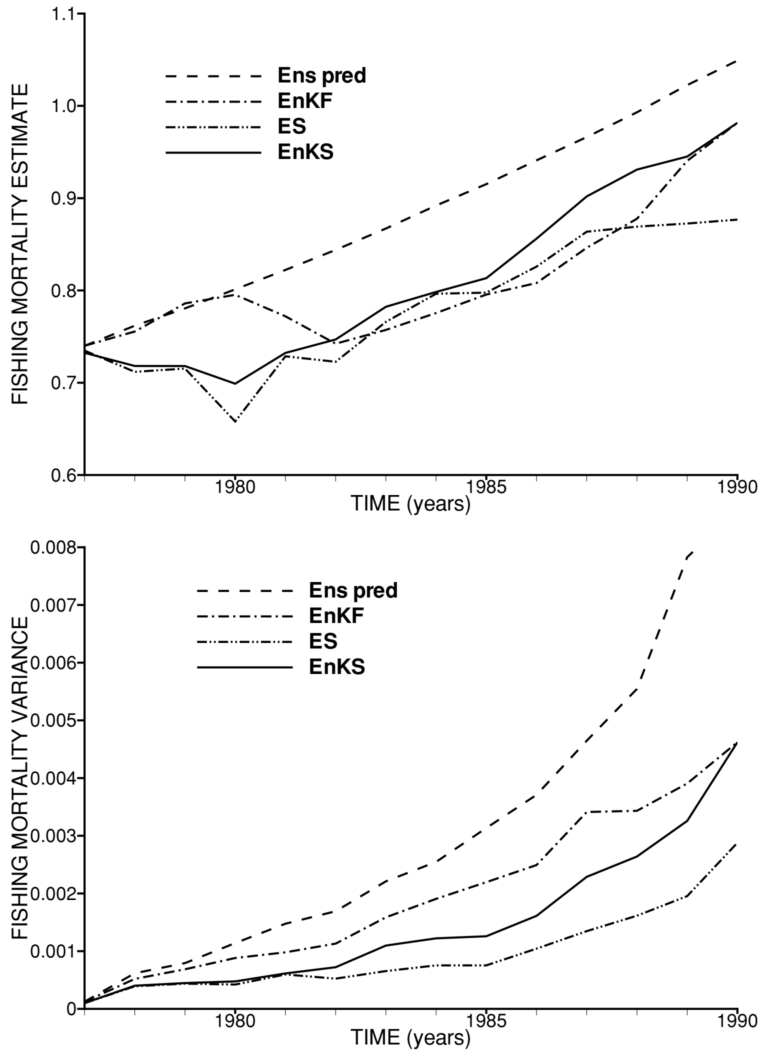

Fig. 3. The upper plot shows estimated fishing mortality for 7 years old fish. The lower plot shows the corresponding variances of the estimated fishing mortalities for 7 years old fish.

with those found by Gudmundsson (1987, 1994).

The upper plot in Fig. 2 shows the estimated abundance of seven years old cod from the different simulations and the lower plot shows the corresponding error variances. The year 1980 is particularly interesting. Both smoother solutions results in very high and probably incorrect values for this year, which are due to an immigration of 8 years old cod observed to occur in 1981. Since immigrational effects are neglected in the model formulation, the assimilation system interprets this immigration as fish that have survived from 1980, and therefore decreases the fishing mortality rate for this year, as seen in Fig. 3, in addition to increasing the abundance. Thus, by analyzing the filter and the smoother solutions in conjunction it may be possible to detect immigration and/or emigration, and the effects of these may be added to the model for a second simulation as it was done by Gudmundsson (1994).

From the filter solution, which does not take later values into account, we see that 1980 would have been a good year for 7 years old cod, anyway. Thus, only a part of the peak in the smoother should be seen as an error due to immigration. Gudmundsson (1994) took explicitly into account immigrational effects and for the year 1980 he got a result which lies between the EnKF solution and the two smoother solutions.

Fig. 3 shows the estimated fishing mortality rates and the corresponding variances in the upper and lower plot, respectively. The fishing mortality rates are not allowed to vary to the same extent as those obtained by Gud-

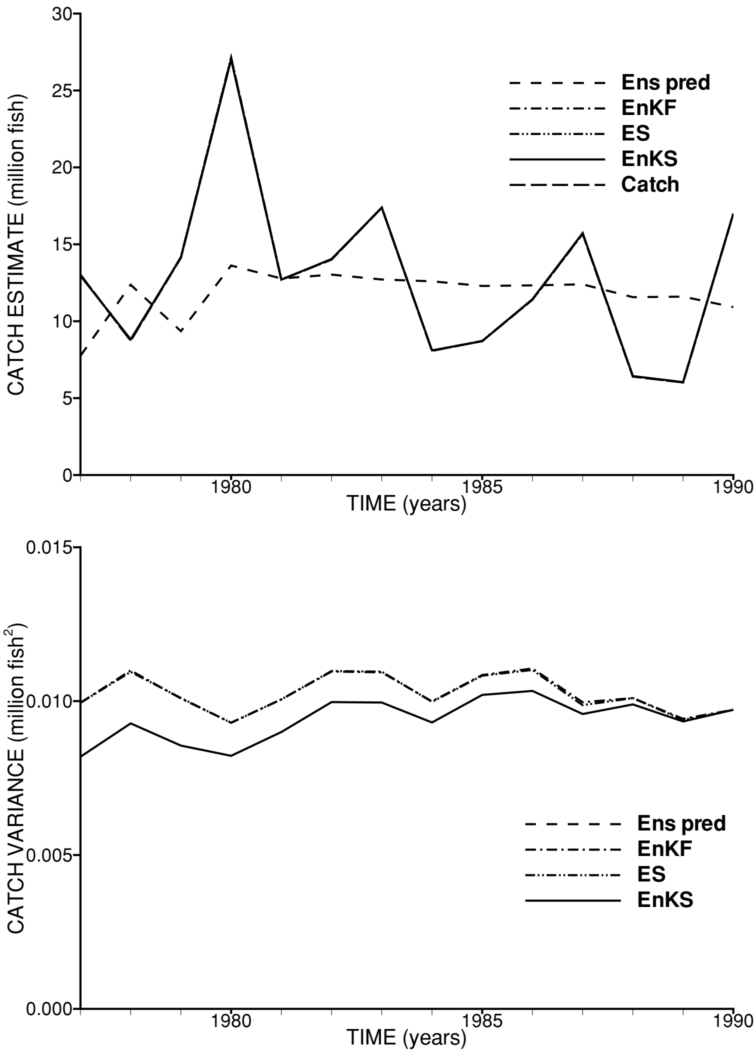

Fig. 4. Upper plot shows estimated catch in million fish for 7 years old fish. Note that the estimated catches for the filter and the smoothers are very similar and the curves overlap in the plot. The lower plot shows the corresponding variances of the estimated catch for 7 years old fish. The variance of the pure ensemble prediction where no data are assimilated is large compared to the variances of the filter and the smoothers (typically between 20 and 50 and therefore not shown), indicating a positive impact of the data.

mundsson (1994), although they are otherwise similar. A constant trend term was included in the model and this could perhaps been chosen smaller or omitted altogether, since both the filter and smoother estimated mortalities are lower than the pure ensemble prediction. From the catch plot in Fig. 4 it is seen that the model relies more on the observed data than on the model prediction despite the fact that the measurement variance is equal to the model variance. This can be explained by the large initial variance which gives low weight on the model solution initially. The EnKF, the ES and the EnKS solutions are very similar and about perfectly fitted to the measurements and they all provide estimates with almost equal variance.

5 SUMMARY

This paper has introduced some new ensemble-based assimilation techniques for use in fish stock assessment. The purpose has been to demonstrate the feasibility of the methods in fish stock assessment and to discuss the general properties of them. Three methods, i.e. the Ensemble Kalman Filter (EnKF), the Ensemble Kalman Smoother (EnKS) and the Ensemble Smoother (ES), were implemented with the fish stock model by Gudmundsson (1994). The methods were applied in an experiment with catch at age data of Icelandic cod, in a similar setup to the one used by Gudmundsson (1994).

The model used here is probably too simple for operational fish stock assessment, but has been useful for demonstrating the methodologies. Further, the ensemble techniques can be characterized as model independent, i.e., the data assimilation framework can be used without modifications with any model just by replacing the model integration subroutine and the definition of the model state. Thus, more complex models can be introduced in the system whenever needed.

In the present application, it turned out that the model used is only weakly nonlinear. The strength of the ensemble methods is that they handle strongly nonlinear models just as well. This is in contrast to traditional assimilation methods which are based on linearizations to develop tangent linear operators and/or adjoints. Stronger nonlinearities in the model would lead to greater non-Gaussian contributions for the predicted statistics, and in this case we would expect that the EnKS would outperform the ES as was seen in the application with the Lorenz model in Evensen (1997a).

In summary, we now have a set of filters and smoothers which are designed to work well with strongly nonlinear and high dimensional state space models. Further, the methodologies have successfully been applied with a simple fish stock assessment model. A next step will be to adapt or build a more sophisticated model which can be introduced into the assimilation system and a more extensive validation exercise should be preformed using additional data sets. An ongoing activity involves the introduction of climate variables or information. The underlying hypothesis is that inclusion of appropriate climate time series will reduce the variance attached to the estimated fishing mortality, which in turn will give improved abundance estimates. Such time series could be wind and hydrography for the Faeroe Island cod stock or temperature for the North East Atlantic cod stock.

ACKNOWLEDGEMENTS

The work has been supported by a grant (127432/120) from the Research Council of Norway, and by a grant of computer time from the Norwegian Supercomputing committee (TRU).

REFERENCES

Anderson JL, Anderson SL. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review* 127:2741-2758.

Burgers G, van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review* 126:1719-1724.

Evensen G. 1992. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research* 97:17905-17924.

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

Evensen G. 1997a. Advanced data assimilation for strongly nonlinear dynamics. *Monthly Weather Review* 125:1342-1354.

Evensen G. 1997b. Application of ensemble integrations for predictability studies and data assimilation, in Monte Carlo Simulations in Oceanography. Proceedings 'Aha Huliko'a Hawaiian Winter Workshop, University of Hawaii at Manoa, January 14-17, 1997, edited by P. Muller and D. Henderson, p 11-22, SOEST.

Evensen G, van Leeuwen PJ. 1996. Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasi-geostrophic model. *Monthly Weather Review* 124:85-96.

Evensen G, van Leeuwen PJ. 2000. An Ensemble Kalman Smoother for nonlinear dynamics. *Monthly Weather Review* 128:1852-1867.

Gudmundsson G. 1987. Time series of fishing mortality rates. *ICES Council meeting* 1998/D:6.

Gudmundsson G. 1994. Time series of catch-at-age observations. *Applied Statistics* 1:117-126.

Houtekamer PL, Mitchell HL. 1998. Data assimilation using an Ensemble Kalman Filter technique. *Monthly Weather Review* 126:796-811.

Jazwinski AH. 1970. Stochastic Processes and Filtering Theory. San Diego, CA: Academic Press. 376 p.

Miller RN, Ghil M, Gauthiez F. 1994. Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of Atmospheric Science* 51:1037-1056.

Miller RN, Carter EF, Blue ST. 1999. Data assimilation into nonlinear stochastic models. *Tellus* Ser. A 51:167-194.

Todling R, Sivakumaran NS. 1998. Suboptimal schemes for retrospective data assimilation based on the fixed-lag Kalman smoother. *Monthly Weather Review* 126:2274-2286.

van Leeuwen PJ, Evensen G. 1996. Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review* 124:2898-2913.

Accepted 20 April 2001 – Printed 28 December 2001
Editorial responsibility: Jarl Giske