Nansen Environmental and Remote Sensing Center

Technical Report No. 220

Assimilation of ice concentration into a coupled ice ocean model using the Ensemble Kalman Filter

by

Knut A. Lisæter, Julia Rosanova and Geir Evensen

February 2003

Nansen Environmental and Remote Sensing Center (NERSC)
Edvard Griegsvei 3a
N-5059 Bergen
Norway

phone +47 55297288
fax +47 55200050
email adm@nersc.no

TITLE Assimilation of ice concentration into a coupled ice ocean model using the Ensemble Kalman Filter	**REPORT IDENTIFICATION** NERSC Technical report no. 220
CLIENT European Space Agency (ESA)	**CONTRACT NO.** no. 14992/01/NL/MM
CLIENT REFERENCE M. Drinkwater(APP-FSO) technical matter M- Mezzadri (IMT-CAO) contractual matter	**AVAILABILITY** OPEN
INVESTIGATORS Knut Arild Lisæter Julia Rosanova Geir Evensen email: Knut.Lisaeter@nersc.no	**AUTHORIZATION** Bergen, February 11, 2003 Prof. Ola M. Johannessen Director NERSC

Contents

Assimilation of ice concentration into a coupled ice ocean model using the Ensemble Kalman Filter

Knut A. Lisæter, Julia Rosanova and Geir Evensen
Nansen Environmental and Remote Sensing Center

Abstract

An implementation of the Ensemble Kalman Filter (EnKF) with a coupled ice–ocean model is presented. The model system consists of a dynamic–thermodynamic ice model using the Elastic–Viscous–Plastic (EVP) rheology coupled with the HYbrid Coordinate Ocean Model (HYCOM). The observed variable is ice concentration from passive microwave sensor data (SSM/I).

The assimilation of ice concentration has the desired effect of reducing the difference between observations and model. Comparison of the assimilation experiment with a free–run experiment, shows that there are large differences, especially in summer. In winter the differences are relatively small, partly because the atmospheric forcing used to run the model depends upon SSM/I data. The assimilation has the strongest impact close to the ice edge, where it ensures a correct location of the ice edge throughout the simulation.

An inspection of the model ensemble statistics reveals that the error estimates of the model are too small in winter, a result of too low model ice concentration variance in the central ice pack. It is found that the ensemble covariance between ice concentration and sea surface temperature in the same grid cell is of the same sign (negative) throughout the year. The ensemble covariance between ice concentration and salinity is more dependent upon the physical mechanisms involved, with ice transport and freeze/melt giving different signs of the covariances. The ice transport and ice melt mechanisms also impact the ice concentration variance and the covariance between ice concentration and ice thickness.

The ensemble statistics show a high degree of complexity, which to some extent merits the use of computationally expensive assimilation methods, such as the Ensemble Kalman Filter. The present study focuses on the assimilation of ice concentration, but it is understood that assimilation of other data sets, such as sea surface temperature, would be beneficial.

1   Introduction

In the polar regions ice is an important component of the climate system, and it strongly modifies exchanges of heat, momentum and moisture between the ocean and the atmosphere. This effect is perhaps best seen in wintertime heat fluxes in the interior of the Arctic ice pack where the atmospheric heat fluxes over open water can be two orders of magnitude larger than over ice.

The thermohaline circulation is affected by the presence of sea ice. This is due to the insulating effect of the ice cover, and because of salt fluxes between ice and ocean when sea ice freezes or melts. The ocean stratification is also an important factor in determining the impact of sea ice on the thermohaline circulation. In the Central Arctic Ocean, the stratification is strong, due to the presence of a cold halocline layer. This layer does, to some extent, reduce the thermohaline circulation in this region. The majority of the thermohaline circulation is therefore believed to be connected to processes in the subpolar seas.

Among the subpolar seas where deep water formation takes place are the Siberian and Laptev shelf seas. Here, it is believed, sea ice formation plays an important part in creating dense waters. The brine released by freezing ice creates dense shelf water, which is believed to sink into the Arctic Ocean, entraining water masses on their way to the deep ocean (*Rudels et al.*, 1999). In the Labrador and Greenland Seas, sea ice often has the opposite effect as seen on the shelves. These seas are "production sites" for deep and intermediate water masses of the North Atlantic. Here, ice is transported from other regions, and excessive transport has the capability of producing buoyancy anomalies in these regions, such as the "Great Salinity Anomaly" (*Dickson et al.*, 1988; *Belkin et al.*, 1998). The buoyancy anomalies can reduce the deep water production, and the variability of the ice transport is therefore an important factor in the variability in deep water production. The effect of such variability on the thermohaline circulation has previously been demonstrated by *Holland et al.* (2001).

In addition to the importance of sea ice for the global climate there is also a need for shorter-term predictions of sea ice concentrations. Offshore operations and weather forecasters would be likely to profit from ice concentration estimates in polar regions. At present we have a reasonably good understanding of polar sea ice concentration through data from passive microwave sensors. The data from the Special Sensor Microwave/Imager (SSM/I) carried on board the satellites of the Defense Meteorological Satellite Program makes it possible to compute nearly daily snapshots of the sea ice concentration.

The ability to extend this knowledge, and actually forecast ice variables, is possible through the use of numerical models describing the thermal and dynamic properties of the ice. However, in addition to producing forecasts it is essential also to produce reasonable estimates of the uncertainty in the model forecast. Furthermore, to reduce uncertainties in the model forecast it is essential to have an initial model state which is as close as possible to the truth. This is where the different data assimilation techniques come into play. Furthermore, the use of models and data assimilation is not only related to forecasting, but can also be used to produce a reanalysis.

The focus of this work is to demonstrate the assimilation of sea ice concentration in a coupled sea ice ocean model using the Ensemble Kalman Filter (EnKF; *Evensen*, 1994). The EnKF is a sequential and ensemble–based data assimilation technique; an ensemble of model states is run in parallel and the analysis is computed at discrete times, using only information from that time. Model error statistics, necessary for the analysis, are calculated from the ensemble of model states. This makes the EnKF attractive due to its relatively easy implementation compared to other assimilation schemes. There is, for instance, no need to compute adjoint equations or a tangent linear operator. This point is even more important for coupled models, where different time scales can be important for the different model components. Because of its simple implementation and its capability for describing error statistics in nonlinear models, variants of the Ensemble Kalman filter have been widely used in geophysical applications (*Evensen*, 1994; *Houtekamer and Mitchell*, 1998; *Reichle et al.*, 2002; *Haugen and Evensen*, 2002; *Natvik and Evensen*, in press).

The model system used in this study is the HYbrid Coordinate Ocean Model (HYCOM; *Bleck*, 2002), together with a dynamic-thermodynamic sea–ice model. The sea ice model consists of the Elastic-Viscous-Plastic dynamic model of *Hunke and Dukowicz* (1997), and a thermodynamic module from *Drange and Simonsen* (1996). The system is forced by atmospheric variables from the European Centre for Medium–range Weather Forecasting (ECMWF).

The assimilation is performed with the EnKF, where the assimilated variable is sea ice concentration. The assimilation scheme is multivariate, so both ocean and ice model variables are updated in the analysis. The use of variables from both ocean and ice model components in the assimilation scheme is essential,

as the properties of the oceanic mixed layer are strongly dependent upon the presence of ice. An analysis update in ice concentration with no accompanying modification of surface layer properties could result in a state where newly introduced ice melts immediately.

To our knowledge, there have been no previous attempts to perform multivariate ice concentration assimilation into this type of model system with the EnKF scheme. The focus of the study is therefore largely upon the methodology and the realism of the multivariate scheme. The outline of this study is as follows. In Section 2 we will briefly describe the model system, Section 3 describes the Ensemble Kalman Filter, the procedures related to the assimilation step, and the experimental setup. The impact upon ice concentration and other variables is presented in Section 4, while model ensemble statistics are presented in Section 5. The work is summarized in Section 6.

(a)   (b)
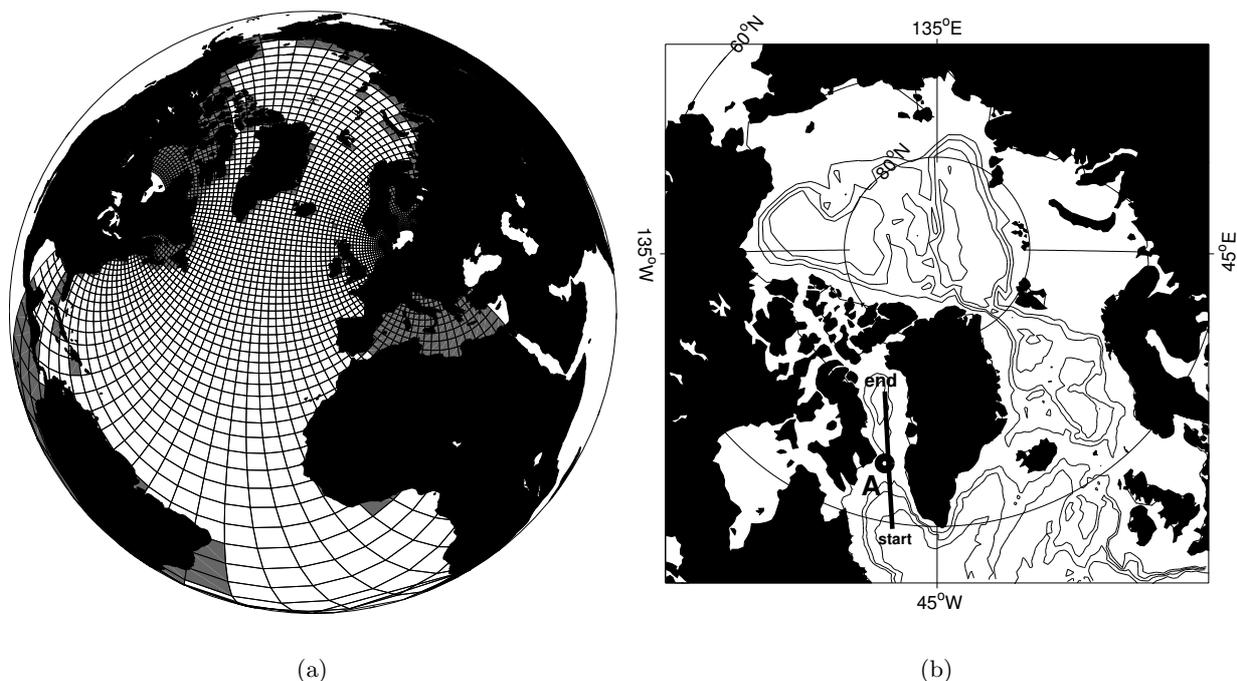
Figure 1: Figure 1(a) Illustration of the model grid. Areas masked in the model are marked as gray, while areas outside the grid are not included in the model. Note that only every second grid line is shown. Figure 1(b) Part of the model domain with the Arctic Ocean. Also shown is a section used in the text (section A), and a point on this section (circle), which we will refer to as point 'O'.

2   Model setup

The ocean model is the HYbrid Coordinate Ocean Model (HYCOM; *Bleck*, 2002), which is a further development of the Miami Coordinate Ocean Model (MICOM; *Bleck and Smith*, 1990). The MICOM model uses density as the vertical coordinate. The main advantage of isopycnic coordinates lies in their ability to maintain the properties of water masses which does not communicate directly with the surface mixed layer. In the interior of the ocean, mixing is believed to mainly occur along neutral surfaces (*Montgomery*, 1938), which for most situations are relatively close to isopycnic coordinate surfaces. One of the drawbacks of the MICOM model is its relatively simple parametrization of the oceanic mixed layer, described by a bulk mixed layer model (*Gaspar et al.*, 1990).

The major improvement in HYCOM relative to MICOM is the introduction of so-called hybrid layers, which does not necessarily need to keep a given target density. This approach does not allow density layers to become massless, and allows for high vertical resolution close to the surface of the ocean. This has in turn facilitated the introduction of more sophisticated vertical mixing schemes, such as the K-Profile Parametrization (KPP; *Large et al.*, 1994), which is used in this study.

The model grid used in this study has a resolution focus in the Nordic Seas with closed boundaries in the Bering Strait and the South Atlantic. A part of the model grid is shown in Figure 1(a). It was created with the conformal mapping tools of *Bentsen et al.* (1999), and has grid sizes ranging from 100 to 150 km in the Arctic. The vertical discretization uses 22 isopycnal layers, which in σ_0–coordinates range from 21.8 to 28.11. Note that the lightest layers in this discretization are primarily used to describe the surface mixed layer, as they are usually too light to describe interior water masses of the ocean. The lightest layers become the surface layers in the hybrid coordinate formulation.

The ice thermodynamics model used has many features in common with the "0–layer" ice thermodynamic formulation of *Semtner* (1976), which ignores the specific heat of the ice. In the limit of zero heat capacity of ice the heat conduction equation gives the vertical temperature profile in the ice as a linear function. The conductive heat flux has, as a result of this, the same absolute value at the surface and bottom of the ice slab. The thermodynamic model also includes a snow layer, and a linear temperature profile is prescribed through the snow as well. A detail which will be commented later on is the thermodynamical formulation

of heat exchange between water and ice. The present model uses a simplified formulation of heat exchange using so-called "infinite diffusivity" (*Holland and Jenkins*, 1999). In this formulation, any heat available for sea ice melt in the upper ocean layer is immediately used to melt ice. The available heat is determined by the upper ocean layer temperature deviation from the freezing point of the ocean. One consequence of this is that no ice will be present in a model grid cell when the sea surface temperature is above the freezing point of the ocean. For full details of the ice thermodynamic model we refer to *Drange and Simonsen* (1996).

The ice dynamics model is the Elastic–Viscous–Plastic (EVP) ice rheology of *Hunke and Dukowicz* (1997). The EVP model presents an alternative to solving the traditional Viscous–Plastic model (VP; *Hibler*, 1979) by introducing an elastic component to the rheology equations. The elastic waves dampen out when solving the dynamical equations, and the resulting solution approach the one obtained by the VP model. The benefits of the EVP rheology is that it allows for an explicit parallel solution of the VP equations. The numerical implementation of the EVP model also shows better response to rapid changes in forcing of the sea ice component, relative to numerical implementations of the VP model (*Hunke and Dukowicz*, 1997). The ice dynamic and thermodynamic models have been solved for the same model grid as the ocean model, shown in Figure 1(a).

The synoptic forcing used were temperature, winds and humidity determined from dew point temperatures, fields which all were acquired from the European Center for Medium–range Weather Forecasting (ECMWF). Clouds are based on climatologies from the Comprehensive Ocean and Atmosphere Data set (COADS; *Slutz et al.*, 1985), while precipitation is based on the climatology of *Legates and Willmott* (1990). River input is modeled as a negative salinity flux, and the river sources in the Arctic include the Lena, Ob, Kotuy, Dvina, Yenisei and the Mackenzie rivers (*Dümenil et al.*, 1993; *Aagaard and Carmack*, 1989). At the surface the ocean model uses temperature and salinity relaxation towards Levitus Climatologies (*Levitus et al.*, 1994; *Levitus and Boyer*, 1994), with a common relaxation time scale of 50 days.

The focus of this study is upon the effects of the Ensemble Kalman Filter Scheme in the coupled model context. The model fields were initialized from a realistic model state used in the EC MAST–III project DIADEM (MAS3-CT98-0167) and integrated for one year prior to the experiment. The generation of the full ensemble of model states used in the EnKF will be described in the next section.

3 The Ensemble Kalman Filter

To assimilate observational data into a model system, it is important to have a knowledge of the errors present in the model and observations. In Kalman Filtering the needed statistics are the error covariance matrices of the observations and the model state. Describing the error covariances correctly is therefore a crucial component to the performance of sequential data assimilation schemes.

The Ensemble Kalman Filter (EnKF; *Evensen*, 1994) uses an ensemble of model states to estimate the model error statistics. It was originally proposed as an alternative to the Extended Kalman Filter (EKF) for solving strongly nonlinear problems. The EKF uses linearized equations to separately solve for the error covariance matrix, while the EnKF members are advanced using the original model equations. An advantage of the EnKF approach is that it retains the effect of nonlinear model behavior on the error covariance matrix. In an EnKF approach, few assumptions are made, and complex correlation patterns will evolve according to the evolution of the ensemble members. The simplified closure of the EKF has, on the other hand, been shown to be problematic when used in some non–linear models (*Evensen*, 1992). A brief explanation of the EnKF is given below, for a more thorough description of the EnKF see *Evensen* (1994); *Burgers et al.* (1998).

Let $\boldsymbol{\psi}_i^f \in \mathbb{R}^{n \times 1}$ be the n-dimensional model forecast of ensemble member number $i \in \{1, 2, ..., N\}$. This forecast evolves in time from a best estimate at time t_k, using the model

$$\boldsymbol{\psi}_i^f(t_{k+1}) = \boldsymbol{g}\left[\boldsymbol{\psi}_i^a(t_k)\right] + \boldsymbol{\beta}_i(t_k). \tag{1}$$

Here $\boldsymbol{\beta}_i \in \mathbb{R}^{n \times 1}$ is a random component drawn from a $N(\boldsymbol{0}, \boldsymbol{\sigma_\beta})$ distribution, and represents the effect of model errors on the evolution of the ensemble members. The covariance $\boldsymbol{\sigma}_\beta \in \mathbb{R}^{n \times n}$ will have to be specified based on knowledge about the model errors. The operator $\boldsymbol{g} : \mathbb{R}^{n \times 1} \to \mathbb{R}^{n \times 1}$ can, as already mentioned, be a nonlinear function of the model state. The individual ensemble members in equation (1) form a Monte Carlo approach for solving the Fokker–Planck equation (*Jazwinski*, 1970), which describe the evolution of the probability density function for a state $\boldsymbol{\psi}$. In order to infer the error evolution of the model state, knowledge is needed of the "truth". As the true model state is unknown, some assumptions must be made regarding its properties. In the EnKF, the best estimate of the truth is represented by the ensemble mean state. It follows that the model state error covariance used in the EnKF is that given by the ensemble covariance. Then, at any time, an estimate of the model state error covariance matrix can be computed from the ensemble of model states as

$$\boldsymbol{P}^f \approx \boldsymbol{P}_e^f = \overline{\left(\boldsymbol{\psi}_i^f - \overline{\boldsymbol{\psi}^f}\right)\left(\boldsymbol{\psi}_i^f - \overline{\boldsymbol{\psi}^f}\right)^T} \tag{2}$$

where $\overline{\boldsymbol{\psi}}$ is the ensemble estimated mean state, and the overbar denotes the expected value. In other words, the ensemble covariance matrix is taken to be representative of the error covariance matrix.

At the time observations are available an analysis is computed. The observations $\boldsymbol{d} \in \mathbb{R}^{m \times 1}$ have an associated uncertainty $\boldsymbol{\epsilon}$, and an observation error covariance matrix $\boldsymbol{R} = \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T$, where the observation error covariance matrix must be based on prior knowledge of the observation errors. Let $\boldsymbol{H} \in \mathbb{R}^{m \times n}$ be a linear operator that transforms the model state to the observation space. Then the analysis update is given by the following variant of the traditional Kalman filter equation (*Jazwinski*, 1970);

$$\boldsymbol{\psi}_i^a = \boldsymbol{\psi}_i^f + \boldsymbol{P}_e^f \boldsymbol{H}^T \left(\boldsymbol{H}\boldsymbol{P}_e^f \boldsymbol{H}^T + \boldsymbol{R}\right)^{-1}\left(\boldsymbol{d}_i - \boldsymbol{H}\boldsymbol{\psi}_i^f\right) = \boldsymbol{\psi}_i^f + \boldsymbol{K}_e\left(\boldsymbol{d}_i - \boldsymbol{H}\boldsymbol{\psi}_i^f\right). \tag{3}$$

$\boldsymbol{K}_e$ is called the Kalman Gain, and is given as

$$\boldsymbol{K_e} = \boldsymbol{P}_e^f \boldsymbol{H}^T \left(\boldsymbol{H}\boldsymbol{P}_e^f \boldsymbol{H}^T + \boldsymbol{R}\right)^{-1}. \tag{4}$$

Equation (3) tells how each ensemble member is updated using the model error covariance matrix estimate (2), and the error covariance matrix of the observations. Special notice should be taken of the observation vector $\boldsymbol{d}_i$ used in equation (3); as indicated by its subscript it is different for each ensemble member. This is because the observations need to be perturbed to get an analysis error covariance matrix consistent with the original Kalman Filter. As shown by *Burgers et al.* (1998) the analysis with $\boldsymbol{d}_i$ taken from a $N(\boldsymbol{d}, \boldsymbol{R})$ distribution gives the following analysis covariance matrix;

$$\boldsymbol{P}_e^a = (\boldsymbol{I} - \boldsymbol{K_e}\boldsymbol{H})\boldsymbol{P}_e^f, \tag{5}$$

Table 1: The table shows some of the parameters used in the assimilation experiment.

Parameter	Explanation	value
N	Number of ensemble members	100
r_0	Observation radius of influence	100 km
σ_c	Obs. error standard deviation	0.05
r_c	Obs. error decorrelation length scale	10 km
r_β	Decorr. length for random forcing	~ 1000 km
τ_β	Atm. stress standard deviation	0.015 N/m^2
T_β	Atm. temp. standard deviation	3.0 K

which is consistent with the covariance of the analysis in the traditional Kalman Filter. Without perturbation of the original observations $\boldsymbol{d}$, the analysed covariance would be systematically underestimated, an effect which could lead to so–called filter divergence. Filter divergence means that the error estimate of the model in the EnKF is too small, and the analysis will have little impact on the model.

Recently there has been a debate on the drawbacks of using perturbed observations in the Ensemble Kalman Filter, connected to noise introduced by the perturbations. This has led to the development of variants of the EnKF which do not need perturbed observations, e.g. *Anderson* (2001); *Whitaker and Hamill* (2002). For our application we note that the effects of random perturbations should only be significant for very small ensemble sizes, and should probably be of lesser importance for our 100 member ensemble.

3.1 Experiment Setup

The state vector $\boldsymbol{\psi}$ was chosen to include both ocean variables and ice variables as these are closely related, particularly for the ocean surface. For each of the 22 layers of the ocean model we included salinity, temperature, layer thickness and velocity. In addition the barotropic velocity and barotropic pressure components were included along with ice concentration and ice thickness from the ice model.

This gives a total of 115 variables for each grid cell. The analysis (3) can be calculated for the full model state vector $\boldsymbol{\psi}$. For this application the dimension for the state vector $\boldsymbol{\psi}$ is

$$ n = \dim \boldsymbol{\psi} = 140 \times 130 \times (22 \times 5 + 5) \approx 2 \times 10^6 \,, \tag{6} $$

which can result in problems if we only have 100 ensemble members forming a basis for this vector space (the analysis update is in reality a linear combination of the ensemble members $\boldsymbol{\psi}_i$). A common practice in data assimilation for high–dimensional systems is therefore to look at the problem locally, meaning that each grid cell value is updated using observation values in a radius of influence r_0 around the grid cell. In this way the 100 ensemble members will better describe the vector space (now with dimension 115). Thus the local analysis should make the problem better behaved and was chosen for this study. Note that the inclusion of variables other than ice concentration will also force an update of these variables because they can be negatively or positively correlated with ice concentration through the ensemble covariance matrix.

During the integration of the ensemble members there is a need to incorporate the effect model errors have on the evolution of the ensemble members. How to best do this is a study in its own right. One approach could be to add random fields to the ocean layer interfaces and ice thickness in the model at the analysis time. The approach used here is to add pseudo–random fields, with a prescribed length and time scale, to the ECMWF forcing fields. This procedure is a crude way of incorporating the effect of model errors $d\boldsymbol{\beta}_i$ in equation (1) into the ensemble, and is the same as assuming that the dominant errors are in the forcing data. Furthermore, this approach does not disturb model dynamics in an unphysical manner. See the appendix in *Evensen* (1994) for a description of the pseudo–random fields added to the forcing, and Table 1 for a description of the random forcing component parameters.

To calculate sea ice concentration, brightness temperatures from the Special Sensor Microwave/Imager (SSM/I) are used. The brightness temperatures are supplied by the National Snow and Ice Data Center (*http://nsidc.org*) on a 304×448 grid using a polar stereographic projection. A variant of the NORSEX algorithm (*Svendsen et al.*, 1983) is used to calculate the ice concentration from brightness temperatures, where channels at 19.35, 22.24, 37.00 and 85.80 GHz are used. The ice concentration was calculated on the same grid as the original brightness temperature grid from the NSIDC, and then averaged over 5×5 grid cells to give an observation vector $\boldsymbol{d}$ of somewhat smaller size and resolution characteristics closer to that of

model grid cells in the Arctic. The observation error variance for the elements of $\boldsymbol{d}$ was set to 0.0025 giving an error standard deviation of 0.05 for the ice concentration. This error variance was fixed throughout the year, although the real error is likely to increase in summer due to the presence of melt ponds on the sea ice. *Emery et al.* (1994) give a value of $\sim 5\%$ for the winter time ice concentration error and bias when using the TEAM and bootstrap algorithms. The error for the NORSEX algorithm is probably of the same magnitude and larger in summer. The observation error decorrelation length scale r_c was set to 10 km for the experiments. Compared to the spacing between the observations this is relatively low, and the result will be an error covariance matrix which is approximately diagonal.

The initial ensemble was generated from a realistic model state, to which we added random fields. For each ensemble member, new ice concentration and ocean model layer thickness fields were generated by adding smooth pseudo–random fields (*Evensen*, 1994) to the original model state. All other variables of the individual ensemble members were identical to the original model state. The random fields added to the ocean model layers were vertically correlated and had a horizontal decorrelation length of ~ 400 km. The standard deviation of the random fields were set to 10% of the original ocean layer thickness, and the ocean layers were finally adjusted so that no barotropic waves were generated by this procedure. In a similar manner, the random fields added to the ice concentration had the same decorrelation length, and the standard deviation were set to 10% of the original ice concentration. The ensemble of model states generated this way has a distribution which through equation (2) describes our confidence in the initial ensemble average.

The ensemble of model states was integrated for a month before starting the experiment. The experiment covers the period from Julian day 289 in 1998 up to January 2000. Parallel to the ensemble run a free-run model was also integrated for comparison. This model run used the standard ECMWF forcing fields with no pseudo–random fields added, and has an initial state equal to the initial ensemble mean. Some of the relevant parameters for the assimilation experiment are given in Table 1.

(a)   (b)
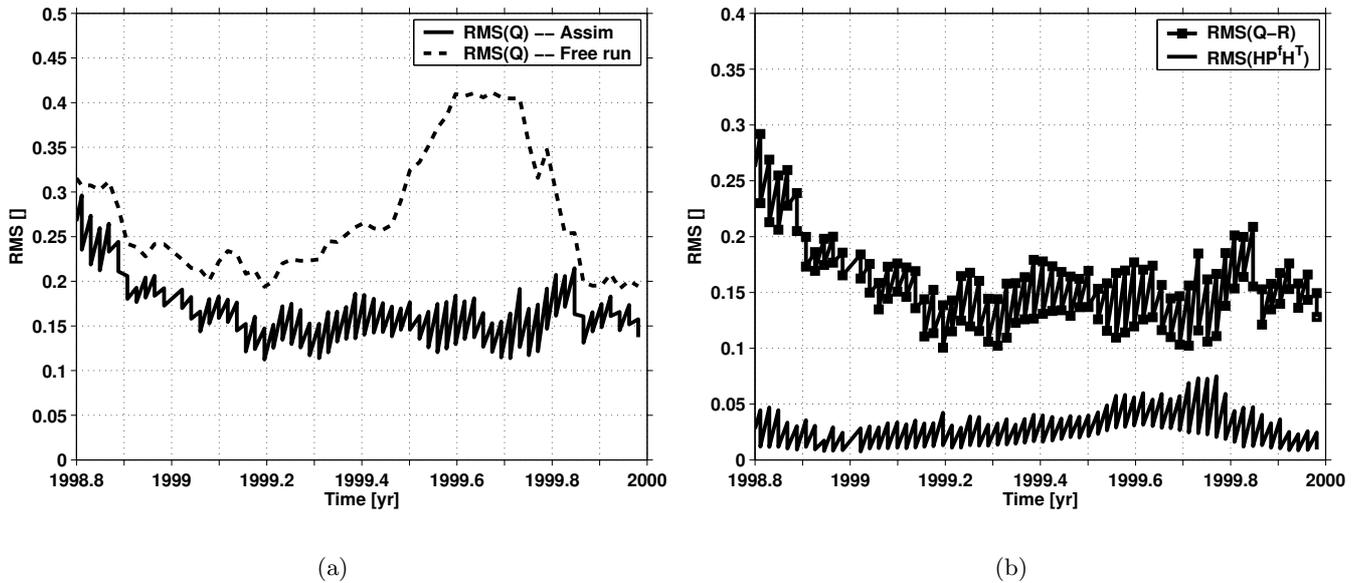
Figure 2: 2(a) The solid line shows the evolution of the innovation RMS for the assimilation experiment, while the dashed line shows the innovation RMS for the free-run model. 2(b) The solid line with square marks shows the RMS of the innovation covariance matrix minus the observation error covariance matrix. The solid line shows the RMS of the error covariance matrix of the model. Note that all RMS estimates are taken only over observation points where either the model or the observations have an ice concentration larger than 0.05

4 Impact of assimilation

The assimilation of ice concentration has an effect upon all the variables of the coupled model. In this section we examine the analysis updates and the cumulative effect of assimilation. The focus is put on the ice model variables, and the surface layer of the ocean model.

4.1 Innovation vector

Given that the object of the assimilation is to introduce observations into the model state, we should see the impact of the data assimilation in the distance between observations and model state. The success of the assimilation can to some extent be examined from this distance. The vector difference between observations and model state in the observation space is frequently referred to as the innovation vector. The innovation vector of member k is given as

$$\boldsymbol{\lambda}_k = \boldsymbol{d}_k - \boldsymbol{H}\boldsymbol{\psi}_k \,. \tag{7}$$

The second–order moment of this becomes

$$\boldsymbol{Q} = \overline{(\boldsymbol{d} - \boldsymbol{H}\boldsymbol{\psi})\,(\boldsymbol{d} - \boldsymbol{H}\boldsymbol{\psi})^T}\,, \tag{8}$$

and a measure of the distance between observations and model is here given as

$$\mathrm{RMS}(\boldsymbol{Q}) = \sqrt{\frac{1}{m}\mathrm{trace}(\boldsymbol{Q})}\,\,, \tag{9}$$

which corresponds to the square root of the mean innovation variance, where the variance is taken in the measurement points. The evolution of the innovation RMS is shown as the solid line in Figure 2(a). To make it easier to interpret the results, the average innovation is taken only over observation points where at least one of the model members, or the original observation has an ice concentration larger than 0.05.

Noticeable in Figure 2(a) are the discontinuities which occur at the times when the analysis is performed, indicative of how the innovation RMS is reduced when the model ice concentration is adjusted by the

analysis. Also to be noted is the strong RMS growth which occurs after the analysis is performed. In many cases the innovation growth between two analyses exceeds the RMS reduction from the first analysis. Further into the time series however, there is a clear cumulative effect of the assimilation. At the beginning of the experiment the innovation RMS exceeded 0.30, but a while into the experiment it is reduced to approximately 0.15, and stays close to this value for the remainder of the experiment.

A free–run experiment was run for the same time period as the assimilation experiment. The evolution of the innovation RMS for the free–run experiment is shown as the dashed line in Figure 2(a). Comparing the assimilation experiment with the free–run experiment, we see that the RMS–values start out the same. They quickly diverge, however, and the innovation RMS of the free-run experiment increases relatively to the assimilation experiment. Note that both the free-run and the assimilation experiment show reduced innovation RMS in the winter of 1999, relative to their initial values.

In winter, the large atmospheric temperature gradient across the ice edge has a strong impact upon the modeled ice concentration fields, with and without data assimilation. This ice edge is prescribed in the ECMWF forcing fields, partly by use of SSM/I data, and as a consequence, the atmospheric forcing alone will serve to pull the model towards the observations. This explains some of the RMS reduction in winter, for both experiments.

Around time point 1999.2, the RMS of the free–run experiment starts to move further away from the observed values, and as summer approaches, we start to see large differences between the experiments. The atmospheric temperature gradient across the ice edge is smaller in summer, meaning that the ECMWF ice edge is not so effective in determining the modeled ice edge. This in turn means that the assimilation will have a larger impact in summer, relative to the free–run experiment. At time point 1999.6 the difference between the model and free-run reaches a maximum, as the free run has an innovation RMS of ~ 0.4 while the assimilation experiment has a innovation RMS of ~ 0.15. Towards the end of the time series the RMS–values of the free–run innovation are rapidly reduced, as the autumn freeze–up starts in the Arctic. We note that both the free–run and the assimilation experiment have lower innovation RMS–values at the start of 2000, than at the start of 1999, with the assimilation experiment being slightly closer to the observations.

The innovations also allow us to make rough estimates of the quality of the estimated errors. Consider the model estimate to be given as a true value plus an error $\boldsymbol{q}'$:

$$\boldsymbol{\psi}^f = \boldsymbol{\psi}^t + \boldsymbol{q}' \,. \tag{10}$$

Likewise, let the observation be given as the sum of the true estimate and an error $\boldsymbol{\epsilon}'$

$$\boldsymbol{d} = \boldsymbol{\psi}^t + \boldsymbol{\epsilon}' \,. \tag{11}$$

If we assume that the observation and model errors are uncorrelated, we have for the second order moment of the innovation vector

$$\boldsymbol{Q}' = \overline{\left(\boldsymbol{\epsilon}' - \boldsymbol{H}\boldsymbol{q}'\right)\left(\boldsymbol{\epsilon}' - \boldsymbol{H}\boldsymbol{q}'\right)^T} = \boldsymbol{R}' + \boldsymbol{H}\left(\boldsymbol{P}^f\right)'\boldsymbol{H}^T \,, \tag{12}$$

where the primes are used to distinct the theoretical value for the innovation covariance matrix from the real innovation covariance matrix, equation (8). By using equation (12), we can now get a simple test on the quality of the model error covariance matrix. This is done by comparing $\mathrm{RMS}\,(\boldsymbol{Q} - \boldsymbol{R})$ against $\mathrm{RMS}\,(\boldsymbol{H}\boldsymbol{P}^f\boldsymbol{H}^T)$. If the theoretical estimate given by equation (12) is correct for $\boldsymbol{Q}$ as well, then we should have

$$\mathrm{RMS}\,(\boldsymbol{Q} - \boldsymbol{R}) = \mathrm{RMS}\left(\boldsymbol{H}\boldsymbol{P}^f\boldsymbol{H}^T\right) \tag{13}$$

The RMS values of the innovation minus the observation covariance matrix, $\mathrm{RMS}(\boldsymbol{Q} - \boldsymbol{R})$, and the RMS–values of the model error covariance matrix, $\mathrm{RMS}(\boldsymbol{H}\boldsymbol{P}^f\boldsymbol{H}^T)$, are shown in Figure 2(b). Apparent in this figure is that $\mathrm{RMS}(\boldsymbol{Q} - \boldsymbol{R})$ is always larger than $\mathrm{RMS}(\boldsymbol{H}\boldsymbol{P}^f\boldsymbol{H}^T)$. This difference is most noticeable at the start of the experiment, when $\mathrm{RMS}(\boldsymbol{Q} - \boldsymbol{R})$ is about five times the size of $\mathrm{RMS}(\boldsymbol{H}\boldsymbol{P}^f\boldsymbol{H}^T)$. The values of $\mathrm{RMS}(\boldsymbol{Q} - \boldsymbol{R})$ are quickly reduced from the initial values of ~ 0.25, and seem to stabilize around a value of ~ 0.15. The RMS values of the model error covariance matrix, on the other hand, remain low throughout winter, with a slight increase in the summer of 1999. At this time, the RMS values of the innovation matrix minus the observation error matrix, are approximately three times the size of the RMS value of the model error covariance matrix. As the freeze–up starts in the fall of 1999, we notice that the RMS of the model error covariance matrix is again reduced to values similar to the winter of 1999.

(a) Ice volume
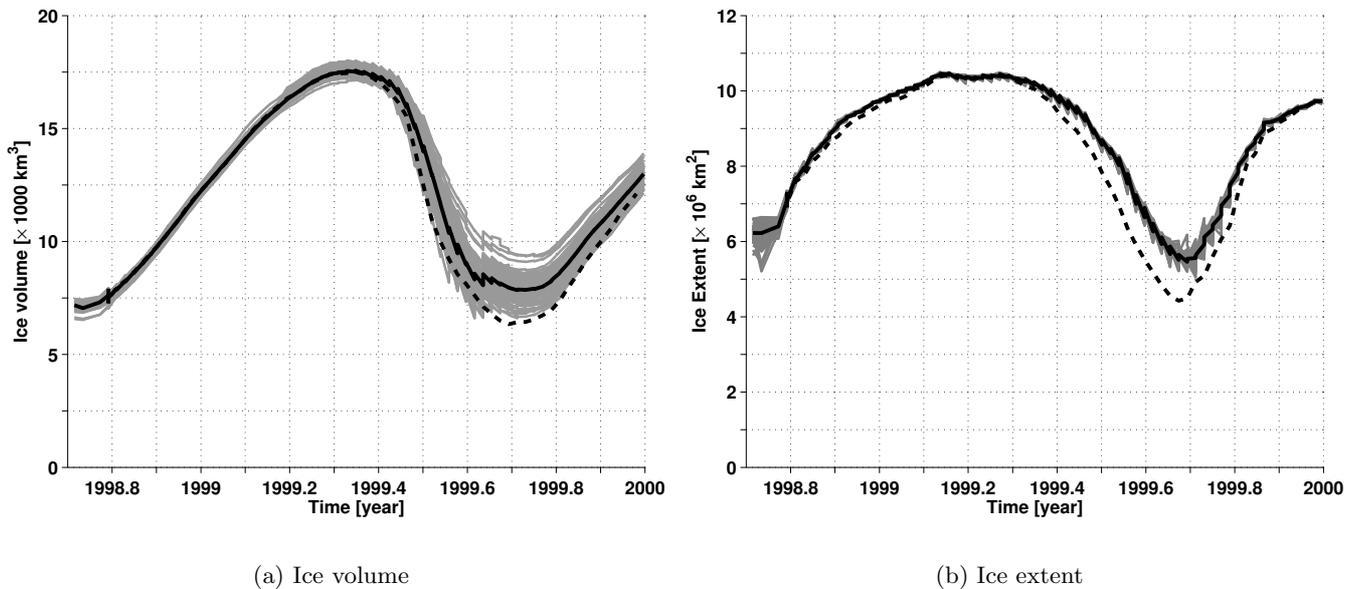
(b) Ice extent

Figure 3: The evolution of ice volume and ice extent for experiments. The gray lines are individual ensemble members, the solid black line is the ensemble mean, while the dashed line shows the evolution of free–run experiment.

All in all, the model error covariance matrix appears to be underestimated compared to the theoretical estimate (13). Note however, that this conclusion can only be drawn if we are confident that the observation errors are of the correct size. We believe the observation errors are of the correct size in winter, while they may be too low in summer. The reason for the underestimate of the model errors can in part be traced back to too low ice concentration variance in the model, especially in winter. The low ice concentration variance results in an underestimate of the model error in the EnKF.

4.2   Large scale Impact on the ice cover

Two convenient measures of the modeled ice cover properties are ice volume and ice extent. Ice volume is the total volume of ice while the ice extent here denotes the area which has an ice concentration of 15% or more. Only the ice cover of the northern hemisphere is considered. The evolution of ice extent and ice volume is shown in Figure 3 and reveals that both the model ice volume and ice extent are changed as a result of the assimilation procedure. The effect of individual analysis updates can be seen through the jumps that occur in ice volume and ice extent for the different ensemble members and for the ensemble mean. Generally, the analysis updates for both ice volume and ice extent increase their values.

The jumps in ice volume are not merely an effect of analysis changes in ice concentration. When the analysis is performed, there is an ice thickness adjustment as well as an ice concentration adjustment brought on by the multivariate assimilation scheme. In other words, the increased ice volume given by the analysis is not just an effect of increasing ice concentration and keeping the ice thickness fixed.

Although the individual updates of ice concentration have a relatively small impact on the ice extent and ice volume, there is a cumulative effect of the assimilation, which is seen by comparison with the free–run experiment. The ice volume and extent is generally higher for the ensemble mean in the assimilation experiment than it is for the free–run experiment, although this difference is small prior to the summer of 1999. The highest ice volume for both experiments is seen in March/April, and as the effect of spring and summer melt sets in, the difference between the free-run and the assimilation experiment starts to increase. The difference is seen clearly in summer, when the ice extent and ice volume are at their lowest values. At this time the ensemble mean in the assimilation experiment shows an ice volume which is $\sim 2.000 \text{km}^3$ larger than in the free–run experiment. The ensemble mean ice extent at this time is $\sim 1 \times 10^6 \text{ km}^2$ greater in the assimilation experiment.

As time progresses and the autumn freeze–up starts in the Arctic, the ice volume differences and ice extent differences between the two experiments start to decrease. The impact of the assimilation has

(a) Forecast ice thickness distribution

(b) Difference Forecast and free-run distribution
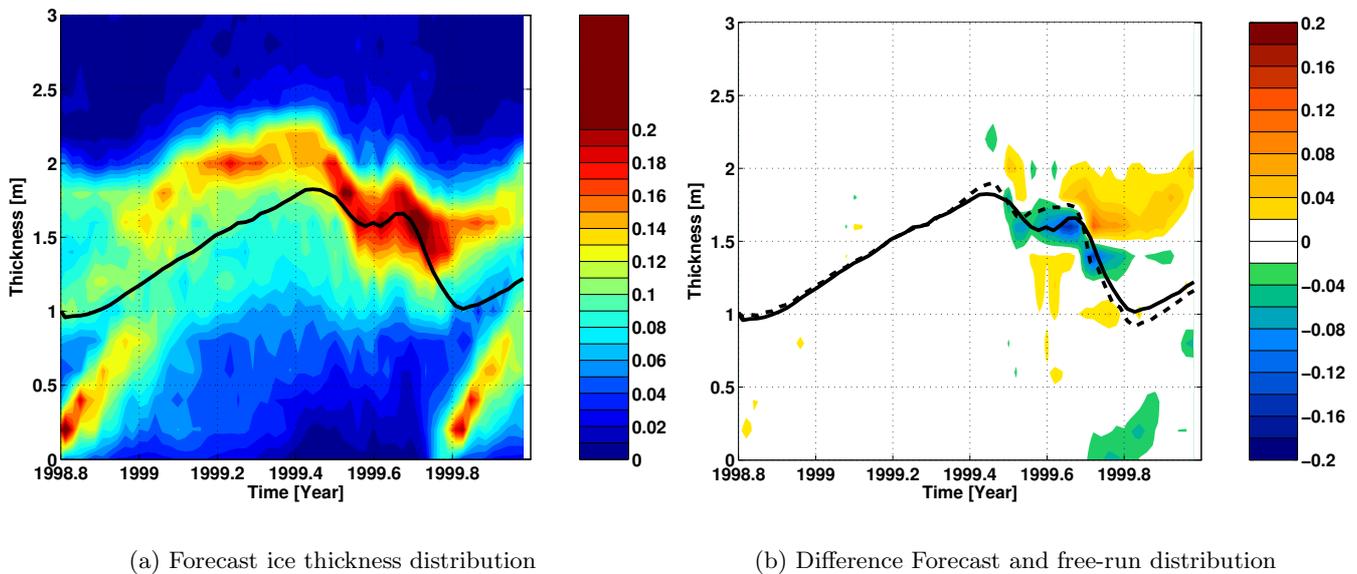
Figure 4: Contour plots of the ice thickness distribution for the entire Arctic. At each time t the Arctic ice thickness pdf can be retrieved as the values along a vertical line. Figure 4(a) shows the ice thickness distribution for the forecast, while the solid line shows the mean ice thickness. Figure 4(b) shows the difference between the forecast and the free–run ice thickness distributions. The solid line shows the forecast mean thickness, while the dashed line shows the free–run mean thickness.

therefore the greatest impact upon the summer time ice cover for the time period considered here. The present model system generally underestimates the ice volume and ice extent as compared to observations, so the assimilation of ice concentration improves these estimates.

Both the ice thickness and the ice concentration are changed by the assimilation steps, and this modifies the Arctic ice thickness distribution. In Figure 4(a) the evolution of the Arctic ice thickness distribution for the ensemble mean forecast is shown. The contour plots show the distribution of ice with different thickness for the entire Arctic, and how it evolves in time. At a particular time t the probability density function for the ice thickness is given by the values on the vertical line extending from the time axis.

The Arctic ice thickness distribution is influenced by the existence of thick, multi–year ice and of thinner seasonal ice. The distribution of thick ice is increased/reduced by thermodynamic processes and increased by ridging of thinner ice. In autumn and early winter the thickness distribution is bimodal with the modes determined by the multi–year ice ~ 2 m and newly frozen seasonal ice < 0.5 m. The freezing and ridging processes change the thin ice into thicker multi–year ice. In spring/summer, as the melt period sets in, the thinnest and seasonal ice generally melts first, so the ice thickness distribution changes to a more uni–modal distribution.

The general features of the Arctic ice thickness distribution is similar in both the free–run and assimilation experiments. The small changes that occur reveal themselves if we look at the difference between the evolution of the distributions. In Figure 4(b) we see the difference between the evolution of the Arctic ice thickness distributions where we have subtracted the free–run experiment from the assimilation experiment.

Figure 4(b) reveals that there are small differences in the thickness distributions before the spring/summer melt sets in. As the melt period sets in, however, the effect of the assimilation start to show. In Figure 4(b), the mean ice thickness in the assimilation experiment (solid line) is initially lower than the mean ice thickness in the free–run experiment (dashed line). In the free–run experiment, the thinner ice generally melts first, while in the assimilation experiment, a lot of this ice is retained through the assimilations. This causes the ice thickness distribution in the assimilation experiment to have more ice in the thickness range 0.5–1.4 m, and less ice of thickness 1.5–2.0 m relative to the free–run experiment, see Figure 4(b) around time mark 1999.6.

Later on, at the end of the summer, the effect of freeze–up in the Arctic begins to show, with the assimilation experiment having more thick ice (~ 2 m) and less thin ice (< 0.5 m) relative to the free-run experiment. This is connected to lower summer ice concentration in the free-run model. When the freezing

sets in for the free-run model, there will be more open water and consequently more thin ice will be created relative to the assimilation experiment. The assimilation experiment has retained a greater fraction of the thick multi–year ice in the central Arctic during the summer, and this shows up as increased occurrences of ice thickness around 2 m.

4.3   Selected ensemble updates

When the analysis is computed, the variables in the ice and ocean models will be adjusted according to the difference between observations and model state, and according to the error covariance statistics. In the multivariate scheme the properties of the ocean model are updated along with the analysis update in ice concentration. Due to changes in model behavior over the season, there are large differences in the analysis update at different days of the year. A controlling factor in this respect is the north–south migration of the ice edge, which has a strong impact upon the spatial distribution of the ice concentration variance. This impacts the magnitude and location of the analysis updates over the season. To illustrate this we show in Figure 5 the analysis update of ice concentration along with updates of upper layer salinity and temperature for the ocean model. The analysis updates are shown for two different times of the year, a typical situation for summer and winter. All plots illustrate the update for the ensemble mean.

For the winter time ice concentration update, Figure 5(a), we see that the biggest impact occurs close to land boundaries or along ice edges, which are marked by the forecast 0.7 ice concentration contour. At this time of year the analysis gives a reduction of ice concentration in the Greenland and Labrador Seas. We also see an increase in ice concentration along the coast of Newfoundland, in the Barents Sea and for a small region in the Denmark Strait. The updates in ice concentration are of the order 0.5 and smaller, and cover a relatively small spatial area compared to the total ice area. For the regions far from the ice edge and land boundaries, the ice concentration update is below the plotting threshold shown in the color scale of Figure 5(a).

The situation for a winter time update of temperature in the upper layer of the ocean model is shown in Figure 5(b), and the updates in surface layer temperature occur over a smaller region than the updates in ice concentration. To some extent the area difference in the figures for ice concentration and temperature updates are due to the plotting threshold, but as will be shown later, it is also due to the statistics of the ensemble, and model limitations. By comparing with the ice concentration analysis update in winter, it should be noted that the sign of the temperature update is mostly opposite to the sign of the ice concentration update. That is, when ice concentration is increased the sea surface temperature is decreased and vice versa. The temperature updates are of order $1°$ Kelvin or less.

The plot of winter time updates of salinity in the upper layer of the ocean model is shown in Figure 5(c). Again we note that the updates occur mainly in the ice edge regions with very small updates far away from the ice edge and land boundaries. If we compare the salinity update with the ice concentration update however, we find that the region with significant salinity updates is somewhat broader along the ice edge. We also see that the connection between the sign of the ice concentration update and the salinity update is less clear than for the connection between temperature and ice concentration update. There is a weak tendency for the sign of salinity updates to be the opposite to that of the ice concentration updates at this time of year. The salinity corrections are of the order 0.1 psu.

The analysis update in ice concentration for a typical summer situation is shown in Figure 5(d). Compared to the winter time situation given in Figure 5(a) we see that the update occurs over a larger area. Although the update is still strongest close to the ice edge, the region with significant updates is much broader as compared to the situation in winter. For the specific time given in Figure 5(d) we see mostly an increase in ice concentration for the region from the Fram Strait to Sevarnaya Zemlya, in the Baffin Bay and close to the Canadian Arctic Archipelago. Reductions in ice concentration are seen in regions of the Beaufort and East Siberian Seas.

Comparing the summer time update in sea surface temperature, Figure 5(e) with the situation in winter, Figure 5(b), we see that the temperature update covers a larger region. However, the winter to summer increase in areas with significant updates are much larger for ice concentration than for sea surface temperature. In fact, the sea surface temperature updates are located mainly along the ice edge, as in winter. The sign of the temperature update is mostly opposite to the sign of the ice concentration update, which is the same behavior as in winter.

Finally we see the summer update in sea surface salinity, Figure 5(f). We note that the regions with a significant update in salinity are larger than the regions with significant update in ice concentration. It is also

(a) Ice concentration update, day 15

(d) Ice concentration update, day 218

(b) Temperature update, day 15

(e) Temperature update day, 218

(c) Salinity update, day 15

(f) Salinity update, day 218
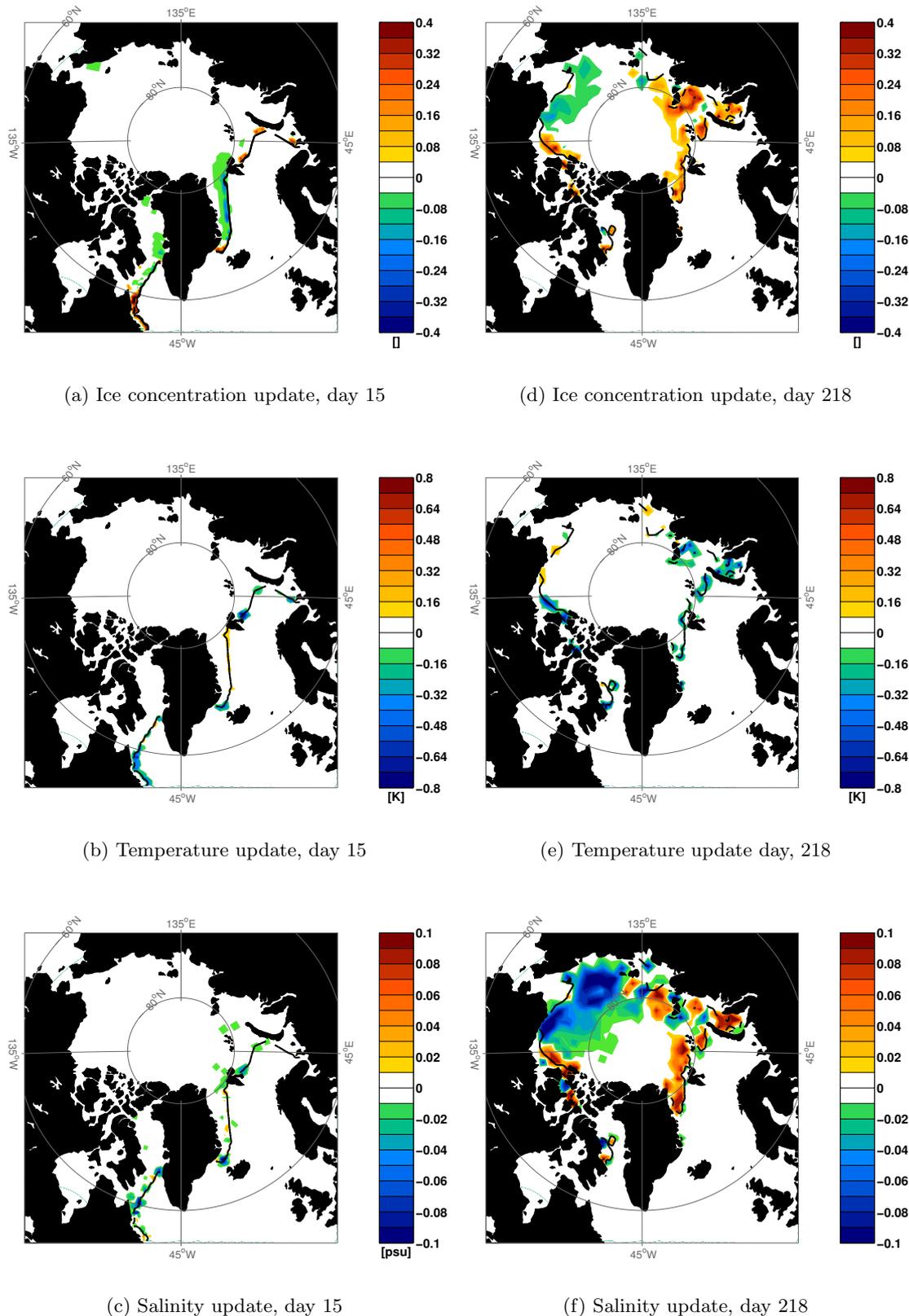
Figure 5: Selected analysis updates for the ensemble average. Typical situations for summer and winter in 1999 are shown. Figure 5(a), 5(b), and 5(c) show typical winter time updates in ice concentration sea surface temperature and salinity, respectively. Figure 5(d), 5(e), and 5(f) show typical summer time updates in ice concentration sea surface temperature and salinity, respectively. The contour lines indicate the 0.7 ice concentration contour of the ensemble mean.

important to note that the magnitude of the updates in summer are larger than the magnitude of updates in winter. Differences can also be seen in the sign of the updates as compared to the situation in winter. Whereas the winter time salinity updates had no clear connection with the sign of the ice concentration updates, we see that the summer time salinity updates have the same sign as the ice concentration updates.

The ensemble update plots give hints of the complex statistical behavior described by the model ensemble. Behavior which show considerable changes both in time and space. This behavior will be considered more closely in following sections.

5 Second order statistics of the model ensemble

The representation of error statistics is an important property of the EnKF. For a multivariate assimilation scheme, as presented here, it is of particular importance that the covariance between different variables, for instance temperature and ice concentration, is realistic. In this section we examine how the variance and covariance fields from the ensemble change in time and space, trying to illustrate some aspects of the ensemble behavior. An understanding of the processes which determine the statistics is important, and can make it easier to evaluate which model improvements should be implemented for the system at a future stage.

Concerning the evolution of the second order ensemble statistics, it is important to realize that the evolution is dependent upon the prior ensemble statistics and the random forcing applied. The random forcing induces most of the variation in the ensemble. A lot of the following discussion will therefore focus upon the effect the random forcing has upon the ensemble.

In this section, we will use the following notation to denote the covariance, variance and correlation:

$$
\begin{aligned}
\mathrm{cov}\left\{F(\boldsymbol{x}_1), G(\boldsymbol{x}_2)\right\} &= \overline{\left(F(\boldsymbol{x}_1) - \overline{F(\boldsymbol{x}_1)}\right)\left(G(\boldsymbol{x}_2) - \overline{G(\boldsymbol{x}_2)}\right)} \\
\mathrm{var}\left\{F(\boldsymbol{x})\right\} &= \mathrm{cov}\left\{F(\boldsymbol{x}), F(\boldsymbol{x})\right\} \\
\gamma\left\{F(\boldsymbol{x}_1), G(\boldsymbol{x}_2)\right\} &= \frac{\mathrm{cov}\left\{F(\boldsymbol{x}_1), G(\boldsymbol{x}_2)\right\}}{\mathrm{var}\left\{F(\boldsymbol{x}_1)\right\}\mathrm{var}\left\{G(\boldsymbol{x}_2)\right\}} \ .
\end{aligned}
\tag{14}
$$

One should note that the positions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ can denote two different locations. The covariance and correlation can therefore be between two different fields at two different points.

5.1 Ice concentration variance

The variance of ice concentration in the model is important for the analysis update since it determines the Kalman gain matrix $\boldsymbol{K}_e$. For a local analysis, which operates on individual grid cells, there will be no analysis update of other variables in a grid cell (temperature, salinity etc) unless we have a nonzero ice concentration variance there. The ensemble ice concentration variance in each grid cell therefore deserves some attention.

An example of the ice concentration standard deviation for winter is shown in Figure 6(a). Along with the variance plot we have also indicated the 0.7 ice concentration contour of the ensemble average. In Figure 6(a) we have masked out the lowest values of ice concentration, leaving only a limited area with significant ensemble variance. Not surprisingly we see that the regions with highest values are the regions which are close to the ice edge, e.g. in the Barents Sea, the Greenland Sea, and the Labrador Sea. The analysis updates of ice concentration, temperature and salinity in winter, e.g. Figure 5(a), illustrated that the changes in these variables happened in proximity of regions with high ice concentration standard deviation in Figure 6(a).

Thorndike and Colony (1982) showed that a lot of the ice motion over small time scales can be explained by the geostrophic wind speeds, especially far from regions with thick ice. Similar observations apply for the ice dynamics model, and a lot of the ice concentration variance along the ice edge can be explained by the random forcing component applied to each member. As the ensemble members have different degrees of transport across the ice edge this will lead to increased ice concentration variance by the following simple mechanism: Ice advection across the ice edge into warmer waters will, through ice melting, cool the surface waters and make them fresher. If this effect is stronger than the warming effect of mixing (due to a possible increase in the applied surface stress) then the net effect will be a cooling and freshening of the surface layers. Consequently, the more ice is transported across the ice edge, the more likely it is that the ice–free waters will be cooled to such a degree that ice can persist there. This effect links the ice concentration variance to the random forcing because different degrees of transport across the ice edge lead to different ice concentrations. A similar argument applies for the ice-pack-side of the ice edge. If the net heat balance for the water interface is close to zero, then different degrees of ice divergence will lead to variations in the local ice concentration. Additionally an increased water fraction could lead to enhanced mixing of warm waters from below the sea ice which could amplify local changes in ice concentration.

Another effect of the varying wind forcing for the ensemble members is the increased ice concentration variance close to land boundaries. This can be seen in the Labrador Sea region, where offshore ice transport reduce the model ice concentration in the western part of Davis Strait. Again, the different degrees of

(a) 1999 day 15   (b) 1999 day 218
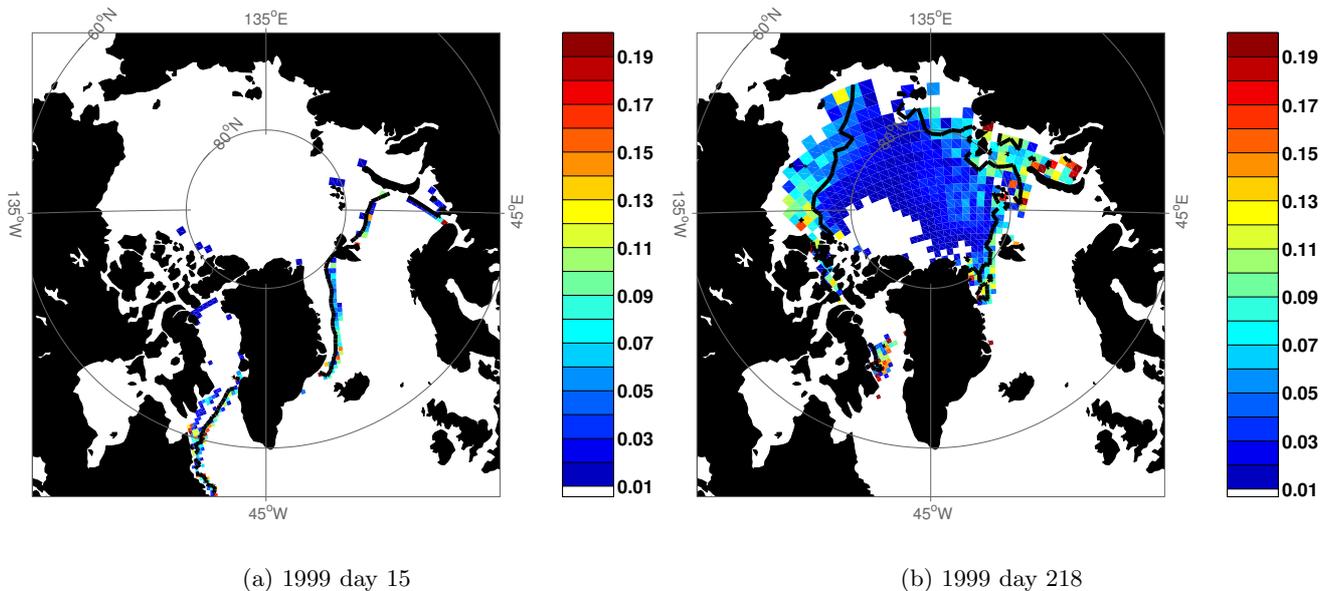
Figure 6: Figure 6(a) show the ice concentration standard deviation for the individual grid cells as calculated from the ensemble. Also shown is the position of the 0.7 ice concentration contour (solid line). The statistics are based on the prior ensemble, and at the time Julian day 15 in 1999 (winter conditions). Figure 6(b) is similar to Figure 6(a) but shows the summer situation (Julian day 218 in 1999).

onshore/offshore winds in the random forcing component of the individual ensemble members will increase the ice concentration variance in the ensemble.

Important in the ice concentration standard deviation in Figure 6(a) are the small values far away from the ice edge, for example in the Central Arctic Ocean. In the present model system there is virtually no ice concentration variance during winter in the Central Arctic Ocean. This is connected to deficiencies in the ice model, where any open water fraction due to ice divergence is quickly closed to the maximum allowed ice concentration, c_{max}, by freezing of new ice. This means that all ensemble members have an ice concentration of c_{max} for large areas in the Central Arctic Ocean, and the ice concentration variance is therefore zero. In reality there are small variations of wintertime ice concentration in the central Arctic Ocean due to small-scale opening of leads (e.g. *Kwok*, 2002). These variations, although small, are important because of the large heat fluxes which occurs over leads in winter.

The low ice concentration variance give ensemble based error estimates which are essentially zero. This means that the EnKF ensemble statistics predict a perfect model and no analysis update is done in these regions, an effect which explain the small analysis updates in the central Arctic Ocean during winter, e.g. Figure 5(a). In the future we hope to improve the wintertime behavior of the ensemble in the Arctic Ocean by improving both the ice dynamics and ice thermodynamics models.

If we look at the situation for ice concentration variance in summer we see a somewhat different picture. Figure 6(b) shows the ice concentration standard deviation for summer, and it is apparent that there is a much larger region with ice concentration standard deviation above the 0.01 threshold. As with ice concentration variance during winter, the highest values are mainly found close to and seawards of the ensemble–average 0.7 ice concentration contour. In summer, the leads opening up in the Central Arctic Ocean do not freeze immediately as they do in winter. This way the ice concentration variance is allowed to increase in response to differences in ice divergence and melting among the members. The ice transport across the ice edge is not as important in summer as in winter. This is an effect of winds being weaker, and of ice concentration variance imposed by ice melt becoming a more important contributor to ice concentration variance.

5.1.1   Ice concentration variance and the ice edge

As illustrated in the wintertime and summertime situations, Figure 6(a) and 6(b), the spatial pattern of ice concentration variance varies strongly through the season. Moreover, the highest values of ice concentration
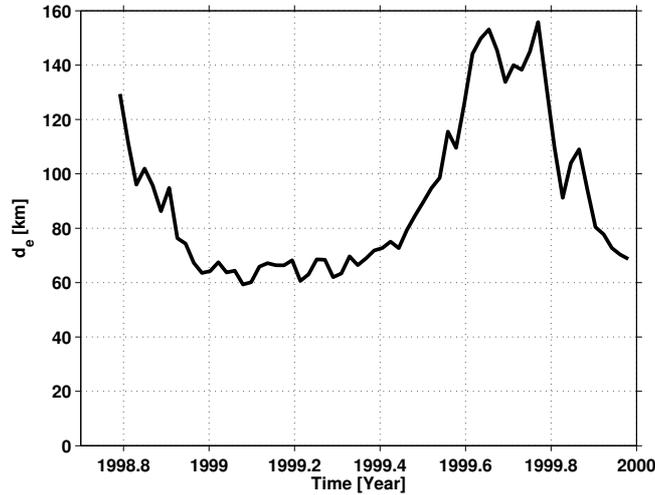
Figure 7: The weighted average distance from the ice edge/land boundaries to regions of high ice concentration variance, $\overline{d_e}$ from equation (16).

variance are largely controlled by the location of the ice edge and by land boundaries. To understand how the ice concentration variance pattern evolves in time it is useful to obtain a simple measure of the distance between the ice edge/land boundaries and regions with high ice concentration variance. To this end we will use the field

$$d_e(\boldsymbol{x}) = \text{distance from point } \boldsymbol{x} \text{ to ice edge or land}\,, \tag{15}$$

where the ice edge is defined by the 0.01 ensemble average ice concentration contour.

A measure of the positions with high ice concentration variance relative to the ice edge or land boundaries can now be obtained, utilizing the field $d_e(\boldsymbol{x})$. Taking the normalized ice concentration variance as weights, we calculate a mean value of $d_e(\boldsymbol{x})$

$$\overline{d_e(\boldsymbol{x})} = \int d_e(\boldsymbol{x}) \, \text{var}\left\{c(\boldsymbol{x})\right\} \, d\boldsymbol{x} \Big/ \int \text{var}\left\{c(\boldsymbol{x})\right\} \, d\boldsymbol{x}\,. \tag{16}$$

Here $\overline{d_e}$ is a weighted average distance from the ice edge/land boundary where regions with high ice concentration variance has been given the most weight. It will be indicative of how far into the ice pack areas of high ice concentration variance are located. Based on Figure 6(b) and 6(a) we anticipate high $d_e(\boldsymbol{x})$ in summer, and low $d_e(\boldsymbol{x})$ in winter. Note that $d_e(\boldsymbol{x})$ will vary throughout the season due to the moving ice edge. Therefore we tried to scale the variable $d_e(\boldsymbol{x})$ with a seasonally dependent parameter describing this effect. However, the analysis did give similar results as without such a scaling, and we therefore show the case without scaling here.

The time series of of $\overline{d_e}$ is shown in Figure 7, and shows how $\overline{d_e}$ varies in the experiment. At the start of the experiment $\overline{d_e}$ is relatively high, but when the winter season sets in, $\overline{d_e}$ is reduced and is close to its minimum value at the end of 1998. As was apparent in Figure 6(a) the regions with high ice concentration variance are close to the ice edge/land boundary during the winter season, giving low values for $\overline{d_e}$. At the onset of the melt season, the ice concentration variance increases. The ice melt in turn affects most of the Arctic ice pack at the peak of the Arctic summer. The increase in d_e indicates that the ice concentration variance is no longer confined to the close proximity of the ice edge, although it is of largest magnitude there, as was seen from Figure 6(b). Towards the next winter the value of $\overline{d_e}$ is again reduced, and is approximately at the same level at the start of year 2000, as at the start of year 1999.

The temporal behavior of the ice concentration variance shown here is largely determined by the background heat fluxes of the ice model. As was mentioned on page 9, the ECMWF forcing has a strong gradient across the ice edge in winter, with very strong negative heat flux in the central ice pack. Because of this, the random forcing component used in the EnKF is not large enough to create any significant ice concentration variance, except close to the ice edge. As a consequence, our ensemble will have a reduced ice concentration variance in winter. In summer the ice cover can develop more freely, which translates into increased ice concentration variance for this time of year.

5.1.2   Ice concentration PDF

In Kalman Filter methods, a variance minimizing estimate is used as the final analysis. It is important to remember that the maximum likelihood estimate is not necessarily equal to the variance minimizing estimate. If the forecast probability density function is Gaussian, however, then the analyzed estimate becomes equal to the maximum likelihood estimate.
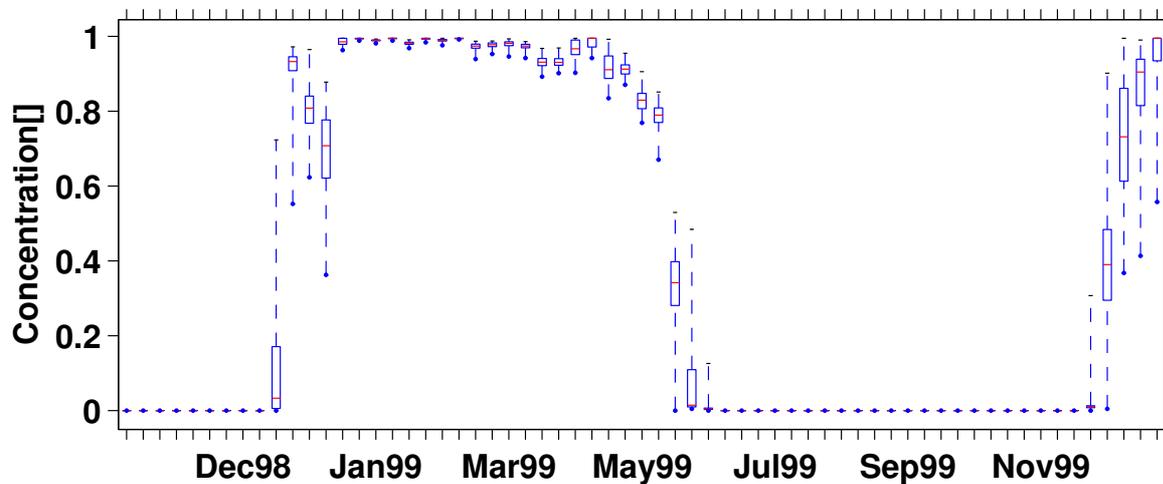
Figure 8: A box–whisker plot for the sample of the ice concentration probability density distribution. The samples are from point O for different times in the assimilation experiment. The plots show the lower, middle (median) and upper quartile, and the lines extending from the box gives the extreme values of the data.

Figure 8 shows a box–whisker plot of the ice concentration in point 'O' in Figure 1(b). One can clearly see the collapse of the ensemble ice concentration for this point in summer. In winter, as the ice edge is at some distance from the point, one can also see that the ice concentration variance is very low.

At the times when the ice edge is closer to point 'O', there is significant ice concentration variance close to the ice edge. Note, however that at many times the sample distribution is skewed, e.g December 1998 and late December 1999. The skewness of the sample is most frequently encountered when the median value is close to the zero or maximum ice concentration. At many times, one can also see that the range of values the samples take is very large compared to the distance between the lower and upper quartile. The skewness, and the large range of values both indicate that the sample is taken from a non–Gaussian distribution.

The distribution of ice concentration can of course never be Gaussian, because it is limited to values between 0 and 1. The behavior of the sample distribution of ice concentration should therefore be quantitatively similar to variables which are constrained by upper and lower limits, be it saturation values or other physical constraints. Examples of variables with this type of behavior are the concentration of biological material in biochemical models (*Natvik and Evensen*, in press) or soil–moisture (*Reichle et al.*, 2002). An improvement of the Gaussianity of the sample could be possible if suitable transformations are applied to the model state variables prior to the analysis (*Bertino et al.*, submitted).

5.2   Ensemble Covariance

An important aspect of the assimilation scheme is how well the surface layers of the ocean model are updated. The temperature of the surface needs to be reduced when new ice is introduced, and the update of surface salinity should take into account that sea ice has a lower salinity than that of the ocean surface. All these calculations for the sea surface should be taken care of by the analysis equation (3). In this section we will focus upon ensemble–based ice concentration covariances for variables taken from the same grid cell (local covariance), for instance ice concentration and sea surface temperature from the same model grid cell. In the notation from (14) the covariance between e.g. ice concentration and temperature of ocean layer 1 in position x can be written as cov $\{c(x), T(x, 1)\}$. Due to the nature of the covariances, which are generally located close to the moving ice edge, and whose position depends upon the season, we will make some use

(a) cov $\{c(\boldsymbol{x}), T(\boldsymbol{x}, 1)\}$ (b) cov $\{c(\boldsymbol{x}), S(\boldsymbol{x}, 1)\}$
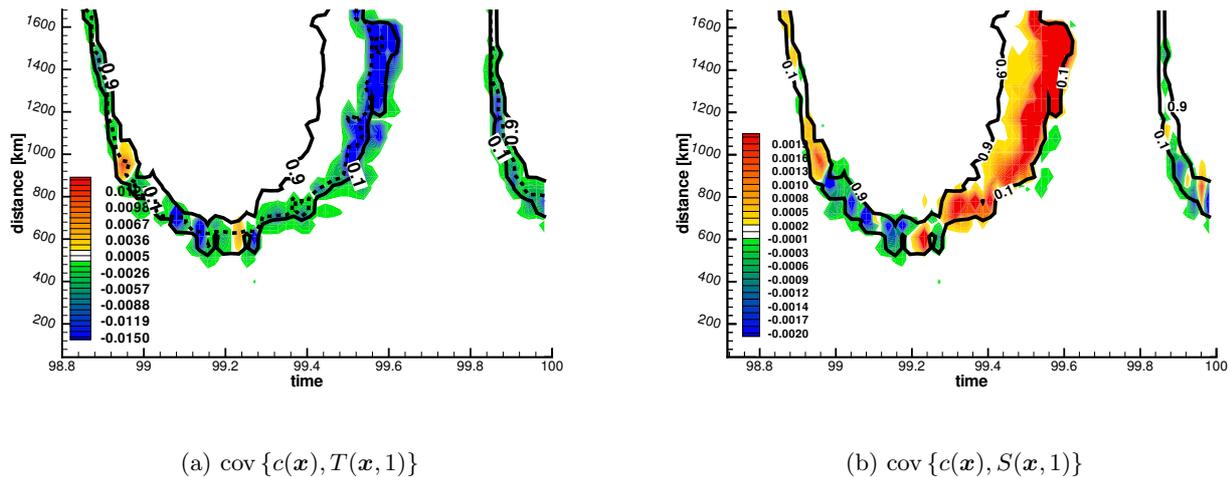
Figure 9: Figure 9(a) shows a time/distance (Hovmöller) plot for the grid cell covariance between ice concentration and surface layer temperature along section A in Figure 1(b). Also shown is the ensemble–mean 0.1 and 0.9 ice concentration contour. Figure 9(b) is similar to Figure 9(a) but shows covariance between ice concentration and surface salinity. For temperature the dashed line denotes the contour for which all ensemble members have an ice concentration of 0.01 or larger.

of Hovmöller plots in the following.

A Hovmöller plot of the grid cell covariance between ice concentration and sea surface temperature is shown in Figure 9(a), while a corresponding plot of the covariance between ice concentration and sea surface salinity is shown in Figure 9(b). These plots are for section A in Figure 1(b), going from the Labrador Sea into Baffin Bay. In both of these figures we have indicated the location of the ensemble–mean 0.1 and 0.9 ice concentration contours.

This particular section in the Baffin Bay was chosen because it appeared best suited to demonstrate the mechanisms causing the covariance signals, which will be explained later.

5.2.1 Ice concentration – SST covariance

The covariances between surface temperature and ice concentration in Figure 9(a) are mostly negative or zero throughout the season, and as for ice concentration we find the highest absolute values close to the ice edge. The ice thermodynamic model used here has some limitations which restrict the covariance between ice concentration and sea surface temperature. The most important restriction is the requirement that the sea surface temperature is at the freezing point in order for ice to exist. This effectively groups the connection between ice concentration and sea surface temperature into three different situations.

1. $c > 0$ and $T = T_f$ for all ensemble members

2. $c = 0$ and $T \geq T_f$ for all ensemble members

3. A mixture of members with properties from 1 and 2

Consider first the situation where all ensemble members in a grid cell have a positive ice concentration, i.e, situation 1 above. In this case, variations in sea surface temperature is determined by the freezing point dependence upon salinity, and the covariance with ice concentration should be small in magnitude. From Figure 9(a) we see that this occurs in summer; As the region with ensemble–averaged ice concentration between 0.1 and 0.9 broadens, there is also a large region with a very low covariance magnitude between sea surface temperature and ice concentration. This region is roughly bounded by the dashed line, and the ensemble 0.9 ice concentration contour, where the dashed line denotes the limit for which all ensemble members have $c > 0$. In other words, this region corresponds to situation 1 above.

Then consider a situation where some ensemble members have $c > 0$ and some members have $c = 0$, situation 3 above. In this situation we can have ensemble–variations in both ice concentration and sea
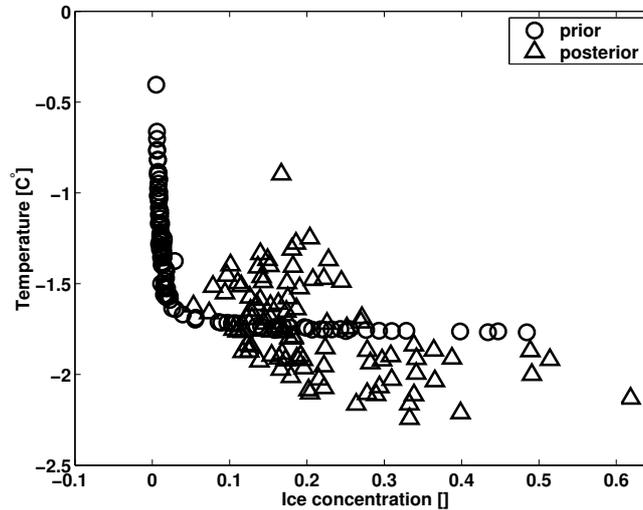
Figure 10: Sea surface temperature plotted against ice concentration for the point x_0=(58.85°W 66.05°N) on Julian day 162 in 1999. Shown are scatter plots for the situation prior to (circles) and after (triangles) the assimilation. Note the nonlinear relation prior to the analysis.

surface temperature, because there are temperatures above the freezing point associated with $c = 0$. Above freezing point temperatures are associated with $c = 0$ and freezing point temperatures are associated with $c > 0$, resulting in a negative covariance. This negative covariance between ice concentration and sea surface temperature, is seen between the dashed line and the 0.1 ice concentration contour in Figure 9(a).

Note that the analysis equation (3) uses the covariances between surface temperatures and ice concentrations to update the model state. This assumes a linear relation between the variables, which is not the case for the relation between ice concentration and sea surface temperature in our model. This is due to the restriction mentioned above, where sea surface temperature is fixed to the freezing point when ice is present. To illustrate this, we have in Figure 10 plotted the ensemble values of sea surface temperature against ice concentration for one of the model grid points along section A (the point marked with 'O' in Figure 1(b)). For the case with an ensemble mixture of ice free and ice covered water in a grid cell the nonlinear ensemble relation between ice concentration and sea surface temperature can lead to unwanted effects. In Figure 10 this is apparent; prior to the analysis the sea surface temperature is fixed at the freezing point of sea water when ice is present, and the deviations from the freezing point only occur when ice is absent. After the analysis however, this is not the case, and there are cases with ice present and temperatures above the freezing point, and even cases with temperatures below the freezing point. These below freezing temperatures are corrected after the analysis is performed. The above freezing temperatures when $c > 0$ are not corrected, and will result in melting of some of the newly introduced ice in Figure 10. As mentioned earlier, the Gaussianity of the model ensemble could be improved by applying suitable transforms prior to the analysis *Bertino et al.* (submitted). This would also improve the nonlinear relation between ice concentration and sea surface temperature.

We note here that the requirement of having sea surface temperature fixed at the freezing point when ice is present is a quite common simplification in sea ice models. It is easy to change this requirement, and introduce more realistic thermal exchanges between the ocean and sea ice (*Holland and Jenkins*, 1999). Note however that more realistic thermal exchange formulas are very sensitive to the sea surface temperature elevation above the freezing point, and this leads to strong melting if ice is present. In the end the net result of such a model improvement might prove to give similar results for the ice concentration – sea surface temperature relation as illustrated in Figure 10.

The negative covariance between ice concentration and sea surface temperature is seen throughout the year, not only in the Hovmöller plot for the section in the Baffin Bay. A plot of the weighted average ice concentration – sea surface temperature correlation is shown in Figure 11. The average is weighted with the ice concentration variance, and shows how this average correlation has negative values throughout the experiment.
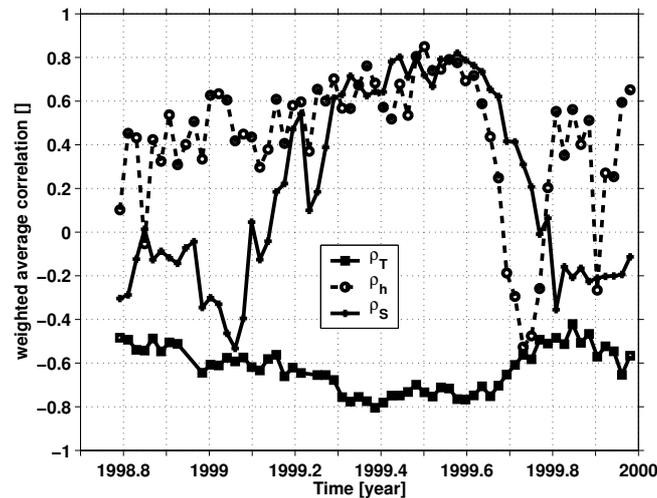
Figure 11: Weighted average correlation between ice concentration and sea surface salinity (solid line and star symbols), between ice concentration and sea surface temperature (solid line and squares), and between ice concentration and ice thickness (dashed and circles). The average has been weighted by the ice concentration variance.

5.2.2 Ice concentration – SSS covariance

For the covariance between ice concentration and sea surface salinity in Figure 9(b), we have a somewhat different situation than for the sea surface temperature–ice concentration covariance. With salinity we are no longer dealing with ice concentration covariance against a variable which is fixed to a lower bound when ice is present. As compared to covariance between ice concentration and sea surface temperature we also have a broader region with relatively high covariance absolute values. Note also the changes in sign for the covariance in the Hovmöller plot. As the ice edge starts to retreat, we see that the covariance changes sign along the ice edge from being mostly negative to being mostly positive (around time-axis point 99.2).

The cause for this change of sign in sea surface salinity–ice concentration covariance can be traced back to the different mechanisms controlling the ice concentration. Consider first the situation where freezing or melting changes the ice concentration. In this case an increased (reduced) ice concentration leads to increased (reduced) sea surface salinity because brine is released from the freezing ice, and melting ice gives a positive freshwater flux to the ocean surface layers. This results in a positive covariance between sea surface salinity and ice concentration. Note that brine released from newly frozen ice induce vertical mixing which could complicate this argument. If the water below the mixed layer is more saline and cold (typical of the Arctic Ocean "Cold Halocline Layer") then the positive covariance becomes even stronger. On the other hand, if the water below the mixed layer is relatively warm this would in turn affect the ice cover and decrease the magnitude of the negative salinity covariance.

As a second factor controlling the ice concentration, consider the case where ice is transported across a thermal boundary separating waters at the freezing point from warmer waters. As the ice reaches the warmer waters it melts, which leads to cooling and freshening of the surface water masses. If the heat content of the mixed layer is low enough, prolonged periods of ice transport help to cool the surface waters to the freezing point and further transport into the region leads to increased ice concentration. Concerning the covariance between sea surface salinity and ice concentration, the result of this transport situation is a reduced sea surface salinity being associated with increased ice concentration, that is, a negative covariance.

Returning to Figure 9(b), it seems like the model ensemble is able to describe the two different mechanisms mentioned above. The ensemble differences in ice transport is likely to be causing the mostly negative sea surface salinity/ice concentration covariance during the advance of the ice edge in the winter of 1999, and to some extent during the fall of 1999 and 2000. A strong support for this mechanism can also be found from covariances with the ice velocity normal to and seawards of the ice edge. Unfortunately we were not able to recover the entire record of ice velocity from our data files. The ice velocity data is missing for the spring and summer of 1999, leaving us with ice velocity data for the autumn and early winter of 1998 and 1999. At these times, and for the Baffin Bay section, the covariance between the ice velocity component normal to the ice edge and sea surface salinity (not shown) is consistently negative. This covariance do suggest

that the negative ice concentration–sea surface salinity covariance is a result of seawards ice transport, at least during the advance of the ice edge in fall and early winter. We believe that the transport mechanism persists throughout the year, while the change in sign of the ice concentration/sea surface salinity covariance in summer is due to ice–melt becoming the more dominant effect.

The positive ice concentration–sea surface salinity covariance in summer, seen in Figure 9(b), and at the start of the time series, indicates that the ice melt mechanism is dominating the covariance. Note however, that the sign of the covariance during the fall of 1999 changes frequently between positive and negative values. Mesoscale events occurring along the ice edge are important effects here. As the ice heat balance often changes from net melting to net freezing across the ice edge, events which bring warmer air in from the south can disturb this balance and cause melting well into the ice pack. As explained, this has the capability of changing the sign of the covariance between ice concentration and salinity.

The seasonally dependent covariance between ice concentration and sea surface salinity is not only important in the Baffin Bay section, were we have focused our attention. It can also be seen from the area and variance weighted average of the correlation between ice concentration and salinity. An illustration of this is given in Figure 11, where the seasonally changing average correlation between ice concentration and sea surface salinity is apparent. At the beginning of the experiment and through much of the winter, the correlation between ice concentration and salinity is negative, with values between -0.5 and 0. The low absolute values indicate that the effect of transport upon the correlation is relatively weak, probably due to spurious warm weather events occurring close to the ice edge. As the summer approaches, the impact of melting starts to dominate over transport along the ice margins, and the correlation between ice concentration and sea surface salinity changes to a positive sign and increases until it peaks at the highest values in summer. It then decreases, and towards the end of the fall freeze-up, negative correlation values are again seen.

5.2.3   Ice concentration – ice thickness covariance

The covariance between ice concentration and ice thickness along the same section as in 9(a) reveals a mostly positive covariance between ice concentration and ice thickness(not shown). This is consistent with the idea that "thicker ice lasts longer", and can explained by reduced horizontal melting for thicker ice. One should note however, that situations arise in which there is a negative correlation, as is evident in Figure 11. During the fall freeze-up, a lot of thin newly frozen ice increases the ice concentration, and consequently reduces the thickness of existing ice. Starting from initially equal conditions, this means that the faster the ice cover freezes, the tinner the ice in the model will become. The result of this is a negative covariance between ice concentration and ice thickness.

5.3   Spatial correlation functions

Another important factor is how the ensemble covariance and correlation varies in the horizontal. We therefore focus the attention on so–called correlation functions. The correlation function used here is given by $\gamma \{F_1(\boldsymbol{x}), F_2(\boldsymbol{x}_0)\}$ where $F_1(\boldsymbol{x})$ and $F_2(\boldsymbol{x}_0)$ denote variables of interest. This gives the correlation between the ensemble members of variable F_2 located at $\boldsymbol{x}$ and variable F_1 located at $\boldsymbol{x}_0$. In the following we keep $\boldsymbol{x}_0$ fixed to a specified value for plotting purposes.

The ensemble correlation functions for ice concentration, temperature and salinity against the ice concentration of the fixed point $\boldsymbol{x}_0$=(58.85°W, 66.05°N) are shown in Figure 12. The correlation functions were acquired from the ensemble for day 162 in 1999. The fixed point $\boldsymbol{x}_0$ is identical to point O in Figure 1(b). It is located close to the ice edge at this time, at a location in the Davis Strait.

In general we see that the correlation functions reveal complex patterns that are dependent upon past flow history of the ice and ocean. If we look at the correlation function for ice concentration, Figure 12(a), we see a positive correlation that decay rather rapidly away from the fixed point $\boldsymbol{x}_0$. Ice concentration is not a conserved quantity in the ice model, and one would therefore expect that the past information from the flow is lost quickly, effects that would lead to such a fast decay of the correlation. The ice concentration correlation function is limited by the ice edges of the ensembles, and is close to zero seawards of the ensemble-average ice edge illustrated in Figure 12(a).

The correlation function for sea surface temperature is shown in Figure 12(b). As suggested by the covariance plots between ice concentration and sea surface temperature, Figure 9(a), we see that there is a negative correlation between ice concentration and sea surface temperature for $\boldsymbol{x} = \boldsymbol{x}_0$. This negative

(a) $\gamma\left\{c(\boldsymbol{x}_0), c(\boldsymbol{x})\right\}$ (b) $\gamma\left\{c(\boldsymbol{x}_0), T(\boldsymbol{x}, 1)\right\}$

(c) $\gamma\left\{c(\boldsymbol{x}_0), S(\boldsymbol{x}, 1)\right\}$
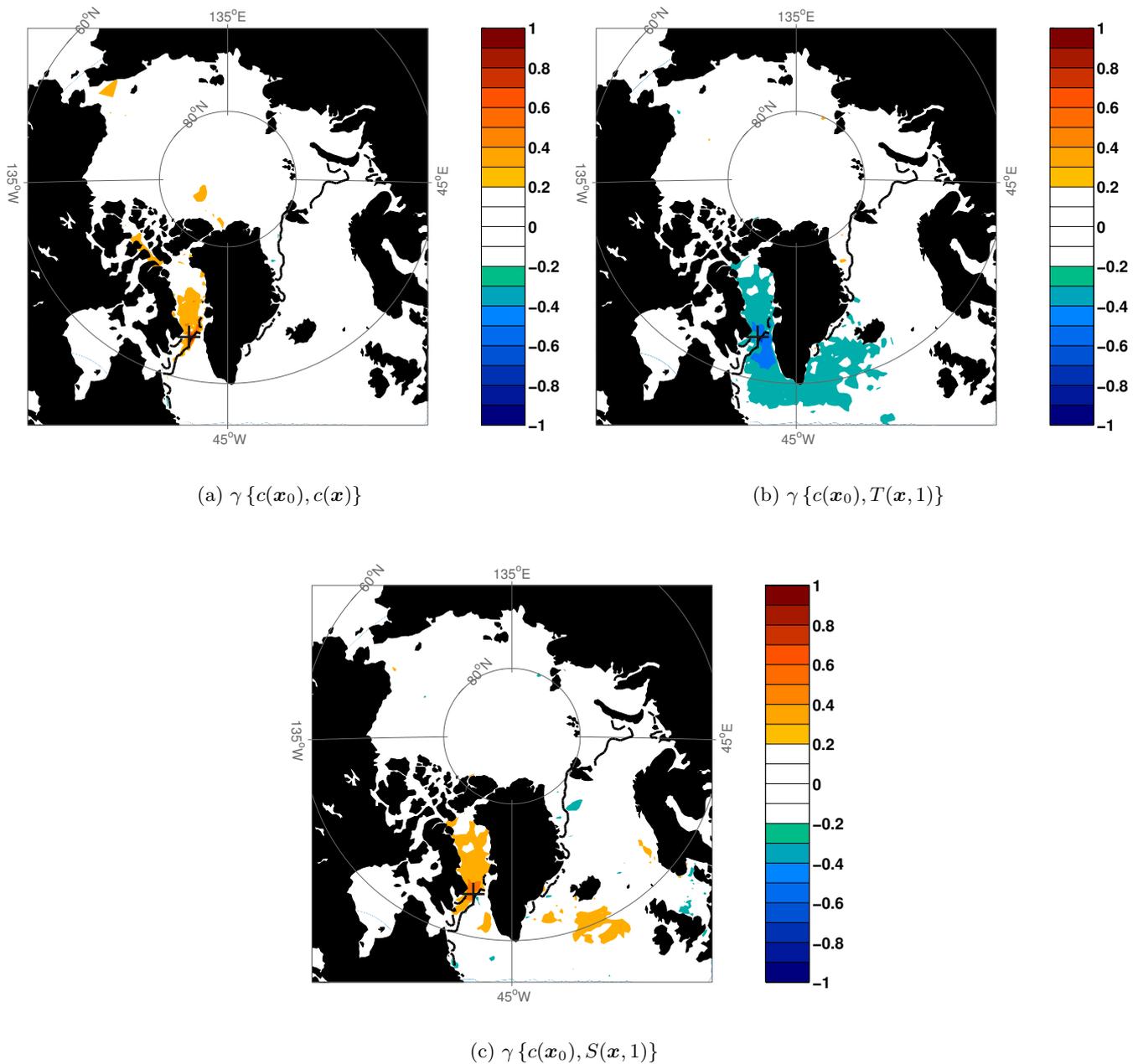
Figure 12: Correlation functions for 1999 day 162, where the fixed point is located at $\boldsymbol{x}_0 = (58.85°\text{W}, 66.05°\text{N})$ and is marked with a plus sign in the figures. The correlation functions are for ice concentration in the fixed point against ice concentration 12(a), sea surface temperature 12(b) and sea surface salinity 12(c).

correlation also extends at some distance away from the fixed point, and relative to the correlation in $\boldsymbol{x} = \boldsymbol{x}_0$, the ice concentration–sea surface temperature correlations decay at a slower spatial rate than the ice concentration–ice concentration correlation. Also notice that the region with relatively high absolute values of temperature–ice concentration correlation extends beyond the ice edge and well into the Labrador Sea. The surface circulation around the southern tip of Greenland, Cape Farewell, generally follows the coastline in an anticyclonic sense, and water masses are transported northwards along the western coast of Greenland (The West Greenland Current). These water masses help to maintain the roughly northeastwards direction of the ice edge from Newfoundland towards the West Coast of Greenland. South of Davis Strait a branch of the West Greenland Current crosses over from the West Coast of Greenland towards Baffin Island and finally join up with the Labrador Current (*Chapman and Beardsley*, 1989).

The point $\boldsymbol{x}_0$ used here is situated in the Davis Strait and the waters passing the point $\boldsymbol{x}_0$ should contain a component originating from the West Greenland Current. As water in the West Greenland Current is

transported northwards, the ensemble differences in the upstream ocean surface temperature at an earlier time (e.g. at Cape Farewell) contribute to differences in ice concentration at the fixed point, where positive temperature anomalies lead to reduced ice concentration at the fixed point x_0. This in turn leads to the negative covariance in Figure 12(b) between ice concentration in x_0 and ocean surface temperatures in the oceanic upstream direction. These negative correlations can be traced back into the Irminger Sea, although with decaying absolute value with upstream distance. Also note that part of the water that circulate within the Labrador Basin show a relatively high correlation with temperature at the fixed point, probably because it originates in the same region as the water that reaches the ice edge at $(58.85°\text{W}, 66.05°\text{N})$.

In Figure 12(b) we note that there is some correlation appearing in the Baffin Bay. At this time of the year, there still is a high fractional ice cover polewards of the point x_0 in the Baffin Bay. Since the model restricts the sea surface temperature to the freezing point when ice is present, this correlation must come from the freezing point dependence upon sea surface salinity. The correlation function for sea surface salinity is shown in Figure 12(c), and close comparison of the ice concentration–temperature and ice concentration–salinity correlations reveals the exact same pattern (with different sign) in the Northern Baffin Bay. This is to be expected, as the model has a linear relationship between freezing point temperature and sea surface salinity.

For salinity we see the strongest correlation with ice concentration from the ice edge and northwards into the ice pack. Roughly speaking this can be attributed to the effect of thicker ice leading to a more persistent ice cover close to the ice edge. As thick ice is transported southwards it is less likely to completely melt away as with thinner ice. Thicker ice from enhanced freezing is again connected to the presence of more saline water in the upper water column leading to the positive correlation pattern polewards of the ice edge. As compared to the temperature correlation function, the correlation with sea surface salinity is much weaker over open water.

Finally we note that spurious correlations occur at regions far from the point of interest, of which some arise due to the spurious correlations in the random forcing fields. This highlights the importance of using local analysis and a radius of influence in the analysis scheme, at least for small ensemble sizes (*Mitchell et al.*, 2002).

6 Summary

In this study we have demonstrated an implementation of the Ensemble Kalman Filter with a coupled sea ice and ocean model. The ocean model used is HYCOM (*Bleck*, 2002), while the ice model uses the Elastic–Viscous–Plastic rheology of *Hunke and Dukowicz* (1997). The observations used in the assimilation were sea ice concentration data, which were calculated using remotely sensed passive microwave data. The ice concentration was assimilated every 7th day for a period longer than a year, allowing us to draw conclusions on the impact of the assimilation over a seasonal cycle. The focus of the study has been on the corrections performed by the EnKF, and on the model ensemble statistics. The statistics of the model ensemble determines the impact of the ice concentration, and a good understanding of this is important in a multivariate ensemble system.

The assimilation experiment was first compared with the results of a model run without assimilation. As expected the difference between modeled and observed ice concentration was smaller in the case with assimilation, than in the case without. This effect was strongest in summer, as can be seen from Figure 2(a). In winter we found that the impact of the assimilation was weaker. This is in part due to the ECMWF forcing fields, which depends on SSM/I derived ice concentrations. In winter the strong atmospheric temperature gradient across the ice edge prescribed in the ECMWF forcing to some extent determines the location of the modeled ice edge. As the ECMWF forcing depends on SSM/I data, it will act to reduce the distance between observed (SSM/I–derived) and modeled (ECMWF–forced) ice concentration. In summer the effect of using SSM/I data in the ECMWF forcing is not as important, because the atmospheric temperature gradient across the ice edge is smaller, and because solar absorption by open water becomes important.

The effect of the assimilation was also seen in the Arctic ice thickness distribution. The main effect of the assimilation was to retain a larger fraction of thick, multi–year ice in summer. The effect of this was seen on the total volume of ice in the Arctic, which was larger at the end of summer in the assimilation experiment than in the experiment with no assimilation. In addition, the effects of individual analysis updates on ice concentration, sea surface temperature and sea surface salinity were illustrated for typical summer and winter situations, Figure 5. The winter and summer situations reveal that the strongest analysis updates take place close to the ice edge.

The quality of the EnKF model error estimate was investigated by looking at the innovation RMS–sequence of the experiment. It was seen that the error estimate of the model was too low when compared with a theoretical estimate. This effect was strongest at the beginning of the experiment, while the most realistic estimate of the model errors was seen for the summer season. A lot of the model error underestimate in winter is connected to a collapse of the ice concentration towards a single value, an effect which was most clearly seen in the Central Arctic Ocean. The reason for this collapse can be traced back to a too simple lead parametrization in the ice model. The effect of this is an underestimated ice concentration variance, and too low model error estimates in the EnKF. Throughout the time period examined in the assimilation experiment, the largest model ice concentration variance, and probably the most realistic error estimates, were found close to the ice edge.

In the EnKF the analysis is given as a variance–minimizing analysis, but often one finds that the maximum–likelihood estimate is a better estimate. If the sample distribution of ice concentration is Gaussian, one will find that the variance–minimizing and maximum–likelihood estimates are equal, so this is a property we would like the sample distributions to have. A box–whisker plot of the sample distribution of ice concentration revealed that the distribution showed signs of skewness, and having a larger range than what is expected from a Gaussian distribution. Evidence of non–Gaussian sample distributions were also found in the covariance between ice concentration and sea surface temperature. One should keep in mind that variables such as ice concentration always have non–Gaussian distribution since they are limited to the interval $[0, 1]$. An approach which is worthwhile considering is using suitable transforms, and then do the analysis on a variable which is closer to being Gaussian (*Bertino et al.*, submitted), but this approach has not been further considered here.

The covariance between ice concentration and variables of the ocean model were also considered. The covariances described by the ensemble statistics appear realistic, and we tried to connect them to model behavior. The temperature–ice concentration covariance was generally positive throughout the experiment, while the salinity–ice concentration covariance changed sign depending on physical mechanisms involved. It was suggested that the change of covariance sign was due to different effects of ice transport and local sea ice formation/melting, where transport resulted in a negative covariance, and local ice melt/freezing resulted in a positive covariance. In winter the transport effect can be seen, while in summer the local melt effect

is dominant. The spatial correlation functions displayed in section 5.3 revealed complex patterns emerging for the correlations between sea ice concentration in a fixed point in Davis Strait and surrounding model grid point temperature, salinity and ice concentration.

The present implementation with a coupled ice–ocean model reveals some of the strengths of the Ensemble Kalman Filter. It is relatively easy to implement, only requiring a number of ensemble members to be run in parallel, and there is no need for computing the adjoint equations. This makes it especially suited in the coupled model context, where the different models can have different temporal and spatial scales. The computational demands of the EnKF can be large, but in the current application, it is felt that the complexity of the ensemble statistics favors the EnKF in place of simpler, less computationally expensive methods.

While this study focuses on the effect of assimilating ice concentration, it is understood that it will be beneficial to include data sets which directly describe the ocean surface, such as sea surface temperature and sea level anomalies (*Haugen and Evensen*, 2002). Although the correlation function between ice concentration and sea surface temperature extended to some distance beyond the ice edge, Figure 12(b), the ice concentration "influence" is still limited when compared with the total model domain. The inclusion of sea surface temperature in the assimilation will produce a more direct control of the ocean properties, and cover a larger area than what can be achieved with assimilation of ice concentration alone. Care must be taken, however, as the ice concentration and sea surface properties are likely to be dependent close to the ice edge. This was clearly demonstrated for sea surface temperature in Figure 12(b).

To our knowledge there are no published results of studies using the Ensemble Kalman Filter to assimilate ice concentration in coupled sea ice/ocean models. Our main conclusion is that assimilation of ice concentration is a viable way of controlling the ice cover in a coupled ice–ocean model. In the current implementation the ice model is not able to correct the generally underestimated ice thickness of the model, but it gives a consistent update of the properties of the modeled ocean surface when ice is introduced or removed.

References

Aagaard, K., and E. C. Carmack, The role of sea ice and other fresh water in the arctic circulation, *J. Geophys. Res.*, *94*, 14,485–14,498, 1989.

Anderson, J. L., An ensemble adjustment kalman filter for data assimilation, *Mon. Wea. Rev.*, *129*, 2884–2903, 2001.

Belkin, I. M., S. Levitus, J. Antonov, and S.-A. Malmberg, Great salinity anomalies in the North Atlantic, *Progr. Oceanogr.*, *41*, 1–68, 1998.

Bentsen, M., G. Evensen, H. Drange, and A. D. Jenkins, Coordinate transform on a sphere using conformal mapping, *Mon. Wea. Rev.*, *127*, 2733–2740, 1999.

Bertino, L., G. Evensen, and H. Wackernagel, Sequential data assimilation techniques in oceanography, *Int. Stat. Rev.*, submitted.

Bleck, R., An oceanic circulation model framed in hybrid isopycnic–cartesian coordinates, *Ocean Modelling*, *1*, 55–88, 2002.

Bleck, R., and L. Smith, A wind-driven isopycnic coordinate model of the north and equatorial atlantic ocean. 1. model development and supporting experiments, *J. Phys. Oceanogr.*, *95*, 3273–3285, 1990.

Burgers, G., P. J. van Leeuwen, and G. Evensen, Analysis scheme in the Ensemble Kalman Filter, *Mon. Wea. Rev.*, *126*, 1719–1724, 1998.

Chapman, D. C., and R. C. Beardsley, On the origin of shelf water in the Middle Atlantic Bight, *J. Phys. Oceanogr.*, *19*, 384–391, 1989.

Dickson, R. R., J. Meincke, S. A. Malmber, and A. J. Lee, The "great salinity anomaly" in th northern atlantic 1968–1982, *Progr. Oceanogr.*, *20*, 103–151, 1988.

Drange, H., and K. Simonsen, Formulation of air-sea fluxes in ESOP2 version of MICOM, *Tech. Rep. 125*, Nansen Environmental and Remote Sensing Center, Bergen, Norway, 1996.

Dümenil, L., K. Isele, H. J. Liebscher, U. Schröder, and K. Wilke, Discharge data from 50 selected rivers for GCM validation, *Tech. Rep. 100*, Max Planck Institute, 1993.

Emery, W. J., C. Fowler, and J. Maslanik, Arctic sea ice concentration from special sensor microwave imager and advanced very high resolution radiometer satellite data, *J. Geophys. Res.*, *99*, 18,329–18,342, 1994.

Evensen, G., Using the extended Kalman filter with a multilayer quasi–geostrophic ocean model, *J. Geophys. Res.*, *97*, 17,905–17,924, 1992.

Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162, 1994.

Gaspar, P., Y. Gregories, and J.-M. Lefevre, A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: Tests at station Papa and long-term upper ocean study site, *J. Geophys. Res.*, *95*, 16,179–16,193, 1990.

Haugen, V. E. J., and G. Evensen, Assimilation of sla and sst data into an ogcm for the indian ocean, *Ocean Dynamics*, *52*, 133–151, 2002.

Hibler, W. D., III, A dynamic thermodynamic sea ice model, *J. Geophys. Res.*, *84*, 815–846, 1979.

Holland, D. M., and A. Jenkins, Modelling thermodynamic ice-ocean interactions at the base of an ice shelf, *J. Phys. Oceanogr.*, *29*, 1787–1800, 1999.

Holland, M. M., C. M. Bitz, M. Eby, and A. J. Weaver, The role of ice–ocean interactions in the variability of the north atlantic thermohaline circulation, *J. Clim.*, *14*, 656–675, 2001.

Houtekamer, P. L., and H. L. Mitchell, Data assimilation using an ensemble Kalman filter technique, *Mon. Wea. Rev.*, *126*, 796–811, 1998.

Hunke, E. C., and J. K. Dukowicz, An elastic-viscous-plastic model for sea ice dynamics, *J. Phys. Oceanogr.*, *27*, 1849–1867, 1997.

Jazwinski, A. H., *Stochastic processes and filtering theory*, Academic Press, 1970.

Kwok, R., Sea ice concentration estimates from satellite passive microwave radiometry and openings from SAR ice motion, *Geophys. Research Letters*, *29*, 10.1029/2002GL014,787, 2002.

Large, W. C., J. C. McWilliams, and S. C. Doney, Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parametrization, *Rev. Geophys.*, *32*, 363–403, 1994.

Legates, D., and C. Willmott, Mean seasonal and spatial variability in gauge–corrected global precipitation, *Int. J. Climatol.*, *10*, 110–127, 1990.

Levitus, S., and T. P. Boyer, *World Ocean Atlas 1994 Volume 4: Temperature*, NOAA Atlas NESDIS 4, Washington, D.C., 1994.

Levitus, S., R. Burgett, and T. P. Boyer, *World Ocean Atlas 1994 Volume 3: Salinity*, NOAA Atlas NESDIS 3, Washington, D.C., 1994.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, Ensemble size, Balance, and Model–Error Representation in an Ensemble Kalman Filter, *Mon. Wea. Rev.*, *130*, 2791–2808, 2002.

Montgomery, R. B., Circulation in upper layers of southern North Atlantic deduced with use of isentropic analysis, in *Papers in Phys. Oceanogr. and Meteorol.*, vol. VI, p. 55 pp., Woods Hole Oceanogr. Inst., Woods Hole, Mass., 1938.

Natvik, L.-J., and G. Evensen, Assimilation of ocean colour data into a biochemical model of the north atlantic. part 1. data assimilation experiments, *J. Mar. Sys.*, in press.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, Hydrologic data assimilation with the Ensemble Kalman Filter, *Mon. Wea. Rev.*, *1*, 103–114, 2002.

Rudels, B., H. J. Friedrich, and D. Quadfasel, The Arctic circumpolar boundary current, *Deep-Sea Res. II*, *46*, 1023–1062, 1999.

Semtner, J., A. J., A model for the thermodynamic growth of sea ice in numerical investigations of climate, *J. Phys. Oceanogr.*, *6*, 379–389, 1976.

Slutz, R., S. L. J. Hiscox, S. Woodruff, R. Jenne, D. Joseph, P. Steurer, and J. Elms, *Comprehensive Ocean-Atmosphere Data Data Set; Release 1,NTIS PB86-105723*, NOAA Environmental Research Laboratories, Climate Research Program, Boulder, Col., 1985.

Svendsen, E., K. Kloster, B. Farrelly, O. M. Johannessen, J. A. Johannessen, W. J. Campbell, P. Gloersen, D. Cavalieri, and C. Mätzler, Norwegian remote sensing experiment: Evaluation of the nimbus 7 scanning multichannel microwave radiometer for sea ice research, *J. Geophys. Res.*, *88*, 2781–2791, 1983.

Thorndike, A. S., and R. Colony, Sea ice motion in response to geostrophic winds, *J. Geophys. Res.*, *87*, 5845–5852, 1982.

Whitaker, J. S., and T. M. Hamill, Ensemble data assimilation without perturbed observations, *Mon. Wea. Rev.*, *130*, 1913–1924, 2002.