



Norwegian  
Meteorological Institute  
*met.no*

*met.no* report

Report no. 2/2009

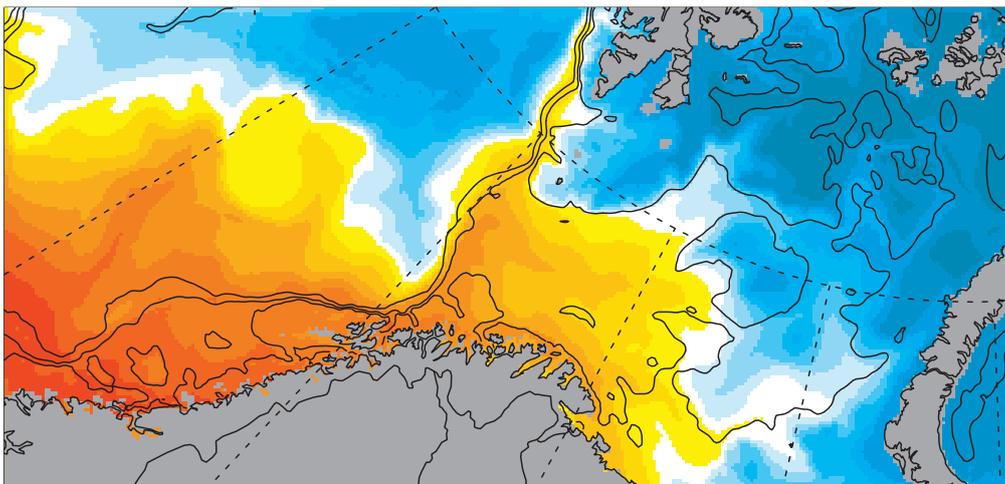
Oceanography

ISSN: 1503-8025

Oslo, February 11, 2009

# Validation of an ocean model ensemble for the Barents Sea and the northeastern Nordic Seas

Arne Melsom and Ingerid Fossum







<b>Number</b> 2/2009	<b>Subject</b> Oceanography	<b>Date</b> February 11, 2009	<b>Classification</b> <input checked="" type="checkbox"/> Open <input type="checkbox"/> Restricted <input type="checkbox"/> Confidential	<b>ISSN</b> 1503-8025
-------------------------	--------------------------------	----------------------------------	---	--------------------------

**Title**

Validation of an ocean model ensemble for the Barents Sea and the northeastern Nordic Seas

**Authors**

Arne Melsom and Ingerid Fossum

**Client(s)**

The Research Council of Norway

**Client reference**

173487/S40

**Abstract**

Results from the validation of an ocean circulation model ensemble are presented. The simulation is performed for a domain that covers the Barents Sea and adjacent ocean regions, with an eddy-permitting resolution of 4 km. The ensemble is generated from “perturbed” atmospheric forcing fields. Model results are validated using data for hydrography, and observations of sea levels from tide gauges. The validation methods that are employed range from basic statistics (bias, rms-errors) to probability density functions and rank histograms. We find that the geostrophic currents flow parallel to the bottom slope as expected, but that the low-frequency variability of sea levels performs poorly. Further, the model is salt deprived, particularly in the Atlantic Water masses, while errors in temperature are moderate. The variability among ensemble members is found to be significantly lower than the observed variability.

**Keywords**

Ocean circulation, ensemble, validation, Barents Sea

**Disiplinary signature**

L.P.R. (sign.)

\_\_\_\_\_  
Lars Petter Røed, Section for Oceanography

**Responsible signature**

Ø.H. (sign)

\_\_\_\_\_  
Øystein Hov, Head of Research



## 1. Introduction

The Barents Sea and the Norwegian continental shelf in the Nordic Seas contain some of the most valuable fish stocks of the Atlantic sector of the world oceans. Another activity in this region which involves large economic investments and profits, is the offshore exploration of hydrocarbons that is presently accelerating in the Barents Sea. Lately, ship traffic has been increasing in volume in this region due to export of Siberian oil from Russian terminals. Should the retreat of the summer ice in the Arctic continue, the prospect for ship traffic from Asia to Europe and the east coast of North America via the Arctic will become increasingly attractive (Hjermann et al. (2007)).

Relatively warm and salty Atlantic Water flows toward the Arctic along the continental slope off Norway as the Norwegian Atlantic Current (NAC). When the Bear Island Channel is encountered by the NAC, it separates into two branches. One branch continues to the north as the West Spitsbergen Current, while the remaining branch flows into the Barents Sea in the Bear Island Channel. Schauer et al. (2002) have explained the special role of the Barents Sea for the transformation of water masses in the Arctic region by noting that it is the deepest of the Arctic shelf seas, it receives the highest salinity water masses of the shelf seas, there are polynyas in various regions of the Barents Sea, and a compensating flow of about 2 Sv out of the Barents Sea to the east. The transport value is estimated by Loeng et al. (1993), and is based on ocean current measurements.

In recent years, the petroleum industry has started to examine the prospects for commercial exploration of oil and natural gas off the Lofoten archipelago and northward along the coast of Norway and into the Barents Sea. In relation to these plans, an environmental impact assessment (EIA) was produced, including a part that describes the statistics of advection and spreading of oil in the ocean after an accidental spill (Johansen et al. (2003)). Regions that may be affected by oil pollution are determined by simulated drift trajectories and weathering of the oil. Traditionally, trajectories were computed using an empirical relation between surface winds and the ocean currents with which the oil spill is advected (Martinsen (1982)). In more recent work, the drift trajectories have been computed with results from ocean circulation models (Martinsen et al. (1994), Wettre et al. (2001)). Available computer resources now allow for such simulations with a grid mesh of  $\leq 5$  km that resolves features such as filaments and eddies. Results from the traditional and the more recent approach were both used by Johansen et al. (2003).

Results from ocean circulation simulations with an eddy resolving resolution are subject to flow instabilities that are of a stochastic nature. Thus, when applying such results based on historical atmospheric conditions in an EIA, the study should be conducted with an approach that incorporates the ocean circulation statistics due to the atmospheric forcing. Unfortunately, the EIA in question applies a deterministic approach in the analyses of results from both the traditional method and for results from the eddy resolving simulations.

Presently, a 10-member, 10-year ocean circulation ensemble has been run in order to examine the consequences of this short-coming. This report is a validation of results from the first year of the 10 year period.

The following section contains information about the numerical model system, the surface and lateral forcing, assimilation and the construction of the model ensemble. In Section 3, the

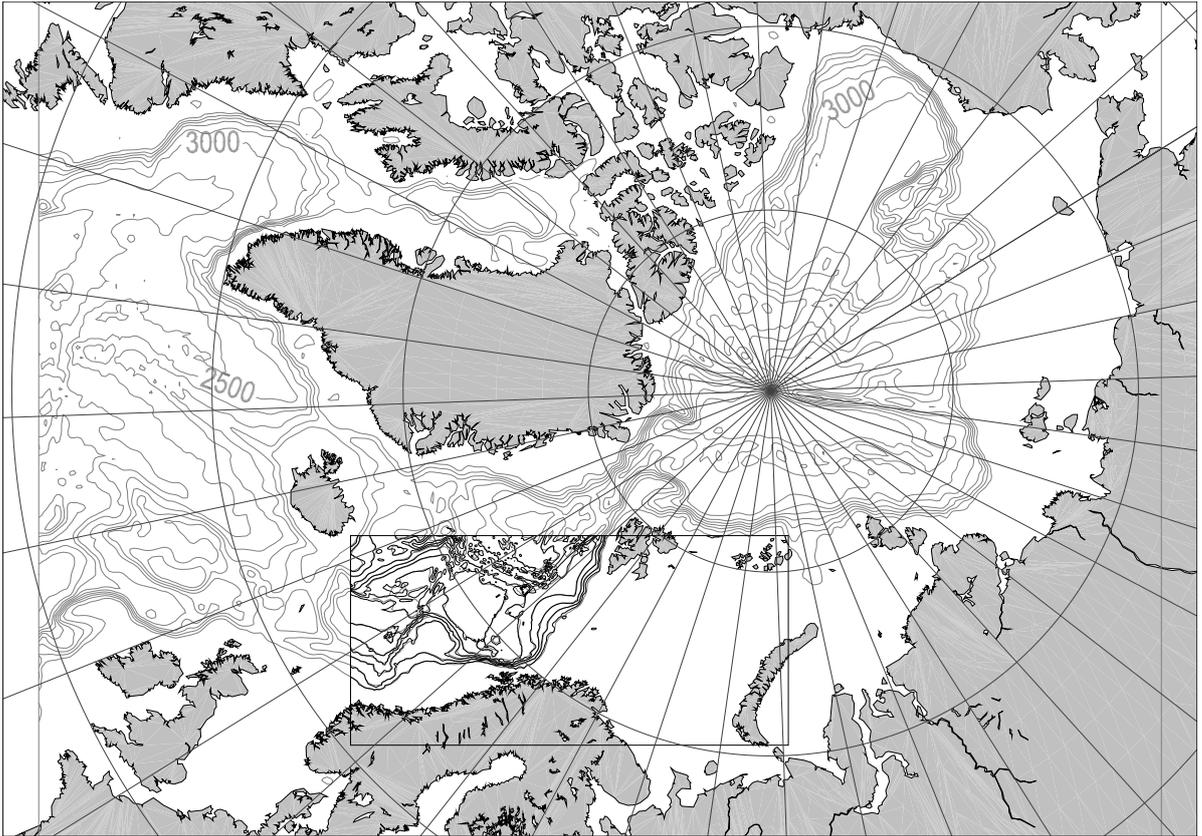


Figure 1: The nested model system. Grey lines are isobaths for every 500 m contour of the coarse mesh model (20 km mesh size). This domain covers the whole Arctic Sea and North Atlantic adjacent seas. Nested into this model is a 4 km mesh size model that covers the Barents Sea and most of the Norwegian Sea, with isobaths in black.

observations that are used in the model validation are presented. The stage is then set for the model validation, which is provided in Section 4, with details from various validation methods given in separate subsections. The final section discusses the validation, and conclusions therefrom are drawn.

## 2. Numerical experiment

The coupled sea-ice model system consists of a physical oceanographic model (MI-POM) and an ice model (MI-IM). MI-POM is Norwegian Meteorological Institute's version of the Princeton Ocean Model, and it is documented by Blumberg and Mellor (1987), Engedahl (1995) and Engedahl et al. (2001). It is a fully three-dimensional, primitive equation, terrain-following ( $\sigma$ -coordinate) model. The dynamic-thermodynamic sea-ice model MI-IM is described by Røed and Debernard (2004).

A model covering the Norwegian and Barents Seas with 4 km mesh size and 21  $\sigma$ -levels

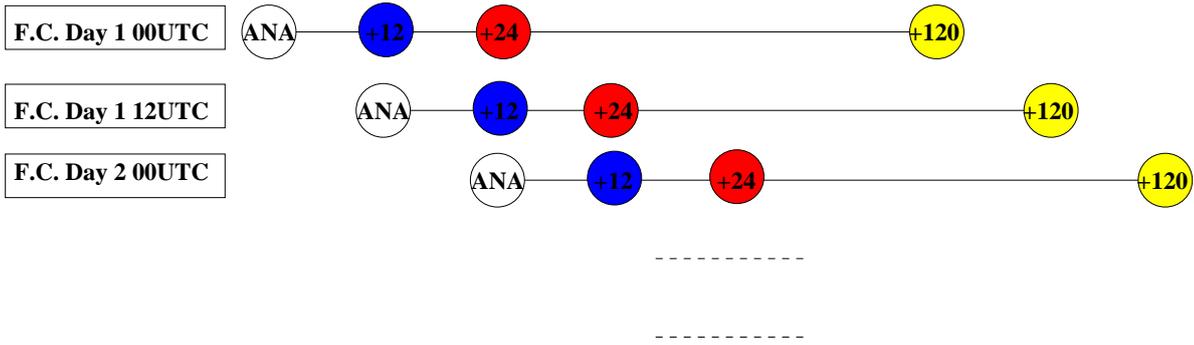


Figure 2: The uppermost line represents the 5 first days of the 10 day forecast produced at ECMWF for an arbitrary day at time 00 UTC. The white circle is the analysis time, the blue, red and yellow circles represent the 12, 24 and 120 hours prognoses, respectively. (The prognoses for every 12th hour between 24 and 120 hours are not shown.) The second line represent the corresponding forecast issued by ECMWF the same day at time 12 UTC, while the third line corresponds to the forecast produced the next day at 00 UTC. The atmospheric forcing for the first ensemble member is created by extracting all the 12 hours prognoses (blue circles), the second ensemble member is forced by the 24 hours prognoses (red circles), and so on.

(NOR-BAR 4) is nested into a coarse mesh model (20 km mesh size and 21  $\sigma$ -levels) covering the whole Arctic Sea and North Atlantic adjacent seas (ARC-N.ATL 20) (Figure 1).

MI-POM was initialized 1981-01-01 with ocean climatology as described by Engedahl et al. (1997, 1998), in the ARC-N.ATL 20 domain. The MI-IM initial fields were set to a motionless ice cover with 2 m thickness and a concentration of 75% in regions where the climatological sea surface temperature (SST) is below 0°C.

A simulation was then performed for 1981-1986 with atmospheric forcing fields from the ERA40 dataset obtained from the European Center for Medium-Range Weather Forecasts (ECMWF). To assimilate SST and ice concentration, a nudging method based on Albretsen and Burud (2006) was applied. The fields that were use in the assimilation were SST from the ERA40 dataset, and merged ice concentration fields from the Ice Sevice at Norwegian Meteorological Institute and from the ERA40 dataset. No tidal forcing was applied in ARC-N.ATL 20.

The nested NOR-BAR 4 model was then initialised on 1984-01-01, by interpolation of ARC-N.ATL 20 results onto its 4 km mesh. NOR-BAR 4 was spun up during 1984 and 1985 using the exteernal forcing as described above, as well as assimilation of sea surface temperature and ice concentration. At the lateral boundaries tidal forcing (8 constituents) were prescribed together with the simulated fields from ARC-N.ATL 20. The resulting fields for 1986-01-01 were used as initial conditions for a set of 10 ensemble simulations for 1986.

The ensemble was constructed by using forcing fields from the 10 day forecast that are produced twice a day at ECMWF. As illustrated in Figure 2 one set of atmospheric forcing fields was created by extracting all the 12 hours prognoses from the ECMWF forecasts. A second set of atmospheric forcing fields was generated by extracting all the 24 hours forecasts,

## mean SSH, 1986

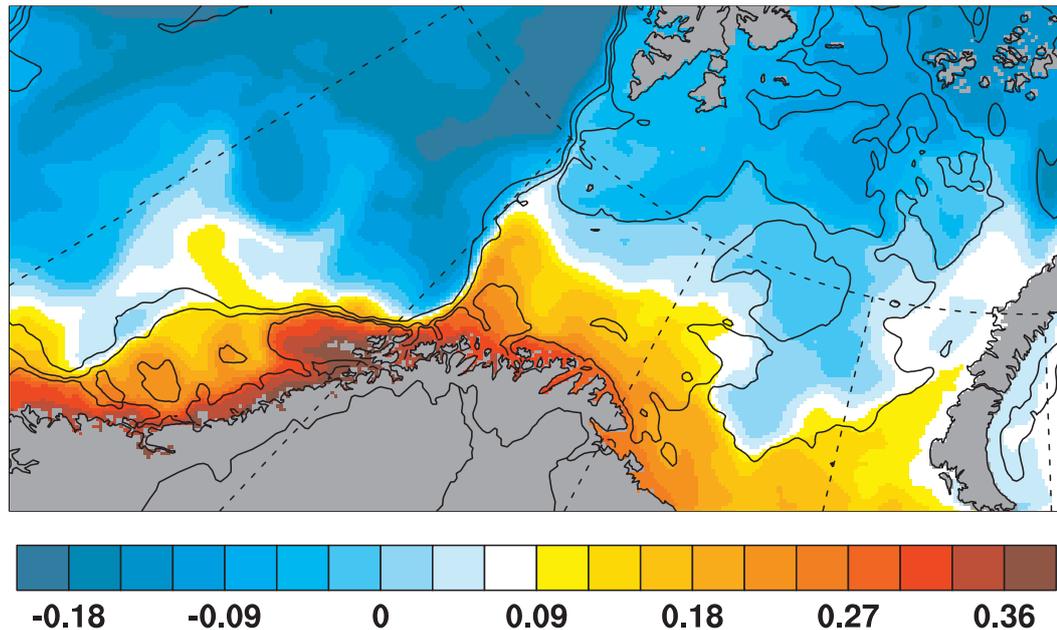


Figure 3: Mean sea surface height averaged over all ensemble members, and one year (1986). Label values are in m. Contour lines are shown for the depth contours 250 m, 500 m, and 1000 m.

a third by extracting all the 36 hours forecasts and so on. To limit the amount computer time that was required, the ensemble size was restricted to 10 members. Hence, only results from the 5 first days of the atmospheric forecast was used. This results in 10 different atmospheric forcing files for one year, which is used to generated the 10 different ensemble members. The atmospheric forcing is the only difference between the ensemble members.

Instantaneous results ("snaphots") from the nested Barents Sea model ensemble were stored with a temporal resolution of two hours. The mean sea surface height (SSH) from 1986 based on this output, averaged over all 10 ensemble members, is displayed in Figure 3. We note that strong gradients of mean SSH is found near the continental slopes, and some places also in the vicinity of other topographic features.

### 3. Observations

Hydrographic data along fixed cruise tracks and cast positions were made available by the Institute of Marine Research (IMR; Kangas et al. (2006)). The data have been subjected to a quality assurance process at IMR, the individual observations have been flagged according to the standard defined by The Integrated Global Ocean Services System (IGOSS). Observations

were made each year in the period 1980 – 2006, and details about the sections and cruise frequencies are listed in Table 1. The cruise tracks are displayed in Figure 4.

Table 1: List of IMR’s cruise tracks from which hydrographic data are used in this study. Note that the values given in the table correspond to the average coverage in space and time, individual years and cruises may differ.

section	cruises/year	month(s)	stations/cruise
Gimsøy northwest	2	3/4, 7/8	14
Bear Island west	1	9/10/11	13
Fugløy – Bear Island	6	all	20
Vardø north	4	all	22

The observations were made as CTD casts, and are available with a vertical resolution of 5 m. In this study, we restrict the analysis to the depths at which model results are stored, which are 0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, and 1000 m. We refrain from removing the seasonal cycle in the observations, since the seasonal cycle is either poorly resolved in the data or not resolved at all.

The CTD casts sometimes miss the target positions by several kilometers. Thus, whenever the processing involves procedures that require each station to be treated separately, observations that have missed their target by more a prescribed distance are discarded. This limit was set to 2 nm (nautical miles) for all stations in Figure 4, except the five westernmost positions in the Bear Island west section and the four northernmost positions in the Vardø north section. At these nine stations, the limit was set to 4 nm, as recommended by Kangas et al. (2006).

The model results have also been validated using hydrographic data from the World Ocean Database (WOD 2005; Conkright et al. (2002)). Only data which had no error flag set have been considered in the present study. Whenever this data set was employed, CTD cast duplicates with the IMR data were removed prior to the analysis.

Finally, hourly sea level data from the tide gauge stations at Rørvik and Vardø were retrieved from the Global Sea Level Observing System (GLOSS; Woodworth et al. (2003)). The positions of these stations are 11° 15' E, 64° 52' N and 31° 6' E, 70° 20' N, respectively, and they are displayed in Figure 4. During some shorter intervals, sea level observations were not available. These intervals are listed in Appendix A.

## 4. Model validation

### 4.1. Bias and standard deviation

Bias and standard deviations provide crude, over-all measures that are relevant for model validation purposes. The bias of property  $p$  from the ensemble is defined as

$$bias_e = \frac{1}{MN} \sum_{n=1}^N \sum_{m=1}^M (p_{m,n}^{mod} - p_n^{obs}) \quad (1)$$

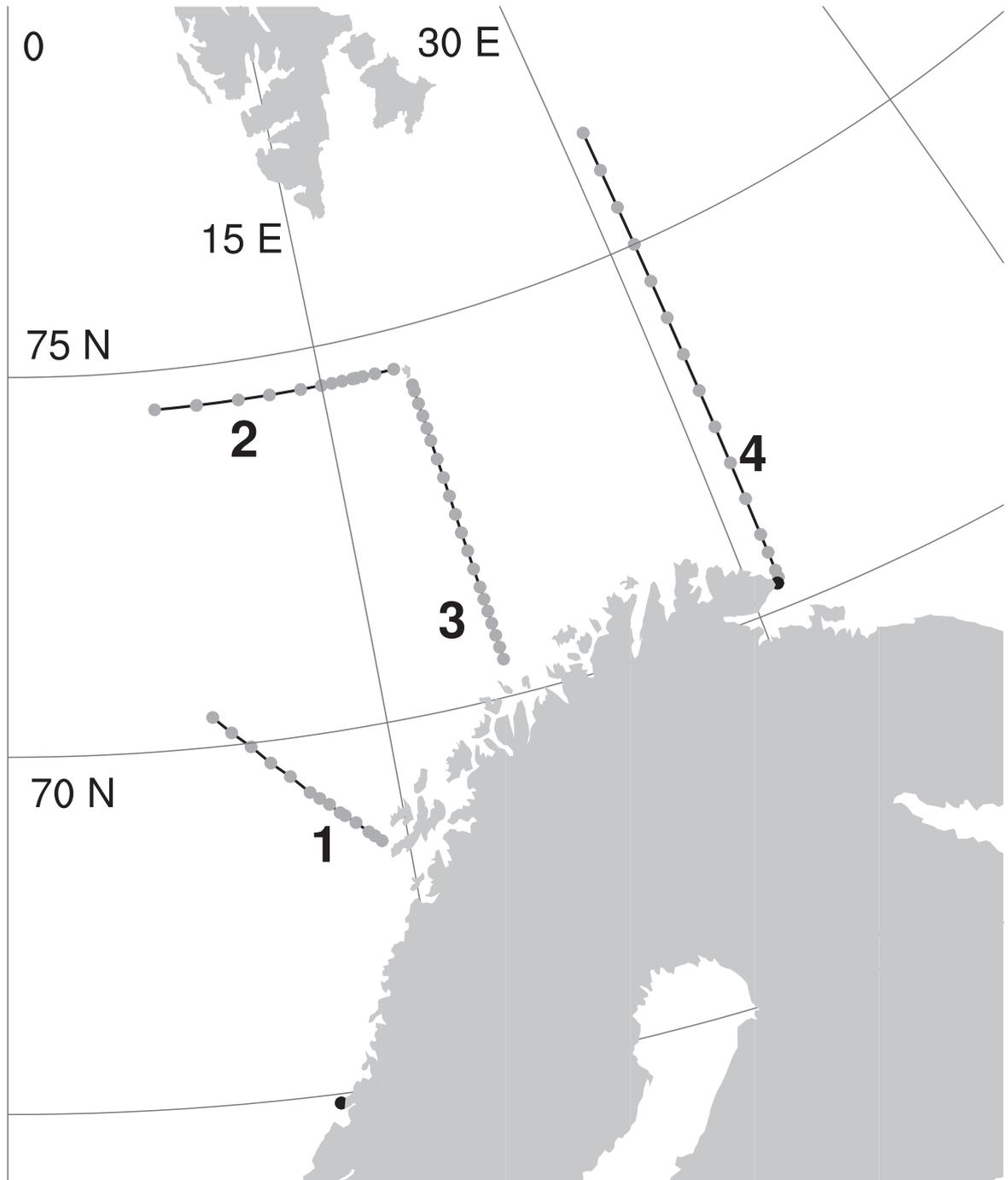


Figure 4: Cruise tracks along which the CTD casts have been made. The sections are **1**: Gimsøy northwest, **2**: Bear Island west, **3**: Fugløy – Bear Island **4**: Vardø north. The filled grey circles correspond to the positions of the fixed cast sites. Bear Island is where sections **2** and **3** meet. The positions of the tide gauge stations at Rørvik and Vardø are also shown as filled black circles (Rørvik is the southernmost of the two stations, while Vardø is at the southern end of cruise track **4**).

Table 2: The model bias and standard deviation when compared to data from the various hydrographic sections, and with the World Ocean Database hydrography. Only values for the levels 0, 10, 20 and 30 m have been considered here. Values given under  $bias_e$  and  $\sigma_{\Delta e}$  are biases and standard deviations of the differences between model values and observations for the entire ensemble, respectively. Biases are positive when the model values exceed the observations. Values in the  $\overline{\sigma_{\Delta 1-\Delta 10}}$  column are averages of the standard deviations when each of the ten ensemble members is treated separately. Temperature and salinity values are in K and PSU, respectively.

data set	temperature			salinity		
	$bias_e$	$\sigma_{\Delta e}$	$\overline{\sigma_{\Delta 1-\Delta 10}}$	$bias_e$	$\sigma_{\Delta e}$	$\overline{\sigma_{\Delta 1-\Delta 10}}$
Gimsøy northwest	-0.403	1.717	1.728	-0.129	0.347	0.350
Bear Island west	0.259	1.789	1.801	-0.454	0.303	0.306
Fugløy – Bear Island	0.714	1.402	1.414	-0.241	0.248	0.249
Vardø north	0.999	1.276	1.291	-0.211	0.169	0.176
World Ocean Database	0.648	1.533	1.548	-0.264	0.503	0.505

where subscripts  $n$  and  $m$  denote observation no. and ensemble member no., respectively. The subscript  $e$  is used for quantities when referring to the entire ensemble. The standard deviation of this property for an individual ensemble member is the square root of the variance, defined as

$$\sigma_{\Delta m}^2 = \frac{1}{N} \sum_{n=1}^N [(p_{m,n}^{mod} - p_n^{obs}) - (\overline{p_m^{mod}} - \overline{p^{obs}})]^2 \quad (2)$$

where the overbar denotes the arithmetic mean value. The standard deviation for the entire ensemble becomes the square root of

$$\sigma_{\Delta e}^2 = \frac{1}{MN} \sum_{n=1}^N \sum_{m=1}^M [(p_{m,n}^{mod} - p_n^{obs}) - (\overline{p_e^{mod}} - \overline{p^{obs}})]^2 \quad (3)$$

Here, and throughout this section, the validation is restricted to results from the NOR-BAR 4 ensemble simulation.

The results for temperature in Table 2 that are based on the cruise data, reveal that the eastern transects ‘‘Fugløy – Bear Island’’ and ‘‘Vardø north’’ are warmer in the model than in the observations, by almost 1 K. The differences for the two western transects are smaller in magnitude. The standard deviations of the model results vs. observation as defined by (3) are between 1 K and 2 K, largest in the western transects.

The corresponding results for salinity reveal that there is a substantial negative bias in the model results. This is particularly the case for the ‘‘Bear Island west’’ transect, which cuts across the northern branch of the Norwegian Atlantic Current. Here the observations are saltier than the model results by 0.45 PSU. As for the standard deviations as defined by (3), the largest values are to the west, similar to the results for temperature.

With one exception, the comparisons of model results and the WOD 2005 data are somewhere in between the corresponding results for the various transects. This is unsurprising, since the WOD 2005 data come from positions that are spread throughout the domain. The

Table 3: The model bias and standard deviation when compared to data from the sea level stations at Rørvik and Vardø. Numbers in the “de-tided data” column refers to values after the two main tidal constituents at these stations (M2 and K1) were removed. Values are given in m, see the text for details.

sea level station	raw data			de-tided data		
	$\sigma_{obs}$	$\sigma_{\Delta e}$	$\overline{\sigma_{\Delta 1 - \Delta 10}}$	$\sigma_{obs}$	$\sigma_{\Delta e}$	$\overline{\sigma_{\Delta 1 - \Delta 10}}$
Rørvik	0.609	0.211	0.215	0.194	0.192	0.196
Vardø	0.758	0.250	0.253	0.156	0.231	0.234

exception is the standard deviation of the differences between model results and observations for salinity. This standard deviation is much larger when the WOD 2005 is used in the analysis. We have not been able to identify why this is so.

We generally find that the model results from the upper 30 m are warmer and less salty than the observations. The density for the corresponding oceanic layer is thus too low in the model.

When this validation is carried out for each ensemble member separately, we find very small differences from validation that is based on the entire ensemble. This is evident from the results displayed under the columns  $\sigma_{\Delta e}$  and  $\overline{\sigma_{\Delta 1 - \Delta 10}}$  in Table 2, which are nearly identical. The implications of the low member-to-member variability will be investigated in subsection 4.5.

Results for the validation of the sea surface height vs. data from the two tide gauge records at Rørvik and Vardø are given in Table 3. In the rightmost columns labelled “de-tided data”, the M2 and K1 tidal constituents with periods of 12.42 h and 23.93 h, respectively, were filtered out prior to the analysis.

The standard deviations under columns labelled  $\sigma_{obs}$  are values for the observational time series. As expected, these are much higher before filtering the two tidal constituents. On the other hand, the standard deviation of the differences between the observations and ensemble results as defined by (3), is only slightly reduced after de-tiding. This suggests that the phase and amplitudes of the model tides are nearly the same as observed.

Furthermore, we find that at Rørvik, the variance of the difference between the ensemble and the observations are nearly identical to the variance of the observational time series. Hence, the ensemble is not more similar to the observations than a time series that represents a pure M2 + K1 tide: The de-tided M2 + K1 time series is 0, yielding an error identical to the de-tided  $\sigma_{obs}$ . The model sea surface height results for Vardø reveal that a pure M2 + K1 tide is closer to the sea level observations than the ensemble. Finally, we note that comparisons between observations and separate ensemble members differs very little from the analysis of the entire ensemble.

## 4.2. Cost functions

In optimization problems, a cost function is generally a function for which extrema are sought. Here, we define the cost function as the difference between a simulation and the corresponding observations, so small values correspond to a simulation that is close to the observations. Further, since the variability is not uniform in space, we investigate differences measured in

Table 4: The cost function  $Dp_\sigma$  for temperature and salinity for the four IMR hydrography transects, for selected levels (0, 10, 20 and 50 m). Values are differences between model results and observations, measured in standard deviations. Further, values are averages for all stations in the relevant transect, and model values are averages from all ensemble members. All values have been computed based on the entire year, except for the row “Fugløy – Bear Is.<sup>8–10</sup>” which is based on data and results for the months August through October only. See the text for details.

data set	#obs.	temperature				salinity			
		0m	10m	20m	50m	0m	10m	20m	50m
Gimsøy northwest	4	0.71	0.81	0.89	1.62	3.61	3.02	3.56	5.97
Bear Island west	14	1.14	0.73	0.82	2.37	20.22	8.80	7.96	6.17
Fugløy – Bear Is.	99	0.88	0.77	0.33	0.61	1.82	1.82	1.92	2.95
Fugløy – Bear Is. <sup>8–10</sup>	39	1.28	1.28	0.63	1.29	1.44	1.36	1.45	1.92
Vardø north	39	0.92	0.82	0.43	0.59	1.31	1.32	1.42	3.01

standard deviations, based on historical observations of the relevant variables.

Let  $p_n^{obs}(s, z)$  be the  $n$ th observation of the property  $p$  at station  $s$  and depth  $z$ . We denote the local variance of observations of  $p$  by  $\sigma_p(s, z)^2$ , and let  $p_n^{mod}(s, z)$  be the corresponding mean value of the model ensemble. Next, we introduce the cost function  $Dp_\sigma$  as

$$Dp_\sigma = \frac{1}{N} \sum_{n=1}^N \frac{|p_n^{obs} - p_n^{mod}|}{\sigma_p} \quad (4)$$

In order to calculate the standard deviation  $\sigma_p$ , we used all observations from the four transects from the period 1980 – 2006. First, we computed standard deviation values at each position in depth-transection station space. Then, the cost function was computed based on observations and model results from 1986, applying weights that correspond to the distance from one station to the next. The number of observations listed in the leftmost column in Table 4 is the number that is available for one depth, distributed among the various stations in each transect. Note that only four CTD casts were available for the “Gimsøy northwest” section from 1986.

The observations from the various transects are not uniform over the year. Hence, we are not able to determine the seasonal variability of the hydrography. Since this variability is likely to significantly contribute to the standard deviation values, we also include an analysis of the data from the “Fugløy – Bear Island” transect for the months from August through October. The period and transect was chosen due to the relatively good coverage of available data.

We find that the values of cost function for salinity is greater than the corresponding values for temperature, for all transects and at all depths listed in Table 4. This was expected, since the air – sea heat fluxes were assimilated in the simulations, while no assimilation was performed for properties relevant to salinity. We must also point out that none of the temperature data from the IMR hydrography was used in relation to the heat flux assimilation, so in the present context, the IMR hydrography constitutes an independent data set.

The all-year cost function values for temperature that are listed in Table 4 are in the range 0.3 – 1.3 in the upper 20 m, and we can’t see an obvious geographical pattern. With the

exception of the “Gimsøy northwest” transect, the value has a minimum between the surface and the 50 m level. Values from the “Fugløy – Bear Island” transect that are based on the August through October results are higher than the all-year results, by 50-100%.

The results for salinity are completely different. The very high cost function values in the “Bear Island west” transect caused by a combination of the large bias of 0.454 (Table 2) and a small standard variation (0.16 in the upper 30 m, compared to *e.g.*, 0.24 in the “Vardø north” transect). This reflects the model’s poor performance with respect to reproducing the salty water masses of the Norwegian Atlantic Current (NAC). In fact, the cost function average for the upper 30 m of the “Bear Island west” transect varies from 0.06 for the station closest to Bear Island, to 22.5 for a station in the core of the NAC.

### 4.3. Probability density functions

The probability density function (p.d.f.) is a representation of the statistical distribution of a variable. It may be used for examining features such as skewed distributions and frequency of extreme events. In the present context, p.d.f.s are used in order to investigate similarities and differences between distributions from observations on the one hand, and model results on the other.

Model results were extracted at the positions of transect stations. Due to the modest amount of observations that are available from 1986, we calculate p.d.f.s from all available data from the transects “Bear Island west”, “Fugløy – Bear Island”, and “Vardø north”. For the two former data sets, we restrict the analysis to the period August-November, while the latter data set is analyzed for the period June-September.

Although the de-correlation time scales exhibit some spatial variability, it is mostly between 6 and 12 days in the model results. Hence, we have sampled model results every 10 days during the relevant four month periods. We discuss results from one of the ensemble members only (+072). Results from the various ensemble members generally give very similar p.d.f.s, exceptions to this is commented below.

We have computed p.d.f.s based on both observational data and model results. The normalization was performed separately for each station and depth. The p.d.f.s for observations and model results from the “Fugløy – Bear Island” transect are shown in Figures 5-8. Values along the top and bottom axes correspond to original and normalized values, respectively. The top axis is positioned so that the overall mean value is placed directly above the zero standard deviation mark on the bottom axis. Moreover, the interval of the top axis is set so that the span for one standard deviation matches the span of one standard deviation along the bottom axis.

p.d.f.s for temperature based on observations and model results from the upper 30 m are displayed in Figure 5 and Figure 6, respectively. We note that prior to normalization, the distribution of the data have a peak at 7-8°C, while the model results peak at around 6-7°C (gray curves). Moreover, the data distribution are left-skewed, whereas the model results are right-skewed. Overall, the model has a warm bias, as also seen from the all-year analysis in Table 2.

There are also distinct differences between the black curves that display the normalized distributions. While the normalized observations follow the normal distribution quite closely, the model results have one peak just above one standard deviation, and another peak at about

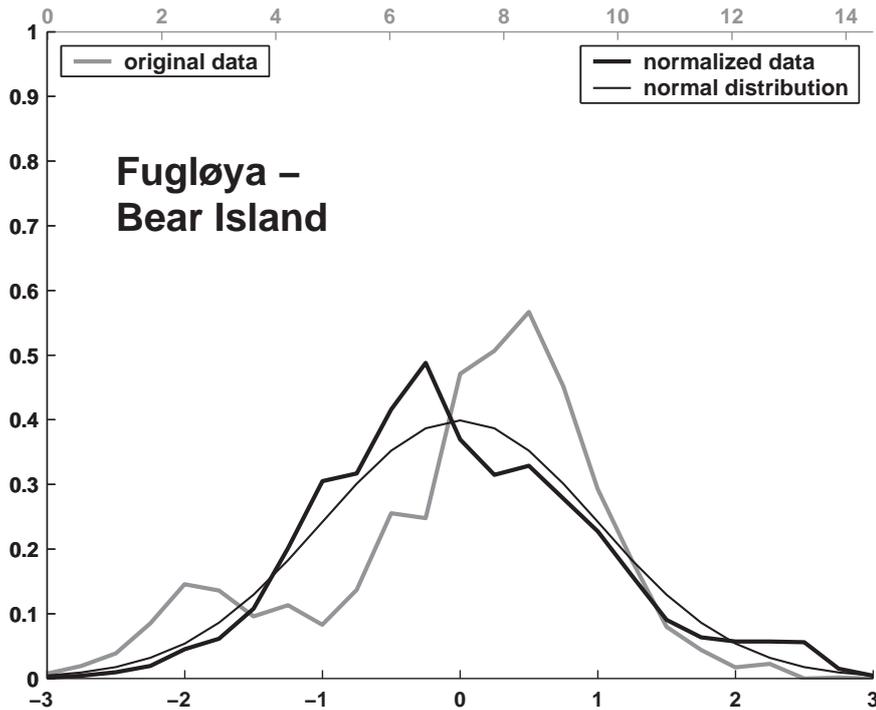


Figure 5: Probability density function for temperature in the upper 30 m, based on observations from the Fugløya – Bear Island transect. Computations were restricted to data from August through October. Numbers in gray along the top axis are values in  $^{\circ}\text{C}$ , for the thick gray curve. Values along the bottom axis are in units of standard deviation. The thick and thin black curves correspond to data and the normal distribution, respectively. See the text for details.

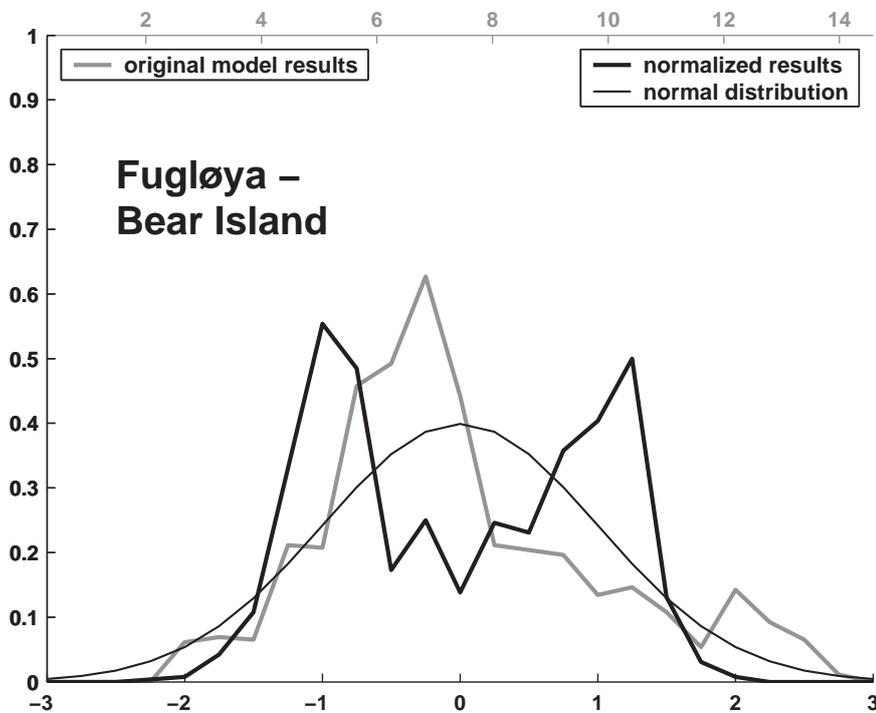


Figure 6: Same as Figure 5, but based on model results.

minus one standard deviations. This finding suggests that the temporal variability in the model results is too high locally. However, the kurtosis of the original model results compares well with the kurtosis of the corresponding observations (2.96 vs. 3.07).

The corresponding p.d.f.s for salinity in the upper 30 m of the “Fugløy – Bear Island” transect are displayed in Figure 7 and Figure 8. The shapes of the original distributions are strikingly similar. The original data and model results are both notably left-skewed. The normalized distributions are also slightly left-skewed, and are otherwise quite close to a normal distribution. The major difference that can be seen from these two figures, is that the standard deviation in the original data are markedly higher than the standard deviation of the model results (0.37 vs. 0.16).

Furthermore, the normalized distributions of the model results for salinity differ somewhat in the ensemble (not shown). One of the ensemble members have no obvious distribution peak, while the others have a two-peak structure. However, most of those that fall within the latter category exhibit one primary peak and one secondary peak, at a positive and a negative offset, respectively. Nevertheless, the variation in the skewness parameter and the kurtosis of the ensemble members is moderate. The skewness parameter is in the range from  $-1.52$  to  $-1.16$ , while the kurtosis is in the range from 5.2 to 6.7.

p.d.f.s for temperature from the “Bear Island west” transect are similar to the distributions from “Fugløy – Bear Island”, with a peak at higher values than the mean in the original observations, and a peak closer to the mean from the model results. Further, the normalized model results from the “Bear Island west” transect exhibit a two-peak structure which is not seen in the normalized data, again suggestive of too large temporal variability in the model results.

The corresponding p.d.f.s for salinity prior to normalization both have elevated peaks at about the level of 1 at a value higher than the mean. The normalized p.d.f.s are both similar to the normal distribution, although the model results have shaved off the maximum value, leaving a nearly flat distribution within the  $\pm 1$  standard deviation interval.

The “Vardø north” temperature p.d.f.s based on the original data and model results are both slightly right-skewed. The normalized p.d.f. from observations exhibit two peaks, whereas the corresponding p.d.f. based on model results are left-skewed with a moderate peak around the 0.5 level. The p.d.f.s for salinity is similar to those from the corresponding “Bear Island west” p.d.f.s, with the exceptions of the normalized data, which are left-skewed with a peak at a level of 0.6.

#### 4.4. Ensemble variability

This subsection is restricted to an analysis of the variability of the model ensemble, without reference to any observations. As such, this is not strictly part of a validation, but we include the topic here since it is relevant for interpretation of the rank analysis in the following subsection.

We intend to examine the variability that is related to the impact of weather on circulation, leaving out the processes that are driven by purely periodic forcing, namely tides and the sun’s elevation. Hence, we filter out the two main tidal constituents in the present region (M2 and K1), a sinusoidal representation of the seasonal variability, and subtract the local time mean. The phase lag of the sinusoid is determined as a function of space.

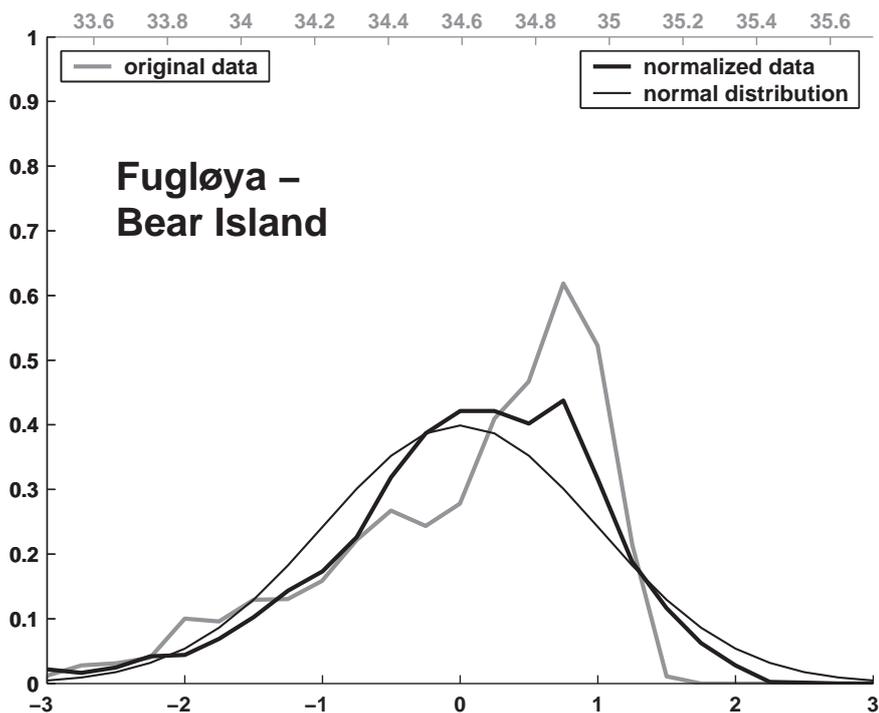


Figure 7: As Figure 5, but for observations of salinity. Values along the top axis are in PSU.

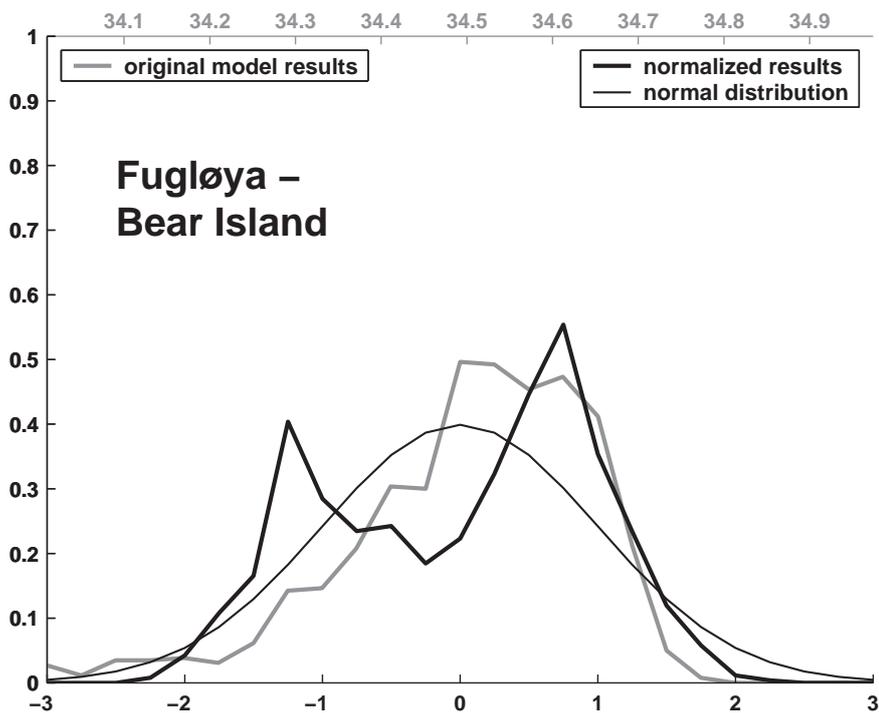


Figure 8: Same as Figure 7, but based on model results.

In order to quantify the role of synoptic atmospheric (weather) forcing and the role of oceanic flow instabilities, we make use of the method that was developed for this purpose by Metzger and Hurlburt (2001). The quantities that are involved in the analysis were described in detail by Melsom (2005), and we repeat the core of the method here. Let  $\eta_m(x, y, z, t)$  be a time series from ensemble member *no. m* which has been processed as described above. Each value can be written as the sum

$$\eta_m = \bar{\eta}^M + \eta'_m \quad (5)$$

where

$$\bar{\eta}^M = \frac{1}{M} \sum_{m=1}^M \eta_m \quad (6)$$

is the ensemble mean, and  $\eta'_m$  is the offset of ensemble member *no. m*. Now, the ensemble's mean square can be written

$$\overline{\eta^2}^M = \bar{\eta}^{M2} + \overline{\eta'^2}^M \quad (7)$$

Then, the variances of these three time series can be written

$$\sigma_{\eta}^2 = \sigma_{\bar{\eta}}^2 + \sigma_{\eta'}^2 \quad (8)$$

where we recognize  $\sigma_{\bar{\eta}}$  as the standard deviation of the ensemble mean time series that represents the deterministic impact due to the synoptic forcing. The fraction of the non-deterministic variability,  $r^{nd}$ , may thus be written

$$r^{nd} = \frac{\sigma_{\eta'}^2}{\sigma_{\eta}^2} \quad (9)$$

All values that are quoted below, are time averages over the year in question.

For construction of an ensemble in which flow instabilities were free to evolve independently in the various ensemble members, we perturbed the atmospheric forcing by applying results from different forecast times in the various realizations, as described in Section 2. Hence, differences in the results from the ensemble members are not entirely due to flow instabilities, they are also affected by differences in the forcing fields. Nevertheless, we shall refer to differences in the model results as being driven by flow instabilities. This topic will be discussed further in Section 5.

After the time series are filtered, the standard deviation of the ensemble mean time series theoretically are representations of the deterministic impact of weather on the ocean circulation. (In practice, the limited size of the ensemble reduces our results to an estimate of the deterministic impact.) The standard deviation for the offsets for the ensemble memberd for SSH,  $\sigma_{SSH'}$ , is displayed in the top panel of Figure 9. The largest values are found in the central Barents Sea and just off-shore of the continental slope off the Lofoten archipelago. The latter region also has a large SST standard deviation (about 0.4 K, not shown), and similar values are found in other regions, notably in the vicinity of sloping bottom topography. The largest standard deviations for salinity at the surface (about 0.3) is found in the northern Barents Sea, where the impact of retreating and expanding ice coverage is highest.

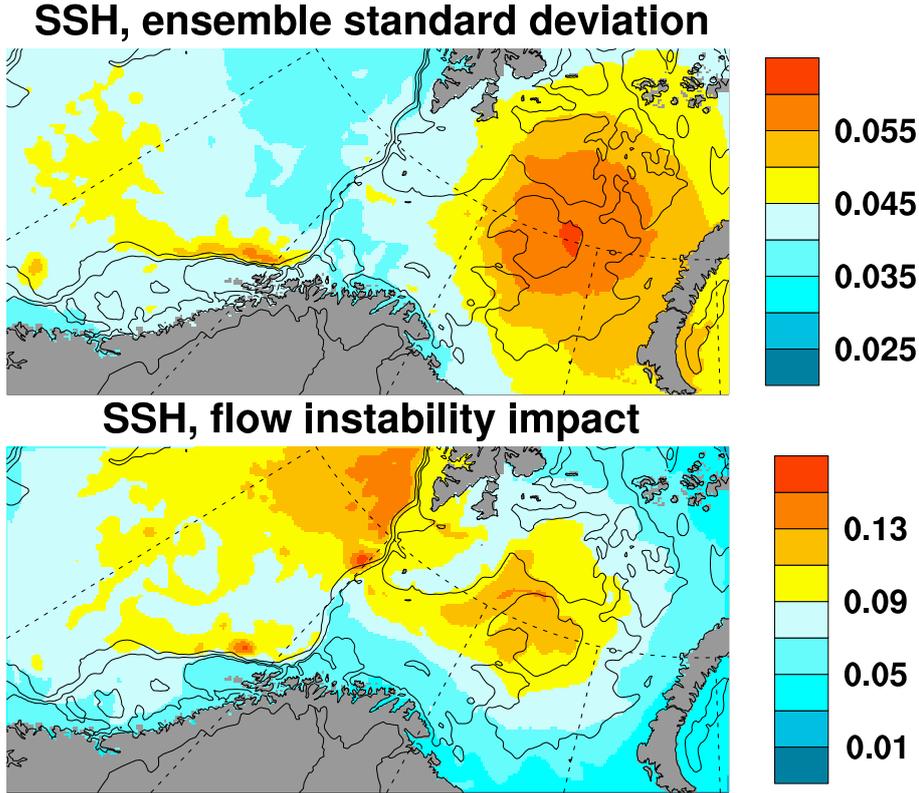


Figure 9: Ensemble variability. The time mean standard deviation of the ensemble offsets of the sea surface height,  $\sigma_{SSH'}$ , is displayed in the top panel. The fraction of non-deterministic variability,  $r_{SSH}^{nd}$ , which describes the relative impact of flow instabilities, is displayed in the bottom panel. This quantity has a value between 0 and 1, where 0 corresponds to a result where all ensemble members are identical. See the text for details. Label values in the top and bottom panels are in m and non-dimensional, respectively. Contour lines are shown for the isobaths at 250m, 500m, and 1000m.

The impact of flow instabilities on sea surface height, as described by the fraction of non-deterministic variability,  $r_{SSH}^{nd}$ , is displayed in the bottom panel of Figure 9. The maximum value of  $r_{SSH}^{nd}$  is approximately 0.15. Thus, even after the effects of the periodic forcing has been filtered out the SSH variability in the model results is predominantly deterministic.

We note that in the Barents Sea and off the continental slope of the Norwegian mainland, both  $\sigma_{SSH'}$  and  $r_{SSH}^{nd}$  are relatively large. However, there is a large region with high fractions of non-deterministic SSH variability in the deep part of the northwestern Norwegian Sea, with maximum values near the continental slope. The SSH standard deviation in this region is low, only about half of the values in the other two regions with high  $r_{SSH}^{nd}$  values.

An inspection of the results for the surface salinity reveals that regions with high values for  $\sigma_{S(0m)'}$  in the northern Barents Sea is a region with minimum  $r_{S(0m)}^{nd}$  values (not shown). Hence, the variability that is likely to be associated by sea-ice processes appears to be of a

highly deterministic nature in the model results. Maximum  $r_{S(30m)}^{nd}$  values are found in the southern Barents Sea, and in patches that are scattered throughout most of the domain, with values exceeding 0.5. Generally, the changes in  $r_S^{nd}$  in the upper 30 m are small.

When we consider the model results for SST, the fraction of non-deterministic variability reaches maximum values of about 0.3 in a banded structure in the deep Norwegian Sea, close to the 2800 m depth contour. The magnitude of  $r_{T(30m)}^{nd}$  is similar to that for  $r_{S(30m)}^{nd}$ , but the region with maximum values cover is shifted slightly northward in the Barents Sea, and also includes much larger regions in the deep Norwegian Sea.

## 4.5. Results from ranking

The method of comparing observations with ranked model results is commonly used for validation of medium-range weather forecast ensembles, see *e.g.* Hammill (2001) for a discussion of how to interpret such an analysis. Ranking is performed by collecting observations and the corresponding results from the model ensemble. For each observation, the results from the various ensemble members are sorted according to their value, and the observation is assigned a value according to the slot it fits into between the ordered ensemble results. If the observation is lower than all ensemble members, it is in the “slot” that covers the range up to the lowest ensemble value, and has rank 1. If the observation falls into the slot between the two lowest values from the ensemble, it has a rank of 2 *etc.* For an ensemble with  $M$  members, observations are assigned ranks in the interval  $1 - (M + 1)$ .

If the spread among the ensemble values is an accurate representation of variability, without a bias, the rank count becomes the same for all ranks. However, as was evident in subsection subsection 4.1 above, the present model result are biased, albeit with biases that vary in space and the variable under consideration. Furthermore, it turns out that the spread among the ensemble members is much too low, so even after the bias is removed, the large majority of observations are either smaller than the lowest value from the ensemble members, or higher than the largest value.

Hence, we restrict this presentation by considering de-biased results that are either outside (lower or higher) or inside the range of the ensemble members. Again, we make use of hydrographic observations from the IMR transects and the WOD 2005, and we subtract the mean bias from each data set, averaged over the depth range and year (1986) that are considered. Further, we take into account the effect of observation errors by adopting the method of perturbing observations that has been described by Saetra et al. (2004). The errors were taken to be Gaussian with standard deviations of 0.02 K and 0.01 PSU for temperature and salinity, respectively (*Ø. Østensen, pers. comm.*).

The results from the ranking analysis are given in Table 5. From the description above, it follows that there are presently 11 ranks, and with an ensemble variability that matches the observed variability, the theoretical probability that an observation has one particular rank, becomes 1/11. This would correspond to a distribution of 0.09 – 0.82 – 0.09 in each row in Table 5. Since the central column in the table has values of the order 0.15-0.20, our ensemble is a substantial underestimate of the observed variability.

We also note that the analysis of the transect data generally have the highest values for positive (de-biased) offsets of observations relative to model results. This is consistent with the

Table 5: Classification of de-biased hydrographic observations from 1986, according to the ensemble rank. The observations are counted in three categories: smaller than the minimum value from the ensemble ( $f < mnm$ ), inside the range of the ensemble ( $mnm \leq f \leq mxm$ ), and larger than the ensemble’s maximum ( $mxm < f$ ). The tabulated frequencies of occurrence,  $f$ , are based on observations and model results from the levels of 0, 10, 20 and 30 m.  $mnm$  and  $mxm$  are the minimum and the maximum values from the 10-member ensemble.

<b>data set</b>	$nO_{obs}$	$f < mnm$	$mnm \leq f \leq mxm$	$mxm < f$
<b>temperature</b>				
Gimsøy northwest	28	0.43	0.00	0.57
Bear Island west	56	0.45	0.21	0.34
Fugløy – Bear Island	432	0.28	0.16	0.56
Vardø north	220	0.33	0.23	0.44
World Ocean Database	11836	0.39	0.17	0.43
<b>salinity</b>				
Gimsøy northwest	28	0.25	0.14	0.61
Bear Island west	56	0.55	0.21	0.23
Fugløy – Bear Island	432	0.29	0.10	0.61
Vardø north	220	0.30	0.24	0.47
World Ocean Database	11836	0.40	0.20	0.40

probability density functions that were presented in subsection 4.3 above: The “original data” for temperatures in the “Fugløy – Bear Island” transect peaks at a value higher than the mean in Figure 5, while the corresponding peak from model results in Figure 6 occurs at a negative offset from the mean. The corresponding results for salinity in Figures 7 and 8 both have higher frequencies for positive offsets. However, the standard deviation in the observations are much higher, leading to a higher value in the rightmost column of the relevant row in Table 5.

## 5. Discussion and conclusions

The ensemble simulation that is investigated here is intended to be a part of an examination of biological activity over a 10 year period, with a focus on the Barents Sea. Hence, we have restricted the present analysis to the upper 30-50 m. The validation of the model results here has revealed that the performance of the model depends on the variable that is examined, and the ocean region that is considered. Generally, the validation reveals that the model temperature are much closer to the observations than the results for salinity. Furthermore, validation that invokes coastal observations of sea level suggests that the representation of the high frequency sea level variability is reliable, while the quality of the low-frequency variability in the model results is not impressive.

Salinity fields from the 10 m level are displayed in Figure 10. We note that the negative salinity bias is even stronger in ARC-N.ATL 20. Hence, the NOR-BAR 4 salinity bias is not

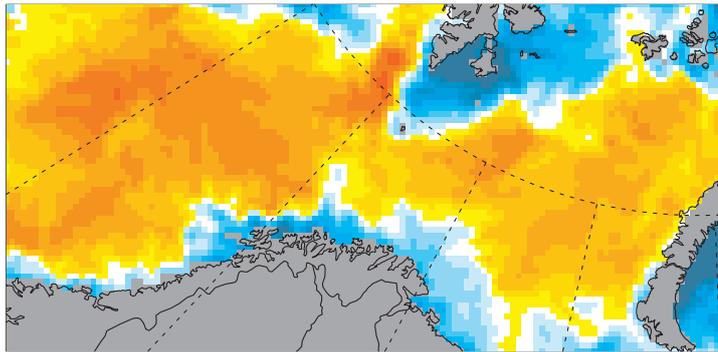
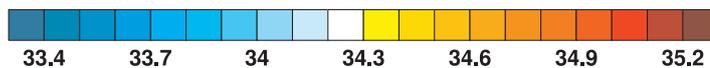
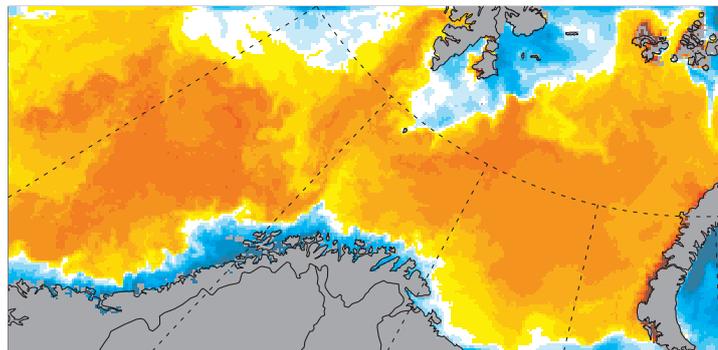
**Salinity, 20km model, 1986-01, 10m****Salinity, 4km model, 1986-01, 10m**

Figure 10: Salinity field at 10 m for January 1986. Results from the ARC-N.ATL 20 and NOR-BAR 4 are displayed in the top and bottom panels, respectively. The 20 km results shown here are for a subdomain that corresponds to the domain of the 4 km model.

due to an incorrect formulation of the conditions imposed on its open boundaries. Rather, it is the climatological salinities at the open boundaries of ARC-N.ATL 20 that are not advected properly into the interior of the domain. We also note that the largest differences in the regional salinities in Figure 10 occur in the Barents Sea, where the negative biases that were reported for NOR-BAR 4 in Section 4 were smaller than in the Norwegian Sea. The cost function (the error measured in units of standard deviations) in the upper 20 m is in the range 1.3-1.9 for the Barents Sea transects. This is far better than the corresponding range (8-20) for the “Bear Island west” transect in the Norwegian Sea sector of the domain. Furthermore, Table 2 reveals that the standard deviation of the difference between the de-biased model results for salinity and the observations is large (0.2-0.5 PSU). An exaggerated variability in the model salinities is also apparent by their larger spread in the p.d.f.s as displayed in Figures 7 and 8 (gray curves) for the “Fugløy – Bear Island” transect.

When discussing the validation of the sea temperature, one must keep in mind that surface heat fluxes are imposed in connection with the assimilation by nudging that was discussed in Section 2. Nevertheless, from Table 4 we find that the cost function decreases from the surface (0.7-1.1) to the 20 m level (0.3-0.9). Obviously, the seasonal variability will inflate the standard deviations for temperature, thus deflating the cost function. In addition to the seasonal changes in solar radiation, and atmospheric temperature changes on a synoptic scale, the transport of water masses by the currents in the region exhibit a seasonal variability (Ingvaldsen et al. (2004)). In order to evaluate the impact of seasonal variability, we repeated the analysis for the “Fugløy – Bear Island” transect for the months August – October only. This led to an increase in the cost function of the order 50-100% (Table 4). The biases reported in Table 2 are generally positive, with the warmest biases occurring in the Barents Sea transects (0.7-1 K).

We are not aware of other publications in which model results for the Barents Sea are validated similarly to the present study. However, model results for a set of simulations for the Skagerrak (the ocean between southeastern Norway and Jutland, Denmark) have been validated by Berntsen and Svendsen (1999). This is of relevance, since the water mass properties in both the Skagerrak and in the Barents Sea are influenced by mixing of Atlantic Water and coastal water masses. Berntsen and Svendsen (1999) found that the model error measured in standard deviations was 1.8 for salinity in the upper 50 m (right column in their Table 2). This is about the same as we find for the two transects in the Barents Sea. Furthermore, in their study which is limited to May and June, they find an error in the temperatures of the upper ocean of 2-2.6 standard deviations. If we extrapolate our results from the “Fugløy – Bear Island” transect for August – October, we estimate a cost function of 1-2 for temperature in the Barents Sea. Note that our simulations incorporate a nudging towards observations by surface heat fluxes, while *Berntsen and Svendsen’s* study does not.

The spatial SSH structures in the model results, which are depicted in Figure 3, are consistent with results from drifter data (Orvik and Niiler (2002)). The geostrophic velocities that are associated with the SSH gradients in the model, are also consistent with the finding that to the lowest order, streamlines in the Nordic Seas follow  $f/H$  contours (*e.g.* Isachsen et al. (2003)). Here,  $f$  is the Coriolis parameter, and  $H$  is the local water depth. The  $f/H$  contours are close to the isobaths at high latitudes.

We found that the model reproduces tidally induced sea levels well, while there are substantial errors in the low frequency SSH variability. A closer examination reveals that this error is manifested by offsets of 0.1-0.5 m that typically lasts for some weeks. Moreover, the offsets broadly occur at the same time and with the same sign for both sea level stations (Rørvik and Vardø). We have not been able to identify precisely why these errors occur, but it is possible that they are due to remote low frequency SSH variability that propagates into the Norwegian and Barents Seas.

On the other hand, the model results for low frequency SSH may be too energetic. We find that the standard variation in the sea level observations at Rørvik and Vardø are 0.19 m and 0.16 m, respectively. The model results have standard deviations of 0.25 m at both locations.

If the model SSH is too energetic, the bias in the corresponding estimate of squared SSH anomalies is positive, leading to deflated values in the fraction of non-deterministic variability  $r_{SSH}^{nd}$  as defined by Equation 9, and displayed in the lower panel of Figure 9. The standard

deviation values that were quoted for Rørvik and Vardø above correspond to an adjustment factor for  $r_{SSH}^{nd}$  of 1.7 and 2.4, respectively. In frontal regions, this quantity is of the order 0.09-0.15, Melsom (2005) found that its magnitude in frontal regions of the North Sea and southern Norwegian Sea is up to 0.3.

The model ensemble from which results are validated here, were generated by differences in the time series of the atmospheric forcing fields. In subsection 4.5, we found that ranked observations in the de-biased, ordered ensemble results for hydrography typically were higher than all ensemble members in 40% of the cases, and lower than all members in about 40% of the cases. Hence, the present approach for generation of an ensemble yields members that span a much smaller interval than that which corresponds to the observational variability. An alternative approach of perturbing initial fields in a decadal simulation of the ocean circulation in the North Sea and the Skagerrak was adopted in Melsom (2005). However, that approach gave rise to a slightly larger mismatch between the ensemble spread and the observational variability. Accordingly, we conclude that more sophisticated methods for production of an ensemble are called for, such as the advanced techniques that are implemented in variational analysis or in applications of the Kalman filter.

## 6. Appendix A: Missing sea level data

A list of intervals from which sea level data were not available from Rørvik and Vardø are given below. The list is limited to the relevant year for the present study, *i.e.*, 1986. If not otherwise is specified, all observations starting at 00 UTC of the first date through 22 UTC of the end date in each interval were missing.

### **Rørvik:**

- 10 January – 11 January
- 21 March – 2 April
- 24 April – 25 April
- 16 September – 18 September
- 1 October – 6 October
- 5 November – 5 November
- 12 December – 16 December
- 22 December – 22 December
- 29 December – 31 December

### **Vardø:**

- 24 March – 30 March
- 18 May 18 – 20 May
- 21 June 10 – 24 June 00

- 38 July – 30 July 06
- 17 August 08 – 19 August 08
- 5 September – 13 September 16
- 19 October 18 – 21 October 18
- 1 December – 6 December

## A. Acknowledgements

Hydrographic cruise data were obtained from the Institute of Marine Research, we thank Dr. Øyvind Strand for making the data available for our research. Additional hydrographic data were available from the National Oceanographic Data Center at the National Oceanic and Atmospheric Administration, and downloaded from the World Ocean Database at <http://www.nodc.noaa.gov/OC5/WOD01/>. Sea level data from Norwegian tide gauge stations were available from the British Oceanographic Data Centre (originally from the Norwegian Hydrographic Service Oceanographic Group), and downloaded from the Global Sea Level Observing System at [http://www.bodc.ac.uk/data/information\\_and\\_inventories/glossary\\_and\\_book/](http://www.bodc.ac.uk/data/information_and_inventories/glossary_and_book/). Thanks also to Dr.s Ana Carrasco, Jens Debernard and Jon Albretsen at our institute for their work related to the model simulations, and to Dr. Lars Petter Røed for many helpful discussions. We are grateful for the financial support for this study, which was provided by the Norwegian Research Council program “Havet og Kysten” under contract 173487/S40: Long-term Effects of Oil accidents. We also acknowledge the generous contribution of computer time allotted to this work by the Norwegian High-Performance Computing Center. Several figures were made with the presentation tool NCAR Command Language (NCL, <http://ncl.ucar.edu/>). NCL is a product of the Computational & Information Systems Laboratory at the National Center for Atmospheric Research (NCAR).

## References

- Albretsen, J. & I. Burud, 2006: Assimilation of sea surface temperature and sea ice concentration in a coupled sea ice and ocean model. In: European Operational Oceanography: Present and Future. Proceedings of the 4th International Conference on EuroGOOS. Ed.: Dahlin, H., C. Flemming, P. Merchand and S.E. Petersson. Brest, France. 661-666.
- Berntsen, J., & E. Svendsen, 1999: Using the SKAGEX dataset for evaluation of ocean model skills. *J. Mar. Sys.*, 18, 313-331.
- Blumberg, A.F. & G.L. Mellor, 1987: A description of a three-dimensional coastal ocean circulation model. In: Three-dimensional Coastal Ocean Models. Coastal and Estuarine Sciences, Vol. 4. Ed.: N. Heaps. American Geophys. Union.

- Conkright, M.E., J.I. Antonov, O. Baranova, T. P. Boyer, H.E. Garcia, R. Gelfeld, D. Johnson, R.A. Locarnini, P.P. Murphy, T.D. O'Brien, I. Smolyar & C. Stephens, 2002: World Ocean Database 2001, Volume 1: Introduction. NOAA Atlas, Coastal and Estuarine Sciences, NESDIS 42. Ed.: S. Levitus. National Oceanic and Atmospheric Administration, U.S. Government Printing Office, Washington, D.C. 167 pp.
- Engedahl, H., 1995: Implementation of the Princeton ocean model (POM/ECOM-3D) at the Norwegian Meteorological Institute (DNMI). Research Report no. 5. Norwegian Meteorological Institute, Oslo, Norway.
- Engedahl, H., G. Eriksrød, C. Ulstad & B. Ådlandsvik, 1997: Climatological oceanographic archives covering the Nordic Seas and the Arctic Ocean with adjacent waters. Research Report no. 59. Norwegian Meteorological Institute, Oslo, Norway.
- Engedahl, H., B. Ådlandsvik & E.A. Martinsen, 1998: Production of monthly mean climatological archives for the Nordic Seas. *J. Mar. Sys.*, 14, 1-26.
- Engedahl, H., A. Lunde, A. Melsom & X.B. Shi, 2001: New schemes for vertical mixing in MI-POM and MICOM. Research Report no. 118. Norwegian Meteorological Institute, Oslo, Norway. 44 pp.
- Hammill, T.M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weather Rev.*, 129, 550-560.
- Hjermann, D.Ø, A. Melsom, G.E. Dingsør, J.M. Durant, A.M. Eikeset, L.P. Røed, G. Ottersen, G. Storvik & N.C. Stenseth, 2007: Fish and oil in the Lofoten-Barents Sea system: a synoptic review of what is known and not known about the effect of oil spills on fish populations. *Mar. Eco. Prog. Ser.*, 339, 283-299 DOI: 10.3354/meps339283
- Ingvaldsen, R.B., L. Aspelin & H. Loeng, 2004: The seasonal cycle in the Atlantic transport to the Barents Sea during the years 1997-2001. *Cont. Shelf Res.*, 24, 1015-1032 DOI: 10.1016/j.csr.2004.02.011
- Isachsen, P.E., J.H. Lacasce, C. Mauritzen & S. Häkkinen, 2003: Wind-driven variability of the large-scale recirculating flow in the Nordic Seas and Arctic Ocean. *J. Phys. Oceanogr.*, 33, 2534-2550.
- Johansen, Ø., K. Skognes, O.Ø. Aspholm, C. Østby & K.A. Moe, 2003: Accidental oil spills – consequences in the water column. (In Norwegian). SINTEF Rapport no. STF66 F03028/ULB 7-c. Foundation for Scientific and Industrial Research at the Norwegian Institute of Technology (SINTEF), Trondheim, Norway. 72 pp.
- Kangas, T.-V., E. Svendsen & Ø. Strand, 2006: Average value of salinity and temperature in the Institute of Marine Research's fixed sections. (In Norwegian). *Fisken og Havet* no. 6/2006, Institute of Marine Research, Bergen, Norway. 51 pp.

- Loeng, H., V. Ozhigin, B. Ådlandsvik & H. Sagen, 1993: Current measurements in the north-eastern Barents Sea. In: Proceeding of the ICES Statutory Meeting 1993. C.M. 1993/C:41. 22 pp.
- Martinsen, E.A., 1982: Operational oil drift model for Norwegian waters and surrounding oceans. (In Norwegian). Technical Report no. 59. Norwegian Meteorological Institute, Oslo, Norway.
- Martinsen, E.A. and H. Engedahl and G. Ottersen and B. Ådlandsvik and H. Loeng and B. Balino, 1992: Metocean MOdeling Project, Climatological and hydrographical data for hindcast of ocean currents. Technical Report no. 100. Norwegian Meteorological Institute, Oslo, Norway.
- Martinsen, E.A., A. Melsom, V. Sveen, E. Grong, M. Reistad, N. Halvorsen, Ø. Johansen & K. Skognes, 1994: The operational oil drift system at DNMI. Technical Report no. 125. Norwegian Meteorological Institute, Oslo, Norway.
- Melsom, A., 2005: Mesoscale activity in the North Sea as seen in ensemble simulations. *Ocean Dyn.*, 3-4, 338-350, DOI: 10.1007/s10236-005-0016-3
- Metzger, E.J. & H.E. Hurlburt, 2001: The nondeterministic nature of Kuroshio penetration and eddy shedding in the South China Sea. *J. Phys. Oceanogr.*, 31, 1712-1732.
- Orvik, K.A. & P. Niiler, 2002: Major pathways of Atlantic water in the northern North Atlantic and Nordic Seas toward Arctic. *Geophys. Res. Lett.*, 29, 1896, DOI: 10.1029/2003EO500009
- Røed, L.P. & J. Debernard, 2004: Description of an integrated flux and sea-ice model suitable for coupling to an ocean and atmosphere model. met.no Report no. 4/2004. Norwegian Meteorological Institute, Oslo, Norway.
- Schauer, U., H. Loeng, B. Rudels, V.K. Ozhigin & W. Dieck, 2002: Atlantic water water flow through the Barents and Kara Seas. *Deep-Sea Res. I*, 49, 2281-2298.
- Saetra, Ø., H. Hersbach, J.-R. Bidlot & D.S. Richardson, 2004: Effects of observation errors on the statistics for ensemble spread and reliability. *Mon. Weather Rev.*, 132, 1487-1501.
- Wettre, C., Ø. Johansen & K. Skognes, 2001: Development of a 3-dimensional oil drift model at DNMI. Research Report no. 133. Norwegian Meteorological Institute, Oslo, Norway.
- Woodworth, P.L., T. Aarup, M. Merrifield, G.T. Mitchum & C. Le Provost, 2003: Measuring progress of the Global Sea Level Observing System. *EOS, Trans Am. Geophys. U.*, 84(50), 565-566, DOI: 10.1029/2003EO500009