Validation of an ocean model ensemble

Arne Melsom*[1]
[1]*Norwegian Meteorological Institute, Oslo, Norway*
* Corresponding author, email: arne.melsom@met.no

Abstract

Validation of a 10-member ocean circulation ensemble with a 4 km resolution reveals that the main discrepancies are too low salinities in the Atlantic Water, and too low variability among the ensemble members. We show that the low salinities are due to insufficient salt transport into an outer 20 km model, while the inadequate ensemble variability is likely due to the construction of the ensemble.

Keywords:

Model validation, ensemble simulations, ocean circulation, ocean variability

1. Introduction

The purposes of a model validation can be manifold. The validation may be undertaken in order to assess the overall quality of a model simulation, and it may be performed in order to pinpoint specific weaknesses of a model. The focus here will be on the latter aspect, i.e., we intend to identify the leading shortcoming of our experiment.

Furthermore, since ensemble simulations for the ocean circulation is still in its infancy, we include a crude validation of how well the observed variability is captured by the differences between the ensemble members.

We will show that the main shortcoming is a bias in the salinity of the Atlantic Water and that the ensemble underestimates the observed variability. Finally, we discuss the origins of these problems, and suggest how they can be amended.

2. Numerical experiment and observations

The coupled model system consists of a three-dimensional, primitive equation, σ-coordinate physical oceanographic model (MI-POM; Engedahl, 1995a) and a dynamic-thermodynamic sea-ice model (MI-IM; Røed and Debernard, 2004). A Barents Sea-Norwegian Sea domain with 4 km mesh size and 21 σ-levels (NOR-BAR4) was nested into a 20 km model that covers the Arctic Sea and adjacent seas (ARC20) (Figure 1).

A simulation was performed for the ARC20 domain for 1981-1986 with atmospheric forcing fields from the ERA-40 re-analysis from the European Centre for Medium-Range Weather Forecasts (ECMWF). SST and ice concentration were nudged towards observations using SSTs from ERA-40 and a merged ice concentration product (Albretsen and Burud, 2006). No tidal forcing was applied to ARC20. The nested NOR-BAR4 model was spun up for 1984 and 1985 using the atmospheric forcing as described above and the SST/ice concentration nudging. Tidal forcing was prescribed for 8 constituents at the lateral boundaries of NOR-BAR4.

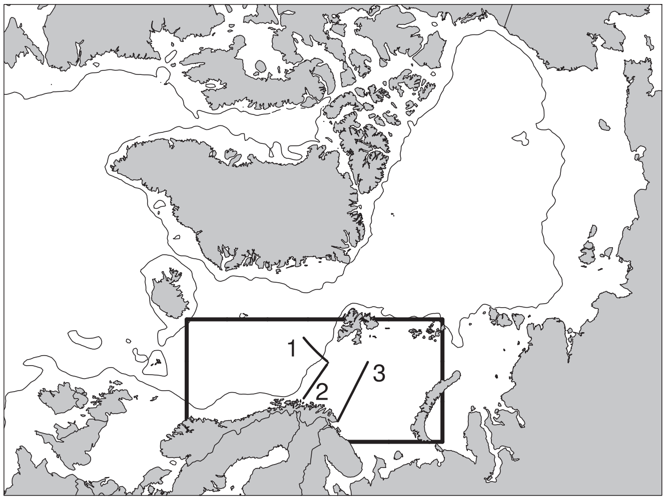

Figure 1:
Model domain for ARC20 with the NOR-BAR4 domain inside the rectangle. The shelf break is shown by the 500m isobath, along with numbered lines that correspond to the IMR transects "Bear Island west", "Fugløya-Bear Island" and "Vardø north", respectively.

The NOR-BAR4 results for 1986-01-01 were used as initial conditions for a 10-member ensemble for 1986. One set of atmospheric forcing fields from the ECMWF forecasts was constructed by extracting all the 12 hour prognoses from the forecasts, from forecasts issued every 12 hours. Thus, the first forecast used for 1986 was issued at 1985-12-31 12:00, since the 12 hour prognosis is valid at 1986-01-01 00:00. The 12 hour prognoses from the next forecast (issued at 1986-01-01 00:00) is valid for 1986-01-01 12:00, and so on. In this way, a series of forcing fields with a temporal resolution of 12 hours was available for the first ensemble member. The procedure was then repeated for the prognoses for 24, 36, .., 120 hours, thus providing different atmospheric forcing fields for each of the ocean ensemble's 10 members. The atmospheric forcing fields constitute the only difference in the generation of the various ensemble members.

Hydrographic data along fixed cruise tracks and cast positions are available from the Institute of Marine Research (IMR; Kangas et al. (2006). The data have been subjected to a quality assurance process at IMR, using the The Integrated Global Ocean Services System standard. Observations were made as CTD casts during 1980-2006, available with a vertical resolution of 5 m. The cruise tracks are displayed in Figure 1.

3. Model validation

3.1 Bias and standard deviation

The bias of property p from the ensemble is defined as

$$(1) \qquad bias_e = \frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\left(p_{m,n}^{mdl} - p_n^{obs}\right)$$

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 variance of p for the entire ensemble is defined as

$$(2) \qquad \sigma_e^2 = \frac{1}{MN}\sum_{n=1}^{N}\sum_{m=1}^{M}\left[\left(p_{m,n}^{mdl} - p_n^{obs}\right) - \left(\overline{p}_e^{mdl} - \overline{p}^{obs}\right)\right]$$

where the over bar denotes the arithmetic mean value. The variance for ensemble member m, σ_m^2, is defined analogously. The temperature statistics in Table 1 that are based on the cruise data, reveal that the eastern transect "Vardø north"' is warmer in the model that 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 above are largest in the western transect.

	temperature			**salinity**		
Data set	$bias_e$	σ_e	$\sigma_{1\text{-}10}$	$bias_e$	σ_e	$\sigma_{1\text{-}10}$
Bear Island west	0.259	1.789	1.801	-0.454	0.303	0.306
Fugløya – 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

Table 1: The model bias and standard deviation when compared to data from the hydrographic sections. Only levels 0, 10, 20 and 30 m have been considered here. Values in the $\sigma_{1\text{-}10}$ column are averages of the standard deviations when each of the ten ensemble members is treated separately. T and S values are in K and PSU, respectively.

There is a substantial negative bias in the model results for salinity. This is particularly the case for the "Bear Island west" transect, which cuts across the northern branch of the Norwegian Atlantic Current (NAC). Here the observations are saltier than the model results by 0.45 PSU. The largest standard deviation values are again found to the west.

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 since σ_e and $\sigma_{1\text{-}10}$ in Table 1 are nearly identical. Further quantification of the low variability among the ensemble members will be given in subsection 3.4.

3.2 Cost function

Let subscript n denote the nth observation of property p at station s and depth z. Then, introduce the cost function Dp_σ as a model error weighted by the observed st. dev.

$$(3) \qquad Dp_\sigma = \frac{1}{N}\sum_{n=1}^{N}\frac{\left|p_n^{obs} - p_n^{mdl}\right|}{\sigma_p}$$

First, we computed σ_p values at each position and depth from the IMR data. Then, the cost function was computed based on data from 1986 and corresponding results from one ensemble member. The observations from the various transects are not uniform over the year, and we are not able to determine the seasonal variability. Since this variability is likely to contribute to the standard deviation values, we also include an analysis of the data from the "Fugløya-Bear Island" transect for the months from August through October only.

The cost function for salinity is greater than the corresponding values for temperature, for all transects and depths in Table 2. This was expected, since air-sea heat fluxes were assimilated in the simulations, while no assimilation was performed for salinity. Note that even though the assimilation was conducted at the surface, any bias in the temperature of the upper ocean will also be reduced due to diffusion and turbulence in the well-mixed region considered here. (None of the temperature data from the IMR hydrography was used in relation to the heat flux assimilation, so in the IMR hydrography constitutes an independent data set.) Values that are based on the August-October results are higher than the whole-year results, due to the inflated standard deviations that arise from including the seasonal temperature cycle in the whole-year results. The large cost function value at 50 m in the "Bear Island west" transect is due to a strong cold bias.

The very high cost function values for salinity in the "Bear Island west" transect caused by a combination of the large bias of 0.454 (Table 1) 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). The "Bear Island west" transect cost function varies from 0.06 closest to Bear Island, to 22.5 in the core of the NAC.

		temperature				**salinity**			
Data set	**#obs**.	0*m*	10*m*	20*m*	50*m*	0*m*	10*m*	20*m*	50*m*
Bear Island west	14	1.14	0.73	0.82	2.37	20.22	8.80	7.96	6.17
Fugløya – Bear Is.	99	0.88	0.77	0.33	0.61	1.82	1.82	1.92	2.95
Fugløya – Bear Is.$^{8\text{-}10}$	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

Table 2: The cost function Dp_σ for temperature and salinity for the three IMR transects, for selected levels. Values are differences between model results and observations, measured in standard deviation units. "Fugløya-Bear Is.$^{8\text{-}10}$" is based on data and results for the months August through October only. See the text for details.

3.3 Probability density function

The probability density function (p.d.f.) is a representation of the statistical distribution of a variable. Here, p.d.f.s are used in order to investigate similarities and differences between distributions from observations and model results. Model results were extracted at the position of transect stations. If results for all months were included in this analysis, the seasonal cycle would lead to a large spread in the p.d.f.s for temperature, which would obscure other aspects of the distributions. Thus, we restrict the p.d.f. analysis to the best sampled periods, which were Jun-Sep for "Vardø north" and to Aug-Nov for the other transects. Further, we sample model results every 10 days (the de-correlation time is 6-12 days, as estimated by the e-folding time of the local autocorrelation for temperature and salinity).

The shapes of the p.d.f.s for salinity in the "Fugløya-Bear Island displayed in Figure 2 are quite similar. Observations and model results are both notably left-skewed, while the major difference is that the standard deviation in the observations are markedly higher than in the model results (0.37 *vs* 0.16).

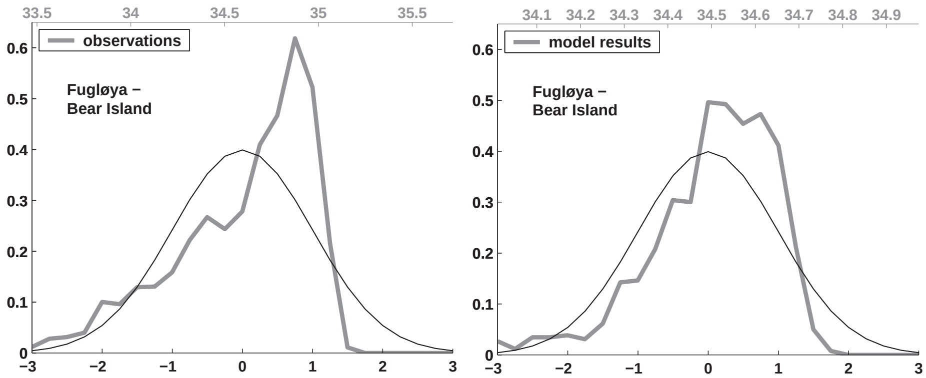

Figure 2: p.d.f. for salinity (upper 30m) of the "Fugløya-Bear Island" transect, from observations (left) and model results (right). Thin lines show the normal distribution (for reference). Values along the top and bottom axes correspond to original and normalized values, respectively. The overall mean along the top axis is placed directly above zero st. dev.s on the bottom axis. The displayed interval is set so that the span for one st. dev. matches on the top and bottom axes.

3.4 Results from ranking

A comparison of observations with ranked results from a model ensemble is commonly used for validation of medium-range weather forecast (Hammill, 2001). 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. For a 10-member ensemble the rank will thus be in the interval 1-11. If the ensemble spread is an accurate representation of the observed variability, the rank count becomes the same for all ranks. However, as discussed in subsections 3.1 and 3.2, our model result are biased and has a spread that is too low.

We restrict this analysis to considering de-biased results that are either outside (lower or higher) or inside the range of the ensemble members. We subtract the bias from each data set, and take into account the effect of observation errors by perturbing observations with Gaussian noise that has standard deviations of 0.02K and 0.01 PSU for temperature and salinity, respectively (Ø. Østensen, pers. comm.).

With matching variability in the ensemble and observations, the theoretical probability that an observation has one particular rank is 1/11, corresponding to a distribution of 0.09-0.82-0.09 in each row in Table 3. The corresponding central values in the table are ~0.15-0.25. Hence, our ensemble substantially underestimates the observed variability.

Data set	**#obs**	***f<mnm***	***mnm<f<mxm***	***mxm<f***
Bear Island west	56	0.45	0.21	0.34
Fugløya – Bear Island	432	0.28	0.16	0.56
Vardø north	220	0.33	0.23	0.44

Table 3: De-biased observations are counted in three categories: smaller than the minimum value from the ensemble (*f<mnm*), inside the range of the ensemble (*mnm<f<mxm*), and larger than the ensemble's maximum (*mxm<f*). The tabulated frequencies of occurrence, *f*, are based on observations and model results for temperature from 0, 10, 20 and 30m.

4. Discussion

The ensemble simulation that is investigated here is part of an examination of biological activity with a focus on the Barents Sea. Hence, we have restricted the present analysis to the upper 30-50 m. Generally, while there are regional differences, the validation reveals that the model temperature is much closer to the observations than the results for salinity. Since we are working with a nested model system using non-conserving open boundary conditions, we first compared the salinity results from ARC20 and NOR-BAR4. While there is a fresh bias in the model as shown in section 3, the results from NOR-BAR4 are actually slightly saltier than those from ARC20 (not shown). Next, we checked the difference between ARC20 5 years after it was initialized, and compared the salinity results with the climatology which was used for initialization. The difference at the surface for regions where S>34.8 is depicted in Figure 3. The model results are generally salt deprived in the core of the Atlantic Water, strongly indicating that there is an insufficient influx of salty Atlantic Water across the open boundary in the North Atlantic Ocean where only weak, thermal wind derived currents push water into the ARC20 domain. Thus, we recommend that a more vigorous volume flux is applied at this boundary, e.g. from a global model simulation. Note also that the flow relaxation scheme that was used here (Engedahl, 1995b) is not property-conserving, but generally work well when the prescribed volume fluxes and distribution of water mass properties at the open boundaries are realistic and consistent (Jensen, 1998).

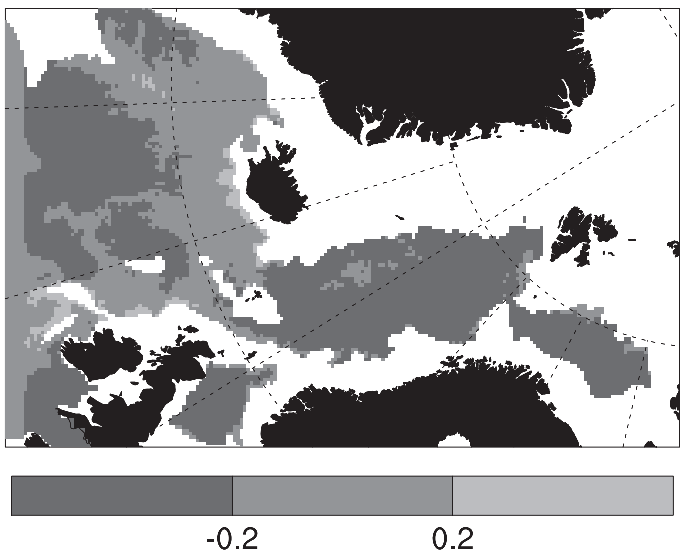

Figure 3: Surface salinity bias in the south-eastern ARC20 relative to climatology. The bias is only displayed in regions where the climatological salinity exceeds 34.8 PSU.

We note that the low variability among the ensemble members that was found in subsection 3.4 is similar to the results of an earlier study (Melsom, 2005). The difference between these two ensembles is the construction technique that was applied. Here, we used sets of perturbed atmospheric forcing fields, whereas perturbations was imposed in the initial conditions in Melsom (2005). Other aspects of the two ensembles, including horizontal resolution and ensemble size, are comparable. Hence, we conclude that a more intrusive means of perturbation (of the ocean circulation variables) should be attempted, such as the techniques that draws upon available observations in ensemble weather forecasting.

Acknowledgements

Hydrographic cruise data were provided by IMR, thanks to Dr. Ø. Strand for making the data available. Financial support was provided by the Norwegian Research Council under contract 173487/S40. HPC computer time was allocated by NOTUR (project NN4559K). Dr.s Ingerid Fossum and Ana Carrasco at the Norwegian Meteorological Institute performed the model simulations.

References

Albretsen, J., and 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 4th EuroGOOS Conference, H. Dahlin et al., ed.s. EuroGOOS Office and European Commission, 661-666.

Engedahl, H. (1995a). Implementation of the Princeton ocean model (POM/ECOM-3D) at the Norwegian Meteorological Institute (DNMI), Research Report no. 5, Norwegian Meteorological Institute, Oslo, Norway, 40 pp.

Engedahl, H. (1995b). Use of the flow relaxation scheme in a three-dimensional baroclinic ocean model with realistic topography, Tellus, 47A, 365-382.

Hammill, T. (2001). Interpretation of rank histograms for verifying ensemble forecasts, Mon. Weather Rev., 129, 550-560.

Jensen, T. G. (1998). Open boundary conditions in stratified ocean models. J. Mar. Sys., 16, 297-322.

Kangas, T.-V., E. Svendsen, and Ø. Strand (2006). Average value of salinity and temperature in the institute of marine research's fixed sections (in Norwegian), Fisken og Havet 6/2006, Institute of Marine Research, Bergen, Norway, 53 pp.

Melsom, A. (2005). Mesoscale activity in the north sea as seen in ensemble simulations, Ocean Dyn., 3-4, DOI: 10.1007/s10236-005-0016-3, 338-350.

Røed, L.P., and J. Debernard (2004). Description of an integrated flux and sea-ice model suitable for coupling to an ocean and atmosphere model, met.no Report 4/2004, Norwegian Meteorological Institute, Oslo, Norway, 56 pp.