Replies to Reviewer #1: Review of "Barents-2.5km v2.0: An operational data-assimilative coupled ocean and sea ice ensemble prediction model for the Barents Sea and Svalbard"

General Comments

This paper presents the development methods and validation of the Barents-2.5km v2.0 forecast model. The new version of the Barents-2.5km model has been upgraded to include a 24 member Ensemble Prediction System. The data assimilation has also been upgraded from only sea ice concentration to additionally including sea surface temperature and hydrographic profiles.

Overall the paper shows the forecast system has promise as a useful tool for various applications that require predictions of ocean variables in the Barents, Norwegian and Greenland Sea. This is particularly important for the marginal ice zone areas and the paper shows the challenge and skill required for accuracy. Some of the description of the model forecast skill is hard to follow. Analysis and validation of currents are presented in a separate paper which means there is not a complete picture here. There is also no comparison to v1.0 of the Barents-2.5km model. There are a few points that need attention that I detail below.

Thank you for your feedback on our manuscript, and a list of specific comments to address. In order to present a fuller assessment of the general model performance for the most relevant state variables, we aim to present the statistical properties of ocean currents and ice drift vectors as compared to observations, and put some focus on validation by non-observed variables.

Specific Comments

Line 7-9: Currents are only briefly mentioned in the paper and seems to be a larger part of another paper. The abstract should be used to discuss what is in the paper.

Parts of the ocean current analysis will be included in the revised manuscript, focusing on the statistical skill of surface currents and ice drift from the reference member. A more comprehensive analysis of surface current predictability and from the same model will be presented in another document, but to avoid confusion we will remove the reference to this work from our abstract.

Line 108-110: Have you considered using the elastic-plastic anisotropic (EAP) rheology scheme in CICE (Heorton et al., 2018)? Why was EVP chosen?

The EVP rheology was chosen in this model setup because it is computationally faster, which has a bearing on our capability to run multiple ensemble members in a forecast mode beside ROMS. A comment on this will be included in the manuscript.

Line 291: You should mention the SST RMSE July peak.

This feature will be mentioned along with the other peaks in the revised manuscript.

Lines 293-296: The model is compared against the same observations that are entrained through the data assimilation system. I understand that the observations are considered independent because they are compared with a previous forecast. This has value in quantifying the quality of the prediction for assimilated variables. However, I think the quality of the paper could be improved if you make comparison with un-assimilated data, perhaps sea surface height which could give an indication of current quality.

We agree that using observation types that were not assimilated at all would allow a more complete assessment of the model performance. We think that both upper ocean current velocities (from HF radar) and low-frequency ice drift vectors (from passive microwave imagery) could fill this gap. For an independent validation of model hydrography, we are able to use float observations during model periods when the EnKF was not successful, hence these data were not assimilated and provide a means for validation. Such results are envisioned for the revised manuscript.

Line 300-305: Please add the SIC ice charts dataset to the methods and reference.

The subjective sea ice concentration charts are another pivot of independent observations used for validation. We will include documentation on this data set in a new method section and add a remark that these data are not assimilated in the model.

Line 312-314: Sea ice concentration tends to increase rapidly when there is no data for assimilation. It seems to be reliant of the SIC assimilation data. I see you've made a point about why this happens in Line 422 but I think it would be better to make that point here or rephrase the bit around Line 422.

We will add this comment in Sec. 5.2 as suggested.

Line 330-334: Why does the model have greater RMSE for SST up to 12h lead time than the persistence, deterministic or trend with observations?

By definition, the persistence forecast delivers zero RMSE at lead time zero (e.g. the persistence reference forecast says that SST of the future is the same as now). The trend forecast same as the persistence plus the time-change from the model, hence it is also zero at lead-time zero. The deterministic is the reference member from the Barents-EPS, it may have an RMSE smaller or larger than the ensemble mean. We clarify these points in the revised manuscript.

Line 347-358: Describe the validation/analysis technique in the methods. Same goes for other validation methods. What do the bins represent?

A new method section that describes our validation methods will be added in the revised manuscript.

Line 374-377: What are analysis increments? This needs to be clearer.
Line 380: How is the spread calculated? Is it the standard deviation of the ensemble?
Line 392-393: What are spread increments? It also does not say what they are in Figure 12 caption.
Line 442-446: I do not understand Figures 10 and 11 enough. Why is the spread in SST satisfactory? Why is the spread in SIC too low?

Analysis increments are the difference between the model analysis (i.e. the model state after assimilation) and the model background (i.e. the model state before data assimilation) at the same time, i.e. the analysis time step. Similarly, spread increments are the difference between the analysis ensemble spread (after assimilation) and the background ensemble spread (before assimilation). We will include all these definitions in the revised text.

The ensemble spread for each state variable is the standard deviation between ensemble members. The spread is typically reduced (i.e. the spread increment is negative) during the assimilation process, as all members are brought closer towards the observations.

Figs. 10,11 and 12 will be improved for clarity and layout, and we will put more emphasis on the text discussing these figures, where the quality of the ensemble spread is assessed. For the rank histogram (Fig.10), an ideal ensemble spread would yield a flat uniform distribution, i.e. all members have an equal probability of realization. Underdispersive ensembles are identified by accumulation of observations in the lowest and highest ranks, meaning that the model values rarely yield as high or low values as seen in observations. We extend the discussion in the revised manuscript and provide references on ensemble modeling techniques. In the reliability diagram (Fig. 11), an ideal spread would be given by a straight diagonal line, indicating that modeled probabilities of exceeding threshold match observed frequencies of exceeding these threshold. Deviations from the diagonal show that the EPS tends to over- or underestimate probabilities to exceed certain ice concentration thresholds. In our case, the EPS underestimates probabilities for occurrence of open water and low ice concentrations.

Line 471-472: How does the Barents-2.5km v2.0 of the model compare against original Barents-2.5km? Can you quantify the improvement in accuracy that users can expect?

We have no period that is suitable for direct comparison as the v1.0 model was prone to failures during the transition to the new model system. However, we will provide a discussion on the changes in model performance that the user may expect:

- The v1 model applied a strong constraint to observed ice concentration using a nudging scheme, and hence matched AMSR maps closely at the analysis time step. Because ocean temperatures were not modified during the analysis, the ice concentration forecast deteriorated rapidly.

- No ocean temperature was assimilated in the v1 model setup. As a consequence, ocean temperature and salinity drifted away from the observed ocean state. Using a nudging towards the parent model state, the Barents v.1 reflected the hydrography of that model (Topaz).

Table 1: This could be quoted in the text instead of a table.

We accept this suggestion.

The Technical corrections will be implemented.

Figure 12: Please improve the caption with more detail. What are the black regions in g and h?.

Fig. 12 will be updated. The black regions in g and h have been removed to make figures clearer. These dark areas corresponded to zones of zero spread increments, thus where the spread was not modified after assimilation. These zones correspond now to white areas with zero as colorbar value.