Response to Reviewer #2: Review for “Barents-2.5km v2.0: An operational data-assimilative coupled ocean and sea ice ensemble prediction model for the Barents Sea and Svalbard”

Recommendations: Major Revision

General Comments and summary

This paper presents the version 2.0 of the operational ocean and sea-ice forecast Barents-2.5km model. This version includes an ensemble prediction system (EPS) with an off-line ensemble-based data assimilation (DA) component. The system routinely assimilates sea ice concentration (SIC), sea surface temperature (SST), and in-situ hydrography observations. With the DA component, the Barents-2.5 km model shows better SST forecast skills, e.g., improvement over the persistence forecast during spring and summer. Although the predictive skill for SIC is not as high, the model still displays skillful performance with DA, as it reduces the model drift away from the truth state. Furthermore, the EPS also provides the uncertainty estimate of the model state. The ensemble spread for SST is generally reasonable, although it may miss some extreme values, whereas the ensemble spread for SIC is too small. Overall, the Barents-2.5km model with its EPS and DA component is a valuable tool for forecasting ocean and sea-ice conditions.

After carefully reviewing the paper, I am impressed with the technical details presented. The work is certainly worthy of publication in GMD. However, I have some concerns with the DA part. In particular, some of the context regarding the ensemble Kalman filter (section 3.1) appears to be incorrect, and in my opinion, some important DA details seem to be missing. Although the paper does include some analysis on the performance of DA, the issue of non-Gaussianity, which can be especially important for sea ice observation, is not much addressed or discussed. While I appreciate the manuscript's primary focus on documenting and demonstrating technical expertise, it's equally important to ensure that the context of the DA component is accurately and completely presented. Therefore, I recommend a major revision of the DA section before publication in GMD.Specific line-by-line comments

Thank you for the appreciation of our work on the model, and the feedback on our manuscript. We would like to improve our manuscript towards a better description of the applied DA methods, making the corrections of errors noted by the reviewer. We also aim to add details about the assimilation of sea ice concentrations using the deterministic EnKF. In particular, we want to present an analysis of the local gaussianity (or lack thereof) in sea ice concentrations. We agree that this aspect is important, as we do see limitations of our model’s capabilities towards constraining sea ice concentrations in an efficient way and future model development has to address this shortcoming.

We also appreciate the reference suggestions provided by the reviewer for several aspects of the EnKF method and sea ice assimilation.

- Line 178 The citation here can be a little misleading, as the EnKF in (Evensen 1994; Burgers et al. 1998) are not usually referred to as the deterministic version of EnKF. I

suggest putting the citation (Sakov and Oke 2008) here, and move the citations (Evensen 1994; Burgers et al. 1998) to line 180. For the EnKF reference, in addition to (Evensen 1994; Burgers et al. 1998), I recommend that also include another reference (Houtekamer and Mitchell, 1998)

This paragraph will be adapted to include the suggested citations and make a clearer difference between the EnKF and DEnKF, mainly regarding formulas and citations. In particular, the approximated ensemble transform matrix used in the DEnKF will be specified in the text.

- Line 183-184: Although it is tangent to the main thread of the paragraph here, van Leeuwen (2020) notes that in the original stochastic EnKF, the perturbations should be added to the ensemble equivalence of the observation H(x) instead of the observation y. This distinction becomes significant when the observation error is non-symmetric (e.g., skewed), which can have important implications for, e.g., bounded observations. van Leeuwen, PJ. (2020) A consistent interpretation of the stochastic version of the Ensemble Kalman Filter. QJR Meteorol Soc., 146: 2815– 2825. https://doi.org/10.1002/qj.3819

This is an interesting detail on how ensemble perturbations are computed. We will add a remark in section 3.1 of the revised manuscript.

- Line 185-190: Equation (2) is the stochastic version of EnKF, while this paper uses the deterministic version of EnKF. Using Equation (2) here can be confusing to the readers. ***Therefore, I suggest, e.g., replacing Equation (2) with the deterministic transform equation in Sakov and Oke (2008) and replacing this paragraph with a new one (or add a new one) for the deterministic EnKF in Sakov and Oke (2008).***

Our revised manuscript will contain a separate paragraph on the deterministic EnKF (DEnKF); in particular, the approximated ensemble transform matrix used in the DEnKF is specified in the text together with the reference to Sakov and Oke (2008). Prior to this paragraph, an introductory paragraph about the EnKF is kept with the corresponding general EnKF equations (Eq. 2 and Eq. 3) and references. The suggested reference of Houtekamer and Mitchell (1998) will be added as an EnKF reference in the text.

- Line 197-198 Is this a reasonable assumption for the observations assimilated in this work? This assumption will introduce larger representation error to the observations that are taken at time points more distant from the analysis time. Although this issue is discussed in Section 6.4, I suggest adding one or two sentences commenting on this assumption here.

Synchronous assimilation of observed variables is clearly a simplification, and currently we are working on experiments to study the benefits of asynchronous assimilation. We will add more details on the implications of this assumption in the revised manuscript. In essence, we argue that the synchronous DA may be sufficient to avoid drift of the model state. Most important shortcomings are in situations with diurnal variation as for summer time SSTs during low winds.

Line 199-202 I suggest extending this paragraph by adding more DA details. Specifically, (1) Including some details about what the "spread reduction factor" and the "global moderation factor" mean, and how they work.

We will add more explanations of the parameters used to configure the EnKF.

(2) It seems that only horizontal localization is applied. Do the sea surface observations have impact on the state variables in the ocean (e.g., the ocean current at 30-m deep)?

It is true that only horizontal localisation is applied for the EnKF in our model. With further investigation of model results following this remark, we note in fact that in some circumstances the lack of vertical localisation combined with relatively low number of ensemble members can lead to artifacts in the model analysis at depth, possibly resulting from spurious correlations in the ensemble. We will discuss this point in a revised manuscript, including possible ways forward to address this issue.

I suggest providing more information on how the observation errors are moderated, since it is one of the key part in DA system. I wonder if the adaptive tuning of the observation error could somehow partially compensate the issue of the time-dependent representation error for the observations.

In `enkf-c`, there is a moderation factor (K factor, see equation below) used to moderate the observation impact by smoothly increasing the observation error in function of the innovation magnitude. For large innovations, the K-factor plays an important role in the increase of the observation error; for weak innovations, the K-factor does not play a big role and observations errors are mostly kept unchanged. A K-factor equal or lower than 2 is recommended; if the K-factor is set too high, then for large innovations we would lose the moderation term to decrease the observation impact and we want to avoid that.

$\sigma^2_{obs} \leftarrow [(\sigma^2_b + \sigma 2obs)2 + \sigma 2b\ d2/K2]½ - \sigma 2b$

Are the observations the same type of variables as the model states in ROMS/CICE? i.e., are the observation operators just interpolating the model state to the observation location?

At present, our model system directly maps the model state variables to observed parameters. We think that this is a fair assumption for in-situ hydrography and sea ice concentration from passive microwave imagery. For sea surface temperature as observed by infrared and visible band imagery, we acknowledge a mismatch during skin temperature and upper model layer temperature. Due to the far north location of the model domain, this mismatch is limited but visible during spring and summer time. We are currently working on the implementation of an observation operator for the skin temperature. During most of the time however, our model domain is either well-mixed due to wind forcing or experiencing relatively low levels of radiative forcing that drive the differences between upper layer and skin temperatures.

- Line 291 Why are the SST validated for these regions (defined in Fig. 6) separately? Are there any important implications from Fig. 7? I suggest adding one or two sentences briefly discussing the results from Fig. 7.

The used areas for SST validation characterize areas with distinct hydrography, e.g. water masses, water depth and sea ice climatology. The same areas are used in a validation of the Arctic CMEMS product. We will add this explanation in the text.

- Line 304 What kind of failures in DA are specifically referred to here?

Most common have been IT-related issues (e.g. lack of sufficient memory allocated to the EnKF task), and lack of exceptions dealt with in the queuing system (e.g. failures of the scheduling system and human mistakes during code updates) . At few instances, the EnKF-c software failed due to observations or model states that resulted in the software exiting.

- Line 347-349 I suggest including the reference(s) for the rank histogram, e.g., (Hamill 2001). Hamill, T.M. (2001) Interpretation of rank histograms for verifying ensemble forecasts.

- Lines 353-355: I suggest incorporating more descriptions on how the reliability diagram is generated, and the way to interpret the reliability diagram.

- Line 366-367: Similar to the previous comment, e.g., why does the reversed S-shape indicate low ensemble spread?- Line 374-396 (general comment for section 5.4)

We will add these references and a more detailed explanation of the ensemble verification methods.

It is interesting to also discuss the analysis increment for the unobserved variables, e.g., ocean current.

The increments for unobserved variables, such as current, will be discussed in a revised manuscript given an example similar to the analysis increments for observed variables in Fig. 12.

- Line 384 I suggest being more specific here. For example, revise "The correlation for SST is..." to "the correlation between ... and ... is"

We will rephrase this part to be more specific.

- Line 481-483 While I do not insist on conducting more DA experiments to address the following issues in this paper, it would be helpful to include some of the discussions regarding the following questions: (1) Could a larger inflation factor lead to a better DA and forecast performance? (2) Exploring whether techniques to address insufficient ensemble spread, e.g., see the list below, can be effective would be an interesting experiment in the future work as well.

During DA experiments we have seen that larger inflation factors for sea ice variables do improve the validation metrics for sea ice concentration. However, as a result of the inflation we also get artifacts in other variables such as sea ice thickness and current velocities. We therefore keep inflation to relatively low levels.

Addressing insufficient ensemble spread in sea ice concentrations, we do not have a final conclusion how this could be improved but are currently experimenting with various configuration setups for the EnKF. We would like to mention some of these strategies in a revised paper, but are not yet ready to provide results. In particular, we have recently noticed that thinning of sea ice concentration observations in the DA step may improve ensemble spread.

(3) Since SIC is a bounded variable, assuming Gaussian error for SIC is inappropriate especially when SIC value is close to the boundary (i.e., zero or one). This non- Gaussianity can make Gaussian DA method, like EnKF, sub-optimal. It would be helpful to check the ensemble distribution of SIC at a single location before and after DA in a single DA cycle, (1) when SIC observation is close to the boundary, e.g., SIC = 0 or 1, (2) when SIC observation is away from the boundary, e.g., SIC = 0.5. Using some non-Gaussian DA techniques (e.g.,

Bishop 2016; Poterjoy 2016; Hu and van Leeuwen 2021; Anderson 2022; Chan et al. 2023, etc) can alleviate this problem and may improve the assimilation and the forecast of SIC.

We follow your suggestion to include a brief analysis of the local distribution of SIC in the ensemble, for the background and analysis. An example of SIC distribution will be provided in a new Figure. We see gaussian-like distributions for intermediate ice concentrations in the marginal ice zone, and certain spread at low ice concentrations. At SIC close to 1, gaussianity of SIC is often broken and we expect deficiencies in the DA method in these areas. It is also worth mentioning that the EnKF may provide increments with SIC > 1 or SIC < 0 , and in those cases we limit the increment such that 0 < SIC < 1. We include a remark on methods to deal with non-Gaussian variables (suggested references) and these are considered in our future model development.