Review of “Novel Arctic sea ice data assimilation combining ensemble Kalman filter with a Lagrangian sea ice model” by S. Cheng *et al.* submitted to “The Cryosphere”

Review by François Massonnet (francois.massonnet@uclouvain.be)
The authors can contact me if clarifications are required.

Note: I have not read the other referee’s report before writing mine, for the sake of independency.

In this study, Cheng and co-authors document the results from a series of numerical experiments conducted with the Lagrangian sea ice model neXtSIM and constrained by the data assimilation (DA) of sea ice concentration (SIC) and/or sea ice thickness (SIT) at two temporal frequencies (1 day or 7 days) with a deterministic Ensemble Kalman Filter. Specifically, the paper reports a case study over winter 2019-2020. The inherent difficulty of performing data assimilation of Eulerian observations in a Lagrangian model is overcome by remapping the models output before the DA step. Different aspects of model performance are discussed.

The novelty of the manuscript resides, I believe, in the first-ever application of an advanced DA method to a Lagrangian sea ice model (note that I am not aware whether DA methods have ever been applied to Lagrangian ocean models, but that could be worth a quick literature review to also position this paper with the literature on that regard). Therefore, I think the paper is eventually suitable for publication. However, I have several concerns and questions that I would like to raise and see addressed, before the manuscript is accepted.

First, I remember from discussion with several authors some years ago, that clean DA with Lagrangian models posed extremely challenging methodological questions as it is clearly not obvious to update an ensemble when each forecast “lives” on its own mesh. I somehow understand that the approach followed here is a fallback solution (and that’s perfectly valid), but I think then that it would be useful for the readership to report on the negative results, what has been tried, etc. so that other groups know that this is not an easy problem. Along the same line, I’m a bit skeptical about the title “Novel Arctic sea ice data assimilation…” because many other groups follow exactly the same approach, namely, remapping the model output to observational space before doing the assimilation. I think the novelty here is to use a new type of sea ice model, hence, I would suggest to place the “novel” besides “Lagrangian” (or to drop it).

Second, as it stands, the manuscript does not offer much physical insights, an aspect that I would expect to feature in a journal like The Cryosphere (otherwise the paper is fine for GMD). Currently, the paper is outlined mostly as a development work: there is a Lagrangian model, a data assimilation method, the two are put together, and measures of performance indicate that the DA works. It would be good, not only for the authors but surely for the entire community, to understand what causes those improvements or lack of improvements. For example, I would have expected to see maps of correlations (in ensemble space) of SIC x SIT, to understand why the SIC7 experiment does not reduce the SIT biases (Fig. 7, row 4), a somewhat surprising result given that other studies have suggested that such cross-improvements are possible. I have seen that this map of correlation was "not shown" in the discussion, so I suspect that the authors have it. Another example: Fig. 6a demonstrates the added value of SIT assimilation. In view of the maps in Fig. 7, it looks like this is possible thanks to a basin-wide reduction of SIT but also to a better representation of sea ice in the transpolar drift area. Do we know why the free model overestimates thickness initially? Can we quantify the process that the DA corrects for, based on those maps? A few more diagnostics (e.g., volume of sea ice created after the assimilation, see e.g. (Mathiot et al., 2012)) would be useful in that respect.

Third, the study is based on experiments spanning half a year (October 2019-2020). I understand that there is an inherent constraint imposed by SIT data unavailability during summer months. Nevertheless, SIC is available throughout the year and this is precisely in May to September months (see, e.g. Fig 1b of (Massonnet et al., 2015)) that SIC assimilation alone might benefit SIT state estimation. The paper would really gain in impact if the SIC7 simulation would be extended until 17 October 2020. I know that it will be difficult (if not impossible) to verify the impact on SIT due to the lack of data during the melting season, but even a SIC performance analysis would be welcome. More generally, it would be good if the paper could cover more than one full annual cycle to make sure that the results are not specific to that 2019-2020 winter. I would propose to amend the title by adding "winter 2019-2020" if the authors choose to not perform those extra experiments.

Fourth, one of the advantages of neXtSIM is its rheology and I am wondering why the paper does not report on any deformation-like metrics, or on linear kinematic features density, etc. Given the improvements in simulated SIT, one could expect such metrics to be better in the DA experiments involving SIT assimilation.

Finally, I have made a few other points (below). I would encourage the authors to implement these changes (along with those mentioned above) to make the manuscript more impactful and relevant for a wide readership. Currently, my own feeling is that it sometimes resembles more a technical report with interesting results, than a paper immediately ready for publication.

Other points

- Line 12: please clarify/rephrase what is meant by "bivariate improvements between SIC and SIT"
- Line 55: "predicting sea ice is more of a boundary condition than an initial value problem". I would be a bit more cautious here. I assume that by "boundary" the authors mean "atmospheric forcing", but in fact, "boundary condition" may mean external forcing from a climate point of view. See, e.g., Blanchard-Wrigglesworth et al. (2011).

- Line 85-87: “The neXtSIM model […] shows remarkable performance”: while I’m convinced that neXtSIM is a great model, the use of “remarkable” is somehow outside what can be expected from a scientific text.
- In relation to the previous comment, I find it odd that the authors cite the Hutter et al. (2022) article but not the companion paper (Bouchat et al., 2022). As I understand the two papers, the Bouchat et al. contribution demonstrates that sea ice model performance for deformation rates is rather independent from the underlying rheological assumptions; while the Hutter al. contribution shows superior performance of the MEB rheology (on which neXtSIM is based) for linear kinematic features. That illustrates that the notion of “performance” always bears some degree of subjectivity – at least in the choice of metric – so that some nuance is always beneficial.
- Line 111-115: The model is forced by an output from IFS, but I could not decide based on the text whether this IFS output is constrained by observations or not. I assume it is since the authors attempt to reproduce observed sea ice conditions for a particular winter. Are the authors then using the output of the ERA5 reanalysis, based on IFS? Clarifications would be welcome.
- Line 172: What exactly does “the sea ice model is nonlinear” mean? Nonlinear in what input?
- Line 180: Regarding the perturbations: I understand that these perturbations bear a spatio-temporal covariance structure, which is a good choice. But do they also bear covariance across variables? Also on that point, why are short-wave radiation, 2m dewpoint temperature, mean sea level pressure, and liquid precipitation not perturbed as well?
- Line 209: add “during the winter season” because the melting can be driven by the atmospheric forcing during spring and summer.
- Line 210: “Recalling…” is not a sentence.
- Line 214: I’m unclear what is the treatment of snow in the model after the DA step, and why it is not included in the list of updated variables. Snow is an important physical parameter that sets the conduction fluxes in winter. If the SIC and / or SIT biases are corrected but the snow depth is left unchanged, it can cause sub-optimal performance of the assimilation, I think. Please clarify this point.
- Line 225: Regarding the mapping procedure, it would be good to know how much interpolation error this procedure introduces to the assimilated fields. One way to do this would be to make a ‘dry run’, i.e., (1) take the SIC and SIT of one member, (2) interpolate them to the observed grid, (3) interpolate them back to the native member mesh, (4) compute statistics between the original SIC and SIT fields and the SIC and SIT fields that have undergone the back-and forth interpolation. To what extent can this interpolation error be included in the DA uncertainty specifications?
- Line 255: What is the physical basis for a radius of localization of 300 km? Does this correspond to a typical scale of spatial variability for SIC, SIT, or both? Please review the studies of, e.g., Blanchard-Wrigglesworth & Bitz (2014) and Lukovich & Barber (2007)
- Line 279: While I understand the principle, I’m unclear how in practice the consistency check is done. Could you be more specific (or document the code in the Appendix?)

- Line 306: What does “noticing that the spread saturates for ensemble sizes above 40” mean? I’d be surprised that nothing more can be learned by adding more ensemble members. The sentence seems to imply that a 41st ensemble member would necessarily be a linear combination of the first 40, but that’s not what I think the authors mean.
- Line 310: “quasi-independent”. I’m not so sure about this, because the IFS output forcing the model has been run with prescribed SIC (and possibly SIT? Not sure) conditions, so that when IFS output forces neXtSIM, it re-introduces observed sea ice information although implicitly.
- Line 372: “expected from Lisaeter et al. 2003”: please clarify or re-explain why this is expected.
- Fig. 4: A few readers won’t be clear how to interpret positive values for underestimation (Fig. 4d). I would clarify this in the caption.
- Line 396: “This is typical of a ‘healthy’ ensemble that the ensemble forecasts and their ensemble mean are statistically undistinguishable”. I would tend to think that in any sequence of number from any distribution, the mean could also be one of the numbers itself (except pathological cases like bi-modal distributions). My (perhaps biased) idea of a healthy ensemble is that the ensemble spread is comparable to the innovations. Clarifications would be good here on what the others really mean.
- Fig. 4: could you please ensure that the y-axis limits are the same, for easy comparison?
- Line 437: is there a way to know why the joint assimilation of SIC & SIT degrades the SIT compared to the SIT assimilation? Related to one of my main comments, some physical understanding going beyond the description of the result would bring value to the paper.
- Line 449: “the relationship between the two variables is nonlinear” → can you clarify? By showing a scatter plot, for example?
- Fig. 8. The figure is, sincerely, very difficult to interpret because the curves are so close to each other. Isn’t there a more effective way to demonstrate the impact of the DA on the simulated drift? Did you consider, for example, showing histograms of ice drift bias instead of time series? I’m not sure the temporal aspect is particularly important here since no obvious seasonality emerges.

Typos

- Line 17: add space before parenthesis
- Line 19: forecast → forecasts
- Line 49-50: “the observations […] observe” is redundant
- Line 60: “construct ensemble” → “construct an ensemble” ?
- Line 87: there is a missing reference “?”
- Line 92: same
- Line 278: missing space before “Especially”
- Fig. 4 caption: “Extend” → “Extent”

Blanchard-Wrigglesworth, E., & Bitz, C. M. (2014). Characteristics of Arctic Sea-Ice Thickness Variability in GCMs. *Journal of Climate*, *27*(21), 8244-8258. https://doi.org/10.1175/JCLI-D-14-00345.1

Blanchard-Wrigglesworth, E., Bitz, C. M., & Holland, M. M. (2011). Influence of initial conditions and climate forcing on predicting Arctic sea ice. *Geophysical Research Letters*, *38*(18), n/a-n/a. https://doi.org/10.1029/2011GL048807

Bouchat, A., Hutter, N., Chanut, J., Dupont, F., Dukhovskoy, D., Garric, G., Lee, Y. J., Lemieux, J.-F., Lique, C., Losch, M., Maslowski, W., Myers, P. G., Ólason, E., Rampal, P., Rasmussen, T., Talandier, C., Tremblay, B., & Wang, Q. (2022). Sea Ice Rheology Experiment (SIREx) : 1. Scaling and Statistical Properties of Sea-Ice Deformation Fields. *Journal of Geophysical Research: Oceans*, *127*(4), e2021JC017667. https://doi.org/10.1029/2021JC017667

Hutter, N., Bouchat, A., Dupont, F., Dukhovskoy, D., Koldunov, N., Lee, Y. J., Lemieux, J.-F., Lique, C., Losch, M., Maslowski, W., Myers, P. G., Ólason, E., Rampal, P., Rasmussen, T., Talandier, C., Tremblay, B., & Wang, Q. (2022). Sea Ice Rheology Experiment (SIREx) : 2. Evaluating Linear Kinematic Features in High-Resolution Sea Ice Simulations. *Journal of Geophysical Research: Oceans*, *127*(4), e2021JC017666. https://doi.org/10.1029/2021JC017666

Lukovich, J. V., & Barber, D. G. (2007). On the spatiotemporal behavior of sea ice concentration anomalies in the Northern Hemisphere. *Journal of Geophysical Research: Atmospheres*, *112*(D13). https://doi.org/10.1029/2006JD007836

Massonnet, F., Fichefet, T., & Goosse, H. (2015). Prospects for improved seasonal Arctic sea ice predictions from multivariate data assimilation. *Ocean Modelling*, *88*, 16-25. https://doi.org/10.1016/j.ocemod.2014.12.013

Mathiot, P., König Beatty, C., Fichefet, T., Goosse, H., Massonnet, F., & Vancoppenolle, M. (2012). Better constraints on the sea-ice state using global sea-ice data assimilation. *Geoscientific Model Development*, *5*(6), 1501-1515. https://doi.org/10.5194/gmd-5-1501-2012