Summary and Decision:

The manuscript “Assimilating CryoSat-2 freeboard to improve Arctic sea ice thickness estimates” by Sievers et al. presents a new study in which satellite-derived radar freeboard (FB) from the Alfred Wegener Institute (AWI), and sea ice concentration (SIC) from the Ocean and Sea Ice Satellite Application Facility (OSI-SAF) are assimilated into the CICE sea ice model in the Arctic, between the period 2018-2020. To benchmark the improvements gained from assimilating FB, comparisons are made to an experiment which assimilates only SIC, and another experiment in which no assimilation is performed. RMSE validation across the three experiments show that modelled FB is improved by assimilating FB and SIC observations, while no improvement in FB is obtained by only assimilating SIC. Comparing observations of sea ice thickness to thickness from the FB assimilation experiment shows that the representation of thicker ice is improved for a test case in March 2020. On the other hand, a snapshot example over the same period suggests that sea ice thickness after FB assimilation is now too low in the Canada basin. Comparisons of sea ice draft are also made with 3 separate moorings from the Beaufort Gyre Exploration Project (BGEP), where sea ice draft in the FB assimilation experiment is consistently improved over the 2018-2020 period, relative to the SIC only assimilation and reference experiments.

The notion of assimilating radar FB, as opposed to sea ice thickness, is well motivated, given the large uncertainties involved when converting FB to thickness, and the authors provide a good overview of this topic in the opening sections. I do have some concerns however relating to the clarity of the methods and the rigor of the validation, which I feel need to be addressed. The methods section in particular is difficult to follow, and the lack of details on the model experiments mean that reproducibility is an issue. Relating to the validation, at present it is difficult to say how well the assimilation is performing in a) different regions of the Arctic, and b) different times of the year. For example, is the thin ice in the Canada basin after FB assimilation a systematic feature throughout the 2018-2020 period? Or does this just occur in the one snapshot? I like the comparison to BGEP moorings as this shows a clear win for the FB assimilation at these locations. It would also be useful however to see e.g., monthly-mean spatial RMSE plots and time series comparisons (see some of my suggestions below). On this note, I’m also unsure why the authors have limited themselves to such a short period (2018-2020), when both CryoSat-2 and BGEP data are available back to 2010. I would strongly encourage the authors to extend their study to this full period in order to give more confidence that modelled thickness is indeed improved by assimilation of FB. I realise that this would create significantly more work and so may be an unrealistic request. Perhaps if some additional analysis shows convincingly that the assimilation is doing a good job between 2018-2020, then extending to 2010 will not be necessary.

In any case, I feel there is a fair bit of work needed before I can recommend this manuscript for publication. Therefore, I recommend major revisions for this article. My thanks to the authors for their work and I look forward to reading the next version!

General Comments:

Introduction

- The authors have done a good job at summarising the various uncertainties/assumptions related to deriving sea ice thickness estimates (L26-86), however one key piece of missing information is the choice of retracking algorithm. The roughness characteristics of the sea ice cause different degrees of scattering of the radar echo, which are then convolved to produce an average height of the snow-ice interface. A retracking algorithm which does not account for changes in scattering due to roughness may therefore produce a freeboard which is too high when sea ice roughness is high, and vice versa. Landy et al., 2019 for example have shown how the use of a 'physically-based' retracker can help mitigate these effects. I think the introduction section here should include a few sentences to highlight this as a source of uncertainty in sea ice thickness estimates.
- L81: I'm a bit wary of saying that by assimilating FB, the effects of snow thickness and density errors are eliminated. Sea ice radar FB assumes that the radar echo is returned from the snow-ice interface, and this generally is not the case (e.g., Willatt et al., 2011; Nab et al., 2023). To appropriately model the scattering surface of the radar echo (and hence reduce uncertainty in FB) we need to account for snow thickness, density and other dieletric properties of the snow. Maybe just worth highlighting this as another source of uncertainty in satellite-derived sea ice thickness.
- Figure 1 (and others throughout the manuscript): I suggest changing colours from red and green (in this case, the BGEP locations) to something more colour-blind friendly.

Methods and data

- I find section 2.2 a little hard to follow and am also struggling to relate it to section 2.6. Is it essential to have these as separate sections? Can section 2.2 not be merged in with section 2.6? In any case, it would be useful to provide more details about the various model runs and how they were initialised etc, and potentially updating figure 2 with more information. For example, what are 'VAR' and 'VARI'? I will summarise what I think I understand, and please correct me if I'm wrong:
    - An initial experiment was run between 1995-2020. This experiment was run as an 80-member ensemble in coupled ice-ocean mode, and forced by ERA5 atmospheric reanalysis
        - What were the initial ice/ocean conditions for this run?
        - How do you e.g., perturb the ice/ocean model parameters to create the ensemble?
        - I'm not sure what is meant by increasing the variance to "account for biases" (L138). Are you not just increasing the variance to prevent ensemble collapse? Ultimately, you're hoping that the data assimilation itself will reduce the biases
    - The 2018-2020 period of the initial experiment corresponds to the refRun

    - The initial experiment at 2018-01-01 was used as initial conditions for both the sicRun and the fbRun. Assimilation over the 2018-2020 period is performed every 7 days.
    - The increments from the assimilation runs are then saved, and then you effectively re-run the sicRun and fbRun experiments over the 2018-2020 period, except that the previously saved increments are now updating the model at every time step (through linear interp of the increments from 7 days to 600 sec).
        - This is to prevent model shock after each assimilation cycle?
    - I'm not sure where the 'static ensemble' fits into all of this? Could you explain?
- More generally, could you explain the motivation for only focusing on the 2018-2020 period rather than the entire CryoSat-2 period (2010-present)? By utilising the entire record I feel that you would be able to derive more rigorous statistics related to the improved FB and thickness from assimilation. For example, time series comparisons of monthly-mean FB and thickness between AWI and the assimilation run over the 2010-present period, for different Arctic regions.
- L135: Could you provide more details on why you choose observational error estimates of 15% for SIC and 0.15m for FB, given that on L175 and L166 you state that the observational uncertainties are 10% and <0.07m, respectively?
- L150-155: Is there any post-processing applied after assimilation to ensure that the updated SIC is bounded between 0 and 1? If so, how is this bounding applied? Particularly to the category terms.
- I believe currently in the fbRun you are updating SIC and FB sequentially. Out of curiosity, is sea ice thickness updated during the assimilation of SIC, and similarly is SIC updated in the assimilation of FB? Do you also expect your results to differ if you first assimilate FB and then SIC?

Results

- L269-270: Suggest clarifying here that by "assimilation period" you mean November-March, as opposed to the whole 2018-2020 period.
- L287: Does "beginning of October" and "end of winter" refer to a single day? Or a weekly average? Please clarify.
- Figure 5: Suggest including either correlation or R^2 values for both refRun and fbRun on each plot to quantify the improvement. I also suggest making this a 4-panel figure, and show the equivalent density plots for freeboard as well.
- L288: Can you also speculate why the "thin bias" problem is not improved with the assimilation of FB? Is this because the FB increment is spread linearly across the ITD categories (Equation 1), whereas in reality more weight should be given to the thinner categories?
- Figure 6: I don't think panels C) and D) are all that informative. Especially considering that our observational estimates of snow thickness and ice density are inherently wrong. I suggest replacing these panels with freeboard comparisons (AWI-refRun) and (AWI-fbRun). It would also be good on each panel to include a mean RMSE value, as it currently looks as if panels A) and B) might actually be similar in terms of RMSE.

- Figure 8: Could you speculate why the differences between AWI and fbRun are significantly large at the beginning of the 2018-2019 period? Is this a spin-up issue?

Discussion

- L400-410: Is the imprint of the FYI/MYI zone on the difference plot not just due to the fact that the AWI snow thickness and ice densities are assigned constant values depending on whether FYI or MYI? In your case you've replaced the density with a linear function after Mallett et al., 2020, but ultimately this imprint of a FYI/MYI mask is just a reflection of what the observations are (hence my suggestion above to remove).
- L412: The panels in Figure 6 are errors in modelled sea ice thickness relative to AWI, not uncertainties in AWI sea ice thickness, right?

Minor suggestions:

Abstract

L4: Define SIC and FB acronyms
L6: Define AWI acronym
L10-11: This last sentence is rather vague. I suggest something like "Modelled sea ice draft errors are in good accordance with that of CryoSat-2 errors at BGEP mooring locations, with mean error differences less than 3 cm over the 2018-2020 period."

Introduction

L17: I'm not sure what is meant here by "need to affect the model variable that the assimilation aims to improve". Does it mean that we need to use an appropriate observational operator to map the model state variable which is being updated, to the space of the observations? E.g., an operator to map thickness to freeboard?
L18: I would clarify here that this statement relates specifically to Arctic sea ice predictability on seasonal-to-interannual timescales.
L24: Suggest stating explicitly that initialising thickness is better for predicting Arctic sea ice on seasonal time scales or longer (sea ice area persistence is more important at short lead times).
L55-57: Would also make reference here to the recently developed SnowModel-LG (Liston et al., 2020; Stroeve et al., 2020), which is being adopted in sea ice thickness products from e.g., Landy et al., 2022.
L65: The phrasing "add up" here feels a little vague. Does this mean that the error in sea ice thickness is equal to a linear sum of the errors in snow/ice/water density and FB?
L71: Suggest rephrasing to "The OSI-SAF ice type product (Aaboe et al., 2021) is one observational data set which aims to distinguish between FYI, MYI and ambiguous ice types."
L76-77: The wording is a bit confusing here. Does it mean that the errors are systematically overestimated in the MYI zone and underestimated in the FYI zone in the OSI-SAF product?
L77-78: As far as I'm aware, sea ice area isn't required to generate thickness anyway, so this sentence seems redundant. Could you explain what you mean here? Unless you're referring to CryoSat-2-derived sea ice volume?

Methods and data

L96: First time using PDAF acronym, please define.

L99: Suggest rephrasing to "An increment is the amount of change in a model state variable after the assimilation of observational data."

L106: First time using WMO acronym, please define.

L109: Suggest removing first sentence on L109 and changing L111 to "the key variables are snow thickness (h_s), snow density (rho_s), sea ice density (rho_i), and ocean water density (rho_w)." Or similarly just defining the terms explicitly on L109.

L142: 80 model states? Or 80 ensemble members?

Results

L299: Suggest clarifying that the date 03-30-2020 is actually a 7-day mean.

L315: change to "independent of the satellite-derived FB data".

L330: change here and elsewhere in manuscript from "data is" to "data are"

Discussion

L410: I'm not sure on Crysophere reference guidelines, but you might need to change "Sievers et. al (in preparation)" to something like "Future work will analyse the effects of different variables…"

References

Landy, J.C., Tsamados, M. and Scharien, R.K., 2019. A facet-based numerical model for simulating SAR altimeter echoes from heterogeneous sea ice surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(7), pp.4164-4180.

Willatt, R., Laxon, S., Giles, K., Cullen, R., Haas, C. and Helm, V., 2011. Ku-band radar penetration into snow cover on Arctic sea ice using airborne data. *Annals of Glaciology*, *52*(57), pp.197-205.

Nab, C., Mallett, R., Gregory, W., Landy, J., Lawrence, I., Willatt, R., Stroeve, J. and Tsamados, M., 2023. Synoptic variability in satellite altimeter-derived radar freeboard of Arctic sea ice. *Geophysical Research Letters*, p.e2022GL100696

Liston, G.E., Itkin, P., Stroeve, J., Tschudi, M., Stewart, J.S., Pedersen, S.H., Reinking, A.K. and Elder, K., 2020. A Lagrangian snow-evolution system for sea-ice applications (SnowModel-LG): Part I—Model description. *Journal of Geophysical Research: Oceans*, *125*(10), p.e2019JC015913.

Stroeve, J., Liston, G.E., Buzzard, S., Zhou, L., Mallett, R., Barrett, A., Tschudi, M., Tsamados, M., Itkin, P. and Stewart, J.S., 2020. A Lagrangian snow evolution system for sea ice applications (SnowModel-LG): Part II—Analyses. *Journal of Geophysical Research: Oceans*, *125*(10), p.e2019JC015900.

Landy, J.C., Dawson, G.J., Tsamados, M., Bushuk, M., Stroeve, J.C., Howell, S.E., Krumpen, T., Babb, D.G., Komarov, A.S., Heorton, H.D. and Belter, H.J., 2022. A year-round satellite sea-ice thickness record from CryoSat-2. *Nature*, *609*(7927), pp.517-522.