Response to reviewer 1 (JF Lemieux)

Dear Jean-François,
Thank you very much for your helpful comments. We will respond to them below.

Review of Presentation and evaluation of the Arctic sea ice forecasting system neXtSIM-F by Williams et al.
In this paper the authors present and evaluate a new sea ice forecasting system based on the neXtSIM sea ice model. They have evaluated drift, sea ice thickness and concentration against different datasets and show that the forecasts, in general, show good agreement with the observations. Being also in the business of ice-ocean predictions, I recognize the amount of work that was done by the authors to develop and evaluate this kind of forecasting system.

I have, however, two major concerns about the paper.
The first one is easy to address. I think the authors need to do a better literature review. We have the impression there are currently only two sea ice forecasting systems (Topaz and GOFS) in the world and that neXtSIM-F is the third one to be proposed. For example, the Danish meteorological institute, the UK Met office, ECMWF, Environment and Climate Change Canada (ECCC) all have operational sea ice forecasting systems. Note also that we (ECCC) have, a few years ago, developed a similar system as neXtSIM-F: a stand alone sea ice model coupled to a slab ocean model (with MLD and SSS initialized from a coupled ice-ocean prediction system and SST from an analysis...). It is not used anymore as it has been replaced by a coupled ice-ocean forecasting system but I think, given the similarities between the two systems, that it should be mentioned. You could then describe what is different (e.g. the ocean heat flux correction you propose...which is interesting by the way.)
We agree that our discussion gives the wrong impression of the field, and are quite sorry about this. We have now scaled back our discussion to refer to the papers of Tonani et al (2015), and also the more recent one by Hunke et al (2020, https://link.springer.com/article/10.1007/s40641-020-00162-y) which also gives an idea about trends in the systems. Thanks for the reference about the ECCC stand-alone sea ice forecast RIPS that you sent - we have included it in our introduction as well.

The second major concern I have is that neXtSIM-F was evaluated for only half a year. And especially, the evaluation was done during the winter months, that is at a time when there is almost no navigation in the Arctic. I would understand why you would choose this period if your system was designed specifically for the Baltic Sea or the Gulf of St-Lawrence (where there is a lot of navigation during winter). I have the impression that the authors have submitted their paper too early and that it would make more sense to show a complete seasonal cycle of the forecast scores. You mention, anyway, that neXtSIM-F will be operational in November of this year...this is coming soon. This means you will soon have the evaluation for the summer months? Then I really think you should include these.

We agree that this was a major short-coming of the previous version of the paper, and have now completed evaluation for a period of 20 months. neXtSIM-F finally went operational in July 2020, and will be updated in December with the version evaluated in the present paper.

This paper will be an important contribution to the field of sea ice forecasting but first the authors need to address these concerns and the minor comments given below.

1 Minor comments

1) p.5 line 19. I guess you use a turning angle for the ocean-ice stress? Please mention it.

Yes, it is 25 degrees. Added to paper (section 4.2)

2) p.5 line 22. this is already mentioned above.

We have simplified this section and added it to section 4.1, where the slab ocean is introduced.

3) sections 3.6-3.7. What is this the atmospheric forcing you will use once the system becomes operational? I hope it will be the same one used for the evaluation because then it does not make sense.

The forcing we will use is the ECMWF forecast described in section 3.6. A line has been added to clarify this.

4) p.8 lines 7-9. You give information that is not needed. We don't need to know that it was first coded in Matlab. Just say that the version of neXtSIM used for neXtSIM-F is described in Rampal et al. 2016 and Samake et al 2017.

OK - we have removed some of the history.

5) p.8 lines 13-14. Mention that the MLD varies spatially and that this spatial field is fixed (I guess) during the 7 days of the forecast.

It actually varies with time according to the ocean forecast. We have added a comment to clarify this.

6) section 4.2. Mention how you initialize the sea ice velocity.

We added a sentence about this: the ice velocity and also the damage are started from zero.

7) section 4.2. What justifies the values of 0.2 and 0.8 for the initial thin and thick ice?

This is a little arbitrary (necessarily so given the difficulty of determining the concentration of thin ice) but the model wasn't too sensitive to this.

8) p.10 line 19. The first condition (i.e. ct > 0) is not needed, right?

Yes you are right - we have removed it.

9) section 5.1. Clearly state how long (time period) is the free run.

10) p. 11, line 6. Typo "the no ice".

Fixed.

11) Fig. 3. Explain how you calculate the concentration and uncertainty. Is it the mean over the whole domain? Then it means you have many grid cells with a concentration of zero and many with a concentration close to 1.0. This means the signal you are interested in is kind of buried because most grid cells have a forecast concentration close to the observed one...which is not surprising.

We actually average over the union of the ice masks. This still leaves the pack ice, which in our case is usually close to 1, while the observations is usually about 0.9-1. So this signal is less interesting since the observations are also uncertain. The extent is actually the main signal we would like to capture, but we present the comparison of the concentration also.

How do you deal with uncertainties when the concentration is 1 or 0...it cannot be gaussian, right?

This is a good question. It turns out that it was quite difficult to estimate uncertainties since we did not know the distribution of the observation errors. Therefore we removed the error shading from most of the plots. For the error plots we kept it as a rough reference level to show when our results were significantly different from the observations. Another interesting thing that came up from the uncertainty analysis was with the drift speed - adding noise to a given set of “true” drift values introduced a bias of 2\sigma^2 to <drift^2> where \sigma^2 is the variance of the drift. Hence one should be a bit careful when comparing drift speeds.

12) p. 13, lines 6-7 and Fig. 6. Is it possible the MEB rheology leads to too much convergence? (e.g. North of the CAA).

Yes. It is slightly improved with BBM, but there are still issues with this near the northeast coast of Greenland and the north coast of Svalbard. This has some unfortunate follow-on effects on the summer ice extent in the Greenland Sea especially.

13) p. 13, lines 18-24. Too many numbers given. I don’t think you need to give all these values.

We have reduced the number of numbers inline throughout.

14) p. 13, line 27. What do you mean by "eroded"? Please clarify.

We mean that by setting a maximum threshold on the uncertainty in the ice drift we had masked the observations in some areas (mainly near the ice edge, coast, and the north pole). We now generally use a higher threshold (10 km/day) so that we can include evaluation in summer, but include a table where we use the lower threshold (1.25 km/day) for 2 winters and show it reduces our errors significantly. However some of the regions with higher errors (eg coast and MIZ) are regions where we have some problems with other variables like thickness and extent so masking them out can help for that reason too.

15) p. 13, line 33. I don’t think the ice is landfast there but it is (very) slowly drifting.

Yes, you are right.

16) p. 14, line 1. remove "very respectable"...just give the number and that’s it.

Fixed.

17) p. 14, line 10-14. Please clarify this paragraph. It is not clear here what are exactly the experiments (especially the sentence"...without assimilation, with assimilation of concentration and with assimilation of concentration and thickness.").

We have now reduced the number of experiments to one free run and one with assimilation of concentration/extent.

18) Fig. 9. I am not surprised persistence is doing so bad here because the beginning of November is a time when there is a lot of ice growth. No wonder the model performs so well. I think it would be good to show another case for example in March. Is it always true that the model beats persistence? Is is always true that assimilation improves the quality of the forecast?

Assimilation always improves the forecast concentration and extent compared to the free run (see the new figure 9). However it only beats the persistence in extent in the months Sep-Feb. We now have 4 examples (2 with positive skill (Jan, Sep) and 2 with negative skill (Mar, Jun)).

19) p. 16, lines 13-14: The captions in Fig. 10 and 11 do not match the text here about the dates.

Figs 10 and 11 have now been removed.

20) p. 18, line 1: East Siberian and Chukchi seas...not obvious to me when I look at Figs 10 and 11.

These figures and the discussion have now been replaced.

21) Fig. 13 and 14 are not discussed. Are they really needed?

They have in fact been replaced.

22) p. 19, lines 11-13: Please rephrase…

(This referred to the evaluation against SMOS thin ice thickness.) We decided to remove evaluation against SMOS since errors were dominated by difference in extent.

23) p. 20, line 1. What drifters are we talking about here?

Every day at 12:00 we place Lagrangian drifters at the grid points of the OSISAF drift product and advect them for 48h at the ice velocity. The total drift is then compared to the OSISAF drift product. We clarify this in the Data Sources section.

24) p. 25, lines 11-17. I think it could also be the thermodynamic model itself. There is clearly too much ice growth...the model might require some tuning. This should be mentioned. What about the way the thin and thick ice categories are initialized? How does it affect the overestimation of the growth?

We have now removed the comparison to SMOS.

25) p. 25, line 19. What do you mean by "fragmentation"?

Technically we mean damaging. We have changed this in the paper.

26) p. 25, line 23. I thought neXtSIM is using our grounding scheme for landfast ice? Would it be worth tuning the k1 parameter?

We now have the opposite problem of too much fast ice in the Laptev and East Siberian sea. We are currently tuning this parameter, but for this paper we were concentrating on the main parameters of the dynamics (cohesion, Pmax, drag)

27) p. 26, lines 20-22. Is it a result you presented in this paper? Or I just don't understand the sentence. Please clarify this.

Yes it is presented. We clarify the different runs better now (eg free run v forecasts)

Congratulations for your work on developing this new sea ice forecasting system.

Jean-François Lemieux

Thank you.