

## Anonymous Referee #2

Dear Reviewer 2,

Thanks very much for your helpful comments. We will respond to them below.

## Interactive comment on “Presentation and evaluation of the Arctic sea ice forecasting system neXtSIM-F” by Timothy Williams et al.

Dear anonymous reviewer, thank you for your thorough review and helpful comments. We will respond to them below.

*In this paper the authors introduce a new sea ice forecasting system neXtSIM-F based on the neXtSIM sea ice model and present an evaluation of the model over a single season - winter 2018-19.*

*I feel that this study will be worth publishing in The Cryosphere (although it would likely fit better in GMD than TC). However several changes will be required before this is possible.*

### General comments

1. *It is not made clear enough what the various runs and systems are that are assessed in this study. In particular there is also no mention of the “free run” before it is evaluated in Section 5.1. Section 3 contains information on the observational datasets used in this study but there is no equivalent for the model datasets. This study needs a summary of exactly which runs and systems are being evaluated with perhaps a table.*

*This has now been clarified in a few places (introduction, start of results section, discussion and conclusions).*

2. *Additionally the names neXtSIM and neXtSIM-F seem to be fairly inter-changeable in the manuscript. I guess the neXtSIM-F forecasting system uses the neXtSIM sea ice model. If so then I think the name of your forecasting system as neXtSIM-F is a bit confusing.*

*This has now been clarified in a few places.*

3. *The evaluation period is only a few months and does not include the late spring/summer period when many sea ice forecasting systems report their poorest performance. This means that it is hard to put the evaluation here into context with other operational systems. The conclusions of this study would be much strengthened if the authors could perform, and evaluate, a complete annual cycle (or preferably 2).*

This is a completely valid criticism and the run is now 20 months. We also have our worst performance in the summer months.

4. *In general I find that there are too many figures and stats in the paper, which makes it hard to understand what the take-home message is.*

We have removed some figures and tables, and replaced some others, and also tried to emphasise the main take-home message more, so hopefully the new figures and tables now help understanding rather than hinder it.

## Specific comments

1. *In many places the language used in the paper is too informal and colloquial (i.e., in the abstract we have “. . .in our system, we obtain. . .” and on P9 we have “the observed ones”).*

We have tried to make the language more formal.

2. *You need to be careful to distinguish between “sea ice concentration”, which ranges from 0 to 100%, and “sea ice area fraction”, which ranges from 0 to 1 throughout this manuscript. For example in Figure 4 the caption says “concentration” but the scale is +/- 0.5%. This is either a low “concentration” or a high “area fraction”. I assumed the former to start with until I noticed that the text talks about an associated reduction in extent. With changes of +/- 0.5% concentration I wouldn't expect to see any departure to the 15% contour (extent) so is it actually “area fraction” plotted here?*

You are correct that it was actually area fraction - we have changed to using concentration with units of % everywhere.

3. *I find the abstract to be rather technical and not very abstract. It reads a bit more like a conclusions/summary section. I would encourage the authors to make the abstract more exciting to make the paper more inviting to potential readers.*

We have removed the numbers from the abstract and tried to make it more inviting.

4. *The introduction section (section #1) is rather disjointed. It starts with some motivation for sea ice forecasting (with background on changing climate) but then jumps straight in to say that neXtSIM-F is based on neXtSIM. It doesn't actually say that neXtSIM(-F) is a sea ice forecasting system! It would be better to include a couple of extra lines to say that this is the case. Perhaps to say something like “Here we introduce a new sea ice forecasting system, neXtSIM-F, that is based upon the neXtSIM model. . .”.*

Thanks - this is a good sentence that we have used in the introduction. Hopefully introduction is now less disjointed.

5. *I find the introduction to operational ocean forecasting systems in Section 2 to be, almost paradoxically, both too detailed and non-existent. I say too detailed because I am left wondering why there is such a thorough introduction provided to the GOFs system when it isn't really used in this study? Of course, GOFs is only one of many global operational ocean-sea ice forecasting systems and you don't mention any others apart from TOPAZ and neXtSIM. The Tonani et al. (2015) GODAE paper provides a nice reference describing the world's operational global forecasting systems. Although several of the systems have doubtlessly moved on since 2015, this reference provides evidence for the breadth of activity in the world. Tonani, M., Balmaseda, M., Bertino, L.,*

Blockley, E. W., Brassington, G., Davidson, F., Drillet, Y., Hogan, P., Kuragano, T., Lee, T., Mehra, A., Paranathara, F., Tanajura, C. A. S. and Wang, H.: Status and future of global and regional ocean prediction systems, *J. Oper. Oceanogr.*, 8, sup2, s201-s220, doi:10.1080/1755876X.2015.1049892, 2015.

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.

7. The data sources section (#3) does not make it clear which datasets are used for assimilation and which are used for evaluation (and hence which are used for both). At the least it is important to note which datasets are independent from the assimilation.

A sentence has been added to the start of each data-source subsection to make it explicit which are used in assimilation and evaluation, and which are used in evaluation only.

8. Related to the above point I find the description of the blended SSMIS+AMSR2 product somewhat confusing. Is this done purely for the evaluation? If not why can't the DA do this blending by waiting the observations with their respective errors?

We now use this only for data assimilation, but it unfortunately became quite inconvenient to use it for evaluation due to missing sections of data in the AMSR2 product particularly.

We clarify this in section 2.3.

9. P3, L4: “. . . profiles from Argo floats.”. Why do you only use Argo floats (if that's true)? Why not CTD/XBT/seals etc.?

Here we were only describing what TOPAZ does. However, it also assimilates ice-tethered profiles as well. Other data like the ones you mention are available too late to assimilate in near-real-time. They are used in the TOPAZ reanalysis though (Laurent Bertino, *personal communication*).

10. I do not understand why a couple of weeks of CFSv2 is used in place of ECMWF. Surely you could get the replacement data from somewhere else (like ECMWF themselves for example)? If not then you should really consider the implications of using CFSv2. Specifically: is this the configuration with unrealistic ice growth caused by the fact that they turned off the stratus cloud formation to improve tropical temperatures and ENSO predictability (as described by Yang et al. 2017 and references therein)?

Yang, Q., M. Wang, J.E. Overland, W. Wang, and T.W. Collow, 2017: Impact of Model Physics on Seasonal Forecasts of Surface Air Temperature in the Arctic. *Mon. Wea. Rev.*, 145, 773–782, <https://doi.org/10.1175/MWR-D-16-0272.1>

We have now been able to access those missing forecasts and have rerun the free run and the relevant forecasts without needing to plug with CFSv2. This also allows us to delete the section describing CFSv2.

11. I don't like your “RMSE” for extent as it is exactly the Integrated Ice Edge Error (IIEE) of Goessling et al., (2016). You cite the ensemble extension of the IIEE (the SPS paper of Goessling and Jung, 2018) and say that your RMSE is like a deterministic version of that, which is misleading. It would be better to just cite the 2016 paper instead and call your metric “IIEE” instead of “RMSE”: Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T.: Predictability of the Arctic sea ice edge, *Geophys. Res. Lett.*, 43, 1642–1650, <https://doi.org/10.1002/2015GL067232>, 2016

Agreed - we have renamed the metric to IIEE to avoid confusion and replaced the reference with your suggested one.

12. *In Figure 3 I note that the neXtSIM concentration evolution is very smooth – more so than the low resolution SSMIS data – which I didn't expect given the resolution of the model. Can you comment on this? Is this caused by the fact that neXtSIM still uses the continuum formulation and so doesn't resolve small scale features?*

For this refer to p10 “For all comparisons we average the model fields in time over an appropriate time window (in practice 1, 2 or 7 days), apply some spatial smoothing (being guided by the size of the satellite footprint), and interpolate onto the observation grid.” When comparing to OSISAF SSMIS we average over 1 day and apply some spatial smoothing, as SSMIS is too coarse to resolve the cracks/leads etc. However, we have applied a bit less smoothing this time around (see fig 11).

13. *I note with interest that MOSAiC forecasting is mentioned as a motivation for improving sea ice forecasts. There is an international project (SIDFEx) currently coordinating operational sea ice drift forecasts specifically to provide guidance to the Polarstern/MOSAiC. Presently the list of models includes TOPAZ but not neXtSIM. Are there any plans to contribute neXtSIM drift forecasts to SIDFEx? This might be an interesting way to show the skill of neXtSIM in this regard.*

This would be interesting if it were still possible, and is worth discussing with the leaders of that project.

14. *Some of the figures (e.g., Figs 10&11) suggest that the data assimilation is having a rather modest impact on the forecasts compared with many of the operational systems that I have seen in the past. Can you comment as to why that might be?*

After 2 quite major updates to our sea ice rheology we had to change the assimilation scheme quite a lot. It is now quite conservative in that we mostly leave the model alone, except where the ice mask is incorrect. There is a noticeable improvement in the IIEE for the first days after doing this, although the improvement is diminishing with lead time.

15. *On page 20 it is mentioned that the “RMSE for drifters placed on the first day. . .” but this is the 1st time in the manuscript that drifters are mentioned. Can you explain this a bit more please?*

Every day at 12:00 we place 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.

## Figures

*As mentioned above I feel that there are too many figures in this manuscript. In particular in Figs 11 & 12 there are 12 panels and each row looks virtually identical. Apart from telling me that the assimilation is having a rather modest impact, I don't understand what I'm supposed to do with all this information.*

*Additionally the next similar set of figures, Figs 13-14, don't even seem to be discussed in the text at all. So are they necessary?*

We have revised the figures substantially now (eg figs 11-12 and figs 13-14 have been replaced.)

*Many of the figure captions are too brief and should be improved. I believe that the*

*Copernicus journal guidelines are that figures should be able to work stand-alone from the text, for which a bit more information is required.*

*We have now checked the captions contain enough information.*

*I find that the x-axis date tick-marks provided on the time-series plots (Figs 3, 5, 7, 12 – less so for Fig 9) are not very useful. With such a short run period it would be better to include more dates. At the very least there minor tick marks should be used to show each day (or 5-days or something). It would also be good to specify this in the figure caption perhaps.*

*We have tried to add enough major and minor tick marks to the time series plots, although the run is a lot longer now.*

*Figure 1 is a bit confusing because I am left wondering whether different time-scales are involved here. Is this a daily schematic or does it depict the whole run? For example the 2 top boxes (initialization) are surely not done each day are they? If not then it should be made clear what is done each day and what isn't – either in the caption or the figure (or both). Perhaps the initialization boxes could be enclosed in a dotted box or something?*

*These were snapshots, but they have actually been removed to reduce the number of figures.*

*I suggest you should also re-think your use of red-blue colour maps for sea ice concentration. I have seen people use red for less ice (as it's hotter) and blue for more ice (colder) in the past as well as the other way around. It might be better to avoid the use of a hot<->cold colour-map therefore.*

*We tried out some other colormaps but decided to keep this one for lack of a better alternative.*

## Technical corrections

*The 1st instance of “SSMIS” is correct but thereafter it has been changed to “SMMIS”.*

*Fixed*

*Also “first day (4th day)” appears in many places, which is not very consistent*

*P2, L26: CMEMS should be “Copernicus Marine Environment Monitoring Service”*

*Fixed*

*P2, L29: “. . .the version 4.1 of the. . .” – suggest to remove the 1st instance of “the”*

*Here*

*Fixed*

*P2, L30: The reference for CICE v4.1 is Hunke and Lipscomb (2010): Hunke, E. C. and Lipscomb, W. H.: CICE: the Los Alamos sea ice model. Documentation and software users manual, Version 4.1 (LA-CC-06-012), T-3 Fluid Dynamics Group, Los Alamos National Laboratory, Los Alamos, 2010*

*Thanks - added this*

*P3, L30: “As specified by the validation reports above. . .” should be “As specified by the validation reports cited above. . .”*

*Changed to this.*

*P4, L5: “metrics” should be “metric”*

*Fixed*

P4, L6: is extent “above 15%” or “at least 15%”? I thought the latter.

In our evaluation script it was >15% - in practice I would expect either to give the same result  
P4, L13: “. . .can be obtained for 48 hours. . .” sounds like only 48 hours of data. Do you mean this or do you mean the data is available 48-hours behind real-time?

In the OSISAF product, the drift is estimated daily from 12:00 on the start date until 48 hours later. In practice this does imply a time lag of at least 48h. “From October to April, however, daily 48-hour ice drift vectors can be obtained at a spatial resolution of 62.5 km.”

P5, L5: calculating volume for each model & obs based on thickness like this will involve different areas of ice won't it?

Yes. In the end we decided to remove the SMOS comparison as errors were greatly affected by errors in extent (which we struggled with).

P5, L26: “Modelling” is spelt incorrectly as “Modeling” in the NEMO acronym

Fixed

P5, L31: “. . .if the temperature is below 0C”. Which temperature – surface skin temperature, or near-surface atmosphere temperature (T2M)? Please be more specific.

It is T2M: we have added a clarification to the paper.

P10, L29: “of these variables” is not adding to this sentence and should be removed

Removed

P11, L6: “. . .predicts the no ice. . .”. Please remove the “the”.

Fixed

P11, L11: “averages values” should be “average values”

Fixed

P11, L19: “underestimation in the Bering Sea”. I presume that you mean the Chukchi Sea here because the Bering Sea is outside your model domain?

P13, L33-4: “land-fast“ should be “land-fast ice”?

Fixed

P13, L35: (as above) I suspect that “Bering Sea” should be “Chukchi Sea” here

That is correct, although with the change of simulation length the figures and discussion relating to the free run are quite changed and this sentence has been removed.

P14, L9: the title “Evaluation of forecasts with assimilation” is confusing because I doubt that you are actually doing assimilation in your forecasts are you? Perhaps this should be changed to something more like “Forecasts performed from analysed ice conditions”?

We are doing a kind of assimilation known as “data insertion” (modifying the initial conditions of the forecast) combined with “nudging” (through the compensating heat flux)

P16, L9: Do you mean “significantly” here in the scientific sense of the word? If so include a p-value, if not I suggest changing to “considerably”.

Changed to ‘considerably’

P18, Fig 9 caption: “(blue)”, “(orange)”, and “(red)” are provided but not “(green)”

This figure has now been removed.

P21, Fig 12 caption: “error bars” should be “shading”

We have removed references to error bars since we use shading.

P26, L7: “limited resources” suggests a deficiency in resource. You should change to “minimal resources” if you wish to suggest that the model is cheap to run.

We decided to remove this discussion.

*P26, L20-21: “. . . forecasts used saved atmospheric and ocean forecasts as forcing. . .”.*  
*What does this mean?*

This is mainly an issue of forecasts vs hindcast winds which we have clarified in a few places (introduction, start of results section, discussion and conclusions)