Review of “Evaluation of Arctic sea-ice drift and its dependency on near-surface wind and sea-ice concentration and thickness in the coupled regional climate model HIRHAM-NAOSIM” by Yu, Rinke, Dorn, Spreen, Lüpkes, Sumata and Gryanik.

This article contains two pieces of work, the first being an assessment of the simulation of sea-ice (in particular sea-ice drift, SID) in a recent ensemble of 10 control members from a coupled regional climate model of the Arctic. The second part is then a sensitivity test, where the parameterization of surface exchange of momentum and heat between the ice and atmosphere is improved based on recent parameterization development documented in other papers. The research is state-of-the-art and an important step forward in the broader aim of trying to improve the fidelity of coupled climate models. The representation of Arctic sea ice is a long-standing known weakness in these models and improving the surface exchange parameterization is tackling one important weakness. The results of this study are mixed. Overall the model CTRL ensemble performs reasonably well compared to observational data sets (although these themselves have deficiencies). The new parameterization acts in a physically realistic way and leads to significant changes in surface variables. However, the authors state that it does not (yet) provide an improved simulation of sea ice, because no model tuning has been carried out yet, and they reserve this for future work.

Overall this is a very commendable study and an important piece of work, so I would like to see it published. I have a number of comments that would improve the manuscript that I’d like to see the authors take on board – some on the presentation that would greatly help new readers and make the work more accessible. The results of the second part of the study seem to end before the punch line! Normally upon introducing a new parameterization, authors invariably tend to find that their new parameterization improves the model. Here we seem to stop short of a full investigation of whether this is the case or not, because some tuning is required. I think this is reasonable, because the paper is already quite long at this point, and I am aware that such tuning is time consuming and opaque. But it does make the paper feel a little unfinished. Have the authors considered making this part 1 of a two-part paper, or at least spelling out in more detail the implied follow up study?

Specific Comments

1. The paper’s title is long and a bit clumsy (three “ands”). I’d maybe try to reword.

2. In the Introduction I would recommend a short discussion on the quality of the surface exchange parameterization you’ve introduced. Around L65 you point out that parameterizations without a form drag element for momentum exchange are “poorly constrained” and that a recent observations-based form-drag parameterization has been implemented in a model by Renfrew et al. Here I think you need a few sentences pointing out that the mathematical parameterizations by Lüpkes et al. 2012 and Lüpkes and Gryanik 2015 were constrained by summertime observations over the sea-ice pack (from Andreas et al. 2010) and by limited aircraft observations over the MIZ (marginal ice zone). Then more comprehensively validated and tuned over the MIZ by a larger set of aircraft observations in Elvidge et al. 2016 [Note, this paper is not in the reference list, but there is a citation for Elvidge et al. 2018 in the manuscript, but no reference, so I think you mean the 2016 paper]. Importantly, I think you also need to point out that most of the validation and tuning has been done for momentum exchange (i.e. C_{DNi}), very little validation has

been done for heat or moisture exchange (i.e. C_{HNi}). The validation and tuning for scalar fluxes is, I think, still something of an open question.

3. Page 4 contains a mathematical description of the new surface exchange parameterization for over sea ice, based on Lüpkes et al. 2012 and Lüpkes and Gryanik 2015. I am familiar with these two papers and I think you are right to leave most of the mathematical details out of this article and refer the reader to these previous articles for details. However, what is tricky is that both of these previous articles are long and technical, with more than 60 and 70 equations in them respectively, and both contain several sets of parameterizations in a hierarchy of complexity. This makes checking the summary you have here difficult, especially as the notation used here is slightly different to the previous papers. I think you need to be more specific and say which equations from the two above papers are implemented and try to use notation that is as close as possible to what is already published (I appreciate this can be difficult). To give one example, equations (3), (4) & (7) all have a '+1', in "$z_{0,i}$+1" – this isn't explained and I don't know what it means. Also equation (8) does not seem to match equation (63) in Lüpkes and Gryanik 2015 – should it? Finally there are a number of parameters set on page 4: C_{e10}, $z_{o,f}$, β, then later on α and $z_{0,i}$. It is not clear where the values for these parameters have come from and I found it difficult to relate them to parameters in the previous studies or in Elvidge et al. 2016. I think this section (2.1.3) could be vastly improved without much additional length or detail. Finally, you don't comment on exchange of moisture, is this changed?

4. In the summary (L440) you state that the SENS simulation is not any better than the CTRL simulations, in terms of sea-ice drift etc. However, you don't really provide evidence for this statement. I think there is evidence in your paper, but you need to discuss it and demonstrate this is the case. Consequently, I would recommend adding another paragraph or two to Section 4, where you discuss the quality of the SENS and CTRL simulations. For example, is it possible to compare the gradients in Fig 6a to Fig 10 and demonstrate whether the CTRL or SENS is better? Could you add some observational data to Fig 10 to show this fact? I appreciate that 'not any better yet' is a bit of a negative result and could be changed by tuning the model, so perhaps you don't want to spend too much time and effort on this aspect. But I think you need to provide a small amount of evidence for this statement.

Minor Points

L14 – "…of the Arctic basin" [insert the]
L28 – "sea ice has experienced…"
L66 You categorise Tsamados et al. 2014 as an ice-ocean model, but that study was actually using only a sea-ice model.

L175 – Have you considered a Cryosat product for sea-ice thickness – probably not worth the effort now, but might be interesting for any follow up studies.

L180 – The description of ERAI resolution is misleading. The resolution of the atmospheric model is T255 equivalent to about 80 km resolution, and you have downloaded it on 0.25 degree grid. So please rephrase.

L185 – "resolutions, a bilinear…"
L191 – "as the study…"
L250 – Maybe swap order to winter then summer to match the order earlier in the sentence, i.e. rephrase sentence.

L258 and L262 – Maybe rephrase to state 'in winter…" and "in summer, …" clearly at the beginning of the statement, rather than hidden in the middle of the sentence.

L335-340 – Can you cite some evidence that PIOMAS is wrong here – I think it is incorrect and it is certainly inconsistent with the model.

L345 – You don't discuss Fig 6a at all. Is it needed? Perhaps it should be discussed later.

Figures

Figures 1, 3, 8, S1, S2, S3, S4
These all use the same colormap which is a blue-white-red (diverging colormap). Such colormaps are ideal for difference plots, e.g. Fig 1b,d, but are an odd choice for non-diverging fields, such as Fig 1a,c. I wonder if you are better changing colormap for the left-hand columns in all of these plots.

Figure 4
Unfortunately, this is really hard to read at this size (printed A4). I also think it has too much unnecessary detail in it. You have 10 wind speed classes. Do you really need this many classes? I think you'd get the same result with 2 m/s bins and it would be much clearer. Also do you include winds >10 m/s in the (9,10] bin? Note I have taken (1,2] to mean winds between 1 and 2 (inclusive) m/s – you should explain this in first caption.

Secondly you have 9 SIC bins – again this is a lot and there seems very little difference in the results between adjacent bins. I'd perhaps recommend fewer bins, perhaps (0,0.1], (0.1,0.3], (0.3,0.5], (0.5,0.7], (0.7,0.9], (0.9,1.0]. This keeps the 'end' bins separate as these are more interesting. At present this Fig 4 and also 9 has so much detail and numbers, that the main message is a bit hidden.

Finally, it may be worth noting how much data is in these bins. Although the bins are the same size (0.2 in ice fraction for example), the distributions mean there could be relatively few points in some bins.

Figure 5
Same comments as above really and note some of the colours are very faint (7,8) class. These plots are more readable but need to be consistent with Fig 4.

Missing Reference:
Elvidge, A.D., I.A. Renfrew, A.I. Weiss, I.M. Brooks, T.A. Lachlan-Cope, and J.C. King 2016: Observations of surface momentum exchange over the marginal-ice-zone and recommendations for its parameterization, *Atmospheric Chemistry and Physics,* **16,** 1545-1563. doi:10.5194/acp-16-1545-2016