Third review of ``Sea-ice deformation in a coupled ocean-sea ice model and in satellite remote sensing data'' by G. Spreen, R. Kwok, D. Menemenlis, and A. Nguyen.

Dear authors,

I'm happy to see that the manuscript continues to improve. In my previous review I noted five major issues I asked the authors to address. Two of those have now been adequately addressed so that even if I don't necessarily agree with your approach I cannot say that it gives wrong or misleading results. Three of the five previously raised issues do, however, still need further work.

First it is the issue of filtering. Your response to my previous comments about filtering the data and model results are based on the misunderstanding that the noise in question originates in the observations themselves. It is of course correct that all observations contain some noise and/or uncertainty. But this is not the point. The point is that for any given velocity field, even artificial and perfectly noiseless ones, using the method you use to calculate the deformation rates gives noisy results. It is the method itself that causes this and that has nothing to do with the data. Indeed Bouillon and Rampal (2015) demonstrate this in their paper using an artificial and smooth velocity field. Crucially, this means that both the data and the model will appear to show stronger scaling due to this noise. Even more importantly you therefore cannot say that your failure to remove this noise can contribute in any way to the differences in deformation rates or scaling between data and model. You do this in a number of places in the manuscript and it is simply wrong.

I would argue that you should filter. It doesn't matter what people did before Bouillon and Rampal published their work; filtering gives more accurate results and should be used. Also, if you actually did the multi-fractal analysis (and I still don't understand why you don't – it's not much extra work) you would see immediately why filtering is unavoidable. But I cannot argue that filtering will change your results in a fundamental way (at least as long as you don't do the multi-fractal analysis) and seeing you seem very reluctant to implement the filter I cannot force you. I do however insist that you not use the noise to explain some of the difference between model and observations - that is simply wrong. I also insist that you should note that the noise inherently produced by the method you use causes an overestimation of the scaling exponent b, both for the observations and the model.

Secondly I still don't know how the LKFs form in the model. Is it one strong event that produces it or are there repeated weaker events at the same location? Your monthly averages don't show this but this is a potentially important difference. Why don't you show snapshots (3-day means)? Snapshots, or possibly the combination of snapshots and monthly means would be more useful in determining how the LKFs form.

Finally, in section 3.2.3 you discuss the regional differences in D and then apply these differences to b without showing or discussing the regional differences in *b*. This is the jump from *D* to *b*. The implicit assumption seems to be that low concentration and thin ice gives high deformation rates and thus more negative *b*. But this is not really the case. We get more negative *b* because the mean deformation rate over a large area is smaller than that averaged over a small area. The deformation rate in the MIZ is indeed larger than in the pack, but that fact alone tells us nothing about *b*. It is therefore a prior not clear (to me at least) that large *D* will give large *b*, nor do you provide references to back such a claim up. Rather, to deduce the regional differences in *b* you should follow a partitioning scheme like Stern and Lindsay (2009, their section 7) use to calculate different values of *b* for different regions (their figures 9 and 10), and *b* as a function of multi-year fraction (their figure 11).

In fact all of section 3.2.3, as it stands seems a bit pointless to me. You do give us the dependence of deformation rate on thickness and concentration, but the link to scaling is not there. You also try to compare results from different resolutions using your equation (7) with a constant *b*, but we know *b* is not constant. You show that it changes seasonally and Stern and Lindsay (2009) show that it also changes depending on ice type, so of course you can't use constant *b* for all seasons and all areas, that much is obvious. I would recommend removing this section; I'm not sure how best to save it.

Now, on re-reading section 3.2.3 I realised that I may have overlooked a much more serious problem with it. Do I understand you correctly that you use the same method to calculate *D* for all the three model incarnations, i.e. for each model resolution you calculate the Lagrangian tracks starting from the RGPS positions etc.? I can't really tell from the text, but if this is the case then L_i is 10 km *for all three model runs*. You can use equation (7) to compare different model resolutions if and only if you also calculate *D* at the model resolution. If *D* is calculated at the same resolution for different resolution models the results should be the same! What you show in figure 10a is then a very nice result, but you completely misinterpret it. It shows that there is in fact a fundamental problem with the VP model in that the deformation doesn't scale when you change the model scale. So if my second thoughts on this section are right then it is based on a misunderstanding of the scaling concepts and is completely wrong.

Aside from those major issues there are some minor ones I noted on my previous review but I don't feel are satisfactorily addressed yet (line numbers refer to the revised paper)

l.13 p.1: You removed the word new from the last sentence, but that doesn't really change the meaning of the sentence much. ``New'' is still implied. This last sentence is anyway a poor way to finish the abstract since it does not really summarise what the paper does. Try something along the lines of ``… this study provides an evaluation of high and coarse resolution VP simulations using existing metrics.''

l.17 p2.: You replaced ``viscous rheology'' with ``any nonlinear rheology''. This doesn't really help and is clearly wrong since *any* nonlinear rheology includes an infinite number of rheologies not suitable for modelling sea at all. You also say that the ``first order mean velocity field as these can be correctly predicted even by simple sea ice models'', but this is also not correct - at least not for my understanding of what constitutes a ``simple sea ice model''. To remove such ambiguity but still motivate considering more advanced metrics you should say that we already know that current sea ice models (i.e. (E)VP class models) are capable of reproducing the first order mean velocity field reasonably well, so now it's time to look at something more difficult.

l.7 p.4 and/or table 1: Please include a reference to the relevant papers by Nguyen et al. for the parameters in table 1. I still think the parameters look strange, but all the more reason to make sure the reader has a proper reference for them. It is true that you cite Nguyen et al (2011) in the following paragraph, but this should be done earlier (and it should be made clearer that you are indeed using their setup, warts and all).

l.13 p.4: Please note that the choice of initial and boundary conditions for the ocean follows Nguyen et al. (2011). Again, this looks jarring to me so it's important to have the reference clear.

l.19 p.5: You use daily outputs to calculate drifter trajectories on a 4.5 km grid. This is insufficient because during one day the drifter will in many cases have drifted out of the grid cell it started in. So if your drifter algorithm looks something like:

1) Calculate (u,v) at (x,y)
2) Move drifter by dx = u*(1 day), dy = v*(1 day)
3) Advance one day and goto 1),

you will incur some error due to the low temporal resolution when dx and dy is large enough for u and v to be substantially different at (x+dx,y+dy) compared to (x,y). I don't pretend this is a major issue (this is the minor issues section after all), but it should be addressed. If re-running at 6 hourly output is not an option say why and say that you don't think this will be a big issue – which is fine.

l.9 p. 9: My previous comment here was perhaps not clear enough. I wonder why the deformation rate is essentially independent of resolution in winter and not in summer – but you surmised as much. There is still no real discussion of this (neither here nor in the discussion section), which would be nice to have. This is a perplexing behaviour because we've seen that the high resolution model appears to give better spatial patterns in the pack ice and so you would expect better mean deformation rate in winter in the high resolution model. But instead it's the summer rates that are better. Does this hold for shear, divergence, and vorticity as well?

l.9 p.13: It's not because RGPS is noisy. More importantly you cannot say that the lack of filtering causes differences between model and observations, as discussed above.

Finally I have some minor remarks following my reading of this latest revision.

l. 2 p.3: Why do you single out Tsamados et al. (2013)? That seems unwarranted by the context.

l.8 p.11: There is no space between the word "section" and the section number.

l.12 p.11: "… we exemplarly use …" is not proper English. Please rephrase.

l.3 p.12: You use averages at 1,000 km in your scaling calculation. This leaves you pushing against the finite size limit. I know Marsan et al and Stern and Lindsay use 1,000 km, but if you look at Bouillon and Rampal you see that they only go up to 700 km. The reason is that once you get too close to the size of the domain the calculated scaling is poorly affected. You can tell that this is happening because the last point on the log-log plot dips below the straight line. This is the case for figure 2 from Marsan et al, for figure 3 from Stern and Lindsay and for your figure 8. So this last point is suspect and should not be used. This will reduce your *b* value in all cases. You can also see that you have a problem if you compare the slope you get between all pairs of points. This should be more or less the same for all pairs, but it will be radically different for the last pair. It will also be different for the first pairs because of the noise I discussed above.

l.19 p.16: This paragraph starts by talking about seasonal vs. perennial ice, but then changes direction to talk about how the 4.5 km model gives more LKFs and then over to the localisation. This is very confusing and should be fixed. You need to split up the paragraph and rewrite.

l.3 p.17: Sentence should say "… scaling exponents … is …" not "are"

l.7 p.17: "… but, however, …" is not proper English. Please rephrase.

l.13 p.17: You never show how the scaling exponent depends on concentration and thickness.

l.18 p.17: You cannot use the results of Bouillon and Rampal (2015) to explain the difference between model and observations.

l.20 p.17: Missing space between the word divergence and an opening bracket

l.23 p.17: The word reasonable is very subjective, but the number of LKFs produced by the model doesn't seem reasonable to me, even at 4.5 km resolution.