

The Cryosphere Discuss., referee comment RC1
<https://doi.org/10.5194/tc-2022-46-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on tc-2022-46

Anonymous Referee #1

Referee comment on "Towards improving short-term sea ice predictability using deformation observations" by Anton Korosov et al., The Cryosphere Discuss., <https://doi.org/10.5194/tc-2022-46-RC1>, 2022

Review of "Towards improving short-term sea ice predictability using deformation observations" by Anton Korosov et al, tc-2022-46

This manuscript describes how including sea ice deformation derived from a satellite data product of sea ice drift may improve short sea ice predictive skills of an Arctic sea ice model. The main "trick" is to connect derived deformation to scalar model variables with some "memory", i.e ice concentration and damage, as assimilating the deformation or drift directly is known to be problematic as this information contains little memory and is usually lost within time steps of the model. The data assimilation (DA) itself is a simple re-initialisation scheme for ice concentration and damage. The authors speculate how their findings can be used in more sophisticated DA systems.

The topic is interesting and important as sea ice forecasts are thought to become more and more relevant for Arctic shipping and exploration. There are some issues with the manuscript that require careful rewriting and even repeating the experiments, so that I cannot recommend publication in the present form.

Main points:

1. According to Fig2 and the text, the data assimilation scheme uses observational data from the period between t_0 and t_1 to initialise the model at t_0 ; more generally the model is initialised at $t(n)$ with data from between $t(n)$ and $t(n+1)$ when in a realistic system, the data is not yet available. The difference between $t(n)$ and $t(n+1)$ is 24h. Then the same data set is used to evaluate the result of the assimilated model, i.e. the evaluation of the "forecast" on the first day is with the same data as the data set that is used to initialise the forecast. Since in the simple scheme, the initialisation is done neglecting the corresponding models value entirely (weight 1 in Section 3.4, 4.1), this comparison only shows how well the model persists the initial conditions. Not surprisingly, the model/data "agreement" is quite good on the first day and quickly deteriorates on day 2-5. A proper scheme would use data from $t(n-1)$ to $t(n)$ to initialise at $t(n)$ and compare to data at $t(n)$, $t(n+1)$, etc. As long as this is not change, the first day of the "forecast" cannot be used for any analysis and shouldn't even be called a forecast.

2. Terminology and language. The use of established terminology is rather liberal in the manuscript. As far as I know (but I may be wrong), the terms prediction, predictive skill, predictability, potential predictability have well defined meanings (I haven't heard of "prediction skill"), and the terms should be used when describing the experiments, otherwise it is hard to relate the work to other DA publications. Similarly, the DA scheme is described as "nudging", whereas it is a weighted re-initialisation scheme (according to eq11), but then the weights are always 1 or 0, so there is no weighting in this manuscript. "Nudging" implies a term in an evolution equation $d v / dt = \text{some terms} + \text{nudgingParamter} * (v_o - v)$, where v_o is the observation. Many smaller problems of similar nature can be found in the text. I marked some of them in the list below.

3. In data assimilation, one can expect that including additional information will improve the result. Therefore, comparing an assimilated model to a free run makes little sense. Essentially the free run in Fig3 is a 22-day forecast that hasn't seen any new data in 22 days. As noted before, comparing the model to observations that have been used in the assimilation cannot say much about the "success" of assimilation. Anything but any small improvement would just be a failure. Similar "mistakes" have been made before, e.g. doi:10.3189/2015AoG69A740

In general, one would have a ctrl-simulation with an established DA scheme and then add new data or new methods and compare the improvements over the ctrl-simulation. This is done in section 4.2, but it would be more interesting to see, if the addition of deformation

data to an existing DA system (which may already assimilate ice concentration or even thickness) would improve the predictive skill. The authors present their work as a "proof of concept", but the evidence they provide does not help in evaluating, if these additional data help in a realistic system, because the framework is so different.

4. When introducing new data and constraining new model variables it is good practice (also in sea ice data assimilation) to test schemes and types of data in twin experiments, where a free run produces "observations", a subset of which is used for assimilation leaving the remaining data for validation. This has been done a lot especially in anticipation of new data (e.g. doi:10.1029/2006JC003786). Here, one could have at least held back some of the observations to be used for model validation. This makes it impossible to check if the DA actually improves the state also away from the observations. Instead, one can only make statements about the plausibility of the solutions outside the areas covered by observations (by no mean can the authors claim, that the LKFs are "corrected" outside of the data coverage).

5. A key point of the procedure is how the deformation derived from satellite ice drift data is connected to the model variables concentration and damage. The derivation of this empirical connection is moved to the "supplemental material/Appendix I", which is not part of the manuscript, nor can it be found online.

minor issues, typos, suggestions, some related to the above points.

page 1

l2: due to the lack -> in the next sentence there are observations to be assimilated.

Please rewrite.

l8: deterministic forecasting with one member -> isn't that a homoioteleuton? A one-member system is always deterministic, or an ensemble with just one member, is not really an ensemble. Remove "with one member", or replace "with a single simulation"

l10: in 3-5 days horizon -> grammar?

l13: reduction in: article missing, or replace by "reducing the", although this sentence is not very clear in general and could be improved

l13: bigger role -> than what?

l20: only -> mainly (sea surface tilt, momentum advection, and you cannot exclude small effects of floe-floe interaction)

or "dominated"

l23: brittle -> I don't think that you can say that. It's driven by complex non-Newtonian mechanics/dynamics, but "brittle" is just one aspect of it, and frankly, only a model for the behavior. Other models of sea ice motion exist (I don't mean numerical models). Please rewrite.

page 2

l25: deforming -> just to illustrate my previous point, this deformation is NOT brittle, but plastic (no restoring force pushes the ice back into the initial state as for elastic behaviour). The brittle part is just the way failure is parameterised in nextsim (which I believe is a good model for this). I think that this general description of sea ice mechanics/dynamics needs to be "decoupled" from the specific model nextsim, that is being used in this manuscript.

l28: "Under divergent ice motion these cracks become open leads, significantly increasing ocean-air heat and mass exchange and modifying local atmospheric boundary layer and ocean mixed layer. Open leads are also key both for marine fauna survival," -> I agree with this, certainly on the local scale of the leads, but this is just a plausibility argument. I have not yet seen that this has been confirmed on large scale heat and mass exchange and budgets. Please give references, if you have them, otherwise marks this as a "plausible assumption".

l37: observe -> isn't RGPS a data product derived from Radarsat on a 12.5km grid? I find "observe" in this context inappropriate. Please rewrite.

l41: (10 - 30 km) -> can be only a result of the "coarse" resolution, or do we really have "cracks" in the Arctic that are 10-30km wide. Those would be large stretches of either open water or vigorous deformation. I assume that the interpretation is important for the DA.

l44: "only one model, neXtSIM" -> Supposedly this is put here to justify the decision to use neXtSIM. The statement is not incorrect, but it is not clear to me, what the authors would like to achieve with this statement. It does not help this paper in any way, because it hides the results of the Bouchat's paper, that other models have similar properties (at finer grid spacing and higher computational cost). Also doesn't the nextsim in Bouchat's paper use the MEB rheology instead of the BBM-rheology? I would rewrite as something like this (I tried to emphasise that this is a useful model for this study, i.e. does the job very well and is comparatively cheap):

In a recent model intercomparison paper (Bouchat et al., 2021), neXtSIM simulations (neXt Generation Sea Ice Model, Bouillon and Rampal, 2015a; Rampal et al., 2016), ranked among the best for simulating the observed probability distribution, spatial distribution and fractal properties of sea ice deformation, even though it operates on a low resolution grid of 10km. All other comparable simulations used higher resolution and were hence more expensive.

l49 skill -> the skill?

l51: observations -> the technical term for this is "potential predictability", which always excludes observations. Why not use that?

l55: "so the assimilation scheme needs to perform a cross- variable update from deformation to sea ice model variables." This is a common "problem" in DA and one would use a proper "observation operator" that maps the model variables to the observations. The dual operation then maps the model-data misfit back to increments of model variables. If you want to talk about "data assimilation", I suggest to use the proper language/terminology. Here you will (according to the abstract) do a nudging experiment (but it turns out to be re-initialisation in reality), which is strictly speaking not really data

assimilation (although total valid as a method).

page 3

l79: "d is the ice damage." maybe it makes sense to clearly state the mean of "d", d=0 is entirely intact and 1 is entirely damaged (which I assume here from the equation) or vice versa. Previous publications use contradicting definitions.

page 4

eq5, what is "P"?

l101: "12 hours frequency" -> 12h is not a frequency, but a period. The frequency is: one record in 12h.

page 5

l110: "reach an equilibrium" in 30 days is hard to believe, usually one would expect a seasonal cycle at least, unless the sea ice models of TOPAZ4 and nextsim are identical (which they are not, I assume). But does the equilibrium matter?

Figure 1 is not really necessary.

l113: this looks like the scheme uses data from the future to correct the model? Does that make any sense? I would expect to update the model variable at $t(n)$ with data collected over $t(n-1)$ (or earlier) to $t(n)$. See also main points above.

l119: in the previous work of (e.g. (Bouillon and Rampal, 2015b)). -> fix parentheses

Fig2, caption: Eps,d and A -> use proper symbols as in figure.

page 6

l125 The "Appendix" should be in the same file as this text, right? Supplementary material is separate. What do we have here?

On the TC-web page I cannot find any supplementary material, so that "Appendix I" is missing for now.

eq9: that would be $1 - 10^{\{k_2\}} * \epsilon_{tot}^{\{k_3\}} - k_1$, right? Now it would be interesting to know at least k_2 and k_3 , because that would show how strongly the total deformation impacts damage, compared to eq10, where the impact is linear but later in the full equations exponential, as argued in the discussion section 5.2

eq10: wouldn't it make sense to treat divergence and shear separately, ie. have two different coefficients: $f_A = 1 - a_1 \epsilon_{div} - a_2 \epsilon_{shear}$, or even differential between divergence and convergence. It is clear the divergence will create open water directly, but convergence will do this to a much smaller degree (e.g. lateral divergence in convergence), and also shear should have a different coefficient.

page 7

l148: simple least-squares nudging approach? Where are the least squares? Deriving eq11 from a least-square formation is possible but a little vain.

I am not sure if I would call eq11 "nudging", as nudging usually implies a time varying equation such as $dv/dt = rhs + nudgingparameter*(v_o-v)$, which is not what eq11 implies.

If in a DA cycle $v_m(n)$ is computed at time $t(n)$, then updated according to eq11, and then $v_a(n)$ is used to initialise the next DA cycle, then this is not nudging, but re-initialisation with a very simplified updating scheme. I am not criticising that, but I think that the description needs to be accurate. See also main comments.

l159: "the very small spatial correlation approximation is reasonable." I disagree. This assumption is valid normal to the fracture, but along the fracture, considering the nearly instantaneous fracture propagation (in nextsim and in observations), this not a "reasonable" assumption. Please rephrase.

l164: value of eps_{min} ? Mention here, that this is part of the sensitivity analysis?

l171: "Since it was difficult to distinguish between the individual impacts of w_v and W in Eq. 12," , unclear, why.

l172: "0 and 1 were tested for w_d and w_A ", but this means that there is no weighted average at all and all that is done is re-initialisation. I think it would help the reader to clarify the scheme: Either, there is pure re-initialisation with a somehow derived value, or no re-initialisation. The entire description of least-square nudging is entirely misleading (and does not describe, what is actually done). See main comments

page 8

l184: "difference in 90th percentile", what is that? Please be more specific.

l191, related to l184. From the explanation is it no clear what is computed here, and in which sense this is different from "MCC". Is MCC a standard statistical method, or something that is only described in Korosov and Rampal, 2017? If it is a standard method, please cite a standard reference/textbook.

Further, there were a few metrics suggested in the cited papers by Bouchat et al 2022, and companion paper Hutter et al 2022, also Mohammadi-Aragh et al 2018. In what sense are the metrics used here related, or do they quantify entirely different properties?

l195: "22nd January 2021" depending on the definition of "free run", I would expect that after 22 days of integration the "free run" has already quite deviated from the observations. What is the point of this comparison? Showing that the model can be "kept on track" even with simple methods, compared to not doing anything?

Normally in DA, one defines a baseline/ctrl with some existing system (not the free run!) and compares how the details of the algorithm affect the solution, as has been done in section 4.2. It would also be interesting how important the observations on the current day are for forecasts, i.e. comparing a run with DA until $t(n-1)$ to a run with DA until $t(n)$ (where, in fact, the observations are not from one day into the future as is the case here). See also main points

- Highlight, page 9

l209: "due to its rheology" -> since only nextsim is used with one rheology, this (part of the) statement is not supported by the experiment and should be removed.

Also, since there is no observational data to check the results in the "unobserved" regions, one cannot claim that "neXtSIM is able to extrapolate and create realistic connections". The model simulation creates connections that look realistic, in the sense that they are not garbage, but that's about it. For a statement like the one in l1209/210, one needs experiments where part of the data is withheld from the DA, to be used later for model validation.

Further, in DA we expect that the results improve with additional data. Any other result would be failure of the DA, so all that figure 3C shows, is that the DA algorithm does,

what is has been designed to do.

This comparison is even further biased, because now (according to the description in Section 3, Fig2) the model has been corrected with data from the future ($t(n)+24h$), and then is compared to the same data from the future. I wouldn't call that prediction, but analysis.

page 11

Figure 4, please add colorbars to make it easier to view the images

page 12

l218: persistent or persistency

l226: fix D_{P90}

l227: assimilation, better : re-initialisation.

l233: sufficient -> a sufficient

l235: consequent -> subsequent

l238 nudging -> re-initialisation

page 13

l242: "The experiments with w_d cannot detect" anything, rewrite as "In the experiment with w_d , one cannot detect ..."

Figure 6 tells me, that the leading order effect is achieved by " a_1 ", (except of $a_1=-2$, where a similar effect is achieved by modifying ϵ_{\min}), so the linear relationship between total deformation and ice concentration. All other parameter appear to have small effects only. Maybe this should be stated somewhere explicitly.

Fig6 is difficult to read, maybe make the bars broader?

In Tab1 a_1 parameters are all positive, here, they have a negative axis, please correct, also there's seem to be experiments with $a_1 < 0$ (i.e. to the right of 0), which are not

listed in Tab1

I246: "first successful attempt", What is meant by "successful" here? This sounds like a conclusion that needs to be backed with evidence. Also since the observations assimilated appear to come "from the future", the results for the first day (which is most "successful") cannot be used.

I250: "it" -> what is "it"? The relationship between deformation fields and model state variables is not shown by the DA, but by a prior correlation analysis, which I cannot evaluate, because it is moved to an appendix/supplemental material that is not accessible at the moment.

Also the damage assimilation had little effect, so that questions both the empirical relation in eq9 and/or the "success" of the DA.

I252: "proves" -> this is clearly too strong.

I253 "corrects" -> to correct means to make it right, but there's no proof for that in the manuscript. All that the experiments show is that the model takes the initialisation information and propagates it sensibly (according to the model dynamics) into areas that have not been re-initialised. This does not mean that we now have "correct" forecasts, just that there's some "dynamical extrapolation" that needs to be evaluated with independent data (and this important step it is missing). See also main points.

page 14

I266: this paragraph sounds like a project proposal with some selling arguments. Not sure if a scientific publication is the right place to advertise one's work in such a way. In my view, a scientific publication in TC should report scientific advances, but not the suitability of a system for tasks that have not yet been performed. Please rewrite or remove.

page 16

I302: skill for 2–5 days -> see earlier comments, I think that the first day cannot be counted because of the data from the future.

I307 to the end of the section: I think that this list of factors impacting the predictive skill of LKFs would be much better placed (slightly modified) in the introduction, to lay out the scope of the manuscript and which of these aspects will be addressed in the manuscript.

I325: Bouillon et al., 2009 -> wrong reference. The correct Bouillon paper is from 2013, where this is called "revised EVP", although I believe that the proper reference would be Lemieux et al 2012, who were the first to modify EVP which then was described as modified EVP in Kimmritz et al (2015).

It is not clear to me, how using a VP rheology (mEVP is a method to solve the VP rheology equations), that has been marked as too slow, etc. in this paper and many other papers of this group, is going to help here at all.

page 17

I349: "neXtSIM is capable of extrapolating the spatially discontinuous satellite observations of deformation by connecting the elements of linear kinematic features in a realistic manner." -> this is a statement, that I think is totally justified from the evidence provided (Fig3). Please rewrite previous statements about "correcting" LKFs etc accordingly.

I351: local -> locally?

page 18

I359: Data availability: TOPAZ data and other forcing data are not mentioned, no code availability.