Replies to Reviewer 1

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.

We appreciate the reviewer's constructive comments. As requested by the reviewer, more experiments for evaluating predictability of sea ice deformation were run. We have also improved the clarity of the mathematical formulation of the problem – the methodology section has been completely re-written.

Main points

1. According to Fig2 and the text, the data assimilation scheme uses observational data from the period between t0 and t1 to initialise the model at t0; 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. The model runs between t(n) and t(n+1).

All the previous and many additional experiments were re-run with assimilation of historical data only. The deformation is now computed from observations of drift at t(n-1) and t(n), then the analysis is computed and the model is initialised at time t(n), then the forecast is compared with deformation computed from t(n) to t(n+1), from t(n+1) to t(n+2), etc. The new experiments confirm that assimilation of data from the day before improves the accuracy of the deformation forecast both in case of synthetic and real data. New results are shown on Fig.6 in the revised manuscript in the Results section.

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 = some terms + 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.

The terminology has been changed according to the reviewer comments. The term "nudging" is replaced with "direct insertion" [Stanev and Schulz-Stellenfleth, 2014]. In the new experiments we tested different values of the weights [0, 0.5, 1].

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.

In the new experiments we evaluate the forecasts on independent data which was not used for assimilation. As explained above, the difference between the forecasts without assimilation and the observations provides the background error. The error of the forecast with assimilation is compared to the background error for evaluating practical predictability. Our previous experiments with assimilation of sea ice concentration [Williams et al., 2021] showed that ice drift is not significantly affected by assimilation. In that sense the ctrl-simulation with assimilation of, e.g., ice concentration is almost equal to the forecasts without assimilation.

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).

The following twin experiments were run as explained above:

- Truth run,
- Forecast initialised from truth,
- Forecast without assimilation,
- Forecast with assimilation of synthetic data,
- Forecast with assimilation of real data.

As explained on Fig. 1 in the manuscript these experiments allowed to evaluate the practical predictability on synthetic observations and real observations.

Satellite observations already have large gaps and artificially reducing the coverage even further will decrease the ability of neXtSIM to connect the pieces of LKFs. Luckily we have a few cases when the coverage of test data exceeds the coverage of assimilated data, and we can test our hypothesis that LKFs are improved also outside the area of assimilation as shown on Fig. 4 in the manuscript. In the new set of experiments we use independent data for testing and can confirm that MCC increases also outside the area of assimilation. Nevertheless, due to lack of sufficient cases of confirmed good extrapolation, the statement on correction of LKFs in the entire basin is rephrased as follows:

"Our experiments illustrate that even if data insertion is spatially limited by satellite observations (or even very localized in high deformation zones) it can realistically extrapolate the deformation pattern by connecting the elements of linear kinematic features."

Experiments with spatially limited assimilation of synthetic data also confirm that accuracy of LKFs is increased not only in the area of assimilation, as shown on Fig. 5 in the manuscript.

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.

Appendix I was not included in the first submission due to a technical problem. It is included in the manuscript.

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.

Rewritten as follows: “Short-term sea ice predictability is challenging despite recent advancements in sea ice modelling and new observations of sea ice deformation, that capture small-scale features (open leads and ridges) at kilometre scale.”

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”

The sentence is rewritten as follows: “This proof-of-concept assimilation scheme uses a data insertion approach and forecasting with one member. We obtain statistics of assimilation impact over a long test period with many realisations starting from different initial times.”

l10: in 3–5 days horizon − > grammar?

The sentence is rewritten as follows: Assimilation and forecasting experiments are run on synthetic and real observations in January 2021 and show increased accuracy of deformation prediction for the first 2 - 3 days.

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

The sentence was removed.

l13: bigger role − > than what?

The sentence was removed.

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

The sentence is rewritten as follows: ”...the speed and direction of the drift are dominated by the atmospheric and ocean drag forces and by the Coriolis force.”

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.

Relation to brittle rheology is removed.

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.

The simplified explanation of brittle rheology is rewritten as follows:

”First, under increasing external forcing the undamaged ice deforms primarily as an elastic material. Internal stresses gradually accumulate in the material until a failure criterion is reached, which corresponds to a limit when sea ice fractures, and then the ice starts deforming along the multiple narrow and elongated cracks, and does so until these later refreeze or when the load (winds and currents) on the ice changes.”

We agree that brittle sea ice deformation is more general than a specific BBM rheology or a specific implementation in neXtSIM. In the paragraph near line 25 neXtSIM is not mentioned.

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”.

A relevant reference (Olason et al., 2021) is added.

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.

There are several RGPS products: a gridded deformation product on 12.5 km resolution and a Lagrangian product with a variable mesh size with a nominal resolution of approx. 10 km. The sentence is rewritten as follows:

"The Radarsat Geophysical Processor System (RGPS) dataset was the first attempt to systematically observe sea ice drift and derive sea ice deformation on high spatial resolution (10 km) and with high frequency (3 days) over a long period of time (winters 1996–2016) (Kwok, 1998)."

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.

Rewritten as follows:

The cracks appear on satellite-derived ice deformation products as narrow (10 - 30 km, depending on resolution of satellite data) and long (up to 1000 km) lineaments and are also called linear kinematic features (LKFs) (Kwok, 2001).

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 45 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.

As pointed out by the second reviewer, some of the models were running at a comparable resolution with quite good results. Therefore, the text is rewritten as follows:

"In a recent model intercomparison paper (Bouchat et al., 2021) neXtSIM results 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 10 km. Analysis of spatial and temporal scaling (Fig. 13 in (Bouchat et al., 2021)) shows that the spatial structure function of neXtSIM matches the RGPS observations very well, whereas the temporal one is overestimated by 3 – 5 %, probably indicating some overestimation of the intermittency by neXtSIM."

l49: skill $->$ the skill?

Corrected

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

Rewritten as follows:

"Mohammadi-Aragh et al. (2018) evaluated the potential predictability of LKFs using an ensemble of sea ice models all using a viscous-plastic rheology, but the practical predictability remains unknown."

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).

The term "nudging" is replaced with "direct insertion", which, according to e.g. [Stanev and Schulz-Stellenfleth, 2014], is one of the data assimilation methods. The sentence is rewritten as follows:

"First, the "direct insertion" method operates in the model state space. However, the observed deformation is not a model prognostic variable, so an operator is used to convert deformation to the model variables. This operator is an inverse of the observation operators used in data assimilation, since it maps from the observation space back to the state space."

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.

Explanation is added: "(with $d = 0$ being completely undamaged ice)."

page 4: eq5, what is "P"?

Added after Eq 5.:

"where P is a constant scaling parameter for the ridging threshold to parameterise P_{max}, following the results of Hopkins (1998), and h is thickness."

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

Rewritten as follows:

"Ice drift is computed from pairs of images separated by approximately 24 hours and the product is delivered every 12 hours."

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?

The entire description of experiments is rewritten.

Figure 1 is not really necessary.

The entire description of experiments is rewritten and a scheme with explanations of several experiments becomes 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.

The experiments were re-run to validate the results on independent data.

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

Corrected.

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

Figure 2 is removed.

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.

Unfortunately, the Appendix was not added due to technical reasons. It is added in the revised manuscript.

eq9: that would be $1 - 10^{k_2} * \varepsilon_{tot}^{k_3} - k_1$, 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

These values are provided in the Appendix.

eq10: wouldn't i make sense to treat divergence and shear separately, ie. have two different coefficients: $f_A = 1 - a_1\varepsilon_{div} - a_2\varepsilon_{shear}$, 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.

We assume that all deformation events (convergence, divergence and shear) indicate presence of weaker ice that may continue to be deformed. Ice weakness is simulated in neXtSIM by decreased concentration or increased damage.

Observation of any of deformation components (including convergence) is interpreted in the assimilation procedure as an increase in ice weakness and, therefore, decrease in concentration or increase in damage. We cannot find reasoning why weakening of ice due to convergence is different from weakening due to divergence or shear and, therefore, suggest that total deformation is a good proxy for detection of weak ice and a single dependence of A (and d) on total deformation can be used. Corresponding explanations are added to the text.

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 $dv/dt = rhs + nudginpgarameter * (v_o - v)$, 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.

Rewritten as :

"We update the damage and concentration variables in the model according to the observed deformation using a simple "direct data insertion" approach as a proof of concept for DA (Stanev and Schulz-Stellenfleth, 2014)."

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.

Rephrased as:

"Since sea ice deformation is accommodated along nearly 1D geometrical features (i.e. fractures), correlation can only usually be seen along the fracture, and so assumption of low spatial correlations in all other directions is reasonable."

l164: value of eps_min? Mention here, that this is part of the sensitivity analysis?

Rewritten as:

"where ε_{min} is a threshold for total deformation found in sensitivity experiments."

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

The number of experiments was increased, and values of 0, 0.5 and 1 were tested. The sentence was removed.

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

The number of experiments was increased, and values of 0, 0.5 and 1 were tested. The sentence was removed.

page 8, l184: "difference in 90th percentile", what is that? Please be more specific.

The 90th percentile is not used anymore for estimating predictability as it shows very similar results to A_{MCC}. In the sensitivity tests it is replaced with the Kolmogorov-Smirnov test applied to PDFs of forecasted and observed deformation.

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.

Explanation of the MCC computation is added to the Appendix. A reference to a relevant textbook (Brunelli et al., 2009) is added.

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?

MCC is slightly different to the LKF metrics used in Hutter et al 2022 and Mohammadi-Aragh et al 2018. The following explanations are added:

"Unlike the LKF evaluation metrics suggested in (Hutter et al, 2020) that compare only statistical properties of LKFs (number, density, length, orientation, etc), the MCC-based metric estimates co-alignment of individual LKFs on model simulations and satellite observations. It is also thought to be more sensitive to LKFs with low deformation magnitude, as no threshold is applied for their detection."

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

As explained above the "free run" has never seen observations of deformation and can be considered as a control simulation. The free run is not expected to match with observations (even if spatial patterns of deformation are remarkably similar). The goal of the comparison is to show that the suggested assimilation puts the model "on track", i.e. the forecasts with DA start to match with observations much better than the free run (without DA or with assimilation of something other than deformation). These aspects are also covered in replies to the major 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.

Relation to rheology is 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 ll209/210, one needs experiments where part of the data is withheld from the DA, to be used later for model validation.

As explained above, in the new set of experiments we use independent data for testing and can confirm that MCC increases also outside the area of assimilation. Nevertheless, due to the lack of sufficient cases of confirmed good extrapolation, the statement on correction of LKFs in the entire basin is rephrased as follows:

"Our experiments illustrate that even if data insertion is spatially limited by satellite observations (or even very localized in high deformation zones) it can realistically extrapolate the deformation pattern by connecting the elements of linear kinematic features."

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.

As explained above, in the revised manuscript observations from t(n-1)-t(n) are used for analysis on t(n) and the forecast on t(n)-t(n+1) is evaluated on independent data from the same period. The goal of Fig. 3 is exactly to illustrate that the DA algorithm does what it is supposed to do.

page 11, Figure 4, please add colorbars to make it easier to view the images

The colorbars are added.

page 12, l218: persistent or persistency

The presidency forecast is not used anymore in evaluations according to reviewers requests.

l226: fix D_P90

Corrected.

l227: assimilation, better : re-initialisation.

Rewritten as "assimilation using direct insertion"

l233: sufficient − > a sufficient

Corrected.

l235: consequent − > subsequent

Corrected.

l238: nudging − > re-initialisation

The entire section is rewritten and 'nudging' is not used anymore.

page 13, l242: "The experiments with w_d cannot detect" anything, rewrite as "In the experiment with w_d, one cannot detect ..."

Rewritten as suggested.

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 eps_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.

This observation is added to Section 4.3

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

Figure 6 is re-plotted with results from more experiments.

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

The sign of a1 on Figure 6 was incorrect and is corrected in the revised manuscript.

l246: "first successful attempt", What is meant by "successful" here? This sounds like a conclusion that needs 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.

The observations from future are not used anymore. Fig.5 evidences the success. The sentences is rewritten as follows:

"We present the first successful attempt to use observed sea ice deformation to increase accuracy of deformation prediction for the first 2 − 3 days."

l250: "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.

Appendix is added and the sentence is rewritten as follows:

"Our study demonstrates in practice that information contained in the observed deformation fields can be used for initialisation of model state variables, and shows the time scales over which the forecast of deformation can be improved."

l252: "proves" − > this is clearly too strong.

Rephrased as follows:

"Our experiments illustrate that even if data insertion is spatially limited by satellite observations (or even very localized in high deformation zones) it can realistically extrapolate the deformation pattern by connecting the elements of linear kinematic features."

l253 "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.

Please see previous reply.

page 14, l266: 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.

The paragraph on practical usefulness is removed.

page 16, l302: skill for 2–5 days $->$ see earlier comments, I think that the first day cannot be counted because of the data from the future.

As seen from Fig. 6, without using data from future the accuracy of forecasts is improved for 2 - 3 days.

l307 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 is addressed in the manuscript.

The list of factors impacting predictability is added to the introduction, but a more detailed description remains in the Discussion section as it leaves many open questions for future research.

l325: 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.

The reference is corrected. The following clarifications are added:

"We expect that the model equipped with the mEVP rheology will not be capable of spatial extrapolation of the assimilated ice weakness (lowered A or enhanced d), and that further tuning of the BBM rheology can improve the practical predictability of LKFs."

page 17, l349: "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.

The statement is rewritten (please see above).

l351: local $->$ locally?

Corrected.

page 18 l359: Data availability: TOPAZ data and other forcing data are not mentioned, no code availability.

The Data availability section is updated correspondingly.

Replies to Reviewer 2

In this paper, damage and concentration – assumed to be a function of satellite-derived total sea ice deformation – are assimilated using a simple nudging method in the Brittle-Bingham- Maxwell (BBM) model – an off-spring of the EB/MEB model – for short-term sea ice forecast of Linear Kinematic Features. Results show that assimilation of sea ice concentration improves the skill of a short-term forecast for up to five days lead-time in comparison with persistence or a free simulation. The assimilation of damage on the other hand does not improve the skill of the forecast. It is argued that that the reason for the lack of sensitivity to damage assimilation is because the healing of ice (that reduces the damage) has a timescale that is too fast (24 hrs) compared with the desired lead time of the forecast (O(days)).
The paper is well written. The organisation and depth of the paper however is lacking. The text refers to a non-existent appendix on several occasions; the sensitivity to assimilation parameters consists of four 1 or 2-line paragraphs; there are multiple unsubstantiated statements; the results section presents inconsistent results (sea ice concentration larger than one) and focuses on small scale features in order to justify broad conclusions. The authors state that neXtSIM is not ready to assimilate deformation

yet and that the paper is a proof-of-concept. There is no problem with this approach, but the authors have pushed this paper out for review much too early. I recommend that the paper be rejected for the moment and that authors be encouraged to resubmit later, a substantially different version of the paper that addresses (not rebut) the comments below. I am recommending a "reject with encouragement to resubmit" simply to give the author time to properly address the comments.

We appreciate the reviewer's constructive comments. As requested by the reviewers, more experiments for evaluating predictability of sea ice deformation were run (please see the replies to the first reviewer for details) in the extra time provided by the editors. The manuscript is undergoing a significant revision that can be accomplished without a rejection. All the requirements and suggestions are duly addressed without a rebuttal.

Overstatements:

Line 304: "The viscous-plastic (VP) rheology used in MITgcm is known to have a less realistic slower time evolution of LKFs (Hutter et al., 2018) than the BBM rheology in neXtSIM (Olason et al., 2022). As a result, the sea ice simulated by the BBM rheology has more rapid error growth (loses skill faster) due to the correctly resolved intermittent ice motion and localised ice deformation." If the intermittent motion is correctly resolved, why is the error growing faster?

Generally speaking, the higher is the intermittency of a stochastic process, the lower is the predictability. Therefore, a model with a low intermittency will have a tendency to overestimate intrinsic predictability.

However, our case neXtSIM is probably overestimating the intermittency. Indeed, looking in details on Fig. 13 in (Bouchat et al.,2021) we can see that the spatial structure function of neXtSIM matches the RGPS observations very well, whereas the temporal one is overestimated by 3 - 5 %. The MITgcm at 4.5 km (same or similar to as used in (Mohammadi-Aragh et al., 2018)) is worse in the spatial domain and seems better in temporal domain.

Line 44: "In a recent model intercomparison paper (Bouchat et al., 2021), only one model, neXtSIM (neXt Generation Sea Ice Model, Bouillon and Rampal, 2015a; Rampal 45 et al., 2016), proved to be capable when run at the same spatial resolution as the available observations (i.e. $\sim$ 10km) to simulate the observed probability distribution, spatial distribution and fractal properties of sea ice deformation." This is incorrect. One 10-km McGill run was able to reproduce the PDF of divergence and multiple 10-km models were able to reproduce temporal scaling and spatio-temporal coupling. The only place where neXtSIM did perform better compared to other 10-km models is the spatial scaling. Intermittency does not come from the brittle parameterization in the EB/MEB/BBM. All models participating in SIREX (VP, EVP, EB) show intermittency (Bouchat et al., 2021). The source of intermittency in observed deformation is still an unresolved issue. The authors must clarify what they mean by "intermittency. If the temporal scaling exponent is used to discuss the LKFs intermittency, this is incorrect. If instead the authors are referring to LKF growth rate and lifetime, this is also incorrect. In SIREX2, it is shown that no apparent link is present between the LKF growth rates, lifetimes and the temporal scaling/multi-fractal parameters. The intermittency is revealed by the quadratic nature of the structure function.

Following (Rampal et al., 2008) we consider the exponent α as a measure of the degree of intermittency.

Indeed, SIREX indicate that all models have some degree of intermittency (a slope of the temporal scaling line in log-log space) but only two low resolution models (RASM-WRF-EVP and RASM-WRF-EAP) are as close to the RGPS observations of total deformation as neXtSIM (Fig. 11.a in Bouchat et al., 2021). It is interesting to note that RASM-POP-EAP, which has exactly the same setup as RASM-WRF-EAP except for using atmospheric forcing with much lower temporal resolution, has much lower deformation rates, probably indicating the role of the atmospheric forcing in intermittency.

Nevertheless, the slope of neXtSIM temporal scaling is higher than the RGPS observations by 3 - 5 % (Fig 11.c and Fig 12.c), probably indicating overestimation of intermittency by neXtSIM. The text is rewritten as follows:

"n a recent model intercomparison paper (Bouchat et al., 2021) neXtSIM results 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 10 km. Analysis of spatial and temporal scaling (Fig. 13 in (Bouchat et al., 2021)) shows that the spatial structure function of neXtSIM matches the RGPS observations very well, whereas the temporal one is overestimated by 3 – 5 %, probably indicating some overestimation of the intermittency by neXtSIM."

Line 209: "Thus, due to its rheology, neXtSIM is able to extrapolate and create realistic connections

between the observed and assimilated pieces of LKFs." This is not demonstrated in the manuscript. I believe any rheology that assimilates sea ice contration (A) will show LKFs in line with observations. See Major Points below.

We agree and the sentence is removed.

Major Points:

On two occasions, the authors are referring to an Appendix that is not included in the paper. The appendix must be included.

Unfortunately, the Appendix was not added due to technical reasons. It is added in the revised manuscript.

Equation 7-10: Sea ice concentration (A) increases in convergence (until A=1) and decreases in divergence; sea ice concentration can also increase or decrease in shear. A single dependency of A on eps_tot is therefore missing events where convergence is present along LKFs. The damage (d) dependency on eps_tot is more realistic, but the authors argue that assimilation of A is useful to increase predictive skill, whereas d is not. See below for more on this topic. The single dependence of A on (eps_tot) must be justified.

We assume that all deformation events (convergence, divergence and shear) indicate presence of weaker ice that may continue to be deformed. Ice weakness is simulated in neXtSIM by decreased concentration or increased damage. Observation of any deformation components (including convergence) is interpreted in the assimilation procedure as an increase in ice weakness and, therefore, decrease in concentration or increase in damage. We cannot find reasoning for why weakening of ice due to convergence should be different from weakening due to divergence or shear and therefore suggest that the total deformation is a good proxy for detection of weak ice and a single dependence of A (and d) on the total deformation can be used.

It should be added, that we have two ice categories in the model: older ice, whose concentration is used in rheology and younger ice, which is formed during water freezing and is converted to older ice only after exceeding a threshold in thickness. Only the older ice is updated in the assimilation procedure and the total ice concentration remains the same.

Corresponding explanations are added in Section 3.3.

Line 145: The value of a1 in F_A(eps_tot) is found through sensitivity experiments using the same BBM model. This is a circular argument. This functional dependence must be derived from sea ice concentration and total deformation derived from passive microwave and Sentinel. I believe the author will find that the functional dependency is not a simple linear relationship. The author must at least show this relationship from observations and acknowledge the simplicity/caveat of the approach.

This is not a circular argument. Indeed, a_1 is found in optimisation of:

$$\min(\varepsilon_{i+1}^{obs} - \varepsilon_{i+1}^{sim}) \Rightarrow a_1 \tag{1}$$

where $\varepsilon_{i+1}^{sim} = f(a_1, A_i)$. Therefore a_1 is a function of observed total deformation and previous concentration and is not a function of a_1.

We must use simulations of the same model for tuning a_1 for two reasons: First, reliable simultaneous observations of ice concentration and deformation at scales of 1d / 10km are not available. Second, weakening of ice (by decreased A or increased d) must be consistent with the rheological model parametrisation. For example, if observations (in case it existed) would show a higher rate of concentration decrease per deformation rate than we obtain from simulations, then the assimilation would decrease the concentration too much, and the model would predict higher deformation than was actually observed.

Corresponding explanations are added in Section 3.4.

Fig 5a: The skill of the model is assessed using the fraction of points where the correlation between observed and simulated deformation is significant. Statistically speaking, there will always be some points that will remain significantly correlated. The statistical significance of the signal must be shown in the figure. I also see no spatial structure in the regions of high correlations which suggest that the high-correlation points are just random events.

It is indeed difficult to illustrate the evaluation method on an example with gaps in observations. Below (Fig. 1) is an example of maximum cross correlation (MCC) maps computed for a free run and a forecast with assimilation of synthetic observations. Only statistically significant (p=0.01) values of MCC are shown. There is an obvious spatial pattern in the MCC values linked with co-alignment of LKFs on the deformation maps. This example also shows that MCC increases in some regions (e.g. in Beaufort Sea) after assimilation where co-alignment is visually better. And in some regions it remains the same (high or low), which indicates that MCC can be used for evaluation of assimilation impact. Examples with masking of insignificant MCC and a more contrast colormap are added in the manuscript.

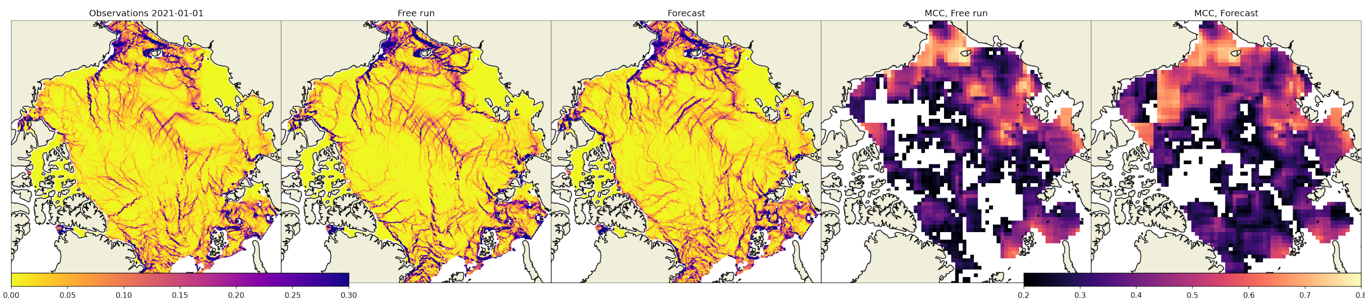

Figure 1: Impact of assimilation of synthetic observations.

Fig 5b: I would have expected that the root mean square difference of the forecast run would asymptote to the free run. The fact that it does not is suspicious. This must be explained.

As we are more interested in improvement of LKF localisation, and the MCC method suits better these needs, and many more experiments (free runs, assimilation of synthetic and real observations) have to be evaluated; it was decided to exclude the D_{P90} metric from evaluation of the predictability. In the sensitivity experiments the D_{P90} metric was replaced by the Kolmogorov-Smirnov test applied to PDFs of forecasted and observed deformation. Corresponding explanations are added to the Methodology section.

The discussion of the sensitivity of the forecast on epsilon_min and the weighting factor w_d appears in two one-line paragraph. I suggest removing them, or a more in-depth discussion should be provided.

The section on sensitivity to assimilation parameters is extended as new results obtained from the assimilation of synthetic observations are achieved.

Figure 6. The constant a1 is negative. This means that the sea ice concentration is larger than 1 (see Equ 10 and since eps_tot > 0). This is not physical. The results in the figure cannot be correct.

The sign of a1 on Figure 6 was incorrect and is corrected in the revised manuscript.

I am assuming that the sign of a1 is incorrect. If so, A = F_A(eps_tot) = 1 – a1 * eps_tot =0.76 for a1 = -1.2 (Line 232) and eps_tot = 0.2 (Fig 3). For A=0.76, P* is scaled down two orders of magnitude and the ice has no strength. Any rheological model where the ice strength is set to nearly zero along a line (an observed LKF in this case) will deform along that line – this can be tested simply. The author instead argue that the brittle rheology is key to the correct simulated location of the LKFs. This is another unsubstantiated statement.

The statement on relation to rhelology is removed.

Results: The simulated concentration and thickness fields after assimilation should be presented. Reading from the deformation fields and the a1 constant derived from sensitivity experiments, we should see concentration of 0.8 along LKFs, something that is not accord with RGPS observations at 10km scale resolution. I suspect assimilating damage would help producing more realistic fields. The authors give reasons for why damage assimilation is not successful, but those are not convincing. See below for more details on this topic.

The fields of concentration and damage after assimilation are added.

The RGPS observations don't provide sea ice concentration estimates, and the current PMW observations cannot be used for accurate estimation of sea ice concentration variations at 10 km. Nevertheless, we agree that such reduction in total ice concentration is not realistic at these scales. As noted above there are two ice categories in neXtSIM: old ice and young ice, with the former being used in rheology and with the sum of both being used in thermodynamics. Reducing the older-ice concentration for increasing of ice weakness is compensated by increasing younger-ice concentration for keeping the thermodynamic balance. Thus, older-ice concentration plays a role of sub-grid parametrisation of ice weakness that is preserved on longer time scales. As opposed to sea ice damage that is working on shorter time scales. Please also see the answer below.

It is argued that the damage does not increase predictive skill of LKFs because ice heals too rapidly (24 hours). In the real world, ice heals through thermodynamic processes on much longer time scale. This choice of short healing time scale must be justified. Perhaps this is the cause of the lack of predictive associated with the assimilation of the damage.

According to the current neXtSIM parametrisation the damage increases due to healing from 0 to 1 in 15 days. This parameter was found as optimal in experiments (Rampal et al., 2019).

In the current study, in the new experiments requested by reviewers we found that, in fact, it is not the fast damage healing that reduces the efficiency of damage assimilation. We hypothesise that concentration and damage act on different time scales. This hypothesis was tested in an idealised twin-experiment: an intact initial ice field with damage=0 and concentration=1 everywhere was 'broken-up' along realistic LKFs. In one experiment, the elements in the LKFs were initiated by reducing concentration to 0.65 and in another one - by increasing damage to 1. Variation of damage and concentration in several thousand elements of broken-up and intact ice was studied (see Fig. 9).

The study shows that in case when LKFs are initiated by reduced concentration the situation is quite simple: concentration of ice in the unbroken elements is stably high, and in the broken elements it is first low and then stably increasing due to freezing (and convergence).

For damage the situation is quite different: in the initially unbroken elements the average damage remains relatively low (0.7 - 0.85), but damage variations are very large with standard deviation reaching 0.2. This happens because in some unbroken elements, that surround the initiated cracks, the internal stress exits the Mohr-Coulomb envelope and damage increases up to 1 at very short time scales (few time steps as discussed on Fig. 7 in the manuscript). Further, a cascade of damage events occurs in the neighbours of these newly broken elements. Probability of a break up (damage increase) is higher in directions of high internal stress. Thus, the information about the initiated damage is almost instantly forgotten - it is masked by many newly damaged elements.

Large scale observations of deformation at hourly frequency could probably confirm or reject the hypothesis of how damage propagates in reality, and illustrate whether or not assimilation of damage indeed leads to a more accurate deformation field **on small time scales**. However, we assimilate and validate against daily observations that show only long term memory in ice weakness expressed in reduced ice concentration.

The details on sea ice categories, the discussion of time scales at which concentration and damage act with the accompanying figures, and the maps showing impact of assimilation on damage and older-ice concentration are added to the manuscript.

The error is shown as 1/4 sigma. This is highly unusual. Typically, one would show an error envelope equal to one sigma (four times larger than what is shown in Figure 5).

The figure is re-plotted to show errors of several forecasts (see Fig. 6).

Minor Points:

Line 23: "Under external forcing the ice deforms primarily as an elastic material." Most deformations in the pack ice are plastic and occurs along LKFs. This sentence also contradicts the next sentence.

The sentences are rewritten:

"As a consequence, sea ice does not drift freely anymore, but instead exhibits an intermittent drift with localised deformation. First, under increasing external forcing the undamaged ice deforms primarily as an elastic material. Internal stresses gradually accumulate in the material until a failure criterion is reached, which corresponds to a limit when sea ice fractures, and then the ice starts deforming along the multiple narrow and elongated cracks, and does so until these later refreeze or when the load (winds and currents) on the ice changes."

Line 26: "...start deforming along multiple narrow and elongated cracks formed and does so until these later refreeze". Or when the load (winds) on the ice changes. This should be added.

Please see the previous reply.

Line 48: :"... the exact timing and position of strong deformation zones, or LKFs, is not yet predicted precisely". The exact position of LKFs will never be located precisely because it depends on unresolved weaknesses within the ice pack. What we can hope to reproduce is the timing, the orientation of the LKFs with respect to the large-scale forcing and their statistical distribution (width, length, density, angle of fracture, etc.). This should be corrected. This is another sentence that suggests incorrectly that BBM could eventually simulate LKFs position correctly.

When we stated the fact that errors still exist in small-scale features in our study, we did not intend to comment on BBM's ability to eventually reproduce these features correctly. The end of the sentence is rewritten as follows:

"Despite the recent advances in the sea ice modeling, the exact timing and spatial distribution (including orientation, width, length and angle of fracture) of strong deformation zones, or LKFs, is not yet predicted precisely."

Line 93: "The observed variables vo (damage and concentration) is computed...". "Damage" is not observed. The word "is" should read "are".

The sentence is rewritten as follows:

"The "observed" model variables damage d_o and concentration A_o are derived from the observed deformation ε_o using the following experimental formulations:"

Line 94: Sometimes, the Greek lowercase epsilon symbol is used and sometimes the lunate epsilon symbols is used. The author needs to choose only one form for consistency. See Line 94 and Equations 6-7 for examples. This needs to be corrected everywhere.

The Greek lowercase symbol (ε) is used everywhere.

Line 102: It is said that CNEMS has a temporal resolution of 12 hours on Line 102; and it is said in the next sentence that it is "observed nearly every day". This sounds contradictory. This should be clarified. See Line 115 as well where 24 hours is specified.

The sentences in Section 3.1 are rewritten as follows:

"The dataset comprises gridded products derived from Sentinel-1 synthetic aperture radar (SAR) images, with 10 km spatial resolution. Ice drift is computed from pairs of images separated by approximately 24 hours and the product is delivered every 12 hours. The spatial coverage of the product is irregular - the East Siberian, Laptev and Kara seas and the polar gap (north of 87°N) are never covered, while other Arctic regions are observed at least once nearly every 24 hours."

Line 105: "The model is forced with the European Center for Medium-Range Weather Forecasts (ECMWF)". The version of ERA should be specified: e.g. ERA5, ERA- interim, etc.

Version is added:

"The model is forced with the latest version (Cycle 45r1) of the Integrated Forecast System European Center for Medium-Range Weather Forecasts (ECMWF) (Owens and Hewson, 2018) and the TOPAZ4 (Sakov et al., 2014) ocean forcing fields (currents, sea surface temperature, sea surface salinity)."

Line 117: "total" should read "total deformation".

Corrected.

Line 203: The units of eps_tot must be given the first time it is introduced (Table 1, Line 197). At present it is only given on Line 203.

The units are specified in Section 2.

Line 235: eps_tot and divergence is used interchangeably; yet they are very different. One includes shear the other not.

The term "divergence" is used only three times: on lines 117, 144 and 233 and every time a correct relation to total deformation is given.