Authors' Response to Reviews of

Tuning parameters of a sea ice model using machine learning

Anton Korosov, Yue Ying, and Einar Ólason

Geophysical Model Development, EGUSPHERE-2024-2527

RC: ***Reviewers' Comment***, AR: Authors' Response, ☐ Manuscript Text

We are grateful to the reviewers for their detailed review of our manuscript and many valuable comments. We agree with all suggestions and have changed the manuscript accordingly.

1. Reviewer #1

1.1. General comments:

RC: ***Just an overall comment about figures. I find it hard to identify the "take-home message" of many of the figures. Some figures contain a lot of panels (figs 3 and 8) and others contain little supporting information, such as color scales or notation to tell the reader what they should be focusing on.***

AR: Figures 1, 8, and 14 have been improved according to suggestions from both reviewers. Figure 3 has been replaced with a table. Figures 1 and 4 have been moved to the Annex. They contain interesting technical information for readers who want to use the suggested method for their research and need reference data. Please see a list of the updated figures at the end of the Response.

RC: ***In Figure 8 it's hard to identify which panels show "good" and "bad" distributions, and looking closely at every single panel will probably give the reader "figure fatigue". Maybe specific good/bad pdfs can be highlighted somehow?***

AR: The figure now contains PDFs only for three cases: a clear impact of P, no effect of P, and a PDF of the RGPS descriptor, which does not match the neXtSIM PDFs (see Figure 2 below).

RC: ***To me, Figure 14 is the key figure of the paper, and I think this should also have panels showing neXtSIM with "default" parameters, so you then have panels: i) obs, ii) control, iii) ML-tuned. With this you can maybe then remove figure 7 altogether. Similarly, I think figures 3 and 4 can probably be moved to SI or removed.***

AR: The figure now contains three columns: observations, default parameters, optimized parameters, and two rows: divergence and shear (see Figure 3 below). Figure 3 is replaced with a Table, Figure 4 is moved to the Annex, and Figure 7 is improved and is kept in Section 5.1 as it visualizes a non-linear response of the deformation fields to the values of the rheology parameters (see Figure 1 below).

RC: ***Section 4.6 describes training the ML model. Here it is mentioned that training/test data are split randomly. I was wondering here if the authors have any thought on whether this could lead to data leakage? I.e. having a training and test samples which are neighboring in time. Also could you provide more details of the DNN architecture? Are neurons fully connected? Which activation functions are used?***

AR: We repeated ML-model training with train and test data sampled from different months. That slightly increases the uncertainty of the inference result but does not add a bias. The text is rewritten as follows.

~~For both models, we randomly split the training dataset (from neXtSIM) into two parts (85:15) for training and validation and then applied it to all RGPS descriptors~~ For both models, we split the dataset from neXtSIM into two parts (85:15) for training and validation. Training and validation data are taken from different months selected randomly. The models are trained on neXtSIM data and then applied to all RGPS descriptors.

AR: Yes, the neurons are fully connected; the ReLu activation function is used on hidden layers and a Linear activation on the output layer. The following sentence is added:

The DNN model contains three hidden dense layers with 32, 16, and 8 fully connected neurons. ReLu activation function is used in the hidden layers and Linear activation in the output layer.

RC: ***Section 5.5. Am I correct that in this section you evaluate the performance of the ML-predicted rheology parameters over the entire 1990-2008 period, based on an ML model which has only been trained on winter 2007-2008 data? I would expect the performance to be quite poor here given the limited training data. It would therefore be interesting to see some analysis of offline training/validation error over this extended period. E.g if you did something like a 80/10/10 training/validation/test split over the 1990-2008 period, does the ML model perform much better than just training on 2006-2007? Also is training on year-round data critical for capturing the seasonal cycle?***

AR: In our experiments, we perturbed the rheology parameters in a wide range of values. We created the data, which allows training an ML model with a high generalization capability. Although we trained the ML model on data from winter 2006/07, this generalization capability enables us to apply it to RGPS data acquired since 1998. We could not train the ML model in earlier years because we didn't have the resources to run an ensemble of neXtSIM members for several years to produce training data. Nevertheless, we can prove that the available training data provides sufficient ML-model generalization.

As discussed in Sections 4.5 and 5.3, we trained an autoencoder on neXtSIM data and applied it to RGPS data from winter 2006/07 for downselecting the deformation descriptors. Application of the same autoencoder to RGPS data from the earlier period could not detect any significant anomalies. Since the encoder section of the autoencoder has the same architecture as the ML model used for inferring the rheology parameters, we can conclude that the ML model trained on data from winter 2006/07 is general enough to be applied to earlier RGPS data. Similar considerations apply to seasonal cycle: no significant anomalies were detected, therefore the ML-model is usable during the entire year and is applicable to detect the seasonal cycle.

The following text is added in Section 5.5.

The PDFs of parameters presented in Figure 13 are derived from all RGPS descriptors computed in the winter of 2006–2007. However, we can apply the ML model trained on neXtSIM data from winter 2006/07 also to the RGPS data acquired since 1998. To test that the trained ML model has sufficient generalization skills, we applied the autoencoder described in Sections 4.5 and 5.3 to the RGPS data from the earlier period. We could not detect any significant anomalies in this data. Since the encoder section of the autoencoder has the same architecture as the ML model used for inferring the rheology parameters, we can conclude that the ML model trained on data from winter 2006/07 is general enough to be applied to the earlier RGPS data.

RC: ***How do the ML-tuned parameters affect biases in other state variables, such as sea ice thickness? Have you checked changes in thickness and compared to observations? E.g. if you run simulations over the 2010—present period with new parameters, is sea ice thickness more in-line with CryoSat-2?***

AR: We evaluated the impact of the rheology parameters on the accuracy of sea ice drift simulation through comparison with the RGPS-derived drift and ICESat-1 sea ice thickness. Our evaluation shows that the optimal parameterization slightly improves the drift (RMSE of drift decreases from 0.05 to 0.04 m/s) and does not significantly impact the ice thickness distribution (RMSE of thickness is approx. 1 m). The following text was added to section 5.4.

> The simulated sea ice drift was validated against the RGPS drift by comparing the velocity vectors of each virtual buoy from RGPS data and the matching node on the neXtSIM mesh. The ice drift root mean square error for the run with optimal parameters is 0.04 m/s, slightly lower than for the run with the default parameters (0.05 m/s).
>
> Sea ice thickness (SIT) from different runs was compared to monthly averaged ice thickness from ICESat-1 in March 2007. SIT RMSE is the highest ($\approx$ 1.3 m) for the runs with $C_{ref} \approx$ 0.5 MPa, but no significant differences between the other runs were found (RMSE $\approx$ 1 m).

RC: ***A last general comment: I believe all simulations in this study are regional (central Arctic), perhaps to align broadly with the zone of observational data covered by RadarSat/RGPS? Do the author have any thoughts on how well they expect the ML parameters to generalize to global simulations? (I'm thinking mainly about the Marginal Ice Zone and also to the Antarctic).***

AR: Our experiments cover various weather and sea ice conditions. Presumably, the ML model trained on such heterogeneous data is general enough to be applied in regions with similar ice conditions not covered by RGPS data. Previously, we used neXtSIM in the Antarctic, and the results were promising (Rafael et al., 2024). In the Marginal Ice Zone, the sea ice concentration is lower, the ice elasticity drops substantially (see Eq. 4), and the rheology is no longer sensitive to c_{ref}, P, or other parameters.

The following text is added to the Conclusions.

> Our experiments cover a wide range of weather and sea ice conditions: from thin young ice in the Eastern Arctic to thick MYI near the Canadian Archipelago, from the beginning to the end of the freezing period, from calm days to winter storms. Presumably, the ML model trained on such heterogeneous data is general enough to be applied in regions with similar ice conditions, e.g., Laptev, Kara, Barents, and Lincoln Seas in the Arctic or Weddel and Ross Seas in Antarctica. Applying neXTSIM in the Antarctic shows that the model reproduces the seasonal cycle of sea ice extent and that BBM rheology simulates the sea ice drift with higher accuracy (Santana et al., 2024). In other regions (e.g., the Marginal Ice Zone), where the conditions are quite different, the sea ice concentration is lower, the ice elasticity drops substantially (see Eq. 4), the rheology is no longer sensitive to c_{ref}, P, or other parameters.

RC: ***L107: I recommend citing some works which have explored Kalman Filters for sea ice parameter estimation (see references below - one of which has been applied in an MEB rheology)***

AR: The suggested references are added.

1.2. Minor comments:

RC: ***L129: are F and H equivalent functions (eq 12 and eq 10)? Can F be swapped for H to be consistent?***

AR: Yes, that was a typo. **H** is now used in Eq. 12.

RC: ***L266/267: Is there not a metric we can use to get a quantitative sense of the similarity between model and obs? E.g. Spatial pattern correlation? Or maybe Spectra?***

AR: As the metric, we used the Kholmogorov-Smirnov (KS) test to compare the PDFs of the deformation descriptors from RGPS and all neXtSIM runs. It confirmed that the simulations with the optimal rheology parameters best match the RGPS observations.

The text on lines L266/267 is rewritten as follows.

> Visually, it is hard to say which neXtSIM maps better match the RGPS data, but we can use the similarity of PDFs of deformation descriptors as the metric. Nevertheless, it is clear that optimization of multiple rheology parameters is required to find the best match.

Results of the KS testing are now introduced in section 5.4.

> We compared the PDFs of the deformation descriptors using the Kholmogorov-Smirnov (KS) test. The KS test is applied to the PDFs of deformation descriptors computed from neXtSIM and RGPS and is averaged over all usable descriptors. The average KS test is the lowest for the neXtSIM run with the optimal parameters (0.41) and is slightly lower than for the default parameters (0.49).

RC: ***The manuscript appears to have very few grammatical errors.***

AR: The text is checked with automated tools, and all grammar errors are corrected.

2. Reviewer #2

2.1. Minor comments:

RC: ***Line 27: Add a sentence on the EAP rheology in this paragraph***

AR: The following sentence is added:

> Additionally, elastic-plastic-anisotropic (EAP) approach was introduced by parameterizing the anisotropy of the ice stress through interactions of diamond-shaped floes (Tsamados et al., 2013; Wilchinsky and Feltham, 2006).

RC: ***Line 49: Can Fig. 2(a) from Olason et al. (2022) be reproduced here in some way? It is hard to visualise it without looking it up.***

AR: The figure is added to the Annex. (see Figure 5 below)

RC: ***Line 54: Definition of "d" (damage) is missing***

AR: The following definition is added:

> , where damage is a single scalar to parametrise the fracture density at the sub-grid scale. The damage value is altered whenever the local stress exceeds the Mohr-Coulomb failure criterion.

RC: ***Line 58: Definition of "Pmax" (elastic limit) is missing. Would be useful for the reader if the term "ice strength" was used in relation to this in the paper as well, for clarity.***

AR: The following definition is added:

> , P_{max} is a compressive ice strength threshold that separates elastic from elastic and stress-dissipative behaviour of sea ice.

RC: ***Line 68: Define "h"***

AR: The following definition is added:

> h is sea ice thickness,

RC: ***Line 79: Missing a closing bracket after 'time'. Define 'α' here too (referred to as damage parameter in Table 1)***

AR: A bracket is added and a definition of alpha is added:

> $\alpha > 0$ is a constant,

RC: ***Line 80: Add comma between 'envelope' and 'or yield curve'***

AR: Corrected.

RC: ***Line 82: Define τ and θ_N (stress invariants), equation 7.***

AR: The definitions of τ and σ_N are already providedon line 62.

2.2. Table 1:

RC: ***Remove ??? after "Reference thickness"***

AR: Corrected.

RC: ***Compaction parameter, C: Should this be negative? In equation (2) there is a minus C, but that means C itself should be positive (I may have misinterpreted here)***

AR: It should be negative. The equation corrected accordingly.

RC: ***For P0 (scaling parameter for ridging), is this not the scaling parameter for the sea ice compressive strength, rather than for the ridging directly? So the resistance to ridging... Suggest renaming "scaling parameter for ice strength" or similar***

AR: Replaced with Scaling parameter for compressive strength

RC: ***Ca also appears as CA in several places in the text, needs to be consistent. Can a line be added to the text to explain why this is included as a rheology parameter?***

AR: The following sentence is added on line 88:

> The ice-atmosphere drag coefficient C_a is also added to the table (although, strictly speaking, it is not a rheology parameter) because it controls the amount of energy transferred from wind and ocean into

sea ice, strongly affects the sea ice drift speed and, correspondingly, sea ice deformation.

RC: ***Line 87: Schulson et al. reference should be in brackets***

AR: Brackets are added.

RC: ***Line 89: What does "proper tuning" mean? Expand on this to link into this study and provide further justification for the work.***

AR: The phrase is rewritten as follows.

> such parameters as P_0 and c_{ref} require ~~proper~~ tuning using satellite observations at large spatial and temporal scales. Given that the rheology parameters nonlinearly affect the field of sea ice deformation, a metric based on satellite-derived deformation should be used, and the tuning should capitalise on nonlinear methods, such as deep learning.

RC: ***Figure 1: Would be useful to include a legend to explain the colours. Correct the duplication of the word "images" in caption.***

AR: Figure 1 is improved and moved to the Annex (see Figure 6 below).

RC: ***Line 127: Define the operator "F"***

AR: F is replaced with H.

RC: ***Line 136: Need a reference for Latin Hypercube***

AR: The reference is added.

RC: ***Line 137: Were these parameters perturbed using the same method?***

AR: Yes, the text is updated accordingly.

RC: ***Figure 2: I don't think this figure is currently referred to in the text. I think "H" between "observed ice drift" and "observed descriptors" should be "F". Suggest also defining M, H, and F in the figure caption as well for clarity.***

AR: The typo in Eq. 12 is fixed, F is replaced wth H.

RC: ***Figure 3: Suggest larger font on figure axes to improve readability, and include more description in the figure caption.***

AR: Figure 3 is replaced with a table.

RC: ***Line 216: add definition of P90***

AR: The definition is added on a previous occurence of P90:

> 90^{th} percentile (P90)

RC: ***Lines 217-8: "mean and P90 of image anisotropy" – says median in the table, is this correct? Also line 163 only refers to mean.***

AR: Replaced with "median".

RC: ***Line 221: Include years with the dates given***

AR: The years are added.

RC: ***Line 229: How many descriptors were rejected by this method?***

AR: Six descriptors are rejected, the text is updated accordingly.

RC: ***Line 254: Need reference for "Adam optimiser"***

AR: The reference is provided.

RC: ***Line 259: "sea deformation" should be "sea ice deformation"***

AR: Corrected accordingly.

RC: ***Figure 7 caption: what is "tree days snapshots"?***

AR: Figure 7 is improved (see Figure 1 below), and the caption is rewritten:

> Maps of total deformation from RGPS and neXtSIM ~~computed from tree days snapshots~~ for three selected dates representing moderate, high and low deformation events. Each map represents a three-day mosaic, i.e., the deformation is derived from pairs of Radarsat-1 images (and corresponding neXtSIM snapshots) accumulated over three days starting from the indicated date.

RC: ***Lines 281-2: This is a repeat of information in lines 106-7, remove one of these instances.***

AR: The sentence on lines 281-281 is removed.

RC: ***Figure 10: What are the error bars? The a90_00 error bar is very large, and this descriptor is excluded by the next test (Figure 11). Would be useful to see this link pointed out in the text and, more generally, it would also be useful to know how much overlap there is in the descriptors eliminated by the different methods. How many are rejected by each method?***

AR: The error bars represent the normalised standard deviation of RGPS data. Six descriptors were eliminated by the first filter (difference of relative mean value) and are not used in the second filter. The text in Section 4.5 has been rewritten as follows:

> We test the applicability of these descriptors ~~using two methods~~ in two steps: comparison of PDFs for descriptors from RGPS and neXtSIM on the one hand and using an autoencoder on the other. In the first ~~method~~ step

The following text is added to Section 5.2.

> The standard deviation of the RGPS descriptor $a90\_00$ is very large compared to the neXtSIM due to noise in the RGPS observations of ice drift and deformation (see Figure 3, left panel).

RC: ***Figure 12: Are the black lines actual trend lines or 1:1 lines? Clarify in figure caption. It would be helpful to show the r values on the plots themselves.***

AR: The black lines are 1:1 lines. The caption has been updated accordingly, and the r values have been added.

RC: ***Lines 315-7: The qualitative assessment using Figure 14 is useful, but the similarity between the optimised neXtSIM run and the RGPS reference data should be quantified (using a metric of your choice). Also, can the optimised values of the parameters be shown in Table 1 alongside the originals for comparison?***

AR: Values of the optimal parameters are added to Table 1. The similarity between neXtSIM and RGPS runs is estimated by comparing the PDFs of the deformation descriptors using the Kholmogorov-Smirnov test. The following text is added.

> We compared the PDFs of the deformation descriptors using the Kholmogorov-Smirnov (KS) test. The KS test is applied to the PDFs of deformation descriptors computed from neXtSIM and RGPS and is averaged over all usable descriptors. The average KS test is the lowest for the neXtSIM run with the optimal parameters (0.41) and is slightly lower than for the default parameters (0.49).
>
> The simulated sea ice drift was validated against the RGPS drift by comparing the velocity vectors of each virtual buoy from RGPS data and the matching node on the neXtSIM mesh. The ice drift root mean square error for the run with optimal parameters is 0.04 m/s, slightly lower than for the run with the default parameters (0.05 m/s).
>
> Sea ice thickness (SIT) from different runs was compared to monthly averaged ice thickness from ICESat-1 in March 2007. SIT RMSE is the highest ($\approx$ 1.3 m) for the runs with $C_{ref} \approx$ 0.5 MPa, but no significant differences between the other runs were found (RMSE $\approx$ 1 m).

RC: ***Line 339: The negative values are seen in which parameters?***

AR: The scaling parameter for compressive strength becomes negative for a short time. This detail is added to the text.

RC: ***Line 341: AC should be CA (or Ca)***

AR: Corrected as C_A.

RC: ***Line 351: Suggest rewording "requires better values of H and C parameters" to "requires optimised tuning of H and C parameters".***

AR: Reworded as suggested.

RC: ***Line 351: Unclear how tuning the parameters would make the rheology independent of ice thickness (also line 371). Would we not expect to see changes in the sea ice deformation patterns related to the thickness (and strength) of the sea ice?***

AR: I principle, the same rheology (same parametrisation) should work for thick and thin ice. However, in the first experiment, we see a long-term drift of P and C_{ref} parameters derived from RGPS data. The intention of including H and C parameters was to reduce the trend in P and C_{ref} by tuning the relation between h and P_{max}.

RC: ***Line 351: Can this motivation be included earlier?***

AR: The text in section 4.2 is modified the following way.

> In the second experiment with 70 members, the following parameters were perturbed: P_0, c_{ref}, H, $\cancel{A}$ C

and C_a. The H and C parameters were added because they control the influence of sea ice thickness on P_{max}.

RC: ***Figure 14: Font size on the colourbars needs to be increased for legibility***

AR: The figure is improved according to the recommendations of two reviewers. It contains three columns: observations, default parameters, optimized parameters; and two rows: divergence and shear (see Figure 3 below).

RC: ***Lines 354-5: Refer to figure 15 here***

AR: A reference to the figure is added.

RC: ***Figure 15: Rotate text on overlapping dates for legibility***

AR: The figure updated accordingly (see Figure 4 below).

3. Updated figures in the main text

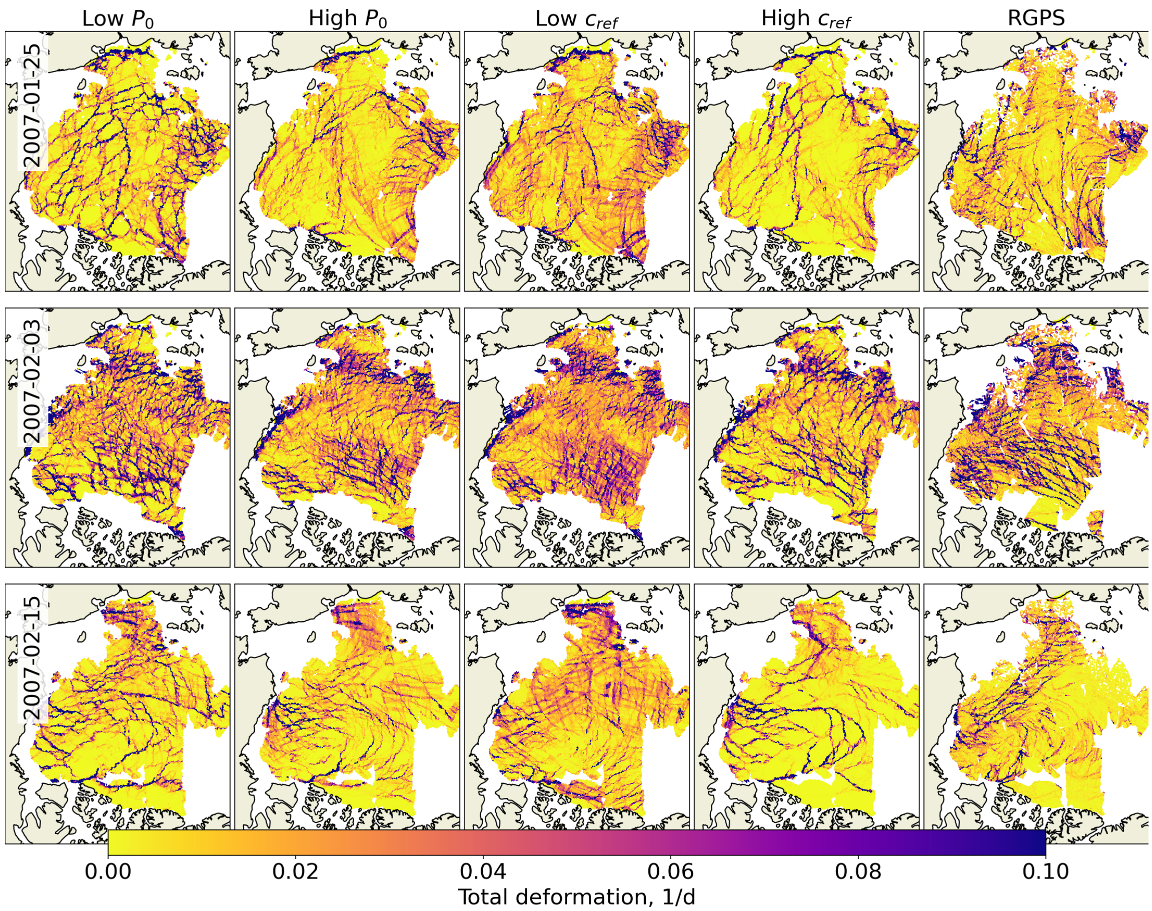

Figure 1: Maps of total deformation from RGPS and neXtSIM for three selected dates representing low, moderate and strong deformation events. Each map represents a three-day mosaic, i.e., the deformation is derived from pairs of Radarsat-1 images (and corresponding neXtSIM snapshots) accumulated over three days starting from the indicated date.

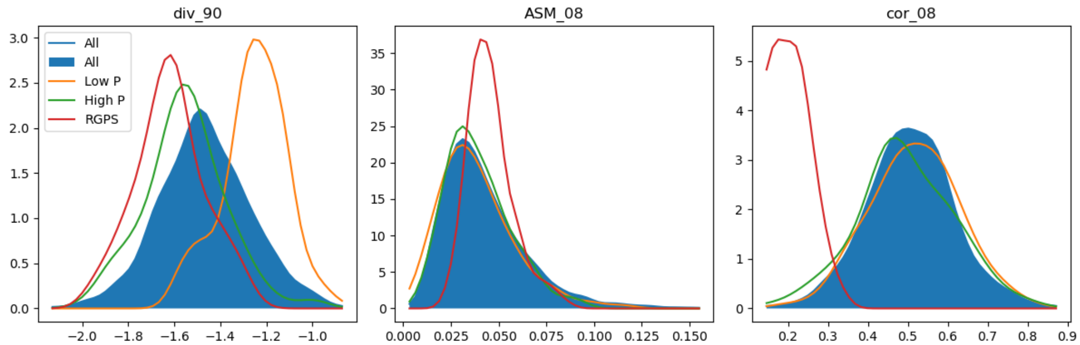

Figure 2: PDFs of few deformation descriptors for RGPS (red), all neXtSIM runs (blue), and runs with lowest (orange) or highest values of P_0. The descriptor div_90 is promising as it shows strong sensitivity to P_0 parameter, and the RGPS values vary within a similar range. The descriptors ASM_08 and cor_08 are less usable as they are either not sensitive to P_0 (i.e., ASM_08) or RGPS values are out of the training range (i.e., cor_08).

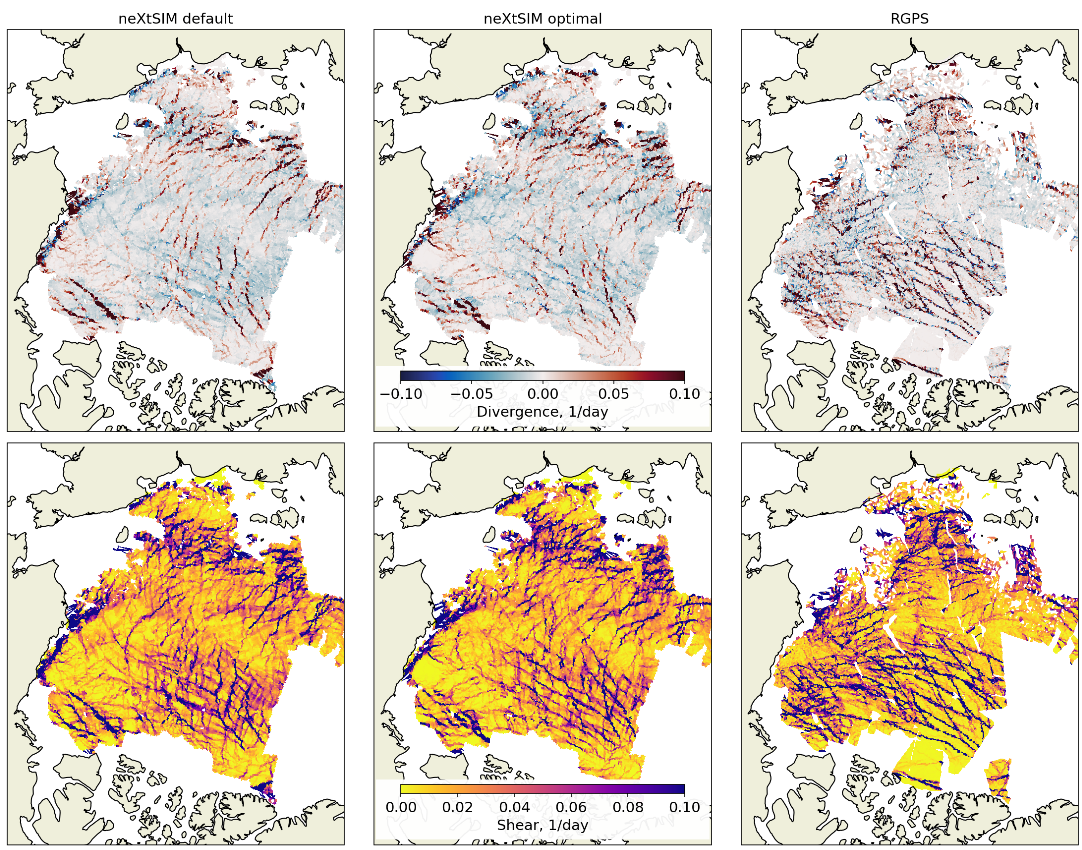

Figure 3: Maps of divergence and shear from neXtSIM run with default parameters (left column), optimal parameters (central column) and from RGPS (right column) for 3^{rd} February 2007.

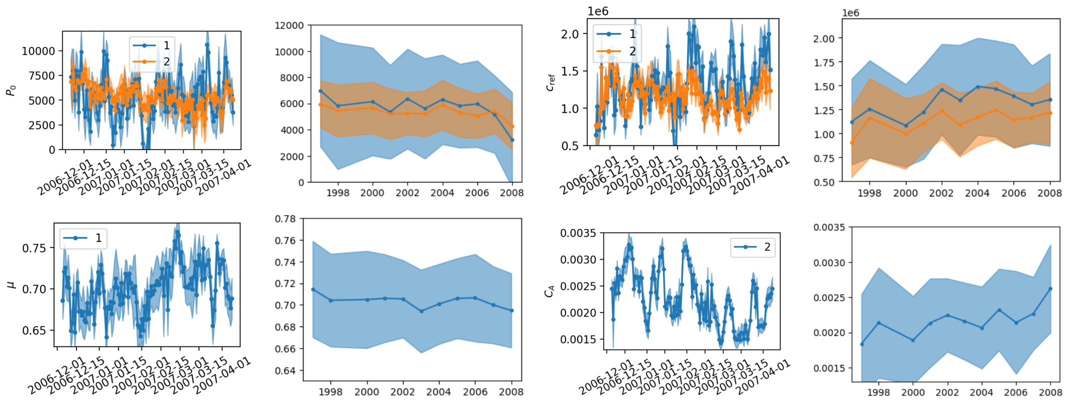

Figure 4: Time series of parameter values derived from RGPS for one year (left column) and several years (right column). Colour denotes the experiment, and the shaded area shows the standard deviation of samples produced by ten neural networks for the daily values (left column) or the samples collected from the entire year (right column).

4. Updated figures in the Annex

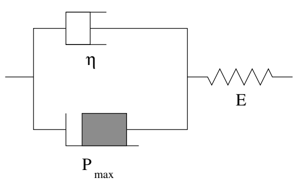

Figure 5: A schematic of the Bingham-Maxwell constitutive model showing a dashpot and a friction element connected in parallel, with both connected to a spring in series. The figure is adapted from Ólason et al., 2022.

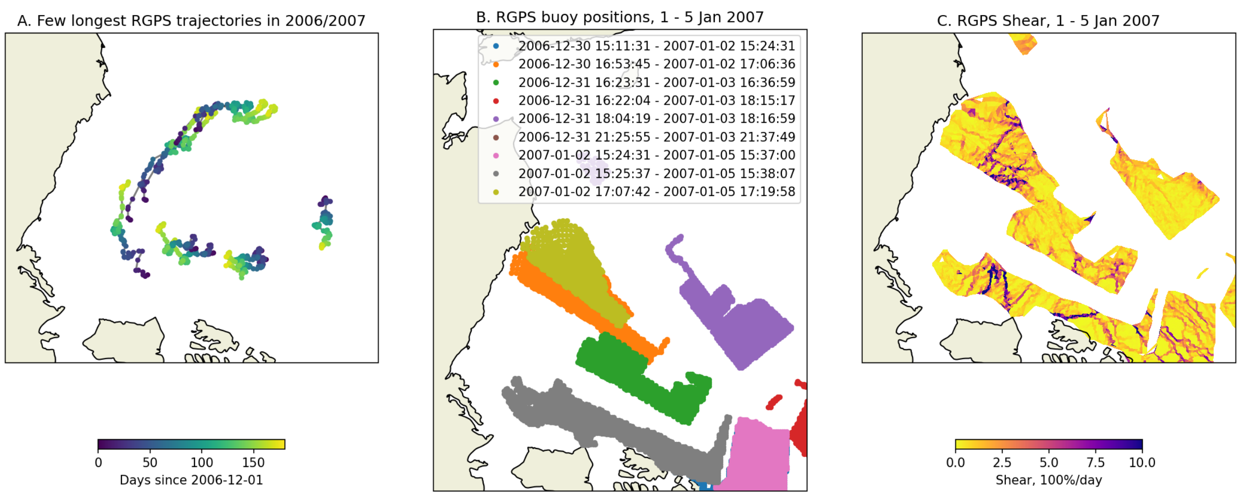

Figure 6: **A.** Example trajectories of virtual buoys detected on Radarsat-1 data by the RGPS system between 1^{st} December 2006 and 15^{th} May 2007. **B.** Position of virtual buoys on SAR images acquired between 1^{st} and 5^{th} January 2007. Points are coloured by the starting and ending image acquisition time, as shown in the legend. **C.** Shear computed from the Radarsat-1 image pairs shown on **B**.