

Response to Reviewer 2 – Super-resolution of Arctic Sea Ice thickness using a conditional diffusion model (egusphere-2026-2318)

The authors would like to thank the editor and the reviewer for their constructive comments and suggestions that have helped improve the quality of this manuscript. The manuscript has undergone a thorough revision according to the reviewer’s comments. Please see below our point-by-point responses.

Reviewer 2

Main comments

Reviewer Comment 2.1 — I believe the discussion section could be substantially more elaborated. At present, it focuses mostly on technical neural network parameters and on the split between the deterministic and probabilistic components. However, I think several points raised in the paper merit more in-depth discussion. In particular, the fact that the model is trained on simulated low-resolution observations from a given model (well known for its ability to represent deformation in the ice pack) and then applied to real observations constitutes a strong hypothesis that warrants further discussion. Would the results still hold if a different model were used? I would also be very interested to know how much the pre-processing step used to match the neXtSIM-simulated low-resolution field to observations affects the results.

Reply: Thank you for this comment. Our response is twofold.

First, the proposed approach relies on the assumption that a model trained on degraded and smoothed model fields can be transferred to observational products such as CS2SMOS. This assumption inherently introduces a distribution shift between the training data and the observational data used during inference. We agree that this is an important limitation, and the discussion of this issue will be expanded in the revised manuscript. In addition, we will evaluate the sensitivity to one aspect of this distribution shift, namely the effect of the observational mask. Please see our response to Reviewer 1, Comment 1.2.

Second, a deliberate choice was made regarding the model used to generate the training dataset. We will add the following justification to Section 2.3, “Low-resolution model dataset”:

“All preprocessing steps were applied to neXtSIM model fields. If a different physical model were used, other preprocessing steps would likely be required to reproduce the characteristics of the satellite observations. In principle, the methodology can be applied with any model, provided that a transformation can be defined to map the model space onto the observation space. In our case, this transformation mainly consists of smoothing and rescaling operations.

The choice of model also affects the quality of the high-resolution fields generated by the diffusion model. We selected neXtSIM because of its relatively realistic representation of sea ice dynamics. However, the quality of the reconstructed fields ultimately depends on the quality of the underlying model at the considered spatial scale. The focus of this study is the methodological framework rather than the specific choice of model. Although neXtSIM was used as the reference model in this work, the same approach could be applied to other models and is, in principle, equally valid provided that an appropriate preprocessing strategy is defined.”

Reviewer Comment 2.2 — The overall quality of the figures could be improved. All figures appear pixelated, which is particularly problematic for subplots containing many maps, especially Figures 1 and 2. The color scheme in Figure 7 could also be improved, in my opinion, to make it easier to distinguish differences between the CryoSat, ML, and LR PSD curves.

Reply: We thank the reviewer for this comment. We will provide higher-quality figures in the revised manuscript.

The colour scheme used in Figure 7 was deliberately chosen so that the CryoSat-2 data and the corresponding super-resolution estimates share the same colour at a given spatial resolution. This facilitates the comparison between the CryoSat-2 observations and the reconstructed fields. For example, the reader can immediately identify the correspondence between the 24 km CryoSat-2 curve and the associated ML-based estimate.

To further improve readability and make the distinction between CryoSat-2 observations and the other estimates clearer, we will replace the dashed lines used for the CryoSat-2 products with dotted lines in the revised figure.

Reviewer Comment 2.3 — I find the question of seasonality quite interesting, and almost puzzling. I understand the explanation offered in terms of increased sea ice thickness, but would it be possible to use a “normalized” metric to further validate this point, for example, by dividing the RMSE by the mean SIT? Additionally, could the use of ECMWF forecasts instead of ERA5 in 2023 be another contributing factor? I think this point deserves a mention somewhere in the manuscript.

Reply: We thank the reviewer for this comment. The advantage of the RMSE is that it provides an error measure in physical units. However, this comes with the drawback of exhibiting a pronounced seasonal cycle. This seasonality is not driven solely by changes in the mean SIT, but also by variations in sea-ice extent and in the spatial variability of SIT throughout the year.

To account for these seasonal effects, we computed a relative RMSE (relRMSE). For each grid point with a reference SIT greater than 0.1 m, we calculated the squared error normalised by the squared reference SIT. These normalised errors were then averaged over all valid grid points. The results are shown in Fig. 1.

The relRMSE exhibits a much weaker seasonal cycle than the RMSE. Using this metric, the baseline and diffusion models show comparable performance. We also observe an increase in relRMSE at the beginning of the time series. This increase is likely related to a short spin-up period in the reference simulation, corresponding to the time required for small-scale features to develop from a model initialised in October.

In the revised manuscript, we have decided to retain the original RMSE as the primary metric because it has a direct physical interpretation and accounts for all grid points, including areas of thin ice. Nevertheless, we will include the relRMSE in an appendix to support the discussion of seasonal effects. We will also add a discussion of the spin-up period and its impact on the results.

Regarding the impact of change in forcing, we expect it to be relatively small. A change in forcing may slightly affect the magnitude of deformation rates and could lead to small differences in the position and orientation of linear kinematic features. However, the dataset forced with ECMWF data is used only for validation. If the forcing mismatch had a substantial impact on the results, we would expect to observe systematic differences in the error metrics between the validation dataset (ECMWF forcing) and the test dataset (ERA5 forcing). This is not the case, as the performance metrics are comparable

across both datasets. Therefore, we conclude that the effect of the forcing mismatch is likely secondary compared to the other sources of uncertainty discussed in the manuscript.

We will add a brief discussion of this point to the Results section of the revised manuscript.

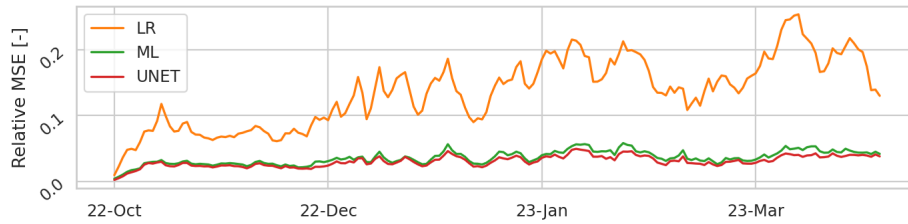


Figure 1: Temporal evolution of the relative mean square error (relRMSE) computed on the test dataset from 10 October 2022 to 15 April 2023.

Reviewer Comment 2.4 — While the overall structure and organization of the manuscript make sense, I believe there is room for improvement in two respects. First, the naming of the sections could be improved, particularly in Section 3: why is the baseline introduced within the neural network section? The same question applies to the metrics. I would suggest either splitting this section or renaming Section 3 to better reflect its content.

Reply: Following the reviewer’s suggestions, we will implement the following modifications:

- Section 3.2 will be renamed “Application of conditional diffusion model to sea ice thickness super-resolution”.
- The current metrics subsection will be moved into a new standalone section entitled “Metrics and Baselines”.
- The description of the baseline methods will be moved to this new “Metrics and Baselines” section.
- The description of the new baseline, based on a deterministic UNET, will also be included in this section.

Specific comments

Reviewer Comment 2.5 — Sea ice / sea-ice: there are inconsistencies in the use of “sea ice” versus “sea-ice” as an adjective (e.g., “sea ice thickness” in L31 versus “sea-ice thickness” in L66). Please check all instances for consistency.

Reply: We will use the term “sea-ice” consistently throughout the revised manuscript.

Reviewer Comment 2.6 — L33: The principles of probabilistic approaches could benefit from a clearer introduction, particularly the idea of learning a distribution and then sampling from it.

Reply: In the revised version, we will reformulate the paragraph insisting on the difference between deterministic and generative models: “This probabilistic approach falls within the field of generative

machine learning [Foster et al., 2022], where the model learns a probability distribution rather than producing a single deterministic output. By running the model multiple times, an ensemble of plausible realisations can be generated, providing samples from the learned distribution.”

Reviewer Comment 2.7 — L49: The introduction could be updated to include section numbers and a brief mention of the conclusion’s content.

Reply: We will add the main conclusion and the reference to the section number in the revised version.

Reviewer Comment 2.8 — L73: This sentence could be rephrased for clarity, as its current meaning is unclear.

Reply: We agree that the sentence was too convoluted. In the revised version we will simplify it: “where u_x , u_y , v_x and v_y are the spatial gradients of the ice motion components.”

Reviewer Comment 2.9 — L123: Equations should be enclosed in parentheses.

Reply: In the revised manuscript, parentheses will be added for references to the equations.

Reviewer Comment 2.10 — L124: Please clarify whether the mean and standard deviation are computed globally or pixel-wise.

Reply: They are computed globally. This mention will be added to the revised version.

Reviewer Comment 2.11 — L131: “Input features” is a familiar term for those accustomed to neural networks, but I would suggest rephrasing this sentence so that readers less familiar with the terminology understand that it refers to the inputs of the NN.

Reply: We will clarify this point in the revised manuscript by specifying it in parentheses: “the input features (i.e. the predictor variables used as inputs to the neural network model)”

Reviewer Comment 2.12 — Algorithms 1 and 2: These describe a standard conditional DDIM algorithm; I suggest moving them to the appendix.

Reply: In the revised version, those two algorithms will be moved to the appendix.

Reviewer Comment 2.13 — Section 3.2: The section title is somewhat confusing.

Reply: Thank you for pointing that out. In the revised version, the corrected title will be “Application of conditional diffusion model to sea ice thickness super-resolution”

Reviewer Comment 2.14 — Section 4.1: The section title could be improved.

Reply: In the revised version, we will modify the title into: “Illustration with specific dates”

Reviewer Comment 2.15 — L245: Two instead of three timesteps, why does this apply only to the validation dataset?

Reply: This was a typo. We will indicate that the figure shows two time steps in the revised manuscript. The figure is intended as an illustrative example: although it was generated using samples from the validation dataset, a similar example could equally have been drawn from the test dataset. We also note that all data and results are publicly available, allowing readers to explore the model outputs for any of the available dates.

Reviewer Comment 2.16 — L267: The PSD acronym has already been defined earlier in the text.

Reply: This is correct. We will remove the acronym expansion the second time.

Reviewer Comment 2.17 — L275: Do you mean “fine-scale bias”?

Reply: Thank you, we will change it to “fine-scale bias”

Reviewer Comment 2.18 — Fig. 4: I am puzzled by the differences in PSD across time steps. Is there high variability from one time step to another? Looking at the three PSD plots, the ML PSD appears fairly consistent across all three, whereas it is the neXtSIM PSD that shows differing slopes.

Reply: This is an interesting discussion. However, after examining the results more closely, we find that the largest variability in the PSD is actually observed in the ML reconstruction (green curve). We interpret this behaviour as a consequence of the stochastic nature of the ML model, which introduces a small amount of random noise in the reconstructed small-scale features. This variability is also visible in the spatial statistics shown in Fig. 4, where the ML reconstruction slightly exceeds the reference values at the smallest scales.

We will add a brief discussion of this effect in the revised manuscript, highlighting that the increased PSD variability is likely associated with the stochastic noise generated by the model and its impact on the reconstruction of fine-scale structures.