

Response to Reviewer 1 – 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 1

Main comments

Reviewer Comment 1.1 — The baseline comparison is limited. Comparing against the low-resolution input field is necessary, but not sufficient to assess the added value of the proposed diffusion model. The paper would be stronger if the authors added at least one stronger baseline, such as a deterministic neural-network super-resolution model, a directly trained U-Net for the SIT increment (since the diffusion model used here has U-Net as backbone), a simpler stochastic baseline, or an ablation showing the added value of the diffusion formulation. This would help distinguish improvements due to the diffusion model from improvements due simply to supervised learning from neXtSIM.

Reply: Thank you for raising this very good point. To enrich the baseline comparison, we trained a deterministic UNET for the super-resolution task. For simplicity, we used the same UNET architecture as in the diffusion model. The comparison was performed on the model data, following the experimental setup described in Section 4 of the original article.

The results are shown in Figs.1, 2 and 3. We find that both the deterministic UNET and the diffusion model substantially reduce the RMSE. The deterministic UNET achieves slightly lower errors (see Fig 2) while requiring significantly less computational cost, being approximately 1000 times faster than the diffusion model. However, the diffusion model performs considerably better for several other metrics. In particular, unlike the deterministic UNET, it naturally provides an ensemble of solutions, enabling uncertainty quantification. In addition, the diffusion-based model reconstructs small-scale structures much more accurately, leading to a more realistic representation of fine-scale variability (see Fig 3).

This new comparison and the updated figures will be included in the revised version of the manuscript.

Reviewer Comment 1.2 — The construction of the synthetic low-res and high-res data seems to be central to this study, because the main quantitative validation is performed in a way that high-resolution neXtSIM fields are smoothed to create low-res inputs, and the original neXtSIM fields are then used as ground truth for evaluation. The observational application reveals a domain-shift issue. The authors state that spurious coastal features occur because OSI SAF deformation masks were not present in the training data. Since the model is intended for CS2SMOS/OSI SAF inputs, the synthetic training data should mimic their missing-data patterns as well if possible. Why were realistic OSI SAF masks not included during training? Perhaps the authors can clarify this point a bit more and address this in the training/evaluation setup and see if it improves. Also, it is not

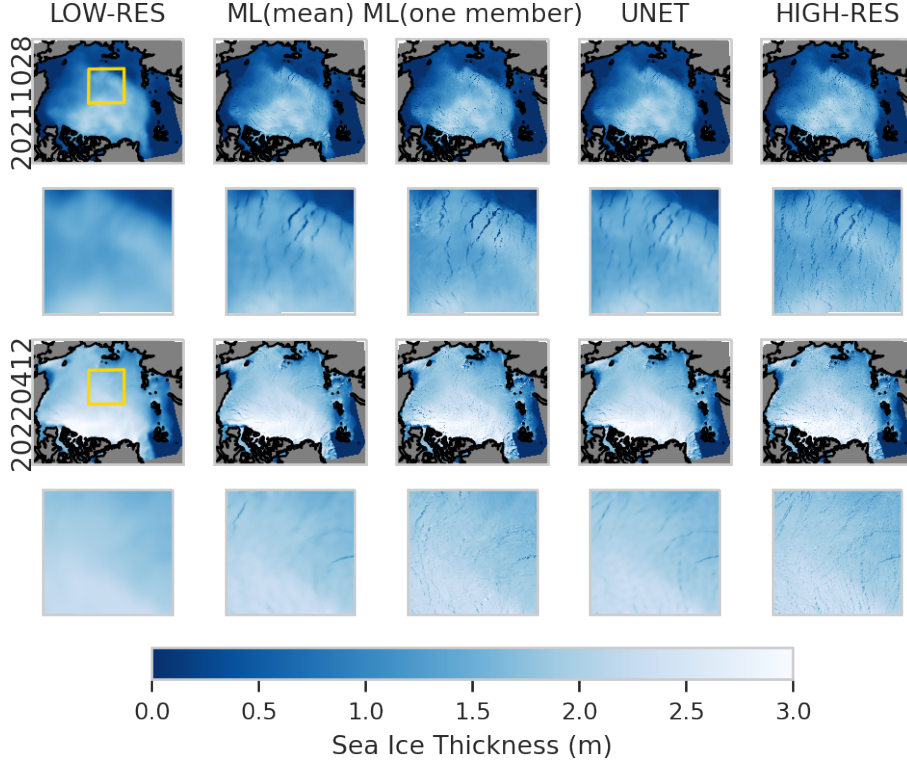


Figure 1: Examples of samples from the validation dataset for two dates chosen for illustration, 28 October 2021 (20211028) and 12 April 2022 (20220412). The column 1 shows the low-resolution product, the column 2 shows the average over the ensemble of 30 ML predictions, the column 3 shows the prediction of one arbitrary member, the column 4 shows the prediction of the deterministic UNET, and the column 5 shows the high-resolution neXtSIM simulation used as ground truth. The second row and fourth row depict a zoom of the figure just above. The zoom region is defined as the yellow box drawn in the first figure of the row.

clear what kinds of CS2SMOS patterns are absent from the training data, like where they occur, or how strongly they affect the results. The authors should elaborate on this point and provide clearer examples.

Reply: Thank you for raising this point. There are many choices involved in generating low-resolution data, and a balance must be struck between realism and generality. In our case, we chose to keep the training dataset relatively generic rather than tailoring it very specifically to the observational dataset. We constrained the preprocessing to reproduce the spatial variability of the target fields, while deliberately excluding other characteristics of the satellite product, such as the observational mask. This choice makes the approach more general and is consistent with the methodological focus of the paper. As stated in the manuscript, the proposed methodology could also be applied to other super-resolution problems, such as reconstructing high-resolution fields from low-resolution model outputs.

Nevertheless, we agree with the reviewer that this choice introduces an obvious domain shift between the training and observational datasets. To assess the impact of this issue, we extrapolated the drift product near the coast to remove the coastal mask present in the OSISAF drift data in the observational dataset. Divergence and shear fields were gap-filled wherever SIC and SIT provided valid data but the

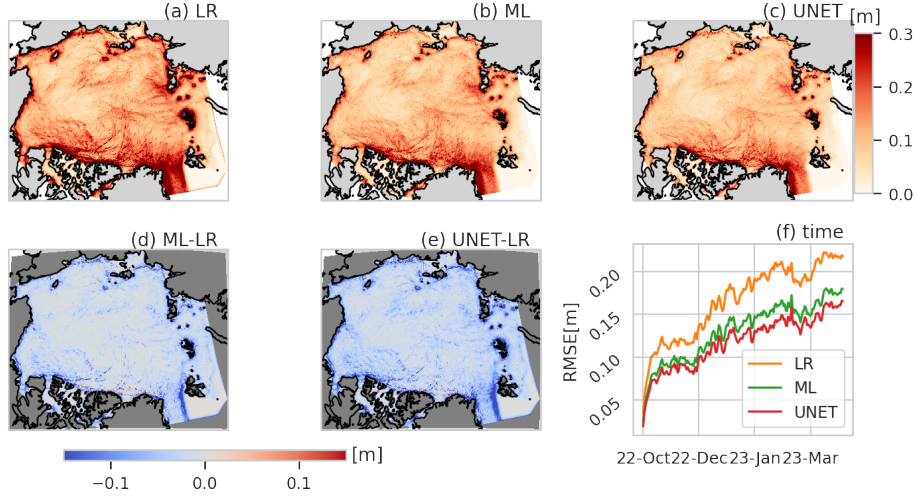


Figure 2: Root Mean Square Error (RMSE, m) computed on the test dataset from 10 October 2022 to 15 April 2023. The upper panel shows the spatial RMSE for (a) the low-resolution (LR) baseline, (b) the ML retrieval, averaged over 30 ensemble members and (c) the retrieval of the deterministic UNET. Panel (d) displays the difference between the ML and LR RMSE and (e) between the UNET and LR, where blue indicates regions where the ML retrieval outperforms the baseline. The temporal evolution of the RMSE is presented in panel (f).

drift-derived fields were missing. Missing values were first filled using nearest-neighbour extrapolation from the closest valid pixel and subsequently smoothed with a NaN-aware Gaussian filter ($\sigma = 1$ pixel, truncated at 2σ). The filter excludes permanently invalid pixels (e.g. land or no-observation areas) from the convolution kernel, preventing contamination of the reconstructed values by non-physical regions. The resulting fields were then masked back to the common SIC/SIT valid-data mask, ensuring identical spatial coverage across all input variables while introducing no modifications outside the originally missing regions.

The results presented in Fig. 4 show a slight reduction in artefacts near the domain boundaries when using the gap-filled drift product. However, the overall results remain largely unchanged. We will discuss this point in more detail in the revised manuscript and include the comparison in an appendix. These results suggest that, although the domain shift is real, it is not primarily caused by the coastal mask. Instead, as already discussed in the manuscript, it is more likely related to other characteristics of the CS2SMOS product that are absent from the training data.

As suggested by the reviewer, we will further elaborate on this point in the revised version:

“The CS2SMOS product is generated by combining observations from multiple satellite tracks collected over several days. Since these tracks are not acquired simultaneously, the resulting product contains artefacts that are not reproduced by our preprocessing procedure, which considers only daily fields. The smoothing and noise-addition steps only approximate the complex aggregation process used to generate CS2SMOS. Furthermore, no preprocessing of NeXtSIM output can fully reproduce the CS2SMOS product because the observed quantity (ice freeboard) differs from the quantity simulated by NeXtSIM. Nevertheless, the closer the preprocessing procedure is to the actual CS2SMOS processing chain, the smaller the expected domain shift should be.”

We consider the development of a more sophisticated preprocessing strategy to be beyond the scope of the present study, especially given the limited availability of independent metrics for validating the

reconstructed fields. In practice, it would be difficult to objectively assess the benefits of a more complex preprocessing approach.

Other comments

Reviewer Comment 1.3 — “These quantities are computed as follow” should be “computed as follows.”

Reply: This will be corrected in the revised version.

Reviewer Comment 1.4 — “high-resolution sea thickness fields” should be “high-resolution sea-ice thickness fields.”

Reply: This will be corrected in the revised version.

Reviewer Comment 1.5 — Section 3.2 is titled “Application of conditional diffusion model to sea ice thickness forecast,” but the manuscript is about super-resolution/reconstruction rather than forecasting. The author should change or clarify this.

Reply: In the revised version, the corrected title will be “Application of conditional diffusion model to sea ice thickness super-resolution”

Reviewer Comment 1.6 — Please check the consistency of Figure 2 and its description. The text says Figure 2 shows examples at “three different dates,” but the caption lists two dates.

Reply: Sorry for this typo, we will change it to “two different dates” in the revised version.

Reviewer Comment 1.7 — The Figure 2 caption says column 1 shows the low-resolution product CS2SMOS, but Figure 2 is presented in the model-data validation section, where the low-resolution input should be the synthetic low-resolution neXtSIM-derived field rather than real CS2SMOS. This should be clarified or corrected.

Reply: Thank you, we will modify to “low-resolution model field”

Reviewer Comment 1.8 — The capitalisation of neXtSIM should be consistent throughout the manuscript. Most occurrences use “neXtSIM,” but Figure 2 uses “NeXtSIM.”

Reply: Thank you, we will correct to “neXtSIM” in the figure 2 caption.

Reviewer Comment 1.9 — The manuscript uses both “realization/realizations” and “realisation/realisations.” Either American or British spelling is fine, but it should be consistent.

Reply: Thank you for pointing that out, we will use British spelling “realisation”.

Reviewer Comment 1.10 — “trade off” should be “trade-off” when used as a noun.

Reply: Thanks, this will be modified in the revised version.

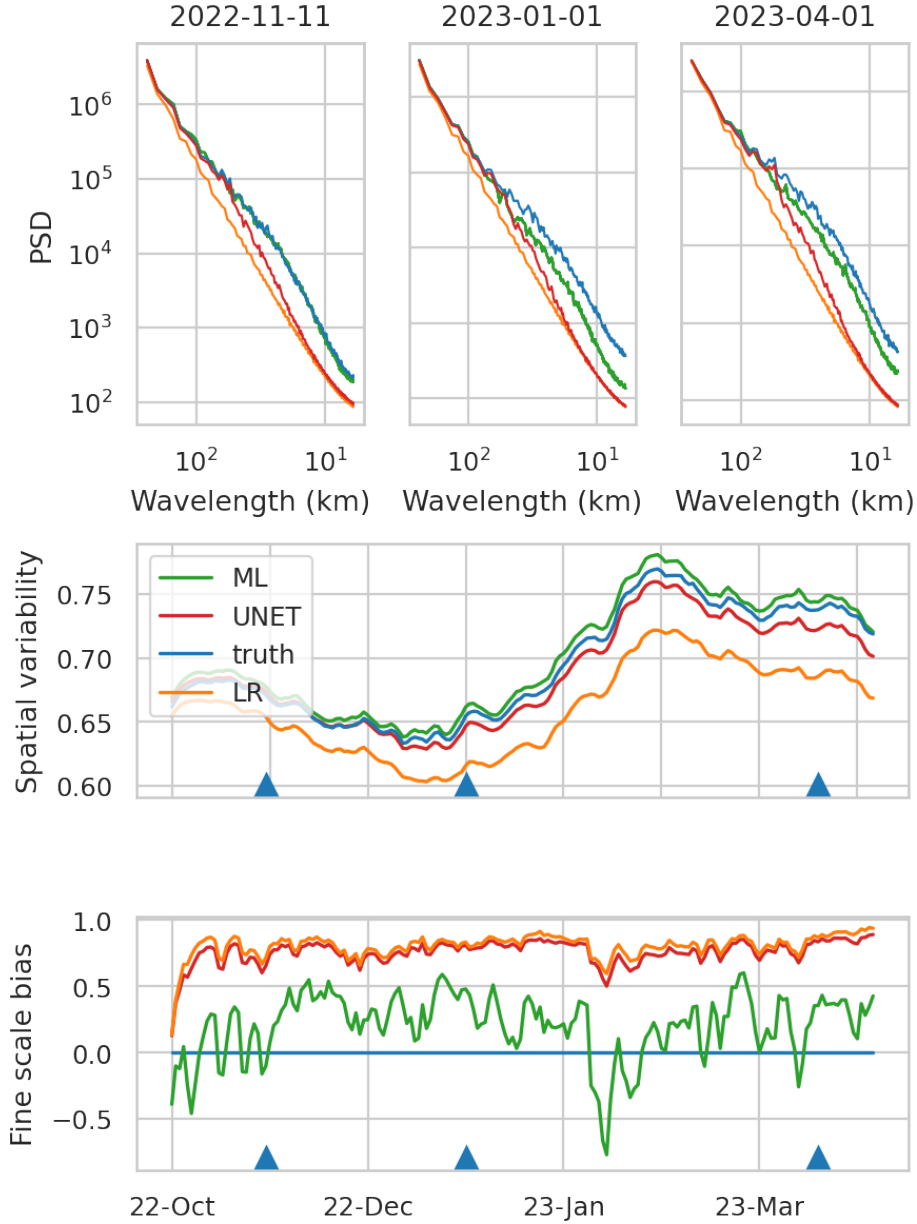


Figure 3: The upper panel shows the Power Spectral Density (PSD) for three selected dates representing the beginning, middle, and end of the test dataset. The middle panel illustrates the spatial variability across the entire test period, while the bottom panel presents the fine-scale bias relative to the true high-resolution field. Metrics are defined in section 4. In all panels, the blue line represents the true high-resolution field, the orange line denotes the low-resolution baseline, the red line depicts the deterministic UNET and the green line corresponds to the ML retrievals. Metrics for the ML retrievals are averaged across all ensemble members. Blue triangles indicate the dates selected in the upper panel.

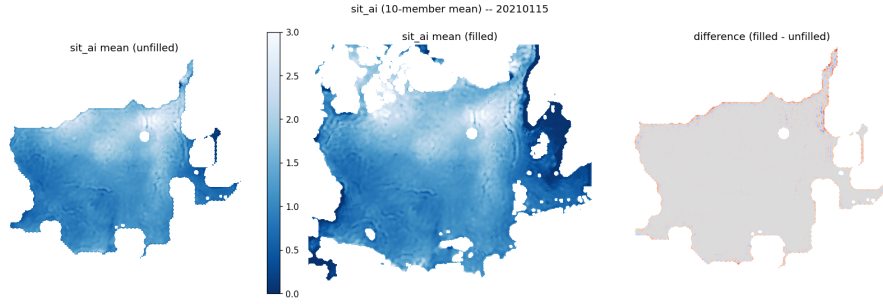


Figure 4: Comparison for an observational sample from 15 January 2021. The SIT field was obtained by averaging 10 realisations generated by the diffusion model. The first column shows the reconstruction obtained using the original OSISAF drift field as input, which contains missing values along the coastline. The second column shows the reconstruction obtained using the gap-filled OSISAF drift field, enabling reconstruction closer to the coast. The third column shows the difference between the two reconstructions over the common valid grid points.