https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

Super-resolution of Arctic Sea Ice thickness using a conditional diffusion model

Julien Brajard, Anton Korosov, Fabio Mangini, Richard Davy, and Yiguo Wang
Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research, Bergen, Norway
Correspondence: Julien Brajard (julien.brajard@nersc.no)

Abstract. Small-scale variability (3–60 km) in Arctic sea-ice thickness plays a crucial role in sea-ice predictability and in the climate system. However, these scales are neither directly observed nor adequately represented in climate models. While coarse-resolution observational products (e.g., CS2SMOS) and some high-resolution model simulations exist, bridging the scale gap remains challenging.

In this work, we use machine learning to develop a super-resolution algorithm that reconstructs small-scale sea-ice thickness features from low-resolution input fields. The algorithm is trained on realistic high-resolution model simulations and is based on diffusion models conditioned on low-resolution observations. This class of models is inherently probabilistic, enabling the generation of an ensemble of plausible high-resolution reconstructions from a single coarse-resolution input.

We apply the method both to model simulation, where high-resolution ground truth is available, and to the CS2SMOS observational product. We demonstrate that the algorithm produces realistic high-resolution sea-ice thickness fields with improved accuracy and provides meaningful uncertainty estimates through the ensemble spread.

1 Introduction

In the Arctic, the sea ice cover exhibits a diverse range of spatial scales. This variety is observable in various sea ice state variables, including ice concentration, drift, and particularly thickness, which is the subject of this study. At large scales, sea-ice thickness and concentration tend to form broad, gradually varying patterns that are primarily shaped by thermodynamic growth and melt, together with large-scale atmospheric and oceanic circulation (Notz et al., 2016). At smaller scales, features such as leads and ridges arise mainly from mechanical deformation of the ice pack, driven by wind and ocean stress and modulated by the ice's local state, including its thickness, concentration, and strength (Olason et al., 2022).

In return, the fine-scale patterns in sea ice thickness are crucial in influencing the dynamics of sea ice and its interactions with oceanic and atmospheric systems. For instance, leads can significantly affect heat flux over extensive areas (Komuro and Suzuki, 2013; Tian et al., 2025), and pressure ridges affect the form drag coefficient (Mchedlishvili et al., 2023). Furthermore, the distribution of sea ice thickness influences the predictability of sea ice over short-term periods of a few days (Sakov et al., 2012; Korosov et al., 2023) up to seasonal timescales (Kimmritz et al., 2019).

Although significant, the fine-scale variations in sea ice thickness are not routinely observed across the Arctic (Ricker et al., 2017) and are rarely incorporated into Earth system models (Notz et al., 2016). This representation gap is expected to introduce

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

biases, particularly within the atmospheric boundary layer (Duffey et al., 2025), due to inaccurate representation of heat fluxes, as well as in the seasonal prediction of sea ice extent (Kimmritz et al., 2019).

We propose a machine learning (ML) algorithm to generate high-resolution sea ice thickness fields from low-resolution observable sea ice variables. We adopt a probabilistic approach since this problem cannot be solved deterministically i.e. large-scale predictors do not contain all information about small-scale features (e.g. precise lead positions). Our goal is to estimate the probability distribution of high-resolution sea thickness fields based on low-resolution predictors.

This probabilistic approach falls within the field of generative machine learning (Foster, 2022), where the model can produce samples from a probability distribution. Running the model multiple times creates an ensemble that samples the probabilistic distribution.

Various generative machine learning techniques have been applied to downscaling tasks, including Generative Adversarial Networks (GANs) (Ledig et al., 2017), normalizing flows (Groenke et al., 2020), and variational autoencoders (Liu et al., 2020). More recently, diffusion models (Song et al., 2020) have also been explored for this purpose, and this is the class of models we employ in the present study. Aich et al. (2025) proposed an approach closely related to ours for downscaling precipitation fields. In their method, both model and observational data are mapped into a shared embedding space, which serves as the domain for the diffusion process. In contrast, our approach, described in detail in Section 3, uses the low-resolution space itself as the embedding space. This allows our model to be directly conditioned on low-resolution sea ice observations without requiring any preprocessing. Our method can also be related to the work of Lockwood et al. (2024), who applied a generative model to tropical cyclones. Their approach involves two stages: a deterministic bias-correction neural network followed by a super-resolution step. In our case, the debiasing is not learned but instead determined using simple filters, as described in Section 2.3.

In this article we use a specific type of diffusion model which is conditioned on the low-resolution sea ice variables (Song et al., 2020). To our knowledge, this is the first application of a conditional statistical downscaling algorithm to the Arctic sea ice thickness.

This paper first introduces the data used for training and applying the algorithm, then details the diffusion algorithm for training and inference. We present results on both model data and real observations, followed by a discussion of our findings.

2 Data

2.1 High-resolution model dataset

For training and validation of our algorithm, high-resolution sea ice variables were generated using the next-generation sea ice model neXtSIM (Olason et al., 2022). neXtSIM is a state-of-the-art model capable of realistically reproducing small-scale sea ice features. The model was run at a spatial resolution of 3 km over ten winter seasons, from October 2013 to April 2023. Oceanic forcing was provided by the TOPAZ reanalysis from 2013 to 2017 and by TOPAZ forecasts from 2018 to 2023 (Sakov et al., 2012). Atmospheric forcing consisted of ERA5 reanalysis data from 2013 to 2022 (Hersbach et al., 2020), and ECMWF forecasts for 2023 (Owens and Hewson, 2018). The model was run with assimilation of the CS2SMOS product

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

in the wintertime (Ricker et al., 2017) and sea ice concentration from the U.S. National Ice Centre daily ice charts (for details, see Williams and Bertino, 2025).

The use of different forcing datasets is due to their availability at the time of simulation. However, we expect the impact of these differences to be marginal. Minor variations may be observed in deformation rates and the orientation of linear kinematic features, but major discrepancies can be ruled out a posteriori. Indeed, a significant distributional shift would be reflected by poor validation performances when evaluating the algorithm (see Section 2.4).

The model outputs include sea ice thickness, concentration, drift, and snow depth, provided on a polar stereographic grid at 3 km resolution, four times daily, averaged over 6-hour intervals. Simulated sea-ice thickness was compared to satellite-derived estimates (see Section 2.2), and due to initialization and assimilation procedures, the bias in modeled thickness remains negligible throughout the season.

Throughout the remainder of this paper, we use the divergence ∇ and shear τ as they are directly related to sea ice deformation and the small-scale structures. These quantities are computed as follow:

$$\nabla = u_x + v_y \tag{1}$$

$$\tau = \sqrt{(u_x - v_y)^2 + (u_y + v_x)^2} \tag{2}$$

where u_x, u_y, v_x and v_y are the spatial gradients in ice motion (Kwok, 2006) components derived from the sea ice drift as input features, as they are directly related to sea ice deformation and the small-scale structures.

2.2 Satellite-derived dataset

The satellite-derived sea ice thickness dataset used in this study is the CS2SMOS product (Ricker et al., 2017), which is available at daily resolution across the entire Arctic during the freezing season (October to April). This product combines data from two satellite sensors: the altimeter CryoSat-2 and the Soil Moisture and Ocean Salinity (SMOS) radiometer. Due to the merging and interpolation procedures involved, CS2SMOS can only resolve low-resolution features (larger than approximately 50 km). Furthermore, the presence of melt ponds prevents reliable retrieval during the melting season, and the product is therefore unavailable from April to October.

In this work, we focus on the super-resolution of CS2SMOS sea ice thickness, and consequently restrict our algorithm and analysis to the freezing season.

Additional satellite products are used to inform the super-resolution process, specifically sea ice concentration and drift. These are obtained from the EUMETSAT Ocean and Sea Ice Satellite Application Facilities (OSI SAF) and are derived from passive microwave satellite data (OSI SAF, 2007; Lavergne et al., 2010, 2019; OSI SAF, 2022a, b), with daily coverage.

All satellite-derived products were interpolated onto the same polar stereographic grid as the neXtSIM model outputs, at a spatial resolution of 3 km. It is important to note that, despite the high grid resolution, the effective resolution of the satellite products remains limited (e.g. larger than 50 km for the thickness), and they do not resolve features at the 3 km scale. Enhancing the effective spatial resolution is precisely the objective of this study. Interpolation for sea-ice concentration, divergence, and shear was performed using nearest Gaussian weighting, whereas cubic spline interpolation was used for sea-ice thickness.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

Table 1. Parameters of the processing applied to each variable of the high-resolution dataset to generate low-resolution fields

Variable name	Standard deviation of the kernel [km]	Further reprocessing
SIT	11.0	None
SIC	4.0	Noise added*
Divergence	9.0	mult. factor 0.49
Shear	15.0	mult. factor 0.59

*See details in the text of section 2.3

2.3 Low-resolution model dataset

The algorithm proposed in this study, described in detail in Section 3.2, is supervised, meaning it is trained using pairs of low-resolution and high-resolution fields. The high-resolution fields are provided by the neXtSIM simulations (see Section 2.1).

Although the algorithm is ultimately applied to the satellite-derived dataset (Section 2.2), this cannot be used directly for training. This is because the discrepancies between satellite-derived and model-generated fields are not solely due to resolution differences. Since neXtSIM is largely free-running, except for the weak nudging to satellite-derived sea ice thickness, the model and observed fields may be temporally desynchronized.

To obtain matching pairs for training, we generated synthetic observations that replicate the characteristics of satellite-derived fields while remaining aligned with the high-resolution model outputs. These synthetic low-resolution fields were produced by smoothing the high-resolution neXtSIM data to mimic the spatial resolution and statistical properties of the satellite products.

Specifically, we applied a low-pass Gaussian filter to each high-resolution variable, adjusting the filter parameters so that the power spectral density (as defined in Section 3.4) of the smoothed field matched that of the corresponding satellite-derived field. In some cases, a multiplicative factor was also applied to match the total variance. Table 1 summarizes the preprocessing parameters used for each variable. Figure 1 illustrates one example from the validation dataset, showing the high-resolution model field, the satellite-derived field, and the preprocessed low-resolution model field. The similarity between the preprocessed model field and the satellite observations is visually apparent and quantitatively confirmed by the spectral comparison shown in the fourth column.

As noted in Table 1, additional processing was required for sea ice concentration (SIC). The satellite-derived SIC exhibited features that appeared to be measurement noise. To replicate this, we added a two-dimensional random noise with a predefined spectral shape to the low-pass filtered neXtSIM SIC. This predefined spectrum was computed as the difference between the average spectra of the observed and low-pass filtered SIC fields.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

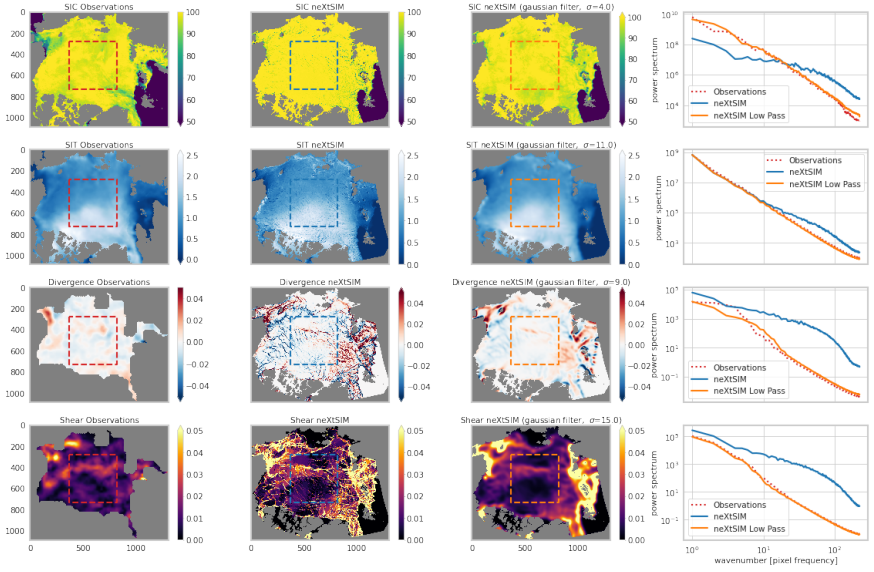

Figure 1. Sample from the satellite-derived dataset (first column), the high-resolution neXtSIM dataset (second column), and the low-resolution model dataset after smoothing (third column) for January 1st, 2021. The first row shows sea ice concentration, the second row sea ice thickness, the third row divergence, and the fourth row shear. The fourth column presents the power spectral density for each corresponding variable: satellite-derived data (dotted red), high-resolution neXtSIM simulation (blue), and low-resolution model data (orange). The dashed squares in each map show the region over which the power spectral density is calculated.

2.4 Prepare data for the machine learning algorithm

The low-resolution and high-resolution model datasets contain fields of sea ice concentration, thickness, divergence, and shear over the Arctic Ocean, spanning the freezing seasons (October 18 to April 15) from 2013 to 2023. All fields are defined on a 3 km spatial grid, with a total size of 1086×1308 grid points. The dataset is split into training, validation, and test subsets: the training set includes samples from 2013 to April 2020 (1157 samples), the validation set covers October 2020 to April 2022 (360 samples), and the test set spans October 2022 to April 2023 (180 samples).

Additional preprocessing is applied to the input features. First, the total deformation is computed and used as an input variable:

$$\psi = \sqrt{\nabla^2 + \tau^2}, \tag{3}$$

where ψ denotes the total deformation, ∇ is the divergence, and τ is the shear of the sea ice drift defined in Eqs 1 and 2.

All input variables are standardized by subtracting the mean and dividing by the standard deviation, both computed over the training dataset.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

Finally, a binary mask is added to the set of input features, with values of 1 over sea ice and 0 elsewhere (ice-free ocean or land). Note that this preprocessing is applied identically to both the model and satellite-derived datasets.

To summarize, the input features for the super-resolution model are sea ice concentration, thickness, total deformation, and the binary mask.

The high-resolution sea ice thickness field is defined on the same grid as the input features. The super-resolution algorithm is trained to reconstruct the sea ice thickness increment, defined as:

$$\mathrm{SIT}_{\mathrm{inc}} = \mathrm{SIT}_{\mathrm{hr}} - \mathrm{SIT}_{\mathrm{lr}}, \tag{4}$$

where $\mathrm{SIT}_{\mathrm{hr}}$ is the high-resolution sea ice thickness (see Section 2.1), and $\mathrm{SIT}_{\mathrm{lr}}$ is the low-resolution sea ice thickness obtained through the procedure described in Section 2.3. This formulation allows the algorithm to focus exclusively on reconstructing the high-resolution component of the signal, as the low-resolution baseline is removed from the target variable.

The increment field is standardized by subtracting its mean and dividing by its standard deviation, both computed over the training dataset. To recover the full high-resolution sea ice thickness from the model output, the predicted increment must first be de-standardized and then added to the low-resolution sea ice thickness field provided as input.

3 Machine-learning model

3.1 Conditional diffusion models

The objective of our algorithm is to generate samples from a conditional probability distribution $p_{\mathrm{data}}(\mathbf{x}_0|\mathbf{c})$, where $\mathbf{x}_0$ represents the field of high-resolution sea ice thickness increment ($\mathrm{SIT}_{\mathrm{inc}}$, see Eq. 4), and $\mathbf{c}$ denotes the low-resolution context available at inference time – namely, the low-resolution sea ice thickness, concentration, deformation, and binary sea ice mask (see Section 2.3).

We employ the Denoising Diffusion Implicit Model (DDIM), which we have adapted to generate conditional samples. The core principles of the algorithm are outlined below; for theoretical foundations, we refer the reader to Song et al. (2020).

The DDIM algorithm first considers a prior sample $\mathbf{x}_1$ of the same dimensionality as $\mathbf{x}_0$, drawn from a standard normal distribution:

$$p(\mathbf{x}_1) := \mathcal{N}(\mathbf{x}_1; \mathbf{0}, \mathbf{I}), \tag{5}$$

where $\mathbf{I}$ is the identity matrix and $\mathcal{N}$ the normal distribution. The goal is to generate a sequence of samples $\mathbf{x}_t$ indexed by a continuous time variable $t \in [0,1]$, such that the final sample $\mathbf{x}_0$ approximates the conditional data distribution $p_{\mathrm{data}}(\mathbf{x}_0|\mathbf{c})$. Hereafter, we refer to t as the diffusion time.

The diffusion process is defined as:

$$\mathbf{x}_t(\mathbf{x}_0, \boldsymbol{\epsilon}) = \sqrt{\alpha(t)}\mathbf{x}_0 + \sqrt{1-\alpha(t)}\boldsymbol{\epsilon}, \quad \text{with } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{6}$$

where $\alpha(t)$ is a variance schedule satisfying $\alpha(0) \approx 1$ and $\alpha(1) \approx 0$, and $\boldsymbol{\epsilon}$ is a noise sample following the prior distribution.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

For each training sample $(\mathbf{x}_0^k, \mathbf{c}^k)$, a noise realization $\boldsymbol{\epsilon}^k$ and a diffusion time t^k are drawn to generate the noisy input $\mathbf{x}_t^k = \mathbf{x}_t(\mathbf{x}_0^k, \boldsymbol{\epsilon}^k)$. A neural network $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}$ is trained to predict the noise $\boldsymbol{\epsilon}^k$ from $\mathbf{x}_t^k$, t^k, and the conditioning features $\mathbf{c}^k$.

The model is trained by minimizing the following loss over a batch of samples:

$$l(t^k, \mathbf{x}_0^k, \boldsymbol{\epsilon}^k, \mathbf{c}^k, \boldsymbol{\theta}) = \left\| \boldsymbol{\epsilon}^k - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{x}_t^k, \mathbf{c}^k, t^k) \right\|^2. \tag{7}$$

The training procedure is detailed in Algorithm 1

Finally, for the inference, we apply an iterative sampling procedure. Given a known context $\mathbf{c}$, we first generate a noise sample $\hat{\mathbf{x}}_1$ and define the number of diffusion steps N_d, corresponding to a time increment $dt = 1/N_\mathrm{d}$. At each step from $t = 1$ to 0, the following quantities are computed:

$$\hat{\boldsymbol{\epsilon}}_t = \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\hat{\mathbf{x}}_t, \mathbf{c}, t), \tag{8}$$

$$\hat{\mathbf{x}}_{0,t} = \frac{\hat{\mathbf{x}}_t - \sqrt{1-\alpha(t)}\hat{\boldsymbol{\epsilon}}_t}{\sqrt{\alpha(t)}}, \tag{9}$$

$$\hat{\mathbf{x}}_{t-dt} = \sqrt{\alpha(t-dt)}\hat{\mathbf{x}}_{0,t} + \sqrt{1-\alpha(t-dt)}\hat{\boldsymbol{\epsilon}}_t, \tag{10}$$

where $\hat{\boldsymbol{\epsilon}}_t$ is the estimated noise and $\hat{\mathbf{x}}_{0,t}$ is the estimated high-resolution field at diffusion time t. The final output of the super-resolution algorithm is $\hat{\mathbf{x}}_{0,t=0}$. The inference procedure is detailed in Algorithm 2

Algorithm 1 Training procedure for conditional diffusion model

Require: Training dataset $\{(\mathbf{x}_0^k, \mathbf{c}^k)\}_{k=1}^N$ shuffled
Require: Noise schedule $\alpha(t)$, Batch size B
for b from the first to the last batch of the training set **do**
  Set batch loss $L \leftarrow 0$
  for $k = 1$ to B **do**
    Sample $(\mathbf{x}_0^k, \mathbf{c}^k)$ from the batch
    Sample $t^k \sim \mathcal{U}(0, 1)$
    Sample $\boldsymbol{\epsilon}^k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
    Compute noisy input: $\mathbf{x}_t^k = \sqrt{\alpha(t^k)}\mathbf{x}_0^k + \sqrt{1-\alpha(t^k)}\boldsymbol{\epsilon}^k$
    Predict noise: $\hat{\boldsymbol{\epsilon}}^k = \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{x}_t^k, \mathbf{c}^k, t^k)$
    Compute sample loss: $l^k = \left\| \boldsymbol{\epsilon}^k - \hat{\boldsymbol{\epsilon}}^k \right\|^2$
    Update batch loss $L \leftarrow L + l^k$
  end for
  Update model parameters $\boldsymbol{\theta}$ using gradient descent on L
end for

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

Algorithm 2 Inference procedure for conditional diffusion model

Require: Context features $\mathbf{c}$ (low-resolution SIT, SIC, deformation, mask)
Require: Trained model $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}$, noise schedule $\alpha(t)$
Require: Number of diffusion steps N_d, time step $dt = 1/N_\mathrm{d}$
Sample initial noise $\hat{\mathbf{x}}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
Set initial noisy field $\hat{\mathbf{x}}_t \leftarrow \hat{\mathbf{x}}_1$
for $i = N_\mathrm{d}$ down to 1 **do**
  $t \leftarrow i \cdot dt$
  Estimate noise: $\hat{\boldsymbol{\epsilon}}_t = \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\hat{\mathbf{x}}_t, \mathbf{c}, t)$
  Estimate high-resolution field: $\hat{\mathbf{x}}_{0,t} = \frac{\hat{\mathbf{x}}_t - \sqrt{1-\alpha(t)}\hat{\boldsymbol{\epsilon}}_t}{\sqrt{\alpha(t)}}$
  Compute next step: $\hat{\mathbf{x}}_{t-dt} = \sqrt{\alpha(t-dt)}\hat{\mathbf{x}}_{0,t} + \sqrt{1-\alpha(t-dt)}\hat{\boldsymbol{\epsilon}}_t$
  Update: $\hat{\mathbf{x}}_t \leftarrow \hat{\mathbf{x}}_{t-dt}$
end for
return Final estimate $\hat{\mathbf{x}}_{0,t=0}$ (standardized SIT increment)

3.2 Application of conditional diffusion model to sea ice thickness forecast

The model $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}$ used is a U-Net architecture (Ronneberger et al., 2015), where the input features are stacked together: the noisy input $\mathbf{x}_t$, the context $\mathbf{c}$, and the diffusion time t. The scalar parameter t is embedded using a positional encoding scheme, which maps it into two-dimensional features of the same spatial dimensions as $\mathbf{x}_t$ (Vaswani et al., 2017). The architecture consists of four residual blocks of convolutional layers for both the encoder and decoder (Siddique et al., 2020). After hyperparameter tuning (see Appendix A), the number of units per block was set to 32, 64, 96, and 128, respectively.

The noise schedule is defined as:

$$\alpha(t) = \cos^2\left(\gamma(t)\right), \quad \text{with } \gamma(t) = \gamma_{\min} + t \cdot (\gamma_{\max} - \gamma_{\min}), \tag{11}$$

where $\gamma_{\min}$ and $\gamma_{\max}$ are chosen such that $\alpha(0) = 0.95^2$ and $\alpha(1) = 0.02^2$, based on sensitivity experiments (Appendix A).

The model is trained for 250 epochs using the Adam optimizer with a learning rate of 10^{-3} and a batch size of $B = 4$, on two AMD MI250x GPUs on the LUMI supercomputer (https://lumi-supercomputer.eu).

During training, we maintain an exponential moving average (EMA) of the model parameters to stabilize inference. At each update step (see Algorithm 1), the EMA parameters $\tilde{\boldsymbol{\theta}}$ are updated as:

$$\tilde{\boldsymbol{\theta}} \leftarrow 0.999\tilde{\boldsymbol{\theta}} + 0.001\boldsymbol{\theta}, \tag{12}$$

where $\boldsymbol{\theta}$ are the current model parameters.

For inference (Algorithm 2), the EMA model is used with $N_d = 60$ diffusion steps. For each low-resolution context $\mathbf{c}$, an ensemble of 30 high-resolution realizations of $\mathbf{x}_0$ is generated. The sensitivity to N_d is discussed in Section 6.1. Each realization is computed on a single GPU and takes approximately 6.3 seconds.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

The model, training, and inference procedures were implemented using the TensorFlow/Keras framework. The code is adapted from the official DDIM example (https://keras.io/examples/generative/ddim/, version 2022/06/24), with modifications to support conditional sampling.

3.3 Baseline

As the objective of this study is to enhance the representation of small-scale features in sea ice thickness from low-resolution input fields, we compare the results of our algorithm against a simple baseline: the low-resolution sea ice thickness itself, either derived from the model (see Section 2.3) or from satellite observations (see Section 2.2).

It is important to note that both high-resolution and low-resolution fields are interpolated onto the same grid, ensuring that the grid resolution is identical across datasets. Therefore, the comparison focuses solely on the effective resolution of the fields, rather than differences in grid spacing.

3.4 Metrics

In the following, we present the metrics used to compare the ground truth sea ice thickness $\mathbf{x}$ with the estimated sea ice thickness $\hat{\mathbf{x}}$. These variables are time series of 2D maps indexed by the time index k and the spatial pixel index i.

3.4.1 Accuracy

The root mean square error (RMSE) is used to assess the accuracy of the reconstructed field:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i,k}\left(\mathbf{x}_{i,k} - \hat{\mathbf{x}}_{i,k}\right)^2}, \tag{13}$$

where the sum can be computed over the spatial index i to produce a time series, or over the time index k to produce an error map. When available, $\hat{\mathbf{x}}_{i,k}$ is computed as the average over an ensemble of realizations.

3.4.2 Realism

Beyond visual inspection, we define several metrics to quantify the realism of the reconstruction.

The first metric is the 2D power spectral density (PSD). It is computed as the squared magnitude of the 2D Fourier transform of a sea ice thickness field, denoted $P(q_x, q_y)$, where (q_x, q_y) are the components of the wave number vector. The PSD is radially averaged over bins of similar wave number magnitude $q = \sqrt{q_x^2 + q_y^2}$. To avoid artifacts due to land presence, the PSD is computed over a land-free sub-region of size 400×400 grid points. The maximum resolved scale using this metric is thus 1200 km ($= 400 \times 3$, given the 3 km spatial resolution).

The second metric is the spatial variability, defined as the spatial standard deviation of sea ice thickness for a given day.

The third realism metric is the fine-scale bias, defined as:

$$\mathrm{Fine-scale\ bias} = 1 - \frac{1}{N_{q<S}}\sum_{q<S}\frac{\mathrm{PSD}^{\mathrm{rec}}(q)}{\mathrm{PSD}^{\mathrm{ref}}(q)}, \tag{14}$$

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

where $\mathrm{PSD}^{\mathrm{ref}}(q)$ is the PSD of the reference high-resolution sea ice thickness, and $\mathrm{PSD}^{\mathrm{rec}}(q)$ is the PSD of the reconstructed field, either from the baseline or the diffusion model. $N_{q<S}$ is the number of PSD points corresponding to small spatial scales, defined here as scales below $S = 20$ km.

The fine-scale bias quantifies the proportion of small-scale variability missing in the reconstruction. A value of 1 indicates that 100% of the small-scale features are *not* recovered, while a value of 0 indicates perfect reconstruction of all small scales. Negative values imply that the reconstruction contains more small-scale variability than the reference high-resolution field.

3.4.3 Reliability

The diffusion-based reconstruction generates an ensemble of 30 realizations. To assess the reliability of the uncertainty represented by this ensemble, we compare the RMSE (as defined above) with the ensemble spread, computed as the standard deviation at each pixel and day across ensemble members.

3.4.4 Dedicated metrics for validation over observational data

There are not enough independent ground-truth measurements to directly validate our ML-based super-resolution approach when applied to real CS2SMOS observations. Because the method generates an ensemble of plausible high-resolution fields rather than a single deterministic estimate, a point-wise matchup with observations is neither expected nor meaningful. Validation must instead focus on assessing the realism of the reconstructed spatial scales.

For this purpose, we use along-track L2P Cryosat-2 data from Sea Ice Thickness from Satellite Radar Altimetry (SIRAL, Hendricks and Paul, 2023). Although these data are not fully independent (CryoSat-2 is one of the inputs to the CS2SMOS product used to drive our ML model) CS2SMOS primarily exploits large-scale information from CryoSat-2. The small-scale variability is effectively filtered out in the CS2SMOS processing chain (Ricker et al., 2017). Consequently, it remains valid to use the small-scale content of CryoSat-2 tracks to evaluate the realism of our high-resolution ML product.

To identify which spatial scales are correctly reproduced, we compute one-dimensional power spectra averaged over all CryoSat-2 along-track sea-ice thickness measurements. The CryoSat-2 data are smoothed using a range of window sizes, from high-resolution (300 m) to strongly smoothed (360 km), allowing a comparison over different scales between observations, the low-resolution baseline, and the ML super-resolved fields.

4 Results on model data

The first experiment evaluates the reconstruction of high-resolution SIT fields obtained from low-resolution inputs on the test dataset. In this setup, the low-resolution fields are obtained by smoothing the corresponding high-resolution fields, as described in section 2.3. Consequently, the original high-resolution simulations serve as ground truth, enabling us to assess the performance of our algorithm using several validation metrics described in section 3.4.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

4.1 One Example

In Figure 2, we display examples of sea ice fields at three different dates from the validation set, corresponding to the beginning, middle, and end of the freezing season.

Visually, the images generated by the ML algorithm appear realistic. In particular, individual samples from the diffusion model are nearly indistinguishable from the true high-resolution fields. When averaged over the ensemble of 30 members, small-scale features are smoothed out as expected, but high-resolution patterns remain visible while they are absent in the low-resolution baseline. This suggests that certain small-scale structures are predictable from low-resolution inputs.

4.2 Deterministic Scores

In Figure 3, we present the root mean square error (RMSE) computed both spatially and temporally over the test dataset. The results show that, at nearly every date and location, the high-resolution reconstruction outperforms the low-resolution baseline.

Note that the ML-generated sea ice thickness used for this comparison is the ensemble mean over 30 realizations, i.e., the same type of image shown as "ML(mean)" in Figure 2. This supports the conclusion from the previous section: robust small-scale features can be inferred deterministically from low-resolution inputs, and the resulting reconstructions are more accurate than the baseline.

We observe that RMSE increases over time during the freezing season. This is primarily a size effect: the spatial extent of sea ice grows later in the season, increasing the area over which errors can accumulate. Additionally, as expected, errors are more pronounced near coastlines and in the Fram Strait. This suggests that recovering deterministic small-scale features is more challenging in these regions. Our primary hypothesis is that sea ice is more prone to small-scale sporadic deformations, including break-up and ridging in these areas, making the localization of small-scale features more uncertain.

4.3 Realism

Following the qualitative assessment of realism through visual inspection in Figure 2, we now quantify the realism of the reconstructed sea ice thickness using the three metrics defined in Section 3.4.2. These metrics focus on the reconstruction of small-scale features, which are the missing components we aim to recover from low-resolution inputs.

The results are presented in Figure 4. The power spectral density (PSD) is computed for three selected days within the test period. The results show variability in the quality of spectrum reconstruction. For example, the reconstruction on 11 November closely matches the true spectrum, while small-scale features are slightly underestimated on 1 January. In all cases, the ML-generated reconstructions are closer to the true high-resolution spectrum than the low-resolution baseline.

The spatial variability metric is consistently better captured by the ML model compared to the baseline. A slight overestimation of spatial variability is observed, which may be attributed to residual uncorrelated noise in the reconstruction.

Finally, the fine-scale bias confirms that the ML model recovers more small-scale variability than the low-resolution baseline. On average, the ML reconstructions slightly underrepresent small-scale features, but this varies across the test period. On some

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

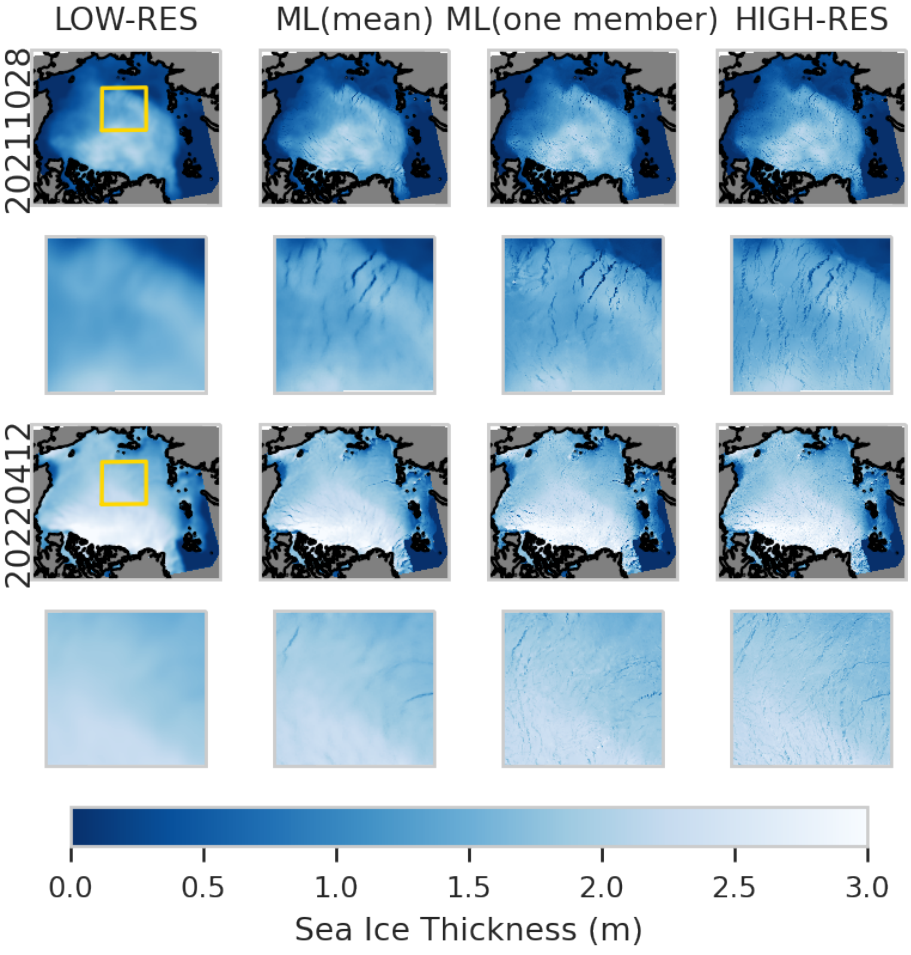

Figure 2. Examples of samples from the validation dataset for two dates chosen for illustration, 28th October 2021 (20211028) and 12 April 2022 (20220412). The column 1 shows the low-resolution product CS2SMOS, the column 2 shows the average over the ensemble of 30 ML predictions, the column 3 shows the prediction of one arbitrary member, and the column 4 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.

275 days, the ML model even reconstructs more small-scale variability than the true field (bias < 0). In contrast, the low-resolution baseline consistently underrepresents small-scale variability throughout the entire test period.

4.4 Probabilistic Scores

An innovative feature of our super-resolution algorithm is its ability to produce an ensemble of high-resolution simulations. This ensemble allows us to assess the uncertainty of the reconstructed sea ice thickness through the ensemble spread. A larger
280 spread indicates higher uncertainty in the reconstruction.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

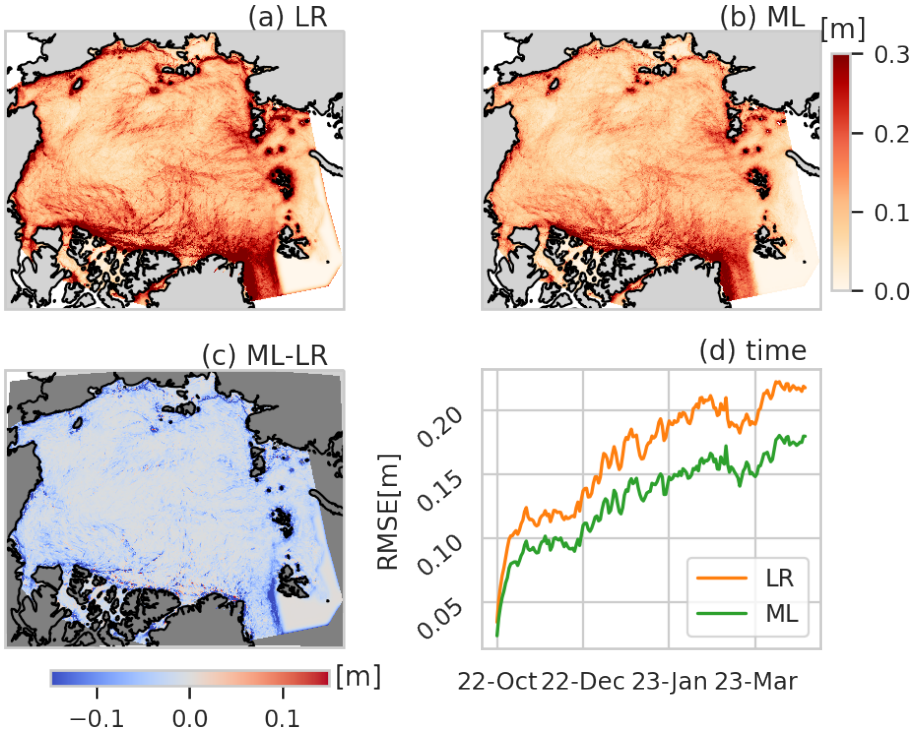

Figure 3. Root Mean Square Error (RMSE, m) computed on the test dataset from 10 October 2022 to 15 April 2023. The upper panel shows the spatial RMSE for (a) the low-resolution (LR) baseline and (b) the ML retrieval, averaged over 30 ensemble members. Panel (c) displays the difference between the ML and LR RMSE, where blue indicates regions where the ML retrieval outperforms the baseline. The temporal evolution of the RMSE is presented in panel (d).

To evaluate the quality of the uncertainty estimates, we compare the ensemble spread to the RMSE. Ideally, the spread should be of the same order of magnitude as the RMSE, indicating a well-calibrated uncertainty estimate.

Figure 5 presents both the spread and RMSE across different categories of sea ice thickness, ranging from thin ice (≤ 0.5 m) to thick ice (> 3.5 m). The results show that the uncertainty is consistently underestimated across all categories, suggesting
285 that the ML model is overconfident in its reconstructions.

Nevertheless, the relative confidence between categories is well captured. For example, in the 1.5–2.5 m thickness range, where the model shows lower spread, the RMSE is also smallest. This indicates that while the absolute uncertainty may be underestimated, the ensemble spread can still be used as a relative indicator of confidence—helping to identify regions where the reconstruction is more or less reliable.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

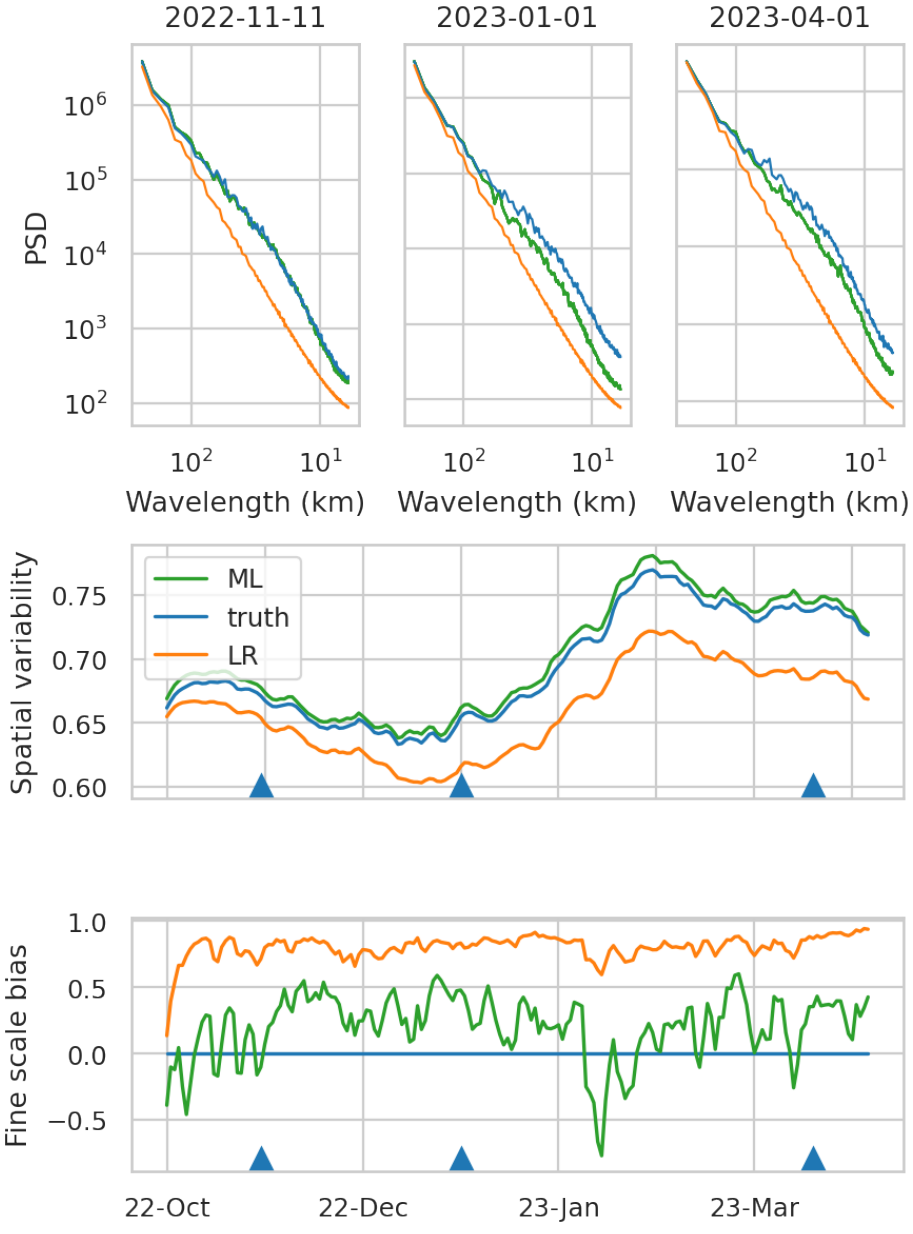

Figure 4. The upper panel shows the Power Density Spectrum (PSD) for three selected dates representing the beginning, middle, and end of the test dataset. The middle panel illustrates the spatial variability across the entire test period, while the bottom panel presents the fine-scale bias relative to the true high-resolution field. Metrics are defined in section 3.4. In all panels, the blue line represents the true high-resolution field, the orange line denotes the low-resolution baseline, 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.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

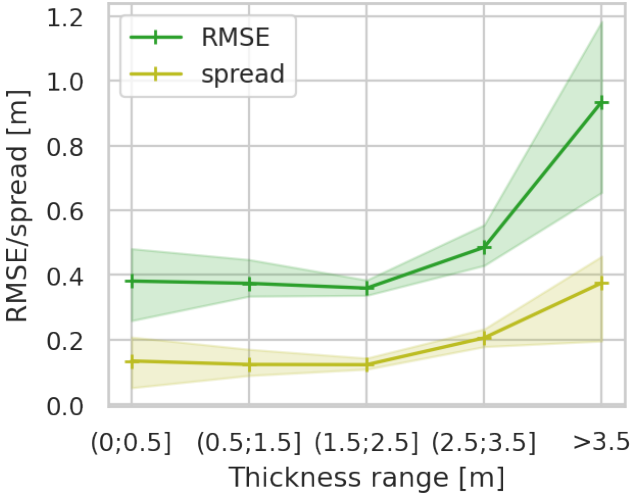

Figure 5. Root-mean-square error (RMSE; dark green) and ensemble-mean spread (light green) computed over the test dataset across sea-ice categories. Both RMSE and spread are spatially averaged. The solid line denotes the temporal mean over all test days, while the shaded area represents the 10th to 90th percentile range across the test period.

5 Results on the observational dataset

Figure 6 illustrates our super-resolution results for two arbitrarily selected dates. A zoomed region is shown for clarity. As in the experiments using model data, the ML algorithm reconstructs spatially coherent small-scale features that are absent from the low-resolution input. Some unrealistic patterns occasionally appear, particularly near coastlines. These artefacts stem from masked regions in the deformation fields from OSI SAF (see Section 2.2). Because the training data did not contain such masks, the model must extrapolate in these regions, leading to spurious features. Additionally, despite careful preprocessing of the high-resolution model data to mimic satellite observations (Section 2.3), certain patterns present in CS2SMOS do not appear in the training dataset, and thus may be imperfectly super-resolved. Nevertheless, the ML-generated high-resolution fields remain qualitatively robust when applied to previously unseen observational data.

To quantify the realism of the reconstructed fields, we compare their power spectrum with that derived from CryoSat-2 observations for January 2022, following the methodology in Section 3.4.4. The results, shown in Figure 7, confirm that the ML product contains substantially more small-scale energy than the low-resolution baseline. The magnitude of this gain is consistent with that obtained from the model-based realism test (Fig. 4). At large scales, the ML product and the low-resolution baseline show similar spectral behaviour, indicating that the ensemble reconstruction remains faithful to the large-scale structure of the input field.

At finer scales, the ML product exhibits power comparable to the CryoSat-2 spectrum filtered at 24 km, whereas the low-resolution baseline corresponds more closely to the 360 km filtered CryoSat-2 spectrum. We therefore conclude that the

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

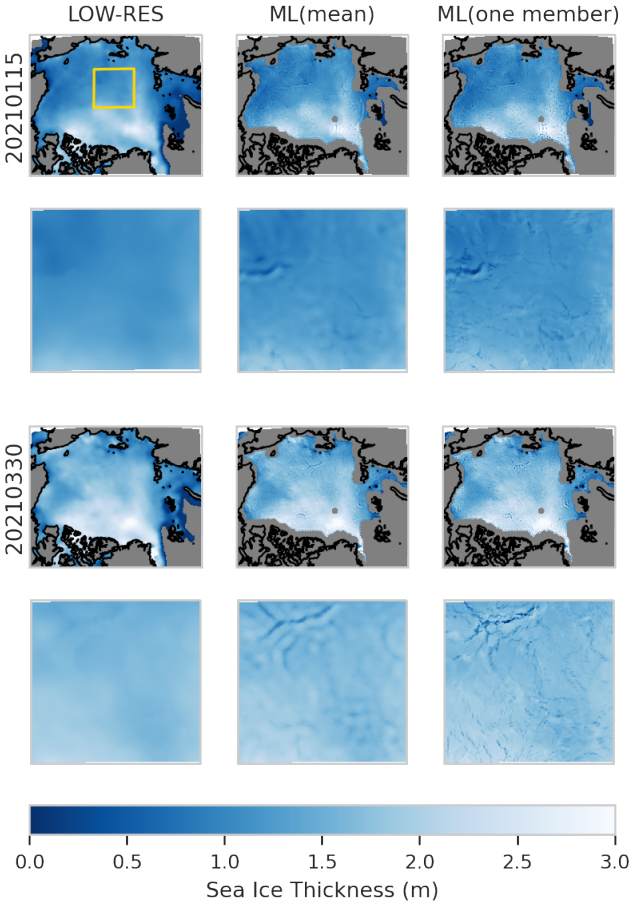

Figure 6. Examples of samples from observational data for two dates chosen for illustration, 15th January 2021 and 30 March 2021. Column 1 shows the low-resolution product CS2SMOS, column 2 shows the average over the ensemble of 30 ML predictions, and column 3 shows the prediction of one arbitrary member. 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 first row.

ML super-resolved product captures realistic variability down to scales of approximately 24 km. At scales finer than this threshold, part of the CryoSat-2 variability is still missing, in agreement with the behaviour seen in Fig. 4.

6 Discussion

6.1 Sensitivity to the number of diffusion steps

It was shown by Song et al. (2020) that the number of diffusion steps used during inference is critical for the quality of the generated samples. In general, there is a trade-off between reconstruction quality and computational cost, which scales linearly

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

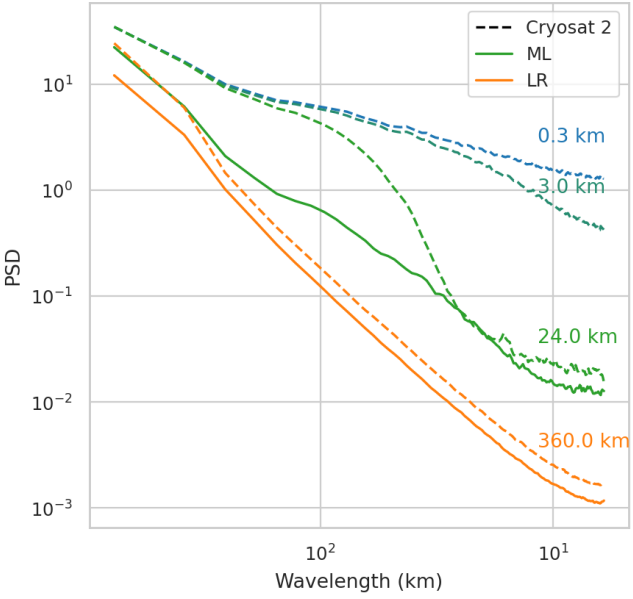

Figure 7. Power spectra computed for January 2022 from observed sea-ice thickness derived from CryoSat-2 tracks (dashed lines), averaged over scales from 0.3 km to 360 km. The corresponding power spectra for the ML reconstruction (ML, green solid line) and the low-resolution baseline (LR, orange solid line) are extracted along the same tracks.

with the number of diffusion steps. In our case, each ensemble member takes approximately 6.3 seconds to generate on a single GPU with 60 diffusion steps.

Figure 8 presents two metrics, RMSE and fine-scale bias, as well as the computing cost for generating one high-resolution sample as functions of the number of diffusion steps, ranging from 10 to 500, computed for January 1st, 2021. To limit computational cost, we did not compute these metrics over the entire validation dataset. However, tests indicate that these metrics are stable over time (see, e.g., Figure 3).

The results show a modest improvement in both RMSE and fine-scale bias with increasing diffusion steps. RMSE decreases slightly, and the fine-scale bias approaches 0, indicating better recovery of small-scale features. Based on these results and practical considerations, we chose 60 diffusion steps as a good compromise between computational cost and reconstruction quality for generating results and sharing the dataset (see the data availability section).

6.2 Split deterministic/probabilistic part

The problem of super-resolving sea ice thickness is fundamentally under-constrained: small-scale features cannot be fully inferred from large-scale observations. The portion of small-scale variability that can be predicted deterministically corresponds to the deterministic component, while the remainder is inherently probabilistic. Our diffusion model is designed to reconstruct both the deterministic part and a specific realisation of the probabilistic component, thereby aiming to produce a realistic high-resolution field. It is well established that realistic high-resolution reconstructions are subject to the double penalty effect,

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

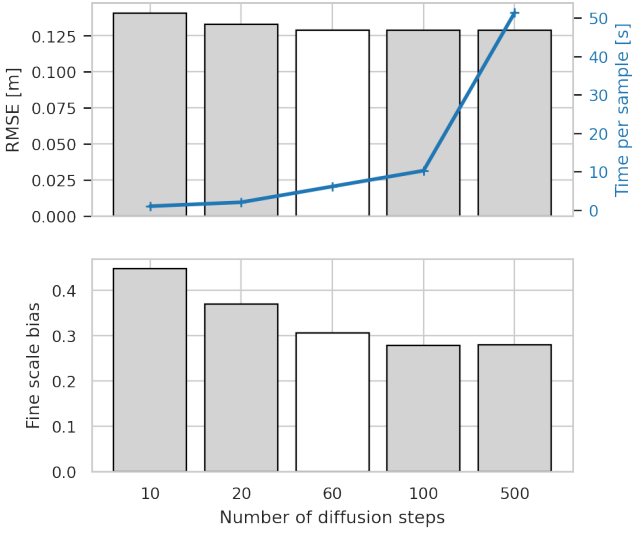

Figure 8. RMSE (upper panel, bar plot), time to generate one sample (upper panel, line plot) and fine-scale bias (lower panel) as a function of the number of diffusion steps in the sampling process. Metrics are computed for January 1st, 2021. The white bar indicates the value used to produce the results of this article.

which leads to higher errors compared to lower-resolution fields. This motivates the use of distinct metrics to evaluate accuracy (associated with the deterministic part) and realism (associated with the probabilistic part).

To obtain more accurate, but less realistic, fields, one can average multiple realisations of the diffusion model. This approach is illustrated in Fig. 9, where we show how accuracy, measured by RMSE, and realism, measured by fine-scale bias, evolve as a function of the number of realisations averaged. Notably, the most realistic reconstructions are obtained using a single realisation, but the RMSE in this case exceeds that of the low-resolution baseline due to the double penalty effect. Conversely, the most accurate reconstruction is achieved by averaging over 30 members, although the fine-scale features below 20 km are then reconstructed no better than in the low-resolution baseline.

Importantly, there is no universally optimal ensemble size; it depends on the intended application of the high-resolution fields. If the goal is to minimize average error, the ensemble mean provides the best estimate. If the objective is to capture realistic small-scale variability, individual realizations are preferable. This last choice is particularly relevant for estimating statistics that depend nonlinearly on sea ice thickness, such as surface heat fluxes.

7 Conclusions

We have developed an algorithm to enhance the resolution of Arctic sea ice thickness fields. The method is based on a generative artificial intelligence model known as a diffusion model. Once trained, the model can generate an ensemble of high-resolution sea ice thickness fields from low-resolution observations.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

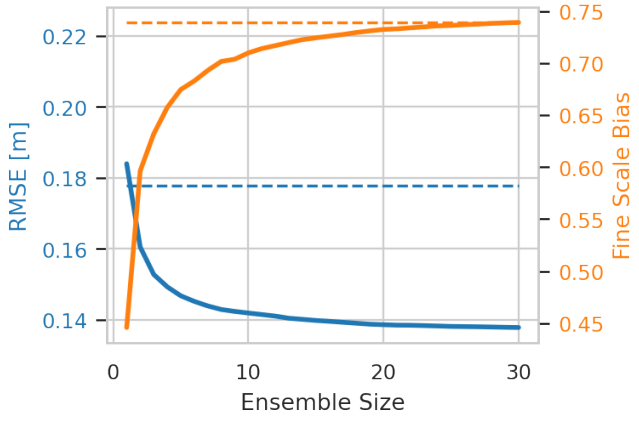

Figure 9. RMSE (blue) and fine-scale bias (orange) computed on the reconstruction averaged over n ensemble members, as a function of n from 1 to 30. Dashed lines indicate the corresponding values of the low-resolution baseline for each metric, using the same color coding.

We demonstrate that the algorithm produces realistic reconstructions, with an average accuracy improvement of 21% compared to a low-resolution baseline. Moreover, the ensemble output enables the estimation of uncertainty, which, while slightly under-dispersive, remains realistic. On average, the reconstructed fields capture 80% of the sea ice thickness variability at scales below 20 km, compared to only 20% for the low-resolution baseline. When applied to real observations, the method is able to reconstruct small-scale features that are otherwise unresolved, while remaining consistent with the large-scale structure of the observational products.

The algorithm was applied to both simulated data and real observations of Arctic sea ice between 2020 and 2023. The resulting high-resolution fields are publicly available to the community (see the Data Availability section).

Beyond sea ice thickness, we show that the diffusion model can generate realistic geophysical fields conditioned on low-resolution inputs. In principle, this approach is applicable to any domain where paired low-resolution and high-resolution fields are available, and will lead to the same trade off between realism and accuracy. In this study, we prioritised robustness over flexibility: the model is designed to upscale fields from a fixed low-resolution to a fixed high-resolution. However, future developments could lead to more versatile algorithms. Improvements in inference efficiency are also possible, for instance by exploring alternative generative approaches such as conditional flow matching (Lipman et al., 2022).

Code availability. The code for the super-resolution algorithm is available at: https://github.com/nansencenter/SID-NN/tree/stoch. It is adapted from the Keras tutorial: https://keras.io/examples/generative/ddim/. A tutorial on how to run the code will be provided online upon acceptance of the article.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

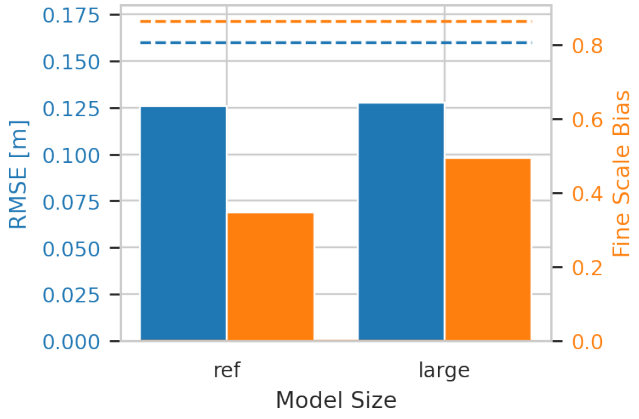

Figure A1. Sensitivity of the model to its size. "Ref" denotes the architecture used in the main model presented in the article (32, 64, 96, and 128 units in the four blocks), while "Large" corresponds to an increased-capacity model (64, 64, 96, and 128 units in the four blocks). Metrics are computed on the validation set. Blue bars indicate the RMSE (m), and orange bars indicate the absolute fine-scale bias. Dashed lines indicate the metrics of the low-resolution dataset (blue: RMSE; orange: fine-scale bias). For both metrics, lower values indicate better performance, with an optimal value of zero.

Data availability. The datasets used for training and the outputs of the machine-learning algorithm are available at: https://github.com/nansencenter/superice_data. Instructions for downloading and using the data are provided in the repository.

Appendix A: Sensitivity to hyperparameters

We selected a set of hyperparameters that we expected to have the strongest influence on the results: the size of the neural network, and the minimum and maximum signal rates used during training, as defined in Eq. 11. We also assessed the variability associated with the random initialization of the model parameters by training three models with identical configurations. All results were computed on the validation dataset and are summarized in the figures below. Figure A1 shows the sensitivity to model size, Figure A2 illustrates the impact of parameter initialization, and Figures A3 and A4 present the sensitivities to the minimum and maximum signal rates, respectively. Overall, these experiments reveal a weak sensitivity in terms of RMSE, although consistent trends emerge that help identify the optimal configuration used in this study. In contrast, the fine-scale bias exhibits stronger and less stable sensitivity, making it more difficult to identify a clear trend. This is particularly evident in Fig. A2, where identical configurations can produce either an overestimation or an underestimation of small-scale variability. Our final strategy was therefore to select the model that achieved the best performance on both the RMSE and the fine-scale bias metrics computed on the validation dataset.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

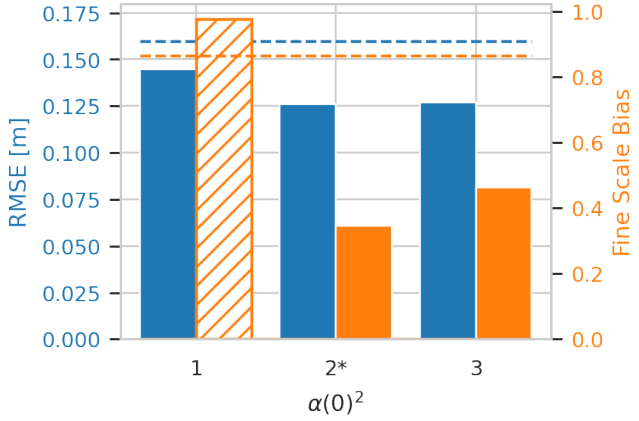

Figure A2. Sensitivity of the model to different parameter initialisations. For each panel, the model was retrained using a different random initialisation. The architecture used for the main model presented in the article is marked with an asterisk ("*") on the x-axis labels. Metrics are computed on the validation set. Blue bars indicate the RMSE (m), and orange bars indicate the absolute fine-scale bias. When the fine-scale bias is negative, the corresponding bar is hatched. Dashed lines indicate the metrics of the low-resolution dataset (blue: RMSE; orange: fine-scale bias). For both metrics, lower values indicate better performance, with an optimal value of zero.

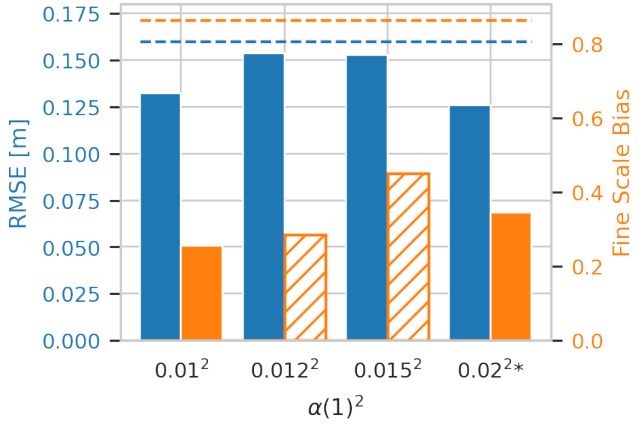

Figure A3. Sensitivity of the model to different minimal signal rate during training $\alpha(1)$ defined in Eq. 11. The value used for the main model presented in the article, $\alpha(1) = 0.02^2$, is marked with an asterisk ("*") on the x-axis labels. Metrics are computed on the validation set. Blue bars indicate the RMSE (m), and orange bars indicate the absolute fine-scale bias. When the fine-scale bias is negative, the corresponding bar is hatched. Dashed lines indicate the metrics of the low-resolution dataset (blue: RMSE; orange: fine-scale bias). For both metrics, lower values indicate better performance, with an optimal value of zero.

Author contributions. JB designed the study and developed the model, ran experiments and supervised the project. JB, AK and FB processed the data and prepared the figures. JB, AK and FB wrote the manuscript. RD and YW provided expertise for this work and edited the manuscript.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

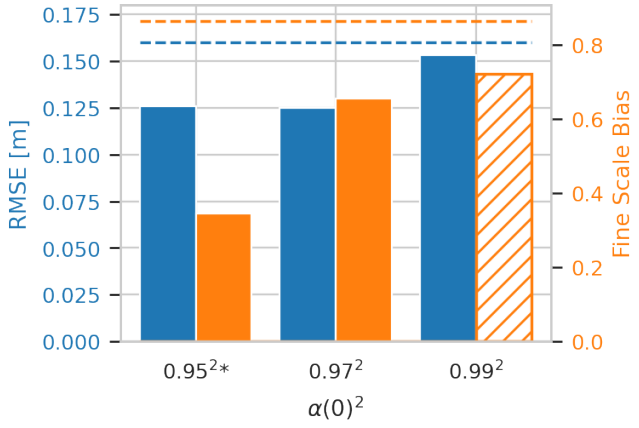

Figure A4. Sensitivity of the model to different minimal signal rate during training $\alpha(0)$ defined in Eq. 11. The value used for the main model presented in the article, $\alpha(0) = 0.95^2$, is marked with an asterisk ("*") on the x-axis labels. Metrics are computed on the validation set. Blue bars indicate the RMSE (m), and orange bars indicate the absolute fine-scale bias. When the fine-scale bias is negative, the corresponding bar is hatched. Dashed lines indicate the metrics of the low-resolution dataset (blue: RMSE; orange: fine-scale bias). For both metrics, lower values indicate better performance, with an optimal value of zero.

Competing interests. The authors declare no competing interests

Acknowledgements. This research has been supported by the European Space Agency (SuperIce project) under the Contract No. 4000142335/231/I-DT and by the Norges Forskningsråd (grant no. 328886 and 325241).

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

References

Aich, M., Hess, P., Pan, B., Bathiany, S., Huang, Y., and Boers, N.: Conditional diffusion models for downscaling & bias correction of Earth system model precipitation, EGUsphere, 2025, 1–21, 2025.

Duffey, A., Mallett, R., Dutch, V. R., Steckling, J., Hermant, A., Day, J., and Pithan, F.: Representation of Arctic winter atmospheric boundary layer stability over sea ice in CMIP6 models, Journal of Geophysical Research: Atmospheres, 130, e2024JD041 412, 2025.

Foster, D.: Generative deep learning, " O'Reilly Media, Inc.", 2022.

Groenke, B., Madaus, L., and Monteleoni, C.: Climalign: Unsupervised statistical downscaling of climate variables via normalizing flows, in: Proceedings of the 10th International Conference on Climate Informatics, pp. 60–66, 2020.

Hendricks, S. and Paul, S.: Product User Guide & Algorithm Specification: AWI CryoSat-2 Sea Ice Thickness (version 2.6), https://doi.org/10.5281/zenodo.10044554, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly journal of the royal meteorological society, 146, 1999–2049, 2020.

Kimmritz, M., Counillon, F., Smedsrud, L. H., Bethke, I., Keenlyside, N., Ogawa, F., and Wang, Y.: Impact of ocean and sea ice initialisation on seasonal prediction skill in the Arctic, Journal of Advances in Modeling Earth Systems, 11, 4147–4166, 2019.

Komuro, Y. and Suzuki, T.: Impact of subgrid-scale ice thickness distribution on heat flux on and through sea ice, Ocean Modelling, 71, 13–25, 2013.

Korosov, A., Rampal, P., Ying, Y., Ólason, E., and Williams, T.: Towards improving short-term sea ice predictability using deformation observations, The Cryosphere, 17, 4223–4240, https://doi.org/10.5194/tc-17-4223-2023, 2023.

Kwok, R.: Contrasts in sea ice deformation and production in the Arctic seasonal and perennial ice zones, Journal of Geophysical Research: Oceans, 111, https://doi.org/https://doi.org/10.1029/2005JC003246, 2006.

Lavergne, T., Eastwood, S., Teffah, Z., Schyberg, H., and Breivik, L.-A.: Sea ice motion from low-resolution satellite sensors: An alternative method and its validation in the Arctic, Journal of Geophysical Research: Oceans, 115, https://doi.org/10.1029/2009jc005958, 2010.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M. A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., and Pedersen, L. T.: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, The Cryosphere, 13, 49–78, https://doi.org/10.5194/tc-13-49-2019, 2019.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681–4690, 2017.

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M.: Flow matching for generative modeling, arXiv preprint arXiv:2210.02747, 2022.

Liu, Z.-S., Siu, W.-C., and Chan, Y.-L.: Photo-realistic image super-resolution via variational autoencoders, IEEE Transactions on Circuits and Systems for video Technology, 31, 1351–1365, 2020.

Lockwood, J. W., Gori, A., and Gentine, P.: A generative super-resolution model for enhancing tropical cyclone wind field intensity and resolution, Journal of Geophysical Research: Machine Learning and Computation, 1, e2024JH000 375, 2024.

Mchedlishvili, A., Lüpkes, C., Petty, A., Tsamados, M., and Spreen, G.: New estimates of pan-Arctic sea ice–atmosphere neutral drag coefficients from ICESat-2 elevation data, The Cryosphere, 17, 4103–4131, https://doi.org/10.5194/tc-17-4103-2023, 2023.

https://doi.org/10.5194/egusphere-2026-2318
Preprint. Discussion started: 5 May 2026

Notz, D., Jahn, A., Holland, M., Hunke, E., Massonnet, F., Stroeve, J., Tremblay, B., and Vancoppenolle, M.: The CMIP6 Sea-Ice Model Intercomparison Project (SIMIP): understanding sea ice through climate-model simulations, Geoscientific Model Development, 9, 3427–3446, 2016.

Olason, E., Boutin, G., Korosov, A., Rampal, P., Williams, T., Kimmritz, M., Dansereau, V., and Samaké, A.: A new brittle rheology and numerical framework for large-scale sea-ice models, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 685, 2022.

OSI SAF: Global Low Resolution Sea Ice Drift, https://doi.org/10.15770/EUM_SAF_OSI_NRT_2007, extracted on: 2024-01-01, 2007.

OSI SAF: Global sea ice concentration interim climate data record (v3.0, 2022), https://doi.org/10.15770/EUM_SAF_OSI_0014, extracted on: 2024-01-01, 2022a.

OSI SAF: Global sea ice concentration climate data record 1978-2020 (v3.0, 2022), https://doi.org/10.15770/EUM_SAF_OSI_0013, extracted on: 2024-01-01, 2022b.

Owens, R. G. and Hewson, T.: ECMWF Forecast User Guide, Tech. rep., ECMWF, Reading, https://doi.org/10.21957/m1cs7h, Replaces previous editions that were available as PDF documents. , 2018.

Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J., and Haas, C.: A weekly Arctic sea-ice thickness data record from merged CryoSat-2 and SMOS satellite data, The Cryosphere, 11, 1607–1623, 2017.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241, Springer, 2015.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K., Oke, P., and Korablev, A.: TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Science, 8, 633–656, 2012.

Siddique, N., Sidike, P., Elkin, C., and Devabhaktuni, V.: U-Net and its variants for medical image segmentation: theory and applications, arXiv preprint arXiv:2011.01118, 2020.

Song, J., Meng, C., and Ermon, S.: Denoising diffusion implicit models, arXiv preprint arXiv:2010.02502, 2020.

Tian, T., Davy, R., Ponsoni, L., and Yang, S.: Impact of modulating surface heat flux through sea ice leads on Arctic sea ice in EC-Earth3 in different climates, The Cryosphere, 19, 2751–2768, 2025.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need, Advances in neural information processing systems, 30, 2017.

Williams, T. and Bertino, L.: PRODUCT USER MANUAL For Arctic Ocean Sea Ice Analysis and Forecasting Products, https://doi.org/10.48670/moi-00004, extracted on: 2026-04-21, 2025.