https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

A prior-regularised heteroscedastic ResUNet for fusing passive-microwave sea-ice concentration products

Fengxin Chen[1,2,3,4,5], Yu-Xuan Fu[1,2,6], Ruibin Xia[1,2], Xiaochun Wang[1,2]

[1]State Key Laboratory of Climate System Prediction and Risk Management, Nanjing University of Information Science and Technology, Nanjing, 210044, China

[2]School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing, 210044, China

[3]Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Center for Ocean Mega - Science, Chinese Academy of Sciences, Qingdao, China

[4]Laboratory for Ocean and Climate Dynamics, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

[5]University of Chinese Academy of Sciences, Beijing, China

[6]Department of Earth and Planetary Science, Graduate School of Science, The University of Tokyo, Tokyo, Japan

Correspondence to: Ruibin Xia (xiarb@nuist.edu.cn)

Abstract. Passive microwave (PMW) sea-ice concentration (SIC) products provide pan-Arctic coverage, but their retrieval errors are often elevated near coastlines and within the marginal ice zone (MIZ), where land spillover, mixed pixels, melt ponds, and atmospheric effects complicate retrieval. We present a prior-regularised two-level residual U-Net (ResUNet) that generates a 12.5 km Arctic SIC field by fusing six PMW SIC products: Bremen-ASI, NSIDC-BT, NSIDC-NT, NSIDC-CDR, SICCI-25km, and OSI-450. The model combines a compact encoder-decoder backbone with auxiliary spatial and seasonal encodings and a heteroscedastic Gaussian negative log-likelihood loss. Empirical relationships between PMW SIC error and distance to land or to the ice edge (defined here as the 0.15 SIC contour), together with product-provided uncertainty estimates, are incorporated into the loss function as pixel-wise reliability information. On the held-out test set, the fused product outperforms all six individual PMW SIC products, reducing MAE by about 55 % and RMSE by about 30 % relative to the best-performing PMW product while maintaining near-zero bias. In the most error-prone regions, RMSE is reduced by about 34 % within 20 km of the coast and by about 25 % within 50 km of the ice edge. Independent validation against Landsat-derived SIC gives the lowest MAE and RMSE for the fused product (0.035 and 0.062), corresponding to improvements of about 13 % and 40 % over the best-performing PMW product, respectively. The fused product also has the lowest errors in the most error-prone regions and smaller interannual RMSE variability. The estimated heteroscedastic uncertainty is informative: on the held-out test set it increases consistently with error and reaches its maximum in coastal and ice-edge regions, while the Landsat validation also shows a consistent ordering of errors with uncertainty. Overall, the proposed framework combines complementary PMW SIC products into a single fused product with improved accuracy and reduced errors near coastlines and the ice edge, providing a useful basis for climate applications and near-real-time sea-ice mapping.

1 Introduction

Rapid Arctic sea-ice change has increased the need for timely, spatially complete, and accurate sea-ice concentration (SIC) information. Arctic sea ice responds sensitively to atmospheric and oceanic forcing (Druckenmiller et al., 2022). Arctic warming has driven pronounced sea ice declines and increased the seasonal accessibility of major passages, creating both

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

economic opportunities and operational risks and amplifying the demand for timely, high-resolution, spatially comprehensive, and accurate sea-ice information (Smith and Stephenson, 2013; Chen et al., 2021).

Sea-ice concentration (SIC), defined as the fraction of a grid cell covered by sea ice, remains a foundational parameter in polar remote sensing and downstream applications. Current SIC products are primarily based on three satellite observing modalities, each with distinctive advantages and limitations. Passive microwave (PMW) instruments provide pan-Arctic, all-weather, day–night coverage and mature long-term SIC products, with some products providing pixel-wise uncertainty estimates, but they suffer from coarse resolution and elevated retrieval errors in summer, near coastlines and within the marginal ice zone (MIZ) (Lavergne et al., 2019; Kern et al., 2020, 2022). Synthetic aperture radar (SAR) can capture small sea-ice features such as floes and leads in much greater detail than coarse-resolution sensors. However, producing high-resolution maps for the whole Arctic every day is still difficult because image coverage is not always complete and routine processing at that scale requires substantial data handling and computing resources (Li et al., 2021; Wulf et al., 2024). Optical or thermal infrared sensors enable high spatial and temporal detail under clear skies yet are constrained by cloud cover and polar night (Ludwig et al., 2020).

Satellite PMW SIC retrievals form the backbone of multi-decadal sea-ice records by exploiting multi-frequency, dual-polarisation brightness temperatures (TB) from sensors such as the Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR), the Special Sensor Microwave/Imager (SSM/I), the Special Sensor Microwave Imager/Sounder (SSMIS), the Advanced Microwave Scanning Radiometer for EOS (AMSR-E), and the Advanced Microwave Scanning Radiometer 2 (AMSR2) (Ivanova et al., 2015; Sandven et al., 2023). Legacy retrieval algorithms include the NASA Team (NT) and Bootstrap (BT) approaches. These algorithms rely on different channel combinations and tie-point definitions, and they show systematic differences across regions and seasons, especially in summer, within the MIZ, and near coastlines. At the product level, long-term SIC products were developed to improve temporal consistency and to reduce errors related to cross-sensor differences, open-water and weather effects, and coastal land-spillover contamination (Peng et al., 2013; Meier et al., 2014). More recent products from the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF) and the European Space Agency Sea Ice Climate Change Initiative (ESA SICCI) also use dynamic tie points and provide per-grid-cell uncertainty estimates (Lavergne et al., 2019). Retrievals based on ~89 GHz channels, including products based on the ARTIST Sea Ice (ASI) algorithm, provide finer nominal grid spacing and a sharper representation of the MIZ. However, they are more sensitive to atmospheric liquid water and weather-related noise than lower-frequency long-term PMW SIC products.

Despite these advances, limitations persist, particularly in low-to-intermediate SIC regions near coastlines and within the MIZ. In these regions, errors are exacerbated by melt ponds, wind-roughened open water, atmospheric effects, and land spillover, which can increase inter-product spread even after filtering. Multi-source data fusion has therefore become an active research direction for balancing spatial resolution, temporal coverage, and accuracy. A common strategy blends a single PMW SIC product with higher-resolution modalities (optical or SAR) to sharpen ice-edge geometry and add fine-scale structure

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

when auxiliary observations are available (Banzon et al., 2016; Ludwig et al., 2020; Malmgren-Hansen et al., 2021). However, cross-modality fusion remains constrained by the availability, noise characteristics, and sampling of the auxiliary observations. Optical inputs are limited by cloud cover and polar night, whereas SAR coverage and processing requirements still hinder daily pan-Arctic mapping. Consequently, large regions and time periods revert to PMW-like effective resolution and error characteristics, particularly in cloudy/polar-night conditions and in dynamically difficult coastal and ice-edge regions.

A key opportunity is that PMW SIC retrieval is represented not by a single baseline product but by a family of mature products that differ in sensor lineage, retrieval algorithm, channel and polarisation combinations, tie-point strategy, and filtering choices. As a result, these products exhibit complementary strengths across seasons, ice conditions, and regions (Cavalieri, 1991; Spreen et al., 2008; Ivanova et al., 2015; Lavergne et al., 2019; Comiso, 2023). Motivated by this diversity, we therefore fuse multiple PMW SIC products instead of relying on intermittently available SAR or optical inputs, thereby preserving the pan-Arctic, all-weather, near-real-time coverage of PMW observations. We further introduce physically informed priors that encode where PMW SIC retrievals are known to degrade, particularly near coastlines, across the MIZ, and in the narrower ice-edge region examined in our diagnostics. To target these error-prone regions, we fit empirical relationships between PMW SIC error and distance to land or to the ice edge, defined here as the 0.15 SIC contour, and combine these relationships with product-provided uncertainty estimates to construct pixel-wise reliability weights for training. The resulting fused SIC field uses the finer-grid, high-frequency PMW information, when it is reliable, to better represent the ice edge; otherwise, the model prioritises accuracy, temporal stability, and quantified uncertainty over nominal grid spacing alone. It can also serve as a stable baseline to which SAR, optical, or reanalysis information may later be added as optional refinements rather than hard dependencies.

We adopt a compact U-Net-style encoder–decoder with residual blocks (ResUNet) to fuse six PMW SIC products because the architecture combines local boundary preservation with broader contextual aggregation, which is important for representing both sharp ice–water and coastal transitions and the larger-scale spatial structure of the Arctic ice cover (Ronneberger et al., 2015; Isensee et al., 2021). Other networks commonly used for similar tasks include DeepLabv3+, HRNet, and Transformer-based models. However, DeepLabv3+ can improve multi-scale context aggregation through atrous spatial pyramid pooling, but typically at higher computational cost (Chen et al., 2018). HRNet maintains high-resolution feature streams, but at the cost of higher memory consumption and increased inference latency (Sun et al., 2019). Transformer-based segmenters can perform strongly when large-scale pre-training is available, but their computational efficiency and performance under limited training data may depend strongly on model scale and training setup in remote-sensing applications (Xie et al., 2021; Wang et al., 2024). For the spatial resolution and image size used in Arctic SIC fusion, these more complex networks may not provide a consistent advantage over a compact ResUNet. Our choice is also consistent with previous cryosphere and coastal remote-sensing studies in which U-Net-type architectures have performed well for SIC products (Hoffman et al., 2021, 2022; Aghdami-Nia et al., 2022; Seale et al., 2022; Huang et al., 2024; Murashkin et al., 2024).

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

Relative to previous SIC fusion studies, this work makes three contributions: (1) an all-PMW fusion strategy that exploits complementary information from six mature SIC products while preserving daily pan-Arctic, all-weather coverage, rather than relying on intermittently available SAR or optical inputs; (2) prior-informed training using empirical error structures with respect to land, the ice edge, and product-provided uncertainty estimates; and (3) joint estimation of SIC and its uncertainty within a heteroscedastic framework.

The paper is organised as follows. Sect. 2 describes the datasets and preprocessing, including the MODIS-derived reference dataset, the six PMW SIC products used for fusion, and the independent Landsat-derived SIC dataset used for validation. Sect. 3 defines the fusion problem and presents the prior-informed ResUNet framework, including the construction of the empirical priors, the input channels, the learning objective, and the training setup. Sect. 4 presents the results, including model selection, comparison with the individual PMW SIC products, distance-stratified error analysis, uncertainty evaluation, and independent validation with the Landsat-derived SIC dataset. Finally, Sect. 5 summarises the main findings and discusses their implications for sea-ice monitoring.

2 Data and preprocessing

2.1 The MODIS-derived SIC reference dataset

The MODIS-derived SIC reference dataset is taken from Rösel et al. (2015) and was also used by Kern et al. (2020) to evaluate PMW SIC products. It covers 2000-05-08 to 2011-09-13 and is provided as an 8-day composite product. The 8-day composite is not an average; for each native 500 m pixel, it retains the best available observation within the 8-day window, prioritising cloud-free conditions, low view angle, and high observation quality. The dataset includes melt-pond fraction, its standard deviation, and open-water fraction (OWF); we derive reference SIC directly from the gridded OWF as SIC = 1 − OWF. The dataset is provided on a 12.5 km Arctic polar stereographic grid and is only available from early May to mid-September, thereby mainly sampling the melt season from pre-melt conditions to late summer. Because optical retrievals provide finer spatial detail under clear-sky conditions, we use the MODIS-derived SIC as an approximate reference for supervised training and evaluation of the PMW fusion model. Because the source product is an 8-day composite and does not provide the exact acquisition date for each retained 12.5 km grid cell, we assign the retained MODIS value to the nominal sample date and use it as an approximate daily label. Because the MODIS-derived SIC is an 8-day composite without exact acquisition dates for all retained grid cells, we treat it as an approximate reference dataset for supervised training and evaluation, rather than as exact same-day truth.

2.2 PMW SIC products used for fusion

PMW SIC products vary in retrieval algorithm, channel and polarisation combination, and filtering procedure. This diversity results in complementary strengths across Arctic regions and seasons. To leverage this complementary information, we selected the PMW SIC products used for fusion based on two criteria. First, they must temporally overlap with the MODIS SIC period to enable collocation with the MODIS reference dataset. Second, they should cover a range of sensors and/or

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

retrieval algorithms so that the selected products exhibit diverse information content and error characteristics, allowing the fusion model to exploit complementary strengths and reducing sensitivity to biases in any single product.

Table 1. PMW SIC products used for fusion. Listed are the sensor/platform (key channels), retrieval algorithm, nominal grid spacing, temporal coverage of the products, and references.

Temporal coverage in the fifth column refers to the entire product record. For model training and evaluation, we use only inputs from the period overlapping the MODIS reference dataset.

ID	Sensor/platform (key channels)	Algorithm	Nominal grid spacing	Temporal coverage (product record)	Literature
Bremen-ASI	AMSR-E/2 (~89 GHz)	ARTIST Sea Ice (ASI)	6.25 km	2002-06-01 to present	Melsheimer and Spreen, 2019
NSIDC-BT	SMMR/SSM/I/SSMIS (19/37 GHz)	Bootstrap (BT)	25 km	1978-10-26 to present	Comiso, 2023
NSIDC-NT	SMMR/SSM/I/SSMIS (19.35/37.0 GHz)	NASA Team (NT)	25 km	1978-10-26 to present	Cavalieri, 1991
NSIDC-CDR	SMMR/SSM/I/SSMIS (19/37 GHz)	CDR (consensus NT/BT)	25 km	1978-12-25 to present	Meier et al., 2017
SICCI-25km	AMSR-E/2 (19/~89 GHz)	ESA SICCI / NASA Team 2 (NT2)	25 km	2002-05-31 to 2017-05-15	Lavergne et al., 2019
OSI-450	SMMR/SSM/I/SSMIS (19/37 GHz)	OSI SAF	25 km	1978-10-25 to 2020-12-31	Lavergne et al., 2019

In Table 1, "Bremen-ASI" refers to the University of Bremen SIC product based on the ASI algorithm, which uses the ~89 GHz channels of AMSR-E/2 and is provided on a 6.25 km grid, enabling a comparatively clearer delineation of the ice edge and narrow leads (e.g., Melsheimer and Spreen, 2019, 2020). "NSIDC-BT", "NSIDC-NT", and "NSIDC-CDR" refer to sea ice concentration products from the National Snow and Ice Data Center (NSIDC), with each using different algorithms: BT, NT, and Consensus (CDR), respectively. "SICCI-25km" is the European Space Agency Sea Ice Climate Change Initiative (ESA

SICCI) product based on an enhanced NASA Team 2 (NT2)-type retrieval and provides pixel-wise uncertainty estimates (Lavergne et al., 2019). "OSI-450" is the reprocessed OSI SAF sea-ice concentration climate data record, produced by the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF) from SMMR/SSM/I/SSMIS observations, providing a long-term, dynamically tuned SIC record with pixel-wise uncertainty estimates (Lavergne et al., 2019). All products are resampled to the common 12.5 km Arctic polar stereographic grid used by the MODIS reference dataset. For brevity, we use the product IDs listed in Table 1 throughout the paper.

2.3 Preprocessing and training-data construction

The MODIS SIC scenes are provided on a native polar stereographic x–y grid whose spatial extent is larger than the Arctic 60–90° N region. However, pixels outside the valid SIC region are set to missing values (NaN). We therefore cropped the native grid using a fixed rectangular box that retains all valid SIC pixels while removing most of the surrounding empty area (Fig. 1a, b). The resulting grid, used throughout this study as the common grid and ResUNet input domain, contains 512 × 512 cells at 12.5 km resolution.

The subsequent preprocessing workflow is summarised in Fig. 1c. To ensure data quality while retaining enough scenes, we retain a MODIS SIC scene only if cloud-masked ocean pixels account for <10 % of valid ocean pixels, where valid ocean pixels are defined by the native land–ocean mask of the MODIS SIC product. This screening yields 156 MODIS SIC scenes. For each 8-day MODIS composite, we use its nominal UTC date as the temporal anchor because the product does not provide an exact acquisition date for every retained 12.5 km grid cell. We then make the land and pole-hole masks consistent across products and rescale SIC from percent to fractional values in [0, 1]. Cloud-masked missing pixels in the MODIS SIC reference dataset are filled with zeros for numerical convenience, while a corresponding missing-data mask is retained so that these pixels are excluded from the loss calculation and evaluation metrics.

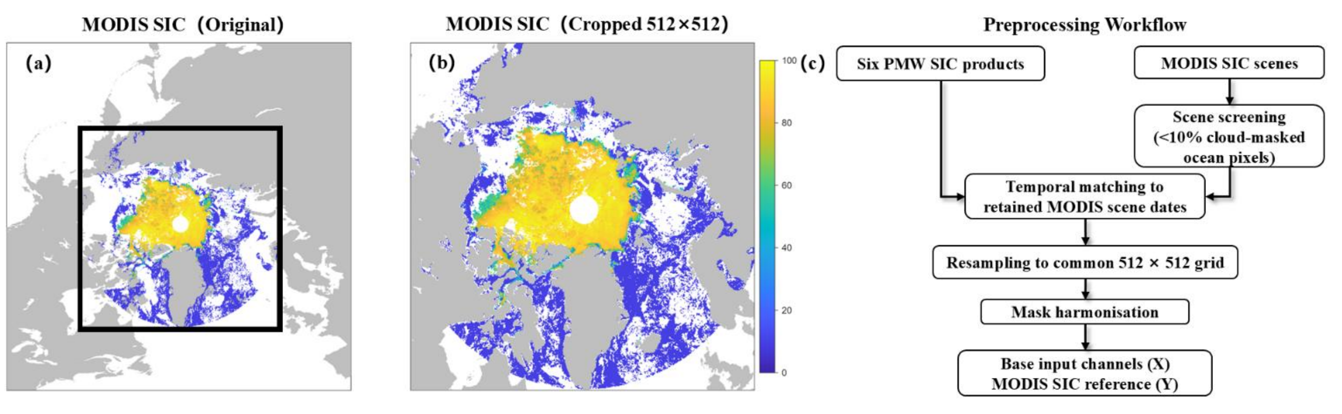


Figure 1. Data preprocessing and construction of matched input–reference pairs. (a) MODIS SIC scene on the native polar stereographic grid. The black rectangle shows the fixed grid used in this study. Pixels outside the valid SIC domain within the native grid are NaN; the

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

fixed 512 x 512 grid contains the retained analysis domain. (b) MODIS SIC scene after cropping to the common 512 × 512 polar stereographic grid (12.5 km) used by our model. (c) Schematic of the preprocessing workflow used to construct the base input channels (X) and the corresponding MODIS SIC reference (Y) on the common grid.

2.4 Landsat-derived SIC dataset for independent validation

For independent evaluation, the Landsat-derived SIC is aggregated to the 12.5 km Arctic polar stereographic grid used by our model. This evaluation dataset is from the ESA Sea Ice Climate Change Initiative Essential Climate Variable (ECV) project (Kern, 2021). Surface broadband albedo is estimated from Landsat spectral channels (2–4 for Landsat-5/-7; 3–5 for Landsat-8) and is then classified into open water, thin ice, and thick ice. SIC is computed as the fraction of thin-ice and thick-ice pixels within each grid cell (Fig. 2). Because it is derived from 30 m surface-type information, this Landsat-based SIC provides a high-resolution independent reference for evaluation. The dataset is provided at 30 m pixel spacing and covers the Arctic for 2003–2010 and 2013–2015; it has been used to evaluate PMW SIC products (Kern et al., 2022).

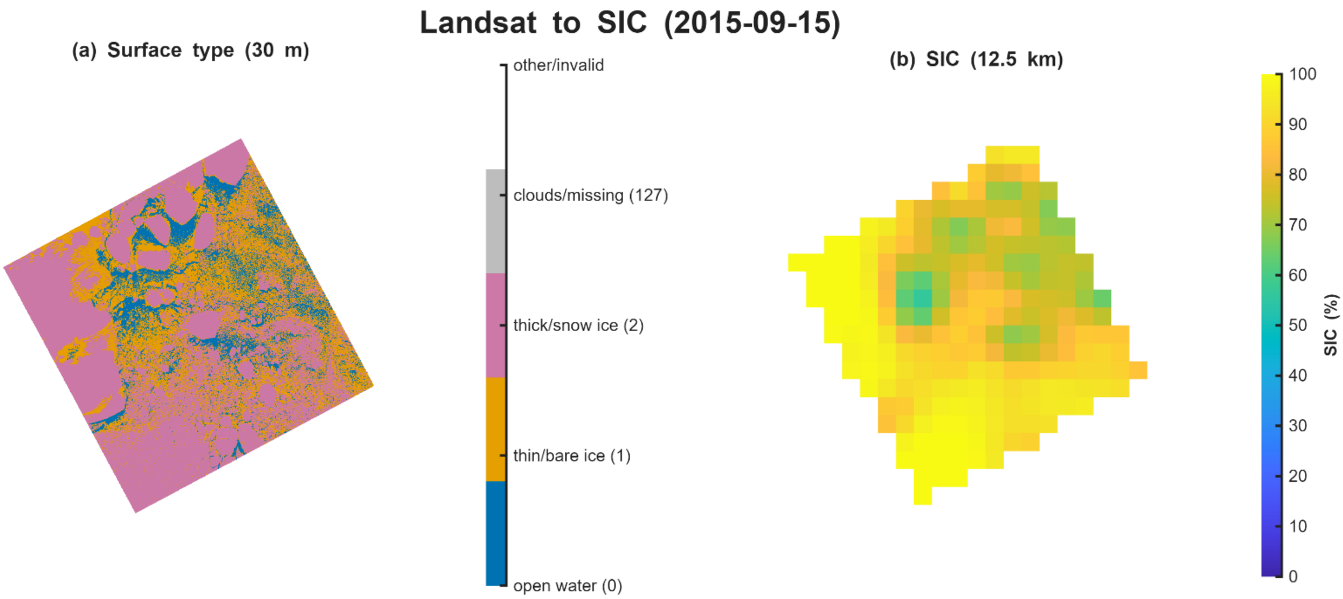


Figure 2. Landsat surface-type classification and derived SIC on the 12.5 km grid for 15 September 2015. (a) Surface type at 30 m pixel spacing in polar stereographic projection; the scene is automatically aligned to the target grid using a flip–then–rotate procedure. Colours denote open water (0; blue), thin/bare ice (1; orange), thick/snow ice (2; purple), clouds/missing data (127; grey), and other or invalid pixels (white). (b) SIC (%) on the 12.5 km grid derived from (a) using a 12.5 km × 12.5 km window centred on each target cell. Only classes 0, 1, and 2 are treated as valid surface classes; windows with no valid pixels or with more than 40 % invalid pixels are set to NaN and shown in white. For visualisation, the SIC field is smoothed with a NaN-aware 3 × 3 filter.

However, after aggregation to the 12.5 km model grid, the matchups are sparse and uneven in both space and time (Fig. 3), and the spatial coverage on any given day is very limited. For this reason, although the dataset is well suited for independent validation, it is not suitable for use as a training label. Because many available samples are located close to land, this dataset

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

is particularly informative for evaluating coastal performance, one of the most challenging settings for PMW SIC retrieval. Because it is independent of the MODIS reference, it also provides a useful check for using the MODIS 8-day composite SIC as the training reference. In our matchups, most Landsat scenes fall in March–May; therefore, this validation mainly reflects spring conditions.

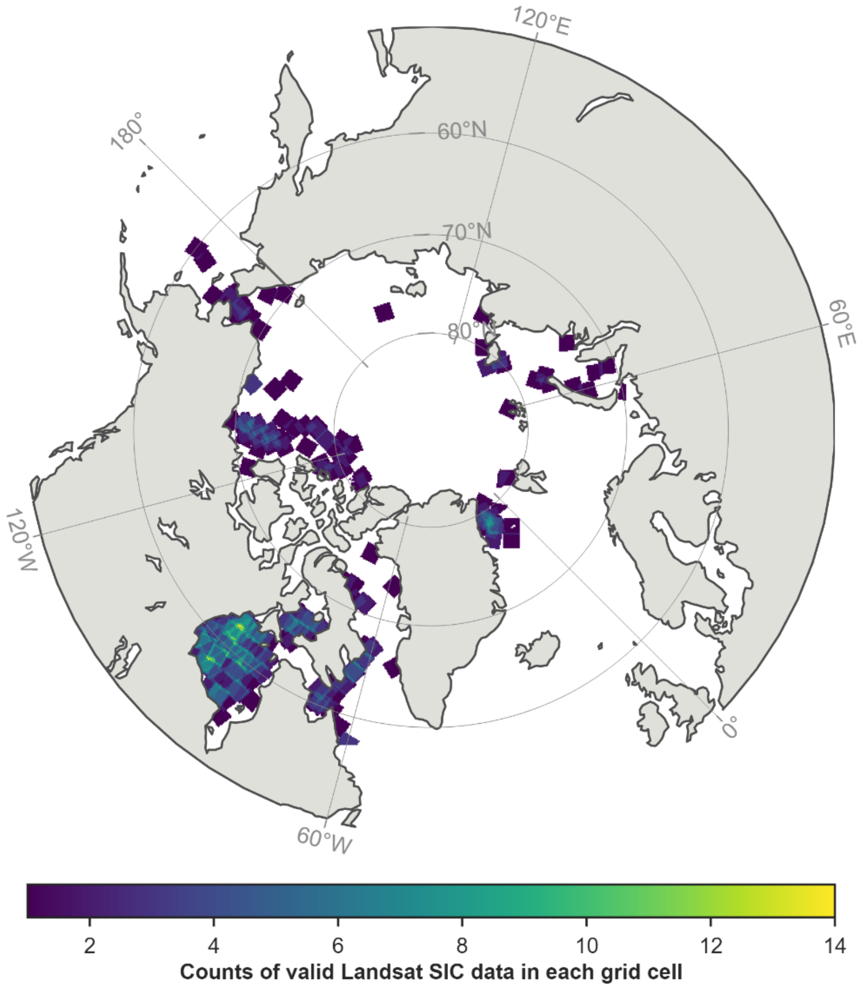


Figure 3. Spatial distribution of the number of valid Landsat-derived SIC retrievals per 12.5 km grid cell used in this study during 2003–2010 and 2013–2015. Colours indicate the number of valid Landsat SIC samples available in each grid cell, and grid cells with no valid retrievals are left blank. The map is shown in a North Polar stereographic projection with coastlines in grey.

3 Methodology

3.1 Problem formulation

We formulate SIC fusion as a masked supervised heteroscedastic regression problem on the 12.5 km Arctic polar stereographic grid of Sect. 2.3. Let $\Omega = \{1, \dots, H\} \times \{1, \dots, W\}$ denote the spatial domain of a single scene, where $H = W = 512$. For any

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

field $A \in \mathrm{R}^{(H \times W)}$, we write A_{ij} for its value at grid cell $(i,j) \in \Omega$. All inputs, reference fields, masks, model outputs, and loss terms are defined on this grid.

For one scene, the model input is

$$X \in [0,1]^{H \times W \times N}, \tag{1}$$

where N denotes the number of input channels. The target field is

$$Y \in [0,1]^{H \times W}, \tag{2}$$

where Y denotes the MODIS SIC reference scene used as the supervised training target. A binary validity mask

$$M \in \{0,1\}^{H \times W} \tag{3}$$

indicates which grid cells have valid MODIS reference values. We define the set of valid grid cells as

$$\Omega_{\text{valid}} = \{(i,j) \in \Omega : M_{ij} = 1\} \tag{4}$$

The model returns values on the full grid, but the loss is computed only over Ω_{valid}. The evaluation metrics for the MODIS reference dataset are also computed over Ω_{valid}unless otherwise noted.

Instead of estimating SIC alone, the model estimates both SIC and pixel-wise uncertainty. Given the input X, the network f_θ with parameters θ outputs a fused SIC estimate μ and a pixel-wise log-variance s,

$$(\mu, s) = f_\theta(X),\ \mu \in [0,1]^{H \times W},\ s \in \mathrm{R}^{H \times W}. \tag{5}$$

Here μ is the fused SIC estimate and s is the estimated log-variance. The corresponding estimated standard deviation is

$$\sigma_{ij} = \exp\left(\frac{1}{2} s_{ij}\right), \tag{6}$$

which is used as the pixel-wise uncertainty estimate. The model is trained using the average loss over all labelled training scenes.

Together, Eqs. (1)–(6) define the masked learning problem used in the remainder of the paper.

3.2 Overview of the proposed framework

The fusion framework is designed around two considerations: the six PMW SIC products provide complementary information, and PMW SIC errors are largest near coastlines and the ice edge. Accordingly, the model takes the six PMW products as the

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

core inputs and can optionally augment them with contextual and diagnostic channels that help the network interpret location, season, and multi-product agreement.

Based on these characteristics, we design a prior-informed fusion framework, summarised in Fig. 4. The six PMW SIC products provide the mandatory observational inputs. The input can optionally be augmented with physically informed contextual channels that encode location, coastal proximity, and seasonal phase, together with PMW-derived diagnostic fields summarising multi-product agreement, spread, and edge-related structure. These additional components are intended to provide contextual and diagnostic support to the fusion, rather than to introduce a new observational modality or redefine the task, and their utility is assessed in the ablation experiments.

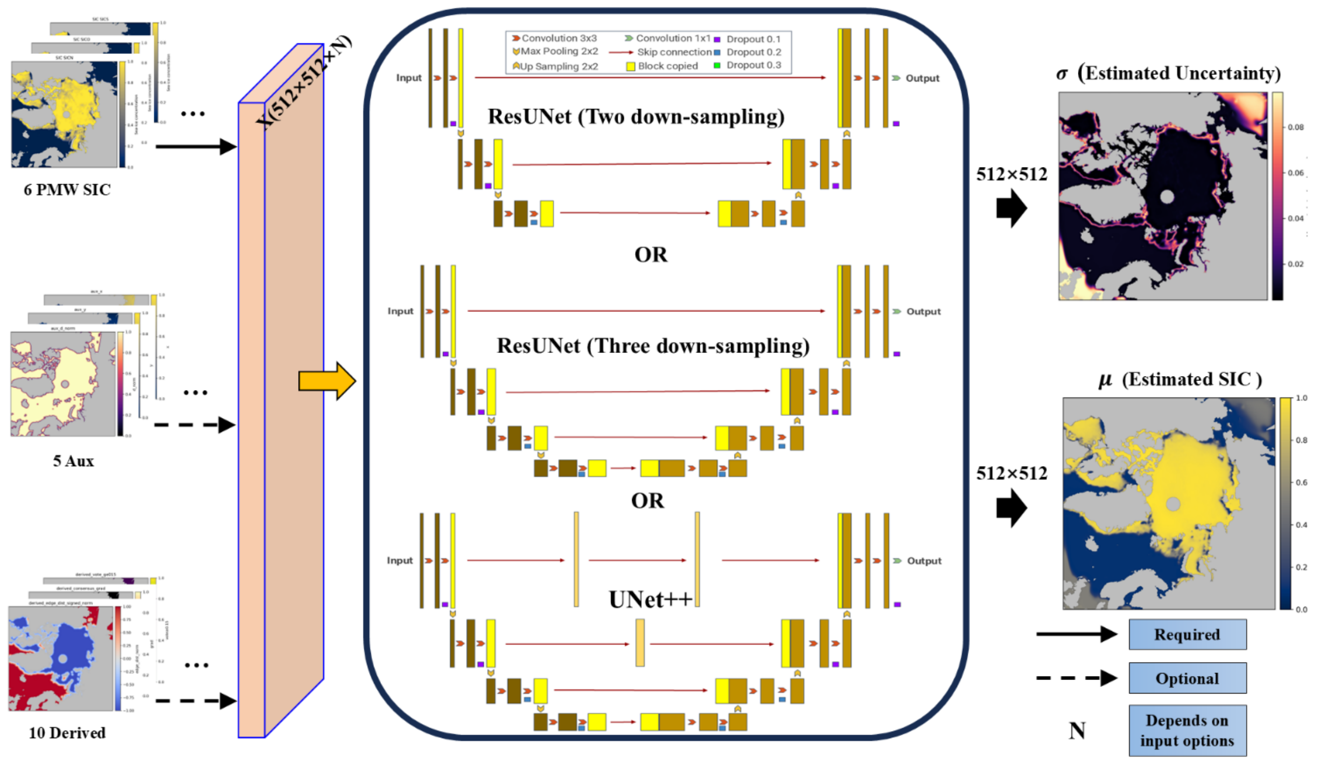


Figure 4. Overview of the proposed PMW SIC fusion framework. Six PMW SIC products on the 512 × 512, 12.5 km Arctic polar stereographic grid form the base input. The input can be optionally augmented with auxiliary channels and with derived PMW diagnostics. A candidate U-Net-based backbone maps the multi-channel input to a fused SIC field and, when enabled, to a corresponding pixel-wise estimated uncertainty field. Empirical priors derived from PMW error analysis are incorporated into training through prior-informed weighting and regularisation. Candidate backbones are compared under controlled ablations, and one reference configuration is selected for the main experiments.

The framework is based on two considerations. First, the model learns the mapping from the multi-product PMW input to a fused SIC estimate and, when enabled, to a corresponding pixel-wise uncertainty estimate using a compact convolutional encoder-decoder. Second, empirically derived error-distance and uncertainty-error relations are used as frozen priors in the

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

loss function. We consider three candidate U-Net-type backbones: a two-level ResUNet, a three-level ResUNet, and a U-Net++ variant. All use residual blocks, encoder-decoder symmetry, and skip connections, with light dropout applied at the bottleneck. The primary output head estimates a fused SIC map through a final sigmoid activation. When heteroscedastic regression is enabled, an additional parallel head estimates a pixel-wise log-variance field, from which the estimated standard deviation is obtained. Auxiliary decoder heads denote deep-supervision output heads attached to intermediate decoder levels and used only during training. Candidate backbones are compared under controlled ablations, and the reference configuration adopted for the main experiments is selected in Sect. 4. We next quantify how PMW SIC retrieval error varies with distance to land, distance to the ice edge, and, where available, product-provided uncertainty, and use the resulting fitted relations as frozen priors in the learning objective.$\mu \in [0,1]^{H\times W} s \in \mathrm{R}^{H\times W} \sigma = \exp\left(\frac{1}{2}s\right)$

The remainder of Sect. 3 follows the structure of Fig. 4. Sect. 3.3 describes how the empirical priors are constructed from MODIS–PMW matchups. Sect. 3.4 defines the model inputs, including the base PMW SIC products and the optional auxiliary and diagnostic channels. Sect. 3.5 presents the prior-informed learning objective. Sect. 3.6 summarises the training setup and reproducibility details. Backbone benchmarking and reference-model selection are reported in Sect. 4.1.

3.3 Empirical prior construction from error analysis

3.3.1 Definition of explanatory variables and fitting strategy

We quantify these error structures using the MODIS–PMW training and validation matchups. Specifically, we quantify how retrieval error varies with distance to land, distance to the ice edge, and product-provided uncertainty, and then convert the fitted relationships into fixed priors for model training.

Let Z_i∈[0,1]^(H×W) denote the SIC field from the i-th PMW product, with i=1,⋯,6. For each valid pixel we define the retrieval error relative to the MODIS reference as

$$e_i = |Z_i - Y|, \tag{7}$$

The held-out test set is excluded from all prior fitting.

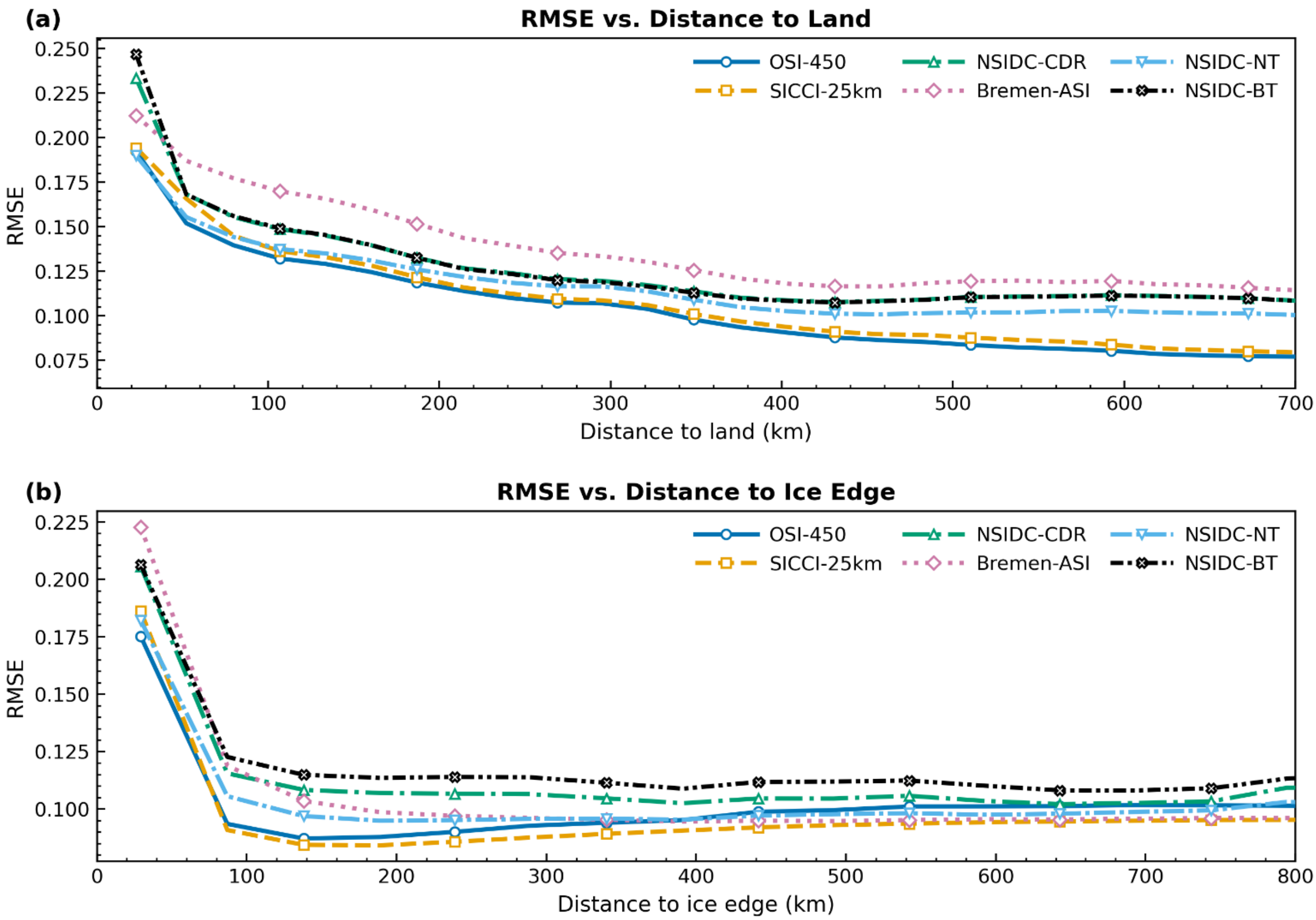
We analyse three variables. The first is the distance to the nearest land pixel, denoted d_{land}. The second is the signed distance to the ice edge, denoted d_{edge}, where the ice edge is defined as the 0.15 SIC contour of the per-date median field across the six PMW products. The third is the product-provided uncertainty U_{i} for products that provide pixel-wise uncertainty maps. For each variable, we quantify the error relationships, fit the response functions, and then use the fitted function to construct frozen prior objects for the learning objective.

3.3.2 Distance-to-land and distance-to-ice-edge error structure

We first quantified retrieval error with distance to land and distance to the ice edge. The ice-edge effect is more localised:

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

errors peak at the ice margin and decay rapidly, whereas the land influence decays more gradually and extends farther offshore. We parameterise these distance–error profiles for later use in training.



265 **Figure 5.** SIC RMSE as a function of distance to (a) the nearest land pixel and (b) the nearest ice edge for the six PMW SIC products. Curves show smoothed RMSE profiles computed from the MODIS–PMW training and validation matchups. RMSE decreases with offshore distance and approaches a low plateau at large distances from land, whereas the ice-edge-related error decreases more rapidly within the first few hundred kilometres from the 0.15 SIC contour. Products are identified by the in-panel legend using a consistent combination of colour, line style, and marker.

270 Using the training/validation MODIS–PMW SIC dataset, we compute RMSE profiles for each PMW product as functions of distance to land, d_{land}, and distance to the ice edge, d_{edge}. RMSE is highest near land or the ice edge, decreases rapidly with increasing distance, and then gradually approaches an approximately constant plateau farther away. We therefore model both dependencies with the same functional form, an offset stretched-exponential decay, while allowing product-specific amplitude and floor parameters. In particular, it captures the elevated errors close to land or the ice edge, the subsequent decay with

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

distance, and the non-zero background error level that remains at large distances. It is also flexible enough to represent both the more gradual offshore decrease in Fig. 5(a) and the sharper near-edge drop in Fig. 5(b).

For boundary type $b \in \{\text{land}, \text{edge}\}$, we model the RMSE of product i as

$$\hat{\sigma}_i{}^b(d) = a_i^b \exp\left[-\left(\frac{d}{\lambda_b}\right)^{\beta_b}\right] + c_i{}^b, \tag{8}$$

where $d = d_{\text{land}}$ for $b = \text{land}$ and $d = d_{\text{edge}}$ for $b = \text{edge}$. Here, a_i^b and c_i^b are product-specific amplitude and floor parameters, while λ_b and β_b are boundary-type-specific scale and shape parameters shared across products. In our implementation, the shared fit gives $(\lambda_{\text{land}}, \beta_{\text{land}})$ = (41.554 km, 0.400) and $(\lambda_{\text{edge}}, \beta_{\text{edge}})$ = (49.365 km, 1.675)

The fitted error curves are then mapped to product-wise reliabilities,

$$r_i{}^b(d) = \frac{1}{(\hat{\sigma}_i{}^b(d))^2 + \epsilon}, \tag{9}$$

where $\epsilon > 0$ is a small constant for numerical stability. These reliability functions are later used to modulate the per-product consensus constraint, so that individual PMW products contribute more strongly where they are empirically more reliable and less strongly where they are expected to be less reliable.

Table 2. Pearson correlation r and Spearman rank correlation ρ between the realised pixel-wise absolute retrieval error and the distance-related explanatory variables. "Raw" values are computed using the original distance variables, whereas "Fitted" values are computed using the corresponding RMSE-based fitted distance–response functions. The correlations are used as dimensionless association scores for prior-family weighting and are not used to refit the RMSE profiles. For each product, the larger absolute value within the land/edge pair for a given metric is bolded, and the last row reports means across products.

ID	Distance-to-land: Raw r	Distance-to-land: Raw ρ	Distance-to-land: Fitted r	Distance-to-land: Fitted ρ	Distance-to-ice-edge: Raw r	Distance-to-ice-edge: Raw ρ	Distance-to-ice-edge: Fitted r	Distance-to-ice-edge: Fitted ρ
OSI-450	-0.255	-0.304	0.291	0.304	-0.068	0.010	0.284	0.121
SICCI-25km	-0.224	-0.288	0.246	0.288	-0.076	0.019	0.286	0.098
NSIDC-CDR	-0.178	-0.191	0.246	0.191	-0.117	-0.035	0.289	0.035
Bremen-ASI	-0.154	-0.172	0.172	0.172	-0.135	-0.057	0.301	0.061

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

NSIDC-NT	-0.188	-0.206	0.212	0.205	-0.100	-0.039	0.280	0.093
NSIDC-BT	-0.196	-0.210	0.288	0.210	-0.115	-0.035	0.274	0.036

To summarise whether the fitted RMSE-based distance response functions provide useful pixel-wise error scores, we compute the Pearson correlation coefficient r and the Spearman rank correlation ρ coefficient between the realised absolute retrieval error and either the raw distance variable or its fitted RMSE-based response value, using all valid analysis pixels (Table 2). These correlations are used only as dimensionless association scores for weighting the prior families; the fitted response functions themselves remain RMSE-based and are not refitted on the test set.

3.3.3 Uncertainty–error relationship

Only three products provide pixel-wise uncertainty fields: SICCI-25km and OSI-450 report total SIC uncertainty, whereas NSIDC-CDR provides the PMW Daily Hemisphere SIC Source Estimated Standard Deviation. To test whether these fields are informative, we analyse how retrieval RMSE varies with the reported uncertainty U_i. For each product, we bin valid pixels by U_i, compute RMSE within each bin, and then fit a smooth response curve, $\hat{\sigma}_i^{\mathrm{unc}}(U_i)$, to reduce bin-to-bin fluctuations. This is appropriate here because RMSE is expected to increase overall with reported uncertainty, while individual bins—especially at the largest U_i—can be noisy due to limited sample counts. As shown in Fig. 6, all three products follow the same main pattern: higher reported uncertainty generally corresponds to larger RMSE, although the rate of increase differs by product.

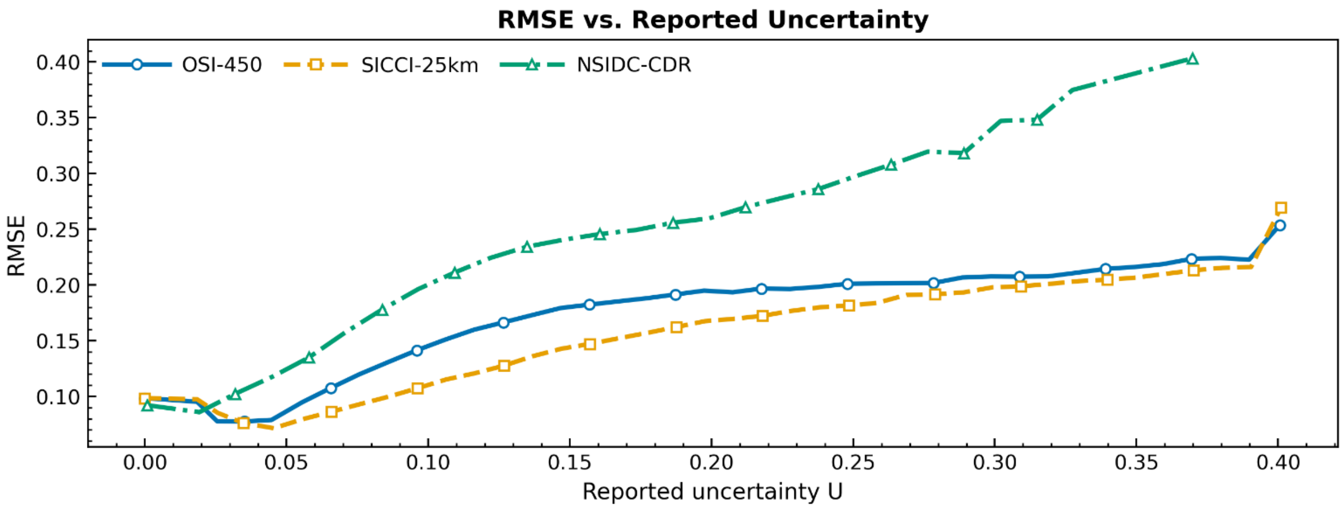


Figure 6. SIC RMSE as a function of reported uncertainty for the three PMW SIC products that provide pixel-wise uncertainty estimates. Markers show binned RMSE values, and lines show the corresponding smoothed response curves. In all three products, larger reported

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

uncertainty is generally associated with larger RMSE, although the growth rate differs among products and the highest-uncertainty range is less stable because it contains fewer samples.

We convert this fitted uncertainty–error relation into an uncertainty-based reliability function,

$$r_i^{\mathrm{unc}}(U_i) = \frac{1}{(\hat{\sigma}_i^{\mathrm{unc}}(U_i))^2 + \varepsilon}, \tag{10}$$

where $\varepsilon > 0$ again ensures numerical stability. For products without uncertainty fields, we set

$$r_i^{\mathrm{unc}} \equiv 1. \tag{11}$$

Table 3. Pearson correlation r and Spearman rank correlation ρ between the realised pixel-wise absolute retrieval error and the product-provided SIC uncertainty U for products that provide uncertainty fields. "Raw" correlations are computed directly from the observed $(|e|, U)$ values; "Fitted" correlations are computed after mapping U through the corresponding RMSE-based fitted uncertainty–response functions. $\Delta r = r_{\mathrm{fitted}} - r_{\mathrm{raw}}$. The correlation coefficients are used as association scores for prior weighting. The final column reports robust slopes of binned MAE and RMSE versus U, which are included only as supplementary diagnostics of error sensitivity to product-provided uncertainty.

ID	r Raw	ρ Raw	r Fitted	ρ Fitted	Δr	Slope (MAE / RMSE)
OSI-450	0.382	0.127	0.405	0.257	0.022	0.328 / 0.456
SICCI-25km	0.388	0.074	0.415	0.212	0.026	0.292 / 0.447
NSIDC-CDR	0.394	0.217	0.405	0.266	0.010	0.647 / 0.958

As in Sect. 3.3.2, we summarise the association between realised pixel-wise absolute retrieval error and product-provided uncertainty U_i using Pearson's correlation r and Spearman's rank correlation ρ, both before and after mapping U through the RMSE-based fitted response curve (Table 3). These coefficients are used later as weighting coefficients, whereas the uncertainty–error relation itself is fitted once on training and validation data and then frozen before test evaluation.

3.3.4 Frozen prior functions and coefficients

The fitted distance–error and uncertainty–error relations are estimated using training and validation data only once and are then frozen. These frozen prior objects are used to construct the pixel-wise product reliabilities, the prior-weighted supervised branches, and the family-level coefficients that control the relative strengths of the prior-informed loss terms.

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

We precompute and keep fixed three types of prior quantities. The first type is the set of product-specific reliability functions $r_i^{\text{land}}, r_i^{\text{edge}}$, and r_i^{unc} defined in Eqs. (8)–(10). These functions are later combined into pixel-wise weights in the PMW consensus regulariser, determining how strongly each PMW product contributes at each pixel.

The second type is the set of scene-level spatial weighting maps used in the supervised reweighting branches. For the coast-weighted branch, we evaluate the fitted land-distance error curve at each pixel to obtain ω_{land}, and then normalise it to unit mean over valid pixels. This ensures that ω_{land} changes the spatial emphasis within a scene rather than simply rescaling the overall loss. For the uncertainty-weighted branch, we convert each available uncertainty field U_i into a fitted error score through $\hat{\sigma}_i^{\text{unc}}(U_i)$, average the available scores pixel by pixel within the scene, and normalise the resulting map to unit mean over valid pixels to obtain ω_{unc}.

The third type is a single fixed weight for each prior family—edge, coast, and uncertainty—which sets how much that family contributes to the total prior-informed loss. Let $r_{(p,i)}^{\text{fit}}$ denote the fitted Pearson correlation for prior family $p \in \{\text{edge}, \text{coast}, \text{unc}\}$ and product i, as reported in Tables 2–3. We first compute an average positive-correlation score for each family,

$$s_p = \frac{1}{N_p} \sum_{i=1}^{N_p} \max\left(r_{(p,i)}^{\text{fit}}, 0\right), p \in \{\text{edge}, \text{coast}, \text{unc}\}, \tag{12}$$

where N_p is the number of products contributing to family p. We then normalise these scores to obtain the corresponding loss weights,

$$\alpha_p = \frac{s_p}{s_{\text{edge}} + s_{\text{coast}} + s_{\text{unc}}}, p \in \{\text{edge}, \text{coast}, \text{unc}\}. \tag{13}$$

In our implementation, this gives $(\alpha_{\text{edge}}, \alpha_{\text{coast}}, \alpha_{\text{unc}}) = (0.301, 0.258, 0.441)$. These coefficients remain fixed during both training and evaluation. They only control the relative strength of the edge-, coast-, and uncertainty-based prior loss terms; they do not modify the fitted reliability functions themselves.

3.4 Input representation and channel design

3.4.1 Mandatory PMW SIC inputs

The mandatory model inputs are the six PMW SIC products. Let $s_i \in [0,1]^{H\times W}$ denote the SIC field from the i-th PMW product. The baseline SIC input block is

$$X_{\text{sic}} = [s_1, \dots, s_6] \in [0,1]^{H\times W\times 6}, \tag{14}$$

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

where the six channels correspond to Bremen-ASI, NSIDC-BT, NSIDC-NT, NSIDC-CDR, SICCI-25km, and OSI-450.

3.4.2 Auxiliary spatial and seasonal channels

In addition to the six SIC channels, we optionally concatenate a compact set of auxiliary channels that encode space and season. These channels are physically meaningful context variables: they provide geographical and seasonal information known to affect PMW SIC retrievals.

Let $x \in \{0, \dots, W-1\}$ and $y \in \{0, \dots, H-1\}$ denote the grid indices. The normalised coordinate channels are defined as

$$C_X = \frac{x}{W-1}, \qquad C_Y = \frac{y}{H-1}. \tag{15}$$

We compute a distance-to-land field d_{land} as the distance, in kilometres, from each valid ocean pixel to the nearest land pixel in the land mask, excluding pole-hole pixels from the distance transform. We then normalise it as

$$d_{\text{norm}} = \min\left(\max\left(\frac{d_{\text{land}}}{50\ km}, 1\right), 0\right). \tag{16}$$

Seasonal phase is encoded from day-of-year (DOY) using sine and cosine channels,

$$d_{\sin} = \sin\left(\frac{2\pi\,\text{DOY}}{N_y}\right), \qquad d_{\cos} = \cos\left(\frac{2\pi\,\text{DOY}}{N_y}\right), \tag{17}$$

where $N_y = 365$ or 366 depending on whether the year is a leap year. When enabled, the auxiliary block is therefore

$$X_{\text{aux}} = [C_X, C_Y, d_{\text{norm}}, d_{\sin}, d_{\cos}] \in \mathrm{R}^{H\times W\times 5}, \tag{18}$$

and the full input becomes

$$X = [X_{\text{sic}}, X_{\text{aux}}] \in \mathrm{R}^{H\times W\times 11}. \tag{19}$$

When auxiliary channels are disabled, the input reduces to the six SIC channels in Eq. (14).

3.4.3 Derived PMW diagnostics

We also compute a set of derived PMW diagnostics on the same grid. These fields depend only on the six PMW SIC inputs and are used for three purposes: (i) analysis and prior construction, (ii) interpretation of multi-product behaviour, and (iii) optional input augmentation in ablation experiments. They are not required in the baseline input configuration.

The default diagnostic set comprises ten maps: (1) the pixel-wise consensus median; (2) inter-product standard deviation; (3) interquartile range; (4) range (maximum minus minimum); (5) valid fraction of finite PMW values; (6) voting fraction for

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

SIC > 0.15; (7) voting fraction for SIC > 0.80; (8) a soft ice-edge membership score, denoted by $\pi_{\mathrm{edge}} \in [0,1]^{H\times W}$; (9) gradient magnitude of the consensus median, normalised and clipped to $[0,1]$; and (10) signed distance to the 0.15 contour of the consensus median, normalised by 100 km and clipped to $[-1,1]$.

For each scene, let $M(p)$ denote the consensus median SIC across the six PMW products at pixel p. We define the soft ice-edge membership score as

$$\pi_{\mathrm{edge}}(p) = \min\left(\max\left(1 - \frac{|M(p) - 0.15|}{0.15}, 1\right), 0\right). \tag{20}$$

Thus, $\pi_{\mathrm{edge}} = 1$ where the consensus median SIC is exactly 0.15, and it decreases linearly to 0 as the consensus median moves away from the 0.15.

These derived diagnostics are especially useful for describing multi-product agreement, edge sharpness, and spatial heterogeneity. In the baseline model, they are not mandatory network inputs. Instead, they support prior construction and are evaluated as optional augmentations in ablation experiments.

3.5 Prior-informed learning objective

3.5.1 Main data-fitting term

Let the MODIS reference SIC field and the set of valid pixels be defined as in Sect. 3.1. In the heteroscedastic setting adopted here, the network estimates both a mean field and a log-variance field, and the corresponding variance is treated as the estimated variance at each pixel. We assume that, for each valid pixel, the reference SIC is conditionally Gaussian,

$$Y_{ij} \mid \mu_{ij}, \sigma_{ij}^2 \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2), (i,j) \in \Omega_{\mathrm{valid}}. \tag{21}$$

Under this assumption, the main data-fitting term is the Gaussian negative log-likelihood, averaged over valid pixels and written up to an additive constant as

$$L_{\mathrm{nll}} = \frac{1}{|\Omega_{\mathrm{valid}}|} \sum_{(i,j)\in\Omega_{\mathrm{valid}}} \left[\exp(-\tilde{s}_{ij})(\mu_{ij} - Y_{ij})^2 + \tilde{s}_{ij}\right], \tag{22}$$

Here, $\tilde{s}_{ij}$ denotes the log-variance after restricting the network output to the interval $[-4.6,1.6]$:

$$\tilde{s}_{ij} = \min\left(\max\left(s_{ij}, -4.6\right), 1.6\right). \tag{23}$$

Since $s_{ij} = \log \sigma_{ij}^2$, this means that the estimated variance is restricted to

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

$$\sigma_{ij}^2 \in [\exp(-4.6), \exp(1.6)] \approx [0.01, 4.95], \tag{24}$$

which corresponds to a standard deviation range of approximately $[0.10, 2.23]$. This restriction is used to keep training stable. The lower bound avoids unrealistically small variances, which would otherwise produce excessively large inverse-variance weights in Eq. (22). The upper bound prevents the model from assigning extremely large uncertainty to difficult pixels and thereby weakening the data-fitting term too much. In this way, the loss encourages the model to learn accurate SIC estimates and uncertainty values that increase where the residual errors are larger.

To discourage pathological variance estimates, we add a weak quadratic prior on the log-variance field,

$$L_{\mathrm{logvar}} = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} (s_{ij} - \bar{s})^2, \tag{25}$$

where $\bar{s}$ is a prescribed mean log-variance.

This heteroscedastic loss formulation allows the model to estimate both the fused SIC field and a corresponding pixel-wise uncertainty estimate, while keeping the variance branch numerically stable during training.

3.5.2 Prior-informed PMW consensus regulariser

The heteroscedastic data-fitting term anchors the fused estimate to the MODIS reference, but it does not by itself encode where individual PMW products should be trusted more or less strongly. To incorporate that information, we introduce a prior-informed PMW consensus regulariser. Let $Z_k \in [0,1]^{H\times W}$ denote the SIC field from the k-th PMW product, with $k = 1, \ldots, 6$. We define

$$L_{\mathrm{pmw}} = \frac{1}{|\Omega_{\mathrm{valid}}|} \sum_{(i,j)\in\Omega_{\mathrm{valid}}} \sum_{k=1}^{6} w_{k,ij}\, \rho_\delta\big(\max(|\mu_{ij} - Z_{k,ij}| - \varepsilon_{\mathrm{free}}, 0)\big), \tag{26}$$

where $\rho_\delta(\cdot)$ is the Huber loss with parameter δ, and $\varepsilon_{\mathrm{free}}$ specifies a free margin within which small deviations from the individual PMW products are not penalised. The pixel-wise product weights are constructed by multiplying the frozen reliability factors from Sect. 3.3 and then normalising them across products:

$$\widetilde{w}_{k,ij} = r_k^{\mathrm{land}}\big(d_{\mathrm{land},ij}\big)\, r_k^{\mathrm{edge}}\big(d_{\mathrm{edge},ij}\big)\, r_k^{\mathrm{unc}}\big(U_{k,ij}\big)\, w_k^{\mathrm{glob}}, w_{k,ij} = \frac{\widetilde{w}_{k,ij}}{\sum_{\ell=1}^{6} \widetilde{w}_{\ell,ij}}. \tag{27}$$

Here w_k^{glob} is an optional product-level trust factor, set to 1 by default. For products without uncertainty maps, $r_k^{\mathrm{unc}} \equiv 1$. This regulariser encourages the fused estimate to remain close to individual PMW products in regions where they are

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

empirically judged reliable, while still allowing the fused output to depart from any single product when the supervised signal supports such a correction.

3.5.3 Prior-weighted loss branches

In addition to the consensus regulariser, we include prior-weighted loss branches that apply prior-derived scalar maps to the squared error. These branches do not change the target labels; they simply place greater emphasis on intrinsically difficult regions. The coast-weighted supervised branch is

$$L_{\mathrm{coast}} = \frac{1}{|\Omega_{\mathrm{valid}}|} \sum_{(i,j)\in\Omega_{\mathrm{valid}}} \omega_{\mathrm{land},ij} \ (\mu_{ij} - Y_{ij})^2, \tag{28}$$

where ω_{land} is the unit-mean emphasis field derived from the fitted land-distance error relation in Sect. 3.3.4. The uncertainty-weighted supervised branch is

$$L_{\mathrm{unc}} = \frac{1}{|\Omega_{\mathrm{valid}}|} \sum_{(i,j)\in\Omega_{\mathrm{valid}}} \omega_{\mathrm{unc},ij} \ (\mu_{ij} - Y_{ij})^2, \tag{29}$$

where ω_{unc} is the unit-mean scene-level emphasis map obtained from the fitted uncertainty–error relations of the available uncertainty-bearing products, as described in Sect. 3.3.4. These branches make the supervised objective pay proportionally more attention to pixels where PMW SIC retrieval is known to be more difficult, especially near coastlines and in locations where the input products themselves report elevated uncertainty.

3.5.4 Optional edge and gradient penalties

Besides the supervised heteroscedastic term and the prior-informed branches above, we consider two optional auxiliary penalties. Let π_{edge} denote the soft ice-edge membership score introduced in Sect. 3.4.3 and defined by Eq. (20). We then define the edge set

$$\Omega_{\mathrm{edge}} = \{(i,j) \in \Omega : \pi_{\mathrm{edge},ij} \geq \tau_{\mathrm{edge}}\}. \tag{30}$$

We further define the valid edge subset as

$$\Omega_{\mathrm{edge}}^{\mathrm{valid}} = \Omega_{\mathrm{edge}} \cap \Omega_{\mathrm{valid}}. \tag{31}$$

The edge-focused loss is an RMSE evaluated only on valid edge pixels:

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

$$L_{\mathrm{edge}}=\sqrt{\frac{1}{|\Omega_{\mathrm{edge}}^{\mathrm{valid}}|}\sum_{(i,j)\in\Omega_{\mathrm{edge}}^{\mathrm{valid}}}(\mu_{ij}-Y_{ij})^2+\varepsilon}. \tag{32}$$

We also consider a gradient-matching loss relative to the consensus median field C:

$$L_{\mathrm{grad}}=\frac{1}{|\Omega_{\mathrm{valid}}|}\sum_{(i,j)\in\Omega_{\mathrm{valid}}}\left(|\partial_x\mu_{ij}-\partial_x C_{ij}|+|\partial_y\mu_{ij}-\partial_y C_{ij}|\right). \tag{33}$$

This term penalises the L_1 difference of gradient components and is intended to discourage excessive smoothing and to preserve plausible spatial transitions. Both L_{edge} and L_{grad} are optional and are used only when enabled in the configuration.

3.5.5 Full objective and coefficient setting

425 Putting all pieces together, the loss for one training pair is

$$L_{\mathrm{sup}}=L_{\mathrm{nll}}+\gamma_{\mathrm{pmw}}(t)L_{\mathrm{pmw}}+\gamma_{\mathrm{edge}}L_{\mathrm{edge}}+\gamma_{\mathrm{coast}}L_{\mathrm{coast}}+\gamma_{\mathrm{unc}}L_{\mathrm{unc}}+\gamma_{\mathrm{logvar}}L_{\mathrm{logvar}}+\gamma_{\mathrm{grad}}L_{\mathrm{grad}}, \tag{34}$$

where $\gamma_{\mathrm{pmw}}(t)$ follows a schedule with an early warm-up phase and a late-stage decay. The term $\gamma_{\mathrm{pmw}}(t)L_{\mathrm{pmw}}$ uses prior-derived reliability weights to tell the model which PMW products should be trusted more at each pixel, so that the fused estimate is preferentially kept close to the more reliable inputs. By contrast, the terms $\gamma_{\mathrm{edge}}L_{\mathrm{edge}}+\gamma_{\mathrm{coast}}L_{\mathrm{coast}}+\gamma_{\mathrm{unc}}L_{\mathrm{unc}}$ do not determine which product to trust; instead, they increase the training emphasis on difficult regions such as ice edges, coastal
430 zones, and high-uncertainty areas.

To reduce manual tuning of the prior-related branch weights, we set the relative magnitudes of the edge-, coast-, and uncertainty-related coefficients from the frozen prior-family statistics in Sect. 3.3.4:

$$(\gamma_{\mathrm{edge}},\gamma_{\mathrm{coast}},\gamma_{\mathrm{unc}})=(\alpha_{\mathrm{edge}},\alpha_{\mathrm{coast}},\alpha_{\mathrm{unc}})=(0.301,0.258,0.441). \tag{35}$$

These coefficients are fixed before test evaluation and are not tuned on the test set.

The full loss therefore combines three complementary signals: the heteroscedastic data-fitting term anchors the fused estimate
435 to the MODIS reference while learning a corresponding pixel-wise uncertainty estimate; the PMW consensus regulariser preserves useful information from the individual PMW products where they are empirically reliable; and the prior-weighted branches and optional penalties place additional emphasis on the most difficult coastal and ice-edge regions.

3.6 Training setup and reproducibility

Each sample comprises the full 512×512 analysis domain. Splits are therefore performed at the acquisition-date level to

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

avoid spatial leakage. The 156 retained MODIS scene dates are split with a fixed random seed of 1337: 20 % are held out as the final test split, and the remaining 80 % are used for model development and further split 80/20 into training and validation subsets.

Unless otherwise noted, all metrics are computed against the co-located MODIS-derived SIC on the held-out test dates and, additionally, against the Landsat-derived SIC where available (Sect. 2.4), using consistent masks and valid-pixel definitions. All loss terms and evaluation metrics are computed using fractional SIC values in [0, 1].

All candidate models are trained under a common optimisation protocol so that architecture and input ablations remain comparable. Training uses AdamW under a cosine learning-rate schedule with 5 % warm-up. The PMW consensus regulariser follows a scheduled weight with an early warm-up phase and a late-stage decay; the schedule is fixed before test evaluation and recorded with the experiment configuration. When heteroscedastic regression is enabled, the variance branch uses a five-epoch warm-up before joint optimisation with the mean branch. Early stopping monitors validation RMSE with a patience of 12 epochs, and the checkpoint with the best validation RMSE is restored at the end of training.

For the candidate backbones, we use a common implementation style: batch normalisation, a base channel width of 32, and bottleneck dropout of 0.1. Training is carried out with distributed data parallelism on three GPUs using bfloat16 mixed precision. All SIC inputs and targets are scaled consistently to [0,1], and land, pole-hole, and missing pixels are masked in all calculations.

For reproducibility, all preprocessing settings, random seeds, runtime hyperparameters, fitted prior parameters, and automatically determined prior-family coefficients are stored together with the experiment outputs. This allows the full training configuration to be reproduced as closely as possible.

4 Results

4.1 Model selection and ablation studies

Table 4 compares the candidate models under a common training protocol using RMSE on the training and validation splits. Model selection is based on validation RMSE for the Arctic, coastal, and ice-edge regions defined in Table 4.

Table 4. Model performance for architecture and input ablations. RMSE is computed against the MODIS reference dataset on the training split (“Train”) and the validation set (“Validation”). Arctic denotes all valid ocean pixels in the Arctic. The coastal mask selects pixels with $d_{\mathrm{land}} < 20$ km, where d_{land} is the distance in kilometres. The ice-edge mask selects pixels with $\pi_{\mathrm{edge}} > 0.70$, where π_{edge} is the soft ice-edge membership score defined in Sect. 3.4.3 by Eq. (20) and indicates a high probability of being at the MIZ.

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

Model	Auxiliary decoder heads	Derived PMW diagnostics	RMSE Train (global)	RMSE Validation (global)	RMSE Train ($d_{\mathrm{land}} <$ 20 km)	RMSE Validation ($d_{\mathrm{land}} <$ 20 km)	RMSE Train ($\pi_{\mathrm{edge}} >$ 0.70)	RMSE Validation ($\pi_{\mathrm{edge}} >$ 0.70)
ResUNet (2 levels)	No	No	0.0804	0.0942	0.1162	0.1323	0.1512	0.1491
ResUNet (2 levels)	Yes	No	0.0858	0.0918	0.1222	0.1316	0.1475	0.1433
ResUNet (2 levels)	Yes	Yes	0.0854	0.0915	0.1214	0.1305	0.1477	0.1434
ResUNet (3 levels)	No	No	0.0782	0.0939	0.1136	0.1332	0.1504	0.1514
ResUNet (3 levels)	Yes	No	0.0904	0.0967	0.1452	0.1494	0.1473	0.1465
ResUNet (3 levels)	Yes	Yes	0.0840	0.0917	0.1198	0.1312	0.1428	0.1442
U-Net++ (3 levels)	No	No	0.0949	0.1012	0.1492	0.1530	0.1577	0.1506
U-Net++ (3 levels)	Yes	No	0.0945	0.0991	0.1466	0.1504	0.1546	0.1457
U-Net++ (3 levels)	Yes	Yes	0.0829	0.0913	0.1182	0.1318	0.1467	0.1433

Across the three evaluation masks, we selected the two-level ResUNet with auxiliary decoder heads as the reference model because it achieved the best coastal validation RMSE, matched the best ice-edge validation RMSE, and remained close to the best global validation RMSE. Because deeper networks and additional derived inputs yield only marginal or inconsistent gains, this configuration provides the best accuracy–complexity trade-off. With optional derived PMW diagnostics, this architecture achieved the lowest coastal validation RMSE (0.1305), whereas without these diagnostics it achieved 0.1316. Because the improvement was small and not consistent across masks, we retained the simpler baseline input without the optional diagnostics for subsequent experiments. One U-Net++ configuration gave a slightly lower global validation RMSE (0.0913), but it did not improve the coastal or ice-edge metrics. Because this study focuses especially on coastal and ice-edge errors, we selected the two-level ResUNet with auxiliary decoder heads as the baseline architecture.

Comparing the two-level ResUNet without optional derived PMW diagnostics, adding auxiliary decoder heads reduced

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

validation RMSE from 0.0942 to 0.0918 for the Arctic-wide region, from 0.1323 to 0.1316 for the coastal region, and from 0.1491 to 0.1433 for the ice-edge region. This indicates that the auxiliary decoder heads, defined as [brief definition], help retain edge- and coast-related information during training.

Increasing architectural depth from two to three levels did not improve RMSE relative to the two-level ResUNet configuration. The three-level ResUNet did not improve the coastal metric, and the U-Net++ variants did not improve the best ice-edge result. Although U-Net++ reached the lowest global validation RMSE in one setting, the difference was very small and did not carry over to the two region-focused masks. We therefore did not use a deeper model in the main experiments.

For the reference two-level ResUNet with auxiliary decoder heads, adding the optional derived PMW diagnostics had only a small effect. For the two-level ResUNet with auxiliary decoder heads, these extra channels changed validation RMSE from 0.0918 to 0.0915 on the global mask, from 0.1316 to 0.1305 on the coastal mask, and from 0.1433 to 0.1434 on the ice-edge mask. Because these changes were minor and not consistently positive across all masks, we do not treat the optional derived PMW diagnostics as part of the baseline input. Accordingly, all subsequent evaluations use the two-level ResUNet with auxiliary decoder heads and the baseline input setting.

4.2 Spatial patterns and error structure

We next evaluate our fused product on the held-out test split using common metrics (Table 5) and three spatial diagnostics: the Taylor diagram (Fig. 7), spatial MAE maps (Fig. 8), and a representative clear-sky snapshot (Fig. 9). All metrics in this subsection are computed against the MODIS SIC reference dataset.

Table 5. Cross-product evaluation of SIC on the held-out test split, using the MODIS SIC as the reference. ResUNet denotes the fused SIC product produced by the selected two-level ResUNet. The metrics include MAE, RMSE, mean bias, and coefficient of determination (R^2). Lower values indicate better performance for MAE, RMSE, and absolute bias, whereas higher values indicate better performance for R^2. The coastal and ice-edge regions are reported as RMSE over region-specific masks: $d_{\mathrm{land}} < 20$ km denotes pixels whose distance to land is less than 20 km; d_{land} is 20–60 km denotes pixels located 20–60 km from the nearest coastline; π_{edge} is the soft ice-edge membership score defined in Sect. 3.4.3 by Eq. (20); and edge $d_{\mathrm{edge}} \leq 50$ km denotes pixels located within 50 km ice edge. This table summarises the overall and region-specific performance of the fused product and the six PMW SIC products.

Product	MAE	RMSE	Bias	R^2	RMSE (d_{land}<20 km)	RMSE (d_{land} is 20–60 km)	RMSE (π_{edge}> 0.70)	RMSE ($d_{\mathrm{edge}}\leq$ 50 km)
ResUNet	0.0429	0.0917	-0.0028	0.9451	0.1364	0.1228	0.1419	0.1577
NSIDC-BT	0.1124	0.1570	0.0229	0.8392	0.2842	0.1957	0.2472	0.2586
NSIDC-NT	0.1018	0.1379	-0.0490	0.8759	0.2101	0.1674	0.1837	0.2136
NSIDC-CDR	0.1101	0.1538	0.0158	0.8456	0.2667	0.1915	0.2382	0.2529

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

OSI-450	0.0957	0.1315	-0.0324	0.8871	0.2122	0.1667	0.1860	0.2088
SICCI-25km	0.0960	0.1364	-0.0441	0.8787	0.2055	0.1821	0.2108	0.2215
Bremen-ASI	0.1130	0.1626	-0.0034	0.8276	0.2285	0.2011	0.2298	0.2546

On the pan-Arctic scale, our fused product performs best among all products in Table 5. It has the lowest MAE (0.0429) and RMSE (0.0917), the highest R^2 (0.9451), and a near-zero mean bias (-0.0028). Relative to OSI-450, which is the best PMW SIC product in terms of MAE and RMSE, MAE is reduced by about 55 % and RMSE by about 30 %.

The region-specific comparisons in Table 5 show that the gains are strongest in the coastal and ice-edge regions. Our fused product reduces RMSE to 0.1364 for $d_{\text{land}} < 20$ km and 0.1228 for d_{land} in 20–60 km, improving on the best product by about 34 % and 26 %, respectively. For the two ice-edge diagnostics, RMSE is 0.1419 for $\pi_{\text{edge}} > 0.70$ and 0.1577 for $d_{\text{edge}} \leq 50$ km, corresponding to improvements of about 23 % and 25 % over the best product. These results show that the main gains occur in the regions where PMW SIC retrievals are most difficult.

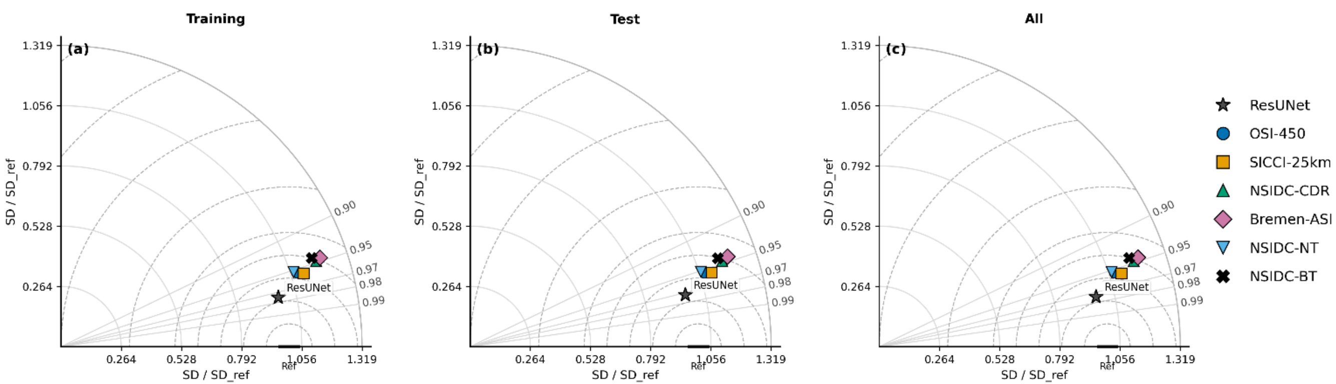


Figure 7. Taylor diagrams comparing the fused ResUNet product and the six PMW SIC products against the MODIS-derived SIC reference. Panels show (a) training, (b) test, and (c) all samples. Radial distance indicates the normalised standard deviation (SD/SD_ref), azimuth denotes correlation, and dashed arcs indicate centred RMSE. The fused ResUNet product lies closest to the ideal reference point in the test panel, indicating the best combined agreement in correlation and variance.

The Taylor diagram in Fig. 7 is consistent with these scalar metrics. In the test panel, our fused product lies closest to the ideal point, indicating both higher correlation with the reference and variance closer to the reference field. This shows that the improvement is not obtained by simply smoothing the field; spatial agreement improves while realistic variability is retained.

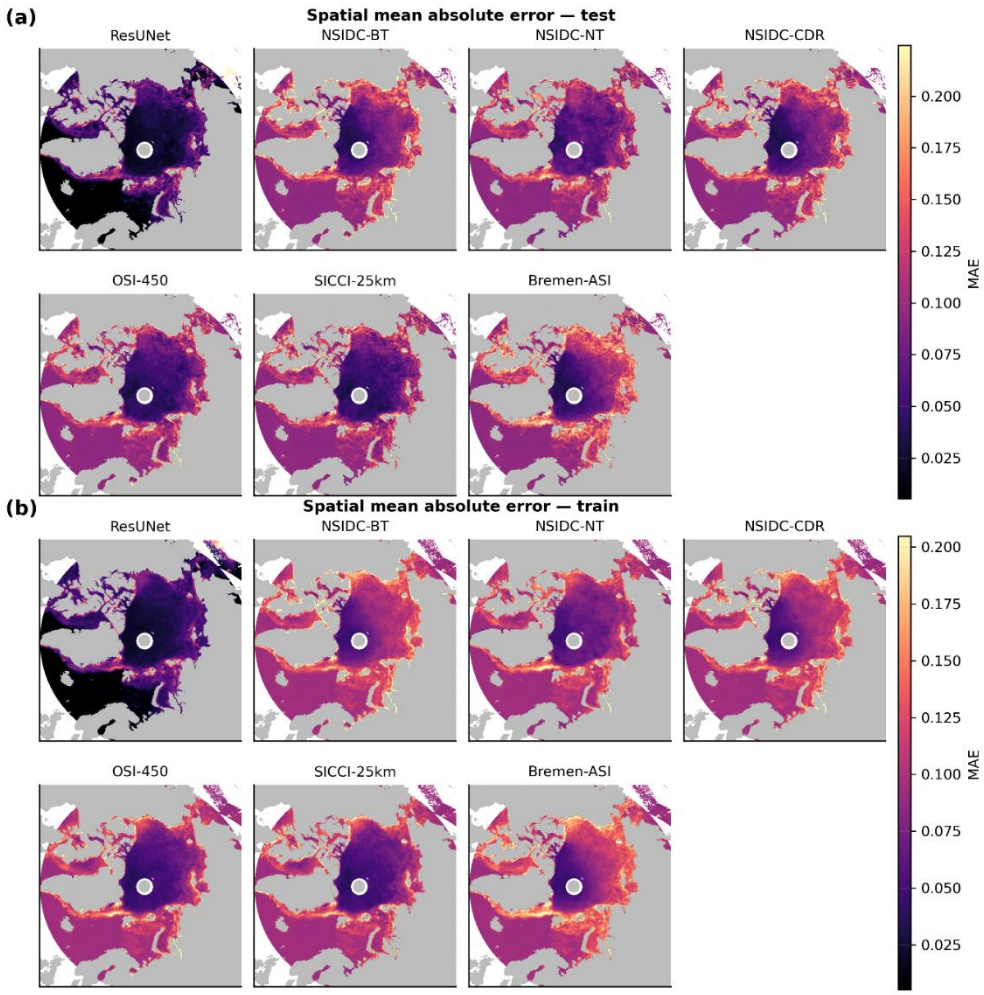


Figure 8. Spatial distribution of MAE in SIC for (a) the held-out test split and (b) the training split, computed relative to the co-located MODIS-derived SIC reference. Panels show, from left to right, the fused product (labelled "ResUNet" in the figure) and the six PMW SIC
520 products: NSIDC-BT, NSIDC-NT, NSIDC-CDR, OSI-450, SICCI-25km, and Bremen-ASI. For each product, the plotted field is the per-pixel mean absolute error averaged over all valid samples in the corresponding split. Land and the pole-hole region are shown in grey. The fused product exhibits broadly reduced errors relative to the products, with the largest improvements concentrated along the MIZ and in coastal sectors, while the interior pack remains uniformly low error in both splits.

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

The spatial MAE maps in Fig. 8 show where these improvements occur. In the test split, the largest reductions appear along the MIZ and in coastal regions such as Fram Strait, the Barents–Kara sector, Baffin Bay, and the Chukchi–Beaufort seas. The high-MAE belt near the ice edge is clearly weakened, while the interior pack remains uniformly low-error. The training split shows a very similar pattern, indicating that this spatial behaviour is stable rather than specific to one split.

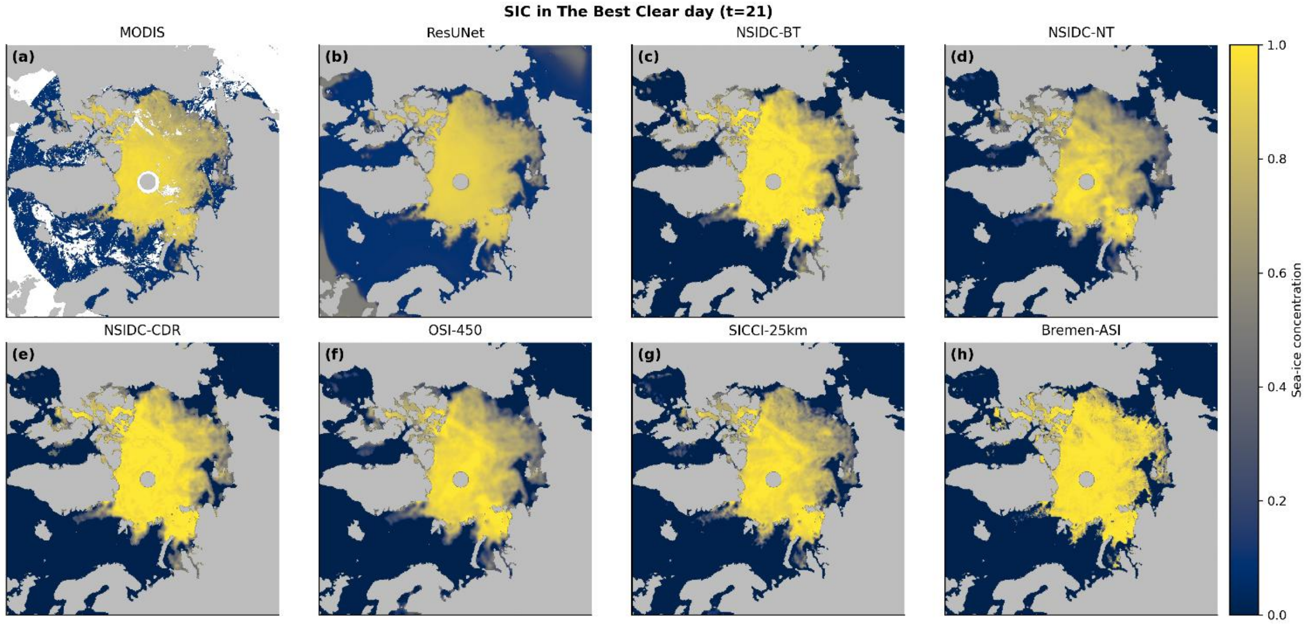


Figure 9. Representative clear-sky day from the held-out test split. Panel (a) shows the MODIS-derived SIC reference; panel (b) shows the fused ResUNet product; and panels (c)–(h) show the six PMW SIC products. SIC is shown as a fraction from 0 to 1; land and the pole-hole region are shown in grey. Compared with the products, our fused product better reproduces the curvature and sharpness of the ice edge and shows reduced coastal contamination.

The representative clear-sky example in Fig. 9 shows the same result for a specific day. Our fused product follows the reference ice-edge geometry and SIC gradients more closely than the PMW SIC products. Several products either retreat the ice edge or extend it too far, whereas our fused product better preserves the curvature of the ice edge and shows less coastal contamination. These scene-scale and basin-scale diagnostics consistently show that the fused product improves both overall accuracy and spatial structure.

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

4.3 Reductions in coastal and ice-edge errors and calibrated uncertainty

When MAE is examined as a function of distance, the fused product shows its largest improvements in the two most difficult regions for PMW SIC retrievals: near coastlines and near the ice edge. Figure 10a presents the corresponding analysis for distance to land. Our fused product yields the smallest near-coastal MAE, and its coastal MAE decreases more rapidly with offshore distance than that of the six PMW products. The training, test, and all-data curves are very similar, indicating that the distance-dependent improvement is stable across data splits. Figure 10b shows the same analysis for distance to the ice edge. Our fused product exhibits a sharp MAE reduction near the ice edge and remains lower than all six PMW SIC products over the distance range shown on the held-out test split. Within the first few tens of kilometres, our fused product reduces MAE to approximately half of the mean MAE across the six products. Beyond about 200 km from the ice edge, the MAE of the fused product becomes nearly constant at a low level.

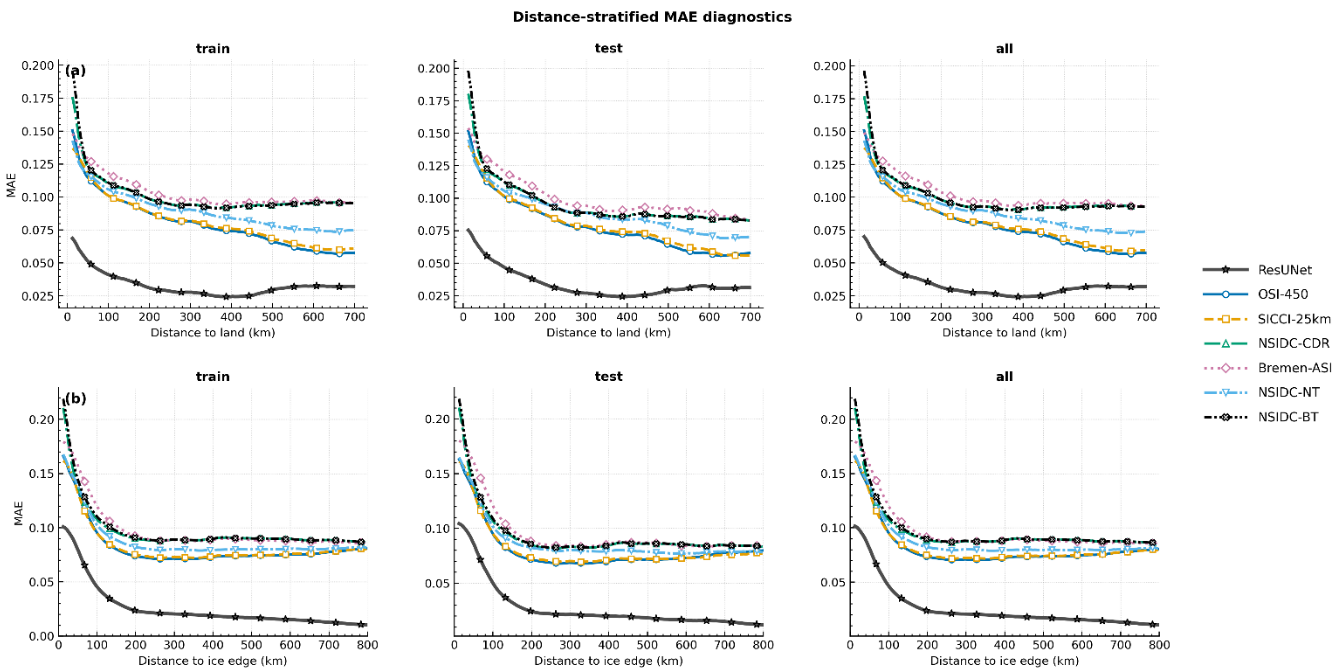


Figure 10. Distance-stratified MAE diagnostics for the fused ResUNet product and the six PMW SIC products, shown for the training, test, and combined samples from left to right. (a) MAE as a function of distance to land. The fused product has the lowest near-coastal MAE and shows the strongest reduction with offshore distance. (b) MAE as a function of distance to the ice edge, defined by the 0.15 SIC contour. The fused product remains below the six PMW products over the distance range shown. Similar train, test, and all-sample curves indicate that the distance-dependent error reduction is stable across data splits.

The estimated pixel-wise uncertainty σ, represented by the estimated standard deviation, tracks the observed errors well. In all three splits, bins with larger uncertainty have larger mean absolute error and a broader error spread (Fig. 11). Figure 12

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

shows the relationship between estimated uncertainty and absolute SIC error; most points cluster on or just below the 1:1 line, indicating that the estimated uncertainty is usually similar to, or slightly larger than, the observed error.

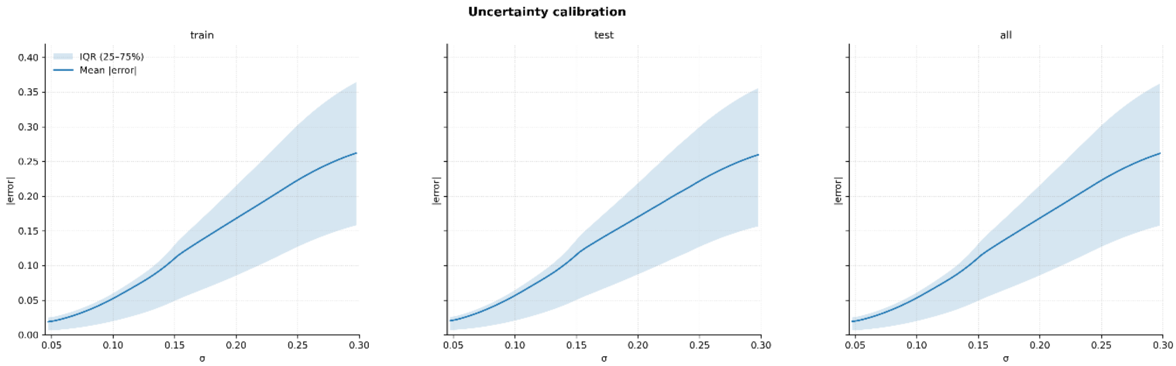


Figure 11. Uncertainty and mean absolute error. The solid line shows the conditional mean absolute error in bins of the estimated uncertainty σ; the shaded band denotes the interquartile range (25–75%). MAE increases almost linearly with σ, and the error spread also widens as σ increases. Thus, larger estimated uncertainty is associated with both larger and more variable errors.

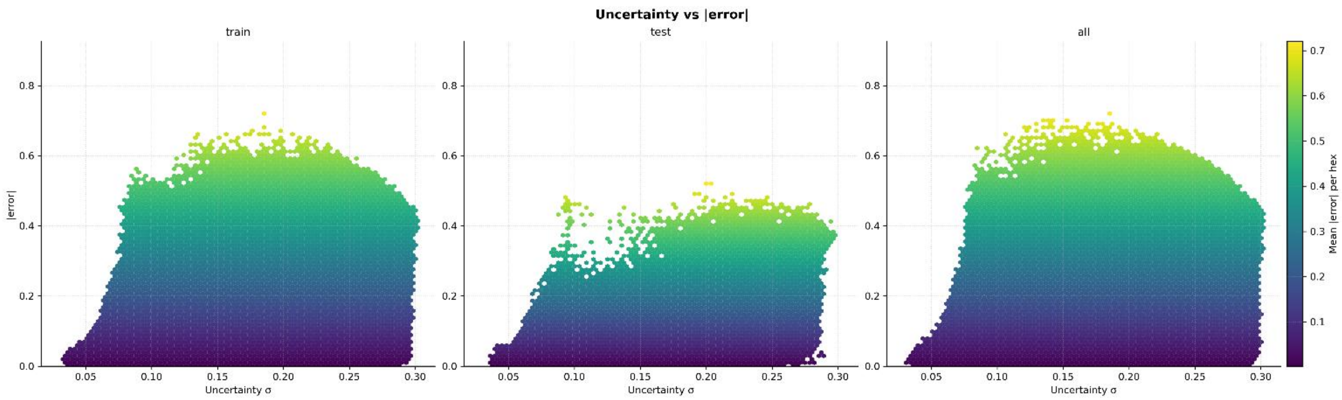


Figure 12. Relationship between estimated uncertainty σ and absolute SIC error for the training, test, and combined samples. In all three panels, larger estimated uncertainty is associated with a wider and generally larger error range. This shows that the estimated uncertainty provides useful information about the likely error size in both the training and test data, although the errors still vary substantially at intermediate uncertainty levels.

Spatially, σ is highest along the ice edge and near land. Additional local maxima appear in export regions such as Fram Strait, while values are low over the central pack (Fig. 13). This spatial pattern matches the error pattern in Sect. 4.2 and Fig. 8: uncertainty is high in the same regions where MAE is high. In other words, the uncertainty map identifies the areas where retrieval errors are most likely to be large.

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

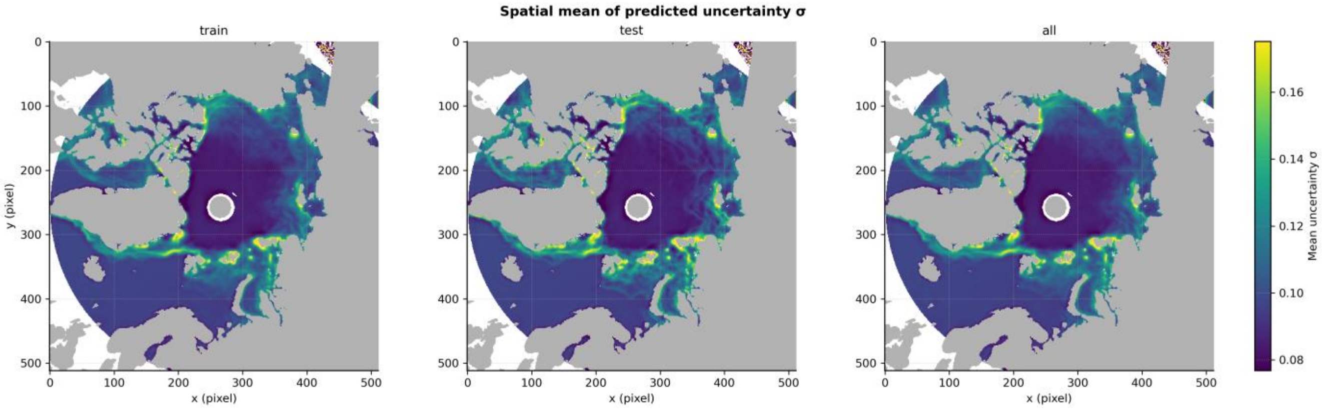


Figure 13. Spatial pattern of the mean estimated uncertainty for the fused SIC product in the training, test, and combined samples from the MODIS reference dataset. The uncertainty shown is the estimated standard deviation from the heteroscedastic ResUNet, averaged at each grid cell over all available samples and expressed in SIC. The training and test panels show broadly consistent spatial patterns, indicating that the estimated uncertainty field is spatially stable across data splits. Land and masked regions are shown in grey or white.

4.4 Landsat-derived evaluation of SIC retrievals

Using the Landsat-derived SIC dataset as an independent reference (Sect. 2.4), we further evaluated our fused product and the six PMW SIC products (Table 6; Fig. 14). As noted in Sect. 2.4, the Landsat matchups are sparse, dominated by spring conditions, and distributed mainly in coastal regions. This validation should therefore be interpreted as an independent spring-dominated benchmark with strong coastal emphasis, rather than as a year-round Arctic assessment.

Table 6. Cross-product evaluation of SIC against the Landsat-derived SIC dataset. The Landsat-derived SIC dataset described in Sect. 2.4 is used as an external reference. ResUNet denotes our fused product. The metrics are defined in the same way as those in Table 5.

Product	MAE	RMSE	Bias	R^2	RMSE (d_{land}< 20 km)	RMSE (d_{land} is 20–60 km)	RMSE (π_{edge} > 0.70)	RMSE (d_{edge}≤ 50 km)
ResUNet	0.035	0.062	-0.020	0.840	0.1155	0.0831	0.1079	0.0709
NSIDC-BT	0.042	0.104	0.0135	0.547	0.1729	0.1546	0.1841	0.1235
NSIDC-NT	0.086	0.134	-0.063	0.256	0.1554	0.1526	0.1247	0.1441
NSIDC-CDR	0.040	0.104	0.0144	0.553	0.1728	0.1533	0.183	0.1228
OSI-450	0.069	0.112	-0.041	0.483	0.148	0.1422	0.1362	0.1248
SICCI-25km	0.064	0.105	-0.045	0.54	0.1399	0.1307	0.1328	0.1172
Bremen-ASI	0.060	0.127	-0.05	0.327	0.1781	0.1471	0.1599	0.1378

Table 6 shows that our fused product achieves the best overall agreement with the Landsat-derived SIC dataset. It yields the lowest MAE (0.035) and RMSE (0.062), the highest R^2 (0.840), and a comparatively small negative bias (-0.020). Among the individual PMW products, NSIDC-CDR provides the lowest overall MAE and RMSE (0.040 and 0.104, respectively), but our fused product still improves upon these values by about 13 % in MAE and about 40 % in RMSE. The best single product is not the same for all diagnostics: NSIDC-CDR performs best for Arctic-region MAE and RMSE, SICCI-25km performs best in the two coastal bands and in the 50 km ice-edge diagnostic, and NSIDC-NT performs best for the high-confidence ice-edge mask.$d_{\mathrm{edge}} \leq 50 \pi_{\mathrm{edge}} > 0.70$ Even relative to these strongest individual comparators, however, the fused product remains consistently superior.

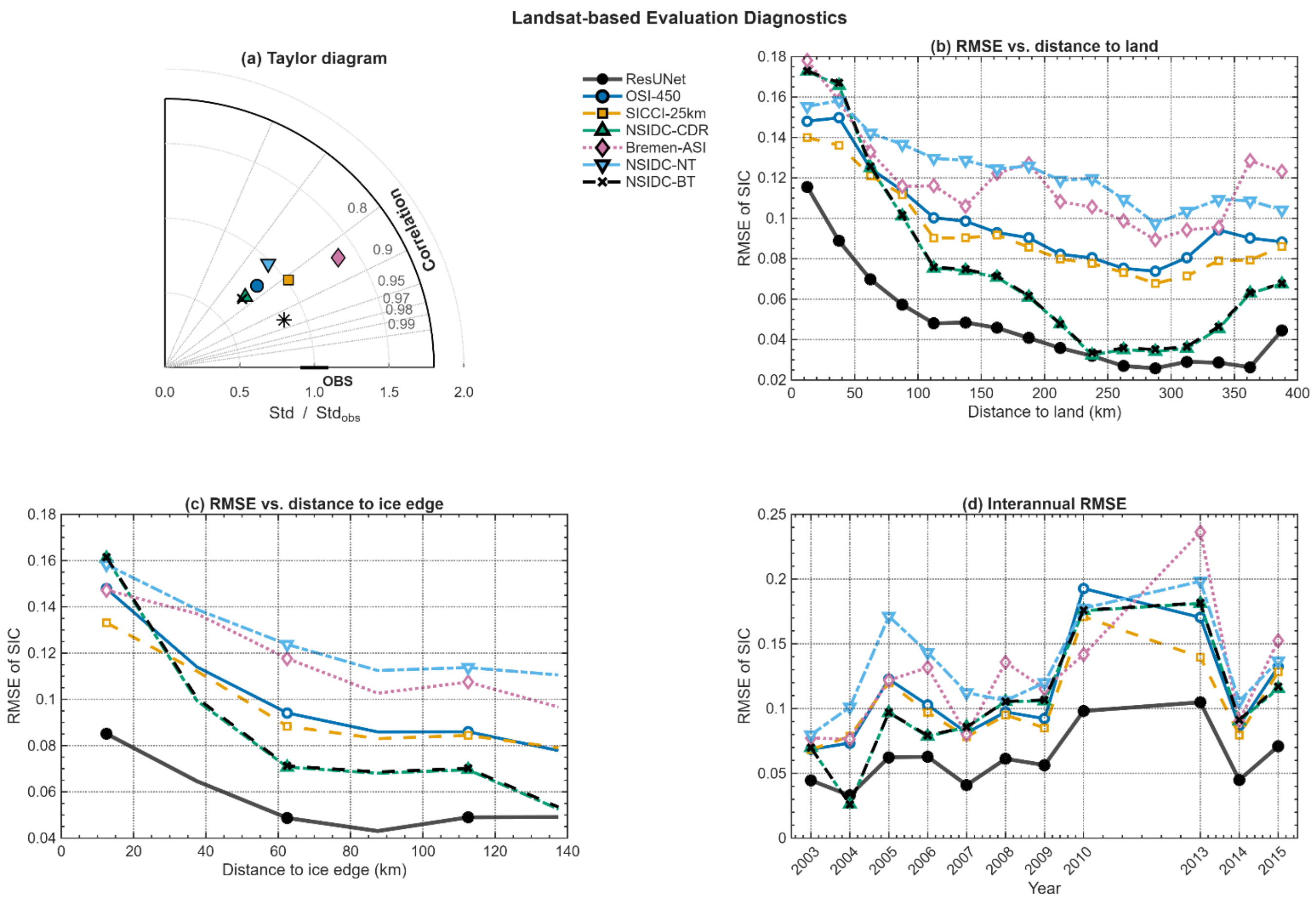


Figure 14. Evaluation against the independent Landsat-derived SIC dataset for the fused ResUNet product and the six PMW SIC products. (a) Taylor diagram summarising correlation, normalised standard deviation, and centred RMSE relative to Landsat-derived SIC. (b) RMSE as a function of distance to land. (c) RMSE as a function of distance to the ice edge over the 0–140 km range shown. (d) Interannual RMSE relative to Landsat-derived SIC. Products are identified by a consistent combination of colour, line style, and marker. The fused product shows the strongest overall agreement with the Landsat-derived reference and generally lower RMSE than the individual PMW SIC products.

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

This overall advantage is also evident in the Taylor diagram (Fig. 14a). The fused product lies closest to the Landsat reference point, indicating the best combined performance in terms of correlation, variance reproduction, and centred RMSE. At the same time, its standard deviation remains somewhat smaller than that of the Landsat reference, suggesting a degree of variance damping. Nevertheless, this moderate variance underestimation is clearly outweighed by the substantial gains in overall fidelity. By contrast, the PMW products lie farther from the reference point, reflecting lower correlation and less favourable variance characteristics.

The Landsat evaluation shows that the improvement is robust across the main error-prone regions sampled by the validation dataset. As a function of distance to land (Fig. 14b), the fused product has the lowest RMSE across essentially the full distance range shown. Its error decreases rapidly from about 11 %–12 % in the nearest coastal bins to below about 5 % by roughly 100–150 km offshore, remains near 3 %–4 % over much of the interior range, and increases only slightly in the outermost bins. By contrast, all six PMW products remain systematically higher and retain a much larger coastal error. This behaviour is reflected in the region-specific statistics in Table 6. Within 20 km of the coast, the fused product reduces RMSE by about 17 % relative to SICCI-25km, the best-performing individual product for this diagnostic. In the 20–60 km coastal band, the fused product attains RMSE = 0.0831, improving upon the best value (0.1307, again from SICCI-25km) by about 36 %. These results indicate that the fusion substantially suppresses the coastal errors that remain evident in all individual PMW retrievals.

A similarly clear separation is found for the ice-edge diagnostics (Fig. 14c). Because the valid Landsat matchups for this diagnostic are confined to absolute distances of approximately 0–140 km from the 15 % ice-edge contour, the comparison is restricted to this available distance range. Within this range, the fused product again yields the lowest RMSE. Its RMSE decreases from about 8 %–9 % in the nearest bin to about 4 %–5 % beyond roughly 60–90 km from the ice edge, and then remains low with only weak variation farther from the edge. The PMW products also improve with increasing distance from the ice edge, but they remain systematically higher over the available 0–140 km range, especially in the immediate edge zone. The comparison for the ice-edge region in Table 6 confirms this advantage. For the high-confidence edge mask ($\pi_{\mathrm{edge}} > 0.70$), the fused product achieves RMSE = 0.1079, compared with 0.1247 for NSIDC-NT, the strongest product for this diagnostic, corresponding to an improvement of about 13 %. For the $d_{\mathrm{edge}} \leq 50$ km mask, the fused product attains RMSE = 0.0709, compared with 0.1172 for SICCI-25km, i.e. an improvement of about 40 %. Thus, the largest gains occur in the narrow ice-edge region, where conventional PMW SIC retrievals are most affected by mixed pixels and strong sub-footprint gradients.

The fused product has the lowest RMSE in every evaluated year and generally remains in the range of about 3 %–10 %, whereas the PMW products exhibit substantially larger year-to-year excursions, with pronounced peaks around 2010 and 2013 that reach roughly 17 %–24 %, depending on the product. The fusion therefore improves not only the general agreement with Landsat, but also the temporal stability of that agreement.

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

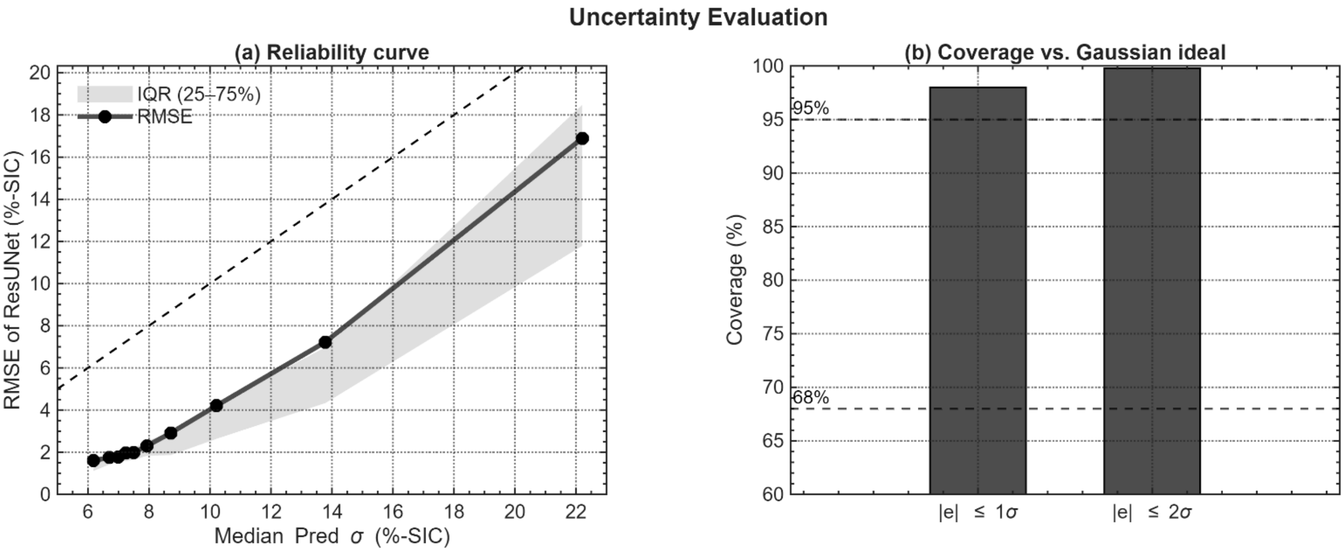


Figure 15. Evaluation of the estimated uncertainty σ from the heteroscedastic ResUNet using the independent Landsat-derived SIC dataset. (a) Reliability curve showing realised RMSE of the fused product as a function of the median estimated uncertainty within uncertainty bins. The shaded region denotes the interquartile range of the absolute error, and the dashed line indicates the 1:1 reference. (b) Empirical coverage of one- and two-standard-deviation uncertainty intervals, with dashed horizontal lines marking the Gaussian reference values of 68 % and 95 %. The monotonic increase in RMSE with σ indicates that the uncertainty estimate is informative, while the high empirical coverage indicates conservative uncertainty estimates.

The reliability curve (Fig. 15a) shows that realised RMSE increases monotonically with estimated uncertainty, indicating that the uncertainty estimate remains informative about error magnitude. RMSE rises from about 1 %–2 % in the lowest-uncertainty bins to about 17 % in the highest-uncertainty bin, and the interquartile range also broadens with uncertainty. Because the curve remains below the 1:1 line, the estimated uncertainty is conservative in absolute scale. The coverage diagnostics in Fig. 15b are consistent with this interpretation: about 97 % of matchups fall within the one-standard-deviation interval, and essentially all matchups fall within the two-standard-deviation interval.

Taken together, the Landsat-based evaluation confirms that the fused product retains its advantage when compared against an independent reference and that its estimated uncertainty remains informative and conservative.

5 Conclusions

In this study, we developed a prior-regularised heteroscedastic two-level ResUNet to fuse six PMW SIC products into a single 12.5 km Arctic SIC product. The fused product outperforms all six individual PMW products in both MODIS-based and Landsat-based evaluations. On the held-out MODIS test set, it reduces MAE by about 55 % and RMSE by about 30 % relative to the best individual PMW product, while maintaining near-zero bias. The largest gains occur in the regions where conventional PMW SIC retrievals are most difficult, with RMSE reductions of about 34 % within 20 km of the coast and about

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

25 % within 50 km of the ice edge. These comparisons show that the fusion improves not only Arctic-wide accuracy but also the accuracy in regions that matter most in practice. The Landsat evaluation confirms that these gains are not limited to the MODIS-based framework and remain consistent in an independent validation dataset. The model also provides a pixel-wise uncertainty estimate that is informative about realised error and therefore useful for quality control and uncertainty-aware applications. The main methodological contribution of this work is the combination of three elements within one PMW-only framework: multi-product PMW fusion, prior-informed training based on empirically fitted error structures, and pixel-wise uncertainty estimation. Rather than relying on intermittently available SAR or optical observations, the framework preserves the intrinsic advantages of PMW observations, namely pan-Arctic, all-weather, and near-real-time coverage, while still reducing the long-standing errors near coastlines and the ice edge. This makes the fused product a practical baseline for climate studies, operational ice monitoring, and future multi-sensor extensions.

The present study has several limitations. The MODIS reference used for training is an 8-day composite and is available only from early May to mid-September, so the fusion experiments are most representative of melt-season conditions rather than of the full annual cycle. The independent Landsat validation is useful but sparse, and it is weighted toward springtime and coastal regions, so it should be interpreted as an independent seasonal benchmark with stronger weighting toward coastal performance rather than as a uniform year-round Arctic assessment. In addition, the Landsat-based Taylor analysis suggests some remaining variance damping in the fused product. Future work should extend the evaluation to a broader seasonal range, test generalisation under winter and transition-season conditions, and explore how SAR, optical, or reanalysis information can be added to the present PMW-only product without sacrificing coverage or stability.

Overall, this study shows that mature PMW SIC products contain substantial complementary information that can be combined effectively when the fusion model is guided by the known spatial structure of PMW retrieval error. The resulting fused SIC product is more accurate, has lower errors in the most difficult coastal and ice-edge regions, and is accompanied by an uncertainty estimate. These properties make it a promising baseline for improved Arctic sea-ice monitoring and for downstream applications that require both wide coverage and quantified confidence in the retrieval.

Code and data availability. The code and data required to reproduce the main results of this study are organized in two reproducibility archives, PMW_SIC_fusion_code_for_review and PMW_SIC_fusion_data_for_review. During peer review, these archives are made available to the editor and reviewers through the journal submission system or reviewer-only access. Upon acceptance, the reproducibility archives will be deposited in a permanent public repository and made openly available with persistent identifiers.

The code archive contains the MATLAB preprocessing scripts used to collocate passive-microwave, MODIS-derived, and Landsat-derived sea-ice concentration data to the target grid; the Python scripts for model training and inference; the scripts used to fit the distance-to-land, distance-to-ice-edge, and uncertainty-based prior functions; and the basic evaluation script

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

used to summarize the inference outputs. The archive also includes example commands, environment files, README files, and a code manifest documenting the workflow.

The data archive contains the preprocessed PMW - MODIS collocation files used for model development, the target-grid files, the processed Landsat-derived sea-ice concentration files already aggregated to the MODIS grid, the fitted-prior outputs, the inference HDF5 files, and the basic evaluation outputs. Raw Landsat files are not redistributed because of their large file size; instead, the processed Landsat-derived validation files used in this study are included. The original third-party satellite products should be obtained from their respective data providers, as documented in the data package.

Only the code and processed data necessary to reproduce the analyses and results reported in this paper are included in the reproducibility archives. Additional auxiliary, exploratory, or project-internal scripts that were not used to generate the reported results are not part of the planned public release. The complete internal project code can be made available by the corresponding author upon reasonable request, where feasible and subject to institutional and third-party data-use restrictions.

Financial support. This research was funded by the National Natural Science Foundation of China (grant no. 42376200).

Acknowledgements. The authors acknowledge the use of the High-Performance Computing Center at the Institute of Oceanology, Chinese Academy of Sciences (IOCAS), and thank the teachers and students of the Deep-Sea Dynamics group at the Key Laboratory of Ocean Circulation and Waves, IOCAS, for their helpful guidance and discussions.

References

Aghdami-Nia, M., Shah-Hosseini, R., Rostami, A., and Homayouni, S.: Automatic coastline extraction through enhanced sea-land segmentation by modifying Standard U-Net, International Journal of Applied Earth Observation and Geoinformation, 109, 102785, https://doi.org/10.1016/j.jag.2022.102785, 2022.

Banzon, V., Smith, T. M., Chin, T. M., Liu, C., and Hankins, W.: A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies, Earth Syst. Sci. Data, 8, 165–176, https://doi.org/10.5194/essd-8-165-2016, 2016.

Cavalieri, D. J.: NASA sea ice validation program for the defense meteorological satellite program special sensor microwave imager, J. Geophys. Res.: Oceans, 96, 21969–21970, https://doi.org/10.1029/91JC02333, 1991.

Chen, J., Kang, S., Du, W., Guo, J., Xu, M., Zhang, Y., Zhong, X., Zhang, W., and Chen, J.: Perspectives on future sea ice and navigability in the arctic, The Cryosphere, 15, 5473–5482, https://doi.org/10.5194/tc-15-5473-2021, 2021.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, in: Computer Vision – ECCV 2018, 833–851, https://doi.org/10.1007/978-3-030-01234-2_49, 2018.

Comiso, J.: Bootstrap sea ice concentrations from nimbus-7 SMMR and DMSP SSM/I-SSMIS, version 4, https://doi.org/10.5067/X5LG68MH013O, 2023.

Comiso, J. C., Cavalieri, D. J., Parkinson, C. L., and Gloersen, P.: Passive microwave algorithms for sea ice concentration: A comparison of two techniques, Remote Sensing of Environment, 60, 357–384, https://doi.org/10.1016/S0034-4257(96)00220-9, 1997.

Druckenmiller, M. L.: NOAA Arctic Report Card 2022: Executive Summary, Global Ocean Monitoring and Observing (U.S.), 2022.

Gloersen, P. and Cavalieri, D. J.: Reduction of weather effects in the calculation of sea ice concentration from microwave radiances, Journal of Geophysical Research: Oceans, 91, 3913–3919, https://doi.org/10.1029/JC091iC03p03913, 1986.

Hoffman, J. P., Ackerman, S. A., Liu, Y., Key, J. R., and McConnell, I. L.: Application of a Convolutional Neural Network for the Detection of Sea Ice Leads, Remote Sensing, 13, 4571, https://doi.org/10.3390/rs13224571, 2021.

Hoffman, J. P., Ackerman, S. A., Liu, Y., and Key, J. R.: A 20-Year Climatology of Sea Ice Leads Detected in Infrared Satellite Imagery Using a Convolutional Neural Network, Remote Sensing, 14, 5763, https://doi.org/10.3390/rs14225763, 2022.

Huang, W., Yu, A., Xu, Q., Sun, Q., Guo, W., Ji, S., Wen, B., and Qiu, C.: Sea ice extraction via remote sensing imagery: algorithms, datasets, applications and challenges, Remote Sens., 16, 842, https://doi.org/10.3390/rs16050842, 2024.

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nat Methods, 18, 203–211, https://doi.org/10.1038/s41592-020-01008-z, 2021.

Ivanova, N., Pedersen, L. T., Tonboe, R. T., Kern, S., Heygster, G., Lavergne, T., Sørensen, A., Saldo, R., Dybkjær, G., Brucker, L., and Shokr, M.: Inter-comparison and evaluation of sea ice algorithms: towards further identification of challenges and optimal approach using passive microwave observations, The Cryosphere, 9, 1797–1817, https://doi.org/10.5194/tc-9-1797-2015, 2015.

Kern, S.: Landsat surface type over water from supervised classification of surface broadband albedo estimates (2021_fv0.01), https://doi.org/10.25592/uhhfdm.9181, 2021.

Kern, S., Rösel, A., Pedersen, L. T., Ivanova, N., Saldo, R., and Tonboe, R. T.: The impact of melt ponds on summertime microwave brightness temperatures and sea-ice concentrations, The Cryosphere, 10, 2217–2239, https://doi.org/10.5194/tc-10-2217-2016, 2016.

Kern, S., Lavergne, T., Notz, D., Pedersen, L. T., and Tonboe, R.: Satellite passive microwave sea-ice concentration data set inter-comparison for arctic summer conditions, The Cryosphere, 14, 2469–2493, https://doi.org/10.5194/tc-14-2469-2020, 2020.

Kern, S., Lavergne, T., Pedersen, L. T., Tonboe, R. T., Bell, L., Meyer, M., and Zeigermann, L.: Satellite passive microwave sea-ice concentration data set intercomparison using landsat data, The Cryosphere, 16, 349–378, https://doi.org/10.5194/tc-16-349-2022, 2022.

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.

Li, W., Liu, L., and Zhang, J.: Fusion of SAR and optical image for sea ice extraction, J. Ocean Univ. China, 20, 1440–1450, https://doi.org/10.1007/s11802-021-4824-y, 2021.

Lu, J., Scarlat, R., Heygster, G., and Spreen, G.: Reducing Weather Influences on an 89 GHz Sea Ice Concentration Algorithm in the Arctic Using Retrievals From an Optimal Estimation Method, Journal of Geophysical Research: Oceans, 127, e2019JC015912, https://doi.org/10.1029/2019JC015912, 2022.

Ludwig, V., Spreen, G., and Pedersen, L. T.: Evaluation of a new merged sea-ice concentration dataset at 1 km resolution from thermal infrared and passive microwave satellite data in the arctic, Remote Sens., 12, 3183, https://doi.org/10.3390/rs12193183, 2020.

Malmgren-Hansen, D., Pedersen, L. T., Nielsen, A. A., Kreiner, M. B., Saldo, R., Skriver, H., Lavelle, J., Buus-Hinkler, J., and Krane, K. H.: A convolutional neural network architecture for sentinel-1 and AMSR2 data fusion, IEEE Trans. Geosci. Remote Sens., 59, 1890–1902, https://doi.org/10.1109/TGRS.2020.3004539, 2021.

Meier, W.: Comparison of passive microwave ice concentration algorithm retrievals with AVHRR imagery in arctic peripheral seas, IEEE Trans. Geosci. Remote Sens., 43, 1324–1337, https://doi.org/10.1109/TGRS.2005.846151, 2005.

Meier, W. N., Peng, G., Scott, D. J., and Savoie, M. H.: Verification of a new NOAA/NSIDC passive microwave sea-ice concentration climate record, Polar Research, 33, 21004, https://doi.org/10.3402/polar.v33.21004, 2014.

Meier, W. N., Fetterer, F., Savoie, M., Mallory, S., Duerr, R., and Stroeve, J.: NOAA/NSIDC climate data record of passive microwave sea ice concentration, version 3 (3), https://doi.org/10.7265/N59P2ZTG, 2017.

Melsheimer, C. and Spreen, G.: AMSR2 ASI sea ice concentration data, arctic, version 5.4 (NetCDF) (july 2012 - december 2019), https://doi.org/10.1594/PANGAEA.898399, 2019.

Melsheimer, C. and Spreen, G.: AMSR-E ASI sea ice concentration data, arctic, version 5.4 (NetCDF) (june 2002 - september 2011), https://doi.org/10.1594/PANGAEA.919777, 2020.

Murashkin, D., Spreen, G., and Huntemann, M.: Pan-Arctic lead maps based on Sentinel-1 SAR images with a convolutional neural network, Preprints.org, https://doi.org/10.20944/preprints202409.0905.v1, 2024.

Peng, G., Meier, W. N., Scott, D. J., and Savoie, M. H.: A long-term and reproducible passive microwave sea ice concentration data record for climate studies and monitoring, Earth Syst. Sci. Data, 5, 311–318, https://doi.org/10.5194/essd-5-311-2013, 2013.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-assisted Intervention – MICCAI 2015, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Rösel, A., Kaleschke, L., and Kern, S.: Gridded melt pond cover fraction on arctic sea ice derived from TERRA-MODIS 8-day composite reflectance data bias corrected version 02, https://doi.org/10.1594/WDCC/MODIS__Arctic__MPF_V02, 2015.

Sandven, S., Spreen, G., Heygster, G., Girard-Ardhuin, F., Farrell, S. L., Dierking, W., and Allard, R. A.: Sea Ice Remote Sensing—Recent Developments in Methods and Climate Data Sets, Surv Geophys, 44, 1653–1689, https://doi.org/10.1007/s10712-023-09781-0, 2023.

https://doi.org/10.5194/egusphere-2026-3101
Preprint. Discussion started: 25 August 2026

Seale, C., Redfern, T., Chatfield, P., Luo, C., and Dempsey, K.: Coastline detection in satellite imagery: A deep learning approach on new benchmark data, Remote Sensing of Environment, 278, 113044, https://doi.org/10.1016/j.rse.2022.113044, 2022.

Smith, L. C. and Stephenson, S. R.: New Trans-Arctic shipping routes navigable by midcentury, Proceedings of the National Academy of Sciences, 110, E1191–E1195, https://doi.org/10.1073/pnas.1214212110, 2013.

Spreen, G., Kaleschke, L., and Heygster, G.: Sea ice remote sensing using AMSR-E 89-GHz channels, J. Geophys. Res.: Oceans, 113, https://doi.org/10.1029/2005JC003384, 2008.

Sun, K., Xiao, B., Liu, D., and Wang, J.: Deep High-Resolution Representation Learning for Human Pose Estimation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5686–5696, https://doi.org/10.1109/CVPR.2019.00584, 2019.

Wang, R., Ma, L., He, G., Johnson, B. A., Yan, Z., Chang, M., and Liang, Y.: Transformers for Remote Sensing: A Systematic Review and Analysis, Sensors, 24, 3495, https://doi.org/10.3390/s24113495, 2024.

Wulf, T., Buus-Hinkler, J., Singha, S., Shi, H., and Kreiner, M. B.: Pan-Arctic sea ice concentration from SAR and passive microwave, The Cryosphere, 18, 5277–5300, https://doi.org/10.5194/tc-18-5277-2024, 2024.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P.: SegFormer: simple and efficient design for semantic segmentation with transformers, in: Advances in Neural Information Processing Systems, NeurIPS 2021, 2021.