Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Variational multivariate data assimilation improves sea-ice analysis and forecasting

Vasily Ivanov

va.ivanov@skoltech.ru

Skolkovo Institute of Science and Technology

Liubov Antiufrieva

Skolkovo Institute of Science and Technology

Vladimir Vanovskiy

Skolkovo Institute of Science and Technology

Evgeny Burnaev

Skolkovo Institute of Science and Technology

Article

Keywords: Data assimilation, Sea ice, NEMO/SI3, Variational autoencoder, AMSR2, Sentinel-3

Posted Date: September 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10699873/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Variational multivariate data assimilation improves sea-ice analysis and forecasting

Vasily Ivanov[1,*], Liubov Antiufrieva[1], Vladimir Vanovskiy[1], and Evgeny Burnaev[1,2]

[1]Artificial Intelligence Center, Skolkovo Institute of Science and Technology, Moscow 121205, Russia
[2]Autonomous Non-Profit Organization Artificial Intelligence Research Institute (AIRI), Moscow 105064, Russia
[*]va.ivanov@skoltech.ru

ABSTRACT

This study presents a neural data assimilation method based on a variational autoencoder (VAE) for sea-ice state estimation and short-range forecast initialization in a high-resolution regional ocean model. Specifically, we propose VAE Multivariate Data Assimilation (VAE-MDA), an approach that jointly processes several physical fields linked to sea-ice through a shared VAE representation. The VAE-MDA method is evaluated using AMSR2 sea-ice concentration sampled along Sentinel-3 SRAL tracks, with background fields provided by the NEMO ocean model coupled to the SI3 sea-ice component. We show that VAE-MDA, assimilating only sparse track observations, reduces sea-ice concentration errors against both dense AMSR2 and sparse SRAL-conditioned AMSR2 validation targets. We also develop a restart-compatible algorithm to transfer the assimilated fields into the NEMO/SI3 ocean–sea-ice model. Furthermore, we introduce a physically bounded thermodynamic upper-ocean restart adjustment procedure. We demonstrate that applying this procedure alongside the sea-ice correction yields a substantial additional improvement in 72-hour forecast skill. The proposed VAE-MDA provides a trainable nonlinear latent-space alternative to traditional methods such as 3D-VAR in practical ocean modeling workflows.

Keywords: Data assimilation, Sea ice, NEMO/SI3, Variational autoencoder, AMSR2, Sentinel-3

Introduction

A major trend in Earth science is the growing interest in the Arctic, driven by the rapid decline in sea-ice cover [1] and the profound environmental and economic changes associated with it [2–4]. Numerical models for the Arctic and Antarctic are difficult to run and calibrate [5,6]. To produce useful sea-ice forecasts, one must account for the coupled effects of wind forcing, ocean dynamics, and ice thermodynamics and rheology. Such forecasts have both scientific importance and practical value for navigation in ice-covered waters.

Accordingly, Earth science is moving toward higher-resolution climate models [7–9] and larger volumes of observational data [10, 11] to improve the prediction of extreme events and circulation patterns. This trend is driven by advances in computational power and by improved observing technologies.

Numerical models are a cornerstone of ocean simulation because they incorporate current knowledge of the complex physical processes that govern ocean dynamics. Owing to nonlinear coupled dynamics, small uncertainties in the initial conditions can significantly degrade forecasts, especially near rapidly evolving ice edges. Data assimilation is, therefore, essential for constraining numerical models using observations and improving forecast quality. Classical data assimilation algorithms typically rely on assumptions of linearized model dynamics and Gaussian-distributed background and observation errors. However, as model resolution increases, these assumptions become less valid, motivating the development of approaches that can better capture the complexity of high-resolution systems [12].

Data assimilation iteratively updates the model physical fields x_b (background) using the observed data y. The goal is to move closer to the unknown true state of the system, represented by the analysis fields x_a [13,14]. The assimilation process must account for the spatial and temporal relationships among the physical fields.

There are several approaches to data assimilation. In this paper, we refer to non-neural approaches as classical data assimilation methods. In most cases, these methods are based either on filtering techniques such as the Best Linear Unbiased Estimator (BLUE) and the Kalman filter [15–18], or on the optimization of a cost function, as in three- and four-dimensional variational data assimilation (3D-VAR and 4D-VAR) [13,14,19,20]. Neural networks have shown a remarkable ability to learn complex nonlinear relationships between oceanographic physical fields [21]. Given their capacity to capture high-dimensional, nonlinear dependencies directly from data, they offer a natural complement to classical assimilation schemes, which often face challenges with such complexity. Several studies have, therefore, combined classical data assimilation algorithms with

modern neural networks; these approaches range from accelerating classical solvers [22, 23] to replacing covariance matrices with learned representations [24–26].

Most existing neural data assimilation studies rely on simplified benchmark systems, such as the Lorenz-63 or Lorenz-96 models [25, 27]. These toy models are useful testbeds for evaluating assimilation techniques because they capture essential features of chaotic dynamics while remaining computationally tractable. However, their simplicity often limits the direct applicability of the resulting methods to more complex real-world systems characterized by high dimensionality and intricate spatial correlations.

Observational data can be sparse and subject to measurement errors. The true physical field is unknown, and so are the exact model and observation errors. As a result, practical assimilation and validation procedures must rely on estimates of error statistics. This challenge is especially pronounced when constructing the background-error covariance matrix B, since real-world geophysical errors often violate the Gaussian assumptions underlying classical methods. For example, sea ice concentration is strictly bounded between 0 and 1, leading to inherently non-Gaussian error distributions; yet the Ensemble Kalman Filter (EnKF) has been used to assimilate this field [28].

3D-VAR and 4D-VAR are variational data assimilation methods based on cost-function optimization. The cost function for 3D-VAR, used here as a classical baseline, is

$$J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y - Hx)^T R^{-1} (y - Hx) \tag{1}$$

where B is the background-error covariance matrix, R is the observation-error covariance matrix, and $H : \mathbb{R}^n \to \mathbb{R}^m$ is the observation operator that maps the model state to observation space. For track-based assimilation, H is implemented as a sparse spatial interpolation from the model grid to the observation locations. 4D-VAR extends this objective over an assimilation window using the forward model operator, and the Kalman filter minimizes an equivalent cost function in the linear-Gaussian case [29].

An alternative to classical data assimilation schemes is to integrate artificial neural networks (ANNs) into traditional systems to replace or enhance specific components. There are two main strategies for incorporating ANNs into data assimilation. The first uses a neural network to distill classical methods [22, 23, 30], such as 3D-VAR, 4D-VAR, and the Ensemble Transform Kalman Filter (ETKF), and capture their key features in a more computationally efficient form. The second integrates ANNs directly into the assimilation process by replacing particular components of traditional algorithms. Specific approaches include learning covariance matrices or latent representations [24–26], training neural tangent-linear and adjoint models [31], and applying super-resolution techniques to enhance observation impact [32]. This strategy potentially improves adaptability and provides more accurate error estimates. The second approach is of greater interest here because it enables a more flexible and adaptive assimilation process. This study explores the potential of such an approach to improve accuracy, reduce computational cost, and enhance robustness under dynamically changing ice conditions.

Classical data assimilation algorithms face two major challenges. The first is the specification of the background-error (B) and observation-error (R) covariance matrices, which can be extremely large and may fail to capture nonlinear interactions between variables. The second challenge is temporal error propagation: using the full forecast model is computationally prohibitive, whereas simplified linear approximations fail to capture complex nonlinear dynamics.

The variational autoencoder (VAE) [33] is a generative probabilistic model that learns a low-dimensional latent representation of high-dimensional data while capturing complex nonlinear dependencies. By imposing a structured prior on the latent space, VAEs provide a flexible framework for modeling intricate multivariate distributions, making them particularly suitable for physical fields with strong cross-variable interactions. In [24], a VAE is applied within a 3D-VAR algorithm to reduce the dimensionality of physical fields while capturing inter-variable relationships; the method is evaluated on high-dimensional spatiotemporal environmental data. Fan et al. [34] analyze multivariate latent-space variational assimilation and show that, for an affine decoder, it is equivalent to model-space assimilation with a decoder-induced background covariance, providing a theoretical motivation for multivariate latent corrections. Similarly, [25] applied EnKF in the latent space of an autoencoder and tested it on the Lorenz-96 model. Furthermore, [26] used an autoencoder to completely replace the covariance matrix in a 3D-VAR algorithm. That approach was tested on temperature data at the 850-hPa pressure level from the ERA5 reanalysis. The latter study is conceptually similar to our work, but it considers only a single physical field and was not tested in an operational forecasting pipeline.

This paper makes the following contributions:

- We propose VAE Multivariate Data Assimilation (VAE-MDA), a ResNet-based VAE method for assimilating multiple geophysical fields that employs an attention bottleneck and a shared spatial latent feature map. We evaluate it in several experimental settings against classical baselines (3D-VAR, BLUE, and EnKF) and a VAE-based approach [26].
- We develop a physically bounded thermodynamic upper-ocean restart transfer procedure and demonstrate that coupling it with restart-based sea-ice corrections significantly enhances short-range forecast skill compared to sea-ice-only updates.

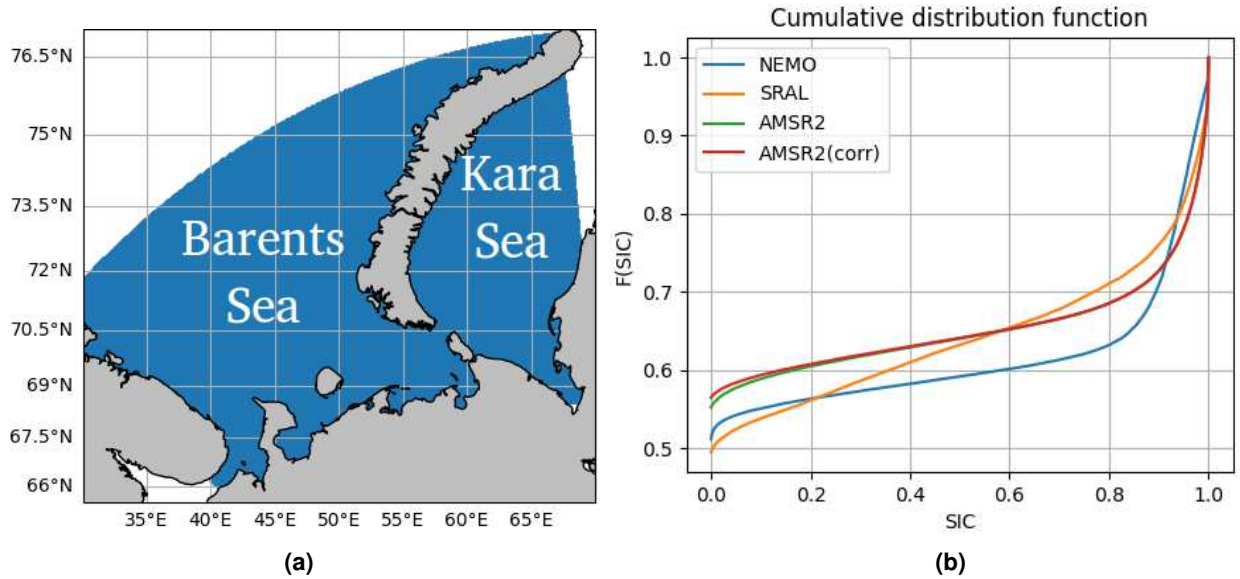


Figure 1. (a) Study area; blue indicates the NEMO simulation domain. (b) Cumulative distribution functions of sea-ice concentration from different data sources. The SRAL-conditioned AMSR2 curve uses the surface-type conditioning described in the Methods section.

- We implement a restart-compatible algorithm that integrates VAE-MDA into a regional forecasting workflow based on the Nucleus for European Modelling of the Ocean (NEMO) and its sea-ice component SI3, to assimilate Advanced Microwave Scanning Radiometer 2 (AMSR2) sea-ice concentration sampled along Sentinel-3 SAR Radar Altimeter (SRAL) tracks.

Results

Experimental setup

The study domain encompasses parts of the Barents Sea and the Kara Sea (Figure 1a). The study regions are free of multi-year ice [35]. In the Kara Sea sector, freeze-up begins rapidly in December and melting starts in April, whereas the Barents Sea sector remains largely ice-free for most of the year. This diversity in seasonal cycles ensures that the proposed method is evaluated across a wide range of thermodynamic conditions and marginal ice-zone dynamics.

To set the data context, we compared sea-ice concentration distributions along the Sentinel-3 track between December 2019 and April 2020 across sources (Figure 1b). Data were collected for identical dates and spatial domains across all sources. The NEMO model (blue line) exhibits a lower probability of zero ice concentration than AMSR2 (green line), which indicates an overestimation of ice cover in the model output. The SRAL-derived `sea_ice_concentration_20_ku` variable (orange line) shows a markedly different slope from both NEMO and AMSR2; this divergence is likely due to the coarse resolution of the OSI-430-a product, which is the source of this variable and effectively smooths the distribution in space.

The multivariate VAE architecture used throughout this study, together with the VAE-MDA procedure that couples background fields with sparse satellite observations, is summarized in Figure 2. The architecture uses a spatial latent feature map rather than a vector, ResNet-based blocks throughout the encoder and decoder, and attention in the middle layers; the main four-field model has four input/output channels (*siconc*, *sithic*, *sitemp*, and *sosstsst*), four encoder/decoder stages with channel widths 128, 256, 512, and 512, and two ResNet blocks per stage. Training, baseline and day-of-year (DOY)-conditioned variants, and the full assimilation procedure are described in the Methods section. Because our goal is to evaluate sparse-track sea-ice assimilation in a regional NEMO/SI3 forecasting workflow, we organized the evaluation in several stages. First, we assessed the reconstruction performance of the VAE on the 2022 validation year. Second, using the same development year, we performed model-to-model and satellite-to-model assimilation experiments to guide architecture, observation-window, and assimilation-parameter selection. The model-to-model experiment served as a controlled pseudo-observation stress test, in which interannual same-season differences typically produced larger innovations and therefore provided a more discriminative test of whether sparse concentration updates induced coherent multivariate increments. The satellite-to-model experiment then evaluated performance in the more realistic observational regime using AMSR2 concentration sampled along SRAL tracks.

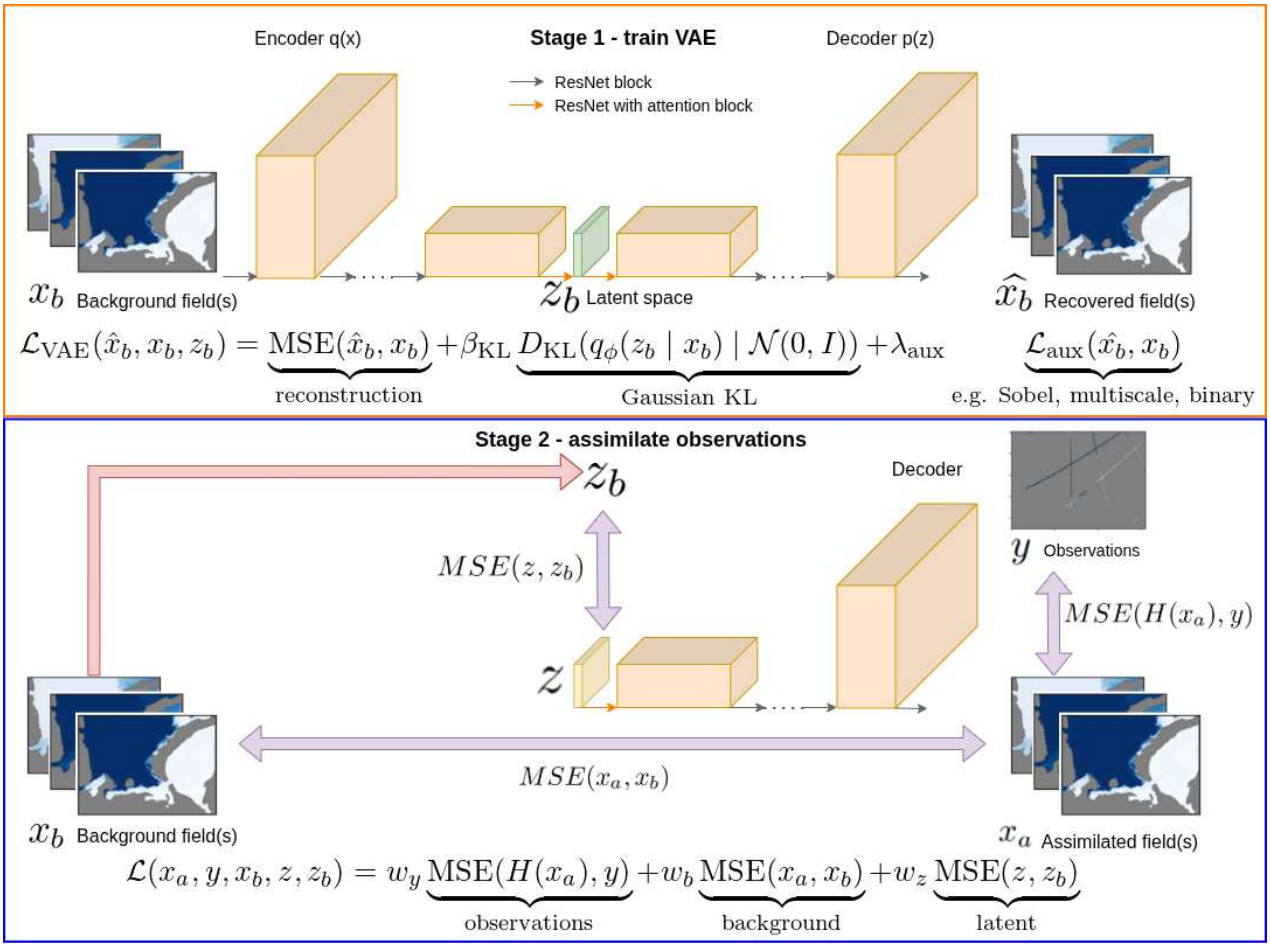


Figure 2. Four physical fields (*siconc*, *sithic*, *sitemp*, *sosstsst*) are jointly encoded through four ResNet-based stages with channel widths 128, 256, 512, and 512 to an attention bottleneck that produces a spatial latent feature map z; the decoder mirrors the encoder and reconstructs the four fields. During assimilation, the background field x_b is encoded as $z_b = E(x_b)$; x_b and sparse AMSR2 observations y sampled along SRAL tracks enter the loss $\mathscr{L}(z) = w_y\|H(x_a) - y\|_2^2 + w_b\|x_a - x_b\|_2^2 + w_z\|z - z_b\|_2^2$, which is minimized over the latent variable z by Adam steps ($\nabla_z \mathscr{L}$), producing the analysis $x_a = D(z)$.

Finally, after fixing the selected configuration, we evaluated the restart-compatible VAE-MDA within the NEMO forecasting pipeline on the held-out 2023 period.

The VAE is expected to reconstruct the fields reasonably well, although perfect accuracy is not the objective, because VAEs inherently balance reconstruction fidelity against the compactness of the latent representation. The purpose of the model-to-model assimilation experiment is to evaluate assimilation behavior in a controlled setting. Observation data and model data are both noisy and often follow different distributions, which can bias the evaluation. By contrast, the model-to-model setup provides a same-model-distribution benchmark. Its role here is not direct observational validation, but a controlled test of correction capacity and decoded multivariate increments using pseudo-observations that differ substantially from the background, sampled at real observation (SRAL-track) locations.

Reconstruction error

The VAE was trained on data from 2015 to 2021, and data from 2022 were used for validation. The goal was to achieve adequate reconstruction quality for the encoded ocean and ice variables. The input to the VAE consists of multivariate fields (four-field VAE setup): *siconc* (sea-ice concentration), *sithic* (sea-ice thickness), *sitemp* (sea-ice temperature), and *sosstsst* (sea-surface temperature). The quantitative reconstruction results are summarized in Supplementary Table S3.

The 2022 M2M and S2M comparisons were used for architecture and parameter selection and are reported as development results; the 2023 forecast-cycling experiment was held out from this selection.

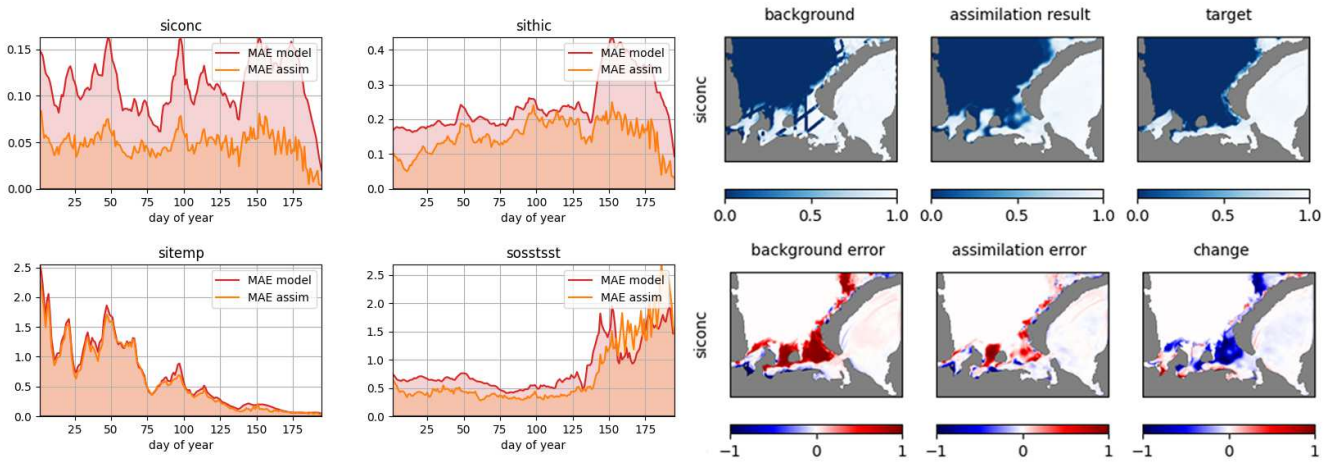


Figure 3. Model-to-model assimilation performance of the *vae_4f_um* model. The left panel shows background and analysis MAE values for different physical fields by day of the year. The right panel shows an example of sea-ice concentration assimilation on 12 April.

We report reconstruction and assimilation errors using mean absolute error (MAE) and root mean squared error (RMSE):

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{x}_i - x_i|, \qquad \mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2}. \tag{2}$$

MAE measures the average error magnitude, whereas RMSE penalizes large local errors more strongly. For assimilation experiments, skill is reported relative to the background model,

$$\mathrm{skill}_m = 1 - \frac{m_{\mathrm{analysis}}}{m_{\mathrm{background}}}, \tag{3}$$

where m is either MAE or RMSE. The reported mean skill is the arithmetic average of the MAE skill and RMSE skill; positive values indicate improvement over the background model.

The model names encode the main ablation choices. The suffix **_*f** gives the number of input physical fields, **_um** denotes an unmasked reconstruction loss during VAE training, **_d*** denotes a vector latent representation of the specified size, and **_emb** denotes day-of-year conditioning. The suffixes **_multiscale**, **_sobel**, **_binary**, and **_multiscale_block** denote additional sea-ice-concentration loss terms defined in the Methods/Training subsection.

Quality estimation

Model-to-model assimilation

In this experiment, our objective was to assess the ability of a trained VAE to capture relationships both within and across physical fields. In practice, however, the true field values at each time step are unknown: the numerical model suffers from model errors and depends on initial conditions, which introduce additional uncertainty, whereas satellite data are affected by measurement errors and retrieval uncertainties.

We therefore evaluate the method through a model-to-model assimilation experiment. Specifically, we sample sea-ice concentration data along SRAL tracks from the NEMO forecast for the following year, $\tilde{y} = x_{i+365}[m_i]$, and assimilate them into the current data x_i. To evaluate performance, we compare the analysis fields with the model prediction x_{i+365} (Algorithm 1(a)). Because both the pseudo-observations and the verification target are drawn from NEMO, this experiment should be interpreted as a controlled same-model-distribution benchmark rather than as direct validation against reality. Its main role is to test whether the assimilation method can absorb sparse observations and produce coherent multivariate corrections when the imposed innovations are relatively large.

Among the classical baselines, 3D-VAR, BLUE, and EnKF produced broadly similar results in this sparse-track assimilation setting. In the M2M setting, assimilating only sea-ice concentration values induces coupled decoded increments in related fields. The *vae_4f_um* family yields lower error than 3D-VAR and preserves a sharper ice-water boundary, whereas 3D-VAR tends to smooth that boundary. In the *vae_4f_um* results, a decrease in ice concentration in a given region is accompanied

Algorithm 1 Assimilation methods. (a) Model-to-model (M2M): pseudo-observations are NEMO sea-ice concentrations from the following model year sampled at the same SRAL-track locations. (b) Satellite-to-model (S2M): sparse SRAL-conditioned AMSR2 sea-ice concentrations are assimilated at SRAL-track locations and validated against the dense AMSR2 field. Sparse S2M validation is prequential: each newly available track is scored against the existing estimates before that track enters the subsequent three-day assimilation window.

(a) Model-to-model (M2M)

Require: $\{x_1, x_2, \ldots, x_n\}$ – NEMO simulation fields; $\{m_1, m_2, \ldots, m_n\}$ – SRAL-track masks
for day in assimilation cycle **do**
  Select SRAL-track mask m_i for the current day
  Sample pseudo-observation: $\tilde{y}_i = x_{i+365}[m_i]$
  Assimilate: $x_i^a = \text{VAE-MDA}(x_i, \tilde{y}_i; m_i)$
  Validate: compare x_i and x_i^a with x_{i+365}
end for

(b) Satellite-to-model (S2M)

Require: Existing pre-update model and VAE-MDA estimates; sparse observations $\{y_i\}_{i=1}^n$ and track operators $\{H_i\}_{i=1}^n$; dense AMSR2 fields $\{v_i\}_{i=1}^n$
for validation time t_0 in the assimilation cycle **do**
  Validate first: score the existing estimates against y_{t_0} at H_{t_0}, before using y_{t_0} in an update
  After recording the sparse metrics, form $\mathscr{W}_{t_0} = \{(H_t, y_t) : t = t_0, t_0 - 1, t_0 - 2\}$
  Assimilate: update the VAE-MDA estimate using the three-day window $\mathscr{W}_{t_0}$
  Evaluate the updated estimate separately against the dense AMSR2 field v_{t_0}
end for

by a decrease in ice thickness and by increases in both ice temperature and ocean temperature. These patterns are consistent with the target fields in the same-model-distribution benchmark and are plausible within the learned model representation. Overall, this indicates that the shared latent representation captures relationships within individual fields as well as empirical cross-correlations among them, consistent with the latent-space coupling induced by the shared decoder [34]. Theorem 1 provides a boundedness result for coupled decoded increments under the optimization objective, but it does not by itself guarantee forecast improvement or physical validity.

Figure 3 shows the assimilation metrics for the *vae_4f_um* model applied to ice concentration, ice thickness, ice temperature, and upper-layer water temperature over different days of the year. The time series is interrupted in mid-July, when the ice in the Barents and Kara Seas melts completely. The ice-concentration and ice-thickness fields show the strongest correspondence, whereas the ice- and water-temperature fields exhibit a weaker dependence on ice concentration. Table 1 reports sea-ice-concentration skill across models. The main conclusion is that the four-field VAE models form a leading group in this proxy benchmark. Their skill estimates agree within the bootstrap uncertainty, and the paired block-bootstrap difference between the point-best *vae_4f_um_sobel_binary* model and *vae_4f_um_multiscale_sobel_binary* was only 0.004 ± 0.003. In contrast, the difference between *vae_4f_um_multiscale_sobel_binary* and 3D-VAR was 0.124 ± 0.008. We therefore treat the four-field family as the relevant M2M result rather than selecting a unique winner from this proxy benchmark. For subsequent restart-compatible experiments, we use *vae_4f_um_multiscale_sobel_binary* as a representative four-field model, with the final selection further guided by the satellite-to-model evaluation.

Auxiliary-field errors for sea-ice thickness, sea-surface temperature, and sea-ice temperature are reported in Supplementary Table S2; they confirm that the selected four-field models produce decoded multivariate increments rather than isolated sea-ice-concentration corrections.

An additional experiment was conducted to evaluate how the observation-accumulation window affects the quality of the resulting analysis. To this end, we accumulated SRAL-track observations over windows of increasing depth and repeated the model-to-model assimilation experiment. Figure 4 shows that the number of assimilated observations increases almost linearly with window depth, whereas dense-target assimilation skill rises rapidly over the first few days, reaches a plateau at intermediate depths, and then gradually decreases for longer windows. This behavior indicates a trade-off between spatial coverage and temporal consistency: additional older tracks provide more observations but eventually become less representative of the current sea-ice state. This experiment also provides insight into the optimal accumulation window for satellite observations. Note that with real observations this window is additionally limited by evolving ice conditions, whereas in the controlled experiment all tracks were sampled from a single date.

Satellite-to-model assimilation

The next step was the assimilation of AMSR2 ice-concentration data conditioned using SRAL surface-type information (Algorithm 1(b)). We evaluated assimilation performance against both dense AMSR2 data and sparse SRAL-conditioned AMSR2 data at SRAL-track locations. We also tested an empirical quantile correction to map AMSR2 concentrations toward the model distribution, but it did not improve the downstream assimilation results; therefore, the main text reports the S2M

Table 1. Model-to-model sea-ice-concentration assimilation skill on the 200-day evaluation period (2022-01-01 to 2022-07-19; $n = 200$ daily cases). Values report MAE skill, RMSE skill, and their arithmetic mean relative to the background model. Uncertainty is reported as the point estimate ± one standard deviation of the moving-block bootstrap distribution using 14-day blocks.

model	MAE skill	RMSE skill	mean skill
vae_4f_um_sobel_binary	0.592 ± 0.023	0.515 ± 0.022	0.553 ± 0.022
vae_4f_um_multiscale_sobel_binary	0.584 ± 0.023	0.515 ± 0.020	0.550 ± 0.021
vae_4f_um_multiscale	0.552 ± 0.025	0.545 ± 0.019	0.548 ± 0.021
vae_4f_um	0.570 ± 0.024	0.522 ± 0.018	0.546 ± 0.021
vae_4f_um_multiscale_block	0.564 ± 0.020	0.522 ± 0.014	0.543 ± 0.016
vae_3f_um_emb	0.480 ± 0.028	0.434 ± 0.021	0.457 ± 0.024
BLUE	0.455 ± 0.025	0.415 ± 0.018	0.435 ± 0.022
3D-VAR	0.446 ± 0.024	0.406 ± 0.018	0.426 ± 0.021
EnKF	0.429 ± 0.022	0.382 ± 0.018	0.406 ± 0.020
vae_1f	0.435 ± 0.035	0.376 ± 0.030	0.406 ± 0.032
vae_1f_d512	0.408 ± 0.032	0.387 ± 0.027	0.397 ± 0.030

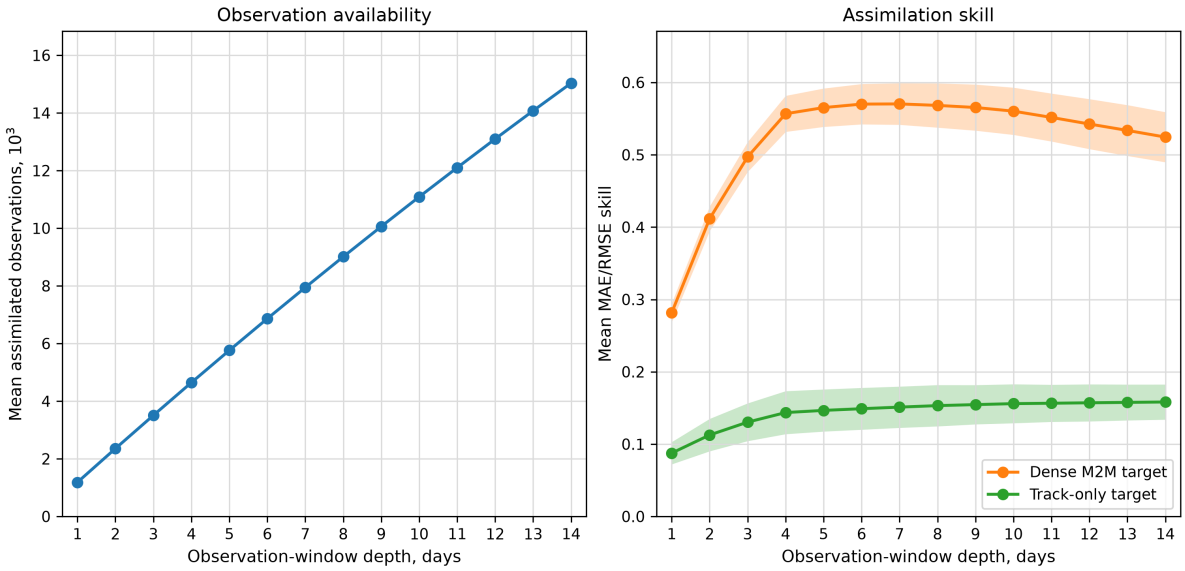


Figure 4. Observation-window-depth experiment. Left: mean number of assimilated SRAL-track observations as a function of accumulation-window depth. Right: mean MAE/RMSE skill for sea-ice concentration against the dense model-to-model target and against the track-only target. Shaded bands show ±1 bootstrap standard deviation using 14-day blocks.

experiment without quantile mapping.

The assimilation results for the sparse SRAL-conditioned AMSR2 sea-ice concentration data are presented in Table 2. We report skill separately against dense AMSR2 and against the sparse SRAL-conditioned AMSR2 target, because these two validation targets favor different model and observation characteristics. The dense AMSR2 field is not used in the latent optimization objective, which only sees sparse SRAL-track samples, but both the sparse observations and the dense validation field come from the AMSR2 retrieval product. The dense score therefore measures consistency with a withheld gridded AMSR2 product rather than validation against an independent observing system. The three-field *vae_3f_um_emb* model achieved the highest dense AMSR2 skill, whereas the single-field *vae_1f_d512* model achieved the highest sparse SRAL-conditioned AMSR2 skill. Forecast cycling was nevertheless restricted to four-field models, because the coupled restart transfer of Algorithm 3 requires an analyzed sea-surface-temperature field, which the single- and three-field variants do not produce. The *vae_4f_um_multiscale_block* model was the strongest four-field variant. Its advantage over the *vae_4f_um_multiscale_sobel_binary* configuration was not resolved by the paired bootstrap, so we retained the latter for

Table 2. Satellite-to-model assimilation skill on the 2022 evaluation period (2022-01-01 to 2022-07-19; $n = 200$ daily cases). Values report the mean MAE/RMSE skill relative to the background model. Uncertainty is reported as the point estimate $\pm$ one standard deviation of the moving-block bootstrap distribution using 14-day blocks. The assimilated sparse observations and sparse validation target use the surface-type conditioning described in Methods. Models are ordered by the arithmetic mean of their dense and sparse skills.

model	dense AMSR2 skill	sparse SRAL-conditioned AMSR2 skill
vae_3f_um_emb	0.379 ± 0.025	0.350 ± 0.025
vae_1f_d512	0.369 ± 0.023	0.356 ± 0.023
vae_4f_um_multiscale_block	0.376 ± 0.025	0.340 ± 0.024
vae_4f_um	0.374 ± 0.025	0.333 ± 0.024
vae_4f_um_multiscale_sobel_binary	0.359 ± 0.027	0.332 ± 0.029
vae_4f_um_sobel_binary	0.361 ± 0.027	0.328 ± 0.029
vae_4f_um_multiscale	0.356 ± 0.024	0.316 ± 0.025
vae_1f	0.346 ± 0.026	0.319 ± 0.028
EnKF	0.304 ± 0.022	0.243 ± 0.021
3D-VAR	0.295 ± 0.023	0.203 ± 0.024
BLUE	0.289 ± 0.023	0.180 ± 0.026

forecast cycling.

Twin experiment

We additionally conducted a twin forecast experiment, a standard validation protocol in data-assimilation studies. A baseline run initialized from the 2022 state was compared against a second run that assimilated 2023 observations into the same initial state, both integrated under 2023 atmospheric forcing and boundary conditions. The assimilated run reduced the MSE and MAE for *siconc* and *sithic* relative to the baseline trajectory; the full setup, metric values, and time series are provided in the Supplementary Note on the twin forecast extension and Supplementary Fig. S1.

Restart-based forecast cycling

In the final series of experiments, we integrated the results of satellite data assimilation into the NEMO pipeline through the restart mechanism. This experiment was conducted on the held-out 2023 period after the VAE architecture and assimilation settings had been selected using 2022 data. The forecast-cycling runs used the selected restart-compatible four-field model with the stronger-observation setting, $w_y = 6.0$, $w_b = 0.1$, and $w_z = 0.01$. The analyzed sea-ice concentration and thickness fields from the *vae_4f_um_multiscale_sobel_binary* model were transferred to the SI3 restart using Algorithm 2. In the coupled ice-ocean restart experiment, this ice correction was followed by the proposed bounded thermodynamic upper-ocean transfer in Algorithm 3, which uses the analyzed sea-surface temperature field as a raw proposal and bounds the imposed temperature and salinity increments by the diagnosed sea-ice-volume change. The ice-temperature field was not written directly to the restart.

The experiment was organized as follows. Starting from the selected initial day, each daily restart of the ocean model was modified by VAE-MDA in a cycling procedure that follows the forecast workflow. Baseline and assimilated forecasts were compared only for initialization dates available in all runs and for which the corresponding validation target was available at the forecast valid date. For each forecast lead, validation used the target field at the forecast valid date; observations from that valid date were not assimilated before evaluation. We evaluated forecast lead times of 0, 24, 48, and 72 hours to assess both immediate analysis impact and short-range forecast persistence.

We compared two restart-transfer variants on a common sample of $n = 200$ forecast cases at each lead time. The same 200 initialization dates (2023-01-01 to 2023-07-19) were used across all four lead times, both restart variants, and both validation targets. The ice-only variant applies the sea-ice restart correction in Algorithm 2, whereas the coupled variant additionally applies the physically bounded thermodynamic upper-ocean transfer in Algorithm 3. Forecasts were evaluated against both dense AMSR2 sea-ice concentration and the sparse SRAL-conditioned AMSR2 target.

Figure 5 shows that the ice-only correction improves the initial state and retains positive but reduced skill through 72 hours. At 72 hours, its mean MAE/RMSE skill was 0.143 ± 0.059 against dense AMSR2 and 0.164 ± 0.066 against the sparse SRAL-conditioned AMSR2 target. The coupled restart with physically bounded upper-ocean transfer preserved higher skill at 72 hours, with mean skill 0.258 ± 0.035 against dense AMSR2, and 0.297 ± 0.036 against the sparse target. This indicates that, in this experiment, forecast persistence depends not only on the VAE-derived sea-ice analysis but also on the accompanying

thermodynamic upper-ocean restart transfer. Examples of the assimilated fields are shown in Figure 6. The evaluated dates cover the ice season, when changes in ice cover can be meaningfully interpreted.

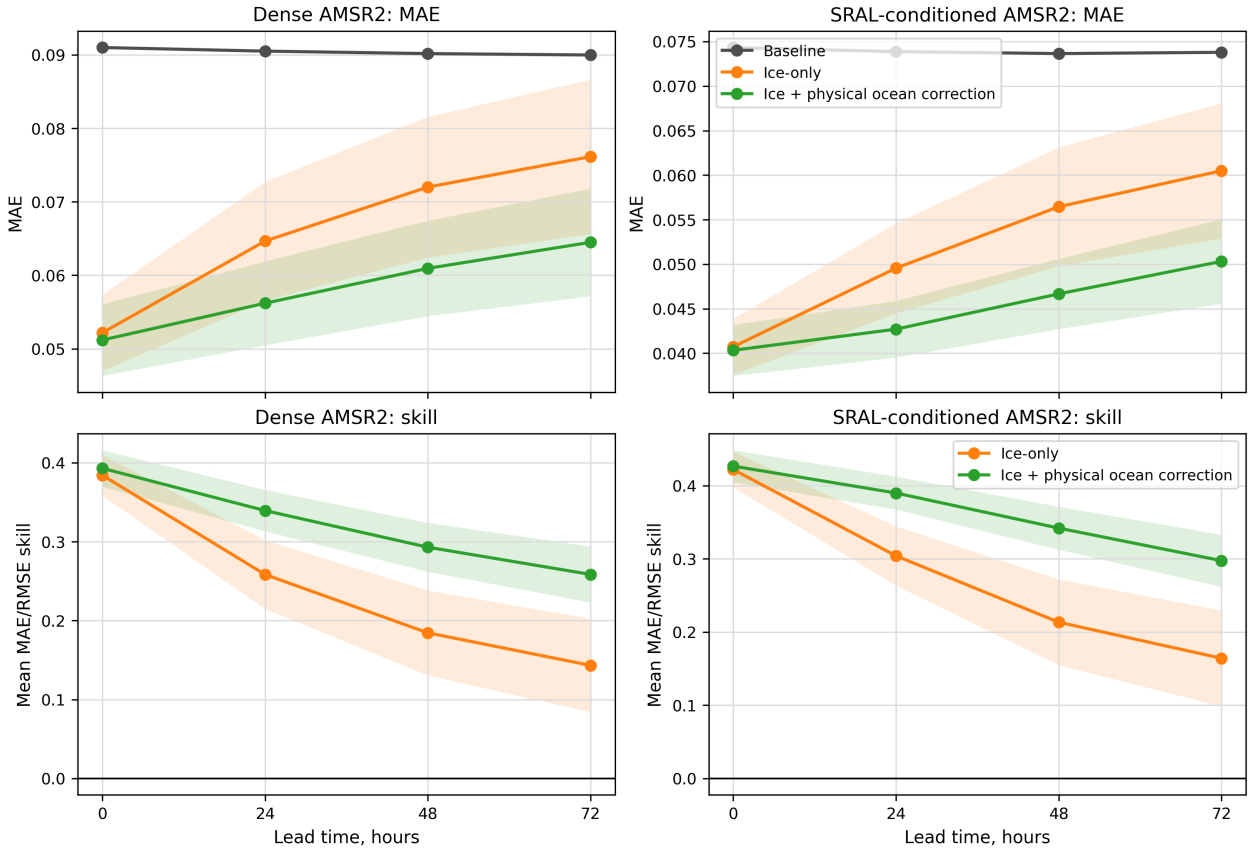


Figure 5. Forecast performance of restart-based satellite-to-model assimilation for a common sample of $n = 200$ forecast cases at each lead time. The top row shows sea-ice-concentration MAE for the baseline forecast, the ice-only restart correction, and the coupled restart with physically bounded upper-ocean transfer. The bottom row shows mean MAE/RMSE skill relative to the baseline forecast. Columns correspond to dense AMSR2 validation and sparse SRAL-conditioned AMSR2 validation. Shaded bands show ± 1 bootstrap standard deviation for the restart experiments using 14-day blocks.

Discussion

This work investigated VAE-MDA for multivariate sea-ice state estimation in a high-resolution regional NEMO/SI3 forecasting workflow. The proposed method uses a VAE decoder as a learned multivariate prior, allowing sparse sea-ice concentration observations to induce spatially coherent corrections in concentration and related decoded fields. Across both the model-to-model and satellite-to-model benchmarks, the four-field VAE models consistently outperformed the classical baselines, while differences within the leading four-field VAE group were small relative to the bootstrap uncertainty. The method is formulated with a multivariate background penalty over all decoded fields. In a paired 200-day ablation (Supplementary Note on the background-penalty ablation), the multivariate penalty improved the unobserved ice-temperature analysis but consistently reduced concentration and, especially, ice-edge skill, while also changing the effective per-field weighting of the loss; this trade-off motivated restricting the background penalty to the observed concentration field in all reported experiments.

The restart experiments show that analysis quality alone is not sufficient to maximize forecast persistence. The ice-only restart correction improves the initial sea-ice state and retains positive but reduced skill through 72 hours. Adding the proposed physically bounded thermodynamic upper-ocean restart transfer preserves larger positive skill through 72 hours for both dense AMSR2 and sparse SRAL-conditioned AMSR2 validation. This suggests that the derived sea-ice concentration and thickness increments are more effective when accompanied by compatible upper-ocean temperature and salinity adjustments. At the same time, the ocean component of the restart update should not be interpreted as a complete dynamically balanced ocean update: the VAE produces a multivariate analysis including sea-surface temperature, but the restart transfer inserts only upper-ocean

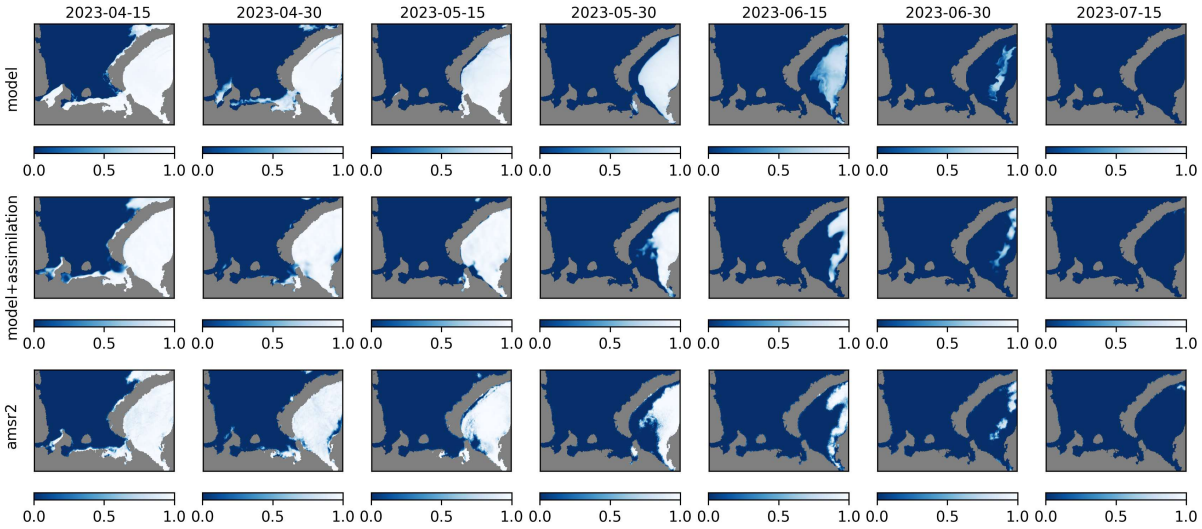


Figure 6. Examples of output fields in the satellite-to-model assimilation regime during the ice season.

temperature and prescribed salinity increments, bounded by local thermodynamic budgets, while velocity, sea-surface height, barotropic variables, and other dynamical restart variables are left unchanged.

Several limitations remain. The SRAL-derived sea-ice concentration product is not yet accurate enough to serve as the sole concentration target, so the satellite-to-model pipeline relies on AMSR2 concentration conditioned using SRAL surface-type information. The forecast comparison was also limited to the common 2023 cases for which all baseline and assimilated forecasts were available. Finally, the theoretical results justify bounded latent-space corrections under an idealized global-minimizer assumption, whereas the implemented algorithm uses local Adam optimization and must be evaluated empirically.

Future work should focus on more dynamically balanced restart coupling and on richer observation products. Improved Sentinel-3 sea-ice products, especially reliable ice-thickness retrievals, could allow the sparse-track assimilation problem to be constrained directly by observed quantities. Another promising direction is to learn latent representations that include a broader coupled ocean–ice state and can propagate corrections through multiple physical fields while respecting dynamical balance, for instance by incorporating physics-informed loss terms that enforce conservation laws and thermodynamic constraints directly during VAE training. A further step would be to extend the proposed VAE-MDA approach toward a four-dimensional variational setting by learning a latent-space evolution operator that propagates the analysis forward in time, analogous to the tangent-linear model in classical 4D-VAR. Extending the geographical domain and incorporating atmospheric-forcing uncertainty may also improve forecast impact in operational Arctic applications.

Methods

Data

We describe below the modeling background that supplies x_b, the SRAL and AMSR2 observational products, the validation targets, and the statistical-evaluation procedure used in this study.

Background

As the assimilation background, we use daily ocean and sea-ice forecasts generated by our operational forecast pipeline incorporating NEMO version 4.0 coupled to SI3, with atmospheric forcing from version 4.4 of the Weather Research and Forecasting (WRF) model. The spatial resolution of the modeling dataset is approximately 3–4 km in the study region; we use the ORCA12 grid and the Drakkar configuration as a high-resolution regional ocean–sea-ice modeling setup. The same WRF–NEMO-SI3 modeling system underlies the Borey regional atmosphere–ocean–sea-ice–wave hindcast dataset [36], which provides a detailed description of the model configuration, atmospheric forcing, and validation against reference products. For the period of interest, the dataset contains daily forecasts for the subsequent 72 hours and spans 2015 to 2023. The reported multivariate VAE experiments use four restart-relevant fields: *siconc* (sea-ice concentration), which is the field constrained directly by observations; *sithic* (sea-ice thickness); *sitemp* (sea-ice temperature); and *sosstsst* (sea-surface temperature). Data from 2015 to 2021 inclusive are used to compute statistical properties and to train the variational autoencoder (VAE). The year 2022 is reserved for VAE reconstruction validation and for selecting the assimilation configuration through the model-to-model

and satellite-to-model experiments. The final forecast experiment is performed on 2023, which is not used for VAE training, model selection, or assimilation-parameter tuning.

Observations

The following observational datasets were considered:

- Sentinel-3 SRAL Sea Ice Thematic Product [37]. This dataset includes satellite tracks with an along-track resolution of up to 330 m. The variable `sea_ice_concentration_20_ku` provides ice-concentration data and is available as an auxiliary product. Its primary source is the OSI-430-a product, which has a 2-day latency and a spatial resolution of 25 km. Therefore, greater scientific interest lies in the variable `surf_type_class_20_ku`, which characterizes surface type (open ocean, floes, lead, or unclassified) on the basis of sea-ice concentration and waveform peakiness, and in the variable `freeboard_20_ku` (radar freeboard). Future work could develop machine-learning methods to estimate sea-ice thickness from the radar freeboard; such a retrieval is outside the scope of this study.

- ASI AMSR2 [38] sea-ice concentration data. These data have a spatial resolution of 6.25 km and are available in near-real time (NRT) with a latency of about 25 hours. A key advantage of AMSR2 is its all-weather, day-and-night operational capability enabled by passive microwave measurements. However, a known limitation is the reduced accuracy of sea-ice concentration retrieval during the melt season because of changes in surface emissivity caused by snow and ice melt [39].

We identified track-data assimilation as the most promising target problem, especially because Sentinel-3 is expected to extend its sea-ice thematic product with ice-thickness information. SRAL tracks are therefore used to define the sparse observation locations. For model-to-model assimilation, pseudo-observations are NEMO sea-ice concentration values from the following model year sampled at the same SRAL-track locations. For satellite-to-model assimilation, AMSR2 sea-ice concentrations sampled along SRAL tracks were conditioned using the SRAL `surf_type_class_20_ku` flag. The SRAL auxiliary concentration field is not used as the main concentration target because it is derived from the coarse-resolution OSI-430-a product. The surface-type flag was rounded to its integer class, and samples with a missing SRAL classification were excluded. At samples classified as open ocean, AMSR2 concentration was set to zero. At samples classified as floes or leads, nonzero AMSR2 values were retained, whereas zero values were replaced by the nearest finite nonzero AMSR2 value in the mapped two-dimensional field. Other samples retained the collocated AMSR2 value. This rule uses SRAL as categorical surface-type information rather than as a quantitative concentration retrieval. We refer to the resulting sparse target as SRAL-conditioned AMSR2; the dense gridded AMSR2 field is retained as a separate validation target.

Validation

For model-to-model assimilation, model fields corresponding to the pseudo-observation dates are used as a proxy target for comparison. For satellite-to-model assimilation, we report two complementary validation targets: dense gridded AMSR2 sea-ice concentration and sparse SRAL-conditioned AMSR2 concentration at SRAL-track locations. For forecast evaluation, the AMSR2 field at the forecast valid date is withheld from assimilation, so validation does not use future observations. The dense AMSR2 score should be interpreted as validation against a withheld AMSR2 product rather than an independent observing system. The sparse SRAL-conditioned AMSR2 target evaluates performance specifically at the track locations after surface-type conditioning.

Statistical evaluation

Statistical uncertainty was estimated with a moving-block bootstrap to account for temporal autocorrelation in the daily evaluation series. For the 2022 model-to-model and satellite-to-model assimilation experiments (2022-01-01 to 2022-07-19), each reported skill used $n = 200$ daily cases. The analysis and background metric series were aligned by common time index and pairs containing missing values were discarded before computing skill. For each bootstrap replicate, contiguous blocks of length $\tau = 14$ days were sampled with replacement, concatenated, and truncated to the original series length; skill was then recomputed as $1 - \overline{m}_{\text{analysis}}/\overline{m}_{\text{background}}$. We used 5000 bootstrap replicates and report the standard deviation of the bootstrap skill distribution as the uncertainty ($\pm$std). The same procedure was used for forecast metrics, with MAE, RMSE, and derived skill scores recomputed from the resampled daily forecast cases. The forecast-restart comparison used a common sample of $n = 200$ forecast cases (initialized 2023-01-01 to 2023-07-19) at each lead time for both validation targets and restart-transfer variants. For sparse forecast verification, non-finite ocean values in the NEMO sea-ice-concentration forecasts, which represent ice-free water, were evaluated as zero concentration; missing values outside the SRAL-track target remained excluded. These uncertainties quantify temporal sampling uncertainty only; they do not include uncertainty from VAE training seeds, model selection, or different evaluation years.

Model

As the single-field baseline, we use a VAE following a similar approach to [26], which was originally applied to temperature data from the ERA5 reanalysis. We also conducted experiments with a date-conditioned autoencoder in which dates were converted to DOY values. DOY was encoded as a cyclic variable using the sine/cosine transformation $(\sin(2\pi \cdot \mathrm{DOY}/366), \cos(2\pi \cdot \mathrm{DOY}/366))$, then transformed by a linear layer and concatenated in the latent space.

Training

The variational autoencoder learns to reconstruct target ocean and sea-ice fields. The base reconstruction term is mean squared error (MSE), whereas the KL divergence is computed for the Gaussian case [33]. Missing data over land were filled with physically plausible values. The reported four-field VAE training runs used the Lion optimizer [40] with learning rate 10^{-4}, batch size 8, and KL weight 0.01; the auxiliary-loss variants reported in the Results were trained for 30 epochs.

The auxiliary VAE variants modify only the training loss. The **multiscale** term adds sea-ice-concentration reconstruction penalties after average pooling at scales 1, 2, and 4, encouraging agreement at both local and coarse spatial scales. The **Sobel** term matches the Sobel-gradient magnitudes of the reconstructed and target concentration fields, encouraging sharper ice-water boundaries. The **binary** term penalizes $p(1-p)$ for reconstructed sea-ice concentration p, discouraging ambiguous intermediate values and favoring concentrations closer to 0 or 1. The **multiscale_block** term uses the same scales as the multiscale loss but applies separate L1 penalties at the fine and pooled resolutions, with larger relative weight on coarse block structure. For the multiscale+Sobel+binary loss, the multiscale term used scales 1, 2, and 4 with relative weights 1.0, 0.5, and 0.25, and the Sobel and binary terms each used weight 0.4. We also explored vector-quantized, hierarchical, mask-supervised, and temporally regularized variants of the VAE; none of these alternatives improved downstream assimilation performance over the simpler continuous-latent model with multiscale and edge-aware losses. We also experimented with adding a discriminator and an SSIM loss, but this did not yield noticeable improvements.

Assimilation

For the classical baselines, we used 3D-VAR, BLUE, and EnKF configurations with explicit covariance and localization choices tuned on the same development setup. These methods provide Gaussian or covariance-based references, whereas VAE-MDA uses the pretrained VAE decoder together with background and latent penalties as a learned regularizing prior. Their background covariances and localization use a compact-support polynomial correlation function [41] with a 100-km length scale; the full parameter settings of the classical baselines are given in the Supplementary Note on classical baseline configurations.

Assimilation is performed in the latent space of the autoencoder: the background state is encoded as $z_b = E(x_b)$, the latent variable is initialized at z_b, and only this latent variable is optimized while the pretrained encoder and decoder remain fixed; the analysis is recovered as $x_a = D(z^\star)$. In the reported 200-day experiments, latent optimization used Adam for 300 steps with learning rate 0.01, cosine annealing to a minimum learning rate of 10^{-4}, gradient clipping with norm 1.0, and batch size 1. Because daily sea-ice concentration changes are typically modest away from rapidly evolving ice-edge and melt-season regions, and the window-depth experiment of Figure 4 indicated an optimal accumulation depth of about three days, observations were accumulated over a three-day window (the current day plus the previous two days of available tracks).

The loss function $\mathscr{L}$ contains three terms: (1) the observation error, (2) the background-field error, and (3) the latent-space error. The second and third terms act as regularizers to prevent excessive deviation from the original model field and are therefore assigned smaller weights. The loss is stated here in its general multivariate form, with the background error evaluated on all decoded fields. In the reported experiments, the background MSE is restricted to the sea-ice-concentration field; this choice is motivated by the background-penalty ablation (Supplementary Note on the background-penalty ablation), in which the restriction improves concentration and ice-edge skill.

$$\begin{aligned}\mathscr{L}(x_a, y, x_b, z, z_b) &= w_y MSE(H(x_a), y) \\ &+ w_b MSE(x_a, x_b) + w_z MSE(z, z_b)\end{aligned} \tag{4}$$

In Eq. (4), w_y, w_b, and w_z correspond to `obs_weight`, `mod_weight`, and `lat_weight`, respectively. For model-to-model experiments we used $w_y = 1.0$ and $w_b = 0.1$. For satellite-to-model experiments we evaluated both $w_y = 1.0$ and a stronger-observation setting with $w_y = 6.0$, while keeping $w_b = 0.1$. The stronger-observation setting improved validation against the sparse SRAL-conditioned AMSR2 target and was therefore used for the reported satellite-to-model comparison and for the 2023 forecast-cycling restart experiments. The latent regularization was $w_z = 0.1$ for the single-field, three-field, and four-field models, and $w_z = 0.01$ for the four-field auxiliary-loss variants.

Bounds on latent-space corrections

We now make the role of the regularization terms in Eq. (4) explicit. Absorbing the MSE normalization constants into the weights w_y, w_b, w_z, the latent optimization problem becomes

$$\mathscr{L}(z) = w_y\|H(D(z)) - y\|_2^2 + w_b\|D(z) - x_b\|_2^2 + w_z\|z - z_b\|_2^2,$$

with $z_b = E(x_b)$ and analysis state $x_a = D(z^\star)$. Because D is a neural network, $\mathscr{L}$ is non-convex; the bounds below are stated for the idealized objective under the assumption that a global minimizer $z^\star \in \arg\min_z \mathscr{L}(z)$ exists. In the experiments, the latent optimization uses Adam initialized at z_b, so the computed analysis is an empirical local solution and the theorem motivates, but does not validate, the optimizer.

For a multivariate decoder, write $D(z) = (D_1(z), \ldots, D_K(z))$ with $D_j : \mathbb{R}^p \to \mathbb{R}^{n_j}$, $x_{0,j} = D_j(z_b)$, $x_{a,j} = D_j(z^\star)$; here $x_{0,j}$ denotes a block of the decoded background reconstruction $D(z_b)$, which differs from the corresponding block of x_b by the reconstruction error. Assume each D_j is Lipschitz on a neighborhood containing z_b and $z^\star$:

$$\|D_j(z) - D_j(z')\|_2 \le L_j\|z - z'\|_2. \tag{5}$$

Theorem 1 (Latent regularization constrains decoded multivariate increments)**.** *Under the assumptions above and the componentwise Lipschitz condition (5), with*

$$E_0 := w_y\|H(D(z_b)) - y\|_2^2 + w_b\|D(z_b) - x_b\|_2^2,$$

the following bounds hold:

$$\mathscr{L}(z^\star) \le E_0, \tag{6}$$

$$\|x_a - x_b\|_2 \le \sqrt{E_0/w_b}, \tag{7}$$

$$\|z^\star - z_b\|_2 \le \sqrt{E_0/w_z}, \tag{8}$$

$$\|x_{a,j} - x_{0,j}\|_2 \le L_j\sqrt{E_0/w_z} \quad (j = 1, \ldots, K). \tag{9}$$

When the encoder–decoder pair satisfies $D(z_b) = x_b$ (so that $x_{0,j} = x_{b,j}$ for every j), the reconstruction term vanishes and E_0 reduces to $w_y\|H(D(z_b)) - y\|_2^2$, so each bound scales linearly with the observation mismatch (full statement in the Supplementary Note on theoretical results). When the background penalty is restricted to the sea-ice-concentration field, as in the reported experiments, the same argument applies with the background terms of $\mathscr{L}$ and E_0 evaluated on that field only: the second bound then controls the concentration component of $x_a - x_b$, while the total-loss, latent, and per-field bounds hold verbatim.

Proof sketch. The total-loss bound follows from $\mathscr{L}(z^\star) \le \mathscr{L}(z_b)$ and $\|z_b - z_b\|_2 = 0$. Nonnegativity of each term then gives $w_b\|x_a - x_b\|_2^2$, $w_z\|z^\star - z_b\|_2^2 \le \mathscr{L}(z^\star)$, and the componentwise bound follows from (5). See the Supplementary Note on theoretical results for the full proof. □

Intuitively, even when the observation operator H acts on only part of the state, the shared latent representation forces every decoded field $x_{a,j}$ to stay close to its background reconstruction unless the observations strongly disagree with the encoded background. The reconstruction and latent regularization terms thus act as a trust region around the background, whose size is set by w_b and w_z relative to the observation weight w_y and the per-field Lipschitz constants L_j. In the concentration-only penalty variant used in the reported experiments, the direct background constraint acts on the concentration field, and the unobserved fields are constrained only through the latent term and the Lipschitz constants; this is consistent with the coupled cross-field increments observed in the experiments. When the encoder–decoder pair reconstructs the background well ($D(z_b) \approx x_b$), the trust region tightens further, since the right-hand side of every bound is then controlled by the observation mismatch alone. Because the per-field Lipschitz constants are not estimated numerically here, these bounds provide a qualitative interpretation of the regularization rather than a calibrated operational trust-region radius.

Physical balance and forecast impact are not implied by these bounds and remain empirical properties that must be assessed in the experiments.

NEMO restart integration

VAE-MDA produces multivariate analysis fields on the model grid, whereas the forecasting system must be initialized from NEMO/SI3 restart files. We therefore converted the analyzed fields into restart variables using two restart-transfer protocols. The first protocol modifies the SI3 sea-ice restart consistently with the analyzed sea-ice concentration and thickness (Algorithm 2). The second protocol, used in the coupled ice-ocean experiment, additionally transfers the analyzed sea-surface-temperature

Algorithm 2 Physical sea-ice restart correction

Require: Analyzed ice concentration and thickness c^a, h_i^a; SI3 restart variables
Clip c^a to $[0,1]$, remove trace ice below 0.01, and set $a_i' \leftarrow c^a$
Set missing or negative h_i^a to zero and then set $h_i^a \leftarrow 0.1$ m wherever $a_i' > 0$ and $h_i^a = 0$
Update ice and snow volumes: $v_i' \leftarrow a_i' h_i^a$ and $v_s' \leftarrow v_s^r a_i'/a_i^r$ where $a_i^r > 0$, with $v_s' = 0$ otherwise
Recompute snow–ice mass from v_i' and v_s' using fixed ice and snow densities, and rescale existing salt-content and ice/snow-energy restart variables in proportion to the corresponding volume changes
Assign a reference salinity of 7.4 psu to newly formed ice and initialize its two layer-energy variables from the prescribed ice volume and fixed coefficients
For remaining ice-dependent restart variables, including internal stress and oa_i, retain restart values where ice persists, set them to zero where analyzed ice is absent, and initialize newly ice-covered cells from the nearest pre-existing ice cell
Leave ocean restart variables unchanged in the ice-only protocol
return Corrected SI3 sea-ice restart files

Algorithm 3 Physically bounded thermodynamic ocean restart transfer

Require: Analyzed fields c^a, h_i^a, T_s^a; restart fields a_i^r, v_i^r, T_k^r, S_k^r; correction parameters defined below
Apply Algorithm 2 using c^a and h_i^a
Compute $\Delta a_i \leftarrow c^a - a_i^r$ and $\Delta v_i \leftarrow v_i' - v_i^r$ from the updated SI3 restart
Apply n_s passes of a normalized 3×3 box filter to $\mathbf{1}\{|\Delta a_i| \geq \tau_a \text{ or } |\Delta v_i| \geq \tau_v\}$ using symmetric boundary extension, clip the result to $[0,1]$, and set weights to zero where T_1^r is nonfinite
Estimate surface freezing temperature as $T_f \leftarrow -0.054 S_1^r$
Form raw $T_1^\star$ from clipped $T_s^a - T_1^r$ and freezing-point buffers using the sign of Δv_i
Clip $w(T_1^\star - T_1^r)$ by the local latent-heat bounds $[-B_T^-, B_T^+]$ and enforce $T_1^r + \delta T_1 \geq T_f - \varepsilon_{sc}$
Prescribe raw salinity increments from ice growth/loss and clip them by the local salt-budget bounds $[-B_S^-, B_S^+]$
for $k = 1, \ldots, K_u$ **do**
    Apply tapered updates $\lambda_k \delta T_1$ and $\lambda_k \delta S_1$, then clip to broad physical T and S ranges
end for
Leave velocity, sea-surface height, barotropic variables, and other dynamical restart variables unchanged
return Corrected coupled ice-ocean restart files

information into a physically bounded thermodynamic correction of the upper-ocean temperature and salinity restart variables (Algorithm 3).

Throughout this section, c^a, h_i^a, and T_s^a denote the analyzed sea-ice concentration, sea-ice thickness, and sea-surface temperature fields, respectively. The original restart variables are denoted with superscript r, and corrected restart variables are denoted with a prime. The SI3 and ocean restart variables used below are the ice concentration a_i, the ice and snow volumes per unit grid-cell area v_i and v_s, the ocean–ice coupling area oa_i, and the ocean temperature and salinity T_k and S_k at vertical level k; ocean level $k = 1$ corresponds to the uppermost active level. Algorithm 3 further uses the ice-concentration and ice-volume increments Δa_i and Δv_i, the activation thresholds τ_a and τ_v with smoothed correction weights $w \in [0,1]$, the local freezing temperature T_f, the freezing-point buffers b_g (ice gain) and b_l (ice loss) entering the raw level-1 temperature proposal $T_1^\star$, the resulting level-1 temperature and salinity increments δT_1 and δS_1, the thermodynamic bounds $B_T^\pm$ and $B_S^\pm$ defined below, the supercooling tolerance ε_{sc}, and the vertical taper coefficients λ_k applied over the K_u uppermost ocean levels.

The local temperature bounds are derived from a latent-heat scale. In the expressions below, ρ_i and ρ_0 are the sea-ice and reference seawater densities, c_p is the seawater specific heat capacity, L_f is the latent heat of fusion of ice, and H_{eff} is the effective depth of the corrected upper-ocean layer. The dimensionless factors γ_Q^+ and γ_Q^- scale the latent-heat budget for ice growth and ice loss, and Q_{min}^- and Q_{min}^+ are minimum heat allowances that permit small corrections where the ice-volume change is negligible. In the salinity bounds, S_w and S_i are the reference seawater and sea-ice salinities, η_S^+ and η_S^- scale the brine-rejection and meltwater salt budgets, and S_{min}^+ and S_{min}^- are the corresponding minimum salinity allowances. For ice growth, cooling is limited by

$$B_T^- = \frac{\gamma_Q^+ \rho_i L_f \max(\Delta v_i, 0) + Q_{min}^-}{\rho_0 c_p H_{eff}},$$

and for ice loss, warming is limited by

$$B_T^+ = \frac{\gamma_Q^- \rho_i L_f \max(-\Delta v_i, 0) + Q_{min}^+}{\rho_0 c_p H_{eff}}.$$

The salinity bounds use the corresponding brine-rejection or meltwater scale,

$$B_S^+ = S_{min}^+ + \frac{\eta_S^+ \rho_i \max(\Delta v_i, 0)(S_w - S_i)}{\rho_0 H_{eff}}, \qquad B_S^- = S_{min}^- + \frac{\eta_S^- \rho_i \max(-\Delta v_i, 0)(S_w - S_i)}{\rho_0 H_{eff}}.$$

Thus the analyzed sea-surface-temperature increment provides a raw thermodynamic proposal, but the imposed restart increment is limited by the diagnosed ice-volume change.

For the reported coupled ice-ocean restart experiment, the raw maximum cooling and warming magnitudes were 8°C and 4°C before applying the physical bounds. The raw salinity increment caps were 0.50 psu in ice-gain cells and 0.25 psu in ice-loss cells. The active-cell thresholds were $\tau_a = 0.0025$ and $\tau_v = 0.001$ m, the mask was smoothed for $n_s = 6$ passes, and the correction was applied to $K_u = 20$ upper-ocean levels with $H_{eff} = 20$ m. No separate land mask was applied during smoothing; exclusion occurred only where the surface restart temperature was nonfinite. We used $\rho_0 = 1025$ kg m^{-3}, $c_p = 4000$ J kg^{-1} K^{-1}, $\rho_i = 900$ kg m^{-3}, $L_f = 334000$ J kg^{-1}, $\gamma_Q^+ = \gamma_Q^- = 1.5$, $Q_{min}^- = Q_{min}^+ = 2.05 \times 10^7$ J m^{-2}, $S_w = 34$ psu, $S_i = 7$ psu, $\eta_S^+ = \eta_S^- = 1$, and $S_{min}^+ = S_{min}^- = 0.02$ psu. The vertical taper coefficients were

$$\lambda = (1.00, 1.00, 1.00, 1.00, 0.95, 0.90, 0.80, 0.70, 0.60, 0.50, 0.42, 0.35, 0.28, 0.22, 0.17, 0.13, 0.10, 0.07, 0.04, 0.02).$$

The freezing-point buffers were $b_g = -0.20$°C for ice-gain cells and $b_l = 1.00$°C for ice-loss cells, with a supercooling tolerance $\varepsilon_{sc} = 0.05$°C. Temperature was clipped to $[T_{\min}, T_{\max}] = [-3, 35]$°C and salinity to $[S_{\min}, S_{\max}] = [0, 40]$ psu. This restart transfer modifies only upper-ocean temperature and salinity while leaving ocean dynamical variables unchanged.

Use of large language models

Large language models were used to polish the writing, check for inconsistencies and grammar mistakes, and adapt the manuscript to the journal format. All scientific claims and analyses were performed and verified by the authors, and the final text was reviewed and approved by the authors.

Data availability

Sentinel-3 SRAL data are available through the Copernicus Data Space Ecosystem. AMSR2 data are available from the JAXA Global Change Observation Mission (GCOM-W). The WRF–NEMO-SI3 modeling system used to produce the assimilation background is documented in the Borey regional hindcast dataset description [36]. The specific NEMO forecast outputs and preprocessed datasets used in this study are not included in the public code repository and are available from the corresponding author upon reasonable request.

Code availability

Source code for VAE training, the VAE-MDA assimilation algorithm, the classical 3D-VAR, BLUE, and EnKF baselines, evaluation, and the NEMO/SI3 restart-transfer protocols (Algorithms 2 and 3) is publicly available in the VAE-MDA GitHub repository.

References

[1] Babb, D. *et al.* The stepwise reduction of multiyear sea ice area in the Arctic Ocean since 1980. *JGR Ocean.* **128**, e2023JC020157, DOI: https://doi.org/10.1029/2023JC020157 (2023).

[2] Kortsch, S., Primicerio, R., Fossheim, M., Dolgov, A. & Aschan, M. Climate change alters the structure of arctic marine food webs due to poleward shifts of boreal generalists. *Proc. Royal Soc. B* **282**, 20151546, DOI: https://doi.org/10.1098/rspb.2015.1546 (2015).

[3] Shu, Y. *et al.* Path planning for ships assisted by the icebreaker in ice-covered waters in the Northern Sea Route based on optimal control. *Ocean. Eng.* **267**, 113182, DOI: https://doi.org/10.1016/j.oceaneng.2022.113182 (2023).

[4] Dvoynikov, M. *et al.* New concepts of hydrogen production and storage in Arctic region. *Resources* **10**, 3, DOI: https://doi.org/10.3390/resources10010003 (2021).

[5] Pan, R. *et al.* Future Arctic climate change in CMIP6 strikingly intensified by NEMO-family climate models. *Geophys. Res. Lett.* **50**, e2022GL102077, DOI: https://doi.org/10.1029/2022GL102077 (2023).

[6] Allende, S. *et al.* Impact of ocean vertical-mixing parameterization on Arctic sea ice and upper-ocean properties using the NEMO-SI3 model. *Geosci. Model. Dev.* **17**, 7445–7466, DOI: https://doi.org/10.5194/gmd-17-7445-2024 (2024).

[7] Moreno-Chamarro, E. *et al.* The very-high-resolution configuration of the EC-Earth global model for HighResMIP. *Geosci. Model. Dev.* **18**, 461–482, DOI: https://doi.org/10.5194/gmd-18-461-2025 (2025).

[8] Olason, E., Rampal, P. & Dansereau, V. On the statistical properties of sea-ice lead fraction and heat fluxes in the Arctic. *The Cryosphere* **15**, 1053–1064, DOI: https://doi.org/10.5194/tc-15-1053-2021 (2021).

[9] Selivanova, J., Iovino, D. & Cocetta, F. Past and future of the Arctic sea ice in High-Resolution Model Intercomparison Project (HighResMIP) climate models. *The Cryosphere* **18**, 2739–2763, DOI: https://doi.org/10.5194/tc-18-2739-2024 (2024).

[10] Copernicus Climate Change Service (C3S). Copernicus Climate Data Store. [Online]. Accessed: 2025-03-19.

[11] National Aeronautics and Space Administration. NASA Earth Observing System Data and Information System (EOSDIS). [Online].

[12] Carrassi, A., Bocquet, M., Bertino, L. & Evensen, G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Clim. Chang.* **9**, e535, DOI: https://doi.org/10.1002/wcc.535 (2018).

[13] Bannister, R. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Q. J. Royal Meteorol. Soc.* **134**, 1951–1970, DOI: https://doi.org/10.1002/qj.339 (2008).

[14] Bannister, R. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Q. J. Royal Meteorol. Soc.* **134**, 1971–1996, DOI: https://doi.org/10.1002/qj.340 (2008).

[15] Kalman, R. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45, DOI: https://doi.org/10.1115/1.3662552 (1960).

[16] Jazwinski, H. *Stochastic Processes and Filtering Theory* (Academic Press, New York and London, 1970).

[17] Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10143–10162, DOI: https://doi.org/10.1029/94JC00572 (1994).

[18] Evensen, G. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean. Dyn.* **53**, 343–367, DOI: https://doi.org/10.1007/s10236-003-0036-9 (2003).

[19] Courtier, P. *et al.* The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. Royal Meteorol. Soc.* **124**, 1783–1807, DOI: https://doi.org/10.1002/qj.49712455002 (1998).

[20] Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J.-F. & Simmons, A. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. Royal Meteorol. Soc.* **126**, 1143–1170, DOI: https://doi.org/10.1002/qj.49712656415 (2000).

[21] Zhao, Q. *et al.* Applications of deep learning in physical oceanography: a comprehensive review. *Front. Mar. Sci.* **11**, 1396322, DOI: https://doi.org/10.3389/fmars.2024.1396322 (2024).

[22] Arcucci, R., Zhu, J., Hu, S. & Guo, Y. Deep data assimilation: Integrating deep learning with data assimilation. *Appl. Sci.* **11**, 1114, DOI: https://doi.org/10.3390/app11031114 (2021).

[23] Farchi, A., Bocquet, M., Laloyaux, P., Bonavita, M. & Malartic, Q. A comparison of combined data assimilation and machine learning methods for offline and online model error correction. *J. Comput. Sci.* **55**, 101468, DOI: https://doi.org/10.1016/j.jocs.2021.101468 (2021).

[24] Mack, J., Arcucci, R., Molina-Solana, M. & Guo, Y. Attention-based convolutional autoencoders for 3D-variational data assimilation. *Comput. Methods Appl. Mech. Eng.* **372**, 113291, DOI: https://doi.org/10.1016/j.cma.2020.113291 (2020).

[25] Peyron, M. *et al.* Latent space data assimilation by using deep learning. *Q. J. Royal Meteorol. Soc.* **147**, 3759–3777, DOI: https://doi.org/10.1002/qj.4153 (2021).

[26] Melinc, B. & Zaplotnik, Z. 3D-Var data assimilation using a variational autoencoder. *Q. J. Royal Meteorol. Soc.* **150**, 2273–2295, DOI: https://doi.org/10.1002/qj.4708 (2024).

[27] Penny, S. *et al.* Integrating recurrent neural networks with data assimilation for scalable data-driven state estimation. *J. Adv. Model. Earth Syst.* **14**, e2021MS002843, DOI: https://doi.org/10.1029/2021MS002843 (2022).

[28] Lisæter, K. A., Rosanova, J. & Evensen, G. Assimilation of ice concentration in a coupled ice–ocean model, using the ensemble Kalman filter. *Ocean. Dyn.* **53**, 368–388, DOI: https://doi.org/10.1007/s10236-003-0049-4 (2003).

[29] Brasseur, P. Ocean data assimilation using sequential methods based on the Kalman filter. In *Ocean Weather Forecasting: An Integrated View of Oceanography*, 271–316, DOI: https://doi.org/10.1007/1-4020-4028-8_10 (Springer Science+Business Media B.V., 2006).

[30] Zavala-Romero, O., Bozec, A., Chassignet, E. & Miranda, J. Convolutional neural networks for sea surface data assimilation in operational ocean models: test case in the Gulf of Mexico. *Ocean. Sci.* **21**, 113–132, DOI: https://doi.org/10.5194/os-21-113-2025 (2025).

[31] Hatfield, S. *et al.* Building tangent-linear and adjoint models for data assimilation with neural networks. *J. Adv. Model. Earth Syst.* **13**, e2021MS002521, DOI: https://doi.org/10.1029/2021MS002521 (2021).

[32] Barthélémy, S., Brajard, J., Bertino, L. & Counillon, F. Super-resolution data assimilation. *Ocean. Dyn.* **72**, 661–678, DOI: https://doi.org/10.1007/s10236-022-01523-x (2022).

[33] Kingma, D. & Welling, M. Auto-encoding variational Bayes. *Int. Conf. on Learn. Represent. (ICLR)* DOI: https://doi.org/10.48550/arXiv.1312.6114 (2014).

[34] Fan, H. *et al.* Physically consistent global atmospheric data assimilation with machine learning in latent space. *Sci. Adv.* **12**, eaea4248, DOI: https://doi.org/10.1126/sciadv.aea4248 (2026).

[35] de Gelis, I., Colin, A. & Longepe, N. Prediction of categorized sea ice concentration from Sentinel-1 SAR images based on a fully convolutional network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **14**, 5831–5841, DOI: 10.1109/JSTARS.2021.3074068 (2021).

[36] Ivanov, V. *et al.* Borey: A high-resolution regional atmosphere–ocean–sea ice–wave forecasting system and hindcast dataset for the Barents and Kara seas. arXiv preprint arXiv:2608.09957, DOI: https://doi.org/10.48550/arXiv.2608.09957 (2026).

[37] Aublanc, J. *et al.* Sentinel-3 altimetry thematic products for hydrology, sea ice and land ice. *Sci. Data* **12**, 714, DOI: https://doi.org/10.1038/s41597-025-04956-3 (2025).

[38] Spreen, G., Kaleschke, L. & Heygster, G. Sea ice remote sensing using AMSR-E 89 GHz channels. *J. Geophys. Res.* **113**, C02S03, DOI: 10.1029/2005JC003384 (2008).

[39] Pang, X. *et al.* Comparison between AMSR2 sea ice concentration products and pseudo-ship observations of the Arctic and Antarctic sea ice edge on cloud-free days. *Remote. Sens.* **10**, 317, DOI: https://doi.org/10.3390/rs10020317 (2018).

[40] Chen, X. *et al.* Symbolic discovery of optimization algorithms. arXiv preprint arXiv:2302.06675, DOI: 10.48550/arXiv.2302.06675 (2023).

[41] Gaspari, G. & Cohn, S. Construction of correlation functions in two and three dimensions. *Q. J. Royal Meteorol. Soc.* **125**, 723–757, DOI: https://doi.org/10.1002/qj.49712555417 (1999).

Acknowledgements

The work was supported by the grant for research centers in the field of AI provided by the Ministry of Economic Development of the Russian Federation in accordance with the agreement 000000C313925P4F0002 and the agreement with Skoltech № 139-10-2025-033.

Author contributions

V.I. led the research and carried out the main body of the work. V.I. formulated the initial concept, formalized VAE-MDA, implemented the computational prototype, conducted the main analysis, prepared part of the visualizations, and substantially revised and developed the manuscript into its current form. L.A. contributed to the latent-space assimilation idea, performed a substantial part of the experiments and parameter tuning, supported data preparation, prepared part of the visualizations, contributed to interpretation of the results, prepared the initial manuscript draft, and revised the manuscript. V.V. supervised the study, provided research guidance, contributed to interpretation of the results, and edited the manuscript. E.B. supervised the study and contributed to scientific discussion and manuscript revision. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf