Four-dimensional variational data assimilation with a sea-ice thickness emulator

Charlotte Durand[1], Tobias Sebastian Finn[1], Alban Farchi[1,*], Marc Bocquet[1], Julien Brajard[2], and Laurent Bertino[2]

[1]CEREA, École des Ponts and EDF R&D, Institut Polytechnique de Paris, Île-de-France, France
[2]Nansen Environmental and Remote Sensing Center, 5007 Bergen, Norway
[*]Now at European Center for Medium-Range Weather Forecasts, Bonn, Germany

Correspondence: Charlotte Durand (charlotte.durand@enpc.fr)

Abstract. Developing operational data assimilation systems for sea-ice models is challenging, especially using a variational approach due to the absence of adjoint models. NeXtSIM, a sea-ice model based on a brittle rheology paradigm, enables high-fidelity simulations of sea-ice dynamics at mesoscale resolution (~ 10 km) but lacks an adjoint. By training a neural network as an Arctic-wide emulator for sea-ice thickness based on mesoscale simulations with neXtSIM, we gain access to an adjoint.

5 Building on this emulator and its adjoint, we introduce a four-dimensional variational (4D–Var) data assimilation system to correct the emulator's bias and to better position the marginal ice zone (MIZ). Firstly, we perform twin experiments to demonstrate the capabilities of this 4D–Var system and to evaluate two approximations of the background covariance matrix. These twin experiments demonstrate that the assimilation improves the positioning of the MIZ and enhances the forecast quality, achieving an average reduction in sea-ice thickness root-mean-squared error of 0.8 m compared to the free run. Secondly, we

10 assimilate real CS2SMOS satellite retrievals with this system. While the assimilation of these rather smooth retrievals amplifies the loss of small-scale information in our system, it effectively corrects the forecast bias. The forecasts of our 4D–Var system achieve a similar performance as the operational sea-ice forecasting system neXtSIM-F. These results pave the way to the use of deep learning-based emulators for 4D–Var systems to improve sea-ice modeling.

1 Introduction

15 Combining observational data with sea-ice models by data assimilation can improve the accuracy of sea-ice forecasts for practical applications such as maritime routing but is computationally expensive. Deep learning offers a solution by providing efficient neural network emulations that serve as auto-differentiable alternatives to costly physical models, which often have no adjoint available. This enables the implementation of four-dimensional variational data assimilation systems (4D–Var), which can potentially enhance the accuracy and scalability of sea-ice forecasts.

20 In this paper, we introduce a 4D–Var system based on the sea-ice thickness (SIT) emulator previously developed by Durand et al. (2024). This emulator aims at reproducing the evolution of the sea-ice thickness as modeled by ~~neXtSIM~~the state-of-the-art sea-ice model neXtSIM (Rampal et al., 2016; Ólason et al., 2022). We initially demonstrate the feasibility of employing this emulator and its adjoint within a 4D–Var framework through twin experiments. Subsequently, we present

promising outcomes derived from applying this approach to real observational data, with the assimilation of SIT retrievals

25 from the merged ~~CryoSat-SMOS product~~ product of Cryosat-2 altimeter and the Soil Moisture and Ocean Salinity (SMOS) radiometer, called CS2SMOS (Ricker et al., 2017). Forecasts obtained with our system are comparable to those of the operational neXtSIM-F system (Williams et al., 2021)~~.~~, with -F standing for "Forecast".

~~Various data assimilation techniques are currently implemented in operational sea-ice forecasting systems (Liu et al., 2019).~~
~~Ensemble Kalman filters with flow-dependent covariances are predominantly utilized in sea-ice forecasting systems, (Sakov et al., 2012; Ki~~
30 ~~, particularly for assimilating key variables like sea-ice concentration (Massonnet et al., 2015). Other techniques use covariances~~
~~that are static in time such as nudging (Lindsay and Zhang, 2006; Tietsche et al., 2013), optimal interpolation (Wang et al., 2013; Ji et al., 2~~
~~, and three-dimensional variational assimilation (Hebert et al., 2015; Toyoda et al., 2015; Lemieux et al., 2015) and are employed~~
~~in diverse systems (Caya et al., 2010; Donlon et al., 2012; Zuo et al., 2019).~~

Arctic sea-ice thickness is heavily dependent on the sea ice rheological model. ~~The state-of-the-art sea-ice model neXtSIM~~
35 ~~(Rampal et al., 2016; Ólason et al., 2022)~~ NeXtSIM has been developed around brittle rheologies (Girard et al., 2011; Dansereau et al., 2016) to better parameterize the observed small-scale processes of sea ice. Run at a mesoscale resolution of approximately $10\,\mathrm{km}$, the model successfully simulates the observed scaling and multifractal properties of sea ice across space and time (Rampal et al., 2019; Bouchat et al., 2022). Until now, two data assimilation systems have been developed for neXtSIM: an ensemble Kalman filter (Cheng et al., 2023b) has been tested to assimilate observations of the sea-ice thickness and concen-
40 tration~~, but for~~. For operational forecasting, the neXtSIM-F system resorted instead to a simple nudging technique (Williams et al., 2021).

Various data assimilation techniques are currently implemented in operational sea-ice forecasting systems (Liu et al., 2019). Ensemble Kalman filters with flow-dependent covariances are often utilized in sea-ice forecasting systems, (Sakov et al., 2012; Kimmritz et , particularly for assimilating key variables like sea-ice concentration (Massonnet et al., 2015). Other techniques use covariances
45 that are static in time such as optimal interpolation (Wang et al., 2013; Ji et al., 2015), or three-dimensional variational assimilation (Hebert et al., 2015; Toyoda et al., 2015; Lemieux et al., 2015) and are employed in diverse systems (Caya et al., 2010; Donlon et al., 2012 . Initialization techniques using nudging (Lindsay and Zhang, 2006; Tietsche et al., 2013) have also been implemented.

In this work, we focus on four-dimensional variational data assimilation methods (4D–Var), which rely on a cost function whose efficient minimization requires gradients and the adjoint of the model~~as well as a background term that incorporates~~
50 ~~prior information about the state of the system~~. The idea behind (strong constraints) four-dimensional variational methods is to estimate a model trajectory that fits ~~the observations at best~~ best the observations throughout a time period (Sasaki, 1970; Talagrand and Courtier, 1987), called data assimilation window (DAW). ~~To propagate~~ The propagation of the gradient information from the observational time backwards in time within the DAW ~~, the cost function minimization implicitly depends~~ relies on the model's adjoint during the cost function minimization. Hence, by updating the analysis at initialization time, the
55 data assimilation accounts for all parts of information up to the end of the DAW.

Adjoints for sea-ice models, including its rheology, are rarely developed because of potential numerical instabilities ~~(Fenty and Heimbach~~ (Kauker et al., 2009; Fenty and Heimbach, 2013). While adjoints for simplified free-drift models are achievable (Koldunov et al., 2017), they yield limited realism for full Arctic sea-ice simulations. Usui et al. (2016); Toyoda et al. (2015, 2019) de-

veloped an adjoint of their sea-ice model, which relies on an ~~slightly simplified~~ elasto-visco-plastic (EVP) rheology scheme (Hunke and Dukowicz, 1997) within a coupled ocean-sea ice model framework to avoid numerical instabilities. Their approach includes a sensitivity analysis of the adjoint, particularly targeting grid-cells in the Marginal Ice Zone (MIZ) and the central Arctic. Koldunov et al. (2017) ~~incorporated an adjoint based on a viscous-plastic (VP) rheology~~ studied sea-ice concentration assimilation with an adjoint (Hibler, 1979) in the MIT-gcm model. The rheology in those adjoints is either discarded or simplified.

Recent advances in deep learning for sea-ice modeling encompass a range of applications, from the full emulation of variables such as sea-ice thickness (Durand et al., 2024), probabilities of sea-ice coverage (Andersson et al., 2021), or sea-ice concentration itself (Liu et al., 2021), to more specialized tasks like model error correction (Finn et al., 2023) or the emulation of melt ponds (Driscoll et al., 2024). Additionally, techniques to integrate neural networks with data assimilation have been proposed, including learning data assimilation increments for model bias correction (Gregory et al., 2023, 2024a), as well as the calibration of sea-ice forecasts (Palerme et al., 2023).

In the present work, we exploit the sea-ice thickness emulator developed by Durand et al. (2024) for both forward and adjoint modeling, as needed for the 4D–Var data assimilation system. The principle of using an emulator for 4D–Var was introduced by Hatfield et al. (2021) and showcased in Lorenz 63 toy examples (Chennault et al., 2021). Recently, this principle was extended to a Numerical Weather Prediction (NWP) emulator based on ERA5 data (Xiao et al., 2023). However, our study represents the first 4D–Var data assimilaton system that is built around an emulator specifically designed to capture the evolution of sea ice and the adjoint of the dynamics.

We explore two setups in this paper. Firstly, we focus on twin data assimilation experiments where observations are artificially generated by introducing noise into neXtSIM simulation outputs. Secondly, we assimilate real satellite observations from the merged CS2SMOS product. The paper is organized as follows. Section 2 presents the sea-ice model ~~,~~ and the additional forcings~~, and the observations~~. Section 3 provides a brief description of the emulator. Section 4 outlines the 4D–Var framework and the metrics used for the results evaluation. Results from twin experiments and with real observations are presented in Section ~~??,~~ 5 and Section 6, respectively, followed by a discussion in Section 7 and conclusions in Section 8.

2   Physical model simulation and ~~observations~~preprocessing of atmospheric forcings

In this section, we begin by introducing the geophysical sea-ice model, neXtSIM, which serves as the ground model for our results~~(Sec. ??). Following this, we describe the observations employed in the 4D–Var scheme. Specifically, we outline the observations simulated based on neXtSIM in Sec. 5.1, followed by real CS2SMOS retrievals in Sec. 6.1.~~

2.1   ~~NeXtSIM sea-ice thickness~~

. We then present the atmospheric forcings used as additional inputs for the emulator. NeXtSIM is a state-of-the-art sea-ice model (Rampal et al., 2016) built around brittle rheologies (Girard et al., 2011; Dansereau et al., 2016). In the here-used simulations (Boutin et al., 2023), neXtSIM is employed with the brittle Bingham-Maxwell rheology ~~(?)~~ (Ólason et al., 2022)

to replicate the observed subgrid-scale behavior of sea ice. Originally run on a Lagrangian triangular mesh, neXtSIM's outputs are projected onto a Eulerian curvilinear grid, forming the basis of our surrogate model (Durand et al., 2024). Additionally, the sea-ice model is coupled with the ~~ocean component of the modeling framework NEMO~~Nucleus for European Modelling of the Ocean (NEMO) framework's ocean model, OPA (version 3.6, Madec et al., 1998; Rousset et al., 2015). For further information about the model and its setup, we refer to Boutin et al. (2023).

In this study, we predict the sea-ice thickness with a neural network which is trained on simulations spanning from 2009 to 2016. The simulations are run on the regional CREG025 mesh configuration (Talandier and Lique, 2021), a regional subset of the global ORCA025 configuration developed by the Drakkar consortium (Bernard et al., 2006). The simulated area covers the Arctic and parts of the North Atlantic down to $27°$N latitude, with a nominal horizontal resolution of $0.25°$ ($\simeq 12$ km in the Arctic basin). The data is cropped in lower latitudes and areas in Eastern Europe and America where no sea ice is present, and then coarse-grained by averaging over a 4×4 window, resulting in a final resolution of 128×128 grid-cells. A simulated sea-ice thickness snapshot is displayed in Fig. 1a).

Let $\mathbf{x} \in \mathbb{R}^{128 \times 128}$ be the sea-ice thickness. The normalized sea-ice thickness $\tilde{\mathbf{x}}$ is then defined by

$$\tilde{\mathbf{x}} = \frac{\mathbf{x} - \mu_{\text{SIT}}}{\sigma_{\text{SIT}}}, \tag{1}$$

with μ_{SIT} the globally averaged sea-ice thickness and σ_{SIT} the global standard deviation, computed over all the coarse-grained grid-cells in the training dataset (2009 to 2016). The subtraction and division are pointwise operations. Masking land-covered cells within the original 128×128 grid-cells, $N_z = 8871$ remain unmasked covered by either open water or sea ice. Hence, the data assimilation is performed on the 1-D state vector $\tilde{\mathbf{x}}^{1D} \in \mathbb{R}^{N_z}$ which represents the normalized sea-ice thickness on the unmasked grid-cells. We will further drop the 1D superscript for the sake of readability.

We also consider the 2 m temperature (T2M), and the atmospheric u- and v-velocities in 10 m height (U10 and V10) from the ERA5 reanalysis dataset (Hersbach et al., 2020). Interpolated onto the native Eulerian curvilinear grid with nearest neighbors, forcings at time t, $t + 6$h and $t + 12$h are then coarse-grained, normalized and added as predictors to the input of the neural network, as commonly done in sea-ice forecasting (Grigoryev et al., 2022). The normalization of all input fields in the neural network is a common practice to stabilize and speed up the training (Ioffe and Szegedy, 2015).

2.1 ~~Observations~~

2.0.1 ~~Simulated observations from neXtSIM~~

~~In the first approach, a twin experiment setup is employed, wherein synthetic observations are generated by adding noise to neXtSIM simulations from 2017 and 2018. Using the observation error variance $\sigma_{\text{obs}}^2 = 0.4^2$, we define several types of~~

perturbed observations,

$$\tilde{\mathbf{x}}_{t,obs}^{G} = \tilde{\mathbf{x}}_t + \boldsymbol{\epsilon}^{G}, \qquad\qquad\qquad \boldsymbol{\epsilon}^{G} \sim \mathcal{N}(0, \sigma_{obs}^2),$$

$$\tilde{\mathbf{x}}_{t,obs}^{LN} = \mu_{SIT} + \sigma_{SIT}\mathbf{x}_t \exp\left(\boldsymbol{\epsilon}^{LN}\right), \qquad\qquad \boldsymbol{\epsilon}^{LN} \sim \mathcal{N}\left(0, \sigma_{obs} - \frac{1}{2}\sigma_{obs,}^2\right),$$

$$\tilde{\mathbf{x}}_{t,obs}^{cond-clipped} = \min\left(\tilde{\mathbf{x}}_t\left(1 + \boldsymbol{\epsilon}^{G}\right), \min_{SIT}\right).$$

with exp the exponential function. The Gaussian observation noise, as defined in Eq. (13a), is an idealized case, tailored to the common assumptions of 4D-Var, to test an adaptive inflation scheme, which will be defined later. Equation (13b) specifies a log-normal distribution for the noise, as more commonly encountered in sea-ice observations from satellites. Furthermore, a variant to log-normal noise is introduced in Eq. (13c) by adding a fraction of the sea-ice thickness and incorporating clipping to the normalized minimum ($\min_{SIT}$). This approach ensures that the observations remain confined to the physical bound of sea ice, unlike Gaussian noise. Examples of the different noises are shown in Fig. 1b)-d).

Snapshots of neXtSIM SIT (a) and different type of observations (b) Gaussian noise, (c) log-normal noise (LN) and (d) conditioned noise (cond-clipped). The colorbar for panels (b), (c) and (d) is shared and displayed on the right.

These noise definitions yield different noise magnitudes. The log-normal noise, defined in Eq. (13b), provides a more significant spread, especially for thicker ice. In average, the log-normal noise definition results in a standard deviation of 0.35 m, because of the skewness of the log-normal law, whereas the conditioned noise, defined in Eq. (13c), results in a smaller standard deviation of 0.29 m.

2.0.1 Real observations: combined Cryosat2-SMOS retrieval

The dataset of CS2SMOS (Ricker et al., 2017) retrievals provides real observations that are assimilated into our surrogate model. The retrievals merge observations from CryoSat-2 (Kurtz and Harbeck, 2017), known for its observations of thick and perennial sea ice, and from SMOS (Tian-Kunze et al., 2014), used to infer the thickness of thin ice. Merged weekly to account for the different temporal resolution of CryoSat-2 and SMOS observations, the retrievals are available in a daily moving window average. Note that the CS2SMOS is the result of Kriging and has been considerably smoothed in the process, even when compared to a weekly average of neXtSIM, as illustrated in Fig. 6.

CS2SMOS retrievals are only available on grid-cells covered by sea ice, and no information is available on grid-cells with open water. This creates a temporally changing mask, and we assume that grid-cells without information contain no sea ice.

Additionally, the CS2SMOS retrievals come with their own errors and uncertainties (Ricker et al., 2017). Based on the diagnostics of Desroziers et al. (2005), Xie et al. (2018) proposed an empirical formula for the observation error variance $\sigma_{obs,CS2SMOS}^2$, as an increasing function of ice thickness h_{ice},

$$\sigma_{obs,CS2SMOS}^2 = \begin{cases} \min(0.2, 0.02e^{1.8(h_{ice}-3)}) & \text{if } h_{ice} > 3\text{m}, \\ \max(0.02, 0.1e^{-1.5h_{ice}}) & \text{otherwise.} \end{cases}$$

3 Surrogate model

In this section, we describe our surrogate model, which has the same structure as the emulator previously developed in Durand et al. (2024), with the only update being that it is trained to account for the positivity of sea-ice thickness.

The surrogate model g_θ predicts the full sea-ice thickness $\tilde{\mathbf{x}}_{t+\Delta t}$ with a $\Delta t = 12\,\mathrm{h}$ lead time. The neural network f_θ with its weights and biases θ is trained to predict the evolution of the SIT after $\Delta t = 12\,\mathrm{h}$ based on the initial conditions $\tilde{\mathbf{x}}_t$ and given atmospheric forcings $\mathbf{F}$, $f_\theta(\tilde{\mathbf{x}}_t, \mathbf{F}_t) \approx \tilde{\mathbf{x}}_{t+\Delta t} - \tilde{\mathbf{x}}_t$. Added to ~~the initial conditions~~ $\tilde{\mathbf{x}}_t$, this results in the prediction of the full SIT field $\tilde{\mathbf{x}}_{t+\Delta t}$,

$$\tilde{\mathbf{x}}_{t+\Delta t} = g_\theta(\tilde{\mathbf{x}}_t, \mathbf{F}_t) \tag{2a}$$

$$= \mathrm{Relu}\big(\tilde{\mathbf{x}}_t + f_\theta(\tilde{\mathbf{x}}_t, \mathbf{F}_t)\big), \tag{2b}$$

with the point-wise activation function, ~~$\mathrm{Relu}(\tilde{\mathbf{x}}) = \max(\min_{\mathrm{SIT}}, \tilde{\mathbf{x}})$~~ $\mathrm{Relu}(\tilde{\mathbf{x}}) = \max(\mathrm{SIT}_{\min}, \tilde{\mathbf{x}})$, limiting the output to the lower physical bound in the normalized space. Note that f_θ is an intermediate emulator. As similarly done in Durand et al. (2024), the idea is to simplify the problem by focusing on the tendencies that happen in $12\,\mathrm{h}$ instead of predicting directly the full state, where small changes can be harder to predict. By splitting the whole learning process, we ensure that the emulator learns the SIT evolution, and in a second time that g_θ respects the positiveness constraint.

The neural network is trained with a mean-squared error loss between the predicted sea-ice thickness and the targeted sea-ice thickness as simulated by neXtSIM. The main part of the loss function is defined by a pixel-wise mean-squared error (MSE) on all ~~$N_x \times N_y$~~ $N_x \times N_y = 128 \times 128$ grid-cells, multiplied by the land-sea mask,

$$\mathcal{L}_{\mathrm{local}}(\mathbf{x}, \hat{\mathbf{x}}) = \mathrm{MSE}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{N_x \cdot N_y} \sum_{i}^{N_x} \sum_{j}^{N_y} (x_{i,j} - \hat{x}_{i,j})^2. \tag{3}$$

Note that, to simplify the equation, we did not include here the ~~land~~ land-sea mask, which is applied in the numerical implementation to compute the loss solely on grid cells with open ocean and sea ice. To address the systematic bias of the surrogate model and to mitigate its influence, which is already accounted for in the MSE loss, we introduce an additional penalty term

to the loss function,

$$\mathcal{L}_{\text{global}}(\mathbf{x}, \hat{\mathbf{x}}) = \left(\frac{1}{N_x \cdot N_y} \sum_i^{N_x} \sum_j^{N_y} (x_{i,j} - \hat{x}_{i,j}) \right)^2 . \tag{4}$$

Note in particular that the loss in Eq. (4) is squared after averaging, differing from Eq. (3). The total loss is then given by

$$\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}}) = \mathcal{L}_{\text{local}}(\mathbf{x}, \hat{\mathbf{x}}) + \lambda \mathcal{L}_{\text{global}}(\mathbf{x}, \hat{\mathbf{x}}), \tag{5}$$

with λ weighting the two terms. First, we pre-train the model f_θ, omitting the clipping in Eq. (2b) during training, with a first loss. As in Durand et al. (2024), we use $\lambda = 100$. Trained until convergence, we select the best model in the validation dataset. Secondly, we fine-tune g_θ to account for the clipping with a second loss. Here, we select $\lambda = 10$ as penalty weight, striking a good balance between $\mathcal{L}_{\text{local}}$ and $\mathcal{L}_{\text{global}}$ which have a different magnitude during second training. Results of the training and inference of the surrogate are presented in Appendix A.

4   Four-dimensional variational data assimilation

In this section, we describe the experimental setup of our 4D–Var, which is based on the surrogate model and its adjoint, as previously discussed in Sec. 3. Note that we are focusing on full 4D–Var rather than incremental 4D–Var.

4.1   4D–Var setup

The ~~length of the DAW is set to $N_{\text{daw}} = 16$ days which corresponds to 32 iterations of the surrogate model. In the twin experiment setup, observations are acquired every 2 days (every $N_f = 4$ iterations). The truth is given by neXtSIM, and the~~ forecast model is our data-driven emulator of neXtSIM. $\tilde{\mathbf{y}}_k$ represents the ~~k-th observation~~ observation at time k, $\tilde{\mathbf{x}}^b$ represents the background state, $\tilde{\mathbf{x}}^a$ represent the analysis and $\tilde{\mathbf{x}}_0$ the first guess. $\mathbf{B}$ is the background error covariance matrix and its specific parameterizations will be described below. $\mathbf{R}$ is the observation error covariance matrix ~~. $\mathbf{R}$ and~~ is defined by $\mathbf{R} = \sigma_{\text{obs}}^2 \mathbf{I}$, with $\mathbf{I}$ the identity matrix. To initialize the cycling of the data assimilation, we start with a field as simulated by neXtSIM for the 1st of January 2016, which is contained in our training dataset and which can be seen as sample from the climatology for the starting day.

For twin experiments, observations are generated directly in the model space, requiring no special preprocessing, as opposed to real observations, as explained in Sec. 6.1. ~~After that, in both cases, the~~ The observation operator $\mathcal{H}$ is simply defined as a diagonal matrix $\mathbf{H}$.

In this study, the 4D–Var is cycled across N_{cycle} cycles. The state at the end of each DAW is used as first guess and background state for the next DA cycle. To evaluate the 4D–Var system, forecasts are run for 45 days after the end of the DAW in the twin experiment case, and for 9 days in the real observations case.

Two types of 4D–Var are evaluated. 4D–Var-diag, which ~~correspond~~ corresponds to the use of a diagonal matrix as ~~background matrix~~ covariance matrix of background errors $\mathbf{B}$ (see Sec. 4.2), and 4D–Var-EOF in which the minimization is carried out in the empirical orthogonal function (EOF) space (see Sec. 4.3).

4.2 4D–Var with diagonal B matrix: 4D–Var-diag

The cost function associated to the 4D–Var minimization problem is

$$\mathcal{J}(\tilde{\mathbf{x}}_0) = \mathcal{J}^{\mathrm{b}} + \mathcal{J}^{\mathrm{o}} \tag{6a}$$

$$= \frac{1}{2} \left\| \tilde{\mathbf{x}}_0 - \tilde{\mathbf{x}}^{\mathrm{b}} \right\|^2_{(\lambda_{\mathrm{inf}}^2 \mathbf{B})^{-1}} + \frac{1}{2} \sum_{k=1}^{K} \left\| \tilde{\mathbf{y}}_k - \mathbf{H}_k \tilde{\mathbf{x}}_k \right\|^2_{\mathbf{R}_k^{-1}}, \tag{6b}$$

with

$$\tilde{\mathbf{x}}_k = \mathcal{M}_{0 \mapsto k \cdot \Delta t}(\tilde{\mathbf{x}}_0) = \underbrace{g_\theta \circ \cdots \circ g_\theta}_{k-\text{times}}(\tilde{\mathbf{x}}_0), \tag{7}$$

and where $\|\mathbf{x}\|_{\mathbf{A}} = \sqrt{\mathbf{x}^\top \mathbf{A} \mathbf{x}}$ is the Mahalanobis norm, and $\mathbf{B}$ is defined by $\mathbf{B} = \sigma_{\mathrm{b}}^2 \mathbf{I}$. Through all experiments, $\sigma_{\mathrm{b}}^2 = 0.4^2$ (non-dimensional, as $\tilde{\mathbf{x}}$ has been normalized). The coefficient λ_{inf} is a ~~background~~ multiplicative inflation term for background errors, which is set to 1 when no inflation is used and is further described in Appendix E. Note that the observation error covariance matrices $\mathbf{R}_k$ are not inflated because, with diagonal matrices, this is equivalent to background inflation, as far as the cost function minimization is concerned. Within the DAW of 16 days, with observations taken every second day starting on day 2, the total number of ~~observations~~ observation times is $K = 8$. The results of the minimization of the cost function is the analysis $\mathbf{x}_0^{\mathrm{a}}$.

4.3 4D–Var projected onto the EOFs basis: 4D–Var-EOF

Empirical orthogonal functions (EOFs) are a set of orthogonal state vectors derived from data, which form a basis of the full state space. Details about the computation and the analysis of the EOFs are given in Appendix B. A reduced strategy to enhance 4D–Var by projecting onto the EOFs has been proposed by Robert et al. (2005) for ocean models. The projection onto the EOFs enables access to cross-covariances and ~~improves~~ may improve the numerical conditioning of the ~~B matrix, thereby enhancing the minimization of the cost function~~ minimization of $\mathcal{J}$. The minimization is carried out with respect to the control variable $\mathbf{w} \in \mathbb{R}^m$ defined by the projection of the vector $\tilde{\mathbf{x}}$ onto the matrix $\boldsymbol{\varphi}_m \in \mathbb{R}^{N_z \times m}$:

$$\tilde{\mathbf{x}} = \bar{\mathbf{x}} + \boldsymbol{\varphi}_m \mathbf{w}, \tag{8}$$

with $\bar{\mathbf{x}}$ representing the temporal average sea-ice thickness field over the dataset used to construct the EOFs and m standing for the truncation index corresponding to the number of EOFs which are kept in the definition of the minimization subspace. The goal is to run the 4D–Var minimization in this affine truncated space spanned by the $\boldsymbol{\varphi}_m$.

In this subspace, the cost function reads, at cycle n,

$$\mathcal{J}(\mathbf{w}_0) = \frac{1}{2\lambda_{\mathrm{inf}}^2} \left\| \mathbf{w}_0 - \mathbf{w}_0^{\mathrm{b}} \right\|^2 + \frac{1}{2} \sum_{k=1}^{K} \left\| \tilde{\mathbf{y}}_k - \mathbf{H}_k \underline{\tilde{\mathbf{x}}_{k \times N_{\mathrm{f}}}} {}_k g_\theta^{k \times N_{\mathrm{f}}} (\bar{\mathbf{x}} + \boldsymbol{\varphi}_m \mathbf{w}_0, \mathbf{F}_{k \times N_{\mathrm{f}} \mapsto 0}) \right\|^2_{\mathbf{R}_k^{-1}}. \tag{9}$$

The details of the 4D–Var cost function computation are presented in Alg. C1 and Alg. C2. The result of this minimization is the control variable at the beginning of the DAW, $\mathbf{w}_0^{\mathrm{a}}$. The ~~choice~~ value of the truncation index is set to $m = 7000$ for all experiments and is further discussed in the Appendix B while further details on the optimization are given in Appendix C.

5 ~~Results~~

~~In this section, the 4D-Var results obtained first on simulated observations (Sec. 5), then on real observations (Sec. 6) are presented.~~

4.1 Metrics for the evaluation of the experiments

To evaluate the efficiency of the data assimilation scheme, we compute the root-mean-squared error (RMSE) between the predicted state performed by applying the emulator with the analysis as initial state, $\mathbf{x}^{\mathrm{f}}(t) = \mathcal{M}_{0 \mapsto t\Delta t}(\mathbf{x}_0^{\mathrm{a}})$, and the truth $\mathbf{x}_{\mathrm{t}}$, which corresponds to the neXtSIM simulation in the twin experiment case, and to CS2SMOS fields in the real observations case, for each cycle n, at the lead time t, and over all unmasked pixels $i \in N_z$. Inside the DAW (up to 16 days for the twin experiments case and 8 days for the real observations case), this corresponds to an analysis RMSE and afterwards a forecast RMSE.

$$\mathrm{RMSE}_n(t) = \sqrt{\frac{1}{N_z} \sum_{i=1}^{N_z} (\mathbf{x}_{i,n}^{\mathrm{f}}(t) - \mathbf{x}_{i,n}^{\mathrm{t}}(t))^2}. \tag{10}$$

This RMSE is defined for each cycle of each experiment, and can then be averaged across all cycles to get the mRMSE, which becomes a function depending on the lead time t only.

$$\mathrm{mRMSE}(t) = \frac{1}{N_{\mathrm{cycle}}} \sum_{n=1}^{N_{\mathrm{cycle}}} \mathrm{RMSE}_n(t). \tag{11}$$

For the evaluation of CS2SMOS assimilation, see Sec. 6, we use two additional metrics, the bias error,

$$\mathrm{bias}_n(t) = \frac{1}{N_z} \sum_{i=1}^{N_z} \mathbf{x}_{i,n}^{\mathrm{f}}(t) - \frac{1}{N_z} \sum_{i=1}^{N_z} \mathbf{x}_{i,n}^{\mathrm{t}}(t), \tag{12}$$

and the Ice Integrated Edge Error (~~IIEE~~IIEE$_{\mathrm{SIT}}$), as initially introduced by Goessling et al. (2016), but slightly modified by using sea-ice thickness instead of sea-ice concentration as the threshold. Specifically, we define a metric that counts the grid cells where the surrogate model disagrees with CS2SMOS on the presence of sea ice. A grid cell is considered to be covered by sea ice if the thickness exceeds 0.1 m ~~(Durand et al., 2024)~~(cf Durand et al., 2024, B1 for the threshold justification), analogous to the sea-ice concentration threshold of 0.15 defined in Goessling et al. (2016). The ~~IIEE~~ IIEE$_{\mathrm{SIT}}$ is expressed as ~~a fraction of grid-cells (not multiplied by the sea-ice covered area).~~

4.2 ~~Twin experiments~~

the total area (km^2) of disagreement between the observations and the forecast.

5 Twin experiments

In this section, the results from the twin experiments are ~~discussed, with the observations simulated as explained~~ presented in Sec. 5.2, after the description of the simulated observations in Sec. 5.1. ~~In~~

5.1 Simulated observations from neXtSIM

In the first approach, a twin experiment setup is employed, wherein synthetic observations are generated by adding noise to neXtSIM simulations from 2017 and 2018. Using the observation error variance $\sigma_{\mathrm{obs}}^2 = 0.4^2$ (non-dimensional), we define several variants of perturbed observations,

$$\tilde{\mathbf{x}}_{\mathrm{t,obs}}^{\mathrm{G}} = \tilde{\mathbf{x}}_{\mathrm{t}} + \boldsymbol{\epsilon}^{\mathrm{G}}, \qquad\qquad \boldsymbol{\epsilon}^{\mathrm{G}} \sim \mathcal{N}(0, \sigma_{\mathrm{obs}}^2), \tag{13a}$$

$$\tilde{\mathbf{x}}_{\mathrm{t,obs}}^{\mathrm{LN}} = \frac{\mathbf{x}_{\mathrm{t}} \exp\left(\boldsymbol{\epsilon}^{\mathrm{LN}}\right) - \mu_{\mathrm{SIT}}}{\sigma_{\mathrm{SIT}}}, \qquad\qquad \boldsymbol{\epsilon}^{\mathrm{LN}} \sim \mathcal{N}\left(0, \sigma_{\mathrm{obs}} - \frac{1}{2}\sigma_{\mathrm{obs}}^2,\right), \tag{13b}$$

$$\tilde{\mathbf{x}}_{\mathrm{t,obs}}^{\mathrm{cond-clipped}} = \min\left(\tilde{\mathbf{x}}_{\mathrm{t}}\left(1 + \boldsymbol{\epsilon}^{\mathrm{G}}\right), \mathrm{SIT}_{\mathrm{min}}\right). \tag{13c}$$

with exp the exponential function. The Gaussian observation noise as defined in Eq. (13a), is an idealized case tailored to the common assumptions of 4D–Var, to test an adaptive inflation scheme which will be defined later. Equation (13b) specifies a log-normal distribution for the noise, as more commonly encountered in sea-ice observations from satellites (Landy et al., 2020) . Furthermore, a variant to log-normal noise is introduced in Eq. (13c) by adding a fraction of the sea-ice thickness and incorporating clipping, based on $\mathrm{SIT}_{\mathrm{min}}$ the corresponding 0m thickness in the normalized space. This approach ensures that the observations remain confined within the physical bounds of sea ice, unlike Gaussian noise, but similarly to log-normal noise. Examples of the different noises are shown in Fig. 1b)–d).

These noise definitions yield different noise magnitudes. The log-normal noise, defined in Eq. (13b), provides a more significant spread, especially for thicker ice. In average, the log-normal noise definition results in a standard deviation of 0.35 m, because of the skewness of the log-normal law, whereas the conditioned noise, defined in Eq. (13c), results in a smaller standard deviation of 0.29 m.

5.2 Results

The length of the DAW is set to $N_{\mathrm{daw}} = 16$ days which corresponds to 32 iterations of the surrogate model. In the twin experiment setup, observations are acquired every 2 days (every $N_{\mathrm{f}} = 4$ iterations). In all twin experiments, we initialize the ~~data assimilation cycles with the same past field.~~ first data assimilation cycle with a past neXtSIM SIT field from the 1st of January 2016. In this section, the multiplicative inflation coefficient is set to 1. Experiments are conducted throughout years 2017 and 2018, which are after the training dataset (years $2009 - 2016$). To compute averaged results, only the year 2018 is used, i.e. 2017 is used as spin-up. The 4D–Var is run on a single trajectory starting the January 1st 2017, for ~~45~~ $N_{\mathrm{cycle}} = 45$ cycles, until December ~~22th~~21th, 2018.

As shown in Tab. 1, the choice of noise distribution—whether Gaussian, log-normal, or cond-clipped as defined in Sec. 5.1, influences the efficiency of the assimilation process. However, all experiments remained stable, meaning that there was no divergence in the results. While the assimilation of perfect observations yields the best results, the different noise definitions produce comparable outcomes. Using the cond-clipped perturbations provide the best RMSE among the three different ~~type~~ types of perturbations. This could be linked to the lower RMSE of its observations. As shown in Tab. 1, projecting the 4D–Var
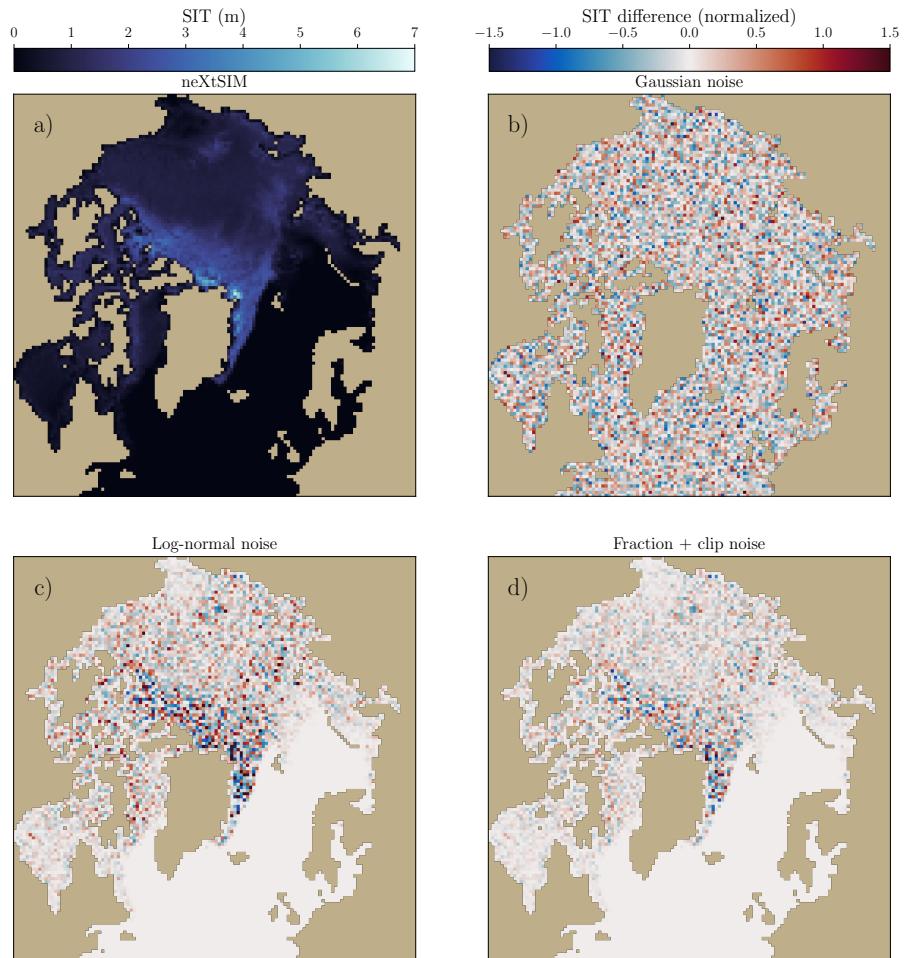
Figure 1. Snapshots of neXtSIM SIT (a) and different type of observations (b) Gaussian noise, (c) log-normal noise (LN) and (d) conditioned noise (cond-clipped). The colorbar for panels (b), (c) and (d) is shared and displayed on the right.

Table 1. Comparison of RMSE for different types of simulated observation noise (cf. Sec. 5.1) for the two types of 4D–Var algorithms (with diagonal **B** matrix and with projection onto the EOFs). Results are presented with the mRMSE computed inside the DAW across 2018. The RMSEs between neXtSIM and the perturbed observations, are outlined in the second column. They correspond to the averaged RMSE between the observations and their associated non-perturbed SIT field.

mRMSE (m)	Observations	4D–Var-diag	4D–Var-EOF
No noise	0.000	0.305	0.256
Gaussian noise	0.638	0.321	0.272
LN noise	0.587	0.333	0.281
Cond-clipped noise	0.527	0.318	0.264

onto the EOFs yields improvements in all cases, with relative improvements in the range $15\% - 17\%$ for the different types of noise. This improvement can be attributed to the ~~preconditioning of~~ non-diagonal terms in the **B** matrix~~and its non-diagonal~~

295 ~~terms~~. This systematic improvement is also observed in Fig. 2 during forecast. When extending the forecast beyond the DAW, the advantage of the 4D–Var-EOF over the 4D–Var-diag remains noticeable but diminishes as the lead time increases. On average, both forecasts show an improvement of 0.8 m over the emulator's free run (initialized on January 1st, 2017) across all lead times. Additionally, when comparing the 4D–Var-EOF forecast to the emulator forecast (initialized with perfect conditions at the end of each DAW, red dashed curve), we observe ~~a slight improvement of 1.1 cm, demonstrating a gain in forecast skill~~

300 ~~with our assimilation system~~comparable RMSE, showing the stability of the forecast produced by the analysis.

Figure 3 displays fields to illustrate the benefit of data assimilation over a cycle. The first guess field is smoother because of the emulator, whereas the analysis is actually noisier than the truth, because of the observations' noise. The largest corrections are applied to the MIZ to correct its position, especially in the Beaufort Sea, Chukchi Sea, and Hudson Bay. In the other places, we observe a positive correction, which is consistent with the negative bias ($\simeq -0.015$ m) of the surrogate model at the time

305 of the depicted cycle.

As seen in Fig. 4, at the start of each cycle, the initial analysis RMSE is generally lower than the first guess RMSE, which corresponds to the end of the previous DAW, indicating that the analysis improves over the first guess, of 2 cm in average across all cycles. In the specific case on late February 2018 (cycle 26), large differences between the truth and the analysis are located near the Canadian Archipelago, leading to a higher RMSE compared to the first-guess RMSE, where such differences do not

310 occur. We hypothesize that, in order to minimize the RMSE throughout the DAW, the 4D–Var removes sea ice in this region. This is consistent with the emulator's positive bias during this period. The analysis from Fig. 4(top) reveals a strong seasonality in results, with RMSE peaking in ~~May, dropping, then rising again in December. These peaks align~~ July. The peak aligns with significant changes in sea-ice extent ~~: decreasing in summer (May) and increasing rapidly by winter (December)~~at this period, with an important decrease due to the warmer temperatures in the Arctic. This suggests that 4D–Var possibly struggles more

315 with dynamic shifts. See additional multi-year results in Appendix. D. Additionally, the bias error of the emulator (Fig. 4,
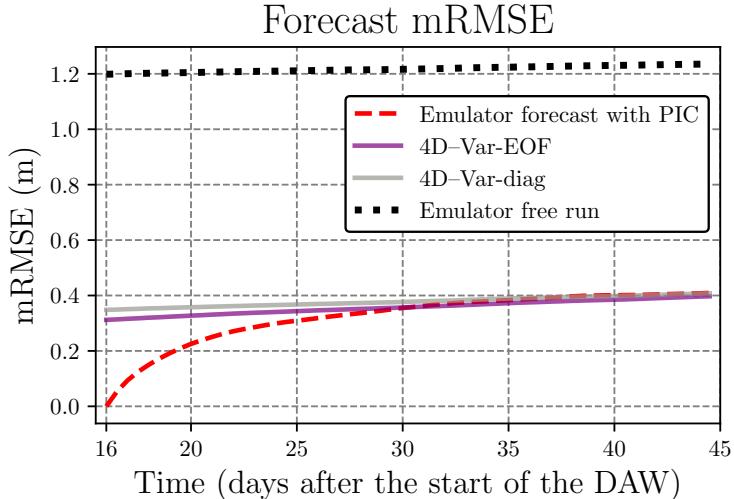
Figure 2. Cycle-averaged mRMSE of $\tilde{x}_f$ over 2018 during forecast stage. The dashed red line represents the free run of the emulator started at the end of each DAW (with perfect initial conditions, PIC). The mRMSE of the 4D–Var-EOF forecast is shown in purple, and the 4D–Var-diag forecast is shown using in grey. Cond-clipped noise is used in both cases. The black dotted line corresponds to the emulator free run, initialized on January 1st, 2017.

lower panel), indicative of model error, shows a similar seasonal pattern, transitioning from a positive to negative bias around summer and reverting in December. The key factor affecting assimilation accuracy appears to be not just the amplitude but also the seasonal variation of this bias.

5.3 ~~CS2SMOS assimilation~~

320 An important factor influencing the quality of the assimilation is the frequency of the observations. We denote the number of observations during one cycle of 16 d (32 model iterations) as N_{obs}. As shown in Fig 5, when there are too few observation times per cycle, no improvement is observed in the analysis compared to the forecast at the end of the previous cycle (this is the case for both 2 and 4 observation times per cycle). Once 8 observation times per cycle are reached, this divergence disappears, which is why we choose 8 observation times per cycle in our setup.

325 ## 6 CS2SMOS assimilation

In this section, the results from the real observations are presented in Sec. 6.2, after the description of the observations used in Sec. 6.1.

6.1 Real observations: combined Cryosat2-SMOS retrieval
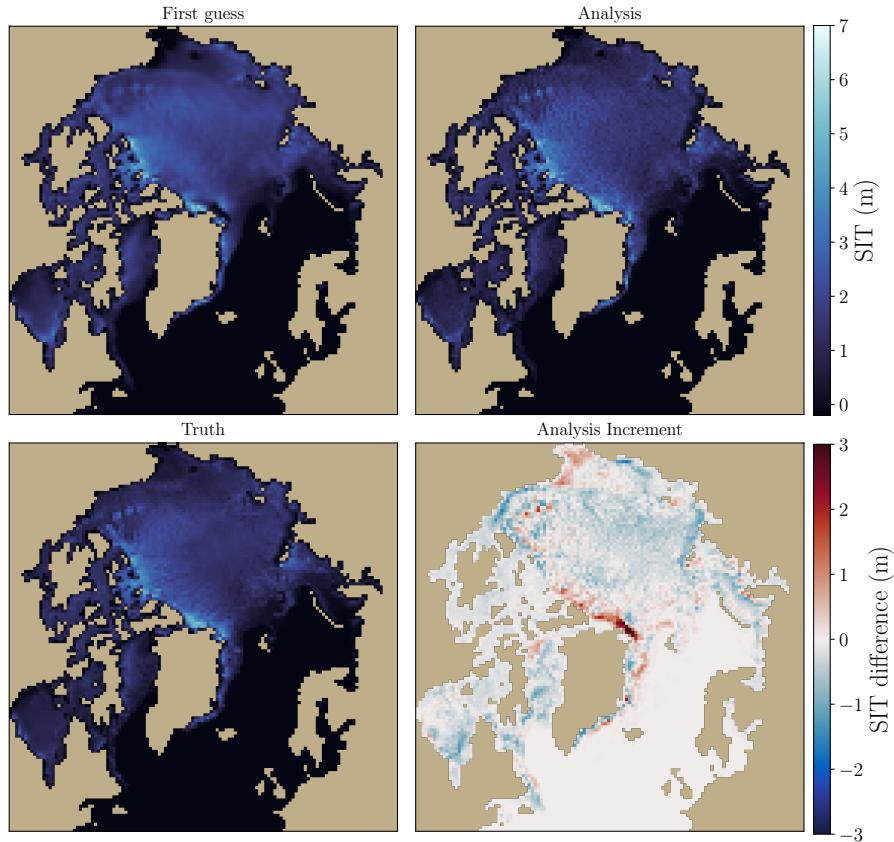
Figure 3. Fields of the SIT ~~in~~ at the ~~10th~~ beginning of the 25th cycle of the DA, corresponding to ~~2017-06-10~~2018-02-05, are shown in the 4D–Var-EOF case. The upper left panel represents the first guess, which is the output of the forecast from the previous minimization. The upper right panel corresponds to the analysis of the ~~10th~~ 25th cycle. For comparison, the associated neXtSIM field, considered as the truth, is displayed in the lower left panel. Note that these three fields share the same colormap and scale. The lower right panel shows the analysis increment, which represents the analysis minus first guess.
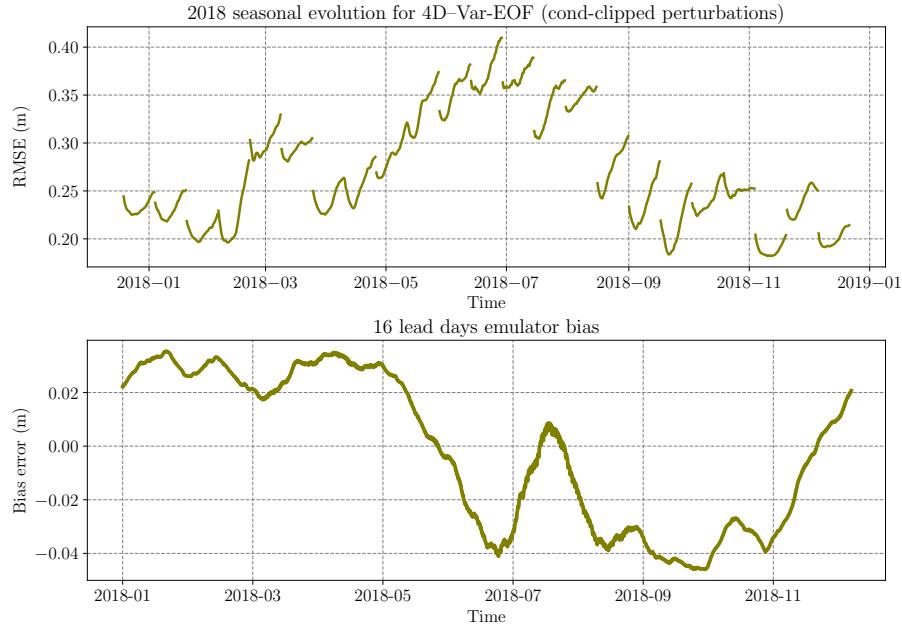
Figure 4. Time evolution of the 4D–Var-EOF analysis RMSE (inside the DAW) in 2018 (upper panel), with cond-clipped simulated noise. Corresponding bias error of the emulator is represented below, as defined in Eq. (A2) with a 16 days lead time, with a new forecast starting every 6 h, at the given time of the x-axis.
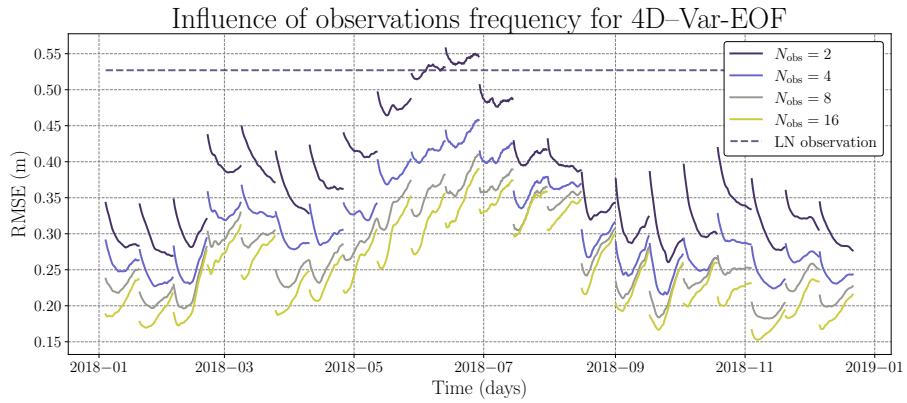
Figure 5. 2018 RMSE of 4D–Var-EOF depending on the number of observations in every assimilation cycle in the twin experiment case. Curves indicated the RMSE of the 4D–Var analysis with regard to neXtSIM SIT, with 2 observations per cycle (purple curve), 4 observations per cycle (blue curve), 8 observations per cycle (gray curve) and 16 observations per cycle (green curve). The dashed grey line correspond to the averaged RMSE between log-normal (LN) noise and the truth.
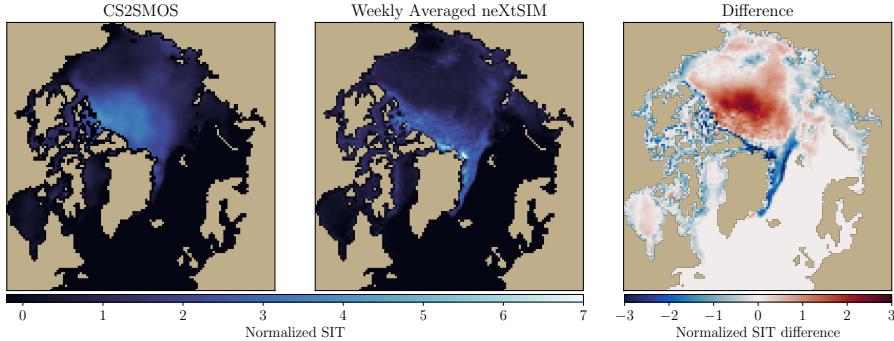
Figure 6. Difference (right) between daily CS2SMOS (left) and neXtSIM SIT (middle). CS2SMOS is interpolated on neXtSIM reduced grid. neXtSIM SIT is averaged over one week in order to mimic CS2SMOS weekly averaging, centered on 2018-01-04.

The dataset of CS2SMOS (Ricker et al., 2017) retrievals provides real observations. The retrievals merge observations from CryoSat-2 (Kurtz and Harbeck, 2017), known for its accurate observations of thick and perennial sea ice, and from SMOS (Tian-Kunze et al., 2014), used to infer the thickness of thin ice. Merged weekly to account for the different temporal resolution of CryoSat-2 and SMOS observations, the retrievals are available as daily moving window average. Note that the CS2SMOS is the result of Kriging and has been considerably smoothed in the process, even when compared to a weekly average of neXtSIM, as illustrated in Fig. 6.

CS2SMOS retrievals are only available on grid-cells covered by sea ice, and no information is available on grid-cells with open water. This creates a temporally changing mask, and we assume that grid-cells without information contain no sea ice.

Additionally, the CS2SMOS retrievals come with their own errors and uncertainties (Ricker et al., 2017). Note that Nab et al. (2025) showed that modifying SIT observation uncertainties introduces significant sensitivities during SIT assimilation. Based on the diagnostics of Desroziers et al. (2005), Xie et al. (2018) proposed an empirical formula for the observation error variance $\sigma^2_{\mathrm{obs,CS2SMOS}}$ as an increasing function of ice thickness h_{ice}, with the coefficient $0.2, 0.02, 0.1$ in m^2, 3 in m and 1.5 without unit,

$$
\sigma^2_{\mathrm{obs,CS2SMOS}} = \begin{cases} \min(0.2, 0.02 e^{1.8(h_{\mathrm{ice}}-3)}) & \text{if } h_{\mathrm{ice}} > 3\mathrm{m}, \\ \max(0.02, 0.1 e^{-1.5 h_{\mathrm{ice}}}) & \text{otherwise.} \end{cases} \tag{14}
$$

This observation error variance is also used in Cheng et al. (2023a). We will rely on this assessment to introduce observation error statistics for the real observation setup. Note that, unlike the usual approach in data assimilation where the model state is projected onto the observation space using $\mathcal{H}$, we simplify the process by doing the other way around. In a preprocessing step, real observations are interpolated onto the model space, making them retrievals. This is feasible because the observations are at a higher resolution in their native grid, and then coarse-grained by a factor 2. Even in their original resolution, they remain smoother than the forecasts of our surrogate model.

6.2 **Results**

350 In this section, instead of assimilating simulated observations, we assimilate CS2SMOS retrievals daily (every two iterations of the emulator) within an 8-day window for $N_{\text{cycle}} = 20$ cycles. We use the observations for the winter 2020-2021, since there are no CS2SMOS retrievals during summer. We consider the truth now as CS2SMOS. Note that ground truth is a quantity that is assimilated, and hence that the analysis score should not be over-interpreted. To initialize the cycling of the data assimilation, we start with ~~a field as simulated by neXtSIM for the 10th of October 2016.~~ the CS2SMOS retrieval from the date of the first

355 assimilation cycle. The observation error variance used is defined in Eq. (14). Only results performed with the 4D–Var-EOF are presented in this section. Note that no multiplicative inflation scheme is used.

In order to compare the results to a practical forecast benchmark, we use the past forecasts from neXtSIM-F (Williams et al., 2021), a forecasting system that consists of a stand-alone version of neXtSIM, forced by the TOPAZ ocean forecast (Sakov et al., 2012) and ECMWF atmospheric forecasts. The past forecasts have been obtained in 2023, corresponding to the

360 version of the forecasting system released in November 2023 (European Union-Copernicus Marine Service, 2020). ~~NeXtSIM-F assimilates~~ In neXtSIM-F, the model is nudged to CS2SMOS sea-ice thickness observations weekly~~with a simple nudging~~. It produces a 9-day forecast, which we ~~will use for~~ systematically use for numerical comparison with our data assimilation scheme, starting the neXtSIM-F forecast at the end of each DAW. Both data assimilation systems are compared directly to CS2SMOS observations for forecasts beyond the DAW.

365 The 4D–Var analysis is using atmospheric reanalysis $\mathbf{F}$ from ERA5 as input for the emulator. For the sake of fairness, for the 9-day forecast that ~~are~~ is run beyond the DAW, we use atmospheric forcings from the ECMWF atmospheric model HRES, which provides 10-day forecasts at a $16\,\text{km}$ resolution. By contrast, ERA5 is a reanalysis product that assimilates observations, making it unsuitable for an operational forecasting setup where the aim is to predict the future. Therefore, following the approach used in neXtSIM-F, we rely on atmospheric forecasts as forcings of the emulator during the forecast period. These

370 forecasts are interpolated onto the neXtSIM grid and normalized using the same processing method as the ERA5 forcings preparation, replacing them when applying the emulator during forecast.

As seen in Fig. 7, the assimilation of real data into our 4D–Var works and yields RMSEs similar to those of the neXtSIM-F assimilation forecast. The free run, initialized with a SIT field from October 2018 (due to the lack of simulation outputs in 2020), produces a stable trajectory but significantly deviates from the CS2SMOS fields. Assimilating real observation yields a

375 substantial decrease in RMSE, of $-0.49\,\text{m}$ during forecast. Results after the 9-day forecast are compared with the neXtSIM-F 9-day forecast and show similar outcomes. ~~In~~ On average, neXtSIM-F has a $0.34\,\text{m}$ RMSE while our forecast has a $0.36\,\text{m}$ RMSE. Initially, neXtSIM-F exhibits lower RMSEs; however, we see improved results in term of forecast towards the last cycles of the assimilation. The RMSE ~~score~~ metric penalizes the high level of details of the neXtSIM model more than the emulator that smooths gradually with time~~, an effect known as "double penalty" in weather forecasting~~.

380 As seen in the twin experiment results, especially with the Fig. 4, we can infer that our data assimilation system is less efficient during periods of strong dynamic change. This is also shown here, with better results after January, at the end of the refreezing period when comparing the end of each cycle forecast (end of the purple dashed lines) with corresponding neXtSIM-
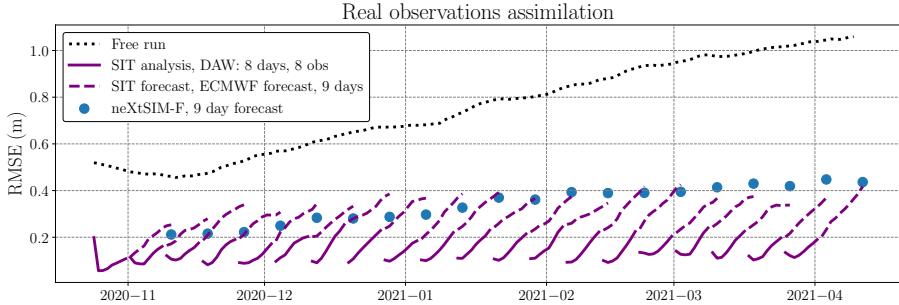
Figure 7. RMSE results for CS2SMOS assimilation across several DAWs throughout the full CS2SMOS observation period in 2020-2021 are shown. The black dotted line represents the free run of the emulator through all cycles, initialized with a SIT field from October 2018. The solid purple line corresponds to the analysis of the 4D–Var over the DAW, while the associated dashed purple lines represent the additional forecasts using ECMWF atmospheric forecasts for 9 days. The RMSE values from neXtSIM-F, corresponding to a 9-day forecast, are displayed as blue dots and should be compared with the end of each corresponding dashed line. All RMSE values are computed with CS2SMOS considered as the truth.

F. Conversely, ~~when the system is less dynamic~~after the refreezing period, it faces fewer difficulties in predicting the optimal state. Yet, the high initial RMSEs could also be linked to the spin-up of the assimilation. Similarly, to evaluate our assimilation
385 framework, we use the bias error and the ~~IIEE~~ IIEE$_{SIT}$ as defined in Sec. 4.1 in the same data assimilation cycles, see Fig. 8. The ~~bias error of the trajectory is significantly reduced~~absolute value of the bias of the emulator free run is generally reduced, with the exception of December 2020, where the free run bias is smaller than the forecast bias. At the end of the forecast, the assimilation run brings a bias reduction of $6.7\,\mathrm{cm}$ compared to the free run. The average bias of the assimilation run compared to CS2SMOS at the end of the forecast is $-0.22\,\mathrm{cm}$, while the bias of neXtSIM-F is $-1.5\,\mathrm{cm}$. ~~During the forecast, the bias~~
390 ~~systematically decreases, which might indicate that the emulator considers the analysis to increase~~ The biases at the beginning of the DAWs are systematically positive and then generally decrease. Since the emulator tends to reduce the amount of sea ice~~.~~ ~~neXtSIM~~, as indicated by its negative bias, the analysis overestimates sea ice to compensate for this loss. NeXtSIM, on which the emulator is trained, has a different SIT distribution than CS2SMOS, which could explain this phenomenon~~. The IIEE~~, as it can be seen with the different order of magnitude of the bias between the emulator initialized with neXtSIM fields in Fig. 4b)
395 (around $0.02\,\mathrm{m}$) and the emulator initialized with CS2SMOS (around $0.1\,\mathrm{m}$). The IIEE$_{SIT}$ serves as a reliable indicator of how accurately the MIZ is positioned. The IIEE of neXtSIM-F is 5.29% and is slightly better than that of the assimilation run (5.88%). At the end of each forecast, the assimilation run shows a 15% improvement in IIEE compared to the free run, highlighting a significant enhancement in MIZ positioning achieved through the 4D–Var-EOF assimilation.

The data assimilation analysis acts as a bias correction for the emulator. However, it relies on smooth observations, thereby
400 losing the small-scale information available in neXtSIM and neXtSIM-F, as illustrated in Fig. 9. In Fig. 3, we observe that the first guess, corresponding to the end of the previous cycle forecast, appears smoother than the truth. This smoothing is linked to the deterministic nature of the emulator (Durand et al., 2024), as it optimizes its MSE loss by smoothing fine-scale
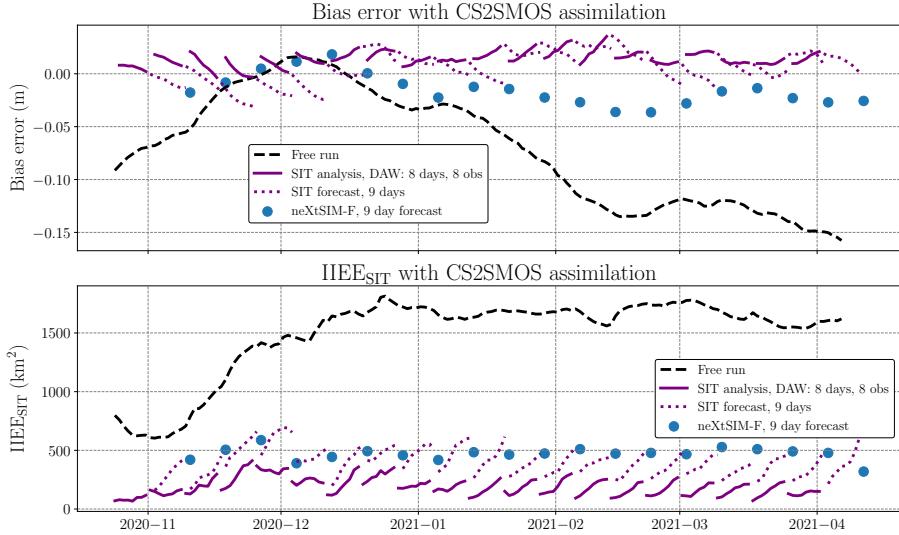
Figure 8. Bias (upper panel) and ~~IIEE~~ $IIEE_{SIT}$ (lower panel) results for CS2SMOS assimilation across several DAWs throughout the full CS2SMOS observation period in 2020-2021 are shown. The black ~~dotted~~ dashed line represents the free run of the emulator through all cycles, initialized with a SIT field from October 2018. The solid purple line corresponds to the analysis of the 4D–Var over the DAW, while the associated ~~dashed~~ dotted purple lines represent the additional forecasts using ECMWF atmospheric forecasts for 9 days. The bias errors and IIEE from neXtSIM-F, corresponding to a 9-day forecast, are displayed as blue dots and should be compared with the end of each corresponding dashed line. All bias errors and IIEE are computed with CS2SMOS considered as the truth.

dynamics. While in the twin experiments, the use of observations closely aligned with neXtSIM could help recover some small-scale dynamics, smooth observations result in a complete loss of these finer details. Nevertheless, our data assimilation
405 system effectively acts as a model error correction mechanism.

7 Discussion

We have demonstrated that using an emulator as forecast model in a 4D–Var framework for sea-ice forecasting is feasible, thanks to its numerical efficiency and auto-differentiability. However, this approach raises several important questions: one key observation is that while replacing the physics-based model with an emulator allows for the benefits of adjoint optimization
410 in a 4D–Var system, the success of the data assimilation is inherently tied to the emulator's accuracy. As illustrated in Fig. 3, the emulator tends to smooth out SIT during forecasts extending up to 16 days, a behavior previously noted in Durand et al. (2024).

In the case of real observations, we observe a significant bias correction in the assimilation run. Yet, it is worth noting that this study does not consider weak-constraint 4D–Var, which incorporates model error into the cost function minimization.
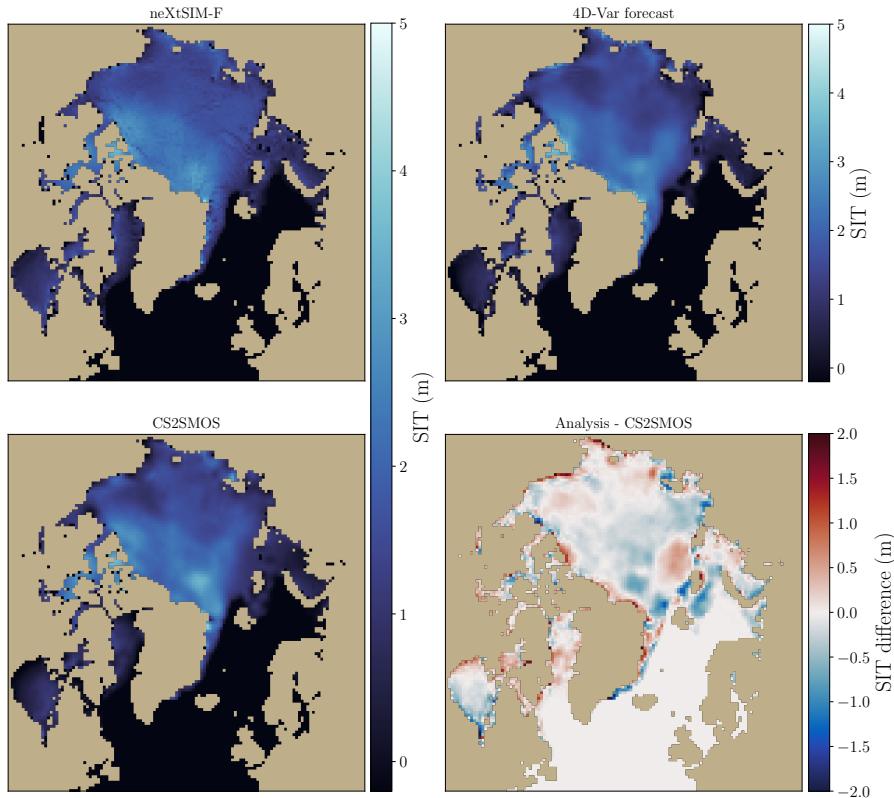
Figure 9. Visualization of the forecast for 2021-03-26 is shown. The upper left panel displays the neXtSIM-F 9-day forecast, while the upper right panel shows the 9-day forecast from the 4D–Var. The lower left panel presents the CS2SMOS observations, and the lower right panel illustrates the difference between the 4D–Var forecast and the CS2SMOS observations.

415   However, we ~~can infer~~ speculate that improving the emulator's quality—addressing both bias and smoothing issues—would result in more accurate 4D–Var analyses.

     An analysis comparing the numerical efficiency of the 4D–Var method between 4D–Var-diag and 4D–Var-EOF is presented. For twin experiments, all computations were executed on a single NVIDIA A100 SXM4 80 GB GPU. On average, over a full data assimilation run of 45 cycles, 4D–Var-EOF takes $155\,\mathrm{s}$ per cycle, while 4D–Var-diag takes approximately $229\,\mathrm{s}$ per

420 cycle, making the 4D–Var-EOF around 32% faster. Although the gradient computation times differ slightly between the two methods — $1.66\,\mathrm{s}$ for 4D–Var-EOF versus $1.36\,\mathrm{s}$ for 4D–Var-diag — the forward pass through the DAW is similar, with times of $578\,\mathrm{ms}$ and $552\,\mathrm{ms}$, respectively. The computation of the gradient involves the emulator adjoint evaluation and reshaping the one-dimensional vector of assimilated grid cells into the two-dimensional field required to run the emulator, as well as computing the cost function terms gradients. Note that the emulator adjoint is evaluated 8 times in each DAW. Interestingly,

while individual operations are faster in 4D–Var-diag (as expected due to the absence of projections to and from the EOF basis), the complete cycles are faster with the EOF-based method. In other words, fewer iterations of the L-BFGS optimizer are required on average for 4D–Var-EOF to achieve the analysis at each cycle, indicating better conditioning of the minimizations.

A significant reduction in RMSE (16%) was indicated by our results, achieved through the projection of the minimization onto the EOFs basis. Although a substantial number of EOFs is retained in the ensemble to preserve information, it is conceivable that computational time could be further reduced by decreasing the number of EOFs. However, it is important to note that the forward pass of the emulator and the computation of its adjoint are inherently performed in the (128×128 grid-cells) physical space. Consequently, the computational time savings from reducing the number of EOFs may not be substantial. On the other side, it might lead to a loss of the small scale information, which can also be observed in Fig. B2, where the RMSE increases as the truncation index decreases, down to a certain threshold.

We introduce in Appendix E an inflation scheme to tune the background cost function term. Two versions are evaluated: a more often used constant model inflation and an adaptive one, based on the χ^2_p (with p degrees of freedom) diagnostic (Michel, 2014). The adaptive background error inflation can be easily implemented in twin experiment scenarios, under a Gaussian noise simulation. We observe a modest improvement in the time-averaged RMSE, with most of the gains occurring at the beginning of each assimilation window. Interestingly, in both cases, we have to deflate the B-matrix for an optimal analysis RMSE (around 0.5 to 0.6 with a seasonal dependency).

A major factor for the quality in terms of RMSE is the frequency of the observations. While most of the results above focus on a fixed number of observations per cycle, additional experiments using varying frequencies in a twin experiment setup were presented in Fig. 5, showing important improvement in RMSE when increasing the frequency of the observations. However, current satellite data either provide spatially and temporally sparse observations (like Cryosat-2) or smooth, time-averaged full coverage products, which introduces inherent correlations between the observations errors (like CS2SMOS). It is important to note that our comparison with neXtSIM-F occurs 9 days after the last batch of assimilated observations, thus falling outside the average observation window. Yet, in the case of real observations, there is currently no SIT retrieval dataset that provides daily, non-smoothed, and non-time-correlated SIT measurements to validate our experiments against. Ongoing works are proposing ML-based approaches to derive complete daily sea ice freeboard fields from satellite altimetry at fine spatial resolution (5 km), by modeling the spatio-temporal covariance of daily fields rather than relying solely on temporal averaging (Gregory et al., 2024b; Chen et al., 2024).

Let us note that neXtSIM, as well as neXtSIM-F exhibits more small-scale features than our emulator that tends to smooth out the fields. By assimilating CS2SMOS observations, we lose small scale information. Yet, as the observations are extremely smooth, the comparison is not as fair to neXtSIM-F, which provides more small scale dynamics, and hence suffers from double penalty effects. Implementing a stochastic emulator (Finn et al., 2024a, b) could yield more physically consistent results by conserving spectral energy. This approach may enhance the performance of the model within the 4D–Var minimization framework, although it raises questions about the reliability of the associated gradients and the increase in computation time. Newer satellites will hopefully provide data at higher resolution and better accuracy, but we expect that the current trade-off between resolution (by altimeters) and coverage (by passive microwaves) will remain an issue

in the foreseeable future. Another option is to apply super-resolution algorithms to enhance local scale dynamics in the data assimilation system (Barthélémy et al., 2022). Exploring these possibilities could significantly improve the efficacy of data assimilation with emulators for sea ice modeling.

One important aspect to monitor is the evolution of the cost function during minimization, as well as its associated gradient. More details are provided in the Appendix C. As seen in Fig. C2, there is more than an order of magnitude difference in the gradient norm during minimization within a single cycle. It is worth noting that the only stopping criterion consistently achieved is related to the tolerance of the cost function, where its decrease becomes smaller than a given value. Occasionally, a second criterion related to the gradient norm may also be considered, which requires that the maximal value of the gradient for the entire field be below another specified threshold. However, this gradient criterion is never met in our case. As shown in Fig. C2, there are still some grid-cells, often located in the MIZ, where the gradient norm remains non-negligible. Yet, as seen in Fig. C1, at each cycle, the cost function attains a stable minimum which indicate that the L-BFGS optimization worked correctly.

An important point to discuss is the realism of the emulator's adjoint. In deep learning, the high number of degrees of freedom often results in ~~poor-quality gradients~~ noisier gradients (Sitzmann et al., 2020). In the 4D–Var framework, this issue is mitigated by the background term, which regularizes the emulator's potentially noisy gradient. This noise is evident in the analysis, as shown in Fig. 3 and Fig. C2, particularly near the MIZ. While a dedicated training procedure for the emulator could potentially improve this aspect, it does not appear to hinder the current 4D–Var setup. In fact, the successful results achieved with this system indirectly validate the adjoint of both the emulator and the cost function. Furthermore, correctness checks for both adjoints are provided in Appendix C4.

Since the emulator provides fast forecasts of the SIT dynamics, and due to the fast data assimilation, we can investigate the possibility to run ensemble data assimilation (EDA) (Raynaud et al., 2008; Isaksen et al., 2010) by running an ensemble of 4D–Vars, with perturbations of the observations and the background term. This ensemble can be used to build an ensemble of trajectories, but also to improve the flow-dependency of the background covariances in the EOF space. Interestingly, this approach also allows for the comparison with other data assimilation methods using this emulator, such as the ensemble Kalman filter. This would enable a more straightforward comparison with current state-of-the-art data assimilation methods for sea ice. However, it would be important to further discuss the capability of our deterministic surrogate model to generate a state ensemble in this context.

These results are promising and demonstrate the potential for using model emulators in data assimilation, particularly with classical methods in real-world applications. ~~However, the main challenge observed is the smoothness of both the emulator and the observations.~~ Furthermore, it could be interesting to see the impact of assimilating several variables, like SIT and sea-ice concentration (SIC) onto an emulator.

8   Conclusions

In this paper, we introduced the first 4D–Var system based on a surrogate model that is trained to fully emulate the evolution of the sea-ice thickness. This work is a preliminary step towards the use of fully emulated models in data assimilation. Through twin experiments, we initially demonstrate the ability of the surrogate model to leverage its automatic gradient in a 4D–Var minimization. The 4D–Var system can be efficiently implemented by using EOFs extracted from the model's climatology. The assimilation in EOF space improves the system compared to a diagonal background covariance. Inflation techniques bring small improvement. In the second part of the study, we investigate the assimilation of real observations with the 4D–Var system. Assimilating real observations improves the positioning of the MIZ. These observations act as a bias correction, highlighting the potential need for weak-constraint 4D–Var to address such biases. This could also create opportunities to train a bias correction model or refine the emulator using analysis increments. With limited resources, such as emulating only sea-ice thickness and assimilating CS2SMOS observations, the developed 4D–Var system performs comparably to the ~~ensemble Kalman filter-based~~ operational neXtSIM-F system. Moreover, the emulator-based 4D–Var system is significantly more computationally efficient than ensemble Kalman filter systems relying on geophysical models. Although these results are derived from a coarse-grained emulator of the available neXtSIM dataset, no major issues are anticipated in increasing the resolution of both the surrogate model and the observations. The computational cost of this data assimilation approach remains well within the standards of current sea-ice data assimilation systems.

Appendix A: Surrogate modeling

In this section, we present the forecast ability of the emulator. The root-mean-squared error (RMSE) between the prediction $\mathbf{x}^{\mathrm{f}}(n+k\Delta t)$ and the simulation $\mathbf{x}^{\mathrm{t}}(n+k\Delta t)$ is computed over all pixels (i, j) of the field of size (N_x, N_y), for each sample n of the validation set containing $N_s = 1470$ trajectories, initialized at time t_n,

$$\mathrm{RMSE}(k) = \frac{1}{N_s} \sum_{n=1}^{N_s} \sqrt{\frac{1}{N_x \cdot N_y} \sum_{i,j}^{N_x,N_y} \left(\mathbf{x}^{\mathrm{f}}_{t_n+k\Delta t} - \mathbf{x}^{\mathrm{t}}_{t_n+k\Delta t}\right)^2} \sqrt{\frac{1}{N_x \cdot N_y} \sum_{i,j}^{N_x,N_y} \left(\mathrm{x}^{\mathrm{f}}_{i,j}(t_n+k\Delta t) - \mathrm{x}^{\mathrm{t}}_{i,j}(t_n+k\Delta t)\right)^2}. \quad \text{(A1)}$$

In order to quantify systematic errors of the surrogate model, we compute its mean error (bias). This metric tells about the ability of the neural network to correctly estimate the total amount of sea ice in the full domain,

$$\mathrm{bias}(k) = \frac{1}{N_s} \sum_{n=1}^{N_s} \frac{1}{N_x \cdot N_y} \sum_{i,j}^{N_x,N_y} \left(\mathrm{x}_{i,j}{}^{\mathrm{f}}_{t_n+k\Delta t}(t_n+k\Delta t) - \mathrm{x}_{i,j}{}^{\mathrm{t}}_{t_n+k\Delta t}(t_n+k\Delta t) \right). \quad \text{(A2)}$$

The code structure of the surrogate model g_θ is presented in Alg. A1. Let us note that f_θ maps $\tilde{\mathbf{x}}_t$ to $\tilde{\mathbf{y}}_{t+\Delta t} = \tilde{\mathbf{x}}_{t+\Delta t} - \tilde{\mathbf{x}}_t$ which corresponds to the normalized difference in sea-ice thickness over 12 hours. The normalization is defined as in Eq. (1) with μ_{out} and σ_{out} the associated global mean and standard deviation of $\mathbf{y}_{t+\Delta t} = \mathbf{x}_{t+\Delta t} - \mathbf{x}_t$, computed over the training dataset.

Algorithm A1 Full-state surrogate model g_θ mapping from $\tilde{\mathbf{x}}_t$ to $\tilde{\mathbf{x}}_{t+\Delta t}$, using the previously trained f_θ. This algorithm describe exactly how the state $\tilde{\mathbf{x}}_{t+\Delta t}$ is obtained by the application of the fine-tuned f_θ neural network as defined in Eq (2b).

Require: $f_\theta(\tilde{\mathbf{x}}_t, \mathbf{F}, \theta), \tilde{\mathbf{x}}_t, \mathbf{F}, \sigma,$ and normalization values ($\mu_{\mathrm{SIT}}, \sigma_{\mathrm{SIT}}, \mu_{\mathrm{out}}, \sigma_{\mathrm{out}}$)

$\tilde{\mathbf{y}}_{t+\Delta t} \leftarrow f_\theta(\tilde{\mathbf{x}}_t, \mathbf{F}, \theta)$

$\mathbf{y}_{t+\Delta t} \leftarrow \sigma_{\mathrm{out}} \tilde{\mathbf{y}}_{t+\Delta t} + \mu_{\mathrm{out}}$

$\mathbf{x}_t \leftarrow \sigma_{\mathrm{SIT}} \tilde{\mathbf{x}}_t + \mu_{\mathrm{SIT}}$

$\mathbf{x}_{t+\Delta t} \leftarrow \mathbf{x}_t + \mathbf{y}_{t+\Delta t}$

$\tilde{\mathbf{x}}_{t+\Delta t} \leftarrow \frac{\mathbf{x}_{t+\Delta t} - \mu_{\mathrm{SIT}}}{\sigma_{\mathrm{SIT}}}$

$\tilde{\mathbf{x}}_{t+\Delta t} \leftarrow \sigma(\tilde{\mathbf{x}}_t) = g_\theta(\tilde{\mathbf{x}}_t)$ $\tilde{\mathbf{x}}_{t+\Delta t} \leftarrow \mathrm{ReLU}(\tilde{\mathbf{x}}_{t+\Delta t}) = g_\theta(\tilde{\mathbf{x}}_t)$

The training of the emulators are shown in Fig. A1, with the display of the training losses and the validation losses. The training loss measures how well the emulator fits the training data, while the validation loss assesses its performance on unseen data to detect overfitting and ensure generalization. The baseline consists of the UNet trained by learning f_θ. By transfer learning, f_θ weights are fine-tuned in order to learn g_θ, with a constrain ($\lambda = 10$) inside the loss function, see Eq. (5). Results in term of forecast ability of those emulators are presented in Fig. A2. The forecast skill of g_θ is compared to the one of f_θ, and the persistence, which consist to take the initial condition as the constant state of the system. We can see that in terms of RMSE, g_θ is slightly worse than the baseline, but in terms of bias, the fine-tuned constrained emulator display a smaller bias error compared to f_θ.
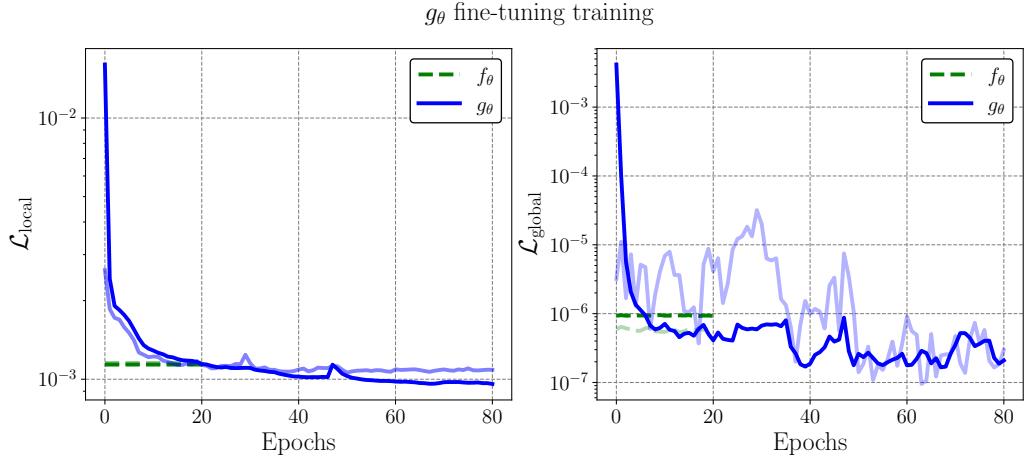
Figure A1. Left: Training and validation losses of the surrogate model. Right: Training and validation global losses of the surrogate models. ~~Lines~~ Light blue and light green lines in transparency indicate the validation losses. The green dashed line indicates the experiment where the weights of f_θ are frozen and a linear activation function is applied after the renormalization process. Blue line corresponds to the training of g_θ. The weights of f_θ are retrained with the new learning objective, with $\lambda = 10$.
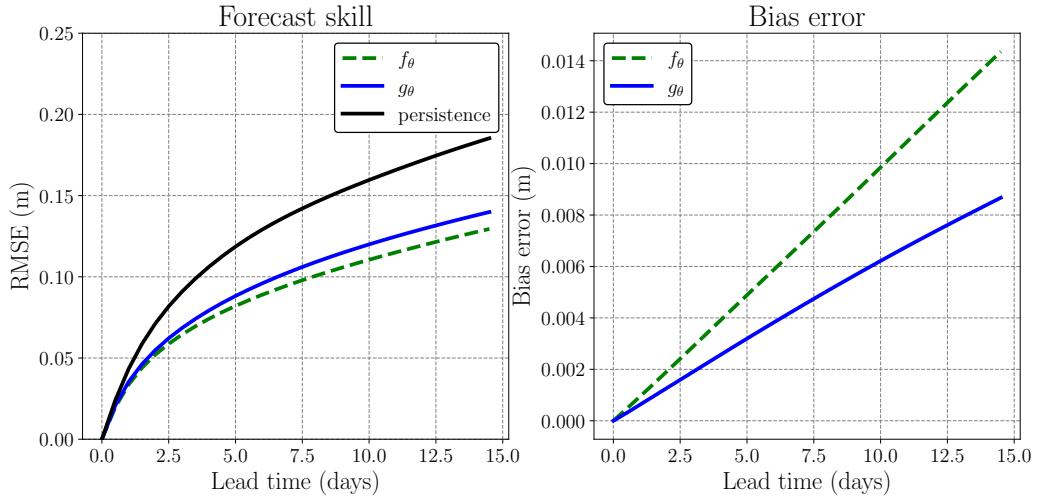
Figure A2. Left: Forecast skill (RMSE) of the surrogate model. Right: Bias errors of the surrogate models. The green dashed line indicates the results for f_θ. Blue line corresponds to the training of g_θ. The weights of f_θ are retrained with the new learning objective, with $\lambda = 10$.

Appendix B: Empirical Orthogonal Functions

B1 EOFs definition

We build an ensemble of perturbations using neXtSIM simulation outputs, $\mathbf{X}$ of size $\mathbb{R}^{N_t \times N_z}$, with $N_z = 8871$ the number of unmasked pixels and N_t the number of state in the ensemble, this number depends on the number of years taken to compute the ensemble and varies from 1500 to 11000. After the removal of the temporal mean from this ensemble:

$$\tilde{\mathbf{X}} = \mathbf{X} - \bar{\mathbf{x}}, \tag{B1}$$

we can compute the singular value decomposition of $\tilde{\mathbf{X}}$,

$$\tilde{\mathbf{X}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top, \tag{B2}$$

with $\mathbf{U}$ an orthogonal matrix of size $(N_t \times N_t)$, $\boldsymbol{\Sigma}$ a diagonal matrix of size $(N_t \times N_z)$ containing the N_t singular values of $\tilde{\mathbf{X}}$ and $\mathbf{V}$ an orthogonal matrix of size $(N_z \times N_z)$. We then define the EOFs $\boldsymbol{\varphi}$ as the columns of $\mathbf{V}$.

One advantage of using EOFs is the ability to reduce the dimensionality of the minimization space by projecting the state vector onto a truncated set of EOFs. We denote this truncation by $\boldsymbol{\varphi}_m$, where m represents the truncation index. In practice, this involves limiting the projection of the orthonormal matrix to the first m EOFs, ordered by explained variance, thereby reducing the computational burden while retaining the most significant modes of variability. The four predominant EOFs are displayed in Fig. B1, as well as their associated variance.

B2 Choice of the truncation index m

In order to validate the best truncation index, we run an experiment with 4D–Var-EOF, in the twin experiment setup, with a value of m ranging from 10 to 8871. We then compute the total RMSE over all cycles. Results are presented in Fig. B2. We observe that for $m > 5000$, the RMSE has reached a minimum. Let us note that using a smaller value of m reduces the minimization time, as it is reducing the dimension of the minimization space. Based on these results, and while we wanted to maintain a good reconstruction capacity, we chose to maintain $m = 7000$ for all experiments.

Appendix C: 4D–Var optimization

C1 4D–Var algorithms

We present here the algorithms for the computation of the 4D–Var optimization, in the 4D–Var-EOF case. The computation of the cost function, for a given cycle, is shown in Alg. C1. The state $\mathbf{w}_0$ is mapped back to the physical space, forecasted throughout the DAW using the emulator, where the observation term of the cost function is computed and then transformed back into the affine space of the EOFsto compute the observation term of the cost function. The total computation across all DAW is presented in Alg. C2.
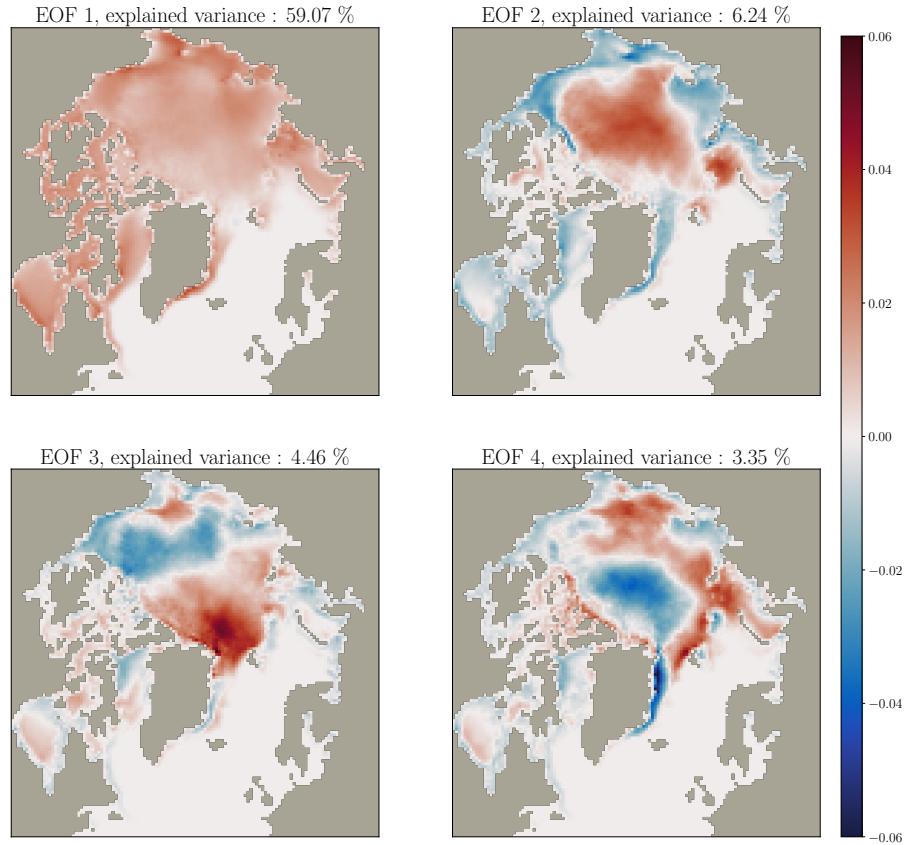
Figure B1. Four predominant EOFs of SIT, at the top left of each EOF is indicated the associated explained variance.
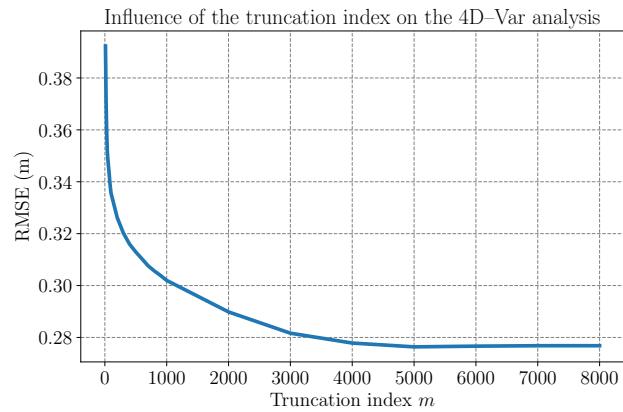
Figure B2. Average of the RMSE between $\mathbf{x}_a$ and $\mathbf{x}_t$ across all cycles and all timesteps for different values of k with Gaussian noise for the observations.

Algorithm C1 Cost function ($\mathcal{J}$) computation for the 4D–Var-EOF

Require: $g_\theta, \mathbf{w}_0, \tilde{\mathbf{y}}_{k:1,\ldots,K},$ the number of application of the emulator between two observations $N_f, K, \sigma_{obs}, \lambda_{inf}, \mathbf{H}, \boldsymbol{\varphi}_m, \mathbf{F}$

$\mathcal{J}_b = \frac{1}{2\lambda_{inf}^2}(\mathbf{w}_0 - \mathbf{w}_0^b)^\top (\mathbf{w}_0 - \mathbf{w}_0^b)$

$\mathcal{J}_o = 0$

$\tilde{\mathbf{x}}_0 = \bar{\mathbf{x}} + \boldsymbol{\varphi}_m \mathbf{w}_0$

for i in range $(1,\ldots,K)$ **do**

~~$\tilde{\mathbf{x}}_i = g_\theta^{N_f}(\tilde{\mathbf{x}}_{i-1}, \mathbf{F}_{i-1\to i}, \theta)$~~ $\tilde{\mathbf{x}}_i = g_\theta^{N_f}(\tilde{\mathbf{x}}_{i-1}, \mathbf{F}_{i-1\to i})$

$\mathcal{J}_o = \mathcal{J}_o + \frac{1}{2\sigma_{obs}^2}\|\tilde{\mathbf{y}}_i - \mathbf{H}\tilde{\mathbf{x}}_i\|^2$

end for

$\mathcal{J} = \mathcal{J}_o + \mathcal{J}_b$

Algorithm C2 Wrapper for cycling the 4D–Var-EOF~~minimization~~, using the loss ($\mathcal{J}$) defined in Alg. C1.

Require: $g_\theta(\tilde{\mathbf{x}}_t, \mathbf{F}), \tilde{\mathbf{y}}_{k:1,\ldots,K}, N_f, N_{cycle}, K, \mathbf{w}_0, DOF, \lambda_{inf}, \boldsymbol{\varphi}_m$

for n in range $(1,\ldots,N_{cycle})$ **do**

$\mathbf{w}_{0,n}^a = L-BFGS(\mathbf{w}_{0,n}, loss(\mathbf{w}_{0,n}, \tilde{\mathbf{y}}_{k:1,\ldots,K}))$

$\tilde{\mathbf{x}}_{0,n}^a = \bar{\mathbf{x}} + \boldsymbol{\varphi}_m \mathbf{w}_{0,n}^a$

$\tilde{\mathbf{x}}_{N_f(n\to n+1)}^a = g_\theta(\tilde{\mathbf{x}}_{0,n}^a, \mathbf{F}_{n\to n+1})$

$\tilde{\mathbf{x}}_n^b, \tilde{\mathbf{x}}_n^0 = \tilde{\mathbf{x}}_{N_f(n+1)}^a$

$\mathbf{w}_{0,n+1} = \boldsymbol{\varphi}_m^\top(\tilde{\mathbf{x}}_n^0 - \bar{\mathbf{x}})$

$n \leftarrow n+1$

end for

The L-BFGS-B (Broyden, 1967; Liu and Nocedal, 1989) algorithm is used to minimize the cost function $\mathcal{J}$, defined in
Alg. C1. The optimization is constrained by bounds defined over all the variables of $\tilde{\mathbf{x}}$ with a minimal value set to ~~min~~$_{SIT}$
SIT_{min} as defined previously, in the case of the 4D–Var-diag. Two criteria are used to stop the minimization: f_{tol}, which
corresponds to a threshold below which the cost function improvement is considered sufficient, and a gradient norm threshold
g_{tol}, below which the norm of the gradient must fall. Note that we did not define a maximal number of iterations of the
L-BFGS-B and the criteria f_{tol} was systematically reached.

In practice, in our case the only stopping criterion used is f_{tol}, as the gradient is significantly decreasing, yet, some instabilities at each iteration, especially on the MIZ are still observed.

C2 Cost function ~~analysis~~diagnostics

The value of the cost function $\mathcal{J}$ as defined in Eq. (6b) or Eq. (9) and minimized with a L-BFGS-B can be followed across
all cycles, as shown in Fig. C1. ~~During the first cycles, the seasonality~~ The RMSE increase observed in Fig. 4 can also be
observed in the cost function. The increase in May comes after 9-10 cycles, which corresponds to 2 cycles before the time
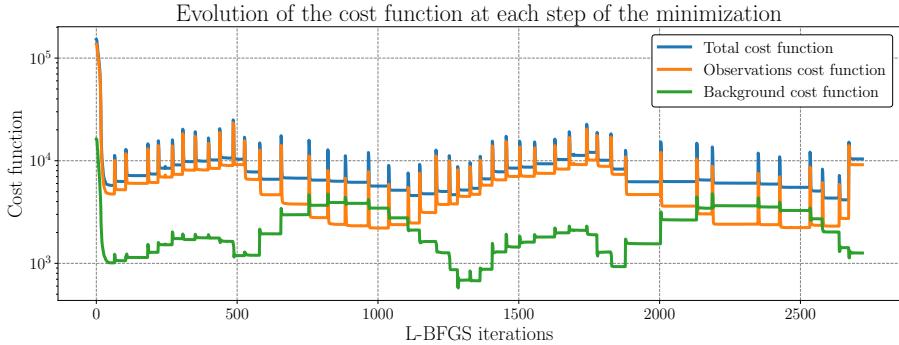
Figure C1. Cost function minimization with L-BFGS optimizer across all cycles for 4D–Var-EOF with log-normal noise, the total cost function is shown in blue, this term can be decomposed with the background loss term (green term) and the observation loss term (orange curve). Note that the y-axis is in log scale.

where the observation cost function decreases and the background cost function increases and becomes predominant. Based on this observed seasonality, we introduce an adaptive background strategy to offer an estimation of the order of magnitude of the background cost function. This implementation is further described in ~~See~~Appendix. E.

C3 Gradient analysis

We can also investigate the gradient of the cost function, and its evolution between the beginning and the end of the ~~DAW~~cycle, as shown in Fig. C2. We can observe a global decrease of the gradient across the full Arctic. Yet, some grid-cells keep a strong gradient, especially on the MIZ.

C4 Tests of the ~~adjoint of the~~ emulator and the cost function gradients

First of all, we test the gradient of the cost function. With Taylor's expansion

$$\mathcal{J}(x + \epsilon h) = \mathcal{J}(x) + \epsilon \mathcal{J}'(x) \cdot h + \mathcal{O}(\|\epsilon\|^2), \tag{C1}$$

we look at the ratio

$$\mathcal{I}(\epsilon) = \left\langle \left| \frac{\mathcal{J}(x + \epsilon h) - \mathcal{J}(x - \epsilon h)}{2\epsilon \mathcal{J}'(x) \cdot h} \right| \right\rangle_{\|h\|=1}, \tag{C2}$$

with **x** a state on the trajectory of the emulator. Note that we have

$$\mathcal{J}(x + \epsilon h) = \mathcal{J}(x) + \epsilon \mathcal{J}'(x) \cdot h + \mathcal{O}(\|\epsilon\|^2) \tag{C3a}$$

$$\mathcal{J}(x - \epsilon h) = \mathcal{J}(x) - \epsilon \mathcal{J}'(x) \cdot h + \mathcal{O}(\|\epsilon\|^2). \tag{C3b}$$

By making the difference of those two equations we obtain

$$\mathcal{J}(x + \epsilon h) - \mathcal{J}(x - \epsilon h) = 2\epsilon \mathcal{J}'(x) \cdot h + \mathcal{O}(\|\epsilon\|^2). \tag{C4}$$

Gradient evolution during one cycle
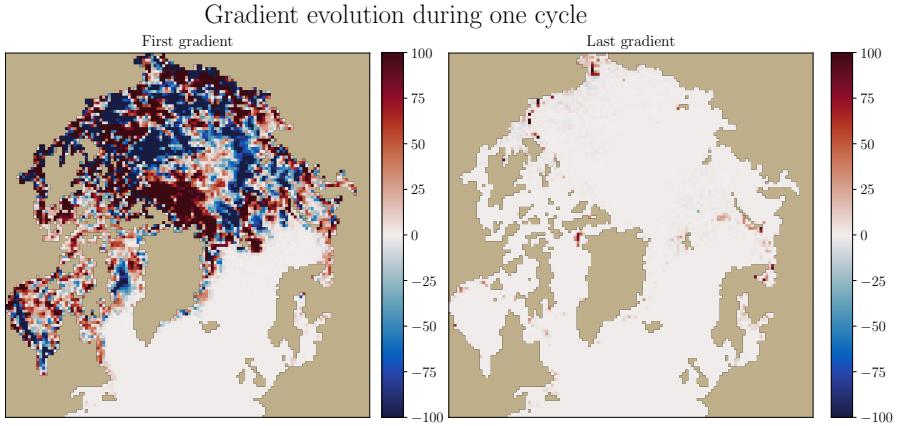
Figure C2. Evolution of the gradient ~~under minimization~~ during the ~~4D–Var~~ 4D–Var-diag minimization. ~~In~~ The left panel ~~is displayed~~ shows the gradient at the end of the first minimization ~~and on~~ of one cycle (cycle 31), while the right panel ~~is outlined~~ displays the gradient at the end of the last minimization of the same cycle, when the minimization criterion is reached. In this case, the synthetic observations are perfect observations (no noise addition).

By averaging, we expect

$$\mathcal{I}(\epsilon) \simeq 1 + \mathcal{O}(\|\epsilon\|). \tag{C5}$$

In practice, we take for the trajectory the full 2 years free run without assimilation, with the same parameters as defined in Sec. 5, with observations perturbed with cond-clipped noise. We evaluate several values $\mathbf{h}$, including canonic vectors to focus only on single pixel, especially in the MIZ, and in Central Arctic. We define the canonic vectors $\mathbf{h}_{\text{zone}} \in \mathbb{R}^{8871}$ as

$$\mathbf{h}_{\text{zone}} = (0, \cdots, 0, 1, 0, \cdots, 0), \tag{C6}$$

with the 1 at the position corresponding to the chosen area. We select 5 pixels for the MIZ and 5 pixels for Central Arctic. The results are shown in Fig. C3. While the residual errors remain small, they increase for $\epsilon < 10^{-7}$. Above the ϵ value, we observe the expected behavior, but the residual errors are slightly noisy. Interestingly, we obtain similar behavior for the different values of $\mathbf{h}$.

To test the ~~adjoint~~ gradient of the emulator, we evaluate the test function

$$\mathcal{L}(\mathbf{z}, \mathbf{x}) = \mathbf{z}^\top \mathbf{M} g_\theta(\mathbf{x}, \mathbf{F}), \tag{C7}$$

whose Taylor's expansion in $\mathbf{x}$ is

$$\mathcal{L}(\mathbf{z}, \mathbf{x} + \epsilon \mathbf{h}) = \mathbf{z}^\top \mathbf{M} g_\theta(\mathbf{x} + \epsilon \mathbf{h}, \mathbf{F}) \tag{C8a}$$

$$= \mathcal{L}(\mathbf{z}, \mathbf{x}) + \epsilon \mathbf{z}^\top \mathbf{M} g_\theta{}'(\mathbf{x}, \mathbf{F}) \cdot \mathbf{h} + \mathcal{O}(\|\epsilon\|^2) = \mathcal{L}(\mathbf{z}, \mathbf{x}) + \epsilon \mathbf{h}^\top \mathbf{M}'^\top \cdot \mathbf{z} + \mathcal{O}\|\epsilon\|^2. \tag{C8b}$$
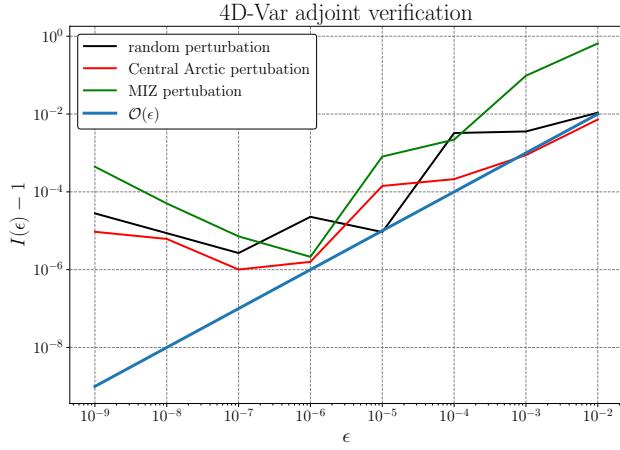
Figure C3. Logarithm of the absolute value of $\mathcal{I}(\epsilon)$ for several values of ϵ. Black line corresponds to the choice of a random perturbation of the cost function, with 10 experiments performed. Red line corresponds to a perturbation inside the Central Arctic region, with 5 experiments performed. Green line corresponds to a perturbation in the MIZ, with 5 experiments performed. Blue line corresponds to the ~~expected evolution~~ identity $\mathcal{O}(\|\epsilon\|)$.

We study the ratio

$$\mathcal{I}_{\mathbf{z}}(\epsilon) = \left\langle \left| \frac{\mathcal{L}(\mathbf{z}, \mathbf{x} + \epsilon\mathbf{h}) - \mathcal{L}(\mathbf{z}, \mathbf{x} - \epsilon\mathbf{h})}{2\epsilon\mathbf{z}^\top \mathbf{M}'(\mathbf{x}) \cdot \mathbf{h}} \frac{\mathcal{L}(\mathbf{z}, \mathbf{x} + \epsilon\mathbf{h}) - \mathcal{L}(\mathbf{z}, \mathbf{x} - \epsilon\mathbf{h})}{2\epsilon\mathbf{z}^\top g'_\theta(\mathbf{x}) \cdot \mathbf{h}} \right| \right\rangle_{\|h\|=1}, \tag{C9}$$

600   with $\mathbf{x}$ in the trajectory of the emulator. We expect the same behavior as for $\mathcal{I}(\epsilon)$. In this case, we look at several values for $\mathbf{z}$, defined exactly as the previous $\mathbf{h}$. The emulator is evaluated onto the full 2017-2018 dataset. Results are presented in Fig. C4. We obtain satisfying results, with a divergence of the residual for $\epsilon < 10^{-8}$. The curves for all different $\mathbf{z}$ vectors depict similar behavior, following the expected evolution in $\mathcal{O}(\|\epsilon\|)$. The displacement in the MIZ and in Central Arctic yields lower residual results than with a random $\mathbf{z}$ vector.

Appendix D: ~~Number of observations in the assimilation~~ Multi-year experiment

605

~~An important factor influencing the quality of the assimilation is the frequency of the observations. We denote the number of observations during one cycle of 16 d (32 model iterations) as N_{obs}. As shown in Fig 5, when there are too few observations per cycle, no improvement is observed in the analysis compared to the forecast at the end of the previous cycle (this is the case for both 2 and 4 observations per cycle). Once 8 observations per cycle are reached, this divergence disappears, which is why we~~
610   ~~adopt 8 observations per cycle in our setup.~~ In this section, we present additional results showing the full two-year assimilation, including the one year spin-up cycles which are not included in the metrics computation. This shows the seasonality of the analysis RMSE, with increase during highly thermodynamics-driven period (spring melting and winter refreezing).
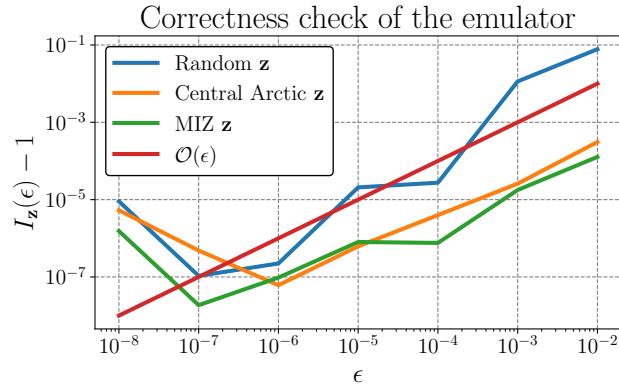
Figure C4. Logarithm of the absolute value of $\mathcal{I}(\epsilon)$ for several values of ϵ. ~~Black~~ Blue line corresponds to the choice of a random perturbation of the cost function, with 10 experiments performed. Orange line corresponds to a perturbation inside the Central Arctic region, with 5 experiments performed. Green line corresponds to a perturbation in the MIZ, with 5 experiments performed. ~~Blue~~ Red line corresponds to the ~~expected evolution~~ identity $\mathcal{O}\big(\|\epsilon\|\big)$.
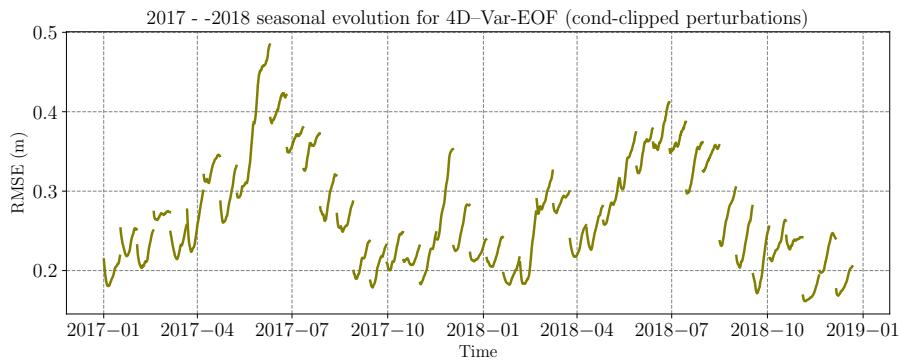
Figure D1. 2017 and 2018 RMSE of 4D–Var-EOF ~~depending on the number of observations~~ in ~~every assimilation cycle~~ the twin experiment case, with cond-clipped perturbations. ~~Curves indicated~~ The curve represents the RMSE of the 4D–Var analysis with regard to neXtSIM SIT~~, with 2 observations per cycle (purple curve), 4 observations per cycle (blue curve), 8 observations per cycle (gray curve) and 16 observations per cycle (green curve).~~

Appendix E: Background inflation

We adopt an adaptive multiplicative inflation scheme on the background error term, materialized by $\lambda_{\text{inf},n}$ and which is evaluated on each cycle n. It can be decomposed in two terms:

$$\lambda_{\text{inf},n} = \lambda_{\text{m}} \times \lambda_{\text{a},n}. \tag{E1}$$

λ_{m} corresponds to the inflation term associated with the model error. It is constant throughout the experiments. In order to test the consistency of the solution of the innovation vector, we perform the χ_p^2 test. Under Gaussian assumption, the minimum value of the cost function has a χ_p^2 distribution with p the number of observations assimilated (Michel, 2014). It means that the average of the cost function minimum should stay around p, in our case $8871 \times \text{N}_{\text{obs}}/2$, with N_{obs} the number of observations per window. Yet, with the different noise investigated in our study, this test does not work under other assumptions.

Following the χ^2 assumption, for each cycle, the minimal value of the cost function $\mathcal{J}$ should in average be equal to the number of degree of freedom (DOF), which is equal to ~~the total number of observations~~p. In practice, we define for each cycle n,

$$\lambda_{\text{a},n+1} = \sqrt{\frac{\mathcal{J}_{\text{b}}^n}{|\text{DOF} - \mathcal{J}_{\text{o}}^n|}}, \tag{E2}$$

where $\mathcal{J}_{\text{b}}^n$ corresponds to the background term of the cost function at cycle n and $\mathcal{J}_{\text{o}}^n$ corresponds to the observation term of the cost function at cycle n; with for the initial cycle, a value of 1 for $\lambda_{\text{a},0}$. The multiplication of the two terms λ_{m} and $\lambda_{\text{a},n}$ gives us the total inflation $\lambda_{\text{inf},n}$ in the adaptive case.

In order to help the optimization of the 4D–Var, we investigate the use of inflation scheme as defined in Eq. (E2). Two schemes are evaluated, a constant inflation, only modeled by λ_{m} and an adaptive inflation scheme with the ~~multiplicative addition~~ multiplication of λ_{a} based on the χ^2 estimation of $\mathcal{J}$ and λ_{m}. Only Gaussian noise for the observations are considered and the value of λ_{m} ranges between 0.6 and 3. Results with the mRMSE are outlined in Fig. E1. As observed before we can see that 4D–Var-EOF is clearly better than ~~4D-Var-diag~~4D–Var-diag, but both inflation schemes have similar behavior in both cases. Using an inflation scheme ($\lambda_{\text{inf}} \neq 1$) yields better results in terms of mRMSE in both cases if we correctly tune λ_{m}. Interestingly, in the case of constant inflation, the mRMSE increases almost linearly with λ_{m}. This indicates that, when using constant inflation, emphasizing the ~~$\mathcal{J}_{\text{b}}$~~$\mathcal{J}_{\text{o}}$ term tends to undermine the performance of the 4D–Var. Conversely, in the adaptive case, assigning a significant value to λ_{m} while still optimizing the ratio between $\mathcal{J}_{\text{b}}$ and $\mathcal{J}_{\text{o}}$ results in improved mRMSE.

As shown in Fig. E2, in which is plotted the ratio between the adaptive inflation scheme with λ_{m} set to 3 and the run with Gaussian noise without any inflation scheme, in 2018, we can see that the benefit from inflation is primarily observed at the start of the DAW and more prominently at the beginning of the year. On average, the gain from the adaptive inflation is mitigated, as some end of cycle ratio are above 1. By using the inflation, we favor the improvement at the beginning of the DAW to the detriment of the end of the DAW. Let us also note that the major gain from the inflation come from March to July, which corresponds to the period where the RMSE of the analysis $\mathbf{x}_{\text{a}}$ is higher~~, hence, where the 4D–Var is struggling the most~~.
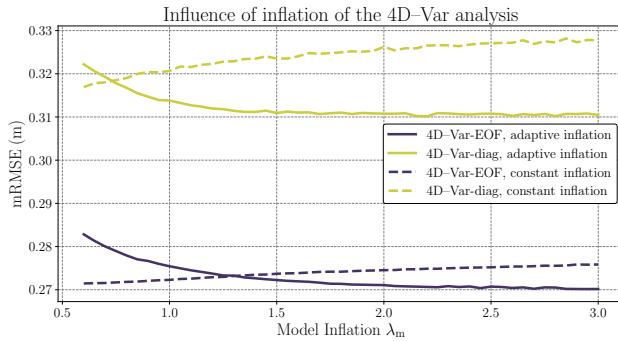
Figure E1. mRMSE of x_a for both 4D–Var-EOF (dark blue lines) and 4D–Var-diag (green lines), for the two types on background inflation proposed, the constant inflation (dashed lines) and the adaptive inflation (solid lines). The results are plotted as a function of the model inflation λ_m. In the case of the constant inflation, λ_a is automatically set to 1 whereas in the adaptive case, it is defined as per Eq. (E2).
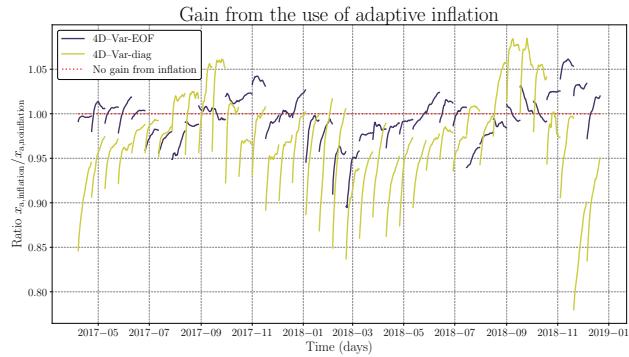
Figure E2. Evaluation of the gain from the adaptive inflation, compared to no inflation, for both 4D–Var-EOF (dark blue lines) and 4D–Var-diag (green lines). The ratio between the RMSE of $x_{a,inflation}$ and $x_{a,noinflation}$ is plotted with respect to time. The dotted red line indicates the position where there is no gain from the inflation. A value of this ratio below 1 means a gain from the adaptive scheme whereas a value above 1 corresponds to a loss in RMSEs.

To conclude, adaptive background inflation can be easily implemented in twin experiment scenarios. We observe a modest
645 improvement in the total average RMSE, with most of the gains occurring at the beginning of the assimilation windows. In the case of adaptive inflation, the inflation values fluctuate around 0.5, exhibiting a seasonal pattern.

Code and data availability. The outputs of neXtSIM model used here (Boutin et al., 2022) are available at https://ige-meom-opendap. univ-grenoble-alpes.fr/thredds/catalog/meomopendap/extract/SASIP/model-outputs/OPA-neXtSIM_CREG025/OPA-neXtSIM_CREG025-ILBOXE140- catalog.html. Forcings data from ERA5 are publicly available in the Copernicus Data Store (C3S, 2018). All the codes to build the datasets,
650 train the emulator and build the 4D–Var system are provided, the jupyter-notebook used to create the figures, as well as the post-processed datasets and neural network weights (Durand, 2024). Additionally, for real observations assimilation, CS2SMOS retrievals are publicly

available (European Space Agency, 2023). ECMWF forecast (Owens and Hewson, 2018) were provided by Alban Farchi from the ECMWF. neXtSIM-F forecast archives are not publicly available but can be requested by e-mail from nextsimf@nersc.no.

Author contributions. CD, TSF, AF, and MB refined the scientific questions and prepared an analysis strategy. JB and LB provide technical
655   knowledge on sea-ice data assimilation. CD performed the experiments. CD, TSF, AF, and MB analyzed and discussed the results. CD wrote the manuscript with TSF, AF, MB, JB, and LB reviewing.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors acknowledge the support of the project SASIP (grant no. $G-24-66154$) funded by Schmidt Sciences – a philanthropic initiative that seeks to improve societal outcomes through the development of emerging science and technologies. This
660   work was granted access to the HPC resources of IDRIS under the allocations 2021-AD011013069, 2022-AD011013069R1, and 2023-AD011013069R2 made by GENCI. The authors would like to thank Timothy Williams for its availability and access to neXtSIM-F forecast. CEREA is a member of the Institut Pierre-Simon Laplace (IPSL). This manuscript was grammatically revised using ChatGPT.

References

Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., Phillips, T., Byrne, J., Tietsche, S., Sarojini, B. B., Blanchard-Wrigglesworth, E., Aksenov, Y., Downie, R., and Shuckburgh, E.: Seasonal Arctic sea ice forecasting with probabilistic deep learning, Nature Communications, 12, https://doi.org/10.1038/s41467-021-25257-4, 2021.

Barthélémy, S., Brajard, J., Bertino, L., and Counillon, F.: Super-resolution data assimilation, Ocean Dynamics, 72, 661–678, https://doi.org/10.1007/s10236-022-01523-x, 2022.

Bernard, B., Madec, G., Penduff, T., Molines, J.-M., Treguier, A.-M., Sommer, J. L., Beckmann, A., Biastoch, A., Böning, C., Dengg, J., Derval, C., Durand, E., Gulev, S., Remy, E., Talandier, C., Theetten, S., Maltrud, M., McClean, J., and Cuevas, B. D.: Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution, Ocean Dynamics, 56, 543–567, https://doi.org/10.1007/s10236-006-0082-1, 2006.

Bouchat, A., Hutter, N., Chanut, J., Dupont, F., Dukhovskoy, D., Garric, G., Lee, Y. J., Lemieux, J.-F., Lique, C., Losch, M., Maslowski, W., Myers, P. G., Ólason, E., Rampal, P., Rasmussen, T., Talandier, C., Tremblay, B., and Wang, Q.: Sea Ice Rheology Experiment (SIREx): 1. Scaling and Statistical Properties of Sea-Ice Deformation Fields, Journal of Geophysical Research: Oceans, 127, e2021JC017 667, https://doi.org/10.1029/2021JC017667, 2022.

Boutin, G., Regan, H., Ólason, E., Brodeau, L., Talandier, C., Lique, C., and Rampal, P.: Data accompanying the article "Arctic sea ice mass balance in a new coupled ice-ocean model using a brittle rheology framework", https://doi.org/10.5281/ZENODO.7277523, 2022.

Boutin, G., Ólason, E., Rampal, P., Regan, H., Lique, C., Talandier, C., Brodeau, L., and Ricker, R.: Arctic sea ice mass balance in a new coupled ice–ocean model using a brittle rheology framework, The Cryosphere, 17, 617–638, https://doi.org/10.5194/tc-17-617-2023, 2023.

Broyden, C. G.: Quasi-Newton methods and their application to function minimisation, Mathematics of Computation, 21, 368–381, https://doi.org/10.1090/s0025-5718-1967-0224273-2, 1967.

C3S: ERA5 hourly data on single levels from 1940 to present, https://doi.org/10.24381/CDS.ADBB2D47, 2018.

Caya, A., Buehner, M., and Carrieres, T.: Analysis and Forecasting of Sea Ice Conditions with Three-Dimensional Variational Data Assimilation and a Coupled Ice–Ocean Model, Journal of Atmospheric and Oceanic Technology, 27, 353–369, https://doi.org/10.1175/2009jtecho701.1, 2010.

Chen, W., Mahmood, A., Tsamados, M., and Takao, S.: Deep Random Features for Scalable Interpolation of Spatiotemporal Data, https://doi.org/10.48550/ARXIV.2412.11350, 2024.

Cheng, S., Chen, Y., Aydoğdu, A., Bertino, L., Carrassi, A., Rampal, P., and Jones, C. K. R. T.: Arctic sea ice data assimilation combining an ensemble Kalman filter with a novel Lagrangian sea ice model for the winter 2019–2020, The Cryosphere, 17, 1735–1754, https://doi.org/10.5194/tc-17-1735-2023, 2023a.

Cheng, S., Quilodrán-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., Fablet, R., Lucor, D., Iooss, B., Brajard, J., Xiao, D., Janjic, T., Ding, W., Guo, Y., Carrassi, A., Bocquet, M., and Arcucci, R.: Machine Learning With Data Assimilation and Uncertainty Quantification for Dynamical Systems: A Review, IEEE/CAA Journal of Automatica Sinica, 10, 1361–1387, https://doi.org/10.1109/jas.2023.123537, 2023b.

Chennault, A., Popov, A. A., Subrahmanya, A. N., Cooper, R., Rafid, A. H. M., Karpatne, A., and Sandu, A.: Adjoint-Matching Neural Network Surrogates for Fast 4D-Var Data Assimilation, https://doi.org/10.48550/ARXIV.2111.08626, 2021.

Dansereau, V., Weiss, J., Saramito, P., and Lattes, P.: A Maxwell Elasto-Brittle Rheology for Sea Ice Modelling, The Cryosphere, 10, 1339–1359, https://doi.org/10.5194/tc-10-1339-2016, 2016.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, https://doi.org/10.1256/qj.05.108, 2005.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, Remote Sensing of Environment, 116, 140–158, https://doi.org/10.1016/j.rse.2010.10.017, 2012.

Driscoll, S., Carrassi, A., Brajard, J., Bertino, L., Bocquet, M., and Ólason, E. O.: Parameter sensitivity analysis of a sea ice melt pond parametrisation and its emulation using neural networks, Journal of Computational Science, p. 102231, https://doi.org/10.1016/j.jocs.2024.102231, 2024.

Durand, C.: Code and data for 'Four-dimensional variational data assimilation with a sea-ice thickness emulator', https://doi.org/10.5281/ZENODO.14418068, 2024.

Durand, C., Finn, T. S., Farchi, A., Bocquet, M., Boutin, G., and Ólason, E.: Data-driven surrogate modeling of high-resolution sea-ice thickness in the Arctic, The Cryosphere, 18, 1791–1815, https://doi.org/10.5194/tc-18-1791-2024, 2024.

European Space Agency: SMOS-CryoSat L4 Sea Ice Thickness, https://doi.org/10.57780/SM1-4F787C3, 2023.

European Union-Copernicus Marine Service: Arctic Ocean Sea Ice Analysis and Forecast, https://doi.org/10.48670/MOI-00004, 2020.

Fenty, I. and Heimbach, P.: Coupled Sea Ice–Ocean-State Estimation in the Labrador Sea and Baffin Bay, Journal of Physical Oceanography, 43, 884–904, https://doi.org/10.1175/jpo-d-12-065.1, 2013.

Finn, T. S., Durand, C., Farchi, A., Bocquet, M., Chen, Y., Carrassi, A., and Dansereau, V.: Deep learning subgrid-scale parametrisations for short-term forecasting of sea-ice dynamics with a Maxwell elasto-brittle rheology, The Cryosphere, 17, 2965–2991, https://doi.org/10.5194/tc-17-2965-2023, 2023.

Finn, T. S., Durand, C., Farchi, A., Bocquet, M., and Brajard, J.: Towards diffusion models for large-scale sea-ice modelling, https://doi.org/10.48550/ARXIV.2406.18417, 2024a.

Finn, T. S., Durand, C., Farchi, A., Bocquet, M., Rampal, P., and Carrassi, A.: Generative diffusion for regional surrogate models from sea-ice simulations, https://doi.org/10.22541/au.171386536.64344222/v1, 2024b.

Girard, L., Bouillon, S., Weiss, J., Amitrano, D., Fichefet, T., and Legat, V.: A new modeling framework for sea-ice mechanics based on elasto-brittle rheology, Annals of Glaciology, 52, 123–132, https://doi.org/10.3189/172756411795931499, 2011.

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T.: Predictability of the Arctic sea ice edge, Geophysical Research Letters, 43, 1642–1650, https://doi.org/10.1002/2015gl067232, 2016.

Gregory, W., Bushuk, M., Adcroft, A., Zhang, Y., and Zanna, L.: Deep Learning of Systematic Sea Ice Model Errors From Data Assimilation Increments, Journal of Advances in Modeling Earth Systems, 15, https://doi.org/10.1029/2023ms003757, 2023.

Gregory, W., Bushuk, M., Zhang, Y., Adcroft, A., and Zanna, L.: Machine Learning for Online Sea Ice Bias Correction Within Global Ice-Ocean Simulations, Geophysical Research Letters, 51, https://doi.org/10.1029/2023gl106776, 2024a.

Gregory, W., MacEachern, R., Takao, S., Lawrence, I. R., Nab, C., Deisenroth, M. P., and Tsamados, M.: Scalable interpolation of satellite altimetry data with probabilistic machine learning, Nature Communications, 15, https://doi.org/10.1038/s41467-024-51900-x, 2024b.

Grigoryev, T., Verezemskaya, P., Krinitskiy, M., Anikin, N., Gavrikov, A., Trofimov, I., Balabin, N., Shpilman, A., Eremchenko, A., Gulev, S., Burnaev, E., and Vanovskiy, V.: Data-Driven Short-Term Daily Operational Sea Ice Regional Forecasting, Remote Sensing, 14, 5837, https://doi.org/10.3390/rs14225837, 2022.

Hatfield, S., Chantry, M., Dueben, P., Lopez, P., Geer, A., and Palmer, T.: Building Tangent-Linear and Adjoint Models for Data Assimilation With Neural Networks, Journal of Advances in Modeling Earth Systems, 13, https://doi.org/10.1029/2021ms002521, 2021.

Hebert, D. A., Allard, R. A., Metzger, E. J., Posey, P. G., Preller, R. H., Wallcraft, A. J., Phelps, M. W., and Smedstad, O. M.: Short-term sea ice forecasting: An assessment of ice concentration and ice drift forecasts using the U.S. Navy's Arctic Cap Nowcast/Forecast System, Journal of Geophysical Research: Oceans, 120, 8327–8345, https://doi.org/10.1002/2015jc011283, 2015.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hibler, W. D.: A Dynamic Thermodynamic Sea Ice Model, Journal of Physical Oceanography, 9, 815–846, https://doi.org/10.1175/1520-0485(1979)009<0815:adtsim>2.0.co;2, 1979.

Hunke, E. C. and Dukowicz, J. K.: An Elastic–Viscous–Plastic Model for Sea Ice Dynamics, Journal of Physical Oceanography, 27, 1849–1867, https://doi.org/10.1175/1520-0485(1997)027<1849:aevpmf>2.0.co;2, 1997.

Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, https://doi.org/10.48550/ARXIV.1502.03167, 2015.

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M., and Raynaud, L.: Ensemble of data assimilations at ECMWF, https://doi.org/10.21957/OBKE4K60, 2010.

Ji, Q., Zhu, X., Wang, H., Liu, G., Gao, S., Ji, X., and Xu, Q.: Assimilating operational SST and sea ice analysis data into an operational circulation model for the coastal seas of China, Acta Oceanologica Sinica, 34, 54–64, https://doi.org/10.1007/s13131-015-0691-y, 2015.

Kauker, F., Kaminski, T., Karcher, M., Giering, R., Gerdes, R., and Voßbeck, M.: Adjoint analysis of the 2007 all time Arctic sea-ice minimum, Geophysical Research Letters, 36, https://doi.org/10.1029/2008gl036323, 2009.

Kimmritz, M., Counillon, F., Bitz, C., Massonnet, F., Bethke, I., and Gao, Y.: Optimising assimilation of sea ice concentration in an Earth system model with a multicategory sea ice model, Tellus A: Dynamic Meteorology and Oceanography, 70, 1435 945, https://doi.org/10.1080/16000870.2018.1435945, 2018.

Koldunov, N. V., Köhl, A., Serra, N., and Stammer, D.: Sea ice assimilation into a coupled ocean–sea ice model using its adjoint, The Cryosphere, 11, 2265–2281, https://doi.org/10.5194/tc-11-2265-2017, 2017.

Kurtz, N. and Harbeck, J.: CryoSat-2 Level 4 Sea Ice Elevation, Freeboard, and Thickness, Version 1, https://doi.org/10.5067/96JO0KIFDAS8, 2017.

Landy, J. C., Petty, A. A., Tsamados, M., and Stroeve, J. C.: Sea Ice Roughness Overlooked as a Key Source of Uncertainty in CryoSat-2 Ice Freeboard Retrievals, Journal of Geophysical Research: Oceans, 125, https://doi.org/10.1029/2019jc015820, 2020.

Lemieux, J., Beaudoin, C., Dupont, F., Roy, F., Smith, G. C., Shlyaeva, A., Buehner, M., Caya, A., Chen, J., Carrieres, T., Pogson, L., DeRepentigny, P., Plante, A., Pestieau, P., Pellerin, P., Ritchie, H., Garric, G., and Ferry, N.: The Regional Ice Prediction System (RIPS): verification of forecast sea ice concentration, Quarterly Journal of the Royal Meteorological Society, 142, 632–643, https://doi.org/10.1002/qj.2526, 2015.

Lindsay, R. W. and Zhang, J.: Assimilation of Ice Concentration in an Ice–Ocean Model, Journal of Atmospheric and Oceanic Technology, 23, 742–749, https://doi.org/10.1175/jtech1871.1, 2006.

Liu, D. C. and Nocedal, J.: On the limited memory BFGS method for large scale optimization, Mathematical Programming, 45, 503–528, https://doi.org/10.1007/bf01589116, 1989.

Liu, J., Chen, Z., Hu, Y., Zhang, Y., Ding, Y., Cheng, X., Yang, Q., Nerger, L., Spreen, G., Horton, R., Inoue, J., Yang, C., Li, M., and Song, M.: Towards reliable Arctic sea ice prediction using multivariate data assimilation, Science Bulletin, 64, 63–72, https://doi.org/10.1016/j.scib.2018.11.018, 2019.

Liu, Q., Zhang, R., Wang, Y., Yan, H., and Hong, M.: Daily Prediction of the Arctic Sea Ice Concentration Using Reanalysis Data Based on a Convolutional LSTM Network, Journal of Marine Science and Engineering, 9, 330, https://doi.org/10.3390/jmse9030330, 2021.

Madec, G., Delecluse, P., Imbard, M., and Levy, C.: OPA 8 Ocean General Circulation Model - Reference Manual, Tech. rep., LODYC/IPSL Note 11, 1998.

Massonnet, F., Fichefet, T., and Goosse, H.: Prospects for improved seasonal Arctic sea ice predictions from multivariate data assimilation, Ocean Modelling, 88, 16–25, https://doi.org/10.1016/j.ocemod.2014.12.013, 2015.

Michel, Y.: Diagnostics on the cost-function in variational assimilations for meteorological models, Nonlinear Processes in Geophysics, 21, 187–199, https://doi.org/10.5194/npg-21-187-2014, 2014.

Nab, C., Mignac, D., Landy, J., Martin, M., Stroeve, J., and Tsamados, M.: Sensitivity to Sea Ice Thickness Parameters in a Coupled Ice-Ocean Data Assimilation System, Journal of Advances in Modeling Earth Systems, 17, https://doi.org/10.1029/2024ms004276, 2025.

Owens, R. and Hewson, T.: ECMWF Forecast User Guide, https://doi.org/10.21957/M1CS7H, 2018.

Palerme, C., Lavergne, T., Rusin, J., Melsom, A., Brajard, J., Kvanum, A. F., Macdonald Sørensen, A., Bertino, L., and Müller, M.: Calibration of short-term sea ice concentration forecasts using deep learning, The Cryosphere, https://doi.org/10.5194/egusphere-2023-2439, 2023.

Rampal, P., Bouillon, S., Ólason, E., and Morlighem, M.: neXtSIM: A New Lagrangian Sea Ice Model, The Cryosphere, 10, 1055–1073, https://doi.org/10.5194/tc-10-1055-2016, 2016.

Rampal, P., Dansereau, V., Olason, E., Bouillon, S., Williams, T., Korosov, A., and Samaké, A.: On the multi-fractal scaling properties of sea ice deformation, The Cryosphere, 13, 2457–2474, https://doi.org/10.5194/tc-13-2457-2019, 2019.

Raynaud, L., Berre, L., and Desroziers, G.: Spatial averaging of ensemble-based background-error variances, Quarterly Journal of the Royal Meteorological Society, 134, 1003–1014, https://doi.org/10.1002/qj.245, 2008.

Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J., and Haas, C.: A weekly Arctic sea-ice thickness data record from merged CryoSat-2 and SMOS satellite data, The Cryosphere, 11, 1607–1623, https://doi.org/10.5194/tc-11-1607-2017, 2017.

Robert, C., Durbiano, S., Blayo, E., Verron, J., Blum, J., and Le Dimet, F.-X.: A reduced-order strategy for 4D-Var data assimilation, Journal of Marine Systems, 57, 70–82, https://doi.org/10.1016/j.jmarsys.2005.04.003, 2005.

Rousset, C., Vancoppenolle, M., Madec, G., Fichefet, T., Flavoni, S., Barthélemy, A., Benshila, R., Chanut, J., Levy, C., Masson, S., and Vivier, F.: The Louvain-La-Neuve sea ice model LIM3.6: global and regional capabilities, Geoscientific Model Development, 8, 2991–3005, https://doi.org/10.5194/gmd-8-2991-2015, 2015.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K. A., Oke, P. R., and Korablev, A.: TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Science, 8, 633–656, https://doi.org/10.5194/os-8-633-2012, 2012.

Sasaki, y.: SOME BASIC FORMALISMS IN NUMERICAL VARIATIONAL ANALYSIS, Monthly Weather Review, 98, 875–883, https://doi.org/10.1175/1520-0493(1970)098<0875:sbfinv>2.3.co;2, 1970.

Sitzmann, V., Martel, J., Bergman, A., Lindell, D., and Wetzstein, G.: Implicit neural representations with periodic activation functions, Advances in neural information processing systems, 33, 7462–7473, 2020.

Talagrand, O. and Courtier, P.: Variational Assimilation of Meteorological Observations With the Adjoint Vorticity Equation. I: Theory, Quarterly Journal of the Royal Meteorological Society, 113, 1311–1328, https://doi.org/10.1002/qj.49711347812, 1987.

Talandier, C. and Lique, C.: $CREG025.L75 - NEMO_r3.6.0$, https://doi.org/10.5281/ZENODO.5802028, 2021.

Tian-Kunze, X., Kaleschke, L., Maaß, N., Mäkynen, M., Serra, N., Drusch, M., and Krumpen, T.: SMOS-derived thin sea ice thickness: algorithm baseline, product specifications and initial verification, The Cryosphere, 8, 997–1018, https://doi.org/10.5194/tc-8-997-2014, 2014.

Tietsche, S., Notz, D., Jungclaus, J. H., and Marotzke, J.: Assimilation of sea-ice concentration in a global climate model – physical and statistical aspects, Ocean Science, 9, 19–36, https://doi.org/10.5194/os-9-19-2013, 2013.

Toyoda, T., Fujii, Y., Yasuda, T., Usui, N., Ogawa, K., Kuragano, T., Tsujino, H., and Kamachi, M.: Data assimilation of sea ice concentration into a global ocean–sea ice model with corrections for atmospheric forcing and ocean temperature fields, Journal of Oceanography, 72, 235–262, https://doi.org/10.1007/s10872-015-0326-0, 2015.

Toyoda, T., Hirose, N., Urakawa, L. S., Tsujino, H., Nakano, H., Usui, N., Fujii, Y., Sakamoto, K., and Yamanaka, G.: Effects of Inclusion of Adjoint Sea Ice Rheology on Backward Sensitivity Evolution Examined Using an Adjoint Ocean–Sea Ice Model, Monthly Weather Review, 147, 2145–2162, https://doi.org/10.1175/mwr-d-18-0198.1, 2019.

Usui, N., Wakamatsu, T., Tanaka, Y., Hirose, N., Toyoda, T., Nishikawa, S., Fujii, Y., Takatsuki, Y., Igarashi, H., Nishikawa, H., Ishikawa, Y., Kuragano, T., and Kamachi, M.: Four-dimensional variational ocean reanalysis: a 30-year high-resolution dataset in the western North Pacific (FORA-WNP30), Journal of Oceanography, 73, 205–233, https://doi.org/10.1007/s10872-016-0398-5, 2016.

Wang, K., Debernard, J., Sperrevik, A. K., Isachsen, P. E., and Lavergne, T.: A combined optimal interpolation and nudging scheme to assimilate OSISAF sea-ice concentration into ROMS, Annals of Glaciology, 54, 8–12, https://doi.org/10.3189/2013aog62a138, 2013.

Williams, T., Korosov, A., Rampal, P., and Ólason, E.: Presentation and evaluation of the Arctic sea ice forecasting system neXtSIM-F, The Cryosphere, 15, 3207–3227, https://doi.org/10.5194/tc-15-3207-2021, 2021.

Xiao, Y., Bai, L., Xue, W., Chen, K., Han, T., and Ouyang, W.: FengWu-4DVar: Coupling the Data-driven Weather Forecasting Model with 4D Variational Assimilation, https://doi.org/10.48550/ARXIV.2312.12455, 2023.

Xie, J., Counillon, F., and Bertino, L.: Impact of assimilating a merged sea-ice thickness from CryoSat-2 and SMOS in the Arctic reanalysis, The Cryosphere, 12, 3671–3691, https://doi.org/10.5194/tc-12-3671-2018, 2018.

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M.: The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment, Ocean Science, 15, 779–808, https://doi.org/10.5194/os-15-779-2019, 2019.

Ólason, E., Boutin, G., Korosov, A., Rampal, P., Williams, T., Kimmritz, M., Dansereau, V., and Samaké, A.: A New Brittle Rheology and Numerical Framework for Large-Scale Sea-Ice Models, Journal of Advances in Modeling Earth Systems, 14, https://doi.org/10.1029/2021ms002685, 2022.