https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

Reconstruction of Arctic sea ice thickness (1992-2010) based on a hybrid machine learning and data assimilation approach

Léo Edel[1,2], Jiping Xie[1,2], Anton Korosov[1,2], Julien Brajard[1,2], and Laurent Bertino[1,2]

[1]Nansen Environmental and Remote Sensing Center, Bergen 5007, Norway
[2]Bjerknes Center, Bergen 5007, Norway

Correspondence: Léo Edel (leo.edel@nersc.no)

Abstract. Arctic sea ice thickness (SIT) remains one of the most crucial yet challenging parameters to estimate. Satellite data generally presents temporal and spatial discontinuities, which constrain studies focusing on long-term evolution. Since 2011, the combined satellite product CS2SMOS enables more accurate SIT retrievals that significantly decrease modelled SIT errors during assimilation. Can we extrapolate the benefits of data assimilation to past periods lacking accurate SIT observations? In this study, we train a machine learning (ML) algorithm to learn the systematic SIT errors between two versions of the model TOPAZ4 over 2011-2022, with and without CS2SMOS assimilation, to predict the SIT error and extrapolate the SIT prior to 2011. The ML algorithm relies on SIT coming from the two versions of TOPAZ4, various oceanographic variables, and atmospheric forcings from ERA5. Over the test period 2011-2013, the ML method outperforms TOPAZ4 without CS2SMOS assimilation when compared to TOPAZ4 assimilating CS2SMOS. The root mean square error of Arctic averaged SIT decreases from 0.42 to 0.28 meters and the bias from -0.18 to 0.01 meters. Also, despite the lack of observations available for assimilation in summer, our method still demonstrates a crucial improvement in SIT. Relative to independent mooring data in the Central Arctic between 2001 and 2010, mean SIT bias reduces from -1.74 meters to -0.85 meters when using the ML algorithm. Ultimately, the ML-adjusted SIT reconstruction reveals an Arctic mean SIT of 1.61 meters in 1992 compared to 1.08 meters in 2022. This corresponds to a decline in total sea ice volume from 19,690 to 12,700 km^3, with an associated trend of -3,153 km^3/decade. These changes are accompanied by a distinct shift in SIT distribution. Our innovative approach proves its ability to correct a significant part of the primary biases of the model by combining data assimilation with machine learning. Although this new reconstructed SIT dataset has not yet been assimilated into TOPAZ4, future work could enable the correction to be further propagated to other sea ice and ocean variables.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

1 Introduction

In this study, we investigate an original approach combining data assimilation and machine learning to correct past model estimations of sea ice thickness using present observations. While in situ observations offer unparalleled accuracy, they lack global coverage, contrasting with remote sensing observations that, although global, are associated with large uncertainties due to necessary assumptions for estimation. At present, the best estimation is commonly obtained by integrating remote sensing observations into models to reduce their biases. However, this approach relies on the availability of observations and, as a result, cannot help retrieve historical sea ice thickness. Studies focusing on long-term evolution, particularly those oriented toward climate research, demand extensive and accurate time series of sea ice thickness, given the essential role sea ice plays as the interface between ocean and atmosphere.

Arctic sea ice acts as a multifaceted and vital interface between the ocean and the atmosphere, playing a major role in regulating energy exchange, reflecting sunlight, and influencing local weather patterns. Sea ice significantly influences marine ecosystems, providing habitat and migration routes for diverse species (Kahru et al., 2011; Frainer et al., 2017). As sea ice melts, it injects freshwater into the ocean, affecting salinity levels and exposes ocean to the atmosphere. Moreover, as Arctic sea ice extent is declining due to warming (Comiso et al., 2008), the Arctic is becoming more navigable, opening up new opportunities for maritime transportation and resource exploration, but also raising concerns about environmental impacts and sustainable management of the region's fragile ecosystems (Aksenov et al., 2017). Notably, the thickness of Arctic sea ice stands as a major unknown parameter as thicker ice, usually older and deformed, resists better to melting and mechanical stresses. Its variations are intricately tied to the heat and freshwater budget, the sea ice dynamics, and the ecosystem.

The current deficiency in a comprehensive and accurate climate record for sea ice thickness (SIT) is attributed to the sparse availability of SIT observations and the relatively recent integration of satellite technology. Although SIT observations have been taken in situ (Lindsay and Schweiger, 2015) and by no less than five satellites, they generally suffer from severe representativity issues, high uncertainties (Zygmuntowska et al., 2014) and lack both the temporal and spatial continuity that long-term climate studies need. Consequently, model reanalyses of Arctic SIT diverge substantially (Uotila et al., 2019) and lack credibility. An extended reconstruction of Arctic sea ice thickness, along with its uncertainty estimates, is essential to unlock investigations on the Arctic climate, including heat budgets (Trenberth et al., 2019), freshwater fluxes (Solomon et al., 2021) and its ecosystem (Arrigo, 2014).

Physical-based sea ice models (e.g. Hunke and Dukowicz (1997)) can simulate reasonable sea ice thickness, yet SIT biases in numerical models remain important, originating from various factors, including external components like atmospheric or ocean fluxes and internal aspects intrinsic to the model itself. Intercomparisons of SIT between state-of-the-art models thus exhibit large deviations from one model to another in terms of spatial distribution (Johnson et al., 2012; Uotila et al., 2019; Watts et al., 2021). This is also the case when comparing satellite products (Sallila et al., 2019) or diverse in-situ datasets, mostly due to the difference in spatial and temporal distributions (Lindsay and Schweiger, 2015; Labe et al., 2018).

Since 2010, the merged remote sensing product CS2SMOS provides continuous SIT every winter combining data from SMOS and CryoSat-2 for thin and thick ice respectively (Ricker et al., 2017), yet longer time series are required to conduct

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

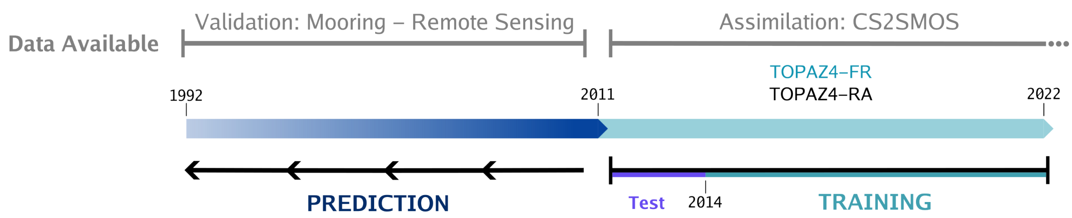

Figure 1. Chronological conception of our study. Development of the ML algorithm is based on 2011-2022. Prediction by the ML algorithm is done from 2011 backward in time until 1992. CS2SMOS serves for the development of our ML algorithm, while ULS and ICESat provide for the evaluation of its prediction.

climate studies. Assimilating CS2SMOS data in the coupled ocean-sea-ice model TOPAZ corrects a low SIT bias of about 16 cm and thus decreases RMS errors from 53 cm to 38 cm and down to 20 cm in March (Xie et al., 2018; Xiu et al., 2021). Can we extrapolate the benefits of data assimilation to past periods without SIT observations? Brajard et al. (2020) introduced a method to iteratively combine data assimilation (DA) with machine learning (ML) to built a chaotic model. The present study is applying this approach to 'rewind' a climate record, focusing on the first iteration only.

Machine learning has advanced to a point where it can effectively address the high dimensionality, complexity, and nonlinearity inherent in dynamical systems (Rolnick et al., 2022), especially when combined with DA (Cheng et al., 2023). Recent investigations demonstrated the potential of machine learning for sea ice, focusing on various objectives such as parameterizing subgrid-scale dynamics (Finn et al., 2023), emulating sea ice melt ponds (Driscoll et al., 2023), or skillfully predicting DA increments of sea ice concentration across all seasons (Gregory et al., 2023). In the present study, our assumption is that a suitable compression of the variables at play (e.g. via Empirical Orthogonal Function, EOF) identifies the complex nonlinear relationships between physical variables, without distorting them (Liu et al., 2023).

In the present investigation, we train a machine learning algorithm to learn the systematic SIT errors between two versions of the model TOPAZ4 over 2011-2022, with (TOPAZ4-RA) and without CS2SMOS assimilation (TOPAZ4-FR). Then, we use the algorithm to predict the SIT error and extrapolate the SIT prior to 2011 (Fig. 1). For this work, the training period (2014-2022) supports algorithm development and includes a validation period (20% of the training period, in chronological order without randomization) to optimize hyperparameters. The test period (2011-2013) enables us to verify our algorithm performances with the data held specifically for this purpose. The evaluation period (1992-2010) allows us to assess the ML-adjusted SIT, called TOPAZ4-ML, compared to independent datasets.

Section 2 describes various datasets and the model TOPAZ4. Section 3 further explains the method used to combine DA and ML. Section 4 presents the results and evaluation of the ML algorithm, as well as an assessment of the extended SIT time series with independent datasets, and highlights unprecedented outcomes from this brand-new product. Section 5 discusses the limitations and uncertainties of this investigation.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

2 Datasets

2.1 CS2SMOS

The CS2SMOS sea ice thickness (SIT) product (Ricker et al., 2017) combines measurements from two satellite missions: CryoSat-2 (CS2) and Soil Moisture and Ocean Salinity (SMOS). CryoSat-2 (Wingham et al., 2006) utilizes radar altimetry to measure the height of the ice surface above the water level, which is converted to sea ice thickness assuming hydrostatic equilibrium. SMOS (Kaleschke et al., 2012) measures microwave emissions at 1.4 GHz, allowing to derive sea ice thickness in thin ice. The first is able to measure thick ice (>~1m) while the latter is primarily designed for thin ice (<~1m), so this advanced merged product provides the first estimate of the total spectrum of sea ice thickness. As neither CS2 nor SMOS can measure SIT during the melting season, the period of observation is limited from October to April, starting in 2010. The average uncertainty is typically around 0.50m, with CS2 uncertainties ranging from 0.1 to 1m and SMOS uncertainties inferior to 1.1m in thin ice (Ricker et al., 2017). The novel year-round processing of CS2 by Landy et al. (2022) was not considered here due to artefacts in the transitions from summer to winter.

2.2 TOPAZ4

TOPAZ is a regional coupled ocean-sea-ice data assimilation system successfully implemented into the Arctic Ocean operational forecast, and version 4 is described in Sakov et al. (2012) and Xie et al. (2017). It is built on the HYCOM ocean model (Bleck, 2002), coupled with a sea ice model based on elastic-viscous-plastic (EVP) rheology (Hunke and Dukowicz, 1997) and rudimentary thermodynamics Drange and Simonsen (1996). The data assimilation is based on a deterministic formulation of the ensemble Kalman filter (DEnKF, detailed in Sakov and Oke (2008)), using 100 dynamical members to assimilate various ocean and sea ice observations (see Xie et al. (2018) and Xie et al. (2023)). Historically, the system has used the atmospheric forcing fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) to drive the model. To archive the real-time forecast, the system is forced by the operational weather forecast product. But for a long-time model run, such as obtaining the Arctic reanalysis, we use the corresponding atmosphere reanalysis product released by ECMWF as well.

Rather than learning from the winter-only satellite observations, which would not provide any information in the summer season, two model runs have been produced: without and with assimilation, covering the years 1992-2022. Both of them are forced by the latest ECMWF Reanalysis Version 5 (ERA5, Hersbach et al. (2020)) and provide daily outputs on regular grids with a spatial resolution of 10 km. In this study, the raw version of TOPAZ4, without assimilation, is hereafter called free run or TOPAZ4-FR. For TOPAZ4 with assimilation, called TOPAZ4-RA, we weekly assimilated Sea Level Anomalies (SLA, doi:10.48670/moi-00146), Sea Surface Temperatures (SST, doi:10.48670/moi-00169), in situ profiles of temperature and salinity (doi:10.17882/46219 and doi:10.48670/00036), Sea Surface Salinity (SSS, Version 3.1 from the Barcelona Expert Center), sea ice concentrations (SIC, doi:10.48670/moi-00136) and sea ice drift (SID) from the Ocean and Sea Ice Satellite Application Facility (OSISAF), and Sea Ice Thickness (SIT, doi:10.48670/moi-00126) from CS2SMOS (see Ricker et al. (2017)). The assimilation is performed weekly and SIT assimilation is only carried out from October to April after 2011. All the observations except for SSS and SID are downloaded from the Copernicus Marine Environment Monitoring Service (CMEMS). Since

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

2004, the Ice-Tethered Profilers (ITP) can provide more density-layered profiles under sea ice and provide rare information for measuring polar marine environments. However, its appearance in the TOPAZ4 system within a limited representative ensemble brings considerable interference to the SIT update, especially in the summer absence of SIT observation. To overcome this nonphysical response of sea ice updating to ITP, TOPAZ4-RA implements some specific changes that have been made posterior to Xie et al. (2017). In each assimilation cycle, the final optimization of the model state consists of two steps. First, all ocean variables are updated as before. In the second step, the sea ice variables are updated but switching off the covariance contributions from the in-situ profiles. As a preprocessing step, if the sea ice concentration in TOPAZ4 free run falls below 15%, we interpret this as the absence of sea ice (SIC = 0 and SIT = 0) in TOPAZ4-RA. This decision was made to maintain consistency between the two TOPAZ4 runs and ensure uniformity in sea ice extent across both datasets.

2.3 ERA5

In this study, the atmospheric fields from the latest ECMWF Reanalysis ERA5 (Hersbach et al., 2020) are used as predictors for our ML algorithm. They bring valuable information about environmental conditions that improve the bias prediction. The following variables are used at the surface level: air temperature, mean sea-level pressure, total precipitation, and wind speed East-West and North-South. Daily averaged fields at the horizontal resolution of 31 km are projected onto the TOPAZ4 grid. Next, they are processed following the methodology described in section 3.

2.4 Sea ice age

The observed mean sea ice age (Korosov et al., 2018) is used as a predictor, which we consider more precise than a modeled one. In this product, the advection scheme predicts the subsequent creation or loss of new ice by taking into account the observed divergence or convergence, freezing, or melting of sea ice. Sea ice concentration and daily gridded drift products from the OSISAF are used by the algorithm. The primary benefit of the new technique lies in its capacity to produce unique ice age fractions for every pixel in the output result, providing the ice age's frequency distribution which allows us to obtain the mean, median, or weighted average. This feature should aid the machine learning model, as the sea ice age is a proxy of thickness; older ice has undergone more growth, freezing, and compression processes (Liu et al., 2020).

2.5 Validation data: Mooring data

In-situ observations have been gathered to evaluate our ML-adjusted daily SIT at different times and places in the Arctic. Upward Looking Sonars (ULS) are the most statistically robust instruments deployed in the Arctic for measuring the sea ice draft from underneath the drifting ice pack (Krishfield and Proshutinsky, 2006). The sea ice thickness can then be derived assuming the hydrostatic equilibrium. In this work, SIT is computed by multiplying the sea ice draft by a factor of 1.12, which corresponds approximately to the ratio of mean seawater density and sea ice density (Sumata et al., 2023; Johnson et al., 2012; Bourke and Paquette, 1989). The datasets listed in Table 1 are used during the validation period prior to 2011. They have been collected by the Beaufort Gyre Exploration Project (BGEP, https://www2.whoi.edu/site/beaufortgyre/), and the North Pole

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

Environmental Observatory (NPEO, http://psc.apl.washington.edu/northpole/). Their locations are exhibited in Supplements Fig. S3. We employ a 7-day running mean to smooth the mooring data for a more consistent comparison. To do so, we choose the nearest grid point to the mooring location, from which we extract daily SIT values from the model.

Table 1. Mooring data used in this study. Beaufort Gyre Exploration Project is abbreviated as BGEP, North Pole Environmental Observatory as NPEO. ULS stands for Upward Looking Sonars.

Name	Sensor	Location	Number of buoy	Frequency of measure	Years	Length	Accuracy of ice draft
BGEP	ULS	Beaufort Gyre	4	2-second	2003-2011	3 to 7 years	± 5/10 cm
NPEO	ULS	North Pole	1	5 to 10-minute	2001-2010	9 years	± 5 cm for level ice

2.6 Validation data: Remote sensing

ICESat-1 (Ice, Cloud, and land Elevation Satellite) emerged as a pioneering instrument for the assessment of sea ice thickness, specifically designed for polar regions (Schutz et al., 2005). Despite its innovative approach, the Geoscience Laser Altimeter System (GLAS) encountered dysfunction that forced it to operate only for one-month periods out of every three to six months to extend the time series of measurements. It operated from January 2003 to October 2009, resulting in 15 campaigns in the Arctic. The process of converting the retrieved freeboard and the snow cover climatologies is further explained in Kwok and Cunningham (2008), allowing ICESat-1 to provide mean SIT for each campaign at a spatial resolution of 25 km x 25 km. The satellite orbital configuration causes a data gap at latitudes north of 86 °N, which is filled through interpolation.

Envisat, the European Space Agency's (ESA) satellite launched in 2002, has played a crucial role in advancing our understanding of Earth's polar regions. The dataset provides sea ice thickness derived from the Radar Altimeter-2 instrument, developed by the ESA Climate Change Initiative (CCI) project. It provides monthly gridded sea ice thickness data for the freezing period (October-March) from 2002 to 2012. The spatial resolution is 25 km x 25 km in the Arctic, with the pole hole north of 81,5 °N.

Previous studies utilizing these satellites drew the following conclusions. Envisat, with its sensor's coarse resolution (~2 km footprint), primarily samples larger and thicker sea ice (Tilling et al., 2019), whereas ICESat-1's sensor has a much finer footprint (~170 m), enabling more detailed measurements. In comparison to airborne and ULS data, ICESat-1 SIT was consistently less than that of CryoSat-2 by ~50 cm Kim et al. (2020).

3 Method for Sea Ice Thickness Adjustment

Our adjustment method is applied as a post-processing operation on the TOPAZ4 free run, dependent on the state of the sea ice but also the external forcing variables. The approach is based on the Empirical Orthogonal Functions (EOF) decomposition, to reduce the dimensionality of our problem. This data compression enables us to apply a strong adjustment that requires minimal

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

computational resources and remains unaffected by static geographic features, such as the coastlines. During the training period (2014-2022), we compute the Empirical Orthogonal Functions (EOF, spatial component of the statistical patterns) and associated Principal Components (PC, temporal evolution of the statistical patterns) of the SIT biases. Applying the method outside of the training period assumes that this EOF decomposition is stable back in time. Consequently, the EOFs of the SIT biases are assumed invariant, while the target variables to predict are the PCs of SIT biases back in time (Fig. 2). The increments from data assimilation give the best estimates of SIT biases, and we learn to emulate these increments, similarly to what is done in e.g., Brajard et al. (2020); Gregory et al. (2023). Using the increments rather than the innovations means that the algorithm can be used with irregular observations while the data assimilation takes care of their interpolation. Likewise, each input feature (listed in Tab. 2) is decomposed independently using EOF with several components ranging from four to height. At first, 14 a priori relevant features are used as inputs, and then an arbitrary threshold on the variable importance enables an adequate selection of the best-suited variables. Considering that using eight components for the EOF decomposition of the SIT bias yields satisfactory results (Supplements Fig. S5), the subsequent results will exclusively focus on this configuration.

Table 2. List of variables used as inputs of the machine learning algorithm. A crossed cell indicates that the variable is used for the corresponding PC. The lower part of the table displays the parameters used to train each model.

Variable	Source	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Sea ice thickness	TOPAZ4 Free run	x	x	x	x	x	x		x
Sea ice concentration			x	x	x	x	x	x	x
Snow depth on top of sea ice		x	x	x	x	x	x	x	x
Sea surface height above geoid		x	x	x	x	x	x	x	x
Sea ice drift x velocity		x	x	x	x	x	x	x	x
Sea ice drift y velocity		x	x	x	x	x	x	x	x
Sea ice age	Korosov et al. (2018)	x	x	x	x	x	x	x	x
Air temperature at 2 meters	ERA5	x	x	x	x	x	x	x	x
Mean sea level		x	x	x	x	x	x		x
10m wind U		x	x	x	x	x	x		x
10m wind V			x	x	x		x	x	x
Total precipitation		x	x	x	x			x	
Surface net Solar Radiation		x		x	x	x	x		x
Surface net Thermal Radiation		x	x	x	x	x	x		x
Number of input features		12	13	14	14	12	13	9	13
Number of epochs		100	40	60	70	50	60	100	100

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

Long Short-Term Memory (LSTM) is a recurrent neural network designed to model chronological sequences and store information on a long time range (Hochreiter and Schmidhuber, 1997). LSTM estimates the current prediction using data from its own prior prediction and enables the propagation of the bias backward in time like a nonlinear type of autoregressive process. A unique model is developed for every single PC, as each depends on different input variables and time lag. The architecture is composed of three backward-prediction LSTM layers alternated by dropout layers, which prevent overfitting by randomly deactivating neural connections during training. The hyperparameters such as the number of components of the inputs, the input variables, and their time lags can change between models, while the overall architecture remains the same. Details regarding the differences between each model can be found in Tab. 2. Since certain PCs proved more challenging to predict than others, a comprehensive analysis of PC prediction is provided in appendix A to better understand the performances of each model. Throughout this investigation, we discovered that the input variables have a much greater impact on the prediction than the ML architecture.

The uncertainty associated with the nonlinear estimation is computed by introducing random walk processes to perturb the inputs of the LSTM. Multiple perturbation instances are employed to compute the ML-adjusted SIT, and the standard deviation of the resulting ensemble of SIT predictions is used as uncertainty estimate. It is important to note that this uncertainty solely characterizes the sensitivity of the algorithm to its inputs, and does not encompass the uncertainty associated with the training process of the ML algorithm. The final uncertainty is computed using 50 members, with a random walk perturbation of the inputs set at 100% of the original values scaled between -1 and 1.

To predict SIT biases in the past, our method is the following. We project the values of each input variable onto its principal components. As a result, we obtain a time series of each principal component for each variable. Then, the ML algorithm predicts the PCs of the SIT bias, and thus SIT biases can be retrieved by inverting the EOF projection. Lastly, TOPAZ4-ML SIT is reconstructed by adding SIT biases to TOPAZ4-FR. For an integrated sanity check, we evaluate the total sea ice volume as the product of sea ice thickness with concentration and the area of each grid cell.

We introduce a trivial bias correction as a baseline to evaluate the efficiency of our ML adjustment. Monthly biases between TOPAZ4-RA and TOPAZ4-FR are averaged from 2014 to 2022. The daily baseline SIT, called TOPAZ4-BL, is then obtained by adding the monthly biases to the SIT from TOPAZ4-FR at each grid point for the corresponding month. Considering that TOPAZ4 generally has too thin SIT in areas of thick ice, even this simple baseline a priori constitutes a solid benchmark.

4 Results

After a brief analysis of the SIT biases of TOPAZ4, the sections below follow the standard steps of ML applications, first testing the algorithm on withdrawn data (2011-2013) and then predicting SIT biases outside of the training and testing windows, extending in our case the SIT data into the past (1992-2010).

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

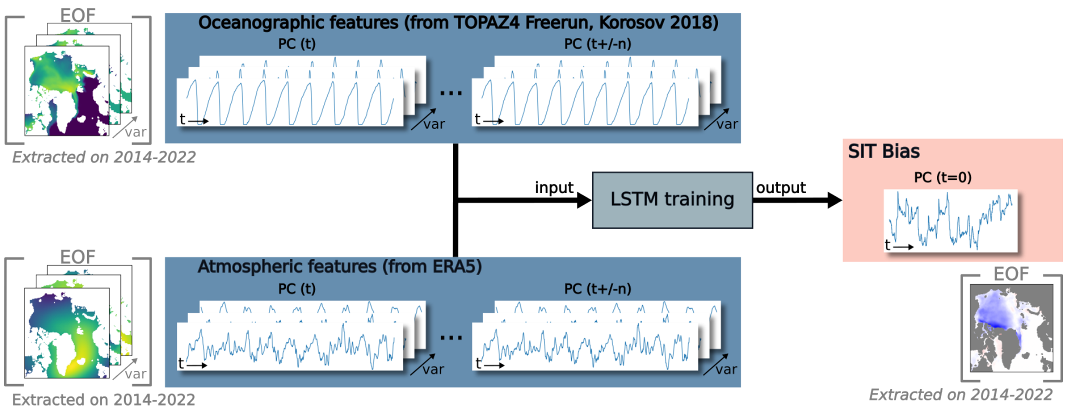

Figure 2. Illustration of LSTM prediction for one component of the EOF decomposition. The PCs of oceanographic and atmospheric features are used as inputs (blue boxes) to predict one PC of the SIT bias (red box), while the EOFs are not used as predictors (in brackets). Multiple variables var are used as input features at different times t and t plus/minus time lag n (because the LSTM can use input features backward or forward in time).

4.1 Features of the SIT bias in TOPAZ4 on 2011-2022

Between 2011 and 2022, the mean Arctic sea ice thickness (SIT) within the ice edge (sea ice concentration (SIC) above 15%) ranges between 0.6 and 2 meters (Fig. 4 upper panel) for the 2 versions of TOPAZ4 used in this study. Both SIT simulations show a yearly cycle that is consistent with the observations at hand. When assimilating CS2SMOS, TOPAZ4-RA SIT gets closer to the observations (Fig. 4a) and the spatial distribution improves drastically. The bias (Fig. 4 lower panel) varies from year to year, and shows extreme peaks (mostly negative) often at the end of summer, as SIT errors accumulate in the absence of SIT data for assimilation. The three most recent years (2020-2022) show lower SIT bias compared to earlier years, both against TOPAZ4-RA and in CS2SMOS datasets. The recent decline of SIT is less pronounced in the free running model, where the ice is already thin.

A systematic bias of SIT can be noted all year round (Fig. 3a): TOPAZ4-FR shows too thin ice in all areas of thick ice: the central Arctic, close to the north of Greenland and the Canadian Archipelago, while it depicts too thick sea ice in the Beaufort Gyre. The amplitude of this error varies slightly according to seasons, yet remains observable at all times. The underestimation of thick ice is widespread among other models (Johnson et al., 2012; Uotila et al., 2019) and can be explained by too strong ice drift along the north of Greenland, advecting the multiyear ice westwards into the Beaufort Gyre, whereas observations show a dense and stable area of multiyear ice to the north of Greenland. The complex geography of the Arctic region, notably in the Beaufort Gyre, is prone to sea ice entrapment either because ocean currents or winds are inaccurate or due to deficiencies of the sea ice rheology.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

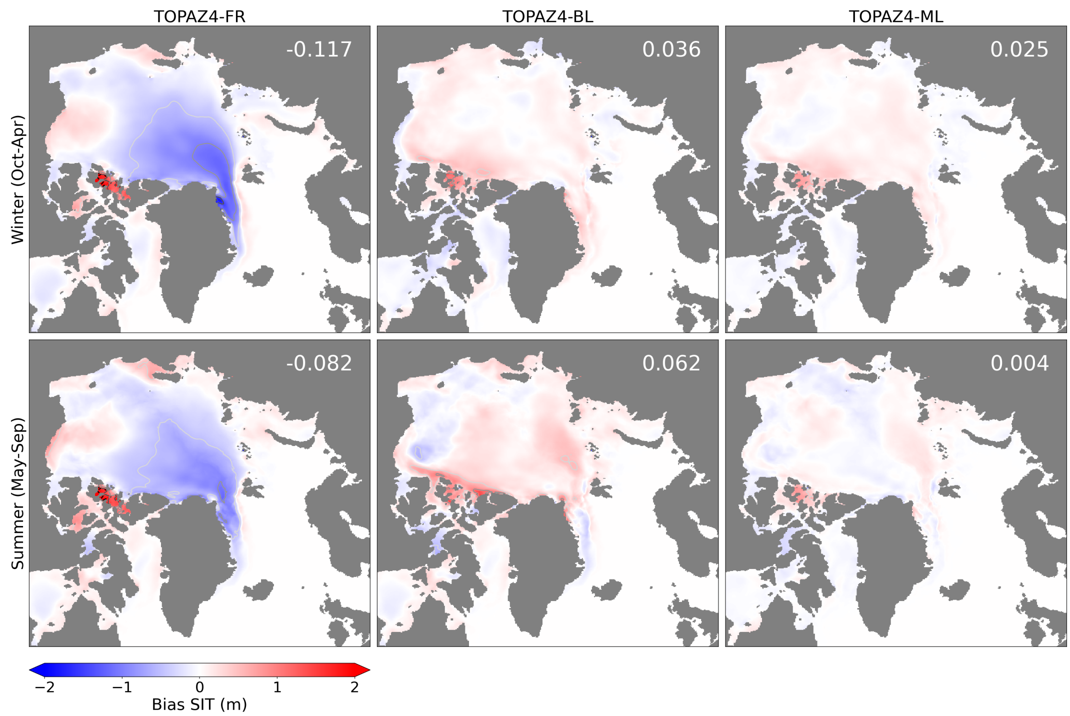

Figure 3. Seasonal bias of SIT (m) averaged over the test period (2011-2013), between **left column)** TOPAZ4-FR and TOPAZ4-RA, and **middle column)** TOPAZ-BL and TOPAZ4-RA, and **right column)** TOPAZ4-ML and TOPAZ4-RA. The blue colour indicates that the TOPAZ4 reanalysis SIT is thicker. The freezing period (**top row**) extends from October to April, while the melting season (**bottom row**) spans from May to September.

4.2 Evaluation of the ML performance on 2011-2013

After training the algorithm, we apply it to the period 2011-2022 and evaluate its performance on the test dataset from 2011 to 2013, which was excluded from the calculation of the EOFs and therefore from the training. We chose a contiguous period for the test dataset to avoid as much as possible the dependencies caused by the temporal autocorrelation in the SIT data. Hereupon the SIT predicted by our algorithm will be called TOPAZ4-ML for the sake of brevity. Our models predict the PC for each EOF (further analyzed in Appendix A), which are then converted to SIT following the methodology presented in section 3. Within this section, we will exclusively focus our evaluation on sea ice thickness.

As dimensionality reduction leads to an ineluctable loss of information due to truncation, the EOF decomposition introduced an inherent error into our SIT retrieval. The EOF error (Fig. 5b) represents a lower bound that even optimal ML performance cannot mitigate. The highest RMSE values (0.5 m) are obtained in the marginal seas, particularly in the East Greenland Sea, Beaufort Gyre, and Laptev Sea regions. Conversely, the error obtained by the baseline (Fig. 5c) is considered as our upper bound, being a trivial bias correction. In contrast with the lower bound, the RMSE values are much higher (up to 1.5 m) from the Fram Strait to the whole Central Arctic as well as the Canadian Archipelago, areas where the free run is most biased.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

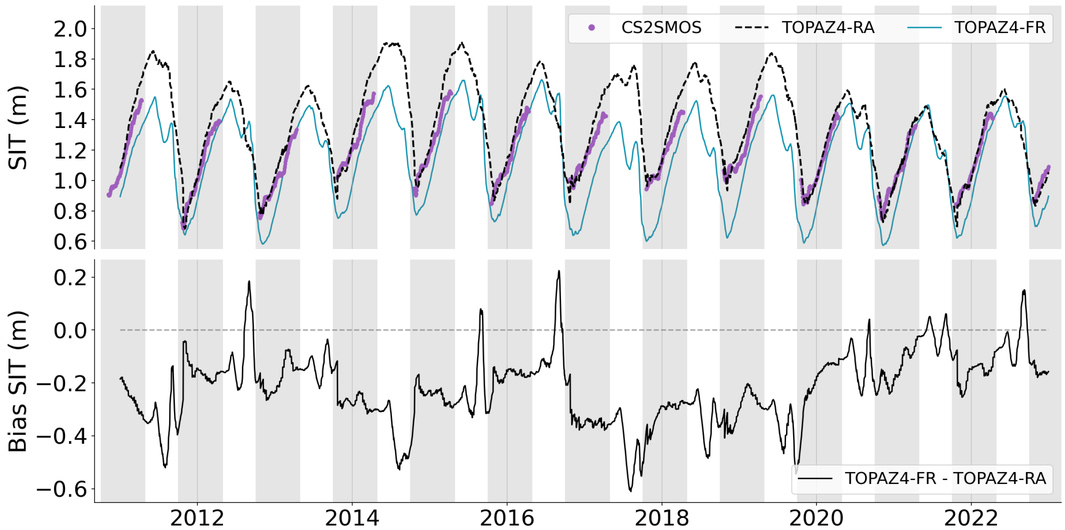

Figure 4. Top) Daily sea ice thickness (m) averaged over the Arctic within the ice edge. TOPAZ4-RA, TOPAZ4-FR, and the CS2SMOS merged-satellites product are displayed. **Bottom)** Bias of sea ice thickness (m) computed as follows: TOPAZ4-FR - TOPAZ4-RA.The freezing periods from October to April are highlighted with a grey background.

The baseline RMSEs are however small in the marginal seas where sea ice is thin. The ML-adjusted error reveals patterns more similar to the EOF error (Fig. 5a). This can be interpreted as the residual error being predominantly influenced by the truncation of the EOF rather than the ML error. The ML-adjusted RMSE increases by 0.2 m compared to the EOF truncation RMSE, mostly visible in the Central Arctic, as well as the Beaufort Gyre. On average, the mean RMSEs of 0.24 m (ML-adjusted), 0.21 m (EOF), and 0.31 m (baseline) attest that the ML algorithm is outperforming the baseline, with a performance close to the optimal EOF capability. Despite the large RMSE observed in the Central Arctic, the baseline manages to provide an acceptable correction on average.

Similar behaviors have been noted for other error indicators, such as the bias and the correlation (not shown). This demonstrates that our methodology can reconstruct the SIT with a relatively small error induced by the ML algorithm itself and that the correction goes beyond a trivial monthly bias adjustment.

Over the test period, TOPAZ4-ML SIT is in strong concordance with TOPAZ4-RA SIT (Fig. 6), while still showing noticeable differences, specifically during the melting period of 2011 and the end of the growth period of 2013. The temporal evolution of the mean SIT for all methods, including TOPAZ4-RA used as our reference, is shown for the entire training period in Fig. 6. These time series show the artefacts related to the experimental setup throughout the summer, mostly due to the lack of sea ice thickness assimilation. As anticipated, the ML algorithm closely aligns with TOPAZ4-RA during the training period, although the degree of agreement varies from year to year, supporting the assumption that the test period is largely independent of the training period. The baseline presents more substantial differences, mostly during the melting period as well as in the later years of thinner ice. In particular, a secondary peak of SIT stands out at the beginning of each melting period, at times

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

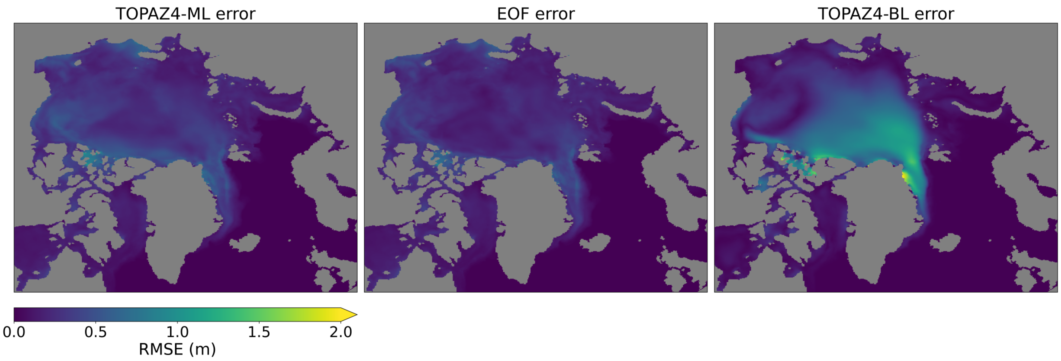

Figure 5. RMSE of SIT bias (m) over the test period of a) ML-adjusted error, b) EOF error, c) baseline error.

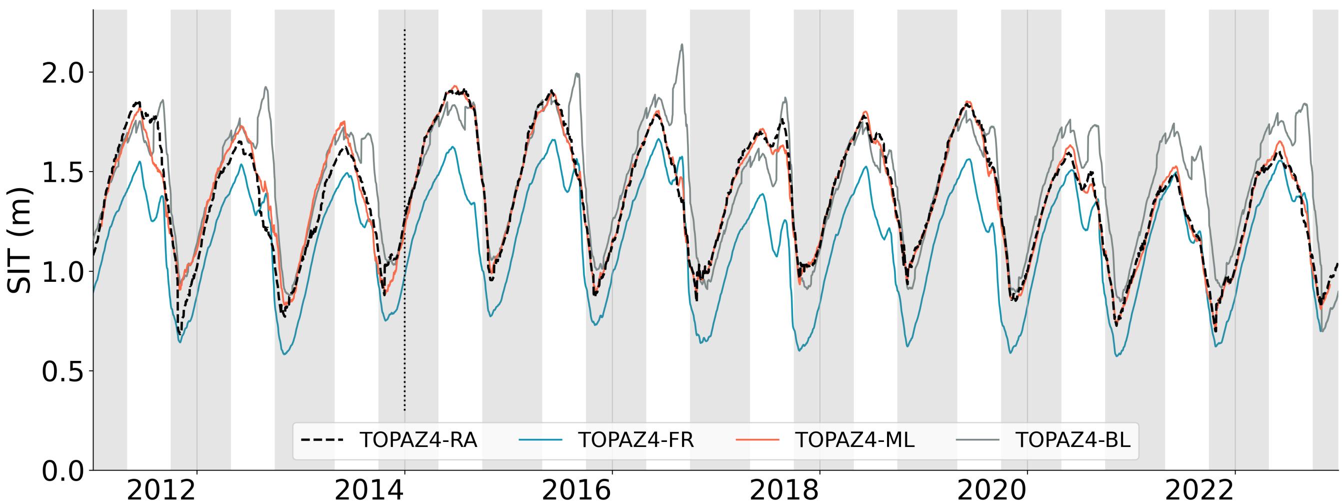

Figure 6. Daily SIT (m) averaged over the Arctic for SIC>15%. SIT for TOPAZ4-RA (considered as our truth), TOPAZ4-FR, TOPAZ4-ML, and TOPAZ4-BL are shown. A vertical line in 2014 separates the test (2011-2013) from the training sets (2014-2022). The freezing periods from October to April are highlighted with a grey background.

higher than the main peak. This eye-catching feature is also observed simultaneously in the TOPAZ4-FR, albeit to a lesser extent, as a statistical artefact of computing the average thickness: the thin ice melts first and the surviving thick ice causes the average to increase where the ice is still present. It will be further addressed in section 5. The baseline however agrees robustly with TOPAZ4-RA during the growth season. This indicates that the spatially averaged SIT bias repeats identically every year during the freezing season and could be improved by tuning a model parameter like the thickness of new ice (Wang et al., 2010) or more preferably by upgrading to a more advanced thermodynamical model.

The application of the ML algorithm results in a drastic bias reduction, outperforming the baseline. Over the test period, the mean bias between TOPAZ4-FR and TOPAZ4-RA is -10.0 cm. The year-round bias reduces to 1.4 cm after ML adjustment,

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

with a seasonal modulation of 2.5 cm (October-April), and 0.4 cm (May-September) (Fig. 3 right column). Regarding the baseline, the averaged remaining bias is 4.9 cm and the seasonal values are the following: 3.6 cm (October-April), and 6.2 cm (May-September) (Fig. 3 middle column). Although the baseline constitutes a clear improvement during the test period, particularly during the winter season, the errors remain large in some areas, see Fig. 5.

4.3 Application of the ML adjustment on 1992-2010

Since the ML algorithm performed well during the test period, it is further extrapolated to predict SIT biases before both the CryoSat-2 and SMOS missions were launched in 2011. As suggested by Lam et al. (2023), a greater performance is anticipated when training on the whole dataset, so in this section, we retrained the ML algorithm taking into account all years starting in 2011, without adjusting any parameters. Undeniably, three additional cycles of growth and melt are valuable information, especially considering that our full dataset only spans 12 years.

4.3.1 Validation with independent datasets

Our first step is to assess the performance of our prediction against *in situ* datasets during the first decade of prediction (2000-2010). For this task, the most representative year-round data is from moorings rather than floe-tethered Lagrangian observations (i.e., buoys measuring the thickness of a specific sea ice floe to which they are attached).

TOPAZ4-ML SIT overall improves slightly the agreement with *in situ* data, while both TOPAZ-BL and TOPAZ-FR display similar inaccuracy, albeit more severe in TOPAZ-FR. A representative case is the mooring A from BGEP (Fig. 8), situated within the Beaufort Gyre, shows a clear enhancement over the summer season when contrasting the TOPAZ4-FR with TOPAZ4-ML SIT. Baseline SIT always underestimates SIT at the onset of the melting season, potentially specific to the Beaufort Gyre region as it is the only area where the free run systematically overestimates SIT. The SIT data from the buoy exhibit considerable variability, particularly towards the end of 2006 and the transition between 2007 and 2008. This variability might be attributed to the specific climatic conditions during those years, notably 2007, which marked a record-setting ice retreat characterized by the flushing of old and thick sea ice and we do not expect a coarse resolution model like TOPAZ4 to render this level of variability.

The mooring is occasionally in open water while the free run still has ice covering it. Since both the baseline and the ML algorithms are not trained to reduce ice edge discrepancies, their performance is poor during these periods. On the positive side, the time series does not indicate that the adjustment methods degrade further back in time so the extrapolation is yielding reasonable values.

TOPAZ4-ML SIT exhibits similar performance to the baseline when considering all BGEP in-situ datasets, as summarized in table 3. An improved performance is noted on the NPEO mooring, located at the North Pole, closer to the perennial sea ice. However, the improvement of the ML compared to the baseline is less striking than in the test period, mostly because assessing one specific location over a brief time period may not provide sufficient representativity to distinguish between these two adjustment methods. All the scores are poorer than those obtained with full spatial coverage in the test period but the adjustments are never worse than the free run.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

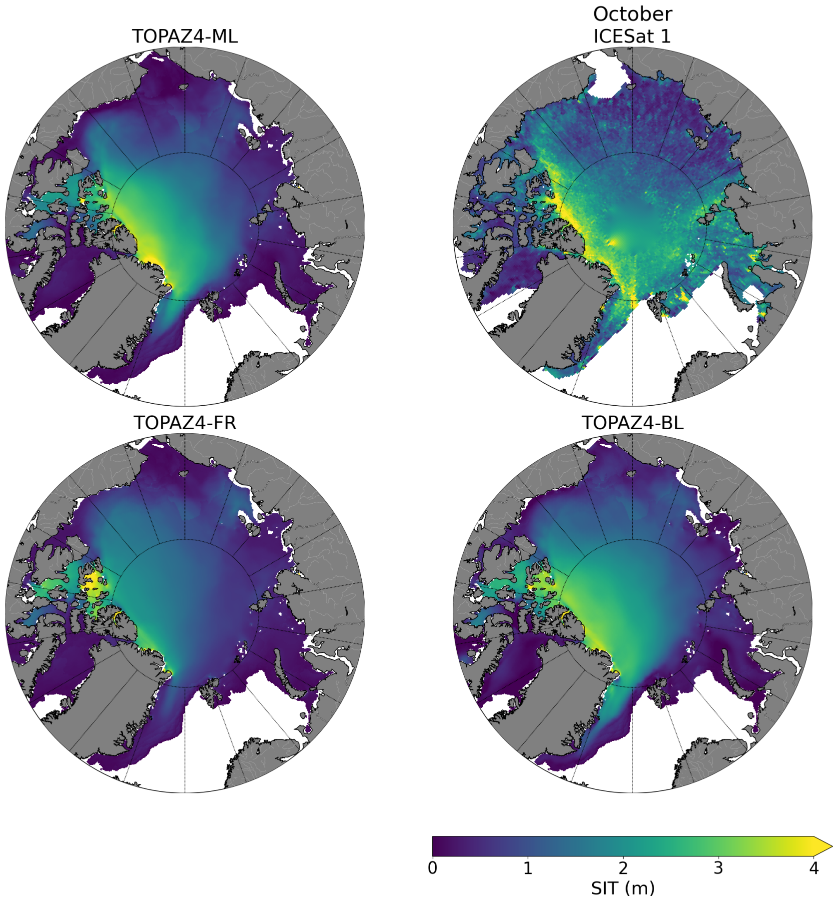

Figure 7. Sea ice thickness (meters) for TOPAZ4-ML, ICESat-1, Envisat, TOPAZ4-FR, and TOPAZ4-BL averaged over October 2003-2007. ICESat-1 observation period varies, extending into November depending on the year.

A qualitative comparison between remote sensing data and TOPAZ4-ML confirms the improvement of SIT agreement compared to TOPAZ4-FR (Fig. 7). Firstly, monthly values averaged over October between 2003 and 2007 are considered for TOPAZ4 and Envisat, while ICESat-1 campaigns do not precisely align with calendar months. TOPAZ4-ML enhances the SIT gradient from Greenland to the North Pole, addressing the well-known issue of a flattened gradient of sea ice thickness as one moves away from the northern coast of Greenland. Moreover, TOPAZ4-ML SIT is effectively reduced along the Siberian coast in agreement with satellite observations, contrasting with the baseline's inability to do so. TOPAZ4-ML and remote sensing show similar patterns within the Beaufort Gyre and Canadian Archipelagos, whereas TOPAZ4-BL displays comparable correction but with insufficient intensity. Considering Envisat SIT, we observe significantly less young and thin sea ice around the periphery of the Central Arctic when compared to other datasets. As a consequence, Envisat shows high SIT (> 2m) in March (Supplements Fig. S6) near the sea ice edge in the Barents Sea, a scenario considered unrealistic and consistent with past reports of Envisat's tendency to overestimate SIT compared to other datasets.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

Table 3. Sea Ice Thickness bias in meters, RMSE and Pearson correlation coefficient (R) between SIT from TOPAZ4-ML (ML), TOPAZ4-BL (BL) and TOPAZ4-FR (FR) and in situ datasets. The highest score is highlighted in bold.

	Buoy	Freezing			Melting			All time		
		ML	BL	FR	ML	BL	FR	ML	BL	FR
Bias	BGEP A	-0.33	**-0.13**	0.19	**0.03**	-0.19	0.23	**-0.15**	-0.16	0.21
	BGEP B	-0.29	**-0.03**	0.10	-0.13	-0.09	**0.05**	-0.21	**-0.06**	0.07
	BGEP C	-0.54	-0.19	**-0.11**	-0.29	-0.19	**-0.16**	-0.42	-0.19	**-0.14**
	BGEP D	-0.33	-0.12	**0.07**	0.31	0.04	**0.28**	**-0.01**	-0.04	0.17
	NPEO	**-0.78**	-0.84	-1.77	**-0.93**	-0.99	-1.72	**-0.85**	-0.92	-1.74
RMSE	BGEP A	0.67	**0.50**	0.55	0.53	**0.49**	0.60	0.60	**0.49**	0.58
	BGEP B	0.53	**0.34**	0.39	0.55	**0.50**	0.56	0.54	**0.42**	0.48
	BGEP C	0.75	**0.44**	0.45	0.70	**0.55**	0.58	0.73	**0.49**	0.51
	BGEP D	0.73	**0.58**	0.60	0.91	**0.72**	0.93	0.82	**0.65**	0.76
	NPEO	**1.11**	1.18	1.95	**1.25**	1.29	1.91	**1.18**	1.24	1.93
R	BGEP A	0.53	**0.71**	0.67	0.80	**0.85**	0.76	0.67	**0.78**	0.72
	BGEP B	0.80	**0.89**	0.88	0.88	0.90	**0.92**	0.84	**0.90**	**0.90**
	BGEP C	0.63	**0.83**	0.82	0.60	0.82	**0.83**	0.61	0.82	**0.83**
	BGEP D	-0.01	**0.42**	0.28	0.35	**0.59**	0.20	0.17	**0.51**	0.24
	NPEO	**0.77**	0.75	0.75	0.63	**0.64**	0.63	**0.70**	**0.70**	0.69

4.3.2 Interpretation of the reconstructed data

The previous section has validated the reconstructed SIT, we will evaluate here the consistency of TOPAZ4-ML SIT over the whole Arctic, for which there are no observations available. This section will quantify various trends and changes identified with this new dataset over time.

The May (October) mean sea ice thickness in 1992 is estimated at 2.16 meters (1.08 meters), while in 2022, it has shrunk down to 1.54 meters (0.57 meters). In total volume, this corresponds respectively to 26,605 km^3 (12,575 km^3) and 18,804 km^3 (6,258 km^3), with linear trends of -3,274 km^3/decade (-3,002 km^3/decade). The year-round trend is -3,153 km^3/decade according to our reconstruction, which is compatible with the estimate from the PIOMAS model reconstruction (Schweiger et al., 2014). We observe a significant downward trend in the mean SIT from 2002 to 2012, surrounded by two periods without distinct trends (Fig. 9b). Our ML-adjusted SIT respects this behaviour qualitatively and does not introduce unrealistic trends by extrapolation.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

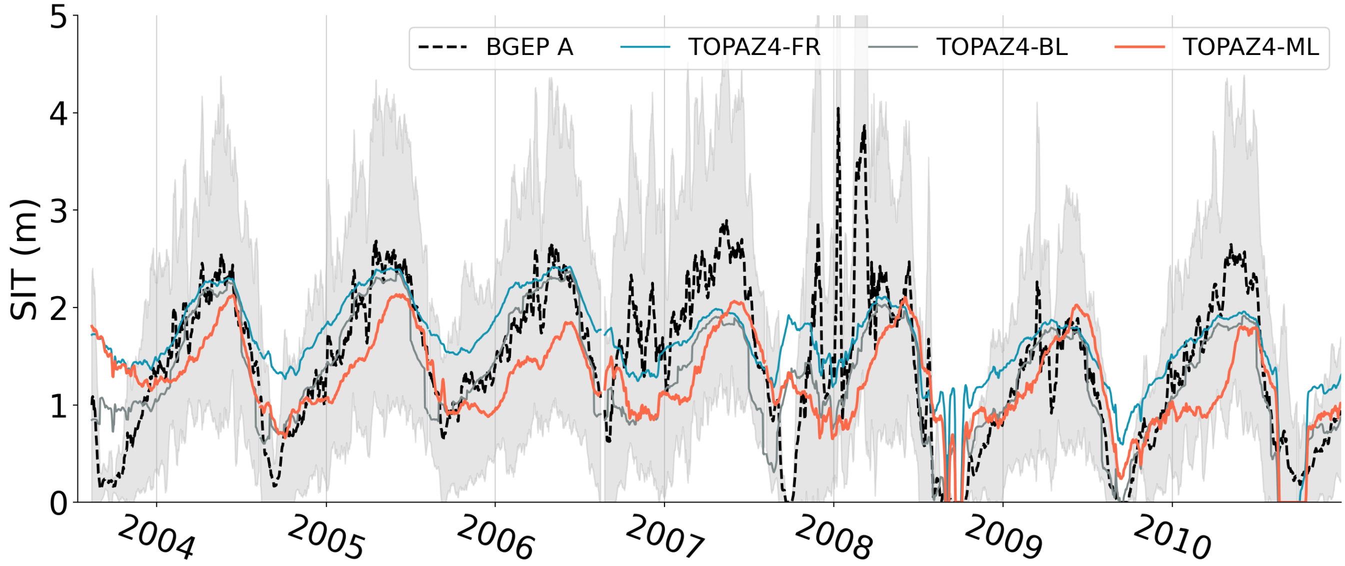

Figure 8. Daily SIT (m) for buoy BGEP A, TOPAZ freerun, baseline, and ML-adjusted. The standard deviation of SIT for ULS BGEP A is displayed in grey.

Sumata et al. (2023) show how the distribution of SIT exiting the Arctic through the Fram Strait changed throughout the past two decades as observed from moored upward-looking sonars. They reveal a bimodal distribution and a regime shift following the sea ice minimum of Summer 2007. Since the Transpolar Drift brings sea ice from large stretches of the Arctic into the Fram Strait, the representativeness of these moorings is higher than in most other locations. Some delay should however be expected due to the advection time to the Fram Strait, which can take from months to a couple of years depending on the origin of the ice.

The yearly cycles of the main modes of SIT look generally continuous in TOPAZ4-ML (see Fig. 9a), except for a few occasional discontinuities. So the combination of DA and ML did not seem to cause much distortions of the physical signals. TOPAZ4-ML SIT distribution of the whole Arctic also exhibits a more gradual transition from a bimodal distribution (before 2007) during the growth period to an unimodal distribution (after 2007) as depicted in Fig. 9a. Prior to the 2007 minimum, a significant portion of the ice is thicker than 2 meters. However, after 2008, only thinner sea ice is observable year-round. At the end of the melting period in the years before 2007, when most of the first year ice has melted, the median sea ice thickness falls within the 1 to 2-meter range. In contrast, after 2007, the median sea ice thickness is almost consistently below 1 meter. Moreover, the distribution of the thickest sea ice (depicted in green on Fig. 9a) is notably diminished when comparing the periods before 2007 (4-5 meters) and after 2007 (3-4 meters). The area-average SIT is broadly similar between TOPAZ4-RA and TOPAZ4-ML, all lying consistently about 20 cm above TOPAZ4-FR throughout the whole time series. Contrary to sea ice extent time series, the record minima of SIT are somewhat less spectacular, indicating that significant ridging may occur during years of lowest ice cover (Regan et al., 2022), piling up sea ice vertically rather than horizontally. The years 2011 and 2012 are

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

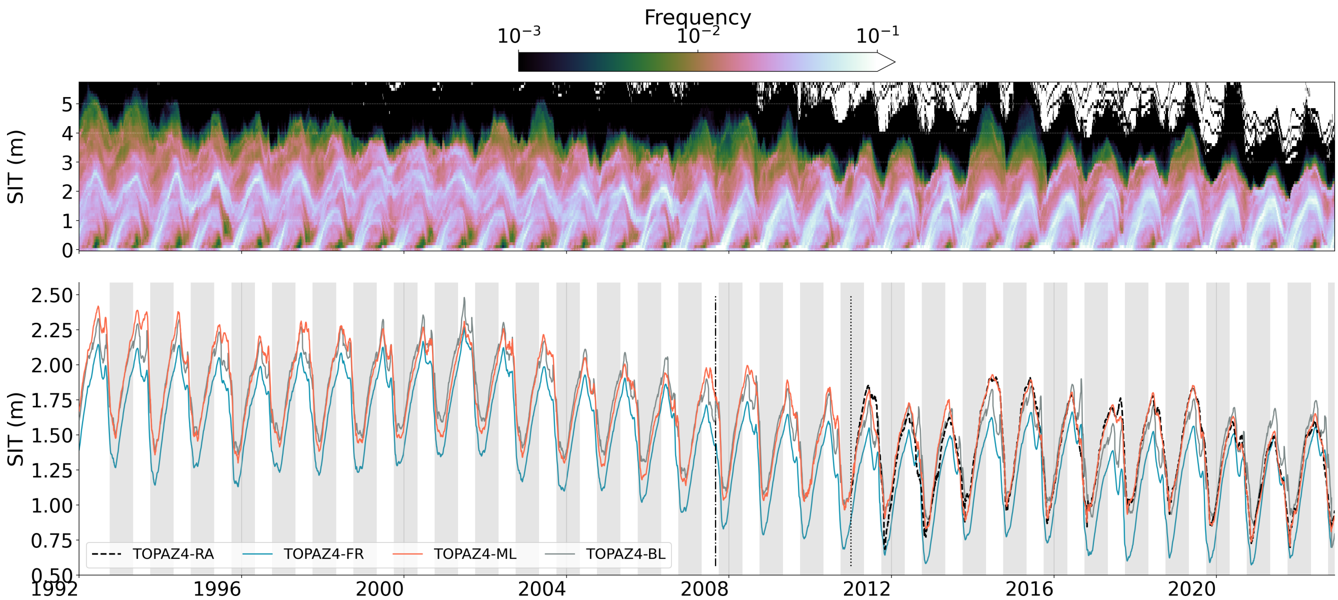

Figure 9. Top) Distribution of daily SIT (m) from 1992 to 2022. Bins of 0.1m are used and the color bar is a log scale. **Bottom)** Daily SIT (m) averaged over the Arctic for SIC>15% for the same period. The ML algorithm has been retrained including 2011-2013, as indicated by the vertical line in 2011. The dot-dashed vertical black line marks September 2007. The freezing periods from October to April are highlighted with a grey background.

335 clear minima of the SIT in all datasets, in agreement with the PIOMAS model. The disagreements between the free run and other datasets are more important in the later years, as the free run indicates minimum years between 2017 and 2021, while TOPAZ4-RA and TOPAZ4-ML datasets rather point to 2021 and 2022 as mimimum SIT years. Surprisingly, the summer 2007 does not stick out in the area-averaged SIT time series as the regime shift seems to spread over a few years. In the Discussion section, we will compare various trends reported in the literature.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

5 Discussion

The novelty of the present study lies in the combination of ML and DA to adjust sea ice thickness backward in time over a long period, longer than the training period. Since 1990, the sea ice thickness distribution in the Arctic has shifted drastically towards thinner sea ice (Sumata et al., 2023; Lindsay and Schweiger, 2015) as documented by both satellite and *in situ* data. With our adjusted dataset (TOPAZ4-ML), the mean sea ice thickness in May (October) 1991 is 2.16 meters (1.08 m), while in 2022, it is 1.54 m (0.57 m), resulting in a decrease of 29% (47%). Using independent data in the Arctic Basin, Lindsay and Schweiger (2015) found that the annual mean SIT over the period 2000–2012 has declined from 2.12 to 1.41 m (34%), while September thickness has declined from 1.41 to 0.71 m (50%). When including all the marginal seas until the 15% isoline of concentration, we find that the annual SIT is generally lower but the trends are compatible, reducing from 1.51 m in 2000 to 1.01 m in 2012 (33%), while September thickness declined from 1.42 m to 0.81 m (43%). In our estimation, the annual mean sea ice thickness is lower compared to Lindsay and Schweiger (2015), primarily due to differences in the area considered. These disparities diminish in September as the residual sea ice shrinks toward the Central Arctic. Kwok (2018) reported losses of 2870 km^3/decade in winter (Feb-March) for 15 years of satellite record (2004-2018) from the non-overlapping ICESat and CryoSat-2 periods. For the same period, the TOPAZ4-ML data indicates losses of 2941 km^3/decade (Fig. B1), which falls well within the uncertainties caused by the lack of snow depths data (Zygmuntowska et al., 2014). Compared to another data assimilative model, PIOMAS, the sea ice volume trends between 1979 and 2018 is -2700 (April) and -3200 (September) km^3/decade (Johannessen et al., 2020)[Fig. 5.24.]. TOPAZ4-ML data indicates -3120 (April) and -2960 (September) km^3/decade on a shorter period, which is not significantly different. Drawing from the range of available data, the ML-adjusted trends correspond closely with those documented in the existing literature. Although TOPAZ4-FR and TOPAZ4-ML differ significantly in the total amounts of SIT, their respective trends are close.

By training our algorithm over the latest decade to predict the past, we assumed the following: the EOFs obtained from the SIT bias between 2011-2022 are representative of the statistical behavior of the errors made by the model over a longer time period, including a dramatic regime shift. To probe the robustness of this assumption, we extracted the EOFs over two subperiods of our dataset: the training period with and without the test period. We only found differences in the least important components (from the 6^{th} and further) while showing similar patterns overall (Supplements Fig. S2). The time series of the differences only shows unstructured noise.

Moreover, since we lack summer SIT observations, we assess the differences in SIT between two versions of TOPAZ (with and without assimilation) and not the SIT directly, so the data assimilation residuals may also have caused some loss of signal for the ML. However, the ML algorithm can adjust the thickness even of the thickest sea ice (>6 meters) with less than 20% of error (Fig. C1), which explains its performance in an earlier period dominated by multi-year ice.

Our approach based on EOF decomposition enables a drastic reduction of dimensions, leading to fewer parameters in the ML algorithm, thus reducing the costs required to train and apply the algorithm. This method is fast to implement and execute (around 1 hour on a personal laptop), requiring minimal computational resources. Given its effectiveness, it demonstrates a strong ability to correct a large share of the biases. In comparison, approaches relying on more intricate 2D neural network

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

layers produced comparable outcomes but at a higher cost (at least 12 hours to train), and in a more complex setup. Additionally, it is possible that with higher dimensional features, the training set would be too small, increasing the risk of overfitting.

Multiple ML models (LSTM, Convolutional Neural Network, Dense, eXtreme Gradient Boosting, and Random Forest) were tested, yielding small local variations, but without visible advantages in overall performances between models. The decision to select the LSTM was thus driven by its robust time series prediction capabilities and its slightly better results. Throughout this study, the ML architecture (i.e. number of layers and hyperparameters) only played a minor role in achieving the optimal prediction; instead, the prediction accuracy is considerably dependent on the input variables, i.e. the choice of variables and associated time lags.

Three distinct sources of errors are identified when predicting SIT before 2011: ML reconstruction error, errors in the free run of TOPAZ4, and errors induced by regime shifts in sea ice conditions. Since the two latter are out of reach, we can only provide uncertainty estimates related to the ML method itself. Note that the uncertainty obtained here only characterizes the sensitivity of the algorithm to its inputs (details in section 3). The areas exhibiting the highest uncertainty encompass the Fram Strait, the Canadian Archipelagos, and the Beaufort Gyre, and with a lower degree of uncertainty the East Siberian Sea (Supplements Fig. S7). Upon examining the temporal evolution of uncertainty (Supplements Fig. S8), it appears that uncertainty diminishes during both the growth and melt phases of sea ice, likely attributable to the strong sea ice thickness seasonal cycle. Moreover, higher uncertainty is noticed during the peak of winter and summer seasons, when sea ice thickness is less affected by predominant freezing or melting, potentially leading to divergence among individual members.

Despite the baseline yielding good average results, the trivial bias correction displays strong regional biases and mediocre performance during outlier years. In addition, we expect the performance of the baseline to decrease even further as we extrapolate back in time. Indeed, the correction of the baseline is applied once and solely relies on the patterns of mean bias observed during 2014-2022 with no ability to accommodate different environmental conditions. On the contrary, our ML adjustment method proves more adaptable when predicting back in time since it takes into account the past state of environmental variables and the variable relative importance of each component (as independent errors identified by the EOF decomposition).

A distinct feature appears in the SIT averaged (Fig. 4) at the onset of the melting season: a second peak, brief compared to the first, occurs shortly after the SIT maximum. It is observed almost yearly in TOPAZ4-FR while only twice (2017 and 2020) in TOPAZ4-RA. The phenomenon can be explained as follows: the relatively thin sea ice melts first, decreasing the area faster than the thickness, thus increasing the average SIT as a case of "survivor bias". This survivor bias may intensify in cases of erroneously thinner sea ice in the Central Arctic. Such instances can arise from either thinner sea ice in the Central Arctic (TOPAZ4-BL) or misplaced thick sea ice in the Beaufort Gyre (TOPAZ4-FR). To prevent this artefact, many studies prefer to use the total volume or a geographic restriction to an area of perennial ice in the Central Arctic.

Comparing TOPAZ4 to *in situ* datasets is challenging, primarily due to representation errors. Knowing the "true" sea ice thickness remains a major issue for evaluation, particularly when considering historical data from older satellite missions such as ICESat-1 and Envisat. This issue becomes more pronounced as we delve further into the past. Large uncertainties linked to *in situ* observations and model ultimately lead to differences in SIT and difficulties in evaluating our product. Adding to this

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

point, the limited availability of global datasets over extended periods in the Arctic restricts the scope of possible comparisons. One notable advantage of our methodology is its capacity to bridge data gaps when mooring observations are unavailable.

6 Conclusions

In this investigation, we demonstrated that machine learning (ML) can be combined with data assimilation (DA) to predict sea ice thickness (SIT) errors backwards in time to 1992, using the ice-ocean model TOPAZ4 and atmospheric variables from the reanalysis ERA5. The SIT biases are the results of accumulated increments from the assimilation of sea ice thickness data from CS2SMOS every 7 days between 2011 and 2022 during the ice growth period (October-May). Then, we reduced the dimensionality of the DA increments using Empirical Orthogonal Functions (EOF). The LSTM learned to predict SIT biases using as inputs Principal Components (PC) of various sea ice, ocean, and atmospheric variables. This study demonstrates that our PC-based approach is effective in providing a major sea ice thickness adjustment.

Our approach significantly reduced sea ice thickness biases throughout the test period (2011-2013) from a low bias of -10.0 to 1.4 centimeters. Significant improvements are noted during the melting period, likely attributable to substantial errors in TOPAZ4 with assimilation, as sea ice thickness data assimilation is unavailable during summer. After applying our algorithm before 2011, the evaluation with independent mooring data shows an overall improvement compared to the free run of TOPAZ4. However, the scarcity of in-situ datasets and the often limited continuity of observations restrict the comparison to only a few locations. Remote sensing data from Envisat and ICESat-1 were primarily utilized for qualitative assessment due to their inherent high uncertainties and temporal-spatial discontinuities. Our approach demonstrates a general improvement in SIT despite the challenge of selecting a reliable "truth" for validation.

Furthermore, this prolonged time series brings new insights into various aspects of SIT, including distribution, spatial patterns, and changes through time. The estimated May (October) mean sea ice thickness in 1992 was 2.16 meters (1.08 meters), whereas it was 1.54 meters (0.57 meters) in 2022, resulting in a 29% (47%) decline. This amounts to a decrease in total sea ice volume from 19,690 to 12,700 km^3, with a corresponding trend of -3,153 km^3/decade, corroborating previous model estimates. A decrease of the thickest sea ice is observed throughout the years, with the proportion of sea ice thicker than 2.5 meters going from 28% in 1992 to 7% in 2022. In the ML-adjusted data, the transition in 2007 is however less abrupt than deduced from moored observations from the Fram Strait.

The ML-adjustment technique can be implemented for other variables, as long as equivalent resources are available: two model runs with and without assimilation of the target variable and some auxiliary data related to the target variable but in complex ways. Further work is required to compare our SIT time series with the novel year-round processing of CS2 (Landy et al., 2022), especially regarding the summer sea ice thickness. The ML-adjustment method was originally introduced in the framework of an iterative method combining DA and ML techniques (Brajard et al., 2020). In a subsequent investigation, a second iteration of DA using the reconstructed SIT and its uncertainty will be performed with TOPAZ4, improving the initial conditions of SIT of the latter decade.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

Code availability. The code is available at https://github.com/LeoEdel/tardis-ml-paper1.

Data availability. Our ML-adjusted SIT dataset (TOPAZ4-ML) is available at https://zenodo.org/records/11191854 (doi: 10.5281/zenodo.11191853) and can be visualized at https://av.tib.eu/media/68161 (doi: https://doi.org/10.5446/68161). The following datasets are used as inputs or for evaluation. ERA5 data are available at https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5. TOPAZ4b reanalysis data are available at https://doi.org/10.48670/moi-00007. SID used in TOPAZ4b reanalysis is available at 10.15770/EUM_SAF_OSI_NRT_2007. CS2SMOS data are available at ftp://ftp.awi.de/sea_ice/product/cryosat2_smos/v204/. ICESat-1 data are available at https://nsidc.org/data/nsidc-0393/versions/1. Envisat data are available at https://catalogue.ceda.ac.uk/uuid/f4c34f4f0f1d4d0da06d771f6972f180. ULS BGEP data are available at https://www2.whoi.edu/site/beaufortgyre/data/mooring-data/. ULS NPEO data are available at https://arcticdata.io/catalog/view/doi:10.5065/D6P84921.

Appendix A: PC prediction

For a deeper understanding of our method, the original values predicted by our algorithm are displayed (Fig. A1) and commented on in this section. The corresponding EOFs are plotted in Supplements Fig. S1. The quality of the final sea ice thickness reconstruction relies on the accuracy of predicting each component. A large error in one PC may be observed in the resulting SIT. PCs showing a yearly cycle (such as #1 and #2) show better predictability than the more irregular PCs (#4 and #7). The prediction of the ML shows a slight smoothing of the signal. It is beneficial in the sense that the ML is not trying to update SIT every week like DA does, thus avoiding a noisy reconstruction. We notice some difficulties in the prediction of the test period: major differences (#7), and light offsets (#1, 4, 8). While PCs #2, #3, #5, #6 show a more consistent and reliable prediction.

Appendix B: Sea Ice Volume

Sea ice volumes (Fig. B1) are obtained by multiplying the sea ice thickness by the area in each grid cell and by the sea ice concentration. It is then summed over the whole model domain. It is insensitive to high SIT values in areas of low ice concentrations and therefore a more convenient quantity than the average SIT to compare between models, although it is not as easily obtained from observations. For a clearer view of the decadal difference in sea ice thickness, refer to Supplements Fig. S4.

Appendix C: Capability of the adjustment method as function of sea ice thickness

To evaluate our method performance across various sea ice thicknesses, we analyze the bias obtained from our method with the true bias as a function of the SIT (Fig. C1). Over the test period, our ML algorithm overestimates the adjustment (SIT bias difference is positive) for sea ice thickness between 3 and 5 m and underestimates the adjustment (SIT bias difference is negative) for thickness above 6m.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

Author contributions. L.E. created the database, developed the machine learning algorithm, and wrote this paper. J.X. initiated the original idea and produced the two versions of TOPAZ. A.K. provided the sea ice age product. J.B. and L.B. supervised the study and closely followed the writing of the final draft. All co-authors contributed to the discussion of the results and the improvement of the final draft.

Competing interests. The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgements. This research was supported by the project TARDIS funded by the Research Council of Norway grant number 325241. The model runs and data storage were supported by the Norwegian supercomputing infrastructure Sigma2 (project numbers nn2993k and NS2993K respectively).

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

References

Aksenov, Y., Popova, E. E., Yool, A., Nurser, A. G., Williams, T. D., Bertino, L., and Bergh, J.: On the future navigability of Arctic sea routes: High-resolution projections of the Arctic Ocean and sea ice, Marine Policy, 75, 300–317, 2017.

Arrigo, K. R.: Sea Ice Ecosystems, Annual Review of Marine Science, 6, 439–467, https://doi.org/10.1146/annurev-marine-010213-135103, 2014.

Bleck, R.: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates, Ocean modelling, 4, 55–88, 2002.

Bourke, R. H. and Paquette, R. G.: Estimating the thickness of sea ice, Journal of Geophysical Research: Oceans, 94, 919–923, 1989.

Brajard, J., Carrassi, A., Bocquet, M., and Bertino, L.: Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model, Journal of computational science, 44, 101 171, 2020.

Cheng, S., Quilodrán-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., Fablet, R., Lucor, D., Iooss, B., Brajard, J., et al.: Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review, IEEE/CAA Journal of Automatica Sinica, 10, 1361–1387, 2023.

Comiso, J. C., Parkinson, C. L., Gersten, R., and Stock, L.: Accelerated decline in the Arctic sea ice cover, Geophysical research letters, 35, 2008.

Drange, H. and Simonsen, K.: Formulation of air-sea fluxes in the ESOP2 version of MICOM, Nansen Environmental and Remote Sensing Center, 1996.

Driscoll, S., Carrassi, A., Brajard, J., Bertino, L., Bocquet, M., and Olason, E.: Parameter sensitivity analysis of a sea ice melt pond parametrisation and its emulation using neural networks, arXiv preprint arXiv:2304.05407, 2023.

Finn, T. S., Durand, C., Farchi, A., Bocquet, M., Chen, Y., Carrassi, A., and Dansereau, V.: Deep learning of subgrid-scale parametrisations for short-term forecasting of sea-ice dynamics with a Maxwell-Elasto-Brittle rheology, EGUsphere, pp. 1–39, 2023.

Frainer, A., Primicerio, R., Kortsch, S., Aune, M., Dolgov, A. V., Fossheim, M., and Aschan, M. M.: Climate-driven changes in functional biogeography of Arctic marine fish communities, Proceedings of the National Academy of Sciences, 114, 12 202–12 207, 2017.

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, e2023MS003 757, 2023.

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

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural computation, 9, 1735–1780, 1997.

Hunke, E. C. and Dukowicz, J. K.: An elastic–viscous–plastic model for sea ice dynamics, Journal of Physical Oceanography, 27, 1849–1867, 1997.

Johannessen, O. M., Bobylev, L. P., Shalina, E. V., and Sandven, S.: Sea Ice in the Arctic, Past, Present and Future, 2020.

Johnson, M., Proshutinsky, A., Aksenov, Y., Nguyen, A. T., Lindsay, R., Haas, C., Zhang, J., Diansky, N., Kwok, R., Maslowski, W., et al.: Evaluation of Arctic sea ice thickness simulated by Arctic Ocean Model Intercomparison Project models, Journal of Geophysical Research: Oceans, 117, 2012.

Kahru, M., Brotas, V., Manzano-Sarabia, M., and Mitchell, B.: Are phytoplankton blooms occurring earlier in the Arctic?, Global change biology, 17, 1733–1739, 2011.

Kaleschke, L., Tian-Kunze, X., Maaß, N., Mäkynen, M., and Drusch, M.: Sea ice thickness retrieval from SMOS brightness temperatures during the Arctic freeze-up period, Geophysical Research Letters, 39, 2012.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

Kim, J.-M., Sohn, B.-J., Lee, S.-M., Tonboe, R. T., Kang, E.-J., and Kim, H.-C.: Differences between ICESat and CryoSat-2 sea ice thicknesses over the Arctic: Consequences for analyzing the ice volume trend, Journal of Geophysical Research: Atmospheres, 125, e2020JD033 103, 2020.

Korosov, A. A., Rampal, P., Pedersen, L. T., Saldo, R., Ye, Y., Heygster, G., Lavergne, T., Aaboe, S., and Girard-Ardhuin, F.: A new tracking algorithm for sea ice age distribution estimation, The Cryosphere, 12, 2073–2085, 2018.

Krishfield, R. and Proshutinsky, A.: BGOS ULS data processing procedure, Woods Hole Oceanographic Institution, 2006.

Kwok, R.: Arctic sea ice thickness, volume, and multiyear ice coverage: losses and coupled variability (1958–2018), Environmental Research Letters, 13, 105 005, 2018.

Kwok, R. and Cunningham, G.: ICESat over Arctic sea ice: Estimation of snow depth and ice thickness, Journal of Geophysical Research: Oceans, 113, 2008.

Labe, Z., Magnusdottir, G., and Stern, H.: Variability of Arctic sea ice thickness using PIOMAS and the CESM large ensemble, Journal of Climate, 31, 3233–3247, 2018.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al.: Learning skillful medium-range global weather forecasting, Science, 382, 1416–1421, 2023.

Landy, J. C., Dawson, G. J., Tsamados, M., Bushuk, M., Stroeve, J. C., Howell, S. E., Krumpen, T., Babb, D. G., Komarov, A. S., Heorton, H. D., et al.: A year-round satellite sea-ice thickness record from CryoSat-2, Nature, 609, 517–522, 2022.

Lindsay, R. and Schweiger, A.: Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations, The Cryosphere, 9, 269–283, 2015.

Liu, G., Bracco, A., and Brajard, J.: Systematic Bias Correction in Ocean Mesoscale Forecasting Using Machine Learning, Journal of Advances in Modeling Earth Systems, 15, e2022MS003 426, https://doi.org/https://doi.org/10.1029/2022MS003426, 2023.

Liu, Y., Key, J. R., Wang, X., and Tschudi, M.: Multidecadal Arctic sea ice thickness and volume derived from ice age, The Cryosphere, 14, 1325–1345, 2020.

Regan, H. C., Rampal, P., Ólason, E., Boutin, G., and Korosov, A.: Modelling the evolution of Arctic multiyear sea ice over 2000–2018, The Cryosphere Discussions, 2022, 1–28, 2022.

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

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., et al.: Tackling climate change with machine learning, ACM Computing Surveys (CSUR), 55, 1–96, 2022.

Sakov, P. and Oke, P. R.: A deterministic formualtion of the ensemble Kalman filter: an alternative to ensemble square root filters, Tellus, 60, 361–371, 2008.

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

Sallila, H., Farrell, S. L., McCurry, J., and Rinne, E.: Assessment of contemporary satellite sea ice thickness products for Arctic sea ice, The Cryosphere, 13, 1187–1213, 2019.

Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., and DiMarzio, J. P.: Overview of the ICESat mission, Geophysical research letters, 32, 2005.

Schweiger, A., Zhang, J., Lindsay, R., Steele, M., and Stern, H.: PIOMAS Arctic Sea Ice Volume Reanalysis, https://psc.apl.uw.edu/research/projects/arctic-sea-ice-volume-anomaly/, 2014.

Solomon, A., Heuzé, C., Rabe, B., Bacon, S., Bertino, L., Heimbach, P., Inoue, J., Iovino, D., Mottram, R., Zhang, X., Aksenov, Y., McAdam, R., Nguyen, A. T., Raj, R. P., and Tang, H.: Freshwater in the Arctic Ocean 2010-2019, Ocean Science, 17, 1081–1102, https://doi.org/10.5194/os-17-1081-2021, 2021.

Sumata, H., de Steur, L., Divine, D. V., Granskog, M. A., and Gerland, S.: Regime shift in Arctic Ocean sea ice thickness, Nature, 615, 443–449, 2023.

Tilling, R., Ridout, A., and Shepherd, A.: Assessing the impact of lead and floe sampling on Arctic sea ice thickness estimates from Envisat and CryoSat-2, Journal of Geophysical Research: Oceans, 124, 7473–7485, 2019.

Trenberth, K. E., Zhang, Y., Fasullo, J. T., and Cheng, L.: Observation-Based Estimates of Global and Basin Ocean Meridional Heat Transport Time Series, Journal of Climate, 32, 4567 – 4583, https://doi.org/10.1175/JCLI-D-18-0872.1, 2019.

Uotila, P., Goosse, H., Haines, K., Chevallier, M., Barthélemy, A., Bricaud, C., Carton, J., Fučkar, N., Garric, G., Iovino, D., et al.: An assessment of ten ocean reanalyses in the polar regions, Climate Dynamics, 52, 1613–1650, 2019.

Wang, Z., Lu, Y., Wright, D. G., Dupont, F., and Division, O. S.: Sea ice sensitivity to the parameterisation of open water area, Journal of Operational Oceanography, 3, 1–7, 2010.

Watts, M., Maslowski, W., Lee, Y. J., Kinney, J. C., and Osinski, R.: A spatial evaluation of Arctic sea ice and regional limitations in CMIP6 historical simulations, Journal of Climate, 34, 6399–6420, 2021.

Wingham, D., Francis, C., Baker, S., Bouzinac, C., Brockley, D., Cullen, R., de Chateau-Thierry, P., Laxon, S., Mallow, U., Mavrocordatos, C., et al.: CryoSat: A mission to determine the fluctuations in Earth's land and marine ice fields, Advances in Space Research, 37, 841–871, 2006.

Xie, J., Bertino, L., Counillon, F., Lisæter, K. A., and Sakov, P.: Quality assessment of the TOPAZ4 reanalysis in the Arctic over the period 1991–2013, Ocean Science, 13, 123–144, 2017.

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, 2018.

Xie, J., Raj, Roshin P.and Bertino, L., MartInez, J., Gabarro, C., and Catany, R.: Assimilation of sea surface salinities from SMOS in an Arctic coupled ocean and sea ice reanalysis, Ocean Science, 19, 269–297, 2023.

Xiu, Y., Min, C., Xie, J., Mu, L., Han, B., and Yang, Q.: Evaluation of sea-ice thickness reanalysis data from the coupled ocean-sea-ice data assimilation system TOPAZ4, Journal of Glaciology, 67, 353–365, 2021.

Zygmuntowska, M., Rampal, P., Ivanova, N., and Smedsrud, L. H.: Uncertainties in Arctic sea ice thickness and volume: new estimates and implications for trends, The Cryosphere, 8, 705–720, 2014.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

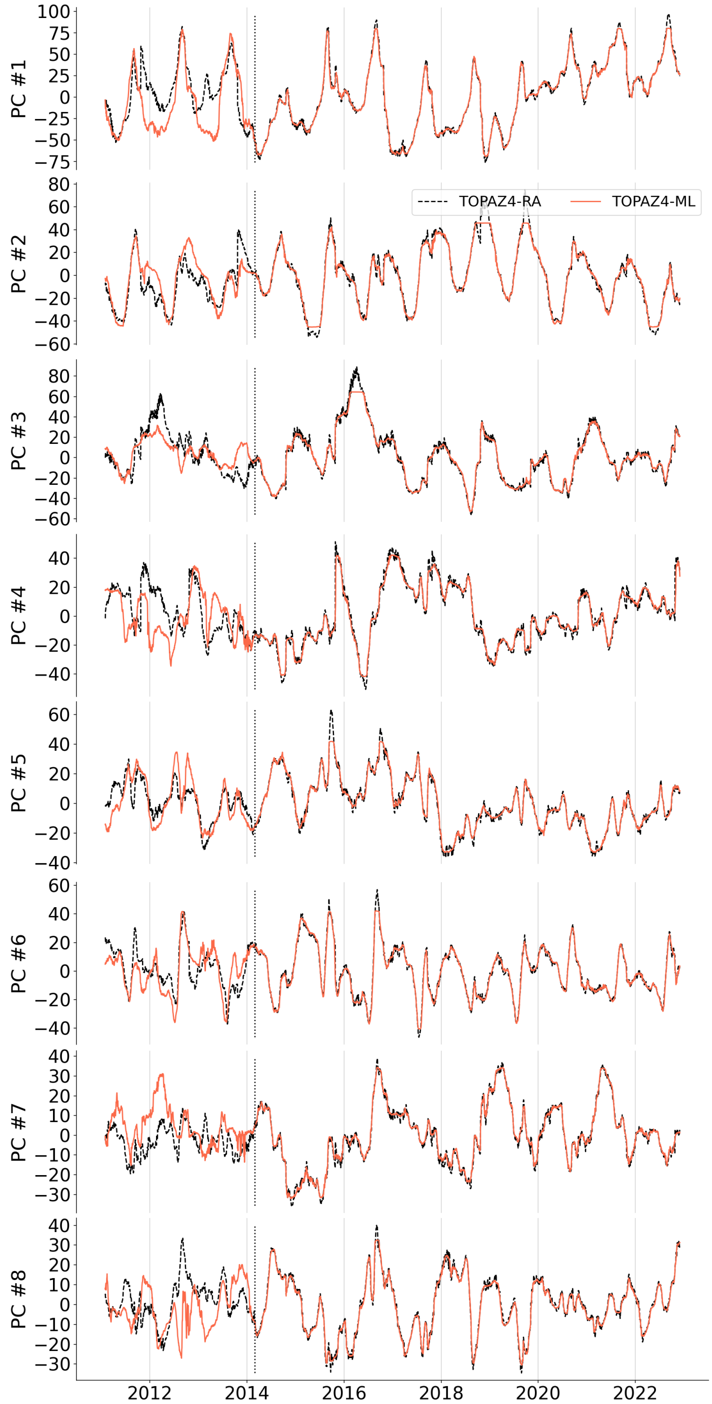

Figure A1. Time series of principal component (PC) for each component in this study. TOPAZ4-RA (considered as truth) and TOPAZ4-ML predicted values are presented. A vertical line in 2014 indicates the separation of the test period from the training period.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

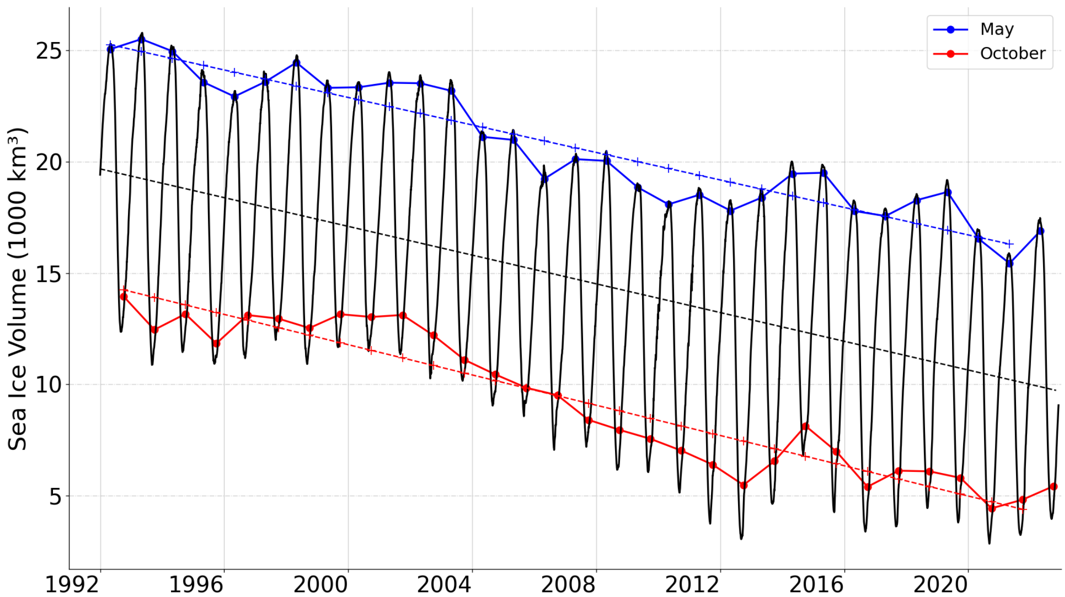

Figure B1. Total sea ice volume (1000 km^3) for the entire Arctic domain from 1992 to 2022. The monthly average in May (October) is indicated in blue (red). Trends are depicted for the entire period in dotted lines. It should be noted that the TOPAZ domain excludes the Pacific Seas south of the Bering Strait.

https://doi.org/10.5194/egusphere-2024-1896
Preprint. Discussion started: 3 July 2024

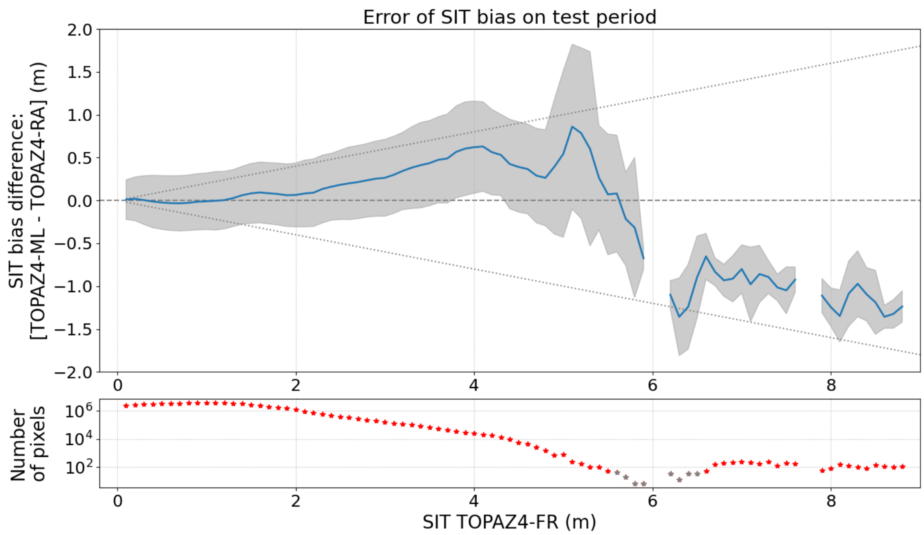

Figure C1. Top) Difference of bias correction between the ML prediction and the true bias correction as a function of the sea ice thickness from TOPAZ4-FR over the test period (2011-2013). The true bias correction is obtained from TOPAZ4-RA - TOPAZ4-FR. Bins of 10 cm are used to average the differences (blue) and their standard deviation (grey). The two oblique lines represent 20% of sea ice thickness for each bin. Positive values indicate that the ML algorithm predicts an excessively high adjustment of sea ice thickness compared to the correction applied by the CS2SMOS data assimilation in TOPAZ4. **Bottom)** Number of pixels collected in each bin as a function of sea ice thickness estimated by TOPAZ4-FR. Grey stars indicate bins with fewer than 50 pixels.