MET-AICE v1.0: an operational data-driven sea ice prediction system for the European Arctic

Cyril Palerme[1], Johannes Röhrs[2], Thomas Lavergne[2], Jozef Rusin[2], Are Frode Kvanum[1, 4], Atle Macdonald Sørensen[2], Arne Melsom[2], Julien Brajard[3], Martina Idžanović[2], Marina Durán Moro[2], and Malte Müller[1, 4]

[1]Development Centre for Weather Forecasting, Norwegian Meteorological Institute, Oslo, Norway
[2]Ocean Department, Norwegian Meteorological Institute, Oslo, Norway
[3]Nansen Environmental and Remote Sensing Center, Bergen, Norway
[4]Section for Meteorology and Oceanography, Department of Geosciences, University of Oslo, Oslo, Norway

Correspondence: Cyril Palerme (cyril.palerme@met.no)

Abstract. There is an increasing need for reliable short-term sea ice forecasts that can support maritime operations in polar regions. While numerous studies have shown the potential of machine learning for sea ice forecasting, there are currently only a few operational data-driven sea ice prediction systems. Here, we introduce MET-AICE, a prediction system providing sea ice concentration forecasts for the next 10 days in the European Arctic. To our knowledge, it is the first operational data-driven prediction system designed for short-term sea ice forecasting. MET-AICE has been trained to predict sea ice concentration observations from the Advanced Microwave Scanning Radiometer 2 (AMSR2) at 5 km resolution. After one year of operation, we show that MET-AICE considerably outperforms persistence of AMSR2 observations (~~root mean square error about 31~~ errors about 30 % lower on average)~~and forecasts from the~~, as well as forecasts from several dynamical models such as TOPAZ5, Barents-2.5km ~~physically-based model (root mean square error about 50 % lower on average) .~~ and the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System.

1 Introduction

There are growing economic and geopolitical interests in the Arctic affecting various sectors such as shipping, tourism, fishing, and resource extraction (Stocker et al., 2020; Gunnarsson, 2021; Müller et al., 2023). This has resulted in a recent increase of the maritime traffic of about 7 % per year over the past decade (Müller et al., 2023). However, sea ice still remains a source of hazards, in particular due to the remoteness of the polar regions. In order to reduce the risk of accidents, seafarers going to ice infested waters are required to get sea ice information before their journeys by the International Code for Ships Operating in Polar Waters since 2017 (Polar Code; Deggim, 2018). The primary source of sea ice information used by seafarers navigating in the Arctic usually consists of sea ice charts produced by national ice services and satellite products (Wagner et al., 2020; Copeland et al., 2024). While the current conditions can be assessed using such observations (though some of these data might already be outdated in areas with fast changing conditions), they do not allow to anticipate the sea ice evolution during the next few days. For route planning, skillful sea ice forecasts could

1

be used (Veland et al., 2021), but dynamical sea ice prediction systems are often not able to outperform persistence of sea ice concentration (SIC) observations for short lead times ~~(Melsom et al., 2019; Röhrs et al., 2023; Palerme et al., 2024; ?)~~ (Melsom et al., 2019; Röhrs et al., 2023; Palerme et al., 2024; Kvanum et al., 2025). Furthermore, the need for accurate sea ice

25 forecasts is also growing due to increasing sea ice motion caused by sea ice thinning (Tandon et al., 2018; Tschudi et al., 2020).

Several studies have recently shown that data-driven sea ice forecasting systems trained on satellite observations can be skillful ~~(e.g. Grigoryev et al., 2022; Ren et al., 2022; Chen et al., 2023; Keller et al., 2023; Koo and Rahnemoonfar, 2024)~~(e.g. Grigoryev et a, and can provide more accurate forecasts than dynamical models ~~(Andersson et al., 2021; Palerme et al., 2024; ?)~~(Andersson et al., 2021; F. However, most operational sea ice prediction systems are still based on dynamical models (e. g. Smith et al., 2016; Barton

30 et al., 2021; Williams et al., 2021; Ponsoni et al., 2023; Röhrs et al., 2023; Paquin et al., 2024), and there have only been a few based on machine learning approaches. The first operational data-driven sea ice prediction system developed has been IceNet (Andersson et al., 2021), which provides forecasts of the probability that SIC exceeds 15 % for the next 6 months. For lead times from 2 to 6 months, IceNet outperforms the forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) SEAS5 seasonal prediction system (Johnson et al., 2019), while running 2000 times faster on a laptop

35 than SEAS5 on a supercomputer. Furthermore, Palerme and Müller (2021) developed post-processed sea ice drift forecasts for lead times up to 10 days using machine learning that have been delivered on the commercial application IcySea from 2020 to 2024 (https://driftnoise.com/icysea.html, von Schuckmann et al., 2021).

Sea ice changes on short-time scales are primarily driven by the atmosphere, and in particular by the wind (Mohammadi-Aragh et al., 2018; Yu et al., 2020). Hence, it is crucial to include predictors from weather forecasts when developing data-

40 driven sea ice prediction systems, as suggested by previous studies. Grigoryev et al. (2022) reported an improvement of 5 to 15 % when ~~forecasts~~ using forecasts of 2-meter temperature, surface pressure and wind from the National Centers for Environmental Prediction (NCEP) operational Global Forecast System (GFS)~~are used, whereas~~. Moreover, Palerme et al. (2024) assessed an error reduction of 7.7 % when using ECMWF ~~weather forecasts~~forecasts of 2-meter temperature and 10-meter wind in addition to sea ice predictors. In this paper, we introduce MET-AICE, a data-driven sea ice prediction

45 system producing 10-day SIC forecasts with daily time steps in the European Arctic. The predictors used in MET-AICE are derived from SIC observations provided by the Advanced Microwave Scanning Radiometer 2 (AMSR2), ECMWF weather forecasts, and a land sea mask. The forecasts have been produced daily since March 2024 and are publicly available (see code and data availability section). In section 2, the datasets used in MET-AICE and for verification are described. Then, the deep learning prediction system and the verification methods are presented in section 3. ~~The performance of~~ MET-AICE is evaluated

50 and compared to the ~~Barents-2.5km Ensemble Prediction System (EPS)~~ forecasts from three dynamical models in section 4, followed by the discussion and conclusions in section 5.

Table 1. List of predictors used in MET-AICE.

Source	Variable	Time
AMSR2	SIC observations	Day preceding the forecast start date
AMSR2	Land sea mask	Constant predictor
ECMWF	2-meter temperature	Mean value between the forecast start date and the predicted lead time
ECMWF	10-meter x wind component	Mean value between the forecast start date and the predicted lead time
ECMWF	10-meter y wind component	Mean value between the forecast start date and the predicted lead time

2 Data

2.1 Datasets used for MET-AICE

MET-AICE has been trained to predict daily SIC observations at about 5 km resolution derived from AMSR2 data using a three-step algorithm called reSICCI3LF (Rusin et al., 2024). The first step combines the AMSR2 microwave imagery channels at 19 and 37 GHz to derive a SIC field at about 15 km resolution with relatively low uncertainties. It then computes a higher resolution SIC field (about 5 km) using the two 89 GHz channels. This higher-resolution SIC has larger uncertainties due to the atmosphere being less transparent at these higher frequencies. Finally, the high resolution details obtained from the 89 GHz channels are added to the coarser resolution field to produce a final SIC field at 5 km resolution with relatively low uncertainties. For more details about this algorithm and product, we refer to Rusin et al. (2024) and to Palerme et al. (2024) where this product was shown to be more accurate for the ice edge position than the OSI-408-a product from the Ocean and Sea Ice Satellite Application Facility (OSI SAF), also derived from AMSR2.

The AMSR2 SIC observations from the day preceding the forecast start date are used as a predictor in MET-AICE, and can be considered as the initial SIC conditions of the prediction system (table 1). The land grid points in these observations are considered as ice-free ocean (0 % SIC) since only valid values can be provided to the neural networks. MET-AICE also uses a land sea mask as a predictor, as well as ECMWF weather high-resolution forecasts (HRES, grid spacing of 9 km) in 3 predictors that are the 2-meter temperature and the 10-meter wind (x and y components on the MET-AICE grid). For the predictors from ECMWF weather forecasts, the mean values between the forecast start date and the predicted lead time are used. Due to this choice, 10 different models were developed for predicting the SIC evolution during the next 10 days. All the predictors are normalized using the statistics from the training dataset (values ranging between 0 and 1), and projected onto the MET-AICE grid (Lambert azimuthal equal area at 5 km resolution) using nearest neighbor interpolation before providing them to the neural networks. In Palerme et al. (2024) and Kvanum et al. (2025), the SIC trend from passive microwave observations during the 5 days preceding the forecast start date was used as a predictor, but it was shown to have a negligible impact on the predictions. Due to the limited importance of this predictor and the potential impact of missing data on the production of the forecasts, no SIC trend from satellite observations is used in MET-AICE.

2.2 Datasets used for verification

2.2.1 Observations

The AMSR2 SIC product from Rusin et al. (2024) is used as reference in this study since MET-AICE has been trained to predict these observations. However, passive microwave products sometimes indicate some spurious sea ice along the coasts due to the difference in brightness temperatures over open water and land, as well as the spatial resolution of several kilometers (Lavergne et al., 2019; Kern et al., 2020). In order to avoid this issue, the oceanic grid points closer than 20 km from the coasts are not taken into account when the forecasts are evaluated using AMSR2 SIC observations as reference.

In addition, the ice charts produced by the ice service of the Norwegian Meteorological Institute (JCOMM Expert Team on sea ice, 2017; Copeland et al., 2024) are used as an independent dataset for verification. The ice charts, which represent sea ice classes of concentration, are manually drawn by ice analysts during week days using several types of observations. Synthetic aperture radar (SAR) observations are used where they are available due to their high spatial resolution. In other areas, the analysts prioritize visible and infrared data, whereas passive microwave observations are only used where no higher resolution data are available. ~~Since~~ Therefore, land contamination is much less present in the ice charts ~~are not produced during weekends and public holidays~~than in passive microwave observations, and we decided to take into account all oceanic grid points when the ice charts are used as reference (no coastal grid points excluded). Furthermore, there are fewer ice chart data accessible for verification than ~~using~~ AMSR2 observations because the ice charts are not produced during weekends and public holidays.

~~Sea ice forecasts from the Barents-2.5km EPS (Röhrs et al., 2023)~~

2.2.2 Dynamical models

SIC forecasts from three dynamical models are used as ~~a benchmark~~ benchmarks for evaluating the performances of MET-AICE. The forecasts from ECMWF Integrated Forecasting System starting at 00:00 UTC (hereafter referred to as "ECMWF IFS") are evaluated in this study. They are produced on an octahedral reduced Gaussian grid (O1280) at about 9 km resolution and cover lead times up to 15 days. ECMWF IFS is a coupled numerical weather prediction system with the ocean component consisting of the NEMO ocean model and the LIM2 sea ice model which uses a viscous-plastic rheology (ECMWF, 2023, 2024). The initial oceanic conditions are based on the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system (Donlon et al., 2012) as well as the OCEAN5 ocean analysis (Zuo et al., 2018). ECMWF IFS is constantly developed and forecasts from different model cycles are used in this study ~~.~~(https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecm

The TOPAZ5 Arctic Ocean system (hereafter referred to as "TOPAZ5") produces daily operational ocean and sea ice forecasts that are delivered by the Copernicus Marine Service (Ali et al., 2025). It is based on the Hybrid Coordinates Ocean Model (HYCOM, Bleck, 2002; Chassignet et al., 2006) version 2.2.98 coupled with the Los Alamos Sea Ice Model CICE version 5.1 (Hunke et al., 2017) and the ECOSMO-II biogeochemical model (Yumruktepe et al., 2022). Data assimilation is performed weekly using a 100-member ensemble Kalman filter (EnKF). The forecasts consist of 10 members covering lead

times up to 10 days with hourly time steps and a spatial resolution of 6.25 km. The atmospheric forcing comes from ECMWF weather forecasts, and the forecasts are available from September 2023.

110   The Barents-2.5km ~~EPS~~ Ensemble Prediction System (hereafter referred to as "Barents-2.5") is a regional ocean-sea ice dynamical model producing hourly forecasts with lead times up to 96 hours (Röhrs et al., 2023). It consists of the Regional Ocean Modeling System (ROMS) version 3.7 (Shchepetkin and McWilliams, 2005) coupled with the Los Alamos Sea Ice Model (CICE) version 5.1 (Hunke et al., 2017). Barents-2.5 has 42 vertical layers and a spatial resolution of 2.5 km. The ensemble consists of 24 ensemble members with daily assimilation cycles (AMSR2 SIC from Rusin et al. (2024), sea surface

115   temperature, in-situ temperature and salinity observations) as well as updates of atmospheric files four times a day (six members are respectively updated at model hours 00:00, 06:00, 12:00, and 18:00 UTC). Only ~~daily means of~~ the members produced at 00:00 UTC are used ~~in this study~~ here for a consistent comparison with MET-AICE. In five of the six ensemble members produced at 00:00 UTC, the atmospheric forcing comes from ECMWF ensemble forecasts (ECMWF-ENS) at 9 km resolution. The remaining member is forced by the regional weather prediction system AROME-Arctic (Müller et al., 2017) developed

120   at the Norwegian Meteorological Institute, which has the same domain and spatial resolution as Barents-2.5. We chose to compare MET-AICE to Barents-2.5 since the same AMSR2 SIC observations are used to initialize both prediction systems. An ensemble Kalman filter is used to assimilate these observations in Barents-2.5, though the model's ensemble is underdispersed (lack of ensemble spread, see Idžanović et al. (2023)), which indicates an overly high confidence in the forecast and thus limits the impact of the assimilation.

125   This study primarily focuses on evaluating the MET-AICE operational forecasts that are deterministic and have daily time steps. Therefore, the daily means of SIC fields from the dynamical models are evaluated. These datasets consist of the high-resolution ECMWF IFS (HRES) forecasts, the TOPAZ5 ensemble mean, the Barents-2.5 member forced by AROME-Arctic, and the mean of Barents-2.5 members forced by ECMWF-ENS forecasts. In addition, we also investigated producing probabilistic forecasts with MET-AICE in this study. For this experiment, MET-AICE forecasts were compared to the first 10 ensemble

130   members from ECMWF IFS (ENS), the 10 TOPAZ5 ensemble members, and the 5 Barents-2.5 ensemble members forced by ECMWF-ENS produced at 00:00 UTC.

3   Methods

3.1   Deep learning prediction system

MET-AICE is based on convolutional neural networks with a U-Net architecture (Ronneberger et al., 2015), including residual

135   connections (He et al., 2016) and spatial attention blocks (Oktay et al., 2018). This architecture (Fig. 1) was shown to outperform the original U-Net model for short-term sea ice prediction (Palerme et al., 2024). While the residual connections are used in every layer of the neural network, the spatial attention blocks are only present in the layers of the decoder. The residual connections are used to avoid vanishing issues and ease neural network training (He et al., 2016), whereas the spatial attention blocks are designed to identify and give more attention to relevant areas during training (Oktay et al., 2018). The MET-AICE

140   grid (480 x 544 grid points) has a spatial resolution of 5 km, and the neural networks are composed of five downsampling and
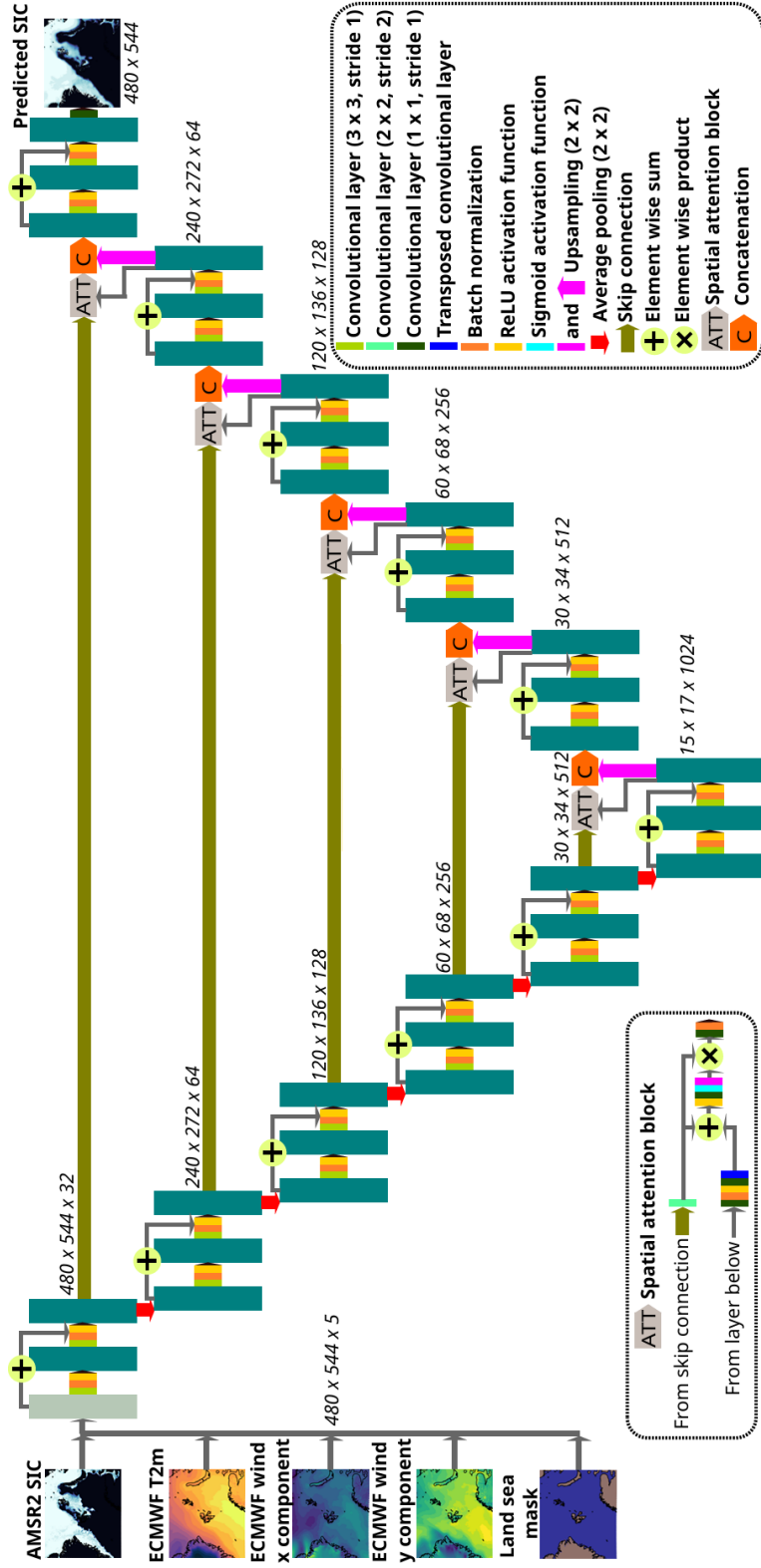
Figure 1. Architecture used in MET-AICE. The big green rectangles represent multi-channel feature maps. The dimensions of the feature maps (y, x, number of channels) are written next to each convolutional block.

6

five upsampling operations. 32 convolutional kernels are used in the first layer, and then the number of convolutional kernels is doubled at every layer in the encoder (and divided by two in the decoder). The models were trained to minimize the mean squared error during 100 epochs with a batch size of 4. An initial learning rate of 0.005 was used with an Adam optimizer, which was then reduced by a factor of 2 every 25 epochs. In order to avoid overfitting, the model version with lowest validation loss was selected during training for each lead time. The models contain about 39 million parameters.

The deep learning models were trained using ~~weekly data~~ one forecast per week during the period 2013 - 2020 (about 400 forecasts for each lead time). While using weekly data decreases the size of the training dataset, this also reduces the overlap between the forecasts for long lead times, which improved the performances for lead times longer than 1 day during the tuning phase (probably due to less overfitting). Nevertheless, all the forecasts available (daily data) produced in 2021 were used for validation (for tuning the models), and all the forecasts from 2022 were used for the test dataset during the development of MET-AICE. ~~However, this study focuses on evaluating the forecasts that have been operationally produced since March 2024.~~ Furthermore, ECMWF weather forecasts produced by different model cycles were used for training the deep learning models due to the 8-year training period (https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model).

3.2 Verification

3.2.1 Verification scores

In this study, the SIC and the ice edge position (defined by the 10 % SIC contour and excluding coastlines) are evaluated ~~using two different verification scores~~. In order to assess the quality of SIC forecasts, the root mean square error (RMSE) of the SIC is calculated over all oceanic grid points using AMSR2 observations as reference. Since the ice charts do not provide SIC as a continuous variable but as sea ice classes, the ice charts are only used for evaluating the ice edge position. The ice edge position in deterministic forecasts is evaluated using the ratio of the Integrated Ice Edge Error (IIEE, Goessling et al., 2016) over the length of the ice edge in the observations used as reference (hereafter referred to as "ice edge distance error", equation 1). This metric was introduced by Melsom et al. (2019) and provides the mean distance between the ice edges in the forecasts and in the observations. It is worth noting that normalizing the IIEE by the ice edge length (calculated using the method described in Melsom et al. (2019)) allows to use a metric which is not seasonally dependent, contrary to using the IIEE without normalization (Goessling et al., 2016; Palerme et al., 2019).

~~The~~

$$Ice\ edge\ distance\ error = \frac{IIEE}{Observed\ ice\ edge\ length} \tag{1}$$

While deterministic SIC forecasts are evaluated in most of this study, an experiment was performed with probabilistic forecasts. The ice edge position in probabilistic forecasts is assessed using the sea ice probability ($SIP_{forecasts}$ in equation 2), which is defined as the fraction of ensemble members with a SIC higher or equal to 10 %, whereas a binary sea ice probability is used for the observations ($SIP_{observations}$ in equation 2). Then, the forecasts are evaluated using the ratio of the Spatial Probability Score (SPS, Goessling and Jung, 2018) over the ice edge length in the observations used as reference (hereafter referred to as

7

"probabilistic ice edge distance error", equation 2). This metric was introduced by Palerme et al. (2019) and can be considered as the probabilistic extension of the ice edge distance error.

175
$$Probabilistic\ ice\ edge\ distance\ error = \frac{\int_x \int_y \left(SIP_{observations}(x,y) - SIP_{forecasts}(x,y) \right)^2 dydx}{Observed\ ice\ edge\ length} \qquad (2)$$

3.2.2 Benchmark forecasts and evaluation period

In this study, the forecasts are considered skillful if they outperform persistence of AMSR2 SIC observations from the day before the forecast start date (hereafter referred to as "persistence of AMSR2 observations"). When the ice charts are used as reference, persistence of the ice charts from the day before the forecast start date (hereafter referred to as "persistence of ice
180 charts") is used as an additional benchmark forecast, though this reduces the number of days available for verification due to the lack of ice charts during weekends and public holidays.

In addition, we also developed a similar bias correction method for ~~Barents-2.5~~ the dynamical models as the one used for TOPAZ4 in Palerme et al. (2024), where we calculate the difference between the first hourly time step from ~~Barents-2.5~~ the dynamical models and the AMSR2 observations from the day before. Then, we subtract this difference from ~~Barents-2.5~~ the
185 forecasts for longer lead times in order to produce a benchmark forecast called "~~Barents-2.5 bias corrected~~ bias corrected forecasts" (the values lower than 0 % and higher than 100 % are then replaced by 0 % and 100 %, respectively). Due to the limitations of passive microwave SIC observations along the coasts (where spurious sea ice can be reported), this bias correction was not applied to the grid points that were closer than 20 km from the coasts. Instead, the raw forecasts were used for areas close to land. This improves the impact of the bias correction when the forecasts are evaluated using the ice charts
190 as reference, probably due to land contamination in the passive microwave observations. The bias correction used in this study is slightly different from the one proposed by Palerme et al. (2024) since the bias is calculated using the first hourly time step ~~from Barents-2.5~~ here whereas the first daily time step from TOPAZ4 was used in Palerme et al. (2024), and because the bias correction was also applied to coastal grid points in Palerme et al. (2024).

The forecasts are primarily evaluated during the period April 2024 - March 2025, which corresponds to the first year during
195 which the MET-AICE forecasts have been produced operationally. During this period, the forecasts from the three dynamical models are available. While TOPAZ5 and ECMWF IFS cover the pan-Arctic, the domain of Barents-2.5 does not cover the full domain of MET-AICE (Fig.2 a). Due to the difference in model domains, the SIC forecasts are evaluated in the shared domain between Barents-2.5 and MET-AICE after projecting all the datasets on the MET-AICE grid using nearest neighbor interpolation. It is also worth noting that a common land sea mask is used for all the datasets during verification (land if there
200 is land in at least one of the dataset). In addition, we also evaluated MET-AICE and ECMWF IFS forecasts during the period 2022 - 2024 since only data from 2013 to 2021 were used during the development of MET-AICE (2013 - 2020 for training, and 2021 as validation dataset). TOPAZ5 and Barents-2.5 are not included in this comparison because TOPAZ5 forecasts have only been available since September 2023, and Barents-2.5 forecasts have had lead times up to 4 days since December 2023.
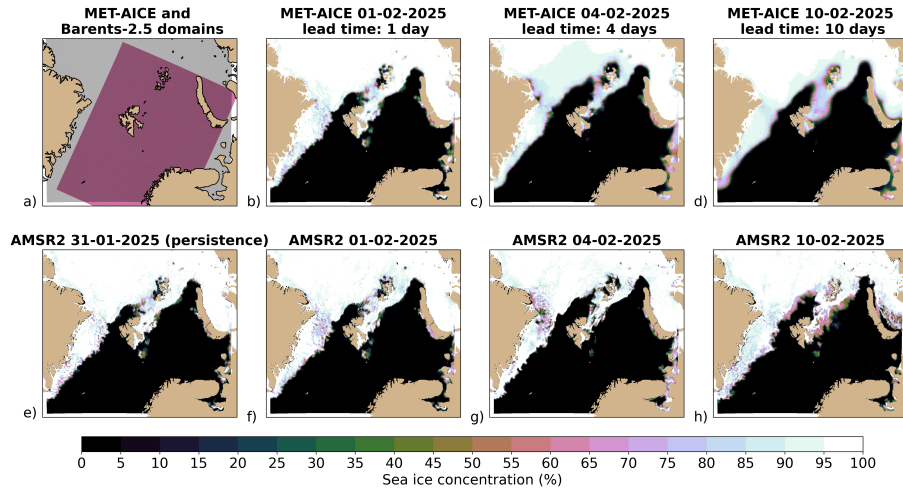
Figure 2. a) Domains of MET-AICE (gray) and Barents-2.5 (purple). Comparison between the MET-AICE forecasts initialized on the 01/02/2025 at 00:00 UTC (b to d) and the AMSR2 sea ice concentration observations for the corresponding days (f to h). e) The AMSR2 observations from the day preceding the forecast start date that are used as a predictor in the MET-AICE forecasts initialized on 01/02/2025.

During the period 2022 - 2024, MET-AICE and ECMWF IFS forecasts are evaluated over the full MET-AICE domain after
205   projecting all the datasets onto the MET-AICE grid.

4   Results

Fig. 2 shows the MET-AICE forecasts initialized on 01/02/2025 for lead times of 1, ~~5~~4, and 10 days. During the first ~~five~~ four days of February 2025, the surface of a polynya located around Franz Josef Land grew before closing between ~~05~~04/02/2025 and 10/02/2025. MET-AICE was able to reproduce relatively well the growth and the closing of this polynya, though the
210   magnitude of the closing was underestimated. The retreat of the ice edge north of Svalbard during the first ~~five~~ four days of February 2025 was also predicted successfully. There was also a decrease of the sea ice extent around Novaya Zemlya which was predicted by MET-AICE, though it was slightly underestimated. However, the sea ice patches located between Svalbard and Bjørnøya remained longer than in MET-AICE predictions. Furthermore, the MET-AICE forecasts become smoother with increasing lead times~~, which is expected for deterministic deep learning models~~. This is due to the use of the mean squared
215   error as loss function, as well as ~~due~~ to the temporal averaging of the predictors from ECMWF weather forecasts.

~~The performances of MET-AICE and Barents-2.5 for the period April~~ In order to investigate to what extent the verification scores used in this study are influenced by the effective spatial resolution of the SIC fields, a sensitivity experiment was performed (Fig. 3). The AMSR2 SIC observations from 2013 to 2024 were used to calculate the RMSE of the SIC and the ice edge distance error for 1-day persistence of AMSR2 retrievals. The observations from the first day were smoothed using
220   various Gaussian filters with a standard deviation ranging from 5 km (1 grid point) to 30 km (6 grid points), and compared to
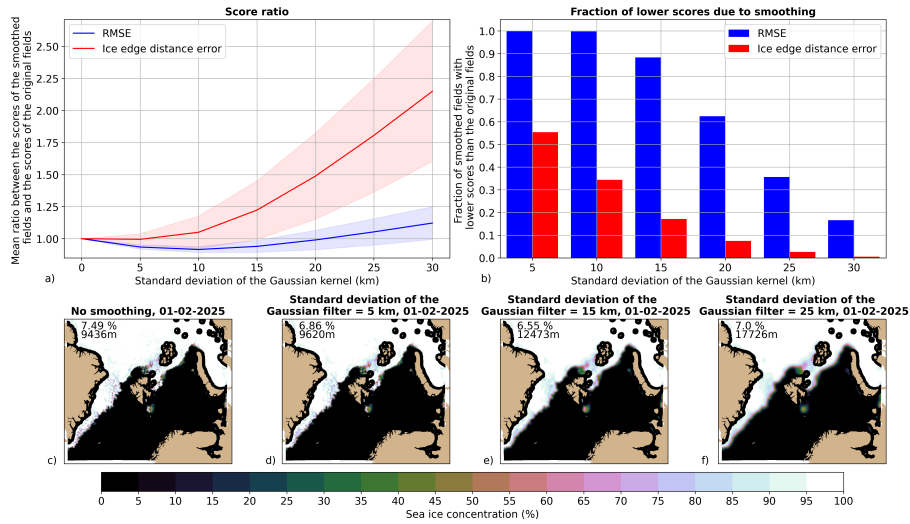
9

Figure 3. ~~Performances~~ Influence of ~~MET-AICE and Barents-2.5~~ the effective spatial resolution on the verification scores. The AMSR2 observations during the period ~~April 2024~~ 2013 - ~~March 2025. a) Root mean square error (~~2024 were used to calculate the RMSE ~~)~~ of the sea ice concentration and the ice edge distance error for 1-day persistence of AMSR2 observations over the ~~shared domain between~~ MET-AICE ~~and Barents-2.5~~ domain (4327 days evaluated). The AMSR2 observations from the first day were smoothed using various Gaussian filters with a standard deviation ranging from 5 km (1 grid point) to 30 km (6 grid points), while the original AMSR2 observations from the next day were used as reference for computing the scores. ~~b) Evaluation of~~ The ice edge distance error was computed using the ice edge ~~position~~ length from the observations of the second day at the original spatial resolution (~~defined by~~ 5 km). The grid points closer than 50 km from the ~~10 %~~ coastlines were excluded from this analysis. a) The mean ratio between the scores after smoothing and the scores using the original data (a score lower than 1 means that the smoothing results in a better score). The shaded areas represent the standard deviations of the verification scores. b) The fraction of scores that are better after smoothing the AMSR2 sea ice concentration ~~contour~~ fields. c) ~~using~~ AMSR2 sea ice concentration observations ~~as reference. c~~ at the original spatial resolution (5 km) ~~Evaluation~~ on 01/02/2025. d, e, f) AMSR2 sea ice concentration observations on 01/02/2025 smoothed using Gaussian filters with a standard deviation of 5 km (d), 15 km (e), and 25 km (f). The scores on the top left corners of the maps show the RMSE of the sea ice ~~edge position using~~ concentration (top, in %) and the ice ~~charts as reference.~~ edge distance error (bottom, in meters) for the comparison between the AMSR2 observations from 01/02/2025 and 02/02/2025.

the original AMSR2 SIC field from the second day. Only the ice edge length from the observations of the second day at the original spatial resolution (5 km) was used to compute the ice edge distance error regardless of the size of the Gaussian filter. In order to avoid some biases resulting from smoothing coastal grid points, the grid points closer than 50 km from the coastlines were not taken into account in this analysis. Smoothing the AMSR2 SIC fields results in lower RMSE if the standard deviation of the Gaussian filter is between 5 and 20 km. The lowest RMSE is obtained for a Gaussian filter with a standard deviation of 10 km, with a reduction of 8.4 % compared to the RMSE obtained using the original SIC fields at 5 km resolution. This is in contrast with the ice edge distance error, which is only reduced by 0.5 % when a Gaussian filter with a standard deviation of 5 km is applied compared to the score obtained with the original resolution. Moreover, the ice edge distance error constantly
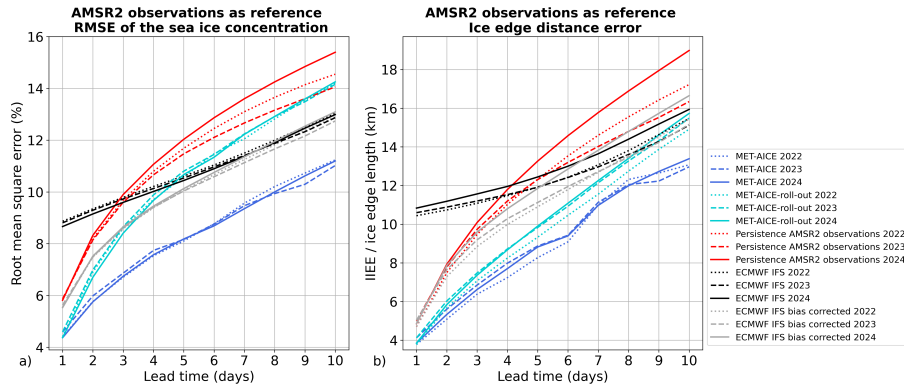
225

Figure 4. Performances of MET-AICE and ECMWF IFS during the years 2022, 2023, and 2024 over the full MET-AICE domain using AMSR2 observations as reference. MET-AICE-roll-out corresponds to using the MET-AICE model for 1-day lead time auto-regressively to predict the sea ice concentration for lead times up to 10 days. a) Root mean square error (RMSE) of the sea ice concentration. b) Evaluation of the ice edge position (defined by the 10 % sea ice concentration contour).

increases when larger Gaussian filters are applied. Therefore, the ice edge distance error does not really favor smoother SIC
230   fields contrary to the RMSE of the SIC.

MET-AICE and ECMWF IFS forecasts are evaluated during the period 2022 - ~~March 2025 are shown in Fig. ??.~~ 2024 over the full domain of MET-AICE ~~considerably outperforms persistence of~~ using AMSR2 observations ~~for all lead times. When AMSR2 observations are used as reference~~ as reference (Fig. 4). In addition, we evaluated the MET-AICE model developed for 1-day lead time in an auto-regressive mode to predict SIC up to 10 days ahead (hereafter referred to as "MET-AICE-roll-out").
235   In this configuration, the SIC prediction from the previous lead time is used as a predictor to predict the next time step. While this configuration allows to improve the consistency between the time steps, it results in poorer performances than the operational version of MET-AICE. On average over all lead times, the RMSE of ~~SIC for~~ the SIC is about 19 % larger for MET-AICE-roll-out than for the operational version of MET-AICE ~~is about 31 % lower than for~~, and the error for the ice edge position is about 10 % larger for MET-AICE-roll-out. Therefore, we decided to only present the results for the operational
240   version of MET-AICE for the rest of this study. For all lead times and all the years evaluated, MET-AICE considerably outperforms persistence of AMSR2 observations ~~on average (between~~ (RMSE of the SIC and ice edge distance error about 28 % smaller on average), ECMWF IFS forecasts (RMSE of the SIC and ice edge distance error about 25 % and ~~33 % depending on lead time). The ice edge position is also more accurate in~~ 30 % smaller, respectively), as well as ECMWF IFS bias corrected (RMSE of the SIC and ice edge distance error about 18 % and 21 % smaller, respectively). Furthermore,
245   there is little interannual variability in the performances of MET-AICE~~, with errors about 32 % lower than for persistence of AMSR2 observations (between 25 % and 37 % depending on lead time). Barents-2.5 forecasts are less accurate than~~, which suggests that re-training MET-AICE does not have to be done every year. While ECMWF IFS does not outperform persistence of AMSR2 observations for ~~all lead times . It is worth noting that the member forced by the regional weather~~
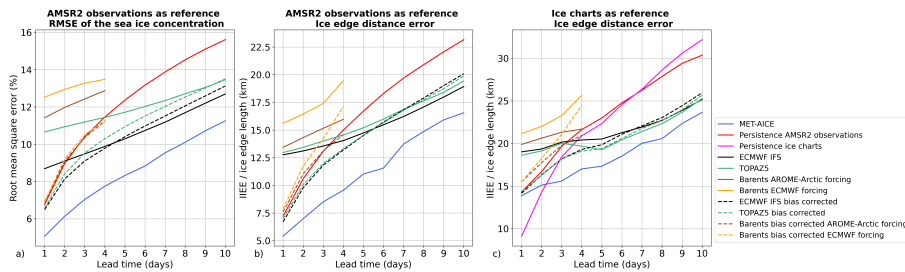
Figure 5. Performances of MET-AICE, ECMWF IFS, TOPAZ5 and Barents-2.5 during the period April 2024 - March 2025 over the shared domain between MET-AICE and Barents-2.5. a) Root mean square error (RMSE) of the sea ice concentration using AMSR2 observations as reference. b) Evaluation of the ice edge position (defined by the 10 % sea ice concentration contour) using AMSR2 observations as reference. c) Evaluation of the ice edge position using the ice charts as reference.

prediction system AROME-Arctic outperforms the members forced by ECMWF weather forecasts, which is consistent with
250 the findings of Idžanović et al. (2023) regarding surface currents. Nevertheless, lead times up to 3 days, the bias correction allows a considerable improvement of the forecast skill for lead times up to 6 days.

In Fig. 5, the bias-corrected forecasts from performances of MET-AICE and the three dynamical models (ECMWF IFS, TOPAZ5, and Barents-2.5achieve considerably better performances, comparable to persistence of) during the period April 2024 - March 2025 are evaluated over the domain shared by MET-AICE and Barents-2.5. When the forecasts are evaluated
255 using AMSR2 observations.

When the ice edge position is evaluated using the ice charts as reference observations as reference (Fig. 5 a and b), MET-AICE is able to outperform outperforms all benchmark forecasts for all lead times. Compared to persistence of AMSR2 observationsfor , MET-AICE has a RMSE of the SIC about 31 % smaller on average over all lead times (by 24 % on average), but not between 25 % and 33 % depending on lead time), and an ice edge distance error about 32 % smaller on average
260 (between 25 % and 37 %). Nevertheless, MET-AICE is not able to outperform persistence of the ice charts for 1-day lead times of 1 and 2-day lead times 2 days when the ice charts are used as reference due to the differences between the two observational products. However, MET-AICE outperforms persistence of ice charts has considerably lower errors for the ice edge position than all benchmark forecasts for longer lead times , with an (ice edge distance error about 8 % lower on average for all lead times. Although Barents-2.5 forecasts do not outperform persistence of 23 % lower than for persistance of the ice
265 charts for any lead time, they outperform persistence of AMSR2 observations from 3-day lead time with the member forced by AROME-Arctic, and from 4-day lead time with the members forced by ECMWF weather forecasts. Nevertheless, the Barents-2.5 bias corrected forecasts outperform persistence of AMSR2 observations for all lead times, while it is not the case lead times from 3 to 10 days). Among the dynamical models, while ECMWF IFS has the best performances when AMSR2 observations are used as reference. This is probably due to the coastal grid points (distance from the coast shorter than 20
270 km) which are taken into account when the ice charts are used as reference, but which are excluded when using AMSR2 observations , TOPAZ5 has slightly better performances than ECMWF IFS when the forecasts are evaluated using ice charts
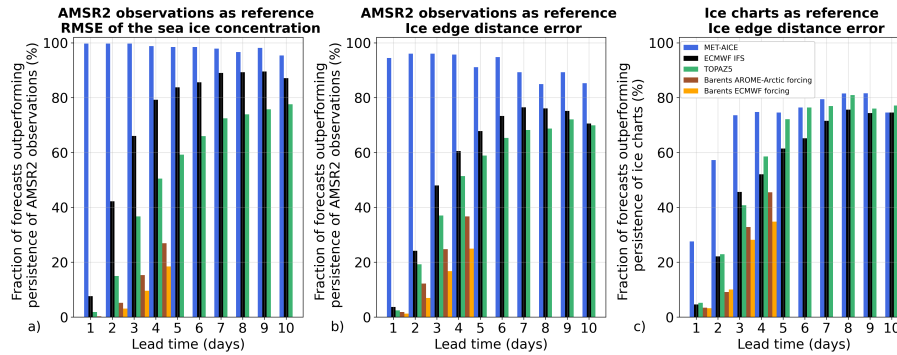
Figure 6. Fraction of days during which the forecasts outperform persistence of AMSR2 observations for the root mean square error of the sea ice concentration (a) and the ice edge position (b) using AMSR2 observations as reference during the period April 2024 - March 2025. c) Fraction of days during which the forecasts outperform persistence of the ice charts for the ice edge position using the ice charts as reference. The forecasts are evaluated over the shared domain between MET-AICE and Barents-2.5.

as reference. ~~Furthermore, the~~ It is also worth noting that Barents-2.5 ~~bias corrected forecastsachieve similar performances as MET-AICE for 1-day lead time, but have larger errors for longer lead times~~performs considerably better when the model is forced by AROME-Arctic weather forecasts than when it is forced by ECMWF weather forecasts, which is consistent with the findings of Idžanović et al. (2023) regarding surface currents.

In Fig. 6, the fraction of days during which the forecasts can be considered skillful are shown. MET-AICE outperforms persistence of AMSR2 observations for the RMSE of the SIC between ~~96~~ 95 % and 99.7 % of the days depending on lead time (98 % on average), and between ~~86 % and 97~~ 85 % and 96 % of the days (92 % on average) for the ice edge position using AMSR2 observations as reference. When the ice edge position is evaluated using the ice charts as reference, MET-AICE outperforms persistence of the ice charts only ~~26~~ 28 % of the days for 1-day lead time. Nevertheless, it outperforms persistence of the ice charts ~~55~~ 57 % of the time for 2-day lead time, and at least ~~68~~ 74 % of the days for longer lead times. While ECMWF IFS is the best dynamical model for all lead times when AMSR2 observations are used as reference, TOPAZ5 performs slightly better than ECMWF IFS when the ice charts are used as reference. Despite the assimilation of AMSR2 SIC observations, Barents-2.5 outperforms more often persistence of the ice charts using the ice charts as reference than persistence of AMSR2 observations using AMSR2 observations as reference. This is consistent with Durán Moro et al. (2024) who also reported a better agreement between Barents-2.5 and the ice charts than with AMSR2 observations due to a higher SIC in the ice charts than in the AMSR2 product, and an over-generation of ice by the Barents-2.5 model (in particular in ~~certain~~ some areas such as the Fram Strait).

The seasonal variability in the performances of MET-AICE and ~~Barents-2.5~~ the dynamical models for 1-day and 3-day lead times are shown in Fig. 7 and 8. When AMSR2 observations are used as reference (Fig. 7), MET-AICE considerably outperforms persistence of AMSR2 observations for all ~~months, while Barents-2.5 has larger errors than persistence~~ the months for ~~all months~~1-day and 3-day lead times. This is in contrast with the dynamical models since none of them outperform persistence
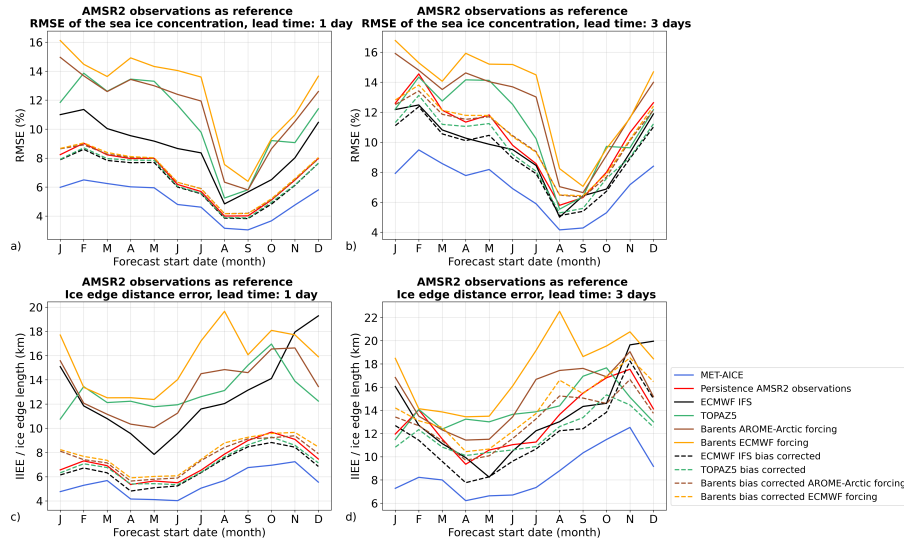
13

Figure 7. Seasonal variability of the performances of MET-AICE, ECMWF IFS, TOPAZ5, and Barents-2.5 for 1-day and 3-day lead times using AMSR2 observations as reference during the period April 2024 - March 2025. a, b) Root mean square error (RMSE) of the sea ice concentration. c, d) Evaluation of the ice edge position (defined by the 10 % sea ice concentration contour). The forecasts are evaluated over the shared domain between MET-AICE and Barents-2.5.
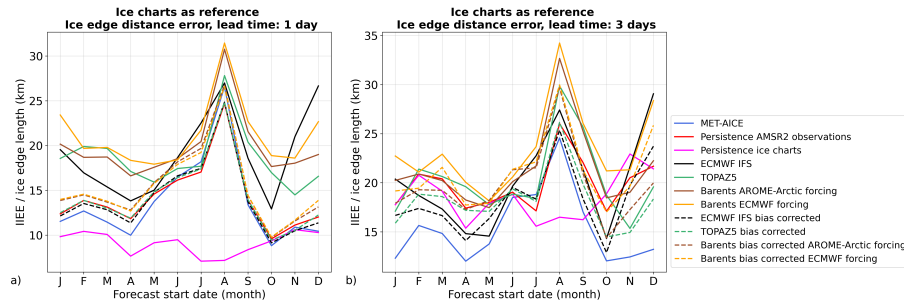

Figure 8. Seasonal variability of the performances of MET-AICE, ECMWF IFS, TOPAZ5, and Barents-2.5 for 1-day (a) and 3-day (b) lead times using the ice charts as reference during the period April 2024 - March 2025. The forecasts are evaluated over the shared domain between MET-AICE and Barents-2.5.

of AMSR2 observations for any month without bias correction for 1-day lead time. However, there is a much stronger seasonal variability when the ice charts are used as reference (Fig. 8) due to the differences between the two observational products (Palerme et al., 2024). For 1-day lead time, MET-AICE ~~never~~ only slightly outperforms persistence of the ice charts in October, and has similar errors for the ice edge position as persistence of ~~AMSR2 observations in July and August~~ice charts in November and December. Nevertheless, MET-AICE outperforms persistence of the ice charts most of the year for 3-day lead time, but has larger errors than persistence of the ice charts between July and September.
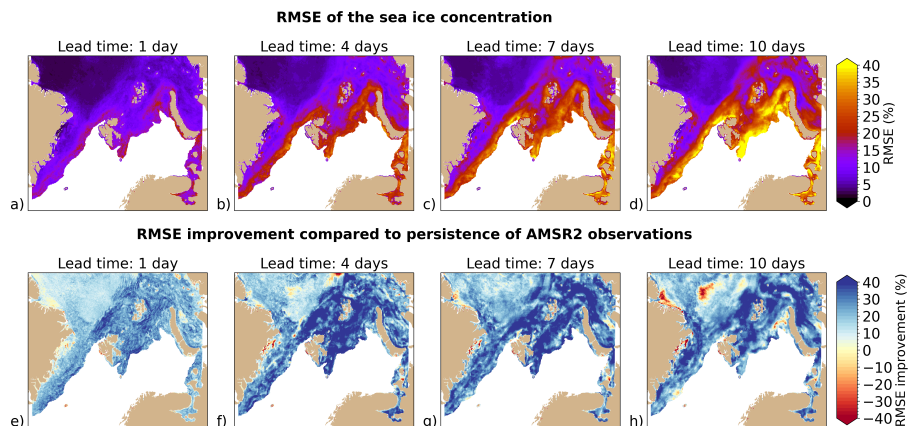
14

RMSE of the sea ice concentration

RMSE improvement compared to persistence of AMSR2 observations

Figure 9. Spatial variability in the performances of MET-AICE using AMSR2 observations as reference. The top row shows the root mean square error (RMSE) of the sea ice concentration in MET-AICE forecasts during the period April 2024 - March 2025. The bottom row shows the relative improvement for the RMSE of the sea ice concentration of MET-AICE forecasts compared to persistence of AMSR2 observations. Positive values (blue) mean that MET-AICE outperforms persistence of AMSR2 observations. The grid points with less than 50 days during which AMSR2 observations indicate a sea ice concentration higher than 0 % are excluded from this analysis.

Fig. 9 shows the spatial variability in the performances of MET-AICE when AMSR2 observations are used as reference. It is worth noting that this figure is the result of forecasts from different seasons, and that the location of the marginal ice zone varies considerably depending on the season. In order to keep only meaningful information, the grid points with ~~less~~ fewer than 50 days during which sea ice is present (SIC higher than 0 %) between April 2024 and March 2025 are excluded from this figure. As expected, the largest errors occur where the marginal ice zone is often located, with the errors growing with increasing lead times. Furthermore, MET-AICE is able to outperform persistence of AMSR2 observations almost everywhere and for all lead times. The relative improvement compared to persistence of AMSR2 observations is larger for long lead times, in particular where the marginal ice zone is often located.

In addition to evaluating the operational MET-AICE deterministic forecasts, we investigated if we could use several ensemble members from ECMWF IFS (ECMWF-ENS) for producing probabilistic SIC forecasts with MET-AICE. We used the first 10 ensemble members of ECMWF-ENS for the atmospheric predictors (10-meter wind and 2-meter temperature) to produce a set of 10 ensemble members with MET-AICE. In Fig. 10, the forecasts starting on Mondays, Wednesdays, and Fridays from April 2024 to March 2025 are evaluated in the shared domain between MET-AICE and Barents-2.5 in order to include the three dynamical models. Overall, MET-AICE and the three dynamical models produce ensemble forecasts that are overconfident (not enough ensemble spread), which means that low sea ice probabilities (SIP) are observed more frequently than predicted whereas high SIP are observed less frequently than predicted. While the MET-AICE forecasts are particularly overconfident for 1-day lead time, they have a similar spread as the TOPAZ5 forecasts for 10-day lead time. Among the prediction systems, ECMWF IFS produces the largest ensemble spread. Furthermore, the probabilistic ice edge distance error is lower in MET-AICE forecasts than in all dynamical models, except for lead times of 9 and 10 days when the ice charts are used as reference.
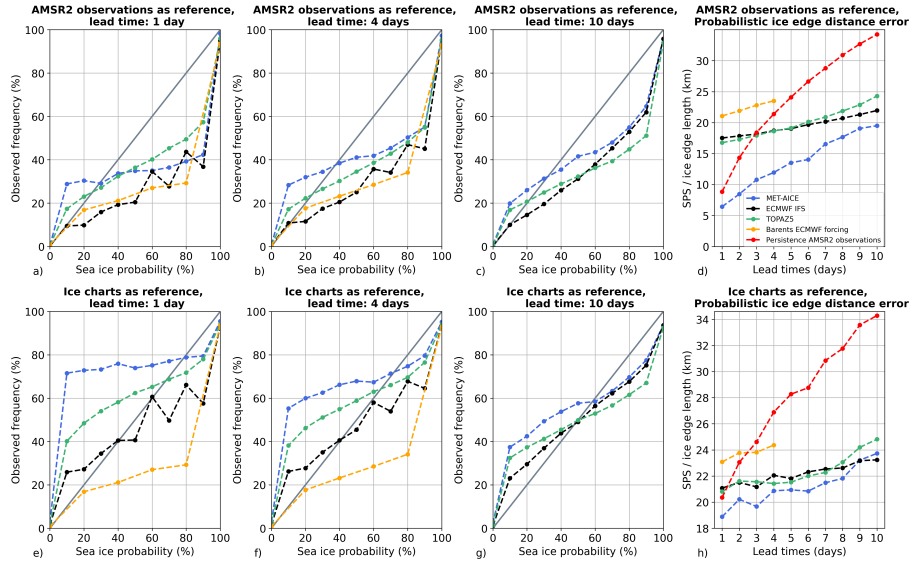
Figure 10. Evaluation of probabilistic sea ice forecasts. The forecasts starting on Mondays, Wednesdays, and Fridays from April 2024 to March 2025 are evaluated in the shared domain between MET-AICE and Barents-2.5. a, b, c) Comparison between the sea ice probability (probability that the sea ice concentration exceeds 10 %) and the observed frequency of occurrence in AMSR2 observations. d) Probabilistic ice edge distance error (ratio of the Spatial Probability Score (SPS) over the observed ice edge length) depending on the lead time using AMSR2 observations as reference. e, f, g) Comparison between the sea ice probability and the observed frequency of occurrence in the ice charts. h) Probabilistic ice edge distance error depending on the lead time using the ice charts as reference.

5 Discussion and conclusion

While many studies have demonstrated that data-driven sea ice forecasts can be skillful and more accurate than dynamical models ~~(e.g. Andersson et al., 2021; Grigoryev et al., 2022; Ren et al., 2022; Chen et al., 2023; ?)~~(e.g. Andersson et al., 2021; Grigoryev et al., , there are currently only a few of these systems running operationally. To our knowledge, MET-AICE is the first operational data-driven sea ice forecasting system focused on short and intermediate time scales (1 to 10 days). The forecasts are produced daily and are publicly available (see code and data availability section). MET-AICE is skillful for predicting AMSR2 SIC observations (that were used for training the deep learning models), and produces more accurate forecasts than the dynamical models evaluated in this study (ECMWF IFS, TOPAZ5, and Barents-2.5~~prediction system~~), even after bias correction. It considerably outperforms persistence of AMSR2 observations for all lead times (the RMSE of the SIC is about 31 % lower on average, and the ice edge position is about 32 % more accurate on average).

~~Nevertheless, passive~~ Passive microwave SIC observations have some limitations such as melt pond ambiguity and land contamination (Kern et al., 2016, 2020; Lavergne et al., 2019). In particular, it was shown by Palerme et al. (2024) that the AMSR2 product used in MET-AICE has larger discrepancies with the ice charts close to the sea ice minimum (in August, September, and October) compared to the rest of the year. Therefore, MET-AICE has lower skill when the ice charts are used

16

as reference and particularly close to the sea ice minimum. We also noticed that MET-AICE sometimes ~~predict~~ predicts too much sea ice along the coastlines, which is likely due to the land contamination of the passive microwave observations. It would be relevant to use other types of satellite SIC observations along the coasts in order to mitigate this issue in the future.

335   Whereas many dynamical sea ice models produce sea ice forecasts with hourly time steps (e.g. Williams et al., 2021; Ponsoni et al., 2023; Röhrs et al., 2023; Paquin et al., 2024), it is challenging to develop data-driven sea ice forecasts with such a high temporal resolution due to the limited number of observations available at this time scale. SIC and drift products with a pan-Arctic coverage are typically made available as daily averaged products (e.g. Tonboe et al., 2016; Lavergne et al., 2019; Tschudi et al., 2020; Wulf et al., 2024), limiting their usefulness for short lead times. One of the current limitations of dynamical sea

340 ice models seems to be their inaccuracy at the initial state despite using data assimilation. ~~In~~ For example, in the case of Barents-2.5, the large initial bias is likely due to the lack of ensemble spread, which mitigates the impact of data assimilation. A relatively simple bias correction at the initial state can greatly improve the accuracy of the forecasts as it was already shown by Palerme et al. (2024). Furthermore, this is also one of the main advantages of data-driven prediction systems which can have low biases if they are trained on a period long enough that captures the vast majority of the situations encountered.

345   The development of MET-AICE will continue in the future. It is planned to add predictions of sea ice drift and sea ice thickness since several studies have already shown the potential of machine learning for predicting these variables (e.g. Palerme and Müller, 2021; Hoffman et al., 2023; Durand et al., 2024; Koo and Rahnemoonfar, 2024; Zaatar et al., 2025). MET-AICE is currently a deterministic prediction system that was developed using the mean squared error as loss function. As a consequence, the forecasts become smoother with increasing lead times. In order to avoid this smoothing, generative artificial intelligence

350 could be used to increase the resolution of the forecasts and create a set of ensemble members. For example, Finn et al. (2024) recently showed that diffusion models can be used to produce probabilistic sea ice forecasts without reducing the effective spatial resolution.

    Furthermore, it is common practice to evaluate machine learning models using verification scores that are strongly correlated with the loss function. This study highlights that this can lead to spurious conclusions if no independent verification score is

355 used in addition. Here, we show that the RMSE of the SIC can be reduced by 8.4 % due to the smoothing of the SIC fields. Therefore, we strongly recommend using more independent verification scores such as the ice edge distance error which does not favor smoother SIC fields as the RMSE of the SIC does.

References

Ali, A., Burud, A., Williams, T., Xie, J., Yumruktepe, C., Wakamatsu, T., Melsom, A., Bertino, L., and Gao, S.: Product User Manual for Arctic Ocean Physical and BGC Analysis and Forecasting Products, Tech. rep., Copernicus Marine Environment Monitoring Service (CMEMS), https://doi.org/10.48670/moi-00001, 2025.

Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., Phillips, T., et al.: Seasonal Arctic sea ice forecasting with probabilistic deep learning, Nature communications, 12, 5124, https://doi.org/10.1038/s41467-021-25257-4, 2021.

Barton, N., Metzger, E. J., Reynolds, C. A., Ruston, B., Rowley, C., Smedstad, O. M., Ridout, J. A., Wallcraft, A., Frolov, S., Hogan, P., Janiga, M. A., Shriver, J. F., McLay, J., Thoppil, P., Huang, A., Crawford, W., Whitcomb, T., Bishop, C. H., Zamudio, L., and Phelps, M.: The Navy's Earth System Prediction Capability: A New Global Coupled Atmosphere-Ocean-Sea Ice Prediction System Designed for Daily to Subseasonal Forecasting, Earth and Space Science, 8, e2020EA001 199, https://doi.org/10.1029/2020EA001199, 2021.

Bleck, R.: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates, Ocean Modelling, 4, 55–88, https://doi.org/10.1016/S1463-5003(01)00012-9, 2002.

Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Hogan, P. J., Wallcraft, A. J., and Bleck, R.: Ocean Prediction with the Hybrid Coordinate Ocean Model (HYCOM), pp. 413–426, Springer Netherlands, Dordrecht, ISBN 978-1-4020-4028-3, https://doi.org/10.1007/1-4020-4028-8_16, 2006.

Chen, X., Valencia, R., Soleymani, A., and Scott, K. A.: Predicting Sea Ice Concentration With Uncertainty Quantification Using Passive Microwave and Reanalysis Data: A Case Study in Baffin Bay, IEEE Transactions on Geoscience and Remote Sensing, 61, 1–13, https://doi.org/10.1109/TGRS.2023.3250164, 2023.

Copeland, W., Wagner, P., Hughes, N., Everett, A., and Robertsen, T.: The MET Norway Ice Service: a comprehensive review of the historical and future evolution, ice chart creation, and end user interaction within METAREA XIX, Frontiers in Marine Science, 11, https://doi.org/10.3389/fmars.2024.1400479, 2024.

Deggim, H.: The International Code for Ships Operating in Polar Waters (Polar Code), pp. 15–35, ISBN 978-3-319-78424-3, https://doi.org/10.1007/978-3-319-78425-0_2, 2018.

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, advanced Along Track Scanning Radiometer (AATSR) Special Issue, 2012.

Durán Moro, M., Sperrevik, A. K., Lavergne, T., Bertino, L., Gusdal, Y., Iversen, S. C., and Rusin, J.: Assimilation of satellite swaths versus daily means of sea ice concentration in a regional coupled ocean–sea ice model, The Cryosphere, 18, 1597–1619, https://doi.org/10.5194/tc-18-1597-2024, 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.

ECMWF: IFS Documentation CY48R1, chap. Part V: Ensemble Prediction Sytem, p. 24 pp, ECMWF, Reading, United Kingdom, https://www.ecmwf.int/en/elibrary/81371-ifs-documentation-cy48r1-part-v-ensemble-prediction-system, 2023.

ECMWF: IFS Documentation CY49R1, chap. Part V: Ensemble Prediction Sytem, p. 24 pp, ECMWF, Reading, United Kingdom, https://www.ecmwf.int/en/elibrary/81627-ifs-documentation-cy49r1-part-v-ensemble-prediction-system, 2024.

Finn, T. S., Durand, C., Farchi, A., Bocquet, M., Rampal, P., and Carrassi, A.: Generative Diffusion for Regional Surrogate Models From Sea-Ice Simulations, Journal of Advances in Modeling Earth Systems, 16, e2024MS004 395, https://doi.org/10.1029/2024MS004395, 2024.

415 Goessling, H. F. and Jung, T.: A probabilistic verification score for contours: Methodology and application to Arctic ice-edge forecasts, Quarterly Journal of the Royal Meteorological Society, 144, 735–743, https://doi.org/10.1002/qj.3242, 2018.

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.

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

Gunnarsson, B.: Recent ship traffic and developing shipping trends on the Northern Sea Route—Policy implications for future arctic shipping, Marine Policy, 124, 104 369, https://doi.org/10.1016/j.marpol.2020.104369, 2021.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and
425 Pattern Recognition (CVPR), pp. 770–778, https://doi.org/10.1109/CVPR.2016.90, 2016.

Hoffman, L., Mazloff, M. R., Gille, S. T., Giglio, D., Bitz, C. M., Heimbach, P., and Matsuyoshi, K.: Machine Learning for Daily Forecasts of Arctic Sea Ice Motion: An Attribution Assessment of Model Predictive Skill, Artificial Intelligence for the Earth Systems, 2, 230 004, https://doi.org/10.1175/AIES-D-23-0004.1, 2023.

Hunke, E., Lipscomb, W., Jones, P., Turner, A., Jeffery, N., and Elliott, S.: CICE, The Los Alamos Sea Ice Model, https://www.osti.gov/
430 biblio/1364126, 2017.

Idžanović, M., Rikardsen, E. S. U., and Röhrs, J.: Forecast uncertainty and ensemble spread in surface currents from a regional ocean model, Frontiers in Marine Science, 10, 1177 337, https://doi.org/10.3389/fmars.2023.1177337, 2023.

JCOMM Expert Team on sea ice: Sea ice information services of the world, Edition 2017, Tech. Rep. WMO-No 574, World Meteorological Organization, Geneva, Switzerland, https://doi.org/10.25607/OBP-1325, 2017.

435 Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M.: SEAS5: the new ECMWF seasonal forecast system, Geoscientific Model Development, 12, 1087–1117, https://doi.org/10.5194/gmd-12-1087-2019, 2019.

Keller, M. R., Piatko, C., Clemens-Sewall, M. V., Eager, R., Foster, K., Gifford, C., Rollend, D., and Sleeman, J.: Short-Term (7 Day) Beaufort Sea Ice Extent Forecasting with Deep Learning, Artificial Intelligence for the Earth Systems, 2, e220 070, https://doi.org/10.
440 1175/AIES-D-22-0070.1, 2023.

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

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

445 Koo, Y. and Rahnemoonfar, M.: Hierarchical Information-Sharing Convolutional Neural Network for the Prediction of Arctic Sea Ice Concentration and Velocity, IEEE Transactions on Geoscience and Remote Sensing, 62, 1–13, https://doi.org/10.1109/tgrs.2024.3501094, 2024.

Kvanum, A. F., Palerme, C., Müller, M., Rabault, J., and Hughes, N.: Developing a deep learning forecasting system for short-term and high-resolution prediction of sea ice concentration, The Cryosphere, 19, 4149–4166, https://doi.org/10.5194/tc-19-4149-2025, 2025.

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

Lin, Y., Yang, Q., Li, X., Yang, C.-Y., Wang, Y., Wang, J., Liu, J., Chen, S., and Liu, J.: Optimization of the k-nearest-neighbors model for summer Arctic Sea ice prediction, Frontiers in Marine Science, Volume 10 - 2023, https://doi.org/10.3389/fmars.2023.1260047, 2023.

455 Lin, Y., Yang, Q., Li, X., Dong, X., Luo, H., Nie, Y., Wang, J., Wang, Y., and Min, C.: Ice-kNN-South: A Lightweight Machine Learning Model for Antarctic Sea Ice Prediction, Journal of Geophysical Research: Machine Learning and Computation, 2, e2024JH000 433, https://doi.org/10.1029/2024JH000433, 2025.

Melsom, A., Palerme, C., and Müller, M.: Validation metrics for ice edge position forecasts, Ocean Science, 15, 615–630, https://doi.org/10.5194/os-15-615-2019, 2019.

460 Mohammadi-Aragh, M., Goessling, H., Losch, M., Hutter, N., and Jung, T.: Predictability of Arctic sea ice on weather time scales, Scientific reports, 8, 6514, https://doi.org/10.1038/s41598-018-24660-0, 2018.

Müller, M., Batrak, Y., Kristiansen, J., Køltzow, M. A. Ø., Noer, G., and Korosov, A.: Characteristics of a Convective-Scale Weather Forecasting System for the European Arctic, Monthly Weather Review, 145, 4771 – 4787, https://doi.org/10.1175/MWR-D-17-0194.1, 2017.

Müller, M., Knol-Kauffman, M., Jeuring, J., and Palerme, C.: Arctic shipping trends during hazardous weather and sea-ice conditions and 465 the Polar Code's effectiveness, npj Ocean Sustainability, 2, https://doi.org/10.1038/s44183-023-00021-x, 2023.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas, arXiv preprint arXiv:1804.03999, https://doi.org/10.48550/arXiv.1804.03999, 2018.

Palerme, C.: Code accompagning the article "MET-AICE v1.0: an operational data-driven sea ice prediction system for the European Arctic", 470 https://doi.org/10.5281/zenodo.17327120, 2025.

Palerme, C. and Müller, M.: Calibration of sea ice drift forecasts using random forest algorithms, The Cryosphere, 15, 3989–4004, https://doi.org/10.5194/tc-15-3989-2021, 2021.

Palerme, C., Müller, M., and Melsom, A.: An Intercomparison of Verification Scores for Evaluating the Sea Ice Edge Position in Seasonal Forecasts, Geophysical Research Letters, 46, 4757–4763, https://doi.org/10.1029/2019GL082482, 2019.

475 Palerme, C., Lavergne, T., Rusin, J., Melsom, A., Brajard, J., Kvanum, A. F., Macdonald Sørensen, A., Bertino, L., and Müller, M.: Improving short-term sea ice concentration forecasts using deep learning, The Cryosphere, 18, 2161–2176, https://doi.org/10.5194/tc-18-2161-2024, 2024.

Paquin, J.-P., Roy, F., Smith, G. C., MacDermid, S., Lei, J., Dupont, F., Lu, Y., Taylor, S., St-Onge-Drouin, S., Blanken, H., et al.: A new high-resolution Coastal Ice-Ocean Prediction System for the east coast of Canada, Ocean Dynamics, 74, 799–826, https://doi.org/10.1007/480 s10236-024-01634-7, 2024.

Ponsoni, L., Ribergaard, M. H., Nielsen-Englyst, P., Wulf, T., Buus-Hinkler, J., Kreiner, M. B., and Rasmussen, T. A. S.: Greenlandic sea ice products with a focus on an updated operational forecast system, Frontiers in Marine Science, 10, 979 782, https://doi.org/10.3389/fmars.2023.979782, 2023.

Ren, Y., Li, X., and Zhang, W.: A Data-Driven Deep Learning Model for Weekly Sea Ice Concentration Prediction of the Pan-Arctic During 485 the Melting Season, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–19, https://doi.org/10.1109/TGRS.2022.3177600, 2022.

Röhrs, J., Gusdal, Y., Rikardsen, E. S. U., Durán Moro, M., Brændshøi, J., Kristensen, N. M., Fritzner, S., Wang, K., Sperrevik, A. K., Idžanović, M., Lavergne, T., Debernard, J. B., and Christensen, K. H.: Barents-2.5km v2.0: an operational data-assimilative coupled

ocean and sea ice ensemble prediction model for the Barents Sea and Svalbard, Geoscientific Model Development, 16, 5401–5426, https://doi.org/10.5194/gmd-16-5401-2023, 2023.

490 Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Rusin, J., Lavergne, T., Doulgeris, A. P., and Scott, K. A.: Resolution enhanced sea ice concentration: a new algorithm applied to AMSR2 microwave radiometry data, Annals of Glaciology, p. 1–12, https://doi.org/10.1017/aog.2024.6, 2024.

Shchepetkin, A. F. and McWilliams, J. C.: The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-

495 following-coordinate oceanic model, Ocean Modelling, 9, 347–404, https://doi.org/10.1016/j.ocemod.2004.08.002, 2005.

Smith, G. C., Roy, F., Reszka, M., Surcel Colan, D., He, Z., Deacu, D., Belanger, J.-M., Skachko, S., Liu, Y., Dupont, F., Lemieux, J.-F., Beaudoin, C., Tranchant, B., Drévillon, M., Garric, G., Testut, C.-E., Lellouche, J.-M., Pellerin, P., Ritchie, H., Lu, Y., Davidson, F., Buehner, M., Caya, A., and Lajoie, M.: Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System, Quarterly Journal of the Royal Meteorological Society, 142, 659–671, https://doi.org/10.1002/qj.2555, 2016.

500 Stocker, A. N., Renner, A. H., and Knol-Kauffman, M.: Sea ice variability and maritime activity around Svalbard in the period 2012–2019, Scientific Reports, 10, 17 043, https://doi.org/10.1038/s41598-020-74064-2, 2020.

Tandon, N. F., Kushner, P. J., Docquier, D., Wettstein, J. J., and Li, C.: Reassessing Sea Ice Drift and Its Relationship to Long-Term Arctic Sea Ice Loss in Coupled Climate Models, Journal of Geophysical Research: Oceans, 123, 4338–4359, https://doi.org/10.1029/2017JC013697, 2018.

505 Tonboe, R. T., Eastwood, S., Lavergne, T., Sørensen, A. M., Rathmann, N., Dybkjær, G., Pedersen, L. T., Høyer, J. L., and Kern, S.: The EUMETSAT sea ice concentration climate data record, The Cryosphere, 10, 2275–2290, https://doi.org/10.5194/tc-10-2275-2016, 2016.

Tschudi, M. A., Meier, W. N., and Stewart, J. S.: An enhancement to sea ice motion and age products at the National Snow and Ice Data Center (NSIDC), The Cryosphere, 14, 1519–1536, https://doi.org/10.5194/tc-14-1519-2020, 2020.

Veland, S., Wagner, P., Bailey, D., Everet, A., Goldstein, M., Hermann, R., Hjort-Larsen, T., Hovelsrud, G., Hughes, N., Kjøl, A., Li, X.,

510 Lynch, A., Müller, M., Olsen, J., Palerme, C., Pedersen, J., Rinaldo, Ø., Stephenson, S., and Storelvmo, T.: Knowledge needs in sea ice forecasting for navigation in Svalbard and the High Arctic, Svalbard Strategic Grant, Svalbard Science Forum. NF-rapport 4/2021, https://doi.org/10.13140/RG.2.2.11169.33129, 2021.

von Schuckmann, K., Le Traon, P.-Y., Smith, N., Pascual, A., Djavidnia, S., Gattuso, J.-P., Grégoire, M., Aaboe, S., Alari, V., Alexander, B. E., Alonso-Martirena, A., Aydogdu, A., J., A., Bajo, M., Barbariol, F., Batistić, M., Behrens, A., Ben Ismail, S., Benetazzo, A., Bitetto,

515 I., Borghini, M., Bray, L., Capet, A., Carlucci, R., Chatterjee, S., Chiggiato, J., Ciliberti, S., Cipriano, G., Clementi, E., Cochrane, P., Cossarini, G., D'Andrea, L., Davison, S., Down, E., Drago, A., Druon, J., Engelhard, G., Federico, I., Garić, R., Gauci, A., Gerin, R., Geyer, G., Giesen, R., Good, S., Graham, R., Greiner, E., Gundersen, K., Hélaouët, P., Hendricks, S., Heymans, J. J., Holt, J., Hure, M., Juza, M., Kassis, D., Kellett, P., Knol-Kauffman, M., Kountouris, P., Kõuts, M., Lagemaa, P., Lavergne, T., Legeais, J.-F., Libralato, S., Lien, V. S., Lima, L., Lind, S., Liu, Y., Macías, D., Maljutenko, I., Mangin, A., Männik, A., Marinova, V., Martellucci, R., Masnadi,

520 F., Mauri, E., Mayer, M., Menna, M., Meulders, C., Møgster, J. S., Monier, M., Mork, K. A., Müller, M., Nilsen, J. E. O., Notarstefano, G., Oviedo, J. L., Palerme, C., Palialexis, A., Panzeri, D., Pardo, S., Peneva, E., Pezzutto, P., Pirro, A., Platt, T., Poulain, P. M., Prieto, L., Querin, S., Rabenstein, L., Raj, R. P., Raudsepp, U., Reale, M., Renshaw, R., Ricchi, A., Ricker, R., Rikka, S., Ruiz, J., Russo, T., Sanchez, J., Santoleri, R., Sathyendranath, S., Scarcella, G., Schroeder, K., Sparnocchia, S., Spedicato, M. T., Stanev, E., Staneva, J., Stocker, A., Stoffelen, A., Teruzzi, A., Townhill, B., Uiboupin, R., Valcheva, N., Vandenbulcke, L., Vindenes, H., Vrgoč, N.,

525   Wakelin, S., and Zupa, W.: Copernicus Marine Service Ocean State Report, Issue 5, Journal of Operational Oceanography, 14, 1–185, https://doi.org/10.1080/1755876X.2021.1946240, 2021.

Wagner, P. M., Hughes, N., Bourbonnais, P., Stroeve, J., Rabenstein, L., Bhatt, U., Little, J., Wiggins, H., and Fleming, A.: Sea-ice information and forecast needs for industry maritime stakeholders, Polar Geography, 43, 160–187, https://doi.org/10.1080/1088937X.2020.1766592, 2020.

530   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.

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

Yu, X., Rinke, A., Dorn, W., Spreen, G., Lüpkes, C., Sumata, H., and Gryanik, V. M.: Evaluation of Arctic sea ice drift and its dependency
535   on near-surface wind and sea ice conditions in the coupled regional climate model HIRHAM–NAOSIM, The Cryosphere, 14, 1727–1746, https://doi.org/10.5194/tc-14-1727-2020, 2020.

Yumruktepe, V., Samuelsen, A., and Daewel, U.: ECOSMO II(CHL): a marine biogeochemical model for the North Atlantic and the Arctic, Geoscientific Model Development, 15, 3901–3921, https://doi.org/10.5194/gmd-15-3901-2022, 2022.

Zaatar, T., Cheaitou, A., Faury, O., and Rigot-Muller, P.: Arctic sea ice thickness prediction using machine learning: a long short-term
540   memory model, Annals of Operations Research, pp. 1–36, https://doi.org/10.1007/s10479-024-06457-9, 2025.

Zuo, H., Balmaseda, M. A., Mogensen, K., and Tietsche, S.: OCEAN5: the ECMWF ocean reanalysis system and its real-time analysis component, Tech. rep., https://www.ecmwf.int/sites/default/files/elibrary/2018/18519-ocean5-ecmwf-ocean-renalysis-system-and-its-real-time-analysis-component.pdf, 2018.