# Assessment of high-resolution dynamical and machine learning models for prediction of sea-ice concentration in a regional application

**Sindre Fritzner**[1], **Rune Graversen**[1,2], **Kai H. Christensen**[3,4]

[1]UiT The Arctic University of Norway, Tromsø, Norway
[2]The Norwegian Meteorological Institute, Norway
[3]The Norwegian Meteorological Institute, Oslo, Norway
[4]The University of Oslo, Oslo, Norway

**Key Points:**

- Both dynamical and machine learning methods are applied for sea-ice modelling
- We demonstrate the potential of machine learning in sea-ice forecasting
- The dynamical model utilises data assimilation of high-resolution sea-ice concentration and sea-surface temperature satellite observations

Corresponding author: Sindre Fritzner, `sindre.m.fritzner@uit.no`

## Abstract

In this study, the potential for sea-ice concentration prediction using machine-learning methods is investigated. Three different sea-ice prediction models are compared: one high-resolution dynamical assimilative model and two statistical machine-learning models. The properties of all three models are explored, and the quality of their forecasts is compared. The dynamical model is a state-of-the-art coupled ocean and sea-ice ensemble-prediction system with assimilation. The observations assimilated are high-resolution sea-ice concentration from synthetic aperture radar (SAR) and sea-surface temperature from infrared instruments. The machine learning prediction models are a fully convolutional network and a $k$-Nearest Neighbours method. These methods use several variables as input for the prediction: sea-ice concentration, sea-surface temperature and 2-m air-temperature. Earlier studies have applied machine-learning approaches primarily for seasonal ice forecast. Here we focus on short term predictions with a length of 1-4 weeks which are of high interest for marine operations. The goal is to predict the future state of the sea ice using the same categories as traditional ice charts. The machine-learning forecasts where compared to persistence, which is the assumption that the sea-ice does not change over the forecasting period. The machine-learning forecasts where found to improve upon persistence in periods of substantial change. In addition, compared to the dynamical model, the k-Nearest Neighbour algorithm was found to improve upon the 7-day forecast during a period of small sea-ice variations. The fully convolutional network provided similar quality as the dynamical forecast. The study shows that there is a potential for sea-ice predictions using machine-learning methods.



## Plain Language Summary

This study investigates the use of statistically-based models and compares them to a physically-based model for sea-ice prediction. The physical model uses assimilation of observations to improve the forecast. When substantial changes in the sea ice are observed, the machine learning models show skilful forecasts compared to assuming that the sea ice does not change during the forecasting period (persistence). A comparison between the dynamical and statistical forecast shows that the statistical model may be a simple alternative to the physical model during periods of small variations in the sea-ice extent.


## 1 Introduction

Since the start of the satellite era about 40 years ago, there has been a decline in the Arctic sea-ice extent. Cavalieri og Parkinson (2012) reported that during 1970-2010, the Arctic sea-ice extent declined by on average 4% per decade. The decline has continued so that the last 12 years have been those with the lowest sea-ice minimums recorded during the 40-year satellite period (Scott, 2018). The decrease in sea-ice extent has led to an increase of marine operations in the Arctic from several industries such as shipping, tourism, fishing, and oil and gas exploration (Stephen, 2018). As the Arctic sea-ice continues to melt, and the Arctic becomes more accessible for marine operations, the human presence in this region will likely increase further. The recent sea ice decline is much smaller than the seasonal variations, which is one of the main challenges for the operators close to the ice edge. Therefore, safety requirements for future marine operations close to the ice edge demands for accurate sea-ice predictions with a high spatial and temporal resolution both for daily and seasonal predictions.

Since the 1950s there has been a continuous development of sea-ice models, *e.g.* The Los Alamos Community sea-ice model (CICE; Hunke & Dukowicz, 1997), Louvain-la-Neuve sea-ice model (LIM3; Vancoppenolle et al., 2009) and coupled ocean-sea-ice models such as Towards an Operational Prediction system for the North Atlantic coastal Zones (TOPAZ; Sakov et al., 2012) and the Massachusetts Institute of Technology Ge-

neral Circulation Model (MITgcm; Marshall et al., 1997) to mention a few. Numerous sea-ice forecast studies in the Arctic have been performed using these models *e.g.* (Caya et al., 2010; K. Wang et al., 2013; Sakov et al., 2012; Buehner et al., 2013; Yang et al., 2014; Posey et al., 2015; Shlyaeva et al., 2016; Xie et al., 2016; Mu et al., 2018; Fritzner et al., 2018, 2019). Common for many of the Arctic sea-ice models used in these studies is that the model resolution is typically coarse, on the order of 10-20 km. Coarse-resolution models are often satisfactory for climate studies on a global scale, but less useful for maritime operations where detailed forecasts of the sea ice are important.

In a forecasting system, the initial state of the model forecast is essential. When observations are available, the best possible initial state can be achieved through data assimilation. Satellite-based passive microwave observations of sea-ice concentration have been available for the last 40 years, and several studies have investigated how the assimilation of these observations impacts the models, *e.g.* (Lisæter et al., 2003; Sakov et al., 2012; K. Wang et al., 2013; Buehner et al., 2013; Posey et al., 2015; Fritzner et al., 2018, 2019). Sea-ice concentration (SIC) is by far the most used variable in sea-ice data assimilation studies, however other types of observations have become available in recent years. In the last decade, there have been efforts to extract more sea-ice information from satellites, and now observational products of sea-ice thickness (Kurtz & Harbeck, 2017; Tian-Kunze et al., 2016; Ricker et al., 2017), sea-ice drift (Ninnis et al., 1986; Lavergne et al., 2010; Kræmer et al., 2015), and snow depth (Rostosky et al., 2018) are available. The SIC products derived from passive microwave are generally provided with a resolution of 10-25 km depending on the instrument, method and measurement frequencies used. Deriving SIC from the brightness temperatures observed by the passive microwave instruments can be done in several different ways with various benefits and uncertainties (Andersen et al., 2007).

The steady increase in computing power is facilitating for more complex numerical models with higher spatial and temporal resolutions. High-resolution observations of sea-ice are available through active microwave measurements such as those of Synthetic Aperture Radars (SAR), *e.g.* onboard the European Space Agency's (ESA) Sentinel constellation which consists of two SAR satellites, Sentinel 1a and b. In the Arctic, the Sentinel-1 satellites use extra-wide swath, acquiring measurements with a resolution of about 20x40 m covering a width of approximately 400 km (Torres et al., 2012). This resolution provides detailed information regarding the sea-ice-edge variability with a higher spatial resolution compared to that obtained from passive microwaves. Currently, manual products based on SAR observations, such as the ice charts provided by the Norwegian Meteorological Institute (MET Norway) for the area around Svalbard (`http://wms.met.no/icechart/`), are produced. This operational product consists of hand-drawn maps combining several different sea-ice retrievals such as SAR, passive microwave and optical instruments, into a high-resolution (1 km grid spacing) SIC product. As far as is known to the authors, there is not yet any operational high-resolution automatic Arctic sea-ice maps from SAR.

In recent years some high-resolution sea-ice assimilation studies have been performed using the 3-D variational method for data assimilation (Buehner et al., 2013; Posey et al., 2015). Posey et al. (2015) investigated the effect of assimilating sea-ice concentration observations with a resolution of 4 km into a coupled model with an approximate resolution at the North pole of 3.5 km. In their study, a blended sea-ice concentration product with data from AMSR-2 and the Interactive Multisensor Snow and Ice Mapping System (Helfrich et al., 2007) was applied. These observations were assimilated into the Arctic Cap Nowcast/Forecast system produced by the US Navy (Metzger et al., 2014). By assimilating the high-resolution observations (4 km), a smaller ice-edge error was obtained compared to assimilating coarser (25 km) resolution observations. Buehner et al. (2013) provided an alternative method for high-resolution sea-ice forecasting without applying a dynamical model. In that study, several types of sea-ice observations we-

re assimilated to provide a 5 km sea-ice concentration analysis every six hours. Their results showed an improved sea-ice extent compared to the Canadian Meteorological Centre global ice analysis.

The Arctic SIC is assumed to be strongly related to the upper ocean temperature, the past sea-ice concentration, and the ice-edge location. In addition, for a large portion of the year, the day-to-day sea-ice variations on a general model scale (1-20 km) are small, and for these periods persistence (assuming no change) provides a sufficiently accurate forecast. Therefore, instead of a dynamical model, a prediction could potentially be performed with statistical-based models. A forecast from such a method may likely be performed with a significantly lower computational cost than a dynamical model system. In recent years several alternatives to dynamical models using different methods for sea-ice prediction have been introduced. L. Wang et al. (2019) applied the vector auto-regressive model and the vector Markov model for sea-ice prediction. For sub-seasonal predictions the Markov model provided the best result by more effectively capturing the underlying sub-seasonal dynamics. An ensemble method taking into account nonlinearities was applied by Comeau et al. (2019). With this ensemble method, they found improvements compared to a damped persistence forecast of sea-ice area and volume in the Arctic. More advanced statistical methods include machine-learning methods that use historical data for model training. J. Kim et al. (2019) proposed a deep neural network for sea-ice prediction for the forthcomming 10 to 20 years. Y.J. Kim et al. (2020) proposed a convolutional neural network for 1-month predictions. They compared the model to both a random-forest-based model and persistence. The neural network was found to improve upon both. Chi og Kim (2017) used a deep neural network to perform one and two-month forecasts of the Arctic sea-ice based on past observations of monthly observed SIC. The September sea-ice extent was found to be reasonably well predicted compared to an average of the dynamical model forecasts submitted to the Sea Ice Prediction Network (SIPN) (https://www.arcus.org/sipn). Compared to the previous studies discussed above primarily focusing of seasonal forecast, in this study short term predictions (1-4 weeks) are applied. These are of interest for marine operations in the Arctic.

In the present study, the use of machine-learning prediction models for sea-ice concentration forecasts is investigated and compared to a dynamical model. However, as none of these model setups have previously been described in the literature they are investigated individually before they are compared. Therefore this study consists of three parts, one where the assimilation system is investigated, one where the machine learning predictions are investigated, and in the last part the two methods are compared for weekly SIC prediction.

With the ice charts described above, there now exist daily high-resolution observational products of SIC in the Arctic. In addition, there exist high-resolution observations of sea-surface temperature from the Multi-scale Ultra-high Resolution (MUR) product (Chin et al., 2017). It has previously been shown that by using high-resolution models, the assimilation of high-resolution observations improves the results (Buehner et al., 2013; Posey et al., 2015). Both of these previous studies applied the 3-D variational method for data assimilation. In this study, the Ensemble Kalman Filter (EnKF) is applied for assimilation. With the EnKF, the model covariance matrix is continuously updated for multivariate assimilation, and the ensemble provides a probabilistic forecast. Besides, we apply a higher model and observational resolution compared to the previous studies (Posey et al., 2015; Buehner et al., 2013).

The machine-learning model part of this study builds upon previous studies that applied machine learning for sea-ice forecasting (Chi & Kim, 2017; Y.J. Kim et al., 2020; J. Kim et al., 2019). However, in this study, we apply both a fully convolutional network method and a k-nearest neighbour method for prediction. In addition, our models use input from two SIC products, an sea-surface temperature (SST) product and a 2-m air-temperature (T2) product. We hypothesise that the use of SST observations and T2 re-

analysis in addition to SIC observations will have a positive impact on the forecast skill of the machine-learning methods, as these represent two of the main drivers of the sea-ice variations. Finally, we compare the machine learning models with a dynamical model to see whether machine learning can provide an alternative to complex and computationally costly dynamical models.

In section 2, the dynamical and the two machine learning models are presented. In section 3, the observations used for verification and assimilation are introduced. This includes several observational products for sea-ice concentration and sea-surface temperature. In section 4, the setup of the model experiments are described, both for the machine-learning experiments and the experiments with the dynamical model. In section 5, the assimilation system of the dynamical model is investigated. In section 6, machine learning models are investigated and tested for different forecast lengths. In section 7, the two machine learning methods are compared to the dynamical model. And finally, in section 8, a summary and a conclusion are presented.

## 2 The models

### 2.1 The dynamical model

A coupled ocean and sea-ice model (Kristensen et al., 2017) with a horizontal resolution of 2.5 km is used. This model is similar to that applied in Fritzner et al. (2019). However, here a high resolution regionally downscaled version covering the ice infested areas in the Barents Sea, Greenland Sea and the Kara Sea is used (grid size: 739x949). An overview of the model domain is shown in Figure 4. The study area is chosen such that it covers the ice edge around Svalbard which is the most important for Norwegian marine operations in the Arctic.

The model ocean component is the Regional Ocean Modeling System (ROMS; Shchepetkin & McWilliams, 2005) version 3.6, and the sea-ice component is the Community sea-ice model (CICE; Hunke & Dukowicz, 1997) version 5.1.2. The ocean component has 42 terrain-following sigma layers, and a second-order turbulence closure model is used to parametrise the eddy diffusivity and viscosity. The sea-ice component uses a mechanical redistribution scheme with five ice-thickness categories, seven ice layers and a single snow layer. This state-of-the-art model includes both melt pond and ridging parameterisations, as well as models for thermodynamics, ice dynamics, and transport.

The dynamical model framework includes an ensemble prediction system (EPS) with ten ensemble members and EnKF assimilation every seven days. The ensemble model system is forced by an ensemble of low-resolution (18 km) atmospheric forecasts from the European Centre for Medium Ranged Weather Forecast (ECMWF; Owens & Hewson, 2018) Integrated Forecast System (IFS). The ocean boundary conditions are based on an ensemble from the TOPAZ4 model (Sakov et al., 2012). Generating the ensemble from ensemble forcing is a preferable alternative to *ad-hoc* forcing perturbations, as the ensemble forcing input already contains a well-established and tested method for ensemble generation. The TOPAZ4 forcing data are available from 2018-03-15 to 2018-05-15.

The dynamical model does not include nesting of ice boundary conditions, only ocean boundary conditions. The lack of an ice boundary leads to errors along the northern and western boundaries due to ice transport. In this study, we avoid these boundary problems by omitting the first 15 edge grid cells on the northern and western boundaries for verification. This study primarily focuses on the sea-ice edge location; thus, the results are not effected by the lack of ice boundary conditions. For brevity, the dynamical model will in the rest of this work be referred to as Metroms (Kristensen et al., 2017).

### 2.2 The Ensemble Kalman Filter

In Metroms, we use the Deterministic Ensemble Kalman Filter (DEnKF; Sakov & Oke, 2008) for assimilation; the same setup was also used in Fritzner et al. (2019). The DEnKF is a version of the Ensemble Kalman Filter (EnKF; Evensen, 1994; Burgers et al., 1998) which has been applied in a wide range of geophysical models (Houtekamer & Zhang, 2016). In contrast to the traditional EnKF, the DEnKF is not dependent on perturbation of observations to maintain ensemble spread. Perturbing observations introduces additional sampling error in the analysis, which for applications with few ensemble members might be significant (Sakov & Oke, 2008; Whitaker & Hamill, 2002).

The standard analysis equation solved by the EnKF is given by (Jazwinski, 1970; Evensen, 2003):

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{P}_b\mathbf{H}^T \left( \mathbf{H}\mathbf{P}_b\mathbf{H}^T + \mathbf{R}\right)^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}_b), \tag{1}$$

where $\mathbf{x}_a \in \mathbb{R}^{n\times N}$ is the analysis vector representing the updated variables after assimilation, $\mathbf{x}_b \in \mathbb{R}^{n\times N}$ the model first guess (background), and $\mathbf{y} \in \mathbb{R}^{m\times N}$ is the observation vector. $N$ is the number of ensemble members, $n$ the number of variables multiplied by the number of spatial grid points in our model, $m$ the total number of observations of all variables, $\mathbf{R} \in \mathbb{R}^{m\times m}$ the observation covariance, and $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the observation operator. The key property of the EnKF is that the background error covariance matrix $\mathbf{P_b} \in \mathbb{R}^{n\times n}$, providing the model uncertainty, is estimated as the variance of the ensemble of background states,

$$\mathbf{P_b} = \overline{((\mathbf{x}_b - \overline{\mathbf{x}_b})(\mathbf{x}_b - \overline{\mathbf{x}_b})^T)}. \tag{2}$$

In the equation above, the overbars signify the average operator. The implementation of the assimilation is done offline with the use of the enkf-c software package (Sakov, 2015).

### 2.3 Machine learning methods

The growing field of machine learning includes numerous approaches ranging from simple, transparent methods such as those based on regression to more sophisticated variants based on, for instance, deep neural networks. In this work, a straightforward approach, the $k$-Nearest Neighbours ($k$-NN), and a deep neural network, a fully convolutional network (FCN), is applied for sea-ice prediction. These methods have traditionally been used for image segmentation, where an image is separated into different classes based on pixel properties. A classification is, for example, that a pixel is a part of a car. Then this pixels' class is "car". Other classification can be for example cars, persons, dogs and bicycles. In this study, the SIC intervals defined by the World Meteorological Organization (WMO) total concentration standard (table 1) used by the ice charts are the output classes, while the input is sea-ice related variables. Because the machine-learning methods applied are an attempt at predicting the future ice state as defined by the ice charts, the area covered by the Norwegian ice charts is used as the study area for these methods. In comparison the dynamical model study area is a sub area of the ice chart area.

Both the $k$-NN and FCN are supervised methods. This means that they are dependent on labelled training data, containing input-output pairs. During a machine-learning-training process, the methods apply the labelled training data to learn functions that map the input to output. After training, the models can be used on new input data, for example for sea-ice prediction. In this study, the $k$-NN method was chosen both because of its theoretical simplicity and its ease of implementation. As mentioned, this is a supervised method, however, no training process is needed.

**Table 1.** The WMO Total concentration standard

| Concentration | description | value in ice chart |
|---|---|---|
| 0 | Ice free | 0 |
| $<$ 10 % | Open water | 0.05 |
| 10-30% | very open ice | 0.2 |
| 40-60% | open ice | 0.5 |
| 70-80% | close ice | 0.75 |
| 90-100% | very close ice | 0.95 |
| 100% | Fast ice | 1.00 |

In contrast, the more intricate FCN is a deep neural network with many layers that requires extensive training. Deep learning methods have received much attention in recent years due to several beneficial properties when it comes to image processing, *e.g.* learning of intricate patterns and features (Guo et al., 2018). In general, a prediction performed by the trained FCN model is significantly faster than a prediction with the $k$-NN model. However, the one time cost of the training process can be substantial. Since both machine-learning methods are based on relatively simple relations and do not require a small time step for stable solutions, they are both, generally, computationally less costly than a dynamical model. Another essential difference between the two machine-learning methods applied is that the $k$-NN does not incorporate a spatial context in the prediction.

#### *2.3.1 k-Nearest Neighbours*

The $k$-Nearest Neighbours ($k$-NN) classifier is a supervised machine-learning method (Cover et al., 1967), where labelled data are required. However, no training procedure as such is necessary since the training data are used as a reference dataset only. For each prediction, the input variables are compared to the input of the training dataset based on a distance. The prediction is obtained from the classification of the $k$ nearest training samples. In this study, the Euclidean distance, $d$, is used to find the nearest samples in the training data,

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}, \tag{3}$$

where $\boldsymbol{x}$ is the model input vector, $\boldsymbol{y}$ the input vector of a single training data sample and $n$ is the number of input variables. Thus, for each pixel, the input variables are compared to those of the training dataset, and the prediction is the median class of the $k$ nearest neighbours (lowest $d(\boldsymbol{x}, \boldsymbol{y})$). The input variables in this case of SIC prediction are the sea-ice related variables, SIC, SST and 2-m temperature.

#### *2.3.2 Fully convolutional network*

The FCN is based on the work done by (Long et al., 2015; Shelhamer et al., 2017). This method is a particular type of a neural network that is widely used to address segmentation tasks. In an artificial neural network, a hierarchy of transformations structured in multiple layers is used, where the transformations are parametrised by a set of weights that are learned from data. As mentioned, the FCN is a supervised learning method dependent on labelled (input with known output) training data. The FCN uses a hierarchy of layers (transformations) that perform convolution, pooling and upsampling operations, where the convolutional and upsampling layers consist of learnableparameters. Convolutional layers are further followed by non-linear activation functions.

In a convolutional layer, the input data are convolved with several filters to extract important image features such as edges, vertical lines, horizontal lines and others (Goodfellow et al., 2016). In a pooling layer, the outliers (max pooling) from the convolutional layer output is extracted. A pooling operation provides a larger field of view and improves computational efficiency. In the upsampling layer, the convolved and pooled features are deconvolved for pixel-wise prediction. During the training process of the FCN, the convolutional matrices used in the convolutional and upsampling layers are "learned" based on the labelled training data. More information regarding the individual layers of the FCN can be found in the Appendix :.

In this study, we will use the FCN8 network (Long et al., 2015), and the implementation of the FCN8 in Python with the "Keras" software package (Gupta, 2019; Yumi, 2018). In an FCN with multiple layers of convolution and pooling, the output resolution is in general significantly reduced compared to the input. However, the FCN8 method combines low-resolution deep and high-resolution shallow layers by using so-called skip layers (Long et al., 2015). This combination improves the output resolution, which is essential for the application in the present study. Further information regarding the individual layers of the FCN8 method can be found in the original work (Long et al., 2015).

## 3 Observations

In this study, observations of SIC and Sea-Surface Temperature (SST) are used for machine learning, assimilation and verification. A list of the different observations used and how they are applied is presented in table 2. Three different types of SIC products are used: OSISAF SSMIS, ice charts and OSISAF AMSR-2. These products are based on different observations and they are provided with different resolutions. One product was used for verification only, while the other two provided high- and low-resolution input to the prediction system. The OSISAF SSMIS observations are the Global Sea Ice Concentration product from the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea ice Satellite Application Facility (OSISAF, www.osi-saf.org). In this product, the sea-ice concentration is derived from brightness temperatures measured by the Special Sensor Microwave Imager Sounder (SSMIS; Tonboe et al., 2016), which is a passive microwave instrument. The conversion from brightness temperatures to SIC is done based on a combination of the Bootstrap and the Bristol algorithms (Tonboe et al., 2016). The OSISAF observations include an accompanying uncertainty estimate which is used during the assimilation. The observations are provided on a 10 km grid.

The OSISAF AMSR-2 SIC observations are derived from brightness temperature measurements from the National Aeronautics and Space Administration (NASA) Advances Microwave Scanning Radiometer 2 (AMSR-2) provided on a 10 km grid (Lavelle et al., 2016). This is also a passive microwave instrument. The conversion from brightness temperature to SIC observations is done in the same way as for the SSMIS data. All observations include an estimation of the observation uncertainty (Tonboe et al., 2016).

The ice charts are manually-drawn operational SIC maps provided by MET Norway. The ice charts are based on Synthetic Aperture Radar (SAR) data from Sentinel-1, Radarsat and Envisat, as well as visual and infrared data from MODIS, NOAA and VIIRS. In addition, low-resolution passive microwave observations are used to provide full spatial coverage. This operational product is provided on a 1 km grid (Dinessen & Hackett, 2016). The concentrations in the ice charts are according to the WMO defined total concentration intervals (Table 1). The ice charts do not include an uncertainty estimate for the observations; instead, two times the size of the WMO intervals were chosen as a conservative estimate for the observation uncertainty. Note that the Norwegian ice charts are only available during weekdays, thus to avoid frequent data gaps in our

**Table 2.** Observations used by the forecasting systems and for verification. OSISAF AMSR-2 and OSTIA are used for independent verification, while the other observations are both used for verification and by the forecasting system.

| Name | Type | Assimilated in dynamical model | Verification | Training data for ML |
|---|---|---|---|---|
| OSISAF SSMIS | SIC | yes | yes | yes |
| Ice charts | SIC | yes | yes | yes |
| OSISAF AMSR-2 | SIC | no | yes | no |
| MUR | SST | yes | yes | yes |
| OSTIA | SST | no | yes | no |

dataset, the Friday ice chart is applied for Saturday and Sunday. Other missing days in the dataset are replaced by the previous observed day.

In addition to the SIC observations, two SST observational products are included for assimilation and verification: The Multi-scale Ultra-high Resolution (MUR) product (Chin et al., 2017) and The Operational Sea Surface Temperature and Sea Ice Analysis product (OSTIA; C.J. Donlon et al., 2012). These data products are based on the Group for High-Resolution SST (GHRSST; C.J. Donlon et al., 2009) framework for SST measurements and include SST observations from infrared instruments, microwave instruments, and *in situ* measurements. High-resolution SST observations are observed with the infrared sensors, while microwave observations provide all-weather capabilities to achieve full global coverage. Infrared measurements are profoundly affected by diurnal heating from the sun, and therefore mostly night time measurements are used to derive the SST products. The OSTIA dataset is provided with a resolution of 0.05°x0.05°, while the MUR dataset comes with a resolution of 0.01°x0.01°. The improved resolution in the MUR dataset comes from the inclusion of high-resolution observations from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors, which provide SST observations with 1 km resolution (Chin et al., 2017). Similar to the OSISAF products, the SST products include an uncertainty estimation. As the MUR product consists of several sources of observations, the total uncertainty is a combination of these. For the products from the GHRSST, there is a requirement for the uncertainty estimations C. Donlon et al. (2007), while for those not from the GHRSST a best guess was applied (Chin et al., 2017).

## 4 Methods and model setup

### 4.1 The dynamical model

The quality of the Metroms assimilation system is assessed by several assimilation experiments. These are all started on 20.03.2018 based on an ensemble output from the TOPAZ4 coupled ocean-sea-ice assimilation system (Sakov et al., 2012). The experiments are run for eight weeks, until 15.05.2018, with assimilation every seven days.

Tuning of the assimilation system is performed by two ensemble data assimilation analysis tools: the degrees of freedom for signal (DFS; Cardinali et al., 2004; Sakov et al., 2012) and the spread reduction factor (SRF; Sakov et al., 2012). Tuning is essential in order to avoid ensemble collapse, which occurs when the ensemble spread is reduced too rapidly. More information on ensemble collapse is given in (Sakov & Oke, 2008). The DFS is used to identify potential model rank problems related to an ensemble size which is much smaller than the number of observations in the assimilation system. Without changing the ensemble size, the model rank can be improved, and the DFS decreased by reducing the number of observations used during the assimilation analysis.

**Table 3.** Observation localisation and R-factor.

| Name | Localisation radius | R-factor |
|---|---|---|
| OSISAF SSMIS SIC | 40 | 1.5 |
| Ice chart SIC | 25 | 60 |
| MUR SST | 25 | 70 |

In this study, a local assimilation analysis is performed where the assimilation analysis is calculated for each grid cell individually (Sakov & Bertino, 2011; Houtekamer & Mitchell, 2001). The local analysis uses only local observations within a certain localisation radius of the appropriate grid cell. Thus, the DFS can be changed by varying the localisation radius, effectively changing the number of observations included in the analysis.

The SRF gives a measure of the observation impact on the model during assimilation. More specifically, for the DEnKF, this metric describes the ensemble spread reduction during the assimilation analysis. This metric can be changed by changing the observation impact. For the enkf-c software used in this study, an $R$-factor can be tuned to specify the assimilation impact for each observation. The $R$-factor is defined to be a multiplication factor to the observation error covariance matrix, $R$, defined in Eq. (1). An increased $R$-factor lead to an increased observation variance and a lower effect of the observations in the assimilation analysis, which again gives a reduced SRF.

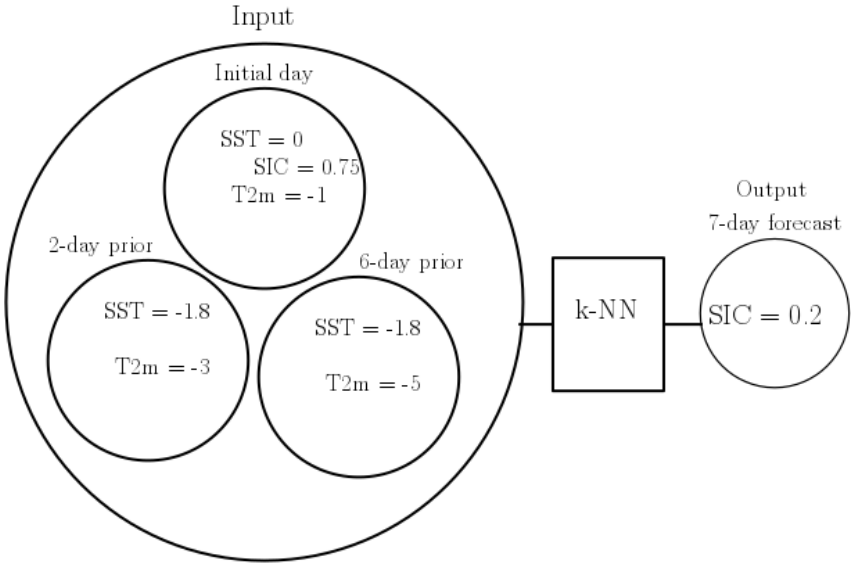
A DFS less than the number of ensemble members divided by three (3.33), and an SRF less than two were used for tuning (Personal correspondence with Pavel Sakov; Sakov et al., 2012). A summary of the individual $R$-factor and localisation radii used for each observation type are given in table 3. The individual $R$-factor values are dependent on both the observation resolution (observation density) and the magnitude of the uncertainties.

The dynamical model has a significantly higher spatial resolution than the OSISAF SSMIS SIC observations. Due to this sizeable spatial difference, the assimilation can lead to a reduced model resolution. In order to avoid this effect, a dummy SIC variable is used in Metroms during the assimilation of low-resolution observations. This dummy variable has the same resolution as the OSISAF SSMIS SIC observations. Based on the analysis update of the dummy variable, the actual model SIC is updated based on the background error covariances.

## 4.2 Machine learning models

### *4.2.1 $k$-Nearest Neighbours*

The $k$-NN model may become impractical if the training data set is too large, depending on the available computational resources. In our case, using a laptop computer, a good compromise between speed and efficiency is obtained when the method is trained on data spanning the year of 2016 only, covering 16,000 randomly chosen grid points primarily in and around the sea-ice edge, which is the only location where SIC is significantly different from 0 and 1. From the full training dataset, 20 % of the data were used for verification only. Recall that for the $k$-NN, $k$ signifies the number of neighbours in the training data used for the prediction. In this study, a $k$ of 15 with uniform weighting is applied, which means that the prediction is the median of the class of the 15 nearest neighbours. The $k$=15, was chosen based on experiments with different values where 15 was found to give the results with lowest errors compared to the verification data. The input data used to calculate the Euclidean distance for the $k$-NN forecast are the initial day ice chart SIC, MUR SST, T2, 2-day prior SST and T2, and 6-day prior SST




**Figur 1.** An example of $k$-NN input and output. The $k$-NN finds the $k$-nearest neighbours in the training data and the output is the median output class of these training data.

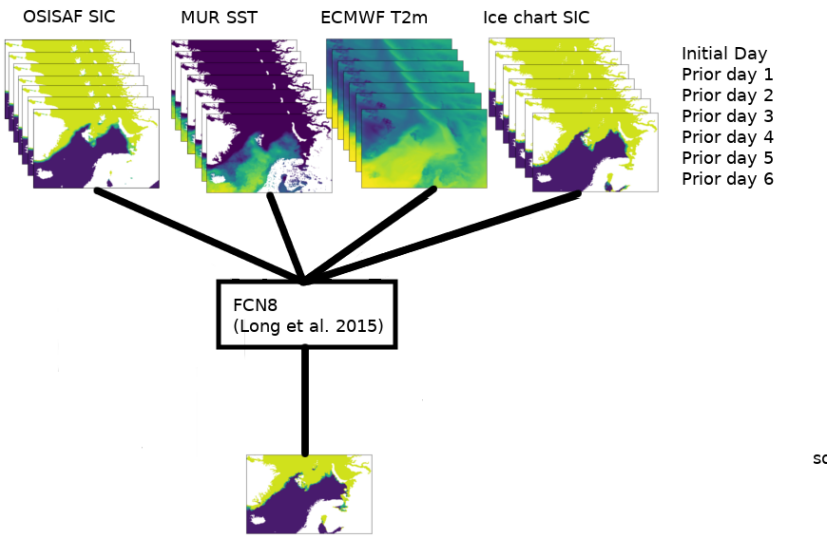
and T2. Input from several days where tested (2-6), but to limit the training dataset only the 2-day (short term information) and 6-day (long term information) observations were used. The T2 observations are from the ECMWF ERA 5 dataset (Copernicus, 2019).

The $k$-NN forecasts become more computer intensive when more input data sources are included. Therefore simple tests were carried out to select the most important input data from the 6-day prior variables. The input variables were chosen based on a combination of best performance and data availability. The idea behind the machine learning prediction is to predict future ice distribution, presented in the same way as the ice charts: WMO total concentration standard for ice classification. A description of the input and output of an example $k$-NN prediction is given in Figure 1. The $k$-NN selects the 15 nearest data points in the training data, and the output is defined as the median over the output classes of these 15 training data points.

#### *4.2.2 Fully convolutional network*

The FCN model provides another method for predicting the future state of the sea ice using the ice chart classification defined by the WMO total concentration standard. The model training data consist of observations from 2016 and 2017. The model uses 28 input data sources for the forecast, which for this model are the six consecutive days prior (in addition to the initialisation day) to the forecast initialisation of T2, MUR SST, ice charts SIC, and OSISAF SSMIS SIC observations.

In order to reduce the computational costs of the training phase, the grid size of all data was reduced to 224x224 pixels. This simplification limits the accuracy of the forecast, especially the short-term forecast. However, we believe this resolution to be sufficient to show the advantage of the FCN for SIC prediction. A figure describing the input and output of an FCN prediction is shown in Figure 2. A more technical description of the internal layers of the FCN8 implementation used in this study can be found in the Appendix (:).




**Figur 2.** An example of FCN8 input and output.

### 4.3 Verification metrics

For verification of the predictions, sea-ice-edge metrics and a variable distance metric are used. The sea-ice edge metrics are only used for verification of the sea-ice edge location. In contrast, the distance metric can be used to verify the whole model area for several model variables. As a distance metric, we utilise the Langenbrunner $D_n$ metric, a variance-based metric for point-to-point verification (Booker, 2006). For sea-ice model verification, this metric was introduced by Urrego-Blanco et al. (2017) to asses modelled sea-ice concentration and thickness. The metric is given by

$$D_n = \frac{1}{N} \sum_{i=1}^{N} \frac{(o_i - m_i)^2}{s_i^2}, \tag{4}$$

where $o$ are the observations, $m$ the model values, $s$ the observation standard deviation, and $N$ is the number of grid cells in the domain. Since the metric is dimensionless, it can be used to aggregate values across different variables. In this study, this metric is used to verify the SIC and SST forecast values of the dynamical model.

For sea-ice concentration verification, generally, the only area where the model and observations are different are along the sea-ice edge. The exact concentration of both the observations and the model are relatively uncertain, therefore instead of evaluating the SIC it can be more instructive to verify the location of the sea-ice edge. There are several metrics available for sea-ice edge verification (Melsom et al., 2019; Goessling et al., 2016; Dukhovskoy et al., 2015; Palerme et al., 2019). The sea-ice edge metrics applied in this study are described in *Melsom et al.* [2019, hereafter MM]. Following this work, several recommended ice-edge displacement metrics are used: (1) The average ice-edge displacement (here called $\hat{E}^{PtP}$, referred to in MM as $\hat{D}^{IE}_{AVG}$), (2) The integrated ice edge error (IIEE) average displacement (here called $E^{Area}$, referred to in MM as $D^{IIEE}_{AVG}$), and (3) The IIEE bias ($\Delta^{IIEE}$) (MM; Goessling et al., 2016). The first two metrics are both used to evaluate the location of the sea-ice edge, although they often provide significantly different results. The average ice-edge displacement metric, $E^{PtP}$, defines the

**Table 4.** The experiments and their properties.

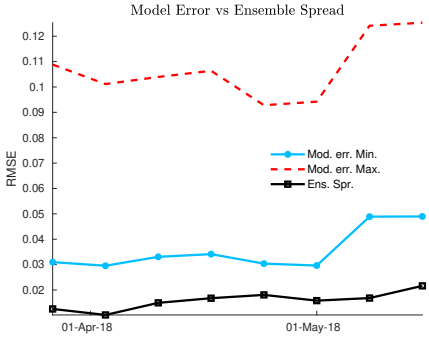
| Experiment | Name | Assimilated variables | Comment |
|---|---|---|---|
| 1 | IC+SST | MUR-SST Ice chart SIC | Metroms |
| 2 | IC | Ice chart SIC | Metroms |
| 3 | PM | OSISAF SSMIS SIC | Metroms |
| 4 | $O_{const}$ | OSISAF SSMIS and ice chart SIC | Metroms, no update of ocean variables during assimilation |
| 5 | Free | None | Metroms, no assimilation |
| 6 | IC pers | N/A | The ice chart SIC from 7 days earlier |
| 7 | PM pers | N/A | The OSISAF SSMIS SIC from 7 days earlier |

ice-edge offset by a point-to-point Euclidean distance between grid cells on the observed ice edge and the shortest distance to the modelled ice edge and *vice versa*. The IIEE average displacement metric, $E^{Area}$, defines the ice edge offset by the area between the observed and modelled ice edge. By utilising the area for error estimation, instead of point-to-point distances, small local ice features such as openings of polynyas have a much lower impact on the total offset (Goessling et al., 2016). The third metric, IIEE bias, is a measure of the integrated amount of ice in the model compared to the observations, where a positive bias means that the ice extent in the model is too large relative to the observations.

## 5 High-resolution dynamical forecasts

In this section, the dynamical model is investigated. The dynamical model system is fundamentally the same as that applied in Fritzner et al. (2019). However, here a high-resolution regional downscaling is used with the assimilation of high-resolution observations. Experiments with this regional model and the assimilation of high-resolution observations have not previously been reported. Therefore a brief assessment of this sea-ice assimilation system is provided here, comparing the assimilation of different variables. A list of the model experiments and observations assimilated in each experiment is given in table 4.

The first four experiments are all assimilation experiments. In the first experiment, both high-resolution SST from MUR and SIC from the ice charts are assimilated. In the second and third experiment, observations of SIC from the ice charts and OSISAF SSMIS are assimilated, respectively. Recall that the OSISAF SSMIS passive microwave (PM) observations have significantly lower resolution than the ice charts, which include high-resolution observations from SAR. In experiment number four, both OSISAF SSMIS and ice charts are assimilated, but the ocean variables are not updated during the assimilation. This experiment is used to asses the importance of multi-variate ocean update for SIC forecast. The fifth experiment is a free run of the Metroms model, *i.e.* the model without assimilation, used to assess the importance of assimilation. The last two experiments represent persistence, where it is assumed that no change has taken place over the forecasting period. Experiment 6 and 7 are persistence defined by the ice charts and OSISAF SSMIS, respectively.




**Figur 3.** The RMSE for maximum (red dashed line) and minimum (blue dotted line) model error compared to the model ensemble spread (black squared line). The model error is calculated from OSISAF AMSR-2 SIC observations.

### 5.1 Ensemble spread

The ensemble spread is a measure of the difference between the individual ensemble members. For efficient data assimilation with the DEnKF, the ensemble spread should represent the model error, which is the difference between the model prediction and the true state. In general, for large-scale geophysical models, the true state is not known. However, observations provide an estimate of the true state. Thus, the model error can be estimated as the difference between the modelled and observed value.

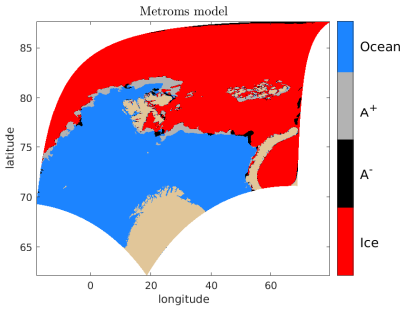
The observation uncertainty can be taken into account by applying an observation interval defined by the observation plus/minus the uncertainty. Thus, a minimum model error can be defined by the distance from the model value to the observation interval. Similarly, a maximum model error is found by the distance to the interval limit the furthest away from the model value. In Figure 3, the maximum and minimum model Root Mean Square Errors (RMSEs) of SIC are compared to the ensemble spread (ensemble standard deviation) before assimilation. The observations used are OSISAF AMSR-2 SIC observations. The ensemble spread is found to be low compared to the model error, but of the same order. A low ensemble spread compared to model error could lead to a lower effect of the observations during assimilation, and potentially a lower model accuracy.

### 5.2 Ice-edge metrics

An example of the Metroms ensemble mean forecast where SIC and SST are assimilated at 2018-05-08 is given in Figure 4. In this figure, $A^+$ represents modelled ice not observed, and $A^-$ observed ice not modelled. For this day, it it seen that the model primarily over-predicts the sea ice extent compared to the observations.

Derived ice-edge distance metrics for the whole period from 2018-03-20 to 2018-05-15 are shown in Figure 5a-c. The observed ice edge used for verification is taken from the independent low-resolution OSISAF AMSR-2 SIC product. The study period can be split into two periods, one period with relatively small changes during the first six weeks, and one period with larger changes in the last two weeks. During the last two weeks there was strong melting along the sea-ice edge, and several polynyas opened around Svalbard and Franz Josef Land. The polynyas at 2018-05-08 can be seen by the grey areas inside the ice in Figure 4.

In the Figures 5a-c the three different sea-ice edge metrics are used to assess different aspects of the forecasts, a) $\hat{E}^{PtP}$, b) IIEE bias and c) $E^{area}$. As described previously, $\hat{E}^{PtP}$ verifies the ice edge by a point-to-point comparison with the observed ice




**Figur 4.** The IIEE area of an ensemble mean of a 7-day forecast with the MUR-ice chart assimilation model verified against an ice chart on 2018-05-08. blue areas represent the ocean in both types of data, and red areas represent ice in both model and observations. Grey areas represent areas where the model has ice, while the ice chart has not, and *vice versa* for black areas.

edge, and $E^{area}$ is based on the integrated area between the modelled and observed ice edge. The IIEE bias gives a measure of the total ice content compared to the observations.

All assimilation experiments (solid lines, see table 4) show an improved modelled ice edge compared to the free-run experiment (dotted blue line). The free-run experiment has higher displacement errors, especially during the last period when there are more substantial changes for both ice edge distance metrics (Figures 5a and c). The IIEE bias (Figure 5b) of all Metroms experiments are similar and seems independent of the assimilation.

The difference between the assimilation experiments (solid lines) is found to be small for all three validation metrics. This result has several implications in our case: there are no significant evidence of local ice edge differences; little effect of assimilating high-resolution observations compared to the conventional low-resolution type; assimilating SST observations does not improve the sea-ice edge forecast; updating the ocean during assimilation has no significant effect when forecasting the sea-ice edge. However, it is important to note that these findings applies to this particular model setup and study period, and may not be valid in general without further research.

A further comparison of the assimilation experiments is performed by using the high-resolution ice charts for verification, note that these were also applied for assimilation. The comparison is provided by using the $E^{area}$ metric and shown in Figure 5d. The use of high-resolution verification data have the potential to reveal the advantages of assimilating high-resolution data. However, also for this test, it is shown that the differences are small. A potential reason for the small differences could be model drift. During the forecast, between the assimilation steps, the model could potentially lose all information from the assimilation due to the model being driven by the forcing. To investigate the drift, the $E^{area}$ was calculated for the assimilation experiments assimilation analysis (Figure 6a). It is found that for the assimilation analysis, the assimilation experiments provide similar $E^{area}$ indicating that the ice-edge information gained during assimilation is of more or less the same level of high-resolution detail for all experiments.

There are several reasons why there does not seem to be any effect of using high resolution compared to low-resolution observations in our study. First of all, the study period is relatively short. Secondly, both SIC products provide approximately the same ice-edge location. The ice charts use the passive microwave observations to both fill the gaps of the SAR observations and to verify ice water in ambiguous situations. Finally,

the small differences could also be related to the assimilation method, and the need to maintain ensemble spread when the EnKF is applied. Remember that the EnKF assimilation impact is tuned such that the ensemble spread (see section 5.1) is maintained. The tuning is performed by reducing the observation impact, which might be why a lower effect than expected is seen when the ice charts are assimilated. Besides, for the ice chart a conservative low-resolution estimate of 5 % sea-ice concentration is applied which could also affect the results.

The lack of improvements when SST is assimilated could be related to the length of the study period. When the SST observations are assimilated, the most significant updates are found far from the ice edge. Thus these updates are not expected to affect the sea-ice forecast immediately. This is also consistent with the lack of impact found when the ocean was kept constant during the assimilation of SIC. In addition, close to the sea-ice edge, the SST in the observational MUR product is derived based on an empirical relationship between OSISAF SIC and SST (Chin et al., 2017), and therefore this information is already taken into account through the SIC assimilation.

Finally, the Metroms model is compared to persistence (dashed black and red lines in Figure 5). Persistence is the ice charts (dashed red) and the OSISAF SSMIS (dashed black) observations from the previous assimilation step. Persistence has comparable and in some cases a lower edge displacement than the assimilated runs for the period with small changes. This is especially seen when the $\hat{E}^{PtP}$ is used (Figure 5a), indicating that small local areas are in fact better predicted by persistence. This could be related to for example polynyas that are not resolved in the model, and because of low ensemble spread, are not updated during the assimilation either. However, for the period with more substantial changes, the dynamical model shows clear improvements over persistence, especially for the $E^{Area}$ metric, when the larger scales are verified.

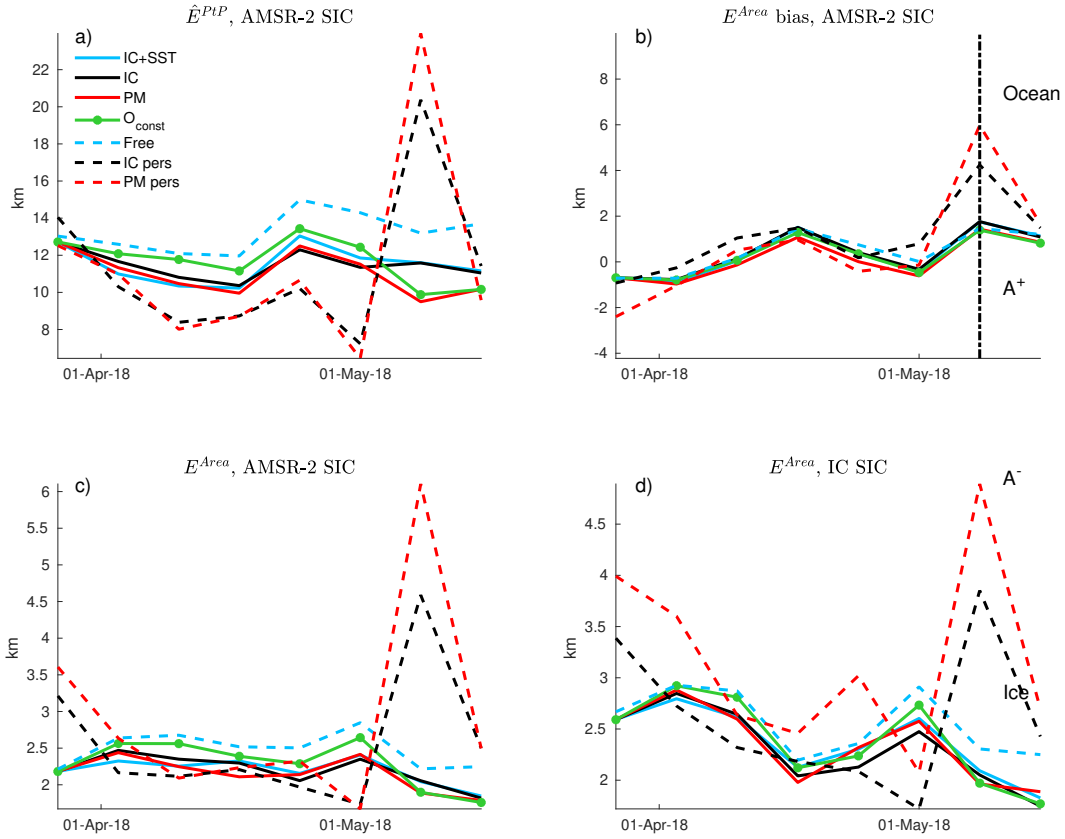
For the experiments shown in Figures 5 and 6a, low-resolution atmospheric forcing data from ECWMF IFS is applied. However, it is expected that for a high-resolution model as applied here, there could potentially be a benefit of using high-resolution atmospheric forcing. In Figure 6b the freerun model forced by atmospheric forcing from Arome Arctic (2.5 km) (Müller, Homleid et al., 2017; Müller, Batrak et al., 2017) and ECMWF IFS (18 km) is compared to the AMSR-2 observations based on the $E^{Area}$ metric. It is shown that when the high-resolution AROME Arctic is used, $E^{Area}$ is significantly reduced for the first period compared to when ECMWF IFS is used, for the last period they are more similar. Thus, utilising high-resolution forcing can potentially have a significant impact on the sea-ice forecast.

### 5.3 $D_n$ metrics

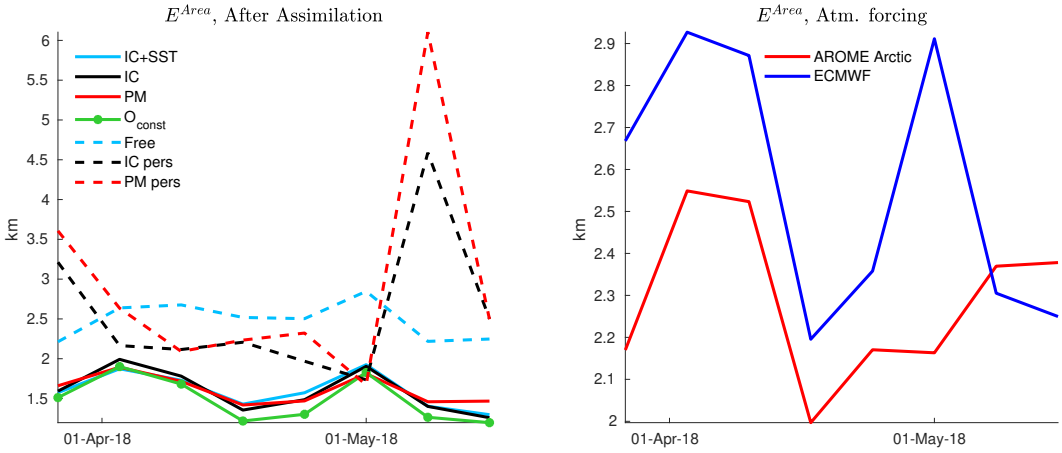
In the previous section the modelled ice-edge offset was analysed, while in this section, the $D_n$ metric (Urrego-Blanco et al., 2017) is used to analyse individual grid cell values. The verification is shown in Figure 7 for two categories of observations: a) SIC and b) SST. For SIC, both OSISAF SSMIS and OSISAF AMSR-2 products are used for verification. Note that, the ice charts are not used for SIC verification in this analysis as these apply discrete values.

Compared to the experiment without assimilation (dashed blue), the assimilation models show significant improvements both for the SIC and SST on the grid-scale. However, there are no significant differences between the individual Metroms assimilation experiments (solid lines) for SIC verification, consistent with the results found previously with the ice edge metrics.

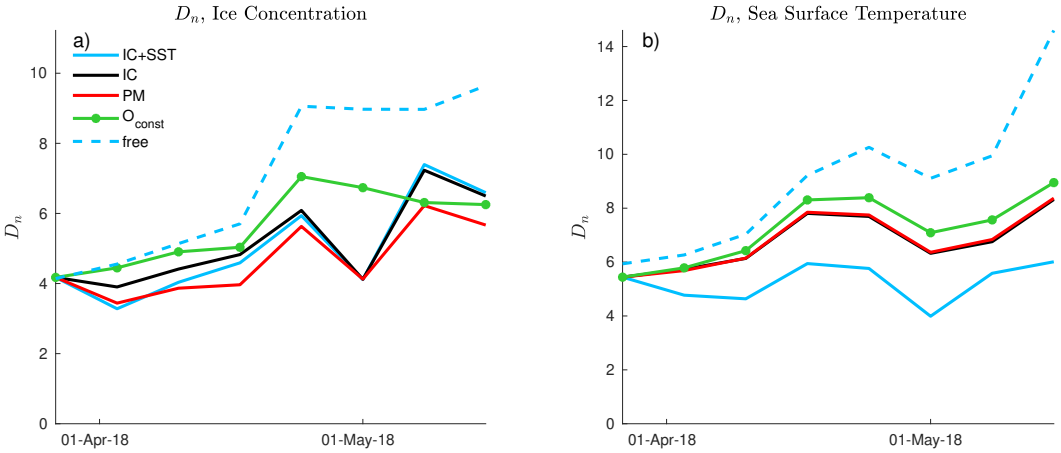
For SST verification, both the MUR SST and OSTIA SST observational products are used. For this verification, the IC-MUR experiment (solid blue), assimilating SST, performs significantly better than the other assimilation systems. In addition, the sys-




**Figur 5.** Ice-edge metrics as a function of time calculated from the 7-day forecasts and verified by observations. The different metrics are a) average ice-edge displacement verified by AMSR-2 SIC, b) IIEE bias verified by AMSR-2 SIC, (c) IIEE average displacement verified by AMSR-2 SIC, and d) IIEE average displacement verified by ice charts. Light blue is assimilation of MUR SST and ice chart SIC, solid black is assimilation of ice chart, dashed black is persistence ice chart forecast, red is OSISAF SIC assimilation, dashed red is persistence OSISAF SSMIS forecast, dashed blue is a free run without assimilation and solid green dotted line is the assimilation of ice chart and OSISAF SSMIS SIC without updating the ocean. The vertical dotted black line represents the date 2018-05-08 used in Figure 4. See also table 4 for more information regarding the experiments.




**Figur 6.** In a) a similar verification as performed in Figure 5c, but here for the model assimilation analysis. In b) $E^{Area}$ for two free-run models with different atmospheric forcing: ECMWF (blue) and AROME Arctic (red). The colour coding in b) is the same as in Figure 5.




**Figur 7.** The $D_n$ metric calculated based on different observations in a) SIC observations from OSISAF AMSR-2 and OSISAF SSMI/S, b) SST observations from MUR and OSTIA. The colour coding follows that of Figure 5.

tem that does not update the ocean during assimilation (solid yellow) shows a small error which is increasing throughout the period.

These results showt an effect of both updating the ocean during assimilation and assimilating SST on the grid-scale. However, as mentioned previously, the effect of assimilating SST is most substantial far from the sea-ice edge. Thus, for more extended temporal simulations, when these parts of the ocean come into contact with the sea ice, the effect of assimilating SST may become significantly more evident.

In summary: In these experiments, little or no improvements in using high-resolution sea-ice concentration observations for assimilation are found. However, it is shown that using high-resolution atmospheric forcing can have a significant impact on the forecast. In general, compared to the coupled model, persistence is a reasonable assumption for forecasts up to 7 days, when there are small changes in the sea ice. However, when more

substantial changes appear, the dynamical forecast model system shows significant skills relative to persistence.

# 6 machine-learning forecasts

## 6.1 FCN forecast

The FCN model was trained with observations from 2016 and 2017, and the model was verified using forecasts for 2018. As mentioned, the FCN is trained on lower resolution observations. Therefore the same resolution is also applied for verification. The trained model is assessed by 1-, 2-, 3- and 4-week forecasts.

In Figure 8, the $E^{Area}$ of the FCN forecasts are shown. The FCN forecasts are compared to reduced-resolution ice-chart persistence. The 7-day FCN forecasts (Figure 8a), have a similar skill as the ice-chart persistence. This similarity is expected since sea ice has small variations on these time and spatial scales, recall the low $E^{Area}$ values for persistence in Figure 5. For the other three forecast lengths, the FCN shows similar skill as persistence during most of the year, except in summer. In summer, the most substantial changes in the Arctic sea ice occur, and the FCN significantly improves upon persistence. These improvements are found for the 2- to 4-week forecasts, and the improvements increase with the forecast length as persistence performs worse for longer timescales.

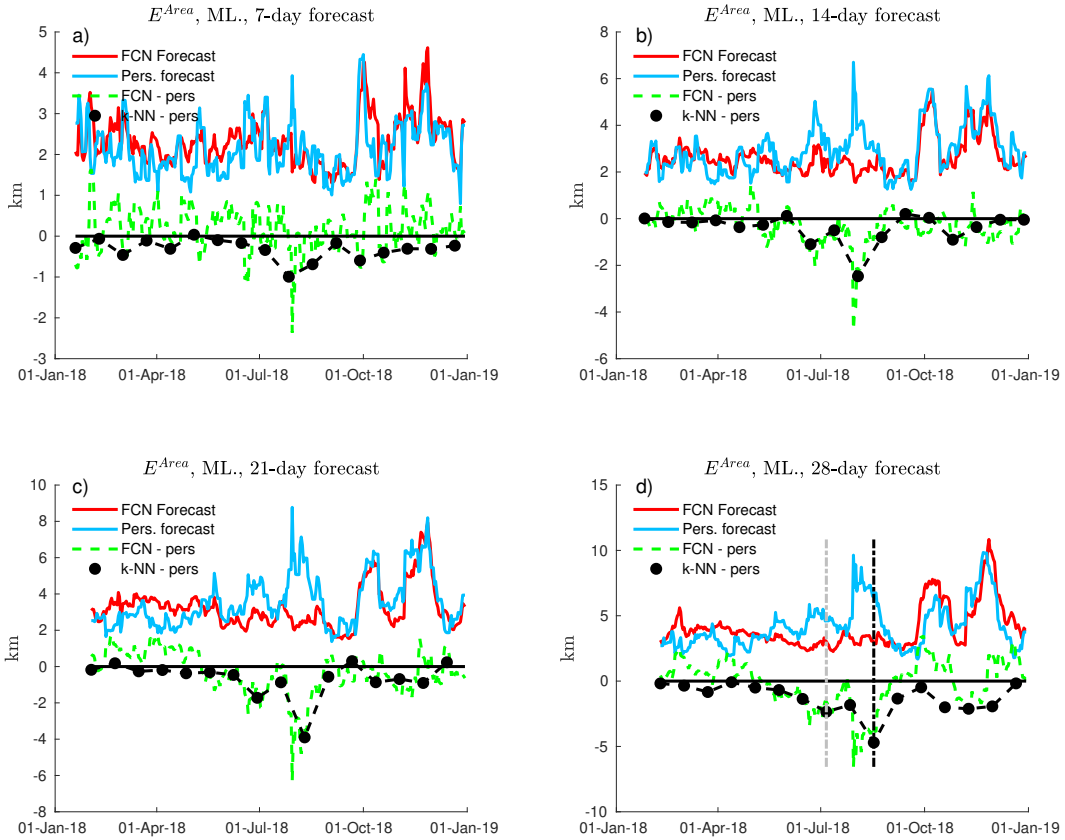
In Figures 9 a) and b) the IIEE areas for an FCN forecast and 4-week persistence, respectively, validated by ice-chart observation 2018-08-17 (black vertical line Figure 8) are shown. This date is in a period of substantial change, where the FCN forecast performs significantly better than the persistence forecast. The most substantial improvements with the FCN for this date are the representation of the melt in the southern Kara Sea.

## 6.2 Verification of the FCN forecast

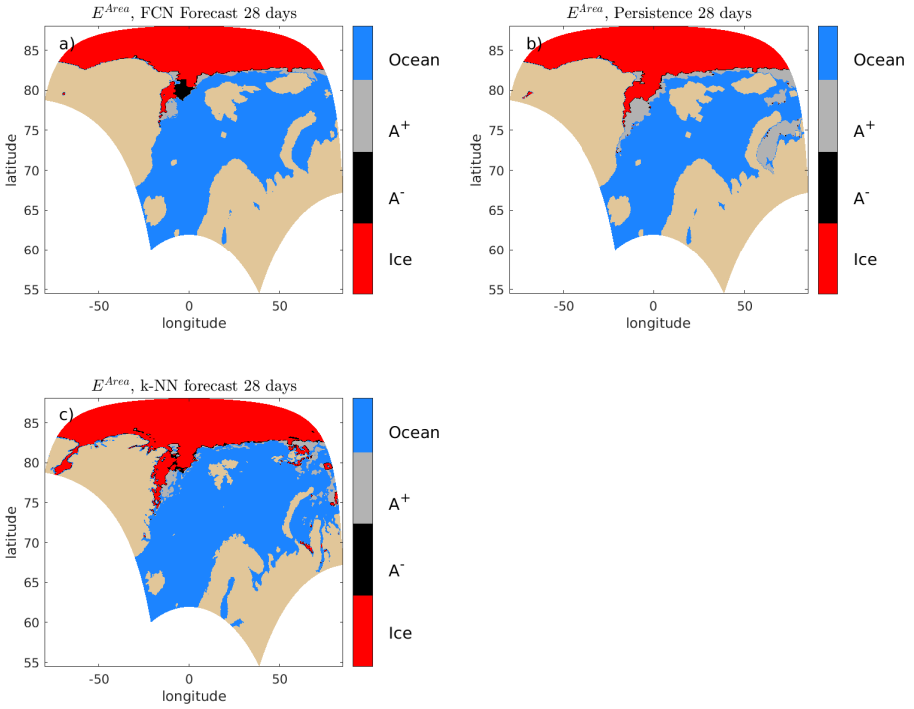
Verification of the FCN 4-week forecast for 2018 is performed by switching the training and verification data. An FCN forecast model for 2016 is trained by data from 2017 and 2018, and an FCN model for 2017 is trained by data from 2016 and 2018. The $E^{Area}$ for the two new 4-week forecasts is shown in Figure 10. As for the 2018 forecasts, the 2016 and 2017 forecasts show improvements in summer, similar to the results for 2018. These results verify that the FCN model well predicts summer melt. In addition to verifying the predictions against regular persistence, two other persistence metrics have been used: persistence of the previous anomaly and damped persistence (not shown; Van den Dool et al., 2007). These two metrics take the climatology into account to give a more accurate persistent prediction estimate. However, by applying these two estimates, the results do not change significantly, the persistence estimate in summer gets slightly better. However, the machine-learning predictions are still much better than persistence. A problem with using climatological SIC data is that because SIC is a bounded variable between zero and one, the climatology will always be biased towards a larger sea-ice extent.

## 6.3 *k*-NN forecast

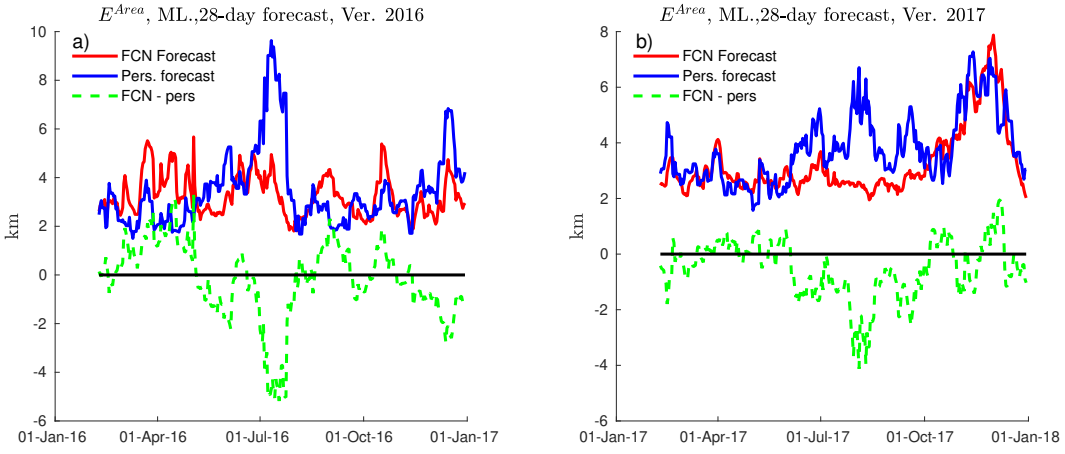
The differences between the $E^{Area}$ for the $k$-NN forecasts and ice chart persistence are shown in Figure 8. The results are compared to the FCN forecast for four forecasting periods, 1-, 2-, 3- and 4-weeks. Forecasting with the $k$-NN model is more computationally expensive than with the FCN. Therefore a $k$-NN forecast is only performed every third week. For comparison purposes, the $k$-NN forecast results are rescaled to the resolution of the FCN forecast, 224×224 pixels. When the changes in the sea ice are small,



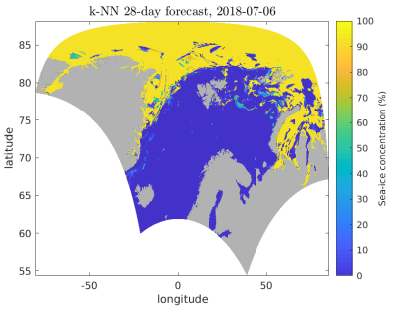

**Figur 8.** IIEE average displacement calculated for FCN and $k$-NN forecasts with a length of a) 7 days, b) 14 days, (c) 21 days and (d) 28 days. The red line represents the FCN forecasts, the blue line a persistence forecast, the green dashed line the difference between FCN and persistence forecast, and the black dashed-dotted line is the difference between $k$-NN and persistence forecasts. The vertical dashed, grey and black lines represent the dates 2018-07-06 and 2018-08-17, respectively.




**Figur 9.** IIEE regions for 28-days forecast of the date 2018-08-17 for a) FCN forecast, b) persistence forecast and (c) $k$-NN forecast. The colours and coding are as in Figure 4.




**Figur 10.** IIEE average displacement plotted as a function of time for 28-day forecasts of a) 2016 and b) 2017. The colours are as in Figure 8.




**Figur 11.** $k$-NN 28-day sea-ice forecast on 2018-07-06.

the $k$-NN forecasts have similar displacements as the persistence forecast, both improve upon the FCN forecasts, while when more substantial changes occur, the $k$-NN forecasts are closer to those of the FCN model.

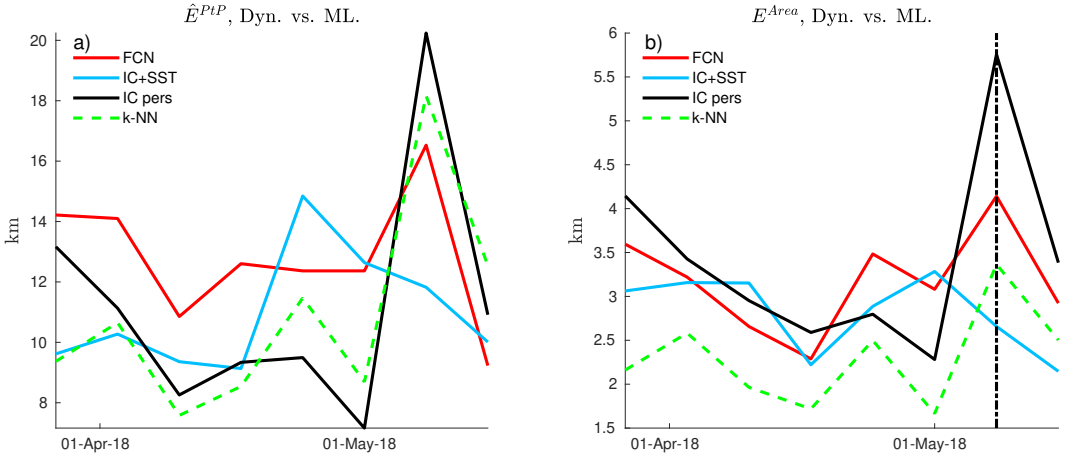
Note that, due to the lack of spatial coherence in the $k$-NN method, small local changes in the marginal ice zone can mean the difference between ice and water in a given pixel. Thus, with this method, the forecast can include unrealistic areas of drifting sea ice and polynyas. An example is shown in Figure 11, where a forecast with the $k$-NN method is shown for 2018-07-06. The forecast shows an unrealistic occurrence of large polynyas and several large areas of drift ice. It can be seen from Figure 8d, that for this particular day (grey dotted vertical line) the $k$-NN forecast has a small $E^{Area}$. This result is related to the fact that $E^{Area}$ considers the integrated values and not local effects.

In Figure 9c the IIEE area for the 4-week $k$-NN forecast at 2018-08-17 is shown. The forecast result for this day is similar to that of the FCN, Figure 9a, with significant improvements compared to the persistence forecast shown in Figure 9b. Again, due to the lack of spatial context, the forecast has more drifting ice around the sea-ice edge, compared to the FCN forecast.

## 7 Comparison between the dynamical Metroms forecast and the two machine learning models for 7-day forecasts

In this section, the 7-day machine-learning forecasts from the $k$-NN and FCN are compared to the Metroms assimilation system assimilating SIC ice charts and SST MUR observations. In Figures 12a and b the $\hat{E}^{PtP}$ and $E^{Area}$ metrics, respectively, are plotted for 7-day forecasts from FCN, $k$-NN, Metroms model and persistence. All forecasts are mapped to the 2.5 km Metroms grid, with the land mask taken from the low-resolution FCN grid.

For the $\hat{E}^{PtP}$, the FCN has on average, higher displacements than the other forecasts. The persistence forecast and $k$-NN are similar and show the lowest displacements for most of the period. The displacements from the Metroms assimilation experiment are also of a similar magnitude in the first part but shows improvements in the second part when substantial changes in the SIC occur. For the $E^{Area}$, the $k$-NN method shows significantly lower displacements than the other forecasts for most of the period, except for the last two weeks when more substantial changes occur, in this period the dynamical forecast is the most skilful. The results show that the FCN method performs worse when it comes to predicting local areas; however, for the larger-scale ice edge, it has similar properties as the other methods. The $k$-NN prediction show reliable results with low errors. However, the most substantial changes are only predicted by the dynamical model.




**Figur 12.** Comparison of a) average ice-edge displacement and b) IIEE average displacement calculated from 7-day forecasts during the Metroms study period 2018-03-20 - 2018-05-15. The red line represents the FCN, blue line the Metroms IC-MUR assimilation model, black line the persistence forecasts, and the green dashed line is the $k$-NN forecast. The vertical dotted black line represents the date 2018-05-08.

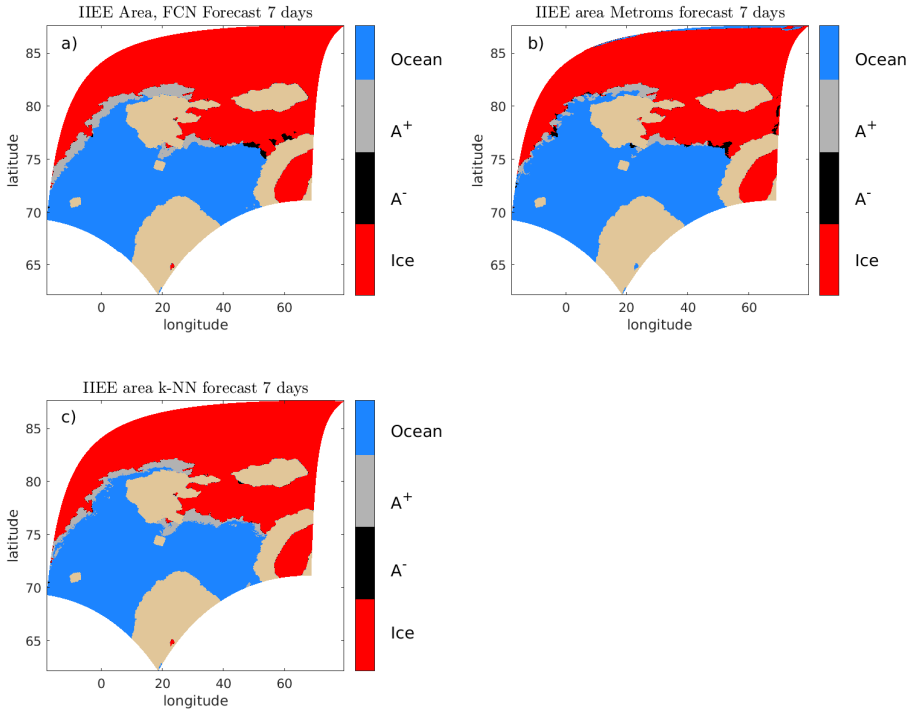
An example of the IIEE areas for 2018-05-08 (black dotted vertical line in Figure 12b) is given in Figures 13a-c for FCN, Metroms and $k$-NN forecast, respectively. During the week leading up this day, there was a significant northward progression of the ice edge, especially close to Svalbard in the local area. From Figure 13a, it is seen that the dynamical forecast shows the best skill with regards to predicting this ice melt, followed by the $k$-NN forecasts. This result is in accordance with that shown in Figures 12.

## 8 Discussion and conclusions

In this study, the applicability of using machine learning for sea-ice concentration forecasting has been assessed. This is done by assessing the machine-learning methods individually and comparing them to a dynamical model.

The improvements of the machine-learning forecasts compared to persistence were found to vary with the forecast length. For short-term forecasts, the FCN provided only small improvements compared to persistence. This was partly due to small sea-ice changes in the SIC on these timescales, implying that persistence becomes a skilful forecast alternative, and partly due to the low resolution of the FCN. The FCN model uses layers of pooling and convolution, which reduce the resolution of the features and provide lower-resolution results. In general, when using such a machine-learning method, it cannot be expected that the output has the same resolution as the input data (see examples in Long et al. (2015)). A potential solution could be to use fewer convolutional blocks to keep more high-resolution features (see appendix :).

For longer timescales, significant improvements were found with the FCN forecast in summer, indicating melt effects being well represented in the trained model. The summer improvements are related to lower accuracy of persistence in this period when the changes in the Arctic sea ice are more substantial. In general, the results found with the FCN method provided limited improvements compared to persistence which might be related to a too-small training dataset and too low resolution. For sea-ice observations, consecutive days have a high correlation. Therefore, the effective training-set size is likely much less than the original 700. In general, as shown by Scher og Messori (2019), a




**Figur 13.** IIEE areas 2018-05-08 for 7-day forecasts of a) The FCN method, b) Metroms assimilation system, and (c) the $k$-NN method. The colours are as used in Figure 4

more extensive training set should be used to obtain improved results. However, the fact that the FCN model can be used to forecasts sea ice quickly on almost any computer as long as a pre-trained model exists makes this model attractive. The results shown here motivate for a more sophisticated version with higher resolution based on a more extensive training dataset. A higher resolution could potentially be achieved by using patching, where smaller areas are used for training and forecasting. In addition, the use of a Graphical Processing Unit (GPU) programming could potentially reduce computational training time.

The $k$-NN method was found to provide forecasts close to that of persistence throughout most of the year but improved upon persistence when the sea-ice changes were more substantial. In general, the $k$-NN method provides forecasts with consistently lower ice-edge displacements than the FCN forecasts. A problem with the $k$-NN forecasts is occasional noise output in the form of ice residuals. These ice residuals are caused by not including the spatial context in the predictions when this method is applied. Compared to the FCN, the $k$-NN forecasts are computationally expensive, and the training data size is limited.

Compared to the dynamical forecasts, the FCN method provided a higher average $\hat{E}^{PtP}$ and a similar $E^{Area}$, and show no significant improvements upon the dynamical forecast. The FCN forecasts have significantly lower resolution than the other forecast models used in this study, which likely affected the results. The $k$-NN method was found to give a significantly lower $E^{Area}$ compared to the dynamical model when there where small changes in the sea-ice concentration during the forecast, while when more substantial changes occurred, the dynamical forecast gave the best SIC forecast. It is important to note that the $k$-NN forecast also has the highest resolution, 1 km, similar to the ice charts, while the Metroms model has a 2.5 km resolution and the FCN 10-20 km.

It is also important to mention that compared to the $k$-NN forecast, both the dynamical model and the FCN use prior information in the forecasts. For the dynamical model, the atmospheric forecast has assimilated data which includes information regarding the future, which likely leads to a better forecast than using an atmospheric prediction. For the FCN, a limitation is the amount of training data. However, a trick to artificially create more training data is to use the same training data several times but shuffled. However, when the same training dataset is used several times in different orders, there is a concern that the model might become overfitted. For an overfitted model, the predictions will be biased towards the training dataset. To avoid overfitting a method of early stopping was applied. With early stopping, the trained model with the lowest error compared to the forecast data was chosen. This method might reduce the generality of the trained model. However, we believe that the results still indicate how well the FCN performs for sea-ice forecasting. Moreover, it is important to note that the FCN is not trained on the forecast data. However, as both the FCN and dynamical model use future information in the forecast, the results of the simple $k$-NN method becomes even more impressive.

The motivation for using machine-learning forecasts is primarily to reduce the computational cost while meeting the requirements for prediction accuracy. Both machine-learning methods presented here are significantly cheaper computationally than the dynamical model. Both the machine-learning predictions are made locally on a desktop computer. However, the training of the FCN model was done on a cluster, but only using 20 CPUs. Ideally, GPUs should be used for FCN training. For comparison, the dynamical model forecast is generally run on more than 80 CPUs. For the FCN method, increasing the resolution would lead to a more costly model training. However, the training of the FCN method is a one-time cost. Thus a computationally costly model training is affordable as the predictions are extremely fast. It was shown in this study, that the simple $k$-NN method outperformed the FCN. However, we recommend that the FCN method is investigated further for sea-ice prediction as it has a wide range of useful proper-

ties. With the FCN method, more than one variable can be predicted, and the method can be used similarly as a dynamical model (Scher & Messori, 2019). In addition, since the forecast is fast, a prediction time step of one day can be used, and more extended forecasts would require several model predictions with model output as input.

In addition to assessing the machine-learning methods, an investigation of the assimilation system was performed. For the dynamical model, a period in spring 2018 was used to investigate the effect of assimilating different observations in the Metroms high-resolution ocean-sea-ice coupled model system. The SIC forecast when assimilating high-resolution observations was found to give similar results as when assimilating lower resolution observations. This result is unexpected with regards to previous results found by Posey et al. (2015). There are several reasons for this, firstly only two months with relatively small changes have been investigated. In addition, the resolution difference between the ice charts and the passive microwave observations are not that large. In Posey et al. (2015), a 25 km product was used as the low-resolution product, while we use a 10 km product.

It was found that neither updating the ocean during assimilation nor assimilating SST have a significant impact on the SIC forecast. However, these ocean-related implementations were found to have a significant effect on the SST forecast. Thus, it is expected that for more extended model simulations, the assimilation of SST and updating ocean variables may have a positive impact on the quality of the SIC forecast. It is also important to note that close to the sea-ice edge, the MUR product uses the OSISAF SIC and an empirical relationship to derive the SST. Thus, these observations provide little new information compared to the SIC observations close to the sea-ice edge, which might be why no significant effect on the SIC forecast is seen when assimilating the MUR SST.

**Acknowledgments**

We want to thank Pavel Sakov for help using and implementing the EnKF-c code and for helpful discussions regarding the EnKF. We want to thank Jostein Brændshøi and Nils Melsom Kristensen for help with setting up and running the coupled model. We would also like to thank Harald Engedahl and Graig Sutherland for providing TOPAZ ensemble output from the operational system, used here for boundary conditions and generating our ensemble. We would also like to thank Michael Kampffmeyer for heplpful tips regarding machine learning.

The observations used for assimilation and verification is available through the following data portals: The AMSR-2 and SSMIS sea-ice concentration products are available from the EUMETSAT OSISAF centre (http://osisaf.met.no/), the ice charts are available from the the Copernicus Climate Change Service Information (http://marine.copernicus.eu/), the MUR and OSTIA SST datasets are available through the NASA EOSDIS Physical Oceanography Distributed Active Archive Center (https://podaac.jpl.nasa.gov) and the ECMWF atmospheric data is available through the ECMWF data portal (https://apps.ecmwf.int). The model data output is published in the NIRD Research Data Archive: DOI: 10.11582/2019.00038, under project NS9063K.

This work was funded through the Center for Integrated Remote sensing and Forecast for Arctic Operations through the Norwegian Research Council grant no. 237906. Two supercomputers provided by the Norwegian Metacenter for Computational Science (NOTUR) was used for the computational work, the Stallo and Fram computers both under project NN9348K.

## Appendix A FCN8

In this section a more technical description of the FCN8 network used in this study is given. The full FCN8 network used is shown in Fig. A1. The network consists of 5 blocks of convolution and max pooling layers.

The FCN8 uses locally connected layers of convolutional (Conv2d, see Figure A1), pooling (MaxPooling2D), upsampling (Conv2DTranspose), and non-linear activation (included in Conv2D) for decision making. The upsampling layer, consists of fractional strided convolutions/deconvolution for pixel-wise prediction of input with reduced spatial dimension due to pooling operations. To improve resolution of the output, skip connections are utilised during the upsampling process (Long et al., 2015; Shelhamer et al., 2017). With skip connections, high-resolution information in early layers is combined with large-scale information in the deep layers for step-wise upsampling. The skip connections are combined in Add, where information from block 3,4 and 5 is combined by individual upsampling.

The activation layers provide an activation function that is performed on the convolution layer output. This activation function introduce non linearity in the model, without the activation function, the network becomes a linear regression model consisting of linear convolution operations. In this study, a rectified linear unit (ReLU) is used (Glorot et al., 2011), which is a function that filters out negative values,

$$g_{ReLU}(\boldsymbol{x}) = max(0, \boldsymbol{x}), \tag{A1}$$

where $\boldsymbol{x}$ is an input and $g$ is the activation function.

The last step in Fig. A1 is an activation layer, which gives the probability for each pixel to be in one of the discrete WMO ice concentration intervals.

## References


Andersen, S., Tonboe, R., Kaleschke, L., Heygster, G. & Pedersen, L.T. (2007). Intercomparison of passive microwave sea ice concentration retrievals over the high-concentration arctic sea ice. *J. Geophys. Res. Oceans*, *112*(C8).

Booker, J.M. (2006). Interpretations of langenbrunner's Dn metric for v&v. *Los Alamos National Laboratory Technical Report*.

Buehner, M., Caya, A., Pogson, L., Carrieres, T. & Pestieau, P. (2013). A new environment canada regional ice analysis system. *Atmos. Ocean*, *51*(1), 18–34. doi: 10.21957/m1cs7h

Burgers, G., van Leeuwen, P. & Evensen, G. (1998). Analysis scheme in the ensemble kalman filter. *Mon. Weather Rev.*, *126*, 1719-1791. doi: 10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2

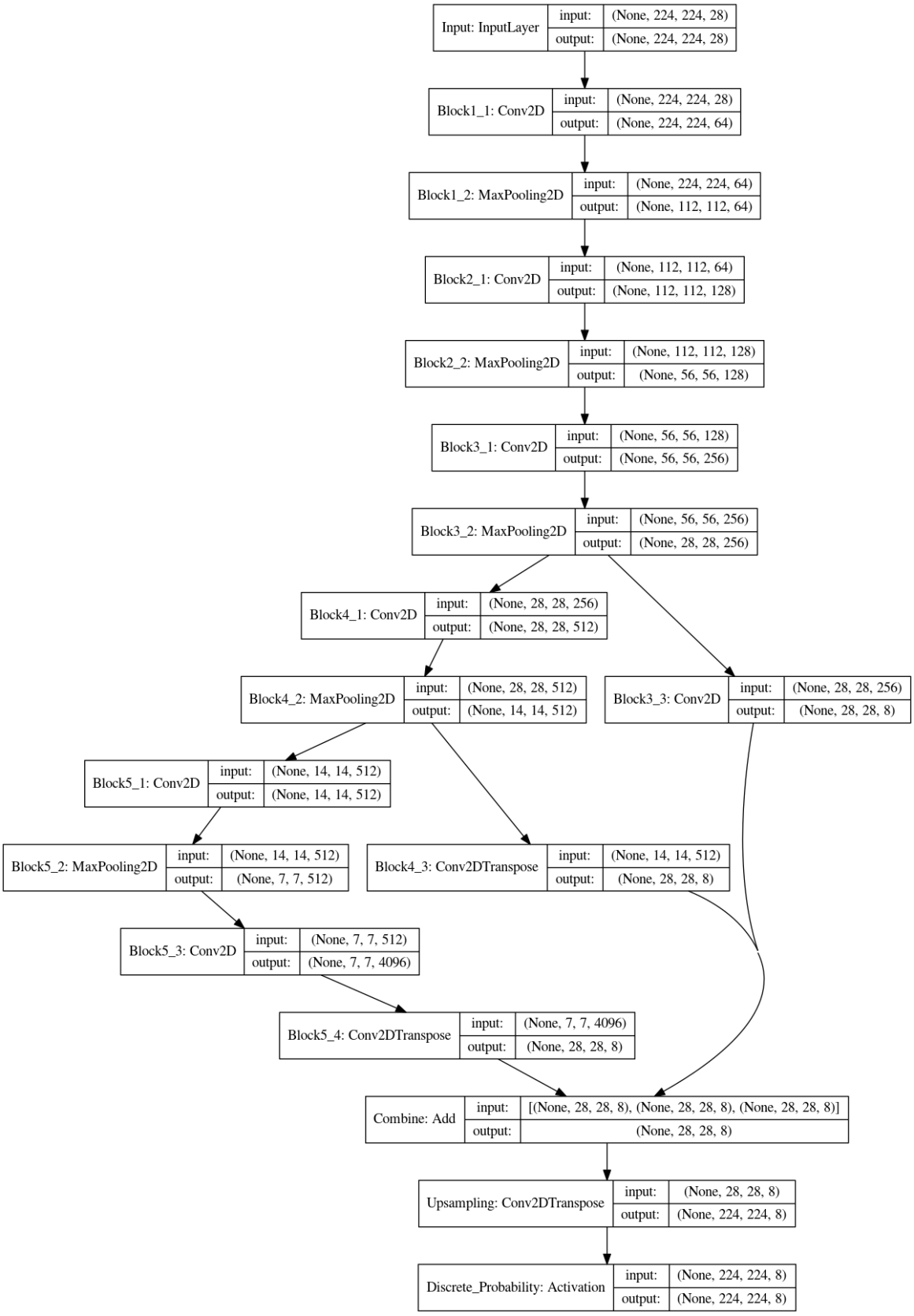
Cardinali, C., Pezzulli, S. & Andersson, E. (2004). Influence-matrix diagnostic of a data assimilation system. *Q. J. Roy. Meteor. Soc.*, *130*(603), 2767–2786.

Cavalieri, D.J. & Parkinson, C.L. (2012). Arctic sea ice variability and trends, 1979-2010. *The Cryosphere*, *6*(4), 881.

Caya, A., Buehner, M. & Carrieres, T. (2010). Analysis and forecasting of sea ice conditions with three-dimensional variational data assimilation and a coupled ice-ocean model. *J. Atmos. Oceanic Technol.*, *27*(2), 353-369. doi: 10.1175/2009JTECHO701.1

Chi, J. & Kim, H.-c. (2017). Prediction of Arctic sea ice concentration using a fully data driven deep neural network. *Remote Sensing*, *9*(12), 1305.

Chin, T.M., Vazquez-Cuervo, J. & Armstrong, E.M. (2017). A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sens. Environ.*, *200*, 154–169.




**Figur A1.** An overview of the internal layers and their dimensions in this study. The numbers in each box are the matrix dimensions for each layer: (time, longitude, latitude, level). None values are the time dimension, which is not applicable here but is included by the the software library used. The level dimension is the channels/features described in the text.

Comeau, D., Giannakis, D., Zhao, Z. & Majda, A.J. (2019). Predicting regional and pan-arctic sea ice anomalies with kernel analog forecasting. *Climate Dynamics*, *52*(9-10), 5507–5525.

Copernicus. (2019). Copernicus climate change service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate . copernicus climate change service climate data store (CDS). *Copernicus Climate Change Service Climate Data Store (CDS*. doi: https://cds.climate.copernicus.eu/cdsapp#!/home

Cover, T.M., Hart, P.E. et al. (1967). Nearest neighbour pattern classification. *IEEE T. Inform. Theory*, *13*(1), 21–27.

Dinessen, F. & Hackett, B. (2016). Product user manual for regional high resolution sea ice charts svalbard region. *http://resources.marine.copernicus.eu/documents/PUM/CMEMS-OSI-PUM-011-002.pdf*, *2.3*.

Donlon, C., Robinson, I., Casey, K., Vazquez-Cuervo, J., Armstrong, E., Arino, O., ... others (2007). The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *B. Am. Meteorol. Soc.*, *88*(8), 1197–1214.

Donlon, C.J., Casey, K.S., Robinson, I.S., Gentemann, C.L., Reynolds, R.W., Barton, I., ... Evans, R. (2009, September). The GODAE high-resolution sea surface temperature pilot project. *Oceanography*, *22*.

Donlon, C.J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E. & Wimmer, W. (2012). The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sens. Environ.*, *116*, 140–158.

Dukhovskoy, D.S., Ubnoske, J., Blanchard-Wrigglesworth, E., Hiester, H.R. & Proshutinsky, A. (2015). Skill metrics for evaluation and comparison of sea ice models. *Journal of Geophysical Research: Oceans*, *120*(9), 5910–5931.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.*, *99*, 10143-10162. doi: 10.1029/94JC00572

Evensen, G. (2003). The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, *53*, 343-367. doi: 10.1007/s10236-003-0036-9

Fritzner, S., Graversen, R., Christensen, K.H., Rostosky, P. & Wang, K. (2019). Impact of assimilating sea ice concentration, sea ice thickness and snow depth in a coupled ocean–sea ice modelling system. *The Cryosphere*, *13*(2), 491–509.

Fritzner, S., Graversen, R., Wang, K. & Christensen, K. (2018). Comparison between a multi-variate nudging method and the ensemble kalman filter for sea-ice data assimilation. *J. Glaciol.*.

Glorot, X., Bordes, A. & Bengio, Y. (2011). Deep sparse rectifier neural networks. I *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (s. 315–323).

Goessling, H.F., Tietsche, S., Day, J.J., Hawkins, E. & Jung, T. (2016). Predictability of the arctic sea ice edge. *Geophys. Res. Lett.*, *43*(4), 1642–1650.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep learning*. MIT Press. (`http://www.deeplearningbook.org`)

Guo, Y., Liu, Y., Georgiou, T. & Lew, M.S. (2018). A review of semantic segmentation using deep neural networks. *International journal of multimedia information retrieval*, *7*(2), 87–93.

Gupta, D. (2019). *https://github.com/divamgupta/,lastaccessedJune2019*.

Helfrich, S.R., McNamara, D., Ramsay, B.H., Baldwin, T. & Kasheta, T. (2007). Enhancements to, and forthcoming developments in the interactive multisensor snow and ice mapping system (IMS). *Hydrol. Process.*, *21*(12), 1576–1586.

Houtekamer, P.L. & Mitchell, H.L. (2001). A sequential ensemble kalman filter for atmospheric data assimilation. *Mon. Weather Rev.*, *129*(1), 123–137.

Houtekamer, P.L. & Zhang, F. (2016). Review of the ensemble kalman filter for at-

mospheric data assimilation. *Mon. Weather Rev.*, *144*(12), 4489-4532. doi: 10.1175/MWR-D-15-0440.1

Hunke, E. & Dukowicz, J. (1997). An elastic-viscous-plastic model for sea ice dynamics. *J. Phys. Oceanogr*, *27*, 1849-1867. doi: 10.1175/1520-0485(1997)027<1849:AEVPMF>2.0.CO;2

Jazwinski, A. (1970). *Stochastic processes and filtering theory.* Sand Diego, California: Academic.

Kim, J., Kim, K., Cho, J., Kang, Y.Q., Yoon, H.-J. & Lee, Y.-W. (2019). Satellite-based prediction of arctic sea ice concentration using a deep neural network with multi-model ensemble. *Remote Sensing*, *11*(1), 19.

Kim, Y.J., Kim, H.-C., Han, D., Lee, S. & Im, J. (2020). Prediction of monthly arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks. *Cryosphere*, *14*(3).

Kræmer, T., Johnsen, H. & Brekke, C. (2015). Emulating sentinel-1 doppler radial ice drift measurements using envisat ASAR data. *IEEE T. Geosci. Remote*, *53*(12), 6407–6418.

Kristensen, N., Debernard, J., Maartensson, S., Wans, K. & Hedstrom, K. (2017). *metno/metroms.* doi: 10.5281/zenodo.1046114

Kurtz, N. & Harbeck, J. (2017). CryoSat-2 level-4 sea ice elevation, freeboard, and thickness, version 1. *Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, accessed May 2019*. doi: https://doi.org/10.5067/96JO0KIFDAS8

Lavelle, J., Tonboe, R., Tian, T., Pfeiffer, R. & Howe, E. (2016). Product user manual for OSI SAF AMSR-2 global sea ice concentration. *http://osisaf.met.no/docs/osisaf_cdop2_ss2_pum_amsr2\-ice-conc_v1p1.pdf*, *1.1*.

Lavergne, T., Eastwood, S., Teffah, Z., Schyberg, H. & Breivik, L.-A. (2010). Sea ice motion from low-resolution satellite sensors: An alternative method and its validation in the arctic. *J. Geophys. Res. Oceans*, *115*(C10).

Lisæter, K.A., Rosanova, J. & Evensen, G. (2003). Assimilation of ice concentration in a coupled ice-ocean model, using the ensemble kalman filter. *Ocean Dynam.*, *53*, 368-388. doi: 10.1007/s10236-003-0049-4

Long, J., Shelhamer, E. & Darrell, T. (2015, June). Fully convolutional networks for semantic segmentation. I *The ieee conference on computer vision and pattern recognition (cvpr).*

Marshall, J., Adcroft, A., Hill, C., Perelman, L. & Heisey, C. (1997). A finite-volume, incompressible navier stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.*, *102*(C3), 5753–5766. doi: 10.1029/96JC02775

Melsom, A., Palerme, C. & Müller, M. (2019). Validation metrics for ice edge position forecasts. *Ocean Sci.*, *15*(3), 615–630.

Metzger, E.J., Smedstad, O.M., Thoppil, P.G., Hurlburt, H.E., Cummings, J.A., Wallcraft, A.J., ... others (2014). US navy operational global ocean and arctic ice prediction systems. *Oceanography*, *27*(3), 32–43.

Mu, L., Yang, Q., Losch, M., Losa, S.N., Ricker, R., Nerger, L. & Liang, X. (2018). Improving sea ice thickness estimates by assimilating CryoSat-2 and SMOS sea ice thickness data simultaneously. *Q. J. Roy. Meteor. Soc.*, *144*(711), 529–538.

Müller, M., Batrak, Y., Kristiansen, J., Køltzow, M.A., Noer, G. & Korosov, A. (2017). Characteristics of a convective-scale weather forecasting system for the european arctic. *Mon. Weather Rev.*, *145*(12), 4771–4787.

Müller, M., Homleid, M., Ivarsson, K.-I., Køltzow, M.A., Lindskog, M., Midtbø, K.H., ... others (2017). AROME-MetCoOp: A nordic convective-scale operational weather prediction model. *Weather Forecast.*, *32*(2), 609–627.

Ninnis, R., Emery, W. & Collins, M. (1986). Automated extraction of pack ice motion from advanced very high resolution radiometer imagery. *J. Geophys. Res. Oceans*, *91*(C9), 10725–10734.

Owens, R.G. & Hewson, T.D. (2018). ECMWF forecast user guide. reading: ECMWF.

Palerme, C., Müller, M. & Melsom, A. (2019). An intercomparison of verification scores for evaluating the sea ice edge position in seasonal forecasts. *Geophysical Research Letters*, *46*(9), 4757–4763.

Posey, P.G., Metzger, E.J., Wallcraft, A.J., Hebert, D.A., Allard, R.A., Smedstad, O.M., ... Helfrich, S.R. (2015). Improving arctic sea ice edge forecasts by assimilating high horizontal resolution sea ice concentration data into the US navy's ice forecast systems. *The Cryosphere*, *9*(4), 1735–1745. doi: 10.5194/tc-9-1735-2015

Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J. & Haas, C. (2017). A weekly arctic sea-ice thickness data record from merged CryoSat-2 and SMOS satellite data. *Cryosphere Discuss*, 1–27.

Rostosky, P., Spreen, G., Farrell, S., Frost, S., Heygster, G. & Melsheimer, C. (2018). Snow depth retrieval on arctic sea ice from passive microwave radiometers - improvements and extensions to multiyear ice using lower frequencies. *J. Geophys. Res. Oceans*. doi: https://doi.org/10.1029/2018JC014028

Sakov, P. (2015). *EnKF-c user guide, github, retrieved 1. august 2017, https://github.com/sakov/enkf-c*.

Sakov, P. & Bertino, L. (2011). Relation between two common localisation methods for the EnKF. *Comput. Geosci.*, *15*(2), 225–237. doi: 10.1007/s10596-010-9202-6

Sakov, P., Counillon, F., Bertino, L., Lisæter, K., Oke, P. & Korablev, A. (2012). TOPAZ4: an ocean-sea ice data assimilation system for the north atlantic and arctic. *Ocean Sci.*, *8*(4), 633-656. doi: 10.5194/os-8-633-2012

Sakov, P. & Oke, P. (2008). A deterministic formulation of the ensemble kalman filter: an alternative to ensemble square root filters. *Tellus*, *60A*(2), 361–371. doi: 10.1111/j.1600-0870.2007.00299.x

Scher, S. & Messori, G. (2019, 03). Weather and climate forecasting with neural networks: using GCMs with different complexity as study-ground. *Geosci Model Dev.*, 1-15. doi: 10.5194/gmd-2019-53

Scott, M. (2018). "https://www.climate.gov/news-features/featured-images/2018-arctic-sea-ice-minimum-continues-longer-trend", accessed 05.04.2019.

Shchepetkin, A. & McWilliams, J. (2005). The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Model.*, *9*(4), 347 - 404. doi: 10.1016/j.ocemod.2004.08.002

Shelhamer, E., Long, J. & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE T. Pattern Anal.*, *39*(4), 640–651. doi: 10.1109/TPAMI.2016.2572683

Shlyaeva, A., Buehner, M., Caya, A., Lemieux, J.-F., Smith, G.C., Roy, F., ... Carrieres, T. (2016). Towards ensemble data assimilation for the environment canada regional ice prediction system. *Q. J. Roy. Meteor. Soc.*, *142*(695), 1090–1099.

Stephen, K. (2018, 01. Sep). Societal impacts of a rapidly changing arctic. *Current Climate Change Reports*, *4*(3), 223–237. doi: 10.1007/s40641-018-0106-1

Tian-Kunze, X., Kaleschke, L. & Maaß. (2016). SMOS daily sea ice thickness version 3, updated current year. *CDC, icdc.cen.uni-hamburg.de, University of Hamburg, Germany, Digital media*.

Tonboe, R., Lavelle, J., Pfeiffer, R. & Howe, E. (2016). Product user manual for OSI SAF global sea ice concentration. *http://osisaf.met.no/docs/osisaf_cdop3_ss2_pum_ice-conc_v1p6.pdf*, *1.4*.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., ... others (2012). GMES sentinel-1 mission. *Remote Sens. Environ.*, *120*, 9–24.

Urrego-Blanco, J.R., Hunke, E.C., Urban, N.M., Jeffery, N., Turner, A.K., Langen-

brunner, J.R. & Booker, J.M. (2017). Validation of sea ice models using an uncertainty-based distance metric for multiple model variables. *J. Geophys. Res. Oceans*, *122*(4), 2923–2944.

Vancoppenolle, M., Fichefet, T., Goosse, H., Bouillon, S., Madec, G. & Maqueda, M.A.M. (2009). Simulating the mass balance and salinity of arctic and antarctic sea ice. 1. model description and validation. *Ocean Model.*, *27*(1–2), 33 - 53. doi: \url{http://dx.doi.org/10.1016/j.ocemod.2008.10.005}

Van den Dool, H., Cpc, P.S. & Van Den Dool, H. (2007). *Empirical methods in short-term climate prediction*. Oxford University Press.

Wang, K., Debernard, J., Sperrevik, A., Isachsen, P. & Lavergne, T. (2013). A combined optimal interpolation and nudging scheme to assimilate OSISAF sea-ice concentration into ROMS. *Ann. Glaciol*, *54*(62), 8-12. doi: 10.3189/2013AoG62A138

Wang, L., Yuan, X. & Li, C. (2019). Subseasonal forecast of arctic sea ice concentration via statistical approaches. *Climate Dynamics*, *52*(7-8), 4953–4971.

Whitaker, J.S. & Hamill, T.M. (2002). Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.*, *130*(7), 1913–1924.

Xie, J., Counillon, F., Bertino, L., Tian-Kunze, X. & Kaleschke, L. (2016). Benefits of assimilating thin sea ice thickness from SMOS into the TOPAZ system. *The Cryosphere*, *10*(6), 2745–2761. doi: 10.5194/tc-10-2745-2016

Yang, Q., Losa, S., Losch, M., Tian-Kunze, X., Nerger, L., Liu, J., ... Zhang, Z. (2014). Assimilating SMOS sea ice thickness into a coupled ice-ocean model using a local SEIK filter. *J. Geophys. Res.*, *119*(10), 6680-6692. doi: 10.1002/2014JC009963

Yumi. (2018). *"https://fairyonice.github.io/Learn-about-Fully-Convolutional-Networks-for-semantic-segmentation.html, last accessed june 2019.*