

Developing a Deep Learning forecasting system for short term and high resolution prediction of Sea Ice Concentration

Masters' thesis in Computational Science: Geoscience, Spring 2023

Are Frode Kvanum^{1,2}

¹Department of Geosciences, University of Oslo

²Development Centre for Weather Forecasting, Norwegian Meteorological Institute

January 24, 2023

Contents

1	Introduction	3
2	Datasets	7
2.1	Region	8
2.2	Observations	11
2.2.1	Sea Ice Charts	11
2.2.2	Osi-Saf	11
2.2.3	AMSR2	14
2.3	Forecasting systems	16
2.3.1	AROME Arctic	16
2.3.2	NeXtSIM	18
2.3.3	Barents-2.5	18
3	Theory	19
3.1	Convolution	19
3.2	Batch normalization	19
4	Methodology	19

4.1	Pixel classification	19
4.2	Image to image classification	19
5	Forecast verification metrics	20
5.1	Defining the Ice Edge	20
5.2	Integrated Ice Edge Error	21
6	Impact of increased resolution on the IIEE	22
6.1	Computing distances with regards to an Ice Edge	24
7	Developing a U-Net	24
8	Model Architecture	25
8.1	CategoricalCrossEntropy-Loss	25
8.2	FocalLoss	27
8.3	Cumulative probability distribution model	29
8.3.1	Separate convolutional layers as output	29
8.4	Model Selection	29
9	physical connections	32
9.1	Variograms	32
9.2	Case study	32
9.3	Synthetic AA forcing	32
10	Comparing against physical models	32
10.1	Preparing data	32
11	Conclusion and future outlook	32
12	Supporting Figures	38

1 Introduction

The Arctic sea ice extent has continuously decreased since the first satellite observations of the Arctic was obtained in 1978 (Serreze and Meier, 2019), with an average decrease of 4% per decade (Cavalieri and Parkinson, 2012). The summer months are experiencing the greatest loss of sea ice extent (Comiso et al., 2017), with models from the Coupled Model Intercomparison Project Phase 6 (CMIP6) projecting the first sea ice-free Arctic summer before 2050 (Notz and Community, 2020). As a consequence of the sea ice retreat during the summer months, previously inaccessible oceanic areas has opened up causing an increase in maritime operations in the Arctic waters (Eguíluz et al., 2016). The expected influx of operators to the Arctic regions due to the prolonged open water season call for user-centric sea ice products on different spatial scales and resolutions to ensure maritime safety in the region (Wagner et al., 2020; Veland et al., 2021).

Current information on Arctic sea ice concentration can be discerned into several types of products with different spatial and temporal resolutions. Sea ice products designed for climate applications such as OSI-450, SICCI-25km and SICCI-50km provide daily sea ice concentration by merging observations from multiple sensors to create a historical dataset. The purpose of a climatology is to provide accurate reference data (Lavergne et al., 2019a) which can be used for e.g. forecast validation or anomaly detection. Satellite observations are also supplied as daily products, with a timeliness of a few hours on the same day and posing higher spatial resolutions than climatologies. For example, OSI-401-b (Tonboe et al., 2017) and OSI-408 (Lavelle et al., 2016) provide single sensor daily averaged sea ice concentration covering the northern and southern hemisphere, and can be used to force numerical weather prediction systems which only resolve the atmosphere (Müller et al., 2017).

Sea ice models are physically based models resolving for the growth and movement of sea ice forward in time. Standalone models such as CiCE (Hunke and Dukowicz, 1997) and neXtSIM (Williams et al., 2021) can be used in by themselves or coupled with ocean models (Röhrs et al., 2022) to create sea ice forecasting systems for short lead times. Finally, sea ice charts drawn analogously by a sea ice specialist merge recent sea ice observations from different sensors and satellites into a single daily product. The Ice Service of the Norwegian Meteorological Institute (NIS) provides regional ice charts covering the European Arctic. The product consists of polygons which are drawn to match the current resolution of the available observations, which range from 50m to several kilometers, and are assumed to have a low uncertainty due to the quality control exerted by the sea ice specialist (Dinessen et al., 2020).

The previously mentioned sea ice products serve different use cases, and it is possible to infer a correlation between the spatial and temporal resolution of a product and its application scenario for maritime end users. While lower resolution products at larger

temporal time scales can be used in long term planning, regional high resolution products delivered at a high frequency can assist strategic decision making and short term route planning (Wagner et al., 2020). However, it is currently reported by end users that available operational passive microwave satellite products are of a too low resolution, partly due to their insufficient ability to resolve leads and other high-resolution information necessary for maritime safety. Moreover, it is also reported that sea ice forecasting systems lack desired verification, are inadequate for operational use as well as being difficult to integrate with a vessel where computational resources and data-bandwidth are limited (Veland et al., 2021). Though sea ice charts provides personnel in the Arctic with information regarding where sea ice has been observed in the time after the previous ice chart has been published, the ice charts does not provide a description on the future outlook. Thus, the responsibility of interpreting the ice charts and other available sea ice information with a outlook on future development is delegated to the end-user and relies on their experience to ensure a continued safe navigation (Veland et al., 2021).

As such, a different approach to short-range sea ice forecasting may be necessary to deliver short-term sea ice information on a spatial scale that is relevant for end-users. Thus, this thesis proposes an alternative forecasting scheme that applies Convolutional deep learning in the form of a modified U-Net architecture (Ronneberger et al., 2015) to deliver a short lead time (1 - 3 days), 1km resolution forecasting product over a subsection of the European Arctic by utilizing the aforementioned Ice Charts as the ground truth. Moreover, the product is verified with regards to the position of the ice edge, which aims to demonstrate the operational relevance of the product (Veland et al., 2021; Melsom et al., 2019).

There have been made previous attempts to develop deep learning sea ice forecasting systems. Andersson et al. (2021) propose IceNet, a pan-arctic covering U-NET which predicts monthly averaged sea ice concentration (SIC) with 6 month lead time at a 25 km spatial resolution (Andersson et al., 2021). The model classifies sea ice concentration into one of the three classes open-water, marginal ice or full ice. IceNet showed an overall improvement over the numerical SEAS5 seasonal forecasting system (Johnson et al., 2019) for 2 months lead time and more, with the greatest improvement seen in the late summer months. The model is trained on SIC data provided by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea Ice Satellite Application Facilities (OSI-SAF) dataset (Lavergne et al., 2019a), as well as other climate variables obtained from the ERA5 reanalysis (Hersbach et al., 2020).

Similarly, Liu et al. (2021) propose a Convolutional long short-term memory network (ConvLSTM) which forecasts SIC with a lead time up to 6 weeks. The model uses climate variables and SIC from two reanalysis products ERA-Interim (Dee et al., 2011) and ORAS4 (Balmaseda et al., 2013), covering the Barents Sea with a domain size of 24 (latitude) x 56 (longitude). Their results showed skill in beating numerical models as well as

persistence.

Models such as those noted above consider input variables obtained from climatologies, and represent SIC on spatial scales far larger than what is needed for an operational short-term sea ice forecast. The possibility of using higher resolution input data was explored by Fritzner et al. (2020), which combined OSISAF SIC, sea surface temperature from the Multi-scale Ultra-high Resolution product, 2 meter air temperature from the ERA5 reanalysis as well as SIC from sea ice charts produced by the NIS. Fritzner et.al. developed a Fully Convolutional Network (FCN), which achieved similar performance to the Metroms coupled ocean and sea ice model version 0.3 (Kristensen et al., 2017). However, due to computational constraints of training the FCN, the subdomain was reduced to a resolution of 224 x 224 pixels which translates to 10 - 20km (Fritzner et al., 2020). Thus, the product has a limited accuracy for short term operational usage, similar to (Andersson et al., 2021) and (Liu et al., 2021).

Contrary to the authors above, Grigoryev et al. (2022) propose a 10 day lead time regional forecasting system with a 5km spatial resolution trained on a sequential (traditional) and recurrent U-Net architecture. The authors used 5km AMSR-2 sea ice concentration as the ground truth variable, and regrid atmospheric variables from the NCEP Global Forecast System (https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs.php) to match the resolution of the ground truth. Their results showed that the recurrent setup slightly outperformed the sequential architecture for predictions with a lead time up to 3 days, with both architectures significantly outperforming persistence and the linear trend. However, the sequential architecture tended to outperform the recurrent architecture for 10 day forecasts, as the recurrent model was trained without weather data as it only had a lead time of 3 days.

As mentioned in (Andersson et al., 2021; Fritzner et al., 2020), the computational cost of producing a forecast using a pre-trained model is low, such that a laptop running consumer hardware is able to generate a forecast in seconds or minutes depending on the availability of a Graphics Processing Units (GPU). This is in stark contrast to numerical sea ice models, which could run for several hours on high-performance systems (Andersson et al., 2021). Training a model is a one time expense, and can be efficiently performed on a GPU. With the increased complexity, efficiency and availability of high end computing power, smart usage of the available memory allows for model training using high resolution fields. Current GPUs have seen a significant increase in the available video memory, which allows for higher resolution data to be utilized during training. This work will exploit the recent advances in GPU development, as well as incorporating techniques to reduce the floating point precision of the input meteorological variables, circumventing a reduction of the spatial resolution as seen in previous works.

Moreover, the U-Net architecture is part of the supervised learning paradigm of machine

learning, which require labelled samples in order to train the network (Ronneberger et al., 2015). Furthermore, U-Nets perform pixel-level prediction where each pixel is classified according to a category. This work will utilize the image-to-image predictive capabilities of the U-Net to create a semantic segmentation based on its input variables simulating a forward in time propagation of the sea ice concentration akin to a physical model. This allows for the inspection of how changes to the architecture as well as input data configurations affect the behavior of the forecasting system.

In the present work, the development of a deep learning forecasting system will be explored. The choice and tuning of hyperparameters will be reasoned in light of the physical processes affecting sea ice and the surrounding variables. Furthermore, the quality of the machine learning forecasting system will be assessed against relevant benchmarks such as persistence, physical models and linear regression of the observed sea ice concentration. Due to the operational nature of the developed forecasting product, ice edge aware validation metrics such as the Integrated Ice Edge Error (Goessling et al., 2016) will be central to the performance analysis. Furthermore, this thesis aims at providing the framework for which a future operational sea ice prediction system can be built upon. As such, the choice and structure of data will be made with a potential operational transition in mind.

This thesis aims at exploring the following research questions:

- Can a deep learning system resolve regional sea ice concentration for high resolution, short lead time forecasts?
- How does a high resolution, short lead time U-Net forecasting system resolve the translation and accumulation of sea ice compared to a physical based model
- In what sense can a deep learning model be explainable / made transparent to explain the statistical reasoning behind the physical decision-making

The thesis is structured as follows. The First section will describe the datasets used, followed by the second section which will do a rundown of the methodological framework necessary to develop the U-Net as well as validation metrics used to assess forecast skill. The third section will detail the development process behind the U-Net, with the fourth section exploring the physical connections of the model. The fifth section will detail the performance assessment of the forecasts. In the sixth section, a discussion of the findings will be conducted, with the seventh and final section presenting conclusions and future outlook.

2 Datasets

[Training and validating a deep learning system requires data, which can be categorized in two distinct groups. The first group is the data known by the system, which is used during training to increase or validate model performance. Due to developing the model in such a way that it performs well against its validation data, external data is needed to validate the generalizability of the model. I.e., how well does the model perform with unknown data, which is assumed drawn from the same distribution as the data used during training. It is standard practice to arbitrarily split by a given fraction into the three datasets (training, validation, testing), as outlined above. However, due to the variable seasonal dependency of meteorological data, a naive split of the data could result in seasonally unbalanced datasets. As such, the datasets constructed for the purpose of this thesis are each covering at least a full year. Thus, no dataset is assumed to be skewed in the direction of any season.]

To facilitate the development and verification of a high resolution short-term deep learning sea ice forecasting system, several datasets covering observations and physical model forecasting systems have been chosen. When selecting appropriate datasets, their spatial resolution as well as release frequency has been considered. Even though several observational sea ice concentration products exist which cover the region of interest (see Figure 1), a lot of the satellite products based on passive microwave retrievals are of a too coarse resolution (e.g. Lavergne et al. (2019a) or Kern et al. (2019)) to be able to aid in short term decision making (Wagner et al., 2020). On the other hand synthetic aperture radar (SAR) observations such as Sentinel 1A Interferometric Wide swath ($5\text{m} \times 20\text{m}$) or Extra-Wide swath ($20\text{m} \times 40\text{m}$) are on a structure resolving spatial resolution. However the SAR, however Sentinel 1 SAR does not provide pan-Arctic coverage at a daily frequency (<https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar>, Last Accessed 20 Jan 2023).

As such, the highest resolution product with an appropriate frequency available are the sea ice charts produced by the NIS Dinessen et al. (2020). Moreover, the sea ice charts represent a forecasting product directed towards operational users. Thus, the sea ice charts will serve as the ground truth for the model. However, as a deep learning system can increase its skill by combining correlated variables as input, this thesis will explore the impact caused by including several datasets on different spatial resolutions as input predictors.

Finally, forecasts provide the deep learning system with information regarding how the domain will evolve in the period after the forecast has been initialized. Thus allowing the deep learning system to not only rely on past and present sea ice concentration, but also forecasted development. Hence, a regional numerical weather prediction system will be included to encode information about the future sea ice state through relevant variables

Dette
passer
bedre
inn
i en
train-
test-
split
model
devel-
op-
ment

Table 1: Rundown of the products used and their applications. The dashed line separates observational products (above) from forecast products (below)

Product	Variables	Training	Verification
Ice charts	SIC	Yes	Yes
OSI-SAF SSMIS	SIC trend	Yes	Yes
OSI-SAF CDR	Ice edge length	No	Yes
AMSR2	SIC	No	Yes
<hr/>			
AROME-Arctic	T2M, X-wind, Y-wind	Yes	No
NeXtSIM	SIC	No	Yes
Barents-2.5	SIC	No	Yes

which are correlated to the sea ice concentration.

The following section will perform a rundown of the satellite products as well as physical models used. Table 1 presents the different products used for this thesis, and whether the product is used to train or verify the model.

2.1 Region

The domain covered by the deep learning system, covers part of the European Arctic. The region is an intersection between the domain covered by the Ice Charts (Dinessen et al., 2020) and AROME Arctic (Müller et al., 2017) as shown in Figure (1). The domain has a 1km spatial resolution, and contains 1972×1972 equidistant grid points. Compared to the AROME Arctic grid, the model domain has a reduced southern and eastern extent, which is manually subsectioned to conform to the square grid imposed by the deep learning architecture (Ronneberger et al., 2015). Another reason for the reduced domain extent was to limit the amount of memory needed when loading data during training. Both reasons will be thoroughly discussed in later sections. Moreover, the choice of limiting the southern and eastern extent of the domain was deliberate to reduce the amount of likely sea ice concentration containing grid cells lost in the process.

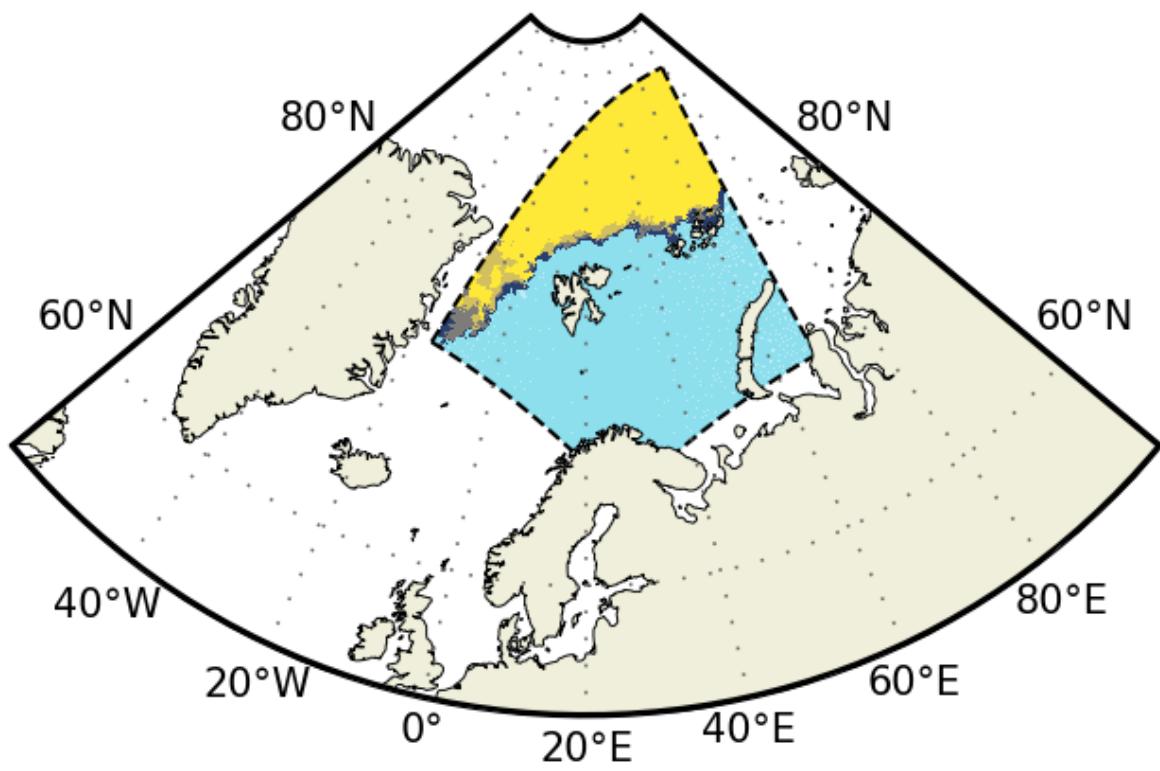


Figure 1: The model domain is shown by sea ice concentration contours retrieved from a sea ice chart (15 Sep 2022). No colorbar is shown, light blue is free open water and yellow is fast ice. Land is displayed atop the domain for visualization purposes only.

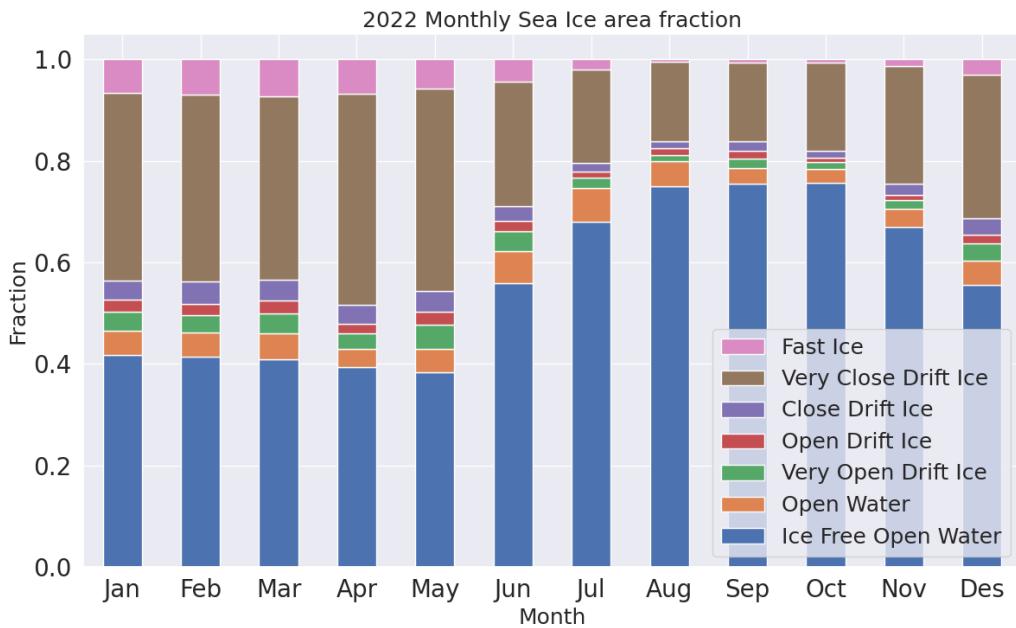


Figure 2: Monthly distribution of each concentration class as respective fraction of the total mean sea ice concentration for the sea ice charts covering 2022. [Could extend to cover larger time period (e.g. from 2011), give a more climate perspective of the sea ice evolution]

Missing figure

Få inn en figur som viser autocorrelation for IceChartsa gitt en viss periode.

2.2 Observations

2.2.1 Sea Ice Charts

The Sea Ice charts is an operational Sea Ice Concentration product provided by MET Norway. The product is manually drawn by a Sea Ice Specialist, and is distributed every workday at 15:00 UTC. The Sea Ice specialist assesses available SAR data from Sentinel 1 and Radarsat 2. However, due to the spatial variability in daily SAR coverage, visual, infrared and low resolution passive microwave observations are supplied to achieve a consistent spatial coverage (?). The Sea Ice charts are drawn in an ArcGIS production environment, and is as such intrinsically not projected onto a defined grid. Yet, the operational product available for download on [Copernicus](#) is provided as mean values on a 1km grid.

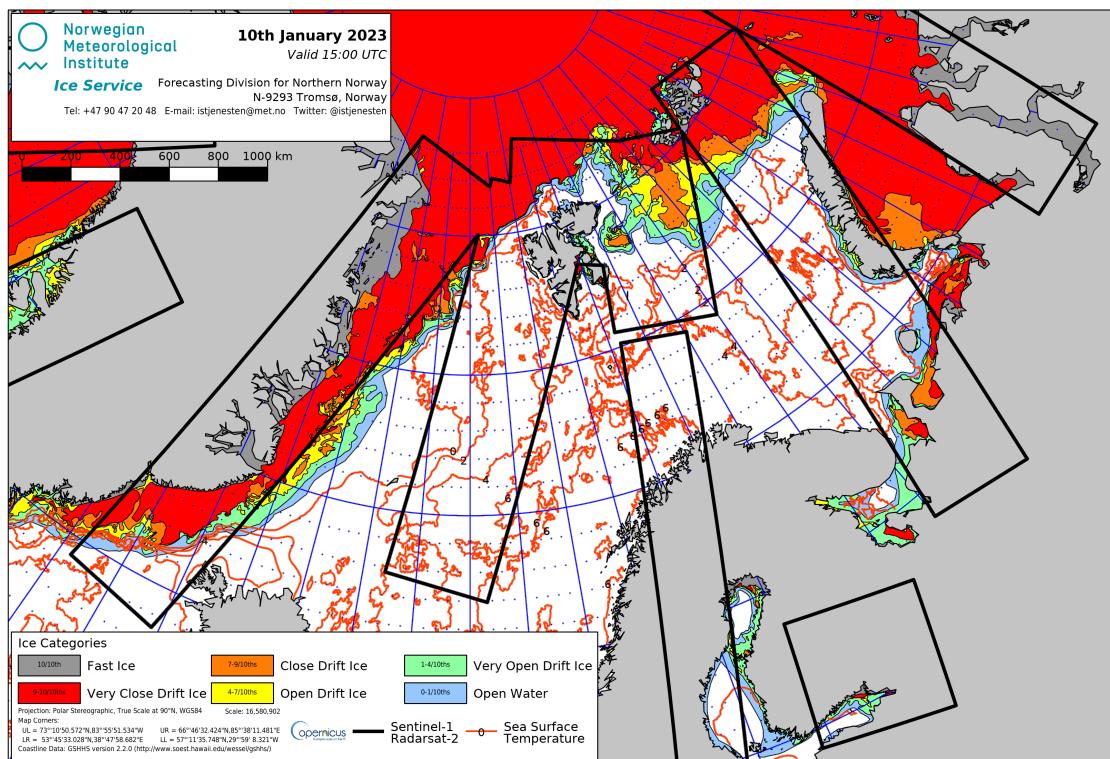
From the description of the Sea Ice charts given above, it is worth addressing the spatial inconsistency following the projection onto a uniformly sized grid. As the Sea Ice specialist draws polygons based on data from different satellite sources with a wide range of spatial resolution (80m from SAR, 1000m from visible / infrared and even lower resolution for passive microwave), the underlying uncertainty and detailed structures in the Sea Ice chart varies (?). Furthermore, I was made aware by one of the Sea Ice Analysts that time constraints also limits the hours different sections of the Ice chart is allotted. Moreover, the Sea Ice charts is an operational product aimed at end users in industries such as fishing, tourism, shipping or other maritime operations. This influences the decision-making when creating the final operational product. . As a consequence, the Sea Ice analyst spends approximately half of the total time draw polygons around the Svalbard archipelago.

In conclusion, concerning the limited resources available both with regards to data availability as well as total hours available, the Sea Ice charts represents a dataset with a spatial uncertainty that is non-uniform across a single sample, and that changes in time. In spite of that, the involvement of a Sea Ice specialist which manually assures each Sea Ice charts, the temporal consistency as well as their high resolution has led us to believe that the Sea Ice charts is the overall best Sea Ice Concentration product available for the current study region.

2.2.2 Osi-Saf

Two different Sea Ice Concentration products are used from OSI-SAF. OSI-SAF SSMIS is an operational product delivering daily sea ice concentration on the northern (and southern) hemisphere. OSI-SAF Climate Data Record (CDR) (Sørensen et al., 2021) deliver sea ice climatology beginning in 1979 (Lavergne et al., 2019a) The operational

ask
Trond
on
email



product will be used as a predictor for the model, whereas the climatology will be used only for validation purposes.

OSI SAF SSMIS

OSI-SAF SSMIS is a passive microwave product derived from the Special Sensor Microwave Imager and Sounder (SSMIS). To convert brightness temperature to estimated sea ice concentration, a hybrid approach combining the Bootstrap algorithm (Comiso et al., 1997) and the Bristol algorithm (Smith, 1996) where the prior is used over open water and the latter used for ice concentrations above 40% (Tonboe et al., 2017). The algorithm uses data from the 19GHz frequency channel (Vertically polarized) and 37GHz channel (Vertically and Horizontally polarized), which are the two lowest spectral resolution channels for the SSMIS Tonboe et al. (2017). The end product is on a 10km polar stereographic grid.

The operational OSI-SAF SSMIS data is used to compute a coarse resolution (with respect to the ice charts) linear sea ice trend in each grid cell, with a length of 3 to 7 days. The idea behind the computed trend is to encode multiple time-steps of sea ice concentration fields into a single 2d-array, in line with the lack of temporal awareness of the deep learning architecture as well as the limiting the amount memory needed as information from multiple large arrays are contained in a single field. Furthermore, the ice concentration trend is computed from a separate sea ice product than the ice chart, with the intent to supply the model with correlated but not overlapping information, as the current day ice chart is already used as a predictor. The coarser resolution also contribute to the OSI-SAF trend serving as complementary information to the ice charts, as the coarse resolution make the trend less resolvent of the local variability which is seen in the ice charts. As such, the trend serves as a indicator of where the sea ice growth is occurring.

The temporal length used when deriving trend will have an impact as to how accurate the computed trend reflects the current growth and retreat zones, especially with regards to the volatile position of the ice edge on a daily timescale but also due to the seasonal variability of the ice area (Holland and Kimura, 2016). Hence, a too large lookbehind would cause a decorrelation between the current sea ice concentration and computed trend. On the other hand, the motivation behind computing the trend is that there exist a large and positive autocorrelation for sea ice concentration on a short time-range.

OSI SAF Climate Data Record

As briefly mentioned in Section (1), OSI-SAF Climate Data record combines observations made by multiple sensors (SMMR, SSM/I, SSMIS) as well as numerical weather prediction fields which are utilized to correct for the atmospheric conditions. Two versions of the dataset has been used, version 2 (OSI-450) which covers (2011 - 2015), and

the interim version (OSI-430-b) which cover (2016 - 2020) (<https://osisaf-h1.met.no/osi-450-430-b-desc>) (Last Accessed 18 Jan 2023). The major differences between OSI-450 and OSI-430-b is that the interim product uses observations from different SSMIS sensors than OSI-450, OSI-430-b also uses operational ECMWF data whereas OSI-450 uses fields from ERA-Interim (Dee et al., 2011). However, the same algorithms are used to process both versions, ensuring consistency (Lavergne et al., 2019b), thus both products will hereby be referred to in tandem as OSI SAF CDR

OSI SAF CDR is provided with a 25km spatial resolution on a Lambert Azimuthal Grid projection (Sørensen et al., 2021). The sea ice concentration data retrieved has been used to compute a climatological ice edge length for each day of the year, applying a daily mean across the time period (2011 - 2020). The ice edge length has been computed according to (Melsom et al., 2019), which will be derived in Section (not yet labelled). Note that though OSI SAF CDR provide a pan-arctic distribution of sea ice concentration, the data has been regridded onto the study region domain with the AROME Arctic projection and a 25km grid spacing before computing the ice edge length.

As can be seen in Figure (3), the Arctic sea ice edge experiences a strong seasonal variability. The computed climatological ice edge will be used as a normalization factor to ensure that the seasonal influence on ice edge length dependant verification scores is mitigated (Goessling et al., 2016; Zampieri et al., 2019; Palerme et al., 2019). Another benefit from utilizing a single ice edge length is to ensure that different sea ice products are normalized according to a common and independent sea ice length. Furthermore, it will be shown in a later section that the Integrated Ice Edge Error (Goessling et al., 2016) (Not yet derived) normalized by the ice edge length is correlated with the resolution of the ice edge length, proving the validity of normalizing using a common, coarse resolution ice edge length.

2.2.3 AMSR2

The Advanced Microwave Scanning Radiometer 2 (AMSR2) data utilized for this thesis is the sea ice concentration product from the University of Bremen (<https://seaice.uni-bremen.de/sea-ice-concentration/amsre-amsr2/>) (Last Accessed 18 Jan 2023). AMSR2 is a passive microwave sensor observing the microwaves emitted by the Earth, similar to **OSI SAF SSMIS**. AMSR2 is located on the JAXA GCOM-W1 satellite Melsheimer (2019), and is retrieved using the ASI algorithm Spreen et al. (2008). The algorithm uses data from the 89GHz channel, which is the band posing the highest spectral resolution, in both polarizations to determine the sea ice concentration. Bands at lower spectral resolutions are only used as weather filters, which can mask out false sea ice detected in the open ocean Spreen et al. (2008). The resulting data is a pan-arctic sea ice coverage with a spatial resolution of 6.25km.

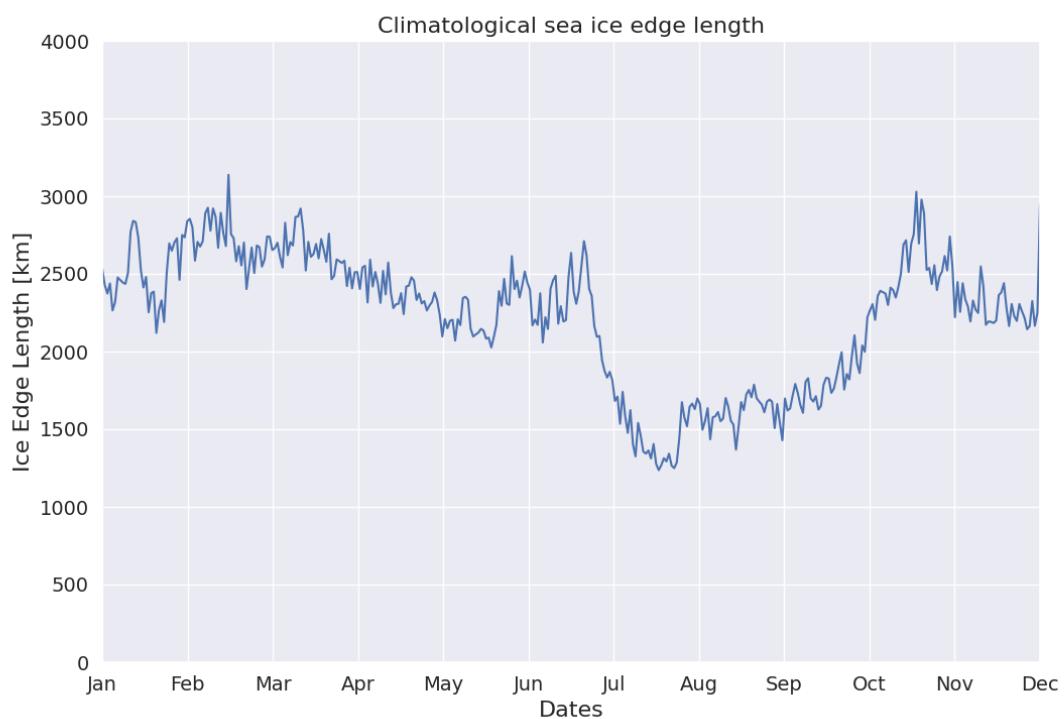


Figure 3: **FIX NONSENSICAL XTICKLABELS** Seasonal variability of the Arctic climatological ice edge length computed from satellite observations during the period 2011 - 2020

As the sea ice charts are treated as the ground truth during training of the deep learning model, it can be assumed that the model is best at predicting sea ice concentration distributions similar to those found in the training data. As such, the AMSR2 data will serve as an external ground truth, and will be used for validation only. Thus, the performance of the deep learning system can be inspected with regards to a ground truth less similar than the sea ice charts, which measures the generalizability of the model.

2.3 Forecasting systems

2.3.1 AROME Arctic

AROME Arctic is a non-hydrostatic, convection resolving high-resolution weather forecasting system which covers the European Arctic (Müller et al., 2017). The model covers the European Arctic similarly to Figure (1) which is the same domain though reduced, with a spatial resolution of 2.5km and 65 vertical levels. AROME Arctic uses different data assimilation techniques for the atmosphere and surface variables from the model background, with 3DVAR combining atmospheric background, HRES and observations and optimal interpolation on the surface background and observations to initialize the forecast analysis Müller et al. (2017). As previously mentioned, variables correlated to the sea ice concentration can aid in improving the predictive capabilities of a deep learning system. While observational products described above such as the ice charts (Dinessen et al., 2020) and OSI SAF SSMIS (Tonboe et al., 2017) describe the condition and dynamics of the sea ice concentration. Integrating weather forecast data as part of the model input can be used to describe the interaction between sea ice and atmospheric variables, thus providing correlated variables to the sea ice concentration. For the scope of this thesis, 2 meter temperature as well as 10 meter wind in the X and Y component has been selected.

Near surface winds influence the sea ice drift speed, with the sea ice in the European Arctic displaying a moderate to strong correlation between the sea ice drift speed and the wind speed during winter (Spreen et al., 2011). Moreover, sea ice drift speed is shown to be inverse proportional to the sea ice concentration (Yu et al., 2020). i.e. low concentration sea ice classes tend to have a higher drift speed than high concentration sea ice classes, though both classes display an increased drift speed given an increased near surface wind speed. Thus, including the X and Y component of the near surface wind from AROME Arctic enables the network (deep learning system) correlate sea ice dynamics with a high resolution proxy for the sea ice drift for the forecasted period.

Similarly, surface temperature influence the sea ice dynamics by melting or facilitating for sea ice growth (Hibler, 1979), for example through the formation of melt ponds on top the sea ice. The 2 meter temperature from AROME Arctic is intended to serve as a proxy

for the sea ice growth, by including a spatial distribution of temperature to the model. Which may be correlated to areas in the model domain experiencing mean positive (melt) or negative (growth) temperatures during the forecast period.

AROME Arctic is shown to have lower RMSE for both 2-meter temperature and 10 meter zonal wind speed than both the deterministic (HRES) and ensemble (ENS) forecast as well as ERA-Interim from ECMWF, for all months when compared to measurements from 89 stations located in Finnmark, Svalbard as well as Jan Mayen and Bjørnøya (Müller et al., 2017). Hence, it is reasonable to assume that extracting the wind and temperature fields from AROME Arctic will provide the most precise information with regards to the strength and spatial location, compared to global medium range numerical weather prediction systems such as the ECMWF Integrated Forecasting System Cycle 47r3 (Haiden et al., 2022). However, it is noted that operational numerical weather prediction systems such as those described by Müller et al. (2017) and Haiden et al. (2022) are in constant development, with new improvements added without any retroactive effect for previous data. Firstly, the comparison made in Müller et al. (2017) was with HRES and ENS as of Cycle 38r2 Bauer et al. (2013) is not necessarily representative of the current state of both products. Secondly, significant advances in model development may cause data before and after the implementation date to be inconsistent, e.g. by introducing a permanent shift in bias for a variable. Problems regarding model updates could be avoided by using variables from a reanalysis product such as CARRA (Køltzow et al., 2022). However, CARRA similarly to other reanalysis products are not delivered with a daily frequency (see <https://climate.copernicus.eu/copernicus-arctic-regional-reanalysis-service>, Last Accessed 21 Jan 2023), which would inhibit the operational aspect of the developed deep learning system. It is also noted that CARRA specifically only pose a 30 hour lead time, which limits the desired "up to 3 day" lead time desired for the developed deep learning system.

With regards to model development, a major development in AROME Arctic in terms of temperature representation over sea ice occurred 10 Oct 2018 (AROME Arctic Changelog, Last Access 21 Jan 2023), in the form of a *snow on ice* variable. As this change is expected to have changed the distribution of 2 meter temperature significantly, especially over sea ice covered grid cells (Yurii Batrak, Pers.Comun.), it has been opted to only consider near surface temperature data from AROME Arctic from 2019 and onwards. This decision is made to avoid having a shift in temperature distribution present in the data, which would exert a negative impact in training the deep learning model.

Though the different datasets in Table (1) has been chosen with the intention to serve as independent products without any intra coupling, it is noted that the sea ice observations used to compute the sea ice concentration trend (Tonboe et al., 2017) is also used to force AROME Arctic with sea ice concentration at the initial timestep (Müller et al., 2017). It is suboptimal to provide input parameters derived from other input parameters, as their

correlation may render one of the input parameters obsolete in terms of additional information the deep learning system will infer from the "redundant" predictor. Nonetheless, it is assumed that the impact of the sea ice concentration forcing is low when combined with other surface forcings during the assimilation process. Furthermore, as the sea ice concentration is kept constant at all timesteps (Müller et al., 2017), the correlation between sea ice concentration and atmospheric variables can be assumed to be decaying with time. Thus, both products will be used as input variables, and their overlap is assumed to tend towards zero.

2.3.2 NeXtSIM

The NeXtSIM-f forecasting system uses a standalone sea ice model (NeXtSIM), and is not coupled with an ocean model (Williams et al., 2021). Furthermore, NeXtSIM differentiates itself from comparative physical sea ice models as it does not apply a rheology based on the Viscous-Plastic scheme. Note that the rheology of a sea ice model refers to how the model relates ice deformation and ice thickness with the internal stresses in the ice (Hibler, 1979). internal stress. Instead, NeXtSIM applies a brittle sea ice rheology, specifically the Maxwell elasto-brittle rheology which treats the sea ice as a brittle material rather than a viscous fluid (Dansereau et al., 2016).

2.3.3 Barents-2.5

Barents-2.5, (hereby Barents) is an in-development operational coupled ocean and sea ice forecasting model at MET Norway (Röhrs et al., 2022). The model has been in operation since September 2021. Barents poses the same resolution and projection as AA, i.e. Lambert Conformal Conic with a 2.5km resolution (Röhrs et al., 2022; Müller et al., 2017). Furthermore, Barents also forecast with a lead time up to 66 hours, which is the same as AROME Arctic. Since Barents covers the same spatial domain as the deep learning system and forecast with a lead time close to three days, its predicted sea ice concentration will be used for validation purposes.

The sea ice model used in Barents is the Los Alamos sea ice model (CICE) version 5.1, which uses an Elastic Viscous Plastic sea ice Rheology (Hunke et al., 2015). Thus, the CICE model represents sea ice as a viscous fluid which creeps slowly given small stresses and deforms plastically under large stress. It is also noted that the elastic behavior was introduced to benefit the numerical aspects of the model, and can be considered unrealistic from a physical point of view (Hunke and Dukowicz, 1997).

Barents includes an Ensemble Prediction System with 6 members executed for each of the four model runs situated at (00, 06, 12 and 18) (Röhrs et al., 2022). As part its

forcing routine, Barents performs non-homogenous atmospheric forcing of its ensemble members, with one member of each ensemble being forced with AA while the rest of the members is forces using atmospheric data from ECMWF. As such, the member forced with AA seem to perform best with regards to ocean currents, but the atmospheric forcing's impact on SIC performance is unknown at the time of writing (Johannes Röhrs, 2022, pers. commun.). However, there is generally little spread within one ensemble with regards to sea ice (Röhrs et al., 2022).

The data assimilation scheme applied for Barents is a Deterministic Ensemble Kalman filter, which solves for the analysis with a background error covariance matrix estimated as the variance of the ensemble of background members (Röhrs et al., 2022). Furthermore, it has been expressed by the developers of Barents that the model performance was unsatisfactory up until May / June 2022 due to spin up time of the data assimilation system (Johannes Röhrs, 2022, pers. commun.). As such, forecasts initiated prior to May 2022 will not be assessed for validational purposes due to the expected shift in performance as expressed by the model authors.

3 Theory

3.1 Convolution

3.2 Batch normalization

4 Methodology

4.1 Pixel classification

4.2 Image to image classification

Architectures such as Krizhevsky et al. (2012) could be used for semantic segmentation, given a sliding windows approach across the image to be classified. However, this approach would prove to time consuming, as each pixel would have to be classified independently, each pixel would only have a receptive field limited by the extent of the sliding window

Wrigglesworth
2011
the persistence of sea ice anomalies are very high at weekly - sub monthly timescales, making it difficult to beat

and the edges would be difficult to classify. Thus, network architectures such as Long et al. (2014) and Ronneberger et al. (2015) provide a translation invariant framework for image to image prediction.

The U-Net architecture was originally proposed by Ronneberger et al. (2015) in 2015.

5 Forecast verification metrics

A robust verification scheme is essential to gain insight into how the developed forecasting product performs. Both from the point of view of a developer which aim to increase the skill of the prediction but also from the user which may utilize the verification score to assess the quality of a given forecast Casati et al. (2008). In the context of Sea Ice forecasting, a spatial field of continuos or discrete sea ice concentration is predicted, the latter being the case for the current work. Given the uneven distribution intra sea ice concentration classes as well as sea ice compared to ice free open water, simply comparing pixels for correctness would be biased by the large portion of open water and result in difficult to interpret values devoid of physical reasoning. Furthermore, as the rate of maritime activity such as commercial shipping increases in the Arctic due to the sea ice decline Ho (2010), having user relevant metrics can aid and alleviate the risks surrounding Arctic navigation. As such, several studies have proposed calculating the position of the ice edge as a user relevant metric which also provides information of the distribution of the Sea Ice Concentration Dukhovskoy et al. (2015); Goessling et al. (2016); Goessling and Jung (2018). However, there is no agreement with how to best calculate the position of the Ice Edge, with the currently available metrics posing different advantages/disadvantages Palerme et al. (2019); Melsom et al. (2019). For the purpose of this thesis, The ice edge position and length will be calculated according to (Melsom et al., 2019, Melsom 2019 et.al), whereas the IIEE originally proposed by Goessling. H. Goessling et al. (2016) will also be utilized.

Syk
udoku-
mentert
pås-
tand,
må
moder-
eres

5.1 Defining the Ice Edge

The ice edge for a given Sea Ice Concentration product is derived on a per pixel basis, and defined as the grid cells which meet the condition

$$c[i, j] \geq c_q \wedge \min(c[i - 1, j], c[i + 1, j], c[i, j - 1], c[i, j + 1]) < c_e \quad (1)$$

i.e. a pixel is marked as a ice edge pixel if the current pixel itself is larger than some given concentration threshold c_e and the minimum of the pixel's 4-neighbors is less than the

same threshold. Following this notation, E is defined as the set of grid cells which constitutes the ice edge Melsom et al. (2019). Moreover, the marked grid cells each contribute to the total length of the ice edge, with each pixel's length contribution determined based on the number neighbors also marked as an ice edge pixel. Consequently, a neighborless pixel is assumed to yield a contribution the length of the diagonal to the ice-edge ($l = \sqrt{2}s$) where s is the side length of the pixel. A pixel with one neighbor a contributes a mixed horizontal - diagonal length $l = \frac{s+\sqrt{2}s}{2}$. Finally a pixel with two or more neighbors contributes with a pixel side-length $l = s$. The final length of the ice edge length then become

$$L = \sum_{e \text{ in } E} l_e \quad (2)$$

where E represent the set of pixels constituting the ice edge as previously mentioned Melsom et al. (2019).

5.2 Integrated Ice Edge Error

The Integrated Ice Edge Length (IIEE) is an error metric which compares the forecast to some ground truth target Goessling et al. (2016). The metric is defined as

$$\text{IIEE} = O + U \quad (3)$$

where

$$O = \int_A \max(c_f - c_t, 0) dA \quad (4)$$

and

$$U = \int_A \max(c_t - c_f, 0) dA \quad (5)$$

with c_t and c_f being the target and forecast concentration respectively, attaining a value of 1 if the concentration for a given pixel i above a set threshold, and 0 elsewhere. From the definition of the metric, it can be seen that the IIEE is a sum of the forecast overshoot and undershoot compared to the ground truth target. For the current work, the IIEE is an easily interpreted metric as it quantifies the total forecast error and reports on the error spatially.

As an additional remark, note that O the current implementation in Equation (4) is given as an area with sidelength dA , and is computed as one product against some other product. Conversely, for this work subscripts f and t denote forecast and target respectively, with the directionality of the computations in Equation (4) and (5) indicating that the forecast is inspected with respect to the target. However, the metric can and has also been used to define the set of pixels which constitutes its area. To clearly distinguish between the area

O and the set of pixels used to compute O , A^+ will be used to note the latter. Similarly, A^- will represent the set of pixels constituting U . Finally, it can be seen that A^+ and A^- represent the False Positive and False Negatives of the forecast respectively.

Furthermore, the IIEE can be combined with the length of the Ice Edge which was derived in the previous section 5.1. Thus, the metric is seasonally normalized, assuming that the IIEE and Ice Edge Length is seasonally correlated.

6 Impact of increased resolution on the IIEE

From both the definition of Equation (1) and (3), it can be seen that there is a dependance on the number of pixels which constitutes the ice edge. However, what effect would a change of resolution, i.e. change in number of pixels, have on the IIEE / ice edge length ratio? To answer this question, the "two day" ice edge targets were compared against persistance on all valid two day forecasts samples for the period 2019 - 2021. Furthermore, the original target resolution of 1km will be assessed, as well as a regridded product downscaled onto a 10km resolution.

The Pearson correlation coefficient was computed directly from the DataFrames containing the metrics for both 1km and 10km resolution, using the statistical Python package Pandas pandas development team (2020); Wes McKinney (2010). For clarity, the correlation between mean_length and IIEE for 1km and 10km was calculated. From the resulting computations, the correlation coefficient gets reported as $r_{mean_length} = 0.9620$ and $r_{IIEE} = 0.9995$. Furthermore, the IIEE divided by the mean length correlation is reported as $r_{normalized_IIEE} = 0.9667$. Finally, Figure (4) display the mean monthly IIEE computed along the inspected three year period.

By inspecting Figure (4) in conjunction with the reported correlation coefficients, it can be seen that increasing the working resolution of the dataset from 1km to 10km has a negligible impact on the reported metrics. Though the 1km ice edge is about 2 times longer than the 10km, coarser resolution ice edge, it does not impact the overall stability of the normalized IIEE metric. Moreover, the normalized IIEE is reduced for the 1km grid. This could be a consequence of each pixel's spatial extent covering less area. Note that the from the definition of the IIEE given in Equation (3), the un-normalized metric is dependent on pixel spatial size. Thus, the number of pixels constituting the sum is inverse proportional with the size of each pixel, hence keeping the stability of the values. However, a discrepancy may arise due to the coarser resolution pixels covering a larger area, thus losing out on the fine details. However, for the current work, the ratio of 1km IIEE and 10km IIEE is 0.9938, indicating that they are close to equal.

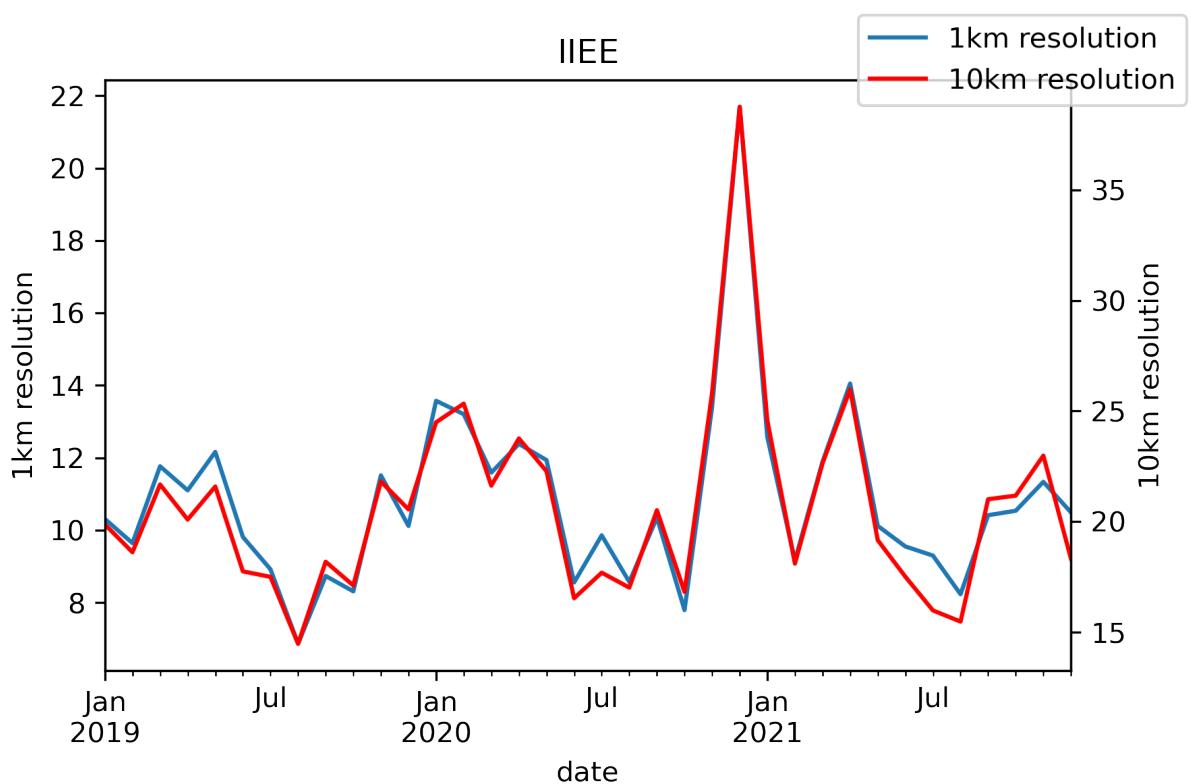


Figure 4: Monthly mean of Normalized IIEE spanning 2019 - 2021

6.1 Computing distances with regards to an Ice Edge

For the scope of this thesis, the above defined set E will be subscripted with f and t which represents *forecast* and *target* respectively. Hence, E_f is the set of pixels which constitutes the forecasted ice edge, with E_t defined analogously. With this definition, the euclidean displacement of the target ice edge E_t from all forecasted ice grid cells E_f has been defined as the following in Melsom et al. (2019)

$$d_t^n = \min \left(\forall e_f \in E_f : [(x_f - x_t^n)^2 + (y_f - y_t^n)^2]^{\frac{1}{2}} \right) \quad (6)$$

where x and y represent the coordinates of the grid cells and n the index of the pixel inspected in E_t . The opposite definition, i.e. the euclidean displacement of the forecasted ice edge in terms of the target ice edge is defined equally.

Equation (6) can also be generalized to hold for other use cases, not only comparing the displacement of two ice edges. Given an IIEE which has been separated into A^+ and A^- , and a target ice edge E_t , the distance between the nearest ice edge pixel e_t two a misclassified pixel (a^+ or a^-) can be defined as, (using a_+ as an example)

$$d_n = \min (\forall e_t \in E_t : \text{distance}(e_t, a_n^+)) \quad (7)$$

where distance is used to define som arbitrary distance metric, n denotes the pixel index in A^+ .

7 Developing a U-Net

The model developed for the two day prediction is based on the SimpleUNET architecture, though with a different sized Input layer to accommodate for the changed dataloader. The dataloader has subsequently been changed to appropriately select the correct fields from the .hdf5 samples and appoint them as input or target variables. As a result of using three variables of two days mean AA forecast, as well as sst, land-sea-mask and current time-step ice chart, the total number of predictors fed into the model is 9. Moreover, the resolution of all fields are kept at 1km, though their spatial extent is limited to (1920 x 1840). This resolution and spatial size conserves (almost) the entirety of the west-east axis of the AA domain. However, the southern border is raised by 450km compared to the AA domain. There are two main motivations behind readjusting the spatial extent of the predictors and targets.

Nevne
stør-
relsen
på
grid,
hva
som
blir
sett
på og
hvor-
for
dette
ble
valgt
som
study
area
(oper-
ational
cov-
erage,
U-Net
down-
sam-
ple
con-

1. The spatial extent of the input domain has to be divisible by the reducing factor enforced by the MaxPooling operation performed in the encoding component of the UNET.
2. The southern latitudes covered by AA has a proportionally skewed Sea Ice / Ice Free open water ratio, as exemplified in Figure (5). Increasing the southern bounding latitude of the subdomain thus decreases the number of guaranteed ice free pixels, which in turn decreases the skewness towards the ice free open water class for the UNET.

8 Model Architecture

The model architecture follows an encoder - decoder structure, commonly referred to as a U-NET Ronneberger et al. (2015) due to its shape funnelling the spatial data to coarser resolution, which resembles the letter "U". The current U-NET implementation follows that of Ronneberger et.al, though it has been modified with batch normalization after each convolution operation to ensure a more stable gradient flow. The weights of the model are Kaiming-He initialized He et al. (2015), as the activation function used throughout the network is the ReLU function Nair and Hinton (2010). The final output of the model is a (1920, 1840, 7) tensor containing softmaxed probabilities along its final axis.

8.1 CategoricalCrossEntropy-Loss

As the title suggests, these runs of the model involved using CategoricalCrossEntropy as the loss function for multi-class image segmentation. Categorical Cross Entropy loss is defined as

$$CE = - \sum_i^C y_i \log (\hat{y}_i) \quad (8)$$

where C denotes the number of available classes, y the ground truth and \hat{y} a prediction of y . Note that as y is onehot-encoded, the formulated function only contributes to the overall loss with the log of the predicted probability of the correct class according to the ground truth.

Two variants of the previously described model have been trained with the CategoricalCrossEntropy described in equation (8). The first model was trained with an encoder consisting of 4 convolutional blocks with channel dimensions (64, 128, 256, 512). The

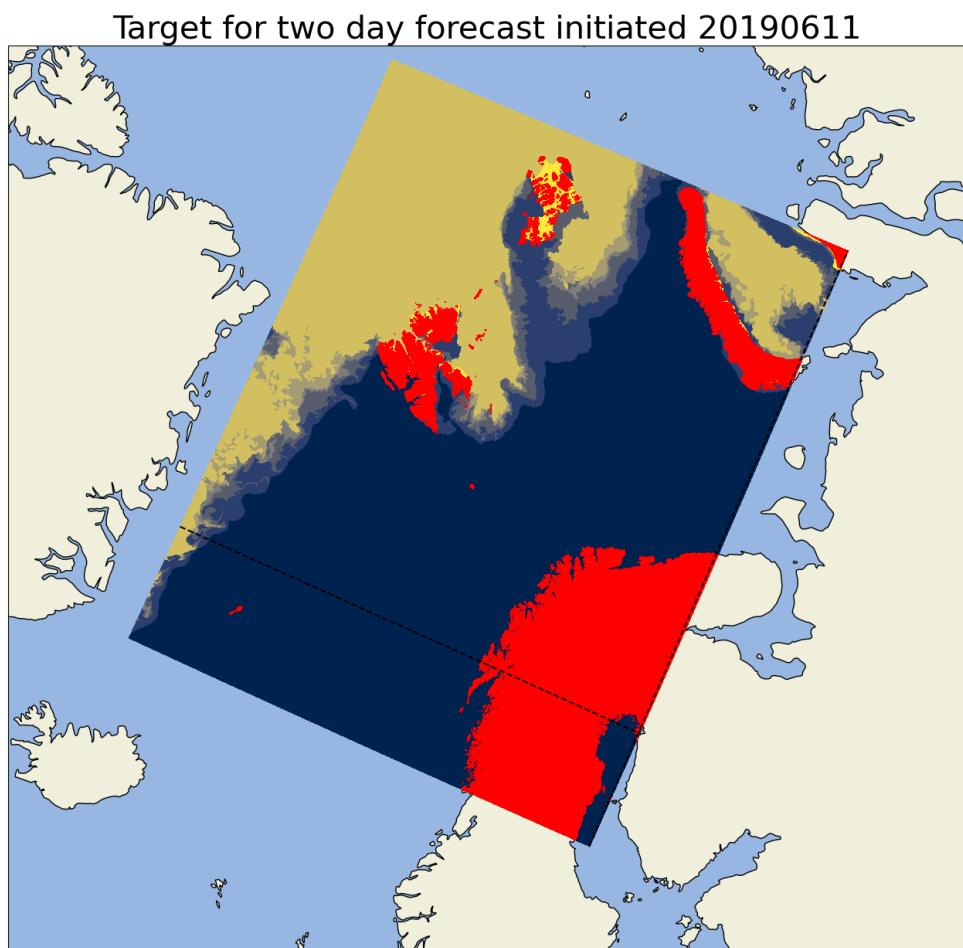


Figure 5: Example sample displaying an Ice Chart on a 1km Arome Arctic projection. Note the horizontal and vertical dashed black line which indicate the domain subsection used by the UNET

second model consisted of 5 convolutional blocks, with an identical architecture except for the last convolutional block increasing the channel dimension to (1024). Example outputs as well as target can be seen in Figure (6).

By inspecting Figure (6), two observations can be made. The first observation is regarding how the model complexity affects how it fit to the data. By comparing Figure (6a) with (6b), it can be seen that the latter is resolving the finer structures of the ice edge to larger extent than the prior. Though the overall correctness is left to be discovered, this shows that increasing the depth of the encoder (increasing the trainable parameter count from 7 million to 31 million) is reflected by the model preserving the details of the ice edge structure. Though it is non-trivial to say why the 1024-model preserves the details to a larger extent than the 512-model, it does follow from the U-Net architecture that a deeper encoder (higher channel count and more convolutional blocks) is better at describing "WHAT" is in the image compared to the shallow-layers, which include a larger amount of spatial information and tells the model to a larger extent "WHERE" things are in the model.

The second observation made from inspecting both forecasts is their inability to represent classes 2 and 3. This likely arises from the general movement-pattern of the sea ice, where the intermediate classes are much less likely to appear than the edge-most classes. Furthermore, the sea ice is much more likely to represent a wider range of concentration classes in the intermediate ice edge region over time, making it more difficult for the network to confidently predict those classes compared to the more probable classes. As can be seen by the network immediately predicting class 4 after class 1, creating an artificial cut-off region. However, to what extent the intermediate classes are predicted has not been inspected directly, though it is likely to assume that they are predicted though with a lower confidence than that of class 4 (which is consequently why it is visualized, as the most probable class is chosen regardless).

This may have a source

They should be

8.2 FocalLoss

The focal loss is derived as a generalization of the Cross Entropy Loss listed in Equation (8). The intent of the loss function is to downweight the easy to predict samples, while focusing on the hard to predict samples by allowing their gradient to have a higher impact on the network Lin et al. (2017). Mathematically, focal loss is defined as

$$FL = - \sum_i^C \alpha_i (1 - \hat{y}_i)^\gamma y_i \log (\hat{y}_i) \quad (9)$$

where α is a balancing parameter, γ is the focusing parameter ($\gamma = 0 \rightarrow CE$), with the

Include figure showing focal loss output, discuss implications of using this loss function

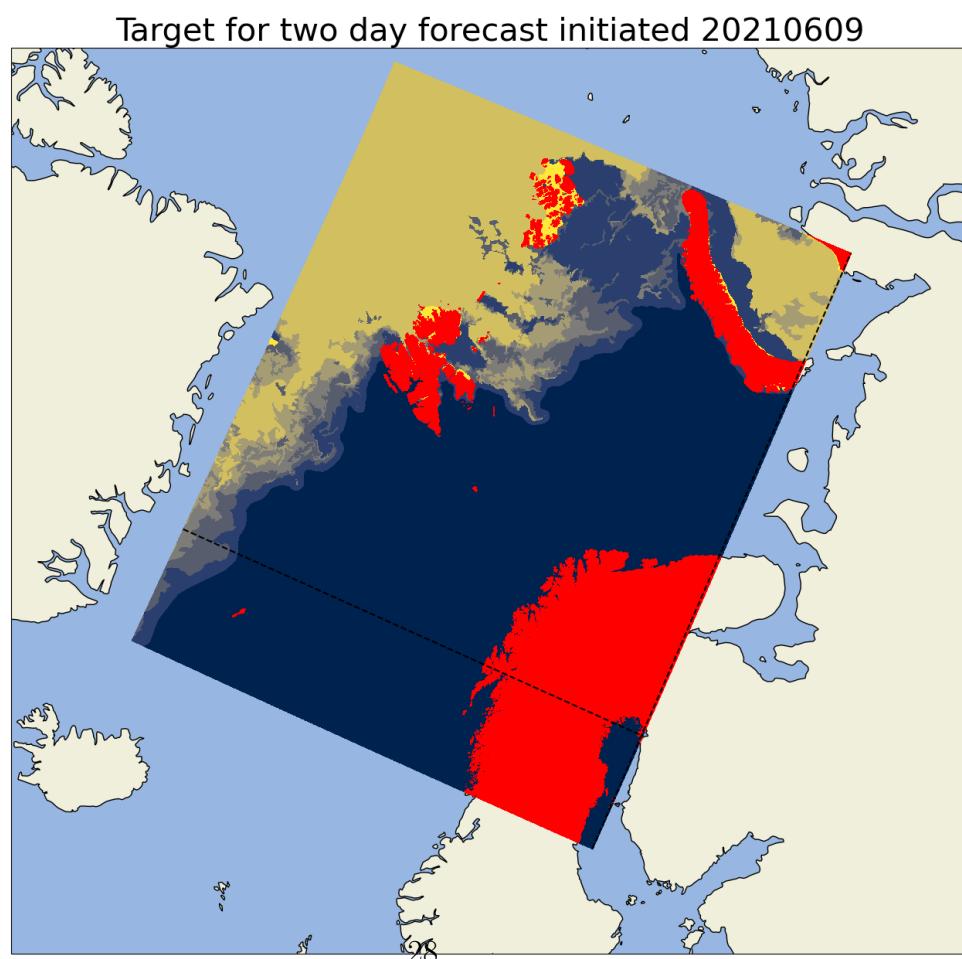
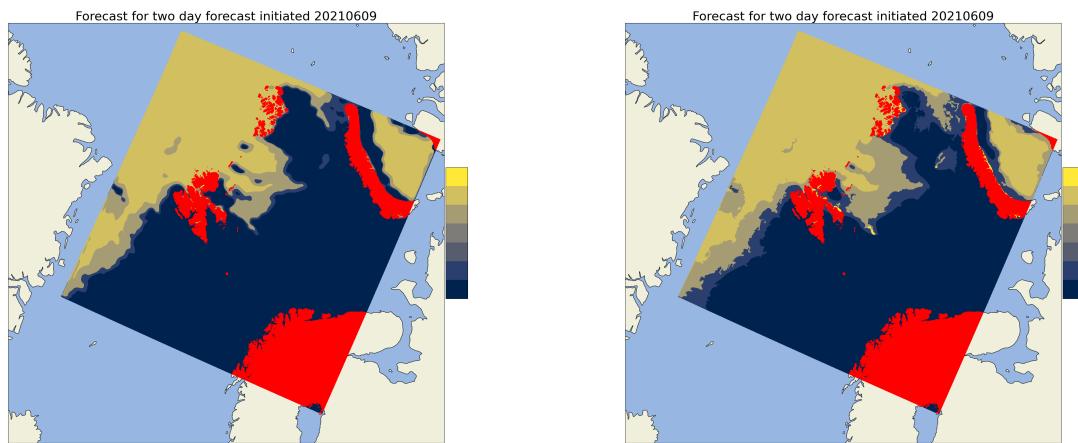


Figure 6: Example forecast attempt made by model_512 and model_1024 09-06-2021

rest similar as Equation (8).

By inspecting Equation (9), it can be seen that predictions that the model is quite confident in making, i.e. $\hat{y}_i \rightarrow 1$ send the Focal Loss towards zero. For the current application, the assumptive motivation is that this affects (by reducing) the contribution made by the Ice Free Open Water pixels as well as the Very Close Drift Ice (class 6), which are the most represented classes in the CE loss model seen in Figure (6). Consequently, as the loss contributions of the most likely (and most represented classes) is reduced, the harder to predict (both due to being less represented and due to sea ice movement) have a larger impact on the overall loss propagating backwards throughout the model. As a result, these intermediate classes should be predicted as the most likely class, resulting in a less sharp ice edge which closer represent the Ice Charts.

8.3 Cumulative probability distribution model

8.3.1 Separate convolutional layers as output

8.4 Model Selection

During the training of a deep learning system, there exists several different ways to save a state of the model during training. A naive approach would be to let the model train all predetermined epochs, and save the weights of the model at the end of the final epoch. However, this approach would be indifferent to whether the model has converged, generalized or overfitted and is thus an inadequate way to save the weights. The Tensorflow Keras API supplies functions which can be used customize the training loop in the form of [callbacks](#), with the EarlyStopping and ModelCheckpoint callbacks relevant for model selection Abadi et al. (2015). EarlyStopping is a technique which ends the training loop when it detects that a monitored values has stopped decreasing. On the other hand, ModelCheckpoint continuously saves the model if a certain condition is met, without terminating the training loop. Both callbacks support monitoring the validation loss as the metric in which to optimize the model. However, a custom metric such as yearly mean IIEE Goessling et al. (2016) could be monitored instead.

To aid in model selection, I developed a custom callback which computed the Normalized IIEE with respect to a climatological Ice Edge length derived from ten years of OsiSaf

Discuss difference in dataloader, same dataset is used differently

Data exists, start writing

data , following the observation in 6 that IIEE is correlated across spatial resolutions. The callback computes said metric for all samples and reduces them to a yearly mean of the validation set. Similar to the aforementioned callbacks, the developed callback is executed at the end of an epoch where it computes the mean Normalized IIEE for all predicted samples from the validation set, which it appends to the *logs* dictionary used by Tensorflow to keep track of other computed metrics, such as loss and validation_loss for the current case. Thus, the newly developed callback would allow for model selection based on Normalized IIEE, as well as the already computed validation loss.

When comparing different models to asses their performance, this project will frequently compare their Normalized IIEE as the metric is Normalized by the ice edge, thus reducing the seasonal variability of the Metric Palerme et al. (2019) . As such, it would be beneficial to select a model based on its Normalized IIEE validation performance. With the above callback, such a selection is possible. However, including the IIEE verification metric as is done in the above callback increases training-time of ten epochs from \approx two hours without the IIEE callback to \approx 24 hours with the IIEE callback. As 20 epochs is currently an adequate number of epochs at the time of writing , it would be too computationally costly to select a model based in its validation Normalized IIEE performance.

On the other hand, it can be seen by inspecting Figure (7) that the Normalized IIEE tend to evolve conjunctually with the validation loss, in the current case defined as the mean cross entropy of all validation samples. Furthermore, the validation loss and Normalized IIEE in Figure (7) have a correlation of 0.82 with regards to epoch. Note that this has been calculated only using the numbers present in Figure (7). As such, there is reason to believe that selecting a model based on its validation loss, which is quick to compute, would result in a generalized model which may also excel at lowering its Normalized IIEE.

When selecting the best model, this project will apply the ModelCheckpoint callback with regards to validation loss as outlined above. ModelCheckpoint is preferred compared to EarlyStopping, as interrupting the training loop early may result in an "undercooked" model. E.g. the weights in earlier model layers are adjusting slower than later weights, giving the impression that the model training has reached a plateau which causes the model to stop. Whereas if the model where to continue training, the later adjustment of earlier weights would cause a later spur in increased model performance. ModelCheckpoint was chosen since behavior such as what was just exemplified is possible with the callback.

Write about the climatological Ice Edge dataset, ref section from here

This citation is actually for SPS_{length} but SPS is reduced to IIEE for a deterministic ice edge Goessling and Jung (2018)

This may change

tmp figure, redo with inspect data notebook in ForecastVer-

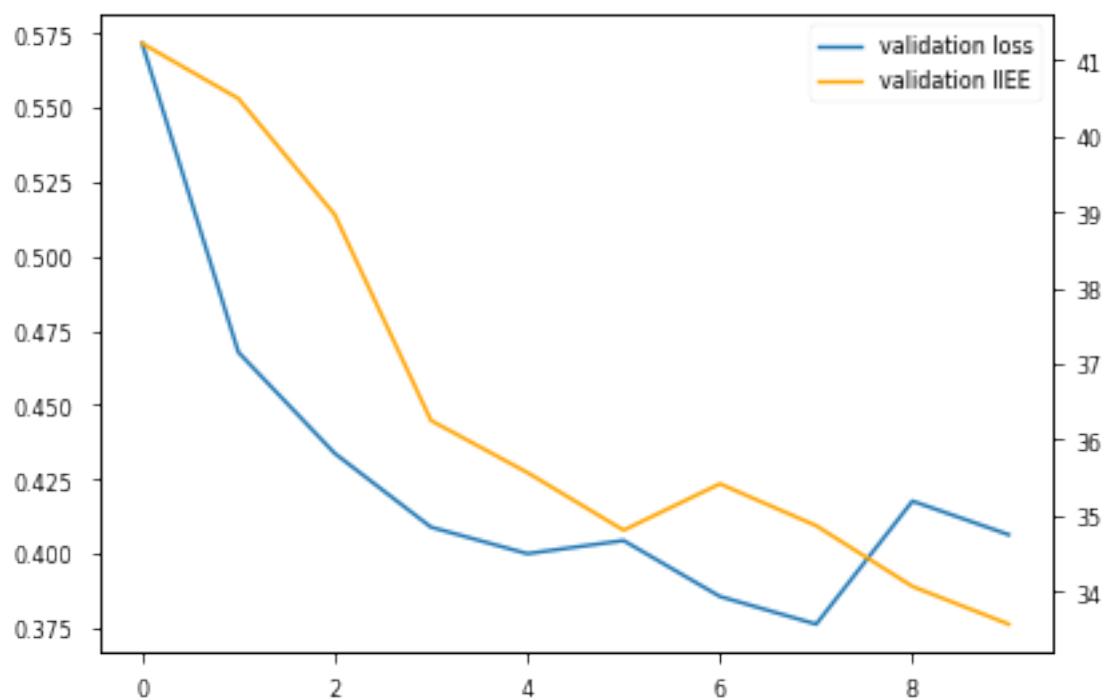


Figure 7: validation loss and Normalized IIEE computed as mean of validation set for each epoch during training

9 physical connections

9.1 Variograms

9.2 Case study

A case study is conducted where the highest reported IIEE value by the machine learning model.

9.3 Synthetic AA forcing

10 Comparing against physical models

The purpose of this section is twofold. Firstly, it aims at describing the process of preparing samples from the Barents-2.5 and NeXtSIM forecasting systems which are comparable to the Machine Learning forecasts at lead times of one, two and three days. Secondly, the performance of the forecasting systems will be assessed against the Sea Ice Charts, which are assumed to be the ground truth.

10.1 Preparing data

The logic behind sample creation is similar for both physical models. The idea is that the bulletin date of the physical forecasting system is +1 the bulletin date of the machine learning forecast. Furthermore, a daily mean is computed from the forecast based on the lead time of the forecast. I.e., a 1 day lead time for the machine learning forecast would constitute a daily mean of the first 24 hours forecasted by a physical forecasting system starting at 00 the following day of the machine learning bulletin date.

11 Conclusion and future outlook

A consequence of the operational aspect is the possibility to force decoupled NWP-systems with updated Sea Ice Concentration.

Figure (8)
to be
made
profes-
sional,
using
e.g.
TiX

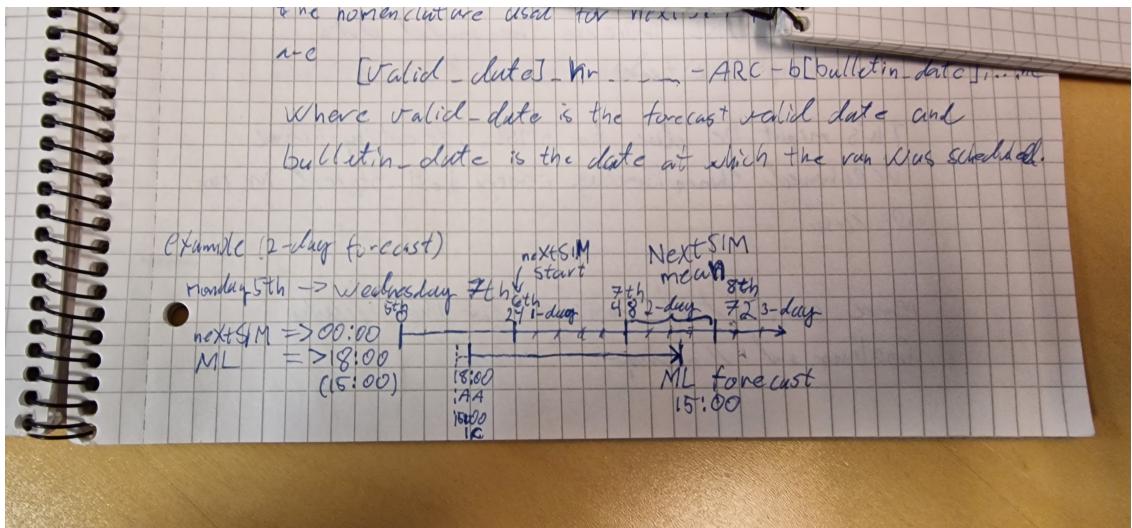


Figure 8: Sketch presenting how physical model forecasts are compared against machine learning forecasts. The axis represents time after 00:00 bulletin date of the machine learning forecast. The machine learning forecast is initiated 6 hours prior to the start of the physical model. The sketch exemplifies how the 2-day lead time machine learning forecast at 15:00 (reality 45 hours) is compared against an entire second day of a physical forecast (lead times 24 - 47).

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, URL <https://www.tensorflow.org/>, software available from tensorflow.org, 2015.
- Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., Phillips, T., Byrne, J., Tietsche, S., Sarojini, B. B., Blanchard-Wrigglesworth, E., Aksenov, Y., Downie, R., and Shuckburgh, E.: Seasonal Arctic sea ice forecasting with probabilistic deep learning, Nature Communications, 12, <https://doi.org/10.1038/s41467-021-25257-4>, 2021.
- Balmaseda, M. A., Mogensen, K., and Weaver, A. T.: Evaluation of the ECMWF ocean reanalysis system ORAS4, Quarterly Journal of the Royal Meteorological Society, 139, 1132–1161, [https://doi.org/https://doi.org/10.1002/qj.2063](https://doi.org/10.1002/qj.2063), URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2063>, 2013.

- Bauer, P., Beljaars, A., Ahlgrimm, M., Bechtold, P., Bidlot, J.-R., Bonavita, M., Bozzo, A., Forbes, R., Hólm, E., Leutbecher, M., Lopez, P., Magnusson, L., Prates, F., Rodwell, M., Sandu, I., Untch, A., and Vitart, F.: Model Cycle 38r2: Components and Performance, <https://doi.org/10.21957/XC1R0LJ6L>, 2013.
- Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Brown, B. G., and Mason, S.: Forecast verification: current status and future directions, *Meteorological Applications*, 15, 3–18, <https://doi.org/10.1002/met.52>, 2008.
- Cavalieri, D. J. and Parkinson, C. L.: Arctic sea ice variability and trends, 1979–2010, *The Cryosphere*, 6, 881–889, <https://doi.org/10.5194/tc-6-881-2012>, 2012.
- Comiso, J. C., Cavalieri, D. J., Parkinson, C. L., and Gloersen, P.: Passive microwave algorithms for sea ice concentration: A comparison of two techniques, *Remote Sensing of Environment*, 60, 357–384, [https://doi.org/10.1016/s0034-4257\(96\)00220-9](https://doi.org/10.1016/s0034-4257(96)00220-9), 1997.
- Comiso, J. C., Meier, W. N., and Gersten, R.: Variability and trends in the Arctic Sea ice cover: Results from different techniques, *Journal of Geophysical Research: Oceans*, 122, 6883–6900, <https://doi.org/10.1002/2017jc012768>, 2017.
- Dansereau, V., Weiss, J., Saramito, P., and Lattes, P.: A Maxwell elasto-brittle rheology for sea ice modelling, *The Cryosphere*, 10, 1339–1359, <https://doi.org/10.5194/tc-10-1339-2016>, 2016.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597, <https://doi.org/https://doi.org/10.1002/qj.828>, URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.828>, 2011.
- Dinessen, F., Hackett, B., and Kreiner, M. B.: Product User Manual For Regional High Resolution Sea Ice Charts Svalbard and Greenland Region, Tech. rep., Norwegian Meteorological Institute, 2020.
- Dukhovskoy, D. S., Ubnoske, J., Blanchard-Wrigglesworth, E., Hiester, H. R., and Proshutinsky, A.: Skill metrics for evaluation and comparison of sea ice models, *Journal of Geophysical Research: Oceans*, 120, 5910–5931, <https://doi.org/10.1002/2015jc010989>, 2015.
- Eguíluz, V. M., Fernández-Gracia, J., Irigoien, X., and Duarte, C. M.: A quantitative assessment of Arctic shipping in 2010–2014, *Scientific Reports*, 6, <https://doi.org/10.1038/srep30682>, 2016.
- Fritzner, S., Graversen, R., and Christensen, K. H.: Assessment of High-Resolution Dy-

namical and Machine Learning Models for Prediction of Sea Ice Concentration in a Regional Application, 125, <https://doi.org/10.1029/2020jc016277>, neural Networks for predicting Sea-Ice concentration are only slightly more accurate than persistence forecasting for short-term predictions., 2020.

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

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

Grigoryev, T., Verezemskaya, P., Krinit斯基, M., Anikin, N., Gavrikov, A., Trofimov, I., Balabin, N., Shpilman, A., Eremchenko, A., Gulev, S., Burnaev, E., and Vanovskiy, V.: Data-Driven Short-Term Daily Operational Sea Ice Regional Forecasting, Remote Sensing, 14, <https://doi.org/10.3390/rs14225837>, URL <https://www.mdpi.com/2072-4292/14/22/5837>, 2022.

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L., Prates, F., and Richardson, D.: Evaluation of ECMWF forecasts, including the 2021 upgrade, <https://doi.org/10.21957/XQNU5O3P>, 2022.

He, K., Zhang, X., Ren, S., and Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, <https://doi.org/10.48550/ARXIV.1502.01852>, 2015.

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

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

Ho, J.: The implications of Arctic sea ice decline on shipping, Marine Policy, 34, 713–715, <https://doi.org/10.1016/j.marpol.2009.10.009>, 2010.

Holland, P. R. and Kimura, N.: Observed Concentration Budgets of Arctic and Antarctic Sea Ice, Journal of Climate, 29, 5241–5249, <https://doi.org/10.1175/jcli-d-16-0121.1>, 2016.

Hunke, E. C. and Dukowicz, J. K.: An Elastic–Viscous–Plastic Model for Sea Ice Dy-

- namics, Journal of Physical Oceanography, 27, 1849–1867, [https://doi.org/10.1175/1520-0485\(1997\)027<1849:aevpmf>2.0.co;2](https://doi.org/10.1175/1520-0485(1997)027<1849:aevpmf>2.0.co;2), 1997.
- Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N., and Elliott, S.: CICE: the Los Alamos Sea Ice Model Documentation and Software User's Manual Version 5.1 LA-CC-06-012, techreport, Los Alamos National Laboratory, Los Alamos NM 87545, 2015.
- Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Møgensen, K., Zuo, H., and Monge-Sanz, B. M.: SEAS5: the new ECMWF seasonal forecast system, Geoscientific Model Development, 12, 1087–1117, <https://doi.org/10.5194/gmd-12-1087-2019>, URL <https://gmd.copernicus.org/articles/12/1087/2019/>, 2019.
- Kern, S., Lavergne, T., Notz, D., Pedersen, L. T., Tonboe, R. T., Saldo, R., and Sørensen, A. M.: Satellite passive microwave sea-ice concentration data set intercomparison: closed ice and ship-based observations, The Cryosphere, 13, 3261–3307, <https://doi.org/10.5194/tc-13-3261-2019>, 2019.
- Køltzow, M., Schyberg, H., Støylen, E., and Yang, X.: Value of the Copernicus Arctic Regional Reanalysis (CARRA) in representing near-surface temperature and wind speed in the north-east European Arctic, Polar Research, 41, <https://doi.org/10.33265/polar.v41.8002>, 2022.
- Kristensen, N. M., JensBDebernard, SebastianMaartensson, Keguang Wang, and Hedstrom, K.: Metno/Metroms: Version 0.3 - Before Merge, <https://doi.org/10.5281/ZENODO.1046114>, 2017.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, in: Advances in Neural Information Processing Systems, edited by Pereira, F., Burges, C., Bottou, L., and Weinberger, K., vol. 25, Curran Associates, Inc., URL <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>, 2012.
- Lavelle, J., Tonboe, R., Tian, T., Pfeiffer, R.-H., and Howe, E.: Product User Manual for the OSI SAF AMSR-2 Global Sea Ice Concentration Product OSI-408, Tech. Rep. 1.1, Danish Meteorological Institute, 2016.
- Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M. A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., and Pedersen, L. T.: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, The Cryosphere, 13, 49–78, <https://doi.org/10.5194/tc-13-49-2019>, URL <https://tc.copernicus.org/articles/13/49/2019/>, 2019a.
- Lavergne, T., Tonboe, R., Lavelle, J., and Eastwood, S.: Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration Climate Data Record OSI-450, OSI-430-b, techreport 1.2, 2019b.

- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P.: Focal Loss for Dense Object Detection, 2017.
- Liu, Y., Bogaardt, L., Attema, J., and Hazeleger, W.: Extended Range Arctic Sea Ice Forecast with Convolutional Long-Short Term Memory Networks, Monthly Weather Review, <https://doi.org/10.1175/mwr-d-20-0113.1>, 2021.
- Long, J., Shelhamer, E., and Darrell, T.: Fully Convolutional Networks for Semantic Segmentation, 2014.
- Melsheimer, C.: ASI Version 5 Sea Ice Concentration User Guide, Tech. rep., Institute of Environmental Physics, University of Bremen, 2019.
- Melsom, A., Palerme, C., and Müller, M.: Validation metrics for ice edge position forecasts, Ocean Science, 15, 615–630, <https://doi.org/10.5194/os-15-615-2019>, 2019.
- Müller, M., Batrak, Y., Kristiansen, J., Køltzow, M. A. Ø., Noer, G., and Korosov, A.: Characteristics of a Convective-Scale Weather Forecasting System for the European Arctic, Monthly Weather Review, 145, 4771–4787, <https://doi.org/10.1175/mwr-d-17-0194.1>, 2017.
- Nair, V. and Hinton, G. E.: Rectified Linear Units Improve Restricted Boltzmann Machines, in: ICML, pp. 807–814, URL <https://icml.cc/Conferences/2010/papers/432.pdf>, 2010.
- Notz, D. and Community, S.: Arctic Sea Ice in CMIP6, Geophysical Research Letters, 47, <https://doi.org/10.1029/2019gl086749>, 2020.
- Palerme, C., Müller, M., and Melsom, A.: An Intercomparison of Verification Scores for Evaluating the Sea Ice Edge Position in Seasonal Forecasts, Geophysical Research Letters, 46, 4757–4763, <https://doi.org/10.1029/2019gl082482>, 2019.
- pandas development team, T.: pandas-dev/pandas: Pandas, <https://doi.org/10.5281/zenodo.3509134>, URL <https://doi.org/10.5281/zenodo.3509134>, 2020.
- Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015.
- Röhrs, J., Gusdal, Y., Rikardsen, E., Moro, M. D., Brændshøi, J., Kristensen, N. M., Fritzner, S., Wang, K., Sperrevik, A. K., Idžanović, M., ThomasLavergne, Debernard, J., and Christensen, K. H.: "in prep for GMD" An operational data-assimilative coupled ocean and sea ice ensembleprediction model for the Barents Sea and Svalbard, p. 20, 2022.
- Serreze, M. C. and Meier, W. N.: The Arctic's sea ice cover: trends, variability, predictability, and comparisons to the Antarctic, Annals of the New York Academy of Sciences, 1436, 36–53, <https://doi.org/https://doi.org/10.1111/nyas.13856>, 2019.
- Smith, D. M.: Extraction of winter total sea-ice concentration in the Greenland and Barents Seas from SSM/I data, International Journal of Remote Sensing, 17, 2625–2646, <https://doi.org/10.1080/01431169608949096>, 1996.
- Spreen, G., Kaleschke, L., and Heygster, G.: Sea ice remote sensing using AMSR-

- E 89-GHz channels, *Journal of Geophysical Research*, 113, <https://doi.org/10.1029/2005jc003384>, 2008.
- Spreen, G., Kwok, R., and Menemenlis, D.: Trends in Arctic sea ice drift and role of wind forcing: 1992–2009, *Geophysical Research Letters*, 38, n/a–n/a, <https://doi.org/10.1029/2011gl048970>, 2011.
- Sørensen, A. M., Lavergne, T., and Eastwood, S.: Global Sea Ice Concentration Climate Data Record Product Uses Manual Product OSI-450 & OSI-430-b, Tech. Rep. 2.1, Norwegian Meteorological Institute, 2021.
- Tonboe, R., Lavelle, J., Pfeiffer, R.-H., and Howe, E.: Product User Manual for OSI SAF Global Sea Ice Concentration, Tech. Rep. 1.6, Danish Meteorological Institute, 2017.
- Velander, S., Wagner, P., Bailey, D., Everett, A., Goldstein, M., Hermann, R., Hjort-Larsen, T., Hovelsrud, G., Hughes, N., Kjøl, A., Li, X., Lynch, A., Müller, M., Olsen, J., Palerme, C., Pedersen, J. L., Rinaldo, ., Stephenson, S., and Storelvmo, T.: Knowledge needs in sea ice forecasting for navigation in Svalbard and the High Arctic, Tech. Rep. NF-rapport 4/2021, Svalbard Strategic Grant, Svalbard Science Forum, 2021.
- Wagner, P. M., Hughes, N., Bourbonnais, P., Stroeve, J., Rabenstein, L., Bhatt, U., Little, J., Wiggins, H., and Fleming, A.: Sea-ice information and forecast needs for industry maritime stakeholders, *Polar Geography*, 43, 160–187, <https://doi.org/10.1080/1088937x.2020.1766592>, 2020.
- Wes McKinney: Data Structures for Statistical Computing in Python, in: Proceedings of the 9th Python in Science Conference, edited by Stéfan van der Walt and Jarrod Millman, pp. 56 – 61, <https://doi.org/10.25080/Majora-92bf1922-00a>, 2010.
- Williams, T., Korosov, A., Rampal, P., and Ólason, E.: Presentation and evaluation of the Arctic sea ice forecasting system neXtSIM-F, *The Cryosphere*, 15, 3207–3227, <https://doi.org/10.5194/tc-15-3207-2021>, 2021.
- Yu, X., Rinke, A., Dorn, W., Spreen, G., Lüpkes, C., Sumata, H., and Grynkiv, V. M.: Evaluation of Arctic sea ice drift and its dependency on near-surface wind and sea ice conditions in the coupled regional climate model HIRHAM–NAOSIM, *The Cryosphere*, 14, 1727–1746, <https://doi.org/10.5194/tc-14-1727-2020>, 2020.
- Zampieri, L., Goessling, H. F., and Jung, T.: Predictability of Antarctic Sea Ice Edge on Subseasonal Time Scales, *Geophysical Research Letters*, 46, 9719–9727, <https://doi.org/10.1029/2019gl084096>, 2019.

12 Supporting Figures

