Reply to Reviewer 1- tc-2021-282 Asadi et al. 2021 - Probabilistic Gridded Seasonal Sea Ice Presence Forecasting using Sequence to Sequence Learning

We sincerely thank the reviewer for the thorough review and excellent comments.
Reviewer comments are shown in black, our responses are shown in blue.

This manuscript presents an innovative forecast tool for sea ice conditions (presence) based on a machine learning approach using sequence to sequence learning for both short term (7 days) and long term (up to 90 days) predictions.
The presented tool is, without a doubt, something that is of interest to the sea ice expert community and is based on novel methods of machine learning that make analysis of massive information datasets possible nowadays.

Even though the pertinence of the presented tool, several improvements must be done on the manuscript. The research design in itself has to be described into more details, especially by providing a more complete description of the different tests and protocols followed in the experiments and model calibration part.
In addition, many paragraphs, especially in the methodology part, could be supported by figures and schematic representations, for example, the ML design and architecture.

We have included preliminary figures (shown later in this response) describing the training and testing sequence (model calibration) and also the ML architecture (encoder and decoder).

Also, as the Hudson Bay region is highly documented and studied, the results obtained by your approach could be compared to data provided in the sea ice atlas from the Canadian Ice Service or results from other probabilistic/modelling approaches applied on the Hudson Bay (Saucier et al. 2004, Hochheim and Barber 2014, Kowal et al. 2017, Gignac et al. 2018, Dirkson et al. 2021.). Even though the comparison may not be quantitative, a qualitative assessment, outlining the differences between the approaches and the strategic advantages you provide using ML would be relevant.

Thank you for this comment. Gignac et al (2019) and Dirkson et al (2019) developed methods for probabilistic forecasting of sea ice concentration based on fitting probability distribution functions (PDFs) to historical passive microwave sea ice concentration data. Gignac et al (2019) used a beta PDF. They found their approach was able to capture freeze-up and break-up within one or two weeks of dates provided by the Canadian Ice Service (CIS) ice atlas, with the exception of Sanirajak (Hall Beach), where there is a polynya. Dirkson et al (2019) developed a related approach but used zero and one inflated beta distribution. This distribution allows the endpoints of their probability distribution to be captured differently than the interior points. Their PDF is fit to output from a prognostic modelling system, CanSIPS, (Canadian Seasonal to Interannual Prediction System), which consists of two coupled atmosphere-ice-ocean models. A novel bias correction approach is applied to their predictions and CanSIPS output before comparison with HadISST2 sea ice and surface temperature dataset. Their predictions show skill in Hudson Bay for forecasts initialized in May and June for 1-2 months (their Fig 10). Building on this, a related approach was used in Dirkson et al (2021), where a PDF was fit to freeze-up and break-up dates. In that case, before bias correction the model (CanSIPS) predictions of freeze up and break up in Hudson Bay were biased by 4 weeks, and 1-2 weeks respectively, where the data used for comparison is passive microwave sea ice concentration. The proposed bias correction methods (Dirsken et al. 2021) lead to improved prediction of these events in Hudson Bay for lead times of 1-6 months, depending on the location. For both of these studies the horizontal grid resolution was 100 km and time scales were monthly.

We also found two other relevant studies, those by Bushuk et al (2017) and Sigmond et al. (2016). Bushuk et al. (2017) use a fully coupled atmosphere-ice-ocean-land model with horizontal grid resolution for the ice-ocean component of 1 degree. They compute the anomaly correlation coefficient (ACC) of detrended sea ice extent from their model with that from passive microwave data and compare with a model of anomaly persistence. Their model shows skill relative to persistence in Hudson Bay, in particular in summer months, with lead times of 3-8 months. Sigmond et al. (2016) use CanSIPS. They examine anomaly correlation coefficients between sea ice advance and retreat dates computed from the model and from passive microwave sea ice concentration data. The grid resolution is relatively coarse (around 100 km), possibly because it is an ensemble forecasting system, which is computationally demanding. They find their model has skill (defined by statistically significant ACC) at lead times of 2-6 months (average 5) for sea ice advance (freeze-up) and 2-3 months (average 3) for sea ice retreat (break-up) (not detrended, results for Hudson Bay).

We will discuss these in more detail, bringing in context from Saucier et al (physical modelling) and Hocchiem and Barber and Kowal (trend analysis), in the revised manuscript.

It should be noted while our approach is similar those by Gignac et al. (2019) and Dirkson et al. (2019, 2021), in that it does not use dynamical sea ice models to produce a prediction, it is different in that it uses a sequence-to-sequence learning approach that is capable of producing forecasts. The present configuration is to used for 90-day forecasts, although it is evaluated in the submitted manuscript in hindcast mode to demonstrate the concept.

Overall, the presented method and tools appear as scientifically sound and clear. They should, however, be described more carefully and more examples of applications of the model shall be presented to the readers.

It is a work of great interest and I hope my comments will guide and help you in improving your manuscript.

Thank you very much. Your comments are very helpful.

Major comments

1. As aforementioned, *comparisons have to be made and applications examples provided*. Especially in areas of high variability or in the presence of particular entities such as polynyas or narrower Bays (Frobisher Bay or Hall Beach polynya, for example).

We are planning to bring in a comparison of freeze-up and break-up dates with ice charts for the ports identified in the manuscript (Churchill, Inukjuak and Quataq) as well as; Sanirajak (formerly known as Hall Beach), and Frobisher Bay, where there are polynyas; and a location in the central part of Hudson Bay (away from shore). We will compare these freeze up and breakup dates to those in Gignac et al, 2019 (their Table 1). We have investigated using daily ice charts of the Canadian Ice Service for this purpose. These are sea ice concentration and ice type analyses derived through manual inspection of SAR data and other data available, with valid time of 18 UTC each day. However, these charts are not available on a daily basis through the entire ice season in the Hudson Bay region. Hence, our analysis will be based on regional ice charts, which are similar to daily ice charts, but bring in data over a week and have less fine spatial detail. We also considered using the CIS ice atlas for this comparison (as was used in Gignac et al) but the CIS ice atlas covers an earlier period. Note that Frobisher bay is not in our study region.

2. *Limitations of the approach shall be discussed*. The model is providing forecasts on a ~31 km grid. How does this affect the usage capabilities for the principal expected users (the mariners)? This should definitely be discussed more in depth.

The resolution is similar to that used in other studies on seasonal forecasting that have been developed with mariners in mind. For example, passive microwave data were used for development of the probabilistic approach of Gignac et al. (2019) and for validation of subseasonal-to-seasonal (S2S) time scales (Zampieri et al. 2018). While passive microwave sea ice concentration data are often gridded to 25 km, the spatial resolutions of the brightness temperature data used to generate the sea ice concentration are typically coarser. The 19.35 GHz channel on the SSMI and SSMI/S sensors (often used to produce sea ice concentration observations) has an instrument field of view of approximately 45 km x 69 km (https://www.remss.com/missions/ssmi). For studies that carry out seasonal forecasting using a dynamic ice-ocean model (or similar) where a sea ice state vector is predicted as a function of time, our resolution is similar to that used in other approaches to be used to support (in part) marine transportation (Sigmond et al., 2016, Askenov et al 2017). Sigmond et al. (2016) assess seasonal forecasts of sea ice advance and retreat at a spatial resolution of 100 km, while Askenov et al. (2017) examine navigation routes using a ¼ degree grid (nominally 28km, 9-14 km in Arctic) coupled sea-ice ocean model. For the latter study, their focus is on the Northern Sea route, and their study region does not include Hudson Bay or Hudson Strait. Nevertheless, the study by Askenov et al. (2017) study highlights potential future changes in Arctic ice cover that will lead increased navigability, but with significant risk

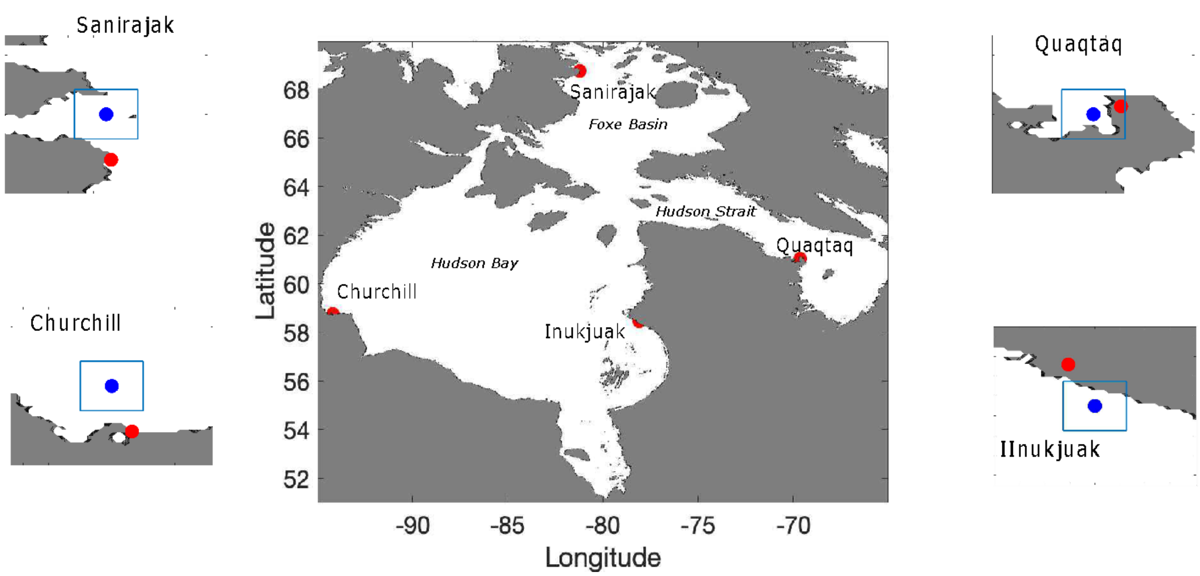
For trend analysis in the Hudson Bay region, which also supports planning activities, Hoccheim and Barber (2014) use passive microwave data (gridded to 25 km) to look at sea ice trends in the Hudson Bay system, while Andrews et al. 2018, use a combination of passive microwave data for offshore trends in ice cover, and regional ice charts from the Canadian Ice Service for nearshore trends.

3. *Sensibility to the input “sea ice normal condition” wasn’t discussed*, when speaking about the augmented model. *Were variable time spans tested?* If so, did they generate similar forecasts? If not, how would you explain this situation? In other words, a certain “sensibility analysis” would be convincing about the model capabilities.

In the version submitted, the augmented model secondary input is based on climate normal computed as the average 2m air temperature (t2m), and average 10m wind components (u10 and v10), where the average is calculated from the first year (1985) until the last year of training (2015). We compared this method of computing the climate normal (CN1) with one that is based on the 10 years previous to the validation year (CN2). There was very little difference in the model predictions when these two different augmented models are used based on heat maps (of the type shown in Fig 1 in the submitted manuscript). This indicates the patterns our model is learning do not vary significantly over the 1985-2015 period in comparison to a decadal time scale. Examining the results on a year to year basis showed that when the shorter time period is used for climate model the year-to-year variability has a more significant impact on model performance, indicating climate normal is given less weight, or is more similar to the training data.

4. I strongly suggest that you *add a map of your validation sites* and provide a short description of each. For example, Quaqtaq is located in a bay, narrower than 31km. How this does affect the results?

We have put together a map of the port locations, including one requested at Sanirajak (formerly known as Hall Beach). The map of the study region is shown below with the port locations shown in red. The Insets show the port location (red) and the nearest point on the model grid (blue) that is outside of the land boundary (where landmask from ERA5 is less than 0.6), in addition to a bounding box that approximates a grid cell. The model grid point near Quaqtaq is located (correctly) in the water region because the landmask from ERA5 has a low value in that region due to the low elevation.



Also, *why were these 3 sites chosen*?

The sites were chosen because they represent locations with significantly different sea ice conditions. Churchill and Inukjuak are located on the east and west coasts of Hudson Bay, with Churchill being a major port as part of the potential Arctic Bridge shipping route. The east coast is significantly impacted by influx of freshwater inflow from rivers draining into Hudson Bay, while the west coast region is impacted by northwesterly winds (there is a latent heat polynya, the Kivalliq polynya, that runs along the northwest shore of Hudson Bay). There are additionally east-west asymmetries in Hudson Bay in terms of ice thickness and sea surface temperature (Saucier, 2014), with counter-clockwise ocean currents leading to thicker ice covers along the eastern shore of the Bay. Quataq is located in Hudson strait, where wind and air temperature patterns are different from those in Hudson Bay, and pressured ice is common in the consolidated ice season. We will be bringing in Sanirajak and a location in the central part of the Bay (away from the coast) in the revised version.

Have you found any irregularities in the ERA-5 sea ice concentration dataset you used to calibrate your model?

We found there were irregularities with the ERA landmask file and the sea ice concentration. There were some locations indicated as land in the landmask file that had a non-zero sea ice concentration value. At these locations the sea ice concentration was set to zero. There were also some locations indicated as non-land in the landmask file that had a zero ice concentration, even when the ice concentration should be non-zero based on the atmospheric conditions and seasons. At these locations (indicated by landmask less than or equal 0.6), the sea ice concentration was set to the non-land average of neighboring pixels.

5. Nowhere in the manuscript have I found a *justification on why it is the presence of ice that is modeled and not the concentration*. That should be discussed since the OSI-SAF OSI409 (which are SIC) data are ingested into ERA-5

The approach used here predicts a grid of (uncalibrated) sea ice probabilities. For seasonal forecasting, probabilistic information is often desired (Gignac et al. 2019), in particular regarding freeze-up and break up (Wagner et. al., 2020). While the model does ingest sea ice concentration, this is combined with other environmental variables to produce a binary probability of ice vs water at the grid cell, or ice presence. This output is similar to probabilistic approaches by Gignac et al. (2019) and Dirkson et al (2019) where their model estimates the probability of sea ice concentration exceeding a certain threshold (e.g., 15%).

Specific comments

Line 10 : Define “high spatial resolution” for the reader. Depending on your field, it differs.

We were referring to the spatial resolution of 5-10 km. For example, the high-resolution ocean and sea ice forecasting system for the Arctic and North Atlantic oceans (Dupont et al., 2015).

Line 46 : Define sea ice presence (SIC > 15%).

We noticed that the term “sea ice presence” was first used at the beginning of the introduction, and have added the information there (sea ice concentration greater than 15%).

Line 51 : This information should be provided way before, otherwise, some will think, as I did, that you use remote sensing data.

The use of ERA5 has been added to the last paragraph of the introduction. The abstract has also been changed to explicitly mention this “Given the recent observations of the declining trend of Arctic sea ice extent over the past decades, seasonal forecasts are often desired. In this study machine learning (ML) approaches are deployed to provide accurate seasonal forecasts based on ERA5 data as input"

Line 57 : Why not starting in 1979 ?

Initially the study was started with a different data set (other than ER5). These data started in 1985, hence that was used in this study as well because it provides a sufficient time series for training and testing.

Line 67 : Remove last "and". Corrected

Line 87 : A schematic representation of the encoder and decoder parts would be useful.

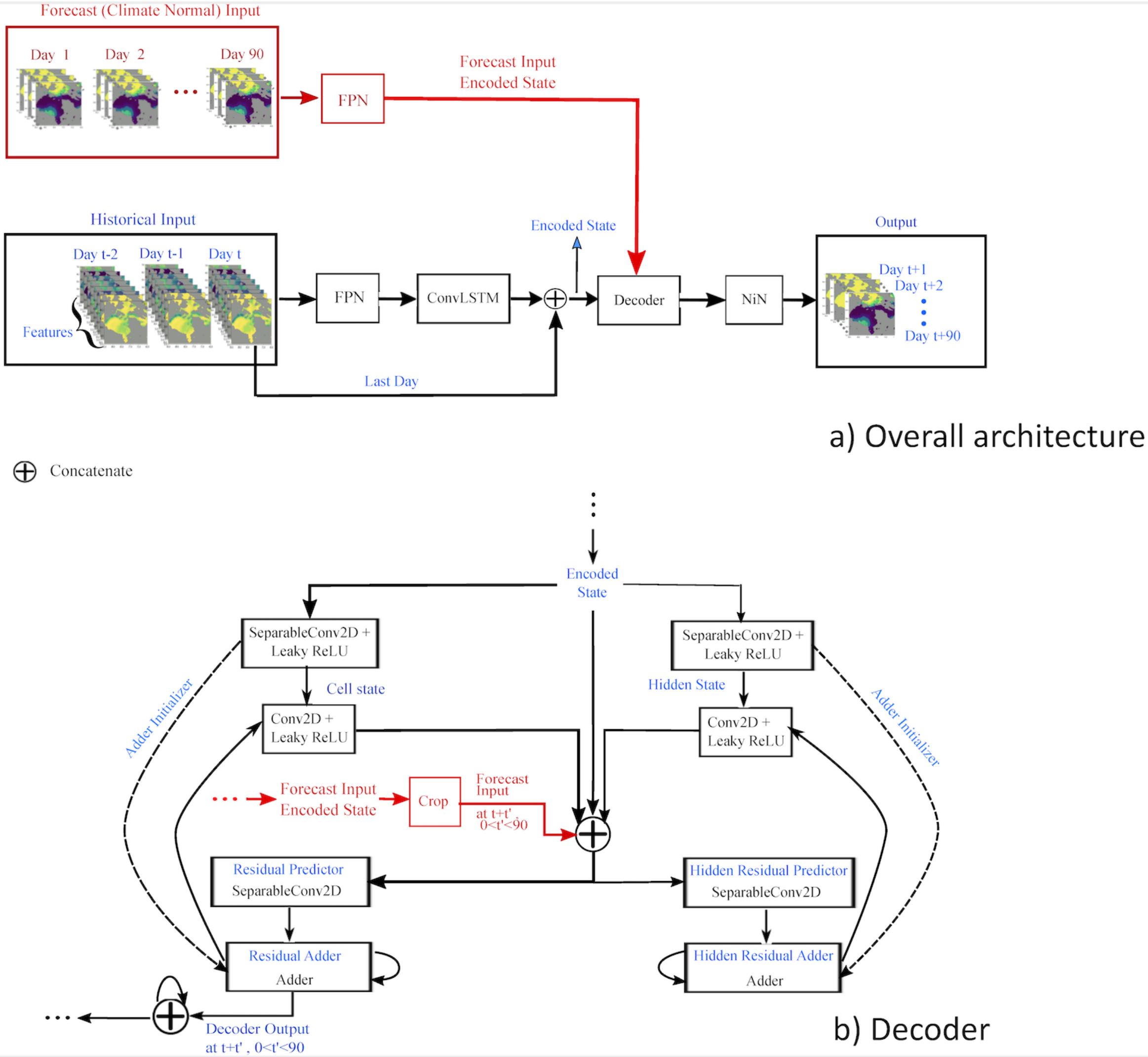
Below is a figure of the architecture which shows the encoder and decoder parts in more detail. The upper panel shows the overall architecture(described on lines 98-105 of the submitted manuscript, modified here).

The overall architecture is shown in the figure below (panel a). The encoder starts by passing each daily sample through a feature pyramid network (Lin et al., 2017) so as to detect environmental patterns at both the local and large scales. Next, the sequence of feature grids extracted from the feature pyramid network are further processed through a convolutional LSTM layer (ConvLSTM) (Hochreiter and Schmidhuber, 1997; Xingjian et al., 2015), returning the last output state. This layer learns a single grid representation of the time series that also preserves spatial locality. Finally, the most recent day of historic input data is concatenated with the ConvLSTM output. The encoder provides as output a single raster with the same height and width as the stack of raster data input to the network, but with a higher number of channels such as to represent the fully encoded system state. The final encoded state is then fed to a custom recurrent neural network (RNN) decoder that extrapolates the state across the specified number of time-steps. It takes as input the encoded state with multiple channels and as output produces a state with the same height and width as the input over the desired number of time-steps in the forecast (here 90 days). Finally, a time-distributed network-in-network (Lin et al., 2013) structure is employed to apply a 1D convolution on each time-step prediction to keep the spatial grid size the same but reduce the number of channels to one, representing the daily probabilities of ice presence over the forecast period (e.g. up to 90 days).

The lower panel shows the decoder (described on lines 106-113 of the submitted manuscript, modified here)

The custom RNN decoder, as is common of many RNN layers, maintains both a cell state and a hidden state (Yu et al.,2019). First, the initial cell state and hidden state are initialized with the input encoded state. Then, at each time-step and for each of the states, the network predicts the difference, or residual, from the previous state to generate the updated states using 2D depthwise separable convolutions (Howard et al., 2017). Depthwise separable convolutions are chosen to preserve the time dimension of the input, which the convolution operates over the two spatial dimensions. The output of the decoder section is the concatenation of the cell states from each time-step (unrolling of the learned time sequence).

The red portion shown in the figure below corresponds to the additional components required for the Augmented model (described on lines 115-122 of the submitted manuscript)



.

Caption: Figure showing (a) the overall network architecture and b) the custom decoder. The red portion refers to the additional components required for the augmented model. The dashed arrows show a process carried out only once (the initialization of the adder). FPN refers to the feature pyramid network, ConvLSTM, the convolutional long short-term memory network, NiN is the network in network module.

Line 92 : What time bin(-s) were used as input (12:00, 00:00, a daily average ?)

Noon samples were used. This information has been added to the last paragraph of the data section.

Line 112 : How does extending to a longer input affects the forecast quality ?

We tested extending to a 5 day input but did not see any improvement in forecast quality. With this longer input the quantity of data to be processed is greater than that for 3 days, which increases the computational expense and data storage requirements. Hence we did not continue with this, or longer inputs. However, we recognize that in a different domain where other processes are important a longer input may show improvements.

Line 126 : Can you describe these “extensive experimentations” ?

Initially we had used a leave-one-out approach for training and testing of the model. For example, given 30 years of data, train with 29 and test with one, and repeat over all 30 years. However, would result in the use of future data for training, which is not desirable for a forecasting approach. Hence, we moved to the current approach of training initially using ten years, updating the model weights for future time periods. We tested different training periods (10 vs 20) and also different numbers of months to include in training our monthly models. The current configuration led to the best results.

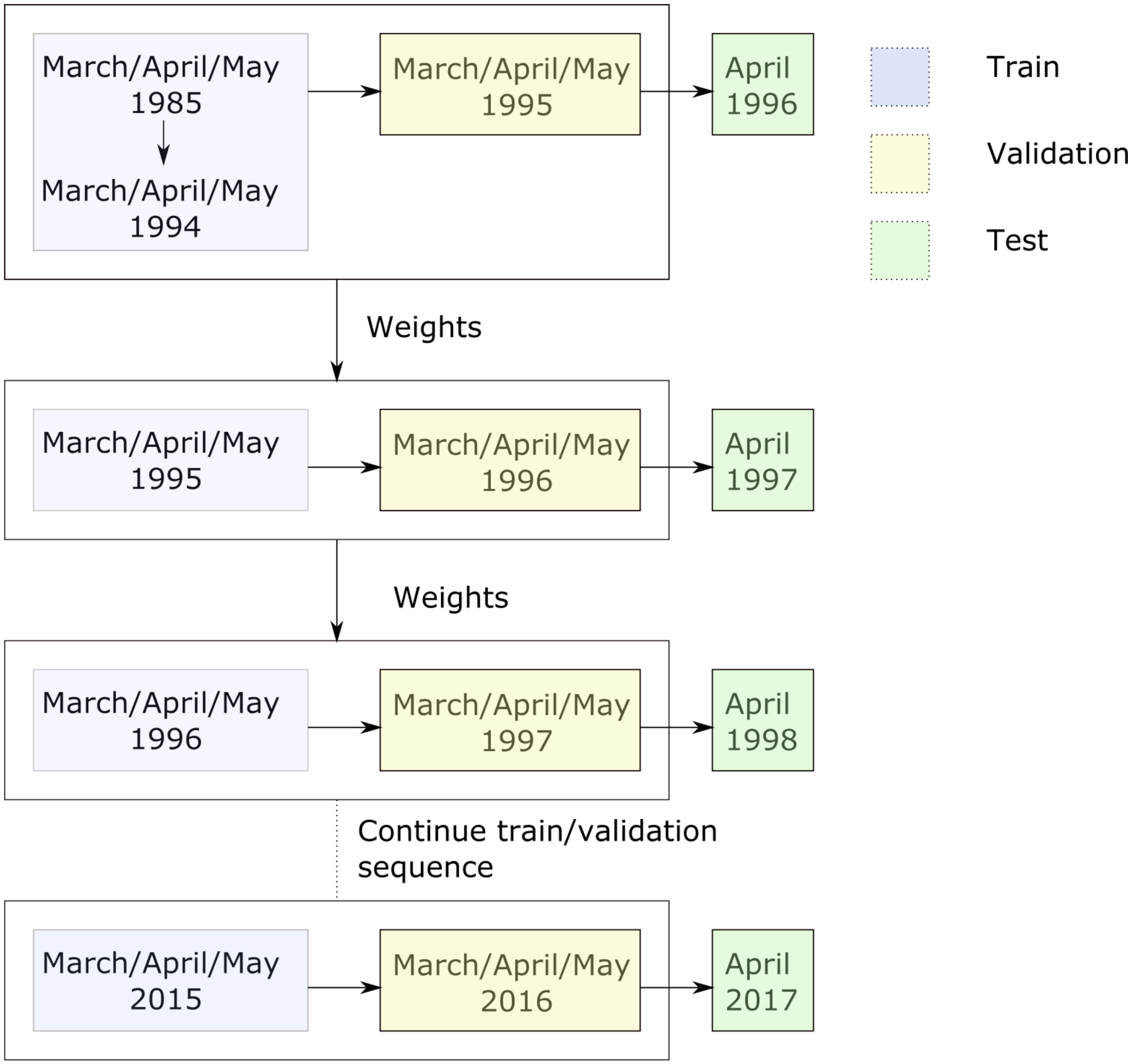
Line 129 : “Chosen to be 10 years”. Why is it so ?

The choice of 10 years is a compromise between having enough data to provide the model with representative conditions from which it can learn, and not having the approach become too data-heavy. Because the model already performs reasonably well with this training approach, it could be used alternatively with other data sets for which a shorter time series is available (eg. an AMSRE/AMSR2 unified data set available from 2002, which would have higher spatial resolution, although also different variables).

Line 135 : This processing logic should definitely be represented in a figure.

Good point. We agree and have prepared a figure (below with text from manuscript lines 127-134, modified here)

For each month of a year a separate model is trained on data from the given month as well as the preceding and following month. For example, the ’April model’ is trained using data from March 1 to May 31. This monthly model is initially trained on data from a fixed number of years, chosen to be 10 years. After this initial experiment, to predict each following test year i, using a rolling forecast prediction, the model from year i−1 is retrained with data from year i−2 and also, data from year i−1 is used as validation for early stopping criteria and to evaluate the training performance. For example, if the initial model is trained on 10 years, data from year 11 is used as validation and first predictions are launched at year 12. The model for year 11 is initialized with weights from the 10-year model and retrained with data from year 10, validated on year 11 and predicts year 12. The model for year 12 is then initialized with weights from the year 11 model, retrained with data from year 11 and validated on year 12 to predict year 13. This process is used to produce forecasts of sea ice presence for years 1996 to 2017.



Line 166 – 167 : The end of the sentence doesn't make sense. Consider reformulating.

This has been changed to "Using additional climate variables for the input of the Augmented model is shown to be beneficial (Fig 1d,e,f). In the periods where the Basic model is worse than climate normal, the Augmented model has better accuracy, for example for the April model at lead days 60-80, and the August model, lead days 75-90..

Line 181 : It seems counterintuitive. Can you explain why ? The reason for the higher Brier score of the Basic model in comparison to augmented here may be because the September model uses training data over August/September and October. We hypothesize that the trend over this period may be less representative of more recent ice conditions than breakup, which may make the additional data used in the Augmented model un-helpful.

Figure 1 : Y-Axes for subfigures d-e-f should be Accuracy differences or ΔAccuracy.

This has been changed.

Line 220 : What would you link the lower accuracy in “central region” ? Is it the higher variability of the freeze-up pattern or to climate variables that are, given the distance to stations, less reliable in such areas ?

We assume the reviewer is referring to the lower performance of Basic model in the central regions relative to the coast in December for a 30 lead day forecast. In this case the degradation was because freeze-up was too late in the model in the central regions. While the climate variables could be less accurate due to their distance to station data (assuming station data are assimilated in a reanalysis and are accurate) our experiments are not set up to evaluate this because we are using ERA5 as our “observations” for comparison.

Lines 236, 246 & 249 + Figures 8 & 9 : From what I know, it should be written Quaqtaq, not Quataq (https://www.makivik.org/quaqtaq/).

Thank you very much. This has been corrected.

References

Andrews, J., Babb, D. and Barber, D.G. (2018), “Climate Change and Sea Ice: Shipping in Hudson Bay, Hudson Strait and Foxe Basin (1980-2016)”, Elementa, 6, DOI: 10.1525/elementa.281.

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

Bushuk, M., Msadek, R. Winton, M. Vecchi, G. A., Gudgel, R., Rosati, A. and Yang, X., (2017), “Skillful regional prediction of Arctic sea ice on seasonal time scales”, Geophysical Research Letters, 44, doi: 10.1002/2017GL073155.

Dirkson, A., Merryfield, W.J., and Monahan, A.H., (2019) “Calibrated probabilistic forecasts of Arctic sea ice concentration”, Journal of Climate, 32, 1251-1271.

Dirkson, A., et al. (2021) "Development and Calibration of Seasonal Probabilistic Forecasts of Ice-free Dates and Freeze-up Dates." Weather and Forecasting 36.1: 301-324.

Dupont, F., Higginson, S., Bourdallé-Badie, R., Lu, Y., Roy, F., Smith, G. C., Lemieux, J.-F., Garric, G., and Davidson, F.: A high-resolution ocean and sea-ice modelling system for the Arctic and North Atlantic oceans, Geosci. Model Dev., 8, 1577–1594, https://doi.org/10.5194/gmd-8-1577-2015, 2015.

Gignac, C., Bernier, M., & Chokmani, K. (2019). IcePAC–a probabilistic tool to study sea ice spatio-temporal dynamics: application to the Hudson Bay area. The Cryosphere, 13(2), 451-468.

Hochheim, K. P. and Barber, D. G. (2014), An update on the ice climatology of the Hudson Bay system, Arct. Antarct. Alp. Res., 46, 66–83.

Saucier, F., Senneville, S., Prinsenberg, S., Roy, F., Smith, G., Gachon, P., Caya, D., and Laprise, R.: Modelling the sea ice-ocean seasonal cycle in Hudson Bay, Foxe Basin and Hudson Strait, Canada, Climate Dynam., 23, 303–326, 2004.

Sigmond, M., Reader, M.C., Flato, G.M., Merryfield, W.J. and Tivy, A. (2016) “Skillful seasonal forecasts of Arctic sea ice retreat and advance dates in a dynamical forecast system”, Geophysical Research Letters, 43, 12,457-12,465, doi:10.1002/2016GL071396.

Zampieri, L., Goessling, H. F., & Jung, T.J. (2018). Bright prospects for Arctic sea ice prediction on subseasonal time scales. *Geophysical Research Letters*, 45, 9731– 9738. https://doi.org/10.1029/2018GL079394