Deep graph neural networks for spatiotemporal forecasting of sub-seasonal sea ice: a case study in Hudson Bay

Gousseau, Zacharie; Lamontagne, Philippe; Jahangir, Mohammad Sina; Scott, K. Andrea

Publisher's version / Version de l'éditeur:
https://doi.org/10.1016/j.apor.2025.104793
Applied Ocean Research, 164, C, pp. 1-17, 2025-10-08

Applied Ocean Research 164 (2025) 104793

Contents lists available at ScienceDirect

Applied Ocean Research

journal homepage: www.elsevier.com/locate/apor

Research paper

Deep graph neural networks for spatiotemporal forecasting of sub-seasonal sea ice: A case study in Hudson Bay

Zacharie Gousseau [a], Philippe Lamontagne [b], Mohammad Sina Jahangir [c], K. Andrea Scott [d,*]

[a] Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, Canada
[b] Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada
[c] Department of Bioresource Engineering, McGill University, Montreal, Canada
[d] Ocean, Coastal and River Engineering Research Centre, National Research Council Canada, Ottawa, Canada

ARTICLE INFO

Dataset link: doi.org/10.24381/cds.adbb2d47, doi.org/10.48670/moi-00021, apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/, doi.org/10.5281/zenodo.15571799

Keywords:
Sea ice
Graph neural network
Forecasting
Deep learning
Attention

ABSTRACT

This study introduces GraphSIFNet (**Graph** **S**ea **I**ce **F**orecast neural **Net**work), a novel graph-based deep learning framework for spatiotemporal sea ice forecasting. GraphSIFNet employs a Graph Long-Short Term Memory (GCLSTM) module within a sequence-to-sequence architecture to predict daily sea ice concentration (SIC) and sea ice presence (SIP) in Hudson Bay over a 90-day time horizon. The use of graph neural networks (GNNs) allows the domain to be discretized into arbitrarily specified meshes, allowing more explicit spatial modeling than approaches based on the convolutional neural network (CNN). This study demonstrates the model's ability to forecast over an irregular mesh with higher spatial resolution near shorelines. The model is trained using atmospheric data from ERA5 and oceanographic data from GLORYS12. Results demonstrate the model's superior skill over a linear combination of persistence and climatology as a statistical baseline. The model showed skill particularly in short- to medium-term (up to 35 days) SIC forecasts, with a noted reduction in root mean squared error (RMSE) by up to 10% over the statistical baseline during the break-up season, and up to 5% in the freeze-up season. Long-term (up to 90 days) SIP forecasts also showed significant improvements over the baseline, with increases in accuracy of around 10% even at a lead time of 90 days. The use of an attention-based convolution offered the additional benefit of interpretability by highlighting the primary direction and magnitude of information flow that aligned with the direction of freezing and melting. The study lays the groundwork for future exploration into dynamic graph-based forecasting, and future work forecasting ice-ocean phenomena.

1. Introduction

The decline of Arctic sea ice volume is one of the most visible and immediate signs of climate change (Stroeve and Notz, 2018). The Arctic is the fastest-warming region on Earth, and this warming is affecting the sea ice cover more than any other component of the climate system (Vihma, 2014; Stroeve et al., 2012; Cavalieri and Parkinson, 2012). According to the National Snow and Ice Data Center (NSIDC), Arctic sea ice extent (SIE)—the total area of the Arctic Ocean with at least 15% ice cover—is seeing a steady decline. This is especially prominent in September when sea ice extent is at its minimum (Serreze and Meier, 2019). Declining sea cover is connected to increasing air temperatures, changes in atmospheric and oceanic circulation, the albedo feedback loop, and the concentration of greenhouse gases in the atmosphere (Stroeve et al., 2012). The Arctic ice cover is of particular importance as it helps regulate the Earth's climate, and the decline in sea ice and subsequent loss of reflectivity directly contribute to the acceleration of climate change (Moon et al., 2019). Changes in Arctic sea ice cover also disturb marine and terrestrial ecological dynamics (Post et al., 2013); create challenges for Northern communities (Meier et al., 2014); and influence human activity as new trade routes become available through the Arctic (Mudryk et al., 2021). Forecasting sea ice conditions is therefore becoming increasingly important as accurate knowledge of these changes would allow for more effective preparation.

In this study, we introduce a deep learning based sea ice forecasting model that employs Graph Neural Networks (GNNs) integrated within a Long Short-Term Memory (LSTM) module to predict daily sea ice concentration (SIC) and sea ice presence (SIP) in Hudson Bay up to 90 days in advance. The choice of Hudson Bay as our study area

* Corresponding author.
E-mail addresses: zacharie.gousseau@uwaterloo.ca (Z. Gousseau), philippe.lamontagne@nrc-cnrc.gc.ca (P. Lamontagne), msjahangir@uwaterloo.ca (M.S. Jahangir), ka3scott@uwaterloo.ca (K.A. Scott).

https://doi.org/10.1016/j.apor.2025.104793
Received 11 February 2025; Received in revised form 27 August 2025; Accepted 29 September 2025
Available online 8 October 2025

is driven by its important role as a shipping hub, the presence of communities living within the region relying on maritime re-supply, and its unique characteristics as an in-land sea largely isolated from the wider Arctic. The 90-day forecasting horizon addresses the needs for planning and decision-making in industries such as shipping operations as well as the planning requirements of local communities residing in the region. This time horizon covers short-term (up to 7 days), medium-term (up to a month) and long-term (up to 3 months) planning needs. The study highlights the effectiveness of GNNs in handling irregular spatial domains by dividing Hudson Bay into a spatially irregular mesh with a higher resolution along shorelines. We evaluate the performance of two types of spatial graph convolutions within the model: the basic Graph Convolutional Network (GCN) and an attention-based transformer convolution. The model was trained using sea ice and oceanographic data from a coupled ice-ocean reanalysis product (GLORYS12 Lellouche et al., 2021), as well as atmospheric data from the ECMWF Reanalysis v5 (ERA5 Hersbach et al., 2020). We validate the model's accuracy by comparing its predictions to a statistical baseline and comparing forecasted and observed freeze-up and break-up dates at ports on Hudson Bay.

2. Background

Sea ice forecasting is a spatiotemporal forecasting task that can be formulated as a next-frame prediction problem. Given a sequence of frames $\mathbf{X} = (\mathbf{X}_{t-n}, \ldots, \mathbf{X}_{t-1}, \mathbf{X}_t)$ with $\mathbf{X_t} \in \mathbb{R}^{w \times h \times c}$ where n is the number of frames in the sequence, w and h are the spatial dimensions of the frames, and c is the number of channels, the objective is to predict the next T frames in the sequence, $X_{t+1}, \ldots, X_{t+T}$.

While traditional time series modeling techniques such as ARIMA (autoregressive integrated moving average) have been used for temporal sea-ice forecasting (Wu et al., 2021), showing promising results in comparison to reanalysis, they are less effective for spatiotemporal forecasting due to their inherent limitations in handling spatial dependencies with complex temporal dynamics. ARIMA models, primarily designed for univariate time series, lack the capacity to effectively model spatial relationships and multi-dimensional data structures, which are critical in spatiotemporal forecasting. To address these limitations, methods like Vector Autoregression (VAR) (Sims, 1980) and Spatial Autoregressive (SAR) (Anselin, 1988) models were developed, offering improved handling of multivariate data and spatial dependencies, respectively. However, these models still struggled with dynamic spatial relationships and non-linear interactions. Space-Time Autoregressive Integrated Moving Average (STARIMA) models (Pfeifer and Deutsch, 1980) were introduced to better integrate spatial dependencies with temporal dynamics. Dynamic Linear Models (DLMs) and State Space Models (Kalman, 1960) offered a framework for handling evolving temporal dynamics but were limited in their spatial modeling capabilities.

With the advent of deep learning, many neural network methods were developed for spatiotemporal problems, largely based on spatial convolutions with fixed-size two- or three-dimensional kernels (Oprea et al., 2022). These convolutional models are particularly well-suited for image data with a gridded structure such as images or video frames and allow for learning rich features that are present in real-world image sequences.

Graph Neural Networks (GNNs) offer a compelling alternative to Convolutional Neural Networks (CNNs) for emulating models of physical processes, such as ice dynamics, for several reasons. One of the primary advantages of GNNs in this context is their inherent ability to capture the spatial relationships between neighboring nodes through graph edges. This is particularly crucial in applications like sea ice dynamics, where the spatial relationships are fundamental in determining heat and momentum exchanges, and other factors influencing ice processes. In GNNs, both nodes and edges can encode information about the system, and graph convolutions update these encodings. This allows GNNs to effectively model the exchange of physical quantities such as heat or ice volume at a given location in space and time while accounting for the directionality of processes, which is represented by directed edges. In contrast, CNNs operate on a fundamentally different principle. They extract features such as edges or gradients from an input image by tuning kernel filters. This process involves convolving these filters over the input image to identify patterns and features at various scales and orientations. While this approach is highly effective for tasks like image recognition, where identifying and categorizing visual patterns is key, it may not be as well-suited for learning the underlying physical laws that govern interactions between points in space. CNNs typically lack the ability to explicitly model directional relationships and complex dependencies between disparate points in a spatial domain, which are critical in understanding and predicting physical phenomena like ice dynamics.

CNNs leverage spatial locality and translation invariance inherent in images through convolutional layers with fixed-size filters that extract local features across the image. Techniques such as the use of pooling operators, stride convolutions, or dilated filters can be used to capture longer-range patterns and hierarchical information (He et al., 2016; Yu and Koltun, 2015). In contrast, message-passing GNNs can natively capture long-range patterns through edge propagation, potentially reaching across the entire graph structure given a sufficiently deep network. Although in most cases the underlying graphs are too large for information to be propagated globally, limited information propagation across can help models gain a holistic view of the spatial domain and learn complex spatial patterns (Wu et al., 2022). Additionally, most types of GNNs exhibit both translation and rotation invariance as convolutions are applied indiscriminately to all nodes and the aggregation operators are most often permutation invariant. Another noteworthy advantage of GNNs over CNNs is their scalability due to the inherent parallelism in their architecture, allowing for efficient processing of data over large regions or with fine resolution. This parallelism however comes at the cost of higher memory usage which may become limiting, though this can be circumvented by partitioning the graph and processing the subgraphs independently before combining the outputs. Overlapping subgraphs can be used to ensure no spatial artifacts or discontinuities arise from the partitioning.

3. Related work

Prior to the advent of deep learning techniques in sea ice forecasting, traditional physics-based and statistical models were the mainstay for both short-term and long-term predictions. Dynamic models, often integrated within data assimilation systems, such as the Pan-Arctic Ice-Ocean Modeling and Assimilation System (PIOMAS) (Zhang and Rothrock, 2003), rely on solving physical equations to simulate the interactions between sea ice, atmosphere, and ocean. These models are computationally intensive and require extensive calibration, but are considered fairly reliable due to their capacity to incorporate well-understood physical processes and parameters. On the other hand, statistical models such as multiple linear regression (MLR) and autoregressive integrated moving average (ARIMA) have been used for their simplicity and computational efficiency relative to physical-based models (Petty et al., 2017). These models often utilize historical sea ice concentration, temperature, and other meteorological variables to make short-term forecasts. However, they lack the ability to adequately capture the complex spatial and temporal patterns inherent in sea ice dynamics needed to forecast over longer timeframes.

The application of deep learning techniques to sea ice forecasting has gained increasing attention in recent years due to their computational efficiency and generalizability, particularly in the face of a changing climate and increased availability of large training datasets. Early studies applying deep learning to sea ice forecasting were limited to either spatial or temporal modeling. For instance, Chi and Kim (2017) used a long-short term memory (LSTM) module to forecast

sea ice on a per-pixel level but did not consider spatial patterns. Kim et al. (2019) later used a deep neural network (DNN) with two fully-connected layers to forecast sea ice concentration considering interactions between pixels through dense layers but did not explicitly account for spatial autocorrelation. Later models based on the convolutional neural network (CNN) were able to leverage spatial patterns. Andersson et al. (2021) used a U-net trained on both climate simulation and observation data to forecast monthly sea ice concentration and was found to out-perform the SEAS5 dynamical model, but did not explicitly model in the temporal dimensions. Spatiotemporal models were then proposed that unify spatial and temporal models. Liu et al. (2021) proposed a model based on the convolutional long-short term memory (ConvLSTM) (Shi et al., 2015) to perform one-step ahead forecasting of sea ice in the Barents sea that showed promise by outperforming statistical baselines. Asadi et al. (2022) built on this work by proposing a sequence-to-sequence model based on the ConvLSTM to forecast sea ice presence in Hudson Bay. The model generally outperformed the European Centre for Medium-Range Weather Forecasts's (ECMWF) subseasonal-to-seasonal (S2S) ensemble predictions (Vitart and Robertson, 2018). A related approach using a Conv-LSTM also for daily sea ice forecasts, with a focus on shorter lead times (without the sequence-to-sequence approach) is that by Liu et al. (2024). A novel contribution of this work is the use of empirical orthogonal functions (EOFs) both for dimensionality reduction and to enforce an additional physical constraint on the loss function. Related to this study, Zheng et al. (2024) developed an approach to forecast sea ice concentration and thickness incorporating EOFs and using a statistical approach to evaluate model outputs with the goal of improving model transparency.

Different from earlier studies, the present contribution evaluates a GNN approach to sea ice forecasting. The motivation is the synergy between a graph approach and the irregular structure coastlines brings to the spatial domain of sea ice forecasting. An earlier study showed such an approach was more efficient than a CNN for river ice prediction (Qu et al., 2024). In addition, GNN-based approaches have recently seen some attention in global climate modeling, motivated in part by successes in GNN-based physics simulation models such as MeshGraphNets (Pfaff et al., 2021) or graph network simulators (Sanchez-Gonzalez et al., 2020; Rubanova et al., 2022). Keisler (2022) first proposed a GNN for forecasting the global climate using an autoregressive encoder–processor–decoder architecture. Gridded reanalysis data was encoded onto an icosohedron graph structure on which a message-passing neural network performed several steps of processing before being decoded back onto the latitude–longitude grid. Results showed that the model is competitive in comparison with state-of-the-art physical models when forecasting geopotential height and temperature over a 6-day rollout with a 6-h temporal step. Lam et al. (2022) built upon this work with GraphCast, a similar model structure with the most notable difference being the use of multiple icosahedron grids at varying spatial resolution. They demonstrated greater skill than operational state-of-the-art physical models when forecasting global temperature, precipitation, and wind patterns over a 10-day rollout at a 6-h temporal step.

4. Methodology

4.1. Data

In this study, ERA5 reanalysis data is used as atmospheric forcing data to train the models along with oceanographic variables from the GLORYS12 reanalysis product. Sea ice concentration estimates from future dates of GLORYS12 are used as the target variable.

4.1.1. ERA5

ERA5 (Hersbach et al., 2020) is a climate reanalysis dataset produced by ECMWF that offers hourly estimates of climatic variables at a spatial resolution of 0.25° from 1979 to present. It is based on the IFS Cycle 41r2 4D-Var data assimilation system and includes a wide range of climatic variables at different pressure levels of the atmosphere. The IFS system assimilates observations from dozens of satellite missions and ground stations to create a physically consistent best representation of atmospheric conditions. Although the model does not have a coupled ocean–atmosphere component, it uses daily passive microwave-derived sea ice concentration estimates from the Ocean and Sea Ice Satellite Application Facilities (OSI-SAF) as boundary conditions (Hersbach et al., 2020). In this study, we follow previous studies (Asadi et al., 2022; Andersson et al., 2021) and use 2-m temperature, 10-m wind speeds, and surface sensible heat fluxes from ERA5 as input features to our model (see Table 1)

4.1.2. GLORYS12

GLORYS12 (Lellouche et al., 2021) is a global ocean and sea ice reanalysis data product developed by the Copernicus Marine Environment Monitoring Service (CMEMS), utilizing the LIM2 EVP NEMO 3.1 platform (Madec et al., 2016) in the ORCA025 configuration designed by the DRAKKAR consortium. This configuration includes a global sea-ice model with a 1/4° Mercator grid. Atmospheric forcing for the ocean surface model is provided by ECMWF's ERA-Interim (Dee et al., 2011) reanalysis data until 2019, and ERA5 data thereafter. The spatial resolution of the ocean and ice models is 1/12°, corresponding to approximately 15 km at this latitude. The data assimilation component of GLORYS12 includes in-situ temperature and salinity (T&S) profiles, satellite sea surface temperature (SST), and along track sea-level anomalies derived from satellite altimetry. The assimilation of oceanic observations occurs using a singular evolutive extended Kalman (SEEK) filter. The SEEK filter utilizes a three-dimensional multivariate background error covariance matrix and operates on a 7-day assimilation cycle. The system also integrates sea ice concentration observations from IFREMER/CERSAT. Sea ice concentration from GLORYS12 is comparable to that from other reanalyses in terms of seasonal cycle and extent in the Arctic. However, GLORYS12 does exhibit thicker ice in the Beaufort Sea than other reanalyses (Chevallier et al., 2016). We are not aware of a specific compariosn between sea-ice reanalyses for Hudson Bay. Historical records for the GLORYS data are available from 1993 to present. For model training, validation and testing we use only the data from 1993–2019 because the data source for atmospheric forcing changed in 2019. This study uses GLORYS12 sea ice concentration, thickness, velocities and sea surface temperatures.

4.2. Meshing

Within a GNN framework, the mesh structure allows for greater flexibility in defining the model's spatial basis. Unlike two-dimensional convolutional approaches, which require defining a regular two-dimensional grid of pixels over a region, meshes are comprised of cells of arbitrary sizes, allowing the modeler to control which areas are modeled in higher resolution (e.g., around ports or passages of interest). Since cells are only defined in regions of interest we also avoid the need to apply a land mask as a post-processing step, unlike in CNN-based approaches which most often model over the whole region before applying a mask to exclude land pixels from the output.

Fig. 1 shows possible meshes for Hudson Bay using a 1/12° grid as the base resolution when trying to balance resolution and computational requirements. The mesh shown in (a) uses the base resolution as a regular mesh comprised of 32,856 cells, while the mesh in (b) uses a regular four-times coarsened version of the same mesh comprised of 2425 cells. As a compromise between resolution and computational efficiency, an irregular mesh can be defined with the same four-times coarsened resolution refined near shorelines such that no cell overlaps

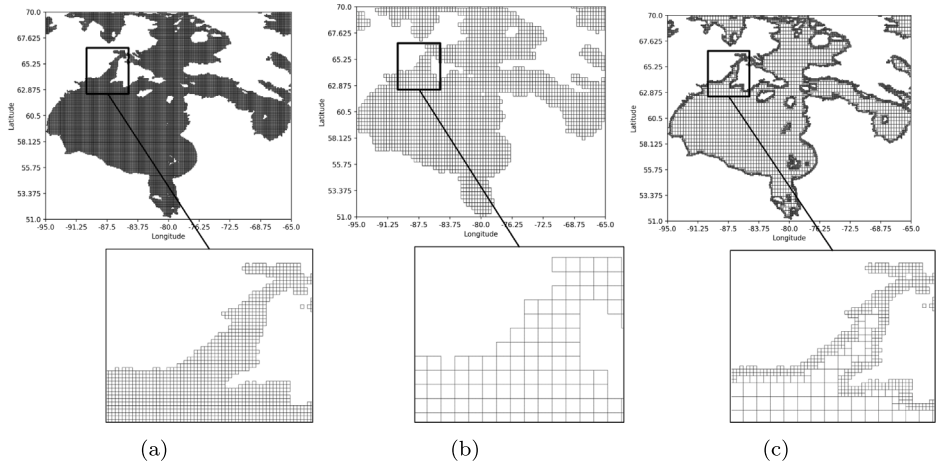


Fig. 1. Comparison of different mesh definitions for modeling Hudson Bay. (a) A high-resolution regular mesh with 32,856 cells, computationally intensive but highly detailed; (b) a four-times coarsened regular mesh with 2425 cells lacking sufficient detail along land interfaces; (c) irregular mesh with 9422 cells, a compromise for both computational efficiency and high resolution at land interfaces. This approach ensures no cell overlaps land while providing high-resolution data for critical regions like ports, passages, and areas of meteorological interest such as the Kivalliq latent heat polynya.

land. This is shown in (c). This can be done by recursively splitting the cells of the base (coarsened) mesh in four equal parts until no cell overlaps land. The result is a mesh with 9422 cells. A secondary advantage of this technique is that modeling around shorelines at a higher resolution may be of interest to port operators or local communities. For shipping and freight purposes in Hudson Bay, there is a keen interest in knowing the state of the ice near shipping ports since some operations might required ice free conditions.

To convert gridded data from a grid representation $X \in \mathbb{R}^{W\times H\times C}$ for data with C channels and $W \times H$ spatial dimensions to a mesh representation $G \in \mathbb{R}^{C\times N}$ with N cells, we first construct a sparse mapping tensor $M \in \mathbb{R}^{N\times W*H}$ where entry (n, p) is assigned 1 if the pth pixel of the flattened grid $Y \in \mathbb{R}^{C\times W*H}$ should be mapped to cell n. We also construct a tensor $P \in \mathbb{R}^{N}$ which stores the number of pixels which are mapped to each cell. Then, to convert a sample from a grid to a mesh representation, for each node we find the mean value of each of its constituent pixels with

$$G = Y M^T \oslash P \tag{1}$$

where $\oslash$ represents an element-wise or Hadamard division. G can be converted back to a grid representation by splitting the cells back into its constituent pixels as

$$\hat{Y} = GM. \tag{2}$$

Since Eq. (1) takes the mean of the constituent pixels of each cell, it cannot be perfectly inverted, instead Eq. (2) simply assigns the cell value to each of its constituent pixels. Formulating these transformations as matrix multiplications allows for greater GPU acceleration.

A graph can then be defined based on this mesh by assigning a node to each cell and placing edges between any two neighboring cell as in Fig. 2. To preserve spatial awareness, the positions of each node and size of each cell are added as node features, and the length and angle of the edges are stored as edge features. The edges are therefore considered to be directed edges as the edge features are direction-dependent, that is, for two nodes x_i and x_j, the edge from x_i to x_j (e_{ij}) is not equivalent to the edge from x_j to x_i (e_{ji})

4.3. Model architecture

The proposed model uses graph convolutional long-short term memory (GCLSTM) modules within a sequence-to-sequence architecture. The GCLSTM module and the overall architecture are shown in Fig. 3, and described in the subsections below.

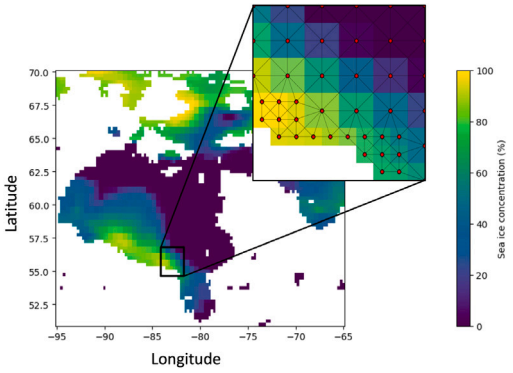


Fig. 2. Input images are represented as graphs by relating each neighboring pixel with edges. In this figure, a spatially irregular mesh is used to represent SIC in Hudson Bay, where red dots represent graph nodes and black lines represent edges. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.3.1. GCLSTM

The graph convolutional long-short term memory (GCLSTM) module used in this work is a modified version of the model from Seo et al. (2018), which is in turn inspired by the ConvLSTM first proposed by Shi et al. (2015). The module closely resembles the peephole LSTM introduced by Gers et al. (2002), with the only modification being the addition of graph convolution operators over the hidden and input states at each of the input, forget, cell and output gates in the place of weight matrices. This is represented as the $*\mathcal{G}$ block in Fig. 3b. The graph convolution operators allow information exchange between nodes through the directed edges. The model proposed by Seo et al. (2018) uses a single graph convolution, which has limited spatial expressivity since a single convolution can only exchange information between immediate neighbors. Since the processes dominating ice formation and break-up are physical processes occurring across space, we wish to increase the model's ability to recognize spatial patterns, and therefore use K stacked convolutions followed by leaky ReLU activations, which provides information exchange over K hops. The peephole variant of the LSTM is used here as it has been shown to outperform the vanilla LSTM (Joshi et al., 2022), particularly for video understanding (Srivastava et al., 2015). The convolution operator taking the place of GraphConv in Fig. 3b can be arbitrarily selected from

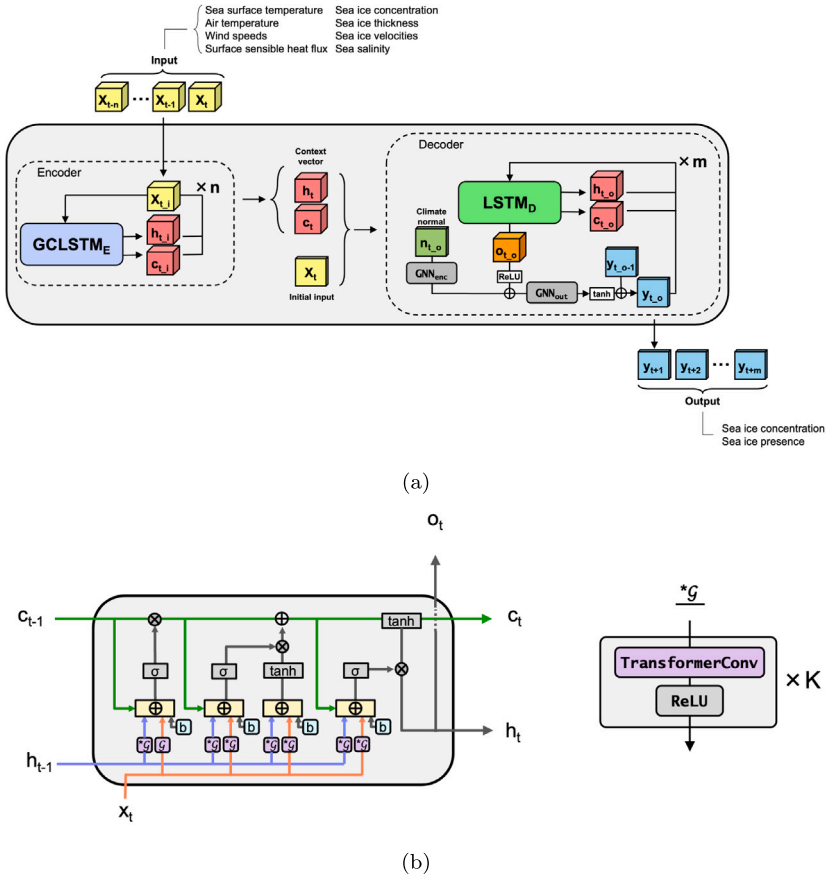


Fig. 3. (a) Overall model architecture. The last hidden (h_t) and cell (c_t) states of the encoder act as the context vectors and are used as the initial states of the decoder. The encoder learns features from the n input timesteps, and the last hidden (h_t) and cell (c_t) states are retained as the context vector used to initiate the decoder, which unrolls over the fixed m desired output timesteps. The initial input to the decoder X_t is the ice channel of the last input timestep. GNN_{enc} and GNN_{out}, used to encode climatology at each output timestep (n_{t_o}) and reduce the dimensionality of the output (o_{t_o}), respectively, are stacked spatial convolutions with leaky ReLU activations. (b) A single graph convolutional long-short term memory (GCLSTM) cell. $\oplus$ represents element-wise addition and $\otimes$ represents element-wise multiplication.

the myriad graph convolution operators that have been proposed. In this work, we evaluate both the graph transformer convolution from Shi et al. (2021), and the more basic Graph Convolutional Network (GCN) first proposed by Kipf and Welling (2017).

In the graph transformer convolution, the feature vector of a given node i, x_i, is updated by aggregating information from its neighbors $j \in \mathcal{N}(i)$, and the node itself, using edge features from i to j, e_{ij}. The governing equation for the graph transformer convolution is

$$x_i' = W_1 x_i + \sum_{j \in \mathcal{N}(i) \cup i} \alpha_{ij}(W_2 x_j + W_3 e_{ij}) \tag{3}$$

where $\mathcal{N}(i)$ denotes the neighbors of node i, W are weight matrices that project the inputs to their latent representation where the attention coefficients α_{ij} are given by

$$\alpha_{ij} = softmax\left(\frac{(W_4 x_i)^T (W_4 x_j + W_3 e_{ij})}{\sqrt{d}}\right). \tag{4}$$

The attention weights allow the model to selectively attend to a given node's neighbors based on their node and edge feature vectors. The inclusion of edge features and an edge specific weight matrix allows the model to learn to relate the edge features to better reflect anisotropic evolution of the model state.

We compare the transformer convolution with the Graph Convolutional Network (GCN) proposed by Kipf and Welling (2017), as it is a commonly used and simpler convolution operator. The GCN operator is defined by the equation

$$x_i' = W^T \sum_{j \in \mathcal{N}(i) \cup i} \frac{e_{ij}}{\sqrt{\hat{d}_j \hat{d}_i}} x_j \tag{5}$$

where X is a weight matrix, $\hat{d}_i = 1 + \sum_{j \in \mathcal{N}(i)} e_{ij}$ and e_{ij} are the edge weights from i to j. Since e_{ij} must be a scalar, here we use the normalized distance between nodes as the edge weights. Note that this limits the spatial awareness of the model as it does not receive information about the nodes' relative positions, unlike the transformer convolution.

4.3.2. Sequence-to-sequence architecture

The GCLSTM module is used within a sequence-to-sequence encoder–decoder structure to learn features from the inputs and evolve the sea ice state forward in time. The overall architecture is shown in Fig. 3a. Since navigation and offshore operations are affected at various degree by the presence and concentration of sea ice, our model forecasts both SIC and SIP as a multi-task learning approach. Although sea ice presence, defined as any pixel where SIC is greater than 15%, can be derived from the forecasted SIC values, a model trained without the secondary SIP forecasts would not be optimized for this 15% threshold. It was also found through experimentation that including SIP as a secondary task improved SIC forecasts in the break-up and freeze-up seasons.

The encoder is responsible for learning rich spatiotemporal features from the input sequence while the decoder is responsible for evolving

the state forward in time from these learned features. The encoder therefore acts as an information bottleneck, meaning it is crucial that the encoder is sophisticated enough to distill the inputs into a context vector with sufficient information for the decoder to use in the unrolling process. Given a sufficiently rich context vector, the decoder does not necessarily need to learn additional spatial features within the context vector, nor during the unrolling process. Therefore, in this work we use a spatiotemporal GCLSTM module in the encoder block, and a simple LSTM in the decoder block. Although the decoder block also contains graph convolutions (e.g., in GNN_{out}), the distinction between the two is that the GCLSTM in the encoder block integrates graph convolutions within the temporal model allowing for simultaneous spatial and temporal modeling, while the decoder block models temporal and spatial dynamics separately, with GNN_{out} being used mainly for dimensionality reduction. Using an LSTM rather than a GCLSTM module in the decoder block also greatly reduces training time in the case where there are fewer input timesteps than output timesteps. Note that experiments with a GCLSTM in the decoder were also run but showed no improvements over using an LSTM.

The encoder processes each input timestep sequentially, updating the hidden and cell states at each timestep with layer normalization (Ba et al., 2016) applied to the hidden and cell states after each timestep to increase model stability. The final hidden and cell states are the high-dimensional vectors that are taken as the context vectors that contain the learned features from the input and are used to initialize the hidden and cell state of the decoder. The last input ice state is used as the initial input to the decoder (or start token) since we wish to evolve the state forward from this initial state. The decoder is run recurrently for the desired number of output timesteps in a similar fashion to the encoder but using the last step's prediction (y_{t-1}) as the input for the current step (y_t).

Since sea ice is highly seasonal, the model is susceptible to a form of modal collapse wherein the model converges to a local minimum, predicting only the average sea ice conditions for a particular day of the year. These daily averages are known as the climate normals or climatology. For long-term forecasting of climatological variables, climatology can perform reasonably well compared to dynamic or statistical models due to strong seasonality. Since we wish to outperform climatology and expect the model to learn to use it as a heuristic, we choose to include it as an input to the decoder such that model can focus on learning departures from normal conditions. This was shown to be beneficial for sea ice forecasting in a previous study (Asadi et al., 2022). Climate normals are calculated as the mean ice concentration values for each day of the year over the entire training set and are encoded into latent space using a shallow multi-layer GNN before being combined with the decoder output by element-wise addition. We include climatology in the decoder so that the model can directly condition each forecast timestep on the climatology for that specific day-of-year, avoiding the need for the model to learn temporal alignment across from the encoded inputs. The result is then fed through a multi-layer GNN with leaky ReLU activations to reduce the dimensionality to two, and finally through a hyperbolic tangent activation to map the values between −1 and 1. This output represents the change in sea ice conditions and is added to the last timestep's prediction. Since both SIC and SIP should be bound between 0 and 1, the output is passed through a sigmoid layer that produces the final predictions.

4.4. Experimental set-up

4.4.1. Mesh definition

Experiments were run on both the coarsened regular mesh described in Section 4.2 and shown in Fig. 1b and the irregular mesh shown in Fig. 1c. The irregular mesh is refined to a higher resolution at the land edges by splitting the base 1/3° mesh if a cell intersects a one-cell buffer around land. The resulting irregular mesh contains 1/12° 1/6° and 1/3° sized cells.

4.4.2. Data partitioning

The Hudson Bay region, including Hudson Strait, James Bay and Foxe Basin, undergoes a cyclical transformation in its ice cover characterized by complete freezing during the winter months and total melt in the summer, with some multi-year ice possible in Foxe Basin. This seasonal cycle is subject to considerable inter-annual variability, both in terms of the rate at which these processes occur and the timing of these transitions. Fig. 4 illustrates this variability by showing monthly SIC anomalies between 1993 and 2020. These anomalies are computed as the mean differences between observed SIC and the long-term average concentration for each corresponding month. The data reveals distinct periods of anomalous behavior in SIC. Specifically, the years 1993 to 1997 were marked by higher-than-average SIC, indicating that during these years, Hudson Bay experienced an earlier freeze-up and a delayed break-up season. In contrast, the period from 2010 to 2012 exhibited anomalously low SIC, characterized by a late onset of freeze-up and an earlier melting season. Including data from both these anomalous periods along with years that exhibit more typical ice conditions is critical for enhancing model robustness in the face of varying environmental conditions. This is particularly important in the context of climate change, where shifts in temperature and weather patterns could further exacerbate the variability in sea ice conditions. The data is therefore partitioned into a sequential 20-year, 1-year, 6-year split, wherein data from 1993–2012 is used for training, 2013 is used for validation, and 2014–2019 is used for testing. This type of split of training, validation and test data is common in deep-learning based sea-ice forecasting (Zheng et al., 2024; Liu et al., 2024). Note that the test period includes years with normal or lower-than-normal ice conditions. While these test years may be representative of future ice conditions in the Hudson region (Stroeve and Notz, 2018), the model has fewer years with these conditions in comparison to above-normal ice concentration in the training data, in particular for the break-up period.

One model is trained for each month of the year, each denoted as a 'monthly model'. Each monthly model was trained using data from the respective month with a 15-day buffer before and after the beginning and end of the month respectively. For example, the April model is trained with input data for each day between March 16 and May 15 over all training years. A longer buffer of one month was tested but did not lead to significant improvements in model performance. In inference mode, each model is used only to produce a forecast with inputs from its respective month. For example, to generate 90 day forecasts for April, a 90 day forecast is launched for each day between April 1 and April 30. Training separate model for each month of the year was done since we expect the dynamics that must be learned for one time of the year to be sufficiently different from other times of the year such that each model will have greater accuracy by concentrating efforts in learning specific ice dynamics (Asadi et al., 2022). As a secondary benefit, this also allows training to be carried out more efficiently as each monthly model can be trained in parallel.

4.4.3. Input features

Sea ice concentration data from GLORYS12 serve as the target variable, while atmospheric variables from ERA5, combined with oceanographic variables from GLORYS12, are used as input features. Sea ice dynamics are primarily influenced by factors such as air and sea temperature (Wang et al., 2019), wind (Stammerjohn et al., 2003), heat fluxes (Ivanov et al., 2012), and ocean salinity (Yao et al., 2000), thus we include these variables as input features. The 10 chosen input variables are listed in Table 1, along with the rationale for their selection. It should be noted that ERA5 hourly variables are re-gridded from their original 1/4° grid to match the GLORYS12 1/12° grid, and resampled to match the GLORYS12 daily temporal resolution. This is achieved through spatial linear interpolation and aggregation from an hourly to a daily resolution using a simple mean. The input sequence length is 10 days and the spatial domain as a grid is 229 × 361. A 10-day input window was selected as a practical balance between

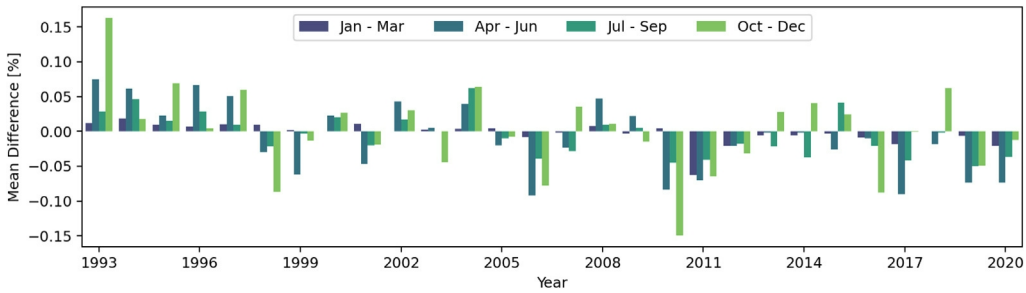


Fig. 4. Monthly sea ice concentration anomalies in Hudson Bay from 1993–2020. Highlights periods of higher and lower-than-average sea ice concentrations.

Table 1
List of variables used as input to the architectures evaluated.

Short name	Full name	Source	Rationale for inclusion
`sic`	Sea ice concentration	GLORYS12	Direct measure of what is being forecasted; crucial for temporal dynamics and initial conditions.
`sit`	Sea ice thickness	GLORYS12	Provides insights into the resiliency and robustness of the ice, affecting its likelihood to melt or deform.
`siuv`	Sea ice velocities	GLORYS12	Indicates the direction and speed of sea ice movement.
`so`	Sea water salinity	GLORYS12	Affects the freezing point of sea water.
`sst`	Sea surface temperature	GLORYS12	Affects ice melt and formation rates.
`t2m`	2-meter temperature	ERA5	Provides additional context for the thermal conditions affecting the sea ice surface.
`u10v10`	10-meter wind velocity	ERA5	Influences the motion and deformation of sea ice.
`sshf`	Surface sensible heat flux	ERA5	Indicator of heat exchange between the atmosphere and the sea surface, affecting ice melt and formation.
`x`	x-position of each node	–	Provides the latitudinal spatial context for each data point.
`y`	y-position of each node	–	Provides the longitudinal spatial context for each data point.
`doy`	Day of the year	–	Provides temporal context.
`csize`	Cell size	–	Provides the relative area covered by each cell for additional spatial context.

computational efficiency and predictive performance, and because the use of monthly models and climatology inputs means the models are implicitly seasonally contextualized without requiring longer input sequences. Since the model operates over the mesh domain rather than the grid domain, the dimensionality of the inputs to the encoder as (input steps, number of nodes, input features) is $10 \times 9422 \times 10$ for the irregular mesh and $10 \times 2425 \times 10$ for the regular mesh. The input to the decoder is the context vectors provided by the encoder as well as the climatology for each forecast day. The output dimensionality is $90 \times 9422 \times 2$ for the irregular mesh, and $90 \times 2425 \times 2$ for the regular mesh.

4.4.4. Baseline model

As a baseline model with which to compare the GNN approaches, we use a combination of two common statistical baselines: persistence and climatology. Persistence refers to persisting the most recent sea ice conditions and tends to perform well at very short forecast lengths particularly outside of the freezing and melting seasons. Climatology refers to the pixel-wise average SIC for each day of the year where the average is taken over the historical period of interest. Climatology tends to perform best relative to forecast models at longer lead times. For forecasts produced over a seasonal scale, a stronger baseline than either persistence and climatology can be derived by combining the two using a weighted average with the relative weights varying by lead time, where more weight is given to persistence than climatology at short lead times and more weight is given to climatology than persistence at long lead times. The form chosen for the baseline model is

$$F = (1 - \gamma)P + \gamma C, \tag{6}$$

where

$$\gamma(t) = \gamma_0 \times e^{-\lambda t}. \tag{7}$$

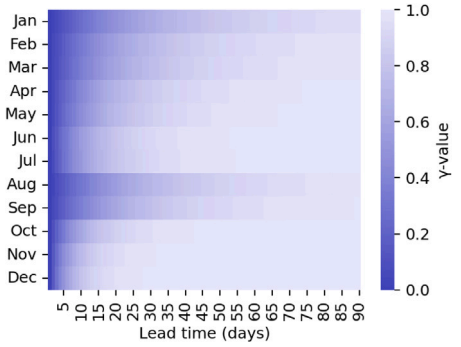


Fig. 5. $\gamma(t)$ values for the baseline model (Eq. (6)) showing the balance between persistence and climatology by the month of the launch date and lead time. Gamma values near 0 favor persistence while values near 1 favor climatology. Less variable ice seasons such as January/February and August/September rely more on persistence for longer lead times.

γ_0 is set to 1 since we know persistence to be a strong predictor at short lead times, and λ is optimized by minimizing the mean squared error over the training dataset for each month. The resulting $\gamma(t)$ values are shown in Fig. 5.

4.4.5. Model hyperparameter configurations and implementation

This study evaluates three distinct models, listed in Table 2. Our primary focus is the GraphSIFNet-Att model, which incorporates three TransformerConv spatial convolutions in the GCLSTM block and is trained on the irregular mesh for 35 epochs. That is, in Fig. 3b, $*\mathcal{G}$ uses the TransformerConv as the GraphConv block with $K = 3$. For comparison, we examine the GraphSIFNet-Att-Reg model which is

Table 2
Summary of developed model configurations. The models differ in their spatial convolutions and their underlying meshes, with the aim of contrasting the attention-based transformer convolution with the graph convolutional network, as well as demonstrating the model's ability to model over an irregular mesh.

Name	Convolution (# layers)	Mesh	Approximate training time
GraphSIFNet-Att	TransformerConv (3)	Irregular	10 h (30 epochs)
GraphSIFNet-Att-Reg	TransformerConv (3)	Regular	8 h (30 epochs)
GraphSIFNet-GCN	GCN (6)	Irregular	10 h (45 epochs)
Baseline	N/A	N/A	N/A

identical in architecture but trained on the coarsened regular mesh from Section 4.2 for 35 epochs. Additionally, we compare with the GraphSIFNet-GCN model, which employs six GCN convolutions within the GCLSTM module, that is, the GraphConv block is the GCN with $K = 6$. GraphSIFNet-GCN is trained over the irregular mesh for 45 epochs. Each of these models have the same number of parameters (approximately 123,000).

Each model uses a 10-day input sequence to predict the subsequent 90 days, A hidden dimension size of 32 is used for each of the hidden state and cell state of the encoder and decoder LSTMs, as well as in all graph convolutional layers. The GNN used to encode climatology (GNN_{enc}) is comprised of a single graph convolution layer, and the output GNN (GNN_{out}) is comprised of 3 stacked convolution layers with leaky ReLU activations. The hidden size, number of spatial convolutions and number of GCLSTM/LSTM layers were chosen based on small-scale experiments which aimed to keep the model simple yet effective. The optimizer is the Adam optimizer with an initial learning rate of 0.001 reducing by 10% every 5 epochs. An L2 regularization value of 0.01 is applied to the weights reduce the risk of overfitting, and gradient clipping with a value of 1.0 is applied to mitigate the risk of gradient explosion due to the extended forecast length. Early stopping was used if no improvement in the validation loss was observed for 10 epochs. Since the model produces two outputs, SIC and SIP, a custom loss function was used that combines a mean square error (MSE) loss from the continuous SIC prediction and binary cross-entropy (BCE) loss from the probabilistic SIP prediction. The BCE loss is scaled by a factor of 0.1 and added to the MSE loss before back-propagation. Losses also include scaling to account for the variable size of each grid cell. This prevents the model from over-valuing correct predictions in areas of higher spatial resolution. The models are implemented in Pytorch using the pytorch-geometric (Fey and Lenssen, 2019) package and trained on a single Tesla V100 GPU. A summary of models tested and training times is given in Table 2.

5. Results

5.1. Baseline and GraphSIFNet-Att model performance

The performance of the baseline statistical model defined by Eq. (6) for the SIC forecasting task is shown in Fig. 6a. These heatmaps are generated by calculating the spatial average of the root mean squared error (RMSE) over the domain using only the test years (2016–2019). The errors are grouped by the month of the launch dates and lead times. For instance, the value in the top right corner of the error heatmaps (January, 90-day lead time) indicates the mean RMSE for all 90-day forecasts launched in January, that is, forecasts for dates spanning April 1st to May 1st. The two clearly visible bands of higher RMSE values correspond to the break-up and freeze-up seasons, the former normally spanning from the beginning of May to mid-July and the latter normally spanning from the beginning of November to the end of December. These seasons are the most difficult to forecast as the timing and pattern of the break-up and freeze-up vary between years. Conversely, August to beginning of October are largely ice-free, thus the errors are near zero. In the winter months, that is, mid-December to the beginning of April, ice is present throughout the Hudson Bay system though some open water can sporadically be found around shorelines, for example due to offshore winds, thus SIC RMSE values during the winter months are small but not zero.

The performance of GraphSIFNet-Att model and the difference in performance relative to the baseline model is shown in Fig. 6b and c, respectively. For the majority of the months and lead times, the GraphSIFNet-Att model exhibits improvements in SIC forecasts over the baseline, with minor exceptions. The model exhibits the largest improvements over the baseline in its short- to medium-term forecasts of the break-up season (lead times 5 to 45 launched in May to July). These show up to a 10% improvement over the baseline. At longer timesteps, the improvements over the baseline during the break-up period (launched in March and April) are less pronounced, hovering around 2%–3%. However at these long lead times even small improvements demonstrate forecast skill and can provide value to users of the system. During the winter months when the region is almost entirely frozen, the model still exhibits a 2%–3% improvement over the baseline at all lead times. During freeze-up, the model only shows skill over the baseline at short lead times from 0 to 25 days. Longer forecasts beyond 25 days perform on par with the baseline or only marginally better. Forecasts launched in November with a 30 to 55 day lead time perform worse than the baseline, indicating difficulty in capturing the final stages of ice formation.

The SIP accuracy heatmaps in Fig. 7 show similar patterns, with increases in accuracy of up to 20% from the GraphSIFNet-Att model over the baseline during the break-up process, and more modest increases during the freeze-up process. Notably, however, GraphSIFNet-Att outperforms quite significantly (>10%) even at long lead times. This indicates that although the model may struggle to forecast the precise SIC at these lead times, it still has skill in forecasting the point at which the ice will completely melt or break up.

5.2. Comparison between model configurations

To better understand the differences in the model configurations, spatial monthly SIC RMSE maps for 30-, and 60-day lead times for forecasts launched in May and November are presented in Fig. 8 and Fig. 9. These correspond to parts of the break-up and freeze-up periods, respectively. Panels (a) and (c) show the impact of the convolution operator, while panels (b) and (d) show the impact of the mesh resolution. At both 30-day and 60-day lead times GraphSIFNet-Att shows improvements in May in comparison to GraphSIFNet-GCN (Fig. 8a), indicating the incorporation of attention is helping the forecasts at this time. The impact of the refined mesh is less clear, although in Fig. 8b decreases in RMSE (indicating improved performance) can we seen near the coastlines in some regions, more so in the southern parts of the domain where the ice may be melting, or moving away from shore. During freeze-up the results are more mixed (Fig. 9a). In Hudson Bay GraphSIFNet-Att demonstrates clear improvements over GraphSIFNet-GCN at both 30-day and 60-day lead times. However, there is some degradation in Hudson Strait. The impact of mesh refinement is mixed at 30-day lead times and not significant at 60-day lead times (Fig. 9b). Since all three models exhibits slightly improved performance, with additional interpretability granted by the attention weights in GraphSIFNet-Att, this may motivate the use of GraphSIFNet-Att over the others.

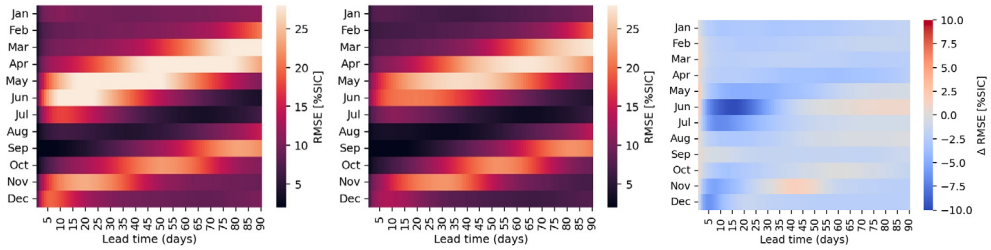


Fig. 6. Performance of the baseline statistical model and GraphSIFNet-Att over the test years aggregated by the month of the launch date and lead time for SIC; (a) Baseline model (b) GraphSIFNet-ATT, (c) GraphSIFNet-Att-Baseline. Blue indicates a reduction in RMSE and therefore an improvement. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

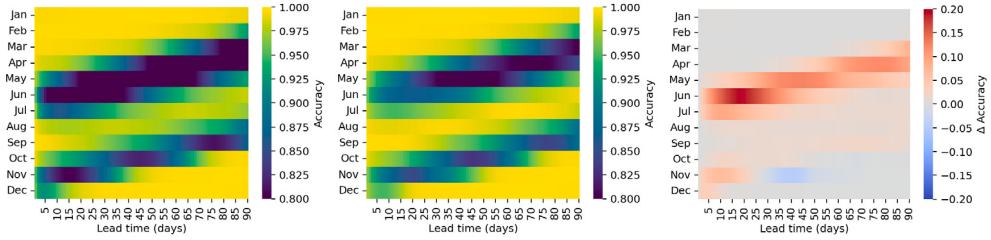


Fig. 7. Accuracy heatmaps for the SIP forecasting task by month and lead time for the GraphSIFNet-Att model (a), and the RMSE differences between the GraphSIFNet-Att model and the baseline (b), the GraphSIFNet-Att-Reg (c) and GraphSIFNet-GCN models (d). Positive values (red) in the difference plots indicate an increase in model accuracy from the GraphSIFNet-Att model relative to the respective baseline model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

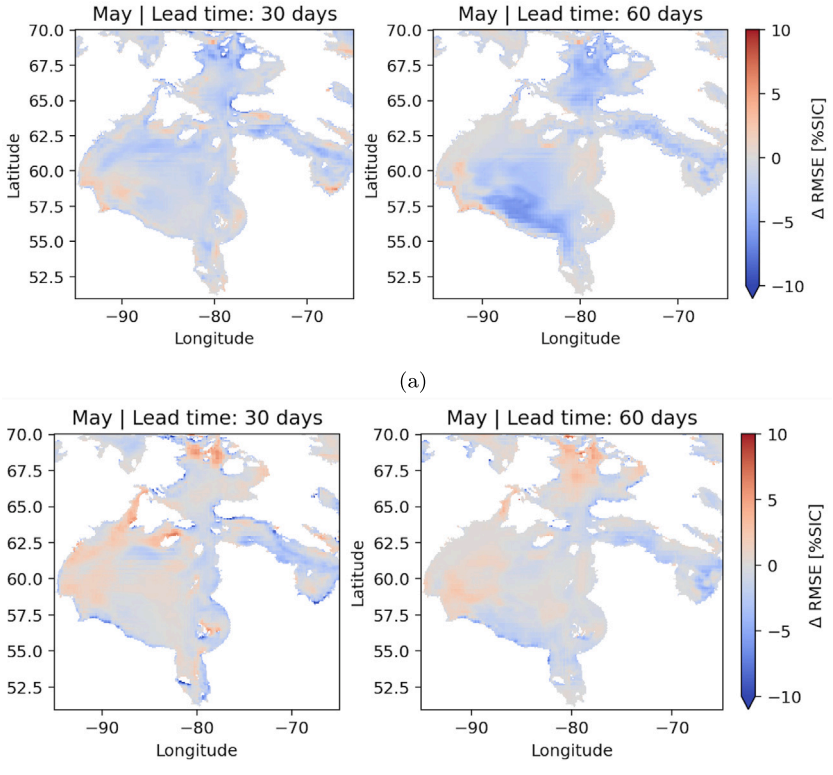


Fig. 8. Comparison of SIC RMSE for GraphSIFNet-Att, GraphSIFNet-Att-Reg, and GraphSIFNet-GCN models at 30- and 60-day forecast lead times, initiated in May. The figure shows the difference in RMSE between GraphSIFNet-Att and both GraphSIFNet-Att-Reg and GraphSIFNet-GCN. Negative values (blue) indicate a reduction in error in the GraphSIFNet-Att relative to the other indicated model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

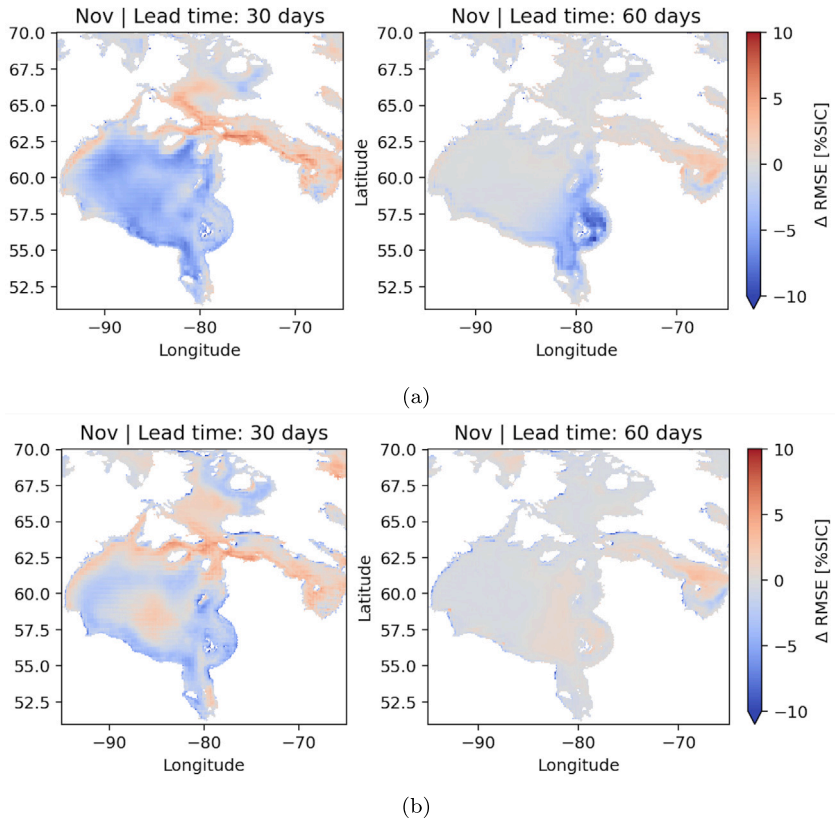


Fig. 9. Comparison of SIC RMSE for GraphSIFNet-Att, GraphSIFNet-Att-Reg, and GraphSIFNet-GCN models at 30- and 60-day forecast lead times, initiated in November. The figure shows the difference in RMSE between GraphSIFNet-Att and both GraphSIFNet-Att-Reg and GraphSIFNet-GCN. Negative values (blue) indicate a reduction in error in the GraphSIFNet-Att relative to the other indicated model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

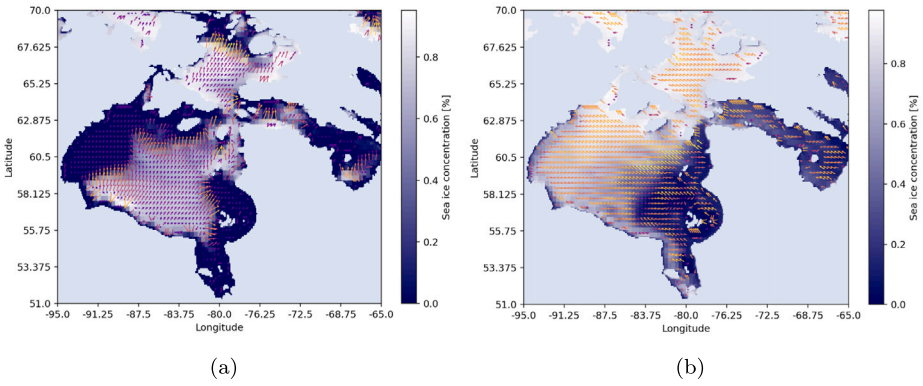
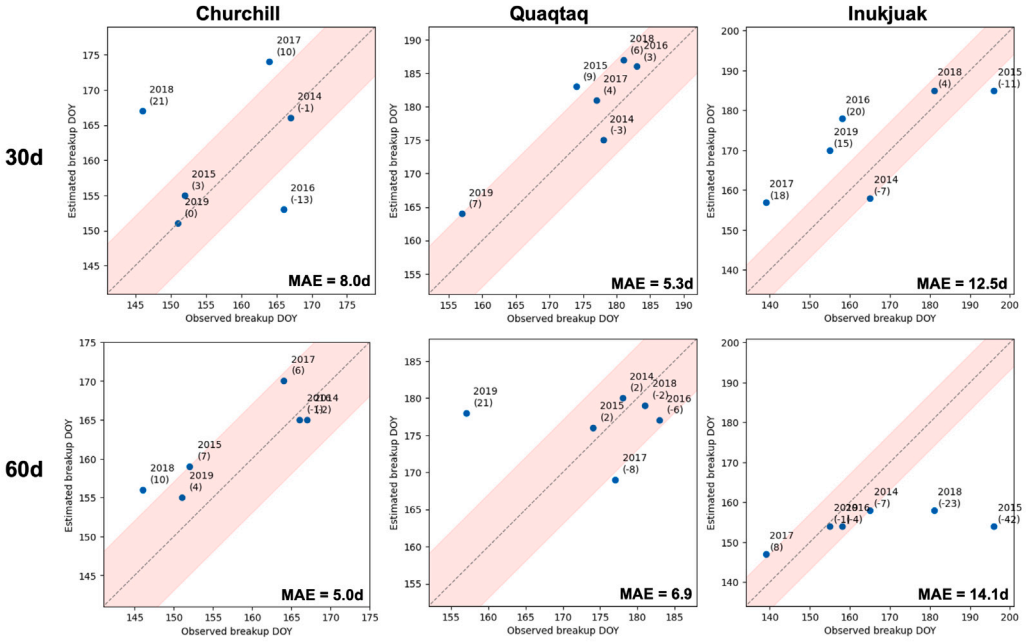


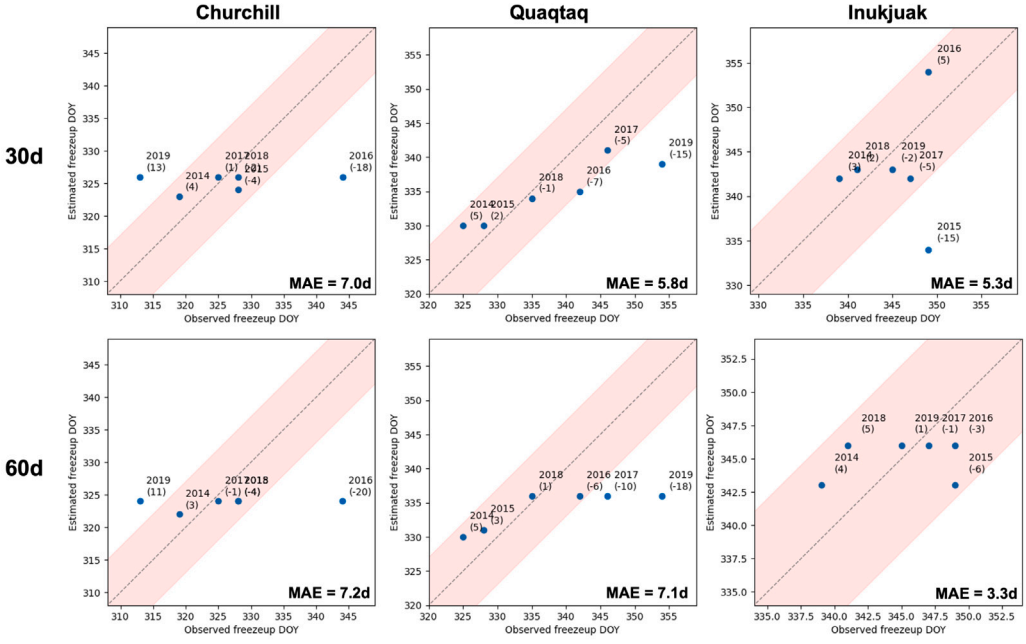
Fig. 10. Visualization of attention weights of the input gate applied to the input tensors during the freeze-up (a) and melting (b) seasons overlaid on the sea ice concentration input. Arrows indicate the primary direction and magnitude of information flow based on the learned attention weights. Attention weights at the land interfaces are omitted for clarity. The attention weights appear to be largely influenced by sea ice concentration, but other input variables also influence the weights, for example surface sensible heat flux in (a), and sea ice thickness in (b).

5.3. Attention maps

The use of transformer convolution in the model enhances its interpretability. By examining the attention weights in the encoder's first layer of graph convolutions, insights can be gleaned into how the model encodes the input data. According to Eqs. (3) and (4), each node is assigned attention weights for its neighboring nodes based on learned weight matrices in each transformer layer. The softmax function ensures that the sum of all attention weights for a given node's neighbors equals 1. Consequently, the node is updated using a weighted average of its neighbors' features, which are projected into a latent space. Due to the large number of edges, visualizing these weights on a simple map is challenging. A simpler approach for visualization involves calculating the primary direction from which each node is updated. This can be done by summing the attention weights as vectors (α values in Eq. (3) with the direction of their respective edges) for each node. These can be represented by arrows, the magnitude of which is proportional to the difference in weights. For example, a node with evenly distributed attention weights among eight neighbors would be represented as a single dot, whereas a node with a dominant westward neighbor would have a large arrow pointing westward. These arrows can be interpreted



(a) Break-up date estimates



(b) Freeze-up date estimates

Fig. 11. Break-up and freeze-up dates predicted by GraphSIFNet-Att at Churchill, Inukjuak, and Quaqtaq ports for lead times of 30 and 60 days for the years 2014 to 2019 compared to the observed dates from GLORYS12. The pink shaded area represents a 7-day buffer around a perfect forecast. Samples which fall within this buffer are deemed correct forecasts. The annotated numbers in parentheses are the error for each year. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as indicating the direction of information flow through the graph as the model processes the input maps.

Fig. 10 provides examples of attention weights of the input gate for a date during both freeze-up (Fig. 10a) and melting (Fig. 10b). Other dates show similar results. Although the attention mechanism is applied to the hidden and input tensors at each of the LSTM gates, it is most informative to visualize the weights that are applied to the inputs since the inputs are physically interpretable. In the freeze-up condition, the model directs information flow generally from the southeast to the northwest, i.e., the nodes are 'looking to the northwest'. This suggests that the model learns the importance of understanding the sea ice and atmospheric conditions of nodes to the northwest, aligning with the direction of freezing. It is logical that a node that contains water should know the condition of its 3-hop neighbor to the northwest, as if this neighbor is frozen, it is likely that this node will freeze in the near future. Conversely, during the melting season, arrows point towards open water, indicating that nodes with icy conditions but with water-containing neighbors should consider these neighbors important as they indicate the node is likely to melt soon. Notably, the magnitude of the arrows are larger at the ice edge and nearly zero in the consolidated ice

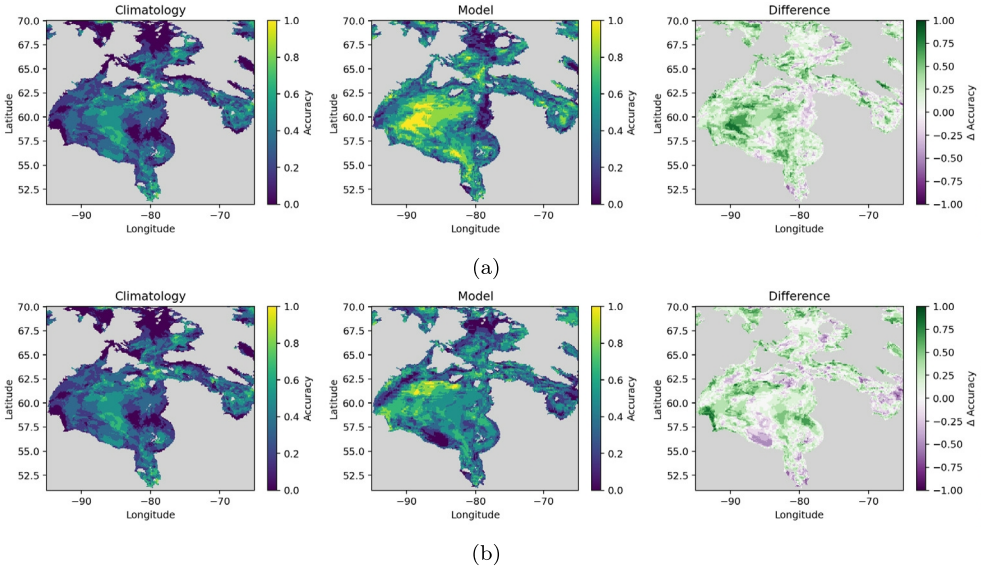


Fig. 12. Break-up date estimate maps for (a) 30 days, (b) 60 days. From left to right, estimates from climatology, GraphSIFNet-Att model predictions, the difference between the two. Positive values in the difference plots indicate an increase in accuracy from the model relative to the baseline, where accuracy is defined as the proportion of predictions falling within 7 days of the observed date for the years 2014 to 2019.

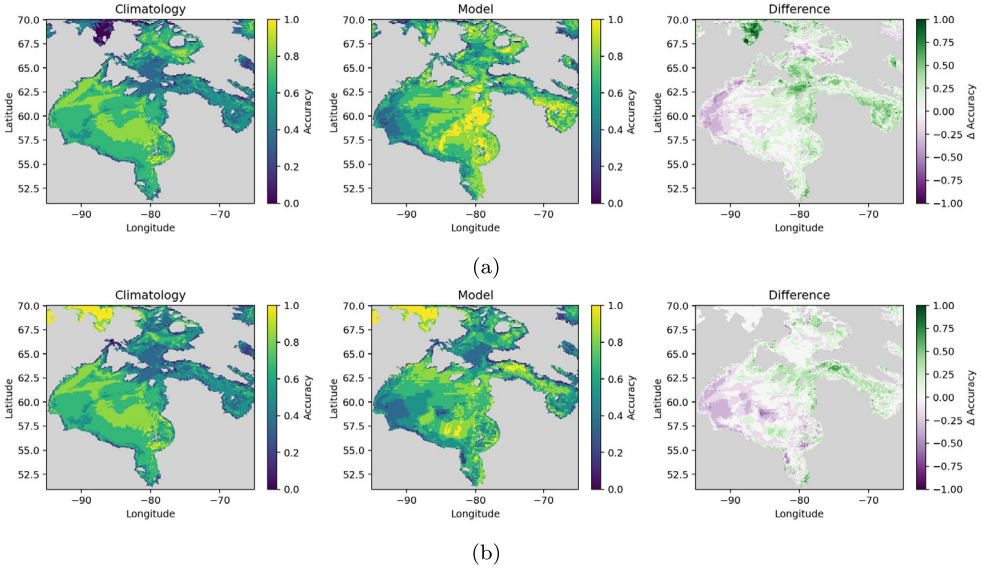


Fig. 13. Freeze-up date estimate maps for (a) 30 days, (b) 60 days. From left to right, estimates from climatology, GraphSIFNet-Att model predictions, the difference between the two. Positive values in the difference plots indicate an increase in accuracy from the model relative to the baseline, where accuracy is defined as the proportion of predictions falling within 7 days of the observed date for the years 2014 to 2019.

region, which could reflect the localized nature of the break-up process compared to the more gradual freeze-up, although they are still looking towards the open water. Nodes in the open water region during the melting season are less likely to change and, therefore, do not require attention to specific neighbors, as is indicated by the small arrows in open water.

5.4. Estimating break-up and freeze-up dates

A potential use-case for sea ice prediction in Hudson Bay is the estimation of break-up and freeze-up dates in key locations, as well as across the region. These dates have significant implications for marine navigation and local communities for planning purposes. We evaluate the GraphSIFNet-Att model's performance in estimating the freeze-up date at three key ports in Hudson Bay: the ports of Churchill, Quaqtaq and Inukjuak. The port of Churchill is mostly used to export grain while the ports of Quaqtaq and Inukjuak are regularly used for community resupply. These three ports were chosen as their locations are representative of the varying sea ice conditions found in the Hudson Bay region.

In this study, the test years (2014 to 2019) serve as the period for assessing the predicted break-up and freeze-up dates. These dates are determined using the same criteria as the previous study, which follows the definition given by the Canadian Ice Service (CIS). That is, the break-up date at a given site is defined as the initial day when open water persists for 15 consecutive days, with open water being defined as a SIC of less than 15%. Conversely, the freeze-up date is defined as first day at which SIC exceeds 15% for 15 consecutive days. The 30-day and 60-day predicted break-up and freeze-up dates are determined using the same criteria, but with open water and ice conditions being

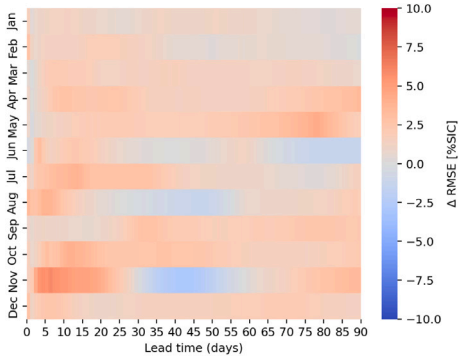


Fig. 14. Difference in SIC RMSE between CNN-LSTM and GraphSIFNet-Att by forecast month and lead time. Positive values (red) indicate GraphSIFNet-Att has lower error; negative values (blue) indicate CNN-LSTM performs better. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

defined as a sea ice presence probability less than and greater than 50%, respectively. For each port, we take the mean pixel value of a 3×3 window around the nearest pixel to the port locations.

Fig. 11 displays the predicted dates of freeze-up/breakup at the three ports with 30 and 60 days of lead time compared to the dates for the validation and test years along with the mean absolute error. Predicted dates falling within 7 days of those from validation and test are considered correct, visualized by the pink shaded area (Asadi et al., 2022). The 30-day forecasted break-up and freeze-up dates for Churchill are noticeably inferior to the other two ports, likely due to challenges presented by the latent heat polynya in the Northwest of Hudson Bay. Break-up predictions at Inukjuak also pose a challenge for the model. Quaqtaq sees the most successful predictions, with all freeze-up dates falling within 7 days of the test and validation data.

In Figs. 12 and 13, the break-up and freeze-up accuracies are shown spatially for the entire region. These accuracies are calculated as the proportion of years with predicted break-up or freeze-up dates within 7 days of that from the validation and test data. These are compared to predictions made using the climate normals. The model performs equally or better than climatology for most of the region in predicting break-up dates at both 30-days and 60-days of lead time, with stronger performance during break-up, consistent with other results shown. However, there is a strong pattern in the freeze-up maps where the model performs worse than climatology in the western half of the bay but still outperforms climatology in the eastern half and in Hudson Strait. The poor performance during freeze-up in the Western part of Hudson Bay was also noted in Asadi et al. (2022). There is a large polynya in this region and there may not be a strong enough signal in the atmospheric conditions to allow the model to accurately predict the heat exchange between the ocean and atmosphere.

5.5. Comparison with CNN-LSTM

To evaluate the benefit of using a graph-based model over more conventional gridded approaches, we compare GraphSIFNet-Att with a CNN-LSTM baseline model trained on the same input data for the SIC forecasting task, and evaluated against the same test period. The CNN-LSTM model uses the same sequence-to-sequence architecture, replacing the graph convolutions with 3×3 spatial convolutions applied to a fixed 2D grid over Hudson Bay. Fig. 14 shows the SIC RMSE difference between the CNN-LSTM and GraphSIFNet-Att models, disaggregated by forecast launch month and lead time. Blue indicates conditions under which CNN-LSTM outperforms GraphSIFNet-Att; red indicates lower error for GraphSIFNet-Att. GraphSIFNet-Att achieves consistently lower error across most launch months and lead times, with the exception of some mid- to long-range forecasts launched in June, August, and November. The improvements are most pronounced during transitional phases (break-up and freeze-up), where spatial heterogeneity increases and the ability to represent directed relationships becomes more important. In these regimes, the use of directed graph relationships likely helps the model capture localized dynamics.

5.6. Comparison with subseasonal-to-seasonal forecasts

The subseasonal-to-seasonal (S2S) system by ECMWF (Vitart and Robertson, 2018) was used to further assess the model predictions. This system uses a dynamic–thermodynamic sea ice model coupled to an ocean model (NEMO-LIM2 Nucleus for European Modeling of the Ocean Louvain-la-Neuve sea ice model version 2), which are coupled to ECMWF's Integrated Forecasting System (IFS). The study by Day et al. (2022) compared the SIC and ice edge with data from passive microwave sensors as well as MODIS (Moderate Resolution Imaging Spectroradiometer) and found the coupling improved predictions of the atmospheric boundary layer, with some improvement needed for the ice edge at some locations (there are not detailed analyses for Hudson Bay). For the present comparison we used sea ice states from the GraphSIFNet-Att model, since it was the best performing of the models. Since the S2S systems provides SIC forecasts from 0 to 45 days, we validated the model outputs over the years of 2016–2019 at 15, 30 and 45 days, in a similar manner as described in Asadi et al. (2022). It can be seen in Fig. 15 that GraphSIFNet-Att model has lower SIC RMS difference (RMSD) with the test data than the S2S data for all months at 15 days. At 30 and 45 days, the GraphsSIFNet-Att model continues with lower RMSD from January-August, with comparable performance from September–December. These results are consistent, but improved, in comparison to those from Asadi et al. (2022). In that study a related deep learning model performed better during break-up than freeze-up when sea ice presence was compared against a thresholded SIC from the ECMWF S2S system.

5.7. Comparison with passive microwave sea ice concentration

The GLORYS system assimilates daily SIC from CERSAT (Ezraty et al., 2007), which are used here to further assess the model predictions. A comparison of the GLORYS forecasts with the CERSAT SIC as well as two other passive microwave SIC products (Lellouche et al., 2021) found the Arctic sea ice extent from CERSAT was lower than the others, indicating an underestimate of SIC. The CERSAT SIC is derived from the 85-GHz channels on the Special Sensor Microwave Imager (SSM/I) using the ASI (Artist Sea Ice) algorithm. A comparison between SIC forecasts for 0–30, 31–60, and 61–90 days and the monthly averaged SIC from CERSAT is shown in Fig. 16 for both the GraphSIFNet-Att model and the CNN-LSTM. For the first 30 days the GraphSIFNet-Att model is much closer to the observations than the CNN-LSTM, with the results more comparable beyond 30 days. An inspection of the SIC maps for these models and the observations indicated a sharper ice/water boundary for the CERSAT SIC during freeze up, likely due to the weather filters used when SIC is derived from high frequency passive microwave channels, similar results have been found in other intercomparisons (Radhakrishnan et al., 2021). Note the results from the two deep learning approaches are different from those shown in Section 5.5 because here the comparison is with observational data, whereas in Section 5.5 the comparison is with GLORYS data.

6. Conclusion

The study presented in this paper demonstrated the effectiveness of using a GNN-based spatiotemporal forecasting model for predicting daily sea ice concentration and sea ice presence in Hudson Bay over a 90-day time horizon. To demonstrate the ability of GNNs to handle

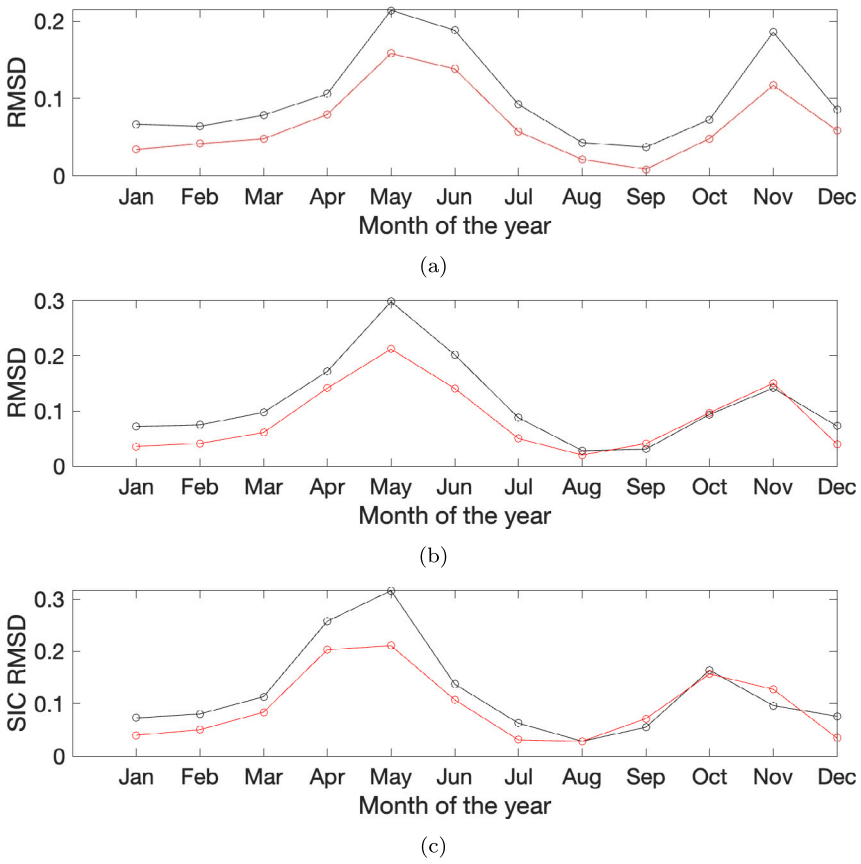


Fig. 15. RMS difference (RMSD) of SIC from the ECMWF S2S system (black) and the GraphSIFNet-Att model (red) at (a) 15 days (b) 30 days and (c) 45 days. Both differences are taken with respect to the test data (from GLORYS). Results are averaged over 2016–2019. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

spatially irregular meshes, models were trained on both a uniform regular mesh and an irregular mesh with higher resolution near shorelines. The proposed model uses an attention-based transformer spatial convolution to learn spatial features from the input, which was shown to have improved performance compared to the more basic graph convolutional network. The attention-based convolution offered the additional benefit of interpretability by highlighting the primary direction and magnitude of information flow in the encoder, which aligned with known physical processes such as the direction of freezing and melting.

Results from this study highlighted the model's skill in predicting sea ice concentration and presence, with particular success noted in short- to medium-term forecasts during the break-up season when compared to a linear combination of persistence and climatology as a statistical baseline. Comparisons with the ECMWF S2S SIC also showed the model performed better during break-up at 15, 30 and 45 days. In general the model performed as well or better on the irregular mesh as on the regular mesh, with the exception of some difficulty capturing the initial freeze-up in the Northwest region of Hudson Bay as well as the polynya formation at longer lead times. This suggests that improvements could be made in refining the model's sensitivity to complex spatial–temporal features in areas where ice dynamics are highly variable. This could involve more sophisticated positional and spatial encoding, perhaps by projecting the positional, cell size, distance and angle encodings into higher dimensional latent space. It could also involve a different selection of input features.

Several potential avenues for future work exist. To better use the mesh refinement, the underlying mesh could be dynamic in time, evolving as the underlying data changes (e.g. as the ice conditions evolves). For example, one could define a dynamic mesh which has a higher resolution at the ice edge where the ice conditions are known to be more dynamic. As the ice conditions evolve, so too would the underlying mesh. The advantages are two-fold. First, it allows for a reduction in data volume with minimal information loss, contrary to the static mesh used in this work which has information loss where the data has high spatial variance. Second, the dynamic mesh could help the model learn more sophisticated dynamics and is more consistent with physical simulation software. This idea was explored in Pfaff et al. (2021). Another avenue for future work could be a deeper investigation of the adjacency matrix. In this study, edges were placed between any two directly spatially adjacent cells. However, edges could also be placed between distant cells thereby widening the receptive field without adding convolutions. This could be investigated by transforming the adjacency matrix into a learnable matrix optimized during training. Although attention-based convolutions already permit re-weighting of messages across a fixed set of edges, learning the adjacency matrix would allow the graph structure itself to be optimized, enabling the model to discover connections beyond spatial adjacency that may better capture relevant physical or dynamical relationships. Furthermore, node sampling strategies could also be used to reduce training time. Specifically, adaptive sampling techniques could be employed where nodes in dynamic regions, such as the ice edges known for their fluctuating conditions, are sampled with higher frequency compared to the more static areas. Incorporating long-term weather forecasts from third party sources such as the Canadian Global Ice Ocean Prediction System (GIOPS) could also be beneficial, particularly in forecasting freeze-up. Lastly, multi-resolution modeling either through an ensemble of models operating over meshes of different resolution or using multiple meshes of varying resolutions within a single model could be explored. This may help the model better capture both large-scale and small-scale phenomena.

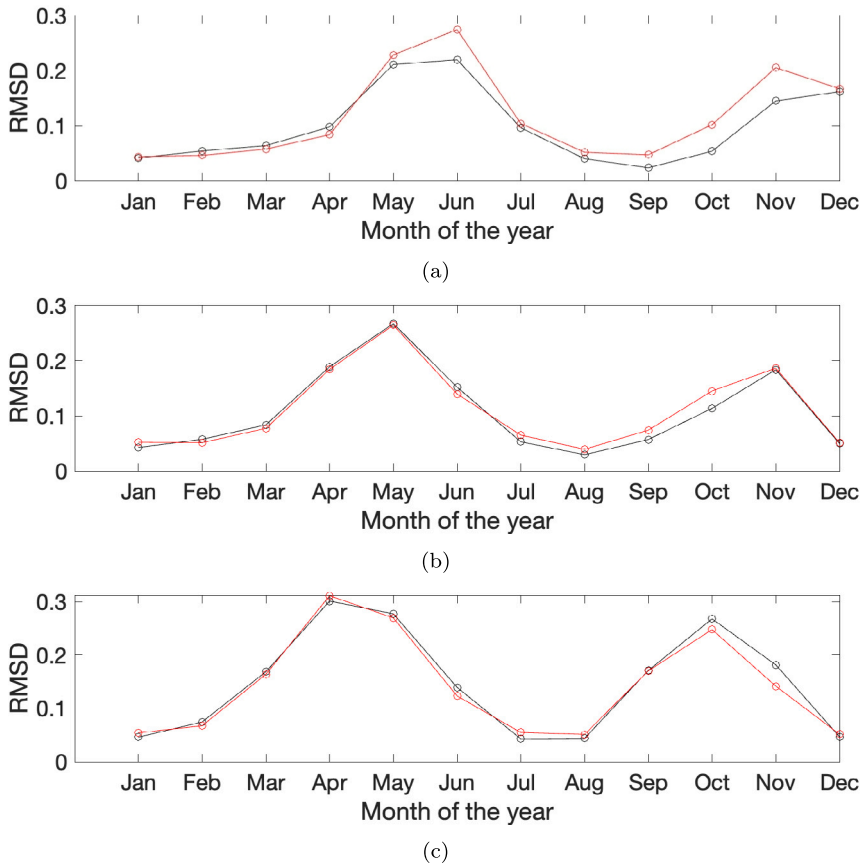


Fig. 16. RMS difference (RMSD) of sea ice concentration passive microwave data (CERSAT) and the GraphSIFNet-Att model (black) and CNN-LSTM (red) over (a) 0–30 days (b) 31–60 days and (c) 61–90 days. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

CRediT authorship contribution statement

Zacharie Gousseau: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Philippe Lamontagne:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Conceptualization. **Mohammad Sina Jahangir:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **K. Andrea Scott:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization.

Code and data availability

ERA5 atmospheric reanalysis data (Hersbach et al., 2020) are available at doi.org/10.24381/cds.adbb2d47, GLORYS12 ocean reanalysis data (Lellouche et al., 2021) are available at doi.org/10.48670/moi-00021, and the S2S forecast data are available at apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/. The model source code can be downloaded from the repository website at doi.org/10.5281/zenodo.15571799.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge funding from the National Research Council of Canada through the AI4Logistics and Ocean Programs via AI4L-121-1 and computing resources provided by the Digital Resource Alliance of Canada.

Data availability

ERA5 atmospheric reanalysis data Hersbach et al. (2020) are available at doi.org/10.24381/cds.adbb2d47, GLORYS12 ocean reanalysis data Lellouche et al. (2021) are available at doi.org/10.48670/moi-00021, and the S2S forecast data are available at apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/. The model source code can be downloaded from the repository website at doi.org/10.5281/zenodo.15571799.

References

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., Shuckburgh, E., 2021. Seasonal Arctic sea ice forecasting with probabilistic deep learning. Nat. Commun. 12 (1), http://dx.doi.org/10.1038/s41467-021-25257-4.

Anselin, L., 1988. A typology of spatial econometric models. In: Spatial Econometrics: Methods and Models. In: Studies in Operational Regional Science, Springer Netherlands, Dordrecht, pp. 32–40. http://dx.doi.org/10.1007/978-94-015-7799-1_4.

Asadi, N., Lamontagne, P., King, M., Richard, M., Scott, K.A., 2022. Probabilistic spatiotemporal seasonal sea ice presence forecasting using sequence-to-sequence learning and ERA5 data in the Hudson Bay region. Cryosphere 16 (9), 3753–3773. http://dx.doi.org/10.5194/tc-16-3753-2022.

Ba, J.L., Kiros, J.R., Hinton, G.E., 2016. Layer normalization. arXiv:1607.06450 [cs, stat].

Cavalieri, D.J., Parkinson, C.L., 2012. Arctic sea ice variability and trends, 1979–2010. Cryosphere 6, 881–889. http://dx.doi.org/10.5194/tc-6-881-2012.

Chevallier, M., Smith, G., Dupont, F., et al., 2016. Intercomparison of the Arctic sea ice cover in global ocean–sea ice reanalyses from the ORA-IP project.

Chi, J., Kim, H.-c., 2017. Prediction of Arctic Sea Ice concentration using a fully data driven deep neural network. Remote. Sens. 9 (12), 1305. http://dx.doi.org/10.3390/rs9121305, URL: https://www.mdpi.com/2072-4292/9/12/1305, Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

Day, J.J., Keeley, S., Arduini, G., Magnusson, L., Mogensen, K., Rodwell, M., Sandu, I., Tietsche, S., 2022. Benefits and challenges of dynamic sea ice for weather forecasts. Weather. Clim. Dyn. 3 (3), 713–731. http://dx.doi.org/10.5194/wcd-3-713-2022.

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., K˚allberg, 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., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. http://dx.doi.org/10.1002/qj.828.

Ezraty, R., Girard-Ardhuin, F., Piolle, J.F., Kaleschke, L., Heygster, G., 2007. Arctic and antarctic sea ice concentration and Arctic Sea Ice drift estimated from special sensor microwave data.

Fey, M., Lenssen, J.E., 2019. Fast graph representation learning with PyTorch geometric. http://dx.doi.org/10.48550/arXiv.1903.02428, arXiv:1903.02428 [cs, stat].

Gers, F., Schraudolph, N., Schmidhuber, J., 2002. Learning precise timing with LSTM recurrent networks. J. Mach. Learn. Res. 3, 115–143. http://dx.doi.org/10.1162/153244303768966139.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, Las Vegas, NV, USA, pp. 770–778. http://dx.doi.org/10.1109/CVPR.2016.90.

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., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., 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., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. http://dx.doi.org/10.1002/qj.3803.

Ivanov, V.V., Alexeev, V.A., Repina, I., Koldunov, N.V., Smirnov, A., 2012. Tracing atlantic water signature in the Arctic Sea Ice cover East of Svalbard. Adv. Meteorol. 2012, e201818. http://dx.doi.org/10.1155/2012/201818.

Joshi, A., Deshmukh, P.K., Lohokare, J., 2022. Comparative analysis of Vanilla LSTM and Peephole LSTM for stock market price prediction. In: 2022 International Conference on Computing, Communication, Security and Intelligent Systems. IC3SIS, pp. 1–6. http://dx.doi.org/10.1109/IC3SIS54991.2022.9885528.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82 (1), 35–45. http://dx.doi.org/10.1115/1.3662552.

Keisler, R., 2022. Forecasting global weather with graph neural networks. Number: arXiv:2202.07575 arXiv:2202.07575 [physics].

Kim, J., Kim, K., Cho, J., Kang, Y.Q., Yoon, H.-J., Lee, Y.-W., 2019. Satellite-based prediction of Arctic Sea Ice concentration using a deep neural network with multi-model ensemble. Remote. Sens. 11 (1), 19. http://dx.doi.org/10.3390/rs11010019, Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

Kipf, T.N., Welling, M., 2017. Semi-supervised classification with graph convolutional networks. http://dx.doi.org/10.48550/arXiv.1609.02907.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., Ravuri, S., Ewalds, T., Alet, F., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Stott, J., Vinyals, O., Mohamed, S., Battaglia, P., 2022. GraphCast: Learning skillful medium-range global weather forecasting. arXiv:2212.12794 [physics].

Lellouche, J.-M., Eric, G., Romain, B.-B., Gilles, G., Angélique, M., Marie, D., Clément, B., Mathieu, H., Olivier, L.G., Charly, R., Tony, C., Charles-Emmanuel, T., Florent, G., Giovanni, R., Mounir, B., Yann, D., Pierre-Yves, L.T., 2021. The copernicus global 1/12° Oceanic and Sea Ice GLORYS12 reanalysis. Front. Earth Sci. 9, URL: https://www.frontiersin.org/articles/10.3389/feart.2021.698876.

Liu, Y., Bogaardt, L., Attema, J., Hazeleger, W., 2021. Extended range Arctic Sea Ice forecast with convolutional long-short term memory networks. Mon. Weather Rev. 149, http://dx.doi.org/10.1175/MWR-D-20-0113.1.

Liu, Q., Wang, Y., Zhang, R., Yan, H., 2024. Incorporating physical constraints in a deep learning framework for short-term daily prediction of sea ice concentration. Appl. Ocean Res. 148, 104007. http://dx.doi.org/10.1016/j.apor.2024.104007.

Madec, G., et al., 2016. NEMO ocean engine. URL: https://www.nemo-ocean.eu/doc/ (Accessed November 2024).

Meier, W.N., Hovelsrud, G.K., van Oort, B.E., Key, J.R., Kovacs, K.M., Michel, C., Haas, C., Granskog, M.A., Gerland, S., Perovich, D.K., Makshtas, A., Reist, J.D., 2014. Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity. Rev. Geophys. 52 (3), 185–217. http://dx.doi.org/10.1002/2013RG000431.

Moon, T.A., Overeem, I., Druckenmiller, M., Holland, M., Huntington, H., Kling, G., Lovecraft, A.L., Miller, G., Scambos, T., Schädel, C., Schuur, E.A.G., Trochim, E., Wiese, F., Williams, D., Wong, G., 2019. The expanding footprint of rapid arctic change. Earth's Futur. 7 (3), 212–218. http://dx.doi.org/10.1029/2018EF001088.

Mudryk, L., Dawson, J., Howell, S., Derksen, C., Zagon, T., Brady, M., 2021. Impact of 1, 2 and 4 °C of global warming on ship navigation in the Canadian Arctic. Nat. Clim. Chang. 11, 1–7. http://dx.doi.org/10.1038/s41558-021-01087-6.

Oprea, S., Martinez-Gonzalez, P., Garcia-Garcia, A., Castro-Vargas, J.A., Orts-Escolano, S., Garcia-Rodriguez, J., Argyros, A., 2022. A review on deep learning techniques for video prediction. IEEE Trans. Pattern Anal. Mach. Intell. 44 (6), 2806–2826. http://dx.doi.org/10.1109/TPAMI.2020.3045007.

Petty, A.A., Schröder, D., Stroeve, J.C., Markus, T., Miller, J., Kurtz, N.T., Feltham, D.L., Flocco, D., 2017. Skillful spring forecasts of September Arctic sea ice extent using passive microwave sea ice observations. Earth's Futur. 5 (2), 254–263. http://dx.doi.org/10.1002/2016EF000495.

Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A., Battaglia, P.W., 2021. Learning mesh-based simulation with graph networks. URL: http://arxiv.org/abs/2010.03409.

Pfeifer, P.E., Deutsch, S.J., 1980. A STARIMA model-building procedure with application to description and regional forecasting. Trans. Inst. Br. Geogr. 5 (3), 330.

Post, E., Bhatt, U.S., Bitz, C.M., Brodie, J.F., Fulton, T.L., Hebblewhite, M., Kerby, J., Kutz, S.J., Stirling, I., Walker, D.A., 2013. Ecological consequences of sea-ice decline. Sci. 341 (6145), 519–524. http://dx.doi.org/10.1126/science.1235225.

Qu, Y., Soleymani, A., Sudom, D., Scott, K.A., 2024. Learnable weight graph neural network for river ice classification. Proc. 110 (1).

Radhakrishnan, K., Scott, K.A., Clausi, D.A., 2021. Sea ice concentration estimation: Using passive microwave and SAR Data With a U-Net and curriculum learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 14, 5339–5351. http://dx.doi.org/10.1109/JSTARS.2021.3076109.

Rubanova, Y., Sanchez-Gonzalez, A., Pfaff, T., Battaglia, P., 2022. Constraint-based graph network simulator. In: Proceedings of the 39th International Conference on Machine Learning. PMLR, (ISSN: 2640-3498) pp. 18844–18870, URL: https://proceedings.mlr.press/v162/rubanova22a.html.

Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J., Battaglia, P., 2020. Learning to simulate complex physics with graph networks. In: Proceedings of the 37th International Conference on Machine Learning. PMLR, (ISSN: 2640-3498) pp. 8459–8468, URL: https://proceedings.mlr.press/v119/sanchez-gonzalez20a.html.

Seo, Y., Defferrard, M., Vandergheynst, P., Bresson, X., 2018. Structured sequence modeling with graph convolutional recurrent networks. In: Cheng, L., Leung, A.C.S., Ozawa, S. (Eds.), Neural Information Processing. In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 362–373. http://dx.doi.org/10.1007/978-3-030-04167-0_33.

Serreze, M.C., Meier, W.N., 2019. The Arctic's sea ice cover: trends, variability, predictability, and comparisons to the Antarctic. Ann. New York Acad. Sci. 1436 (1), 36–53. http://dx.doi.org/10.1111/nyas.13856.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., Woo, W.-c., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Vol. 28, Curran Associates, Inc..

Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., Sun, Y., 2021. Masked label prediction: Unified message passing model for semi-supervised classification. In: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, Montreal, Canada, pp. 1548–1554. http://dx.doi.org/10.24963/ijcai.2021/214.

Sims, C.A., 1980. Macroeconomics and reality. Econom. 48 (1), 1–48. http://dx.doi.org/10.2307/1912017, Publisher: [Wiley, Econometric Society].

Srivastava, N., Mansimov, E., Salakhudinov, R., 2015. Unsupervised learning of video representations using LSTMs. In: Proceedings of the 32nd International Conference on Machine Learning. PMLR, pp. 843–852, ISSN: 1938-7228.

Stammerjohn, S.E., Drinkwater, M.R., Smith, R.C., Liu, X., 2003. Ice-atmosphere interactions during sea-ice advance and retreat in the western Antarctic Peninsula region. J. Geophys. Res.: Ocean. 108 (C10), http://dx.doi.org/10.1029/2002JC001543.

Stroeve, J., Notz, D., 2018. Changing state of Arctic sea ice across all seasons. Environ. Res. Lett. 13 (10), 103001. http://dx.doi.org/10.1088/1748-9326/aade56, Publisher: IOP Publishing.

Stroeve, J.C., Serreze, M.C., Holland, M.M., Kay, J.E., Malanik, J., Barrett, A.P., 2012. The Arctic's rapidly shrinking sea ice cover: a research synthesis. Clim. Change 110 (3), 1005–1027. http://dx.doi.org/10.1007/s10584-011-0101-1.

Vihma, T., 2014. Effects of Arctic Sea Ice decline on weather and climate: A review. Surv. Geophys. 35 (5), 1175–1214. http://dx.doi.org/10.1007/s10712-014-9284-0.

Vitart, F., Robertson, A.W., 2018. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. Nat. Clim. Atmospheric Sci. 1 (1), 1–7. http://dx.doi.org/10.1038/s41612-018-0013-0.

Wang, C., Graham, R.M., Wang, K., Gerland, S., Granskog, M.A., 2019. Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: effects on sea ice thermodynamics and evolution. Cryosphere 13 (6), 1661–1679. http://dx.doi.org/10.5194/tc-13-1661-2019, Publisher: Copernicus GmbH.

Wu, Z., Jain, P., Wright, M.A., Mirhoseini, A., Gonzalez, J.E., Stoica, I., 2022. Representing long-range context for graph neural networks with global attention. arXiv:2201.08821 [cs].

Wu, D., Lang, X., Zhang, D., Eriksson, L., Mao, W., 2021. A statistical Arima model to predict arctic environment for NSR shipping. In: Proceedings of the 40th International Conference on Ocean, Offshore and Arctic Engineering.

Yao, T., Tang, C.L., Peterson, I.K., 2000. Modeling the seasonal variation of sea ice in the Labrador Sea with a coupled multicategory ice model and the Princeton ocean model. J. Geophys. Res.: Ocean. 105 (C1), 1153–1165. http://dx.doi.org/10.1029/1999JC900264.

Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. CoRR abs/1511.07122.

Zhang, J., Rothrock, D., 2003. Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. Mon. Weather Rev. 131, http://dx.doi.org/10.1175/1520-0493(2003).

Zheng, Q., Wang, R., Han, G., Li, W., Wang, X., Shao, Q., Wu, X., Cao, L., Zhou, G., Hu, S., 2024. A spatiotemporal multiscale deep learning model for subseasonal prediction of Arctic Sea Ice. IEEE Trans. Geosci. Remote Sens. 62, 1–22. http://dx.doi.org/10.1109/TGRS.2024.3355238.