Icy Waters: Developing a Test-Suite to Benchmark Sea Ice Concentration Forecasting

Kiernan McGuigan
Systems Design Engineering
University of Waterloo
Waterloo, ON
kmcguiga@uwaterloo.ca

K. Andrea Scott
Mechanical and Mechatronics Engineering
University of Waterloo
Waterloo, ON
ka3scott@uwaterloo.ca

Sirisha Rambhatla
Management Science and Engineering
University of Waterloo
Waterloo, ON
sirisha.rambhatla@uwaterloo.ca

Abstract

Artificial intelligence (AI) for Climate Change efforts have made significant progress in forecasting atmospheric weather patterns and events. Despite this, translating these gains in the context of phenomenon on earth surface, e.g. sea-ice concentration, has been limited because of differences in how these physical processes evolve. Sea ice concentration is one of the key indicators of climate change and is also critical for a number of different applications and indigenous peoples. Consequently, there is an acute need to develop a baseline of a diverse set of modern machine learning techniques within the Arctic. Our work aims to fill this gap, with the goal of both informing current research, as well as pointing out limitations with certain architectures. We achieve this by providing baselines for a number of different convolutional LSTMs, transformer based, and neural operator based machine learning methods.

1 Introduction

While popular imagination might view the Arctic as a cold and barren region, it is home to a complex and important ecosystem [1]. Unfortunately, in recent years the Arctic has experienced a rapid warming due to climate change resulting in unprecedented changes to the sea-ice dynamics [1, 2]. Many remote and native Arctic communities depend upon the ice for transportation, hunting, and general way of life. These communities have developed inter-generational knowledge of sea-ice patterns to thrive in the Arctic climate [3, 4]. However, the changing climate and ice conditions have started to threaten their way of life by making the use of the ice more dangerous [4]. For these reasons short term high resolution ice forecasts are not only critical for day-to-day operations and weather forecasting [5, 6], but for these remote communities as well [7].

Despite the significance of this problem, there is a lack of works focusing on providing strong baselines for different model architectures for sea-ice forecasting. Many sea-ice forecasting works elect to focused on either a singular or small subset of architectures. This is alongside the fact that many works focus on different regions within the Arctic, at different data resolutions, and for different forecast horizons. These differences make comparing results between works incredibly difficult, hence the motivation for a baseline of architectures all evaluated under the same conditions. Besides

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022.

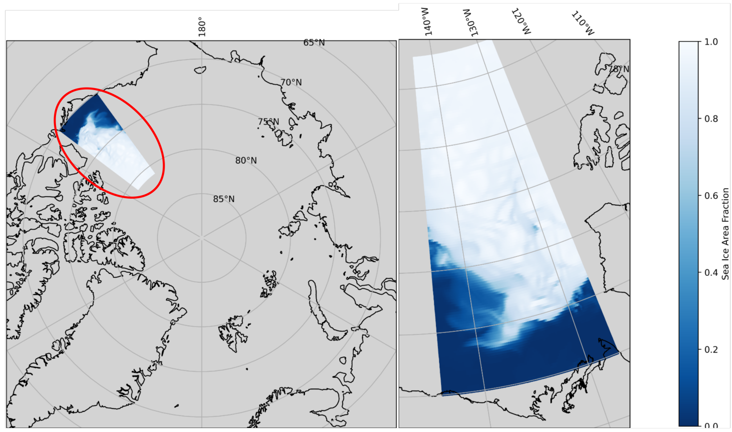

Figure 1: Polar-centric view (left) shows our selected forecast coordinates highlighted with a red rectangle. The domain-centric view (right) displays a sea-ice concentration sample.

providing for a good starting point for performance evaluation, baselines can also help uncover patterns of success and limitations between architectures inspiring future research.

2 Dataset

The data source uses the GLORYS12V1 product [8] and is largely based on the CMEMS system. This study used twenty years of daily historical sea ice concentration data within the Beaufort Sea in the Canadian Arctic (see 1). We focus on the freeze up period of September to December. We use the years 2000 to 2016 for training, 2017 to 2018 for validation, and 2019 to 2020 for testing, which was selected based upon the most recent ice concentration values in the data product. The data spatial resolution is $0.083° \times 0.083°$ and we manually selected a region between Latitude 70.00 and 80.66 and Longitudes -144.00 and -133.33 which results in images of size 128×128 for the models to ingest.

3 Methodology

3.1 Problem Setting

The goal of this task is to produce short range high resolution sea ice forecasts, ingesting five days of ice concentration to predict the following fourteen days, selected based upon existing literature [5, 6]. We could formally define this as $f : X \rightarrow Y$ where $X \in \mathbb{R}^{1\times5\times128\times128}$ and $Y \in \mathbb{R}^{1\times14\times128\times128}$ for our 1 variable (sea ice concentration) and 128×128 spatial grid.

3.2 Previous Works

Past works have focused on linear models [9, 10] to predict sea-ice concentration. While intuitive, restricting models to a linear relationship can be too restrictive for the complex dynamics of sea-ice. Other works have focused on using convolutional neural networks, alone and in a U-Net structure, to capture spatial dynanmics [11, 12, 13, 14]. To model temporal dynamics, convolutional (Conv) LSTMs [15] have been applied to great effect [16, 17, 18, 19, 20].

In recent years, transformers have been gaining popularity for a number of tasks [21, 22, 23] including time series forecasting [24, 25]. Despite this, criticism has been raised as to the efficiency and effectiveness of transformers for time series applications [26, 27]. Existing works have applied transformers to forecast sea-ice extent (total area of sea ice in a region) [28], however this does not produce the ice-charts that are critical to our applications. Other works have applied transformers in a spatial-temporal nature, however these either predict ice presence rather than concentration [19], focus on a very small and specific shipping pathway [29], or work with low resolution data [30].

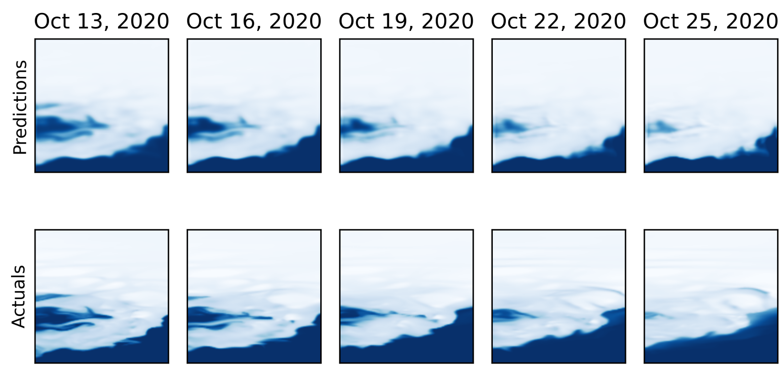

Figure 2: Samples for a forecast generated from our Conv LSTM model between Oct 13th and Oct 25th 2020 (outside training window). We compare the predictions (top) and the actuals (bottom).

A commonality between many of these works is that they focus on only a singular or select subset of architectures, and are being applied to different regions at different data resolutions. This makes comparing performance for insights amongst works incredibly difficult and uninformative.

3.3 Baselines

We present a baseline for a number of common forecasting architectures for the sea-ice domain. The first group are **Conv LSTM models** [15]. We implement both a standard Conv LSTM predicting sea-ice concentration as well as a second one predicting the daily change in sea-ice concentration at each time step (Conv LSTM Residual).

We present three **transformer based models**, each using a different prediction strategy. The first being a decoder only approach (Transformer Decoder) which predicts ice concentration one day ahead with a causal attention mask that prevents data-leakage at training and testing done auto-regressively. The second approach (Transformer Residual) is similar to the decoder-only approach however it trains and tests auto-regressively. The third is an encoder-decoder (Transformer Single Shot) architecture which encodes the historical ice observations before passing them to the decoder mechanism. A combination of self-attention and cross-attention layers allow the decoder tokens to attend to the historicals as well as each other. This approach produces predictions for every token within the forecast horizon simultaneously.

An issue with transformers is the computational complexity of the attention mechanism [31]. We separate spatial and temporal attention into separate layers to limit the number of tokens being attended to at any one time. We also use a sliding window approach [32] for the spatial attention with window sizes of 5x5 or 7x7 to reduce the sequence length of attention.

The final baseline we present comes from the family of **operator based methods**, which aim to parameterize resolution invariant operations [33]. This removes a model's dependency on a specific input resolution. To ensure fair training conditions, we continue to train and test at the same resolution as the other models. The operator selected is the Fourier neural operator (FNO) which learns to parameterize a global convolutional kernel in the Fourier domain [33].

4 Results

To **quantitatively compare** the forecasting performance of the different architectures we report a number of common forecasting metrics described in table 4. We observe that the Conv LSTM models perform best, closely followed by the FNO. It should be noted that the FNO is a very computationally efficient model, especially if trained at lower resolutions. Interestingly, we observe that the transformer architectures struggled compared to other architectures. They all shared a similar lead time to error curve profile to that of the persistence model, displaying a difficulty in capturing the evolving dynamics of the sea-ice.

Table 1: The MAE, MSE, the MAE of predicted versus actual day-to-day ice concentration change (Step MAE), and sea-ice extent error (SIE) measuring the error in total ice area predicted per time step. All metrics are calculated from test set forecasts.

Model	MAE	MSE	Step MAE	SIE
Persistance	7.041×10^{-2}	3.035×10^{-2}	1.778×10^{-2}	7.77×10^{2}
Transformer Decoder	7.098×10^{-2}	3.034×10^{-2}	1.794×10^{-2}	7.01×10^{2}
Transformer Residual	7.016×10^{-2}	2.786×10^{-2}	1.863×10^{-2}	6.36×10^{2}
Transformer Single Shot	6.922×10^{-2}	2.734×10^{-2}	1.838×10^{-2}	6.44×10^{2}
FNO	6.403×10^{-2}	2.195×10^{-2}	1.851×10^{-2}	5.04×10^{2}
Conv LSTM Residual	5.721×10^{-2}	2.231×10^{-2}	1.864×10^{-2}	5.61×10^{2}
Conv LSTM	5.628×10^{-2}	2.129×10^{-2}	1.912×10^{-2}	5.35×10^{2}

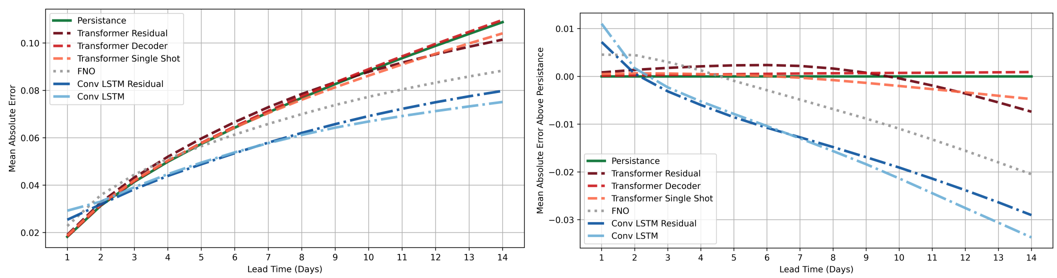

Figure 3: The lead time in days versus the MAE averaged over our test set. Left shows the raw MAE while right shows the MAE above the persistence model (lower is better).

To **qualitatively analyze** our results we present a sample forecast in figure 2 from the Conv LSTM model during mid to late October. This period was selected as it sits in between the periods of ice freeze-up beginning and when the ice has frozen over resulting in limited changes inside the domain. We can observe that the model performs well at forecasting the freeze up, and we specifically observe the accurate forecast of the freeze up on the lower left hand side of the domain.

5 Conclusion

We present a study comparing the performance of a number of modern machine learning approaches for sea-ice forecasting. Our goal is to provide a baseline that will both highlight a number of different forecasting methodologies, but also inspire future research into the sea-ice forecasting domain. We intend to provide open source code upon acceptance in a readable and reproducible format.

With all the attention around transformers, we want to specifically comment on their poor performance under our experiment setup. This is not to claim that transformers are ineffective for sea-ice forecasting, but highlight that some limitations exist that need to be addressed for their proper application. One important factor is data availability and model scaling for success [34, 26], something that can be difficult to achieve with rapidly changing conditions, making historical Arctic patterns less predictive of future ones.

6 Acknowledgements

This study has been conducted using E.U. Copernicus Marine Service Information; Global Ocean Physics Reanalysis. E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS). DOI: 10.48670/moi-00021 (Accessed on 23-07-2024)

References

[1] Toru Hirawake, Masaki Uchida, Hiroto Abe, Irene D. Alabia, Tamotsu Hoshino, Shota Masumoto, Akira S. Mori, Jun Nishioka, Bungo Nishizawa, Atsushi Ooki, Akinori Takahashi, Yukiko Tanabe, Motoaki Tojo, Masaharu Tsuji, Hiromichi Ueno, Hisatomo Waga, Yuuki Y. Watanabe, Atsushi Yamaguchi, and Youhei Yamashita. Response of arctic biodiversity and ecosystem to environmental changes: Findings from the arcs project. *Polar Science*, 27:100533, March 2021.

[2] Roberto. Climate explained: why is the arctic warming faster than other parts of the world?, June 2021.

[3] Jackie Dawson, Natalie Carter, Nicolien van Luijk, Colleen Parker, Melissa Weber, Alison Cook, Kayla Grey, and Jennifer Provencher. Infusing inuit and local knowledge into the low impact shipping corridors: An adaptation to increased shipping activity and climate change in arctic canada. *Environmental Science & Policy*, 105:19–36, March 2020.

[4] Chris Baraniuk. The inuit knowledge vanishing with the ice, October 2021.

[5] Tom Carrieres, Mark Buehner, Jean-François Lemieux, and Leif Pedersen. *Sea ice analysis and forecasting towards an increased reliance on automated prediction systems*. October 2017. Citation Key: book.

[6] F. Dupont, S. Higginson, R. Bourdallé-Badie, Y. Lu, F. Roy, G. C. Smith, J.-F. Lemieux, G. Garric, and F. Davidson. A high-resolution ocean and sea-ice modelling system for the arctic and north atlantic oceans. *Geoscientific Model Development*, 8(5):1577–1594, May 2015.

[7] Rod Boyce. New research benefits arctic sea ice forecasting, Dec 2023.

[8] E.U. Copernicus Marine Service Information (CMEMS). Global Ocean Physics Reanalysis. Marine Data Store (MDS), 2024. DOI: 10.48670/moi-00021 (Accessed on 23 Jul 2024).

[9] Frank Kwasniok. Linear inverse modeling of large-scale atmospheric flow using optimal mode decomposition. September 2022.

[10] Sean Horvath, Julienne Stroeve, Balaji Rajagopalan, and William Kleiber. A bayesian logistic regression for probabilistic forecasts of the minimum september arctic sea ice cover. *Earth and Space Science*, 7(10):e2020EA001176, 2020.

[11] Timofey Grigoryev, Polina Verezemskaya, Mikhail Krinitskiy, Nikita Anikin, Alexander Gavrikov, Ilya Trofimov, Nikita Balabin, Aleksei Shpilman, Andrei Eremchenko, Sergey Gulev, Evgeny Burnaev, and Vladimir Vanovskiy. Data-driven short-term daily operational sea ice regional forecasting. *Remote Sensing*, 14(22):5837, November 2022.

[12] Tobias Sebastian Finn, Charlotte Durand, Alban Farchi, Marc Bocquet, Yumeng Chen, Alberto Carrassi, and Véronique Dansereau. Deep learning subgrid-scale parametrisations for short-term forecasting of sea-ice dynamics with a maxwell elasto-brittle rheology. *The Cryosphere*, 17(7):2965–2991, July 2023.

[13] Sahara Ali and Jianwu Wang. Mt-icenet – a spatial and multi-temporal deep learning model for arctic sea ice forecasting. In *2022 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*, December 2022. arXiv:2308.04511 [physics].

[14] Tom R. Andersson, J. Scott Hosking, María Pérez-Ortiz, Brooks Paige, Andrew Elliott, Chris Russell, Stephen Law, Daniel C. Jones, Jeremy Wilkinson, Tony Phillips, James Byrne, Steffen Tietsche, Beena Balan Sarojini, Eduardo Blanchard-Wrigglesworth, Yevgeny Aksenov, Rod Downie, and Emily Shuckburgh. Seasonal arctic sea ice forecasting with probabilistic deep learning. *Nature Communications*, 12(1):5124, August 2021.

[15] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting.

[16] Eliot Kim, Peter Kruse, Skylar Lama, Jamal Bourne, Michael Hu, Sahara Ali, Yiyi Huang, and Jianwu Wang. Multi-task deep learning based spatiotemporal arctic sea ice forecasting. In *2021 IEEE International Conference on Big Data (Big Data)*, page 1847–1857, Orlando, FL, USA, December 2021. IEEE.

[17] Quanhong Liu, Ren Zhang, Yangjun Wang, Hengqian Yan, and Mei Hong. Daily prediction of the arctic sea ice concentration using reanalysis data based on a convolutional lstm network. *Journal of Marine Science and Engineering*, 9(3):330, March 2021.

[18] Nazanin Asadi, Philippe Lamontagne, Matthew King, Martin Richard, and K. Andrea Scott. Probabilistic spatiotemporal seasonal sea ice presence forecasting using sequence-to-sequence learning and era5 data in the hudson bay region. *The Cryosphere*, 16(9):3753–3773, September 2022.

[19] Mary Ruth Keller, Christine Piatko, Mary Versa Clemens-Sewall, Rebecca Eager, Kevin Foster, Christopher Gifford, Derek Rollend, and Jennifer Sleeman. Short-term (seven-day) beaufort sea-ice extent forecasting with deep learning. *Artificial Intelligence for the Earth Systems*, page 1–36, August 2023.

[20] Junhwa Chi, Jihyun Bae, and Young-Joo Kwon. Two-stream convolutional long- and short-term memory model using perceptual loss for sequence-to-sequence arctic sea ice prediction. *Remote Sensing*, 13(1717):3413, January 2021.

[21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[22] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. (arXiv:2010.11929), June 2021. arXiv:2010.11929 [cs].

[23] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J. Fleet. Video diffusion models. (arXiv:2204.03458), June 2022. arXiv:2204.03458 [cs].

[24] Bryan Lim, Sercan Ö. Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, October 2021.

[25] Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. October 2022.

[26] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? (arXiv:2205.13504), August 2022. arXiv:2205.13504 [cs].

[27] Zhijian Xu, Ailing Zeng, and Qiang Xu. Fits: Modeling time series with $10k$ parameters. (arXiv:2307.03756), January 2024. arXiv:2307.03756 [cs].

[28] Bin Mu, Xiaodan Luo, Shijin Yuan, and Xi Liang. Icetft v1.0.0: interpretable long-term prediction of arctic sea ice extent with deep learning. *Geoscientific Model Development*, 16(16):4677–4697, August 2023.

[29] Da Wu, Xiao Lang, Wengang Mao, Di Zhang, Jinfen Zhang, and Rong Liu. Vae based non-autoregressive transformer model for sea ice concentration forecast. June 2022.

[30] Zhuoqing Jiang, Bing Guo, Huihui Zhao, Yangming Jiang, and Yi Sun. Sicformer: A 3d-swin transformer for sea ice concentration prediction. *Journal of Marine Science and Engineering*, 12(88):1424, August 2024.

[31] Feyza Duman Keles, Pruthuvi Mahesakya Wijewardena, and Chinmay Hegde. On the computational complexity of self-attention. (arXiv:2209.04881), September 2022. arXiv:2209.04881 [cs].

[32] Iz Beltagy, Matthew E. Peters, and Arman Cohan. Longformer: The long-document transformer. (arXiv:2004.05150), December 2020. arXiv:2004.05150 [cs].

[33] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. (arXiv:2010.08895), May 2021. arXiv:2010.08895 [cs, math].

[34] Xiaohua Zhai, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer. Scaling vision transformers. (arXiv:2106.04560), June 2022. arXiv:2106.04560 [cs].