STYRELSEN FÖR

**VINTERSJÖFARTSFORSKNING**

WINTER NAVIGATION RESEARCH BOARD

Research Report No 135

Henri Vuollekoski

# FORECASTING SEA ICE IN THE BALTIC SEA USING DEEP LEARNING

Finnish Transport and Communications Agency

Finnish Transport Infrastructure Agency

Finland

Swedish Maritime Administration

Swedish Transport Agency

Sweden

Talvimerenkulun tutkimusraportit — Winter Navigation Research Reports
ISSN 2342-4303
ISBN 978-952-425-021-4

## FOREWORD

In this report no 135, the Winter Navigation Research Board presents the results of WiseIce - Forecasting sea ice in the baltic sea using deep learning. The project goal was to implement deep learning tools for generating short term ice forecasts in the Baltic Sea.

The Winter Navigation Research Board warmly thanks the authors for this report.

Helsinki

April 2026


Ville Häyrynen

Finnish Transport and Communications Agency

Amund Lindberg

Swedish Maritime Administration

Helena Orädd

Finnish Transport Infrastructure Agency

Jonas Gustaffson

Swedish Transport Agency

# Forecasting sea ice in the Baltic Sea using deep learning

Henri Vuollekoski[a,*]

[a]*Finnish Meteorological Institute, P.O. Box 503, FI-00101 Helsinki, Finland*

## Abstract

Practically all transport of goods to and from Finland is marine. Since the sea areas experience seasonal ice cover, accurate and timely forecasts of sea ice conditions are vital. During the past few years there have been several attempts in the deep learning community to forecast sea ice type, concentration, or thickness, in the Arctic and Antarctic oceans, but few studies have focused on the Baltic Sea. In contrast to polar regions, the Baltic Sea benefits from abundant monitoring and observations present and historical, and all sea ice is seasonal with little brine. On the other hand, the complex coastline and archipelago mean that high-resolution forecasts are required. Here, we present a short-term forecast model for sea ice thickness based on the common U-Net architecture. On average, the model appears to perform well without numerical issues at least up to two weeks. Less common phenomena such as shore lead formation remain inadequately reproduced, and would likely benefit from more training data. The model can be run within seconds on a regular laptop, and shows promise for more development.

*Keywords:*
sea ice, deep learning, forecast

## 1. Introduction

While a warming climate is likely to decrease overall sea ice volume in the Baltic Sea, the operation environment for winter-time navigation could actually become more difficult due to higher variability, potentially higher coverage by thinner ice and its ridging, and decreasing trend in ship engine power. Although short-term forecasting is fairly reliable using numerical process models (Haapala, 2000), their usage requires significant computational power. Also, the resolution of operational process models is inadequate for particularly difficult ice-breaking conditions.

Several studies have utilized deep learning to forecast sea ice. To name a few, Andersson et al. (2021) presented a U-Net-based system (Ronneberger et al., 2015) to forecast 6 months of monthly-averaged sea ice concentrations in the Arctic, which they have later extended to daily resolution, and the Antarctic. Liu et al. (2021) developed a model to forecast Arctic sea ice concentration skillfully on weekly to monthly time scales using an unsupervised approach, a convolutional long short-term memory application. More recently, Durand et al. (2024) trained another U-Net-based deep learning model as a surrogate model for the neXtSIM process model, and used it to forecast sea ice thickness in the Arctic from a lead time of 12 h up to a full year. Palerme et al. (2024) likewise used a U-Net-based deep learning model to post-process and improve forecasts from an operational sea ice forecast model, focusing on sea ice concentration with lead times from 1 to 10 days. Perhaps the latest development was shown by Ren et al. (2025), who applied a transformer model to predict Arctic sea ice concentration in seasonal scale.

*henri.vuollekoski@fmi.fi

To our knowledge, none of the previous studies have speficifically focused on the Baltic Sea; perhaps due to its relatively small size and, thus, lesser climatic impact. However, the smaller area should enable higher resolution forecasts, and observations of not only sea ice concentration but its thickness are abundant. In this study, we decided to apply the U-Net architecture (Ronneberger et al., 2015) that has already worked so well in many related studies, and use daily ice charts as the input, with ice thickness as the target. The result is a proof-of-concept type framework that shows promise for future improvements and potential operational use.

## 2. Methods

### *2.1. Input Data*

The input data, used in both training and validation, consists of regional meteorological variables as well as ice charts produced by Finnish Meteorological Institute. More speficifically, we used 10-metre wind speeds, zonal and meridional, from the CERRA sub-daily regional reanalysis (Schimanke et al., 2021), as well as 2 meter air temperature (the CERRA dataset does not provide sea surface temperature). This 5.5 km $\times$ 5.5 km, 3-hour resolution dataset is provided by the European Centre for Medium-Range Weather Forecasts, and its data assimilation system is optimized for the European area with surrounding sea areas. The data was averaged to daily resolution, and bilinearly interpolated to a rectangular $440 \times 260$ grid, which made it simple to feed into the model. Note that in operational use, the reanalysis data can be directly replaced with weather forecasts.

Ice charts are issued daily during ice season. The parameters include sea ice concentration, type and degree of deformation as well as minimum, average and maximum thickness; for some years also sea surface temperature is included. The charts are nowadays in NetCDF format, and produced as polygons fit to images created from satellite data. As such, the represented time is inexact, and should be interpreted as "yesterday's" situation. The polygon-nature of the charts causes some artificial smoothening, as very small features cannot be represented. The full resolution ice charts were downsampled to match the $440 \times 260$ resolution of the CERRA data.

It should be noted that the input data are treated as ground truth, even though there is certainly large uncertainty especially in sea ice thickness. The thickness reported in ice chart represents average level ice, and its minimum and maximum estimates are provided as well. These remain unutilized in the model for now.

Daily ice charts are available since 2007. The training data therefore constituted years 2007-2021, with years 2022-2024 used in validation. CERRA renanalysis data was not yet available for 2025.

### *2.2. Model description and training*

The implemented forecast model is based on the commonly used U-Net architecture (Ronneberger et al., 2015), where the input and output data can be interpreted as multi-channel images. The U-Net is a convolutional neural network consisting of an encoder followed by a decoder, and *skip connections* help convey finer details. Land areas, which actually represent most of the model domain enclosing the Baltic Sea, are masked and dealt with a modified operation called *partial convolution* (Liu et al., 2018; Durand et al., 2024). The model architecture is visualized and explained in Figure 1. Here, the model consists of 1,865,793 parameters.

The model was implemented within the Lux.jl deep learning framework (Pal, 2023) in the Julia programming language, which provides a flexible and efficient development environment. The model was trained in random batches of 10 samples using the Adam (Adaptive Moment

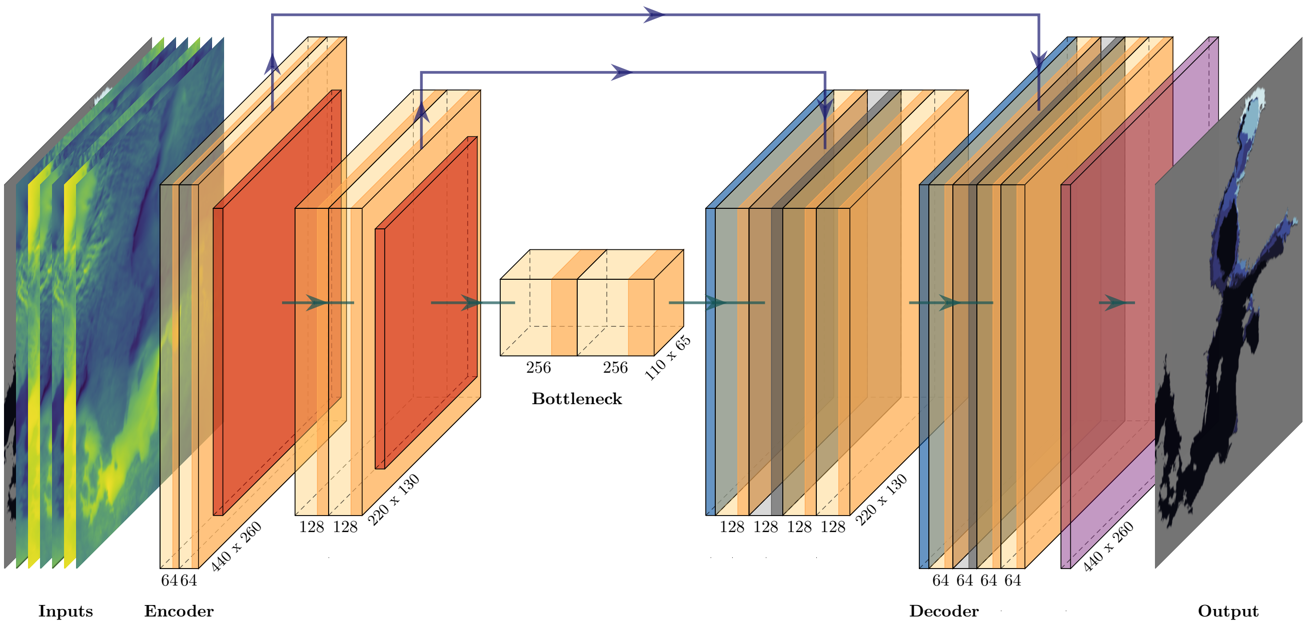


Figure 1. Visualization of model architecture with data flowing to the right. Starting from left, there are 7 input fields representing ice thickness in the previous ice chart, wind speeds as well as air temperature for the previous and current day. The inputs all have a resolution of 440×260, and they are fed into the encoder via a 64-channel partial (3,3)-convolution layer with *mish* activation to introduce non-linearity. Another layer with similar function follows, and *BatchNorm* normalization is applied. The output branches to the *skip connection* bypass between encoder and decoder parts of equal size, and to a (2,2) max pooling layer that effectively halves the spatial resolutions. A similar block follows with 128 channels. The 256-channel *bottleneck* block comprises of just two similar convolutional layers as before. In the decoder part, the first layer in each block is a transpose convolution operation, which doubles the spatial dimensions while halving the number of channels. Its output is concatenated with the feed from skip connection, which doubles the number of channels, but the next partial convolution layer once again halves it. Otherwise the convolution blocks in encoder and decoder parts are similar. The final layer is another partial convolution operation which combines the remaining 64 channels into one that represent the increment in sea ice thickness.

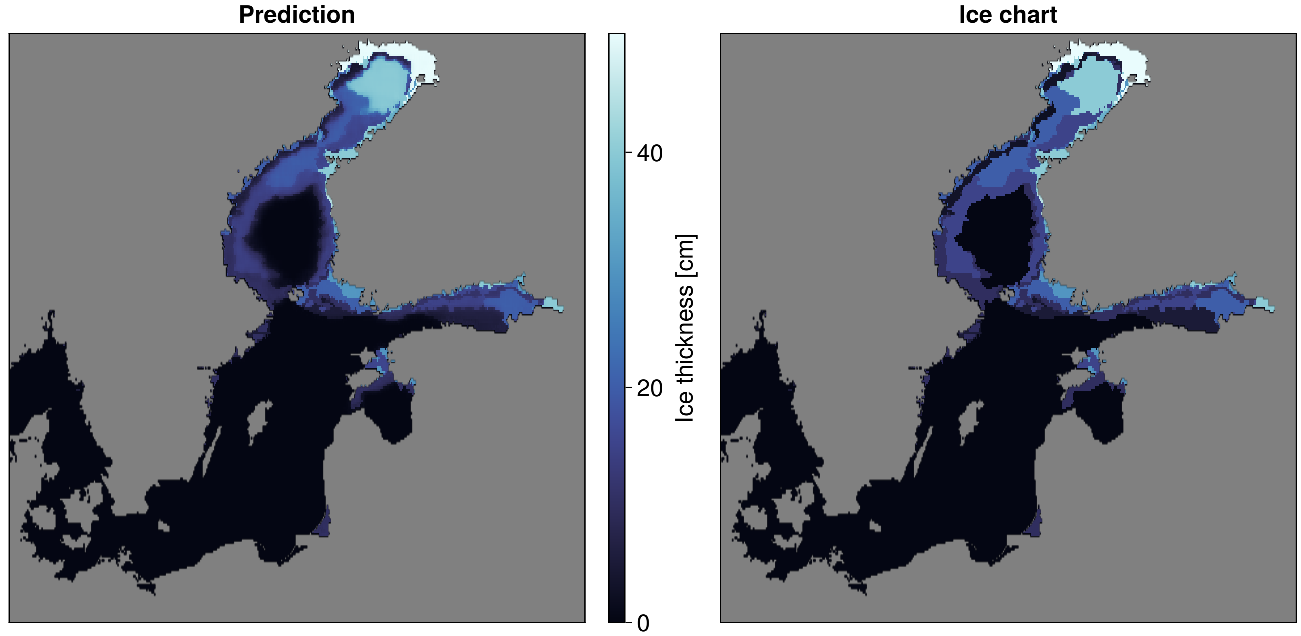


Figure 2. Example of predicted sea ice thickness on 20 February 2024 with lead time of 1 day. The right-hand-side figure shows the actual ice chart.

Estimation, Kingma and Ba (2014)) optimizer, which is a popular modern choice for gradient-based optimization due to its efficiency and self-adapting learning rate. The target of optimization, the loss function, was the typical root-mean-square penalized with squared mean error for more physical relevance (Durand et al., 2024)

$$L(y, \hat{y}) = \frac{1}{N_x \times N_y} \sum_{i,j}^{N_x,N_y} (y_{i,j} - \hat{y}_{i,j})^2 + \lambda \big( \frac{1}{N_x \times N_y} \sum_{i,j}^{N_x,N_y} y_{i,j} - \hat{y}_{i,j} \big)^2 \tag{1}$$

where $y$ is the ground truth matrix (i.e. ice chart), $\hat{y}$ is the prediction matrix, $N_x$ and $N_y$ are the matrix dimensions, and $\lambda$ is a parameter for adjusting the relative importance of terms, here set to 10. This loss function is non-negative and differentiable, making it convenient for gradient-based optimization. The minimum loss in the validation data set was observed after only about 40 iterations, after which the training loss kept decreasing as expected due to over-fitting. Model training took only about one hour using a NVIDIA A100 GPU hosted by CSC. Once trained, and assuming input data loaded in memory, a forecast can be run in about 2 seconds locally on a typical laptop.

## 3. Results and Discussion

The trained model can be used to rapidly forecast the ice thickness field in an ice chart, and it works very well for short-term forecasts, see Figure 2 (a randomly chosen date in the validation data set). By eye, the only relevant difference in the prediction and ice chart is the softness of borders in the prediction — this smearing out of details is typical of AI weather models. However, in the context of ice charts, this feature is perhaps less important, as the exact locations of ice features can be argued to be inexact in daily charts anyhow.

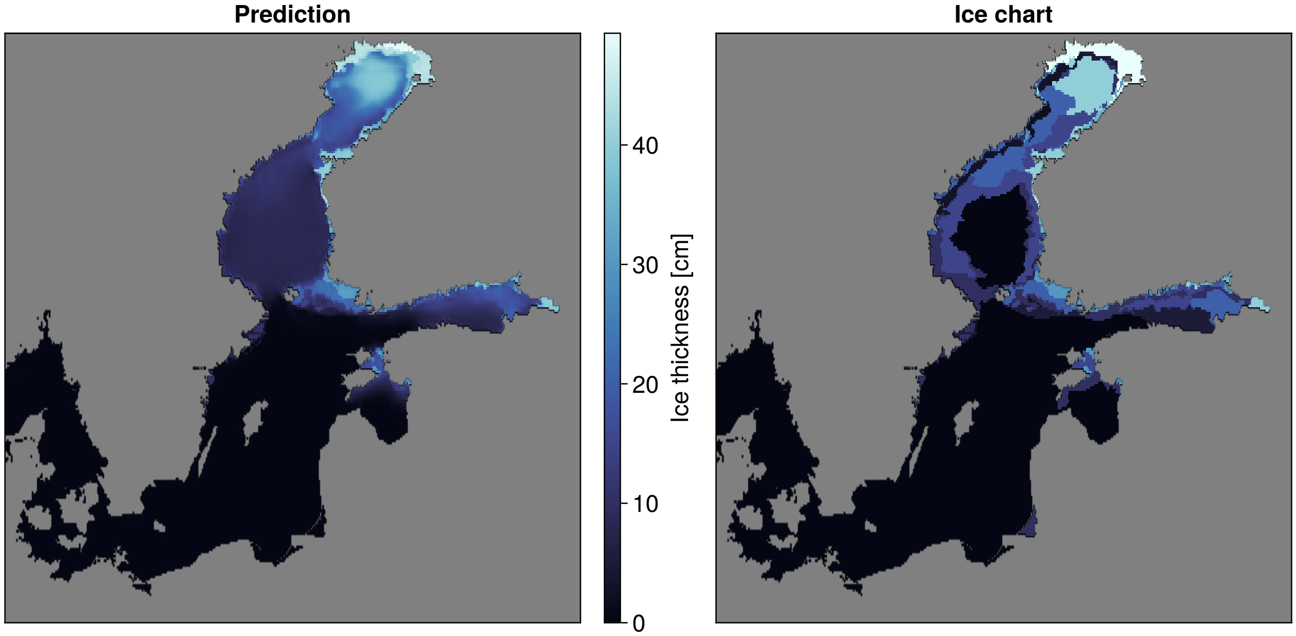


Figure 3. Example of predicted sea ice thickness on 20 February 2024 with lead time of one week. The right-hand-side figure shows the actual ice chart.

The prediction can be iterated arbitrarily far into the future by using the forecast as input ice chart for the next day's prediction (assuming the meteorological input data are available). In practice though, the accuracy of the forecast decreases with increasing lead time, sometimes substantially; see example in Figure 3. Here, the prediction is for the same day as in Figure 2 but made one week in advance. The model fails to predict the formation of a shore lead on Swedish coast, and on the other hand wrongly predicts freezing over of the Bothnian Sea. This week is investigated in more detail in Figure 4, which presents the absolute differences in model prediction and ice chart. Most of the errors in longer forecasts appear to be simply due to accumulating errors. However, both long forecasts display failure to predict the formation of shore lead on the Swedish coast 19 - 20 February, and particularly the longer forecast has inflated the error of level ice thinning on 17 February. However, the AI does sometimes also succeed in predicting shore lead formation, for example on 6 February (not shown here), so its limitations are not as obvious, and it was able to take advantage of the high wind speeds during that day. It is likely that the entire data contains few examples of such shore leads opening, and thus this behaviour cannot be learnt as easily. Clearly, this is a topic for more research. The other obvious issue with the longer forecast, too rapid formation of sea ice in Bothnian Sea, can probably be remedied with regularization, and is also a good topic for more research.

A day-by-day comparison is tedious, and a more objective measure is for example the mean error over the validation data set. Figure 5 compares this loss against two common references: persistence and climatology. A persistence forecast means that the situation is assumed to persist i.e. stay constant over the forecast window, for example one week. Although very simple, a persistence forecast is usually very good short-term. A climatology forecast, on the other hand, means that the situation is assumed to follow a climatic daily mean, here calculated for years 2007-2024. A climatology forecast tends to be good in long time scales. In the figure we see that the mean forecast error for the model beats both persistence and climatology in short-term forecasts up to the tested 14 days.

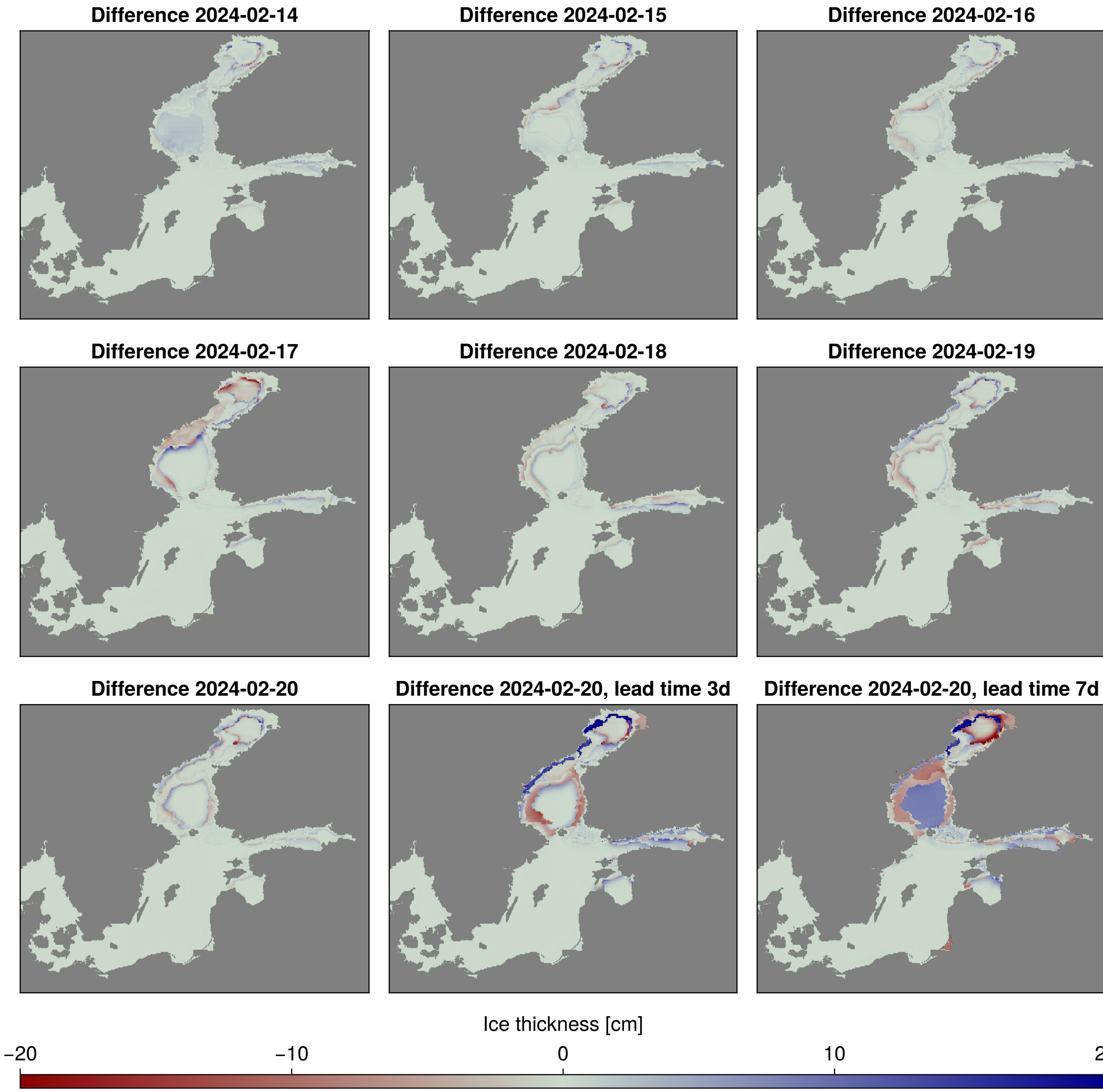


Figure 4. Differences (model - ice chart) in daily forecasts for 14–20 February 2024, and also 3 day and 7 day forecasts for the final day.

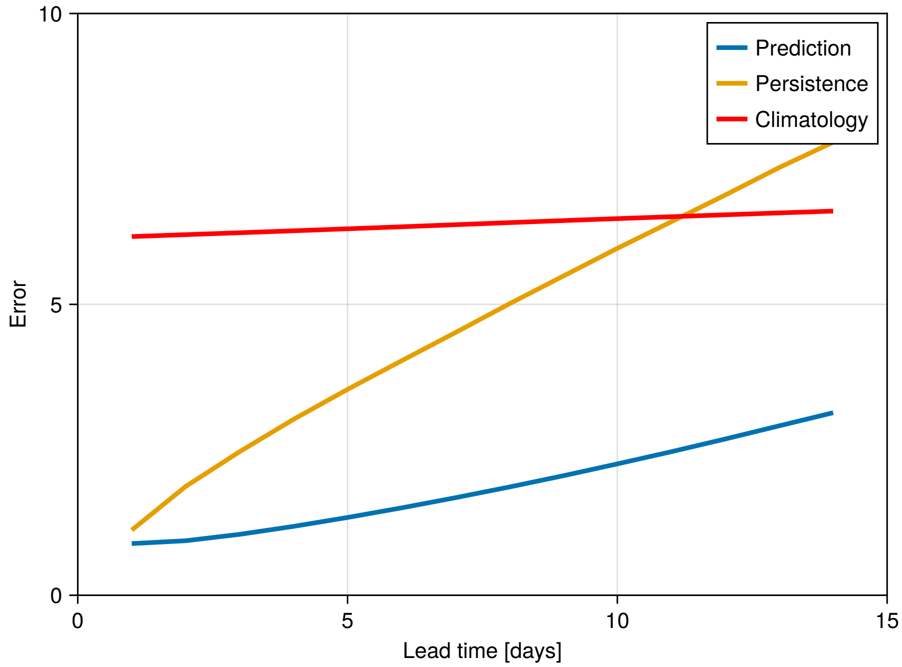


Figure 5. Mean error in predicted sea ice thickness vs. lead time in days. For comparison, persistence and climatology are also shown. (The error in climatology varies due to the days used in calculation varying by lead time. In general, error in climatology should be constant.)

Figure 6 presents the errors of each individual forecast run for 14 days. While most of the runs seem to perform well against the reference forecasts, there are some runs that are clearly off, but at least within the 14 day window no prediction seems to run into numerical errors that would cause the error to explode exponentially. The large errors mostly occurred late in ice seasons and can occur for example when the model fails to capture movement of ice: as the ice field relocates, the pixel-wise loss function (1) penalizes for existing ice as well as non-existing ice. Spotting these automatically is a good topic for research.

Pinpointing the very highest error did not reveal anything particularly seriously wrong with the long 14-day prediction; much of the fine details are simply softened out, see Figure 7.

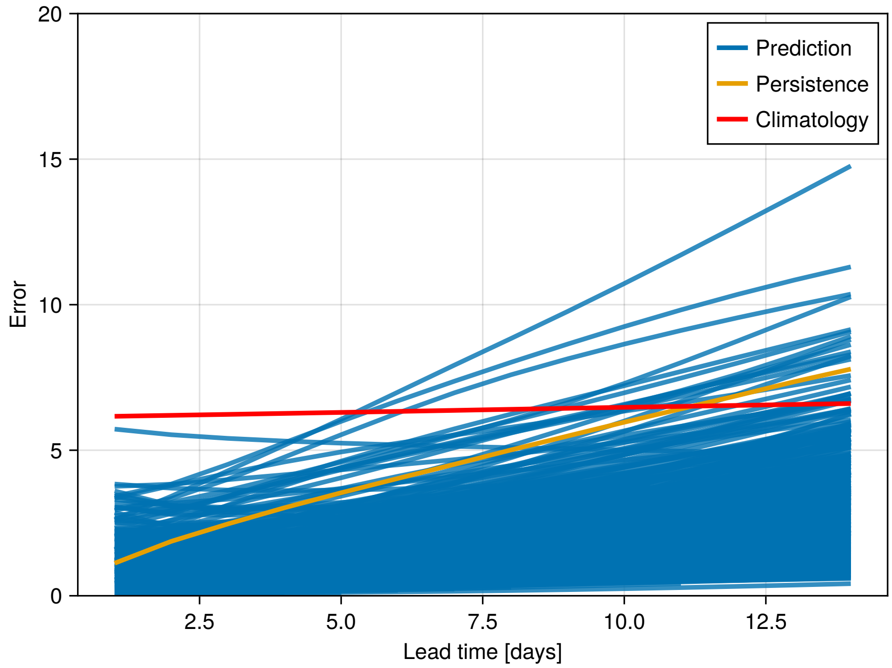


Figure 6. Error in predicted sea ice thickness for all days in the validation data set.

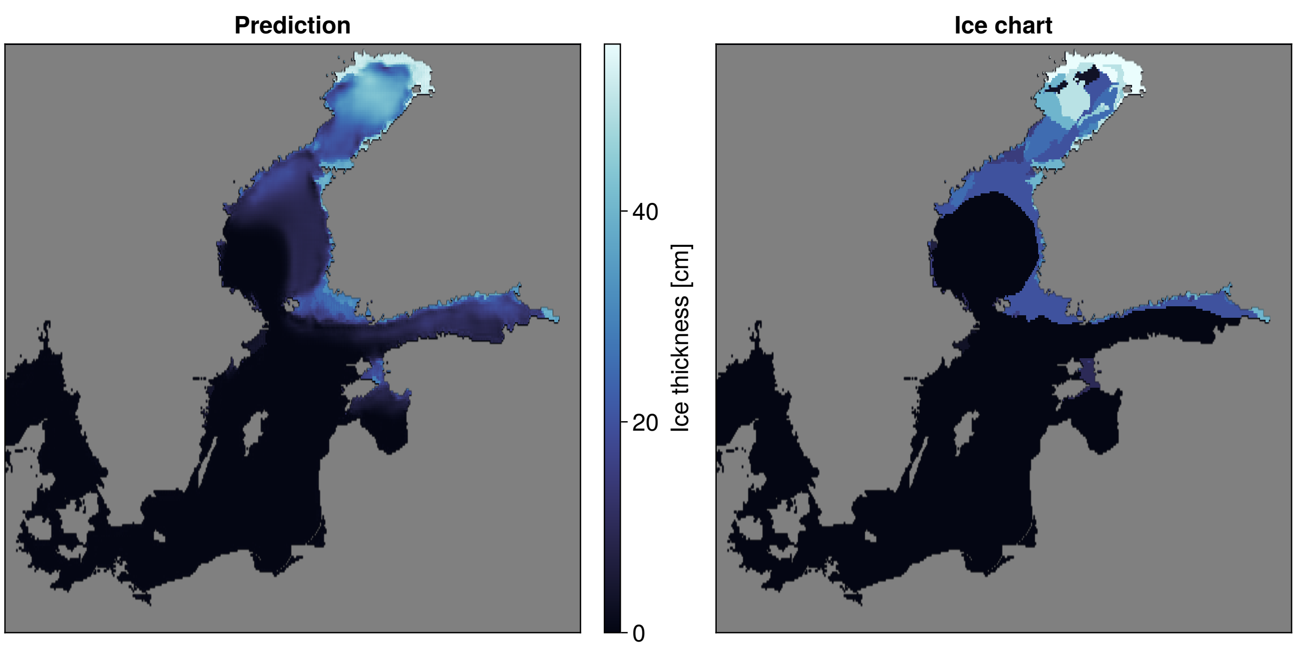


Figure 7. The very highest prediction error in the data set: sea ice thickness on 3 April 2024 with lead time of two weeks. The right-hand-side figure shows the actual ice chart.

## 4. Conclusions

An AI model was developed for predicting sea ice thickness in the Baltic Sea. The U-Net-based model performs well for short-term forecasts, outperforming persistence and climatology up to a lead time of at least 14 days, and allows for generating a forecast in a matter of just a few seconds. However, challenges remain, particularly in predicting rapid changes such as shore lead formation and particularly detecting when the prediction is off.

Future work should focus on incorporating physical constraints to address unrealistic predictions, improving the representation of events such as shore leads with few examples in training data, and exploring additional input features such as ice deformation or historical trends. Regularization techniques and ensemble methods could also be investigated to enhance model robustness.

Overall, this proof-of-concept framework highlights the promise of AI-driven approaches for operational sea ice forecasting, with potential applications in navigation safety and resource optimization in the Baltic Sea.

## Acknowledgements

The authors wish to acknowledge funding from the Winter Navigation Research Board, project W25-3 WiseIce.

## 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. Nature Communications 12, 5124. URL: https://doi.org/10.1038/s41467-021-25257-4, doi:10.1038/s41467-021-25257-4.

Durand, C., Finn, T.S., Farchi, A., Bocquet, M., Boutin, G., Ólason, E., 2024. Data-driven surrogate modeling of high-resolution sea-ice thickness in the arctic. The Cryosphere 18, 1791–1815. URL: https://tc.copernicus.org/articles/18/1791/2024/, doi:10.5194/tc-18-1791-2024.

Haapala, J., 2000. On the modelling of ice-thickness redistribution. Journal of Glaciology 46, 427–437. doi:10.3189/172756500781833106.

Kingma, D., Ba, J., 2014. Adam: A method for stochastic optimization. International Conference on Learning Representations .

Liu, G., Reda, F.A., Shih, K.J., Wang, T.C., Tao, A., Catanzaro, B., 2018. Image inpainting for irregular holes using partial convolutions, in: Proceedings of the European Conference on Computer Vision (ECCV). URL: https://arxiv.org/abs/1804.07723.

Liu, Y., Bogaardt, L., Attema, J., Hazeleger, W., 2021. Extended-range arctic sea ice forecast with convolutional long short-term memory networks. Monthly Weather Review 149, 1673 – 1693. URL: https://journals.ametsoc.org/view/journals/mwre/149/6/MWR-D-20-0113.1.xml, doi:10.1175/MWR-D-20-0113.1.

Pal, A., 2023. Lux: Explicit Parameterization of Deep Neural Networks in Julia. URL: https://doi.org/10.5281/zenodo.7808904, doi:10.5281/zenodo.7808904.

Palerme, C., Lavergne, T., Rusin, J., Melsom, A., Brajard, J., Kvanum, A.F., Macdonald Sørensen, A., Bertino, L., Müller, M., 2024. Improving short-term sea ice concentration forecasts using deep learning. The Cryosphere 18, 2161–2176. URL: https://tc.copernicus.org/articles/18/2161/2024/, doi:10.5194/tc-18-2161-2024.

Ren, Y., Li, X., Wang, Y., 2025. Sicnet$_{\text{season}}$ v1.0: a transformer-based deep learning model for seasonal arctic sea ice prediction by incorporating sea ice thickness data. Geoscientific Model Development 18, 2665–2678. URL: https://gmd.copernicus.org/articles/18/2665/2025/, doi:10.5194/gmd-18-2665-2025.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham. pp. 234–241.

Schimanke, S., Ridal, M., Le Moigne, P., Berggren, L., Undén, P., Randriamampianina, R., Andrea, U., Bazile, E., Bertelsen, A., Brousseau, P., Dahlgren, P., Edvinsson, L., El Said, A., Glinton, M., Hopsch, S., Isaksson, L., Mladek, R., Olsson, E., Verrelle, A., Wang, Z.Q., 2021. CERRA sub-daily regional reanalysis data for Europe on single levels from 1984 to present. URL: https://doi.org/10.24381/cds.622a565a, doi:10.24381/cds.622a565a.