

1 Conclusions and future outlook

This thesis aimed to develop a high-resolution (1km) deep learning sea ice concentration forecasting system for 1 to 3-day lead time. This was done through tuning the hyperparameters of the developed Deep learning system, comparing the developed Deep learning system against baseline-forecasts and dynamical models and finally applying XAI techniques to increase the transparency of the Deep learning model. Sea ice concentration from the sea ice charts, 2-meter temperature and the grid-adjusted x and y components of the 10-meter winds from AROME Arctic and passive microwave OSI SAF SSMIS sea ice concentration observations were pre-processed and structured into three datasets (one for each lead time). Furthermore, the Deep learning model was presented as a variation of the U-Net architecture, with the major architectural difference to the U-Net being a novel reformulation of the prediction task through the introduction of cumulative contours.

It was shown that a U-Net architecture which closely resembled the original formulation of Ronneberger et al. (2015) was not able to capture the intermediate sea ice concentration classes. This was shown to be a result of the sea ice concentration distribution highly favouring the $< 10\%$ and $90 - 100\%$ ice categories, as well as the intermediate classes also constituting the MIZ which changes extent at a daily frequency.

The introduction of cumulative contours was made possible by taking into consideration the underlying properties of spatial concentration distributions, and showed an immediate improvement in terms of resolving all ice categories. Hence the approach demonstrated the effectiveness and importance of custom-tailoring the deep learning network architecture to the physical quantity to be predicted.

Increasing the depth of the U-Net from 256 to 1024 feature maps in the bottleneck was shown to have an indifferent to detrimental effect on forecast performance. This was explored from different perspectives, first the 256-depth Deep learning system quickly converged to the training dataset, probably due to the limited number of samples. Thus increasing the number of parameters by a factor of 4 (512-depth) or 16 (1024-depth) would only add superfluous trainable parameter increasing the risk of underfitting. Second, the depth of the U-Net was explored with regards to the theoretical and effective receptive field of the feature maps in the bottleneck. Here it was argued that despite the 1024-depth U-Net facilitating a scene encompassing theoretical receptive field in the bottleneck, this was not reflected in the effective receptive field causing the features to still be mostly effected by a local neighborhood. This has implications which should be considered when designing deep learning architectures for high-resolution predictors in the future, since regardless of U-Net depth computed feature maps will be predominantly dominated by a limited surrounding area. This implies that a U-Net will be restricted in terms of how extensive the captured spatial dependencies of the computed features can be, such that

higher resolution predictors are to a higher degree influenced by local processes. Future work should assess the relationship between predictor resolution and the effective receptive field to determine how it affects forecast performance.

This work successfully applied the NIIEE for a 1km resolution grid. Furthermore the NIIEE was demonstrated to be conserved for varying resolutions, both in terms of sea ice concentration grid resolution and sea ice edge grid resolution. Finally, it was shown that NIIEE and binary cross entropy loss have a strong correlation, allowing model selection to be performed 92% faster while still achieving high NIIEE during testing.

When assessing the confidence for the predictions, it was seen that the Deep learning system tended to predict similar cumulative contours for all ice categories except fast-ice. This was discussed in terms of the loss function, and the shared decoder used for the cumulative contours formulation of the prediction task. And it was explored in the discussion the ramifications of all cumulative contours contributing to a common computed loss which is backpropagated throughout a shared decoder. It was noted that the shared high confidence North of Svalbard towards Greenland for all the cumulative contours (except for fast-ice) could adjust the weights towards predicting mostly predicting this region with confidence for all cumulative contours. It was also noted that the fast-ice cumulative contour potentially degrades the quality of the computed loss, since the sea ice exerts different physical properties and spatial distributions. Future work should avoid predicting ice categories which deviate from the expected physical behavior of the desired cumulative contours, especially fast-ice but also $< 10\%$. It is also advised to reformulate the U-Net architecture following principles from multitask learning as described in Crawshaw (2020), since such an architecture would preserve the intention behind the cumulative contours while also avoiding negative interactions during training.

Despite the limitations of limited and inconsistent data, it has been shown that the Deep learning system outperforms baseline-forecasts and dynamical state-of-the-art forecasting systems in terms of short term sea ice forecasting. It was also shown that the Deep learning system retains the high performance when validated against independent AMSR2 sea ice concentration observations. These results demonstrates that significantly cheaper deep learning forecasts (in terms of computational expenses) are still meeting the requirements for forecast accuracy. Furthermore, when the predictors are published the Deep learning system can produce forecasts in less than a minute. This means that the forecasts can be sent to maritime operators in the Arctic a day before the forecast valid date.

This work attempted three different techniques to increase the transparency of the Deep learning system. First, individual predictors or samples were permuted. Second, a the possibility to conduct a sensitivity analysis with a Deep learning system was investigated. Finally, Seg-GradCAMs of individual contours were analyzed. One result from the experiments was the importance of AROME Arctic predictors, and that the model was able to

fit strongly to 2-meter temperature. This motivates the inclusion of additional forecasting systems as predictors, since there is currently a lack of ice-ocean interactions in the Deep learning system. Based on the discussion, it is recommended to investigate the possibility to include variables from a ocean wave prediction system such as Carrasco et al. (2022) due to the important interactions between surface waves and MIZ dynamics.

However, the strong fit to 2-meter temperature from AROME Arctic was shown to potentially limit the potential of the Deep learning system through the Seg-GradCAMs. When 2-meter temperature was present, the results indicated that the 2-meter temperature caused the Deep learning system to only assign pixels as important for making a prediction when they were located above the sea ice. However, these results are based on a case to case comparison, and lack statistical robustness. Future work on explainable deep learning forecasting systems should explore the possibility to systematically process deep learning output using XAI techniques such that results are directly comparable and interpreted using statistical methods.

Finally, although the Deep learning system showed signs of inferring physical connections between variables, and it was demonstrated that the model is able to infer seasonality only from the physical predictors, it is noted that the Deep learning system is a statistical model without the necessary framework to simulate the physical processes. It was noted in the discussion that as input data becomes more out-of-distribution, less trust can be attributed to the potentially convincing result. Hence, further work should inspect different approaches to constructing synthetic samples for a sensitivity analysis, which closer resembles in-distribution predictors.

Three research questions were stated in the Introduction

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

The first question was explored through measuring NIIEE performance at different lead times, against multiple dynamical models as well as baseline-forecasts and by validating performance against different sea ice concentration datasets.

The second question was exemplified through the use of a case study. The Deep learning forecast is mainly adjusting the recent sea ice chart used as predictor, based on the physical interpretation it correlates from the atmospheric predictors.

The third question motivated the use of XAI techniques, with this work presenting a

novel attempt at understanding the predictions from a deep learning sea ice forecasting system using Seg-GradCAM. The techniques gave a possible explanation as to why 2-meter temperature have a detrimental effect on the Deep learning system performance, which motivates further work on the subject to develop a consistent methodology for diagnosing deep learning forecasting systems to achieve explainable predictions.

References

- Carrasco, A., Øyvind Saetra, Burud, A., Müller, M., and Melsom, A.: PRODUCT USER MANUAL For Arctic Ocean Wave Analysis and Forecasting Products ARCTIC_ANALYSIS_FORECAST_WAV_002_014, Tech. rep., EU Copernicus Marine Service, 2022.
- Crawshaw, M.: Multi-Task Learning with Deep Neural Networks: A Survey, <https://doi.org/10.48550/ARXIV.2009.09796>, 2020.
- Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Lecture Notes in Computer Science, pp. 234–241, Springer International Publishing, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.