Developing data-driven ocean models for the Norwegian coast and fjords using graph neural networks

Ina K. B. Kullmann[1], Mateusz Matuszak[1], Johannes Röhrs[1], Thomas Nils Nipen[1], Ivar Ambjørn Seierstad[1], Kai H. Christensen[1,2], Ann Kristin Sperrevik[1]

[1]Norwegian Meteorological Institute, Henrik Mohns Plass 1, 0371 Oslo, Norway
[2]University of Oslo, 0371 Oslo, Norway

Abstract
Recent advancements in using graph neural networks (GNN) to create global and regional weather forecasting models have demonstrated the ability to produce highly skilled forecasts, comparable to traditional deterministic numerical weather prediction models. In this study, we extend these techniques to train a high-resolution ocean model using the Anemoi framework, which has been successfully employed in developing ECMWF's and MET Norway's latest data-driven weather forecasting models. Our aim is to integrate novel machine learning methods with operational ocean general circulation models (OGCM), achieving resolutions down to 160 meters. A data-driven model trained at this resolution will enable us to produce cost-effective, high-resolution forecasts for the entire Norwegian coast, necessary to properly resolve ocean dynamics in Norway's countless fjords, where the majority of human marine activities take place.

MET Norway has a long-standing tradition of running operational OGCMs, which are crucial for applications such as search and rescue operations and oil-spill preparedness. Our primary coastal forecasting tool, the Norkyst model, provides predictions of ocean temperatures, salinity, and currents at 800-meter horizontal resolution across multiple depth layers. This model has been operational for over a decade, and since early 2024, we have established two new, smaller model domains with 160-meter resolution. Additionally, we have access to a 10-year hindcast of the Norkyst 800m model. This extensive repository of high-resolution data serves as an invaluable resource for training our data-driven models.

We first explore the ability of a GNN to emulate the forward time-integration of the OGCM on coastal scales for short lead-times. Our training methodology involves a two-step process: first utilizing 10 years of 800m data, followed by fine-tuning on 1 year of 160m data. If these steps do not yield satisfactory results, we plan to explore transfer-learning techniques from larger model domains, such as the TOPAZ system from the Copernicus Marine Service.

In this talk, we will present preliminary results after training on the ocean model archive at 800m resolution, with forecast lead times extending up to several days. We will also provide a comparative analysis of our preliminary results against existing operational forecasts to highlight the potential of our approach. Furthermore, we will discuss the challenges and insights encountered in adopting a data-driven approach using the Anemoi framework. Finally, we outline our plans for future research aimed at enhancing our capabilities for high-resolution coastal ocean forecasting in complex coastlines.