HIGH PERFORMANCE COMPUTING MODERNIZATION PROGRAM RESEARCH PROJECT #: HPCMP-HIP-25-038

Data-driven Ocean Forecast Model for Ensemble Prediction

About U.S. Naval Research Laboratory (NRL) Ocean Sciences Division:

The U.S. Naval Research Laboratory is the corporate laboratory for the U.S. Navy. The Ocean Sciences Division is located at Stennis Space Center, MS. Researchers in the Ocean Dynamics and Prediction Branch seek to understand and represent the fundamental dynamics of the ocean with the long-term goal to transition prediction systems for operational use. With access to major high-performance computing resources, we utilized over 300 million CPU hours in the last year. Research results transition to operational systems at ocean forecast centers and are used in daily decisions covering a broad range of applications around the globe.

RESEARCH LOCATION: Stennis Space Center, MS

PROJECT DESCRIPTION:

This is an opportunity to research with the nation's largest group of dedicated ocean prediction researchers, in building numerical systems that forecast ocean dynamics, developing techniques to process satellite and in situ information, and assimilating observations into numerical models.

In this project the aim is to demonstrate an AI/ML data-driven model for ensemble ocean prediction in the Gulf of Mexico and other areas of interest presently forecast using physics-based regional data assimilative forecast systems. We propose to begin with the OceanNet approach in Chattopadhyay, et al. 2023 (OceanNet; arXiv:2310.00813v1) and Gray, et. al. 2024 (https://doi.org/10.5194/egusphere-2024-1238) with training and validation performed on the nVIDIA GPU resources on the Navy DSRC systems narwhal and nautilus, but the plan is to allow for additional/alternative approaches based on a review of the most recent literature. The exact plan will depend on the background and interests of the intern.

- Update the literature review for the most recent publications.
- Configure appropriate training and validation datasets on the DSRC system (e.g., 3 years of weekly 90-day 32-member ensemble forecasts of ocean SSH and SST for the GoM).
- Install required software in the user account (miniconda python installation with additional TensorFlow/Keras or pytorch software).
- Demonstrate the capability to submit single-GPU and multiple-GPU jobs to the queue Install the OceanNet software modified for the Navy dataset (prior work) and port it to run efficiently on the DSRC GPU resources.

Under the guidance of a mentor, the intern will research mostly with python coding and some shell scripting on Linux systems to assemble the training dataset(s), install python and software packages, interpret the OceanNet code and research options to parallelize the model training, complete the ML training and validation, compare forecast skill with the numerical model using basic skill metrics, document the results in a written report and an oral presentation.

The simplest demonstration of OceanNet on the HPC system should be straightforward. Enabling parallelization to access more GPU memory will likely require significant independent research to understand the problem and PyTorch fundamentals and implement a solution.

Successful completion of the internship will demonstrate skill in application of HPC resources to machine learning problems of broad interest to the science and engineering community. Student will have access to NRL personnel for ocean data/physics/numeric questions, and onsite PETTT support can answer questions and point to resources for specific hardware/software issues. Local topical seminars and site/facility tours organized for NRLSSC SEAP/NREIP students will be open to the HPC intern.

ANTICIPATED START DATE:

June 2025 – Exact start dates will be determined at the time of selection and in coordination with the selected candidate.

QUALIFICATIONS:

Computer science coursework with interest in machine learning. A basic knowledge of Python and Linux shell scripting will be required before the start of the internship.

ACADEMIC LEVEL:

Degree received within the last 60 months or currently pursuing:

- Bachelor's
- Master's
- Doctoral

DISCIPLINE NEEDED:

- Computer, Information, and Data Sciences
- Engineering
- Science & Engineering-related