ML-based ocean emulator as a surrogate for ocean forecasting and reanalysis systems

Jungwon LEE[1], Ronan Fablet[1], Charles-Emmanuel Testut[2]

[1]IMT Atlantique, Lab-STICC, Brest, France
[2]Mercator Ocean International, Toulouse, France

Abstract

Numerical Ocean models are computationally expensive, motivating the development of ML (Machine Learning)-based emulators as efficient alternatives. However, ML-based surrogates often struggle to represent high-frequency signals. Addressing this methodological gap, this study investigates whether reconstructing a physically consistent initial condition using an ML-based surrogate—given uncertain and incomplete observations—can accurately reproduce ocean trajectories.

An optimal initial condition is reconstructed using the ML-based ocean dynamical model **GLONET**, leveraging the conceptual similarity between ML optimization and weak 4D-Variational Data Assimilation (4D-Var). The optimized initial condition, obtained via gradient-based optimization of the surrogate model, is propagated within both ocean forecasting and reanalysis frameworks. To assess how prior knowledge of the ocean state affects reconstruction, uncertainties in the initial condition are treated as control variables within OSSE and OSE configurations.

Our preliminary experiments focus on sea surface variables for the global ocean. Assuming the reanalysis ocean state as a fully accessible reference, results show that RMSE can be reduced by up to **51%**, with a corresponding **4-day gain in predictability** at the first forecast lead time in global ocean forecasting, relative to the reference. In an OSSE setting, we also assess how to the 4DVar scheme with the ML-based surrogate for short-term forecasting and reanalysis tasks from sparse and noisy observations.

Overall, the results support the potential of ML-based dynamical surrogates in 4DVar schemes for deterministic ocean forecasting and reanalysis tasks. We discuss further how these findings could contribute to a transition from computationally intensive numerical ocean models toward efficient ML-based emulators in ocean data assimilation systems.