A comparison of data-driven approaches to build low-dimensional ocean models

Niraj Agarwal¹, Dmitri Kondrashov², Peter Dueben³, Eugene Ryzhov¹, and Pavel Berloff¹

¹Imperial College London, Department of Mathematics, London, United Kingdom of Great Britain – England, Scotland, Wales (n.agarwal17@imperial.ac.uk)
²Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, USA
³ECMWF, Shinfield Road, Reading, UK

I will present a comprehensive inter-comparison of linear regression, stochastic and deep-learning-based models for reduced-order statistical modelling of the simplified ocean circulation. The reference dataset is provided by the top 150 empirical orthogonal functions (EOFs) and principal components (PCs) of an idealized, eddy-resolving, double-gyre ocean model. Our goal is to have a systematic and comprehensive assessment of the skills, costs and complexities of all the models considered.

The model based on linear regression is considered as a baseline. Additionally, we investigate stochastic models (linear regression plus additive-noise and a multi-level approach), deep-learning models (a feed-forward Artificial Neural Network (ANN), a Long Short Term Memory (LSTM)), and deep-learning augmented linear regression models (also called hybrid models). We also explored stochastically improved deep learning methods by adding spatially correlated white noise in the deep learning models to account for the residuals and left out variance in the discarded PCs. The assessment metrics considered are climatology, variance, RMSE, instantaneous correlation coefficients, frequency map, prediction horizon, and computational costs for training and predictions.

Until now, we found that the hybrid LSTM models perform the best, followed by the multi-level linear stochastic model and multiplicative white noise model. Additionally, hybrid models found to perform better when augmented by spatially correlated white noise. This suggests that an amalgam of physics, memory effects, and stochasticity provides the best strategy for low-order representation of oceanic process. However, LSTM was also found to be most expensive to train and forecast amongst all. Skills of simple stochastic models are similar to those of the linear regression model but superior to those of the pure deep learning models, as evidenced by relatively better frequency maps, infinite prediction horizon, and low running cost.

Overall, our analysis promotes multi-level stochastic methods, with memory effects, and stochastic hybrid methods for low-dimensional ocean models as a more practical option when compared to pure deep-learning solutions as they are more accurate, stable, and low-cost. Furthermore, this is an ongoing research project and more updated results will be discussed at the time of
presentation.