Assimilation of SST retrievals via ML and application to a global ocean eddying forecasting system (GOFS16)

A multitude of sea surface temperature (SST) datasets have been produced and disseminated in the recent years. These datasets vary in processing levels, treatments for cloud/rain contamination, probe different depths, and so on. Although these SSTs are strongly correlated with the SST from Ocean General Circulation Models (OGCM), there are different at the same time. Simple L3 SST assimilation strategies, based on the equivalence between those SSTs and the model SST, are therefore suboptimal; proper assimilation requires ad-hoc, dataset-dependent operators. In this context, machine learning (ML) approaches can be useful for constructing such operators and different SST products can be employed by simply re-training the network. In the presentation, several ML algorithms (random forest, CNN, cGAN) will be proposed and compared, showing the importance of convolutional approaches, and resulting in a neat preference for the cGAN. Such operator will be used in the context of a global high-resolution ocean forecasting system (GOFS16). GOFS16 is a short-term prediction system (6-day) whose purpouse is to represent the full dynamics and life cycle of baroclinic eddies in the majority of the global ocean. It has been running daily in operational mode at CMCC since August 2017 and it is based on a global eddy-resolving configuration at 1/16° resolution that corresponds to 6.9km at the Equator, increasing toward high latitudes. In 2021, GOFS16 joined the OceanPredict inter-comparison project that gathers and compares the global prediction systems from diverse institutions. The ocean engine is coupled with a 3dvar assimilation system, called OceanVar, that ingests multiple and heterogeneous data sources (insitu, altimetry and SST data) on the same model grid. The L3 SST data are particularly noteworthy for their abboundancy and highest resolution: presently they are placed directly on the first model level with an increased observation error. The application of the ML operator is meant to improve the skill of the system by reducing the systematic bias that can occur, especially within the first model vertical levels.

Andrea Cipollone (1), Matteo Broccoli (1), Doroteaciro Iovino (1) (1) CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy