**Downscaling of ocean fields by fusion of heterogeneous observations using Deep Learning algorithms**

For several decades, a large variety of satellite sensors has allowed us to dramatically improve the quality of Earth's climate monitoring through satellite remote sensing imagery. Satellite sensors provide global coverage of the ocean. These sensors are diverse both in terms of remote sensing technology and in geometrical sampling. They provide us with observations of a multitude of geophysical parameters at a multitude of spatiotemporal resolutions. Products of this abundance of sensors allow for the recovery of multiple physically important fields of the ocean state:

- *Sea Surface Temperature* (SST) with high-resolution radiometers such as the VHRR sensors launched onboard meteorological satellites (EumetSat…),
- *Ocean Circulation* with Altimeters (Topex Poseidon and then Jason altimeters),
- *Ocean Biological Productivity* with multispectral ocean color sensors (SeaWIFS, Meris, MODIS, VIIRS),
- *Ocean Salinity* (SMOS and Aqua).

These fields provided by the satellite sensors play a crucial role in monitoring the ocean's response to global warming. The ocean is a major driver of our climate state, via processes air-sea exchanges (radiative processes, latent and contact heat fluxes), latitudinal heat transport via ocean circulation, climate regulator via its enormous heat stockage capacity.

The aim of the research is to use the existing fine resolution of the SST to increase the coarse SSH resolution provided by a satellite altimeter and to retrieve simultaneously the 2 components of the current with this improved resolution. Owing to the high-resolution data fields we deal with and the large area we want to study, Deep Learning technique (DL) proposes suitable methods to solve this complex problem. A first study, conducted by the supervisory team, uses simulated data provided by the output of the NATL60 high-resolution ocean numerical model (https://meom-group.github.io/swot-natl60/virtual-ocean.html) and shows the ability of Convolutional Neural Networks (CNN) to deal with this problem.

Recent developments in artificial intelligence, combined with the availability of large datasets, high resolution simulations, and computational power, provide enormous potential to learn the hidden structure of geophysical phenomena. In the last decade, statistical and machine learning methods have started to allow the investigation of multi-dimensional data sets. They are increasingly used to extract patterns and insights into causal links from the ever-increasing stream of geospatial and climate data. Specifically in the domain of downscaling there are exploratory works that show the effectiveness of such architectures (T. Bolton, L. Zanna (2019). Applications of Deep Learning to Ocean Data Inference and Subgrid Parameterization. JAMES, 11, 376–39). Problematics of downscaling are also closely related to super resolution problems encountered in Image Processing and who has benefited greatly from deep learning in recent years (Wang et al (2020), Deep Learning for Image Super-Resolution: A Survey. 10.1109/TPAMI.2020.2982166). Furthermore, these approaches are nowadays the object of challenges by the international community (https://platform.ai4eo.eu/challenge/air-quality-and-health).

The thesis will focus on the Gulf-stream region (figure 1) which is the most energetic region of the North Atlantic Ocean (26°N, 45°N; 40°W, 65°W).

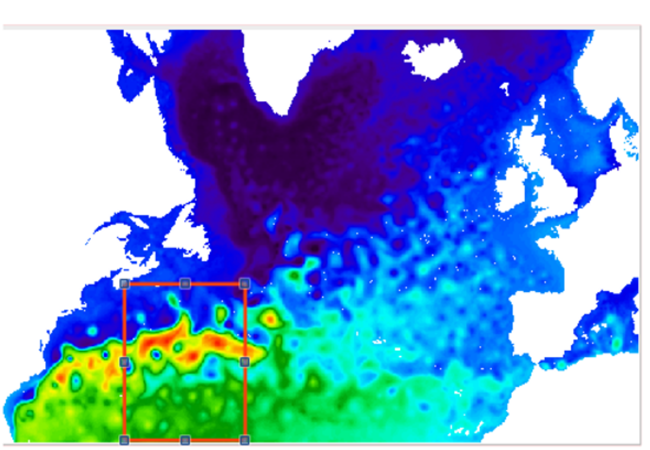

**Figure 1** *The domain under study*

As mentioned above a  proof of concept has been proposed: low resolution of altimeter data has been merged with high resolution of sea temperature surface data in a CNN architecture that can retrieve the ocean circulation with a good accuracy, even when data are noisy.  It was possible to estimate the current at a resolution 12x15 km starting from a coarse resolution of 120x122 km. This preliminary study can be the starting point of this thesis, a first challenge being applying this approach to real satellite data. In our approach, the PhD student would be called upon to use Convolutional Neural Networks and other Deep Learning architectures in order to downscale satellite images to very fine resolutions. The work would include the conception of new architectures, prototyping them, and expanding this work from twin experiments with simulated data to applying it to satellite data.

For this area, satellite data are available on coarse resolutions that do not exceed 12x15 km, so it will be necessary to combine satellite and simulated data at different scale and to use transfer learning in order to combine observation and physical knowledge.

The results will be validated not only by the traditional Data Science values, such as a R2 correlation coefficient and Root Mean Square Error, but also by preserving geophysical properties of the ocean. The preservation of these will require prior analysis of the dynamics of the fields, and probably require intelligent constraining of the architecture selected.

This work has additional potential applications such as the study of the Gulf Stream, the prediction of primary production, but could also showcase a methodology to adopt with other downscaling problems that have to preserve intrinsic physical properties.

The qualities sought in a candidate are, in order of importance:

- a real interest in the subject,
- a solid grasp of machine learning theoretical basics (backpropagation, architectures),
- applied experience in deep learning with Python (PyTorch/Keras/TensorFlow, since testing and creating new architectures will be important),
- a comprehension of ocean circulation dynamics,
- fluency in French and English.

This thesis is thus at the junction of two disciplines: geosciences and artificial intelligence. The PhD student will share his working time between LIP6 and LOCEAN, both laboratories being located on the Pierre and Marie Curie campus of Sorbonne University.

Dominique Béréziat is an associate professor at Sorbonne Université. His research is led at LIP6 laboratory and focused on data assimilation methods applied to image processing problems. More recently, he is working to combine data assimilation and deep learning methods.
Anastase Charantonis is an associate professor of Deep Learning at ENSIIE, Evry, and does his research at the LOCEAN - IPSL laboratory of Sorbonne Université. His interests are focused on developing ML/DL methodologies to solve geophysical problems.
Sylvie Thiria is professor emeritus at LOCEAN. She is specialist in Machine Learning and their applications to environment and climate.