# Model Artificial Intelligance Ocean Observations
# MAIOO

## Learning-based calibration of ocean carbon models using emerging data from new observing devices and platforms

**Context**. The Ocean Carbon Pump represents the processes that regulate the absorption and storage of atmospheric $CO_2$ in the deep ocean. This Pump plays a major role in climate and biogeochemical cycles (C, $O_2$, nutrients, ..). The biological part of this pump reduces unperturbed atmospheric $CO_2$ by 35 to 50% [1]. It is driven by photosynthesis at the ocean surface, which creates particles that are exported by gravity in the deep ocean (export production). These particles are partially remineralized in the ocean by bacterial activity and zooplankton metabolism: when they reach the deep ocean, the carbon they carry is isolated from the atmosphere for centuries.

Until the 2000s, observations of the deep ocean carbon pump were limited to a few dozen sediment traps on fixed moorings. Therefore, although these sparse data provide important first-order information in terms of vertical carbon flux, with a rough estimate of seasonal and regional variability, they are unable to strongly constrain biogeochemical models. This statement is largely based on three inescapable facts: i) these models involve a very large number of parameters, many of which are poorly known [ref]; ii) the variability of the export flux and its fate in the deep ocean is not only determined by large basin scale processes, but also by much smaller scales associated with meso and submeso scales related to ocean dynamics; iii) the vertical carbon flux is driven by a strong heterogeneity of particles, in terms of size, density, shape, composition, sinking velocity.

Since the 2010s, deep ocean observation has undergone a tremendous qualitative and quantitative evolution. In fact, new high-frequency devices on autonomous platforms are shedding light on processes that take place in the ocean depths, which have been difficult to access until now. Argo biogeochemical floats are thus deployed intensively at an increasing rate in the world ocean [2]. Thanks to technological advances in engineering sciences, critical parameters of the carbon cycle are now observable. Observations as different as oxygen, nitrate, fluorescence (chlorophyll indicator), carbon content of particles (by optical techniques), distribution of large particles and zooplankton (by imaging) are now or will soon be available as standard Argo data [ref].

Although there are still important issues to be addressed, at first order, BGC models correctly simulate the processes that govern carbon cycling in the euphotic. Nevertheless, the fate of carbon after its transfer from the surface layer remains quite coarse and highly under-parameterized [3]. The depth of remineralization is of critical importance for the carbon cycle and climate, as well as for deep ecosystem functioning [4]. Together with acoustic data, the observations obtained with Argo floats are new data, non-existent until now, that should constrain these little considered processes (more specifically particle dynamics and zooplankton distribution [4]). Given the budget and intensity of coverage expected from these BGC Argo floats, it is crucial to quantify how these new observations can better describe the processes involved and constrain the ocean carbon models used in IPPC-type simulations.

**Objectives.** The main objective of this thesis is to develop a new and efficient methodology to better constrain the parameters of BGC models, in particular for biogeochemical processes in the mesopelagic layer, using these emerging observations. From a methodological point of view, model calibration problems are classically stated as minimization problems. The complexity of models has been increasing steadily over the last decades. This raises concerns about the complexity of the model/observation system in terms of the number of variables/data, the estimation of an increasing number of parameters, the quantification of uncertainties, the confidence and robustness of the model results, the nonlinear behavior of natural systems, and the cost in terms of computational power. New approaches are therefore needed. They must simplify the models while retaining the relevant processes and scales, develop efficient tools to allow exchanges between heterogeneous observations and models, and rigorously quantify uncertainties. In this regard, Artificial Intelligence has recently become an emerging field in oceanography [6], offering promising avenues both for the analysis of complex heterogeneous data and for the implementation of model-data interaction. There is a growing consensus that solutions to complex scientific problems require new methodologies that can integrate traditional physics-based modeling with Machine Learning approaches [7]. Future climate models are likely to rely on a tight integration of physical and statistical modeling paradigms, which is a fundamental change. Recent advances in AI, particularly differentiable physics, open up new avenues for addressing model parameterization issues using so-called differentiable emulators. Overall, these differentiable emulators can bridge the gap between current operational systems and AI approaches.
The proposed methodology will be applied to the NEMO/LIM3/PISCES model. This model is a national ocean/sea ice/carbon cycle dynamics model currently used in IPCC simulations and in operational oceanography (by Mercator Ocean), with a strong European dimension [5]. PISCES is considered as a model of intermediate complexity, taking into account five limiting nutrients, two types of phytoplankton, zooplankton, detritus and dissolved organic matter. Although particle dynamics and zooplankton behavior are considered in the mesopelagic layer, the processes leading to the attenuation of carbon flux with depth are nevertheless taken into account in a coarse way, as observations were lacking until now to constrain them effectively.
**Proposed approach and work plan.** The proposed approach will first rely on Observing System Simulation Experiments (OSSEs) to design and evaluate the proposed methodology before its application to real data. Overall, the proposed work plan includes three main tasks:

- **Task 1: OSSE design.** The PhD candidate will design an OSSE for 1D vertical model configurations for at least three contrasting target regions in the North Atlantic Ocean, Mediterranean Sea, characterized by long historical time series and intense observational coverage using new platforms and sensors: in the inter-gyre region of the Northeast Atlantic at the UK Porcupine Abyssal Plain station, in the subtropical ocean near Bermuda at the US BATS station, and at DYFAMED in the Ligurian basin. Depending on the availability of additional data in other ocean regions, these three studies could be complemented by other stations. The pseudo-observations will be simulated by the PISCES 1D model according to different model parameterizations. This pseudo-observation dataset will be representative of observations measured on BGC Argo floats, such as temperature, oxygen, fluorescence, as well as less classical data (vertical particle size profiles, or vertical nekton migration), complemented by satellite data such as Sea Surface Temperature and surface chlorophyll.
- **Task 2: Learning-based calibration methodology.** We will exploit this OSSE context to design and evaluate the emulator approach for parameterizing BGC models from

observational data. The evaluation framework will measure the influence of the following key factors: emulator architecture, sampling models, and learning criterion. With respect to emulator architecture, both direct inversion schemes and data assimilation-based architectures will be considered [8].

- **Task 3: Application to real data.** The proposed emulator-based approach will then be applied to real data in the North Atlantic, gathered at PAPS and BATS stations, and in the Mediterranean Sea (DYFAMED). Furthermore, in addition to the standard historical data, there are many observations from process cruises, which study the whole water column. Two main concrete questions will be addressed: Is it necessary to vary the parameters for different ocean regions / production regimes? Do the data require changes in the way processes are represented in PISCES? The final step of the thesis will be to adapt and extrapolate the methodology to the global ocean using the full Argo float array, supplemented by satellite data.

**Added values, Synergies, Hosting teams.** This project is fundamentally at the crossroads of two disciplines: ocean biogeochemistry and computer science. It contributes to the emergence of Artificial Intelligence in marine sciences. By developing innovative approaches (based on emulators), it aims to use observations (highly heterogeneous) that have not been taken into account in modeling so far, especially from the Argos BGC floats (collaboration with Laboratoire d'Océanographie de Villefranche, LOV) that are deployed worldwide at an increasing rate. Finally, we will estimate the poorly constrained parameters of a model that is intensively used in IPCC simulations and in operational oceanography (collaboration with Mercator), and with an approach capable of quantifying errors and uncertainties in a rigorous manner. The PhD student will benefit from the multidisciplinary environment developed in the framework of the AI OceaniX Chair at the crossroads of AI and oceanography (http://cia-oceanix.github.io) on the Brest campus, as well as in the Isblue University Research School (https://www.isblue.fr), which brings together the UBO, Ifremer, and the engineering schools of western Brittany. Beyond the grant, the PhD student will benefit from state-of-the-art AI resources as well as financial support for stays abroad (generally up to 6 months) in partner laboratories. These partners are Pierre Lermusiaux, MIT, Boston, a world-renowned expert in ocean data assimilation and ocean modeling using Artificial Intelligence approaches, and Adrian Martin, NOCS, UK, who has an undeniable international reputation for his expertise in the carbon pump, and in physical - biological interactions. This thesis is very strongly linked to the ANR project APERO (PI: L. Memery), aiming at constraining the fate of carbon in the mesopelagic layer on the basis of an ambitious cruise in 2023 around the PAP station. Moreover, APERO is part of JETZON, an international project supported by the UN (decade of the ocean), with Adrian Martin as PI.

**References**.


[1] Follows, M., and T. Oguz (2012), *The ocean carbon cycle and climate*, Springer Science & Business Media.

[2] Claustre, H., K. S. Johnson, and Y. Takeshita (2020), Observing the Global Ocean with Biogeochemical-Argo, in *Annual Review of Marine Science, Vol 12*, edited by C. A. Carlson and S. J. Giovannoni, pp. 23-48, doi:10.1146/annurev-marine-010419-010956.

[3] Le Moigne, F. A. C. (2019), Pathways of Organic Carbon Downward Transport by the Oceanic Biological Carbon Pump, *Frontiers in Marine Science*, *6*, doi:10.3389/fmars.2019.00634.

[4] Martin, A., *et al*. (2020), The oceans' twilight zone must be studied now, before it is too late, *Nature*, *580*(7801), 26-28, doi:10.1038/d41586-020-00915-7.

[5] Aumont, O., C. Ethe, A. Tagliabue, L. Bopp, and M. Gehlen (2015), PISCES-v2: an ocean

biogeochemical model for carbon and ecosystem studies, *Geoscientific Model Development*, *8*(8), 2465-2513, doi:10.5194/gmd-8-2465-2015.
[6] Malde, K., *et al.* (2020). Machine intelligence and the data-driven future of marine science, J. Mar. Sci., 77(4), 1274-1285.
[7] Willard, J., *et al*. (2021). Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems, ArXiv: 2003.04919
[8] Fablet, R., B. Chapron, L. Drumetz, E. Mmin, O. Pannekoucke, and F. Rousseau (2020), Learning Variational Data Assimilation Models and Solvers, *Journal of Advances in Modeling Earth Systems*, doi:https://doi.org/10.1029/2021MS002572.
[9] Schartau, M., et al., (2017), Reviews and syntheses: parameter identification in marine planktonic ecosystem modelling, *Biogeosciences*, *14*(6), 1647-1701, doi:10.5194/bg-14-1647-2017
[10] Kriest, I., S. Khatiwala, and A. Oschlies (2010), Towards an assessment of simple global marine biogeochemical models of different complexity, *Progress in Oceanography*, *86*(3-4), 337-360, doi:10.1016/j.pocean.2010.05.002.