Action PPR Océan & Climat "Postdoctoral Fellowship for AI-native approaches for ocean modelling, forecasting and monitoring"

Keywords: Digital Twin of the Ocean (DTO), Artificial Intelligence (AI), Ocean Forecasting & Reanalysis, End-to-End Neural Learning, Observation-Driven and Generative Modelling, GPU Acceleration, Science-driven demonstrations.

Scope. There is growing agreement that AI, particularly deep learning, provides new means to develop improved earth system models. Recent achievements like short-to-medium term weather forecasts using neural networks (Bi et al., 2023; Lam et al., 2023) reaching top performance are remarkable examples. These applications are still in their early stages for the digital twins of the ocean (DTO) (Wang et al., 2024). The PPR "Océan & Climat" aims at advancing these capabilities and at supporting significant breakthroughs in ocean simulation, monitoring, and forecasting**.**

Challenges of Existing Ocean GCMs and Data Assimilation Systems. Current ocean forecasting and reanalysis tools are valuable for global and regional applications. However, they fall short of what's expected for "digital twins" of the ocean, especially regarding data ingestion and user needs. Here are some key limitations:

- ***Limited Ocean Dynamics Representation*:** Existing models struggle to capture the full range of upper ocean-atmosphere processes and to sample the distribution of plausible states (incuding. extremes);
- ***Underused Observational Data*:** Because of their lack of flexibility, these systems often miss the full potential of existing and emerging high-resolution data from satellites, e.g., SWOT, and in-situ observing systems;
- ***Computational Bottlenecks and Compatibility Issues*:** Traditional Fortran-based high-performance-computing (HPC) is not natively adapted to GPUs and does not integrate well with modern AI environments built on Python and GPUs.

AI-Native Solutions for the DTO. Recent successes in weather forecasting (Rasp et al., 2019) and initial ocean applications (Wang et al., 2024; Johnson et al., 2023) suggest exploring AI-native solutions to replace entire components of ocean modeling, forecasting and reanalysis schemes. One promising approach is "end-to-end neural learning". This allows the entire system, from observation and simulated data, to be optimized by learning from the relationships between data points. However, the specific components of these methods need further exploration to handle the complexities of the ocean. This initiative identifies **key breakthroughs** to be addressed through novel AI-native solutions such as**:**

- **Comprehensive Probabilistic Representation, Monitoring, and Forecasting:** Creating AI models that not only predict and infer the most likely state of the ocean but also account for the uncertainties involved and efficiently sample realistic states;
- **Efficient Use of Limited and Unevenly Distributed Observational Data:** Designing AI models to make the most of sparse and irregularly collected data;
- **GPU-Accelerated and "Plug-and-Play" Ocean Simulations:** Developing computationally-efficient AI models that can run on powerful GPUs and seamlessly integrate with DTO platforms.

These advancements will pave the way for the creation of AI-native DTO components.

Topical Demonstrations. Scientifically-motivated demonstrations will showcase these breakthroughs through real-world datasets and benchmarks (see below). These demonstrations will contribute to the scientific priorities of the PPR "Océan & Climat"[1], for instance towards improved monitoring and forecasting of tropical cyclones ("Défi 1") and the Arctic system ("Défi 2"). These demonstrations will leverage existing observational data (e.g., ARGO floats, satellite data), simulations (e.g., NEMO, CROCO), and reanalysis datasets (e.g., CMEMS products).

[1] See here the description of the scientific objectives of the PPR "Océan-Climat" (in French): https://www.ocean-climat.fr/Le-PPR/Les-defis-du-PPR

Methodology and Workplan. This action promotes structuring a collaborative research effort through the design of **open "data challenges"**. Data challenges have been key to rapid advancements in neural weather forecasting with a variety of approaches (Rasp et al., 2019; Lam et al., 2023; Johnson et al., 2023). The proposed workplan involves a first phase dedicated to the definition of two-to-four complementary data challenges according to the targeted AI-related breakthroughs and topical demonstrations. These data challenges will provide the experimental ground for postdoctoral fellowships addressing the development, evaluation, and exploitation of novel AI-native solutions for ocean modeling, forecasting and monitoring.
Each data challenge will specifically involve: (i) a specific problem (e.g., short-term forecasting of a specific variable), (ii) performance metrics (e.g., mean square error) with independent evaluation data (iii) baselines for benchmarking purposes, (iv) training datasets, (v) standardized ML-ready workflows for training and evaluation.
The proposed methodology will implement a workplan based on one-year cycles. Each cycle will comprise progress evaluations within each data challenge, informing the next round of scientific and technological development. The proposed 3-year workplan is as follows:

- ***Year #1 (Sept.2024 - Sept. 2025)***:
  - Sept. 30, 2024: Open call for proposals for ocean data challenges
  - Oct. 4 & 11, 2024: Webinars, presentation of the action, Q&A
  - Oct. 18, 2024: Deadline for the proposals of ocean data challenges
  - Oct. 18-Nov. 29, 2024: Collaborative review and revision of candidate ocean data challenges
  - Dec. 13, 2024: Publication of the selected ocean data challenges
  - Jan. 8, 2025: Publication of the call for proposals for 2-year postdoctoral fellowships
  - March 14, 2025: Deadline for the applications to 2-year postdoctoral fellowships
  - March 31, 2025: Selection of 2-year postdoctoral fellowships (team level)
  - July 2025: Publication of a first version of the targeted data challenges
  - Sept. 2025: 1st Ocean-AI workshop (in person), presentation of the ocean data challenges, hands-on session, presentation of the postdocs
- ***Year #2 (Sept.2025 - Sept. 2026)***:
  - March. 2026: 2nd Ocean-AI workshop (remote), review of the progress for each data challenge, proposal for evolutiogns of the data challenges
  - Sept. 2026: 3rd Ocean-AI workshop (in person), synthesis of the progress for each data challenge, distribution of updated versions of the data challenges
- ***Year #3 (Sept.2026 - Sept. 2027)***:
  - March. 2027: 4th Ocean-AI workshop (remote), review of the progress for each data challenge, proposal for evolutions of the data challenges
  - Sept. 2027: 5th Ocean-AI workshop (in person), synthesis of the progress for each data challenge, distribution of updated versions of the data challenges

Integration with French and European DTO Ecosystems. The proposed ocean data challenges and solutions will support the uptake of AI-native solutions into existing French and European DTO platforms. In particular, we will ensure full compatibility with the EDITO platform led by Mercator Ocean International.

Ressources:

- Up to 9 2-year postdoctoral fellowships
- Four-year full-time equivalent engineer positions for the design and implementation of the data challenges
- Additional operational budget (e.g., workshop venues)

Scientific committee: The scientific committee is composed of:

- R. Fablet, Prof. IMT Atlantique, Lab-STICC, Odyssey, France
- D. Durand, Senior Scientist, Covartec, Norway

- Y. Shin, Senior Scientist IRD, Marbec, France
- Y. Drillet, Mercator Ocean Intl, France

The scientific committee is in charge of the scientific supervision of this initiative, in particular regarding the selection of data challenges, the selection of the postdoctoral scholarships, and the follow-up of the progress of the implementation of the targeted workplan.

Technical coordination: Ifremer and the Odyssey team will be in charge of the technical coordination of this initiative. It includes the specification of the general architecture of the data challenges as well as the associated online platform for sharing related datasets and benchmarks.