

---

## AAP Post-Doc PPR “AI-native approaches for ocean modeling, forecasting and monitoring” 2025

### Context

The research program (PPR) PPR Océan & Climat, led by CNRS and Ifremer, has been established by the government for a period of 7 years, from 2021 to 2028. The PPR is designed around 7 challenges (<https://www.ocean-climat.fr/Le-PPR/Les-defis-du-PPR>) to address research issues related to both the advancement of climate and ecological knowledge, as well as blue economy, law, geostrategy, global management of socio-ecosystems, and societal well-being. Since 2022, two major project calls, three PhD calls, and one postdoctoral call have supported ambitious research projects addressing part of these challenges (<https://www.ocean-climat.fr/Les-actions-et-projets>).

In addition to the 7 challenges, the scientific council of the PPR has identified specific needs for the development of AI-native solutions for ocean modeling, forecasting and monitoring. There is growing agreement that AI, particularly deep learning, provides new means to develop improved earth system models. Applications for the digital twins of the ocean (DTO) (Wang et al., 2024, El Aouni et al., 2024) are still in their early stages. The PPR “Océan & Climat” aims at advancing these capabilities and at supporting significant breakthroughs in ocean simulation, monitoring, and forecasting.

The open call for proposals aims to support up to 9 fully funded two-year postdoctoral fellowships, covering both the salary up to a maximum of 150,000 euros and an operating budget up to 24,000 euros per postdoctoral position. The postdoctoral contracts must start no later than December 1, 2025. No management fees will be funded.

### Objectives

As described in the roadmap of the proposed initiative (see [here](#)), recent advances ([Rasp et al., 2019](#), [Wang et al., 2024](#); [Johnson et al., 2023](#)) suggest exploring AI-native solutions to replace entire components of ocean modeling, forecasting and reanalysis schemes. This allows the entire system, from observation and simulated data, to be optimized by learning from the relationships between data. However, these methods need further exploration to handle the complexities of the ocean. Through the call for ocean data challenges, the PPR Océan & Climat has selected five data challenges (DC), which shall serve as demonstration testbeds for the submitted postdoctoral projects:

- DC1: Emulation of global ocean reanalyses
- DC2: Probabilistic short-term forecasting of global ocean dynamics
- DC3: Arctic Sea ice forecasting
- DC4: High-resolution monitoring and forecasting of tropical cyclones
- DC5: Marine biodiversity prediction

A one-page summary of each data challenge is appended below and all information, including the description of the original data challenge proposals is available [here](#)<sup>1</sup>.

This call invites postdoctoral projects that will develop or exploit AI-native solutions for one of the data challenges listed above. They may focus on original methodological contributions for the considered

---

<sup>1</sup> <https://github.com/ppr-ocean-ia/data-challenges-info>

---

ocean data challenge(s) as well as on the exploitation of state-of-the-art solutions calibrated on a data challenge to address relevant scientific ocean-related challenges.

## Eligibility criteria

- The principal investigator(s) who will submit the proposal will be the (co-)supervisors of the postdoctoral candidate.
- At least one of the principal investigators shall hold a permanent position in a French research institute or university.
- The postdoctoral fellow will be hosted in French research institutes or universities.
- Each principal investigator may submit only one project.
- The postdoctoral project must address one of the ocean data challenges listed above (see the Annex for a short description of each data challenge).
- The proposal must comply with the submission template provided in Annex and be no longer than 7 pages (Calibri 11, 1.5 line spacing) without references and the CVs of the principal investigator(s).
- The principal investigator(s) cannot be members of the scientific committee of the PPR "Océan & Climat" or of the coordination committee of the associated AI-native initiative.

## Evaluation criteria

- The relevance of AI-native solutions proposed for the ocean data challenge(s) considered,
- The relevance of the scientific questions addressed,
- The novelty and the transformative potential of the project,
- The expertise of the (co-)PIs to develop the proposed research project,
- The integration into national and international collaboration networks,
- The coherence of the resources with the objectives and scope of the project.

## Evaluation procedure

The proposals will be submitted to Céline Dégremont ([celine.degremont@ifremer.fr](mailto:celine.degremont@ifremer.fr)) **before March 14, 2025**. The evaluation process will be supervised by the coordination committee of the AI-native initiative of the PPR "Océan & Climat" (see here). The assistance of external reviewers may be considered.

**Contact:** For any questions regarding this call, you can contact Ronan Fablet, [ronan.fablet@imt-atlantique.fr](mailto:ronan.fablet@imt-atlantique.fr).

## Agenda

- January 29, 2025: opening of the call for proposals
- March 14, 2025: submission deadline
- March 31, 2025: publication of the selected postdoctoral projects

---

## Appendices:

- A. One-page summaries of the 5 selected ocean data challenges
- B. Submission templates in French and English

## Terms of payment

Depending on the selection panel's decision on the project and depending on acceptance of the terms and conditions of the payment agreement between the research organisation carrying out the project and Ifremer, Ifremer will pay out its share of the funding totalling 174,000 euros net of tax, including 24,000 euros for operating costs (missions, purchase of equipment related to the position) according to the following schedule:

- A first instalment corresponding to 40% of the total amount subject to (cumulative conditions) the signature of the Funding Agreement, the payment by the ANR of the first instalment, the signature of the payment agreement and finally the presentation of the postdoctoral fellow's employment contract;
- A second instalment, halfway through the project, corresponding to 40% of the total amount, on presentation of a statement of expenses signed by the accounting officer.
- The final 20% of the total amount on presentation of a summary statement of expenditure, proof of salaries paid and a final postdoc report.

Management fees are not funded.

---

## Appendice A

### One-page summaries of the 5 selected ocean data challenges

---

## DC1: Emulation of Global Ocean Reanalyses

### Task description

Creating neural emulators of global ocean dynamics, which reproduce the time evolution of the global 3D state vector, given perfectly known initial conditions and time-varying forcings.

### Training dataset

- **GLORYS12 reanalysis data:** from 1993 to 2019, at 1/12° (50 levels) and 0.5° (30 levels) resolution – the latter used for prototyping.
- **Initial conditions and time-varying forcings used in operational contexts:** from CMEMS global prediction system and from ECMWF analyses.

### Evaluation metrics / data

Three categories of metrics will be computed:

- **Short-term accuracy in 3D space:** forecasting ability on short (<30 days) time horizons (also used for training).
- **Accuracy in observation space:** from interpolated emulator predictions.

- **Physical consistency:** assessing the potential of emulators to be used for longer timescales, consisting of physical diagnostics computed from multi-year simulations.

**Evaluation Data:** the last 5 years of GLORYS12 (2020-2024) reanalysis dataset

### Baseline solutions

Two state-of-the-art neural emulators will be available as baselines: [GloNet](#), [XiHe](#).

### References

- [Bora et al. \(2023\)](#)
- [El Aouni et al., \(2024\)](#)
- [Griffies et al. \(2016\)](#)
- [Lellouche et al. \(2021\)](#)
- [Ryan et al. \(2015\)](#)
- [Wang et al. \(2024\)](#)

---

## DC2: Probabilistic short-term forecasting of global ocean dynamics

### Task description

Probabilistic short-term (a few tens of days) forecasting of upper ocean dynamics (scalar and vector fields) from gridded satellite-derived data (2D+T and sparsely sampled fields) and ARGO float data (sets of vertical profiles).

### Training datasets

- **Satellite data:** Nadir altimetry and SWOT measurements for SSH, SST, SSS
- **ARGO data:** MLD, with added T, S, U, V for different depths in the 3D case.
- **Ocean simulation data:** e.g. [eORCA36](#).
- **Atmospheric forecasts:** up to 10 days.

### Evaluation metrics / data

**Standardized outputs for all solutions:** global scale at daily 0.25° resolution for the following variables

- **Case I:** SST, SSH, SSS, U, V, MLD
- **Case II:** T, H, S, U, V, MLD from surface to bottom

**Usual metrics for probabilistic forecasting:** CRPS, spread-skill ratio, for surface variables, and then in 3D.

**Data:** data from 2023 to 2024 will be available as inputs during inference

- gridded atmospheric forecasts up to 10 days lead time
- sparse (L3) satellite data for sea surface variables
- ARGO float data

L3 satellite-derived and ARGO data will also be used to evaluate the considered performance metrics.

### Baseline solutions

Operational product (CMEMS GLO Forecast) and deterministic ML baselines for the emulation of ocean dynamics ([XiHe](#), [GLONET](#)) using the initial conditions delivered by the operational system (CMEMS GLO Forecast).

### References

- [Price et al. \(2023\)](#)
- [Wang et al. \(2024\)](#)
- [Lam et al. \(2023\)](#)
- [Pauthenet et al. \(2022\)](#)
- [Chattopadhyay et al. \(2024\)](#)
- [El Aouni et al. \(2024\)](#)

---

## DC3: Arctic Sea Ice Forecasting

### Task description

Forecasting the Arctic sea ice's response to external forcings on an hourly to seasonal time frame, by predicting key sea ice parameters such as extent, thickness, concentration and drift over short (i.e. hours) and long (i.e. seasonal) time horizons. The proposed solution should produce outputs that can be compared against gridded observations (at 1, 10 or 60 km resolution) and drifter trajectories.

### Training datasets

- **Satellite data:** sea ice concentration, thickness and drift from missions such as [Sentinel-1 and 2](#), [CryoSat-2](#), [SMOS](#) and [AMSR2](#) (except 2016-2020).
- **Reanalysis data:** information on [oceanographic](#) and [atmospheric](#) variables.
- **Hi-res sea ice simulations:** e.g. NextSIM-F archives.

### Evaluation metrics / data

**Forecast accuracy:** comparison of predicted vs observed variables.

**Sea Ice Physics:** realism of several predicted statistics

**User-oriented metrics:** relating to activities like shipping, pollutant tracking, etc.

**Computational efficiency:** runtime and computational resource requirements

**Data:** the following from 2016 to 2020

- sea ice concentration and drift from [AMSR2](#) at 10 to 60 km resolution
- sea ice lead fraction from [MODIS](#) gridded at 1km resolution
- ice embedded drifter data from [IABP](#)

### Baseline solutions

[NextSIM-F](#), [IceCastNet](#), and a [sea ice thickness emulator](#) for short-term forecasting. For seasonal forecasts, [TOPAZ](#) and the aforementioned sea ice emulator.

### References

- [Williams et al. \(2021\)](#)
- [Sakov et al. \(2012\)](#)
- [Durand et al. \(2024\)](#)
- [Boutin et al. \(2023\)](#)
- [Du et al. \(2017\)](#)
- [Reiser et al. \(2020\)](#)

---

## DC4: High-resolution monitoring and forecasting of tropical cyclones

### Task description

With the objective of better understanding and forecasting the evolving structure of the cyclone (radii, intensity, inflow circulation) and its interaction with the ocean during and after its passage (i.e. the TC wake signatures), the task will exploit various observations co-located with TC tracks in order to estimate ocean-atmosphere parameters such as surface winds, waves, sea surface height (SSH), sea surface temperature (SST), sea surface salinity (SSS), ocean colour and upper ocean SST and SSS, during the TC lifecycle. Three subtasks will be explored: filtering ( $[t - inf, t]$ ), smoothing ( $[t - inf, t+inf]$ ) or forecasting ( $[t - inf, t-24h]$ ) the observations in order to estimate this set of variables at time  $t$ .

### Training Data

A first version (scheduled September 2025) will rely on the existing MAXSS databases (ranging from 2010 to 2020, references MAXSS 1, 2, 3 and 4) as training data. These databases aggregate multivariate ocean and atmosphere data from various sources (satellite and in situ observations, forecasts and reanalyses from atmosphere and ocean models), with different sampling, resolutions and noise will be used as training input. In a second phase (scheduled September 2026), this dataset will be extended and updated in 2025-2026 making it possible to improve this DC by integrating the latest data.

### Evaluation data and metrics

The evaluation will use a subset from the same datasets used for training. Representative TC events with relevant collocated data will be selected as evaluation data. Remaining data will be assigned to the training dataset. Evaluation metrics are of two types :

- Wind metrics : intensity spectral score, radius errors, wind profiles, ... RMSE, wrt SAR, wind direction error wrt scatterometer and SAR,
- Ocean metrics: scaling laws, RMSE and spectral score for surface waves, surface currents, SSH, SST and SSS signatures in the TC wakes, Brunt-Vaisala parameter and mixed layer depth changes.

### Baseline solutions

Baselines will be based on forecasts and reanalyses from wind and ocean models:

- Baselines for wind parameters: MAXSS 1 (see references section below), ERA5, ECMWF forecasts.
- Baselines for ocean parameters: SST, SSS, ocean colour merged L4, SSH merged AVISO, Interior ocean ARGO profiles and ISAS, Mercator-GLORYS re-analysis

### References

MAXSS 1 : <https://doi.org/10.12770/35002607-3546-412b-8c5d-9c182a16ffea>

MAXSS 2 : <https://doi.org/10.12770/447aa88f-0c0b-4607-afe2-9c77e95a14b8>

MAXSS 3 : <https://doi.org/10.12770/6c56bcde-050f-42eb-92b8-8e882e1f4db9>

MAXSS 4 : <https://doi.org/10.12770/cc0577e4-55d6-4aa9-a938-b4965be121ab>



---

## DC5: Marine Biodiversity Prediction

### Task description

The goal is to predict the abundance of multiple taxa (reef fish and plankton) and marine particulate matter across the ocean using socio-environmental predictors (e.g., physical and biogeochemical ocean parameters delivered by operational products, descriptors of the habitats and of socio-economic activities such as fishing pressure). Once tested and validated, using independent observations, the models will be used to extrapolate the spatiotemporal distribution of these taxa in unobserved parts of the ocean.

### Training datasets

**Response data:** abundance per size class of ~700 reef fish species in 20k surveys, biovolume of ~10 morphotypes of particulate matter at 30k locations, biovolume of ~20 taxa of plankton at 50k locations.

**Predictors:** bathymetry ([GEBCO](#)), climatologies of environmental variables ([World Ocean Atlas](#)), standard resolution satellite data ([CMEMS L4 ocean colour](#), [CMEMS L4 physics](#)), ocean reanalyses ([GLORYS](#)), high resolution satellite images (Sentinel-1 and 2), possibly high resolution biogeochemical model outputs ([CMEMS PISCES](#))

### Evaluation metrics / data

The response variables have distributions with very long tails (rare occurrences of large values) which make usual regression metrics (e.g. R2) inappropriate. L1 or L2 loss on log transformed data will be used instead, for a given target.

To define the metric for all targets, micro and macro averages will be used, to also consider targets of low abundance.

**Data:** For reef fish, data from 2006 to 2023 will be used to train the models while the 2024-2025 surveys will constitute independent evaluation datasets. For particulate matter and plankton, an evaluation dataset as independent of the training set as possible will be constructed by clustering observations in lat., lon., time space.

### Baseline solutions

The baseline solutions are classical machine-learning based models (e.g. gradient boosted trees, random forest) fitted (i) on local observations (not at the seascape level), and (ii) per individual target (not in a multivariate way).

- [code for Drago et al 2022](#): modelling of plankton biomass.
- [code for Haute et al \(in prep\)](#): modelling of fish abundance.

### References

- [Clements et al \(2022\)](#)
- [Drago et al \(2022\)](#)
- [Kaneko et al \(2023\)](#)
- [Schickele et al \(2024\)](#)

---

## Appendice B

### Submission templates in French and English

The docx version of the templates are available here: <https://www.ocean-climat.fr/Le-PPR/Actualites/Intelligence-artificielle-Appel-a-postdoc-et-data-challenges>

<b>Titre du sujet de postdoc</b>	
<b>Laboratoire(s) d'accueil</b> (si 2 laboratoires, précisez la durée dans chacun)	
<b>Organisme porteur du postdoctorat et contact administratif</b>	
<b>Data challenge(s) associé(s)</b>	

	Porteur.se 1	Co-porteur.se 2 éventuel
<b>Prénom Nom</b>		
<b>email</b>		
<b>Organisme</b>		
<b>Laboratoire</b>		
<b>Discipline(s) scientifique(s)</b>		

**Résumé du sujet de postdoc** (2000 caractères max):

**Mots-clés :**

## **Description du projet postdoctoral**

Cette partie devra comprendre les éléments suivants :

- Contexte et objectifs scientifiques
- Approche proposée
- Originalité et Innovation
- Contribution au(x) data challenge(s) visé(s)
- Calendrier des travaux

## **Conditions du postdoctorat**

Il s'agit ici de préciser :

- Les conditions matérielles et financières (budget, lieu d'accueil avec localisation géographique) de réalisation du postdoctorat
- L'expertise de l'équipe encadrante pour mener à bien le projet
- Les modalités de suivi de l'avancement des recherches du/de la postdoctorant.e
- Les mobilités prévues (durée, où, quand) et les collaborations nationales et internationales envisagées

## **Annexes:**

### **Bibliographie**

**Curriculum Vitae des porteur.ses du postdoctorat (2 pages max. par CV)**

---

<b>Title</b>	
<b>Hosting teams(s)</b>	
<b>Hosting organization for the postdoctoral fellowship (incl. administrative contact)</b>	
<b>Targeted data challenge</b>	

	<b>Principal Investigator 1</b>	<b>Principal Investigator 2 if any</b>
<b>First name and Last name</b>		
<b>email</b>		
<b>Organization</b>		
<b>Laboratory</b>		
<b>Scientific discipline(s)</b>		

**Summary of the postdoctoral project (2000 characters max):**

**Keywords :**

## **Description of the Postdoctoral Project**

This section should include the following elements:

- Context and scientific objectives
- Proposed approach
- Originality and Novelty
- Contribution to the targeted data challenge(s)
- Work Schedule/Timeline

## **Environment of the postdoctoral project**

This section specifies:

- the material and financial conditions (budget, a reception area with geographical location) for carrying out the postdoctoral work
- the expertise of the hosting team(s) for carrying out the postdoctoral work
- the procedures for monitoring the progress of the research conducted by the postdoctoral researcher
- planned mobility if any (duration, location, timing) and possible international collaborations

## **Annex:**

### **Bibliography**

**Curriculum Vitae of the principal investigator(s)** (2 pages max per CV)