LATE REQUEST FOR A SPECIAL PROJECT 2026–2028

MEMBER STATE: Belgium

Principal Investigator[1]: Victor Mangeleer

Affiliation: University of Liège, MAST / SAIL / FOCUS / AGO

Address: Bâtiment B5c, allée du 6 Août 19, 4000 Liège, Belgium

Other researchers: Luc VANDENBULCKE (luc.vandenbulcke@uliege.be), Gilles LOUPPE (g.louppe@uliege.be) and Mathurin CHOBLET (mchoblet@uliege.be)

Project Title: Black Sea Ensemble for Neural Emulator

To make changes to an existing project please submit an amended version of the original form.)

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SP** N/A	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2026	
Would you accept support for 1 year only, if necessary?	☑ YES	NO ☐

Computer resources required for project year:		**2026**	**2027**	**2028**
High Performance Computing Facility	[SBU]	3 300 000	0	0
Accumulated data storage (total archive volume)[2]	[GB]	55 000	0	0

EWC resources required for project year:		**2026**	**2027**	**2028**
Number of vCPUs	[#]	0	0	0
Total memory	[GB]	0	0	0
Storage	[GB]	0	0	0
Number of vGPUs[3]	[#]	0	0	0

[1] The Principal Investigator will act as contact person for this Special Project and, in particular, will be asked to register the project, provide annual progress reports of the project's activities, etc.
[2] These figures refer to data archived in ECFS and MARS. If e.g. you archive x GB in year one and y GB in year two and don't delete anything you need to request x + y GB for the second project year etc.
[3] The number of vGPU is referred to the equivalent number of virtualized vGPUs with 8GB memory.

Principal Investigator: Victor MANGELEER

Project Title: Black Sea Ensemble for Neural Emulator

Extended abstract

All Special Project requests should provide an abstract/project description including a scientific plan, a justification of the computer resources requested and the technical characteristics of the code to be used. The completed form should be submitted/uploaded at link Following submission by the relevant Member State the Special Project requests will be published on the ECMWF website and evaluated by ECMWF and its Scientific Advisory Committee. The requests are evaluated based on their scientific and technical quality, and the justification of the resources requested. Previous Special Project reports and the use of ECMWF software and data infrastructure will also be considered in the evaluation process. Requests exceeding 10,000,000 SBU should be more detailed (3-5 pages).

1. Scientific Context and Motivation

The Black Sea is one of the world's most oxygen-depleted enclosed seas. Its permanent anoxic zone extends from ~100 [m] to the seafloor and is expanding under the combined effects of climate warming, nutrient loading, and reduced ventilation. Seasonal coastal hypoxia events, i.e. periodic collapses of near-bottom oxygen, severely impact fisheries and coastal ecosystems. Despite their ecological and socioeconomic importance, these events remain difficult to monitor and forecast: direct in-situ observations are sparse, and no operational tool currently provides real-time estimates of the three-dimensional (3D) oxygen field.

This project addresses this gap by developing a neural emulator of the Black Sea. The emulator is a deep generative neural network trained to reconstruct and forecast the full 3D sea state (temperature, salinity, oxygen, nutrients) from satellite observations alone, providing a fast and physically consistent monitoring tool. This work is carried out in collaboration with Prof. Laure Zanna's group at New York University (NYU), a world-leading team in the application of machine learning to ocean modeling.

2. Plan of Action

The project is structured around three axes. **AXIS 1** (this proposal) generates a 16-member ensemble of numerical ocean simulations using the NEMO-BAMHBI coupled model. Each member is initialized with distinct initial conditions and integrated freely for 20 simulation years, providing a rich and diverse dataset for our neural network training. **AXIS 2**, led at NYU, trains the neural emulator on the full ensemble, develops a high-compression latent architecture for joint representation of physical and biogeochemical variables, and establishes rigorous validation diagnostics. **AXIS 3** uses the trained emulator to address three scientific questions:

(1) Can hypoxic events be detected from space using satellite observations alone?

(2) How do in-situ and satellite observations complement each other for 3D state estimation?

(3) Can historical deoxygenation trends be reconstructed from sparse historical records?

3. Numerical Model and Technical Characteristics

The simulations are performed with the following software stack:

- NEMO (v4.2.0, CeCILL license): Simulate sea physics dynamics at a 3 [km] resolution.
- XIOS (v2.5, CeCILL license): Asynchronous parallel I/O server, decoupled from NEMO.
- BAMBHI (v1.0, MIT license): Simulate sea biogeochemistry dynamics, online to NEMO.

The code is MPI-only (no GPU acceleration). It is actively used in the CMEMS Black Sea Monitoring and Forecasting Centre (operational production, daily updates). Performance on ECMWF infrastructure was validated by Luc Vandenbulcke in prior special projects on the ECMWF infrastructure.

4. Resource Justification

The 16-member ensemble is generated sequentially, one member at a time, each using 256 CPU cores for approximately 30 wall-clock days. More precisely:

Resource	Unit	Per Member	16 Members
CPU cores	cores	256	4096
Wall-clock time	days (total, sequential)	30	480
SBU	core-hours	184320	2949120
Output (archived)	TB	3.1	49.6

The SBU estimate (3 300 000 SBU) is grounded in the direct core-hour count (2,949,120 core-hours, see table above) and cross-checked against a comparable Black Sea NEMO-BAMHBI project by Luc Vandenbulcke (ECMWF BSENSE, 2022: 2,400,000 SBU for a Black Sea ensemble generation phase with the same code stack). The increase relative to that reference reflects the longer per-member integration (20 simulation years vs. 10 in the BSENSE project).