REQUEST FOR A SPECIAL PROJECT 2025–2027

MEMBER STATE: Italy

Principal Investigator[1]**:** Andrea Storto

Affiliation: CNR ISMAR (National Research of Italy, Institute of Marine Sciences)

Address: Via del Fosso del Cavaliere 100, I-00133 Rome, Italy

Other researchers: Giovanni Nunziante
Aqsa Riaz
Vincenzo de Toma

Project Title: Enhancing regional and global ocean data assimilation

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** itstor	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2025	
Would you accept support for 1 year only, if necessary?	**YES**	NO ☐

Computer resources required for project year:		**2025**	**2026**	**2027**
High Performance Computing Facility	[SBU]	9M	9M	9M
Accumulated data storage (total archive volume)[2]	[GB]	30	30	40

EWC resources required for project year:		**2025**	**2026**	**2027**
Number of vCPUs	[#]			
Total memory	[GB]			
Storage	[GB]			

[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.

Number of vGPUs[3]	[#]	1	1	1

Continue overleaf.

Principal Investigator: Andrea Storto

Project Title: Enhancing regional and global ocean data assimilation

Extended abstract

CNR-ISMAR is devoting several efforts in global and ocean data assimilation activities, thanks to a newly formed group on oceanic modelling and assimilation, targeting basin-scale ocean dynamics, stochastic ocean physics, and the impact of advanced data assimilation methods for historical (centennial) reanalyses of the ocean.

The overarching goal of this project, which is in full continuity with the previous Special Project run by the P.I. (spitstor-2019, spitstor-2022) is to allow us to experiment several new ensemble generation and assimilation techniques, which are well aligned with the priorities outlined by the ocean data assimilation community and may also contribute to advancing the ocean data assimilation science at European level.

The project focusses on several development tasks, which are detailed in the following section. CNR-ISMAR has two ocean model configurations based on NEMO, which cover, respectively, the global ocean (ORCA1 and ORCA025 configurations) and the North Atlantic – Arctic oceans (CREG025, CREG12 and CREG36 configurations) and are detailed in

https://www.ismar.cnr.it/en/infrastructures/modelling/oceanographic-modelling

The global ocean model configuration, in combination with variational data assimilation and several bias-correction schemes, is used within centennial reconstructions in the CIGAR (The Cnr ISMAR Global historicAl Reanalysis) reanalysis system (http://cigar.ismar.cnr.it ; https://doi.org/10.1038/s41467-024-44749-7).

Additionally, we run a regional coupled model, based on NEMO-WRF-HD, called MESMAR (Mediterranean Earth System model at ISMAR) whose details and relevant links are available at:

https://www.ismar.cnr.it/en/infrastructures/modelling/earth-system-models

The coupled model is used to assess increasingly complex coupled data assimilation schemes (see e.g. https://doi.org/10.5194/gmd-16-4811-2023).

With these systems, we plan to run several assessment experiments, where the individual configurations are chosen depending on the specific research action (see below).

All configurations use the NEMO version 4.0.7 (also in use at ECMWF for ORAS6-lie experiments) and, although upgrade to newer version may be considered during the project execution, we plan to run NEMO on the aa/ab/ac/ad clusters, which was already optimized during the previous special projects. In addition, however, we ask for a virtual GPU through the EWC cloud environment, as some research tasks may be strongly benefit for the use of GPU for AI training tasks. While we do have GPU availability in the local cluster at CNR-ISMAR, their use in the EWC will avoid heavy data transfer procedures and allow a smoother transition between the GPU-based AI training tasks and the CPU-based inference tasks within the NEMO and data assimilation software.

Project Tasks

The proposal is articulated in different tasks, which are detailed below.

1. **Enhancing and updating stochastic physics schemes for NEMO.** In this task, we will devote efforts to test new ways to add stochastic elements in NEMO. In particular, we have developed a stochastic scheme package for NEMO (STOPACK; https://doi.org/10.1002/qj.3990); the scheme includes perturbation of tendencies (SPPT), of model parameters (SPP) and stochastic modulation of kinetic energy backscattered to the ocean large scales (SKEB). STOPACK is being used by ECMWF, UKMO, and ECCC, within a number of projects (HE ACCIBERG; C3S2_601), and will be part of the schemes endorsed by the NEMO Working Group on Machine Learning and Uncertainty Quantification. The scheme needs to be updated to allow more efficient random field generation, which in turn will enable independent perturbations for any SPP parameter. In order to do so, we have implemented a Perlin noise random field generation capability to the scheme, which was found extremely efficient. The new random field generation needs to be assessed extensively through a number of experiments. We will primarily use the CREG025 and CREG12 configurations of NEMO to assess the benefit of independent perturbations for each SPP parameter, separately for ocean and sea-ice parameters. These tests, performed on the aa/ab/ac/ad cluster, will allow us to quantify the relative merits of different random field generation techniques in terms of computational performances and ensemble reliability metrics.

2. **Coupled data assimilation experiments.** The MESMAR system is formed by NEMO, WRF and HD models, and will be used to test different ideas to introduce elements of strong assimilation coupling in the coupled data assimilation system that is run for the Mediterranean Sea. So far, the system has been widely tested with a weakly coupled data assimilation scheme, where the ocean implements a variational scheme and the atmosphere a spectral nudging scheme; such a scheme has proven successful in preliminary studies, for instance for the representation of hurricane-like events in the Mediterranean Sea (medicines) where both data assimilation schemes concur to improve both the tracks and intensity of the medicine events. Here, we plan to test different cross-component balance operator in the ocean variational scheme to adjust the near-surface atmospheric fields and assess their benefits for reducing coupled imbalances during the first 1-2 days of forecasts. We plan to use EOFs (empirical orthogonal functions) and ANNs (artificial neural networks) to design and compare these cross-component balance operators. These operators will be first used off-line (i.e., as an additional step between the analysis and the forecasts steps), and later inserted online in the variational scheme, augmenting the data assimilation control vector to include the near-surface atmospheric parameters. This task includes running coupled experiments on the ECMWF cluster, but also training steps within the EWC vGPU.

3. **Data assimilation and bias correction for historical reanalyses**. Historical reanalyses are known to be mostly affected by biases and drifts, provided that the poor observing network can hardly correct model inaccuracies in early periods. The early periods reanalysis exercise is thus mostly a bias-correction problem. This task aims to advance our capability of bias-correcting model simulations for long-term and historical periods. We will first estimate the degree of bias in the CIGAR system and assess its uncertainty through the use of analysis increments aggregation and use of independent datasets. We will also analyse differences in early periods between the use of ERA5 and 20CRv3 as atmospheric forcing data for the ocean simulations. Next, the task will focus on two fundamental aspects for the production of historical reanalyses: i) bias correcting state-dependent systematic errors through the use of column neural networks to correct either the air-sea fluxes, or the temperature-salinity

profiles, using either nudging increments or variational data assimilation increments from preliminary experiments as training dataset, in combination with relevant predictors of the oceanic and atmospheric states; ii) test the impact of hybrid covariances in early periods, whose ensemble component is derived offline from a previous ensemble reanalysis (CIGAR). The previous ideas will be tested within the CIGAR reanalysis systems, in periods of observational network transitions, notably during the transition to the altimetry era (i.e. during the decade 1987-1997), or earlier during the XBT increase period (1955-1965). This task is expected to use both the ECMWF cluster for CPU-based NEMO simulations, and the EWC vGPU for training tasks.