SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2025
Project Title:	Machine learning techniques for ocean spin-up acceleration
Computer Project Account:	spitsozz
Principal Investigator(s):	Alessandro Sozza
Affiliation:	Institute of Atmospheric Sciences and Climate, National Research Council of Italy (CNR-ISAC)
Name of ECMWF scientist(s) collaborating to the project (if applicable)	Paolo Davini, Susanna Corti, Matteo Nurisso, Renata Coppo
Start date of the project:	January 2025
Expected end date:	December 2026

Computer resources allocated/used for the current year and the previous one

(if applicable)
Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)			15,000,000	3,162,952
Data storage capacity	(Gbytes)			20,000	9,800

Summary of project objectives

The MARTINI project aims to accelerate the spin-up of the EC-Earth4 climate model using both deterministic data-driven methods and machine learning (ML) techniques. By reducing the time and computational cost needed to reach climate equilibrium, the project seeks to improve the efficiency of long-term climate simulations. Key objectives include: (1) testing low-resolution configurations of EC-Earth4, (2) developing and validating spin-up acceleration strategies based on trend extrapolation, and (3) extending these approaches to ML methods such as ARX and GANs. The ultimate goal is to establish a robust framework for efficient and physically consistent model initialization.

Summary of problems encountered

Initial challenges included configuring the low-resolution version of EC-Earth4 on the ECMWF Atos HPC system, with particular focus on load balancing and optimal parallel layout. Strong and weak scaling tests were performed to ensure efficient resource usage. Long spin-up runs revealed technical issues in OIFS after several centuries of simulated time, likely due to improper memory handling or hardcoded accumulation of diagnostic arrays, ultimately leading to fatal crashes. NEMO outputs showed high-frequency wiggles in upper-ocean fields. Tuning is still ongoing to address these artefacts and improve model stability over long integrations. Efforts are also being dedicated to developing robust diagnostics to assess convergence toward equilibrium and to identify most effective metrics. Additionally, restart-field manipulation posed technical challenges, particularly in preserving vertical stability within the ocean model.

Summary of plans for the continuation of the project

The next phase will focus on consolidating the "stop-and-go" protocol by systematically testing deterministic spin-up acceleration techniques (e.g., linear trend extrapolation, EOF-based reconstruction). These methods will be benchmarked against standard long spin-up runs to quantify performance gains. An automated tool for managing restart manipulation and relaunching simulations will be developed and integrated within EC-Earth's workflow. Diagnostics to monitor equilibrium convergence will be refined and extended. Next, the implementation of ML-based strategies (e.g., ARX, GANs) will begin, using existing simulation data for training and validation. Finally, the most promising techniques will be scaled up and tested on higher-resolution EC-Earth4 configurations, paving the way for their potential application in CMIP7-class simulations.

List of publications/reports from the project with complete references

Alessandro Sozza, Paolo Davini, and Susanna Corti, Data-driven approaches for accelerating ocean spin-up in coupled climate simulations, EGU25-4573, https://doi.org/10.5194/egusphere-egu25-4573, EGU General Assembly 2025, Session ITS1.2/OS4.8 "Machine Learning for Ocean Science"

Summary of results

During the first six months of the MARTINI project, substantial progress has been made in setting up the EC-Earth4 climate model, conducting performance benchmarks, and initiating the development of spin-up acceleration strategies. The low-resolution EC-Earth4 configuration (TL63L31-ORCA2Z31) was successfully ported and tested on the ECMWF Atos HPC system.

Strong and weak scaling tests were conducted to assess the parallel performance of the coupled EC-Earth4 model, leading to an optimized layout with improved load balancing between the ocean (NEMO) and atmosphere (OIFS) components. A detailed analysis of domain decomposition was carried out, particularly for NEMO, exploring various MPI configurations while considering the geometric constraints of the ORCA2 grid and the coupling frequency with OIFS. The best performance was achieved when the number of MPI tasks assigned to NEMO and OIFS was kept approximately equal (i.e. with a ratio of order one), ensuring balanced throughput and minimizing idle time and I/O delays. To support this effort, a Python-based diagnostic tool was developed to automatically collect timing and memory usage metrics across simulations, facilitating the identification of bottlenecks and the tuning of launch configurations.

Long-duration spin-up simulations were launched using pre-industrial forcing, and preliminary results revealed technical issues in OIFS after several centuries of model time. The problem was traced to hardcoded array growth in the atmospheric component, which leads to memory overflows and simulation crashes. As a workaround, simulations are now segmented into shorter chunks of a few hundred years. Restart files from NEMO are used to initialize the ocean state, while the atmosphere is re-initialized each time from a standard 1990 state. This introduces a small inconsistency, but it is considered acceptable given that atmospheric dynamics have limited influence on the ocean spin-up process.

A prototype of the deterministic “stop-and-go” method was tested. This procedure involves stopping the simulation at fixed intervals (e.g., every 10 years), performing trend extrapolation on temperature and salinity fields, and restarting the model with modified initial conditions. Initial tests demonstrated technical feasibility, though additional refinements are needed to ensure vertical stability and consistency of the restart fields.

In parallel, a full post-processing and analysis suite was developed to evaluate model convergence toward equilibrium. Several diagnostics have been implemented, including bias evolution, radiative imbalance, and cost function minimization. A GitHub repository has been created to host these tools and ensure reproducibility.

In summary, the project is progressing as planned. The low-resolution configuration of EC-Earth4 (TL63L31-ORCA2Z31) is now fully operational in coupled mode and technically able to run multi-century simulations. Key bottlenecks and technical challenges have been identified and initial solutions have been implemented. The groundwork for spin-up acceleration strategies, including deterministic methods and automation tools, is well underway.