REQUEST FOR A SPECIAL PROJECT 2025–2027

MEMBER STATE:	ITALY
Principal Investigator:	Alessandro Sozza
Affiliation:	National Research Council, Institute of Atmospheric Sciences and Climate (CNR-ISAC)
Address:	Corso Fiume 4, 10133, Turin, Italy
Other researchers:	Paolo Davini (CNR-ISAC, Torino) Susanna Corti (CNR-ISAC, Bologna) Matteo Nurisso (CNR-ISAC, Torino)
Project title:	Machine learning techniques for ocean spin-up acceleration (MARTINI)

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** …………..	
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]	15,000,000	15,000,000	\
Accumulated data storage (total archive volume)[2] [GB]	20,000	45,000	\

EWC resources required for project year:	**2025**	**2026**	**2027**
Number of vCPUs [#]	0	0	/
Total memory [GB]	0	0	/
Storage [GB]	0	0	/
Number of vGPUs[3] [#]	0	0	0

Continue overleaf.

Principal Investigator:	Alessandro Sozza
Project Title:	Machine learning techniques for ocean spin-up acceleration (MARTINI)

Extended abstract

Overview & Objectives

Within MARTINI, we aim to explore the best strategies to accelerate the spin-up towards climate equilibrium using the newly developed version4 of the EC-Earth4 global climate model (GCM). By conducting a series of coupled ocean-atmosphere simulations at low and high resolution, we will leverage diverse techniques based on deterministic protocols and machine-learning (ML) algorithms both aiming at reducing the amount of computational resources required to spin up a GCM. MARTINI aims to provide a deeper understanding of the relaxation towards equilibrium in the coupled climate system, offering insightful guidance for future model spin-up strategies and sustainable usage of computational resources. Main objectives of the project are:

1) to conduct validation tests and tuning of the low-resolution EC-Earth4,
2) to implement and test diverse deterministic and ML-based techniques of spin-up acceleration to expedite the model's convergence to equilibrium,
3) to examine the asymmetry in model responses to cold/hot initial conditions, exploring different pathways to equilibrium.

Overall, MARTINI seeks to bridge the knowledge gap in the spin-up strategies, ensuring computational efficiency without compromising model accuracy and precision.

Introduction

In the framework of any complex couple climate simulation, ranging from millennial paleoclimate integrations up to CMIP DECK protocol, a crucial element to the accuracy and reliability of these simulations is the "spin-up" process. Indeed, the spin-up phase in ocean and climate modelling allows the model to evolve from initial conditions to a state of equilibrium, ensuring that variables like temperature, salinity, ocean currents, and atmospheric conditions stabilise realistically. In an ideal model, this will be represented by a simulation without any drift (in absence of energy leaks or sources). Without a proper spin-up, simulations could produce misleading results since a remnant drift might affect the interpretation of the results (Bryan, 1984).

Various methodologies have been developed to optimise the spin-up process. Early approaches relied on iterative techniques, requiring significant computational resources and time (Bryan, 1984). A notable advancement was the adoption of Newton-Krylov methods, combining Newton's method for solving nonlinear equations with Krylov subspace methods for solving large linear systems. These techniques significantly reduced the computational burden by accelerating convergence to equilibrium states (Bernsen et al., 2008; Bernsen et al., 2010; Khatiwala, 2008).

In recent years, machine learning (ML) has revolutionised spin-up methodologies. ML algorithms, particularly deep learning models, have been used to predict and expedite equilibrium states based on historical data and patterns (Dueben & Bauer, 2018; Sun et al., 2023). These advanced techniques can identify complex relationships within the model, providing faster and more accurate spin-up processes. Hybrid approaches that combine traditional numerical methods with ML techniques are also being explored.

Despite these advancements, a significant knowledge gap remains in applying ML techniques to Earth System Models (ESMs) and General Circulation Models (GCMs) at low resolution. The challenge is to reduce computational resources without sacrificing accuracy and precision. While intermediate complexity models like PlaSim and SPEEDY-NEMO have shown promise (Angeloni et al., 2020; Ruggieri et al., 2024), there is a pressing need to extend these advancements to more comprehensive ESMs and GCMs. This gap underscores the necessity for further research to harness ML's potential in enhancing the efficiency of these models, ensuring high-fidelity simulations with reduced computational demands.

This project proposal aims to address this gap by developing and integrating traditional iterative approaches and more advanced ML techniques into low-resolution ESMs and GCMs. The goal is to significantly reduce computational resources while maintaining or improving the accuracy of climate and ocean simulations.

Methodology

The MARTINI project aims to optimise the spin-up process in Earth System Model (ESM) by leveraging advanced machine learning (ML) techniques, traditional spin-up methods, and innovative strategies for initial conditions. The methodology encompasses a structured approach to simulations, validation, and integration of traditional and ML approaches for the accelerating spin-up phase.

The Earth System Model (ESM) adopted in this project is EC-Earth4, the successor to EC-Earth3, which participated in the CMIP6 campaign (Döscher et al., 2021) and is now adapted for CMIP7. It’s designed to simulate the climate system by integrating the atmosphere and land surface model through OpenIFS (OIFS cy43r3), a modified version of ECMWF's Integrated Forecasting System (IFS), coupled with the ocean and sea-ice model (NEMO v4.2). Thanks to adaption of the XIOS server, developed within the EC-Earth consortium, the model output is directly written into cmor-like NETCDF format. Additionally, EC-Earth4 can incorporate modules for simulating dynamic vegetation, atmospheric chemistry, aerosols, ocean biogeochemistry, the carbon cycle, and dynamical ice sheets.

EC-Earth4 low resolution setup

We will primarily utilise the low-resolution version of EC-Earth4 (ECE-FAST), employing a linear grid TL63 grid for the atmosphere (~2.8° and ~300 km of horizontal resolution at the equator) and ORCA2 grid for the ocean (~2° and ~222 km of horizontal resolution). Both grids feature 31 vertical levels. The model has been already ported on ECMWF Atos HPC2020 where a few technical scalability tests have demonstrated the efficiency of ECE-FAST (see Fig.1).

The cost of the ECE-FAST configuration is estimated to be around 200 simulated years per day (SYPD) and 50 core-hours per simulated year (CHPSY), making it an excellent configuration to explore different spin-up strategies, lasting thousands of years with relatively low usage of resources. Still, some issues in numerical stabilities seem to be arising after long integrations and must be addressed with proper testing. Similar efforts are needed in paleoclimatic simulations, fostering synergies with the proposed Special Project EPOCHAL, coordinated by Dr. M. Nurisso.

Validation and Tuning

Our team has experience running simulations on Atos ECMWF supercomputer, where ECE-FAST is already operational although only in a rather "technical"configuration. However, validation and tuning will be carried out in a companion proposed special project (EPOCHAL, lead by Dr. M. Nurisso) to ensure model accuracy and stability. For our goals, however, even a model which is not tuned is not affecting the strategy and development of our approach.

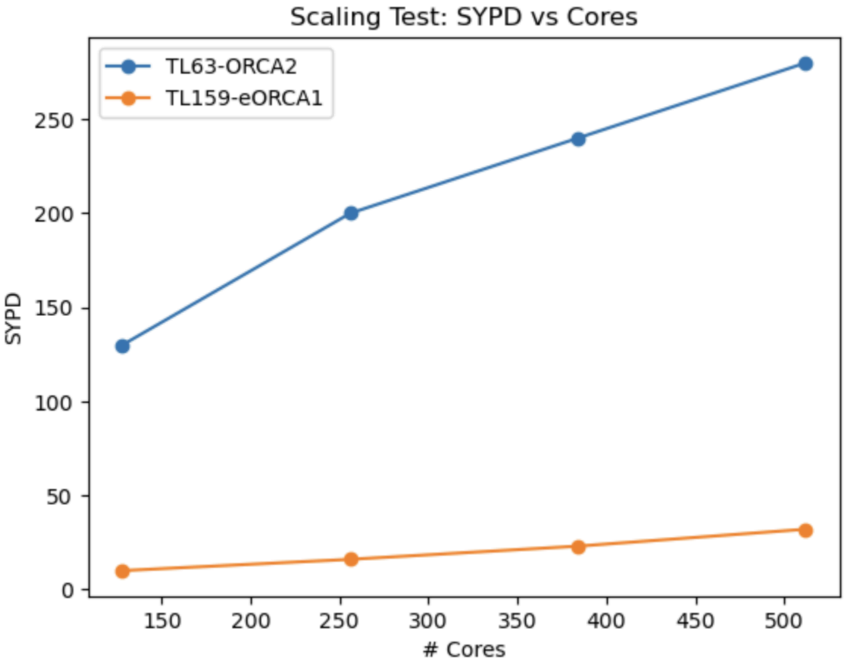

Fig.1: Scalability test for EC-Earth4 low resolution TL63L31-ORCA2Z31 configuration (ECE-FAST), compared with an higher resolution TL159L91-eORCA1Z75, showing simulated years per day (SYPD) vs number of cores. For reference, a 2,000-year simulation runs in 10 days.

Once a stable model configuration is reached, a minimal CMIP6 DECK simulation following CMIP6 protocol (Eyring et al. 2016) is planned:

- 5,000-year pre-industrial “piControl”
- 5,000-year “historical”

These will serve as baseline simulations and benchmarks for assessing the effectiveness of accelerated spin-up techniques.

Cold/Warm Scenarios

During this initial phase, utilising the capabilities of the low-resolution ECE-FAST, we will investigate the implications of different initialisations, either warmer or cooler than the equilibrium state. To achieve these initial conditions, we will employ two distinct methods. The first method involves starting from rest but rescaling all temperatures by a multiplicative factor. For the warm initial condition, we will scale the initial temperature profile by a factor greater than one, while for the cold initial condition, we will scale it by a factor less than one. Salinity profiles will be adjusted reciprocally to maintain density stability. The second method involves running simulations under altered CO2 scenarios: the "4xCO2" scenario to create a warmer initial state, and the "1/4xCO2" scenario to generate a cooler initial state. Once equilibrium is reached in these scenarios, we will turn off the additional forcing and observe which simulations return to the former equilibrium state first. Two simulations are planned, starting from the equilibrium of the “piControl”:

- 2,000-year warm scenario
- 2,000-year cold scenario

This approach will help us understand the transient behaviour and stability of the system under different thermal initializations, and elucidate the potentially asymmetrical responses of the climate system to various pathways towards equilibrium.

Spin-Up Strategies (SUS)

Achieving equilibrium in ocean models involves addressing the varying time scales of different ocean layers. Equilibrium is primarily constrained by radiative fluxes directed toward the ocean surface. Consequently, the mixed layer equilibrates within ~1,000 years, while the deep ocean can take up to ~5,000 years. Although small, the imbalance in deep-ocean heat uptake keeps the ocean in a quasi-steady state for an extended period (Li et al. 2013). This understanding is crucial for developing effective spin-up techniques. To address these challenges, our project will explore various spin-up strategies.

SUS1: Deterministic Approaches

The idea is to run small chunks of simulations (e.g., 10 years), detect long-term trends, and then project these trends into the future to create new initial conditions for the ocean (please keep in mind that for all the approaches we are going to test in the entire project, the atmosphere remains untouched since its contribution to the energy budget is limited). The exact extension of the projection of the trend must be defined according to a procedure of trial and error, and it might be expected to be dependent on the magnitude of the trend itself. The model is then run again for another small chunk to adjust for potential instabilities, and this process is repeated iteratively until equilibrium is achieved. Initially, the focus will be on monitoring and controlling temperature and salinity (e.g., density). Different protocols can be adopted, tested, and compared to study their responses and determine the most efficient method:

- Linear Regression: This can be applied locally at each grid point or in subregions which share similar trends. However, a prerequisite to this approach is the identification and classification of oceanic regions based on common features such as temperature and density, or through basin and layer divisions, exploiting the advantages of dimensionality reduction.
- Empirical Orthogonal Functions (EOF): Applying EOF analysis to obtain new initial conditions, projecting the dominant modes of variability in the system in the future and then rebuilding the new status of the system.

One big issue to be addressed is the exclusion of internal variability which might alter the long-term oceanic trend. In this sense filtering done with EOF might be a good option. Multiple approaches will be developed based on the above two concepts, trying to develop a robust code that takes care of modifying the initial condition of the ocean in a consistent way. For example, given that trend projection might generate unstable conditions in the vertical column of water, a filtering to ensure vertical stability might be added.

SUS2: Machine learning Approaches

By the project's end, we aim to develop machine learning techniques based on:

- Auto-regressive models with exogenous inputs (ARX). These techniques are able to emulate and enhance the performance of traditional deterministic approaches such as linear regression and empirical orthogonal functions (EOF) in accelerating the time evolution of the climate system. By training the model with a set of auxiliary variables, such as radiative forcing and heat content, it will be possible to effectively predict the relaxation towards equilibrium and expedite the spin-up phase of climate models.
- Generative Adversarial networks (GAN). Other approaches may focus on harnessing the ML algorithm's capability to effectively mine extensive yet sparse datasets. This involves establishing intricate relationships between variables, such as temperature changes and radiative flux, to facilitate the mapping of initial states to final equilibrium states based on specific radiative forcing and initial conditions. Crucially, these methods aim to achieve this mapping without relying on the temporal evolution of the Earth System Model (ESM).

Regarding diagnostics, an essential role is devoted in assessing the distance from equilibrium using several key metrics. The Gregory plot is used to gain insights into how effectively climate models simulate the balance between incoming and outgoing radiation across different layers of the Earth system (Gregory et al., 2004). This diagnostic tool correlates the radiative net flux with temperature changes, traditionally applied to the top of the atmosphere (TOA) but now extended to incorporate oceanic variables like temperature and salinity. Additionally, metrics such as forecast error and cost functions provide further indicators of model performance, particularly when compared against the equilibrium state of the control/reference simulations. These diagnostics collectively contribute to refining and validating the accuracy of climate model predictions.

A plan for the "SUS" simulations has been established. Although the convergence of the proposed acceleration methods is not yet known, we estimate an upper bound of 500 years per simulation. We will conduct simulations to reproduce both the piControl and the historical scenarios using the four types of techniques proposed (resulting in a total of 8 simulations and 4,000 simulated years):

- Two 500-year simulations of SUS1-LR “linear regression”
- Two 500-year simulations of SUS1-EOF “empirical orthogonal functions”
- Two 500-year simulations of SUS2-ARX “autoregressive model”
- Two 500-year simulations of SUS2-GAN “generative adversarial network”

Once tested for the low resolution, the developed techniques will be applied to a higher resolution configuration, ideally the one that will be used for EC-Earth4 in the CMIP7 DECK integrations. It is though unlikely we will be able to provide an alternative/faster spinup than the usual one that will be used by the EC-Earth4 consortium, but the official high resolution spin up will be fundamental validation for the approach developed within MARTINI. So, in the second part of the project will utilise a TL255L91-eORCA1Z75 grid (~75 km and 1-degree of horizontal resolution, ~1,200 CHPSY) or whichever configuration will be used for running CMIP7 simulation by the EC-Earth4 consortium.

The outcomes and performance of protocols described in this project (e.g., linear regression, EOF, and ML techniques) are likely to be model-dependent and resolution-dependent. However, these methods offer a generalizable framework that can be employed to expedite equilibrium predictions across different climate models and resolutions. By systematically evaluating their effectiveness, we aim to establish a robust methodology that accelerates the spin-up phase.

Workflow

The project timeline is organised to ensure systematic development, testing, and validation:

- Month 1-3: Validation of ECE-FAST configuration.
- Month 3-6: Tuning of ECE-FAST for CMIP6. Production of the reference dataset (“piControl”, “historical” simulations & “cold/warm” scenarios).
- Month 3-12: Implementation and execution of the deterministic approaches (SUS1).
- Month 9-18: Implementation of ML techniques (SUS2).
- Month 12-23: Completion of the simulations using all accelerated spin-up techniques. Upscaling to higher resolutions.
- Month 6-24: Post-processing data analysis, data reduction and transfer to CNR-ISAC cluster machines for more detailed analysis.

Resources and Technical Development

Given the details of the model performance and the proposed plan of simulations that will be performed, we summarise here the resource budget estimated for this special project:

- **"piControl" and "historical" (TL63L31 - ORCA2Z31):** two 5,000-year simulations, for a total of 10,000 years, consuming 5 million SBU and generating 13 TB of data.
- **"Cold/Warm" scenarios (TL63L31 - ORCA2Z31):** two 2,000-year simulations, for a total of 4,000 years, consuming 2.5 million SBU and 5.5 TB of storage.
- **"SUS" simulations (TL63L31 - ORCA2Z31):** eight 500-year simulations for a total of 4,000 years, consuming 2.5 million SBU and 5.5 TB of storage.
- **"SUS" simulations at higher resolution (TL255L91 - eORCA1Z75):** two 500-year simulations, consuming 20 million SBU and storing 21 TB of data.

Storage of simulations is estimated to be of 1.3 GB/year (ECE-FAST) and ~21 GB/year for the higher resolution (TL255L91-eORCA1Z75). However, efforts in the direction of a massive data reduction policy will be pursued in order to lighten the data analysis and the occupied storage space. For example, one solution is to save only yearly output since we are not interested in monthly output but rather in the long term trend of the climate simulations.

Overall, we aim at publishing a paper on low resolution configuration ECE-FAST, showing the validation and tuning of the model, and at least one peer-reviewed article detailing the different spin-up strategies explored in the MARTINI project.

References

Angeloni, M., Palazzi, E., & von Hardenberg, J. (2020). Evaluation and climate sensitivity of the PlaSim v. 17 Earth System Model coupled with ocean model components of different complexity. *Geoscientific Model Development Discussions*, *2020*, 1-23

Bernsen, E., Dijkstra, H. A., & Wubs, F. W. (2008). A method to reduce the spin-up time of ocean models. *Ocean modelling*, *20*(4), 380-392.

Bernsen, E., Dijkstra, H. A., Thies, J., & Wubs, F. W. (2010). The application of Jacobian-free Newton–Krylov methods to reduce the spin-up time of ocean general circulation models. *Journal of Computational Physics*, *229*(21), 8167-8179.

Bryan, K. (1984). Accelerating the convergence to equilibrium of ocean-climate models. *Journal of Physical Oceanography*, *14*(4), 666-673.

Dueben, P. D., & Bauer, P. (2018). Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, *11*(10), 3999-4009.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., ... & Zhang, Q. (2021). The EC-earth3 Earth system model for the climate model intercomparison project 6. *Geoscientific Model Development Discussions*, *2021*, 1-90.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937-1958.

Fabiano, F., Davini, P., Meccia, V. L., Zappa, G., Bellucci, A., Lembo, V., ... & Corti, S. (2024). Multi-centennial evolution of the climate response and deep-ocean heat uptake in a set of abrupt stabilisation scenarios with EC-Earth3. *Earth System Dynamics*, *15*(2), 527-546.

Fletcher, C. G., McNally, W., Virgin, J. G., & King, F. (2022). Toward Efficient Calibration of Higher‐Resolution Earth System Models. *Journal of Advances in Modeling Earth Systems*, *14*(7), e2021MS002836.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., ... & Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical research letters*, *31*(3).

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P. A. B., Caron, L. P., Castrillo, M., ... & Wyser, K. (2020). HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR. Description, model performance, data handling and validation. *Geoscientific Model Development Discussions*, *2020*, 1-37.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., ... & Zaehle, S. (2016). C4MIP–The coupled climate–carbon cycle model intercomparison project: Experimental protocol for CMIP6. *Geoscientific Model Development*, *9*(8), 2853-2880.

Khatiwala, S. (2008). Fast spin up of ocean biogeochemical models using matrix-free Newton–Krylov. *Ocean Modelling*, *23*(3-4), 121-129.

Khatiwala, S. (2024). Efficient spin-up of Earth System Models using sequence acceleration. *Science Advances*, *10*(18), eadn2839.

Li, C., von Storch, J. S., & Marotzke, J. (2013). Deep-ocean heat uptake and equilibrium climate response. Climate Dynamics, 40, 1071-1086.

Liao, C., Lu, X., Huang, Y., Tao, F., Lawrence, D. M., Koven, C. D., ... & Luo, Y. (2023). Matrix Approach to Accelerate Spin‐Up of CLM5. *Journal of Advances in Modeling Earth Systems*, *15*(8), e2023MS003625.

Lofverstrom, M., Fyke, J. G., Thayer‐Calder, K., Muntjewerf, L., Vizcaino, M., Sacks, W. J., ... & Bradley, S. L. (2020). An efficient ice sheet/Earth system model spin‐up procedure for CESM2‐CISM2: Description, evaluation, and broader applicability. *Journal of Advances in Modeling Earth Systems*, *12*(8), e2019MS001984.

Pudig, M. P., Yung, C. K., Zika, J. D., & Holmes, R. M. (2023). Rectified Ocean Heat Uptake from Oscillatory Surface Forcing. *Journal of Climate*, *36*(8), 2663-2680.

Qu, Y., Maksyutov, S., & Zhuang, Q. (2018). An efficient method for accelerating the spin-up process for process-based biogeochemistry models. *Biogeosciences*, *15*(13), 3967-3973.

Rose, B. E., & Rayborn, L. (2016). The effects of ocean heat uptake on transient climate sensitivity. *Current climate change reports*, *2*, 190-201.

Ruggieri, P., Abid, M. A., García-Serrano, J., Grancini, C., Kucharski, F., Pascale, S., & Volpi, D. (2024). SPEEDY-NEMO: performance and applications of a fully-coupled intermediate-complexity climate model. *Climate Dynamics*, 1-19.

Sanderson, B. M., Wehner, M., & Knutti, R. (2017). Skill and independence weighting for multi-model assessments. *Geoscientific Model Development*, *10*(6), 2379-2395.

Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., ... & Romanou, A. (2016). Inconsistent strategies to spin up models in CMIP5: Implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, *9*(5), 1827-1851.

Sun, Y., Goll, D. S., Huang, Y., Ciais, P., Wang, Y. P., Bastrikov, V., & Wang, Y. (2023). Machine learning for accelerating processes‐based computation of land biogeochemical cycles. *Global Change Biology*, *29*(11), 3221-3234.

Xia, J. Y., Luo, Y. Q., Wang, Y. P., Weng, E. S., & Hararuk, O. (2012). A semi-analytical solution to accelerate spin-up of a coupled carbon and nitrogen land model to steady state. *Geoscientific Model Development*, *5*(5), 1259-1271.

Yool, A., Palmiéri, J., Jones, C. G., Sellar, A. A., De Mora, L., Kuhlbrodt, T., ... & Rumbold‐Jones, S. (2020). Spin‐up of UK Earth system model 1 (UKESM1) for CMIP6. *Journal of Advances in Modeling Earth Systems*, *12*(8), e2019MS001933.