EMI R&D PROJECT FINAL REPORT

All the following mandatory information needs to be provided.

Project Title:	Speeding up ocean spin-up using stochastic parametrisations
Computer Project Account:	spgbchri
Start Year - End Year :	2023 - 2026
Principal Investigator(s)	Dr Hannah Christensen
Affiliation/Address:	University of Oxford Atmospheric, Oceanic and Planetary Physics Clarendon Laboratory Sherrington Rd Oxford OX1 3PU
Other Researchers (Name/Affiliation):	Dr Kristian Strommen (Oxford) Dr Robert (Bobby) Antonio (Oxford) Dr Pablo Ortega (Barcelona Supercomputing Centre) Dr Valentina Sicardi (Barcelona Supercomputing Centre)

The following should cover the entire project duration.

Summary of project objectives

(10 lines max)

The primary aim of this project was to use recent advances in machine learning (ML) to speed up the ocean spin-up in a coupled earth system model, to improve the quality of climate simulations as part of the EERIE (European Eddie-Rich ESMs) project. Our approach to do this was by coupling a ML emulator of the atmosphere to a standard ocean model (NEMO), to significantly reduce the compute time required per year of spin up. Our first task was to therefore construct a coupling module between an ML atmosphere (ACE2) running on GPU and NEMO running on multiple CPUs. Having created a scheme to couple these models and stabilised the coupled model, the final task was to produce spun-up ocean states using this hybrid model, and investigating the quality of these states by running a conventional climate model (EC-Earth3) using the spun-up state and assessing the drift and biases. Secondary objectives were to use this setup to probe the physical realism of the ML emulator, and to assess how well it performs when coupled to an 'out-of-sample' ocean that it has not been fine-tuned to interact with.

Summary of problems encountered

(If you encountered any problems of a more technical nature, please describe them here.)

Initially we had difficulty finding an ML atmospheric emulator that was suitable for the task; we began by experimenting with GraphCast, since at the beginning of the project this was one of the few publicly-available models with the correct surface variables for ocean interaction. Our attempts at coupling GraphCast to the sea surface temperature (SST), by creating a model that learns corrections based on SSTs, did not demonstrate enough evidence of ocean-atmosphere coupling. We therefore pivoted to experimenting with two models which had become available in the meantime; GenCast and ACE2, both of which are much more stable at longer rollouts than GraphCast and accept SSTs as input. Whilst GenCast showed stability and reasonable skill at seasonal forecasting, we still encountered stability issues after around 4 months, which motivated us to use ACE2 instead.

We also encountered challenges of how to couple ACE2 to NEMO, as they operate in different environments (CPU vs GPU, which meant that it was not possible to make the two models interact in the same MPI environment in the HPC) and with different programming languages (FORTRAN and Python). To our knowledge, coupling models in this way was not something that had not been tackled before, and so there was no existing framework that we could base our model on. We resolved this by using polling mechanisms and passing files through shared storage, but we believe there should be more elegant solutions.

We also encountered errors in the final spin-up simulations, where NEMO restart files were not being loaded correctly; this led to the ACE2-NEMO spin-up simulations being invalid. Fortunately, through applying to a late project we were able to rerun the simulations in time. A different problem occurred during simulations run by our colleagues at Barcelona Supercomputing Centre, who were running a conventional model (EC-Earth3) using outputs from our model. This has limited the number of years of evaluation we currently have to evaluate the spin-up, however we plan to extend these simulations in future work.

Experience with the EMI R&D Project framework

(Please let us know about your experience with administrative aspects like the application procedure, progress reporting etc.)

Application for late units was an efficient and easy process, which enabled us to respond quickly to discovering problems in our initial spin-up runs, in time to reach a deliverable milestone.

Progress reporting has not been too much overhead, perhaps because the template documents are simple and not too prescriptive. Also we appreciate the option of sharing an existing report in this summary rather than replicating the same thing in two reports.

Adding additional users to the project was also straightforward and did not take too long.

Many thanks for supporting our research through providing computer resources

Summary of results

(This section should comprise up to 10 pages, reflecting the complexity and duration of the project, and can be replaced by a short summary plus an existing scientific report on the project.)

Below is a short summary, please see attached document (Deliverable 12.1 from the Horizon Europe EERIE project) for a more in-depth scientific report.

Achieving efficient and physically consistent ocean spin-up remains a major computational bottleneck in coupled climate modelling. This is especially the case for computationally expensive eddy-rich Earth System Models. To address this problem, we have presented ACE2-NEMO, the first coupling of a pure machine-learnt atmospheric emulator (ACE2) with a dynamical ocean model (NEMO). Apart from changes to the sea ice fluxes, no fine tuning of either component is performed. This creates a unique test of the robustness of the ACE2 atmospheric emulator to out-of-sample inputs and provides a novel test of how well the machine learning emulator has learned the appropriate feedbacks to force a dynamical ocean. It also presents an unexplored approach to integrating machine learning models with dynamical ocean models. Given the relative sparsity of ocean data, integrating with ocean models that model the physical equations may provide an advantage over ‘full' machine-learning-based ocean emulators in how faithfully they can represent long-term ocean variability.

Achieving the coupling requires overcoming several technical challenges; the models run in different environments, in different programming languages, and the atmospheric model does not produce all of the fluxes required by the ocean. Our solution creates a bridge between the two environments, and leverages the existing OASIS coupling infrastructure together with the AirSeaFluxCode library to achieve realistic coupling. Through developing the coupling, several problems were identified relating to the ACE2 fluxes over sea ice; the sensible heat fluxes produced by ACE2 over sea ice did not seem to be consistent with the temperature differences, and there were large drifts in sea ice volume and sea surface temperature. These problems were solved by using simple bulk formulae to calculate heat fluxes over ice.

The resulting model is stable over multidecadal timescales and shows realistic mean states and fluxes comparable to those in EC-Earth3 control simulations. These results demonstrate that the machine learning emulator is capable of driving a dynamical ocean model in a physically consistent manner for long integrations. This indicates that ACE2–NEMO is a viable option for replacing (or partially replacing) the long coupled spin-up simulations needed to initialise EC-Earth3.

We estimate the computational efficiency of the ACE2–NEMO configuration relative to EC-Earth3 and find a factor of three reduction in energy usage at this particular resolution. This is due to the extremely high efficiency of running the ACE2 atmospheric emulator compared to the dynamical IFS atmospheric model used in EC-Earth3. This would enable significantly longer integrations for a fixed computational budget, which could improve the spun-up state used for NEMO initialisation.

We have also defined and implemented spin-up protocols in which the ocean undergoes a first equilibration using the ACE2–NEMO system, to produce the initial conditions to subsequently initialise EC-Earth3. Due to technical issues in EC-Earth that were only recently overcome, the testing of this approach has been done with relatively short simulations, which are insufficient to determine its efficiency and suitability for spin-up acceleration purposes. An extension of those experiments is currently underway.

We note some limitations of the study. The initial conditions used for ACE2-NEMO are warmer than those used for EC-Earth3, which may introduce additional biases in the coupled system. The efficiency of the model is limited by the flux calculations, since they are not optimised for this application; significant speed-up of this module would therefore be expected with the appropriate optimisation. This additional efficiency gain would only strengthen the potential of this hybrid approach as a practical method to accelerate ocean spin-up in dynamical climate models.

Our findings motivate further investigation of this approach, particularly in the context of longer integrations and higher-resolution configurations.

List of publications/reports from the project with complete references

Antonio, B., Strommen, K., & Christensen, H. M. (2026). Seasonal forecasting using the GenCast probabilistic machine learning model. *Climate Dynamics*, *64*(4), 148

Antonio, B., Strommen, K., Ortega, P., & Christensen, H. M. (2026). ACE2-NEMO: Coupling an ML atmospheric emulator to a full-depth dynamical ocean model. *arXiv preprint arXiv:2603.28704*.

Antonio, B., Strommen, K., & Christensen, H. M. (2026). Role of the ocean for fast atmospheric evolution revealed by machine learning. *arXiv preprint arXiv:2602.01904*.

EERIE Deliverable 12.1 (not yet published)

Future plans

(Please let us know of any imminent plans regarding a continuation of this research activity, in particular if they are linked to another/new EMI R&D Project.)

The first priority is to extend the duration of the ACE2-NEMO and EC-Earth3 spin-up experiments to better understand the convergence behaviour. Our results indicate that the current simulations are too short to robustly assess equilibration, particularly in the deep ocean where adjustment timescales are very long. Within the final months of 2026 we intend to produce longer simulations with ACE2-NEMO, before branching off EC-Earth3, to quantify the extent to which we can reduce drift.

The simulations presented have used ACE2 and NEMO at 1 degree resolution. A key question is how applicable the framework is to the eddy-rich models used and developed in EERIE. The resolution of current ML atmospheric emulators is relatively coarse (~ 1 degree) because they have been trained on ERA5. This means that coupling an eddy-rich ocean to a resolution-matched atmospheric emulator is not currently possible. Instead, an interesting avenue of future research is to couple a coarse atmospheric emulator to a high-resolution ocean model. This would test the spin-up approach in a modelling regime closer to that used in EERIE. We note that for the IFS-NEMO model used in EERIE (with Tco1279-eORCA12 grids), the atmosphere and ocean components have a similar computational cost: replacing the IFS with an emulator could therefore achieve speed-ups for a factor of 2 at these resolutions. In addition, the development of future high resolution reanalyses (e.g. ERA6 at 0.1 degree) is expected to enable future training of higher-resolution ML atmospheric models, which could be explored in this framework. Work using higher resolution ocean models and atmospheric emulators will be explored in future projects.

EERIE

European Eddy-Rich ESMs

Using machine learning to achieve more efficient coupled ocean spin-up

Deliverable ID	D12.1
Work Package Reference	WP12
Issue	Version 1
Due Date of Deliverable	31/04/2026
Submission Date	29/04/2026
Dissemination Level	Public
Lead Partner	UOXF
Contributors	Bobby Antonio Hannah Christensen Pablo Ortega Valentina Sicardi
Grant Agreement No	101081383
Call ID	HORIZON-CL5-2022-D1-02-02

Funded by the European Union

UK Research and Innovation

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Prepared by	Reviewed by	Approved by
Bobby Antonio	Hannah Christensen	EERIE Coordination Team (ECT)
	Pablo Otega	

Issue	Date	Description	Author
1.0	17/04/2026	First issue of the deliverable	Bobby Antonio University of Oxford
			Hannah Christensen University of Oxford
			Pablo Ortega Barcelona Supercomputing Center
			Valentina Sicardi Barcelona Supercomputing Center

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

Index

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

Executive Summary

Coupled climate simulations require a stable initialisation of the ocean state. This is essential, as insufficient spin-up leads to model drift and biases that contaminate the forced signals, hindering their interpretation. However, ocean spin-up is computationally expensive, with very long simulations required to reach equilibrium. This is a key problem for eddy-rich Earth System Models (ER-ESMs) due to their computational cost, which can be prohibitive for high-resolution configurations. We therefore need to explore approaches to reduce the computational cost of ESM spin-up.

Here we investigate the use of machine learning (ML) to accelerate coupled ocean spin-up. The approach we take is to replace the atmospheric component of a conventional climate model with a fast, data-driven emulator. Specifically, we couple the ACE2 atmospheric emulator to the physical NEMO ocean model, forming a hybrid ML–physics system (ACE2–NEMO). This is expected to enable longer (or more efficient) spin-up integrations.

We first describe the approach taken to compute and communicate the necessary fluxes to couple NEMO and ACE2. We show that the coupled ACE2–NEMO system is stable over multi-decadal timescales: it produces realistic large-scale fields, including sea surface temperature, heat fluxes, and variability patterns, when compared to a reference EC-Earth3 configuration, which couples NEMO to a fully dynamical atmosphere.

We estimate the improvements in computational efficiency. ACE2-NEMO runs at a rate of around 10 simulated years per day (SYPD), compared with around 12 SYPD for NEMO alone using the same resources. EC-Earth3 requires around three times as many node-seconds per simulated day as NEMO alone: the ACE2–NEMO configuration is therefore approximately three times faster and uses one third of the energy per simulated day compared to EC-Earth3. This energy reduction would enable significantly longer spin-up for a fixed computational budget.

We define and implement spin-up protocols in which the ocean is first spun up using the ACE2–NEMO system and then used to initialise EC-Earth3. We present preliminary results from these spin-up experiments. While they indicate that longer simulations (both spin-up and test period) are needed to robustly assess the approach, deep ocean temperatures show potential improvements when an ACE2-NEMO simulation is used to provide initial conditions compared to the conventional approach.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

To sum up, these experiments indicate that ML-based emulators could be used to reduce the cost and, depending on how scalable they are (to be investigated in the future with high-resolution configurations), enable a substantial speed-up of ocean spin-up in coupled climate models. This provides a clear motivation for further development and evaluation of hybrid ML–physics approaches.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

Introduction

Achieving a stable initialisation of the ocean model in a climate simulation is important to reduce spurious trends (‘drifts’). Typical approaches to achieving this initialisation is to ‘spin-up’ the ocean using a brute force approach, whereby the ocean is initialised with a best guess ocean state (such as climatology), and run for several years. Whilst it can take thousands of years for the deep ocean to reach equilibrium, computational resources and time limit the typical spin-ups to 50-100 years. In the advent of coupled atmosphere-ocean models that use higher resolution ocean models, this spin-up cost becomes even larger, and so this motivates exploring ways to reduce the computational burden of spinning up the ocean. In this work, we explore a method to reduce the resources required to perform coupled spin-up by leveraging the recent developments of stable, accurate and efficient machine-learning based atmospheric emulators. These atmospheric emulators are computationally cheap and extremely fast compared to physical atmosphere and ocean models, enabling compute resources to be redeployed from the physical atmosphere model to the ocean model to be spun up. This would enable longer and/or faster spin-ups, which would reduce model drift in the full ESM simulation. We therefore couple an atmospheric emulator (ACE2) to a dynamical ocean (NEMO), and explore how this coupled system simulates the mean state, coupled variability, and seasonal cycle compared to a conventional climate model. We then perform a spin-up experiment using this coupled setup, and analyse the drift and biases resulting from using this spun-up state compared to spin-up using a conventional model.

1. Experimental setup

1.1. Atmospheric Model

We use ACE2 as the atmospheric model (Watt-Meyer et al., 2025). ACE2 is a machine-learnt climate emulator that has been demonstrated to be stable out to thousands of model years. The model checkpoint we use is trained on ERA5 data (Ai2, 2025; Hersbach et al., 2020) forced by global mean atmospheric CO_2, top of atmosphere short-wave radiation flux, and sea surface temperature, although without any aerosol forcing. The input and output of the model is on a regular 1°x1° grid at 8 different levels, with predictions made autoregressively in 6-hour timesteps. ACE2 predicts sensible, latent, short-wave and long-wave fluxes, as averages over 6 hours.

In order to pass NEMO outputs to ACE2, changes to ACE2 are required. For ingestion of sea surface temperature (T_{oce}) and sea ice temperature (T_{ice}) we introduced a modified Ocean class in ACE2, to allow reading ocean information from a file using a polling mechanism. Once the data was read, the surface temperature $T_{surf,oce}$ over ocean points was overwritten using a weighted sum of T_{oce} and T_{ice}, with weighting given by the sea ice fraction f_{ice} output from NEMO:

$$T_{surf,oce} = f_{ice}\, T_{ice} + (1 - f_{ice})\, T_{oce}$$

To pass sea ice fraction from NEMO to ACE2 required a different mechanism, since the Ocean model class in ACE2 is not set up to handle sea ice information. We instead made use of the \`perturbation\` mechanism created for e.g. perturbing the sea surface temperatures; by constructing a custom perturbation function in ACE2 that reads data from disk, we can overwrite the sea ice fraction forcing data sent to ACE2 from the forcing files with the data from NEMO.

1.2. Ocean model

We use the Nucleus for European Modelling of the Ocean (NEMO) model (Madec et al., 2023), which includes a sea ice model. The ocean grid is chosen as ORCA1, to closely match the 1° grid of the atmosphere, with 75 depth levels, and a model time step of 45min.

To align the model as closely as possible with EC-Earth3, we also use NEMO version 3.6. Namelists for NEMO are taken from the PRIMAVERA configuration (Haarsma et al., 2020), from which the standard climatological run-off and geothermal heating inputs for NEMO are also taken. NEMO is initialised from the 1951 checkpoint from a historical run performed in EC-Earth3's CMIP6 version (Bilbao et al., 2021), which starts in 1850 (see Section 1.8).

To facilitate robust compilation and setup of the model, we use the infrastructure created for EC-Earth4[1]. This utilises python packages such as scriptengine[2] and rdy2cpl[3] in order to reliably set up and compile the files required for NEMO and OASIS to run.

1 https://ec-earth-4-docs.readthedocs.io/
2 https://scriptengine.readthedocs.io
3 https://github.com/uwefladrich/rdy2cpl

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

1.3. Coupling Infrastructure

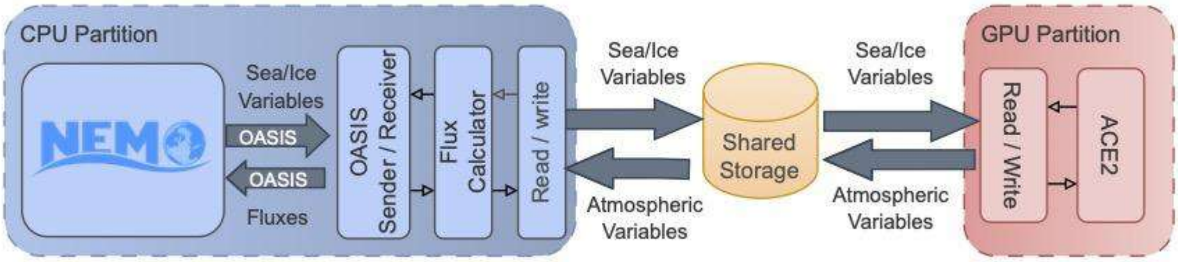

Figure 1: Schematic diagram of the infrastructure created to couple NEMO to ACE2

A fundamental challenge is that NEMO runs on multiple CPUs, whilst ACE2 runs on a single GPU, and the conventional HPC environments do not allow multi-node hybrid CPU-GPU jobs within the same MPI environment, thus ruling out a straightforward heterogeneous batch job. As a workaround we used shared storage to mediate data passing between ACE2 and NEMO (Figure 1). This adds some read/write overhead; however this was small compared to the overhead of running the models and calculating the fluxes. We note that in the future, NEMO will run on GPU, and candidate configurations already exist for this, simplifying this communication challenge.

To communicate with NEMO, we use the existing integration of NEMO with the OASIS library (Craig et al., 2017). Using PyOASIS (Gambron et al., 2021), SST, bulk ice temperature, sea ice fraction, sea ice thickness, ocean currents, and ice velocities from NEMO are intercepted by a Python router module, within which atmospheric variables are also read from shared storage. The coupling frequency is set to the timestep of ACE2, 6 hours, considerably longer than the 45 minutes used in EC-Earth at 1° resolution.

The freshwater fluxes that NEMO receives are evaporation (over ocean and ice points) and precipitation (liquid and solid). The heat fluxes NEMO receives are non-solar heat flux (sum of sensible, latent, and long-wave fluxes), solar heat flux, and momentum fluxes, all provided for ocean and sea ice separately. The version of ACE2 available to us at the time of writing (Ai2, 2025) produces most of the fluxes, except for momentum fluxes, evaporation, and solid precipitation. In order to calculate momentum fluxes and evaporation, we use the AirSeaFluxCode package (Biri et al., 2023). This is a python implementation of the bulk formulae used within NEMO for calculating fluxes when forced by atmospheric variables (Brodeau et al., 2017): our coupling approach is therefore consistent with EC-Earth. The output of AirSeaFluxCode provides wind stress, and latent heat of evaporation which allows conversion of latent heat flux to evaporation.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

To provide NEMO with solid precipitation, we implemented a simple heuristic model to estimate solid precipitation. Based on observations of the relationship between solid precipitation and surface temperature with ERA5, we set the solid precipitation to be equal to the total precipitation over sea points whenever the 2-metre temperature is ≤ 273K. River discharge is provided by standard climatology files from within the EC-Earth3 data.

1.4. Iterative model development

Here we report on results from intermediate ACE2-NEMO model versions. These early model versions showed unacceptable drifts and instabilities which were fixed before the model was used to produce the final control run and spin-up simulations. Figure 2 and 3 show results from these intermediate model versions, demonstrating the need for each development step: for the performance of the final model, we refer the reader to Section 2 and Figures 4-14.

1.4.1 Freshwater Fluxes

Early testing of NEMO coupled to ACE2 showed large unrealistic downward trends in global sea surface height, and large upwards trends in global sea ice volume (Figure 2). This needed to be fixed. We observed that this decrease was partly removed when freshwater fluxes were removed

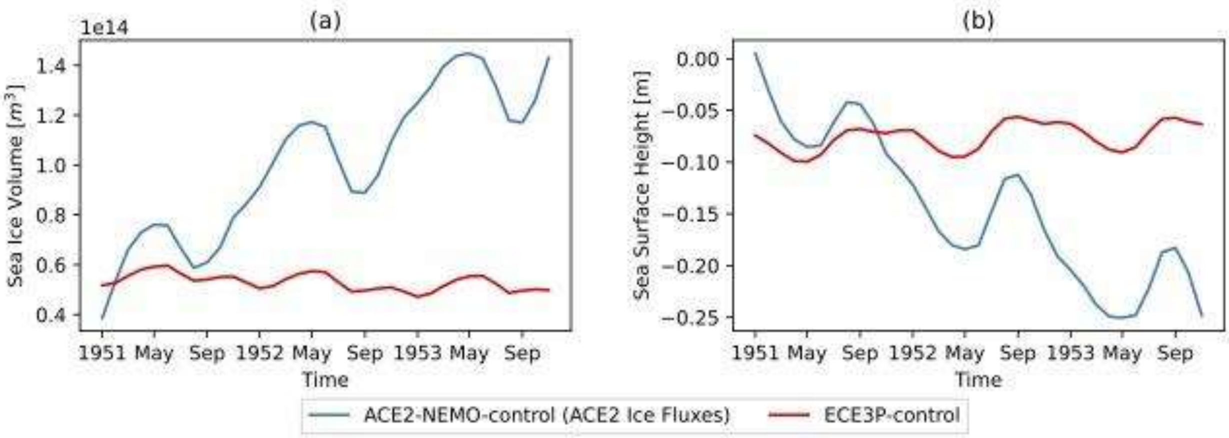

Figure 2: Evaluation of an intermediate model configuration. Evolution of (a) sea ice volume and (b) sea surface height for an ACE2-NEMO simulation using the fluxes from ACE2 over sea ice points. The large drift in Sea Ice Volume and Sea Surface Height was fixed by using fluxes calculated using the CORE bulk formulae - see Figure 6 for timeseries from final model version.

(intermediate model version not shown), and therefore applied the setting in NEMO that enforces

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

freshwater flux conservation `nn_fwb=1` in the NEMO ocean namelist), so that evaporation minus the sum of precipitation and run-off is adjusted to be zero at each time step. Although ACE2 has freshwater fluxes conserved at each time step, such a freshwater imbalance is to be expected, because the model uses a run-off climatology rather than a coupled run-off approach. Future development of the hybrid ACE2-NEMO model will implement a simple catchment scheme, such that run-off is related to precipitation over land, which should close the moisture budget.

1.4.2 Sea Ice Coupling

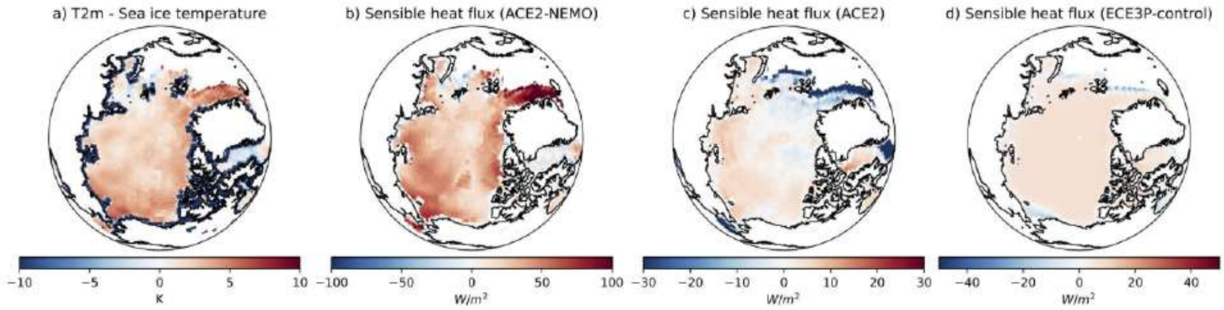

Figure 3: Comparison of sensible heat fluxes over sea ice points, over the first month of simulation only (a) difference between monthly 2-metre temperature and sea

Despite imposing a freshwater flux conservation constraint (see Section 1.4.1), the model still showed substantial and unacceptable drift in sea surface height and sea ice volume. This needed to be fixed. We observed that over many ice areas, the direction of the sensible heat flux did not align with the gradient between 2-metre temperature and sea ice temperature (Figure 3). We found a solution which removed the drift and aligned the flux direction with the temperature gradient: calculating the fluxes directly over ice using the simple CORE bulk formulae (W. G. Large & Yeager, 2009; W. Large & Yeager, 2004), as implemented in the NEMO 3.6 code (within the `sbcblk_core` module).

The bulk formulae to calculate momentum flux, sensible heat flux, and latent heat flux over ice, denote τ_{ice}, H_{ice}, and L_{ice} respectively, are:

$$\tau_{ice} = \rho C_{ice}(u_{10m} - u_{ice})|u_{10m} - u_{ice}|$$

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

$$H_{ice} = \rho C_{ice} c_p (T_{2m} - T_{ice})|u_{10m} - u_{ice}|$$

$$L_{ice} = L_s \rho C_{ice}(q_{2m} - 11637800\ \ exp\,(-5897.8/T_{ice})/\rho)|u_{10m} - u_{ice}|$$

Where ρ = 1.22 kg m^{-3} is the air density, c_p=1005.0 J kg^{-1}K^{-1} the specific heat capacity of moist air, L_s=2.839 x 10^{6} J kg^{-1} the latent heat of vaporisation, and C_{ice}=1.4 x 10^{-3} is the heat transfer coefficient. T_{2m}, u_{10m} and q_{2m} are atmospheric fields, namely 2-metre temperature, 10-metre wind velocity, and 2-metre specific humidity respectively, predicted by ACE2. T_{ice} is the bulk temperature of the ocean, and u_{ice} is the velocity of the ice, both output from NEMO. Fluxes are defined such that a positive flux goes downwards (i.e. from air to ocean) to be consistent with NEMO conventions. Latent heat fluxes are constrained to be negative.

Net long-wave and short-wave radiation, $R_{lw,ice}$ and $R_{sw,ice}$ respectively, are calculated as:

$$R_{lw,ice} = 0.95\,(R_{lw\downarrow} - \sigma_B T_{ice}^4)$$

$$R_{sw,ice} = (1 - \alpha_{ice})R_{sw\downarrow}$$

where σ_B is the Stefan-Boltzmann constant, $R_{lw\downarrow}$, $R_{sw\downarrow}$ are the downwelling long- and short-wave radiative fluxes provided by ACE2, and α_{ice} is the ice albedo provided by NEMO. Moving from ACE2 fluxes (Figure 2) to bulk formula computed fluxes led to stable long term simulations with no erroneous trends in sea ice volume or sea surface height. Results from the final module configuration are shown in Figure 6 - including for Sea Ice Volume and Sea Surface Height.

Despite the realistic global trends, large local values of sea ice thickness near coastal points were still observed, causing the model to crash due to large sea surface height values. These appeared to be due to large fluxes produced because of the temperature in the shallow coastal waters, and perhaps as a result of regridding the fluxes to the ORCA1 grid. To fix this local instability, we removed heat fluxes for any point within 1 grid square of the coast that also had more than 10% coverage of sea ice.

1.5. Computational Cost

ACE2-NEMO runs at a rate of around 10 simulated years per day (SYPD), compared with around 12 SYPD for NEMO alone using the same resources: the cost of ACE2 is very small compared to NEMO. With this setup, NEMO typically runs at around 20s per simulated day, whilst ACE2 takes less than 1s per simulated day. The interface between ACE2 and NEMO is dominated by the speed of flux calculations using AirSeaFluxCode (20-40s per simulated day) which is not

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

currently optimised for this application; note that if ACE2 provided momentum fluxes and an estimate of latent heat of vaporization, this cost would be removed. More recent versions of ACE2 predict momentum fluxes (Duncan et al., 2025) and so removing this cost for future simulations is very achievable.

The NEMO model is run in parallel across 2 CPU nodes, with each node running 16 tasks. The Python router runs on a separate node, although this router is lightweight and so has negligible cost compared to NEMO. ACE2 is run on a single NVIDIA Ampere A100 GPU. Following a similar calculation to Pathak et al. (2022), the NEMO model uses approximately as many cores as are available on one Cray XC40 node, and so takes 20 node-seconds per simulated day, equating to around 5.4kJ per simulated day. ACE2 takes 1 node-second per simulated day, equating to around 0.4kJ, based on a thermal design power of 300W (NVIDIA, 2026); therefore ACE2-NEMO consumes around 5.8kJ per simulated day. EC-Earth3 requires around three times as many node-seconds per simulated day as NEMO by itself (Acosta et al., 2023), so that EC-Earth3 consumes around 16.2kJ per simulated day. ACE2-NEMO is therefore significantly more efficient to run than EC-Earth3, requiring around a third of the energy per simulated day at this resolution, under this simple energy proxy based on node-seconds and thermal design power. Further efficiency gains are expected after optimisation of the flux calculation.

This factor of three reduction in the energy cost of ACE2-NEMO compared to EC-Earth3 enables substantially longer spin-up for a fixed computational budget.

1.6. Control Run Experiment

To explore how realistic the coupled ACE2-NEMO model is, we performed a simulation using constant 1951 CO_2 forcing (ACE2-NEMO-control). This has a 3-member lagged ensemble created by performing runs with atmospheric initial conditions shifted by 1 day. These are compared with a 1950s control run from the PRIMAVERA experiment (Haarsma et al., 2020) denoted ECE3P-control. Further details on the data used is given in Section 1.8.

1.7. Spin up protocol

For the spin-up experiments, we have developed two protocols. The first is a ‘Like-for-like’ comparison, whereby the ocean is spun-up for the same amount of time using ACE2-NEMO and EC-Earth 3. The second approach (‘Extended’) makes use of the fact that ACE2-NEMO uses less than half the CPU resources of EC-Earth3, and so we can perform a longer spin-up using still

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

fewer compute resources than for the EC-Earth3 spinup. In each case, the spun-up ocean state from ACE2-NEMO is provided as initial conditions to EC-Earth3 to investigate how much drift the post-spin-up periods present. It is also expected that there will be some 'shock' when providing EC-Earth3 with initial conditions from ACE2-NEMO, because feedback from the atmospheric model will not be identical to the IFS model in EC-Earth3. We therefore build in a 20 year buffer at the end of the ACE2-NEMO spin up to account for this adjustment. Results are compared against a spin-up using EC-Earth3 only ('Baseline').

The full spin-up protocol for these three runs (Like-for-like, Extended, and Baseline) is provided in Table 1.

Simulation name	ACE2-NEMO spin-up Length	EC-Earth3 spin-up length	EC-Earth3 evaluation period
Like-for-like (aall)	40	20	40
Extended (aalm)	70	20	40
Baseline (aaex)	-	60	40

Table 1: Summary of the spin-up experiments performed

1.8. EC-Earth3 Comparison Datasets

EC-Earth3 is used for comparison with the ACE2-NEMO control experiment because it uses the same NEMO version, and because the atmospheric model used in EC-Earth 3 is the Integrated Forecast System (IFS) from the European Centre for Medium-Range Weather Forecasts (ECMWF), which is the same atmospheric model (albeit not exactly the same version) used to construct the ERA5 dataset. We use the 1950s control run from the PRIMAVERA experiment (Haarsma et al., 2020) denoted ECE3P-control. This is initialised from a 50-year spin-up run, and has the same ocean configuration as ACE2-NEMO. The ocean initial condition for our experiments is taken from a 1850s historical run presented in Bilbao et al. (2021). Using initial conditions from the 1850s historical run means that we are likely to have a well-spun-up ocean, reducing ocean drifts in NEMO due to residual spin-up, compared to using the 50-year spun-up initial conditions from the PRIMAVERA experiment. The EC-Earth3 simulations use the T255 atmospheric grid (~80km) and the ORCA1 ocean grid ~100km. ACE2 atmospheric initial conditions are taken from the pre-prepared initial conditions detailed in(Watt-Meyer et al. (2025)

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

2. Results

2.1. Control run experiment

By coupling ACE2 to NEMO, we provide it with ocean inputs that are out of sample from what it has been trained on. Therefore, we begin by examining the stability of the coupled model, and investigate the extent to which it has erroneous behaviour, or persistent long-term trends. We compare ACE2-NEMO-control with ECE3P-control. These simulations are not expected to line up precisely, since the ECE3P-control run starts from different ocean initial conditions, and uses a dynamical atmosphere model that ACE2 has not been trained to emulate. However, we can still compare the mean state behaviours to look for differences in patterns or trends.

2.1.1. Mean state

Plots of mean daily precipitation, 2-metre temperature, sea surface temperature and sea surface height are shown in Figure 4. The spatial distribution of these variables is in good agreement with that seen for ECE3P-control. However, 2-metre temperature and SST are warmer for ACE2-NEMO-control in the tropics; one reason for this is the ocean initial conditions used are around 0.5-1 K warmer than the ECE3P-control initial state, and additionally there is warming seen during the first year that appears to be due to an imbalance in evaporation and the air-sea temperature gradient. There are also differences in the sea surface height, particularly noticeable in the western Pacific; this may be related to differences in balance of freshwater fluxes. Heat and freshwater fluxes averaged over the control run show good visual agreement with ECE3P-control (Figure 5). This confirms that ACE2 is still able to produce realistic fluxes when coupled to NEMO instead of receiving prescribed boundary conditions from ERA5.,, This is despite the fact that NEMO boundary conditions will be systematically different (biased) compared to the ERA5 fields on which ACE2 was trained, making this an out-of-sample generalisation test for the ML model ACE2.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

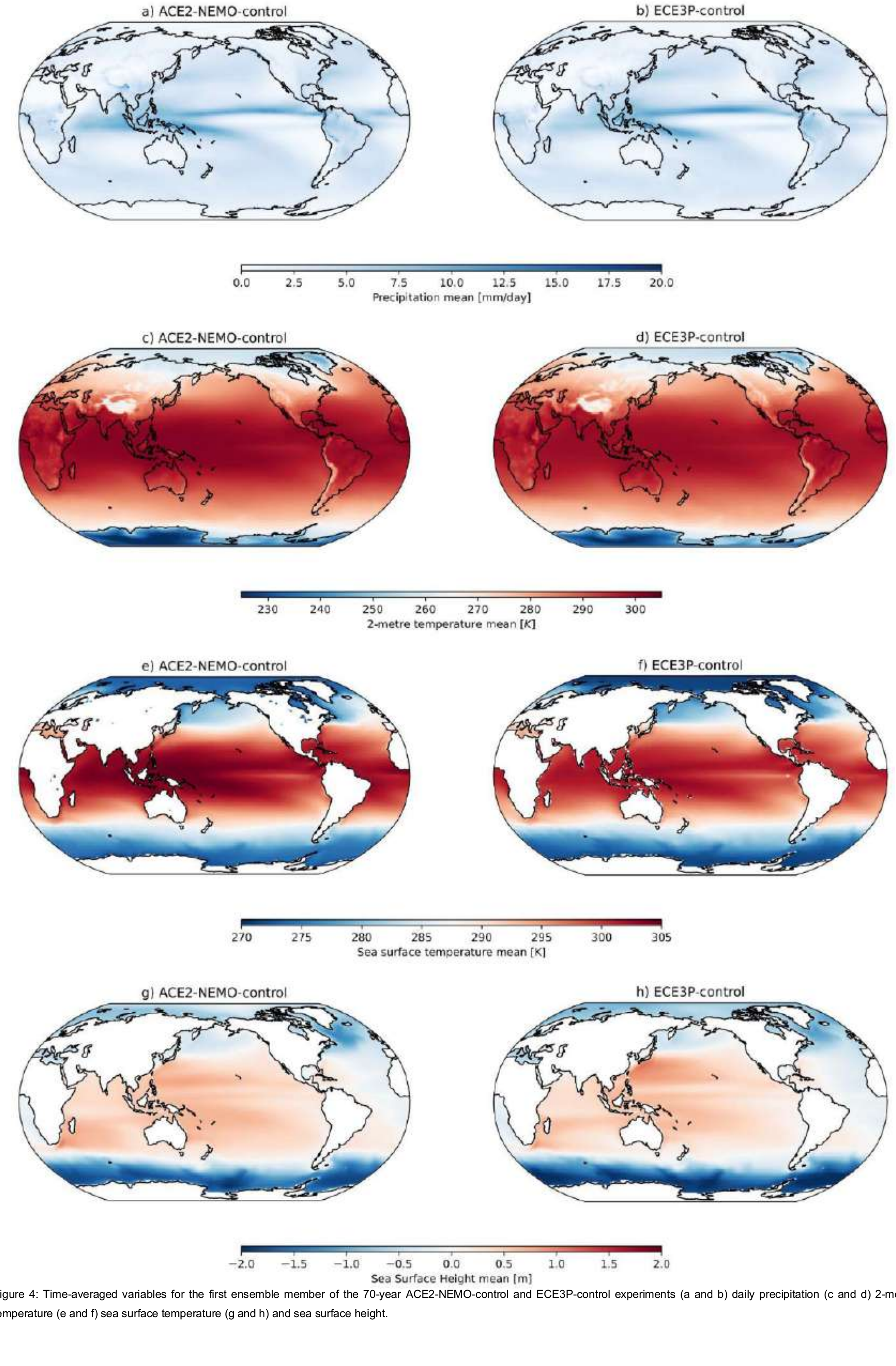

Figure 4: Time-averaged variables for the first ensemble member of the 70-year ACE2-NEMO-control and ECE3P-control experiments (a and b) daily precipitation (c and d) 2-metre temperature (e and f) sea surface temperature (g and h) and sea surface height.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

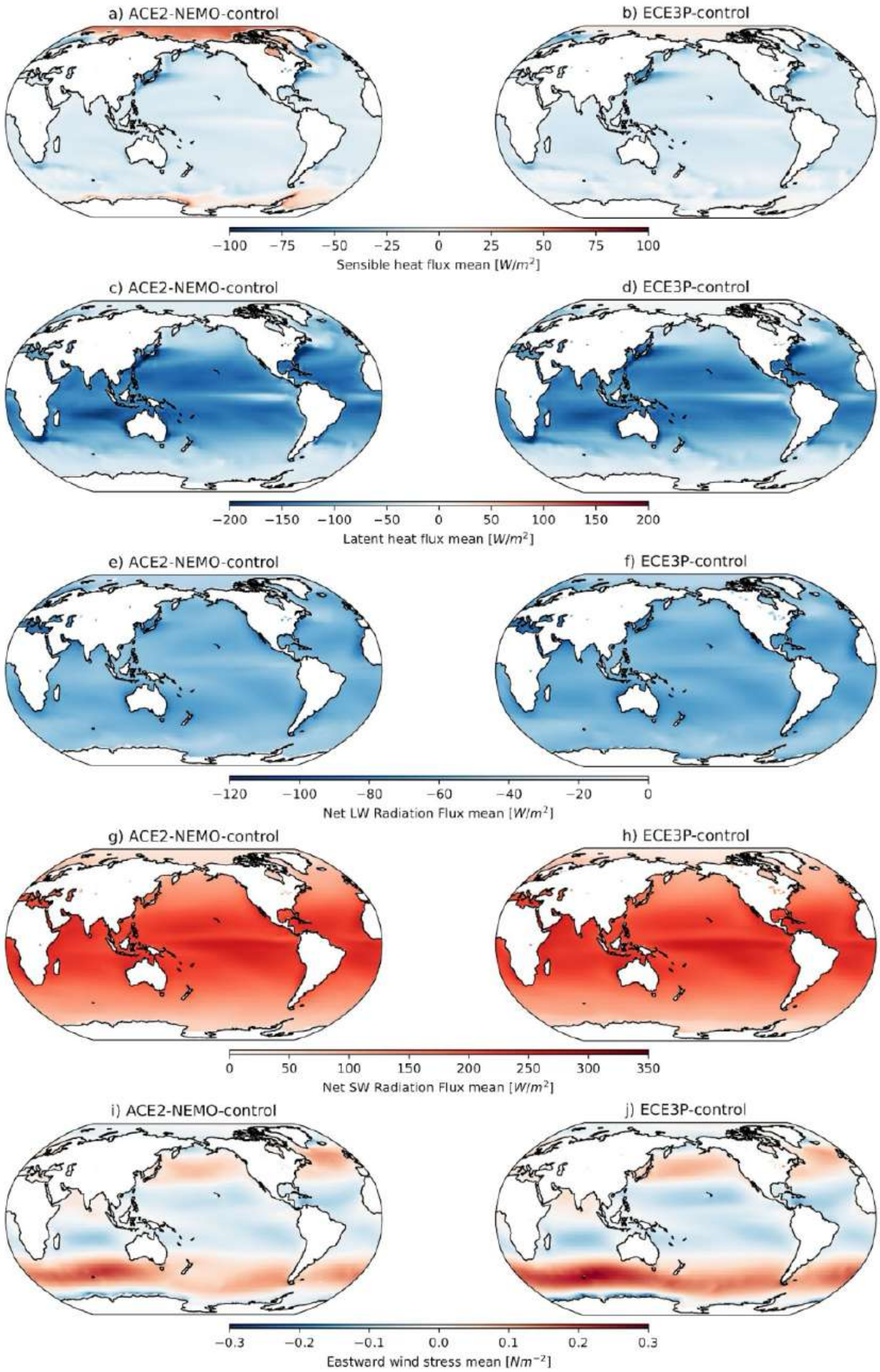

Figure 5: Mean fluxes for ACE2-NEMO-control and ECE3P-control: (a) and (b): sensible heat flux, (c) and (d): latent heat flux, (e) and (f): net surface long-wave radiation flux, (g) and (h): net surface short-wave radiation flux, (i) and (j) Eastward wind stress.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

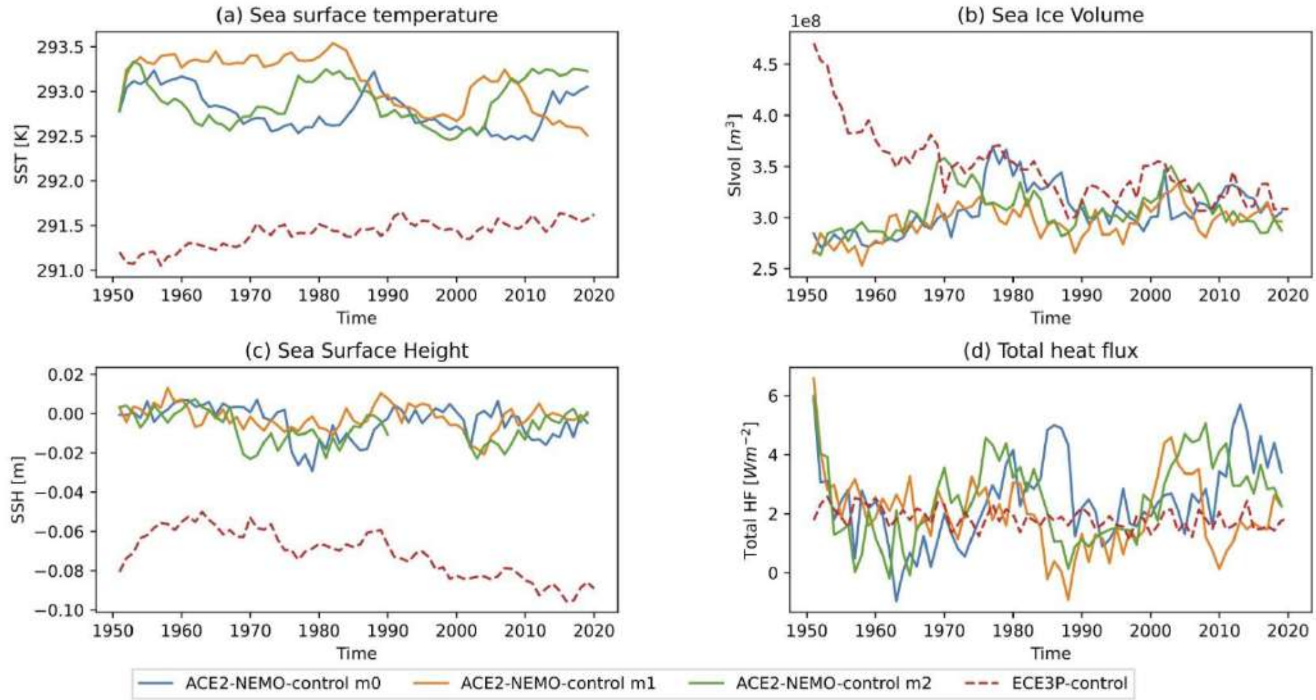

Figure 6: Globally averaged time series for the ACE2-NEMO-control 3-member ensemble and EC-Earth3P-control. (a) sea surface temperature, (b) sea ice volume, (c) sea surface height (d) total heat flux

2.1.2. Global mean evolution

Time series of global averages of several variables are shown for the control runs in Figure 6. The global sea surface temperatures (panel a) show little overall trend, although a larger multi-decadal variability than ECE3P-control is evident with a possible oscillation between two different modes. The ACE2-NEMO SSTs are also warmer than ECE3P-control by around 1-2 K, due to the difference in initial conditions and increased evaporation mentioned above. We note that the magnitude of this difference is within the typical CMIP6 range of errors - e.g. EERIE Deliverable 1.4 reports on simulations made with two different versions of the same MPI-M ICON model: the same model with updated physics shows a comparable difference in globally averaged SST of 1.4 K. Sea ice volume (panel b) shows a much larger change for ECE3P-control than ACE2-NEMO, with an initial decrease in volume of around one third over 20 years, perhaps due to the shorter spin-up of ECE3P-control compared to the initial conditions used for ACE2-NEMO, which had spun-up from several centuries. ACE2-NEMO shows a slight increase in volume over the same period, such that both models converge on approximately the same amount. Sea surface height (panel c) shows similar variation and stability in both models. Total heat flux (panel d)

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

shows larger variability for ACE2-NEMO, but a similar mean value overall, with little trend. Overall, these time series indicate stable control simulations have been achieved.

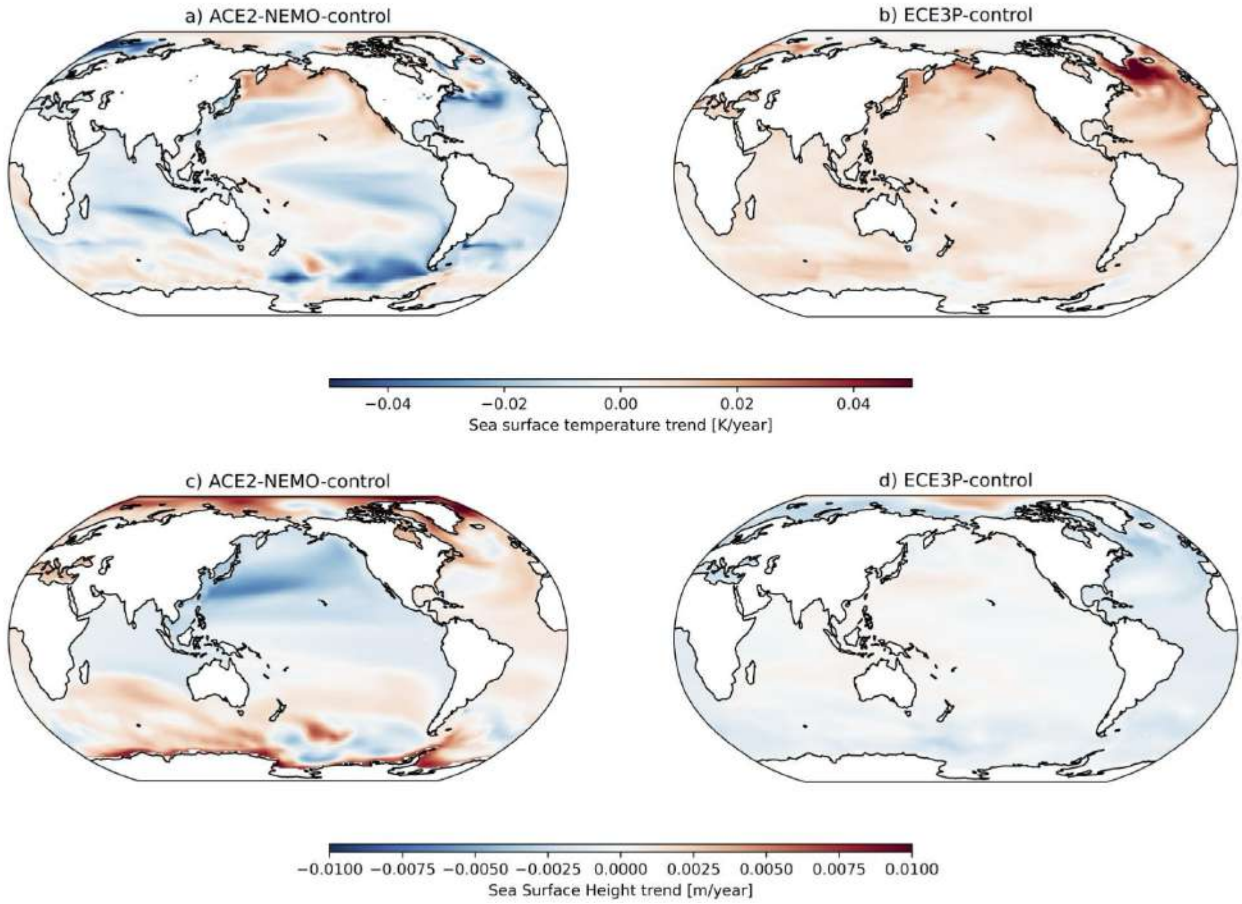

Figure 7: Linear drift in surface variables, estimated by fitting a regression model to each grid cell separately, on

Although the global changes are similar, the pattern of SST and SSH changes are quite different between ACE2-NEMO and ECE3P-control (Figure 7), with SSH changes likely driven by modifications made to the air-ice coupling (see Section 1.4). Temperatures in the deep ocean tend to decrease in ACE2-NEMO, except for towards the poles (Figure 8 a), whilst in E3P-control the trends are more towards warming (Figure 8 b).

Despite these differences, these results indicate that ACE2 can run stably with unfamiliar boundary conditions provided by NEMO. In turn it can provide NEMO with realistic outputs including fluxes, which drive realistic ocean evolution.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

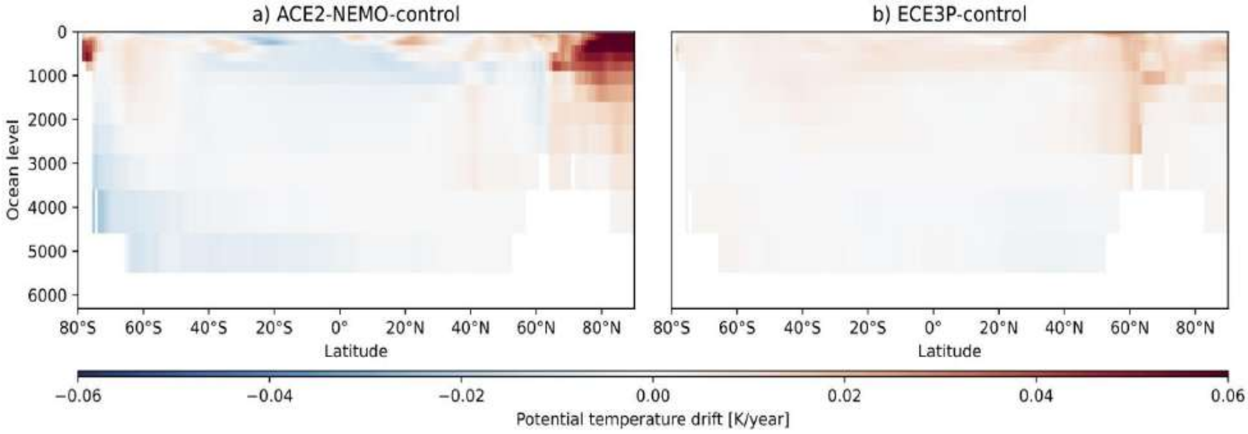

Figure 8: Drift in ocean potential temperature, averaged zonally for each latitude for (a) the first ensemble member of ACE2-NEMO-control and (b) ECE3P-control.

2.1.3. Seasonal cycle and interannual variability

The seasonal cycle of sea surface temperature, sensible heat flux, latent heat flux, and eastward momentum flux are shown in Figure 9. These are calculated by first calculating the average monthly value for each year, and dividing the monthly values by the average for their respective year. This then gives a representation of the seasonal cycle excluding interannual variability. Shading indicates 2 standard deviations either side of the mean seasonal cycle. Results are shown over sea points only, and are split into three regions: Tropics (20°S-20°N), northern extratropics (30°-70°N), and southern extratropics (70°-30°S). Overall we can see good agreement between the annual cycles of ACE2-NEMO-control and ECE3P-control, with similar levels of variance. The seasonal cycle for sea surface temperature in the extratropics (panels e and i) shows lower variance for ECE3P-control, and covers a smaller range in the southern extratropics, possibly related to the large variability of the sea ice (Section 2.1.2). Sensible and latent heat fluxes in the tropics also show differences in the relative strengths of fluxes in March and October.

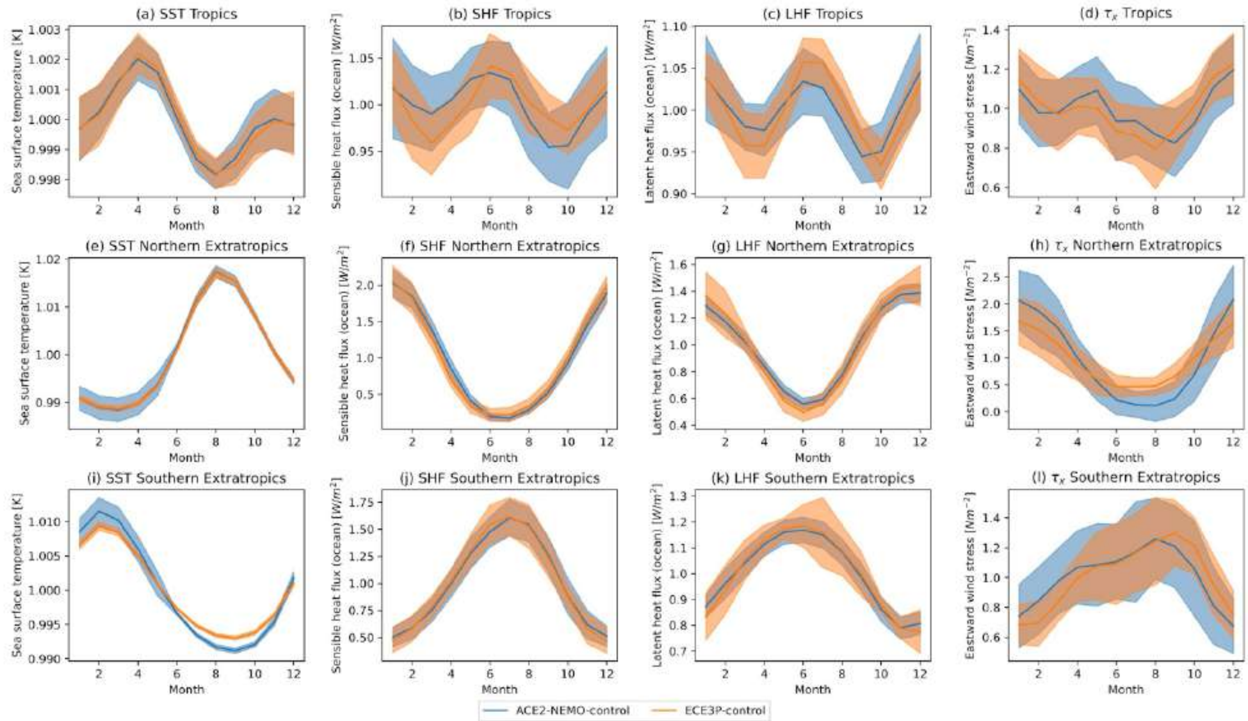

Figure 9: Seasonal cycle of variables from ACE2-NEMO-control (first ensemble member)

Plots of the interannual variability for the same variables are shown in Figure 10. These show the distribution of the annual mean of monthly values, with median values shown as white dots, and interquartile ranges shown as the thick black lines. Most of the data points for ACE2-NEMO-control and ECE3P-control have similar magnitudes of ranges, albeit with noticeable offsets in the median values. Sea surface temperature (panels a, e, and i) is consistently higher for ACE2-NEMO-control than ECE3P-control, as expected from the warmer ocean initial conditions used by ACE2-NEMO. Interannual variability in the southern extratropics is particularly high, perhaps because of the larger variability in sea ice due to the use of bulk formulae in the air-ice fluxes (Section 1.2). Sensible and latent heat fluxes in the extratropics (panels f, g, j, and k) are consistently lower for ACE2-NEMO-control in the extratropics, and generally show a greater range of values over all areas. The momentum fluxes (panels d, h, and l) show similar ranges of interannual variability, but much higher fluxes in the tropics (panel d) whilst much lower fluxes in the extratropics (h and l).

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

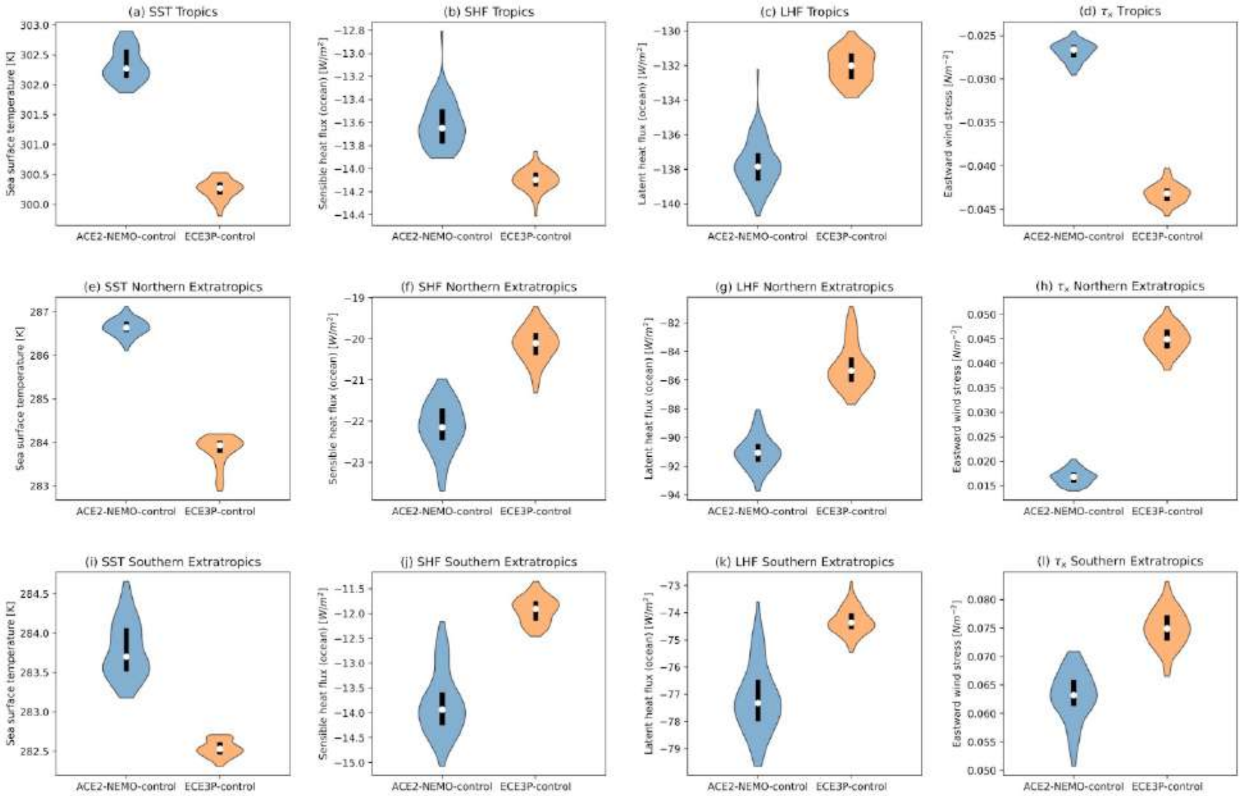

Figure 10: Interannual variability of variables from ACE2-NEMO-control (first ensemble member) and ECE3P-control. Median values shown as white dots, and interquartile ranges shown as the thick black lines. See main text for a description

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

2.1.4. Sea ice

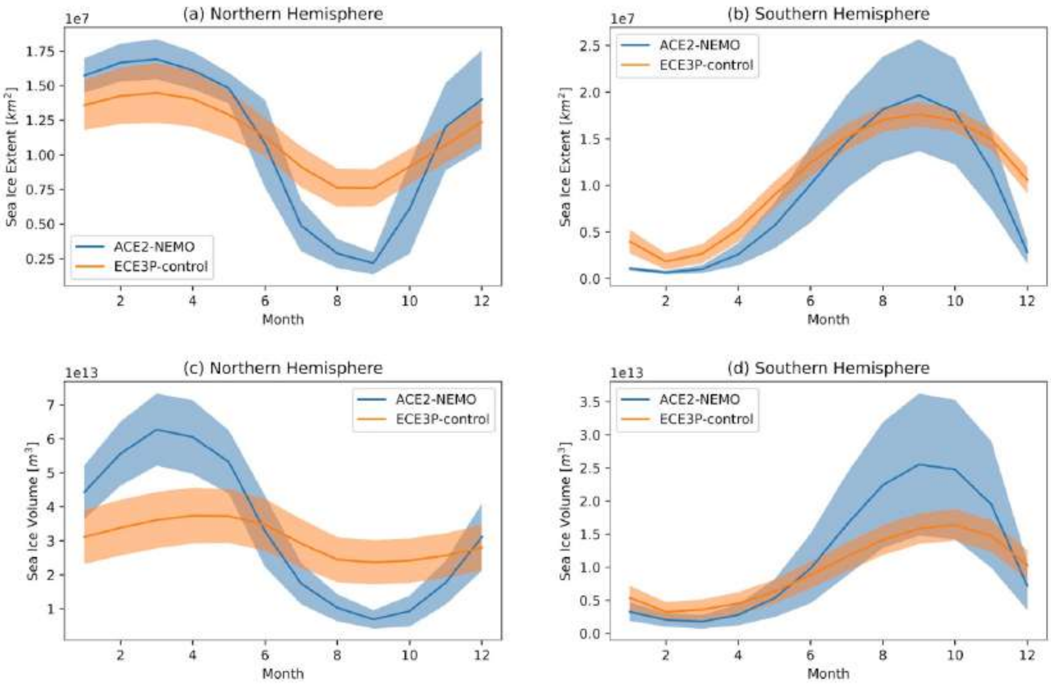

Figure 11: Average annual cycle of sea ice extent (top row) and sea ice volume (bottom row) for the northern

The annual cycle of sea ice volume for the northern and southern hemisphere is shown in Figure 10 for ACE2-NEMO-control and ECE3P-control. The maxima and minima of the cycle for ACE2-NEMO-control align well with ECE3P-control, indicating the ice fluxes are correctly modulating the sea ice, however it is clear that the variation in sea ice volume and extent is much larger for ACE2-NEMO. For comparison, the observed Northern Hemisphere Sea ice extent varies between 6 and 15 million square km such that the actual annual cycle lies between the two models (Copernicus marine services). The variation in sea ice volume in Figure 11 (c) and (d) indicates that the sea ice becomes much thicker in winter and much thinner in summer: the amplitude in ACE2-NEMO is much greater than observed. This difference may be explained by the use of bulk sea ice temperature in the calculation of sensible fluxes over ice for ACE2-NEMO, since this is the field output from NEMO. This may give larger temperature gradients than using surface ice temperature. Plots comparing sensible heat flux in Figure 3 also confirm that this heat flux is much larger than the original ACE2 fluxes or the ECE3P-control experiment.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

2.1.5 Summary

The control simulations indicate that the ACE2-NEMO system is able to produce stable coupled simulations over multi-decadal timescales with a substantial computational saving compared to the full coupled EC-Earth simulations. This indicates the potential for using this approach to spin-up the dynamical NEMO model.

2.2. Spin up analysis

This section presents a first analysis of the spin-up behaviour of the three EC-Earth3 simulations (baseline, like-for-like, and extended), comparing the ocean state evolution over the available evaluation period (2001-2040) for a selection of key variables. It must be noted that these are relatively short experiments with limited spin-up phases and lingering equilibration trends are found over the evaluation period, so the results should be interpreted with caution. Since all three experiments use the same model version and use the same fixed radiative forcings, they would be expected to eventually converge toward the same climatological state; the key question explored here is whether the ACE2-initialized (like-for-like and extended) experiments can reach that state faster than the baseline, reducing the need for a long traditional spin-up.

Figure 12 shows the globally averaged time series of sea surface temperature (tos), mixed layer depth (MLD), and sea surface height (zos) for the three EC-Earth3 simulations over the full spin-up and evaluation periods. For tos, the three simulations enter the evaluation period at clearly distinct values, with the like-for-like and extended experiments starting approximately 0.5-1°C cooler than the baseline. While the baseline shows a rather stable evolution both ACE2-initialized experiments exhibit strong positive trends over 2001-2050, indicating they are still warming toward the equilibrium climatology of the model established by the baseline, and have therefore not yet completed their adjustment within the available simulation length. The mixed layer depth time series presents a more encouraging picture: while small offsets persist and the linear trends differ somewhat between experiments, interannual variability is broadly compatible across all three simulations during the evaluation period, suggesting that the upper-ocean mixing layer adjusts relatively quickly and that the ACE2-initialized experiments are already sampling a similar dynamical regime to the baseline in this respect. The sea surface height, by contrast, remains distinctly offset throughout the evaluation period, with the extended experiment consistently higher than the other two. While all three simulations exhibit negligible trends in the globally averaged zos, this apparent stability is misleading, as the spatial trend maps in Figure 13

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

(computed over the evaluation period) reveal strong regional trends of opposite sign across different basins (with zos increases in the Indo-Pacific sector and decreases in the North Atlantic) that largely cancel in the global mean. This indicates that, despite the flat global-mean evolution, the ocean circulation is still undergoing a large-scale dynamical adjustment, and that persistent steric and dynamic height differences have not equilibrated within the experiment length.

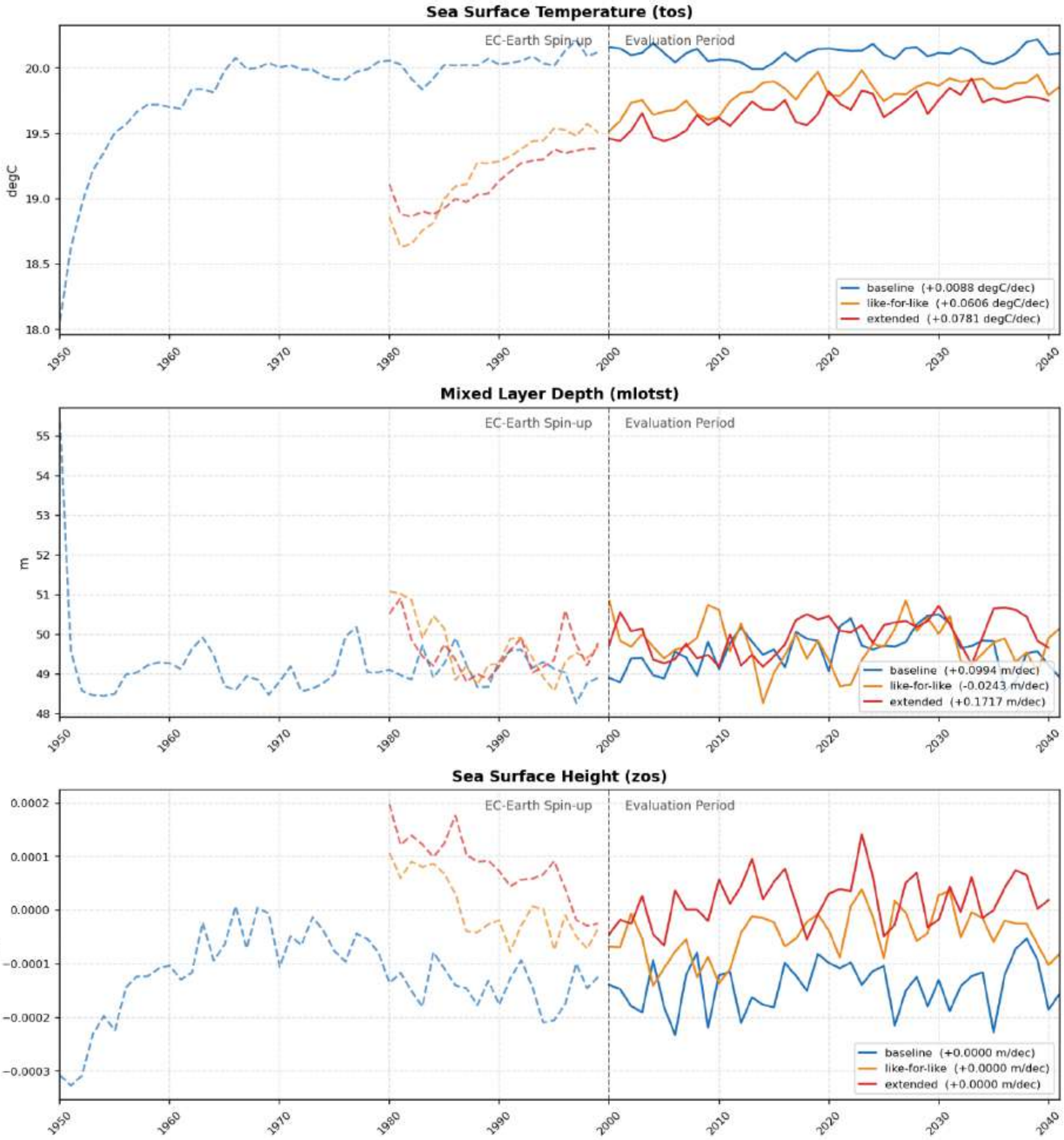

Figure 12: Timeseries of the globally averaged annual sea surface temperature, mixed layer depth and sea surface height in the three EC-Earth simulations. Linear trends over the evaluation period are shown in the legend.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

The tos trend maps in Figure 13 also show a clear ongoing adjustment in the like-for-like and extended experiments, characterized by a broad, quasi-uniform warming trend across most ocean basins, particularly strong in the tropical and subtropical Pacific and Indian Oceans, leading to an overall drift toward a warmer equilibrium state. The baseline, in contrast, shows a more spatially heterogeneous pattern, including some cooling in regions such as the North Atlantic subpolar gyre, which likely reflects internal variability superimposed on a much weaker overall trend. The MLD trend maps are more similar in spatial structure across the three experiments, with the large trends concentrated in the regions of deep water formation of the North Atlantic, the Arctic and Southern ocean. The coherent deepening of the mixed layer depth in the Subpolar North Atlantic across models indicates that all three simulations are probably undergoing a common large-scale strengthening of the Atlantic Meridional Overturning circulation, which has yet to reach its equilibrium state.

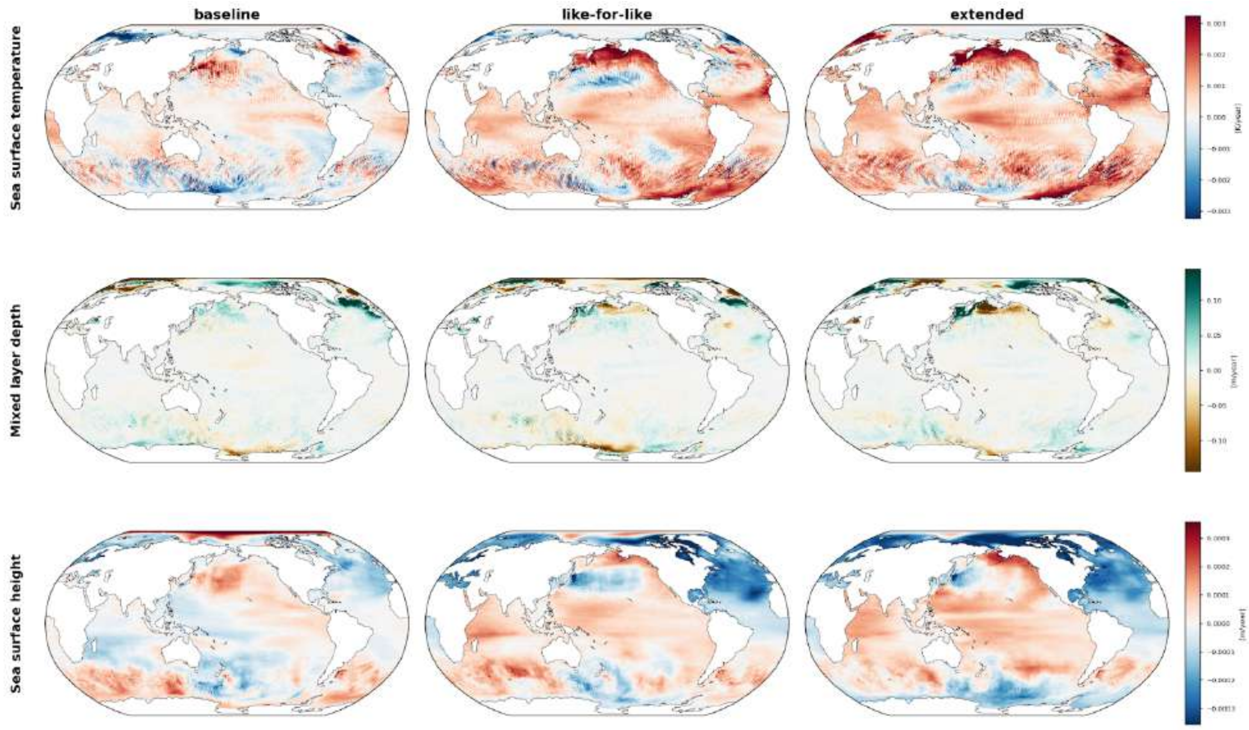

Figure 13: Linear trends over the evaluation period 2001-2040 for the annual sea surface temperature, mixed layer depth and sea surface height in the three simulations with EC-Earth.

Figure 14 now shows the globally averaged ocean potential temperature time series at six depth levels (0-100m, 100-500m, 500-1000m, 1000-2500m, 2500-4000m, and 4000m-bottom) for the three EC-Earth3 simulations. At the shallowest levels (0-100m and 100-500m), the ACE2-initialized experiments start the evaluation period considerably colder than the baseline and

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

display substantially larger positive trends, confirming that they are still warming toward the baseline's thermal state. At intermediate depths (500-1000 m and 1000-2500 m), a similar picture emerges, with the ACE2-initialized experiments being warmer than the baseline at the start of the evaluation period but with diverging rather than converging trends, suggesting that these experiments are still far from reaching a common equilibrium with the baseline. Interestingly, between 1000 and 4000m the trend magnitudes in the ACE2-spun up experiments are markedly weaker than at shallower levels (e.g. −0.0023 and −0.0027 °C/dec at 1000-2500m for like-for-like and extended, and near-zero to slightly negative values at 2500-4000m), in contrast to the strong positive trends that persist in the upper ocean. This reduction in trend magnitude at intermediate depths in the ACE2-initialized simulations encouragingly points to a possible earlier stabilization of the intermediate ocean. Below 2500m, the three simulations track each other more closely in terms of trend magnitude, though they remain offset in absolute values, which could be explained by the very long timescales associated with deep ocean ventilation and adjustment. These results confirm that deep ocean temperature is one of the variables most strongly affected by the short duration of these experiments, and that longer integrations will be necessary to properly assess convergence at depth.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

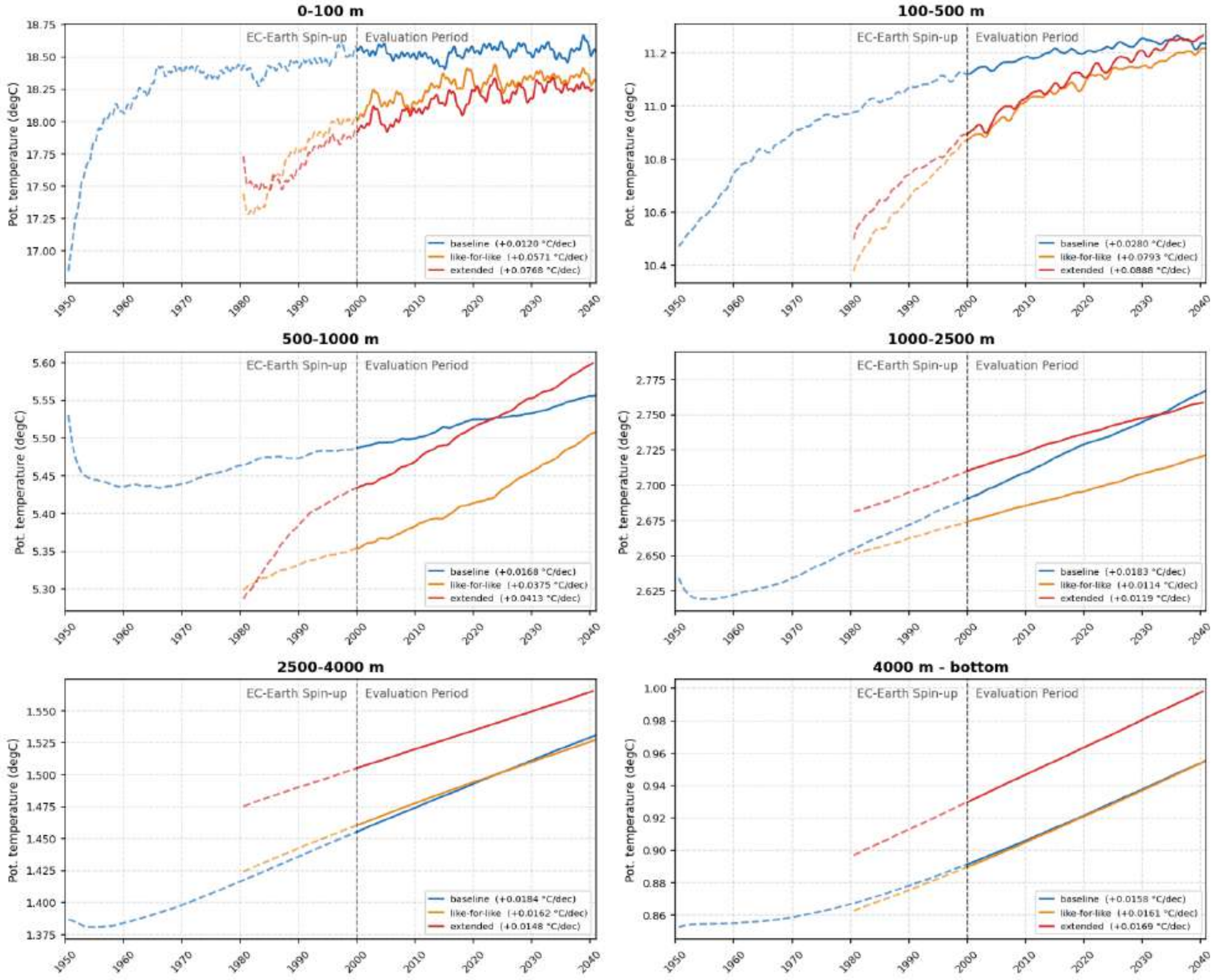

Figure 14: Timeseries of the globally averaged ocean temperature at different selected levels in the three EC-Earth simulations. The linear trends over the evaluation period are included in the legend.

Overall, the results presented in this section provide a first assessment of the potential of ACE2-NEMO as a tool for accelerating coupled ocean spin-up. While the baseline and ACE2-initialized experiments present different evolutions over the evaluation period, all of them present lingering drifts across most variables and depth levels, which suggests that substantially longer integrations will be required to reach robust conclusions about convergence. Based on the timescales implied by the current trends, particularly in the deep ocean where adjustment can take centuries, we estimate that evaluation periods of at least 200 years will be necessary to properly characterize convergence across the full water column and to determine whether the ACE2-initialized

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

experiments ultimately approach the same equilibrium state as the baseline at all depths. This in turn also underscores the need to extend the ACE2-NEMO spin-up phase substantially beyond the 40 and 70 years currently available, so that the ocean initial conditions passed to EC-Earth3 are closer to the model's coupled equilibrium. Such extended experiments are planned and will constitute the core of the remaining analyses from task 12.1, which will be presented in the final project report.

3. Conclusion

Achieving efficient and physically consistent ocean spin-up remains a major computational bottleneck in coupled climate modelling. This is especially the case for computationally expensive eddy-rich Earth System Models. To address this problem, we have presented ACE2-NEMO, the first coupling of a pure machine-learnt atmospheric emulator (ACE2) with a dynamical ocean model (NEMO). Apart from changes to the sea ice fluxes, no fine tuning of either component is performed. This creates a unique test of the robustness of the ACE2 atmospheric emulator to out-of-sample inputs and provides a novel test of how well the machine learning emulator has learned the appropriate feedbacks to force a dynamical ocean. It also presents an unexplored approach to integrating machine learning models with dynamical ocean models. Given the relative sparsity of ocean data, integrating with ocean models that model the physical equations may provide an advantage over 'full' machine-learning-based ocean emulators in how faithfully they can represent long-term ocean variability.

Achieving the coupling requires overcoming several technical challenges; the models run in different environments, in different programming languages, and the atmospheric model does not produce all of the fluxes required by the ocean. Our solution creates a bridge between the two environments, and leverages the existing OASIS coupling infrastructure together with the AirSeaFluxCode library to achieve realistic coupling. Through developing the coupling, several problems were identified relating to the ACE2 fluxes over sea ice; the sensible heat fluxes produced by ACE2 over sea ice did not seem to be consistent with the temperature differences, and there were large drifts in sea ice volume and sea surface temperature. These problems were solved by using simple bulk formulae to calculate heat fluxes over ice (Figure 6).

The resulting model is stable over multidecadal timescales and shows realistic mean states and fluxes comparable to those in EC-Earth3 control simulations. These results demonstrate that the machine learning emulator is capable of driving a dynamical ocean model in a physically consistent manner for long integrations. This indicates that ACE2–NEMO is a viable option for

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

replacing (or partially replacing) the long coupled spin-up simulations needed to initialise ECEarth3.

We estimate the computational efficiency of the ACE2–NEMO configuration relative to EC-Earth3 and find a factor of three reduction in energy usage at this particular resolution. This is due to the extremely high efficiency of running the ACE2 atmospheric emulator compared to the dynamical IFS atmospheric model used in EC-Earth3. This would enable significantly longer integrations for a fixed computational budget, which could improve the spun-up state used for NEMO initialisation.

In this deliverable, we have also defined and implemented spin-up protocols in which the ocean undergoes a first equilibration using the ACE2–NEMO system, to produce the initial conditions to subsequently initialise EC-Earth3. Due to technical issues in EC-Earth that were only recently overcome, the testing of this approach has been done with relatively short simulations, which are insufficient to determine its efficiency and suitability for spin-up acceleration purposes. An extension of those experiments is currently underway.

We note some limitations of this study. The initial conditions used for ACE2-NEMO are warmer than those used for EC-Earth3, which may introduce additional biases in the coupled system. We have performed the above simulations at 1 degree resolution. This is because current atmospheric emulators operate at this resolution, enabling a resolution match at the coupling interface. However the experiments must therefore be viewed as a proof of concept when it comes to eddy-rich model spin up – future work will consider applying this approach to higher-resolution models. Finally, the efficiency of the model is limited by the flux calculations, since they are not optimised for this application; significant speed-up of this module would therefore be expected with the appropriate optimisation. This additional efficiency gain would only strengthen the potential of this hybrid approach as a practical method to accelerate ocean spin-up in dynamical climate models.

Our findings motivate further investigation of this approach, particularly in the context of longer integrations and higher-resolution configurations. This will be outlined in the following section.

4. Next steps

We have demonstrated that coupling a ML-atmospheric emulator to a dynamical ocean model is technically feasible and more computationally efficient than a full dynamical model. The ML-emulator is capable of producing fluxes to drive stable ocean simulations. This is a strong proof-of-concept that hybrid ML-physics models could be used as a tool to accelerate ocean spin up.

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

The first priority is to extend the duration of the ACE2-NEMO and EC-Earth3 spin-up experiments to better understand the convergence behaviour. Our results indicate that the current simulations are too short to robustly assess equilibration, particularly in the deep ocean where adjustment timescales are very long. Within the final months of EERIE we will focus on producing longer simulations with ACE2-NEMO, before branching off EC-Earth3, to quantify the extent to which we can reduce drift. We will report on these results in the final project report.

The simulations presented have used ACE2 and NEMO at 1 degree resolution. A key question is how applicable the framework is to the eddy-rich models used and developed in EERIE. The resolution of current ML atmospheric emulators is relatively coarse (~ 1 degree) because they have been trained on ERA5. This means that coupling an eddy-rich ocean to a resolution-matched atmospheric emulator is not currently possible. Instead, an interesting avenue of future research is to couple a coarse atmospheric emulator to a high-resolution ocean model. This would test the spin-up approach in a modelling regime closer to that used in EERIE. We note that for the IFS-NEMO model used in EERIE (with Tco1279-eORCA12 grids), the atmosphere and ocean components have a similar computational cost: replacing the IFS with an emulator could therefore achieve speed-ups for a factor of 2 at these resolutions. In addition, the development of future high resolution reanalyses (e.g. ERA6 at 0.1 degree) is expected to enable future training of higher-resolution ML atmospheric models, which could be explored in this framework. Work using higher resolution ocean models and atmospheric emulators will be explored in future projects.

5. References

Acosta, M. C., Palomas, S., & Tourigny, E. (2023). Balancing EC-Earth3 Improving the Performance of EC-Earth CMIP6 Configurations by Minimizing the Coupling Cost. *Earth and Space Science*, *10*(8), e2023EA002912. https://doi.org/10.1029/2023EA002912

Ai2. (2025). *ACE2-ERA5 (Revision a4ca6cc)*. Hugging Face. https://doi.org/10.57967/hf/5377

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., Cruz-García, R., Cvijanovic, I., Doblas-Reyes, F. J., Donat, M., Dutra, E., Echevarría, P., Ho, A.-C., Loosveldt-Tomas, S., Moreno-Chamarro, E., Pérez-Zanon, N., Ramos, A., … Vegas-Regidor, J. (2021). Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth. *Earth System Dynamics*, *12*(1), 173–196. https://doi.org/10.5194/esd-12-173-2021

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

Biri, S., Cornes, R. C., Berry, D. I., Kent, E. C., & Yelland, M. J. (2023). AirSeaFluxCode: Open-source software for calculating turbulent air-sea fluxes from meteorological parameters. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.1049168

Brodeau, L., Barnier, B., Gulev, S. K., & Woods, C. (2017). Climatologically Significant Effects of Some Approximations in the Bulk Parameterizations of Turbulent Air–Sea Fluxes. https://doi.org/10.1175/JPO-D-16-0169.1

Craig, A., Valcke, S., & Coquart, L. (2017). Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0. *Geoscientific Model Development*, *10*(9), 3297–3308. https://doi.org/10.5194/gmd-10-3297-2017

Duncan, J. P. C., Wu, E., Dheeshjith, S., Subel, A., Arcomano, T., Clark, S. K., Henn, B., Kwa, A., McGibbon, J., Perkins, W. A., Gregory, W., Fernandez-Granda, C., Busecke, J., Watt-Meyer, O., Hurlin, W. J., Adcroft, A., Zanna, L., & Bretherton, C. (2025). *SamudrACE: Fast and Accurate Coupled Climate Modeling with 3D Ocean and Atmosphere Emulators* (arXiv:2509.12490). arXiv. https://doi.org/10.48550/arXiv.2509.12490

Gambron, P., Ford, R., Piacentini, A., & Valcke, S. (2021). *pyOASIS - a python and C interface for OASIS3-MCT* [Technical Report]. CECI, Université de Toulouse, CNRS, CERFACS, Toulouse, France, TR-CMGC-21-56. https://cnrs.hal.science/hal-04739704

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., Corti, S., Davini, P., Exarchou, E., Fabiano, F., Fladrich, U., Fuentes Franco, R., García-Serrano, J., von Hardenberg, J., Koenigk, T., Levine, X., Meccia, V. L., van Noije, T., van den Oord, G., … Wyser, K. (2020). HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR – description, model computational \hack\breakperformance and basic validation. *Geoscientific Model Development*, *13*(8), 3507–3527. https://doi.org/10.5194/gmd-13-3507-2020

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Large, W. G., & Yeager, S. G. (2009). The global climatology of an interannually varying air–sea flux data set. *Climate Dynamics*, *33*(2), 341–364. https://doi.org/10.1007/s00382-008-0441-3

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

Large, W., & Yeager, S. (2004). Diurnal to decadal global forcing for ocean and sea-ice models: The data sets and flux climatologies (No. NCAR/TN-460+ STR). *University Corporation for Atmospheric Research: NCAR*, *2044*, 111.

Madec, G., Bell, M., Blaker, A., Bricaud, C., Bruciaferri, D., Castrillo, M., Calvert, D., Chanut, J., Clementi, E., Coward, A., Epicoco, I., Éthé, C., Ganderton, J., Harle, J., Hutchinson, K., Iovino, D., Lea, D., Lovato, T., Martin, M., … Wilson, C. (2023). *NEMO Ocean Engine Reference Manual*. https://doi.org/10.5281/zenodo.8167700

NVIDIA. (2026). *A100 80GB PCIe GPU Product Brief, date accessed 26-02-2026*. https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/PB-10577-001_v02.pdf

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K., Hassanzadeh, P., Kashinath, K., & Anandkumar, A. (2022). *FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators* (arXiv:2202.11214). arXiv. https://doi.org/10.48550/arXiv.2202.11214

Watt-Meyer, O., Henn, B., McGibbon, J., Clark, S. K., Kwa, A., Perkins, W. A., Wu, E., Harris, L., & Bretherton, C. S. (2025). ACE2: Accurately learning subseasonal to decadal atmospheric variability and forced responses. *Npj Climate and Atmospheric Science*, *8*(1), 205. https://doi.org/10.1038/s41612-025-01090-0

Funding

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101081383

All UK Partners in EERIE are funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (grant numbers 10057890, 10049639, 10040510, 10040984).

a) T2m - Sea ice temperature b) Sensible heat flux (ACE2-NEMO) c) Sensible heat flux (ACE2) d) Sensible heat flux (ECE3P-control)

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

ETH Zürich’s contribution to EERIE is funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract #22.00366.