SPECIAL PROJECT PROGRESS REPORT

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

Reporting year	………2025……………………………………..……
Project Title:	Speeding up ocean spin-up using stochastic parametrisations …………………..……
Computer Project Account:	……spgbchri……………………………………..……
Principal Investigator(s):	Dr Hannah Christensen Dr Kristian Strommen Dr Robert (Bobby) Antonio
Affiliation:	University of Oxford
Name of ECMWF scientist(s) collaborating to the project (if applicable)	…………………………………………………………..…… …………………………………………………………..……
Start date of the project:	September 2023
Expected end date:	March 2026

Computer resources allocated/used for the current year and the previous one
(if applicable)
Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	50000000	15600000	50000000	172000
Data storage capacity	(Gbytes)	15500	500	15500	2000

Summary of project objectives (10 lines max)

The primary aim of this project is 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 is by coupling a machine learning emulator of the atmosphere to a standard ocean model (NEMO), which will significantly reduce the compute time required per year of spin up. Once we have successfully created a scheme to couple these models and stabilised the machine learning model, we will produce one or more spun-up ocean states and assess the quality of these states by running a conventional climate model (EC-Earth) using the spun-up state and assessing the drift and biases. We also hope to use this novel setup to explore scientific questions about how the atmospheric dynamics affect the ocean.

Summary of problems encountered (10 lines max)

The main problem we have had is constructing a robust and meaningful way to provide GraphCast with sea surface information; our analysis showed GraphCast's errors are related to the sea surface in a very location dependent manner. This means it is not possible to simply apply flux from the ocean into GraphCast, as we had first envisaged, but that a model needs to be learned as to where the sea surface corrections should be applied. So far, our experiments at learning such a model to perform seasonal forecasts did not demonstrate enough evidence of ocean-atmosphere coupling.
In December we also attempted to set up and run a simple seasonal forecasting experiment with GraphCast coupled to NEMO by simply applying fluxes. We encountered stability issues in NEMO at short run times, likely because of the initial conditions and config, which meant we could not run the experiment before the end of the year, and thus were unable to use the allocated units in time.

Summary of plans for the continuation of the project (10 lines max)

We have pivoted to using different emulators, GenCast or ACE2, both of which are much more stable at longer rollouts than GraphCast and accept SSTs as input. Having run some results on seasonal forecasting using GenCast, we are setting up NEMO 3.6 to couple with GenCast or ACE2 in seasonal forecast experiments (using NEMO 3.6 to align with our colleagues in EERIE). Once this is setup, we will run seasonal forecasting experiments using GenCast/ACE2 coupled to NEMO and assess how stable and accurate these forecasts are. Provided this step is successful, we will then move to performing coupled spin-up runs and assessing the quality of the spun-up ocean. We expect the spin-up and seasonal forecasting runs to use the remaining computational units available. In the meantime, we will also write up and submit for publication our results exploring how GraphCast's errors relate to properties of the sea surface, and on the seasonal forecasting using GenCast.

List of publications/reports from the project with complete references

None

Summary of results (July 2024 – June 2025)

Relationship between GraphCast's errors and the sea surface

To calculate the relationship between GraphCast's 2-metre temperature errors ε and sea surface properties, we created a dataset of GraphCast errors in 2-metre temperature at 24h lead time (to avoid diurnal effects), as well as a dataset of ERA5 variables averaged over a 24hr period starting from 30h before the forecast target time. We then analysed covariances between sea surface temperature (SST), sea surface temperature minus 2-metre temperature (SST-2mt), 2-metre temperature, mean surface sensible heat flux, and mean surface latent heat flux. Our results showed that by far the strongest relationship was between ε and SST or SST-2mt. In June there is particularly strong negative covariance between ε and SST-2mt over the Northern Pacific, tropical eastern Pacific, off the eastern coast of Africa, and in the Arabian Sea (Fig. 1, left panel). In December there are larger covariances around the Kuroshio extension and Gulf Stream, and weaker covariances around the Agulhas Return Current and Falkland-Malvinas current (Fig. 2, left panel). These results are consistent with the typical behaviour in these areas, where ocean-driven variability drives the air-sea interaction.
For June covariance with SST (Fig. 2, left panel) there is a dependence on SSTs in the southern hemisphere, which represents the lagged response of the ocean to heating during the boreal summer, such that heat is released from the ocean into the atmosphere in the boreal winter. The reverse is seen in December, where there are more significant correlations between ε and SST in the Northern hemisphere.

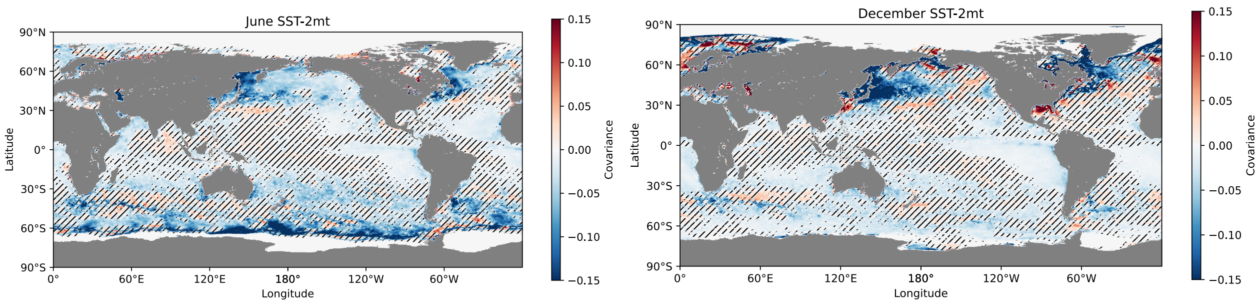

Fig. 1: Covariance between GraphCast 2mt errors at 24h lead time, and SST-2mt, (left) for June months only and (right) for December months only in 2004-2013. Hatching indicates where the correlation is not significant according to a two-tailed t-test at the 95% confidence level.

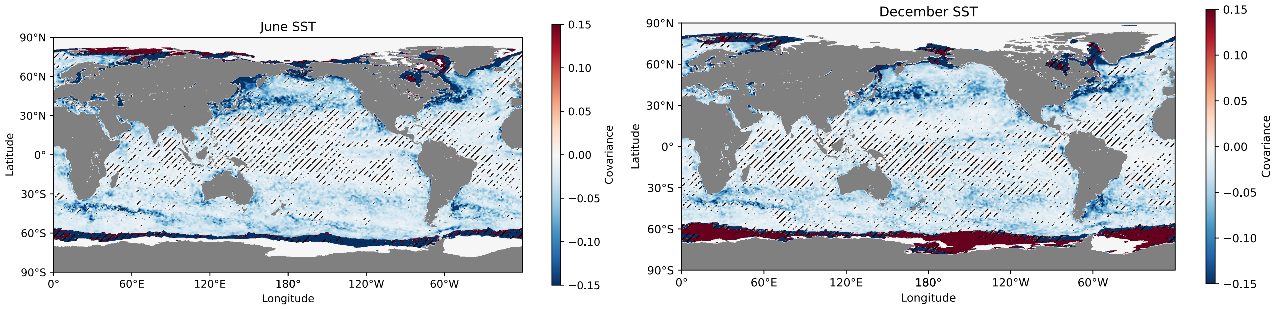

Fig. 2: Covariance between GraphCast 2mt errors at 24h lead time, and SST, (left) for June months only and (right) for December months only in 2004-2013. Hatching indicates where the correlation is not significant according to a two-tailed t-test at the 95% confidence level.

Seasonal Forecasting with GenCast

GenCast is trained to produce forecasts at up to 14-day lead times using autoregressive rollouts; it is therefore not clear how the model will perform when rolled out to seasonal timescales (several months). This experiment provides crucial information about how well GenCast will perform at longer timescales and will highlight any biases present in the model that we should be aware of. We perform seasonal forecasting experiments by running GenCast from 1st November to the end of the following February, for 2009-2024, with a 20-member ensemble. The SST data provided consists of either ERA5 ('Forced') or SST anomalies at 1st November persisted on the ERA5 climatology ('Persisted-anomaly'). The latter is closer to a true forecast since we do not know the SST information in advance, whilst the former provides a useful indication of where the lack of SST information is to blame for a lack in skill.

We begin by exploring patterns of 12hr-precipitation for seasonal forecasts initialised on 1st November 2021 and 1st November 2023. The former date had a moderately high La Niña event, whilst the latter date had a strong El Niño event. We limit ourselves to these years since GenCast is trained on data up to 2018. Ensemble mean monthly 12hr precipitation anomaly maps are shown in Fig. 3 and Fig. 4 for the 2021 and 2023 start dates, respectively, with equivalent data shown for ERA5 and a 20-member SEAS5 ensemble. In both cases, the wetting and drying signal over the tropical pacific and maritime continent broadly aligns with ERA5 and SEAS5, suggesting that GenCast is capturing some of the air-sea interactions at this long timescale. The 'Forced' experiment also appears to correct some of the biases of the 'Persisted-anomaly' experiment over e.g. the tropical Atlantic and over Southern Africa for the 2021 run.

Anomaly correlation coefficients calculated over the 15-year period (Figs. 5 & 6) also show some skill in forecasting 2-metre temperature (Fig. 5) and z500 (Fig. 6), although with significantly reduced skill compared to SEAS5, particularly at forecasting z500 over the high latitudes, and forecasting 2-metre temperature over the land. It is likely that a significant amount of this skill difference is due to SEAS5 having interactive ocean, ice, and land models, and because GenCast lacks sea ice as an input variable.

Overall, the results are encouraging that GenCast responds in broadly the correct way to the sea surface; it is an open question how much these seasonal forecasts results will carry over to the case of coupling to a dynamic ocean, where the model may encounter 'out-of-sample' ocean states.

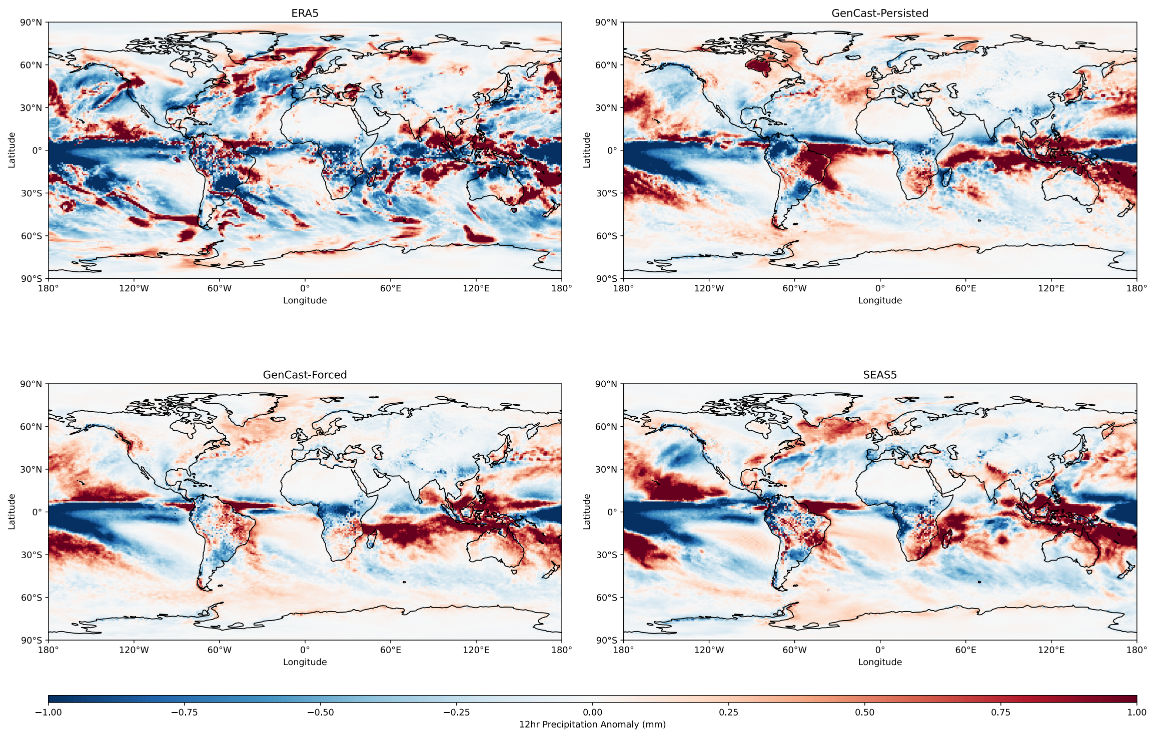

Fig. 3: 12hr precipitation anomalies averaged over December-January, for a forecast initialised on 1st November 2021.

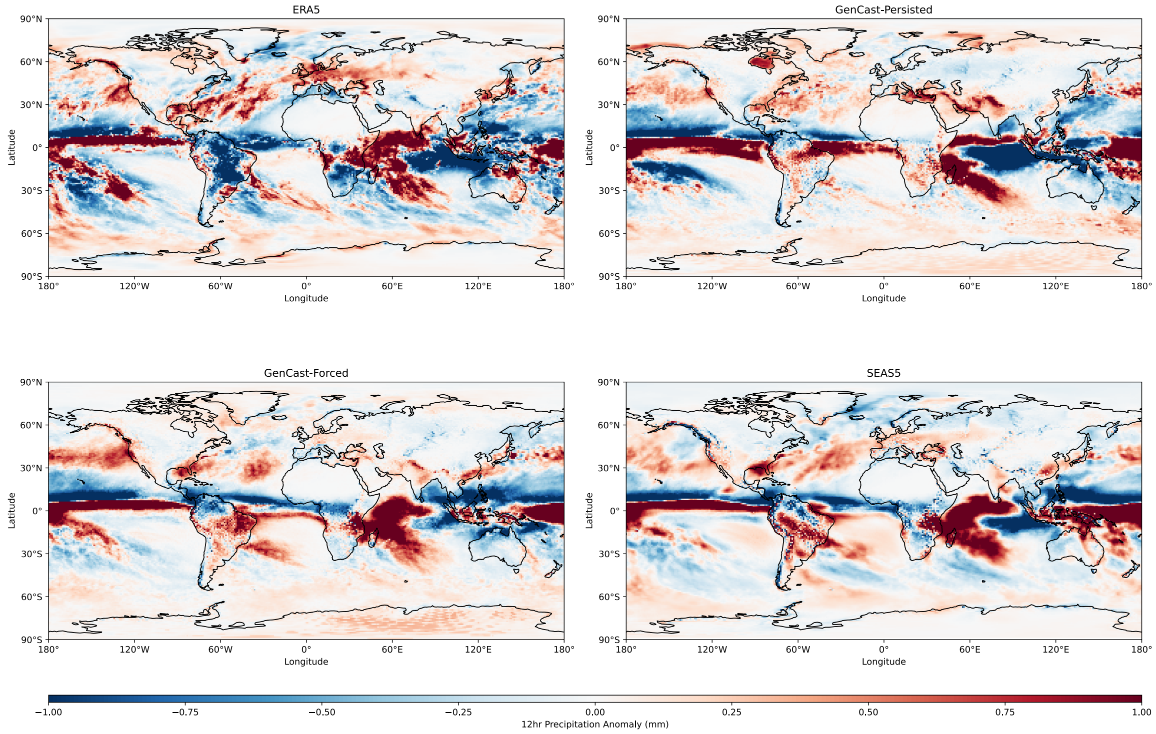

Fig. 4: 12hr precipitation anomalies averaged over December-January, for a forecast initialised on 1st November 2023.

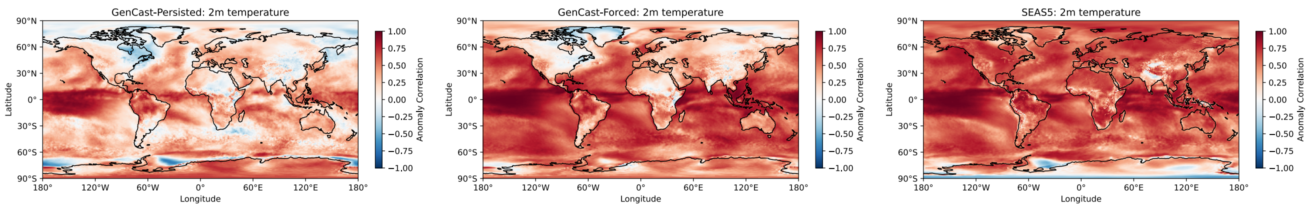

Fig. 5: Anomaly correlation coefficient between the different seasonal forecasts (at monthly level) and ERA5, for 2-metre temperature.

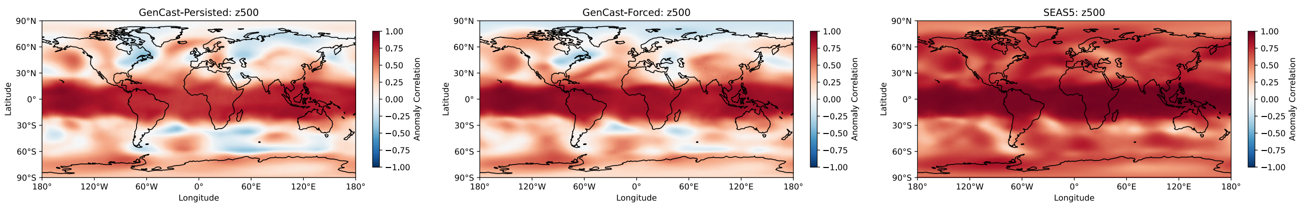

Fig. 6: Anomaly correlation coefficient between the different seasonal forecasts (at monthly level) and ERA5, for geopotential height at 500hPa.