
A modular framework to run AI-based models from high-resolution climate projections

Aina Gaya-Àvila

Earth Sciences Department
Barcelona Supercomputing Center
Plaça d'Eusebi Güell, 1-3, Les Corts, 08034 Barcelona
aina.gayayavila@bsc.es

Oriol Tintó

Earth Sciences Department
Barcelona Supercomputing Center
Plaça d'Eusebi Güell, 1-3, Les Corts, 08034 Barcelona

Amirpasha Mozaffari

Earth Sciences Department
Barcelona Supercomputing Center
Plaça d'Eusebi Güell, 1-3, Les Corts, 08034 Barcelona

Amanda Duarte

Earth Sciences Department
Barcelona Supercomputing Center
Plaça d'Eusebi Güell, 1-3, Les Corts, 08034 Barcelona

Abstract

Recent advances in AI-based weather and climate models promise major improvements in forecasting. However, their integration with state-of-the-art climate simulations are still constrained by data heterogeneity. High-resolution projections, such as those from the most important digital twins initiatives, are produced at 5 km resolution with specialized grids, formats, and variable sets that are incompatible with most AI models. Current integration efforts are ad hoc, model-specific, and difficult to reproduce, which slows progress and limits large-scale evaluation.

In this work, we propose a modular, source-agnostic framework that builds on our experience in climate workflows. It enables systematic and reproducible execution of AI-based climate models across diverse datasets and high-performance computing environments. The framework standardizes data preparation, automates model execution through containerized workflows, and includes post-processing and evaluation tools. Preliminary experiments show that it reproduces and extends recent studies on AI model robustness in future climates with minimal technical overhead.

1 Introduction

A key challenge in climate simulations is the volume and complexity of the data. The climate community relies on a variety of formats and storage solutions, making dataset handling and processing very demanding, especially at scale. At the same time, AI-driven weather and climate models are

emerging as tools that broaden access to accurate forecasts and climate projections (1; 2). Yet the lack of systematic tools to connect heterogeneous data and evolving AI architectures has become a major bottleneck for research and operational use. This heterogeneity, combined with the need for fast and reproducible analysis, highlights the importance of developing efficient workflows.

To address this issue, we propose a modular and reproducible framework that uses established workflow practices from climate science and modern containerization tools. The framework allows users to initialize diverse AI-based models from multiple data sources, run ensemble forecasts on High Performance Computing (HPC) platforms, and automatically generate diagnostics and evaluations. Its modular design ensures that new AI models and datasets can be integrated with minimal effort and scale with the fast evolution of machine learning in climate applications. The framework is also designed to be aligned with the FAIR principles (3). This will allow researchers to focus on scientific questions rather than the complexities of HPC and reproducibility.

2 Methodology

Our goal is to provide a flexible framework that enables researchers to initialize, run, and evaluate AI-based climate models from heterogeneous data sources in a systematic and reproducible way. The design emphasizes modularity, portability, and ease of extension, allowing new models and datasets to be integrated with minimal effort. The framework is structured as a workflow composed of four modular stages:

- **Setup:** prepares the sources and the environment in the HPC.
- **Data preparation:** retrieves, unifies, and transforms the initial conditions to be ready to be ingested by the AI-model.
- **Simulation:** runs the AI-model simulation using the initial conditions prepared in the previous step.
- **Post-processing:** automatically generates plots and evaluations.

These stages are orchestrated using Autosubmit (4), a workflow manager widely used in climate science, ensuring robust execution across HPC platforms.

The framework is designed to be both model-agnostic and source-agnostic. In order to add a new AI-based model, the user specifies (i) the required input variables, (ii) the accepted data format, (iii) a container recipe, and (iv) a standard interface (typically a Python call). These specifications are incorporated into configuration files and run scripts, making model integration straightforward. Currently, the framework supports two different weather forecast models: AIFS (5) and NeuralGCM (6). To integrate a new dataset, the steps to convert the data from its native representation into the intermediate uniform format must be defined. This guarantees interoperability across models. The framework currently supports ERA5 (7) and the Destination Earth Climate Digital Twin (8).

HPC environments differ widely in architecture, libraries, and compilers, which poses significant portability challenges. To overcome these, we implemented the entire simulation pipeline within Singularity containers. This encapsulation isolates the software from the host system, simplifying deployment and guaranteeing consistent model execution across heterogeneous HPC platforms.

3 Pathway to impact

We illustrate the potential of the proposal by presenting two use cases where this framework can be applied.

3.1 Use-case 1: Uncertainty quantification

The current Destination Earth Climate Digital Twin generates high-resolution projections at 5 km resolution but faces a key limitation since it produces deterministic forecasts without uncertainty bounds. Traditional ensemble methods, which rely on running multiple simulations under slightly different conditions, are computationally infeasible at this scale. Within the proposed framework, AI-based models are initialized from high-resolution simulations to generate ensembles that explore

alternative trajectories to those obtained from physics-based simulations. This approach enables the estimation of uncertainty associated with high-resolution predictions.

3.2 Use-case 2: Evaluating the fairness of AI weather models across global regions

Most AI models are trained on historical observations and numerical simulations, which often encode substantial spatial and temporal biases, particularly due to the underrepresentation of data-scarce regions in the Global South. These imbalances can manifest in uneven forecast skill, raising serious concerns for climate adaptation and disaster risk management. Recent studies have begun to explore these disparities. For instance, (9) compared model performance during extreme weather events, while (10) demonstrated regional sensitivity in model accuracy for extremes. Additionally, (11) assessed the fairness of several AI models, highlighting inequities in their predictive capabilities. Building on these insights, the proposed framework in this study offers a structured approach to evaluating and comparing model skill across geographic regions, under both average and extreme conditions. Similar in spirit to benchmarking efforts in traditional numerical modeling (12), our approach is scalable to a wide range of AI-based systems. Our framework will enable systematic research aimed at assessing the fairness of AI-driven weather and climate predictions, and provide a foundation for testing solutions toward more equitable forecasting across regions.

4 Preliminary results

A preliminary implementation of this framework allowed us to generate results using AIFS-single and NeuralGCM with initial conditions generated from Destination Earth Climate Digital Twin. Two experiments were set up to check whether the outputs of the AI models were able to follow similar trajectories as the physics-based models, if the models under consideration showed a cold drift towards the training state, and if the model was performing similarly over the globe.

Each model was configured to produce 10-day forecasts starting from initial conditions generated for January 2039. Both AIFS and NeuralGCM were configured to run for 10 days, producing outputs every six hours.

The time series for temperature for both NeuralGCM (10-member ensemble) and AIFS (single member) can be seen in Fig. 1. It can be seen how AIFS drifts to a colder state, more similar to where it was trained. These results are consistent with the ones published in (13). On the other hand, the different ensemble members of NeuralGCM tend to follow a trajectory similar to the one described in Climate Digital Twin.

This example shows that users can conduct valuable scientific experiments, such as reproducing and extending state-of-the-art research on the robustness of AI models in future climates, with minimal technical effort. This demonstrates the framework’s potential to accelerate research and operational adoption of AI in climate science.

Acknowledgements

AD and OT acknowledge their AI4S fellowship within the “Generación D” initiative by Red.es, Ministerio para la Transformación Digital y de la Función Pública, for talent attraction (C005/24-ED CV1), funded by NextGenerationEU through PRTR. AM acknowledges his JdC fellowship within the Grant JDC2023-051208-I (MICIU/AEI/10.13039/501100011033 and ERDF “A way of making Europe”).

References

- [1] G. T. Reporters, “Ai-driven weather prediction breakthrough reported,” *The Guardian*, Mar. 2025. Accessed: 2025-08-01.
- [2] I. Khadir *et al.*, “Democracy of ai numerical weather models: An example of global forecasting with fourcastnetv2 made by a university research lab using gpu,” *arXiv preprint arXiv:2504.17028*, 2025.

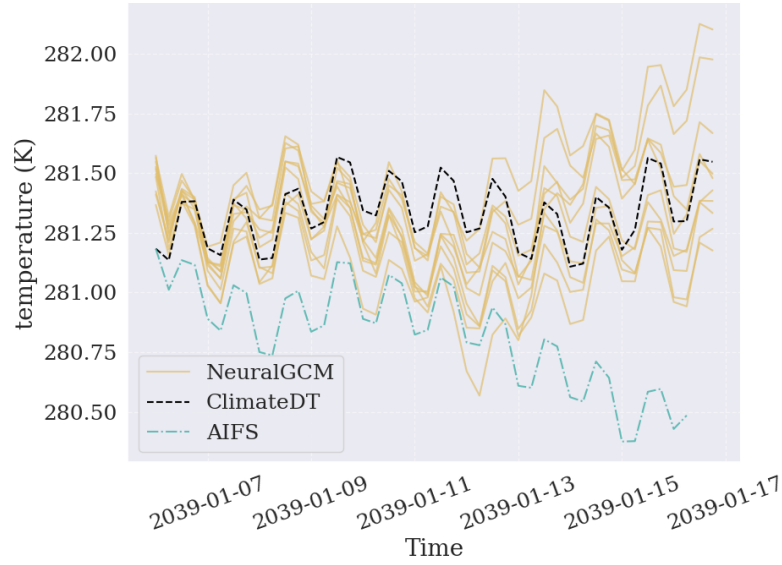


Figure 1: Temperature time series comparison of ClimateDT (data used as reference), AIFS (deterministic model), and NeuralGCM (stochastic model).

- [3] M. D. Wilkinson *et al.*, “The fair guiding principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, p. 160018, 2016.
- [4] D. Manubens-Gil *et al.*, “Seamless management of ensemble climate prediction experiments on HPC platforms,” pp. 895–900, July 2016.
- [5] S. Lang *et al.*, “Aifs – ecmwf’s data-driven forecasting system,” 2024.
- [6] D. Kochkov *et al.*, “Neural general circulation models for weather and climate,” *Nature*, vol. 632, pp. 1060–1066, Aug. 2024.
- [7] H. Hersbach *et al.*, “ERA5 hourly data on single levels from 1940 to present.” <https://doi.org/10.24381/cds.adbb2d47>, 2023. Accessed on 03-07-2025.
- [8] F. J. Doblas-Reyes *et al.*, “The destination earth digital twin for climate change adaptation,” *EGUsphere*, vol. 2025, pp. 1–41, 2025.
- [9] M. Feldmann *et al.*, “Lightning-fast thunderstorm warnings: Predicting severe convective environments with global neural weather models,” *arXiv e-prints*, pp. arXiv–2406, 2024.
- [10] L. Olivetti and G. Messori, “Do data-driven models beat numerical models in forecasting weather extremes? a comparison of ifs hres, pangu-weather, and graphcast,” *Geoscientific Model Development*, vol. 17, no. 21, pp. 7915–7962, 2024.
- [11] L. Olivetti and G. Messori, “Whose weather is it? a fairness framework for data-driven weather forecasting,” *Authorea Preprints*, 2025.
- [12] T. A. Duc *et al.*, “Skill validation of high-impact rainfall forecasts over vietnam using the european centre for medium-range weather forecasts (ecmwf) integrated forecasting system (ifs) and dynamical downscaling with the weather research and forecasting model,” *Atmosphere*, vol. 16, no. 2, p. 224, 2025.
- [13] T. Rackow *et al.*, “Robustness of AI-based weather forecasts in a changing climate,” arXiv, Sept. 2024. arXiv:2409.18529 [physics].