eMONARCH: An AI-Driven Emulator of the MONARCH CTM

David Mathas[1], Michael Orieux[1], Isidre Mas Magre[1], Hervé Petetin[1]

[1]Barcelona Supercomputing Center (BSC-CNS), Earth Sciences Department, Barcelona, Spain

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Background

Chemical Transport Models (CTMs) like MONARCH are essential for understanding atmospheric pollutant dynamics and designing mitigation strategies, but their high computational cost limits applications requiring many forward simulations.

- Improved emulation of CTMs enables rapid "what-if" emission scenario exploration for policy support at a fraction of the computational cost.
- Fast, reliable air quality predictions support better risk management, adaptation, and mitigation planning.

eMONARCH addresses this challenge by developing a meteorology- and emission-sensitive ML emulator of MONARCH over the Iberian Peninsula.

Dataset & Training Setup

An extensive synthetic dataset was generated using MONARCH over the **Iberian Peninsula** domain at $\mathbf{0.1^\circ \times 0.1^\circ}$ resolution:

- **Ensemble:** 5 emission-perturbed members of 4-year MONARCH simulations (anthropogenic emission scalings), split as 3 members for training, 1 for evaluation, and 1 for testing.
- **Variables:** Surface-level pollutant concentrations (NO_2, NO), gridded emissions, and meteorological fields (T, U, V, P, PBLH). The NO_x configuration uses **18 input channels** (14 physical + 4 cyclical time encodings) and **2 output channels**.
- **Data pipeline:** Zarr-compressed stores with TensorStore access and optimized chunking, enabling efficient I/O on MareNostrum 5 (BSC-CNS).
- **Training:** Trained on up to 32 NVIDIA H100 GPUs. Performed hyperparameter tuning with grid search via AutoML over, among others, normalization strategies and learning rates.

Model Design

The emulator learns a one-step mapping from the current atmospheric state plus forcings to the next hourly concentration field:

$$\hat{C}_{t+1} = \mathcal{F}_\theta(C_t,\ E_{t+1},\ M_t,\ M_{t+1})$$

where C is the chemical concentration state, E represents gridded emissions, and M meteorological fields. Cyclical time encodings (hour-of-day, day-of-year) are appended as additional input channels.

- **U-Net** with BatchNorm and Softplus activations, trained with MSE loss and the Adam optimizer (lr $= \mathbf{10^{-3}}$).
- **Normalization:** Z-score standardization of all input and output channels consistently selected as best across all tuning experiments.

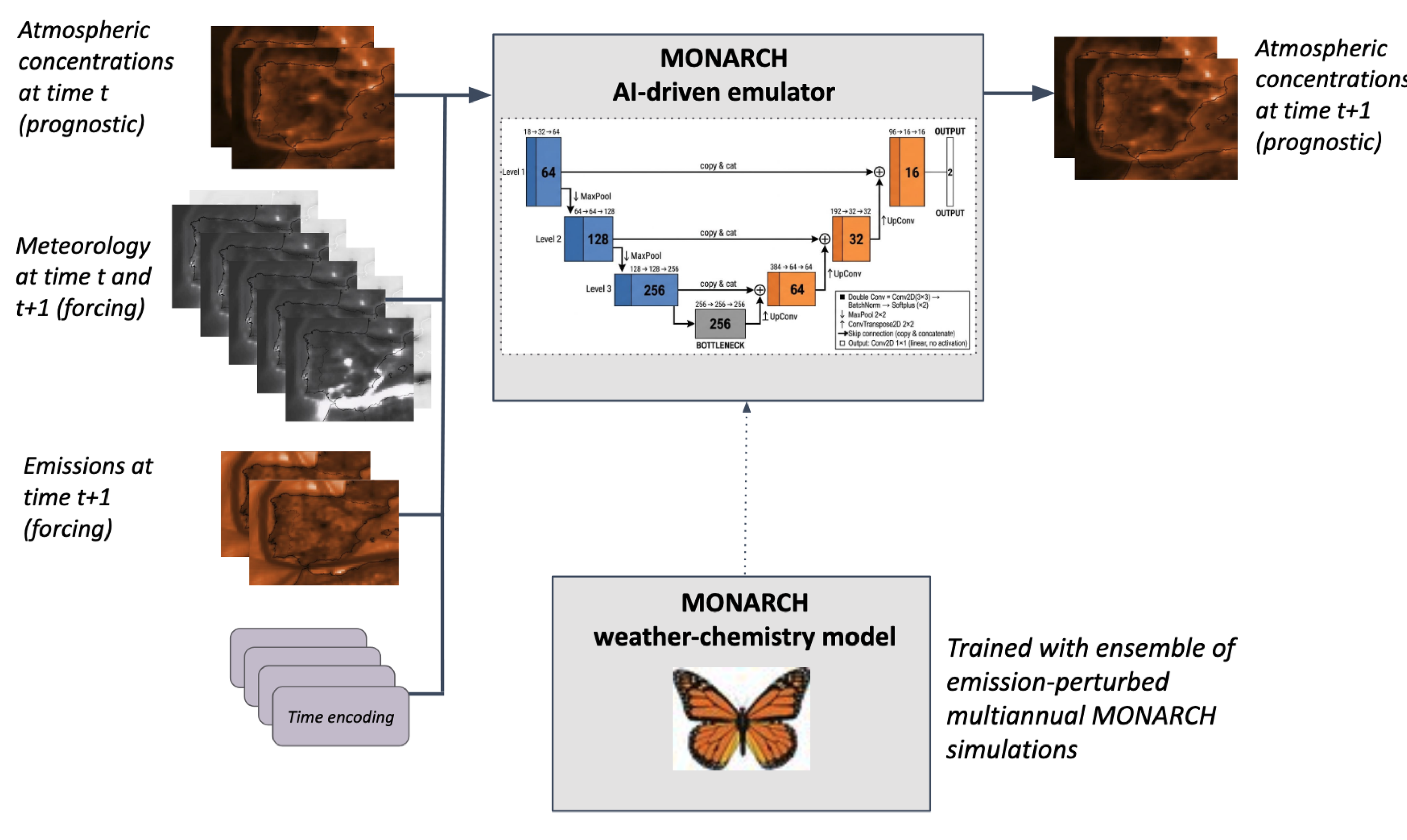

Figure 1: eMONARCH prediction pipeline: emissions, meteorology, and previous concentrations are fed to the U-Net, which predicts the next-hour concentration field.

Results

Single-step prediction (one-hour-ahead, U-Net) for NO_2 and NO on the *unseen test scenario* achieves:

Species	R^2	Slope	PCC	nRMSE	nMB
NO_2	97.6%	0.979	0.988	22%	≈0%
NO	90.5%	0.866	0.963	142%	−10%

Table 1: One-step U-Net evaluation metrics on the test emission scenario (member 0). Best tuning trial, 652 training epochs.

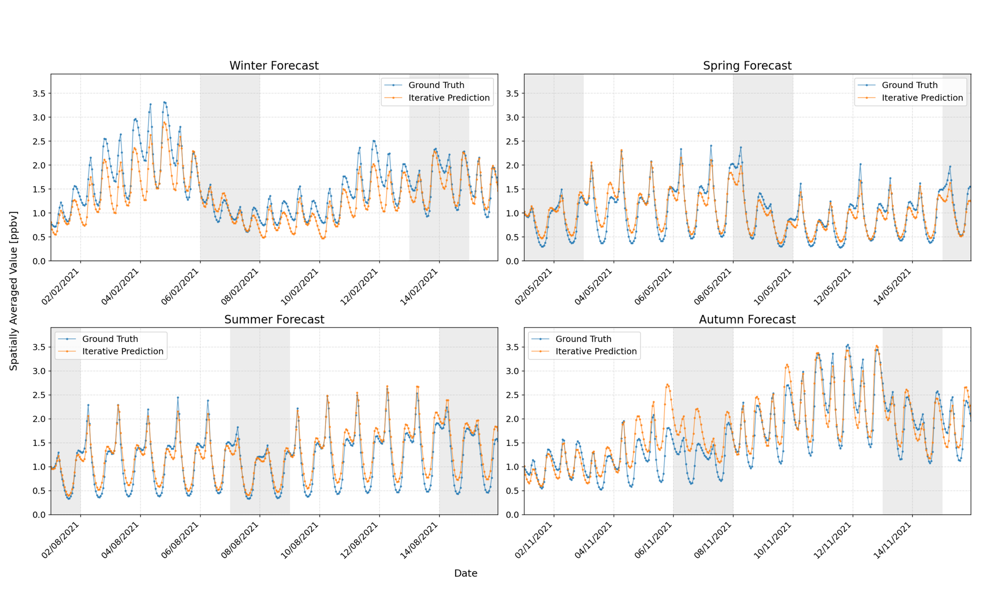

Figure 2: Autoregressive NO_2 seasonal cycle on the test scenario (member 0): domain-averaged MONARCH ground truth vs. 15-day rolling forecasts.

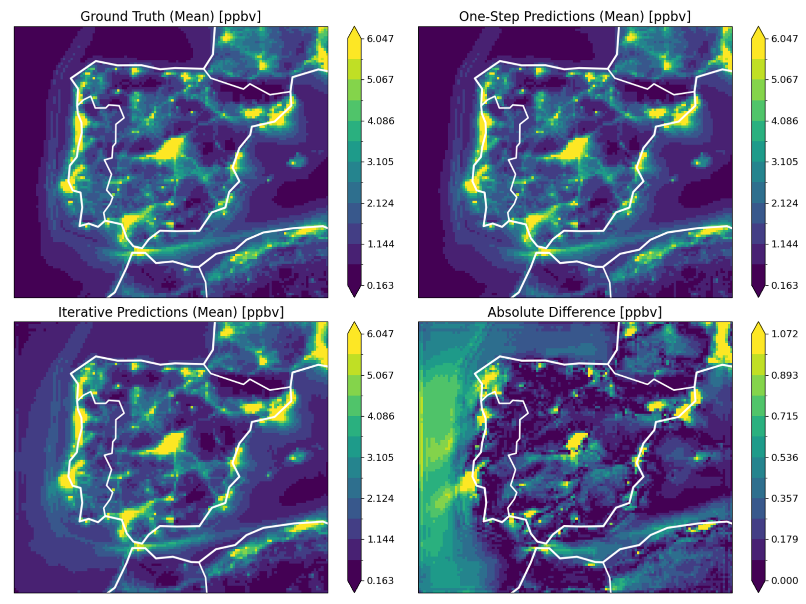

Figure 3: Autoregressive NO_2 map on the test scenario (member 0). Spatial distribution of predicted vs. MONARCH concentrations, averaged over a 15-day forecast window.

Autoregressive rollout (15-day iterative forecasts initialized every day of the year at hourly frequency) reveals a systematic **positive bias** in the most frequent concentration range (∼0.5–3.5 ppbv) and a **negative bias** for extreme values, resulting in a regression slope < 1.

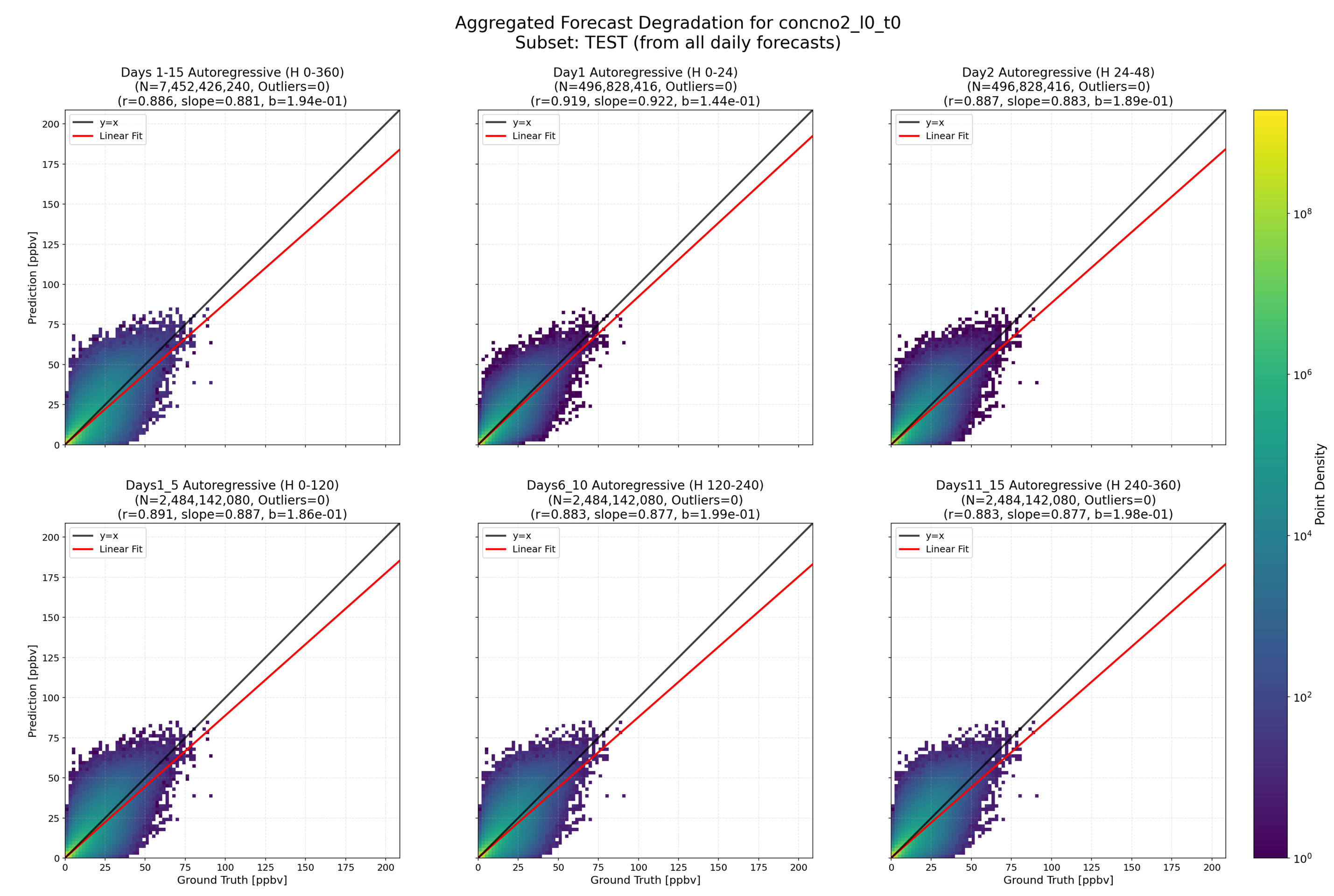

Figure 4: Predicted vs. MONARCH scatter density for NO_2 from 15-day autoregressive forecasts (all grid cells, all daily initialisations over the year). Panels decompose each forecast into individual early days and 5-day blocks. A clear drop in correlation (0.919→0.887) and slope (0.922→0.883) occurs between Day 1 and Day 2; beyond Day 2 the degradation plateaus.

Outlook & Next Steps

- **Bias mitigation:** AdamW with cosine learning rate scheduling and land-surface mask channels to reduce autoregressive drift.
- **Full-species extension:** Scaling from 2 output species (NO_2, NO) to 10. First full-species experiments show strong one-step performance on the test scenario (see table below), though autoregressive stability for the full set remains an open challenge.

Species	R^2	Slope
O_3	97.4%	0.974
$PM_{2.5}$	97.5%	0.995
PM_{10}	97.8%	0.966

Table 2: Full-species one-step metrics on the test scenario.

- **Transformer-enhanced U-Net:** Investigating self-attention and cross-attention mechanisms within the U-Net processor.
- **GNN track:** Early GNN results already show promising multi-step stability.

This poster is part of the R&D project AIRE, funded by the grant "Proyectos de Investigación en el Ámbito de la Inteligencia Artificial - 2025", Grant AIA2025-163866-C21 funded by MICIU/AEI/10.13039/501100011033. Additionally, the present work was funded through the RESPIRE project, which is part of the Recovery, Transformation and Resilience Plan (Plan de Recuperación, Transformación Resiliencia, PRTyR) funded by the European Commission to repair the harm caused by the Covid-19 crisis. We acknowledge the RES (AECT-2025-1-0012 and AECT-2025-2-0012) and EuroHPC (EHPC-AI-2024A06-068 and EHPC-REG-2025R01-219) for providing CPU and GPU computing resources. HP has received funding from the Ministerio de Ciencia e Innovación (MCIN) and Agencia Estatal de Investigación (AEI) through the Ramon y Cajal grant (RYC2021-034511-I, MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR).

Funded by the European Union