



EMEPai 1.0.0: A fast regional data-driven air-quality forecast model for Europe

Erlend Lima^{1,★}, Halvard Sutterud^{1,★}, Lewis Blake^{1,2}, Magnus Ulmoen^{1,2}, and Erik Askov Mousing¹

¹Norwegian Meteorological Institute, Oslo, Norway

²Statkraft, Oslo, Norway

★These authors contributed equally to this work.

Correspondence: Halvard Sutterud (halvard.sutterud@met.no)

Abstract. Air-quality forecasting over Europe relies largely on chemical transport models, whose computational cost limits forecast resolution and frequency. This study presents EMEPai, a regional data-driven forecasting system designed to emulate atmospheric composition fields produced by the EMEP/MSC-W modeling and data-assimilation framework. The model uses a graph-based encoder–processor–decoder architecture to generate three-hourly forecasts for NO₂, O₃, PM₁₀, PM_{2.5}, and SO₂ over Europe at 0.1° resolution, with lead times up to 96 hours.

Against surface observations, its performance exceeds that of the corresponding EMEP hindcast while approaching the skill of the assimilated reanalysis. Reduced skill is found for SO₂ and during high-concentration events. In an operational forecast setting, EMEPai performs comparably to or better than the CAMS regional ensemble across the evaluated pollutants and forecast lead times. A 96-hour forecast requires only seconds on a single GPU, with the full forecast workflow approximately 230–460 times faster than the corresponding EMEP/MSC-W workflow.

1 Introduction

Air quality refers to the composition of the atmosphere in terms of gases and airborne particles that affect human health and the environment (World Health Organization, 2021). Exposure to pollutants such as nitrogen dioxide (NO₂), ozone (O₃), particulate matter (PM_{2.5} and PM₁₀) and sulfur dioxide (SO₂) remains a major environmental health risk, contributing to excess mortality, increased incidence of cardiovascular and respiratory disease, and substantial social and economic costs (Targa et al., 2025; CAMS, 2025; GBD 2019 Risk Factors Collaborators, 2020; EEA, 2025). These pollutants arise from a combination of various activities and processes such as traffic, industrial activity, residential combustion and agricultural emissions. They undergo chemical reactions and are transported through the atmosphere by meteorological processes (CAMS, 2025; Seinfeld and Pandis, 2016). Air quality therefore exhibits strong, heterogeneous, spatial and temporal variability, with urban hotspots and episodic pollution events that can lead to regulatory exceedances and acute health impacts (CAMS, 2025; Baklanov and Zhang, 2020; European Parliament and Council of the European Union, 2024). Reliable information on present and future air quality is consequently important for vulnerable individuals, public-health advisories, regulatory compliance, and long-term environmental planning.



Operational air-quality forecasting is primarily based on chemical transport models (CTMs), which simulate the processes governing atmospheric composition (Baklanov and Zhang, 2020; Marécal et al., 2015). These models simulate emissions, advection, turbulent transport, chemical reactions and deposition through large coupled systems of differential equations, often involving hundreds of chemical reactions (Seinfeld and Pandis, 2016; Baklanov et al., 2014; Simpson et al., 2012).

In Europe, the Copernicus Atmosphere Monitoring Service (CAMS) employs a regional multi-model system in which eleven independent CTMs provide daily forecasts of key pollutants (Marécal et al., 2015; Flemming et al., 2015; Inness et al., 2019; Colette et al., 2025; Copernicus Atmosphere Monitoring Service, 2026). Each model produces daily forecasts of up to 96 hours on a 0.1° (~ 10 km) grid, initialized from an analysis that assimilates recent observations. The median of the eleven model forecasts constitutes the CAMS regional ENSEMBLE and serves as the operational multi-model air-quality forecast for Europe.

CTMs allow for a detailed representation of the underlying processes, and the CAMS regional product receives regular updates. However, their complexity comes with limitations. For example, their computational cost constrains both spatial resolution and forecast frequency, and systematic model errors and drift can degrade forecast skill. These limitations motivate data-driven forecasting approaches.

In the field of numerical weather prediction (NWP), data-driven models have provided effective for atmospheric forecasting. Models such as GraphCast (Lam et al., 2023) and AIFS (Lang et al., 2024a, b) use graph neural networks with encoder–processor–decoder architectures to produce autoregressive forecasts of meteorological fields with skill comparable to physics-based numerical weather prediction systems at a fraction of their computational cost.

Atmospheric-composition forecasting presents additional challenges. Air pollutants are governed by nonlinear chemistry involving large reaction networks and highly heterogeneous emissions (Harder et al., 2026). Consequently, machine-learning methods in air-quality applications have until recently been used mainly to augment existing numerical products (Méndez et al., 2023). For example, Shetty et al. (2025) used stacked gradient-boosted trees to downscale and bias-correct CAMS regional forecasts of $PM_{2.5}$ to daily 1 km grid, while Guion et al. (2026) combined CAMS reanalysis with observations and fine-scale spatial predictors to produce annual maps of several air-quality indicators on a 500 m grid. Such approaches can improve the spatial detail and accuracy of existing products, but remain dependent on an underlying CTM.

Beyond post-processing, several recent studies have explored data-driven atmospheric-composition forecasting directly. Aurora (Bodnar et al., 2025) showed that a large pretrained Earth-system model can be adapted to atmospheric composition by adding pollutant-specific input and output components and fine-tuning on CAMS global ensemble reanalysis and analysis. AIFS-COMPO (Harder et al., 2026) takes a more direct approach, jointly predicting meteorological variables and chemical species at three-hourly resolution on an ~ 80 km grid in a global forecasting system. For several species, AIFS-COMPO achieved performance comparable to or better than IFS-COMPO. AI-GAMFS provides global forecasts of aerosol properties such as optical depth and dust components, with a performance surpassing that of CAMS global forecast (Gui et al., 2026).

A parallel development is taking place at regional scales. Zeeman (Pang et al., 2026) is a deep-learning surrogate for the LOTOS–EUROS chemistry transport model that uses prescribed meteorological fields and emissions to autoregressively predict hourly values of multiple atmospheric constituents on a 0.1° grid. It reproduces the corresponding CTM simulations at



Table 1. Forcings used as inputs to EMEPai.

Category	Variables
Meteorology	Wind at 10 m height; temperature at 2 m height; surface stress; surface flux of sensible heat; surface flux of latent heat; surface pressure; planetary boundary layer height; convective precipitations; large scale precipitations.
Emissions	EMEPwRef2_v2.1C summed sectoral emissions for NO_x , SO_x , $\text{PM}_{\text{coarse}}$, and $\text{PM}_{2.5}$.
Spatiotemporal encodings	cos/sin of (Julian day, latitude, longitude, local time); cos of solar zenith angle.

substantially lower computational cost, but has so far been demonstrated only over a relatively small domain centered on the Netherlands. It is also limited by the inherent biases of the LOTOS–EUROS model, as it is trained on forecasts without assimilated observations. Taken together, these developments establish the feasibility of data-driven atmospheric-composition forecasting, but leave a gap between coarse-resolution global systems and regional models demonstrated over limited geographical domains.

In this work, we present EMEPai, a lightweight data-driven forecasting system designed for regional air quality forecasting over Europe. EMEPai is a deep-learning surrogate of the EMEP/MSC-W model and assimilation system (Simpson et al., 2012, 2025) based on the same broad class of encoder–processor–decoder architecture as AIFS-COMPO and Aurora, implemented within the Anemoi framework (Lang et al., 2024a; Keisler, 2022). Whereas existing global models resolve atmospheric composition on comparatively coarse grids, EMEPai predicts surface concentrations of $\text{PM}_{2.5}$, PM_{10} , O_3 , NO_2 , and SO_2 at the CAMS regional grid, with skill comparable to the CAMS regional ensemble. To our knowledge, EMEPai is the first purely data-driven, multi-pollutant forecasting system demonstrated across Europe at the same grid as the operational CAMS regional ensemble. The following section describes the forecasting setup, training procedure, and evaluation framework. We then present the retrospective and operational evaluation results and discuss their implications for regional data-driven air-quality forecasting.

2 Methods

2.1 Model setup and data

The EMEPai model is an autoregressive air-quality forecast model, predicting surface concentrations on a regional scale. At each forecast step, it takes as input forcings and prognostic variables from two consecutive times, t_{-3h} and t , and predicts prognostic variables at t_{+3h} . The prognostic variables are surface pollutant concentrations of $\text{PM}_{2.5}$, PM_{10} , O_3 , NO_2 , and SO_2 . The forcings comprise surface meteorological variables, sector-aggregated gridded emissions, and spatiotemporal encodings, as detailed in Table 1.

EMEPai uses a graph-based encoder–processor–decoder architecture implemented within the Anemoi framework (Lang et al., 2024a). The encoder maps fields from the regular latitude–longitude grid to a multiresolution latent graph, where graph-



transformer blocks exchange information spatially, before the decoder maps the latent representation back to the original grid. The model predicts an increment to the current pollutant state, which is added through a residual connection to obtain the next 3-hourly forecast state. Further architectural details are provided in Appendix E.

All fields are represented on a regular $0.1^\circ \times 0.1^\circ$ latitude-longitude grid covering (30° – 76° N, 30° W– 45° E). The domain corresponds to the standard EMEP CAMS 0.1° subdomain and extends slightly beyond the CAMS regional domain. The grid spacing was chosen to align with the CAMS Regional forecast product (Colette et al., 2025). All fields are sampled at 3-hourly intervals. This temporal resolution was chosen as a compromise between the computational and storage cost of producing the training data, and the stability of long forecasts (Schloer et al., 2026), as well as matching the temporal frequency of the meteorological input forcings.

The pollutant concentrations are generated by EMEP/MSC-W v5.4 Simpson et al. (2012). The EMEP/MSC-W fields used for training employ climatological boundary conditions. The meteorological fields are produced by ECMWF-IFS (ECMWF, 2025) using the default 2025 EMEP/MSC-W configuration, as documented in EMEP/MSC-W et al. (2025); Simpson et al. (2012). The emission forcings are specified in Appendix A.

The full dataset consists of 11 years of validated reanalyses (VRA; 2013–2023) and 22 years of hindcasts (1990–2012), totaling 425 GB. The data are split with 1990–2021 used for training, 2022 for validation during model training, and 2023 held out for final testing.

For each VRA year, 30% of the observation stations are withheld from the data assimilation, following the CAMS validation-station selection. These withheld stations are subsequently used to evaluate both EMEPai and the corresponding EMEP/MSC-W fields against observations.

2.2 Training

The model was trained on the LUMI high-performance computing cluster, with each training run using 32 Instinct MI250X GPUs. Training followed a two-stage procedure.

In the first stage, the model was trained to predict a single 3-hour forecast step from two preceding states, encouraging the model to learn consistent short-term dynamics before exposing it to errors generated by its own predictions.

In the second stage, the model was fine-tuned with autoregressive rollouts. Predictions from one forecast step were fed back as input to the subsequent step, and the loss was averaged over all rollout steps. This exposed the model to distributional shift arising from compounding forecast errors and improves stability during multi-step inference. Longer rollouts did not improve validation performance.

Optimization used the AdamW optimizer (Loshchilov and Hutter, 2019) together with a single-cycle cosine learning-rate schedule. The computational details are given in Table 2.

To reduce interference from SO_2 during training, its contribution to the loss was weighted by a factor of 0.2. This choice was pragmatic since reducing the SO_2 weight improved performance for the other target species without degrading SO_2 performance. Removing SO_2 from the loss entirely produced no further improvement.



Table 2. Overview of computational details for the two training stages of the model. All learning rates are per batch.

Aspect	Stage 1: pretraining	Stage 2: rollout finetuning
Training steps	150,000 steps	29,000 steps (10 epochs)
Prediction target	$(t_{-3h}, t) \rightarrow t_{+3h}$	$(t_{-3h}, t) \rightarrow t_{+3h}$ and $(t, t_{+3h}) \rightarrow t_{+6h}$
Learning rate	5×10^{-4}	5×10^{-6}
<i>Shared settings</i>		
Minimum LR	3.125×10^{-7}	
Batch size	32	
Optimization	AdamW, $\beta_1 = 0.9$, $\beta_2 = 0.95$	
LR scheduler	350-step linear warm-up followed by a cosine single cycle	
Loss function	Area-weighted mean-squared error	
Input scaling	Mean-std standardization	

2.3 Inference

At inference time, EMEPai is initialized from two consecutive atmospheric states and advanced autoregressively at 3-hour intervals. Forecasts are produced out to 96 hours, matching the forecast range of the CAMS Regional system (Colette et al., 2025). Meteorological and emission forcings for the corresponding forecast times are supplied externally throughout the roll-
 120 out. Boundary conditions are not explicitly imposed during EMEPai inference.

2.4 Evaluation

The 2023 test-year evaluation considers both agreement with the corresponding EMEP/MS-CW fields and performance against surface observations. For the 2023 test-year, EMEPai is compared with the corresponding EMEP/MS-CW hindcast (EMEP-hi), and validated reanalysis (EMEP-vra). The observational comparison uses station observations provided through the CAMS
 125 Regional evaluation service. The observations originate from the European Environment Agency up-to-date (EEA-UTD) air-quality data stream (European Environment Agency, 2023).

The CAMS Regional evaluation service applies a station filtering using the pollutant-specific objective station-classification system developed by Joly and Peuch (Joly, 2026, 2017; Joly and Peuch, 2012). For the 2023 test-year, evaluation is performed on the 30% of stations not part of the EMEP-vra assimilation process. No additional station filtering was applied in this study.
 130 The resulting evaluation set comprises 1598 stations for O_3 , 1576 for NO_2 , 1032 for PM_{10} , 626 for $PM_{2.5}$, and 666 for SO_2 .

2.5 Operational forecast

For the operational forecast, the model is run with the same setup as the regional CTMs constituting CAMS regional ENSEMBLE. On the morning of each day d , a five day forecast is run using the 12UTC ECMWF-IFS meteorology from day



$d - 1$. We use as initial fields the last two time steps of the EMEP/MSC-W analysis run for day $d - 2$, which incorporates the latest available observations. The same emission datasets as CAMS regional ENSEMBLE are used as forcings to ensure equal comparison in forecast mode, with anthropogenic emissions taken from CAMS-REG-ANT-v8.1 (Denier van der Gon et al., 2025), using the CAMS-TEMPO_v5.1 temporal profiles. See Appendix D for an illustration of the the setup. For evaluation of the operational forecast, all EEA-UTD stations are used.

3 Results

We evaluate EMEPai in two complementary settings. First, we assess its performance over the 2023 test year against the EMEP/MSC-W validated reanalysis reference fields (EMEP-vra), the corresponding EMEP/MSC-W hindcasts (EMEP-hi), and surface observations (EEA-UTD), considering spatial structure, temporal variability, and station-based statistics. We then examine its forecast skill in the operational setting, where EMEPai is compared with the CAMS regional ENSEMBLE forecast under operational initialization conditions.

3.1 Overall forecast skill

Figure 1 provides a representative example of the spatial structure of forecasts at 24 h lead time. Across all species, EMEPai reproduces the main spatial patterns of the corresponding EMEP-vra reference fields. For NO_2 , $\text{PM}_{2.5}$ and PM_{10} , EMEPai captures the key hotspots over western and central Europe, including the Benelux region, northern Italy, parts of south-eastern Europe, as well as the contrast between these regions and the background over northern and maritime areas. The O_3 fields show large-scale north–south and land–sea gradients, with EMEPai capturing the broad distribution of lower concentration over continental Europe compared to northern Africa, northern Europe, and over the ocean. SO_2 concentrations are generally low, with EMEPai identifying the main regions of higher concentration, but underestimating their magnitude. Across species, the EMEPai fields are smoother than the EMEP-vra reference fields, with weaker small-scale variability and less pronounced peaks.

Figure 2 shows time series of daily and monthly mean concentrations for the 2022 validation year and the 2023 test year, comparing EMEPai with EMEP-vra, EMEP-hi, and EEA-UTD observations. For O_3 , NO_2 , $\text{PM}_{2.5}$, and PM_{10} , EMEPai broadly reproduces the seasonal evolution and longer-term temporal variability of EMEP-vra. Several prominent pollution episodes are also captured, although differences remain in the magnitude of individual peaks. The monthly mean concentrations of EMEPai are generally close to those of EMEP-vra, with some periods of overprediction for particulate matter. The clearest exception is SO_2 , for which EMEPai exhibits a persistent negative bias, particularly during the 2022 validation year. Overall, the temporal behavior of EMEPai is generally closer to EMEP-vra than that of EMEP-hi across the two-year evaluation period.

The station-based statistics in Table 3 quantify the agreement with EEA-UTD observations during the 2023 test year, using only stations excluded from VRA assimilation. Across the evaluated pollutants, EMEPai performs better than EMEP-hi in correlation, RMSE, and absolute bias, and generally approaches the performance of EMEP-vra. For NO_2 , EMEPai achieves

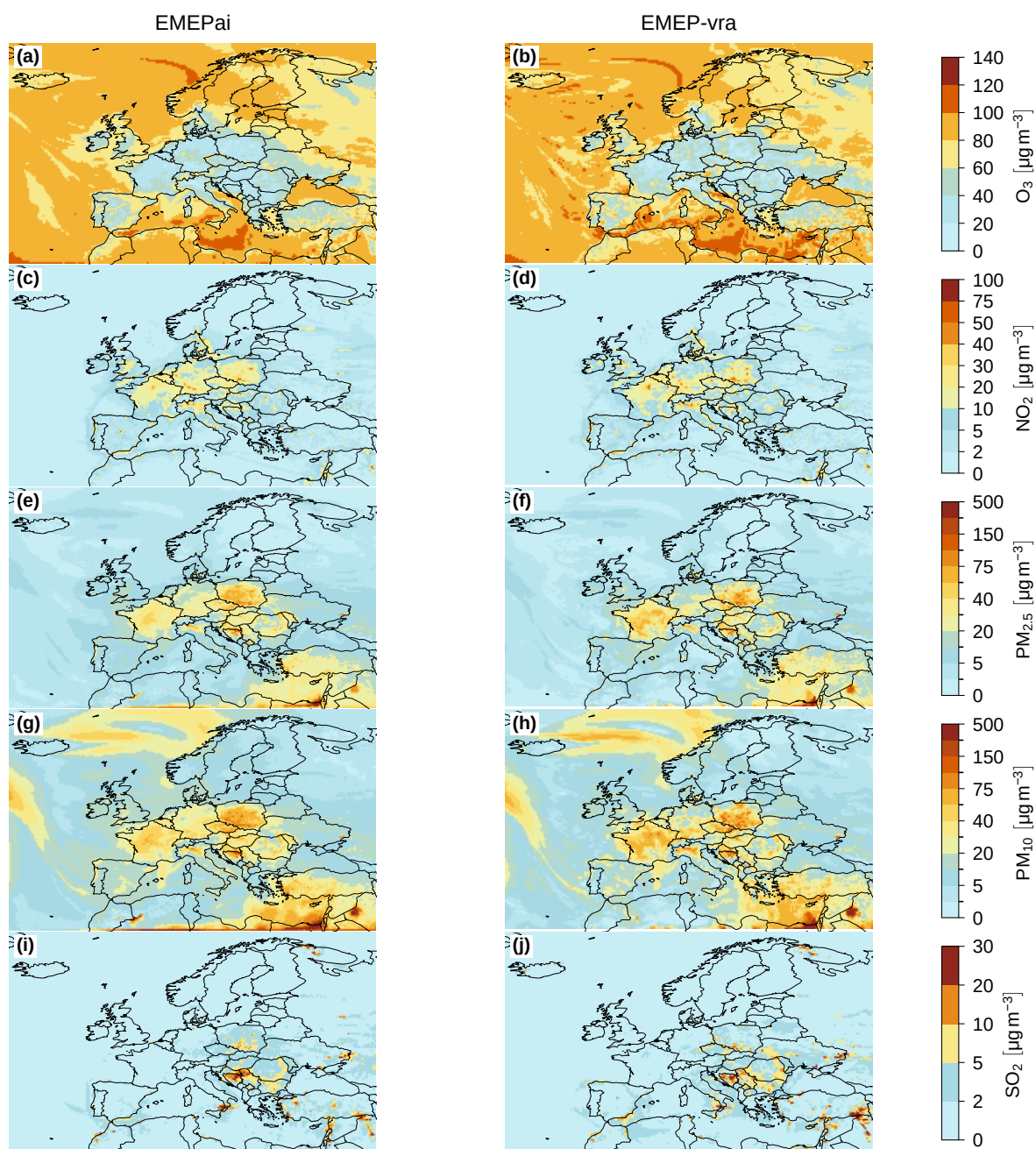


Figure 1. Comparison of EMEPai predictions and corresponding EMEP-vra reference fields over Europe at a 24 h lead time for 3 February 2023. Each row shows one chemical species, with EMEPai predictions on the left and EMEP-vra on the right. This date was selected as a representative case, with daily RMSE closest to the median daily RMSE over 2023.

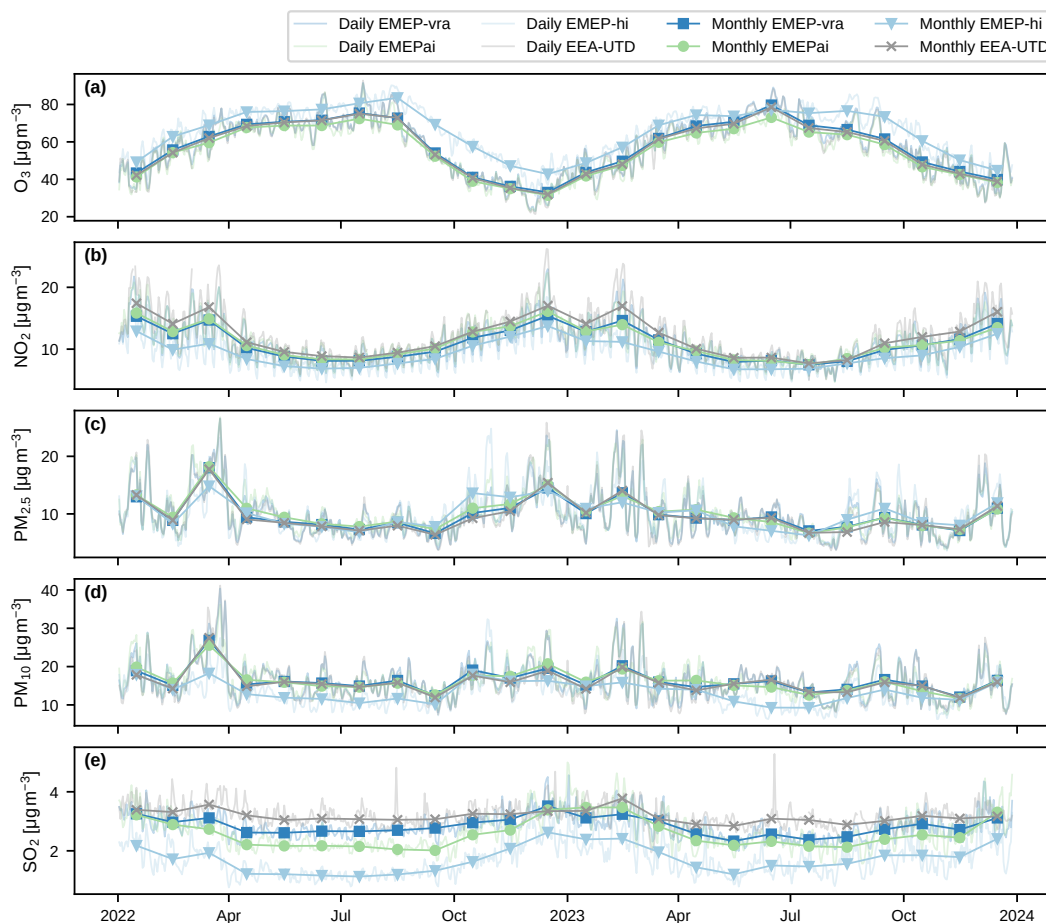


Figure 2. Intercomparison of EMEPai, EMEP-vra and EMEP-hi for both the validation year 2022 and test year 2023.

the highest correlation and the lowest RMSE, while also reducing the negative bias relative to EMEP-hi. For O_3 , the correlation and RMSE of EMEPai are close to those of EMEP-vra, despite a slight negative bias. For $PM_{2.5}$ and PM_{10} , the correlations of EMEPai lie between those of EMEP-hi and EMEP-vra, while its RMSE values are close to EMEP-vra and its mean biases remain small. Performance is weaker for SO_2 : EMEPai has a lower correlation and higher RMSE than EMEP-vra, consistent with the negative offset visible in Fig. 2. Taken together, the results show that EMEPai reproduces much of the observed temporal variability and concentration magnitude, with performance consistently exceeding EMEP-hi and, for most pollutants, approaching that of the assimilating EMEP-vra system.

3.2 Operational forecast skill

Table 4 shows evaluation of the operational setup described in Sect. 2.5 for the period November 16 2025 to June 31 2026. EMEPai matches or outperforms CAMS regional ENSEMBLE for all variables under the common forecast setup. For $PM_{2.5}$



Table 3. Pearson correlation coefficient r , RMSE, and mean concentration of EMEPai, EMEP hindcast (EMEP-hi), and EMEP validated reanalysis (EMEP-vra) relative to EEA-UTD observations (OBS) for the test period 2023. Normalised bias shown in parenthesis. The statistics are calculated over daily means against the evaluation stations not part of the EMEP-vra assimilation.

	r			RMSE [μg^{-3}]			Mean concentration [μg^{-3}]			
	EMEP-vra	EMEPai	EMEP-hi	EMEP-vra	EMEPai	EMEP-hi	EMEP-vra	EMEPai	EMEP-hi	OBS
O ₃	0.69	0.68	0.55	12.19	12.20	16.54	58.9 (2%)	56.0 (-3%)	65.3 (13%)	57.6
NO ₂	0.51	0.53	0.43	7.28	6.98	8.77	10.5 (-9%)	10.4 (-9%)	9.0 (-22%)	11.5
PM _{2.5}	0.69	0.61	0.52	3.66	4.19	5.33	9.3 (2%)	9.3 (3%)	9.4 (3%)	9.1
PM ₁₀	0.70	0.64	0.52	5.59	6.19	7.99	15.4 (2%)	15.1 (0.4%)	12.7 (-15%)	15.0
SO ₂	0.27	0.20	0.17	3.65	4.19	4.68	2.8 (-11%)	2.7 (-12%)	1.8 (-42%)	3.1

and PM₁₀, EMEPai has a positive bias, larger than the one seen in the evaluation set by $2.5\mu\text{g m}^{-3}$ and $3.7\mu\text{g m}^{-3}$, respectively, and overpredicts already large values from the single physics-based model member, represented by EMEP.

Figure 3 evaluates the forecast skill of the operational model using the median station-level temporal correlation at each forecast lead time. EMEPai outperforms or is competitive with the CAMS regional ENSEMBLE on all lead times. Because the forecasts are initialized at midnight, variations in performance with lead time also reflect the diurnal cycle profile of the models. For NO₂, all models exhibit reduced correlation during the morning and evening hours, possibly related to increased traffic emissions during rush hours, and at night for O₃, NO₂ and SO₂, with EMEPai correcting for some of the nocturnal performance loss. For PM₁₀, EMEP exhibits reduced forecast skill during the morning rush hour, while EMEPai maintains its performance.

Table 4. Results from evaluating operational EMEPai forecast against the operational CAMS forecast from EMEP/MS-CW (EMEP) and CAMS regional ENSEMBLE (ENS). Pearson correlation coefficient r , RMSE and mean concentration calculated against all EEA background stations (OBS) over daily values calculated from the first 24 hours of each forecast in the operational data period.

	r			RMSE [μg^{-3}]			Mean concentration [μg^{-3}]			
	ENS	EMEPai	EMEP	ENS	EMEPai	EMEP	ENS	EMEPai	EMEP	OBS
O ₃	0.81	0.85	0.81	15.10	12.42	15.13	59.9 (10%)	54.0 (-0.6%)	59.5 (10%)	54.3
NO ₂	0.64	0.68	0.61	8.16	7.00	8.33	8.8 (-30%)	11.3 (-9%)	9.4 (-25%)	12.4
PM _{2.5}	0.74	0.78	0.71	6.28	6.19	7.81	10.8 (3%)	12.5 (19%)	12.8 (22%)	10.5
PM ₁₀	0.69	0.71	0.66	9.82	9.97	10.71	15.3 (-10%)	19.6 (15%)	17.5 (3%)	17.0
SO ₂	0.22	0.31	0.19	5.59	5.26	6.42	1.9 (-51%)	3.2 (-19%)	2.5 (-37%)	4.0

3.3 Computational performance

A 96 h EMEPai forecast consists of 32 autoregressive steps and requires approximately 2.2 s of GPU execution time on a single NVIDIA A100 GPU. Including data loading and workflow overhead, the full forecast requires approximately 11 s

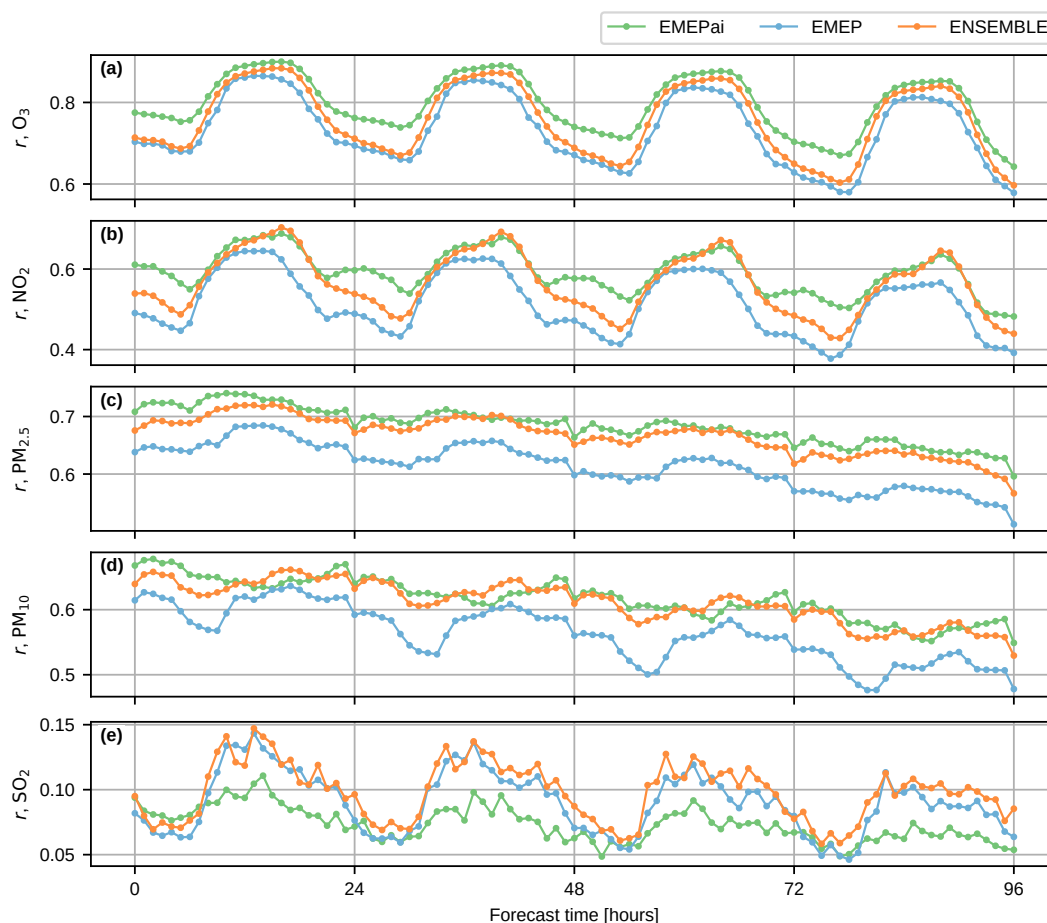


Figure 3. Median score of hour-of-day temporal Pearson correlation coefficient R of EMEPai, CAMS ENSEMBLE and EMEP forecasts for the operational data period, Nov 2025–June 2026. Correlation is calculated over the entire time period for each station and each hour of the 96 hour forecast runs, and the median is taken over EEA-UTD stations.

per initialization. For comparison, the corresponding EMEP/MSC-W forecast workflow requires approximately 0.7–1.4 h, depending on the computing system, making the full EMEPai workflow approximately 230–460 times faster. Further details of the computational comparison are given in Appendix B.

4 Discussion

By learning from historical assimilated fields, EMEPai can approximate the benefits of data assimilation without requiring concurrent observations at forecast time. Its performance relative to EMEP-hi suggests that these learned relationships recover information beyond that contained in the physics-based hindcast alone. The operational evaluation further indicates that this advantage largely persists under realistic forecast conditions.



As to the operational forecast, the quality of any forecast model depend on the forecast time. During the initial hours, the fields still contain information from the initial state, which incorporates observations from analysis and as such exhibits less bias than a hindcast. At later forecast times the forecast skill is degraded both due to model bias and due to meteorological uncertainty. In forecast mode the meteorological uncertainty affects the quality of the results more than for hindcast, as indicated by the median score plots 3 and C1 where the performance of both EMEPai and EMEP degrades faster in the operational runs than in the evaluation time period.

The large positive bias in particulate matter for operational forecast might be attributed to differences in anthropogenic emissions between training and inference, due to the difference in emission datasets for training and operational mode. During training the model has learned to correct for biases in the EMEP emission inputs, which might not be correct for the CAMS emissions inventories.

The weaker performance for SO₂ appears to reflect the difficulty of the target. SO₂ concentrations are generally low, with comparatively noisy observations close to the detection limits, which weakens the training signal. Under equal weighting, SO₂ degraded the performance of the other pollutants. Its contribution to the loss was therefore reduced to 20% in the final model, without noticeably degrading SO₂ performance.

The results position EMEPai as a deliberately specialized model for short-range regional surface air-quality forecasting. AIFS-COMPO matches or improves upon IFS-COMPO at some short lead times when evaluated against independent observations, with its strongest and most consistent gains found for aerosol optical depth. A clearer dependence on lead time appears in comparisons against CAMS analysis: for several aerosol variables, AIFS-COMPO tends to overtake IFS-COMPO after approximately the first forecast day, while results for surface gases are more mixed. Aurora shows a stronger increase in relative skill with lead time. CAMS retains an advantage for several near-surface species initially, whereas Aurora outperforms CAMS for a larger share of targets by day three (Bodnar et al., 2025).

When comparing EMEPai to AIFS and Aurora, EMEPai is competitive with or outperforms the CAMS regional ensemble from the start of the forecast. This is not a direct benchmark because the systems differ in spatial scale and evaluation setup. It nevertheless suggests that focusing model capacity on European surface concentrations can provide strong short-range skill. The narrower scope also allows EMEPai to operate at substantially finer resolution than the global models while retaining a lightweight architecture. Its use of explicit time-dependent emissions may also be beneficial at short lead times, since neither Aurora nor AIFS-COMPO currently receives equivalent emission forcing during inference.

Zeeman provides a closer regional comparison, but it is primarily designed to emulate LOTOS–EUROS simulations without assimilation. Instead, assimilation is proposed as a future development (Pang et al., 2026). EMEPai is trained partly on assimilated fields, allowing corrections represented by the reanalysis to persist during forecasts without requiring observations at inference time. This gives EMEPai a different objective from a direct CTM surrogate: it seeks to transfer some of the skill of an assimilated system into an operational forecast.

Explicit emissions are central to this specialization because they inform the model about source changes that cannot always be inferred from preceding concentration fields, particularly for short-lived pollutants. Zeeman also includes emissions explicitly (Pang et al., 2026), whereas AIFS-COMPO learns their influence indirectly and identifies explicit emission forcing as



a future development (Harder et al., 2026). EMEPai remains lightweight by prescribing operational meteorology rather than forecasting it jointly with atmospheric composition. Its intended product consists only of surface concentrations over Europe, so reproducing the full meteorological state would add substantial complexity without directly serving that objective. Developing and maintaining a joint meteorological forecasting system would also require broader expertise and a larger development team, whereas the present work can build on established operational meteorological forecasts and concentrate its effort on atmospheric composition.

4.1 Outlook

Future development of EMEPai will extend its chemical scope through additional species and a broader emission inventory. Its vertical representation may also be expanded by adding more model levels, while ensemble forecasting with a CRPS-based objective would allow forecast uncertainty to be represented explicitly and improve the fidelity of fine scale spatial structure. A separate version trained as a direct emulator of the EMEP system, to support scenario-based applications. Further increases in spatial resolution would broaden the range of processes and applications represented by the model. For operational forecasting, better consistency between the emission inventories used during training and inference may help reduce systematic biases.

5 Conclusions

EMEPai demonstrates that systematic corrections learned from assimilated EMEP-vra fields can be transferred to an autoregressive forecast model. The improvement over EMEP-hi indicates that EMEPai learns corrections contained in the reanalysis targets, while its performance relative to CAMS regional ENSEMBLE shows that these corrections remain useful when the model is run with forecast meteorology in a operational setting. Thus, EMEPai can reproduce some of the benefits of an assimilated modeling system without requiring observations from the forecast period, and can do so at low computational cost. A 96 h forecast requires approximately 11 s for the full EMEPai workflow, compared with approximately 0.7 h for the EMEP/MS-CW workflow.

The remaining errors show the limitations of the present approach. EMEPai produces smoother concentration fields than EMEP-vra, as is expected with deterministic, data-driven models trained with an area-weighted MSE-based loss function (Garg et al., 2026; Rasp and Thuerey, 2021). This reduces its ability to represent local maxima and high-concentration episodes. Performance is also weaker for SO₂. These deficiencies are particularly relevant for applications concerned with pollution exceedances rather than average forecast skill.

Within these limitations, EMEPai provides a computationally efficient approach to producing 96-hour air-quality forecasts over Europe while retaining much of the skill associated with the EMEP model and assimilation system.



Appendix A: Emission data sources

The anthropogenic emissions were based on the gridded EMEP emissions prepared for modelers by the Centre on Emission Inventories and Projections (2025), as used for the EMEP 2025 status report (MSC-W et al., 2025, Appendix B). Land-based emissions are based on the 2025 official national submissions to UNECE CLRTAP for 1990–2023. The EMEPwRef2_v2.1C emission set was used for primary particulate matter, in which condensable organics from small-scale combustion are included using expert estimates from the TNO Ref2 v2.1 inventory for GNFR sector C in selected countries. International shipping emissions within the EMEP domain are based on CAMS-GLOB-SHIP v3.2 for 2000–2023, with emissions for 1990–1999 extrapolated backwards using EDGAR v8.1 shipping trends. For each anthropogenic emission sector and each country, CAMS-TEMPO_v5.1 hour-of-day, day-of-week and monthly emission factors are used to convert annual totals to hourly emission values, as well as heating degree days. Biogenic emissions are not included explicitly, so the model must infer their influence on atmospheric concentrations from the training data.

Appendix B: Computational performance

This appendix summarizes the computational cost of applying EMEPai after training, and compares this cost with the corresponding runtime of the EMEP/MS-CW. The purpose of the comparison is not to treat EMEPai and EMEP/MS-CW as equivalent modeling systems, as EMEP/MS-CW solves a substantially more complete chemical transport problem. The comparison instead characterizes the marginal computational cost of producing forecasts for the target species and forecast task evaluated in this paper, once the learned model has been trained.

EMEPai inference was run on a single NVIDIA A100 GPU, with each forecast initialization processed as an independent job. For a 96 h forecast with 3-hourly output, the rollout consists of 32 autoregressive steps. The model execution time was 56 ± 1 ms per step, corresponding to 2.2 s of GPU time per forecast initialization. The full workflow required 11 s per forecast initialization, with the extra overhead coming from loading the python environment and setting up the dataset loaders. A serial run of daily 96 h forecasts for one year takes approximately 70 min.

For comparison, a CAMS2_40 4-day EMEP/MS-CW forecast requires approximately 1.4 h on Nebula (National Supercomputer Centre, Linköping University) using 128 physical CPU cores from Intel Xeon Gold 6130 CPUs at 2.1 GHz. The corresponding runtime on Fahrenheit is approximately 0.7 h using 144 physical CPU cores on one dual-socket AMD EPYC 9565 node. The EMEP/MS-CW forecast timing includes one additional spin-up day without output, and is not a one-to-one comparison with EMEPai inference.

Using the faster Fahrenheit timing as reference, the full EMEPai workflow is approximately 230 times faster than the corresponding EMEP/MS-CW forecast workflow. Using the Nebula timing gives a factor of approximately 460. Equivalently, a serial year of daily 96 h forecasts would require approximately 70 min with EMEPai, compared with approximately 10.6 day for the Fahrenheit EMEP/MS-CW workflow and 21.3 day for the Nebula workflow.

The training cost of EMEPai depends on the model configuration and hardware. The final model was trained for 150 000 one-step optimization steps, followed by 29 000 additional optimization steps using a two-step rollout, requiring 45h on 32 AMD



MI250X GPUs on LUMI. This offline cost does not include the cost of producing the EMEP/MSC-W-based training archive. The computational advantage of EMEPai is therefore primarily an inference advantage. Once trained, the model produces new forecasts at low marginal cost.

295 **Appendix C: Additional figures**

Figure C1 shows median scores for the evaluation period, comparing EMEPai with EMEP-vra and EMEP-hi. For the two latter, each run spans a full year which is split into daily forecasts and colocated with the EMEPai runs. The result is a diurnal profile rather than a forecast evaluation. EMEPai runs starting from both hindcast fields and validated reanalysis fields are shown, which shows that the effect of initial conditions quickly disappears.

300 Figure C2 shows time series of daily and monthly mean concentrations for the operational time period, comparing EMEPai with EMEP-vra, EMEP-hi, and EEA-UTD observations. The figure shows temporal consistency of the model biases.

Appendix D: Operational setup

Figure D1 shows the operational EMEPai setup and its dependency on the CAMS regional forecast. Operational EMEP forecasts are produced from the latest available EMEP analysis and meteorological forecast fields. The meteorological forecast
 305 initialized at 12:00 UTC on day -2 is used until the next forecast becomes available at 12:00 UTC on day -1 . EMEPai is initialized directly from the latest available EMEP analysis state, using the two most recent 3 h concentration states as input, corresponding to x_t and x_{t-1} . Following a 24 h spin-up period, the first valid forecast fields are available at 03:00 UTC on day 0. Model performance is evaluated over day 0, corresponding to the first forecast day after initialization.

Appendix E: Model architecture

310 The EMEPai model uses an encoder–processor–decoder architecture, following the general design introduced in AIFS (Lang et al., 2024a) and illustrated in Fig. E1. The model represents the atmospheric and chemical state as feature vectors attached to nodes of a graph. The encoder, processor, and decoder are distinct graph-neural networks applied sequentially, each built from graph transformer blocks acting on graphs. The encoder maps information locally from the data grid to a latent graph, the processor updates this latent representation and propagates information spatially, and the decoder maps the latent state back
 315 locally to the data grid.

The architecture uses two sets of nodes. The data nodes are constructed from the data grid, with one node associated with each grid point. The input fields are represented as a feature vector at each data node, obtained by stacking the input variables and concatenating the node’s attributes. The latent nodes are defined on a fixed, multi-resolution icosahedral discretization of the sphere, following Lam et al. (2023). Each latent node carries a learned latent state vector, whose feature dimension is
 320 independent of the resolution of the data grid. Edges from all refinement levels are retained, forming a multi-scale graph that allows information to be exchanged across spatial scales.

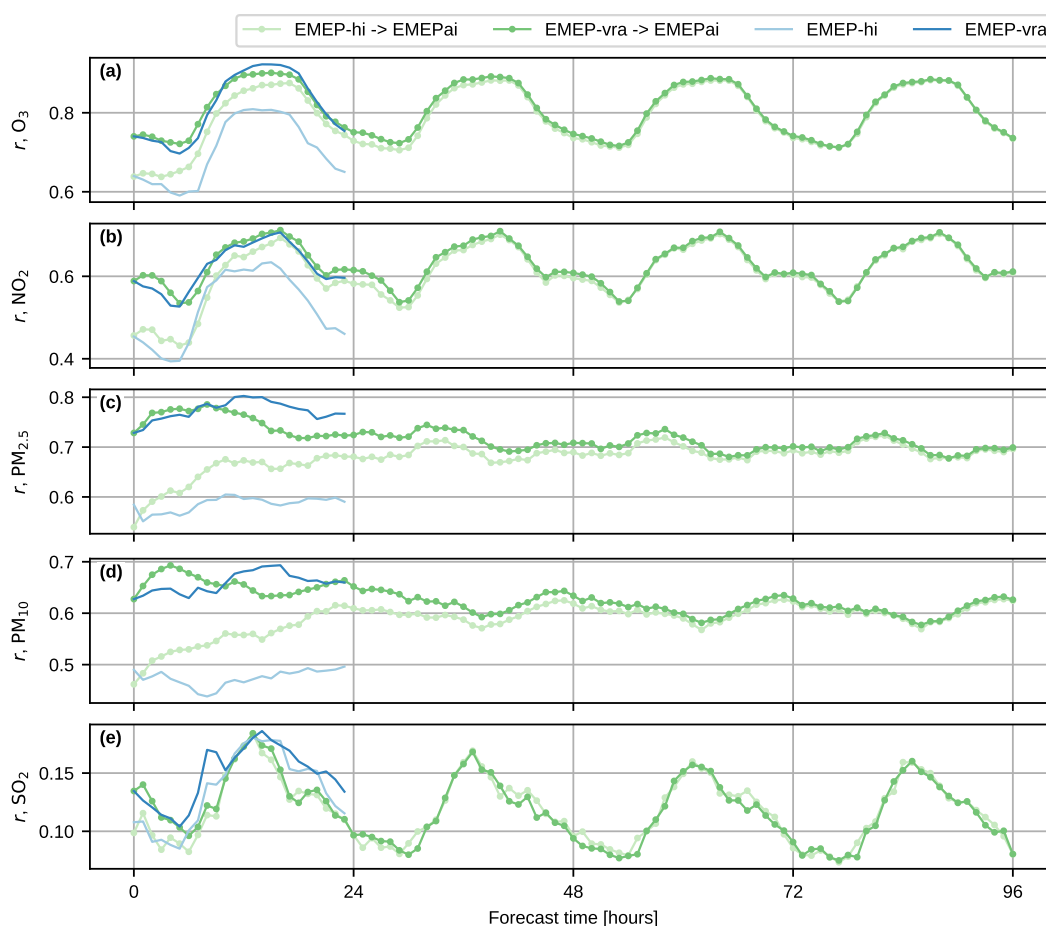


Figure C1. Median score of hour-of-day temporal Pearson correlation coefficient R of EMEPai, EMEP-vra and EMEP-hi for the test year 2023. Correlation is calculated over the entire time period for each station and each hour of the 96 hour forecast runs for EMEPai. The median is taken over the EMEP validated reanalysis evaluation station set for 2023. EMEP-hi -> EMEPai and EMEP-vra -> EMEPai indicates EMEPai with initial fields from EMEP-hi and EMEP-vra, respectively.

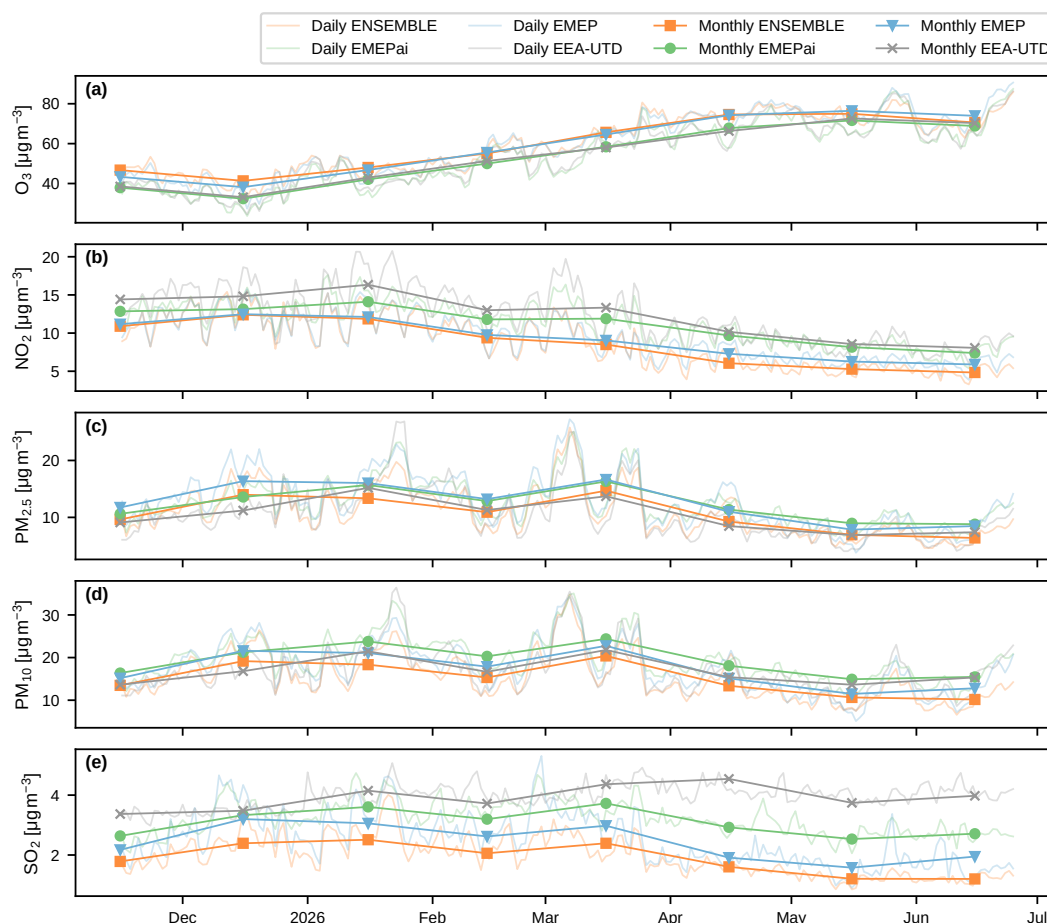


Figure C2. Intercomparison of ENSEMBLE, EMEPai and EMEP against EEA-UTD observations.

The graph transformer blocks can be viewed as learned, state-dependent message-passing operators. At each block, a node updates its feature vector by aggregating information from neighboring nodes, where the aggregation weights depend on the current node features and edge attributes. Stacking several such blocks allows information to propagate over increasing distances on the graph. In this sense, the model learns effective spatiotemporal correlations that map the current atmospheric state and forcing fields to the next forecast state.

The encoder maps the data-grid state to the latent nodes using a bipartite encoder graph. Each latent node is connected to all data nodes within a specified great-circle distance. The encoder aggregates information from this local data-node neighborhood and edges, and produces an initial latent state at each latent node.

The processor then acts on the latent graph. It updates the latent state at each latent node by attending to neighboring latent nodes on the multi-scale mesh. Information that was initially local on the data grid can be exchanged through the latent graph representation.

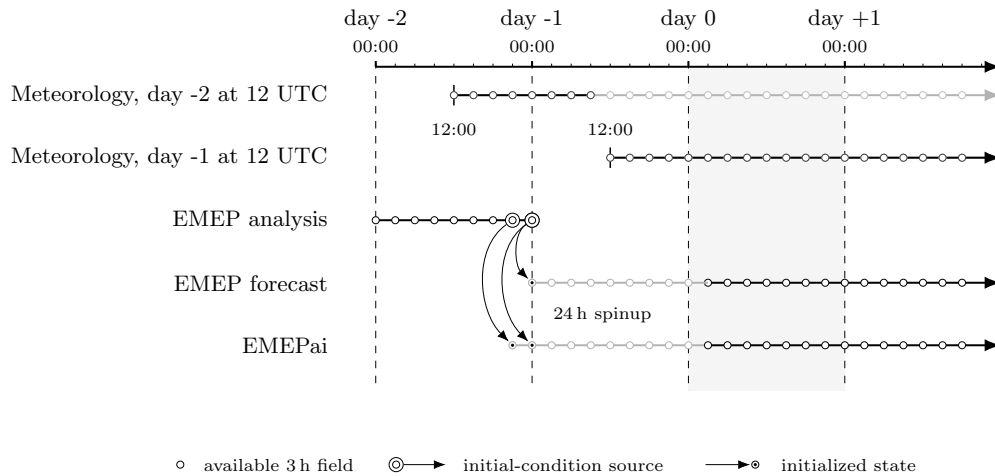


Figure D1. Schematic of the operational forecasting workflow used for model evaluation. Open circles denote available 3 h states, double circles mark states used to initialize downstream models, filled circles mark the resulting initial-condition states, and arrows indicate initialization transfers.

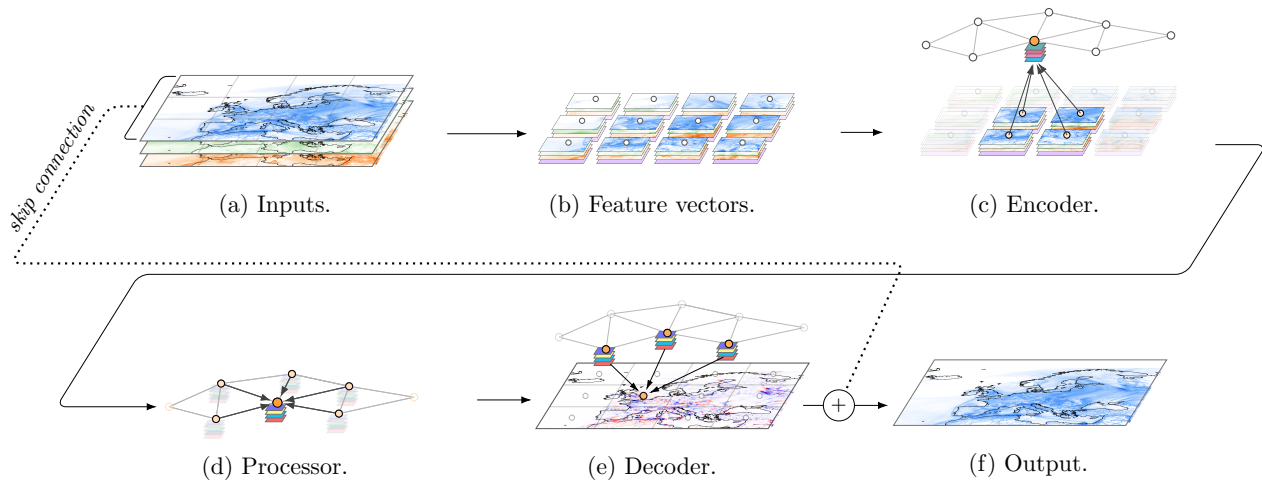


Figure E1. Schematic overview of the encoder–processor–decoder architecture. (a) Input variables are defined as fields on the data grid. (b) Input fields and node attributes are concatenated into a feature vector at each induced data node. (c) The encoder maps from data nodes to latent nodes using a bipartite graph. (d) The processor updates the latent state on the latent graph through message passing between neighboring latent nodes. (e) The decoder maps the updated latent state back to the data grid using a bipartite graph and predicts an increment $\widehat{\Delta x}$ on the data grid. (f) The predicted increment is added to the target input state through a skip connection, yielding the updated state $x_t + \widehat{\Delta x}$.



Finally the decoder maps the updated latent state back to the data grid using a second bipartite graph. Each data node receives information from its $k = 3$ nearest latent nodes, giving a localized mapping from the latent representation to the data-grid state. The decoder output is interpreted as a predicted state increment, $\widehat{\Delta x}$, which is added to the input state through a skip connection. The model therefore predicts $x_t + \widehat{\Delta x}$, avoiding reconstructing the full output state directly.

Appendix F: Ablation

To assess the sensitivity of the model to different input fields, an ablation study was performed. Selected variables or variable groups were removed from the input, and the model was retrained using the same configuration. Figure F1 shows the annual change in nRMSE relative to the mean performance of three baseline seeds. Uncertainty associated with the composition of the validation period was estimated by paired bootstrap over forecast dates and is shown as confidence intervals.

In addition to single-variable ablations, grouped ablations were performed for selected sets of physically related and potentially correlated predictors. These groups were intended as a diagnostic test of broader information classes where individual variables may be partly redundant. The mixing group contains PBLH and surface stress. The precipitation group contains convective and large-scale precipitation. The energy-exchange group contains latent heat flux, sensible heat flux, and 2 m temperature. The model relies strongly on wind, emissions, temporal encodings, insolation, and spatial encodings. Wind removal gives the largest degradations for all species except for SO_2 , as expected from dependence on transport. Emissions are important for NO_2 , PM_{10} , $\text{PM}_{2.5}$, and SO_2 , while the weaker response for O_3 is consistent with its secondary and chemically nonlinear character.

The grouped ablations have much larger effects than the corresponding single-variable ablations. This is clearest for the mixing group, where removing either PBLH or surface stress has little effect for most species, while removing both variables is strongly detrimental. At the same time, removing surface stress improves O_3 , PM_{10} , and $\text{PM}_{2.5}$. Similar compensation effects are seen for the precipitation and energy-exchange groups, although with smaller magnitudes. The model can compensate for the removal of some individual predictors, but not for the removal of the broader information class.

Overall, the ablation study shows that the current predictor set is not minimal. Further refinement of the input set is left for future model development.

Code and data availability. The model used to produce the results used in this paper, EMEPai 1.0.0, is archived on repository under 10.5281/zenodo.22955196 (Sutterud et al., 2026). This includes the training configurations, model checkpoint, input data and scripts to run the model for all the simulations presented in this paper. Data and products produced by the Norwegian Meteorological Institute (MET Norway) are licensed under the Norwegian License for Open Government Data (NLOD) 2.0 and under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. The emission data are based on the EMEP/CEIP gridded emission inventory, which is published under the Creative Commons Attribution 4.0 (CC BY 4.0) license (EMEP Centre on Emission Inventories and Projections, 2026). The EEA-UTD observation data is published by the EEA under the ODC-by license, and can be retrieved from <https://www.eea.europa.eu/en/datahub/>

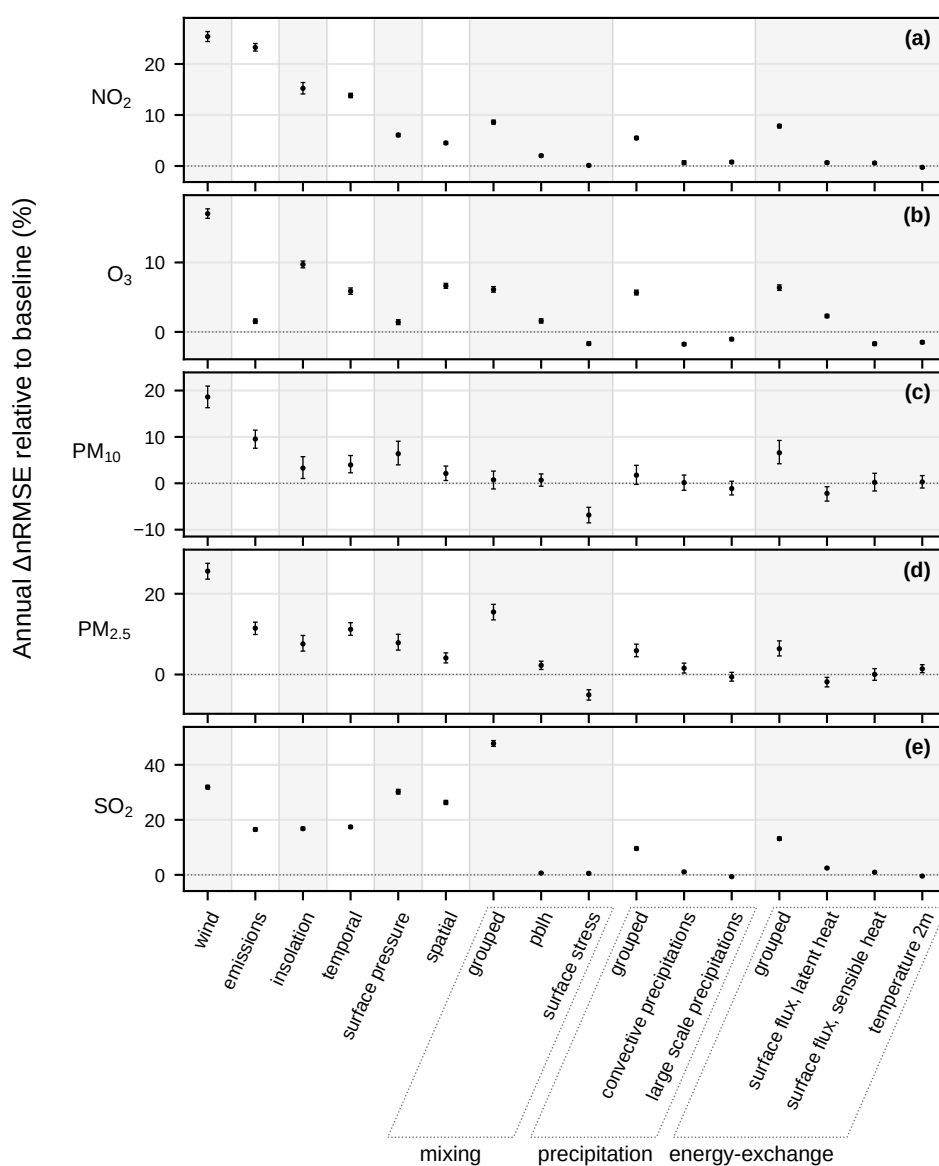


Figure F1. Ablation effects with paired-bootstrap confidence intervals. The x-axis shows the ablated input variables or variable groups, and each panel shows the annual change in nRMSE relative to the mean baseline performance. Dotted boxes indicate grouped ablations, with the grouped case shown alongside its constituent variables. Negative values indicate improved performance relative to baseline.



365 datahubitem-view/778ef9f5-6293-4846-badd-56a29c70880d. CAMS forecast model output is available under the CC-BY license with DOI
10.24381/a4005cee (Copernicus Atmosphere Monitoring Service, 2020).

Author contributions. EM, EL, HS, MU, and LB contributed to the conceptualization and methodology of the study and to software development. EL and HS performed the formal analysis, validation, and visualization. EM, HS, MU, and LB contributed to data curation. EM supervised and administered the project, provided resources, and acquired the funding. EM, EL, and HS prepared the manuscript draft. All authors reviewed and edited the manuscript.

370 *Competing interests.* The authors declare that they have no conflict of interest.

Acknowledgements. The training of EMEPai was performed on resources provided by Sigma2 – the National Infrastructure for High-Performance Computing and Data Storage in Norway. We acknowledge the EuroHPC Joint Undertaking for awarding project ID EHPC-AIF-2025SC03-032 access to the EuroHPC supercomputer LUMI. LUMI is hosted by CSC (Finland) and the LUMI consortium. We would like to thank the operational team at MET Norway for help generating the training data and setting up the pre-operational pipeline. The
375 authors acknowledge the use of products and tools provided or partly supported by the Copernicus Atmosphere Monitoring Service (CAMS / <https://atmosphere.copernicus.eu>), and in particular CAMS Regional production (ENSEMBLE analyses/reanalyses) and the CAMS Regional evaluation service.



References

- Baklanov, A. and Zhang, Y.: Advances in air quality modeling and forecasting, *Global Transitions*, 2, 261–270, <https://doi.org/https://doi.org/10.1016/j.glt.2020.11.001>, 2020.
- Baklanov, A., Schlünzen, K., Suppan, P., Baldasano, J., Brunner, D., Aksoyoglu, S., Carmichael, G., Douros, J., Flemming, J., Forkel, R., Galmarini, S., Gauss, M., Grell, G., Hirtl, M., Joffre, S., Jorba, O., Kaas, E., Kaasik, M., Kallos, G., Kong, X., Korsholm, U., Kurganskiy, A., Kushta, J., Lohmann, U., Mahura, A., Manders-Groot, A., Maurizi, A., Moussiopoulos, N., Rao, S. T., Savage, N., Seigneur, C., Sokhi, R. S., Solazzo, E., Solomos, S., Sørensen, B., Tsegas, G., Vignati, E., Vogel, B., and Zhang, Y.: Online coupled regional meteorology chemistry models in Europe: current status and prospects, *Atmospheric Chemistry and Physics*, 14, 317–398, <https://doi.org/10.5194/acp-14-317-2014>, 2014.
- Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Allen, A., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J. A., Dong, H., Gupta, J. K., Thambiratnam, K., Archibald, A. T., Wu, C.-C., Heider, E., Welling, M., Turner, R. E., and Perdikaris, P.: A foundation model for the Earth system, *Nature*, 641, 1180–1187, <https://doi.org/10.1038/s41586-025-09005-y>, 2025.
- CAMS: CAMS Assessment Report on European Air Quality in 2024, Tech. rep., Copernicus Atmosphere Monitoring Service, <https://doi.org/10.24380/pdtn-dc12>, 2025.
- Centre on Emission Inventories and Projections: Emissions as used in EMEP models, <https://www.ceip.at/webdab-emission-database/emissions-as-used-in-emep-models>, accessed 23 June 2026, 2025.
- Colette, A., Collin, G., Besson, F., Blot, E., Guidard, V., Meleux, F., Royer, A., Petiot, V., Miller, C., Fermond, O., Jeant, A., Adani, M., Arteta, J., Benedictow, A., Bergström, R., Bowdalo, D., Brandt, J., Briganti, G., Carvalho, A. C., Christensen, J. H., Couvidat, F., D’Elia, I., D’Isidoro, M., Denier van der Gon, H., Descombes, G., Di Tomaso, E., Douros, J., Escribano, J., Eskes, H., Fagerli, H., Fatahi, Y., Flemming, J., Friese, E., Frohn, L., Gauss, M., Geels, C., Guarnieri, G., Guevara, M., Guion, A., Guth, J., Hänninen, R., Hansen, K., Im, U., Janssen, R., Jeoffrion, M., Joly, M., Jones, L., Jorba, O., Kadantsev, E., Kahnert, M., Kaminski, J. W., Kouznetsov, R., Kranenburg, R., Kuenen, J., Lange, A. C., Langner, J., Lannuque, V., Macchia, F., Manders, A., Mircea, M., Nyiri, A., Olid, M., Pérez García-Pando, C., Palamarchuk, Y., Piersanti, A., Raux, B., Razinger, M., Robertson, L., Segers, A., Schaap, M., Siljamo, P., Simpson, D., Sofiev, M., Stangel, A., Struzewska, J., Tena, C., Timmermans, R., Tsikerdekis, T., Tsyro, S., Tyuryakov, S., Ung, A., Uppstu, A., Valdebenito, A., van Velthoven, P., Vitali, L., Ye, Z., Peuch, V.-H., and Rouil, L.: Copernicus Atmosphere Monitoring Service – Regional Air Quality Production System v1.0, *Geoscientific Model Development*, 18, 6835–6883, <https://doi.org/10.5194/gmd-18-6835-2025>, 2025.
- Copernicus Atmosphere Monitoring Service: CAMS European Air Quality Forecasts, <https://doi.org/10.24381/A4005CEE>, 2020.
- Copernicus Atmosphere Monitoring Service: CAMS European air quality forecasts, <https://ads.atmosphere.copernicus.eu/datasets/cams-europe-air-quality-forecasts>, accessed 2026-08-28, 2026.
- Denier van der Gon, H., Gauss, M., Granier, C., Guevara, M., Arellano, S., Sindelarova, K., Liaskoni, M., Markova, J., Prieto Perez, A., Bartik, L., Jalkanen, J.-P., Majamäki, E., Quack, B., Bouarar, I., Li, C., Liousse, C., Simpson, D., Olivíé, D., Benedictow, A., Segers, A., Zilbermann, N., Dellaert, S., Kuenen, J., Schoenmakers, E., Super, I., and Krueger, K.: Documentation of CAMS Emission Inventory Products, Documentation, Copernicus Atmosphere Monitoring Service, <https://doi.org/10.24380/UAG-0SVT>, 2025.
- ECMWF: IFS Documentation, Tech. rep., European Centre for Medium-Range Weather Forecasts, available from <https://www.ecmwf.int/en/publications/ifs-documentation>, 2025.
- EEA: Preventing Cardiovascular Disease Through a Healthy Environment, Tech. Rep. Briefing 12/2025, European Environment Agency, ISBN 978-92-9480-738-0, ISSN 2467-3196, <https://doi.org/10.2800/5535347>, 2025.



- 415 EMEP Centre on Emission Inventories and Projections: EMEP/CEIP 2026 Gridded emissions, <https://www.ceip.at/the-emep-grid/gridded-emissions>, 2026.
- EMEP/MSC-W, CCC, CEIP, and CIAM: Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components, Tech. Rep. EMEP Status Report 1/2025, EMEP MSC-W, https://www.emep.int/mscw/mscw_publications.html, joint MSC-W, CCC, CEIP & CIAM Report, 2025.
- 420 European Environment Agency: Up-to-date air quality data, <https://www.eea.europa.eu/en/analysis/maps-and-charts/up-to-date-air-quality-data>, accessed 23 June 2026, 2023.
- European Parliament and Council of the European Union: Directive (EU) 2024/2881 of the European Parliament and of the Council of 23 October 2024 on ambient air quality and cleaner air for Europe (recast), <https://eur-lex.europa.eu/eli/dir/2024/2881/oj/eng>, published in OJ L, 20.11.2024, 32024L2881, p. 16, 2024.
- 425 Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A.-M., Diamantakis, M., Engelen, R. J., Gaudel, A., Inness, A., Jones, L., Josse, B., Katragkou, E., Marecal, V., Peuch, V.-H., Richter, A., Schultz, M. G., Stein, O., and Tsikerdekis, A.: Tropospheric chemistry in the Integrated Forecasting System of ECMWF, *Geoscientific Model Development*, 8, 975–1003, <https://doi.org/10.5194/gmd-8-975-2015>, 2015.
- Garg, P., Gergel, D. R., Shao, A. E., and Yacalis, G. J.: The Recipe Matters More Than the Kitchen: Mathematical Foundations of the AI Weather Prediction Pipeline, <https://arxiv.org/abs/2604.01215>, 2026.
- 430 GBD 2019 Risk Factors Collaborators: Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, *The Lancet*, 396, 1223–1249, [https://doi.org/10.1016/S0140-6736\(20\)30752-2](https://doi.org/10.1016/S0140-6736(20)30752-2), 2020.
- Gui, K., Zhang, X., Che, H., Li, L., Zheng, Y., An, L., Miao, Y., Zhao, H., Dubovik, O., Holben, B., Wang, J., Gupta, P., Lind, E. S., Toledano, C., Wang, H., Wang, Z., Wang, Y., Huang, X., Dai, K., Xia, X., Xu, X., and Zhang, X.: Advancing Operational Global Aerosol Forecasting with Machine Learning, *Nature*, 651, 658–665, <https://doi.org/10.1038/s41586-026-10234-y>, 2026.
- 435 Guion, A., Gressent, A., Descombes, G., Janati, Y., Real, E., Ung, A., Meleux, F., Schucht, S., and Colette, A.: High-Resolution Mapping of Air Quality across Europe: An Ensemble Machine and Deep Learning Framework Integrating Multi-Scale Spatial Predictors (CHROMAP v1.0), *EGUsphere*, pp. 1–35, <https://doi.org/10.5194/egusphere-2026-1109>, 2026.
- Harder, P., Flemming, J., Alexe, M., Mertes, G., Raoult, B., and Chantry, M.: AIFS-COMPO: A Global Data-Driven Atmospheric Composition Forecasting System, <https://doi.org/10.48550/ARXIV.2604.03300>, version Number: 1, 2026.
- 440 Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, *Atmospheric Chemistry and Physics*, 19, 3515–3556, <https://doi.org/10.5194/acp-19-3515-2019>, 2019.
- 445 Joly, M.: Actualisation de la classification des sites de mesure, Technical report, CNRM, Météo-France, https://opensource.umr-cnrm.fr/attachments/download/2532/2017_revision_document.pdf, 2017.
- Joly, M.: Objective Classification CAMS2_40_bis Update, Tech. rep., CNRM, <https://opensource.umr-cnrm.fr/attachments/6758>, revision dated 19 February 2026, 2026.
- Joly, M. and Peuch, V.-H.: Objective Classification of Air Quality Monitoring Sites over Europe, *Atmospheric Environment*, 47, 111–123, <https://doi.org/10.1016/j.atmosenv.2011.11.025>, 2012.
- 450 Keisler, R.: Forecasting Global Weather with Graph Neural Networks, <https://doi.org/10.48550/arXiv.2202.07575>, 2022.



- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: Learning Skillful Medium-Range Global Weather Forecasting, *Science*, <https://doi.org/10.1126/science.adi2336>, 2023.
- 455 Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M. C. A., Lessig, C., Maier-Gerber, M., Magnusson, L., Bouallègue, Z. B., Nemesio, A. P., Dueben, P. D., Brown, A., Pappenberger, F., and Rabier, F.: AIFS – ECMWF’s Data-Driven Forecasting System, <https://doi.org/10.48550/arXiv.2406.01465>, 2024a.
- Lang, S., Alexe, M., Clare, M. C. A., Roberts, C., Adewoyin, R., Bouallègue, Z. B., Chantry, M., Dramsch, J., Dueben, P. D., Hahner, S., Maciel, P., Prieto-Nemesio, A., O’Brien, C., Pinault, F., Polster, J., Raoult, B., Tietsche, S., and Leutbecher, M.: AIFS-
460 CRPS: Ensemble Forecasting Using a Model Trained with a Loss Function Based on the Continuous Ranked Probability Score, <https://doi.org/10.48550/arXiv.2412.15832>, 2024b.
- Loshchilov, I. and Hutter, F.: Decoupled Weight Decay Regularization, <https://arxiv.org/abs/1711.05101>, 2019.
- Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., Denier van der Gon, H. A. C., Drouin, A., Elbern, H., Emili, E., Engelen, R. J.,
465 Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillé, E., Josse, B., Kadygrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouil, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, A., van Velthoven, P., van Versendaal, R., Vira, J., and Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, *Geoscientific Model Development*, 8, 2777–2813,
470 <https://doi.org/10.5194/gmd-8-2777-2015>, 2015.
- Méndez, M., Merayo, M. G., and Núñez, M.: Machine Learning Algorithms to Forecast Air Quality: A Survey, *Artificial Intelligence Review*, 56, 10 031–10 066, <https://doi.org/10.1007/s10462-023-10424-4>, 2023.
- MSC-W, CCC, CEIP, and CIAM: Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components, EMEP Status Report 1/2025, Norwegian Meteorological Institute, Oslo, Norway, https://emep.int/publ/reports/2025/EMEP_Status_Report_1_2025.pdf,
475 2025.
- Pang, M., Jin, J., Segers, A., Lin, H. X., Wang, G., Liao, H., and Han, W.: Zeeman: A Deep Learning Framework for Regional Atmospheric Chemistry Forecasting, *Geophysical Research Letters*, 53, e2025GL117 459, <https://doi.org/10.1029/2025GL117459>, 2026.
- Rasp, S. and Thuerey, N.: Data-Driven Medium-Range Weather Prediction With a Resnet Pretrained on Climate Simulations: A New Model for WeatherBench, *Journal of Advances in Modeling Earth Systems*, 13, e2020MS002 405,
480 <https://doi.org/https://doi.org/10.1029/2020MS002405>, e2020MS002405 2020MS002405, 2021.
- Schloer, J., Tietsche, S., Roberts, C. D., Zampieri, L., Lang, S., Mertes, G., Jones, G., Chantry, M., and Vitart, F.: AIFS-SUBS: Extending Data-Driven Forecasting to Sub-Seasonal Timescales, <https://doi.org/10.48550/arXiv.2607.05100>, 2026.
- Seinfeld, J. H. and Pandis, S. N.: *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 3 edn., ISBN 978-1118947401, 2016.
- 485 Shetty, S., Hamer, P. D., Stebel, K., Kylling, A., Hassani, A., Berntsen, T. K., and Schneider, P.: Daily High-Resolution Surface PM_{2.5} Estimation over Europe by ML-based Downscaling of the CAMS Regional Forecast, *Environmental Research*, 264, 120 363, <https://doi.org/10.1016/j.envres.2024.120363>, 2025.
- Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyíri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, Á., and Wind, P.: The



- 490 EMEP MSC-W Chemical Transport Model – Technical Description, Atmospheric Chemistry and Physics, 12, 7825–7865,
<https://doi.org/10.5194/acp-12-7825-2012>, 2012.
- Simpson, D., Hood, C., van Caspel, W., Vieno, M., and Wind, P.: Updates to the EMEP/MSC-W model, 2024–2025, in: Transboundary
particulate matter, photo-oxidants, acidifying and eutrophying components. EMEP Status Report 1/2025, pp. 195–203, The Norwegian
Meteorological Institute, Oslo, Norway, https://emep.int/publ/reports/2025/EMEP_Status_Report_1_2025.pdf, 2025.
- 495 Sutterud, H., Mousing, E. A., Erlend, L., Ulmoen, M., and Blake, L.: EMEPai v1.0.0: Archive of source code for training and inference, and
checkpoint and input data for inference, <https://doi.org/10.5281/zenodo.22955196>, 2026.
- Targa, J., Colina, M., Banyuls, L., González Ortiz, A., and Soares, J.: Status report of air quality in Europe for year 2024, using validated and
up-to-date data. ETC HE Report 2025/1., <https://doi.org/10.5281/zenodo.15168515>, 2025.
- World Health Organization: WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur
500 Dioxide and Carbon Monoxide, World Health Organization, Geneva, ISBN 9789240034228, <https://www.who.int/publications/i/item/9789240034228>, 2021.