GLOBAL PARTICULATE MATTER FORECASTING USING LIGHTWEIGHT, REGION-SPECIFIC DEEP LEARNING MODELS

Ansh Kushwaha
Centre for Interdisciplinary Artificial Intelligence (CAI)
FLAME University
Pune, Maharashtra 412115, India
ansh.kushwaha@flame.edu.in

Kaushik Gopalan
Centre for Interdisciplinary Artificial Intelligence (CAI) &
School of Computing and Data Sciences
FLAME University
Pune, Maharashtra 412115, India
kaushik.gopalan@flame.edu.in

ABSTRACT

Global particulate matter (PM) forecasting is critical for air quality management, yet regional variability in emission sources and atmospheric processes poses challenges for unified modeling approaches. We present a lightweight nested-domain deep learning framework using U-Net architecture for short-range forecasting of PM_1, $PM_{2.5}$, and PM_{10}. Our approach trains separate models for 10 different spatial regions and 3 PM species, using overlapping 256×256 input grids to predict 192×192 forecast regions with explicit spatial context. Using CAMS reanalysis data spanning 2021–2024, we train independent U-Net models for each region/PM species combination using a model size of ≈ 4 million parameters per model, for a total of ≈ 120 million parameters for all the models combined. Evaluated against the ≈ 1 billion-parameter Aurora foundation model, our framework achieves competitive root mean square error at 6–24 hour forecast horizons while consistently resulting in slightly higher structural similarity indices. These results demonstrate that lightweight, regionally-specialized models offer a viable alternative to large-scale foundation models for PM forecasting, providing computational efficiency without sacrificing forecast accuracy.

1 INTRODUCTION

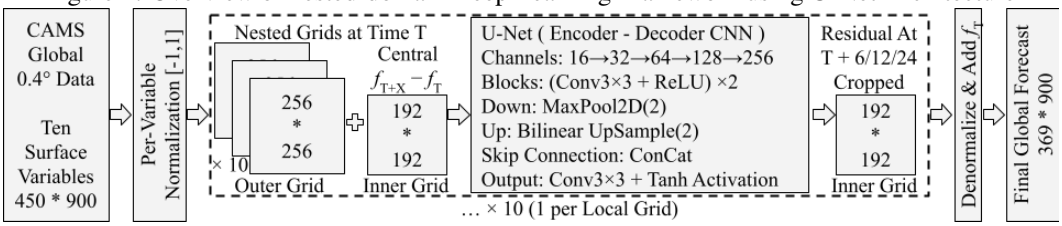
Recent advances in deep learning have enabled global foundation models to forecast atmospheric variables, including particulate matter (PM) concentrations, at planetary scale (Bodnar et al., 2025) as well as regionally (Sharma et al., 2020; Chae et al., 2021; Yang et al., 2020; Kushwaha & Gopalan, 2025). However, the sources and characteristics of PM pollution vary dramatically across regions. In China, severe haze events are driven primarily by industrial and vehicular emissions (Huang et al., 2014; Zhai et al., 2019). In contrast, PM pollution in Indian cities stems from a heterogeneous mix of vehicular emissions, industrial activity, construction dust, and biomass burning (Guttikunda et al., 2014), with seasonal agricultural fires in northwestern India significantly elevating pollution levels across the Indo-Gangetic Plain (Cusworth et al., 2018). North American PM dynamics present yet another pattern: while anthropogenic emissions have declined substantially, wildfire activity has emerged as an increasingly dominant source, particularly in the western United States (McClure & Jaffe, 2018; Ford et al., 2018). Natural dust emissions from arid regions such as northern Africa further contribute to global PM variability through episodic high-wind uplift events (Cowie et al., 2015).

This regional heterogeneity in emission sources, atmospheric chemistry, and transport mechanisms suggests that PM forecasting may benefit from spatially localized modeling approaches rather than unified global architectures. Furthermore, different PM size fractions—PM_1, $PM_{2.5}$, and PM_{10}—exhibit distinct source profiles and atmospheric lifetimes, with fine particles dominated by combustion and secondary formation processes while coarse particles are affected by larger-scale processes. These considerations motivate our development of a nested-domain framework employing separate

models for each spatial region and PM species. We demonstrate that this lightweight, localized approach achieves competitive forecasting performance at substantially lower computational costs.

2 METHODOLOGY

Figure 1: Overview of nested-domain Deep Learning Framework using U-Net Architecture



Analysis data were obtained from the Copernicus Atmosphere Monitoring Service (CAMS) Global Atmospheric Composition Forecasts, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) using the Integrated Forecasting System (IFS) with four-dimensional variational (4D-Var) data assimilation. The dataset is available at four synoptic times per day (00, 06, 12, and 18 UTC) on a uniform $0.4^\circ \times 0.4^\circ$ resolution grid, corresponding to 450×900 spatial points per variable over the global domain.

Ten surface-level variables were considered: 2 m temperature, 2 m dewpoint temperature, 10 m zonal wind, 10 m meridional wind, mean sea-level pressure, land–sea mask, surface geopotential, and concentrations of PM_1, $PM_{2.5}$, and PM_{10}. Data spanning the period 2021–2024 were downloaded in NetCDF format, with each file representing one day of data. Each variable was normalized independently by dividing by the global maximum of its magnitude. This results in a range of $[-1, 1]$ for the wind components (which can be negative), and a range of $[0, 1]$ for the other inputs.

The global CAMS fields are processed using a tiled, nested-domain formulation. The global latitude–longitude grid is partitioned into a set of 10 overlapping regional tiles, and within each tile a nested input–output grid configuration is used. Each tile covers a 256×256 grid at 0.4° resolution, corresponding to an area of approximately $102^\circ \times 102^\circ$. The tiles are arranged in two latitude bands and five longitude sectors. The southern band spans approximately -87° to $+16^\circ$ latitude, and the northern band spans approximately -15.6° to $+87^\circ$ latitude, resulting in overlap around the equatorial region. In longitude, the five tiles are offset by approximately 71°, forming overlapping longitudinal sectors that together span the full 360° domain.

Tiles that extend across the 180° meridian are handled using periodic longitude boundary conditions by concatenating longitude segments across the dateline to form contiguous spatial regions. For each tile, the UNet takes the full 256×256 grid as input and produces predictions on a nested 192×192 grid corresponding to the central portion of the input tile. Only this central region is used in the loss computation, while the surrounding grid points provide spatial context.

A lightweight U-Net architecture was employed, comprising 4,328,273 parameters (Fig. 1) [1]. The network consists of five resolution levels with encoder channel dimensions $16 \to 32 \to 64 \to 128 \to 256$. Each level uses a DoubleConv block composed of two consecutive 3×3 convolution layers, each followed by batch normalization and a ReLU activation.

Downsampling is performed using MaxPool2D with stride 2, while upsampling is carried out via bilinear interpolation with a scale factor of 2. Skip connections between encoder and decoder paths are implemented through channel-wise concatenation. The final layer applies a 3×3 convolution layer and then a Tanh activation to bound the output.

The network input has dimensions $10 \times 256 \times 256$, and the output is $1 \times 256 \times 256$. The central 192×192 region of the output is cropped and retained as the forecast target.

[1] Detailed Architecture Diagram, Complete Training & Inference Code, and Pre-Trained Model Checkpoints are available at `https://anonymous.4open.science/r/ICLR26_ML4RS_Workshop_PM_Forecasting`

Independent U-Net models were trained for each local grid (10 grids spanning the globe) and for each particulate matter species (PM_1, $PM_{2.5}$, and PM_{10}), resulting in a total of 30 models. All models share identical architectures and training hyperparameters.

Models were trained for up to 256 epochs using a batch size of 16 and the AdamW optimizer, with a learning rate $\eta = 10^{-3}$ and weight decay $\lambda = 10^{-4}$. A cosine annealing learning rate scheduler was employed, along with early stopping using a patience of 16 epochs. In practice, convergence typically occurred within approximately 60 epochs.

Training was guided by a composite loss function designed to balance pointwise accuracy, robustness to outliers, structural consistency, and systematic error control:

$$\mathcal{L} = 2^{10} \cdot \mathcal{L}_{\text{LogCosh}} + 2^{10} \cdot \mathcal{L}_{\text{Huber}} + 10^{-4} \cdot (1 - \mathcal{SSIM}) + 10^{-2} \cdot \mathcal{L}_{\text{MAE}},$$

The Log-Cosh and Huber terms emphasize robust pointwise regression and mitigate the influence of extreme concentration values. The MAE term is included to explicitly penalize persistent absolute deviations across the forecast region, encouraging uniform error reduction. The $\mathcal{SSIM}$ term promotes spatial coherence in predicted fields, supporting the preservation of physically meaningful structures.

Data from 2021–2023 were randomly shuffled and split into 90% training and 10% validation subsets. Data from 2024 were held out entirely for testing to assess seasonal and interannual generalization.

The prediction target was reformulated as a temporal difference to stabilize training and improve generalization across forecast horizons (Fig. 1). The prediction target was defined as the difference between the pollutant concentration at forecast time $T + \Delta t$ and the concentration at time T:

$$\Delta f_P^{(I)}(T, \Delta t) = f_P^{(I)}(T + \Delta t) - f_P^{(I)}(T),$$

where P denotes the particulate matter species, I denotes the local grid index, and $\Delta t \in \{6, 12, 24, ...\}$ hours.

For each particulate matter species P and local grid I, the model input consists of the 10-variable 256×256 grid at time T, and the target corresponds to the residual of the central 192×192 region. During inference, predicted residuals from all local grids are spatially stitched and added back to the corresponding initial fields to produce a global forecast of size 369×900, covering all longitudes and latitudes from $74°$N to $73.6°$S.

3 RESULTS

Table 1 presents a quantitative comparison between the proposed lightweight nested-domain U-Net model, the 1.3B-parameter Aurora foundation model (Bodnar et al., 2025), and a persistence (no-change) baseline across three forecast horizons (6, 12, and 24 hours) for PM_1, $PM_{2.5}$, and PM_{10}. Performance is evaluated over the global forecast domain using RMSE, CRMSE, and the structural similarity index (SSIM).

At the 6-hour forecast horizon, the U-Net model demonstrates substantial improvements over both Aurora and persistence across all particulate species. RMSE reductions relative to Aurora are approximately 60% for PM_1, 30% for $PM_{2.5}$, and 40% for PM_{10}, with corresponding CRMSE reductions of 33%, 29%, and 36%. SSIM increases markedly for PM_1 (0.98 vs. 0.70) and shows consistent improvements for $PM_{2.5}$ and PM_{10}. Notably, persistence outperforms Aurora at this lead time for PM_1 in both RMSE and CRMSE, underscoring pronounced short-range bias in Aurora that is mitigated by the U-Net model.

At the 12-hour horizon, performance becomes more heterogeneous. For PM_1, errors are broadly comparable between the two models, while the U-Net model continues to outperform Aurora for $PM_{2.5}$. However, Aurora achieves lower RMSE and CRMSE for PM_{10} by approximately 15%. At the 24-hour horizon, PM_1 errors remain closely matched; $PM_{2.5}$ differences are small; and Aurora maintains an advantage for PM_{10}. Across all horizons and species, the U-Net model consistently achieves higher SSIM values than both Aurora and persistence.

To isolate the contribution of the nested-domain input strategy, we also evaluate a variant of the U-Net model that receives only the 192×192 target-resolution grid as input, without the coarser

Table 1: Comparison Table (CRMSE and RMSE are reported in $\mu g\, m^{-3}$).

		6 Hours Forecast			12 Hours Forecast			24 Hours Forecast		
		PM_1	**$PM_{2.5}$**	**PM_{10}**	**PM_1**	**$PM_{2.5}$**	**PM_{10}**	**PM_1**	**$PM_{2.5}$**	**PM_{10}**
CRMSE	**Persistence**	3.4	6.0	30.2	4.2	8.8	40.2	4.3	7.8	31.5
	Aurora	4.7	7.4	36.4	3.0	6.5	28.0	3.4	6.1	25.9
	UNet	3.2	5.2	23.3	3.2	5.7	32.6	3.3	6.5	30.9
	UNet w/o Nesting	4.8	5.8	24.7	4.5	8.4	35.4	4.8	7.7	28.4
RMSE	**Persistence**	3.4	6.0	30.2	4.2	8.8	40.2	4.3	7.8	31.5
	Aurora	8.3	7.5	39.8	3.1	7.1	28.5	3.4	7.1	26.7
	UNet	3.2	5.2	23.3	3.2	5.8	32.6	3.3	6.5	30.9
	UNet w/o Nesting	4.8	5.8	24.7	4.5	8.4	35.4	4.9	7.8	28.5
SSIM	**Persistence**	0.95	0.94	0.98	0.94	0.94	0.97	0.93	0.90	0.96
	Aurora	0.70	0.93	0.96	0.97	0.97	0.98	0.94	0.89	0.96
	UNet	0.98	0.98	0.99	0.98	0.97	0.99	0.98	0.97	0.99
	UNet w/o Nesting	0.97	0.97	0.99	0.97	0.97	0.99	0.97	0.96	0.99

outer domain (denoted "UNet w/o Nesting" in Table 1). Removing nesting consistently degrades performance. At the 6-hour horizon, CRMSE increases from 3.2 to 4.8 $\mu g\, m^{-3}$ for PM_1 and from 23.3 to 24.7 $\mu g\, m^{-3}$ for PM_{10}, while $PM_{2.5}$ CRMSE rises from 5.2 to 5.8 $\mu g\, m^{-3}$. Similar degradations persist at 12 and 24 hours, with particularly large increases for PM_1 (CRMSE of 4.5–4.8 versus 3.2–3.3 $\mu g\, m^{-3}$) and $PM_{2.5}$ (CRMSE of 7.7–8.4 versus 5.7–6.5 $\mu g\, m^{-3}$). SSIM values also decrease slightly without nesting, though they remain above those of Aurora and persistence.

Overall, the proposed U-Net framework achieves substantially lower errors at the 6-hour forecast horizon across all particulate species—outperforming both Aurora and the persistence baseline—and maintains competitive performance for PM_1 and $PM_{2.5}$ at longer lead times. While Aurora demonstrates advantages for PM_{10} at extended horizons in terms of RMSE and CRMSE, the U-Net model consistently preserves spatial structure, as reflected by higher SSIM values across all forecast horizons and pollutant species compared to both Aurora and persistence.

4 Conclusion

We present a lightweight nested-domain deep learning framework based on U-Net architecture for global PM forecasting. By training separate models for each of 10 spatial regions and 3 PM species, the framework accommodates the heterogeneity in emission sources and atmospheric processes across different regions and particle size fractions.

The experimental results demonstrate that an ensemble of 30 regionally-specialized U-Net models, each containing approximately 4 million parameters (for a total of approximately 120 million parameters), achieves forecast accuracy for PM_1 and $PM_{2.5}$ that is comparable to or better than that of the Aurora foundation model, which contains approximately 1 billion parameters. At the 6-hour forecast horizon, the U-Net framework produces substantially lower errors across all PM species. Aurora achieves lower errors for PM_{10} at longer lead times, consistent with the increasing role of synoptic-scale transport in governing coarse particulate distributions. Spatial structure preservation, as measured by SSIM, is equal to or higher for the U-Net across all PM species and forecast horizons. These findings indicate that a collection of lightweight, regionally-trained models can provide accurate and spatially coherent short-range air quality forecasts at a fraction of the computational cost of global foundation models.

We must however note that our comparison with the Aurora model may be biased due to the fact that Aurora has not been fine-tuned for the specific parameters

REFERENCES

Cristian Bodnar, Wessel P. Bruinsma, Ana Lucic, Megan Stanley, Anna Allen, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan A. Weyn, Haiyu Dong, Jayesh K. Gupta, Kit Thambiratnam, Alexander T. Archibald, Chun-Chieh Wu, Elizabeth Heider, Max Welling, Richard E. Turner, and Paris Perdikaris. A foundation model for the earth system. *Nature*, May 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-09005-y. URL https://doi.org/10.1038/s41586-025-09005-y.

Sangwon Chae, Joonhyeok Shin, Sungjun Kwon, Sangmok Lee, Sungwon Kang, and Donghyun Lee. Pm10 and pm2. 5 real-time prediction models using an interpolated convolutional neural network. *Scientific Reports*, 11(1):11952, 2021.

Sophie M Cowie, John H Marsham, and Peter Knippertz. The importance of rare, high-wind events for dust uplift in northern africa. *Geophysical Research Letters*, 42(19):8208–8215, 2015.

Daniel H Cusworth, Loretta J Mickley, Melissa P Sulprizio, Tianjia Liu, Miriam E Marlier, Ruth S DeFries, Sarath K Guttikunda, and Pawan Gupta. Quantifying the influence of agricultural fires in northwest india on urban air pollution in delhi, india. *Environmental Research Letters*, 13(4): 044018, 2018.

Bonne Ford, Maria Val Martin, SE Zelasky, EV Fischer, SC Anenberg, Colette L Heald, and JR Pierce. Future fire impacts on smoke concentrations, visibility, and health in the contiguous united states. *GeoHealth*, 2(8):229–247, 2018.

Sarath K Guttikunda, Rahul Goel, and Pallavi Pant. Nature of air pollution, emission sources, and management in the indian cities. *Atmospheric environment*, 95:501–510, 2014.

Ru-Jin Huang, Yanlin Zhang, Carlo Bozzetti, Kin-Fai Ho, Jun-Ji Cao, Yongming Han, Kaspar R Daellenbach, Jay G Slowik, Stephen M Platt, Francesco Canonaco, et al. High secondary aerosol contribution to particulate pollution during haze events in china. *Nature*, 514(7521):218–222, 2014.

Ansh Kushwaha and Kaushik Gopalan. A comparison of lightweight deep learning models for particulate-matter nowcasting in the indian subcontinent surrounding regions, 2025. URL https://arxiv.org/abs/2511.11185.

Crystal D McClure and Daniel A Jaffe. US particulate matter air quality improves except in wildfire-prone areas. *Proceedings of the National Academy of Sciences*, 115(31):7901–7906, 2018.

Ekta Sharma, Ravinesh C Deo, Ramendra Prasad, Alfio V Parisi, and Nawin Raj. Deep air quality forecasts: suspended particulate matter modeling with convolutional neural and long short-term memory networks. *Ieee Access*, 8:209503–209516, 2020.

Guang Yang, HwaMin Lee, and Giyeol Lee. A hybrid deep learning model to forecast particulate matter concentration levels in seoul, south korea. *Atmosphere*, 11(4):348, 2020.

Shixian Zhai, Daniel J Jacob, Xuan Wang, Lu Shen, Ke Li, Yuzhong Zhang, Ke Gui, Tianliang Zhao, and Hong Liao. Fine particulate matter (pm 2.5) trends in china, 2013–2018: Separating contributions from anthropogenic emissions and meteorology. *Atmospheric Chemistry and Physics*, 19 (16):11031–11041, 2019.