High-Resolution Downscaling of Satellite-Derived NO_2 Concentrations Using Machine Learning and Multi-Source Data over Delhi NCR

DEEPTI SHRINGARE[1*], SAMRUDDHI SONAWANE[1] , NIRALI SHAH[1], POONAMDEVI YADAV [1] and MAMTA TIKARIA[1]

[1],Electronics and Telecommunication Department, Shah and Anchor Kutchhi Engineering College

Mahavir Education Trust Chowk, W.T. Patil Marg, Chembur, Mumbai - 400088, Maharashtra, India

ABSTRACT:

The presented paper outlines an approach of improving the spatial resolution of NO_2 concentration derived from satellite using machine learning models over Delhi NCR region. Although satellite-based instruments such as Sentinel-5P TROPOMI have the ability to measure wide areas, their poor resolution of about 7 kilometers hampers detection of the local pollution pattern. The authors therefore opted to incorporate a multisource dataset comprising data from satellite, ground observation from CPCB stations, meteorology from ERA5 reanalysis and other auxiliary data sources like population density and nocturnal light intensity.

Three different methods namely Multiple Linear Regression, Random Forest, and Extreme Gradient Boosting were used in training the models in an effort to upscale the coarse resolution of 7km provided by satellites to fine spatial resolution of 500 meters. This study further introduces a lag based technique to handle the problem of missing data due to presence of clouds. A split in space-time data partitioning was further adopted along with seasonal models.

Results indicate that the Lasso regression model (a variant of MLR) with lag features achieved the highest predictive performance, outperforming RF and XGBoost with an R^2 of 0.8404 and lower error metrics. The generated high-resolution NO_2 maps successfully captured fine-scale spatial variability and identified localized pollution hotspots that are not detectable in coarse-resolution data. This framework demonstrates a scalable and efficient approach for urban air quality monitoring and can support targeted pollution mitigation strategies and environmental policy planning.

Keywords: Air Quality, Machine Learning, Downscaling, Sentinel-5P, Environmental Monitoring.

1. INTRODUCTION

This study focuses on the National Capital Region (NCR) of Delhi, India, a region characterized by persistently high levels of air pollution, particularly nitrogen dioxide (NO_2). An Artificial Intelligence/Machine Learning (AI/ML)-based framework is developed to downscale coarse-resolution satellite-derived NO_2 data (7 km, TROPOMI) to a finer spatial resolution of 500 m for enhanced urban air quality assessment. The proposed framework integrates multi-source datasets, including ground-based NO_2 observations, satellite NO_2 retrievals, meteorological variables (U10, V10 wind components, wind speed, boundary layer height (BLH), surface pressure, and temperature), as well as ancillary datasets such as night-time light intensity (VIIRS) and population density.

Recent studies have demonstrated the effectiveness of machine learning and deep learning approaches for high-resolution NO_2 downscaling using satellite observations and auxiliary geospatial data. Kim et al. [3] developed an XGBoost-based framework integrating TROPOMI NO_2 columns, meteorological variables, land use, traffic density, and population data to generate near-surface NO_2 maps at 100 m spatial resolution, achieving an R^2 of 0.84. Liu et al. [5] proposed a machine learning framework for high-resolution tropospheric NO_2 mapping over China using TROPOMI observations and auxiliary meteorological predictors. Similarly, Yu and Liu [7] explored deep learning-based downscaling techniques using satellite and ground-based observations, where their Deep Matrix Network (DMN) achieved lower RMSE and MAE compared to conventional interpolation methods. Zhang et al. [8] further demonstrated the capability of deep ensemble models for generating high-resolution NO_2 pollution maps using Sentinel-5P TROPOMI and VIIRS datasets.

Several studies have also investigated advanced super-resolution and geospatial downscaling techniques for improving spatial detail preservation. Li and Cao [4] explored generative adversarial approaches such as ESRGAN and PhIRE-GAN for extreme geospatial downscaling applications, while Wang et al. [6] introduced ESRGAN for enhanced super-resolution image reconstruction. Wen Ma et al. employed Dense Residual Generative Adversarial Networks (DRGAN) for remote sensing image super-resolution, demonstrating improved preservation of structural and spatial features. In addition, Devarajan et al. [2] proposed a hybrid framework combining XGBoost and neural networks for NO_2 downscaling using satellite observations, meteorological parameters, land-use data, and road-network information, achieving an R^2 value of 0.88. These studies collectively demonstrate the growing potential of AI/ML techniques for atmospheric pollution downscaling; however, challenges related to spatial heterogeneity, cloud contamination, temporal alignment, and generalization across diverse urban environments still remain significant.

To model the complex spatial and temporal variability of NO_2, a comparative analysis is conducted using Multiple Linear Regression (MLR) as a baseline model, along with advanced ensemble learning techniques including Random Forest (RF) and Extreme Gradient Boosting (XGBoost). The study further incorporates lag-based feature engineering to address missing data caused by cloud cover in satellite observations, thereby improving temporal continuity and predictive robustness. Model performance is systematically evaluated under scenarios with and without lag features to quantify their impact on accuracy.

The novelty of this study lies in: (i) the development of a hybrid modeling framework combining baseline statistical methods (MLR) with ensemble machine learning models (RF and XGBoost) for high-resolution NO_2downscaling in the Indian urban context, (ii) the incorporation of lag-based feature engineering to effectively mitigate cloud-induced data gaps in satellite observations, (iii) a comprehensive comparative evaluation of models under different feature configurations, demonstrating the relative effectiveness of simpler models such as MLR under specific conditions, (iv) the adoption of a dual spatial--temporal data splitting strategy to rigorously evaluate model generalization across both space and time, (v) the training of season-specific machine learning models to explicitly account for winter atmospheric stability effects

prevalent in Delhi, and (vi) the implementation of the complete framework within Google Earth Engine to enable scalable, efficient, and reproducible urban air quality mapping.

2. DATA AND METHODOLOGY

2.1 Study Area

The National Capital Region (NCR) of Delhi, situated in northern India between latitudes 28.4°N–28.9°N and longitudes 76.8°E–77.3°E, serves as the geographical domain for this investigation. The region encompasses the urban core of Delhi along with adjoining districts from the states of Haryana, Uttar Pradesh, and Rajasthan. Being one of the most densely populated metropolitan agglomerations in the world, Delhi NCR is subjected to chronic and severe air pollution, with nitrogen dioxide (NO_2) emerging as a pollutant of critical concern. The semi-arid climatic regime of the region is marked by pronounced seasonal contrasts — extreme summer temperatures, a concentrated monsoon period, and cold dry winters. The winter season is of particular relevance to air quality research, as temperature inversions, suppressed boundary layer heights, and diminished wind speeds collectively promote the accumulation of near-surface pollutants. The heterogeneous land-use pattern within the NCR — spanning dense urban settlements, industrial corridors, peri-urban zones, and agricultural tracts — renders it an ideal testbed for spatially resolved AI/ML-based NO_2 downscaling.

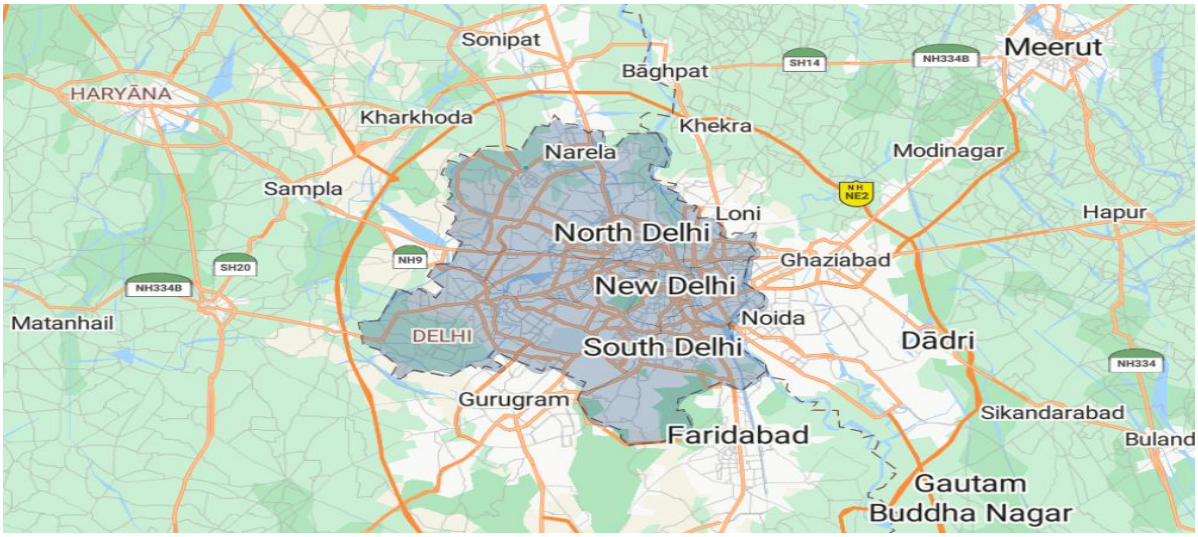

Figure 1. Study area

2.2 Data

Ground-based NO_2 concentration measurements were procured from 39 continuous ambient air quality monitoring stations operated by the Central Pollution Control Board (CPCB) across the Delhi NCR domain. Daily observational records spanning January 2024 to December 2024 were extracted and subjected to quality screening to eliminate erroneous and missing entries.

Tropospheric NO_2 vertical column density retrievals were obtained from the Sentinel-5P TROPOMI sensor via the Google Earth Engine platform. Monthly mean composites at approximately 7 km native resolution were generated to reduce cloud-induced data gaps and retrieval uncertainties.

Meteorological predictor fields were sourced from two ERA5 reanalysis products. Near-surface variables — comprising 10 m zonal wind (U10), meridional wind (V10), 2 m temperature, and surface pressure — were extracted from the ERA5-Land dataset at ~9 km resolution. Planetary boundary layer height (BLH) and total cloud cover fraction were derived from the ERA5 Atmospheric product at ~31 km resolution.

Nighttime light radiance composites from the NOAA/VIIRS sensor at 500 m resolution were incorporated as an urban activity proxy. Gridded population density data at 500 m resolution were acquired from the WorldPop database to represent demographic spatial variability across the study domain.

Table 1. Data Sources, Resolution, and Purpose for the Study

Data	Source	Resolution	Purpose
NO_2 Column	Sentinel-5P TROPOMI	~7 km	Satellite retrievals for downscaling
Meteorology	ERA5 ATMOS	~31 km	Atmospheric variables (wind, temperature, BLH)
Meteorology	ERA5 Land	~9 km	Land surface variables
Remote Sensing	Nighttime Dataset	~500 m	Urban extent and activity
Demographic	WorldPop	100 m	Population distribution
Ground Truth	CPCB (39 stations, Delhi)	Point-based	In-situ validation

Table 2. Variables Used for NO_2 Downscaling and Their Physical Rationale

Dataset	Variable	Reason for NO_2 Downscaling
ERA5-Land (9 km)	10 m u-wind & 10 m v-wind	Surface wind determines the dispersion ("smearing") of NO_2 from emission sources such as roads and industrial regions.
ERA5-Land (9 km)	2 m Temperature	NO_2 chemistry and emission processes are temperature dependent.
ERA5-Land (9 km)	Surface Pressure	Used to convert satellite-derived column density into surface concentration estimates.
ERA5 Atmospheric (31 km)	Boundary Layer Height (BLH)	Low BLH traps NO_2 near the surface, increasing ground-level concentrations.
ERA5 Atmospheric (31 km)	Total Cloud Cover	Helps the model understand uncertainty and missing satellite observations caused by cloud interference.

2.3 Methodology

2.3.1 Data Preprocessing: All multi-source datasets were harmonized to a common 500 m spatial grid using bilinear interpolation within the Google Earth Engine environment. Temporal alignment was performed to match satellite retrievals with corresponding ground station observations and meteorological fields on a monthly basis. Stations with more than 30% missing records were excluded from the analysis to maintain data integrity.

2.3.2 Lag Feature Engineering: To address temporal discontinuities arising from cloud-induced gaps in TROPOMI retrievals, lag-based features were constructed from the NO_2 time series. Specifically, 1-day and 7-day lagged NO_2 values (no2_lag1, no2_lag7) and 3-day and 7-day moving averages (no2_ma3, no2_ma7) were computed for each monitoring station. These engineered features capture temporal persistence in NO_2 concentrations and substantially improve model robustness under data-sparse conditions.

2.3.3 Model Development: Three supervised machine learning models were developed and systematically compared:

(i) Multiple Linear Regression (MLR): A conventional statistical model assuming linear relationships between predictor variables and NO_2 concentrations, employed as a performance baseline.

(ii) Random Forest (RF): A bagging-based ensemble method constructing multiple decorrelated decision trees. Final predictions are obtained by averaging individual tree outputs, which effectively reduces variance and controls overfitting.

(iii) Extreme Gradient Boosting (XGBoost): A gradient-boosted tree ensemble that sequentially minimizes prediction residuals through iterative optimization. XGBoost demonstrates strong capability in modeling non-linear interactions among meteorological and satellite-derived predictors.

2.3.4 Spatial–Temporal Data Splitting: A dual splitting strategy was implemented to ensure rigorous model evaluation free from data leakage:

- **Temporal Split:** Data from March to November (non-winter) were assigned to training (9,067 samples, 34 stations), while December to February observations (2,777 samples) were reserved exclusively for temporal testing.
- **Spatial Split:** Five stations — Shadipur, Sri Aurobindo Marg, Bawana, ITO, and Pusa IMD — were entirely excluded from training and used solely for spatial generalization assessment (1,794 samples).

2.3.5 Season-Specific Modelling : Independent models were trained separately for winter (December–February) and non-winter (March–November) periods. This seasonal stratification explicitly accounts for the atmospheric stability, temperature inversion, and boundary layer suppression characteristic of Delhi winters, thereby reducing seasonal prediction bias.

2.3.6 Evaluation Metrics : Model accuracy was assessed using three statistical indicators:

- $\mathbf{R^2}$ (Coefficient of Determination): Proportion of NO_2 variance explained by the model
- **RMSE** (Root Mean Square Error): Overall magnitude of prediction error ($\mu g/m^3$)
- **MAE** (Mean Absolute Error): Average absolute deviation between predicted and observed values ($\mu g/m^3$).

3. RESULTS AND DISCUSSION

3.1 Comparative Analysis of the Model

Table 3. Comparative analysis RF and XGBoost

Model	With Lag Feature			Without Lag Feature		
	R^2	RMSE	MAE	R^2	RMSE	MAE
RF	0.7332	19.1462	11.0992	0.2246	29.7428	19.867
XGBoost	0.7863	17.1291	9.5335	0.2708	29.2112	17.2015

Table 4. Comparative analysis OLS, Ridge and Lasso (Baseline model MLR analysis)

Model	Without Lag Feature			With Lag Feature		
	R^2	RMSE	MAE	R^2	RMSE	MAE
OLS	0.0315	33.2991	22.661	0.8204	15.7023	10.2461
Ridge	0.0363	33.2178	23.1149	0.8289	15.329	9.76
Lasso	0.0865	32.1628	22.1927	0.8404	14.8016	9.082

Upon comparative analysis between MLR, Random Forest and XGBoost in case of with lag features and without lag features MLR's Lasso model with lag feature was considered for creating downscaled visualization of NO_2 air quality map (Table 3 ,Table 4).

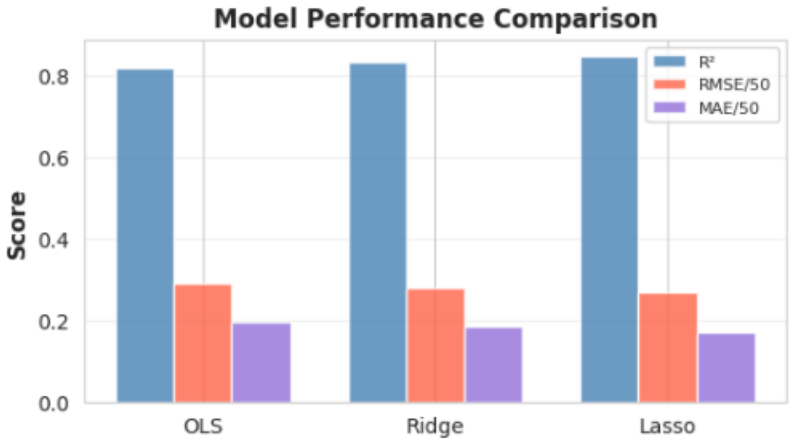

Figure 2. Model Performance Comparison between OLS, Ridge and Lasso with lag features

3.2 Prediction Accuracy and Residual Analysis

A comparative analysis was carried out for the different predictive models used by applying Multiple Linear Regression (MLR) as the benchmark against the more powerful ensemble methods such as Random Forest and XGBoost. Although the machine learning model is complex, the MLR model was found to be a better predictor. The OLS, Ridge, and Lasso regression models were considered under the MLR model. Using quantitative measures, the most accurate model was found to be the Lasso regression model enhanced with lagged predictors (See Figure.3-6).

The advantage of the Lasso model is also proved by the performance of a diagnostic check to examine the accuracy of predictions as well as the residuals. The scatter plot between the actual values and the predicted values of the Lasso model has shown the closest proximity with the identity line, implying that there is little variation in the errors. Besides, thorough residual analysis revealed that the Lasso model with lags was able to capture the pattern of the data set as its residuals were homoscedastic and normally distributed. On the other hand, the ensemble methods appeared to suffer from overfitting/marginal bias.

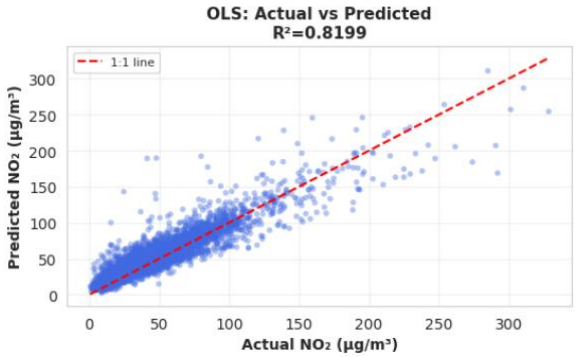

Figure 3. OLS predicted vs. observed NO_2 concentrations

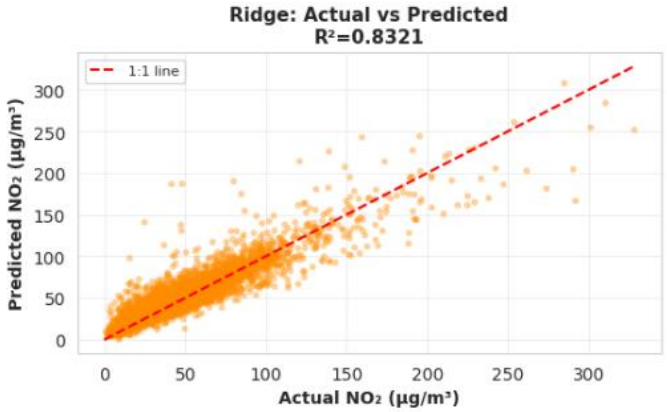

Figure 4. Ridge predicted vs. observed NO_2 concentrations

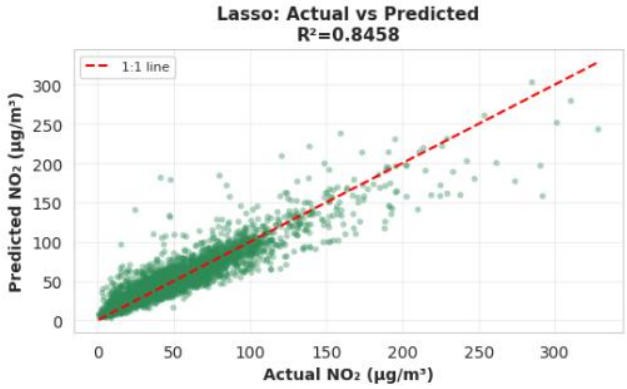

Figure 5. Lasso predicted vs. observed NO_2 concentrations

3.3 Error Analysis by Concentration Range

3.3.1 MAE by Concentration Range: Figure 7 illustrates the Mean Absolute Error (MAE) across six NO_2 concentration bins (0–20, 20–40, 40–60, 60–80, 80–100, and 100–328 µg/m³) for all three regression models, while Figure 6 presents the residual distributions of OLS, Ridge, and Lasso overlaid as a composite histogram.

A consistent and systematic pattern emerges across all three models: MAE increases monotonically with rising NO_2 concentration, with the lowest errors occurring in the 0–20 and 20–40 µg/m³ bins and the highest errors concentrated in the extreme upper range of 100–328 µg/m³. This behaviour is physically expected — high-concentration episodes in Delhi are episodic, short-lived, and driven by compound meteorological and emission factors that are inherently harder to capture through linear regression frameworks.

In the lower concentration bins (0–20 and 20–40 µg/m³), all three models perform comparably well, with MAE values in the range of approximately 6–9 µg/m³, indicating robust predictive skill under moderate pollution conditions. As concentrations progress into the 40–60 and 60–80 µg/m³ ranges, MAE rises gradually to approximately 8–10 µg/m³, remaining within acceptable limits for practical air quality applications. A more pronounced error escalation is observed in

the 80–100 µg/m³ bin (~12–13 µg/m³) and most sharply in the 100–328 µg/m³ extreme bin, where MAE reaches approximately 20–21 µg/m³ across all models. This non-linear error growth at extreme concentrations reflects the well-documented limitation of linear models in capturing the tail behaviour of highly skewed pollutant distributions, particularly during Delhi's severe winter pollution episodes driven by crop residue burning, temperature inversions, and reduced boundary layer ventilation.

Comparing across models, **Lasso** achieves marginally lower MAE in the lower concentration bins (0–40 µg/m³) relative to OLS and Ridge, consistent with its sparser feature structure reducing overfitting to low-pollution conditions. However, all three models converge to nearly identical MAE (~20–21 µg/m³) in the extreme concentration range, suggesting that the high-end prediction error is a structural limitation of the linear modelling framework rather than a regularisation-specific artefact. Ridge and OLS exhibit nearly indistinguishable error profiles across all bins, confirming that the mild L2 penalty of Ridge does not substantially alter concentration-specific error characteristics relative to unregularised OLS in this dataset.

3.3.2 Residual Distribution: The composite residual distribution (Fig 6) reveals that all three models produce approximately symmetric, near-Gaussian residual profiles centred tightly around zero — as indicated by alignment with the red dashed reference line — confirming the absence of systematic bias across the prediction range. The bulk of residuals fall within the ±25 µg/m³ window, with frequency peaking sharply near zero and declining steeply on both sides, consistent with well-calibrated regression models.

A slight positive skew is discernible in the right tail of all three distributions, extending to approximately +100 µg/m³, which corresponds to the underprediction of extreme high-concentration events identified in the MAE analysis above. The left tail, extending to approximately −150 µg/m³, reflects isolated instances of overprediction, likely associated with days where predicted concentrations based on prior lag structure exceeded actual observed values due to rapid meteorological cleansing events such as strong wind episodes or rainfall. The three models overlap almost entirely in their residual histograms, with OLS, Ridge, and Lasso displaying near-identical distributional shapes, reinforcing that all three converge to comparable error structures despite differing regularisation approaches. The near-zero mean residual across all models further confirms negligible systematic bias, satisfying a fundamental assumption of robust regression modelling for air quality applications.

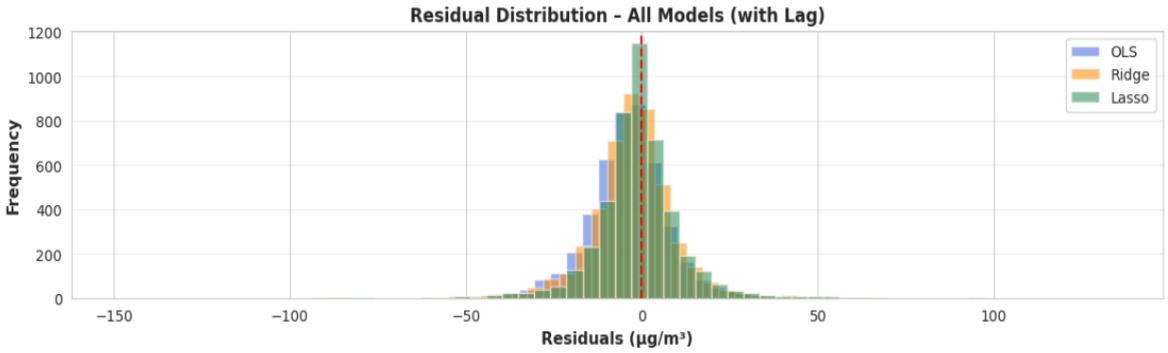

Figure 6. Residual distribution of OLS , Ridge and Lasso

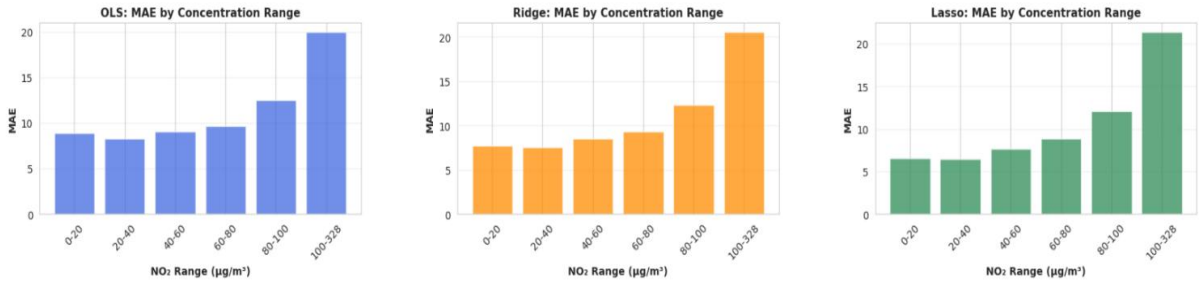

Figure 7. MAE by Concentration Range of OLS, Ridge and Lasso

3.4 Feature Importance Analysis

The Lasso regression model, by virtue of its L1 regularisation penalty, performed simultaneous feature selection and coefficient estimation, yielding a sparse and interpretable representation of the predictors governing surface NO_2 concentrations across Delhi. Figure 8 presents the top 20 features ranked by their absolute standardised coefficients.

The results reveal a pronounced hierarchical structure in feature relevance. Temporal autocorrelation features collectively dominated the model, with the 3-day moving average of ground-level NO_2 (no2_ma3) registering the highest standardised coefficient magnitude of approximately 44 — substantially exceeding all other predictors. This was followed by the 1-day lag (no2_lag1, ~14), the 7-day moving average (no2_ma7, ~7), and the 7-day lag (no2_lag7, ~2), together forming a clear gradient of diminishing temporal memory. The overwhelming dominance of no2_ma3 over no2_lag1 indicates that the accumulated pollution trend over the preceding 72 hours carries greater predictive information than any single prior-day observation, consistent with the multi-day persistence of pollution episodes characteristic of Delhi's wintertime anticyclonic stagnation events.

Beyond the lag group, a secondary tier of meteorological dispersion variables was retained with small but non-zero coefficients. Wind speed, ventilation coefficient (the product of boundary layer height and wind speed), and the meridional wind component (v10) each registered standardised coefficients of approximately 1, confirming the role of atmospheric ventilation in modulating surface NO_2 accumulation. The retention of ventilation_coef as a composite index — rather than its constituent variable BLH — reflects Lasso's tendency to favour combined dispersion metrics over their individual components when multicollinearity is present. Additional variables including day_of_week, is_hot, and is_cold were retained with near-zero weights, providing weak but statistically confirmed contributions from weekly anthropogenic emission cycles and temperature-mediated photochemical effects respectively.

A notable outcome of the Lasso regularisation is the complete elimination — coefficients shrunk to exactly zero — of several physically meaningful but statistically redundant features, including BLH, tropospheric_no2, stability_index, spatial_lag_no2, urban_intensity, dist_from_center, satellite_stability, cloud_temp_interaction, and urban_intensity_log. This does not imply physical irrelevance; rather, it reflects that the information content of these variables is already sufficiently captured by the retained lag and dispersion features. Specifically, BLH's predictive contribution is subsumed by ventilation_coef, and the satellite-derived tropospheric NO_2 column is rendered redundant by the ground-level temporal lag structure. This sparsification resulted in an effective model operating on approximately 10 active predictors out of the full feature set, demonstrating that NO_2 downscaling in Delhi is governed by a compact set of temporal persistence and meteorological dispersion signals rather than a broad ensemble of environmental covariates.

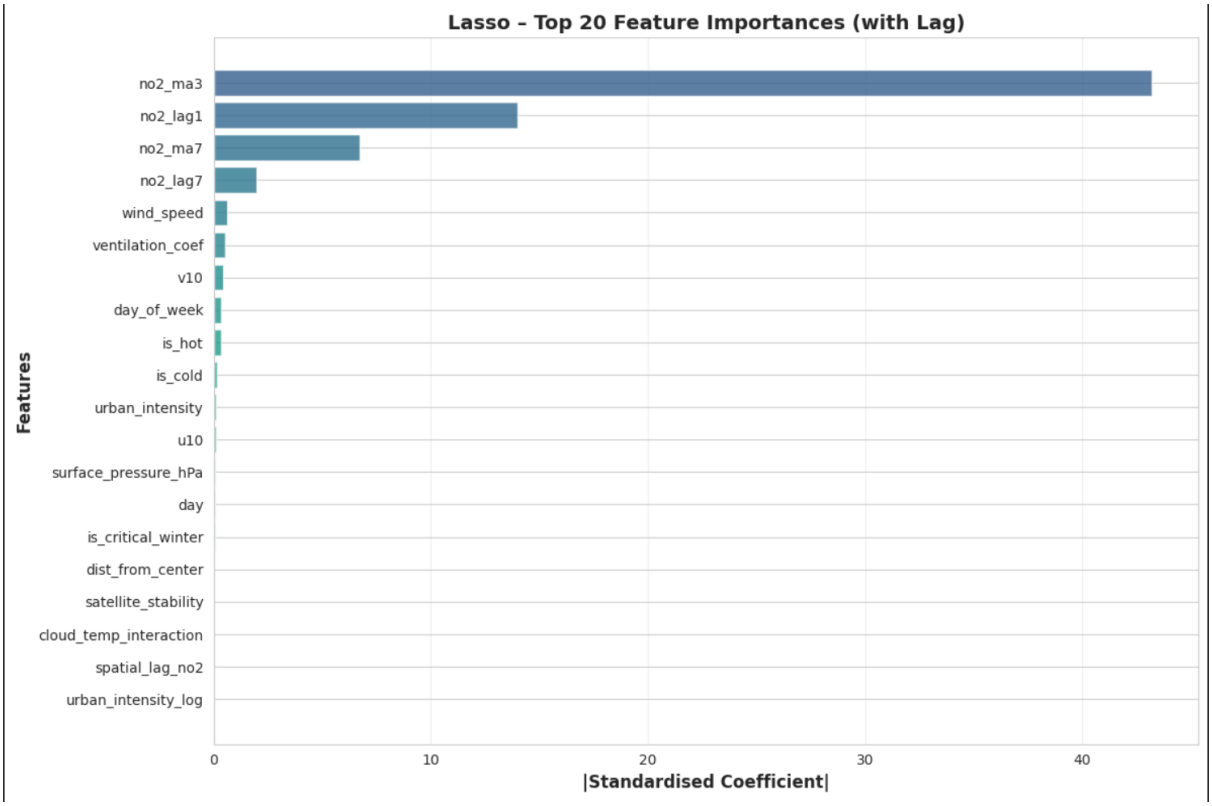

Figure 8. Feature Important Analysis

3.5 Spatial Output and High-Resolution NO_2 Mapping

The comparative visualization systematically demonstrates the limitations of coarse-resolution satellite-derived NO_2 observations and the efficacy of the proposed machine learning–based downscaling framework (Fig 9.).The left-side map shows concentration fields with smooth and homogeneous characteristics, mainly owing to low spatial resolution of the satellite, which intrinsically lacks the capacity to detect intra-urban concentration gradients. On the contrary, the right-side map depicts the downscaled NO_2 concentrations obtained from a machine learning algorithm, which is characterized by an explicit inclusion of auxiliary predictors and nonlinear relationships between inputs and outputs, thus enabling the production of high-resolution concentration maps. Notably, the output map demonstrates high heterogeneity and fine structure of spatial distribution, especially within urban hotspots, which proves that the model succeeds in recovering small-scale concentration variation. This implies that the model is able to learn intricate spatial relationships between variables and processes such as the emission–dispersion dynamics of pollutants, which are ignored by coarse resolution satellite. Additionally, the use of the inspector (console) provides a series of pixel-level diagnostic results, such as geographic coordinates (latitude and longitude), original coarse NO_2 column density, predicted high-resolution NO_2 (proxy) concentration, and predicted surface NO_2 concentration ($\mu g/m^3$).

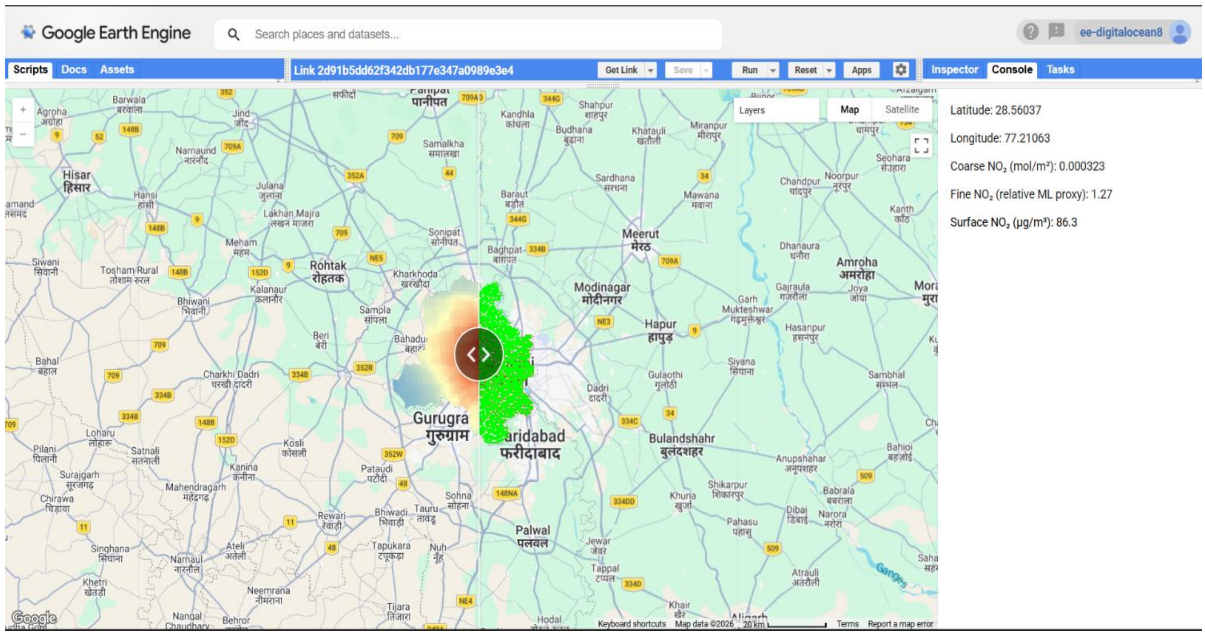

Figure 9. High-resolution downscaled NO_2 map (500 m) over Delhi NCR generated using Google Earth Engine

4. Conclusions

This study successfully developed and evaluated an AI/ML-based spatial downscaling framework for enhancing the resolution of satellite-derived NO_2 observations from approximately 7 km (TROPOMI) to 500 m over the Delhi National Capital Region. The following principal conclusions are drawn from the investigation:

The Lasso regression model with lag features outperformed both Random Forest and XGBoost across all evaluation metrics, achieving a testing R^2 of 0.8404, RMSE of 14.8016 µg/m³, and MAE of 9.082 µg/m³, demonstrating its superior capability in capturing the complex relationships between satellite-derived column densities and surface-level meteorological predictors.

The incorporation of lag-based temporal features — including 1-day and 7-day lagged NO_2 values and 3-day and 7-day moving averages — proved instrumental in addressing cloud-induced data gaps in satellite retrievals, substantially improving model robustness and temporal continuity of predictions.

The dual spatial–temporal data splitting strategy adopted in this study ensured rigorous evaluation of model generalization across both unseen time periods and geographically withheld monitoring stations, thereby providing a reliable and unbiased assessment of predictive performance.

Season-specific model training, with independent sub-models developed for winter (December–February) and non-winter (March–November) periods, effectively mitigated the influence of wintertime atmospheric stability, temperature inversions, and boundary layer suppression on model predictions — conditions that are characteristic of Delhi's severe pollution episodes.
Feature importance analysis consistently identified lag-based NO_2 features — particularly the 3-day moving average (no2_ma3) and 1-day lag (no2_lag1) — as the most influential predictors, with meteorological dispersion variables such as wind speed and ventilation coefficient contributing as secondary predictors, reinforcing the critical role of temporal persistence and atmospheric ventilation dynamics in governing near-surface NO_2 variability over urban environments.
The complete framework, implemented within Google Earth Engine, demonstrated scalability, computational efficiency, and reproducibility, establishing a transferable methodology applicable to other polluted urban domains across India and the broader South Asian region.
The generated high-resolution NO_2 maps successfully resolved fine-scale spatial heterogeneity across the NCR domain, revealing localized pollution hotspots associated with major traffic corridors and industrial zones that remain undetectable in coarse satellite products. These outputs hold significant potential for supporting urban air quality management, targeted emission control strategies, and climate-resilient infrastructure planning.

Data availability

The satellite datasets utilized in this study, comprising Sentinel-5P TROPOMI NO_2 retrievals, ERA5 meteorological reanalysis fields, and NOAA/VIIRS nighttime light composites, are openly accessible through the Google Earth Engine (GEE) platform. Population density data are publicly available from the World Pop repository (www.worldpop.org). Ground-based NO_2 measurements from CPCB monitoring stations are available through the Central Pollution Control Board, Government of India (www.cpcb.nic.in). Processed model outputs and derived datasets are available from the corresponding author upon reasonable request.

[1]*Funding*
No dedicated financial support was received for the conduct of this research or preparation of this manuscript.

Acknowledgement

The authors express sincere gratitude to the Central Pollution Control Board (CPCB), Government of India, for providing ground-based air quality monitoring data utilized in this study. The authors also acknowledge Google Earth Engine for providing the computational platform and access to satellite and reanalysis datasets. The views expressed in this manuscript are solely those of the authors and no conflict of interest exists.

References

[1] M. A. G. Demetillo, A. Navarro, K. K. Knowles, et al., "Observing nitrogen dioxide air pollution inequality using high-spatial-resolution remote sensing measurements in Houston, Texas", Environ. Sci. Technol., vol. 55, no. 21, (2021), pp. 14809–14818, doi:10.1021/acs.est.0c01864.

[2] D. M. S. Devarajan, A. Selvalmar and K. B. Jayanthi, "High-resolution NO_2 mapping via satellite data downscaling using machine learning", Proc. 3rd Int. Conf. Inventive Computing and Informatics (ICICI-2025), IEEE, (2025), pp. 1–6, doi:10.1109/ICICI2025.2025.1099584.

[3] M. Kim, D. Brunner and G. Kuhlmann, "Importance of satellite observations for high-resolution mapping of near-surface NO_2 by machine learning", Remote Sens. Environ., vol. 264, (2021), p. 112573, doi:10.1016/j.rse.2021.112573.

[4] G. Li and G. Cao, "Generative adversarial models for extreme geospatial downscaling", Unpublished manuscript, Department of Geography, University of Colorado Boulder.

[5] J. Liu, X. Meng, H. Zhao and J. Ma, "High-resolution mapping of tropospheric NO_2 using TROPOMI and machine learning: A case study over China", Sci. Total Environ., vol. 817, (2022), p. 152947, doi:10.1016/j.scitotenv.2021.152947.

[6] X. Wang, K. Yu, S. Wu, et al., "ESRGAN: Enhanced super-resolution generative adversarial networks", arXiv preprint, arXiv:1809.00021, (2018).

[7] M. Yu and Q. Liu, "Deep learning-based downscaling of tropospheric nitrogen dioxide using ground-level and satellite observations", Sci. Total Environ., vol. 773, (2021), p. 145145, doi:10.1016/j.scitotenv.2021.145145.

[8] Y. Zhang, Y. Liu, Y. Wang and Q. Zhang, "High-resolution NO_2 pollution mapping from Sentinel-5P TROPOMI and VIIRS data with a deep ensemble model", Remote Sens. Environ., vol. 286, (2023), p. 113336, doi:10.1016/j.rse.2023.113336.