

IDENTIFICATION OF THE DRIVING FACTORS OF NO₂ CONCENTRATIONS IN PAKISTAN USING MACHINE LEARNING

Muhammad Abbas^{*1}, Zia Ul Haq², Shahid Parvez³

¹Remote Sensing, GIS and Climatic Research Lab (National Center of GIS and Space Applications), University of the Punjab, Lahore, Pakistan

^{2,3}Institute of Space Science, University of the Punjab, Lahore

^{*1}abbas.sscs@gmail.com

DOI: <https://doi.org/10.5281/zenodo.23311052>

Keywords

Tropospheric NO₂; NO₂ Drivers; Machine Learning; Random Forest, XGBoost; SHapley Additive exPlanations (SHAP)

Article History

Received: 01 July 2026

Accepted: 22 September 2026

Published: 12 October 2026

Copyright @Author

Corresponding Author: *

Muhammad Abbas

Abstract

Nitrogen dioxide (NO₂) is a short-lived trace gas but highly toxic in nature that causes serious health issues in a country like Pakistan where ground monitoring infrastructure is very poor to support targeted mitigation. Modelling the concentration of tropospheric NO₂ is very important for public health and environmental concerns. This study proposes the use of multi-satellite datasets and machine learning approach to identify the driving factors of elevated NO₂ concentration in Pakistan. High resolution (0.05° × 0.05°) tropospheric NO₂ dataset was collected from Sentinel-5P TROPOMI sensor and combined with 19 other predictor variables selected from various auxiliary domains such as meteorology, land-use/land-cover, anthropogenic emissions and topography. Pearson correlation coefficient 'r' was computed and highly correlated coexisting features were dropped from the feature set using a fair threshold of |r| < 0.85. To model non-linear and complex interactions between tropospheric NO₂ pollution and its drivers, machine learning algorithms, Random Forest (RF) and Extreme Gradient Boosting (XGBoost), were trained on 80% of the dataset and evaluated on the remaining 20% of the dataset. The XGBoost model outperformed the RF with a R²=0.928 (92.8% accuracy in explaining the dataset), RMSE = 3.70×10⁻⁶ mol/m², and MAE = 2.40× 10⁻⁶ mol/m². SHapley Additive exPlanations (SHAP) values were evaluated to find the relative importance and marginal contribution of each predictor variable in the final prediction of the model. Visuals like geospatial maps, scatter plots and beeswarm plots were drawn using Python libraries to understand the underlying dynamics. The results show that road density, surface temperature, surface elevation and industrial emissions are the main driving factors that control the tropospheric NO₂ variation across Pakistan. Geographical differences were observed across the four canonical domains of spatial attribution: across 65.5 % of the grid cells in Pakistan, the first domain of Demographics & Topography remained dominant, providing baseline corridors of accumulation of natural processes in low-lying river plains and high density transportation corridors; the second domain of Meteorology governed the natural processes across 29.8 % of the grid cells in Pakistan, with natural dispersion and the creation of thermal inversion driven winter peaks in pristine northern highlands and coastal belts; the third domain of Direct Anthropogenic Emissions, with extreme degrees of indicators of concentration, governed acute and severe hot spots across 4.7 % of the grid cells across Pakistan, concentrated over the

Lahore – Faisalabad industrial belt and Karachi. These results reveal that boundary layer compression in the presence of dense transport and industrial corridors, and not a surge in emissions alone, trigger recurrent winter smog crises. This research offers an easily understood, airshed-specific methodology for moving from blanket restrictions to emission controls that are coupled to meteorology.

1. Introduction

Nitrogen dioxide (NO_2) is an atmospheric trace gas, short lived but highly toxic in nature. Its chronic exposure may cause obstructive pulmonary issues, cardiovascular problems and birth related abnormalities (Haslina et al., 2024; Johnson et al., 2022; Liu et al., 2021; Meng et al., 2021). Power generation sector, urban industrial units and road transport use fossil fuels as their source of energy and due to their comparatively major contribution these sector are known as primary sources of NO_2 (Bai et al., 2025; Li et al., 2023; Song et al., 2021; Xian & Li, 2026). Apart from these direct contributors, there are also some other natural processes that make indirect contributions to atmospheric NO_2 such as natural lightning that acts like a natural catalyst during NO_x production in upper atmosphere, and soil microbes that release NO into atmosphere during ammonia oxidization process of chemical fertilizers (Jin et al., 2021; Verma et al., 2021). Generally, urban areas having high population density show elevated concentrations of NO_2 because such localities often host the mega industrial units, heavy traffic choke points and commercial centers that employ the heavy consumption of fossil fuels (Pültz et al., 2025; Wang et al., 2022; Zeeshan, 2025). However, these NO_2 emission sources are not the exclusive factor of elevated concentrations over mega cities, rather prevailing meteorological conditions are also a significant factor in creating localized hotspots of NO_2 . When slow moving horizontal wind gets paired with temperature inversion, it becomes stagnant because temperature inversion acts like a cap or barrier against its vertical movement. Consequently, NO_2 gets trapped there to form localized hotspots. In contrast, dynamic weather such as strong wind or rain causes dispersion of pollutants and brings pollution level down. Pakistan is an agricultural country where most of the population lives in rural areas, but current industrial expansion and city centered facilities

have attracted people towards urban localities with great acceleration. This fast urbanization process has ultimately overburdened the urban areas with heavy use of natural resources, especially fuel to run the wheels for their daily needs. The increasing transport volume, road network and urban-industrial corridors are deteriorating the air quality with great speed but, like other developing countries, Pakistan lacks in air quality monitoring infrastructure that is why public is under high exposure risk (Anjum et al., 2021; Carvalho, 2021). Various studies have confirmed an overall increasing trend of NO_2 pollution across Pakistan (Dilawar et al., 2023; ul-Haq et al., 2014). Specially, its major cities like Lahore, Islamabad, Rawalpindi, Faisalabad and Karachi have dense population cores where excessive anthropogenic activities happen to meet daily needs of the population (Ahmed et al., 2022; Bilal et al., 2021). These extreme anthropogenic activities result in NO_2 pollution that is why multiple studies have warned about worsening condition of air quality in these cities, even it has crossed World Health Organization (WHO) recommended limits (Rabbia Rehmat et al., 2023). Winter smog episodes that paralyze major cities of Pakistan are also consequence of increased anthropogenic emissions and prolonged meteorological stagnation (Ghani et al., 2026). Thus, understanding the mechanism of NO_2 hotspot creation in Pakistan and isolated contribution of each key role player in this phenomenon is quite necessary on urgent basis so that policy makers could design effective mitigation strategies to save the humanity from any undesired situation.

Existing NO_2 -driver studies use a wide range of statistical and machine learning practices for driver attribution in Pakistan. Some of these studies have used ground-based monitoring with passive samplers, HORIBA analyzers and Multi-Axis Differential Optical Absorption Spectrometer (MAX-DOAS) for linking elevated NO_2 concentration with roads, emissions and

meteorology in Lahore, Rawalpindi/Islamabad, and Karachi (Ahmad et al., 2011; Ahmed et al., 2022; Khokhar et al., 2017). Several other studies have used satellite-based remote sensing with Ozone Monitoring Instrument (OMI), SCanning Imaging Absorption Spectrometer for Atmospheric CHartography (SCIAMACHY), Global Ozone Monitoring Experiment-2 (GOME-2), and Sentinel 5P/Tropospheric Monitoring Instrument (TROPOMI) to identify spatiotemporal patterns of NO₂ and find their association with urbanization, industry, burning and meteorology (Munawar et al., 2026; Murtaza et al., 2018; Rabbia Rehmat et al., 2023; Zeeshan, 2025). A number of studies have also used mixed approach of using satellite based remote-sensing aided with ground-based monitoring for better comparison and validation purpose (Khokhar et al., 2017). Some studies have used Hybrid Single-Particle Lagrangian Integrated Trajectory model (HYSPLIT) model to trace back the origin of transported pollution to Lahore, Karachi, Peshawar and Islamabad (Naz et al., 2026; Saleem et al., 2024). Other have used Potential Source Contribution Function (PSCF), and Concentration Weighted Trajectory (CWT) maps to locate local and transboundary sources in major cities of Pakistan such as Lahore, Karachi, Islamabad, Quetta and Peshawar (Mariam et al., 2026; Tariq et al., 2023). Some have used statistical models like Mann-Kendall to detect seasonal and long-term trends. Studies also use correlation, regression and wavelet analysis approaches to relate NO₂ with meteorology, nighttime lights and land cover features (Mariam et al., 2026). Recent studies have used machine learning algorithms like convolutional neural networks (CNN), Long Short-Term Memory (LSTM), Random Forest (RF), extreme gradient boosting (XGBoost), CatBoost, Light Gradient Boosting Machine (LightGBM), Support Vector Machine (SVM), Gaussian process regression (GPR) and ensembles to explore complex non-linear relationships of NO₂ concentration with other related features like meteorology, surface features, etc (Naseer et al., 2022).

ul-Haq et al., (2014) employed OMI dataset from 2004-2008 and identified significant increases in NO₂ concentration over major Pakistani cities

(such as Lahore, Rawalpindi, and Karachi). They used statistical mean value approach to quantify a national increase of 3.29% per year and attributed this growth primarily to rapid urbanization, vehicular traffic, and industrial emissions (ul-Haq et al., 2014). Recent satellite-based assessments show that Punjab, the most populated province of Pakistan, faces pronounced burdens of NO₂ with persistent hotspots on its major urban areas: Lahore, Gujranwala, Faisalabad, Rawalpindi and Islamabad. Within Punjab, industrial emission, road transport, power sector, agricultural residue burning and seasonal meteorology have all been identified as important contributors to spatiotemporal dynamics of tropospheric NO₂ (Mariam et al., 2026; Zeeshan, 2025). The smog season extended from October to February is observed as seasonal decline in air quality that poses serious health challenge across the province (Narjis Naqvi et al., 2025). HYSPLIT trajectories show that transboundary air pollution due to crop residue burning across the border significantly affect NO₂ concentration during smog episodes (Zeeshan, 2025). In Lahore, transportation is the leading source of NO_x emissions, contributing 484 3,390 tons compared to 2,296 tons from industry (Zeeshan, 2025), while in Karachi, combustion, transport and biomass burning are the main drivers (Zong et al., 2020). Ahmed et al., (2022) used ground-based local air quality monitoring station data collected from Horiba-A370 and found that vehicular emissions were responsible for compromised air quality in mega city Karachi. In general, coal combustion, traffic, and industry have compromised the air quality situation in mega cities of Pakistan. (Haider et al., 2021).

This study addresses mainly three challenges in the existing air quality research in Pakistan. Conventionally, NO₂ drivers from meteorological, land-use/land-cover (LULC) and emission domains are studied in isolation or at coarse resolution using a linear regression analysis approach that fails to properly capture non-linear relationships among atmospheric variables. This study processes multi-satellite datasets by using machine learning approach and captures non-linear relationships among the variables. The challenges addressed by this study are summarized as follows: (i) Previous studies

use meteorological, sectoral emissions, and LULC-related datasets in isolation that fail to identify an overall impact of inter-related variables. This study integrates 19 parameters from various data sources like Sentinel-5P, ERA5-Land, Copernicus Atmosphere Monitoring Service (CAMS) Emissions, Global Human Settlement Population Grid (GHS-POP), European Space Agency (ESA) WorldCover, Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM), and gROADS into a unified dataset and models the impact of these parameters on tropospheric NO₂ at a large scale. (ii) Most of the studies use linear regression analysis approach for NO₂ studies which fails to identify complex relationships among the highly intertwined set of parameters. This study uses the machine learning model XGBoost that has full potential to automatically understand the individual role of each variable. (iii) In Pakistan, existing studies on NO₂ drivers' analysis are limited to specific cities or division level. Limited study has studied the spatial dominance of various parameters across the whole country. The current study uses SHapley Additive exPlanations (SHAP) analysis and explores what and where parameters are influencing the tropospheric NO₂ across entire Pakistan.

The rest of the paper covers materials and methodology (section 2), results and discussion (section 3) and conclusion (section 4).

2. Materials and Methods

2.1 Study Area

The study area as shown in figure 1 covers entire Pakistan, a South Asian developing country lying between 24° - 37° N and 60° - 77° E, with total land area of approximately 88,913 km². It shares borders with India to the east, Iran to the west, Afghanistan to the north, and China to

northeast, while Arabian Sea forms its southern coastline. It has diverse topography, ranging from northern areas' glaciated peaks to the fertile Indus River plain, finally diffusing into arid plateaus of Baluchistan and southern coastal belt. The presence of snow-capped mountain peaks and Arabian Sea on its two extremes separated by vast plain area, and its proximity to equatorial belt all form the natural basis of existing regional atmospheric circulation patterns that bring four seasons to this region. Generally, summer season is hot and experiences highly dynamic meteorological conditions causing heavy precipitation during monsoon period, while winter season is dry and cold with stagnant meteorology. Consequently, winter season favors concentrated accumulation of NO₂ and summer season tends to disperse or dilute the NO₂ concentrations.

In recent decades, the study area has undergone industrial expansion and fast urbanization, resulting in expanded urban-industrial corridors and increased energy demands. In addition, transformation of rural or agricultural area into residential settlements has also modified LULC. Overall, Punjab and Sindh provinces have higher population densities as compared to Baluchistan, KPK and Gilgit provinces; therefore, rigorous agricultural practices, crop residue burning and anthropogenic activities take place in these regions which in turn deteriorate the air quality of study area. Karachi and Lahore are recognized to have NO₂ concentrations above WHO-recommended limits making air unhealthy for their communities. Thus, understanding spatiotemporal dynamics and the driving factors behind NO₂ anomalies in the region is an urgent need of the time, so that policymakers could design and implement an effective strategy to mitigate or control the worsening situation.

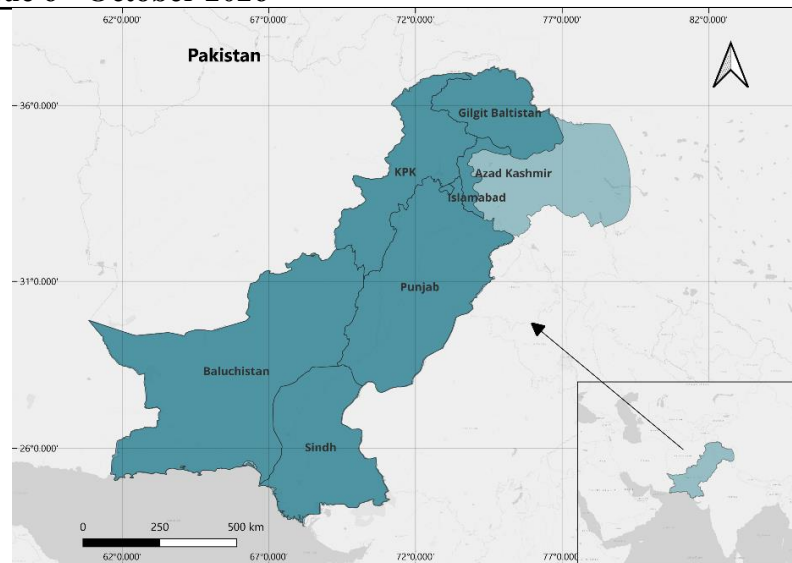


Figure 1: Study area Pakistan along with its provinces

2.2 Datasets

This study uses a unified dataset comprising of tropospheric NO_2 and auxiliary data consisting of 19 most commonly used predictive parameters. The predictive parameters were taken from various domains like meteorology, sectoral emissions, topography, and LULC to see their impact on target NO_2 .

2.2.1 Tropospheric NO_2

Tropospheric NO_2 was collected from a dedicated Earth-observing satellite Sentinel 5-Precursor (Sentinel 5P), for the study period 2019-2025. Sentinel-5P has been orbiting in a sun-synchronous orbit since its first launch in 2017 by ESA as a part of the Copernicus Program. Preferring Sentinel-5P data over other on-orbit satellites is due to its better resolution than others. It carries a TROPOMI instrument for recording Earth observations from space on a daily basis with a high resolution of $3.5 \times 3.5 \text{ km}^2$ (updated to $3.5 \times 7 \text{ km}^2$ by 6 August, 2019). The offline version of level 3 (L3) NO_2 product (COPERNICUS/S5P/OFFL/L3_NO2) was downloaded from the Google Earth Engine (GEE) platform for the study period using its Application Programming Interface (API). The L3 product provides higher quality data than level 2 (L2) NO_2 data because of its quality filter ($\text{qa_value} \geq 0.75$ (Wei et al., 2025)).

2.2.2 Auxiliary Data

Auxiliary data consists of meteorological parameters, demographic/geographic data, anthropogenic emissions, and LULC dataset. Meteorological parameters, including 2m surface temperature (T2M), boundary layer height (BLH), total precipitation (TP) and wind components (U10, V10), were obtained from the fifth-generation reanalysis product ERA5 produced by European Center for Medium Range Weather Forecasts (ECMWF) with spatial resolution of $0.25^\circ \times 0.25^\circ$ and temporal resolution equal to 1 hour. Surface elevation was recorded from DEM, a product of NASA's SRTM, with fine spatial resolution of approximately 30m. Demographic data such as 2020 reference of road density (RD) was collected from OpenStreetMap (OSM) having spatial resolution of $1 \times 1 \text{ km}^2$ whereas Nighttime lighting (NTL) data was collected with a spatial resolution of 500m from Visible Infrared Imaging Radiometer Suite Nighttime Lights (VIIRS). Anthropogenic sectoral emission data related to traffic emissions (TEM), industrial emissions (IEM), power sector emissions (PEM), and agricultural residue burning (AEM) were collected from the CAMS with a spatial resolution of $0.1^\circ \times 0.1^\circ$. LULC dataset originally consists of 9 built-in classes, but only 4 classes were selected for this study, which are cropland, forest cover, barren land, and urban/built-up class. The details of all datasets, along with their

selected parameters, spatial resolution, and temporal resolution, can be seen in Table 1.

2.2.3 Data Preprocessing

The downloaded raw datasets were cleaned and mapped to a common grid using bilinear interpolation and a nearest-neighbor approach (for LULC data), taking TROPMI NO₂ as a spatial reference system. This was necessary because raw datasets were originally not aligned with each other due to the inherent differences in their spatiotemporal resolutions. Raw datasets could not be fed into machine learning models directly unless they were aligned with each other. Therefore, datasets with finer or coarser resolutions were upscaled or downscaled to make them spatially aligned with the primary grid using bilinear interpolation or a nearest-neighbor approach (for LULC). The datasets having coarser resolutions—such as ERA5 meteorological variables (~9km e.g., BLH, T2M, TP, solar radiation, wind components) and CAMS sector-specific emissions (~10km industry, power, traffic, agriculture)—were downscaled to the target NO₂ grid using bilinear interpolation method, whereas finer datasets — such as ESA WorldCover, DEM elevation, gROADS were upscaled using spatial averaging techniques. Temporally, all time-varying meteorological and sector-specific emission layers were processed into uniform monthly aggregations. However, the digital elevation, road density and LULC datasets were static and distinct collections with respect to time, with a 2020 or 2021 reference epoch, therefore their

monthly aggregates were not possible. They were simply added to the monthly aggregated dataset as additional features to prepare a master dataset ('master_ml_dataset.parquet').

Predictor variables were selected using variation inflation factor (VIF), taking default threshold of 10 as the decision point. The variables with VIF value above threshold were dropped. Originally, wind is recorded in its direction components (u,v), we computed wind speed from these vector components by using simple magnitude formula $\text{windspeed} = \sqrt{u^2 + v^2}$ and made it part of our feature set. In such a large and multi-disciplinary feature-set, it was quite possible to have interdependent features accidentally that could influence the results of later processing. Therefore, to ensure that no two highly correlated features co-exist in our dataset, we computed Pearson correlation coefficients as shown in Figure 6. Some of the features definitely showed positive and other exhibited negative correlation. However, we set a fair threshold $|r| < 0.85$ to keep the elements together in dataset. Finally, dataset consisted of overall 24 features, out of which 19 were core predictive features (7 from meteorology, 4 related to emissions, 5 major classes from LULC, surface elevation, road density and population density), while remaining 5 features were related to metadata providing spatiotemporal coordinates like dataset collection time, latitude, longitude and spatial reference (coordinate reference system).

Table 1: Datasets collected from multiple satellite sources showing necessary details of each dataset

Dataset / Product	Variable(s)		Spatial Resolution	Temporal / Period	Provider / Organization
TROPOMI NO ₂	Tropospheric (mol/m ²)	NO ₂	3.5 km × 5.5 km	Monthly (2019–2025)	GEE
ERA5-Land	BLH (m), 2m Temp (K), Total Precipitation (m), Wind U/V (m/s)		0.1° × 0.1° (~9 km × 9 km)	Monthly (2019–2025)	ECMWF
CAMS Emission	Sectoral NOx emissions (kg m ⁻² s ⁻¹): Industry, Power, Traffic, Agri Burning		0.1° × 0.1° (~10 km × 10 km)	Monthly (2019–2025)	CAMS
GHSPOP	Population density (people/km ²)	density	1 km × 1 km	2020 Epoch Reference	European Commission JRC / GEO

ESA WorldCover	LULC Classes (Cropland, Built-up, Forest, Grassland, Bare, Water)	10 m	2021 Map	Reference	ESA
DEM	Surface Elevation (m)	~ 30 m	2020		SRTM
gROADS	Road Density (km/km ²)	1 km Grid	2020	Reference	OpenStreetMap

2.3 Methodology

2.3.1 Random Forest

Random Forest (RF) is a tree-based machine learning algorithm that constructs a large number of independent decision trees known as forest and trains each decision tree on a random sample of training subset. Therefore, each decision tree receives a slightly different subset of observations and predictor variables. RF allows all decision trees to work in parallel, but they remain completely isolated from each other that is why each tree learns a unique perspective of the dataset and makes its own independent decision without being influenced by the prediction of other trees. Individual decisions vary from each other; final decision is made by taking the average of all predictions which improves the prediction accuracy. Its actual strength lies in the random sampling approach called bootstrap sampling which eliminates the chances of overfitting and enables it to learn non-linear and complex interactions of the predictor variables that govern tropospheric NO₂ (Jian et al., 2024).

2.3.2 Extreme Gradient Boosting Algorithm

Extreme Gradient Boosting (XGBoost) Algorithm is also a tree-based machine learning algorithm but adopts a slightly different strategy from RF while building decision trees. Unlike RF that trains all decision trees in parallel and ensures quite isolation among them, XGBoost uses sequential approach and each next decision tree is bound to fix the error of its predecessor. Thus, each next tree corrects the error of previous ensemble of trees and adds its own incremental contribution to the overall accuracy of final prediction. In this way, XGBoost learns from its own errors actively during training process, keeps on improving prediction accuracy on each step and finally may have higher accuracy than RF. Its uses specific control parameters (hyperparameters) to keep a check on its tree structure and learning process, such

as *max_depth* to stop tree growth after achieving a certain depth, *min_child_weight* to require minimum weight or evidence before creating a new tree, *n_estimators* to restrict total number of trees that can be build and *learning_rate* to set the contribution of each new tree in learning process. It uses internal mathematical penalty (regularization) method to avoid unnecessary complexity and overfitting.

2.3.3 SHAP Analysis

Shapley Additive Explanations (SHAP) is a method to evaluate marginal contribution of each predictor variable to final prediction of a model. It is based on cooperative game theory where each player has some contribution to the final achievement of the team. SHAP method quantifies the relative importance of each feature by assessing its change impact on final output of the model. SHAP value may have positive or negative sign depending upon the direction of its impact. Positive value means that the predictor variable or feature tends to move the model output towards the predicted output while the negative value means the predictor variable or feature tends to move the model output away from the predicted output.

2.3.4 Model Building

To capture non-linear and complex interactions between NO₂ pollution and its drivers, two machine learning models RF and XGBoost were trained and tested on the dataset. While building RF and XGBoost models, dataset was cleaned and split into two partitions: first 80% was used as training subset and remaining 20% as testing subset. The best configurations of both models were achieved by optimizing their hyperparameters along with a 5-fold cross validation. RF achieved optimal balance on *n_estimators* = 200, *max_depth* = 16, *max_features* = *sqrt*, *max_samples* = 0.3 while XGBoost achieved that on *n_estimators* = 300, *max_depth* = 8, *learning_rate* = 0.08 and *subsample* = 0.8.

Finally, both trained models were tested on remaining unseen 20% of the dataset, specially reserved for testing purpose. Standard performance metrics R-squared (R^2), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were evaluated to compare performance of the models.

3. Results & Discussion

3.1 Model Evaluation

Performance matrix as shown in table 2 reveals that XGBoost model outperformed the RF with a $R^2=0.928$ (92.8% accuracy in explaining the dataset), $RMSE = 3.70 \times 10^{-6} \text{ mol/m}^2$, and $MAE = 2.40 \times 10^{-6} \text{ mol/m}^2$. Figure 2 provides further deep insight into the performance evaluation by comparing observed NO_2 and predicted NO_2 in form of scatter plots. The scatter plots depict that both models showed outstanding performance in predicting tropospheric NO_2 concentrations accurately under typical baseline conditions; however, XGBoost systematically

outperformed RF as the optimal conditions were reached. XGBoost has a clear performance edge over RF in predicting acute NO_2 spikes, such as localized industrial emissions, dense urban traffic peaks, and severe winter stagnation, more accurately. This is because of both models' internal mechanisms for how they treat acute spikes. RF builds hundreds of trees independent from each other and finally averages their predictions, which makes it resistant to random noise and overfitting. Acute pollution spikes are accurately predicted by certain trees, but when averaged with moderate predictions by other independent trees finally become flattened (Wei et al., 2025). In contrast, XGBoost uses gradient boosting and builds trees sequentially correcting residual errors of prior trees before going ahead. Initially, acute pollution spikes cause large negative residuals that, after accepting correction in subsequent boosting rounds, enable XGBoost to capture pollution spikes successfully.

Table 2: Performance metrics of Random Forest and XGBoost models

Model Name	R^2	RMSE	MAE
RF	0.9008	$4.37 \times 10^{-6} \text{ mol/m}^2$	$2.63 \times 10^{-6} \text{ mol/m}^2$
XGBoost	0.9282	$3.71 \times 10^{-6} \text{ mol/m}^2$	$2.40 \times 10^{-6} \text{ mol/m}^2$

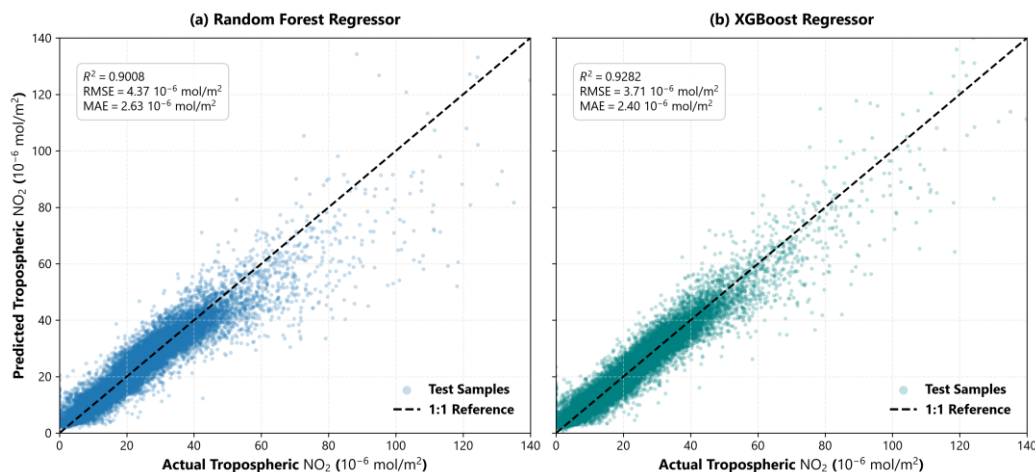


Figure 2: Scatter plots showing performance comparison: a) RF, b) XGBoost

3.2 Spatiotemporal Distribution of Predicted Tropospheric NO_2

XGBoost model has proven to effectively represent the regional variability in the vertical column density (10^{-6} mol/m^2) of tropospheric NO_2 in the region of Pakistan and its unique seasonal behavior. Spatially, high NO_2

concentrations are observed along the high-density anthropogenic corridors primarily in Punjab, Central and North-East plains, which include major industrial/metropolitan cities such as Lahore, Faisalabad, Gujranwala, Rawalpindi / Islamabad, as well as the coastal megacity of Karachi and Port Qasim in the south

as shown in Figure 3. The hotspots are concentrated near heavy traffic emission areas, industrial groups and thermal power plants. In contrast, other areas with high background NO_2 concentrations are predominantly the arid plateaus of Balochistan and the high-altitude topography of northern Khyber Pakhtunkhwa and Gilgit-Baltistan, where vehicular activity and stationary emission sources are very low.

Figure 4 shows seasonal variations in predicted NO_2 concentration over Pakistan using the XGBoost model. The peak of localized pollution events (up to $\sim 329 \times 10^{-6} \text{ mol/m}^2$, with a standard deviation = $15.97 \times 10^{-6} \text{ mol/m}^2$) occurs during winter (DJF) when shallow planetary boundary layer heights, frequent thermal inversion at the surface, and weak wind speeds result in the trapping of primary emissions near the surface, which

intensifies the air pollution during this season across the Indus Basin. The regional intensification is observed during autumn (SON; mean $\sim 20.71 \times 10^{-6} \text{ mol/m}^2$), with the main reason being the post-monsoon atmospheric stabilization and the extensive post-harvest agricultural residue burning in the transboundary Indo-Gangetic Plains. On the contrary, Summer (JJA) has relatively higher country-wise spatial mean ($\sim 22.97 \times 10^{-6} \text{ mol/m}^2$), and relatively lower spatial variability (std dev = $10.11 \times 10^{-6} \text{ mol/m}^2$) because of the deeper vertical mixing and intense thermal convection that cause dispersion of emissions over extensive geographic domains. Finally, Spring (MAM) has moderate concentrations in the transition zone ($\sim 19.45 \times 10^{-6} \text{ mol/m}^2$) representative of greater boundary layer heights and improved ventilation.

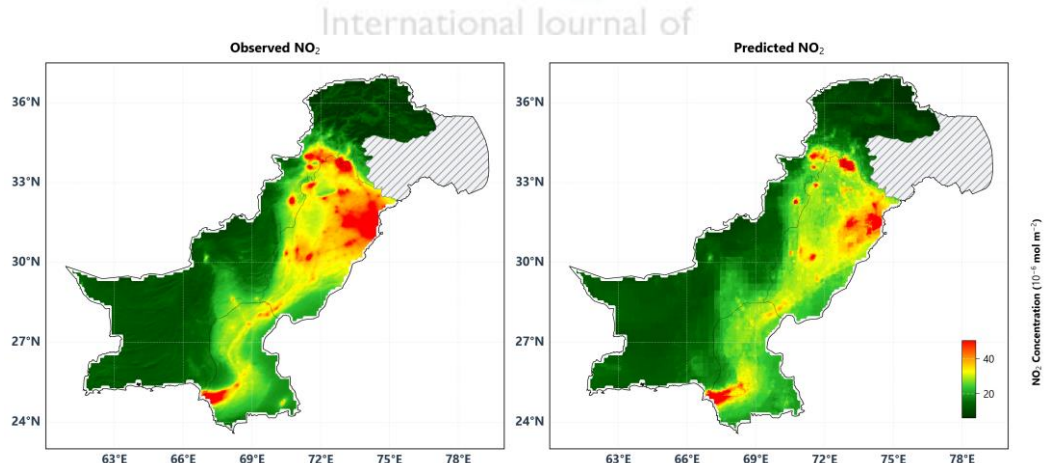


Figure 3: Shows actual NO_2 observed through S5P TROPOMI and predicted NO_2 concentration using XGBoost.

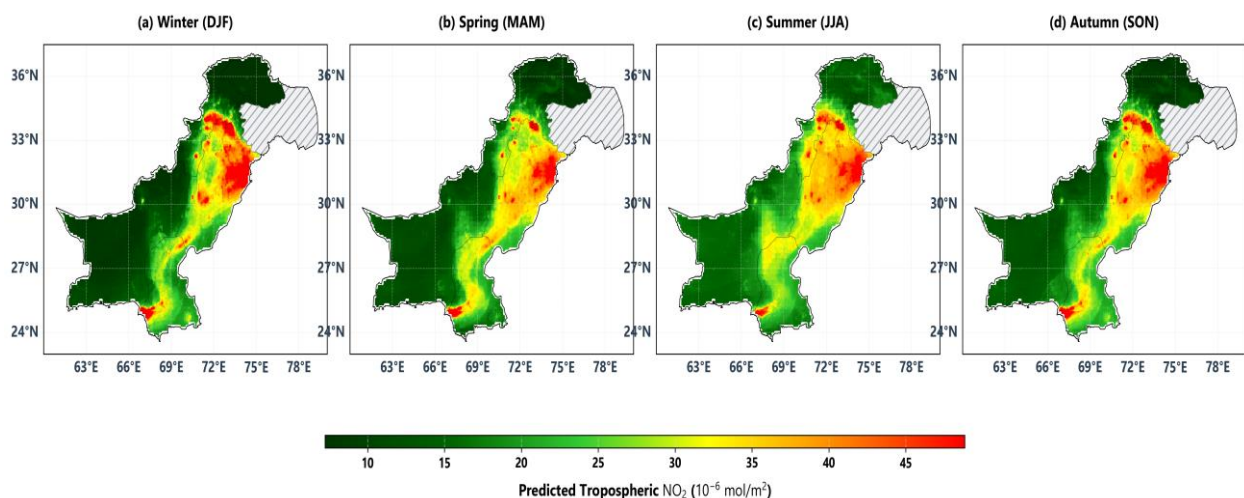


Figure 4: Shows seasonal variations in predicted NO_2 concentration using XGBoost

3.3 Feature Importance Analysis

The relative importance of different environmental and anthropogenic predictors and their directionality in driving tropospheric NO₂ predictions were determined by calculating SHAP values for each of the 19 predictors, which were then summarized with the mean absolute SHAP ranking and beeswarm plot (Figures 5 & 6). The top five factors are anthropogenic infrastructure and meteorology, with road density (ROAD) being the most important, contributing 28.4% of the total model attribution (mean |SHAP| = 4.920×10^{-6} mol/m²), followed by 2-meter air temperature (TEMP; 17.7%, 3.063×10^{-6} mol/m²), digital elevation (DEM; 10.0%, 1.726×10^{-6} mol m⁻²), industrial emissions (IND; 7.8%, 1.350×10^{-6} mol/m²), and agricultural emissions (AGR; 7.2%, 1.246×10^{-6} mol/m²). The ROAD, IND, TRF emissions attributes have clearly differentiated physical and non-linear response profiles: high values of ROAD,

IND and traffic emissions (TRF) have heavy positive tails extending well beyond +30 to +60 $\times 10^{-6}$ mol/m², indicating that dense transportation corridors and industrial point sources are the direct causes of the severe localized NO₂ spikes. Topographically, DEM shows a strong negative correlation, with the magenta/red points at higher elevations in the landscape consistently mapping to SHAP values in the negative range (down to -15×10^{-6} mol/m²), while the basin terrain at lower elevations shows pollutant pooling. Moreover, the meteorological ventilation dynamics are well represented: lower BLH and less strong wind speed are correlated with positive SHAP contributions (up to +20 to +30 $\times 10^{-6}$ mol/m²), indicating that shallow inversion layers and weak winds physically hinder vertical mixing and enhance the accumulation of NO₂ at ground level, whereas stronger wind speeds and higher BLH physically weaken the accumulation of NO₂ in the column.

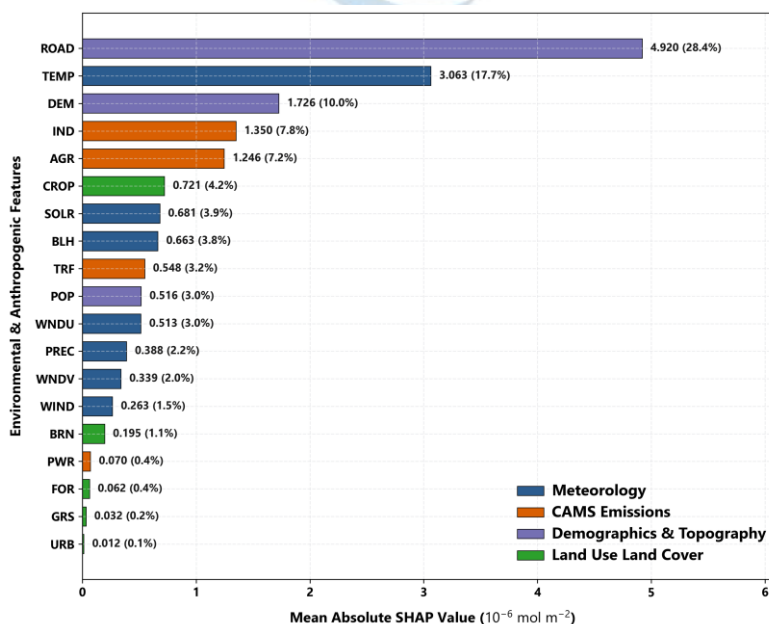


Figure 5: Feature importance ranking using mean SHAP value by XGBoost

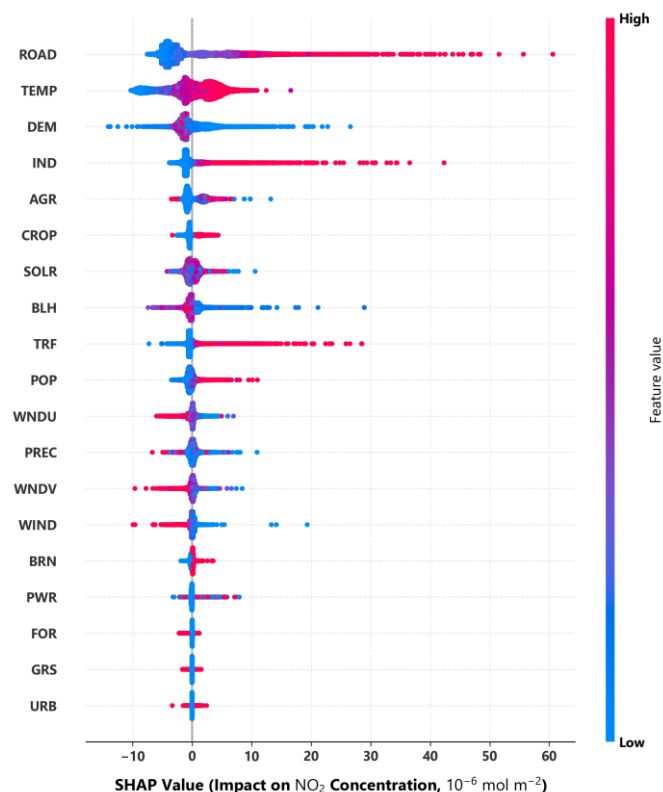


Figure 6: SHAP Feature Impact Distributions and Directional Contributions for NO₂ Predictions in Pakistan

3.3 Features Correlation Analysis

Pearson correlation analysis was performed across all modeling properties (Figure 7) to explore the interaction of all the environmental and human-related predictors. The results demonstrate clear and physically interpretable clusters within the human activities, geography and weather dimensions, while also showing that there are no two features that are overly redundant in data (all $|r| < 0.70$). The highest positive human-activity relationship is found between ROAD and population density (POP) ($r = 0.69$), with a strong relationship between TRF and urban land cover (URB) ($r = 0.51$) also evident. The AGR are also strongly associated with cropland area (CROP) ($r = 0.55$) and road networks ($r = 0.41$) in the agriculture sector. In the meteorological field, TEMP ($r = 0.69$) and solar radiation (SOLR) ($r = 0.62$) show strong positive correlations with BLH, consistent with the direct relationship between the heat of the

Earth's surface due to intense sunshine and the strong thermal convection that extends the vertical depth of atmospheric mixing. Conversely, there are significant negative associations among human settlement, road density ($r = -0.43$), cropland ($r = -0.39$), and agricultural emissions ($r = -0.38$) with topography (DEM) indicating that in northern and western Pakistan high elevations and rugged terrain naturally limit human settlement, the development of roads and cropland, and emissions from agriculture. Last, the inverse relationship between fertile croplands and dry, barren land is illustrated by the very negative correlation ($r = -0.60$). Importantly, none of the correlations between any two predictors exceed the typical multicollinearity warning value of $|r| = 0.80$, so the machine learning models can learn independent, complementary predictors without distortion or instability.

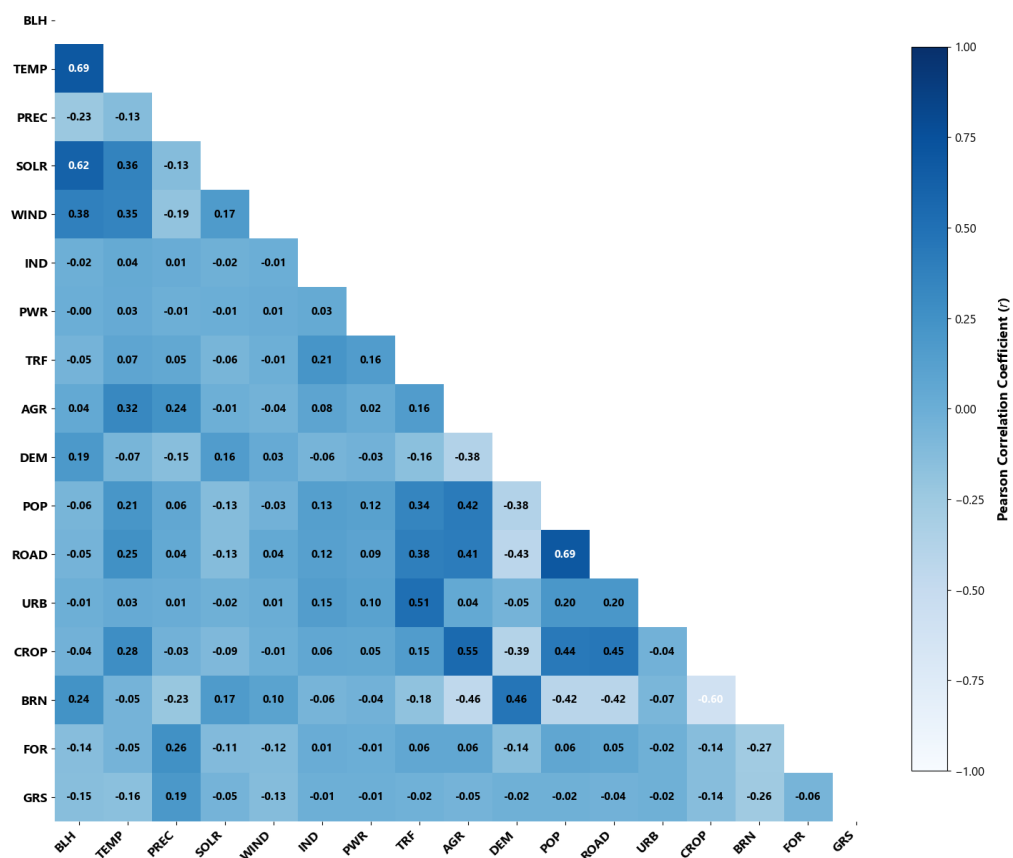


Figure 7: Pearson correlation coefficients of all features

3.4 Spatial Dominance of Environmental & Anthropogenic Drivers

The dominant driver categories are mapped on the 0.05° grid cells across Pakistan, indicating significant regional disparities in the drivers of tropospheric NO₂ distribution, as influenced by both physical terrain and atmospheric dispersion, in addition to local combustion sources (Figure 8). The geographically ubiquitous dominant domain that dominates primary NO₂ attribution is Demographics and Topography (DT) with 65.5% (n = 59,276) of cells across the country. The density of the roads, population, and topography (elevation DEM) affects the movement of humans and their vehicles in low-lying basins throughout the arid Balochistan plateau, western mountain valleys and the central Indus basin, resulting in structural corridors that regulate physical pollutant trapping and human vehicular activity. Second most extensive driver domain is meteorology which accounts for 29.8 % of the landscape (n = 26,917), mainly over the high and

complex topography of northern Pakistan (Gilgit-Baltistan and upper Khyber Pakhtunkhwa) and along the southern coastal belts. Natural atmospheric processes such as compression of the BLH, thermal stratification, photolysis under solar radiation conditions and wind advection are the dominant factors that control the variation of NO₂ in these pristine northern highland and maritime regions, where baseline anthropogenic activity is limited. Anthropogenic Emissions, by contrast, is the dominant source in 4.7% of grid cells (n = 4,244), showing a localized and sharp geographic distribution in Pakistan's main urban-industrial hubs and intensive agricultural regions in central and northern regions of Punjab (especially in the Lahore–Faisalabad–Gujranwala industrial belt) and metropolitan Karachi. Intense point- and mobile-source emissions, including heavy vehicle use, coal/gas-fired thermal power generation and seasonal biomass burning, dominate local environmental baselines in these localized hot spots, resulting

in large-scale anthropogenic influence on columnar NO_2 . The local environmental baseline is overwhelmed by intense point- and mobile-source emissions (heavy vehicle use, coal/gas-fired thermal power generation, and seasonal biomass burning) in these localized hot spots, leading to a dominant anthropogenic influence on columnar NO_2 . Land Use Land Cover accounted for <0.1% of grid cells as primary drivers, showing that LC classes are mainly secondary modulating factors that affect surface roughness and albedo, but are not primary drivers. In addition, regions of low confidence (confidence margin < Q25 = 1.4814) represent transitional areas (e.g., between Balochistan ranges and Indus plain) where strong atmospheric advection and complex surface topography overlap with similar magnitude of attribution, and indicate areas where multi-domain driver interaction is tight. Spatial mapping of dominant categories of drivers across Pakistan (Figure 8) indicates high regionalization as a function of local land use, urban density and topography. Transport infrastructure and traffic emissions are found to be the main dominant driver in major urban

centers/ high density traffic corridors such as the GT Road corridor, Karachi, Lahore, and Faisalabad. In Punjab and Sindh, the NO_2 level is elevated due to agricultural factors (CROP and AGR), which dominate the fertile Indus basin with extensive cultivation, fertilizer application, and biomass burning during the agricultural season. Other localized dominance clusters occur in the vicinity of major manufacturing centers and energy production centers, such as emissions from industrial and power generation sectors. Conversely, the background rural and mountainous areas in Gilgit Baltistan, Khyber Pakhtunkhwa, and western Balochistan, which are dominated by meteorological parameters (BLH, TEMP, and WIND) and topographic elevation (DEM) and have low levels of human activity, are where ambient NO_2 levels are regulated by atmospheric dispersion processes. Low-confidence transition zones are masked, revealing clear spatial demarcations, with anthropogenic transport and agriculture dominating the central plains, and natural atmospheric regulators dominating the northern and western high-altitude areas.

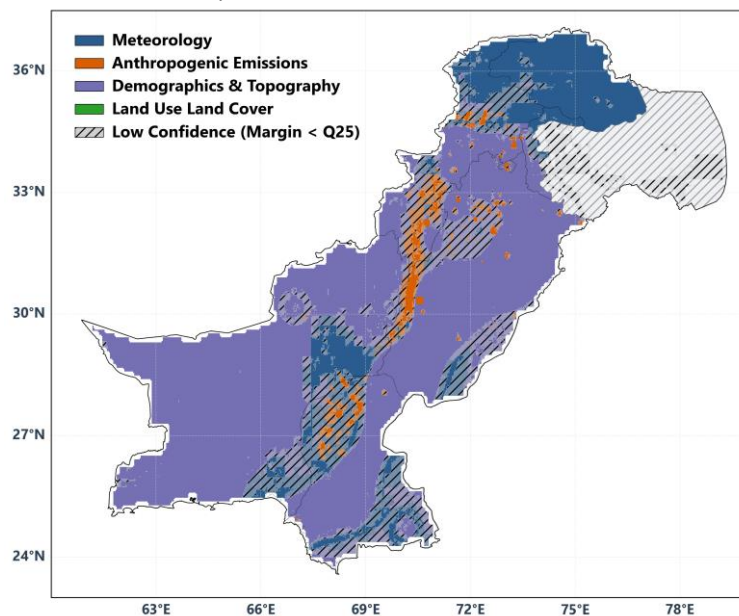


Figure 8: Spatial Classification Map of NO_2 Drivers Across Pakistan

3.5 Model Validation

Figure 9(a) shows the difference between NO_2 concentrations predicted by XGBoost model and the ones observed by the satellite (Predicted - Observed = Residual), and evaluates how well

the model predicts the satellite observations. The resulting histogram and the smooth dark blue density curve are a classic bell-shaped (normal) distribution with a center near the middle of the distribution. The statistical results

are in line with the high accuracy of the model, as the coefficient of determination ($R^2 = 0.9282$) reveals that the model has successfully explained nearly 92.8% of the total variation of NO_2 across Pakistan. The Mean Absolute Error ($\text{MAE} = 2.40 \times 10^{-6} \text{ mol/m}^2$) indicates that on average predictions are within 2.4 units of the actual observations, and the Root Mean Squared Error ($\text{RMSE} = 3.70 \times 10^{-6} \text{ mol/m}^2$) confirms that large errors in predictions are very rare. In addition, the overall mean bias is almost zero ($-0.0040 \times 10^{-6} \text{ mol/m}^2$) as the red dotted line almost coincides with the dashed line of zero error. This balance shows that the model does not have any systematic overestimation or underestimation, and thus it has been able to learn the complex non-linear relationships between NO_2 pollution and the 19 underlying drivers (including industrial emissions, power plants, traffic, population density, wind speed and temperature) correctly.

Figure 9(b) shows the geographical distribution of the errors in the predictions in 90,442 spatial grid cells to see if the model has similar performance on different landscapes and regions of Pakistan. This map shows that the model predicted the NO_2 concentration well in

the white colored regions; it underpredicted NO_2 in the blue regions (Predicted < Observed); and it overpredicted NO_2 in the red regions (Predicted > Observed). The light or white shading covers over 80% of Pakistan's landmass, from the vast deserts of Balochistan, through the rural plains of Sindh and Punjab, to the high mountains of Khyber Pakhtunkhwa. This is true to the extent that this is a neutral model, widely applicable across the country and not biased by region or topography. It can be seen that underestimation values (blue spots) are directly over the dense urban and industrial core of Lahore, Karachi and Faisalabad. This is due to the fact that there are extreme and sudden peaks of pollution in big cities, and these peaks are rare events, so the tree-based machine learning algorithms smooth away the extreme peak values slightly towards the regional average. The semi-colors (light red) around the larger city boundaries represent the natural way that urban pollution plumes spread into rural areas by wind. Overall, the spatial map further supports the notion of small, logical, and limited prediction errors, which are clearly concentrated in the major city centers and not indicative of a general structural error.

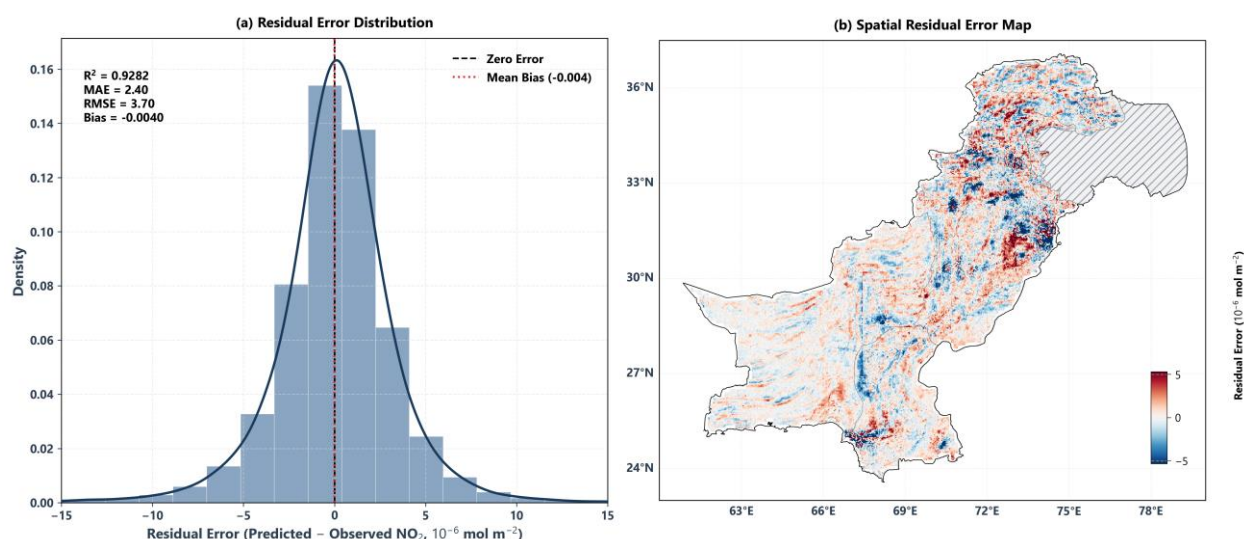


Figure 4: a) Histogram showing residual error distribution and b) showing spatial residual error distribution across Pakistan

4. Conclusion

This study helps to carry out high spatiotemporal attribution of tropospheric NO_2 pollution across Pakistan using satellite

observations and SHapley Additive exPlanations (SHAP) model. Results highlight that anthropogenic emissions are concentrated in intense focal hotspots on about 4.7% of the

land, mostly in the central Punjab industrial corridor and in the Karachi metropolitan basin, with these small areas contributing a much higher percentage of the toxic burden, which is intensified by structural and atmospheric dynamics. Vulnerability to baseline NO₂ exposure is governed by demographic and topographic factors and affects 65.5% of the country, where physical factors of the road network and basin topography cause physical confinement of pollutants. At the same time, the meteorological conditions affect the variability of NO₂ over 29.8% of the territory, indicating that, beyond emission peaks, shallow planetary boundary layers and thermal inversions are responsible for the onset of seasonal winter smog crises. This framework also goes beyond the concept of a nationwide emission cap and introduces precision-targeted airshed management by identifying the spatial borders of emission-dominated hotspots, corridors that are trapped on the terrain, and transition areas that are influenced by the weather.

Based on the study findings, here are a few recommendations for policymakers that might be followed to control NO₂ pollution in Pakistan. The government should put in place systems to permanently and automatically measure emissions in factories, stop the use of old tyres and inferior fuels, and ensure complete conversion to cleaner zig-zag brick kilns. Authorities should be strict while testing vehicles' emission health, and smoke-emitting vehicles should be strictly banned, and electric buses should be introduced within the city. Authorities must also ensure that no fuel station is selling low-quality or high-sulphur fuel. Moreover, the farmers should be facilitated to have access to farm machinery (Happy Seeders) and bio-decomposers, as they need to plant the next crop as soon as possible without burning residue to the ground. The deployment of reliable and calibrated air monitoring stations in smaller towns, in industrial corridors, and in valley areas will give authorities real-time data to ensure they monitor the pollution with accuracy.

Acknowledgment

We are thankful to the Google Earth Engine (GEE) for providing the TROPOMI Sentinel-5P NO₂ satellite data for the analysis. We are also

thankful ECMWF, SRTM, CAMS and OpenStreetMap for producing the data used in this research effort.

Authors Contribution

Muhammad Abbas: Conceptualization, Data curation, Investigation, Formal analysis, Writing—original draft Zia Ul Haq: Project administration, Resources, Supervision, Validation, Writing—review & editing Shahid Pervez: Supervision, Visualization, Methodology, Validation, Writing—review & editing.

Funding This work does not get any funding from any organization.

Data availability Data will be made available on request.

Declarations

Ethical approval Not required.

Consent to publish Not Applicable.

Consent to participate All authors participate in this research.

Competing interests The authors declare no competing interests.

Bibliography

- Ahmad, S. S., Biiker, P., Emberson, L., & Shabbir, R. (2011). Monitoring Nitrogen Dioxide Levels in Urban Areas in Rawalpindi, Pakistan. *Water, Air, & Soil Pollution*, 220(1-4), 141-150. <https://doi.org/10.1007/s11270-011-0741-9>
- Ahmed, D., Shams, Z. I., Ahmed, M., & Siddiqui, M. F. (2022). Spatio-Temporal Variations of Lower Tropospheric Pollutants and Their Relationship With Meteorological Factors in Karachi, Pakistan. *Arab Gulf Journal of Scientific Research*, 118-137. <https://doi.org/10.51758/AGJSR-02-2021-0014>

- Anjum, M. S., Ali, S. M., Imad-ud-din, M., Subhani, M. A., Anwar, M. N., Nizami, A.-S., Ashraf, U., & Khokhar, M. F. (2021). An Emerged Challenge of Air Pollution and Ever-Increasing Particulate Matter in Pakistan; A Critical Review. *Journal of Hazardous Materials*, 402, 123943. <https://doi.org/10.1016/j.jhazmat.2020.123943>
- Bai, X., Wang, Y., Gui, L., Tao, M., & Zeng, M. (2025). Comparing the Influences on NO₂ Changes in Terms of Inter-Annual and Seasonal Variations in Different Regions of China: Meteorological and Anthropogenic Contributions. *Remote Sensing*, 17(1), 121. <https://doi.org/10.3390/rs17010121>
- Bilal, M., Mhawish, A., Nichol, J. E., Qiu, Z., Nazeer, M., Ali, M. A., de Leeuw, G., Levy, R. C., Wang, Y., Chen, Y., Wang, L., Shi, Y., Bleiweiss, M. P., Mazhar, U., Atique, L., & Ke, S. (2021). Air pollution scenario over Pakistan: Characterization and ranking of extremely polluted cities using long-term concentrations of aerosols and trace gases. *Remote Sensing of Environment*, 264, 112617. <https://doi.org/10.1016/j.rse.2021.112617>
- Carvalho, H. (2021). New WHO global air quality guidelines: More pressure on nations to reduce air pollution levels. *The Lancet Planetary Health*, 5(11), e760–e761. [https://doi.org/10.1016/S2542-5196\(21\)00287-4](https://doi.org/10.1016/S2542-5196(21)00287-4)
- Dilawar, A., Chen, B., Ul-Haq, Z., Ali, S., Sajjad, M. M., Junjun, F., Gemechu, T. M., Guo, M., Dilawar, H., Zhang, H., Zicheng, Z., & Lohdi, E. (2023). Evaluating the potential footprints of land use and land cover and climate dynamics on atmospheric pollution in Pakistan. *Frontiers in Environmental Science*, 11, 1272155. <https://doi.org/10.3389/FENV.2023.1272155/BIBTEX>
- Ghani, M. Z., Hassan, M., Chishtie, F. A., Ali, S. S. F., Hussain Shah, S. W., & Bilal, A. (2026). Spatiotemporal analysis of climatic and anthropogenic drivers of winter smog in the Indo-Gangetic Plain with a focus on Lahore. *Environmental Science and Pollution Research*, 33(17), 8311–8334. <https://doi.org/10.1007/s11356-026-37816-7>
- Haider, R., Yasar, A., Tabinda, A. B., & Basharat, A. (2021). ASSESSMENT OF STREET AIR POLLUTION THROUGH OSPM MODEL IN LAHORE, PAKISTAN. *Environmental Engineering and Management Journal*, 20(1), 73–80. <https://doi.org/10.30638/eemj.2021.008>
- Haslina, S., Shafie, M., Wafiy, M., Ramli, A., & Abdullah, A. L. (2024). SPATIOTEMPORAL OF NITROGEN DIOXIDE (NO₂) CONCENTRATION IN THE URBAN ENVIRONMENT OF KLANG VALLEY, MALAYSIA. In *Journal of the Malaysian Institute of Planners* (Vol. 22).
- Jian, L., Xia, X., Wang, Y., Liu, X., Zhang, Y., & Yang, Q. (2024). Spatiotemporal dynamic relationships and simulation of urban spatial form changes and land surface temperature: A case study in Chengdu, China. *Frontiers in Public Health*, 12. <https://doi.org/10.3389/fpubh.2024.1357624>
- Jin, X., Zhu, Q., & Cohen, R. C. (2021). Direct estimates of biomass burning NO_x emissions and lifetimes using daily observations from TROPOMI. *Atmospheric Chemistry and Physics*, 21(20), 15569–15587. <https://doi.org/10.5194/acp-21-15569-2021>

- Johnson, M., Shin, H. H., Roberts, E., Sun, L., Fisher, M., Hystad, P., Van Donkelaar, A., Martin, R. V., Fraser, W. D., Lavigne, E., Clark, N., Beaulac, V., & Arbuckle, T. E. (2022). Critical Time Windows for Air Pollution Exposure and Birth Weight in a Multicity Canadian Pregnancy Cohort. *Epidemiology*, 33(1), 7-16. <https://doi.org/10.1097/EDE.0000000000001428>
- Khokhar, M. F., Nisar, M., Noreen, A., Khan, W. R., & Hakeem, K. R. (2017). Investigating the nitrogen dioxide concentrations in the boundary layer by using multi-axis spectroscopic measurements and comparison with satellite observations. *Environmental Science and Pollution Research*, 24(3), 2827-2839. <https://doi.org/10.1007/s11356-016-7907-3>
- Li, T., Li, J., Sun, Z., Jiang, H., Tian, C., & Zhang, G. (2023). High contribution of anthropogenic combustion sources to atmospheric inorganic reactive nitrogen in South China evidenced by isotopes. *Atmospheric Chemistry and Physics*, 23(11), 6395-6407. <https://doi.org/10.5194/acp-23-6395-2023>
- Liu, S., Lim, Y.-H., Pedersen, M., Jørgensen, J. T., Amini, H., Cole-Hunter, T., Mehta, A. J., So, R., Mortensen, L. H., Westendorp, R. G. J., Loft, S., Bräuner, E. V., Ketzel, M., Hertel, O., Brandt, J., Jensen, S. S., Christensen, J. H., Sigsgaard, T., Geels, C., ... Andersen, Z. J. (2021). Long-term exposure to ambient air pollution and road traffic noise and asthma incidence in adults: The Danish Nurse cohort. *Environment International*, 152, 106464. <https://doi.org/10.1016/j.envint.2021.106464>
- Mariam, A., Haq, Z. U., & Batool, S. A. (2026). Assessment of spatio-temporal dynamics of tropospheric NO₂ and covariates using the wavelet analysis: A Punjab case study. *Environmental Monitoring and Assessment*, 198(8), 831. <https://doi.org/10.1007/s10661-026-15674-7>
- Meng, X., Liu, C., Chen, R., Sera, F., Vicedo-Cabrera, A. M., Milojevic, A., Guo, Y., Tong, S., Coelho, M. D. S. Z. S., Saldiva, P. H. N., Lavigne, E., Correa, P. M., Ortega, N. V., Osorio, S., Garcia, Kysely, J., Urban, A., Orru, H., Maasikmets, M., ... Kan, H. (2021). Short term associations of ambient nitrogen dioxide with daily total, cardiovascular, and respiratory mortality: Multilocation analysis in 398 cities. *BMJ*, n534. <https://doi.org/10.1136/bmj.n534>
- Munawar, S., Miandad, M., Khan, M. A., Bajwa, N., Basharat, M., & Chohan, A. (2026). Impact assessment of land surface temperature on air pollution and smog formation in major cities of Pakistan using Google Earth engine. *Ecological Questions*, 37(1), 1-22. <https://doi.org/10.12775/EQ.2026.012>
- Murtaza, R., Khokhar, M. F., Noreen, A., Atif, S., & Hakeem, K. R. (2018). Multi-sensor temporal assessment of tropospheric nitrogen dioxide column densities over Pakistan. *Environmental Science and Pollution Research*, 25(10), 9647-9660. <https://doi.org/10.1007/s11356-017-1176-7>
- Narjis Naqvi, S. F., Hameed, R., Asghar, B., Busri, A., Husnain, A., Mahmood, M., Riaz, A., & Muhammad Farooq. (2025). Spatiotemporal Analysis of Air Quality Index (AQI) in Pakistan: Implications for Public Health and Policy Interventions. *Indus Journal of Bioscience Research*, 3(10), 141-147. <https://doi.org/10.70749/ijbr.v3i10.2477>

- Naseer, E., Basit, A., Bhatti, M. K., & Siddique, M. A. (2022). Machine Learning for Area-Wide Monitoring of Surface Level Concentration of NO₂ Using Remote Sensing Data. 2022 *International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*, 1–6. <https://doi.org/10.1109/ETECTE55893.2022.10007417>
- Naz, A., Farooq Iqbal, M., Shahzad, M. I., Usman, M., & Hassan, B. (2026). Transboundary smog dynamics: A spatiotemporal assessment of brick kiln emission impacts in Central Punjab, Amritsar, and Ferozepur. *Urban Climate*, 67, 102869. <https://doi.org/10.1016/j.uclim.2026.102869>
- Rabbia Rehmat, Lubna Rafique, Muhammad Irfan, Samreen Riaz Ahmed, Altaf Hussain Lahori, Aqsa Muhammadi, Muhammad Taha, Sergij Vambol, & Mykhailo Shulga. (2023). Space-borne Air Quality Monitoring of Nitrogen dioxide (NO₂) over Karachi and Lahore using Remote Sensing Tools. *Proceedings of the Pakistan Academy of Sciences: B. Life and Environmental Sciences*, 60(3). [https://doi.org/10.53560/PPASB\(60-3\)875](https://doi.org/10.53560/PPASB(60-3)875)
- Saleem, F., Hina, S., Ullah, I., Habib, A., Hina, A., Ilyas, S., & Hamid, M. (2024). Impacts of irregular and strategic lockdown on air quality over Indo-Pak Subcontinent: Pre-to-post COVID-19 analysis. *Chaos, Solitons & Fractals*, 178, 114255. <https://doi.org/10.1016/j.chaos.2023.114255>
- Song, W., Liu, X.-Y., Hu, C.-C., Chen, G.-Y., Liu, X.-J., Walters, W. W., Michalski, G., & Liu, C.-Q. (2021). Important contributions of non-fossil fuel nitrogen oxides emissions. *Nature Communications*, 12(1), 243. <https://doi.org/10.1038/s41467-020-20356-0>
- Tariq, S., Nawaz, H., Mehmood, U., Ul Haq, Z., Pata, U. K., & Murshed, M. (2023). Remote sensing of air pollution due to forest fires and dust storm over Balochistan (Pakistan). *Atmospheric Pollution Research*, 14(2), 101674. <https://doi.org/10.1016/j.apr.2023.101674>
- ul-Haq, Z., Tariq, S., Ali, M., Mahmood, K., Batool, S. A., & Rana, A. D. (2014). A study of tropospheric NO₂ variability over Pakistan using OMI data. *Atmospheric Pollution Research*, 5(4), 709–720. <https://doi.org/10.5094/APR.2014.080>
- Verma, S., Yadava, P. K., Lal, D. M., Mall, R. K., Kumar, H., & Payra, S. (2021). Role of Lightning NO_x in Ozone Formation: A Review. *Pure and Applied Geophysics*, 178(4), 1425–1443. <https://doi.org/10.1007/s00024-021-02710-5>
- Wei, Q., Song, W., Dai, B., Wu, H., Zuo, X., Wang, J., Chen, J., Li, J., Li, S., & Chen, Z. (2025). Spatiotemporal estimation of surface NO₂ concentrations in the Pearl River Delta region based on TROPOMI data and machine learning. *Atmospheric Pollution Research*, 16(3), 102353. <https://doi.org/10.1016/j.apr.2024.102353>
- Xian, J., & Li, G. (2026). Emerging electrocatalytic conversion of NO_x species into value-added nitrogenous chemicals. *Inorganic Chemistry Frontiers*, 13(3), 869–899. <https://doi.org/10.1039/D5QI01917F>
- Zeeshan, M. (2025). Spatiotemporal analysis of tropospheric nitrogen dioxide hotspot over Lahore Division in Pakistan. *Discover Environment*, 3(1). <https://doi.org/10.1007/s44274-025-00323-6>

Zong, Z., Tian, C., Li, J., Syed, J. H., Zhang, W., Fang, Y., Jiang, Y., Nasir, J., Mansha, M., Rizvi, S. H. H., Shafiq, M., Farhan, S. B., & Zhang, G. (2020). Isotopic Interpretation of Particulate Nitrate in the Metropolitan City of Karachi, Pakistan: Insight into the Oceanic Contribution to NO_x . *Environmental Science & Technology*, 54(13), 7787–7797.
<https://doi.org/10.1021/acs.est.0c00490>

