Advances in Engineering Innovation Vol.17 Issue 1 EWA Publishing
Available Online: 26 December 2025 DOI: 10.54254/2977-3903/2025.30894

Long-term spatiotemporal analysis of urban nitrogen oxides in Manchester (2015–2025): statistical and machine learning approaches

Shifei Wang

School of Natural Sciences, The University of Manchester, Manchester, United Kingdom

christy.wangsf@outlook.com

Abstract. Urban Nitrogen Oxide (NO_x) air pollution poses significant public health risks, and therefore spatiotemporal knowledge of NO_x is crucial for air quality regulation. However, few studies have examined long-term NO_x dynamics in Manchester by integrating spatial, temporal, and mechanistic perspectives. This study investigates the long-term trends and drivers of NO and NO_2 in the urban atmosphere of Manchester from 2015 to 2025. Data from five AURN monitoring sites, ERA5 meteorological reanalysis datasets, and UK Department for Transport traffic statistics were analysed using linear regression for trend estimation, seasonal decomposition, and spatial pattern analysis. A hybrid statistical–machine learning framework was additionally employed, combining Ordinary Least Squares (OLS) regression with XGBoost models interpreted through Shapley Additive Explanations (SHAP). The results indicate statistically significant declining trends in both pollutants, with average annual decreases of approximately 4.8%, and dramatic short-term reductions during COVID-19 lockdowns, highlighting the dominant influence of traffic. Seasonal patterns persisted, with winter concentrations 1.9 times greater than summer levels, and spatial analysis revealed strong NO_2 heterogeneity among monitoring sites. Machine learning models performed substantially better than linear regression (R^2 = 0.475 vs. 0.29), and SHAP analysis showed ozone, boundary layer height, and temperature as the main drivers of NO_2 variations. Overall, the findings confirm substantial air-quality improvements while revealing nonlinear processes in urban pollution dynamics, supporting continued emission-reduction policies and enhanced monitoring strategies.

Keywords: Nitrogen Oxides (NO_x), Urban air pollution, Spatiotemporal analysis, Machine learning (XGBoost), SHAP interpretation

1. Introduction

Urban air pollution is among the most critical environmental health concerns of the 21st century, and Nitrogen Oxides (NO_x) are among the major pollutants that require prompt attention. Nitrogen Dioxide (NO_2) is identified by the World Health Organization as responsible for respiratory inflammation [1], cardiovascular disease, and enhanced mortality. Moreover, NO_x ($NO + NO_2$) in cities primarily originates from vehicles (Figure 1). Although Nitric Oxide (NO) is released directly during combustion, nitrogen dioxide (NO_2) arises through complex atmospheric transformations. The significance of NOx pollution reaches beyond health impacts to global environmental and policy relevance. European Union air quality legislation requires annual mean NO_2 levels to be below 40 μg/m3 [2], but exceedances persist in many cities, including Greater Manchester [3].

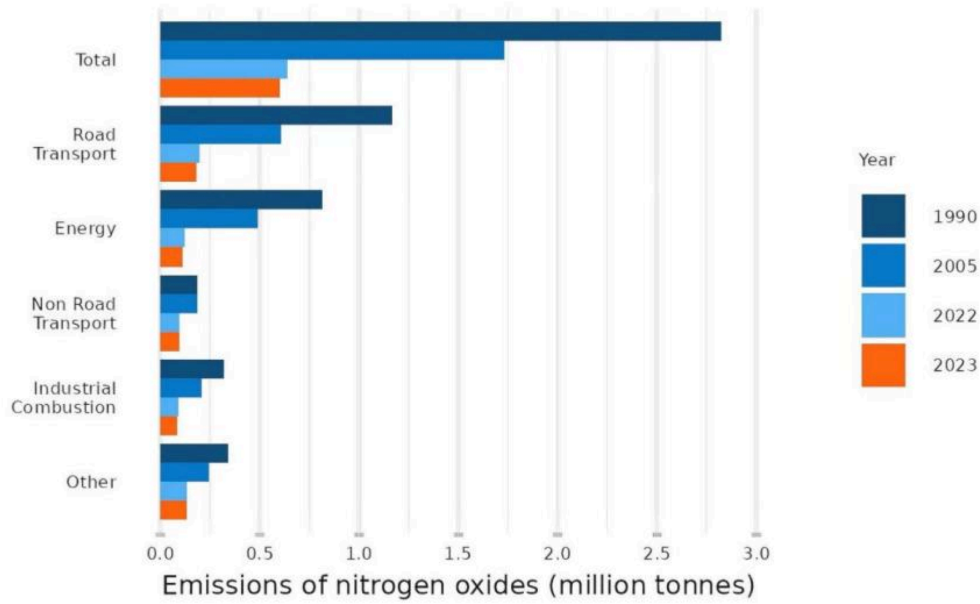


Figure 1. Annual emissions of nitrogen oxides (NO_x) from major sources in the UK (1990, 2005, 2022, and 2023)

The existing NO_x literature is characterised by three elementary limitations that this study aims to address. Firstly, temporal coverage of studies is inclined to be limited to brief periods of observation (typically 1-3 years), making it difficult to identify long-term trends and assess policy impacts. Secondly, spatial coverage is often limited to single monitoring points or small areas, preventing an understanding of urban-scale pollution patterns and their heterogeneity among microenvironments. Thirdly, methodological limitations in current approaches restrict our understanding of NO_x dynamics. Traditional statistical methods, while providing interpretable outputs, impose linearity on relationships between pollutants and drivers. Conversely, although machine learning methods can capture nonlinear trends, they are often applied without proper interpretability, resulting in "black box" models that provide very few clues regarding mechanisms [4]. Together, these restrictions are particularly demanding in Manchester, a larger city characterised by high traffic, extensive diesel use, and complex urban structure, providing a clear rationale for selecting it as the case study for determining long-term NOx trends and their driving factors.

This study addresses these gaps through a comprehensive analysis of NO_x spatiotemporal variability in Manchester utilising an integrative methodological framework. The research focuses on the connection between spatial-temporal pollution patterns and their driving forces: meteorological dispersion and transformation processes and traffic activity controlling emission rate. By blending traditional statistical models with machine learning models, the research provides interpretable linear connections and captures nonlinear interactions. Covering 2015 to 2025, the analysis examines long-term trends and short-term volatility using multiple monitoring sites across Greater Manchester to identify hotspots and spatial heterogeneity in driving forces. This framework combines statistical and interpretable machine learning to quantify the spatiotemporal dynamics and driving mechanisms of NO and NO_2.

This study aims to explore the spatiotemporal characteristics of nitrogen oxides (NO and NO_2) in the Manchester city environment and establish the relative contribution of meteorological and traffic origin driving factors towards variability in pollution. The study addresses three connected research questions outlined below:

• RQ1: What are the long-term trends and seasonal patterns in NO_x concentrations across Manchester from 2015 to 2025?

• Objective: Quantify temporal trends using time series analysis to assess whether pollution levels have improved over the study period and identify seasonal variation patterns that reflect meteorological influences on pollution dynamics.

• RQ2: How do spatial differences in NO_x concentrations relate to local emission sources and urban characteristics?

• Objective: Compare pollution levels at monitoring points in different areas and analyse spatial patterns using emission inventory data to identify pollution hotspots and their underlying drivers.

• RQ3: What are the relative contributions of meteorological conditions and traffic activity to short-term NO_x variability, and how do these relationships differ between NO and NO_2?

• Objective: Apply integrated statistical and machine learning approaches to quantify linear and nonlinear relationships between pollution levels and driving factors, with particular emphasis on understanding differences between primary (NO) and secondary (NO_2) pollutant responses.

2. Results

2.1. Analytical framework and research design

This study employs a multi-scale analytical framework designed to address the three research questions through progressively more detailed studies, ranging from descriptive pattern recognition to mechanistic process understanding. Temporal trend analysis, spatial pattern description, and mechanistic modelling is integrated to provide insight into of NO_x behaviour in Manchester. The analytical approach proceeds from descriptive analysis of long-term temporal trends (RQ1) to testing for spatial heterogeneity (RQ2) to quantitative modelling of driving processes (RQ3). The study period is 2015-2025, spanning routine pollution cycles and extraordinary events. Spatial analysis incorporates six monitoring stations across Greater Manchester, selected to represent a variety of urban environments and examine spatial heterogeneity in pollution patterns and driving factor relationships.

2.2. Data sources and pre-processing

The data were obtained from monthly and annual NO and NO_2 concentrations from the UK AURN network, meteorological variables from the ERA5 reanalysis dataset, and Annual Average Daily Traffic (AADT) statistics from the UK Department for Transport [5]. Pre-processing included format standardisation, time series alignment, and missing value handling.

Air Quality Monitoring Data: Monthly average and annual average concentrations of NO and NO_2 were downloaded from AURN for all available Manchester monitoring stations for the period 2015-2025. For mechanistic modelling (RQ3), pollutant measurements were aligned to hourly meteorological variables, and records below 75% completeness were discarded.

Meteorological Variables: ERA5 data included boundary layer height (BLH), 2-meter temperature (t2m), dew point temperature (d2m), wind speed and direction (derived from u10 and v10 components) and surface shortwave downward radiation (ssrd). Hourly ozone (O_3) was also obtained from the AURN, and oxidant ($O_x = O_3 + NO_2$) was used to represent atmospheric oxidative capacity.

Traffic Activity Data: DfT AADT statistics provide vehicle counts by class for major roads throughout the study region for 2015-2025.

Emission Inventory Integration: NAEI gridded NO_x emission data (1km × 1km) for 2015 and 2022 were used to characterize spatial patterns of emissions.

2.3. Temporal trend analysis methodology

To investigate interannual trends in NO and NO_2 between 2015 and 2025 (RQ1), linear regression was applied to annual mean concentrations, with $p < 0.05$ indicating statistical significance. The coefficient of determination (R^2) quantified the proportion of variability explained, and regression slopes provided rates of change in μg/m^3/year. Visualisation emphasized changes throughout the study period and average annual rates of change.

2.4. Seasonal and spatial pattern analysis

Seasonal analysis focused on NO_2 due to its greater regulatory relevance and more stable concentration profiles compared to variable NO time series. Monthly mean NO_2 concentrations were merged from all monitoring sites to calculate city-wide seasonal patterns. Seasonal patterns were depicted using monthly average line plots and heatmaps of concentration distributions by month and year.

Spatial analysis compared annual mean NO_2 concentrations at five AURN monitoring sites typical of different urban environments, plotting linear trend lines for each site to assess whether pollution reductions are spatially consistent. Spatial trends were supplemented by emission inventory analysis using National Atmospheric Emissions Inventory (NAEI) gridded data for 2015 and 2022, with difference maps showing areas of increasing or decreasing emissions over time.

2.5. Mechanistic modelling framework

2.5.1. Time series decomposition

Seasonal-Trend decomposition using Loess (STL) was applied to hourly NO and NO_2 time series to decompose long-term trends, seasonal cycles, and short-term residual variability. The residual component preserves the impact of meteorological and traffic effects on pollution concentrations after removing seasonal and interannual patterns.

2.5.2. Statistical modelling approach

To assess the impact of meteorological and traffic factors on short-term pollutant fluctuations, a Multiple Linear Regression (MLR) model was constructed using the residual series from the STL decomposition as the dependent variable. Variable selection was based on correlation analysis results and differentiated by pollutant generation mechanism: NO_2 was used the total number of motor vehicles, while NO used separate small-vehicles and heavy trucks classes. All independent variables were z-score standardised, and model performance was assessed using R^2 values and residual diagnostics.

2.5.3. Machine learning implementation

Extreme Gradient Boosting (XGBoost) regression was employed to capture nonlinear relationships and interaction effects not accessible to linear modelling approaches. Hyperparameter optimisation employed GridSearchCV with 5-fold cross-validation. Model interpretation utilized PDPs to reveal functional relationships between predictors and model outputs, and SHAP to quantify feature importance and contribution directions. Performance comparison between linear and nonlinear models quantifies the importance of nonlinear processes in determining NO_2 variability.

3. Results

3.1. Temporal trend analysis

Figure 2 illustrate the annual average concentration trends of NO and NO_2 in central Manchester from 2015 to 2025. The results indicate a clear declining trend for both pollutants, with the linear regression models showing a high coefficient of determination (R^2 = 0.882) and strong statistical significance (p = 0.000). The average NO_2 concentration decreased from approximately 29.5 μg/m^3 to 18.1 μg/m^3, representing a total reduction of 11.4 μg/m^3 and an average annual decrease of around −4.8%. Similarly, NO concentrations also declined by 11.4 μg/m^3, with an average annual decrease consistent with that of NO_2. Both pollutants dropped sharply in 2020 and rose slightly afterward, but the overall trend remains downward.

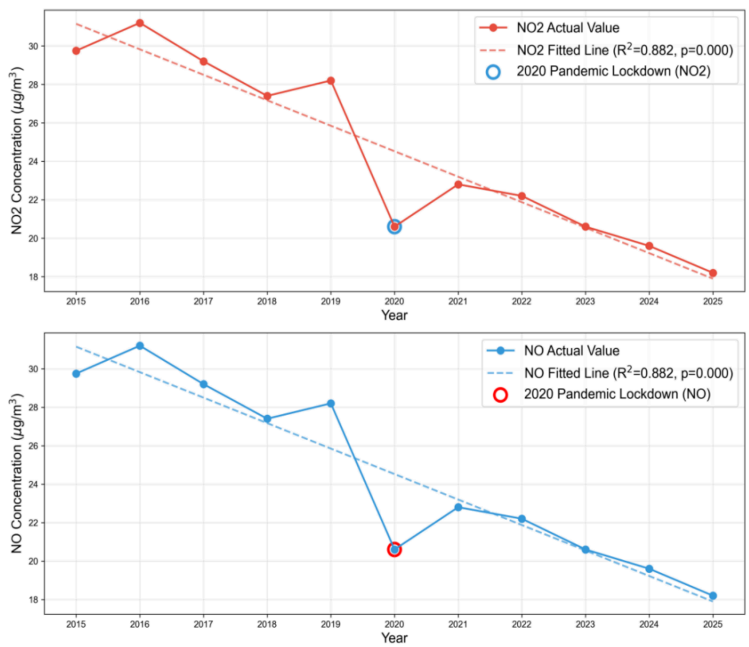


Figure 2. Annual concentration trend of NO_2 (top) & NO (bottom) (2015-2025)

3.2. Seasonal and spatial analysis

3.2.1. Seasonal analysis

Figure 3 and 4 present the monthly average variation curve of NO_2 concentrations in Manchester and its annual-month distribution heatmap. Overall, NO_2 exhibits a typical seasonal pattern characterised by higher concentrations in winter and lower in summer. Figure 3 shows that January has the highest average concentration (33.7 μg/m^3), while July records the lowest (17.5 μg/m^3), forming a clear "U-shaped" curve over the year. Figure 4's heatmap confirms the persistence of this cycle across both high-pollution years (2015, 2019) and low-pollution years (2023, 2024). During the 2020 COVID-19 lockdown, January averaged 30.3 μg/m^3 compared with 14.7 μg/m^3 in July, indicating that the seasonal contrast was still present.

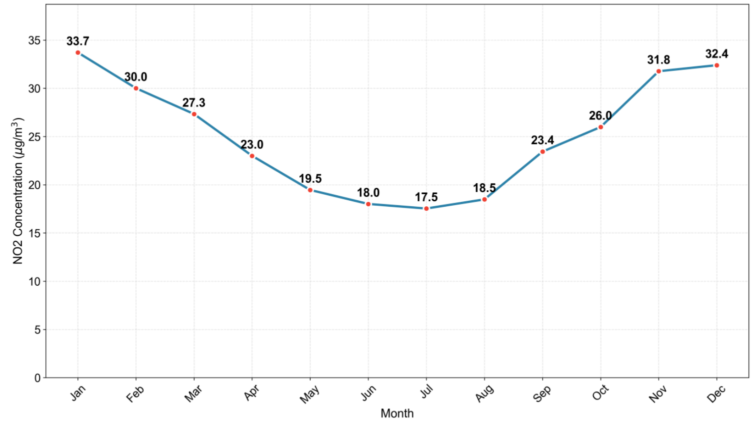


Figure 3. Monthly average NO_2 concentrations in Manchester (2015-2025)

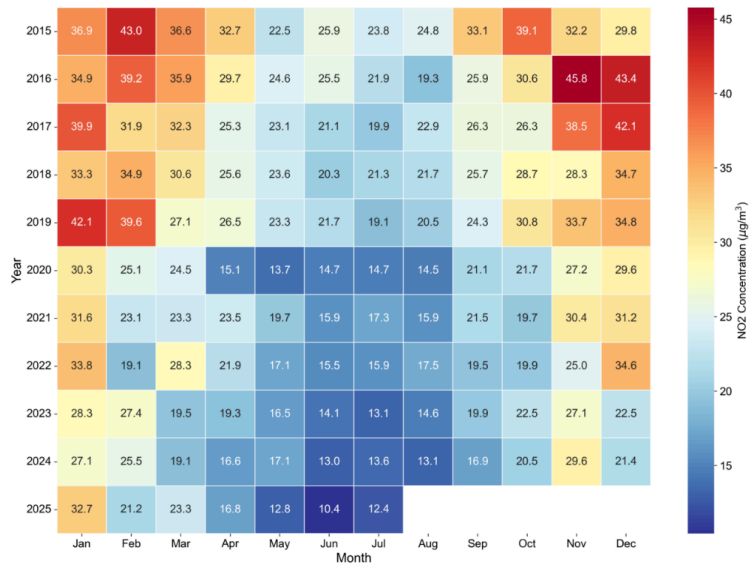


Figure 4. Heatmap of NO_2 concentration by month and year in Manchester (2015-2025)

3.2.2. Spatial analysis

Figure 5 illustrates the annual mean NO_2 concentrations at five AURN monitoring sites across Manchester and surrounding areas from 2015 to 2025. Overall, all sites show a declining trend. The Manchester Piccadilly station recorded the highest concentrations, with levels falling from 38.9 μg/m^3 in 2015 to approximately 25.0 μg/m^3 in 2025. The Shaw Crompton Way and Salford Eccles stations followed suit, with initial concentrations ranging from 28 to 34 μg/m^3. In contrast, the Manchester Sharston station recorded the lowest NO_2 levels, decreasing from 23.7 μg/m^3 in 2015 to approximately 15 μg/m^3 in 2025. During the 2020 lockdown period (highlighted by red circles in the figure), all stations showed a significant reduction in NO_2 levels, followed by a slight rebound after 2021.

To further explore spatial differences in NO_2 distribution, Figure 6 and 7 depict NO_2 emission grid maps published by the NAEI for the years 2015 and 2022. The emission hotspots are concentrated in central Manchester, Salford, and the Old Trafford area, which corresponds with the high-concentration sites observed in the AURN network. Compared to 2015, emission intensity in 2022 has generally diminished.

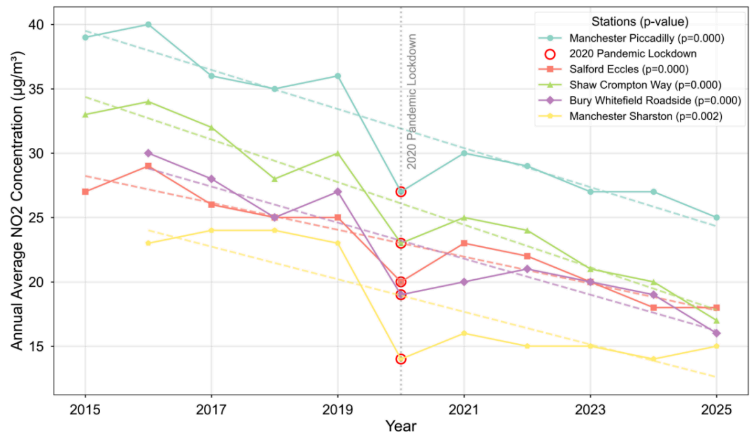


Figure 5. Comparison of annual average NO_2 concentration trends at 5 AURN stations (2015-2025)

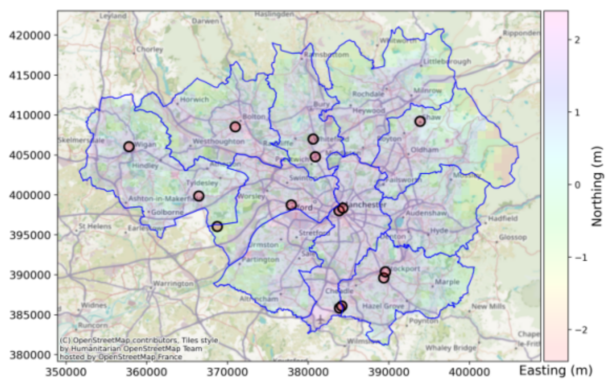


Figure 6. NO_2 emissions from NAEI in 2015

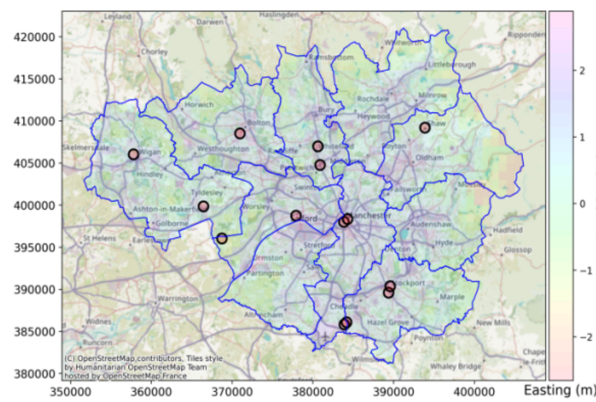


Figure 7. NO_2 emissions from NAEI in 2022

3.3. Pearson correlation and ordinary least squares regression

3.3.1. Pearson correlation

Figure 8 illustrates the correlation between NO and NO_2 concentrations and a range of meteorological variables. According to the heatmap results, both NO and NO_2 exhibit weak negative associations with temperature (t2m) and Boundary Layer Height (BLH), and moderate negative relationships with ozone (O_3). Wind speed (ws) and wind direction (wd) both showed slight negative correlations with NO and NO_2 concentrations.

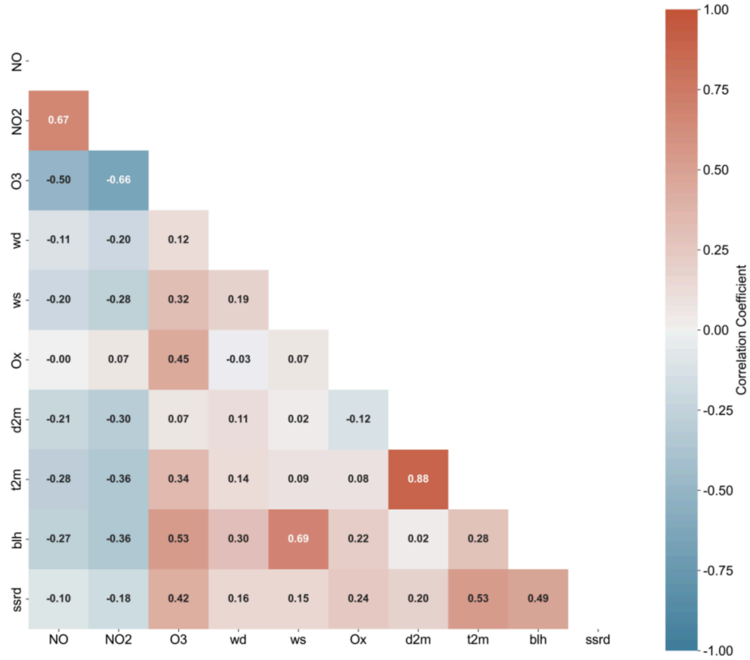
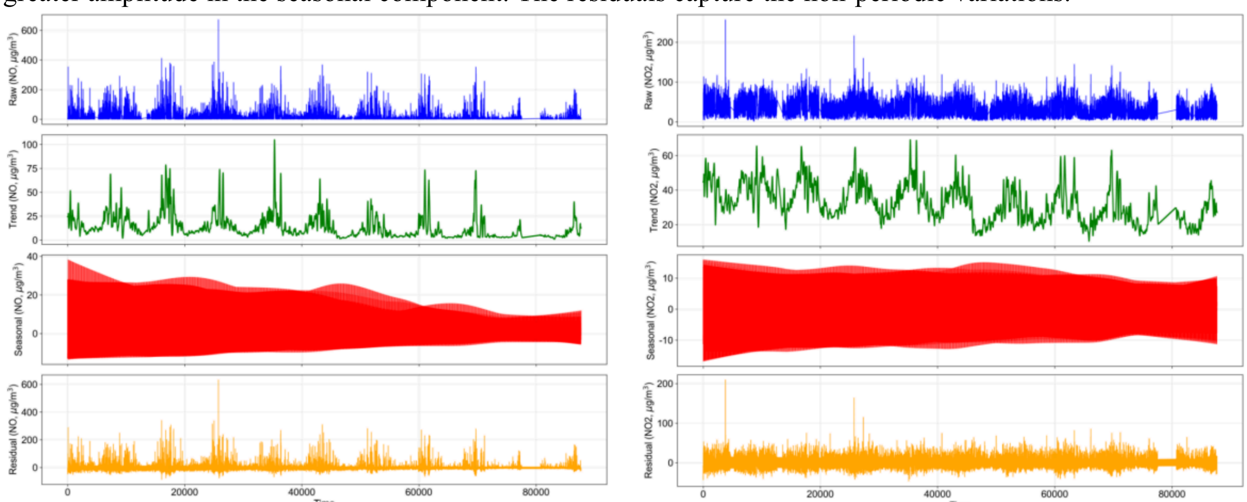


Figure 8. Correlation matrix of pollutants and meteorological variables

3.3.2. Extracting residuals using STL decomposition

Figure 9 illustrates the STL decomposition results for NO and NO_2 concentrations, including the original series (Raw), trend component (Trend), seasonal component (Seasonal), and residual component (Residual). It can be observed that both NO and NO_2 exhibit distinct seasonal variations, while the trend components reflect long-term changes over time. Compared to NO_2, the original NO series shows more marked fluctuations and greater amplitude in the seasonal component. The residuals capture the non-periodic variations.

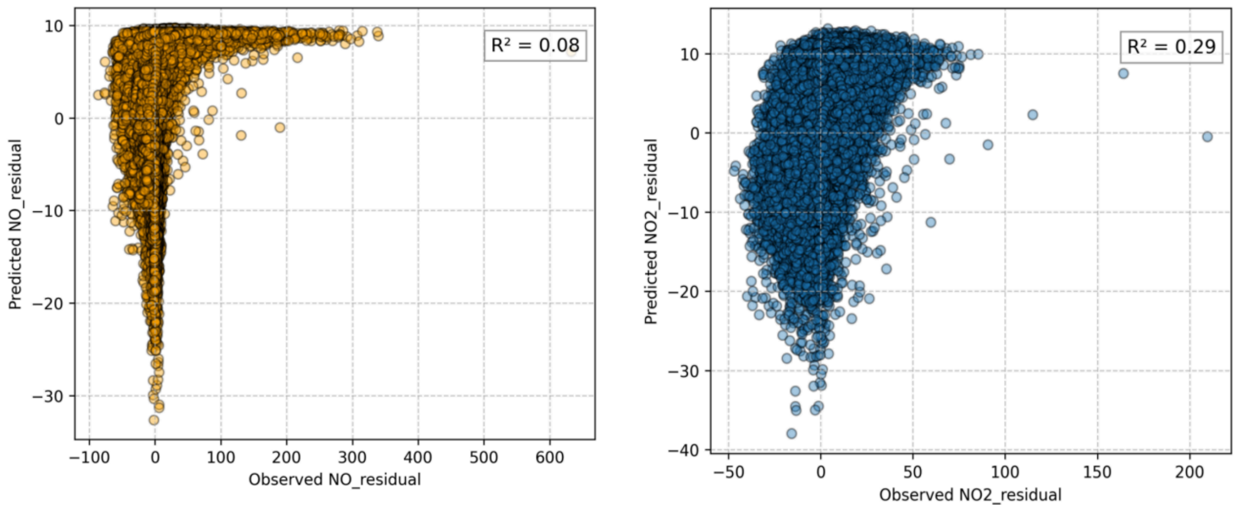


(a) (b)

Figure 9. Monthly STL decomposition of NO (a) & NO_2 (b)

3.3.3. Ordinary least squares regression modelling

Figure 10 depicts the fitting performance of the Ordinary Least Squares (OLS) regression models for the residuals of NO and NO_2. The results show that the explanatory power of the NO model is relatively low, with a coefficient of determination (R^2) of 0.08, indicating that the selected meteorological variables account for only 8% of the residual variability. In contrast, the NO_2 model display better performance than NO model, with an R^2 of 0.29.



(a) (b)

Figure 10. Observed vs. predicted residuals from OLS models for NO & NO_2

In terms of standardized regression coefficients (Table 1), ozone (O_3) exhibits a significant negative effect in both models, with the strongest impact observed in the NO_2 model (−5.6613). Boundary Layer Height (BLH) also shows a negative influence on both pollutants. Temperature (t2m) has a minor positive influence on NO_2 (1.0299) but a minimal effect on NO (0.1379). For traffic-related variables (−0.7433), while the NO model identifies negative impacts from both light vehicles (cars and taxis, -1.1514) and heavy goods vehicles (HGVs, -0.7218).

Table 1. Standardized regression coefficients and R^2 scores of OLS models for NO and NO_2 residuals

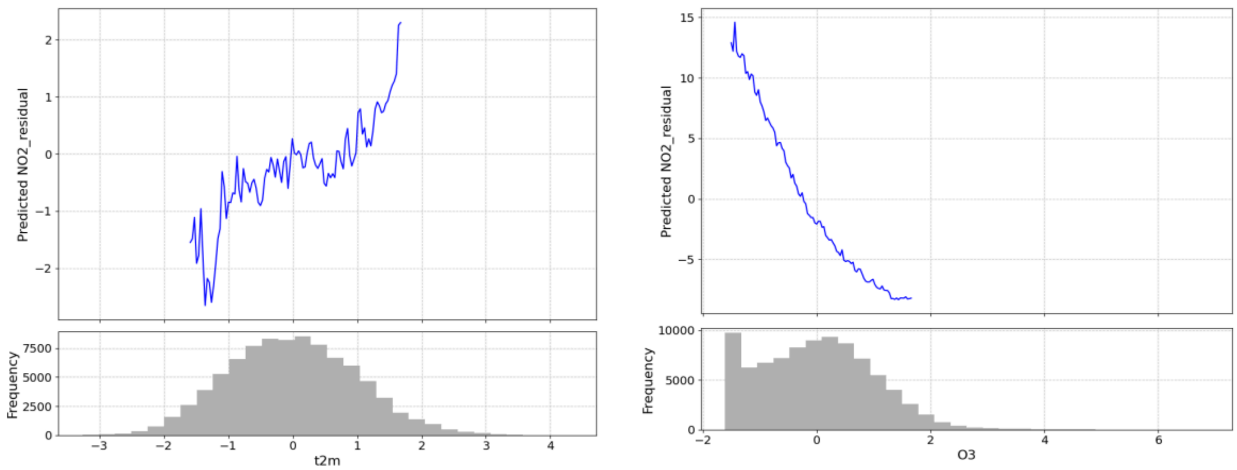
	NO_2_residual Model	NO_residual Model
R^2 score	0.2892	0.0794
BLH	-1.9114	-1.8258
t2m	1.0299	0.1379
O_3	-5.6613	-4.5128
all_motor_vehicles	-0.7433	\
cars_and_taxis	\	-1.1514
all_hgvs	\	-0.7218

3.4. Machine learning models

3.4.1. Variables on NO2 residuals for PDP analysis

This section develops an XGBoost regression model to predict the residual component of NO_2 and employed Partial Dependence Plots (PDPs) to interpret the average effect of individual variables on the model's predictions. The results demonstrate that XGBoost performs well in modelling NO_2 residuals (see Figure 11–12).

Temperature (t2m) exhibits a nonlinear relationship with NO_2 residuals (Figure 11). When the temperature is below 0 °C, NO_2 residuals tend to be negative. However, as temperatures rise, residuals become positive. Ozone (O_3) shows a strong negative correlation with NO_2 residuals (Figure 11). The PDP indicates that higher O_3 concentrations correspond to lower predicted NO_2 residuals.

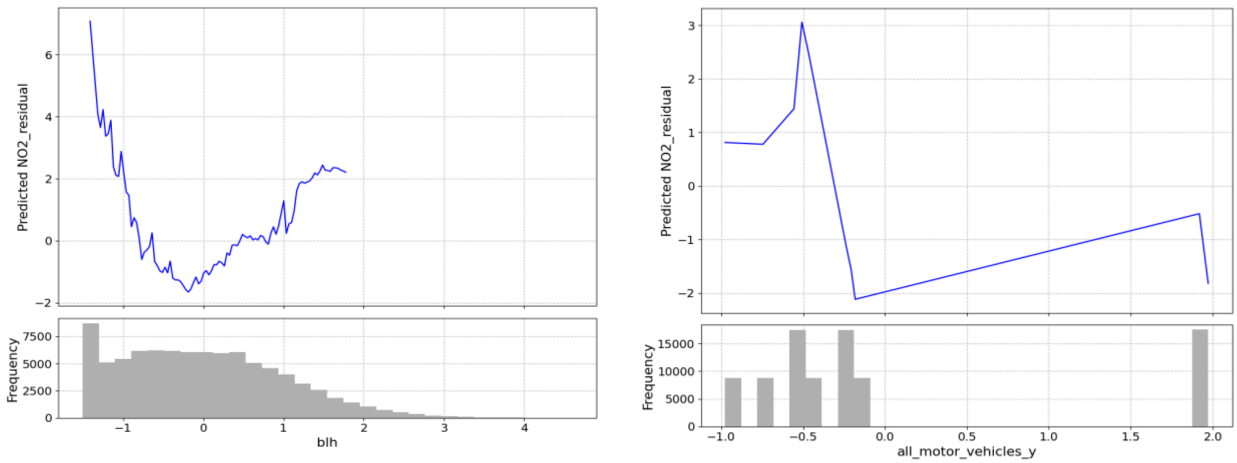


(a) (b)

Figure 11. Partial dependence of t2m (a) & O_3 (b) on NO_2_residual

Boundary Layer Height (BLH) displays a U-shaped nonlinear relationship (Figure 12). Residuals are high under both very low and very high BLH conditions. In contrast, at moderate BLH levels, residuals drop.

The effect of total motor vehicle flow (all_motor_vehicles_y) appears more complex (Figure 12). While the PDP exhibits fluctuations an overall negative trend is observed.



(a) (b)

Figure 12. Partial dependence of BLH (a) & all_motor_vehicles (b) on NO2_residual

3.4.2. SHAP value analysis

To further interpret the prediction mechanism of the XGBoost model for NO_2 residuals, this section utilises the SHAP method to evaluate the magnitude and direction of each feature's contribution to the model output.

Figure 13 display the SHAP summary plot for the training set and the results for the validation set. Overall, ozone (O_3) exhibits the highest absolute SHAP values. Higher concentrations of O_3 are associated with negative SHAP values. Boundary Layer Height (BLH) also shows a moderate negative impact in the lower to mid-range values. Although temperature (t2m) contributes less than O_3 and BLH, it still provides some explanatory power, with higher temperatures generally associated with a positive effect on predictions.

The SHAP values for traffic indicators (e.g., all_motor_vehicles_y) reveals more scattered. A comparison between the training and validation sets (Figure 13) shows consistent directional effects among the major variables.

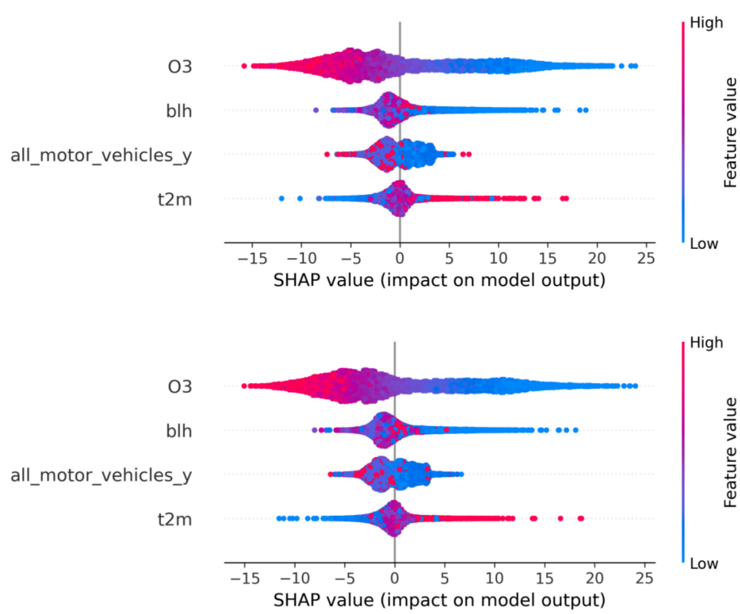


Figure 13. SHAP value summary for training (top) & validation (bottom)

3.4.3. Comparison between xgboost and traditional regression

This section compares the performance of the traditional linear regression model (OLS) and the machine learning model (XGBoost) in predicting NO_2 residuals. The results show that the OLS model yields a coefficient of determination (R^2) of 0.2892, whereas the XGBoost model achieves a higher R^2 of 0.475 on the test set.

As demonstrated by the daily variation boxplot (Figure 14), the XGBoost model captures the diurnal cycle of NO_2 concentrations. The predicted values closely match the observed data in terms of median and distribution, particularly during night-time and off-peak hours. During peak traffic periods, the predictions tend to slightly underestimate the observed values.

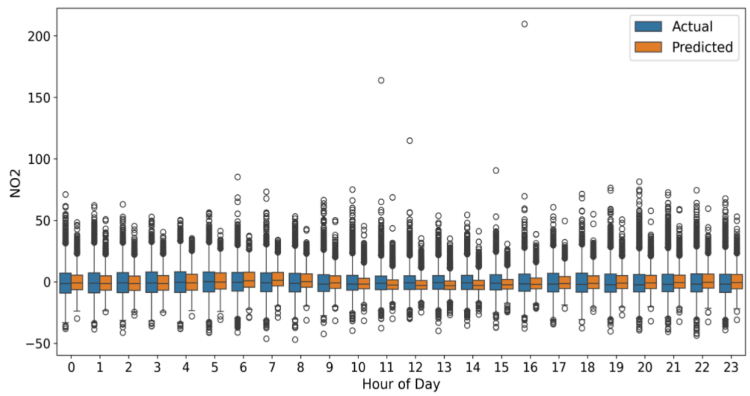


Figure 14. Diurnal variation of observed and predicted NO_2 concentrations

4. Conclusion

This study systematically analysed the changes in nitrogen oxides in the air of Manchester between 2015 and 2025. The results show that air quality has significantly improved during this period and reveal how meteorological conditions, traffic, and policy factors jointly influence urban pollution levels through long-term trend analysis, spatial distribution characteristics, and mechanistic modelling.

4.1. Summary of key findings

Monitoring results show that NO and NO_2 concentrations exhibited a linear downward trend, with an average annual decline of about 4.8% ($p < 0.001$) and a total reduction of more than 11μg/m^3. Consistent downward trends across multiple monitoring sites indicate that the improvement resulted from systematic emission reductions rather than short-term or local factors. The COVID-19 lockdown caused a sharp temporary drop in pollution, followed by a return to the long-term downward trajectory, showing that sustained progress depends on structural transformation of the transport system rather than temporary restrictions.

Seasonal analysis shows that winter concentrations were consistently twice as high as those in summer across all study years, highlighting the critical influence of seasonal meteorological processes on pollutant dispersion and photochemical transformation. Spatial analysis shows that roadside pollution levels were nearly double those in suburban areas, revealing inequities in exposure and supporting the establishment of denser monitoring networks and localised mitigation strategies.

Mechanistic modelling reveals non-linear pollutant–driver relationships. XGBoost improves explanatory power compared to linear regression (R^2 = 0.475 vs 0.29), and SHAP results indicate ozone as the main driver of NO_2, followed by boundary layer height and temperature. The combined statistical and explainable machine-learning framework offers predictive capability while retaining interpretability and is applicable to other cities facing similar air quality issues.

4.2. Policy implications and future research

Observation-based monitoring and modelling systems that integrate real-time observations with forecasting expertise hold considerable potential for improving urban air quality management by enabling early warning and prediction. The findings carry policy implications from urban planning to national regulation, including stricter vehicle emission standards and the promotion of clean-fuel vehicles. Moreover, the spatial variations identified in this study highlight the need for locally tailored interventions, while the seasonal patterns point to the importance of adopting time-sensitive and differentiated measures.

Future research should focus on three key directions: (i) acquiring higher temporal resolution traffic data to better capture emission–concentration relationships and improve short-term forecasting; (ii) incorporating additional meteorological variables, particularly indicators of atmospheric stability and mixing; and (iii) extending the scope of research to other pollutants, such as secondary aerosols and ozone.

References

[1] World Health Organization. (2021). *WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. https: //www.who.int/publications/i/item/9789240034228

[2] European Parliament and Council. (2008, May 21). Directive 2008/50/EC on ambient air quality and cleaner air for Europe. *Official Journal of the European Union, L 152*, 1–44. https: //eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX: 32008L0050

[3] Greater Manchester Combined Authority. (2025). *Greater Manchester five-year environment plan 2025–2030*. https: //www.greatermanchester-ca.gov.uk/media/alnl0fsy/gmca_5-year-plan_final_digital_v3-ua.pdf

[4] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https: //doi.org/10.1038/s42256-019-0048-x

[5] Department for Transport. (2025). *Road traffic statistics: Manchester*. https: //roadtraffic.dft.gov.uk/local-authorities/85