Estimating Nitrogen Dioxide Levels Using Open Data and Machine Learning: A Comparative Modeling Study

Dara Varam[1], Rohan Mitra[1], Furzeen Kamran[2], Diaa Addeen Abuhani[1], Hana Sulieman[2], Imran Zualkernan[1, †]

[1] Dept. of Computer Science & Engineering, American University of Sharjah, Sharjah, UAE
[2] Dept. of Mathematics & Statistics, American University of Sharjah, Sharjah, UAE
[†] Corresponding author - izualkernan@aus.edu

Keywords: NO_2 Prediction, Machine Learning, Sustainability, Open NDVI Data, Temporal Analysis, Spatial Analysis

Abstract

Nitrogen dioxide (NO_2) is a critical pollutant with widespread effects on both air quality and environmental health, recognized as a key concern within the United Nations' Sustainable Development Goals (SDGs). This study investigates NO_2 levels in Italy, analyzing spatial and seasonal variations to better understand pollutant distribution. Using open-source data, we employed machine learning models to estimate NO_2 concentrations, achieving strong predictive accuracy based on the mean absolute percentage error and the root mean-squared error. The results reveal that model performance improves significantly when data is segmented based on seasonal and urban development factors. Specifically, predictions for urban, rural, and mixed cities demonstrated that urban areas exhibited higher NO_2 concentrations, while rural regions showed comparatively lower levels. The analysis underscores the importance of tailoring models to regional and temporal contexts, affirming that open-source data, combined with machine learning techniques, can effectively estimate NO_2 pollution levels across diverse environments.

1. Introduction

Nitrogen dioxide (NO_2) is identified as one of the most lethal pollutants in the atmosphere that has adverse effects on air quality and the ecological environment. The major local sources of nitrogen oxides are emissions from vehicles, fossil fuel power plants, industries and factories, household cooking activities, and natural soil processes (Wei et al., 2022). In addition, transfer of pollutants from other regions contributes to increased ground-level NO_2 contributions (Chi et al., 2022; Fan et al., 2021). One can conclude that human activities are probably the major contributors to increased levels of NO_2 (Cedeno Jimenez et al., 2023). This is further supported by the works of Piccoli and colleagues, which discloses how the influence of mass lockdowns during the COVID-19 Pandemic on significant decrease in NO_2 concentration across various countries, including Italy (Piccoli et al., 2020). Being a major component of chemical processes within the stratosphere and troposphere, NO_2 is affected by natural and man-induced emissions, thereby displaying clear spatial and temporal variations (Chi et al., 2022; Fan et al., 2021; Van Geffen et al., 2015). In particular, high concentrations of NO_2 gather around emission sources allowing large deposits of NO_2 at the bottom of the troposphere (Chi et al., 2022), posing instant adverse threats on the urban population. Recent studies highlight the significant health risks, including but not limited to, premature mortality (Ghahremanloo et al., 2021) and increased risk of various cardiovascular and respiratory diseases (Weinmayr et al., 2010; Zhu et al., 2019) such as the Chronic Obstructive Pulmonary Disease (COPD), which is the fourth leading cause of death globally and can be associated with NO_2 (Zhang et al., 2018). This advocates for rigorous monitoring of NO_2 to negate the adverse effects of NO_2 on the environment and population's health.

There are several initiatives that have been employed to address the hazardous challenge posed by increased NO_2 levels. The United Nations (UN) has introduced Sustainable Development Goals (SDGs) that act as a guide for government entities, organizations, and individuals, offering a united worldwide plan to mitigate numerous issues ranging from poverty, health, climate change, and inequality (Carlsen and Bruggemann, 2022). In particular, air quality is specifically mentioned in two of the 17 SDGs (Carlsen and Bruggemann, 2022). Moreover, the World Health Organization (WHO) has developed Air Quality Guidelines (AQGs) to mitigate the negative effects of NO_2 on human health (Song et al., 2023). Nevertheless, it has been reported that 90% of the world's population resides in areas that are not in compliance with UN established air quality measures (Trushna and Tiwari, 2022), and around 549,715 deaths worldwide have been caused by high NO_2 concentrations which could be prevented if stringent measures had been taken to comply with the AQGs (Song et al., 2023). Recently, Italy has introduced its 2030 sustainability vision prioritizing the alteration of the fundamental biogeochemical cycles including carbon with the aim to decarbonize the country as much as possible.

Current NO_2 monitoring methods rely heavily on ground-based stations, which provide high-accuracy measurements but suffer from sparse spatial coverage and high operational costs. Although satellite remote sensing data offer greater coverage, they are limited by temporal resolution, cloud interference, and lower accuracy near the surface. These limitations highlight the need for a machine learning approach that can integrate multiple data sources to improve spatial and temporal resolution while maintaining accuracy.

In the light of the aforementioned climate risks related to NO_2 levels in the atmosphere, we present an analytical study on the measured NO_2 levels in the region of Italy with respect to spatial and temporal variations. Our contributions are summarized as follows:

1. We study several machine learning based approaches to estimate NO_2 levels based on open data sources.
2. We further analyze the variations of NO_2 with respect to temporal and spatial variations within the region.

2. Related Work

Due to the alarming risks associated with increased NO_2 levels, governments, like the United States Environmental Protection Agency (US EPA) and the European Environment Agency (EEA), have implemented several measures to monitor and control NO_2 concentrations effectively (Cedeno Jimenez and Brovelli, 2023). In general, standard guidelines are proposed by governments to ensure uniformity and accuracy in measurements of NO_2. In their article, Jiminez and colleagues explained how the aforementioned entities use direct analyzers which sample air from decided locations to measure concentrations of nitrogen monoxide and nitrogen dioxide (Cedeno Jimenez and Brovelli, 2023). Due to the large capital investment needed to maintain ground monitoring stations (Pinder et al., 2019), less economically developed countries (LEDCs) are hindered from leveraging these technologies (Cedeno Jimenez and Brovelli, 2023). Moreover, because of the short atmospheric lifespan of NO_2, accurately measuring surface NO_2 concentrations across a wide area is difficult with limited data from ground-based monitoring stations. Despite the high precision and accuracy of ground-stations measuring NO_2, these cannot be placed in urban environments where the spatial distribution of pollutants changes most drastically, calling for the development of accurate strategies to estimate surface NO_2 levels at high spatiotemporal resolutions (Ghahremanloo et al., 2021). Therefore, in lieu of data from ground stations, multiple studies utilize data from satellite instruments, such as the global ozone monitoring experiment (GOME) instrument, Ozone Measurement Instrument (OMI), and the Tropospheric Measurement Instrument for the Sentinel-5P satellite, for estimations of surface NO_2 levels (Cedeno Jimenez et al., 2023; Boersma et al., 2011; Wang and Wang, 2020). For instance, the Sentinel-5P satellite from ESA can measure NO_2 concentrations at a tropospheric level (Cedeno Jimenez and Brovelli, 2023). Clearly, such instruments fail to accurately determine the NO_2 concentration at ground level, which is the ultimate point of interest (Oxoli et al., 2020). Even though there have been attempts to convert satellite tropospheric NO_2 into ground-level NO_2 concentrations using statistical models and chemical transport, the output compromises on the accuracy of NO_2 concentrations (Wei et al., 2022; Lamsal et al., 2008). Moreover, leveraging machine learning algorithms (He et al., 2022; Jiang and Christakos, 2018; Huang et al., 2022; Long et al., 2022), in lieu of regression models, has enabled more reliable estimations of pollutant concentrations (Ghahremanloo et al., 2021). For instance, the work of Zheng and colleagues employed Deep Convolutional Neural Networks (CNN) along with Random Forests for estimating PM2.5 levels in China (Zheng et al., 2020). Despite the exceptional capabilities of machine learning algorithms, a significant limitation is the challenge of interpreting the output (García and Aznarte, 2020), which has limited their applicability.

3. Methodology

3.1 Dataset

The dataset contains information from the Lombardy administrative region in Italy and parts of Veneto (See Fig. 1). Data is collected only from the plains regions (Pianura Padana), and not the mountainous Alpine regions north of the area of study.

The dataset contains both ground-truth and remote sensing data to predict NO_2 (measured in $\mu g/m^3$) levels in Lombardy, Italy. The data has been collected through air quality monitoring stations scattered around the region. In total, there are 79 unique monitoring stations which have each collected data from the beginning of 2019 until the end of 2021. The locations of these monitoring stations can be seen in Fig. 1. On the other hand, the remote sensing data has been collected from open source satellite sources such as Sentinel-2 and Landsat. A full description of the dataset can be found in Table 1.

3.1.1 Pre-processing Some basic pre-processing is applied to the dataset in preparation of training for models. In particular, the dataset is first modified such that all rows containing null values are removed. It is then split into 90% for training and 10% for testing, which is reflected throughout all the experiments conducted within this study. For the purposes of this particular study, the following features are dropped as they will not be contributing to the regression analysis: ID_Zindi, LAT and LON, and the ID (regional identifier). The remaining features are then scaled using a standard Min-Max scaler.

3.2 Regression Analysis

The analysis experiments with the use of classical machine learning models to learn trends within the data to predict the NO_2 levels. The use of deep learning models like LSTM based networks were explored on the aggregated and un-aggregated datasets, achieving MAPEs of 53.3% and 43.51% respectively. Hence, classical machine learning techniques have been applied which produce better results as seen in Section 4. Various types of classical machine learning models are used for comparative purposes in this analysis. Linear models, such as Ridge and Lasso regression, assume a linear relationship between features and the target variable, offering simplicity and ease of interpretation. Non-linear and non-parametric models, like Support Vector Regressor (SVR) and K-Nearest Neighbors (KNN), capture more complex relationships but are generally more difficult to interpret. Tree-based models, including Decision Trees, Random Forests, Gradient Boosting, XGBoost, LightGBM, and CatBoost, use feature splits to form tree structures, effectively handling both linear and non-linear patterns, though often at the expense of interpretability. For fine-tuning, the Optuna framework was used to determine the optimal hyperparameters (Akiba et al., 2019). The models' performances was evaluated using two metrics, namely the mean absolute percentage error (MAPE) and the root-mean square error (RMSE) (Eqs. 1 and 2 respectively.).

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - \hat{Y}_i|}{Y_i} \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{2}$$

In both equations, Y_i is the actual value at index i with $\hat{Y}_i$ represents the forecasted value at index i. The term n refers to the total number of samples. Using these two metrics allows us to see both the average magnitude of prediction errors and a scale-independent measure relative to the actual values at any point. In particular, MAPE gives us a percentage-based measure of a model's performance whilst RMSE is in the same units as the dependent variable, making it more interpretable. Both are heavily used within the field.

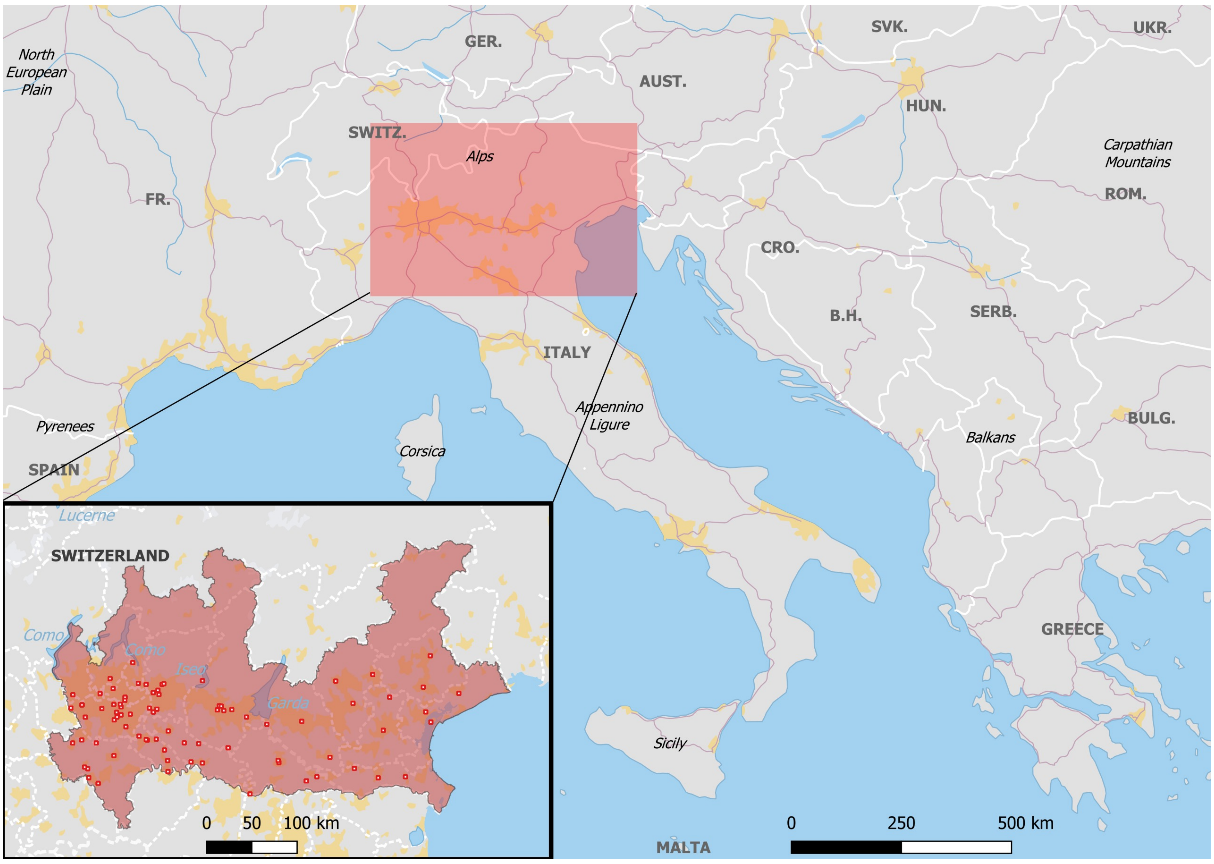

Figure 1. Map of the administrative region of Lombardy containing the locations of the monitoring stations.

The best models are determined by using 10-fold cross validation and then conducting the Kruskal-Wallis Test to determine if the performance of the models are statistically different. After determining the best models using these metrics, the top four best performing models are fine-tuned and put into an ensemble model. This ensemble model is evaluated based on MAPE and prediction accuracy.

Table 1. Summary of Dataset Features

	Description
Features	**ID_Zindi**: Spatial Identifier
	Date: Date of Data Collection
	ID: Region Identifier
	LAT: Latitude
	LON: Longitude
	Precipitat: Precipitation from CHIRPS
	LST: Land Surface Temperature from NOAA
	AAI: Absorbing Aerosol Index
	CloudFract: Cloud Fraction
	NO_2_strat: Stratospheric NO_2 Concentration
	NO_2_total: Total NO_2 Concentration
	NO_2_trop: Tropospheric NO_2 Concentration
	Tropopause: Tropopause Height
Target	**GT_NO_2**: Ground-Truth NO_2 Levels
Total Size	86,584

3.2.1 Aggregating Data Further evaluation is done on the model by aggregating the results based on recorded dates. In other words, all the results for a given day are summed together and the values (after standardizing) are used to train a similar ensemble model. The results of this will be reported to determine whether aggregating data through time is able to improve model performance.

3.3 Geo-spatial Analysis

To further understand the sources of variability in the NO_2 levels with respect to temporal and spatial aspects, we divide the data into sub-categories, particularly pertaining to the seasonality and the degree of urban development. We considered two periods of the year (Spring / summer, Fall / winter), along with two levels of urban development. This is done by cross-referencing the longitude and latitude pairs of each monitoring station with corresponding historical NDVI (normalized difference vegetation index) values: Points with high average NDVI (beyond the range of 0.1-0.2 are categorized as rural, whilst points with average NDVI values lower than that are categorized as urbanized (de la Iglesia Martinez and Labib, 2023). As expected, these points line up with monitoring stations in metropolitan cities such as Milan or Brescia, where we are confident NDVI is lower due to the urbanized environment.

The data is aggregated within each of the four scenarios such that the training data contains the first 90% of data when organized sequentially based on dates, and the testing data contains the remaining 10%. Table 2 shows the sizes of these subdatasets after the null-valued rows have been dropped.

The resulting models from each of the four scenarios are comparatively analyzed to determine whether seasonality and urban development play a factor in the prediction of NO_2 levels. This comparison is done mainly on the basis of feature correlation and importance in prediction.

https://doi.org/10.5194/isprs-annals-X-G-2025-893-2025 | © Author(s) 2025. CC BY 4.0 License.

Table 2. Sizes of Sub-Datasets for Training and Testing

Sub-Dataset	Training Size	Testing Size
Urban - Spring, Summer	428	48
Urban - Fall, Winter	360	41
Rural - Spring, Summer	428	50
Rural - Fall, Spring	378	42

4. Results and Discussion

We begin with a preliminary analysis of the performance of regression models on the full dataset. In particular, we looked at the evaluation metrics for each of the 10 regressors based on the default values. Table 3 shows the corresponding RMSE and MAPE values for each of the regressors. The highlighted values within the Table represent the four lowest MAPE values, which will then be used for further evaluation.

Table 3. Regression Model Performance Metrics

Regressor	RMSE	MAPE
LR	10.929 ± 0.222	0.5 ± 0.0114
Ridge	10.93 ± 0.223	0.501 ± 0.0114
Lasso	14.578 ± 0.317	0.765 ± 0.0147
DT	11.152 ± 0.116	0.431 ± 0.00508
RF	7.7931 ± 0.225	0.328 ± 0.00682
KNN	9.2053 ± 0.224	0.399 ± 0.00882
LGBM	7.7266 ± 0.217	0.335 ± 0.00717
XGB	7.449 ± 0.214	0.314 ± 0.00579
GB	8.7082 ± 0.211	0.379 ± 0.00811
CB	7.3148 ± 0.197	0.311 ± 0.00609
SVR	10.234 ± 0.238	0.401 ± 0.00927
Ensemble	7.1582 ± 0.224	0.306 ± 0.00609

Based on the MAPE values across each of the 10 regressors, it is clear to see that the best-performing models on the full dataset are CatBoost, LightGBM, GB and XGB, with CatBoost performing the best overall. However, their predictive capacity is not particularly strong. MAPE values of over 30% indicate that although the model is adequate in its predictive capacity, it needs to be improved. MAPE values closer to the 10-20% range are much more acceptable regardless of domain. For that an ensemble model based on voting criterion was implemented combing the aforementioned four regressors. The results achieved by the ensemble model outperforms all other regressors scoring in terms of the RMSE score achieving a 10.6784 and a relatively close to best MAPE score of 0.2764 which is slightly higher than that achieved by LGBM regressor.

We determine the best performing model using the Kruskal-Wallis test to determine if the average MAPE across all folds of evaluation were comparable between the ensemble model, CatBoost and XGBoost. The test between XGBoost and CatBoost shows a statistical indifference between the performances ($H(2) = 1.46$, $p = 0.226 > 5\%$). Similarly, the test between ensemble model and CatBoost also shows a statistical indifference ($H(2) = 2.29$, $p = 0.130 > 5\%$). However, comparing XGBoost and the ensemble model leads to a statistical difference ($H(2) = 6.22$, $p = 0.0126 < 5\%$). Hence, we conclude that the best models are CatBoost and the ensemble models.

4.1 Aggregated data

To further investigate the trends in NO_2 levels across northern Italy, we proceed with training models on the aggregated (by date) dataset. A similar process is taken to first determine the best-performing models with default parameters. Table 4 shows the MAPE and RMSE values for the models trained on the aggregated data. It is important to note that the RMSE values here will be significantly higher than in Table 3 since the scale of the target variable has changed (it is now at a country level rather than a weather station level).

Table 4. Regression Model Performance Metrics

Regressor	RMSE	MAPE
LR	473.54 ± 38.9	0.229 ± 0.026
Ridge	474.45 ± 36.4	0.23 ± 0.0259
Lasso	473.25 ± 38.1	0.23 ± 0.0259
DT	618.83 ± 40.6	0.272 ± 0.015
RF	458.15 ± 39.7	0.201 ± 0.0251
KNN	552.34 ± 48.7	0.256 ± 0.0205
LGBM	458.05 ± 43.5	0.195 ± 0.0237
XGB	482.97 ± 46.2	0.207 ± 0.0186
GB	450.05 ± 42.7	0.201 ± 0.0239
CB	447.01 ± 50.5	0.191 ± 0.0271
SVR	970.88 ± 74.8	0.416 ± 0.0216
Ensemble	446.02 ± 43.6	0.194 ± 0.0243

Immediately, it is evident that the aggregation of the data improves the predictive ability across all models. CatBoost and the ensemble model are still part of the top-performers, while XGBoost, Gradient Boosting, Random Forest, and LGBM all show considerable performance too. The Kruskal-Wallis Test ($H(6) = 3.223$,$p = 0.6655 > 5\%$) shows they have statistically indifferent performance in terms of their MAPE across the 10 folds.

Additionally, we conclude that the ensemble model does indeed explain the variance in the data, since the variance in the data for the target variable is $894,343.47$, and the mean squared error (MSE) from the ensemble model is $198,933.84$ ($\approx 22\%$ of dataset variance) which is far less than the variance in the data. The MSE here is calculated by squaring Eq. 2.

With the best model's MAPE dropping to 0.19, we see a clear improvement in inference. To further visualize the performance of this ensemble model, Fig. 2 shows the predictions of the model on the last three months of aggregated data (the testing set).

Other than the peaks, the model is able to accurately predict the direction of the data (or in other words, the NO_2 levels, and the general shape of the curve matches the actual values. This reflects the relatively low MAPE score that the ensemble model was able to produce. We can further evaluate its performance by looking at the percentage difference distribution, as shown in Fig. 3. As the bar chart captures, we can see that the percentage difference is skewed towards the bottom 15%, meaning that the majority of the predictions are within 15% of the actual value. This is a positive indicator for the model, meaning that we are within a confident interval for predicting the data.

The clear indication from the above-mentioned results is that aggregating data based on the date and then ensembling the top-performing fine-tuned models is able to give us the best predictions, especially when it comes to following the trends of the

https://doi.org/10.5194/isprs-annals-X-G-2025-893-2025 | © Author(s) 2025. CC BY 4.0 License.

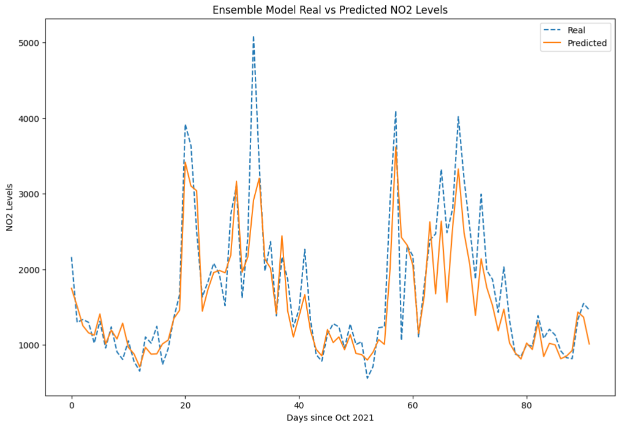

Figure 2. Prediction on the last 3 months of data (From October 2021 onward) for the Ensemble model trained on the aggregated data. NO_2 levels are represented in $\mu g/m^3$.

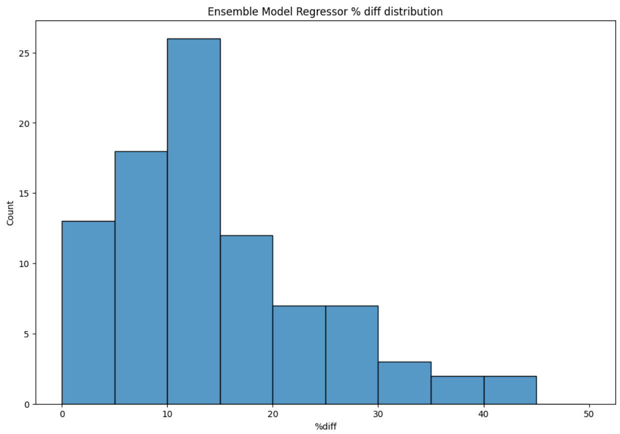

Figure 3. Percentage Difference of the model's prediction on the aggregated data compared to the actual values.

data. Fig. 2 shows us this exactly: although the peaks are not as exaggerated as with the actual data, the trends of the data are captured entirely. This ensemble model is also able to keep the percentage difference of its forecasts from the actual values to below 20%, which is another indication of its strong predictive capacity.

4.2 Geo-spatial Analysis

As mentioned in sec. 3.3, the dataset is divided based on urban development and on the seasonality. The same methodology is followed for fine-tuning models, but due to the high performance, models are not ensembled so as to not increase the complexity. The CatBoost regressor is used here and fine-tuned for each of the four subsets of data. The results are presented in Fig. 5 and Table 5 revealed a minor difference in NO_2 levels between urban and rural areas, though it was not statistically significant. This suggests that NO_2 levels are likely influenced by broader, more impactful factors rather than solely by illegal farming practices in rural areas.

Table 5. MAPE of CatBoost models for each of the 4 sub-datasets.

Model	MAPE
Urban - Spring, Summer	0.21796
Urban - Fall, Winter	0.22182
Rural - Spring, Summer	0.19988
Rural - Fall, Spring	0.22298

It is evident that on average, the models are able to accurately predict the NO_2 levels, facing the same issue of predicting the peaks as with the full dataset presented in the previous section. It is worthwhile to mention that the data was not pre-processed to account for potential outliers in the NO_2 values, which could be the leading cause of this. However, with an average MAPE score of 21.56% across the four models, it is justifiable to say that the model can reasonably predict the NO_2 levels.

The key indication of this set of results is that splitting the data based on the seasons and urban development actually aids in the process of predicting NO_2 levels as compared to the base models. This logically tracks as well, since we cannot expect NO_2 levels to follow the same trend across seasons and regions.

4.2.1 Feature Importance This section presents an analysis of the feature importances for each of the four models presented in the sub-data study. We begin with Fig. 4 which shows the feature importance for urban across seasons. On average, we can see that features such as the precipitation have low importance, especially during the fall and winter season. The two features with the highest importance across both seasons are the tropospheric NO_2 concentration and the total NO_2 concentration. These two features show the highest correlation across seasons. It is interesting to note that the correlation is slightly higher during the fall and winter seasons. Combining this with the fact that the correlation for precipitation is lower, we have a more robust approach to predicting NO_2 levels. On the other hand, the stratospheric NO_2 levels have significantly less importance in predicting the target variable.

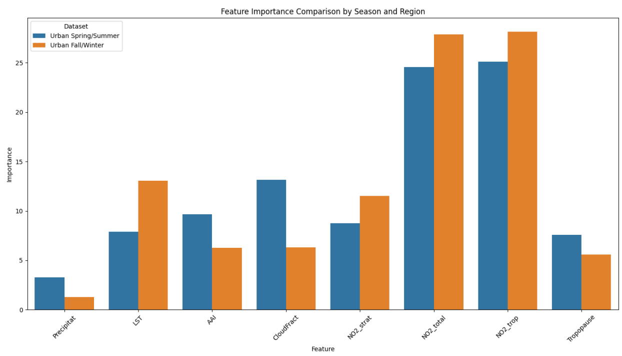

Figure 4. Feature importance expressed as a bar chart across seasons for Urban points.

We can compare the results in Fig. 4 to Fig. 6, which indicates that even for the highest correlated features, there appears to be a drop in the importance. A direct example of this would be the total NO_2 concentration, which significantly drops in importance across both seasons. The tropospheric NO_2 concentration still holds the highest importance, although the importance is reduced during the spring and summer seasons.

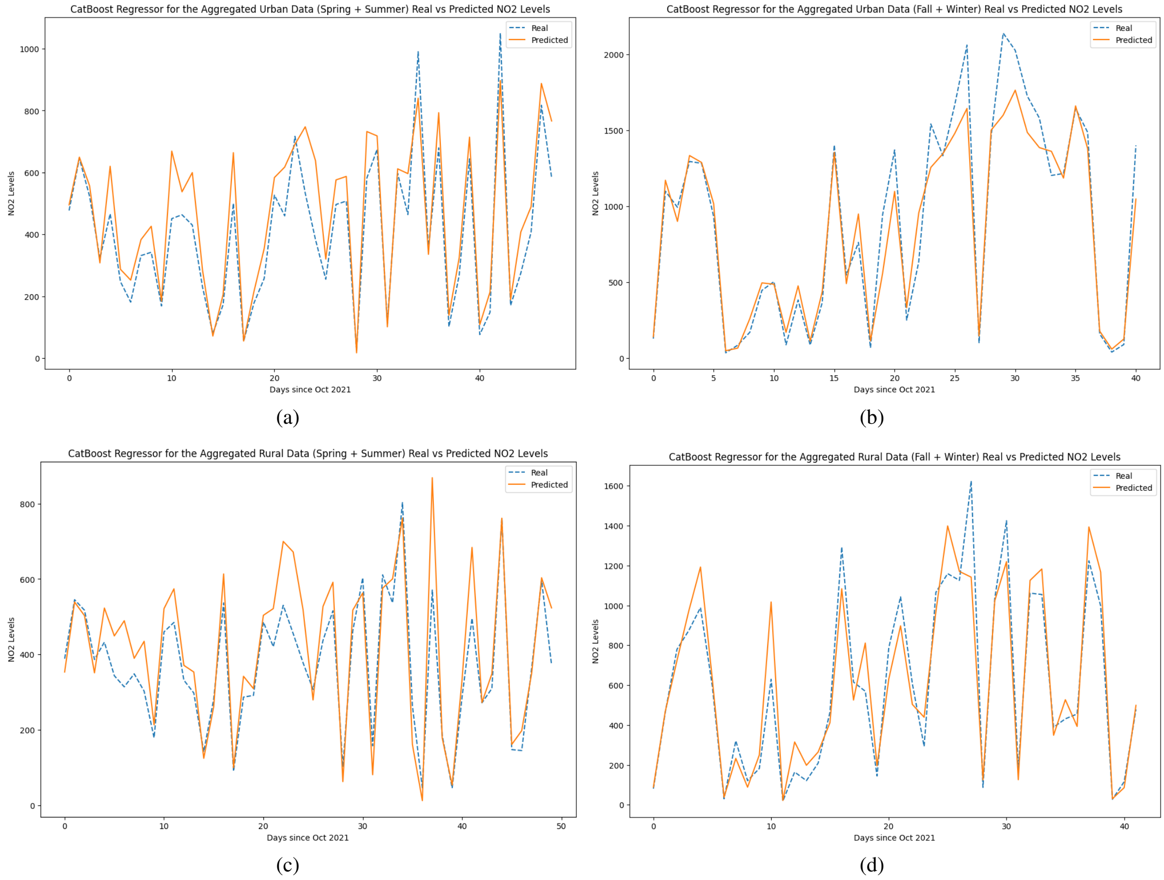

Figure 5. CatBoost predictions on the testing data (last three months of available data) for (a) Urban - Spring and Summer model, (b) Urban - Fall and Winter model, (c) Rural - Spring and Summer model, and (d) Rural - Fall and Winter model. NO_2 levels are represented in $\mu g/m^3$.

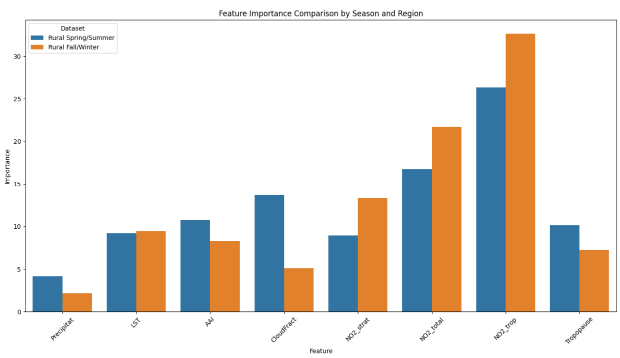

Figure 6. Feature importance expressed as a bar chart across seasons for Rural points.

4.2.2 Correlation Matrices To build on the feature importance study conducted in the previous section (which depends on the models and the importances that they calculate), we can look at the correlation matrix for each of the four sub-datasets between all features and the target variable. Fig. 7 shows the correlation matrix for the urban - spring and summer data. The findings align with the important characteristics in the previous section, particularly when showing the high correlations between the target variable and the various NO_2 indicators available. Changes across the different subsets of the data are minimal and thus the corresponding correlation matrices have been omitted, although they all follow almost the exact same trends.

5. Conclusion

In conclusion, this study demonstrates the feasibility of estimating nitrogen dioxide (NO_2) levels using open-source data across Italy, considering spatial and temporal variations. The results indicate that incorporating seasonal and urban development factors significantly enhances prediction accuracy, with an average MAPE score of 19% achieved using ensemble models. The analysis showed that there is a small difference between NO_2 levels in urban and rural areas but not statistically significant which may indicate that NO_2 levels does not seem to be a result of illegal farming practices in rural areas alone but are a result of a wider and more influential factors. The results showed that ensemble methods are particularly effective in capturing complex patterns in NO_2 distribution. Future work could expand the model's application to other regions to test its generalizability, while incorporating additional environmental variables could further improve accuracy. Advanced machine learning techniques and improved data collection processes also hold potential for refining NO_2 estimations using open-source data.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2623–2631.

Figure 7. Correlation matrix for urban data during spring and summer seasons.

Boersma, K., Eskes, H., Dirksen, R., Van Der A, R., Veefkind, J., Stammes, P., Huijnen, V., Kleipool, Q., Sneep, M., Claas, J. et al., 2011. An improved tropospheric NO 2 column retrieval algorithm for the Ozone Monitoring Instrument. *Atmospheric Measurement Techniques*, 4(9), 1905–1928.

Carlsen, L., Bruggemann, R., 2022. The 17 United Nations’ sustainable development goals: A status by 2020. *International Journal of Sustainable Development & World Ecology*, 29(3), 219–229.

Cedeno Jimenez, J. R., Brovelli, M. A., 2023. NO2 Concentration Estimation at Urban Ground Level by Integrating Sentinel 5P Data and ERA5 Using Machine Learning: The Milan (Italy) Case Study. *Remote Sensing*, 15(22), 5400.

Cedeno Jimenez, J. R., Pugliese Viloria, A. d. J., Brovelli, M. A., 2023. Estimating Daily NO2 Ground Level Concentrations Using Sentinel-5P and Ground Sensor Meteorological Measurements. *ISPRS International Journal of Geo-Information*, 12(3), 107.

Chi, Y., Fan, M., Zhao, C., Yang, Y., Fan, H., Yang, X., Yang, J., Tao, J., 2022. Machine learning-based estimation of ground-level NO2 concentrations over China. *Science of The Total Environment*, 807, 150721.

de la Iglesia Martinez, A., Labib, S., 2023. Demystifying normalized difference vegetation index (NDVI) for greenness exposure assessments and policy interventions in urban greening. *Environmental Research*, 220, 115155.

Fan, H., Wang, Y., Zhao, C., Yang, Y., Yang, X., Sun, Y., Jiang, S., 2021. The role of primary emission and transboundary transport in the air quality changes during and after the COVID-19 lockdown in China. *Geophysical Research Letters*, 48(7), e2020GL091065.

García, M. V., Aznarte, J. L., 2020. Shapley additive explanations for NO2 forecasting. *Ecological Informatics*, 56, 101039.

Ghahremanloo, M., Lops, Y., Choi, Y., Yeganeh, B., 2021. Deep learning estimation of daily ground-level NO2 concentrations from remote sensing data. *Journal of Geophysical Research: Atmospheres*, 126(21), e2021JD034925.

He, S., Dong, H., Zhang, Z., Yuan, Y., 2022. An ensemble model-based estimation of nitrogen dioxide in a southeastern coastal region of China. *Remote Sensing*, 14(12), 2807.

Huang, C., Sun, K., Hu, J., Xue, T., Xu, H., Wang, M., 2022. Estimating 2013–2019 NO2 exposure with high spatiotemporal resolution in China using an ensemble model. *Environmental Pollution*, 292, 118285.

Jiang, Q., Christakos, G., 2018. Space-time mapping of ground-level PM 2.5 and NO 2 concentrations in heavily polluted northern China during winter using the Bayesian maximum entropy technique with satellite data. *Air Quality, Atmosphere & Health*, 11, 23–33.

Lamsal, L., Martin, R., Van Donkelaar, A., Steinbacher, M., Celarier, E., Bucsela, E., Dunlea, E., Pinto, J., 2008. Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument. *Journal of Geophysical Research: Atmospheres*, 113(D16).

Long, S., Wei, X., Zhang, F., Zhang, R., Xu, J., Wu, K., Li, Q., Li, W., 2022. Estimating daily ground-level NO2 concentrations over China based on TROPOMI observations and machine learning approach. *Atmospheric Environment*, 289, 119310.

Oxoli, D., Cedeno Jimenez, J. R., Brovelli, M. A., 2020. Assessment of Sentinel-5P Performance for Ground-Level Air Quality Monitoring: Preparatory Experiments Over the COVID-19 Lockdown Period. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV-3/W1-2020, 111–116. https://isprs-archives.copernicus.org/articles/XLIV-3-W1-2020/111/2020/.

Piccoli, A., Agresti, V., Balzarini, A., Bedogni, M., Bonanno, R., Collino, E., Colzi, F., Lacavalla, M., Lanzani, G., Pirovano, G. et al., 2020. Modeling the effect of COVID-19 lockdown on mobility and NO2 concentration in the Lombardy region. *Atmosphere*, 11(12), 1319.

Pinder, R. W., Klopp, J. M., Kleiman, G., Hagler, G. S., Awe, Y., Terry, S., 2019. Opportunities and challenges for filling the air quality data gap in low-and middle-income countries. *Atmospheric environment*, 215, 116794.

Song, J., Wang, Y., Zhang, Q., Qin, W., Pan, R., Yi, W., Xu, Z., Cheng, J., Su, H., 2023. Premature mortality attributable to NO2 exposure in cities and the role of built environment: A global analysis. *Science of the Total Environment*, 866, 161395.

Trushna, T., Tiwari, R. R., 2022. Establishing the National Institute for Research in Environmental Health, India. *Bulletin of the World Health Organization*, 100(4), 281.

Van Geffen, J., Boersma, K., Van Roozendael, M., Hendrick, F., Mahieu, E., De Smedt, I., Sneep, M., Veefkind, J., 2015. Improved spectral fitting of nitrogen dioxide from OMI in the 405–465 nm window. *Atmospheric Measurement Techniques*, 8(4), 1685–1699.

Wang, Y., Wang, J., 2020. Tropospheric SO2 and NO2 in 2012–2018: Contrasting views of two sensors (OMI and OMPS) from space. *Atmospheric environment*, 223, 117214.

Wei, J., Liu, S., Li, Z., Liu, C., Qin, K., Liu, X., Pinker, R. T., Dickerson, R. R., Lin, J., Boersma, K. et al., 2022. Ground-level NO2 surveillance from space across China for high resolution using interpretable spatiotemporally weighted artificial intelligence. *Environmental science & technology*, 56(14), 9988–9998.

Weinmayr, G., Romeo, E., De Sario, M., Weiland, S. K., Forastiere, F., 2010. Short-term effects of PM10 and NO2 on respiratory health among children with asthma or asthma-like symptoms: a systematic review and meta-analysis. *Environmental health perspectives*, 118(4), 449–457.

Zhang, Z., Wang, J., Lu, W., 2018. Exposure to nitrogen dioxide and chronic obstructive pulmonary disease (COPD) in adults: a systematic review and meta-analysis. *Environmental Science and Pollution Research*, 25, 15133–15145.

Zheng, T., Bergin, M. H., Hu, S., Miller, J., Carlson, D. E., 2020. Estimating ground-level PM2. 5 using micro-satellite images by a convolutional neural network and random forest approach. *Atmospheric Environment*, 230, 117451.

Zhu, F., Ding, R., Lei, R., Cheng, H., Liu, J., Shen, C., Zhang, C., Xu, Y., Xiao, C., Li, X. et al., 2019. The short-term effects of air pollution on respiratory diseases and lung cancer mortality in Hefei: A time-series analysis. *Respiratory medicine*, 146, 57–65.