Article

Not peer-reviewed version

Spatiotemporal Patterns and Random Forest Prediction of Tropospheric NO_2 in Ankara, Türkiye

Fatih Ayhan , Fatih Adiguzel , Enes Karadeniz , Halil Barış Özel , Ioannis Charalampopoulos *

Posted Date: 12 March 2026

doi: 10.20944/preprints202603.0974.v1

Keywords: tropospheric NO_2; satellite data fusion; Sentinel-5P; Random Forest; spatial downscaling; urban air quality; machine learning

 doi:10.20944/preprints202603.0974.v1

Article

Spatiotemporal Patterns and Random Forest Prediction of Tropospheric NO_2 in Ankara, Türkiye

Fatih Ayhan [1], Fatih Adiguzel [2], Enes Karadeniz [3], Halil Barış Özel [4] and Ioannis Charalampopoulos [5,*]

[1] Department of Geography, Faculty of Arts and Science, Afyon Kocatepe University, Afyonkarahisar, Türkiye
[2] Department of Transportation Services, Transportation and Traffic Services Program, Vocational School of Technical Sciences, Bitlis Eren University, Bitlis, Türkiye
[3] Department of Geography, Faculty of Arts and Science, Inonu University, Malatya, Türkiye
[4] Faculty of Forestry, Department of Forestry, Bartın University, Bartın, Türkiye
[5] Laboratory of General and Agricultural Meteorology, Department of Crop Science, Agricultural University of Athens, 75 Iera Odos st., 11855, Athens, Greece
* Correspondence: icharalamp@aua.gr

Highlights

What are the main findings?

- Tropospheric NO_2 concentrations over Ankara exhibited pronounced seasonal variability, with winter peaks reaching monthly means up to 8.93×10^{-5} mol/m² and maximum values exceeding 30×10^{-5} mol/m², forming a persistent urban-rural gradient radiating from the metropolitan core.
- The Random Forest model successfully downscaled Sentinel-5P data to 500 m resolution using only Sentinel-2 spectral bands, achieving consistent predictive performance ($R^2 = 0.30$; RMSE = 2.72×10^{-5} mol/m²), with SWIR (B11, B12) and Red Edge bands emerging as dominant predictors.

What are the implications of the main findings?

- High-resolution NO_2 prediction maps derived solely from multispectral surface reflectance demonstrate that optical satellite data can serve as a scalable and cost-effective alternative for urban air quality assessment in regions with limited ground monitoring infrastructure.
- The identified center periphery pollution gradient and seasonal amplification under basin topography provide spatially explicit evidence for designing targeted emission control and spatial planning strategies in topographically constrained metropolitan regions.

Abstract

Urban air pollution remains a critical challenge in rapidly urbanizing metropolitan regions, where complex topography and uneven monitoring infrastructure limit accurate exposure assessment. Nitrogen dioxide (NO_2), primarily emitted from traffic and combustion sources, exhibits marked spatial heterogeneity that is often underrepresented by sparse ground-based stations. This study examines the spatiotemporal variability of tropospheric NO_2 over Ankara Province, Türkiye, for 2025 and develops and implements a machine learning-based downscaling framework integrating Sentinel-5P TROPOMI observations with Sentinel-2 multispectral surface reflectance data, without relying on ancillary meteorological or emission datasets. After rigorous quality filtering and temporal aggregation, a Random Forest regression model was used to generate annual NO_2 maps at 500 m resolution based solely on spectral predictors. Results indicate strong seasonal variability, with winter monthly means reaching 8.93×10^{-5} mol/m² and peak values exceeding 30×10^{-5} mol/m², alongside a persistent urban–rural gradient radiating from the metropolitan core. The optimized model achieved consistent predictive performance ($R^2 = 0.30$; RMSE = 2.72×10^{-5} mol/m²), with SWIR and Red Edge bands contributing most strongly. These findings demonstrate that high-resolution

 doi:10.20944/preprints202603.0974.v1

urban NO_2 patterns can be reliably inferred from optical satellite data alone, providing a transferable and scalable framework for air quality assessment in data-limited metropolitan environments.

Keywords: tropospheric NO_2; satellite data fusion; Sentinel-5P; Random Forest; spatial downscaling; urban air quality; machine learning

1. Introduction

Cities in the Anthropocene epoch have evolved beyond mere demographic hubs to function as critical tipping points for global biogeochemical cycles and the primary epicenters of atmospheric degradation [1]. This is due to rapid urbanization, unregulated industrial growth, and the expansion of fossil-fuel-dependent transportation networks. Urban metabolism has been dramatically altered by these factors. Such alterations have negatively affected atmospheric chemical composition and air quality to the point at which both human health and ecological integrity are threatened [2–4]. In this regard, nitrogen dioxide (NO_2) is considered as a key indicator of urban air pollution, and is recognized as a major pollutant in urban area, mainly generated from traffic and industrial combustion processes, and also has significant role in photochemical reactions that trigger the formation of secondary pollutants such as ozone (O_3) and fine particulate matter ($PM_{2.5}$) [5,6]. As such, monitoring NO_2 concentrations is not merely an environmental measurement activity but a core constituent of urban sustainability and public health security.

Exposure to NO2 is associated with an increased risk of respiratory tract infections, acute exacerbations of chronic obstructive pulmonary disease (COPD), cardiovascular issues and premature death, as shown in numerous epidemiological research [7,8]. In light of these escalating health risks associated with NO2 exposure, the World Health Organization (WHO) lowered recommended limits for NO_2 emissions in outdoor air from 40 $\mu g/m^3$ to 10 $\mu g/m^3$, which clearly underscoring the hazardous nature of chronic exposure to NO2 [9]. Exposure to airborne pollutants transcends being a mere public health issue; it also imposes a large economic burden on national governments, due to increased healthcare expenditures and reduced labor productivity [10]. Many cost-benefit analyses have demonstrated that the potential benefits to society and the economy resulting from reductions in airborne pollutant emissions far exceed the cost of implementing measures to reduce emissions [11].

Air quality monitoring in metropolitan areas with complex urban morphology, however, often entails inherent challenges of scale and representativeness, which can be difficult to resolve using traditional methodologies. Ground-based air quality monitoring (AQM) stations provide precise data at high temporal resolution but have a sparse spatial distribution due to their high capital and operating costs [12,13]. While these stations effectively measure the level of point-source pollution, they are still too limited to adequately represent the urban heterogeneity where pollutant concentration varies significantly even within a few hundred meters of emission sources [14,15]. The literature has defined this phenomenon as a spatial representativeness gap that leads to over- or underestimation of pollutant exposure levels in vast areas without AQM stations, resulting in poorly informed urban planning decisions based on incomplete datasets [16,17].

Utilizing satellite-derived remote sensing technology to overcome these limitations has catalyzed a paradigm shift in atmospheric monitoring. One of the major advancements made is through the use of the TROPOMI (TROPOspheric Monitoring Instrument) sensor aboard the European Space Agency's (ESA) Sentinel-5P mission. TROPOMI has enhanced atmospheric observation capacity to an unprecedented level, offering daily global coverage and a superior signal-to-noise ratio [18,19]. Although TROPOMI provides the highest accuracy to date for measuring tropospheric NO2 column density values, TROPOMI's spatial resolution is limited to approximately 5.5 km x 3.5 km, limiting the ability to detect changes in urban air quality at the neighborhood or street scale [20,21]. Therefore, there is a need to perform a spatial downscaling on satellite-based data to allow for application at a local level.

 doi:10.20944/preprints202603.0974.v1

Most existing downscaling models that are found in the literature rely heavily on auxiliary data, including meteorological parameters, traffic volumes, and emissions inventory data [22,23]. However, deficiencies in data infrastructure in developing regions and rapidly growing metropolises constrain the applicability of such models. Factors such as the coarse resolution of meteorological data and the outdated traffic datasets further exacerbate model uncertainty. In contrast, atmospheric concentrations of NO2 are not randomly distributed; instead, they exhibit a robust spatial correlation with land use and land cover (LULC) characteristics [24,25]. Emissions can be identified by asphalt/concrete surfaces, and pollutant "sinks" can be identified by vegetated surfaces and bodies of water [6]. As such, high-resolution optical satellite imagery data provide a robust surrogate source for deciphering the spatial patterns of emissions, eliminating the necessity for ancillary meteorological data [26,27]. However, the relationship between surface reflectance properties and NO2 concentrations is inherently nonlinear and multidimensional. Statistical methodologies that conventionally analyze the relationships between variables are typically inadequate to capture such complexity, but Machine Learning (ML) algorithms exhibit generally superior performance in deciphering these intricate interrelationships [28,29]. Specifically, ensemble learning algorithms such as Random Forest (RF), which have been shown to perform well in multiple studies on fusing sensory data, are especially useful due to their robustness to noisy data and their internal mechanisms that prevent overfitting [30,31]. As a result, these methods possess the potential to accurately model the spatial distribution of pollution by leveraging spectral information alone, thereby bypassing the requirement for ancillary datasets [32,33].

The focus of this study, Ankara Province, is a critical natural laboratory for studying the impacts of air pollution on account of its distinctive topography as well as rapid urbanization dynamics. The city's bowl-shaped basin geomorphology, enclosed by prominent mountain ridges, facilitates frequent onset of thermal inversion events, particularly during winter, which suppress atmospheric dispersion and lead to the entrapment of pollutants over the urban core [34,35]. In addition to geographic factors affecting the area, the rapid population growth, the encroachment of industrial zones into urban centers, as well as the intensive traffic loads, have all contributed to establishing Ankara as one of the most vulnerable metropolises regarding particulate matter and NO2 [36]. Nevertheless, the city's intricate topography and spatial heterogeneity cannot be fully captured by the existing network of ground-based monitoring stations, which remains limited in spatial density.

This study proposes a novel ML approach generating high-resolution (500 m) NO_2 prediction maps for the province of Ankara by combining the atmospheric data from Sentinel-5P TROPOMI with surface reflectance data from Sentinel-2 MSI. Importantly, this proposed methodology does not require ancillary atmospheric or social-economic datasets. The research seeks to address a critical knowledge gap in the current literature by (1) downscaling global-scale atmospheric data to the local scale, (2) testing the predictive capability of optical surface data in pollution estimation, and (3) developing a high spatial-resolution air quality monitoring model for metropolises with topographic constraints like Ankara. It is expected that the results of this study will allow urban planners and public health professionals to develop a cost effective and scalable decision-support mechanism to identify hotspots where exposure risk levels are the greatest.

2. Materials and Methods

2.1. Study Area: Ankara Province

The study area encompasses the province of Ankara (39.93° N, 32.85° E), which is the capital and administrative center of the Republic of Türkiye (Figure 1). It is located in the northwestern part of the Central Anatolia Region and has a total area of about 25,000 km^2. As of 2024, the province is roughly home to 5.8 million residents, which makes Ankara the second most populous metropolis in the country. Geomorphologically, the city is situated in a basin shaped like a bowl and is ringed with mountainous terrain. This unique topography limits atmospheric ventilation, thereby hindering both

 doi:10.20944/preprints202603.0974.v1

the horizontal and vertical pollutant dispersion and facilitating frequent temperature inversions, especially in winter months [37].

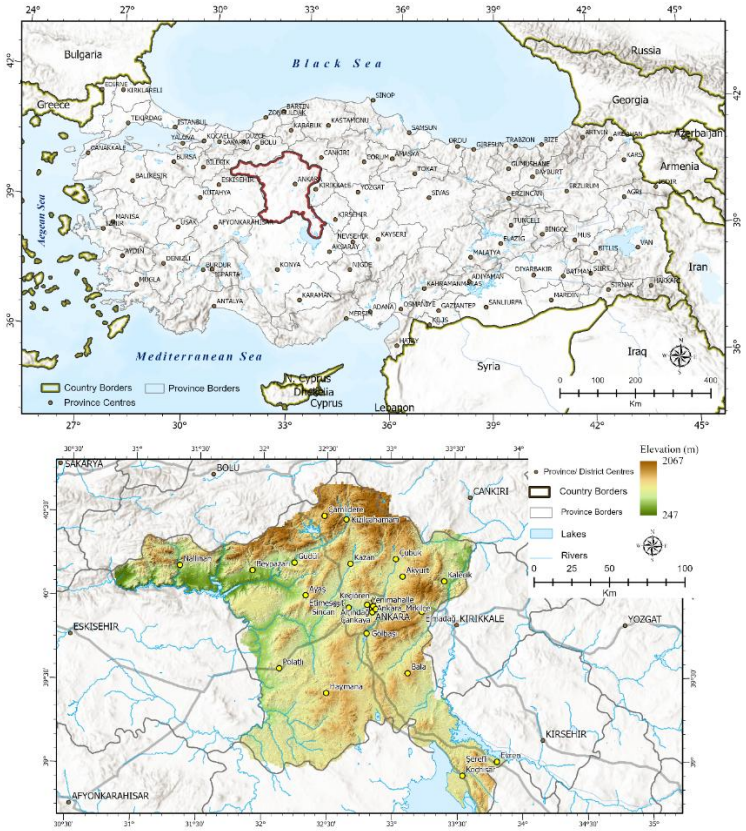

Figure 1. Location Map of the Study Area.

Ankara is characterized by a complex land cover, shaped by intensive urbanization, major industrial zones (Sincan, OSTİM, and İvedik Organized Industrial Zones) as well as an extensive transportation network. As urban growth and industrial activities increase rapidly, this is leading to higher levels of airborne pollutants such as NO2. This study will analyze the spatial and temporal dynamics of tropospheric NO_2 concentrations for the entire year of 2025, from January 1st to December 31st, 2025, by integrating high-resolution optical satellite data and atmospheric sensor measurements.

2.2. Datasets and Sources

This research used a "satellite-based multi-sensor data fusion" approach to address the issue of limited spatial representation at the local scale of individual air quality monitoring stations and to provide high accuracy predictions that are independent of available meteorological conditions. This methodology integrates the satellite-based atmospheric gas sensing capability of Sentinel-5P with the high- spatial-resolution surface reflectance data from Sentinel-2.

2.2.1. Sentinel-5P TROPOMI Atmospheric Data

The primary dependent (target) variable of the model was derived from the TROPOMI sensor on board the Sentinel-5P satellite, operated under the ESA's Copernicus program. TROPOMI is a push-broom type spectrometer that performs measurements across the ultraviolet-visible (UV-VIS: 270-500nm), near-infrared (NIR: 675-775nm), and short-wave infrared (SWIR: (2305-2385nm) spectral ranges, providing daily global coverage (Monzon-Herrera et al., 2025; Srivastava et al., 2026).

 doi:10.20944/preprints202603.0974.v1

For the present research, "Offline" (OFFL) Level-2 tropospheric NO_2 column density (mol/m^2) products were obtained from the Copernicus Open Access Hub for the period between January 1st and December 31st, 2025. The retrieval of atmospheric NO2 concentrations is based on the Differential Optical Absorption Spectroscopy (DOAS) technique, which calculates the slant column density (SCD) using the specific abosorption features of NO2 in a spectral range of 405-465 nm wavelength that are derived through the logarithmic ratio of measured solar radiance to irradiance. The resulting SCD values are then converted into vertical tropospheric column densities (VCD) using an air mass factor (AMF) [38].

The following stringent filtering steps were applied to maximize the reliability of the dataset and minimize atmospheric noise:

1. Quality Assurance (QA) Filtering: This step is essential to minimize the errors that can be caused by cloud cover, uncertainty in surface albedo, and aerosols in satellite data. In order to meet the criteria as defined in the literature, only pixels with a quality assurance value (qa_value) greater than 0.75 were included in the analysis [21,39]. That threshold ensures conditions where the cloud radiance fraction is below 0.5, and no snow/ice cover is present, thus minimizing the error associated with atmospheric scattering.

2. Temporal Aggregation: In order to address the issue of gaps in data resulting from the daily nature of Sentinel-5P overpasses and to determine seasonal trends in pollution, daily observations were aggregated into monthly mean composites within a Python environment. The monthly average NO_2 maps for 2025 are shown in Figure 2, which were derived through the processing of these high-quality datasets.

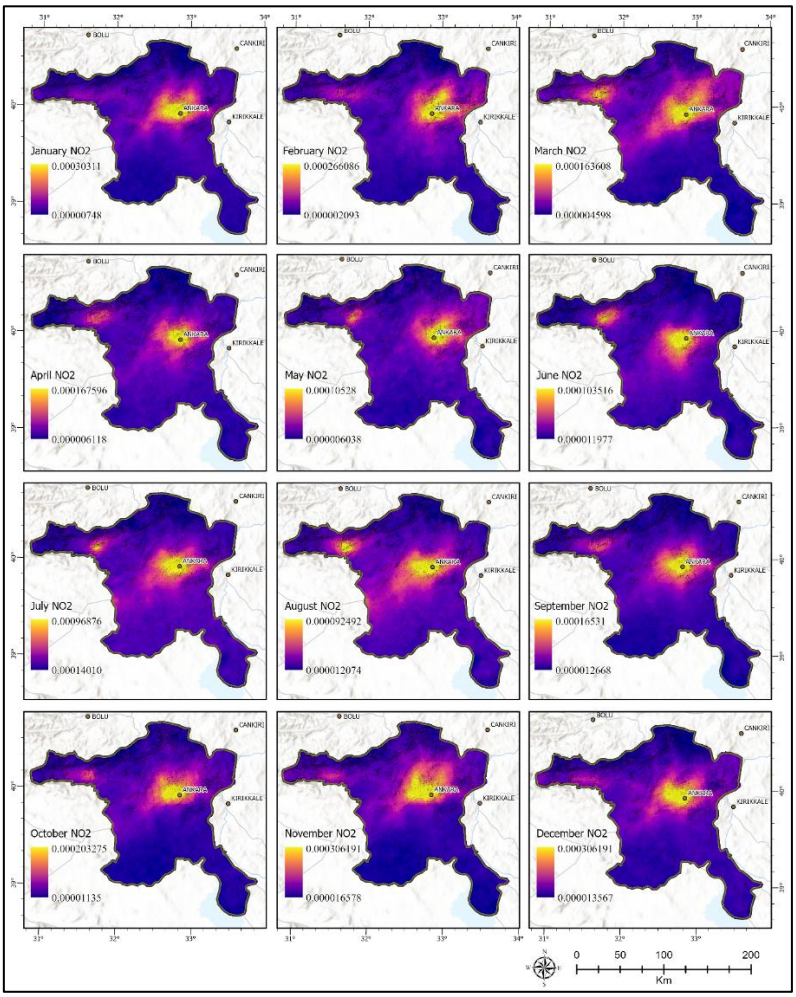

Figure 2. Spatiotemporal distribution of monthly average NO_2 concentrations across Ankara province.

 doi:10.20944/preprints202603.0974.v1

2.2.2. Sentinel-2 Multispectral Surface Reflectance Data

This research uses a novel methodological approach, relying solely on multispectral surface reflectance data from Sentinel-2A/B satellites as the independent variable, rather than using auxiliary meteorological and/or socio-economic data. The theoretical basis for this methodology is the presence of sufficient information contained within the spectral signature of urban (concrete/asphalt/rooftop) and natural (vegetation/water) surfaces relative to identifying and quantifying emissions from specific locations [38].

Given that the spatial resolution of Sentinel-5P data (approximately 5.5 km x 3.5 km) does not allow it to capture urban air pollution dynamics at either street or neighborhood scales, the high spatial resolution of Sentinel-2 (10-20 m) imagery was used for purposes of spatial downscaling. The L2A Sentinel-2 surface reflectance product, which are atmospherically corrected at the bottom of the atmosphere, for the year 2025 with cloud cover threshold of <20% was chosen for analysis. Specific spectral bands were selected from the Sentinel-2's 13 spectral bands as model inputs, based on their established correlations with NO_2 concentrations in the literature.

- Shortwave Infrared (SWIR) Bands (B11, B12): Srivastava et al., [38], demonstrated through their analysis that SWIR bands (1610 nm and 2190 nm) are very sensitive to atmospheric aerosols and gases and therefore exhibit the strongest correlation with the concentrations of NO_2. They are also very useful for differentiating urban construction materials and industrial surfaces, and are capable of detecting anthropogenic heat islands, which is a critical indicator of emissions generated within an urban area.
- Red Edge and Near-Infrared (NIR) Bands (B5, B6, B7, B8, B8A): These bands are strongly associated with vegetation health and chlorophyll content. Since vegetation (e.g., parks and forests) acts as a biological sink for NO2, concentrations represent a significant negative correlation [38].
- Visible Bands (B2, B3, B4): The blue, green, and red bands contribute to the model by representing overall surface albedo and the density of the urban fabric.

2.3. Methodology

The methodologies used in this research are data preprocessing, spatial temporal alignment, ML model development, hyperparameter optimization, and validation stages. All of the data processing and modeling phases of this study were conducted using the researchers' local computer environment with Python (v3.11). The Workflow involved the use of xarray, rasterio, and pandas for the data manipulation phase and scikit-learn for the modeling phase [40].

2.3.1. Data Preprocessing and Integration

The combination of datasets with different levels of spatial (Sentinel-5P: ~5.5 km vs. Sentinel-2: 10-20 m) and temporal resolutions is an essential component of improving the performance of models that estimate air pollution. The integration of the two datasets was accomplished by using the following procedural methodology:

1. Resampling: Bilinear interpolation was used to convert the low spatial resolution Sentinel-5P data into a format that would be compatible with the higher spatial resolution of Sentinel-2. This process also allowed reducing the spatial resolution of the Sentinel-5P data to a 100m x 100m grid, which minimized the impact of pixelation artifacts and ensured the spatial continuity of pollutant concentrations (Monzón Herrera et al., 2025; Li et al., 2022).

2. Temporal Matching: Temporal matching was done by aligning S-2 images having minimal cloud with Sentinel-5P's daily overpasses (around 13:30 local time). Monthly mean composite data were used for filling gaps in data that would allow preserving the temporal relationship. This process also allowed for reducing the level of noise resulting from immediate meteorological changes as well as to create a more stable modeling framework.

 doi:10.20944/preprints202603.0974.v1

3. Feature Extraction: Using the scikit-learn library, a systematic sampling grid was created for the study area. For each sampling point, Sentinel-2 spectral band values (B2–B12) were integrated as independent variables (features), while Sentinel-5P NO2 column density was used as the target variable (label). The objective of this model was to capture the non-linear relationships between the surface reflectance and atmospheric NO2 column density.

2.3.2. Random Forest Regression Model

A nonparametric RF regression model was chosen to estimate the yearly average NO2 concentrations due to its ability to capture both complex and non-linear relationships in a dataset. RF is an ensemble learning technique developed by Leo Breiman (2001) that builds multiple decision trees from random subsets of the training data using bootstrapping (sampling with replacement) [21].

The primary factors behind the selection of RF in this study were based on the following:

- High Performance with Spectral Data: RF can efficiently manage high-dimensional datasets (the 13 bands of Sentinel-2) and handle multicollinearity among variables (e.g., the strong correlation between Red Edge bands) [38].
- Resistance to Overfitting: While an individual decision tree is prone to overfitting noisy data, RF reduces variance and enhances the model's generalization ability by averaging the predictions of multiple trees.
- Feature Importance: The inherent design of the RF algorithm allows for assessing feature importance and which Sentinel-2 spectral bands have the greatest influence on the prediction of NO_2 concentrations. The bands identified as having the greatest contribution to the model's predictive capability included B11 (SWIR-1), B12 (SWIR-2), and the red edge bands. The significant influence of these bands indicates that the spatial distribution of NO_2 is influenced by spectral signatures of surface properties, built environments and human activity.

For training the model, the dataset was divided randomly with an 80/20 split for the training set and test set, respectively [37,38]. The Grid Search technique was used to fine-tune the hyper-parameters of the RF model (i.e., n_estimators and max_depth) so as to achieve the best possible predictive results.

The impact of the number of decision trees for this model's performance was assessed by determining RMSE for different tree counts. This is demonstrated visually in Figure X. It can be seen that as the number of trees increased from 50 to 1,000 there was a considerable decrease in RMSE with the eventual flattening of RMSE values once it reached its lowest value at 1,000 trees. Therefore, given the both the model's overall accuracy and computational cost, the final RF model was built with 1000 decision trees (n_estimators = 1000).

The maximum depth (max_depth) parameter was not explicitly constrained, allowing the model to be able to capture the non-linearity and complexity of the data. During the training phase, it was noted that the effective tree depth converged at around 19-20 levels. This provided a good trade-off between high predictive performance and overfitting.

An annual average NO_2 prediction map for 2025, presented in Figure 3, was derived by deploying the optimized RF model on the Sentinel-2 dataset. The resulting map provides a high-resolution representation of NO_2 spatial distribution across Ankara, effectively delineating pollution intensity and identifying potential hotspots within urban areas.

 doi:10.20944/preprints202603.0974.v1

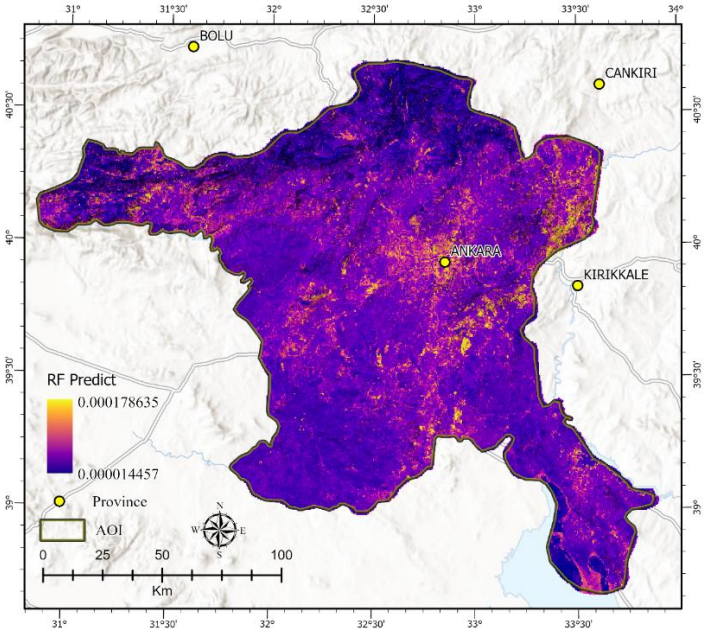

Figure 3. Annual Mean NO_2 Concentrations Predicted by the Random Forest Model (2025).

2.3.3. Validation and Performance Assessment

Three commonly used statistical measures were applied to assess the accuracy of the developed ML model using a 20% test dataset, which was excluded from the training process to determine the model's predictive ability and generalization potential. These measures included Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) [21,37,41]. The equations for each measure are as follows:

The coefficient of determination (R^2) measures how much of the variation in the dependent variable (the NO_2 concentration) is due to variation in the independent variables (Sentinel-2 spectral bands). An R^2 close to 1 would mean that NO_2 pollution has a good fit according to the surface reflectance characteristics:

$$R^2 = 1 - [\sum(yi - \hat{y}i)^2/\sum(yi - \bar{y})^2] \quad (1)$$

RMSE is defined as the square root of the mean of the squared differences between predicted and observed values. The RMSE metric is particularly sensitive to large errors and provides direct information on the overall accuracy of the model predictions:

$$RMSE = \sqrt{[(1/n) \sum(yi - \hat{y}i)^2]} \quad (2)$$

MAE denotes the average of the absolute differences between predicted and observed values, reflecting the model's average error magnitude:

$$MAE = (1/n) \sum|yi - \hat{y}i| \quad (3)$$

Here, n represents the total number of samples, yi is the observed NO_2 value obtained from Sentinel-5P data, $\hat{y}i$ is the NO_2 value predicted by the RF model, and $\bar{y}$ denotes the arithmetic mean of the observed NO_2 values. The statistical significance of the model was also evaluated using a t-test, with a significance level of $p < 0.05$ [38]. This approach was employed to test whether the relationship between model predictions and observational data is non-random. This methodological framework enabled high-resolution NO2 predictions across the entire Ankara province, even where ground-based monitoring stations or other auxiliary data sources (traffic density and meteorology) are lacking. As such, the developed model offers a scalable, reproducible, and cost-effective decision-support tool for urban air quality monitoring and assessment.

3. Results

 doi:10.20944/preprints202603.0974.v1

3.1. Temporal And Spatial Analysis

The data in Table 1 are the monthly averages of the tropospheric NO_2 column density ($\times 10^{-5}$ mol/m^2) over Ankara in 2025. These data show that there is a considerable temporal variability in NO_2 concentrations throughout the year, with evident month to month variations. Moreover, these variations were not uniform over the entire area of the province, but varied heterogeneously with distinct accumulation patterns in specific geographic locations, as illustrated in Figure 2.

The monthly statistical parameters and spatial distribution of tropospheric NO_2 column densities over Ankara for the year 2025 reveal a pronounced seasonal periodicity characterized by significant anthropogenic influence. According to the data in Table 1, the region exhibits a bimodal seasonal cycle where the highest mean concentrations are recorded during the winter months, specifically in January ($8.93 \pm 6.36 \times 10^{-5}$ mol/m^2) and December ($8.84 \pm 5.73 \times 10^{-5}$ mol/m^2). These peak values, which reach maximums exceeding 30.60×10^{-5} mol/m^2, coincide with increased fossil fuel combustion for residential heating and reduced boundary layer heights typical of the cold season.

Conversely, a substantial decline in NO_2 levels is observed during the summer period, with the lowest mean concentration occurring in May ($3.11 \pm 1.70 \times 10^{-5}$ mol/m^2). This reduction, approximately 65% lower than winter peaks, is primarily attributed to enhanced photochemical loss via OH radicals and increased atmospheric mixing during warmer months. Spatially, the distribution is highly heterogeneous; as illustrated in Figure 2, the highest accumulation patterns are consistently localized over the Ankara metropolitan core and along major transportation corridors. The high standard deviation values during winter (e.g., 6.36 in January) further quantify the sharp contrast between the heavily polluted urban center and the relatively stable background levels of the rural periphery. Overall, the transition from low summer concentrations to an upward trajectory starting in September ($5.01 \pm 3.11 \times 10^{-5}$ mol/m^2) highlights the dominant role of localized anthropogenic emissions and meteorological forcing in governing the province's air quality dynamics.

Table 1. Monthly statistical parameters of tropospheric NO2 column densities in Ankara for the year 2025 ($\times 10^{-5}$ mol/m^2).

Month	Minimum	Maximum	Mean	Standard Deviation
January	0.77	30.30	8.93	6.36
February	0.34	26.60	7.16	5.01
March	0.77	16.40	5.16	3.20
April	0.61	16.80	4.65	2.97
May	0.99	10.30	3.11	1.70
June	1.44	10.40	3.41	1.80
July	1.40	9.69	3.57	1.60
August	1.30	9.25	3.36	1.46
September	1.40	16.60	5.01	3.11
October	1.25	20.30	5.79	3.63
November	1.72	22.90	8.00	5.08
December	1.36	30.60	8.84	5.73

The monthly mean NO_2 values range from 3.11 to 8.93. The highest average value was observed in January (8.93), followed by December (8.84) and November (8.00). The lowest average value was recorded in May (3.11). It was observed that average NO_2 concentrations remain relatively low during late spring and summer and begin to increase again starting from the autumn months.

Maximum NO_2 values demonstrate that the pronounced peak concentrations occur primarily during the winter months. As such, the highest maximum concentration of NO_2 was measured in December (30.60) followed by comparable maxima in January (30.30). In contrast, maxima in NO_2 concentrations diminish significantly during the summer months, with the lowest maximum observed in August (9.25). Conversely, the minimum NO_2 values were seen to be confined to a

 doi:10.20944/preprints202603.0974.v1

narrow band throughout the year, ranging from 0.34 (February) to 1.72 (November). This indicates that low concentrations are distributed within a more limited band across the year.

Standard deviation values reveal that NO_2 concentrations exhibit a seasonally dependent spatial heterogeneity. Significantly higher standard deviations were observed during the winter months (e.g., January: 6.36; December: 5.73), whereas these values markedly declined in the summer, reaching their minima in August (1.46). This trend indicates that NO_2 distribution is more heterogeneous during the winter, while it adopts a more homogeneous structure during the summer season.

Maps of the monthly spatial distribution of NO_2 (Figure 2), clearly indicate that NO_2 levels are concentrated in specific regions in Ankara. In comparison to other regions of the province, levels of NO_2 in the city center and in the central districts of the city are generally higher throughout the year. Specifically, the districts of Çankaya, Yenimahalle, Keçiören, Altındağ, and Mamak are among the districts that are shown to have the highest NO_2 concentration values during a majority of the months. The districts exhibiting high concentrations of NO_2 also demonstrated high spatial consistency, as the same districts repeatedly showed elevated concentrations throughout the year.

Conversely, the NO_2 concentrations remained lower throughout the year in northern and northwestern districts of the province, including Beypazarı, Nallıhan, Kızılcahamam, and Çamlıdere. These districts were also characterized by more extensive areas with low NO_2 concentrations and exhibited greater spatial homogeneity. Likewise, low to moderate NO_2 concentrations also dominate in many districts in the south and southeast parts of the province.

It is important to note that the patterns of spatial distribution also exhibit distinct seasonal variations. In the winter months (Dec., Jan., Feb.), elevated NO2 concentrations are spread outward from the urban core to the periphery in such a way that they create a greater gradient between the inner and outer rings of the province than during other seasons. Spatial variation becomes even more heterogeneous, and the contrast between the high- and low-concentration zones is even greater during this period. This contrast gradually diminishes throughout the spring, and by the summer months, low-concentration areas become more prevalent across the entire province.

In the summer months (June-July-August), higher NO_2 concentrations are concentrated within relatively small areas, while large portions of the areas outside of the urban area have lower NO_2 concentrations. August, in particular, represents a period of peak spatial homogeneity, where the contrast between the urban center and the periphery reaches its annual minimum. During September through November, the areas with moderate to high concentrations of NO_2 around the urban core begin to grow outward in the form of rings, creating increasing spatial heterogeneity.

Overall, the temporal statistics presented in Table 1 and the spatial distribution maps in Figure 2 yield highly consistent results. The analysis demonstrates that the concentration of tropospheric NO_2 in the area of the Ankara province has considerable spatiotemporal variability, especially due to a gradient between the central part of the city and its suburban districts. The spatial contrast is affected by the seasons, increasing in the cold season (winter) and decreasing in the warm season (summer).

These findings demonstrate that there is significant spatiotemporal variability for tropospheric NO_2 concentration within the Ankara province, and this can be shown to have a distinct spatial gradient from the center to the outer periphery of the urban area. This spatial contrast undergoes seasonal fluctuations, intensifying during the winter months and diminishing throughout the summer season.

3.2. Annual Mean NO_2 Predictions Using the Random Forest Model

Figure 3 shows predicted annual mean tropospheric NO2 concentrations in 2025 using an RF model. The prediction results reveal that annual mean NO_2 concentrations exhibit pronounced spatial heterogeneity across the Ankara province.

Annual mean NO_2 predictions based on RF indicate that the highest concentrations are located in the urban core as well as its surrounding central districts. The districts of Çankaya, Yenimahalle,

 doi:10.20944/preprints202603.0974.v1

Keçiören, Altındağ, and Mamak had the highest predicted NO_2 concentrations compared to the rest of the province. Predicted high values from the RF model also show a large amount of spatial continuity, which is indicative of a clear high-concentration area across these urban areas.

In comparison, annual average NO_2 levels are found to be significantly lower for the northern and northwest districts of the provinces (e.g., Beypazarı, Nallıhan, Kızılcahamam, and Çamlıdere). Spatial distributions of NO_2 levels appear relatively uniform within these predominantly rural areas, with low-to-moderate levels prevailing throughout. Additionally, it can be seen that rural or periphery areas in all districts of the provinces have low-to-moderate NO_2 concentration ranges.

The yearly average of the RF-based NO_2 prediction map reveals pronounced urban-rural gradient. NO_2 concentration levels are significantly higher and moderate in the city center and its close proximities, but decline gradually as moving further away from the urban core toward the outer limits of the Ankara Province. The overall trend in the concentration distribution in this area indicates a systematic, predictable spatial pattern on an annual scale.

Overall, the RF model's annual mean NO_2 prediction map shows a similar spatial trend as seen in the monthly observational NO_2 distribution maps. The RF-based estimates provided robust spatial verification of the observations regarding annual patterns of NO_2 concentration across the entire Ankara region.

3.3. Random Forest Model Performance and Variable Contributions

Figure 4 shows the connection between the NO_2 column densities as measured by Sentinel-5P, and the NO_2 column density predictions generated by the RF model. Comparison of estimated and observed values for NO_2 demonstrates that the RF model can accurately predict NO_2 concentration levels with some degree of accuracy. Model accuracy is demonstrated with a coefficient of determination and root mean square error (RMSE = 2.72×10^{-5} mol/m^2).

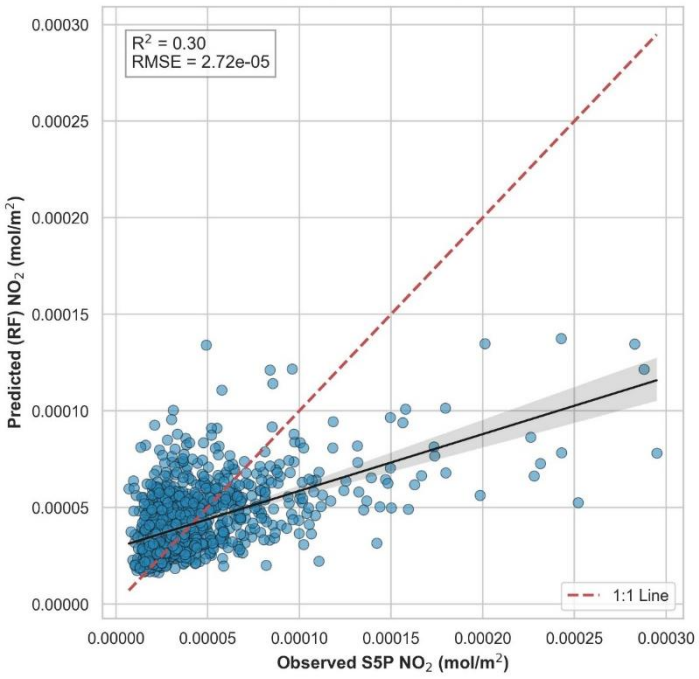

Figure 4. Correlation between Sentinel-5P observed and Random Forest-estimated NO_2 concentrations and assessment of model performance (R^2 vs. RMSE).

An analysis of data in the scatter plot demonstrates that, particularly at lower to moderate NO_2 concentrations, both the observed and predicted values are more densely clustered. Although the predicted values follow the observed values more closely within this interval, a more pronounced dispersion is evident at high NO_2 concentration levels. The deviation of the regression line from the

 doi:10.20944/preprints202603.0974.v1

1:1 identity line indicates that model estimates demonstrate a non-uniform distribution behavior at various concentration magnitudes.

The contributions of each of the spectral bands from Sentinel-2 to predict NO_2 concentrations with the RF model are illustrated in Figure 4. Variable importance analysis shows that bands B8A, B11, and B12 have the greatest relative importance, followed by band B2, B9, B4 and B3. The remaining bands contribute less individually but are crucial in providing the model's complementarity during the prediction process. Also, variable importance scores show that multiple spectral bands are used synergistically to estimate NO2 concentrations, highlighting the model's multivariate nature.

In Figure 5, it can be observed that the effect of the number of decision trees used in the RF model has on the RMSE. The graph clearly illustrates a rapid decline in the RMSE values as the number of trees increase. However, beyond a certain threshold, the rate of decline in the RMSE values slows down and then eventually stops improving, suggesting that the RMSE follows a saturation curve. This curve illustrates that at some point, the performance of the model will reach a plateau and including additional trees will result in no additional improvements in error reduction.

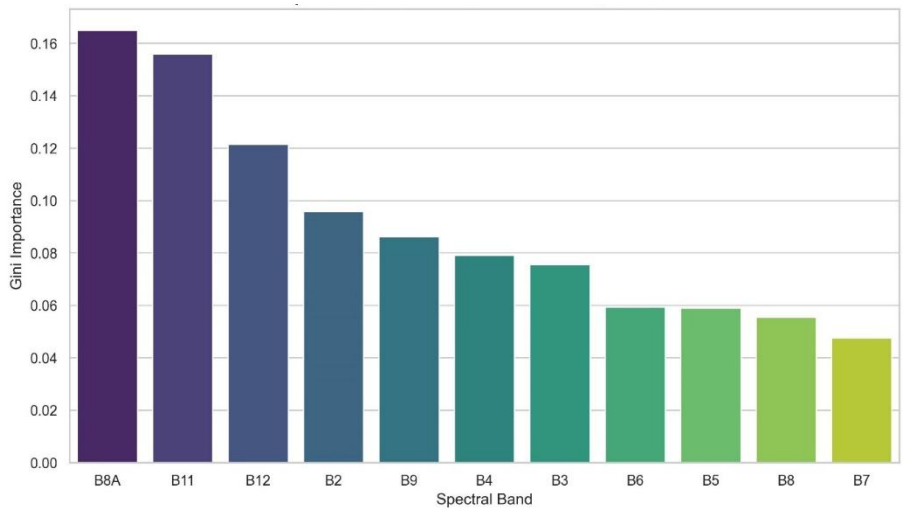

Figure 5. Relative contributions of Sentinel-2 spectral bands to NO_2 prediction within the Random Forest model.

The Pearson correlation matrix among Sentinel-2 bands and NO_2 concentrations is presented Figure 6. The correlation matrix indicates a generally strong positive correlation between the spectral bands. Stronger associations exist between bands in adjacent parts of the spectrum, which are indicative of a multi-collinear structure.

Correlation values for NO_2 levels with Sentinel-2 spectral bands are lower and show a predominantly negative distribution, compared to intra-band correlations. Since the absolute value of the correlation coefficient between NO_2 and all of the spectral bands is low, it follows that no single spectral band can explain NO_2 levels in a strong manner. This suggests that the distribution of NO_2 has a more complex structure than simply linear, direct relationships.

Although more significant negative correlation exists between NO_2 and some spectral bands, those correlations remained generally small in overall magnitude. The results indicate that each individual spectral band will not be able to represent NO_2 levels well enough for an accurate estimate of NO_2 levels. However, their synergistic integration offers significantly more robust information content. Thus, it appears that models for estimating NO_2 will need to incorporate multivariate modeling framework that simultaneously leverages multiple variables, rather than an approach relying solely on univariate band- NO_2 relationships.

 doi:10.20944/preprints202603.0974.v1

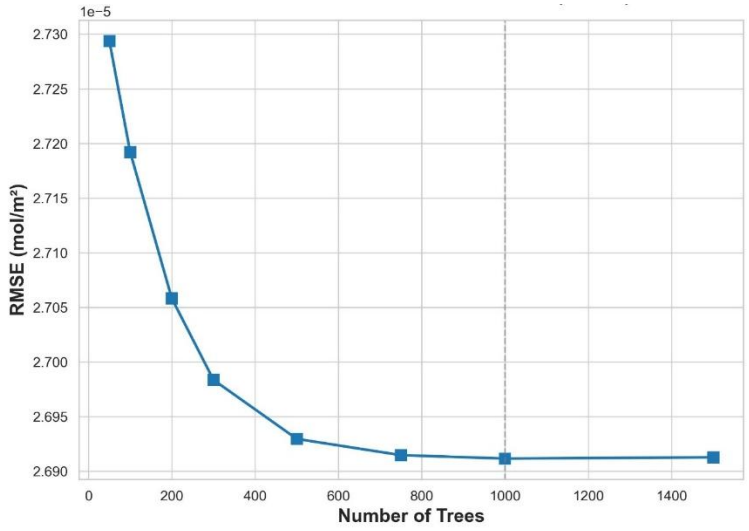

Figure 6. Effect of the number of trees on model error (RMSE) in the Random Forest model.

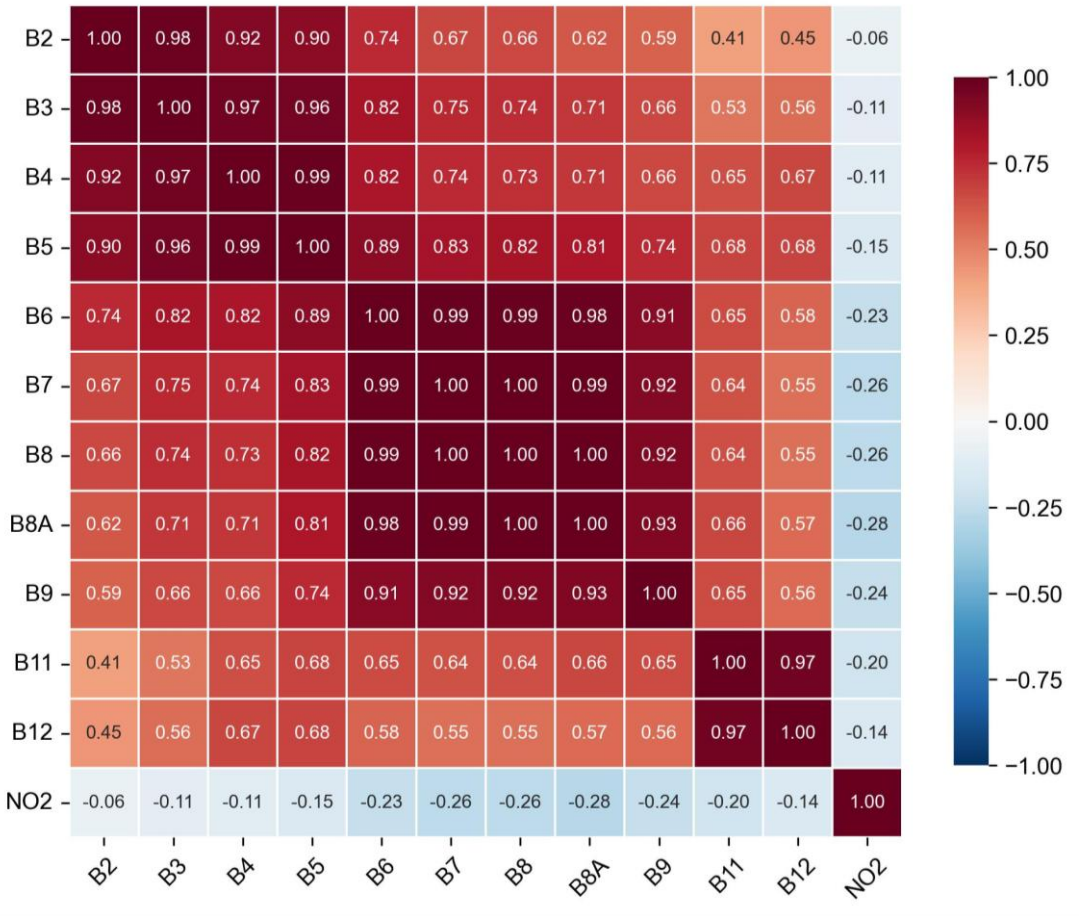

Figure 7. Pearson correlation matrix between Sentinel-2 spectral bands and NO_2 concentrations.

4. Discussion

This paper presents an integrated analysis of spatiotemporal variability of tropospheric NO_2 concentrations across Ankara using an ML method that incorporates remote sensing satellite data. The results derived from the synergy of Sentinel-5P and Sentinel-2 data demonstrate that air quality in urban environments shows marked seasonal and spatial heterogeneity. The study's findings are

 doi:10.20944/preprints202603.0974.v1

consistent with the larger scale spatial and temporal research which have demonstrated that NO_2 levels in major metropolitan areas fluctuate significantly throughout the seasons and follow a gradient in concentration from the center outwards [42].

Temporal analysis clearly indicates that there are high concentrations of NO_2 during the wintertime, with subsequent reductions to relatively lower levels in late spring and throughout the summer season. This seasonal variation in NO_2 concentrations has been found to be consistent with the topography of Ankara's bowl-shaped basin as well as the commonly occurring atmospheric stabilities present during wintertime. Other studies conducted in Türkiye was demonstrated that the relationship between temperature inversions and the low mixing heights, along with the typically stable atmospheric conditions, lead to the accumulation of pollutants during the wintertime (Toros et al., 2013; Ulutaş et al., 2021). As such, temporal patterns of NO_2 concentrations in this study were also consistent with patterns reported for Ankara and similar Central Anatolian cities [34,35].

The results from the spatial analysis indicate that NO_2 concentrations are mostly found within the Ankara city center and surrounding districts. However, NO_2 concentrations were generally lower and evenly spread throughout the peripheral districts and rural communities. The higher NO_2 concentrations observed in the more populated districts that have greater levels of urban activity (i.e., Çankaya, Yenimahalle, Keçiören, Altındağ, and Mamak), illustrate that the urban centers of larger cities are the most vulnerable to NO2 pollution. On the contrary, the lower NO2 concentrations measured in districts such as Beypazarı, Nallıhan, Kızılcahamam, and Çamlıdere provide evidence of an apparent center-periphery gradient across the province. This pattern demonstrates the same type of spatial relationship in terms of urban-rural air quality disparity among the largest metropolitan areas in Türkiye [36].

The annual average NO_2 prediction map produced by the RF model has a strong resemblance to those spatial patterns identified from observational data. The RF model's ability to recognize where the highest concentrations of NO_2 are found, as represented by high predicted values in the city center is impressive. Although the RF model had relatively modest R^2 and RMSE, it demonstrates the complex, non-linear nature of estimating NO_2. Other studies that used satellite-based ML models reported similar moderate explanatory power, particularly highlighting that the margin of error tends to escalate at higher concentration levels [21,41].

The statistical metrics for the performance of the model ($R^2 = 0.30$ and RMSE = 2.72×10^{-5} mol/m^2) for estimating tropospheric NO_2 show that such estimation has an inherent complexity and non-linear nature. Although the R^2 value is moderate, the model's performance is comparable to that of similar studies in the literature, given that it uses only Sentinel-2 optical surface reflectance data and does not include any auxiliary atmospheric or emission inventory data.

Existing literature has demonstrated that downscaling models based solely on satellite-derived spectral data, without including dynamic predictors such as meteorological variables (e.g., wind speed, relative humidity, temperature) or emission inventories, typically achieve explanatory power ranging from 25% to 45%. Therefore, the R^2 value calculated above demonstrates a statistically sound and consistent performance, especially given the types of input data used and the modeling framework.

The use of surface spectral signatures to represent NO_2 levels (a pollutant with short range atmospheric transport, sensitive to photochemistry, and highly dependent upon source emissions) at an approximate rate of 30%, without utilizing proxy variables such as wind field direction, traffic density, or point source emissions as "direct" indicators of NO_2 concentrations indicates a significant spatial correlation between land cover properties and air pollution.

Instead of trying to attain an absolute numerical level of accuracy, the RF model built in this framework provides a cost-effective, scalable preliminary assessment tool to identify spatial and temporal patterns, and localized concentrations, and potentially hotspots of tropospheric NO_2 pollution in metropolises such as Ankara, which are characterized by complex topography and a heterogeneous urban fabric.

 doi:10.20944/preprints202603.0974.v1

Variable importance analysis and correlation matrix of results show that there is no strong linear relationships for the concentration of NO_2 with each of the individual bands of Sentinel-2 spectral bands. Although the high degree of multicollinearity observed among the spectral bands themselves, their weak individual correlations with NO_2 suggest that simple linear models will be unable to capture the distribution of the pollutant. Therefore, the above-mentioned results support the need for multivariate and non-linear modeling frameworks, rendering the selection of ensemble algorithms such as the RF methodologically significant [37,38,43].

4.1. Contributions in the Context of Türkiye and Directions for Future Research

The current study is important for Türkiye because it demonstrates the viability of producing high-resolution predicted NO2 concentration maps from satellite data alone when the number of ground-based monitoring stations is insufficient for spatial representativeness. The approach developed for Ankara also supports previously established findings of global spatiotemporal assessments at a local scale and provides a scalable framework for urban air quality monitoring, exposure assessment, and risk-based planning processes [42].

However, the limited performance of the model suggests that more advanced forecasting frameworks could be established in the future through the integration of supplementary datasets such as meteorological variables, traffic density, land use or population data. Such integrated approaches will contribute to a more detailed understanding of the spatial and temporal dynamics of NO_2 pollution.

5. Conclusions

The spatiotemporal variation of tropospheric NO_2 concentrations was studied over Ankara Province using an integrated approach combining satellite-based observation methods and machine learning (ML) techniques. In temporal analysis, the combination of Sentinel-5P satellite data at 500 m resolution along with 500 m resolution NO_2 predictions produced using Sentinel-2 data and radio frequency (RF) algorithms indicated that large and complex heterogeneous urban areas, such as Ankara, exhibit significant seasonal and spatial variability in air quality.

Findings show there is a significant seasonal variation in NO_2 levels throughout the year and that levels tend to be higher during the winter months. When viewed together with the topography of Ankara and prevailing wintertime atmospheric stability, along with elevated emissions resulting from heating sources in urban settings, it can be concluded that air pollution intensifies seasonally in urban areas. Conversely, the reduced NO_2 levels seen in May and June may be attributed to the increased atmospheric boundary layer mixing and seasonal shifts in urban anthropogenic activities.

The spatial analysis demonstrates differences in the central/peripheral configuration of the region of Ankara. It can be observed from the relatively high levels of NO2 in the urban core and in the central city districts, that there is a greater vulnerability to pollution in these areas compared with all of the remaining parts of the region. On the contrary, it can be seen that the similar, and generally low, levels of NO_2 distributed across rural and periphery zones demonstrate the impact of the functions of the urban areas and their land use configurations on atmospheric quality. Consequently, it has been shown that there is a clear spatial dimension to pollution within the area of Ankara, as well as a temporal dimension.

The annual average NO_2 maps produced by the RF model largely represent the same spatial patterns as those from the observational data and demonstrate that the zones with the highest NO_2 concentrations have spatial continuity. In spite of the fact that the model's performance is still moderate ($R^2 = 0.30$), the result is a direct consequence of the attempt to model an air pollutant that has fast atmospheric transport and is highly sensitive to photochemistry based on optical satellite data alone. However, the approach developed here is a useful tool for obtaining a consistent and reliable preliminary assessment of potential locations of pollution hotspots and spatial risk areas.

These results show that there are no significant linear correlations between individual Sentinel-2 spectral bands and NO_2 concentration. As such, these results reinforce the need to consider

 doi:10.20944/preprints202603.0974.v1

multivariate and nonlinear modeling when estimating NO2 from remote sensing data. The results also show that RF is an appropriate algorithmic tool for analyzing complex environmental problems. However, it is important to acknowledge the limitations of this approach. Outputs from the model should be viewed as a method to identify relative risks and spatial patterns and should not be considered as providing accurate estimates of NO_2 concentration.

Overall, this research demonstrates that satellite-based data are sufficient to analyze spatiotemporal NO_2 tropospheric pollutant concentration patterns when the spatial representativeness of ground-based monitoring stations is limited, as demonstrated in Türkiye. Furthermore, this methodology for analyzing and addressing air pollution in urban environments via satellite-based data represents an alternative and potentially scalable and cost-effective methodology for assessing exposure to urban air pollution, identifying high-risk hotspots, and providing spatially based policy guidance to support a decision-making framework for local governments and policymakers in reducing urban air pollution.

It is suggested that future studies develop better forecasting models that will have better predictive capabilities and can be used to explain NO_2 dynamics through additional datasets (meteorological variables, traffic density, land use, demographics) integrated into the currently developed spectral model. Further comparisons of different machine learning and deep learning algorithms, along with further longitudinal studies over time, will provide a better understanding of NO_2 dynamics. The present study provides an essential benchmark for future methodological and applied research on air quality monitoring.

Author Contributions: Conceptualization, F.Ay., F.Ad. and E.K.; methodology, F.Ad. and E.K.; software, F.Ad. and E.K.; validation, F.Ad., E.K. and I.C.; formal analysis, E.K.; investigation, F.Ay. and E.K.; resources, F.Ad., I.C. and H.B.Ö.; data curation, E.K.; writing-original draft preparation, F.Ad. and E.K.; writing-review and editing, F.Ay., I.C. and H.B.Ö.; visualization, E.K.; supervision, F.Ad.; project administration, F.Ad. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets generated and analyzed during the current study are publicly available in the Zenodo repository at: https://doi.org/10.5281/zenodo.18875156. The repository includes processed satellite data layers, derived NO_2 prediction maps, and supporting datasets used in the Random Forest modeling workflow.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Alberti, M. (2023). Cities of the Anthropocene: Urban sustainability in an eco-evolutionary perspective. Philosophical Transactions of the Royal Society B, 379, 20220264. https://doi.org/10.1098/rstb.2022.0264
2. Velasco, E., Roth, M., Norford, L., & Molina, L. T. (2016). Does urban vegetation enhance carbon sequestration? Landscape and Urban Planning, 148, 99–107. https://doi.org/10.1016/j.landurbplan.2015.12.003
3. Cai, B., Cui, C., Zhang, D., Cao, L., Wu, P., Pang, L., & Dai, C. (2019). China city-level greenhouse gas emissions inventory in 2015 and uncertainty analysis. Applied Energy, 253, 113579. https://doi.org/10.1016/j.apenergy.2019.113579
4. Han, W., Li, Z., Guo, J., Su, T., Chen, T., Wei, J., & Cribb, M. (2020). The urban–rural heterogeneity of air pollution in 35 metropolitan regions across China. Remote Sensing, 12(14), 2320. https://doi.org/10.3390/rs12142320
5. Power, A. L., Tennant, R. K., Jones, R. T., Tang, Y., Du, J., Worsley, A. T., & Love, J. (2018). Monitoring impacts of urbanisation and industrialisation on air quality in the Anthropocene using urban pond sediments. Frontiers in Earth Science, 6, 131. https://doi.org/10.3389/feart.2018.00131

 doi:10.20944/preprints202603.0974.v1

6. Song, J., Wang, Y., Zhang, Q., Qin, W., Pan, R., Yi, W., & Su, H. (2023). Premature mortality attributable to NO_2 exposure and the role of built environment: A global analysis. Science of the Total Environment, 866, 161395. https://doi.org/10.1016/j.scitotenv.2023.161395
7. Khaniabadi, Y. O., Goudarzi, G., Daryanoosh, S. M., Borgini, A., Tittarelli, A., & De Marco, A. (2017). Exposure to PM_{10}, NO_2, and O_3 and impacts on human health. Environmental Science and Pollution Research, 24(3), 2781–2789. https://doi.org/10.1007/s11356-016-8038-6
8. Rodriguez-Villamizar, L. A., Rojas, Y., Grisales, S., Mangones, S. C., Cáceres, J. J., Agudelo-Castañeda, D. M., & Rojas, N. Y. (2024). Intra-urban variability of long-term exposure to $PM_{2.5}$ and NO_2 in five cities in Colombia. Environmental Science and Pollution Research, 31(2), 3207–3221. https://doi.org/10.1007/s11356-023-31306-w
9. World Health Organization. (2021). WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization. https://www.who.int/publications/i/item/9789240034228
10. Guerriero, C., Chatzidiakou, L., Cairns, J., & Mumovic, D. (2016). The economic benefits of reducing the levels of nitrogen dioxide (NO_2) near primary schools: The case of London. Journal of Environmental Management, 181, 615–622. https://doi.org/10.1016/j.jenvman.2016.06.039
11. Vartiainen, A., Mikkonen, S., Leinonen, V., Petäjä, T., Wiedensohler, A., Kühn, T., & Miinalainen, T. (2025). Machine learning-based downscaling of aerosol size distributions from a global climate model. Atmospheric Measurement Techniques, 18(20), 5763-5782. https://doi.org/10.5194/amt-18-5763-2025
12. 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, 116823. https://doi.org/10.1016/j.atmosenv.2019.06.032
13. Fioletov, V., McLinden, C. A., Griffin, D., Zhao, X., & Eskes, H. (2025). Global seasonal urban, industrial, and background NO_2 estimated from TROPOMI satellite observations. Atmospheric Chemistry and Physics, 25(1), 575–596. https://doi.org/10.5194/acp-25-575-2025
14. Souri, A. H., Nowlan, C. R., Wolfe, G. M., Lamsal, L. N., Miller, C. E. C., Abad, G. G., ... & Chance, K. (2020). Revisiting the effectiveness of HCHO/NO2 ratios for inferring ozone sensitivity to its precursors using high resolution airborne remote sensing observations in a high ozone episode during the KORUS-AQ campaign. Atmospheric Environment, 224, https://doi.org/10.1016/j.atmosenv.2020.117341
15. Yu, M., & Liu, Q. (2021). Deep learning-based downscaling of tropospheric nitrogen dioxide using ground-level and satellite observations. Science of the Total Environment, 773, 145145. https://doi.org/10.1016/j.scitotenv.2021.145145
16. Zhao, X., Griffin, D., Fioletov, V., McLinden, C., Cede, A., Tiefengraber, M., & Lee, S. C. (2020). Assessment of the quality of TROPOMI high-spatial-resolution NO_2 data products in the Greater Toronto Area. Atmospheric Measurement Techniques, 13(4), 2131–2159. https://doi.org/10.5194/amt-13-2131-2020
17. Wang, Y., Wang, J., Zhou, M., Henze, D. K., Ge, C., & Wang, W. (2020). Inverse modeling of SO_2 and NOx emissions over China using multisensor satellite data–Part 2: Downscaling techniques for air quality analysis and forecasts. Atmospheric Chemistry and Physics, 20(11), 6651–6670. https://doi.org/10.5194/acp-20-6651-2020
18. Carbone, A., Restaino, R., Vivone, G., & Chanussot, J. (2024). Model-based super-resolution for Sentinel-5P data. IEEE Transactions on Geoscience and Remote Sensing, 62, 1–16. https://doi.org/10.1109/TGRS.2024.3387877
19. Ding, J., Ren, C., Wang, J., Feng, Z., & Cao, S. J. (2024). Spatial and temporal urban air pollution patterns based on limited data of monitoring stations. Journal of Cleaner Production, 434, 140359. https://doi.org/10.1016/j.jclepro.2023.140359
20. Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. C., & Justice, C. (2018). The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sensing of Environment, 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002
21. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

 doi:10.20944/preprints202603.0974.v1

22. Hu, W., Zhao, T., Bai, Y., Shen, L., Sun, X., & Gu, Y. (2020). Contribution of regional $PM_{2.5}$ transport to air pollution enhanced by sub-basin topography. Atmosphere, 11(11), 1258. https://doi.org/10.3390/atmos11111258
23. Zhong, X., Zhao, L., Wang, J., Zheng, H., Yan, J., Jin, R., & Ren, P. (2021). Empirical models on urban surface emissivity retrieval based on different spectral response functions: A field study. Building and Environment, 197, https://doi.org/10.1016/j.buildenv.2021.107882
24. Abdullah, S., Ismail, M., Ahmed, A. N., & Abdullah, A. M. (2019). Forecasting particulate matter concentration using linear and non-linear approaches for air quality decision support. Atmosphere, 10(11), 667. https://doi.org/10.3390/atmos10110667
25. Lu, Y. (2023). Assessing air pollution exposure misclassification using high-resolution $PM_{2.5}$ concentration model and human mobility data. Air Quality, Atmosphere & Health, 16, 2225–2238. https://doi.org/10.1007/s11869-023-01404-2
26. Beauchamp, M., Malherbe, L., de Fouquet, C., & Létinois, L. (2018). A necessary distinction between spatial representativeness of an air quality monitoring station and the delimitation of exceedance areas. Environmental Monitoring and Assessment, 190(7), 441. https://doi.org/10.1007/s10661-018-6788-y
27. Morillas, C., Alvarez, S., Pires, J. C., Garcia, A. J., & Martinez, S. (2024). Impact of the implementation of Madrid's low emission zone on NO_2 concentration using Sentinel-5P/TROPOMI data. Atmospheric Environment, 320, 120326. https://doi.org/10.1016/j.atmosenv.2024.120326
28. Zallaghi, E., Goudarzi, G., Nourzadeh Haddad, M., Moosavian, S. M., & Mohammadi, M. J. (2014). Assessing the effects of nitrogen dioxide in urban air on health of west and southwest cities of Iran. Jundishapur Journal of Health Sciences, 6(4), 38–42. https://doi.org/10.5812/jjhs.23469
29. Lin, C.-A., Chen, Y.-C., Liu, C.-Y., Chen, W.-T., Seinfeld, J. H., & Chou, C. C.-K. (2019). Satellite-Derived Correlation of SO2, NO2, and Aerosol Optical Depth with Meteorological Conditions over East Asia from 2005 to 2015. Remote Sensing, 11(15), 1738. https://doi.org/10.3390/rs11151738
30. Kuik, F., Kerschbaumer, A., Lauer, A., Lupascu, A., von Schneidemesser, E., & Butler, T. M. (2018). Top–down quantification of NOx emissions from traffic in an urban area using a high-resolution regional atmospheric chemistry model. Atmospheric Chemistry and Physics, 18(11), 8203–8225. https://doi.org/10.5194/acp-18-8203-2018
31. Zhang, L., Yang, C., Xiao, Q., Geng, G., Cai, J., Chen, R., Meng, X., & Kan, H. (2021). A satellite-based land use regression model of ambient NO_2 with high spatial resolution in a Chinese city. Remote Sensing, 13(3), 397. https://doi.org/10.3390/rs13030397
32. Zhang, H., Chen, J., & Wang, Z. (2021). Spatial heterogeneity in spillover effect of air pollution on housing prices: Evidence from China. Cities, 113, 103145. https://doi.org/10.1016/j.cities.2021.103145
33. Nejad, M. T., Ghalehteimouri, K. J., Talkhabi, H., & Dolatshahi, Z. (2023). The relationship between atmospheric temperature inversion and urban air pollution characteristics: A case study of Tehran, Iran. Discover Environment, 1(1), 17. https://doi.org/10.1007/s44274-023-00018-w
34. Toros, H., Erdun, H., Çapraz, Ö., Özer, B., Bozyazı Daylan, E., & Öztürk, A. İ. (2013). Air pollution and quality levels in metropolitans of Turkey for sustainable life. Avrupa Bilim ve Teknoloji Dergisi, 1(1), 12–18.
35. Ulutaş, K., Abujayyab, S. K. M., & Abu Amr, S. (2021). Evaluation of the major air pollutants levels and its interactions with meteorological parameters in Ankara. Mühendislik Bilimleri ve Tasarım Dergisi, 9(4), 1284–1295. https://doi.org/10.21923/jesd.939724
36. Bugdayci, I., Ugurlu, O., & Kunt, F. (2023). Spatial analysis of SO_2, PM_{10}, CO, NO_2, and O_3 pollutants: The case of Konya Province, Turkey. Atmosphere, 14(3), 462. https://doi.org/10.3390/atmos14030462
37. Aydin, N. Y. (2025). Machine learning-based ground-level NO_2 estimation in Istanbul: A comparative analysis of Sentinel-5P and GEOS-CF. Applied Sciences, 15(20), 10997. https://doi.org/10.3390/app152010997
38. Srivastava, A., Thakur, A. K., Garg, R. D., & Garg, P. K. (2025). Prediction of tropospheric nitrogen dioxide in Kolkata using Sentinel-5P and Sentinel-2 multispectral data and machine learning algorithm. Journal of the Indian Society of Remote Sensing. https://doi.org/10.1007/s12524-025-02353-2

 doi:10.20944/preprints202603.0974.v1

39. Grzybowski, P. T., Markowicz, K. M., & Musiał, J. P. (2023). Estimations of the ground-level NO_2 concentrations based on the Sentinel-5P NO_2 tropospheric column number density product. Remote Sensing, 15, 378. https://doi.org/10.3390/rs15020378
40. Cedeno Jimenez, J. R., Pugliese Viloria, A. J., & Brovelli, M. A. (2023). Estimating daily NO_2 ground-level concentrations using Sentinel-5P and ground sensor meteorological measurements. ISPRS International Journal of Geo-Information, 12(3), 107. https://doi.org/10.3390/ijgi12030107
41. Monzón Herrera, Y. R., Polanco Gaytán, M., Aquino Santos, R. T., Babu Saheer, L., Mendoza-Cano, O., & Macedo-Barragán, R. J. (2025). Estimating surface NO_2 in Mexico City using Sentinel-5P and machine learning. Atmosphere, 17, 37. https://doi.org/10.3390/atmos17010037
42. Faisal, M., & Jaelani, L. M. (2023). Spatio-temporal analysis of nitrogen dioxide (NO_2) from Sentinel-5P imageries using Google Earth Engine during the COVID-19 social restriction policy in Jakarta. Natural Hazards Research, 3, 344–352. https://doi.org/10.1016/j.nhres.2023.02.006
43. Srivastava, A., Thakur, A. K., Garg, R. D., & Garg, P. K. (2026). A comprehensive assessment of spatiotemporal dynamics of urban NO_2 using Sentinel-5P with evaluation of the predictive capability of Sentinel-2 for pollution estimation. Water, Air, & Soil Pollution, 237, 382. https://doi.org/10.1007/s11270-026-09083-2