Atmospheric Environment 373 (2026) 121904

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Estimating surface-level NO_2 concentrations in the Madrid region using Sentinel-5P observations and ground-based meteorological data with machine learning approaches

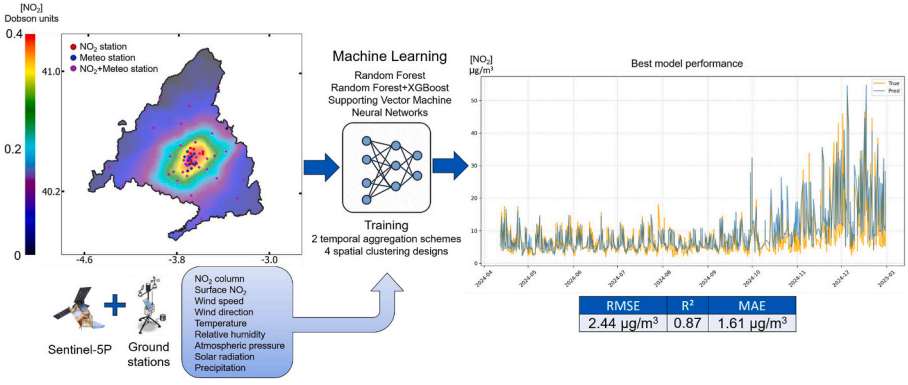
Carlos Morillas [a,*], Rocco Giosa [b], Sergio Alvarez [a], Carmine Serio [b], Guido Masiello [b], Giuliano Liuzzi [b], Carmen Aviles [c], Sara Martinez [a]

[a] Department of Land Morphology and Engineering, Universidad Politécnica de Madrid, Madrid, 28040, Spain
[b] Dipartimento di Ingegneria, Università Degli Studi Della Basilicata, Via Dell'Ateneo Lucano 10, Potenza, PZ, 85100, Italy
[c] Departamento Ingeniería de Organización, Administración de Empresas y Estadística, Universidad Politécnica de Madrid, C/José Antonio Novais, 10, Madrid, 28043, Spain

HIGHLIGHTS

- Surface NO_2 estimated by combining Sentinel-5P and routine meteorological data.
- Dual temporal aggregation captures overpass and antecedent atmospheric conditions.
- Environmentally coherent spatial clustering improves model accuracy.
- High performance achieved without emission inventories or CTM outputs.
- Framework supports air-quality assessment in cities with sparse monitoring networks.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Nitrogen dioxide
Sentinel-5P TROPOMI
Machine learning
Air pollution
Air quality modeling

ABSTRACT

Reliable estimation of surface-level nitrogen dioxide (NO_2) concentrations is critical for air quality assessment in urban regions, where ground-based monitoring networks often provide limited spatial coverage. Satellite observations from Sentinel-5P offer valuable information on tropospheric NO_2 columns, but their translation to surface concentrations remains challenging due to strong spatial heterogeneity and meteorological influences. This study investigates the estimation of daily surface-level NO_2 concentrations across the Community of Madrid (Spain) by combining Sentinel-5P observations with routinely available ground-based meteorological data using machine learning approaches. Four modelling paradigms were evaluated: Random Forest, Support Vector Machine, hybrid ensemble models based on Random Forest and XGBoost, and Artificial Neural Networks. The analysis systematically assessed the influence of temporal and spatial preprocessing by comparing two temporal aggregation strategies—satellite overpass conditions alone and a dual-time window including antecedent atmospheric conditions—and four spatial configurations ranging from regional aggregation to environmentally

* Corresponding author.
E-mail address: c.morillas@upm.es (C. Morillas).

https://doi.org/10.1016/j.atmosenv.2026.121904
Received 23 November 2025; Received in revised form 7 February 2026; Accepted 24 February 2026
Available online 26 February 2026

coherent clustering. Results indicate that incorporating historical meteorological information through an extended temporal window consistently improves predictive performance across all models. Spatial stratification was found to be equally critical: grouping monitoring stations into environmentally coherent clusters substantially outperformed both purely geometric grids and region-wide aggregation. The best-performing configuration, an Artificial Neural Network combined with simplified coherent spatial clusters, achieved an RMSE of 2.44 μg/m^3, an R^2 of 0.87, and a MAE of 1.61 μg/m^3. These findings demonstrate that high predictive accuracy can be achieved through informed temporal and spatial design choices without increasing model complexity or relying on auxiliary emission inventories or chemical transport models, providing a transferable framework for urban air quality monitoring.

1. Introduction

Air quality in large urban areas remains a persistent public health concern. The expansion of urban agglomerations and ongoing economic development intensify anthropogenic activities, leading to increased emissions of atmospheric pollutants. Among them, nitrogen dioxide (NO_2) stands out as one of the most harmful to human health in urban environments. Long-term exposure to NO_2 has been associated to respiratory infections, cardiovascular and immune system diseases, and also has significant effects on ecosystems and cultural heritage (Jonson et al., 2017; Kwon and Cho, 2023; Shaw and Van Heyst, 2022). As urban populations continue to grow, ensuring accurate and spatially representative monitoring of NO_2 concentrations has become an increasingly pressing challenge.

NO_2 in urban areas is primarily generated by road traffic, although residential heating and electricity generation also contribute significantly (McDuffie et al., 2020), especially in colder moths. In many European cities, efforts to transition toward more sustainable urban mobility models are hindered by historical urban planning decisions that prioritized private vehicle use. The case of Madrid, the capital of Spain, is particularly illustrative. Recent epidemiological assessments have identified Madrid as the European city with the highest number of avoidable deaths attributable to long-term NO_2 exposure (Khomenko et al., 2021, 2023), highlighting the critical need for effective air quality assessment and mitigation strategies.

This urgency is further reinforced by increasingly stringent regulatory requirements. Current European legislation sets a NO_2 hourly concentration threshold of 200 μg/m^3 (not to be exceeded more than 18 times per year) and an annual limit of 40 μg/m^3 (European Commission, 2008). According to the Spanish National Air Quality Reports, published since 2001, Madrid consistently exceeded the hourly threshold from 2004 to 2019, and only met the annual standard for the first time in 2022 (MITERD, 2025). The recent Directive (EU) 2024/2881 further tightens the annual limit to 20 μg/m^3 by 2030 (European Commission, 2024), while the World Health Organization (WHO) recommends a stricter guideline of 10 μg/m^3 (WHO, 2021). At the national level, Spain's Climate Change and Energy Transition Law July 2021 requires municipalities with over 50,000 inhabitants to implement sustainable urban mobility plans, including low-emission zones (Morillas et al., 2024a) and other mitigation measures (Government of Spain, 2021), increasing the demand for reliable tools to evaluate policy effectiveness.

Assessing compliance with these regulations and the effectiveness of mitigation measures relies heavily on robust and well-distributed air quality monitoring infrastructure. In the Community of Madrid, which includes 24 municipalities with more than 50,000 inhabitants, the current network consists of 52 air quality monitoring stations: 24 within the city of Madrid and 28 distributed across surrounding municipalities Since 2022, this network has been expanded with four additional air quality monitoring sites. In parallel, air quality observations are supported by a complementary network of 54 meteorological stations distributed throughout the region. While these reference-grade stations provide high-quality measurements, they are costly to install and maintain. A fully equipped station compliant with European Air Quality Index standards typically requires an investment between €100,000 and €250,000, excluding ancillary infrastructure and maintenance. As a result, across Spain, 99 out of 152 municipalities with more than 50,000 inhabitants operate only one monitoring station or none at all, limiting the spatial representativeness of air quality assessments (INE, 2024; MITERD, 2024). This context underscores the urgency of developing complementary solutions that can be implemented promptly without requiring substantial additional public expenditure.

In this context, satellite remote sensing has emerged as a valuable complementary source of air quality information. Since 2018, the Sentinel-5P satellite, which is part of the European Union's Copernicus programme, has provided daily observations of several atmospheric pollutants with unprecedented spatial resolution. Tropospheric NO_2 is of particular relevance, as it has shown strong correlations with ground-based measurements and has become a key variable in studies aimed at identifying emission hotspots and characterizing spatial and temporal pollution pattern (Cersosimo et al., 2020; Garcia Santos and Parés, 2025; Goldberg et al., 2021; Griffin et al., 2019; Jeong and Hong, 2021; Morillas et al., 2024b; Rudke et al., 2023). The relatively short atmospheric lifetime of NO_2 (ranging from approximately 2 h in summer to 12 h in winter), combined with its distinct spectral signature, makes it especially suitable for satellite detection near emission sources (Lange et al., 2022; Laughner and Cohen, 2019; Shah et al., 2020).

In the Community of Madrid, these properties translate into pronounced temporal and spatial variability. NO_2 concentrations typically increase during colder months due to enhanced atmospheric stability and higher heating demand, while stronger photochemical removal processes dominate in summer. Clear weekly cycles are also observed, with higher concentrations on weekdays and marked reductions during weekends, particularly on Sundays, reflecting traffic-related emission patterns (Morillas et al., 2024b, 2025). These temporal and spatial dynamics partly explain the strong correlations reported between satellite-derived and ground-based NO_2 in Madrid city ($r = 0.86$, Morillas et al., 2025), likely due to local topography; a relationship further modulated by local meteorological conditions such as wind speed, wind direction, temperature, and solar radiation (Rahman et al., 2022; Zhang et al., 2015).

Building on these advances, numerous studies have integrated Sentinel-5P observations into atmospheric models and machine learning frameworks to estimate surface-level NO_2 concentrations. Regression-based and machine learning approaches have gained prominence due to their flexibility and computational efficiency, particularly in urban environments (Balamurugan et al., 2023; Cao, 2023; Cedeno Jimenez et al., 2023; Chan et al., 2021; Das et al., 2021; Deng et al., 2022; Grzybowski et al., 2023; He et al., 2022; Kim et al., 2021; Naseer et al., 2022; Shetty et al., 2024; Xing et al., 2024). However, existing studies often rely on complex ancillary datasets (e.g., traffic intensity, land use, or emission inventories), apply limited temporal aggregation strategies focused solely on satellite overpass hours, or adopt simplified spatial treatments that might not represent environmental heterogeneity. As a result, the relative influence of temporal averaging choices and spatial stratification strategies on model performance remains insufficiently explored. This is particularly relevant in heterogeneous urban regions, where emission sources, topography, and meteorological regimes vary substantially over short distances. These limitations suggest that a

one-size-fits-all training strategy may be suboptimal, and that exploring region-specific temporal and spatial training configurations could represent a key pathway for improving the accuracy and robustness of surface-level NO_2 estimation models.

This study addresses the following core questions: (i) to what extent can surface-level NO_2 concentrations be accurately estimated by combining Sentinel-5P observations with ground-based meteorological data alone; (ii) how do different temporal aggregation strategies, including the incorporation of historical atmospheric information, affect predictive performance; and (iii) how does spatial stratification influence model accuracy in a geographically and environmentally heterogeneous region.

Four modelling approaches were implemented and compared: (i) a Random Forest (RF) model used both as an independent predictor and as a feature selection tool; (ii) a hybrid framework combining RF and eXtreme Gradient Boosting (XGBoost), including a stacking configuration and an XGBoost model optimized through Bayesian hyperparameter tuning; (iii) a Support Vector Machine (SVM) regression model to assess the performance of kernel-based methods under different spatial stratifications; and (iv) an Artificial Neural Network (ANN) designed to capture highly nonlinear relationships between predictors and surface-level NO_2 concentrations. All modelling approaches were systematically evaluated under two temporal configurations and four spatial strategies.

The ability to estimate surface NO_2 concentrations from satellite and meteorological data offers the potential to enhance existing monitoring networks, optimize resource allocation, and support more informed air quality policymaking. Once validated in a complex urban setting such as Madrid, the proposed methodology could be extended to other areas of Spain—especially to municipalities with over 50,000 inhabitants that lack reference-grade stations or possess insufficient infrastructure to reliably monitor pollution levels. Such expansion would represent a significant step toward more equitable, efficient, and data-driven environmental governance.

2. Materials and methods

The study area encompasses the Community of Madrid, located in the Central Plateau of the Iberian Peninsula. It comprises 179 municipalities, including the city of Madrid, which serves as both the regional and national capital of Spain. Madrid is situated in a mountainous basin at the foothills of the Central System, and it gathers the highest intensity of anthropogenic activity in Spain.

The analysis spans the period from January 1, 2020, to December 31, 2024, covering full calendar years with satellite imagery of consistent spatial resolution (see Section 2.1).

2.1. NO_2 data

Satellite NO_2 data used in this study were obtained from the TROPOMI sensor onboard the Sentinel-5P satellite. Operational since June 26, 2018, Sentinel-5P follows a sun-synchronous orbit of 824 km altitude, crossing the equator at approximately 13:30 local solar time (Veefkind et al., 2012). The satellite completes 14 orbits per day, enabling daily global coverage of the tropospheric NO_2 column. Overpasses over the Iberian Peninsula typically occur between 11:00 and 14:00 (hereafter expressed in UTC). Due to the sensor's wide swath width of 2600 km, the region may occasionally be observed twice within the same day by consecutive orbits, separated by approximately 100 min. The native spatial resolution of TROPOMI was improved in August 2019 from $3.5 \times 7.5\ km^2$ to $3.5 \times 5.5\ km^2$ at nadir (Van Geffen et al., 2019).

TROPOMI NO_2 data (Table 1) were accessed via Google Earth Engine (GEE), Google's cloud-based geospatial analysis platform. GEE facilitates the processing of large volumes of satellite data without the need for local storage, offering direct access to pre-integrated and preprocessed datasets. The original TROPOMI data are Level-2 products, georeferenced at the pixel level but not aligned to a regular spatial grid. For improved usability within GEE, these datasets are reprocessed to Level-3, involving reprojection and resampling onto a uniform geographic grid (latitude-longitude) with a grid cell size of $0.01^\circ \times 0.01^\circ$. Additionally, quality filters ($QA > 0.75$) are applied to discard unreliable observations, such as those affected by cloud cover or adverse atmospheric conditions. Among the available products—near-real time (NRTI) and offline (OFFL)—the OFFL images were selected for their improved data quality (GEE, 2025).

Table 1
NO_2 data used in the study.

Variable	Unit	Source
Tropospheric NO_2 concentration	$moles/m^2$	Sentinel-5P TROPOMI
Surface NO_2 concentration	$\mu g/m^3$	Regional air quality stations

Of the 1827 days covered by the study period, and after applying the quality filters described above, at least one valid NO_2 observation was obtained for 1573 days, resulting in a total of 2018 Sentinel-5P OFFL products for model training and validation. Of these, 949 (47.0%) corresponded to the October–March period and 1069 (53.0%), largely attributable to reduced cloud cover and improved observation conditions during the warmer months.

Surface-level NO_2 concentrations were obtained from the air quality monitoring network of the Community of Madrid. To ensure temporal continuity and consistency of the dataset, only stations that remained fully operational throughout the entire study period were included. This selection resulted in 24 stations located within the city of Madrid (City of Madrid, 2025a) and 24 stations distributed across the rest of the region (Community of Madrid, 2025a). All devices perform hourly NO_2 measurements, which are publicly available through official open-access platforms (Table 1). The spatial distribution of the air quality network is presented along with meteorological stations in Section 2.2.

2.2. Meteorological data

Meteorological data were obtained from the meteorological stations network operated by the city (City of Madrid, 2025b) and the Community of Madrid (Community of Madrid, 2025b). Following the same criteria applied to the air quality monitoring stations, only stations that were fully operational throughout the entire study period were included. This selection resulted in 26 stations located within the city of Madrid and 24 across the surrounding region. These stations record up to seven meteorological variables and provide hourly measurements, which are publicly accessible through official open-access platforms (Table 2).

The spatial distribution of the monitoring stations and the specific variables measured are presented in Fig. 1 and Table 3.

2.3. Data processing

Before training the prediction models, all input variables were preprocessed to align temporally with the Sentinel-5P satellite overpasses. Consistent with approaches explored in earlier work, two temporal averaging windows were applied to the meteorological variables. First,

Table 2
Meteorological data used in the study.

Variable	Unit	Source
Wind speed	m/s	Regional meteo stations
Wind direction	°	Regional meteo stations
Temperature	°C	Regional meteo stations
Relative humidity	%	Regional meteo stations
Atmospheric pressure	mbar	Regional meteo stations
Solar radiation	W/m^2	Regional meteo stations
Precipitation	l/m^2	Regional meteo stations

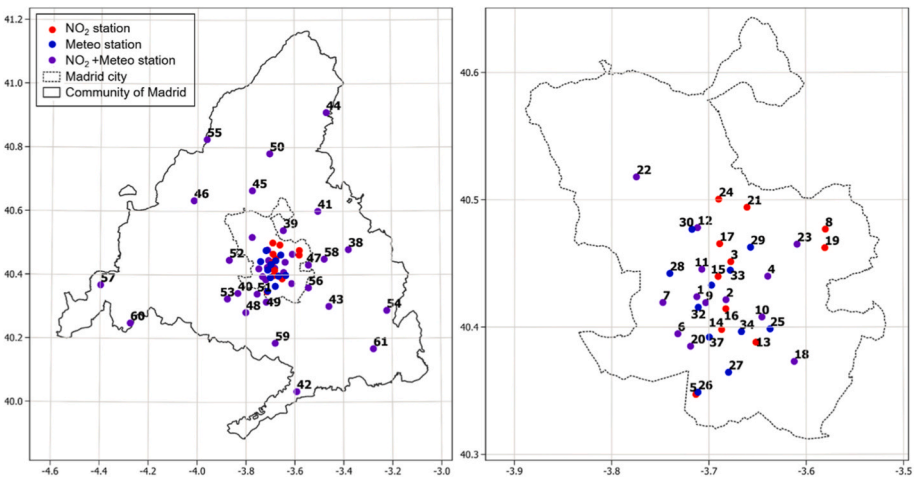


Fig. 1. Air quality (NO_2) and meteorological monitoring networks in the Community of Madrid.

the average of all variables was computed over the 11:00-14:00 time interval, which corresponds to the typical overpass window of Sentinel-5P over the study region (hereafter referred to as H1). Second, an additional average was calculated for the 21-h period immediately preceding the overpass, specifically from 14:00 of the previous day to 11:00 of the current day (H2). This second aggregation captures the cumulative atmospheric conditions that may influence surface-level NO_2 concentrations at the time of satellite observation. This dual-averaging strategy has been shown to reduce short-term variability and enhance model performance by providing a more robust representation of the physical and chemical dynamics of NO_2 within the planetary boundary layer (Cedeno Jimenez et al., 2023). To further assess this assumption, all models were trained under two different temporal configurations: with (H1+H2) and without (H1 only) the inclusion of the extended historical window.

To account for temporal variability in NO_2 concentrations, several calendar-based features were incorporated as additional explanatory variables. The year was included to capture interannual variability, while the day of the year and the day of the week were encoded as cyclic variables using sine and cosine transformations. In addition, all categorical predictors were one-hot encoded. This preprocessing ensured a consistent numerical representation of periodic patterns and discrete factors, preserved the inherent annual and weekly cycles of NO_2, and avoided the introduction of artificial temporal discontinuities.

For the wind-related variables (wind direction and speed), a similar preprocessing was applied. Hourly wind speed w_i and direction θ_i measurements were aggregated over the same two temporal intervals used for the meteorological variables. Within each interval, the mean wind direction was computed using a vectorial averaging method, weighting each direction by its corresponding speed. Specifically, the Cartesian components (u_i, w_i) of the wind vectors were calculated as:

$$u_i = w_i \cos(\theta_i), v_i = w_i \sin(\theta_i)$$

The mean vector components (u_s, w_s) over the interval were then computed:

$$u_s = \frac{1}{N}\sum_{i=1}^{N} u_i, v_s = \frac{1}{N}\sum_{i=1}^{N} v_i$$

where N denotes the number of hourly wind measurements available within the corresponding temporal interval.

The resulting mean direction was obtained as:

$$\theta_{mean} = \arctan(u_s, v_s)$$

Finally, the mean direction was transformed into continuous sine and cosine components providing a smooth numerical representation suitable for statistical and machine learning models.

$$\cos(\theta_{mean}), \sin(\theta_{mean})$$

A summary of all input features used in the modeling process is provided in Table 4.

The dataset was split into training (70%), validation (15%), and test (15%) subsets. To preserve temporal continuity in the evaluation phase, the test set was extracted exclusively from the final portion of 2024 (approximately April to December). This approach ensured that the test data formed a continuous time series, which was necessary for generating the temporal plots and diagnostic figures.

While the temporal partitioning captures seasonal variability, spatial heterogeneity was addressed separately through alternative territorial grouping strategies. Monitoring stations were assigned to spatial sets based on predefined geographical or environmental criteria, without aggregating or averaging station-level observations, in order to evaluate the influence of spatial stratification on model performance.

Based on this rationale, four territorial approaches were implemented and compared, as described below.

2.3.1. *Regional aggregation approach*

All monitoring stations across the Community of Madrid (Fig. 1) were treated as a single regional domain, without applying any spatial grouping or aggregation of station-level observations. Under this configuration, the study area was considered as a homogeneous spatial unit for model training.

2.3.2. *Grid-based approach*

The study area was divided into a regular grid of square bounding boxes, each measuring 5 km per side. This grid size was selected to be consistent with the effective spatial resolution of TROPOMI tropospheric NO_2 column data ($\sim 3.5 \times 5.5\ km^2$), representing the smallest spatial scale at which satellite-derived information can be considered physically meaningful.

Each cell was assessed to determine whether it contained at least one meteorological station and one NO_2 monitoring station. Only those meeting this dual-availability criterion were retained for analysis, resulting in a total of 33 cells, as shown in Fig. 2.

Building on this procedure, satellite data were retrieved from the full rectangular extent of each bounding box. Satellite-derived NO_2 values were then collocated with the corresponding station-level observations based on their spatial assignment, producing the final input dataset used for training and validation of the prediction models.

2.3.3. *Sector-based approach*

The study area was divided into seven broader geographic regions

Table 3
Variables measured by each monitoring station in the Community of Madrid. Numbers in the first column correspond to those displayed in Fig. 1. NO_2 refers to nitrogen dioxide concentration; WS to wind speed; WD to wind direction; T to temperature; RH to relative humidity; AP to atmospheric pressure; SR to solar radiation; and P to precipitation.

Station		NO_2	WS	WD	T	RH	AP	SR	P
1	Plaza España	Yes	No	No	Yes	No	No	No	No
2	Escuelas Aguirre	Yes	No	No	Yes	Yes	No	No	No
3	Ramón y Cajal	Yes	No	No	No	No	No	No	No
4	Arturo Soria	Yes	No	No	Yes	Yes	No	No	No
5	Villaverde	Yes	No	No	No	No	No	No	No
6	Farolillo	Yes	No	No	Yes	No	No	No	No
7	Casa de Campo	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8	Barajas Pueblo	Yes	No	No	No	No	No	No	No
9	Plaza del Carmen	Yes	No	No	Yes	Yes	No	No	No
10	Moratalaz	Yes	No	No	Yes	Yes	No	No	No
11	Cuatro Caminos	Yes	No	No	Yes	Yes	No	No	No
12	Barrio del Pilar	Yes	No	No	Yes	Yes	No	No	Yes
13	Vallecas	Yes	No	No	No	No	No	No	No
14	Méndez Álvaro	Yes	No	No	No	No	No	No	No
15	Castellana	Yes	No	No	No	No	No	No	No
16	Retiro	Yes	No	No	No	No	No	No	No
17	Plaza Castilla	Yes	No	No	No	No	No	No	No
18	Ensanche de Vallecas	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
19	Urbanización Embajada	Yes	No	No	No	No	No	No	No
20	Plaza Elíptica	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
21	Sanchinarro	Yes	No	No	No	No	No	No	No
22	El Pardo	Yes	No	No	Yes	Yes	No	No	No
23	Juan Carlos I	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
24	Tres Olivos	Yes	No	No	No	No	No	No	No
25	J.M.D. Moratalaz	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
26	J.M.D. Villaverde	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
27	E.D.A.R. La China	No	Yes	Yes	No	No	No	No	No
28	C.M. Acústica	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
29	J.M.D. Hortaleza	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
30	Peñagrande	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
31	J.M.D. Chamberí	No	No	No	Yes	Yes	No	No	No
32	J.M.D. Centro	No	No	No	Yes	Yes	No	No	No
33	J.M.D. Chamartín	No	No	No	Yes	Yes	No	No	No
34	J.M.D. Vallecas 1	No	No	No	Yes	Yes	No	No	No
35	J.M.D. Vallecas 2	No	No	No	Yes	Yes	No	No	No
36	Matadero 1	No	No	No	Yes	Yes	No	No	No
37	Matadero 2	No	No	No	Yes	Yes	No	No	No
38	Alcalá de Henares	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
39	Alcobendas	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
40	Alcorcón	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
41	Algete	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
42	Aranjuez	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
43	Arganda del Rey	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
44	El Atazar	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
45	Colmenar Viejo	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
46	Collado Villalba	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
47	Coslada	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
48	Fuenlabrada	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
49	Getafe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
50	Guadalix de la Sierra	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
51	Leganés	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
52	Majadahonda	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
53	Móstoles	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
54	Orusco de Tajuña	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
55	Puerto de Cotos	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
56	Rivas-Vaciamadrid	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
57	San Marín de Valdeiglesias	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
58	Torrejón de Ardoz	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
59	Valdemoro	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
60	Villa del Prado	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
61	Villarejo de Salvanés	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

based on topographic, meteorological, and urbanization patterns known to influence NO_2 dynamics across the Community of Madrid (Fig. 3).

The northern sector encompasses the Sierra de Guadarrama and adjacent high-altitude areas, characterized by steep topographic gradients, elevations frequently exceeding 1500 m, low population density, and a sparse road network. In this sector, NO_2 levels are typically governed by regional background conditions and enhanced atmospheric mixing, with limited influence from local traffic emissions.

By contrast, the central sectors correspond to the metropolitan core of Madrid, where population density and road network intensity are highest. This area is characterized by a dense urban fabric and a concentration of major traffic corridors, which give rise to strong and highly localized NO_2 emissions. Consequently, the monitoring network is densest within the city, reflecting the need to capture pronounced spatial gradients in air composition occurring over relatively short distances.

The eastern sectors are characterized by predominantly flat topography, extensive suburban and peri-urban development, and a dense

Table 4
Description of input variables used for estimating surface-level NO_2 concentrations.

Variable	Description
Tropospheric NO_2 concentration	Tropospheric NO_2 column measured at the time of Sentinel-5P overpass
Mean surface NO_2 concentration H1	Average surface NO_2 concentration between 11:00 and 14:00
Mean surface NO_2 concentration H2	Average surface NO_2 concentration between 14:00 of the previous day and 11:00 of the current day
u_s H1	Average horizontal wind speed between 11:00 and 14:00
u_s H2	Average horizontal wind speed between 14:00 of the previous day and 11:00 of the current day
v_s H1	Average vertical wind speed between 11:00 and 14:00
v_s H2	Average vertical wind speed between 14:00 of the previous day and 11:00 of the current day
cos (θ_{mean}) H1	Horizontal component of average wind direction between 11:00 and 14:00
cos (θ_{mean}) H2	Horizontal component of average wind direction between 14:00 of the previous day and 11:00 of the current day
sin (θ_{mean}) H1	Vertical component of average wind direction between 11:00 and 14:00
sin (θ_{mean}) H2	Vertical component of average wind direction between 14:00 of the previous day and 11:00 of the current day
Temperature H1	Average temperature between 11:00 and 14:00
Temperature H2	Average temperature between 14:00 of the previous day and 11:00 of the current day
Relative humidity H1	Average relative humidity between 11:00 and 14:00
Relative humidity H2	Average relative humidity between 14:00 of the previous day and 11:00 of the current day
Solar radiation H1	Average solar radiation between 11:00 and 14:00
Solar radiation H2	Average solar radiation between 14:00 of the previous day and 11:00 of the current day
Precipitation H1	Average precipitation between 11:00 and 14:00
Precipitation H2	Average precipitation between 14:00 of the previous day and 11:00 of the current day
Pressure H1	Average atmospheric pressure between 11:00 and 14:00
Pressure H2	Average atmospheric pressure between 14:00 of the previous day and 11:00 of the current day
Year	One-hot encoding
Day of the year	Day of the year (sin, cos of month of year)
Day of the week	Day of the week (sin, cos of month of day of week)

network of radial and orbital highways linking residential areas with the metropolitan core. These regions have intermediate emission intensities, with NO_2 variability largely driven by commuter traffic and regional transport processes rather than point-like urban sources.

The southern and western sectors represent transitional zones between the urban core and the surrounding rural or mountainous areas. These regions combine moderate population densities, mixed land-use patterns, and heterogeneous road infrastructures, leading to spatially variable NO_2 concentrations influenced by both local emissions and broader meteorological transport.

This regionalization strategy was designed to match with previous findings for the Madrid region, which show substantial spatial variability in the relationship between satellite-derived and surface-level NO_2 measurements. In particular, rural and high-elevation stations exhibit markedly lower satellite-ground correlations due to enhanced atmospheric mixing, reduced emission intensity, and stronger topographic influences, whereas urban and suburban stations display higher coherence between both data sources (e.g., urban $r \approx 0.79$ vs. rural $r \approx 0.52$; Morillas et al., 2024b). These documented contrasts were considered when defining the broader geographical sectors, allowing monitoring stations to be grouped according to shared atmospheric dispersion conditions, emission regimes, and planetary boundary-layer dynamics.

Following the same procedure described previously, monitoring stations were assigned to their corresponding bounding boxes, and station-level observations were paired with the associated satellite-derived NO_2 column.

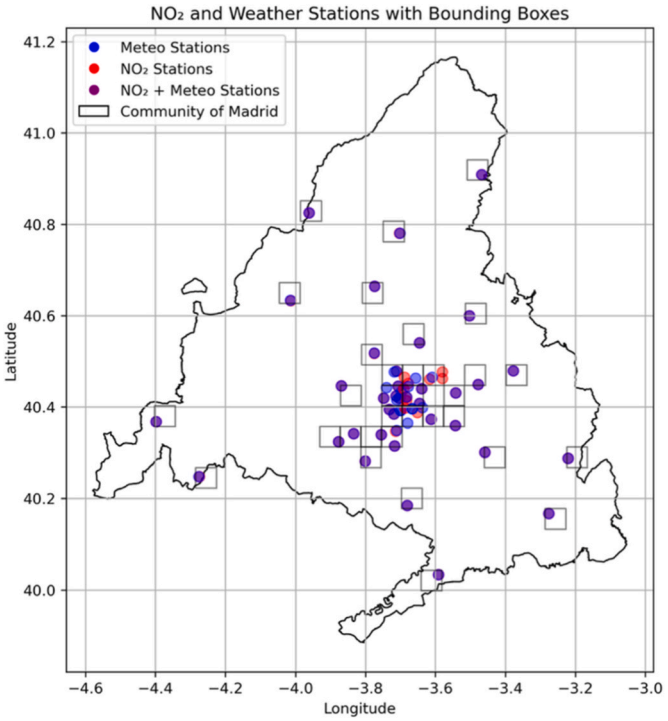


Fig. 2. Spatial distribution of the $5 \times 5\,km^2$ grid cells retained for analysis.

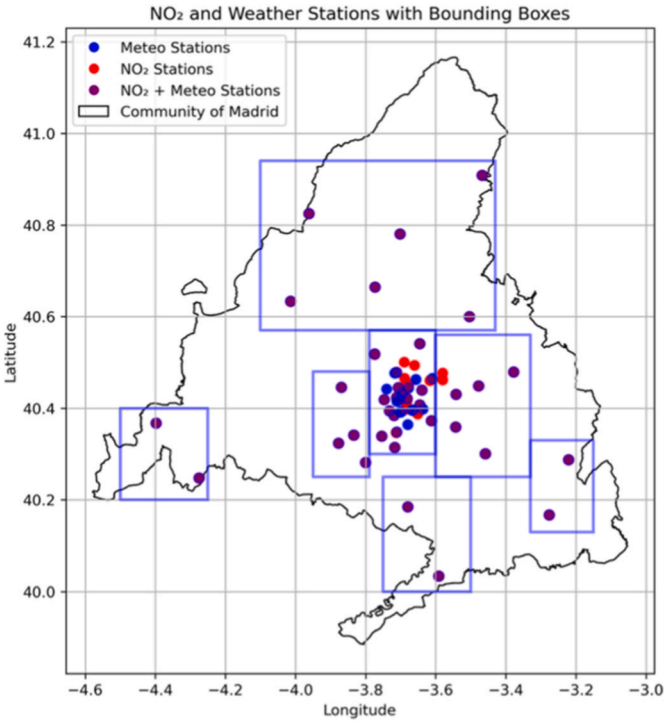


Fig. 3. Spatial clusters used in the sector-based approach.

2.3.4. *Simplified coherent clusters approach*

In line with the previous approach, the study area was further simplified into three spatially coherent clusters defined by contiguous bounding boxes that encompass the main concentration of monitoring stations and dominant emission environments (Fig. 4). This configuration was designed to reflect the strong center-periphery gradient

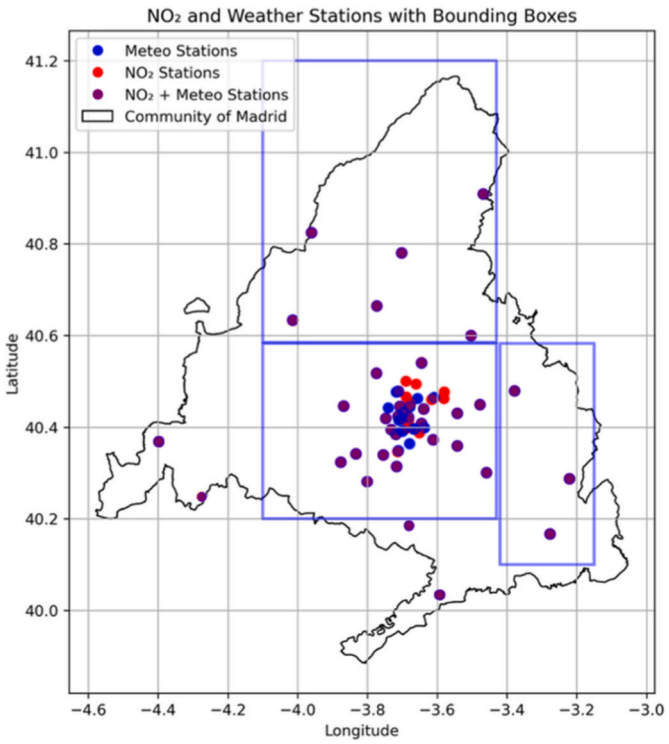


Fig. 4. Spatial clusters used in the simplified coherent clusters approach.

observed across the Community of Madrid, while avoiding excessive spatial fragmentation.

The central cluster covers the metropolitan core, where monitoring stations are densely distributed and where traffic-related emissions, population density, and road network intensity are highest. The northern cluster encompasses predominantly mountainous and rural areas, characterized by higher elevations, lower population density, and distinct meteorological conditions that influence pollutant dispersion and satellite sensitivity. The eastern cluster includes municipalities with flatter topography and a higher degree of urbanization, but located away from the metropolitan core.

In order to increase internal homogeneity, a limited number of monitoring stations were excluded from this configuration because they are located in transitional zones that cannot be consistently associated with any of the three clusters. These stations are influenced by overlapping environmental conditions (e.g., intermediate elevation ranges, mixed urban-rural land use) and their exclusion therefore represents a deliberate trade-off aimed at prioritizing environmental consistency over maximizing data availability.

Comprehensive information on the monitoring stations used in this study, including the variables measured at each site, their functional zone coverage, their geographical location (latitude and longitude), elevation above sea level, and their classification under the different spatial strategies considered, is provided as Supplementary Information (Table S1). In addition, to facilitate interpretation of the station groupings and to better illustrate the rationale behind the different spatial stratification approaches, a supplementary figure is included showing the geographical distribution of monitoring stations overlaid with major traffic corridors, regional topography, and population density across the study area (Fig. S1).

2.4. Machine learning models

The machine learning and deep learning models described in this section were applied to estimate surface-level NO_2 concentrations across the different spatial configurations. To enable a consistent comparison of predictive performance, the models were evaluated using the Root Mean Squared Error (RMSE), which quantifies the overall magnitude of prediction errors; the coefficient of determination (R^2), which measures the proportion of variance explained by the model; and the Mean Absolute Error (MAE), which represents the average absolute deviation between predicted and observed values.

2.4.1. Random Forest

RF was selected as the baseline machine learning model for estimating surface-level NO_2 concentrations. RF is an ensemble learning algorithm that constructs multiple decision trees using bootstrap sampling and random feature selection at each split, and aggregates their predictions to reduce variance and improve generalisation performance (Breiman, 2001).

In this implementation, the RF regressor was trained using 300 decision trees. At each node, a random subset of predictors was considered to introduce diversity among trees and mitigate overfitting. Final predictions were obtained by averaging the outputs of all individual trees. This configuration enables the model to capture nonlinear relationships and complex interactions between satellite-derived, meteorological, temporal, and spatial predictors, while maintaining a relatively low computational cost.

2.4.2. Random Forest and eXtreme Gradient Boosting

In this hybrid configuration, the modelling of NO_2 concentrations was carried out using a two-step strategy combining tree-based methods with gradient-boosting algorithms to enhance robustness and predictive accuracy. In the first step, a RF was used primarily as a feature selection tool.

The RF model was trained using 100 bootstrap-sampled trees, each built on a random subset of features at every split. After training, feature importance scores were derived from the mean decrease in impurity. To avoid arbitrary thresholds, only the variables above the 75th percentile of the importance distribution were retained. This approach removed redundant or weak predictors while improving stability, reducing noise, and lowering overall computational cost.

The selected variables were then used as input for XGBoost (Chen and Guestrin, 2016; Friedman, 2001), which served as the final predictive model. In addition to the standard training procedure, two further optimisation steps were introduced. First, a stacking architecture combining RF and XGBoost was implemented to leverage the complementary strengths of both algorithms: RF provides stable, low-variance predictions, whereas XGBoost captures more complex nonlinear relationships through its incremental boosting framework. Second, XGBoost hyperparameters were tuned using Optuna (Akiba et al., 2019) with a Bayesian optimisation strategy over 100 iterations. The search process was guided by minimization of the RMSE, and the resulting optimal parameter set was adopted for the final model configuration.

2.4.3. Support Vector Machine

SVM was implemented as a kernel-based machine learning approach to estimate surface-level NO_2 concentrations. SVM is grounded in the principle of Structural Risk Minimization and aims to identify a regression function that deviates from the observed targets by no more than a predefined margin ε, while maintaining maximal flatness to promote generalisation performance.

In this study, an ε-SVR formulation with a Radial Basis Function (RBF) kernel was adopted to capture nonlinear relationships between predictors and NO_2 concentrations. Prior to model training, the input data underwent the same preprocessing pipeline applied to the other machine learning models, including one-hot encoding of categorical variables and sinusoidal transformations of cyclical temporal variables to preserve seasonal continuity.

To reduce dimensionality and mitigate the sensitivity of SVM to noisy or redundant predictors, a preliminary feature selection step was performed using a RF regressor. Feature importance scores were

derived, and only variables above the 75th percentile were retained for SVM training. All selected features were subsequently standardised using a z-score normalisation to ensure comparable scaling across predictors.

The SVM model was configured with a regularisation parameter $C = 10$ and an ε-insensitive loss margin of 0.1.

2.4.4. Artificial neural network

An ANN was implemented to capture the highly nonlinear dependencies between meteorological, temporal and chemical variables and ground-level NO_2 concentrations. In this case, the preprocessing phase consisted of normalising all variables using StandardScaler, fitted only on the training data to prevent information leakage. The standardisation was then applied to both validation and test sets to ensure consistency throughout the modelling pipeline.

Unlike the tree-based framework, no feature selection was applied prior to training the ANN. All available predictors were passed to the network, allowing it to autonomously learn which variables were most relevant. This approach leverages the internal weight-learning mechanism of neural networks, which makes them inherently capable of identifying useful patterns in high-dimensional spaces without requiring explicit dimensionality reduction.

The network architecture was not manually designed but generated through AutoKeras, an AutoML tool that explores a broad space of possible structures. The model specification included an input node, a DenseBlock composed of various combinations of fully connected layers and activation functions, and a RegressionHead for the output. AutoKeras evaluated up to thirty different configurations and selected the best performing one based on the validation loss. The resulting model was then exported and further refined through a staged fine-tuning process. Over ten incremental cycles, each comprising twenty additional training epochs, the network was retrained and re-evaluated while systematically monitoring signs of overfitting. This procedure helped identify the point at which the model achieved optimal predictive performance without compromising generalisation.

The network proved particularly effective at reproducing local nonlinear variations in NO_2 concentrations, demonstrating its ability to capture patterns that are more difficult for tree-based methods to model. The combination of AutoML-driven architecture search and incremental optimisation produced a stable and well-adapted neural model.

3. Results

3.1. Model performance

Table 5 summarises the predictive performance of all evaluated models, including RF, SVM, ANN, and the hybrid ensemble configurations combining RF and XGBoost (stacking model and XGBoost optimized model), across the different spatial classification approaches.

As described in Section 2.3, all models were trained under two temporal configurations: using both the satellite overpass window (H1) and the extended historical window (H2), and using the overpass window alone (H1 only). Across all models and spatial configurations, the inclusion of the extended historical window (H1+H2) consistently resulted in improved predictive performance compared to training with H1 alone. For this reason, Table 5 reports the results obtained using the H1+H2 configuration, while the complete set of results, including those based solely on H1, is provided as Supplementary Information (Table S2).

The tree-based feature selection consistently retained Sentinel-5P tropospheric NO_2 columns and key meteorological drivers of dispersion and atmospheric stability, including wind speed, surface temperature, solar radiation and surface pressure. The retained feature set is consistent with the dominant processes governing NO_2 accumulation and dispersion in urban environments, and aligns with previous interpretability analyses conducted for the Madrid region (Morillas et al., 2025).

Table 5
Predictive performance of machine learning models across spatial classification approaches.

Model	Approach	Epochs	RMSE	R^2	MAE
Random Forest	Regional aggregation approach	-	3.85	0.81	2.63
	Grid-based approach	-	4.78	0.82	3.12
	Sector-based approach	-	3.15	0.88	2.07
	Simplified coherent clusters	-	3.25	0.88	2.02
Stacking Model (RF + XGBoost)	Regional aggregation approach	-	4.25	0.81	2.57
	Grid-based approach	-	4.80	0.82	3.10
	Sector-based approach	-	3.24	0.87	2.07
	Simplified coherent clusters	-	3.43	0.86	2.05
Hybrid XGBoost Optimized	Regional aggregation approach	-	3.79	0.81	2.57
	Grid-based approach	-	4.51	0.84	2.95
	Sector-based approach	-	2.93	0.90	1.95
	Simplified coherent clusters	-	3.08	0.89	1.95
Support Vector Machine	Regional aggregation approach	-	3.60	0.83	2.39
	Grid-based approach	-	4.96	0.81	3.09
	Sector-based approach	-	2.90	0.90	1.90
	Simplified coherent clusters	-	3.79	0.83	1.99
Artificial Neural Network	Regional aggregation approach	20	3.54	0.77	2.41
	Grid-based approach	30	4.56	0.76	2.87
	Sector-based approach	30	2.78	0.86	1.67
	Simplified coherent clusters	30	2.44	0.87	1.61

For the ANN models, multiple training configurations with varying numbers of epochs were systematically tested within each spatial classification strategy to identify the optimal level of convergence. Accordingly, Table 5 reports the number of training epochs corresponding to the best-performing ANN configuration in each case.

Overall, ANN achieved the best predictive performance among the evaluated models under the most favorable spatial configurations. In particular, ANN models yielded the lowest RMSE values (down to 2.44 $\mu g/m^3$) and the highest coefficients of determination (up to $R^2 = 0.87$) when combined with the simplified coherent clusters approach. Hybrid ensemble models based on tree-based methods, especially the XGBoost optimized configuration, also exhibited strong performance, reaching RMSE values as low as 2.93 $\mu g/m^3$ and R^2 values up to 0.90 under the sector-based spatial classification.

These results indicate that, under the conditions tested in this study, ANN models exhibit a greater capacity to capture complex nonlinear relationships between atmospheric predictors and surface-level NO_2 concentrations. Nevertheless, ensemble tree-based approaches provide competitive accuracy with lower computational complexity, making them a viable alternative.

3.2. Best performing approach

The ANN model combined with the simplified coherent clusters approach achieved the best overall performance (RMSE = 2.44 $\mu g/m^3$, $R^2 = 0.87$, MAE = 1.61 $\mu g/m^3$), as shown in Table 5. The model was trained under multiple epoch configurations, progressively increasing the number of passes through the training data to assess their impact on

predictive performance. The 30-epoch configuration provided the best balance between predictive accuracy and generalization, yielding the lowest RMSE and MAE and the highest R^2 among all configurations tested. As illustrated in Fig. 5, model performance improved steadily up to this point, whereas additional training led to mild degradation due to overfitting.

To further assess the spatial consistency of model performance, predictive accuracy was evaluated separately for each bounding box within the simplified coherent clusters framework. RMSE, R^2 and MAE were computed at the bounding-box level using the test dataset, allowing the identification of spatial differences in model behavior (Table 6).

Fig. 6 presents the distribution of prediction errors obtained in this case. The error histogram is approximately centered around zero, with most residuals falling within the range of ±5 μg/m^3. The near-symmetry of the distribution indicates the abscence of systematic bias toward overestimation or underestimation and confirms its robustness and reliability. However, the presence of a few extreme values in the tails reflects occasional deviations during specific episodes, likely associated with atypical meteorological conditions or short-term emission peaks.

Fig. 7 shows the scatter plot of observed versus predicted NO_2 concentrations. Overall, predicted concentrations show a clear linear association with observed values, with most data points clustering close to the 1:1 reference line. This alignment is evident at low to moderate NO_2 concentrations, which constitute the majority of the dataset (0-15 μg/m^3).

In the intermediate concentration range (15–35 μg/m^3), predicted values tend to be slightly higher than the corresponding observations, as reflected by the increased number of points located above the 1:1 line. At higher observed concentrations (>35 μg/m^3), the pattern shifts, with predicted values increasingly falling below the reference line, indicating a tendency toward underestimation at the upper end of the concentration range.

Despite this deviation at higher concentrations, the overall slope of the observed–predicted relationship remains close to unity, and the dispersion around the reference line remains moderate across the full concentration range. These results indicate that the model effectively captures the dominant variability in surface-level NO_2 concentrations across the study area, while concentration-dependent differences in predictive behavior become apparent.

The time series comparison shown in Fig. 8 illustrates the temporal evolution of observed and predicted NO_2 concentrations over the test period. As described in Section 2.3, the year 2024 was reserved exclusively for model evaluation in order to preserve temporal continuity in the test set. Consequently, the time series displayed in Fig. 8 corresponds to the available portion of 2024, spanning from April to December. Despite not covering the full calendar year, this period includes episodes of both low and high NO_2 concentrations, providing a representative range of pollution conditions.

Table 6
Bounding-box-level performance metrics (RMSE, R^2 and MAE) for the simplified coherent clusters ANN configuration.

Bounding box	Coordinates		RMSE	R^2	MAE
	Latitude (°)	Longitude (°)			
1	40.20, 40.58	−4.10, −3.43	2.34	3.44	0.86
2	40.59, 41.20	−4.10, −3.43	1.18	1.83	0.42
3	40.10, 40.58	−3.42, −3.15	1.30	1.64	0.67

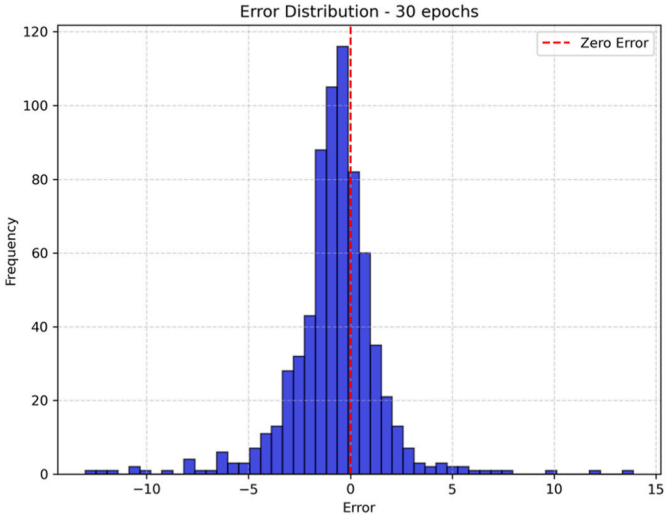


Fig. 6. Frequency distribution of absolute errors (ANN, simplified coherent clusters).

Overall, the model reproduces the main temporal dynamics of surface-level NO_2, capturing both short-term day-to-day variability and the broader seasonal increase toward the colder months. Predicted values closely follow the observed concentrations during much of the period, particularly under low to moderate pollution conditions.

During the latter part of the year, when NO_2 levels exhibit higher variability and more frequent peaks, the agreement between observations and predictions remains generally consistent, although differences become more apparent. Several high-concentration episodes are captured in terms of timing, while their magnitude is occasionally underestimated or overestimated, leading to increased dispersion between the two series. This behavior is consistent with the patterns observed in the error distribution and the observed–predicted scatter plots.

Despite these deviations, the predicted time series preserves the overall temporal structure of the observations, including the transition

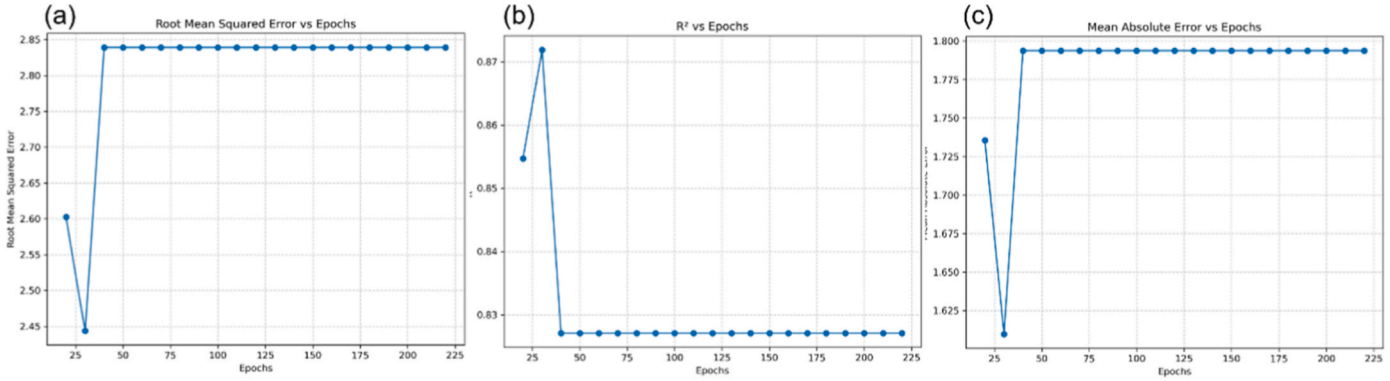


Fig. 5. Evolution of model performance across different training epochs for the ANN with simplified coherent clusters: (a) RMSE, (b) R^2, and (c) MAE.

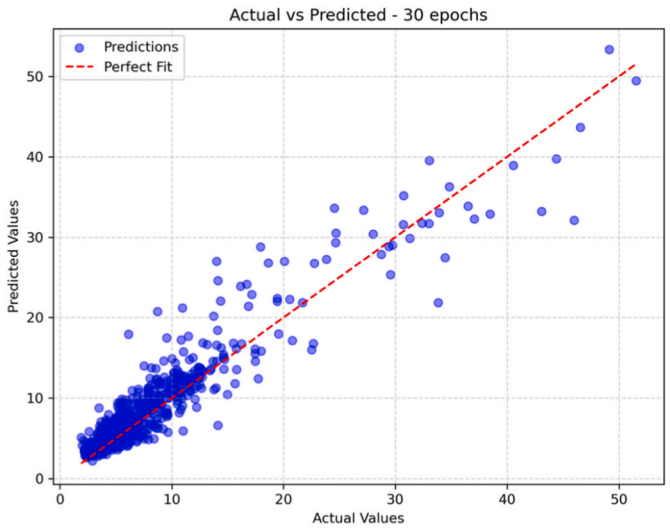


Fig. 7. Scatter plot of observed vs. predicted NO_2 concentrations (ANN, simplified coherent clusters).

from lower to higher pollution regimes. The close temporal alignment between observed (yellow) and predicted (blue) values across most of the test period highlights the model's ability to track NO_2 variability at daily time scales.

For this configuration, additional diagnostic plots, including error distributions, observed–predicted scatter plots, and temporal evolution of NO_2 concentrations at the bounding-box level, are provided as Supplementary Information (Fig. S2).

4. Discussion

4.1. Contextualizing performance with existing literature

The predictive accuracy achieved in this study compares favorably with previous research estimating surface NO_2 from satellite observations (Table 7) Accross Europe, most studies report higher RMSE values (2.89–13.07 μg/m^3), lower R^2 coefficients (0.56–0.88), and higher MAE values (3.10–7.77 μg/m^3) than those obtained here. Similar or even broader performance ranges are reported in studies conducted over China and the United States, where the spatial domains of interest typically encompass entire countries or large multi-regional areas.

Direct comparison among studies is inherently challenging due to differences in spatial extent, temporal coverage, predictor selection, model architecture, and validation strategies. Nevertheless, a general tendency emerges from the literature indicating that deep learning models often outperform tree-based algorithms when applied to large and heterogeneous domains. This pattern, however, does not imply an intrinsic superiority of deep learning approaches. Rather, it reflects the

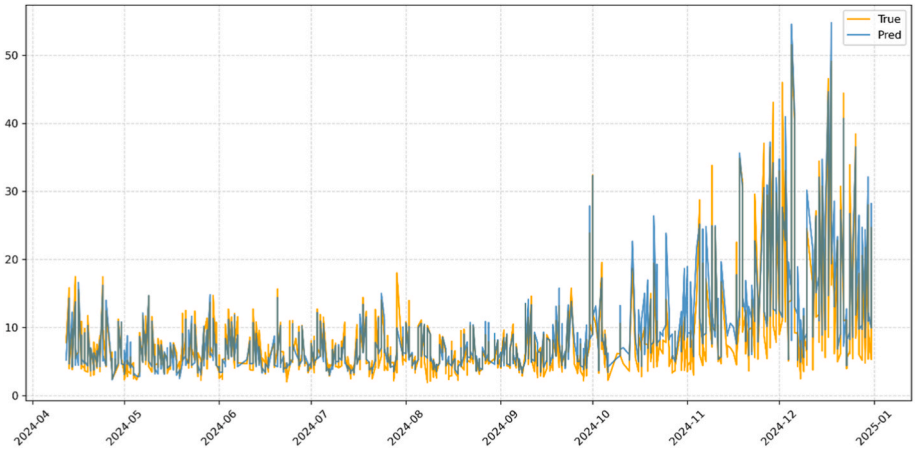


Fig. 8. Observed (yellow) vs. predicted (blue) NO_2 concentrations over the year 2024. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 7
Summary of recent studies estimating ground-level NO_2 concentrations from Sentinel-5P TROPOMI and in situ data across different regions.

Study	Region	Prediction model	Temporal resolution	RMSE [μg/m^3]	R^2	MAE [μg/m^3]
Grzybowski et al. (2023)	Poland	Random Forest	Hourly	5.90	0.53	3.70
			Weekly	4.80	0.60	3.10
Balamurugan et al. (2023)	Germany	Gradient-Boosted Tree	Daily	4.77–8.67	0.68–0.88	–
Shetty et al. (2024)	Europe	XGBoost	Daily	13.07	–	7.77
Chan et al. (2021)	Germany	Artificial Neural Network	Daily	6.32	0.64	4.76
Das et al. (2021)	Ireland	Convolutional Neural Network	Daily	7.20	0.65	4.94
Cedeno Jimenez et al. (2023)	Milan (Italy)	Voting regression: Multi-Layer Perceptron Regressor + Kriging	Daily	2.89	–	–
Kim et al. (2021)	Switzerland and Italy	XGBoost	Hourly	–	0.84	4.74
He et al. (2022)	China	Ensemble: RF + Extra Trees + XGBoost + Deep Neural Network	Daily	5.62	0.89	4.04
Deng et al. (2022)	China	XGBoost	Daily	4.77	0.88	–
Naseer et al. (2022)	Pakistan	Polynomial Regression	Daily	19.27	0.49 – 0.57	–
Cao (2023)	USA	Convolutional Neural Network	Daily	2.26	0.89	1.53
			Annual	0.95	0.99	0.69
Xing et al. (2024)	USA	Deep Model-Measurement Fusion	Daily	2.74	0.98	–

ability of such models to accommodate complex, nonlinear relationships when trained on spatially and temporally diverse datasets. The broader literature on air quality modeling suggests that no single algorithmic class consistently outperforms others across all contexts, and that model performance is strongly conditioned by study design, data representativeness, and regional characteristics.

One important source of methodological divergence across studies concerns the temporal representation of meteorological drivers. Most existing approaches rely exclusively on meteorological variables sampled at or near the satellite overpass time or on daily averages, implicitly assuming that surface NO_2 concentrations are primarily governed by instantaneous atmospheric conditions (Balamurugan et al., 2023; Chan et al., 2021; Deng et al., 2022; Shetty et al., 2024). Alternative strategies have been proposed through deep-learning fusion frameworks that address temporal discontinuities by incorporating chemically consistent simulations, recurrent architectures, or physically constrained training informed by chemical transport models (Xing et al., 2024). While these methods achieve very high predictive accuracy, they require substantial auxiliary information, which may limit their transferability or operational simplicity.

Regarding the spatial treatment of the training data, many previous studies pool all available monitoring stations over large and environmentally heterogeneous domains, particularly in national or continental-scale applications (Cao, 2023; Shetty et al., 2024) or rely on purely geometric representations such as fixed grids or buffer-based averaging (Chan et al., 2021; Deng et al., 2022). Under these configurations, models are required to learn satellite–surface relationships across strongly contrasting emission regimes, topographic settings, and dispersion environments, which increases prediction uncertainty. More recent convolutional or physically constrained deep-learning approaches mitigate this issue by implicitly learning spatial patterns from high-dimensional inputs or by enforcing physical consistency through model-based constraints (Cao, 2023; Xing et al., 2024).

As discussed earlier in this manuscript, the Community of Madrid exhibits pronounced spatial heterogeneity in the relationship between satellite-derived and ground-based NO_2 concentrations. Correlation values reach up to 0.86 within the city of Madrid and 0.81 across the broader metropolitan area, but decrease to approximately 0.5 at rural sites, where topographic complexity, irregular wind and temperature patterns, and lower emission densities weaken the satellite–surface linkage (Morillas et al., 2024b, 2025). Recognizing this variability, and the limitations it poses for model generalization, motivated the adoption of more tailored modeling strategies designed to better capture the temporal and spatial dynamics of NO_2 in the Madrid domain.

Two main methodological approaches distinguish this study from previous research. The first involves the aggregation of meteorological variables over two distinct temporal windows: a short-term average coinciding with the satellite overpass (11:00–14:00) and a long-term average capturing antecedent atmospheric conditions (14:00 of the previous day–11:00 of the current day). Although the general practice is limited to averaging variables during the satellite overpass hours, the procedure applied here made use of two complementary temporal windows that capture both prior and concurrent atmospheric conditions, similarly to the method explored by Cedeno Jimenez et al. (2023).

The results show that including historical meteorological information through the extended temporal window (H2) consistently improves model performance across all tested architectures and spatial configurations. Models trained using both H1 and H2 windows achieved lower RMSE and MAE values and higher R^2 scores than those trained using the overpass window alone, as shown in Table S2. These findings indicate that antecedent atmospheric conditions provide relevant information for estimating surface-level NO_2 concentrations, likely reflecting the cumulative effects of meteorological processes influencing pollutant dispersion, accumulation, and chemical transformation. Given that no universally optimal temporal aggregation strategy has been identified in the literature, incorporating multiple temporal perspectives that characterise the relationships between meteorology and NO_2 may enhance model training, particularly considering that this dynamic can vary substantially across different study areas.

This temporal design is particularly well-suited to the meteorological behavior of the Madrid basin, where nocturnal thermal inversions frequently dominate the diurnal cycle of surface NO_2. Under typical anticyclonic conditions, weak nighttime winds and strong radiative cooling lead to the formation of a shallow boundary layer that traps pollutants close to the ground until late morning. As vertical mixing intensifies around midday, concentrations begin to dilute, coinciding with the Sentinel-5P overpass time. By incorporating both short-term states and antecedent conditions, the model can capture carry-over effects linked to nighttime stability and subsequent boundary-layer development (Martilli et al., 2022; Salvador et al., 2020).

The second methodological contribution concerns the spatial aggregation strategies tested, which represent a relatively underexplored dimension in satellite-based air quality modeling. While many previous studies have relied either on point-based approaches (i.e., pairing individual monitoring stations with their corresponding satellite pixels) or on broad regional aggregation that pools all available data, few have systematically assessed intermediate spatial stratification schemes. The finding that environmentally coherent clusters outperformed both finer (grid-based) and coarser (region-wide) aggregation schemes provides valuable guidance for future model design in other urban contexts. These results suggest that geographically informed spatial aggregation represents an effective and transferable pathway for enhancing prediction accuracy in heterogeneous urban environments such as the Community of Madrid, highlighting that substantial performance gains can be achieved through methodological design choices rather than increased model or data complexity.

4.2. Comparative model performance

The strong performance of the ANN models aligns with recent studies highlighting the increasing effectiveness of deep learning approaches in air quality modeling, particularly when capturing complex atmospheric processes (Chen et al., 2025; Karmoude et al., 2025). While tree-based ensemble methods partition the feature space through sequential binary splits, neural networks learn continuous, differentiable representations that may better approximate the smooth gradients characterizing atmospheric dynamics.

In the context of this study, these methodological differences are clearly reflected in the comparative results. The ANN model combined with the simplified coherent clusters approach achieved the highest overall accuracy, demonstrating a strong ability to capture nonlinear relationships between satellite-derived NO_2, meteorological predictors, and surface concentrations when spatial heterogeneity is reduced.

Nevertheless, this does not preclude the use of traditional machine learning approaches. The SVM model exhibited competitive performance under the sector-based spatial configuration, suggesting that kernel-based methods can effectively approximate nonlinear behavior when spatial stratification reduces heterogeneity within the training data. Similarly, RF models—both as standalone predictors and within the XGBoost-optimized framework—provided stable and consistent performance across spatial configurations, while offering the additional advantage of feature importance ranking, which proved valuable for dimensionality reduction and model interpretability within the hybrid modelling framework.

4.3. Influence of spatial aggregation in model accuracy

The spatial configurations yielding the best predictive performance were neither the complete regional aggregation approach nor the regular grid-based strategy based on fixed 5×5 km bounding boxes. In fact, the grid-based approach did not improve model performance and, in most cases, resulted in higher prediction errors compared to both

regional aggregation and stratification-based schemes. This indicates that a purely geometric subdivision of the study area is insufficient to capture the spatial variability of surface-level NO_2 concentrations.

The best-performing configurations were those that grouped monitoring stations located in relatively homogeneous topographic, urban, and emission characteristics, rather than following arbitrary geometric boundaries. Madrid's geography encompasses pronounced environmental gradients: from mountainous terrain exceeding 2000 m elevation in the north (Puerto de Cotos) to the flat, urbanized plains below 600 m in the south; from densely populated, traffic-dominated central areas to rural zones with minimal anthropogenic influence. Attempting to capture such diverse conditions through a single, uniform training configuration can reduce predictive accuracy, as the model must reconcile conflicting patterns rather than learn from locally coherent relationships.

Both clustering methods structured the monitoring network according to shared atmospheric dynamics—similar elevation ranges, comparable emission profiles, and analogous meteorological regimes—allowing the models to focus on dominant local processes without being distorted by conditions prevailing elsewhere. The simplified coherent clusters configuration further refined this strategy by excluding stations that lacked nearby counterparts with comparable topographic or urban characteristics, thereby increasing the internal homogeneity of each group. This targeted stratification enabled the training process to emphasize locally coherent relationships and ultimately enhanced both the interpretability and the robustness of the predictions.

4.4. Understanding prediction errors and model limitations

While the inclusion of additional spatial predictors—such as traffic intensity, land-use characteristics, nighttime light intensity, or emission inventories—can definitely improve model performance, these variables were deliberately excluded from the present framework. The primary objective of this work was to assess whether high-quality estimates of surface-level NO_2 concentrations can be achieved by combining satellite observations with routinely available meteorological data alone.

This choice was motivated by a methodological emphasis on transferability. Many of the mentioned spatial variables are not consistently available across all cities, are subject to heterogeneous definitions, or require substantial preprocessing and local calibration. By restricting the input space to satellite-derived NO_2 and standard meteorological observations, the proposed approach remains readily applicable to a wide range of urban environments with minimal data requirements.

The strong predictive performance obtained, particularly for the ANN models, demonstrates that a relatively simple input configuration can still capture the dominant drivers of surface-level NO_2 variability. This simplicity may facilitate the extension of the methodology to other Spanish cities, especially medium-sized municipalities that lack detailed emission inventories or high-resolution traffic datasets. Future research could build upon this baseline framework by progressively incorporating additional spatial predictors to further refine predictions during extreme pollution episodes.

Within this methodological scope, the best-performing configuration (ANN combined with simplified coherent clusters) achieved high predictive accuracy ($RMSE = 2.44\ \mu g/m^3$, $R^2 = 0.87$, $MAE = 1.61\ \mu g/m^3$), with most errors concentrated below $5\ \mu g/m^3$, confirming the model's ability to capture the main variability patterns in NO_2 concentrations (Figs. 6–8). Nonetheless, the long-tailed error distribution and the dispersion at higher values suggest systematic difficulties in reproducing extreme pollution episodes.

Several factors may contribute to the underestimation of high-concentration NO_2 events observed in this study. Satellite retrievals of tropospheric NO_2 columns are known to exhibit reduced sensitivity under highly polluted conditions, particularly when the atmospheric NO_2 burden approaches or exceeds the optimal dynamic range of the DOAS (Differential Optical Absorption Spectroscopy) retrieval algorithm implemented in TROPOMI (Van Geffen et al., 2020; Zhu et al., 2022). Evidence from a detailed consistency analysis over the Community of Madrid further supports this behavior: comparisons of TROPOMI tropospheric NO_2 data with surface NO_2 measurements showed high overall correlations but also highlighted limitations in capturing surface concentration extremes, especially under heterogeneous urban conditions (Morillas et al., 2024b).

In addition, highly localized emission plumes (e.g., intense traffic activity or short-lived emission episodes) are not fully resolved within satellite pixels spanning several square kilometres. As a result, peak concentrations measured at individual ground-based stations may be spatially diluted in the satellite signal (Apte and Manchanda, 2024). This limitation is further compounded by the spatial averaging inherent to the simplified coherent clusters strategy adopted here. While this approach improves overall model stability and reduces noise, it also smooths sharp concentration gradients within each cluster, thereby attenuating extreme values recorded at specific monitoring sites (Li et al., 2023; Zhang et al., 2025).

The concentration-dependent pattern of underestimation identified in this study—characterised by a systematic deviation at observed NO_2 levels above approximately $35\ \mu g/m^3$—is consistent with these documented behaviors. The agreement between the behavior observed here and that reported in the literature suggests that the underestimation of extreme pollution events is primarily driven by limitations in the satellite input data and spatial representativeness, rather than by deficiencies in the machine learning framework itself.

Atmospheric conditions themselves also play a key role in shaping when extreme NO_2 episodes occur and how reliably they can be captured by satellite observations. Meteorological conditions during extreme NO_2 pollution episodes in Madrid typically occur in winter, when NO_2 atmospheric lifetime is longer due to reduced photolysis, and emissions increase as a result of both traffic and residential heating. These wintertime stability conditions, characterized by weak winds, frequent temperature inversions, and greater cloud cover, limit vertical mixing and horizontal dispersion, favoring the accumulation of pollutants near the surface (Martilli et al., 2022; Salvador et al., 2020). At the same time, these adverse meteorological situations often reduce the availability or quality of satellite observations (e.g., due to cloud cover or low-quality pixels). As a result, the most polluted days are also those with fewer valid satellite measurements, introducing a potential sampling bias whereby extreme NO_2 episodes may be under-represented in the training dataset.

A further limitation stems from the spatial resolution of TROPOMI, which, although a major improvement over previous instruments, is still too coarse to resolve the fine-scale concentration gradients typical of dense urban environments where exposure variability occurs at sub-kilometer scales. This mismatch between satellite pixel size and urban-scale heterogeneity constrains the model's ability to capture sharp, localized peaks (Goldberg et al., 2021).

Additional variability may arise from stochastic fluctuations in emissions, including unexpected traffic disruptions or isolated events occurring within the aggregated spatial units. Moreover, the exclusion of peripheral stations, while improving internal homogeneity and model performance, limits the generalizability of the approach and suggests that separate models may be required for outlying areas in future operational deployments. Taken together, these factors highlight that, despite its overall robustness, the proposed framework is not free from structural constraints that should be carefully considered when interpreting results or extending the methodology to other regions.

5. Conclusions

This study demonstrates that the integration of Sentinel-5P tropospheric NO_2 observations with ground-based meteorological data enables accurate estimation of surface-level NO_2 concentrations across the Community of Madrid. Three complementary modelling paradigms

were evaluated: tree-based ensemble methods (RF and XGBoost, including hybrid configurations), a kernel-based machine learning approach (SVM), and a deep-learning architecture (ANN). All models were trained using air quality monitoring stations as ground-truth references. The methodological framework systematically assessed the influence of temporal and spatial preprocessing by combining two temporal aggregation windows (11:00–14:00 and the preceding 21-h period, 14:00–11:00) with four spatial configurations (regional aggregation, grid-based approach, sector-based stratification, and simplified coherent clusters).

Across all modelling approaches, the inclusion of historical meteorological information through the extended temporal window (H1+H2) consistently improved predictive performance compared to training based solely on the satellite overpass period, confirming that antecedent atmospheric conditions play a relevant role in shaping surface-level NO_2 concentrations.

The comparative evaluation of modelling approaches showed that ANN models achieved the highest overall predictive accuracy, particularly when combined with the simplified coherent clusters strategy, which groups monitoring stations according to shared environmental characteristics. This result highlights the ability of deep learning models to capture complex and nonlinear relationships between atmospheric predictors and surface NO_2 concentrations when spatial heterogeneity is reduced. Nevertheless, traditional machine learning approaches also demonstrated strong and competitive performance under specific spatial stratifications. In particular, the SVM model achieved its best results under the sector-based configuration, while tree-based ensemble methods (RF and Hybrid XGBoost-optimized models) exhibited stable and robust performance across different spatial schemes.

The results were generally comparable to those reported in other regions, but the spatial design of the training process proved to be a key factor in improving accuracy. Configurations that reduced spatial heterogeneity by grouping monitoring stations into environmentally coherent units yielded substantially better results than approaches that either treated the entire region as a single undifferentiated domain or modeled each station's environment individually. This outcome underscores the importance of aligning spatial stratification with underlying environmental gradients, topographic features, and urbanization patterns when designing air quality modeling frameworks: an aspect that is particularly critical in the Madrid region, where emission profiles are highly variable, population distribution is markedly heterogeneous, and both topography and meteorology exhibit strong spatial contrasts.

The methodology developed here offers significant implications for air quality monitoring in the Community of Madrid and beyond. First, it provides a scalable, cost-effective approach to complement existing ground-based infrastructure, particularly relevant given that the capital investment for a fully equipped reference-grade monitoring station ranges from €100,000 to €250,000. The ability to generate daily NO_2 estimates at spatial resolutions comparable to current monitoring network coverage enables more comprehensive assessment of air quality across the region, including areas where fixed stations are absent or insufficient.

Second, the demonstrated capacity to track temporal variability, including the detection of abrupt changes in emission patterns during the 2020 COVID-19 pandemic, suggests that this methodology could serve as a valuable tool for evaluating the effectiveness of air quality policies. As Madrid and other Spanish municipalities implement low-emission zones and sustainable urban mobility plans in compliance with Climate Change and Energy Transition Law 7/2021, satellite-based monitoring systems could provide near-real-time feedback on the spatial and temporal impacts of these interventions, supporting adaptive management strategies and evidence-based policy refinement.

Third, the approach addresses a critical gap in Spain's air quality monitoring landscape, where 98 out of 151 municipalities with more than 50,000 inhabitants operate only one monitoring station or none at all. The transferability of this methodology to other urban areas, once validated across different geographic and environmental contexts, could substantially improve regulatory compliance and public health protection at the national scale. This is particularly pressing given the tightening of European air quality standards, with Directive (EU) 2024/2881 establishing an annual NO_2 limit of 20 $\mu g/m^3$ by 2030 (half the current threshold) and WHO guidelines recommending even stricter levels of 10 $\mu g/m^3$.

However, several challenges must be addressed to fully realize the operational potential of satellite-based NO_2 monitoring. The tendency to underestimate extreme concentration episodes, indicates that additional refinements are needed to capture highly localized emission events. Future research should explore the incorporation of complementary data sources such as traffic flow information, urban morphology indicators, land use classifications, and emission inventories to improve model performance during high-pollution periods. Hybrid modeling approaches that combine machine learning with deterministic atmospheric transport models may also enhance predictive accuracy for extreme events while maintaining computational efficiency for routine monitoring.

The reliance on satellite observations under favorable atmospheric conditions presents another limitation, as extensive cloud cover can result in temporal gaps in data availability. Although 1573 days with valid observations were obtained during the 2020-2024 period (86% coverage), the development of gap-filling strategies or the integration of multiple satellite platforms with complementary overpass times could further improve temporal continuity. Additionally, while the $3.5 \times 5.5\ km^2$ spatial resolution of TROPOMI represents a substantial advancement over previous satellite instruments, it may still be insufficient to resolve fine-scale concentration gradients in densely populated urban cores where exposure variability occurs at sub-kilometer scales. The forthcoming Sentinel-4 geostationary mission, offering hourly temporal resolution and improved spatial detail, is expected to provide valuable opportunities for overcoming these limitations and strengthening the operational potential of satellite-based NO_2 monitoring systems.

From a broader perspective, this study contributes to the growing body of evidence supporting the operational integration of satellite remote sensing into air quality management frameworks. As Earth observation systems continue to advance in spatial, temporal, and spectral resolution, and as machine learning techniques become increasingly sophisticated, the prospects for developing comprehensive, cost-effective air quality monitoring networks are highly favorable. The methodology presented here adds to these ongoing developments by illustrating how the combined use of satellite data, ground-based measurements, and modern computational tools can improve our ability to monitor and interpret urban air pollution, while complementing existing approaches applied in other regions.

In the specific context of the Community of Madrid, the implementation of this approach could transform air quality monitoring from a network of discrete measurement points into a spatially continuous assessment system. This transition would enable more accurate exposure estimation for epidemiological studies, more targeted identification of pollution hotspots for mitigation efforts, and more equitable environmental governance by ensuring that air quality surveillance extends beyond the most densely instrumented urban centers to include underserved areas. As Madrid continues its trajectory toward becoming a more sustainable and livable city, satellite-based monitoring systems will likely play an increasingly central role in tracking progress, informing policy decisions, and safeguarding public health.

CRediT authorship contribution statement

Carlos Morillas: Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing – original draft. **Rocco Giosa:** Data curation, Formal analysis, Investigation, Software, Validation, Writing – review & editing. **Sergio Alvarez:** Project administration, Resources,

Supervision, Writing – review & editing. **Carmine Serio:** Conceptualization, Formal analysis, Supervision, Writing – review & editing. **Guido Masiello:** Conceptualization, Investigation, Resources, Visualization. **Giuliano Liuzzi:** Data curation, Formal analysis, Investigation, Supervision. **Carmen Aviles:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Sara Martinez:** Formal analysis, Investigation, Methodology, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by:

Project TED2021-130457A-I00, funded by MCIN/AEI/10.13039/501100011033 and European Union "NextGenerationEU"/PRTR".

DIGITWINS4CIUE (project co funded by the European Health and Digital Executive Agency, HADEA, under the Digital Europe Programme, GA 101084054).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2026.121904.

Data availability

Data will be made available on request.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: a next-generation hyperparameter optimization framework. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 2623–2631. https://doi.org/10.1145/3292500.3330701.

Apte, J.S., Manchanda, C., 2024. High-resolution urban air pollution mapping. Science 385, 380–385. https://doi.org/10.1126/SCIENCE.ADQ3678;WEBSITE:WEBSITE:AAAS-SITE;JOURNAL:JOURNAL:SCIENCE;WGROUP:STRING:PUBLICATION.

Balamurugan, V., Chen, J., Wenzel, A., Keutsch, F.N., 2023. Spatiotemporal modeling of air pollutant concentrations in Germany using machine learning. Atmos. Chem. Phys. 23, 10267–10285. https://doi.org/10.5194/acp-23-10267-2023.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324/METRICS.

Cao, E.L., 2023. National ground-level NO2 predictions via satellite imagery driven convolutional neural networks. Front. Environ. Sci. 11, 1285471. https://doi.org/10.3389/FENVS.2023.1285471/BIBTEX.

Cedeno Jimenez, J.R., Pugliese Viloria, A. de J., Brovelli, M.A., 2023. Estimating daily NO2 ground level concentrations using Sentinel-5P and ground sensor meteorological measurements. ISPRS Int. J. GeoInf. 12, 107. https://doi.org/10.3390/ijgi12030107.

Cersosimo, A., Serio, C., Masiello, G., 2020. TROPOMI NO2 tropospheric column data: regridding to 1 km grid-resolution and assessment of their consistency with in situ surface observations. Remote Sens. 12, 2212. https://doi.org/10.3390/RS12142212, 2020, Vol. 12, Page 2212.

Chan, K.L., Khorsandi, E., Liu, S., Baier, F., Valks, P., 2021. Estimation of surface NO2 concentrations over Germany from TROPOMI satellite observations using a machine learning method. Remote Sens. 13, 969. https://doi.org/10.3390/RS13050969, 2021, Vol. 13, Page 969.

Chen, G., Chen, S., Li, D., Chen, C., 2025. A hybrid deep learning air pollution prediction approach based on neighborhood selection and spatio-temporal attention. Sci. Rep. 15, 3685. https://doi.org/10.1038/s41598-025-88086-1, 2025 151.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 785–794. https://doi.org/10.1145/2939672.2939785;CSUBTYPE:STRING:CONFERENCE, 13-17-August-2016.

City of Madrid, 2025a. Hourly air quality data since 2001 [WWW Document]. URL. https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=f3c0f7d512273410VgnVCM2000000c205a0aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default, 11.20.25.

City of Madrid, 2025b. Hourly meteorological data since 2019 [WWW Document]. URL. https://datos.madrid.es/sites/v/index.jsp?vgnextoid=fa8357cec5efa610VgnVCM1000001d4a900aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD, 11.20.25.

Community of Madrid, 2025a. Hourly air quality data since 2005 [WWW Document]. URL. https://datos.comunidad.madrid/dataset/calidad_aire_datos_historico, 11.20.25.

Community of Madrid, 2025b. Hourly meteorological data since 2020 [WWW Document]. URL. https://datos.comunidad.madrid/catalogos/#/dataset/calidad_aire_datos_meteo_historico/?view=info, 11.20.25.

Das, B.P., Pathan, M.S., Lee, Y.H., Dev, S., 2021. Estimating ground-level nitrogen dioxide concentration from satellite data. Prog. Electromagn. Res. Symp. 2021-Nov. 1176–1182. https://doi.org/10.1109/PIERS53385.2021.9694752.

Deng, F., Chen, Y., Li, L., Wang, C., Cao, L., Peng, C., 2022. Estimation of near-surface NO2based on TROPOMI and provincial control stations' data using machine learning. 2022 5th World Conf. Mech. Eng. Intell. Manuf. WCMEIM 2022, 639–644. https://doi.org/10.1109/WCMEIM56910.2022.10021549.

European Commission, 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on Ambient Air Quality and Cleaner Air for Europe.

European Commission, 2024. Directive (EU) 2024/2881 of the European Parliament and of the Council of 23 October 2024 on Ambient Air Quality and Cleaner Air for Europe.

Friedman, J.H., 2001. Greedy Function Approximation: a Gradient Boosting Machine, pp. 1189–1232. https://doi.org/10.1214/AOS/1013203451, 10.1214/aos/101320345129.

Garcia Santos, D., Parés, M.E., 2025. Assessing nitrogen dioxide monitoring techniques: a comparative analysis of Sentinel-5 precursor satellite and ground measurements in Catalonia. Atmos. Meas. Tech. 18, 5569–5590. https://doi.org/10.5194/AMT-18-5569-2025.

Goldberg, D.L., Anenberg, S.C., Kerr, G.H., Mohegh, A., Lu, Z., Streets, D.G., 2021. TROPOMI NO2 in the United States: a detailed look at the annual averages, weekly cycles, effects of temperature, and correlation with surface NO2 concentrations. Earths Future 9. https://doi.org/10.1029/2020EF001665 e2020EF001665.

Google Earth Engine, (GEE), 2025. Sentinel-5P OFFL NO2: offline nitrogen dioxide [WWW Document]. URL. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_OFFL_L3_NO2?hl=es-419#description, 11.20.25.

Government of Spain, 2021. Ley 7/2021, De 20 De Mayo, De Cambio Climático Y Transición Energética.

Griffin, D., Zhao, X., McLinden, C.A., Boersma, F., Bourassa, A., Dammers, E., Degenstein, D., Eskes, H., Fehr, L., Fioletov, V., Hayden, K., Kharol, S.K., Li, S.M., Makar, P., Martin, R.V., Mihele, C., Mittermeier, R.L., Krotkov, N., Sneep, M., Lamsal, L.N., Linden, M. ter, Geffen, J. van, Veefkind, P., Wolde, M., 2019. High-resolution mapping of nitrogen dioxide with TROPOMI: first results and validation over the Canadian oil sands. Geophys. Res. Lett. 46, 1049–1060. https://doi.org/10.1029/2018GL081095.

Grzybowski, P.T., Markowicz, K.M., Musiał, J.P., 2023. Estimations of the ground-level NO2 concentrations based on the Sentinel-5P NO2 tropospheric column number density product. Remote Sens. 15, 378. https://doi.org/10.3390/RS15020378, 2023, Vol. 15, Page 378.

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 Sens. 14, 2807. https://doi.org/10.3390/RS14122807/S1.

Instituto Nacional de Estadística (INE), 2024. Cifras oficiales de población de los municipios españoles: Revisión del Padrón Municipal/Resultados [WWW Document]. URL. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177011&menu=resultados&idp=1254734710990, 11.20.25.

Jeong, U., Hong, H., 2021. Assessment of tropospheric concentrations of NO2 from the TROPOMI/Sentinel-5 precursor for the estimation of long-term exposure to surface NO2 over South Korea. Remote Sens. 13, 1877. https://doi.org/10.3390/RS13101877, 2021, Vol. 13, Page 1877.

Jonson, J.E., Borken-Kleefeld, J., Simpson, D., Nyíri, A., Posch, M., Heyes, C., 2017. Impact of excess NOx emissions from diesel cars on air quality, public health and eutrophication in Europe. Environ. Res. Lett. 12, 094017. https://doi.org/10.1088/1748-9326/AA8850.

Karmoude, M., Munhungewarwa, B., Chiraira, I., Mckenzie, R., Kong, J., Smith, B., Ayana, G., Njara, N., Mathaha, T., Kumar, M., Mellado, B., 2025. Machine learning for air quality prediction and data analysis: review on recent advancements, challenges, and outlooks. Sci. Total Environ. 1002, 180593. https://doi.org/10.1016/J.SCITOTENV.2025.180593.

Khomenko, S., Cirach, M., Pereira-Barboza, E., Mueller, N., Barrera-Gómez, J., Rojas-Rueda, D., de Hoogh, K., Hoek, G., Nieuwenhuijsen, M., 2021. Premature mortality due to air pollution in European cities: a health impact assessment. Lancet Planet. Health 5, e121–e134. https://doi.org/10.1016/S2542-5196(20)30272-2.

Khomenko, S., Pisoni, E., Thunis, P., Bessagnet, B., Cirach, M., Iungman, T., Barboza, E. P., Khreis, H., Mueller, N., Tonne, C., de Hoogh, K., Hoek, G., Chowdhury, S., Lelieveld, J., Nieuwenhuijsen, M., 2023. Spatial and sector-specific contributions of emissions to ambient air pollution and mortality in European cities: a health impact assessment. Lancet Public Health 8, e546–e558. https://doi.org/10.1016/S2468-2667(23)00106-8.

Kim, M., Brunner, D., Kuhlmann, G., 2021. Importance of satellite observations for high-resolution mapping of near-surface NO 2 by machine learning. Remote Sens. Environ. 264, 34–4257. https://doi.org/10.1016/j.rse.2021.112573.

Kwon, H., Cho, N., 2023. Corrosion behaviors of outdoor bronze sculptures in an urban–industrial environment: corrosion experiment on artificial sulfide patina. Met. 13, 2023. https://doi.org/10.3390/MET13061101, 1101 13, 1101.

Lange, K., Richter, A., Burrows, J.P., 2022. Variability of nitrogen oxide emission fluxes and lifetimes estimated from Sentinel-5P TROPOMI observations. Atmos. Chem. Phys. 22. https://doi.org/10.5194/acp-22-2745-2022.

Laughner, J.L., Cohen, R.C., 2019. Direct observation of changing NOx lifetime in North American cities. Science 366, 723–727. https://doi.org/10.1126/SCIENCE.AAX6832/SUPPL_FILE/AAX6832-LAUGHNER-SM.PDF.

Li, L., Wang, J., Franklin, M., Yin, Q., Wu, J., Camps-Valls, G., Zhu, Z., Wang, C., Ge, Y., Reichstein, M., 2023. Improving air quality assessment using physics-inspired deep graph learning. npj Clim. Atmos. Sci. 6, 152. https://doi.org/10.1038/s41612-023-00475-3, 2023 61.

Martilli, A., Sánchez, B., Santiago, J.L., Rasilla, D., Pappaccogli, G., Allende, F., Martín, F., Roman-Cascón, C., Yagüe, C., Fernández, F., 2022. Simulating the pollutant dispersion during persistent wintertime thermal inversions over urban areas. The case of Madrid. Atmos. Res. 270, 106058. https://doi.org/10.1016/J.ATMOSRES.2022.106058.

McDuffie, E.E., Smith, S.J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E.A., Zheng, B., Crippa, M., Brauer, M., Martin, R.V., 2020. A global anthropogenic emission inventory of atmospheric pollutants from sector- and fuel-specific sources (1970-2017): an application of the Community Emissions Data System (CEDS). Earth Syst. Sci. Data 12, 3413–3442. https://doi.org/10.5194/ESSD-12-3413-2020.

Ministerio de Transición Ecológica y Reto Demográfico (MITERD), 2024. Datos oficiales Calidad del Aire 2024 [WWW Document]. URL. https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/evaluacion-y-datos-de-calidad-del-aire/datos/datos-oficiales-2024.html, 11.20.25.

Ministerio de Transición Ecológica y Reto Demográfico (MITERD), 2025. Evaluación De La Calidad Del Aire En España.

Morillas, C., Alvarez, S., Pires, J.C.M., Garcia, A.J., Martinez, S., 2024a. Impact of the implementation of Madrid's low emission zone on NO2 concentration using Sentinel-5P/TROPOMI data. Atmos. Environ. 320, 120326. https://doi.org/10.1016/J.ATMOSENV.2024.120326.

Morillas, C., Alvarez, S., Serio, C., Masiello, G., Martinez, S., 2024b. TROPOMI NO2 Sentinel-5P data in the community of Madrid: a detailed consistency analysis with in situ surface observations. Remote Sens. Appl. Soc. Environ. 33, 101083. https://doi.org/10.1016/J.RSASE.2023.101083.

Morillas, C., Álvarez, S., Pires, J.C.M., García, A.J., Martínez, S., 2025. Linking satellite and ground observations of NO2 in Spanish cities: influence of meteorology and O3. Nitrogen (Switzerland) 6, 32. https://doi.org/10.3390/NITROGEN6020032/S1.

Naseer, E., Basit, A., Bhatti, M.K., Siddique, M.A., 2022. Machine learning for area-wide monitoring of surface level concentration of NO2Using remote sensing data. 2022 Int. Conf. Emerg. Trends Electr. Control. Telecommun. Eng. ETECTE 2022 - Proc. https://doi.org/10.1109/ETECTE55893.2022.10007417.

Rahman, M.M., Shuo, W., Zhao, W., Xu, X., Zhang, W., Arshad, A., 2022. Investigating the relationship between air pollutants and meteorological parameters using satellite data over Bangladesh. Remote Sens. 14, 2757. https://doi.org/10.3390/RS14122757/S1.

Rudke, A.P., Martins, J.A., Hallak, R., Martins, L.D., de Almeida, D.S., Beal, A., Freitas, E. D., Andrade, M.F., Koutrakis, P., Albuquerque, T.T.A., 2023. Evaluating TROPOMI and MODIS performance to capture the dynamic of air pollution in São Paulo state: a case study during the COVID-19 outbreak. Remote Sens. Environ. 289, 113514. https://doi.org/10.1016/J.RSE.2023.113514.

Salvador, P., Pandolfi, M., Tobías, A., Gómez-Moreno, F.J., Molero, F., Barreiro, M., Pérez, N., Revuelta, M.A., Marco, I.M., Querol, X., Artíñano, B., 2020. Impact of mixing layer height variations on air pollutant concentrations and health in a European urban area: madrid (Spain), a case study. Environ. Sci. Pollut. Res. 27 (33), 41702–41716. https://doi.org/10.1007/S11356-020-10146-Y, 2020 27.

Shah, V., Jacob, D.J., Li, K., Silvern, R., Zhai, S., Liu, M., Lin, J., Zhang, Q., 2020. Effect of changing NO*x* lifetime on the seasonality and long-term trends of satellite-observed tropospheric NO2 columns over China. Atmos. Chem. Phys. 20, 1483–1495. https://doi.org/10.5194/ACP-20-1483-2020.

Shaw, S., Van Heyst, B., 2022. An evaluation of risk ratios on physical and mental health correlations due to increases in ambient nitrogen oxide (NOx) concentrations. Atmosphere 13, 967. https://doi.org/10.3390/ATMOS13060967, 2022, Vol. 13, Page 967.

Shetty, S., Schneider, P., Stebel, K., David Hamer, P., Kylling, A., Koren Berntsen, T., 2024. Estimating surface NO2 concentrations over Europe using Sentinel-5P TROPOMI observations and machine learning. Remote Sens. Environ. 312, 114321. https://doi.org/10.1016/J.RSE.2024.114321.

Van Geffen, J., Folkert Boersma, K., Eskes, H., Sneep, M., Ter Linden, M., Zara, M., Pepijn Veefkind, J., 2020. S5P TROPOMI NO2 slant column retrieval: method, stability, uncertainties and comparisons with OMI. Atmos. Meas. Tech. 13, 1315–1335. https://doi.org/10.5194/AMT-13-1315-2020.

Van Geffen, J.H.G.M., Eskes, H.J., Boersma, K.F., Maasakkers, J.D., Veefkind, J.P., 2019. TROPOMI ATBD of the Total and Tropospheric NO2 Data Products.

Veefkind, J.P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H.J., de Haan, J.F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., Levelt, P.F., 2012. TROPOMI on the ESA Sentinel-5 precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. Remote Sens. Environ. 120, 70–83. https://doi.org/10.1016/J.RSE.2011.09.027.

World Health Organization (WHO), 2021. WHO Global Air Quality Guidelines, Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide.

Xing, J., Baek, B.H., Li, S., Wang, C.T., Song, G., Ma, S., Zheng, S., Liu, C., Tong, D., Woo, J.H., Liu, T.Y., Fu, J.S., 2024. A physically constrained deep-learning fusion method for estimating surface NO2 concentration from satellite and ground monitors. Environ. Sci. Technol. 58, 21218–21228. https://doi.org/10.1021/ACS.EST.4C07341.

Zhang, H., Wang, Y., Hu, J., Ying, Q., Hu, X.M., 2015. Relationships between meteorological parameters and criteria air pollutants in three megacities in China. Environ. Res. 140, 242–254. https://doi.org/10.1016/J.ENVRES.2015.04.004.

Zhang, T., Zheng, B., Huang, R., 2025. Adaptive high-resolution mapping of air pollution with a novel implicit 3D representation approach. npj Clim. Atmos. Sci. 8 (1), 180. https://doi.org/10.1038/s41612-025-01044-6, 2025 8.

Zhu, Y., Liu, C., Hu, Q., Teng, J., You, D., Zhang, C., Ou, J., Liu, T., Lin, J., Xu, T., Hong, X., 2022. Impacts of TROPOMI-derived NOX emissions on NO2 and O3 simulations in the NCP during COVID-19. ACS Environ. Au 2, 441. https://doi.org/10.1021/ACSENVIRONAU.2C00013.