

RESEARCH ARTICLE

Nitrogen dioxide pollution in 346 Chinese cities: Spatiotemporal variations and natural drivers from multi-source remote sensing data

Feifei Cheng¹, Qing Sun², Jiaqi Zhang¹, Jiahe Liu², Wenwen Lü¹, Fapeng Shen², Heming Yang^{1*}

¹ Jilin Normal University, Siping, Jilin, China, ² Jilin Agricultural University, Changchun, Jilin, China

☯ These authors contributed equally to this study.

* wosyhm@jlnu.edu.cn



OPEN ACCESS

Citation: Cheng F, Sun Q, Zhang J, Liu J, Lü W, Shen F, et al. (2025) Nitrogen dioxide pollution in 346 Chinese cities: Spatiotemporal variations and natural drivers from multi-source remote sensing data. PLoS One 20(11): e0334535. <https://doi.org/10.1371/journal.pone.0334535>

Editor: Benedetto Schiavo, National Autonomous University of Mexico Institute of Geophysics: Universidad Nacional Autonoma de Mexico Instituto de Geofisica, MEXICO

Received: June 4, 2025

Accepted: September 29, 2025

Published: November 7, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0334535>

Copyright: © 2025 Cheng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/),

Abstract

In this study, tropospheric column concentration of nitrogen dioxide (TNO₂CC) were derived from Sentinel-5P data. We employed statistical and local spatial autocorrelation analyses to investigate the spatiotemporal distribution and variation of TNO₂CC across 346 major Chinese cities from 2019 to 2023. Using Random Forest (RF) and Shapley Additive Explanations (SHAP), we analyzed the influence of 15 natural factors on ambient TNO₂CC levels. The high R² values (0.92 and 0.76), along with the close adherence to the 1:1 line, demonstrate the model's robustness. The most influential natural factors identified include atmospheric pressure, aerosol optical depth, Leaf Area Index, evapotranspiration, and dew point temperature. Additionally, a non-linear response curve approach was applied to examine the independent association between natural driving factors and pollutant concentrations. TNO₂CC varied seasonally across the 346 cities, with the highest levels in winter and the lowest in summer. From 2019 to 2023, TNO₂CC levels exhibited fluctuating trends, with notable regional disparities: higher concentrations were observed in capital cities and in northern and northeastern part of China. TNO₂CC were significantly influenced by temperature-related variables, aerosol optical depth, and leaf area index. The findings of this study identify key natural influencing factors and provide a scientific basis for revealing the causes of urban air pollution in China, informing pollution control strategies, identifying priority areas for remediation, and supporting the natural formulation of protection policies.

Introduction

In recent years, environmental problems caused by air pollutants have attracted increasing attention [1]. Air pollution, which threatens human health and ecosystems, is a global environmental issue [2]. Air quality in any region is directly influenced by

which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All raw data files are available from the figshare database at DOI: [10.6084/m9.figshare.29948741](https://doi.org/10.6084/m9.figshare.29948741).

Funding: This research was supported by the Science and Technology Department of Jilin Province. Approval No.: YDZJ202301ZYS234 (received by Drs. Heming Yang, Feifei Cheng, Jiaqi Zhang, and Wenwen Lü) and YDZJ202401520ZYTS (received by Drs. Heming Yang, Qing Sun, Fapeng Shen, and Jiahe Liu).

Competing interests: The authors have declared that no competing interests exist.

local human activities [3]. Nitrogen dioxide (NO_2) is both a pollutant in its own right and a precursor to other pollutants, such as O_3 through photochemical reactions with volatile organic compounds (VOCs) and fine particulate matter ($\text{PM}_{2.5}$) via nitrate formation [4]. TNO_2CC is an important indicator of air pollution [5], as it contributes to acid rain, acid fog, and photochemical smog [6], increases $\text{PM}_{2.5}$ concentrations [7–8], threatens public health [9–12], and harms both society and the ecological environment [13]. Therefore, it is essential to study NO_2 pollution [14].

To systematically investigate the etiology of atmospheric pollution and develop source-oriented mitigation strategies, scholars have predominantly focused on two key research avenues: (1) the spatiotemporal variations of air pollution, and (2) the identification of multi-scale drivers influencing pollutant dynamics [15]. Traditional ground-based air pollution monitoring and analysis methods are often limited due to (1) insufficient sampling across spatial and temporal dimensions [16], and (2) the uncertainties associated with interference from other gases [17]. Random Forest (RF), a common Machine Learning (ML) method, effectively identifies key features from high-dimensional datasets, making it a robust tool for pollution analysis [18].

Previous studies have identified various factors that may influence air pollution. For example, Wang et al. (2015) observed strong correlations between atmospheric pollutants and meteorological parameters, such as temperature, relative humidity, wind speed, and precipitation, in Wuxi's urban area during 2014 [19]. Liu et al. reported a significant positive correlation between the level of urbanization, human activity intensity, and environmental pollutant concentrations in Fangshan District, Beijing [20]. Additional factors, such as vegetation coverage index [21], have also been widely examined by numerous researchers [22–30]. These influencing factors are widely recognized. Building on this foundation, 15 independent variables were selected to study TNO_2CC based on the principles of data representativeness, quantifiability, and accessibility, while accounting for the systemic interrelationships among multiple factors. The variables include: Pressure (atmospheric pressure), LAI (Leaf Area Index), Dew (dew point temperature), AOD (aerosol optical depth), WS (wind speed), Month (month), ET (evapotranspiration), RH (relative humidity), Pre (precipitation), Fire (fire activity), LST (land surface temperature), Temp (air temperature), GPP (gross primary productivity), Year (year), and Snow (snow cover).

The Tropospheric Monitoring Instrument (TROPOMI) effectively observes global atmospheric trace gases, including NO_2 . This study utilizes the Google Earth Engine (GEE) platform to retrieve TNO_2CC data, integrating a particle swarm optimization-enhanced random forest with Shapley Additive Explanations (SHAP) interpretation algorithms to systematically analyze the spatiotemporal variation patterns and natural driving factors of NO_2 pollution across China's prefecture-level cities. Within this framework, we quantitatively examine the impacts of meteorological drivers on regional air quality dynamics. The integration of satellite remote sensing and explainable machine learning offers a robust analytical framework for identifying pollution control priorities, thereby establishing a theoretical foundation for evidence-based atmospheric pollution mitigation strategies in China.

Materials and methods

Study area

As the world's largest developing country, China covers vast territorial expanses and encompasses diverse natural environments. The nation extends across extensive latitudinal gradients and traverses multiple climatic zones, including tropical, subtropical, temperate, and frigid zones. These distinct climatic regimes exhibit significant variations in temperature, humidity, and wind patterns. For instance, southern cities experience humid subtropical conditions, while northern regions are characterized by continental aridity. China's landforms are complex and diverse, comprising plateaus, mountains, plains, basins, and other landform types. Mountainous and plateau regions facilitate air mass exchange, whereas basin and plain terrains restrict pollutant diffusion. As shown in Fig 1, urban development patterns vary markedly across Chinese 346 study units, which include sub-provincial cities, prefecture-level municipalities, and autonomous prefectures. These urban centers, predominantly concentrated in eastern China, display divergent industrialization trajectories, urbanization rates, and population densities. The selected cities encapsulate representative variations in terrain, climatic, and environmental governance frameworks, ensuring a comprehensive analysis of regional atmospheric dynamics.

Hu Huanyong Line(or Hu Line)is a demographic and geographic dividing line in China, proposed by Chinese geographer Hu Huanyong in 1935. It illustrates the uneven distribution of China's population and economic activity from northeast to southwest. The administrative division data in GeoJSON format is sourced from the National Geospatial Information Public Service Platform (Tianditu), with the website: <https://cloudcenter.tianditu.gov.cn/administrativeDivision>. The map approval number is GS (2024) 0650. The data coordinates are in GCS_WGS_1984. Global Artificial Impervious

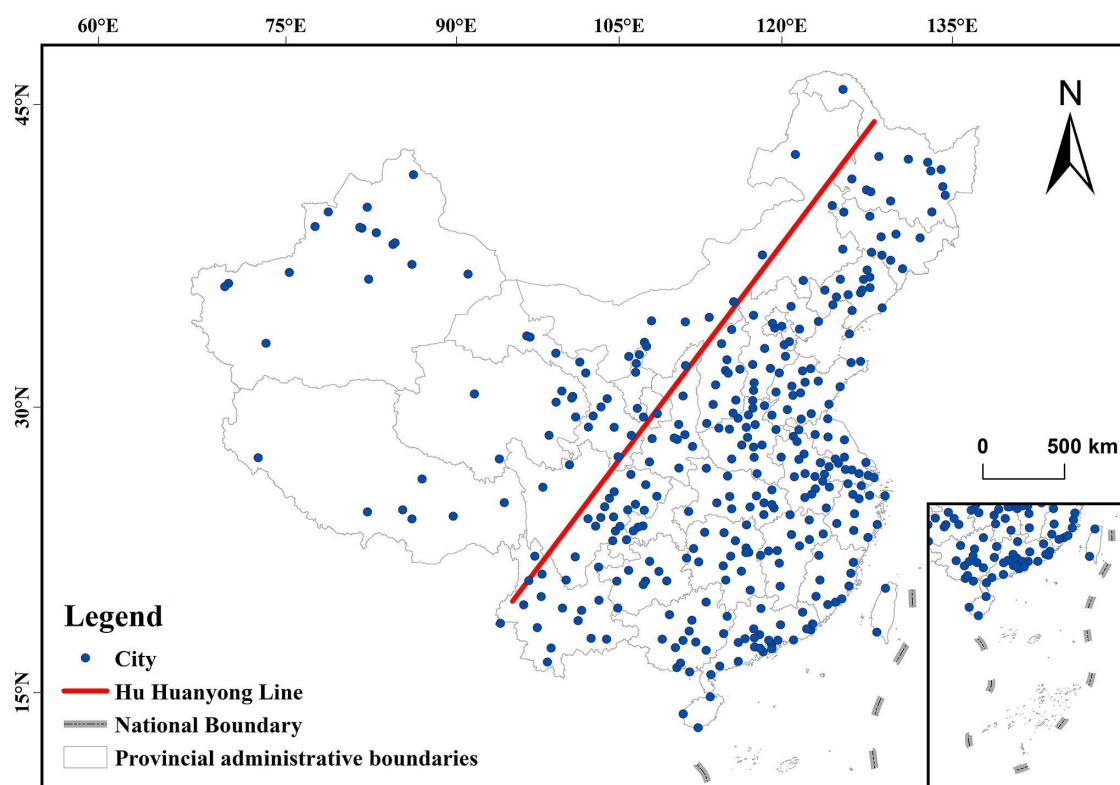


Fig 1. Distribution map of major cities in China.

<https://doi.org/10.1371/journal.pone.0334535.g001>

Area (GAIA) dataset, Version 10 (v10). Developed by: Tsinghua University/Fine Resolution Observation and Monitoring of Global Land Cover (FROM-GLC). <http://data.ess.tsinghua.edu.cn>

Data sources

The Sentinel satellite series is a component of the European Copernicus Programme. The Sentinel-5P (S5P) satellite, launched in 2017, is designed for real-time monitoring of various atmospheric trace gases, aerosols, and cloud distributions on a global scale (<https://dataspace.copernicus.eu/>). Since August 2019, it has achieved a minimum spatial resolution of 5.5×3.5 kilometers. The main data products include Level 1B, which consists of radiometrically calibrated raw spectral data; Level 2: which provides retrieved vertical column concentrations of atmospheric components such as the TNO_2CC ; and Level 3, which offers gridded data after spatiotemporal aggregation, such as global daily/monthly averages. The Level 3 products deliver gridded daily averages measured by the TROPOMI spectrometer. As measurements are taken at same time each day (early afternoon), these data effectively capture the NO_2 pollution characteristics of urban agglomerations in China.

This study uses the OFFL products of Sentinel-5P for research, and uses the “COPERNICUS_S5P_OFFL_L3_NO₂” dataset to count TNO_2CC , which includes annual, seasonal, and monthly TNO_2CC data from 2019 to 2023. The OFFL_L3 product is selected because its update frequency is 1–2 months, which is suitable for scientific research and long-term trend analysis, and the data has undergone complete radiometric calibration, spectral correction, and optimization of retrieval algorithms (such as the update of the TM5 meteorological model) to reduce the impact of cloud cover, etc. In contrast, NRT (Near-Real-Time) data is updated hourly and is mainly used for real-time monitoring. More importantly, the Level 3 data has been aggregated into gridded products (such as 0.01° resolution), which directly supports the calculation of seasonal/annual averages, while NRT requires users to process it themselves. All satellite nitrogen dioxide data used in this study are tropospheric column concentration, with a unit of mol/m^2 . Other data sources are shown in [Table 1](#):

ECMWF/ERA5_LAND/MONTHLY: The ERA5-LAND-MONTHLY dataset available on the Google Earth Engine (GEE) platform combines model data with measured data from countries around the world using physical laws, with a spatial resolution of 10,000 meters and a temporal resolution of 1 month. In this study, bands such as ‘temperature_2m’, ‘total_precipitation’, ‘dewpoint_temperature_2m’, ‘surface_pressure’, ‘u_component_of_wind_10 m’ and ‘v_component_of_wind_10 m’ from the “ECMWF/ERA5_LAND/MONTHLY” dataset in GEE are used for the statistics of meteorological data. <https://doi.org/10.1038/s41558-024-02035-w>

Table 1. Data source for the study area.

Date	Source
Temp	ECMWF/ERA5_LAND/MONTHLY
LST	MODIS/061/MOD11A2
Pressure	ECMWF/ERA5_LAND/MONTHLY
WS	ECMWF/ERA5_LAND/MONTHLY
RH	ECMWF/ERA5_LAND/MONTHLY
Dew	ECMWF/ERA5_LAND/MONTHLY
Pre	ECMWF/ERA5_LAND/MONTHLY
ET	MODIS/006/MOD16A2
Snow	MODIS/061/MOD10A1
GPP	MODIS/061/MYD17A3HGF
AOD	MODIS/061/MOD08_M3
Lai	MODIS/061/MCD15A3H
Fire	MODIS/061/MOD14A1

Date is the data name, and Source is the Source of the data obtained. GEE platform data source link:

<https://earthengine.google.com/>

<https://doi.org/10.1371/journal.pone.0334535.t001>

MOD16A2.061: It is an evapotranspiration and heat flux product with a temporal resolution of 8 days and a spatial resolution of 500 meters. In this study, the “ET” band of the “MODIS/006/MOD16A2” dataset in GEE is used to conduct statistics on ET data.

MODIS/061/MOD11A2: It is a LST product with a temporal resolution of 8 days and a spatial resolution of 500 meters. This study uses the “LST_Day_1km” band of the “MODIS/061/MOD11A2” dataset in GEE to calculate the LST data.

MOD17A2H: It is a gross primary productivity product with a temporal resolution of 8 days and a spatial resolution of 500 meters. In this study, the “Gpp” band of the “MODIS/006/MOD17A2H” dataset in GEE is used for the statistics of gross primary productivity.

MODIS/061/MOD08_M3: This data product contains atmospheric parameters related to atmospheric aerosol particle properties, total ozone load, atmospheric water vapor, cloud optical and physical properties, and atmospheric stability indices. This dataset plays an important role in studying atmospheric environmental changes.

MODIS/061/MCD15A3H: This data product includes LAI with a spatial resolution of 500 meters. The LAI variable is defined as the equivalent number of leaf layers per unit ground area. In this study, the LAI data from the MODIS/061/MCD15A3H dataset on the GEE platform is selected to study the influencing factors.

MOD14 is an important product of the MODIS (Moderate Resolution Imaging Spectroradiometer) fire detection and thermal anomaly dataset. It has a temporal resolution of daily and a spatial resolution of 1 km.

MODIS/061/MOD10A1 data product provides global snow cover information and belongs to the snow cover dataset with a spatial resolution of 500m. Snow cover is usually expressed as a percentage, indicating the proportion of snow in each pixel. Snow cover data has important applications in meteorology, climate research and other fields.

The vector boundaries of urban built-up areas were extracted from the China urban built-up areas 2020 dataset. Using ArcMap 10.8, the geographic coordinates (latitude and longitude) of each city’s built-up area centroid were calculated using spatial analysis tools.

Research methods

Data preprocessing. Remote sensing data processing: The geemap package was used to call and process datasets on the GEE platform in Python 3.8.

Sentinel-5P data processing: Remote sensing data vary in spatial and temporal resolution. Through the GEE platform, Sentinel-5P data were uniformly resampled and exported at a resolution of 1 kilometer. The processing chain involves converting L2 data into L3 data gridded by latitude and longitude using tools such as harpconvert, filtering out low-quality pixels, masking negative values, generating annual and quarterly averages, and performing statistics aggregation within built-up area boundaries.

Data preparation: The Pandas library in Python 3.8 was used for data organization and filtering, and the sklearn package was applied to impute missing values in the original dataset. This ensures the overall quality of the data and facilitates the establishment of a database.

Statistical analysis. This study examined TNO₂CC across 346 major Chinese cities from 2019 to 2023. Raster datasets were homogenized to compute annual and seasonal averages, with seasons defined as follows: spring (March–May), summer (June–August), autumn (September–November), and winter (December–February). For the calculation of interannual mean values, we take into account that satellite observations are affected by factors such as cloud cover and ice/snow, resulting in uneven data availability across months. Therefore, when calculating seasonal mean values and annual mean values, we first compute the seasonal averages, and then use the arithmetic mean of these seasonal averages to derive the annual mean values. This approach ensures that each season contributes equally to the final annual result. ANOVA was applied to assess interannual differences in TNO₂CC, accompanied by characteristic maps illustrating inter annual and seasonal variation patterns.

Spatial analysis. Local spatial autocorrelation analysis: The local spatial autocorrelation method effectively detected spatial heterogeneity in urban atmospheric pollution patterns by identifying the geographic locations of pollution clusters and their aggregation types, such as high-high (HH) and low-low (LL) clusters. The calculation formulas were as follows:

$$I = \frac{(x_i - \bar{x})}{S^2} \sum_{j=1, j \neq i}^n W_{ij} (x_j - \bar{x}) \quad (1)$$

$$S^2 = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2)$$

where I represents the local spatial autocorrelation index, x_i denotes the first element of the attribute, \bar{x} represents the mean value of the nitrogen dioxide, W_{ij} is the spatial weighting matrix, S^2 represents the variance of the attribute, N is the number of elements.

Based on the local Moran's I index, this study used LISA maps to identify regional clustering patterns, categorizing spatial associations into four types: high-high (HH) clusters, low-low (LL) clusters, high-low (HL) spatial outliers, and low-high (LH) spatial outliers [31]. This method enables the detection of statistically significant localized clusters and spatial anomalies, offering critical insights into the spatial heterogeneity of atmospheric NO_2 pollution across China's urban agglomerations [32–33]. LISA cluster analysis was further refined [34]. Each cluster type was defined as follows:

HH clusters: Areas where both the target region and its neighboring zones exhibit elevated TNO_2CC .

HL outliers: Areas with high TNO_2CC levels surrounded by regions with low concentrations.

LH outliers: Areas with low TNO_2CC adjacent to high-pollution neighbors.

LL clusters: Areas where both the target region and its surrounding area consistently demonstrate low TNO_2CC levels.

Non-significant clusters: Regions lacking statistically significant spatial autocorrelation.

Subsequently, based on these four aggregation types, cities were classified and analyzed according to their temporal stability over five years period. This classification distinguished cities that consistently maintained a specific type of aggregation, those that changed once, and those that have undergone multiple transitions.

The Random Forest (RF) model

The RF model is an ensemble learning algorithm [35], proposed by Breiman and Cutler in 2001, which uses decision trees as its base learners. During prediction, the RF algorithm employs the Bootstrap resampling method to draw samples from the original dataset. By iteratively constructing multiple decision trees through sampling with replacement, the model aggregates predictions from each tree and determines the final output via majority voting, thereby ensuring robust generalization performance through ensemble learning [36]. Compared to the traditional linear model, the RF model effectively captures complex interactions among various variables, offers fast training speeds, and does not require a predefined functional form [37]. Additionally, its simple structure and relatively few tuning parameters make it well-suited for multidimensional, multi-factor prediction tasks while delivering highly accurate prediction results [38].

Using the Random Forest (RF) model, follow these steps:

First, determine the 'n_estimators' parameter. Then, further automate parameter tuning using grid search, setting the parameter tuning range. Adjust the 'max_depth' parameter by establishing a tuning interval and using grid search for experimentation. When 'max_depth' reaches the model's highest score, if this score is lower than when only 'n_estimators' is set, the model should not use the 'max_depth' parameter. Adjust the 'min_samples_leaf' parameter with a defined tuning range and grid search. When 'min_samples_leaf' achieves the model's highest score, if this score is lower than that of the model with only 'n_estimators' set, the model should not use the 'min_samples_leaf' parameter. Adjust the

`min_samples_split` parameter by setting a tuning interval and applying grid search. When `min_samples_split` reaches the model's highest score, if this score is lower than the model with only `n_estimators` configured, it indicates that adjusting `min_samples_split` can no longer optimize the model. Finally, adjust the `max_features` parameter using grid search. When `max_features` attains the model's highest score, the model reaches its optimal performance.

This study employed the PSO algorithm to optimize the hyperparameters of the RF model, facilitating an efficient exploration of the hyperparameter space to identify optimal configurations a. The RF model was trained using month average TNO₂CC (2019–2023) and 15 explanatory variables.

Shapley additive explanations. SHAP is an additive explanation model that evaluates the impact of in-put variables on model predictions [39], SHAP quantifies the relative im-portance of input variables by assessing the average variation in model outputs due to changes in those variables [40], This is achieved through scatter plots and SHAP value distributions, which visualize variable contributions, model performance, and any biases in the estimates [41]. Applying SHAP values to interpret the optimized RF model provides deeper insights into the relative contributions of individual factors during training [42].

Assuming the j -th predictor variable of the i -th target variable is denoted as x_{ij} , the model's predicted value for the i -th target variable is y_i , and the average predicted value across all target variables is y_{base} , the SHAP value adheres to the following formula [43]:

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + f(x_{i3}) + \dots + f(x_{ij}) \quad (3)$$

where $f(x_{ij})$ represents the SHAP value of the j -th predictor variable for the i -th target variable, indicating the marginal contribution of this predictor to the model's prediction of the target variable. In this study, the target variable was the TNO₂CC, with 15 explanatory variables as variable. Absolute SHAP values measure the magnitude of influence exerted by each predictor on the model's output, enabling variable importance ranking. A higher absolute SHAP value indicates a greater impact of the corresponding predictor on TNO₂CC variability.

Technical ideas

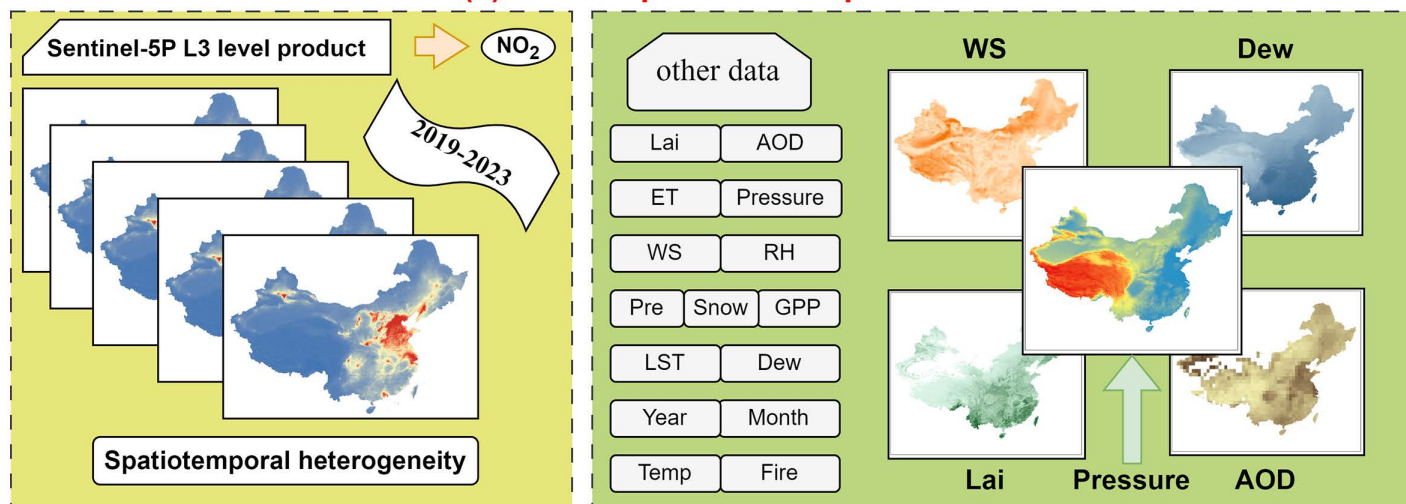
At present, many scholars use TROPOMI remote sensing inversion products to analyze and estimate the spatiotemporal concentration of air pollutants [44]. In this study, TROPOMI-driven TNO₂CC data were used to explore its spatial and temporal distribution and natural influencing factors. The research consists of three main components, as shown in Fig 2:

First, using seasonal and annual time windows, statistical models such as analysis of variance (ANOVA) and Tamhane's T2 post-hoc tests were applied to analyze the temporal evolution patterns of TNO₂CC across 346 major Chinese cities from 2019 to 2023. Second, local Moran's I and LISA cluster analysis were employed to identify high- and low-value spatial agglomerations and autocorrelation patterns.

TNO₂CC (2019–2023) were used as the dependent variable, while 15 natural driving factors, including meteorological, vegetation, and anthropogenic indices, were used as variable. The model was optimized via Particle Swarm Optimization (PSO), resulting in the following hyperparameters: $n_estimators=4953$, $max_depth=18$, $min_samples_split=4$, $min_samples_leaf=1$. By integrating the built-in feature importance ranking of the RF model with SHAP interpretation algorithms, this study identified key variables of air pollution. The top five natural influential factors were selected to generate partial dependence plots (PDPs), illustrating their nonlinear relationships with the dependent variable (TNO₂CC). Analyzing these relationships provides a basis for formulating effective urban air pollution prevention and control strategies in China.

This study used TNO₂CC from 346 cities as the target variable. Based on existing research by domestic and international scholars on natural factors influencing atmospheric pollutant levels, 15 variables were selected, including Pressure, Lai, ET, and others, to train the RF model. Importance was ranked using the SHAP importance metric, quantified by the mean absolute SHAP value for each factor. Key natural drivers were subsequently identified, and their marginal effects on TNO₂CC were analyzed through SHAP value decomposition, revealing nonlinear relationships and threshold behaviors in pollution dynamics.

(a) Data acquisition and pre-treatment



(b) Random Forest training

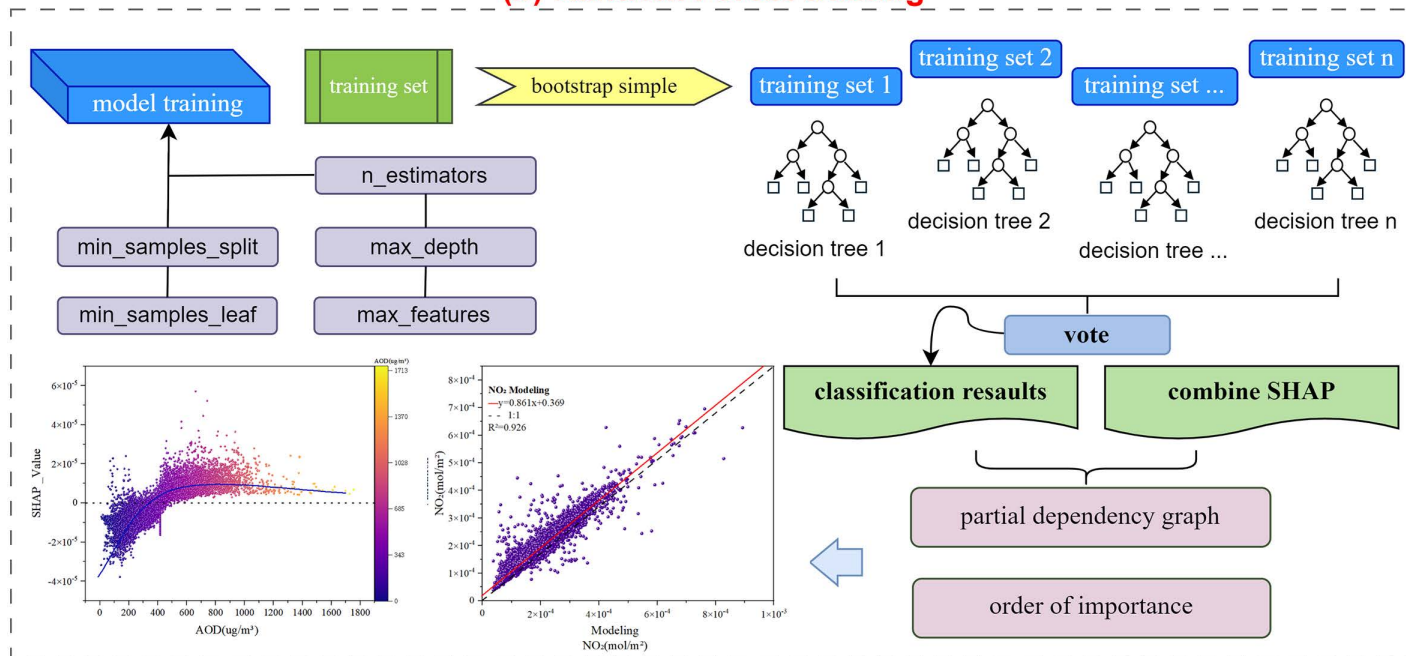


Fig 2. Technology roadmap. The administrative division data in GeoJSON format is sourced from the National Geospatial Information Public Service Platform (Tianditu), with the website: <https://cloudcenter.tianditu.gov.cn/administrativeDivision>. The map approval number is GS (2024) 0650. The data coordinates are in GCS_WGS_1984.

<https://doi.org/10.1371/journal.pone.0334535.g002>

Results

Temporal distribution of TNO₂CC

The interannual variability characteristics of TNO₂CC in major Chinese cities are illustrated (Fig 3). From 2019 to 2023, the average annual NO₂ concentration exhibited a wave-shaped trend, initially increasing, then decreasing, and rising again. Between 2019 and 2021, TNO₂CC gradually increased, with a smaller difference between 2020 and 2021.

compared to 2019. The average annual concentrations were $1.29 \times 10^{-4} \text{ mol/m}^2$, $1.42 \times 10^{-4} \text{ mol/m}^2$, and $1.47 \times 10^{-4} \text{ mol/m}^2$ for 2019, 2020, and 2021, respectively. The highest concentration over these three years occurred in 2021. In 2022, NO_2 concentration dropped significantly to $1.31 \times 10^{-4} \text{ mol/m}^2$. In 2023, concentrations increased slightly to $1.33 \times 10^{-4} \text{ mol/m}^2$.

This study applied ANOVA and Tamhane's T2 post-hoc tests to examine seasonal averages of TNO_2CC across major Chinese cities from 2019 to 2023. This approach enabled a rigorous comparison of interannual and intraseasonal variability in NO_2 pollution patterns, revealing statistically significant differences ($p < 0.05$) across climatic zones and urbanization gradients. The results in Table 2 confirm that significant differences in mean TNO_2CC among the five years within each season. Notably, spring and summer exhibited relatively lower interannual variability, whereas winter demonstrated the most pronounced variability, possibly due to intensified heating emissions and stagnant meteorological conditions during the colder months, autumn is the season with the smallest difference in five years. The average annual concentration in autumn is the second highest value.

The seasonal variability of TNO_2CC in China's major cities is depicted (Fig 4). The TNO_2CC were ranked in descending order: winter > autumn > spring > summer. During 2019–2023, TNO_2CC demonstrated significant seasonal variations, with peak concentrations observed in winter, lowest levels in summer, and moderate values in autumn and spring. The marginal difference between autumn and spring contrasted sharply with the winter-summer disparity, highlighting the dominant influence of seasonal emission patterns such as meteorological stagnation and precipitation is relatively low, making

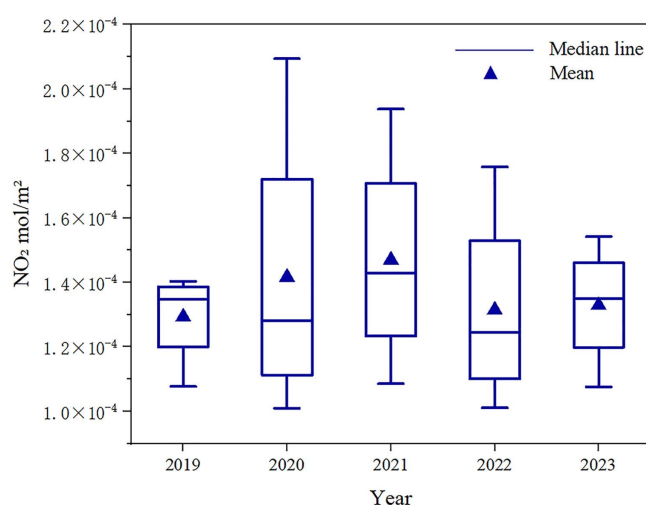


Fig 3. Interannual variation characteristics of troposphere nitrogen dioxide vertical column density in major Chinese cities.

<https://doi.org/10.1371/journal.pone.0334535.g003>

Table 2. Table of seasonal differences in five years.

Season	F	P
Spring	30.66	<0.05
Summer	16.76	<0.05
Autumn	8.38	<0.05
Winter	55.19	<0.05

Season represents the name of each Season, P value (< 0.05) represents the difference between each Season, and F value represents the degree of seasonal difference between each year. The greater the F value, the greater the difference; On the contrary, the smaller the difference.

<https://doi.org/10.1371/journal.pone.0334535.t002>

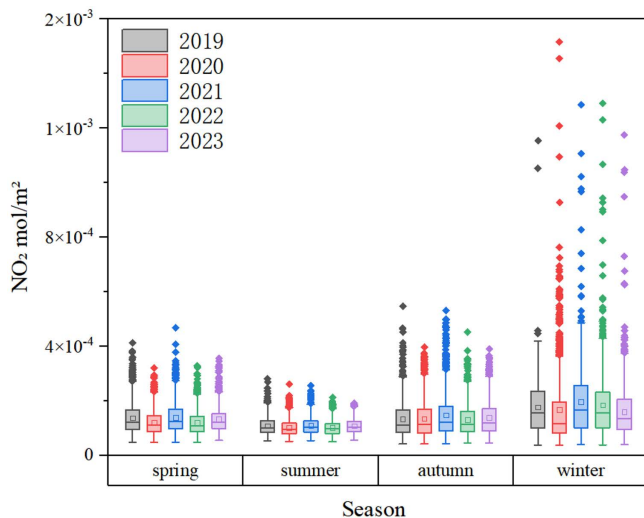


Fig 4. Seasonal variation characteristics of vertical column density of nitrogen dioxide in the troposphere of major cities in China. The points in the figure represent the annual average values of TNO₂CC in each city; The square represents the mean value, and the horizontal line represents the median value.

<https://doi.org/10.1371/journal.pone.0334535.g004>

it difficult for NO₂ pollutants to settle in winter. The figure shows more anomalies in winter, mainly due to seasonal effects, i.e., weaker photochemical sinks.

Spatial distribution characteristics of TNO₂CC

A local spatial autocorrelation analysis model was applied to generate LISA cluster maps, illustrating the spatial heterogeneity of air pollution severity within the study area and its surrounding regions. Fig 5 shows the interannual variation in the number of cities belonging to each cluster types over the five-year period. The number of cities classified as HH and LL clusters remained relatively stable. In contrast, LH clusters exhibited moderate interannual fluctuations, with annual counts of 2, 1, 3, 5, and 2 cities, respectively.

This spatial pattern is influenced by Yulin's unique geographic location. Although local TNO₂CC were relatively low, the city borders Shanxi Province to the east, a heavily industrialized region with elevated TNO₂CC levels, and Yan'an City to the south. It is located in the northernmost part of Shaanxi and serves as a border area connecting five provinces and regions: Shanxi, Gansu, Ningxia, Inner Mongolia, and Shanxi. HL clusters were observed only in 2019, with Harbin City in Heilongjiang Province as the sole representative (Fig 6).

The administrative division data in GeoJSON format is sourced from the National Geospatial Information Public Service Platform (Tianditu), with the website: <https://cloudcenter.tianditu.gov.cn/administrativeDivision>. The map approval number is GS (2024) 0650. The data coordinates are in GCS_WGS_1984.

Significant spatial disparities in TNO₂CC were evident across major Chinese cities (Fig 6), following a general decreasing gradient from east to west. High TNO₂CC zones were primarily located east of the Hu Line (Hu Huanxian Line) and north of the Yangtze River Basin.

Elevated TNO₂CC levels were observed in northern China. Similarly, high concentrations of TNO₂CC were detected in the eastern coastal areas, particularly in the Yangtze River Delta and the Pearl River Delta regions. In Northeast China, significant TNO₂CC accumulation was observed in metropolitan centers such as Harbin and Shenyang. Additionally, distinct regional hotspots of TNO₂CC were identified in northern Xinjiang (including Urumqi) and within the Sichuan Basin. The highlands surrounding the Sichuan Basin hinder the horizontal and vertical dispersion of air pollutants, leading to the

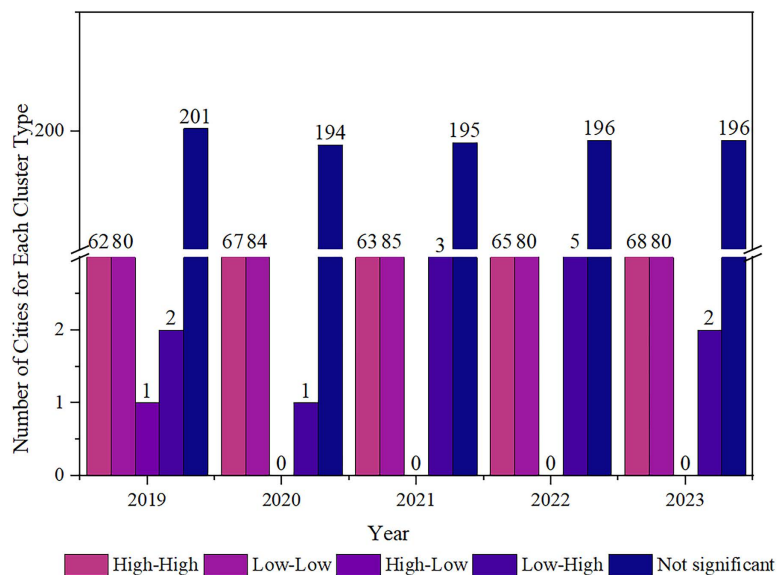


Fig 5. Changes in the number of cities of each agglomeration type in five years.

<https://doi.org/10.1371/journal.pone.0334535.g005>

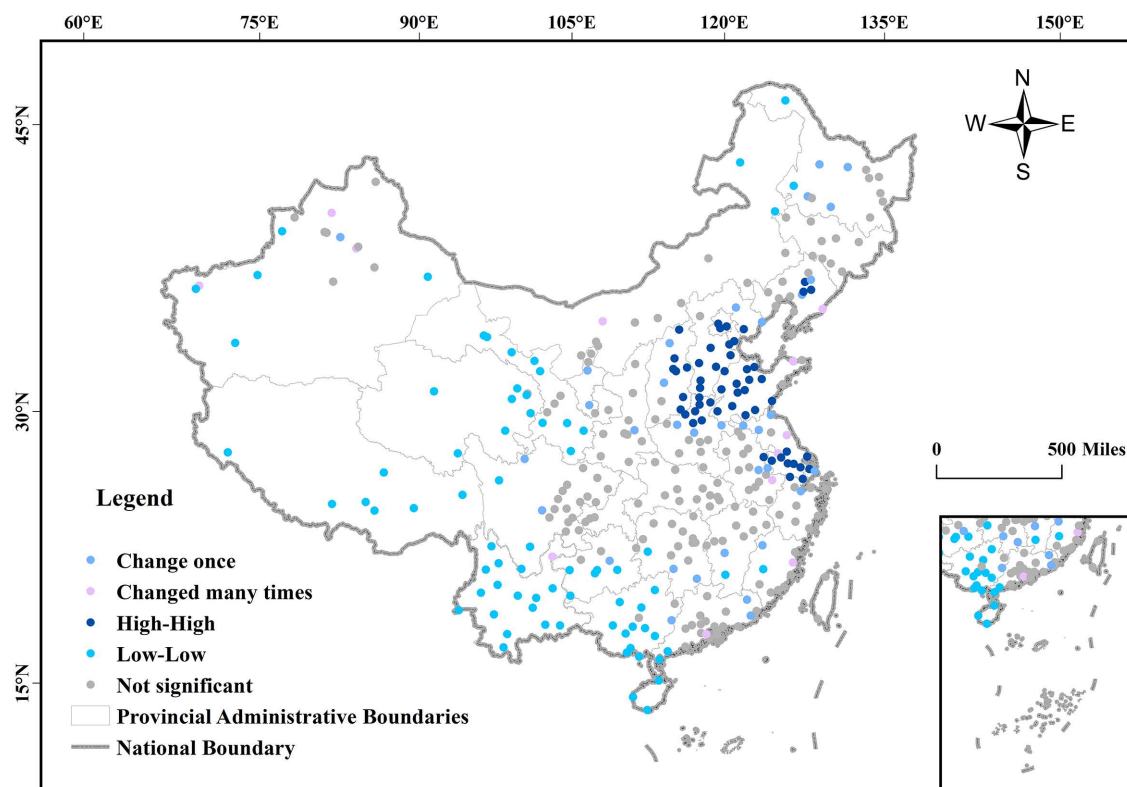


Fig 6. Local spatial autocorrelation analysis of vertical column concentration of tropospheric nitrogen dioxide in major cities of China.

<https://doi.org/10.1371/journal.pone.0334535.g006>

accumulation of pollutants and thus an increase in TNO_2CC . This effect is particularly pronounced in provincial capitals like Chengdu.

Low TNO_2CC are mainly located south of the Yangtze River and west of the Hu Line, covering regions such as Northwest China (including southern Xinjiang, Ningxia, and Qinghai), Southwest China (including Tibet, Yunnan, and Guizhou), Southeast China (such as Guangdong and Fujian), and Northeast China (including Heilongjiang and eastern Inner Mongolia). The spatial distribution of TNO_2CC across China shows a distinct east-west gradient, with significantly higher concentrations in the east and lower levels in the west.

Over the past five years, 54 cities consistently remained classified as HH clustering areas. These were primarily concentrated in Beijing, Shijiazhuang, Jinan, Shanghai, Nanjing, and Hangzhou, as well as Shenyang and Benxi in Liaoning Province. Such regions, particularly the traditional industrial base in northeast China centered on Shenyang, have long maintained persistently high TNO_2CC . In contrast, 74 cities remained LL clusters, characterized by lower industrialization levels, moderate urbanization, and geographic isolation from high-emission zones, minimizing cross-regional pollution influence. Additionally, 34 cities transitioned once in cluster classification, including Qinhuangdao, Chengde, Heihe, and Luoyang. Among these, 20 cities, such as Nanping, Meizhou, and Daqing, exhibited relatively stable clustering patterns for 3–4 years, while 14 cities, including Ma'an Shan, Jieyang, and Ya'an, exhibited non-significant clustering for 1–2 years but maintained a single cluster type in other years. A total of 12 cities, such as Fuzhou, Dandong, and Kizilsu Kirghiz Autonomous Prefecture, underwent multiple transitions in cluster classification. Notably, Xuancheng, Bayannur, Zhaotong, and four prefecture-level cities in Xinjiang, located near provincial capitals, and Zhongshan city near the Pearl River Delta, exhibited unstable clustering due to significant influence from adjacent high-emission zones. Coastal cities such as Fuzhou, Yancheng, Dandong, Yantai, and Yangzhou, which are situated along the Huaihe River's waterway, also exhibited heightened variability in cluster types. This variability was attributed to unstable meteorological conditions, including fluctuations in pressure, temperature, humidity, and WS.

From the perspective of climate zones, as shown in Fig 7, cities with high TNO_2CC are mainly concentrated in the II and III climate zones east of the Hu Huanyong Line, followed by the IV climate zone and the northern part of the V climate zone. Cities with low TNO_2CC are mainly distributed in the II and III climate zones west of the Hu Huanyong Line, as well as the I climate zone, the southern part of the V climate zone, the VI climate zone, and the VII climate zone.

Natural factors influencing tropospheric TNO_2CC in Chinese cities

Using the processed month average TNO_2CC from 2019 to 2023, the following explanatory natural variables were selected: Pressure, Lai, Dew, AOD, WS, Month, ET, RH, Pre, Fire, LST, Temp, GPP, Year, and Snow. These variables formed the basis for constructing an optimized RF model to analyze the spatiotemporal drivers of TNO_2CC variability. Predictor importance was quantified using mean SHAP values, with the top five influential factors selected to generate PDPs. The RF model, combined with SHAP interpretation algorithms, facilitated an analysis of key drivers and used PDPs to isolate the marginal relationships between these natural factors and TNO_2CC while controlling for other variables. As shown in Fig 8, the results showed that the slopes of the modeling group and the validation group were 0.86 and 0.74, respectively, with R^2 values of 0.94 and 0.76. The data points of both the modeling group and the validation group were close to the 1:1 fitting line, as shown in Fig 8, indicating that the obtained RF model has a high degree of fitting.

Fig 9 illustrates the overall importance of each natural variable, with the y-axis representing ranked variable importance and the x-axis indicating mean SHAP values. The analysis demonstrated that Pressure exerted the strongest influence on TNO_2CC , followed by Lai, ET, Dew, and AOD. Based on this ranking, the top five factors, including Pressure, Lai, ET, Dew, and AOD, were identified as key natural drivers for in-depth analysis, while the remaining variables contributed minimally to the model's explanatory power (Fig 9). The parameter tuning results are as follows: S1 File.

The study revealed nonlinear relationships between these key drivers and TNO_2CC by applying PDPs to analyze the top five most influential factors. Pressure was identified as the most important natural variable. As Pressure increased,

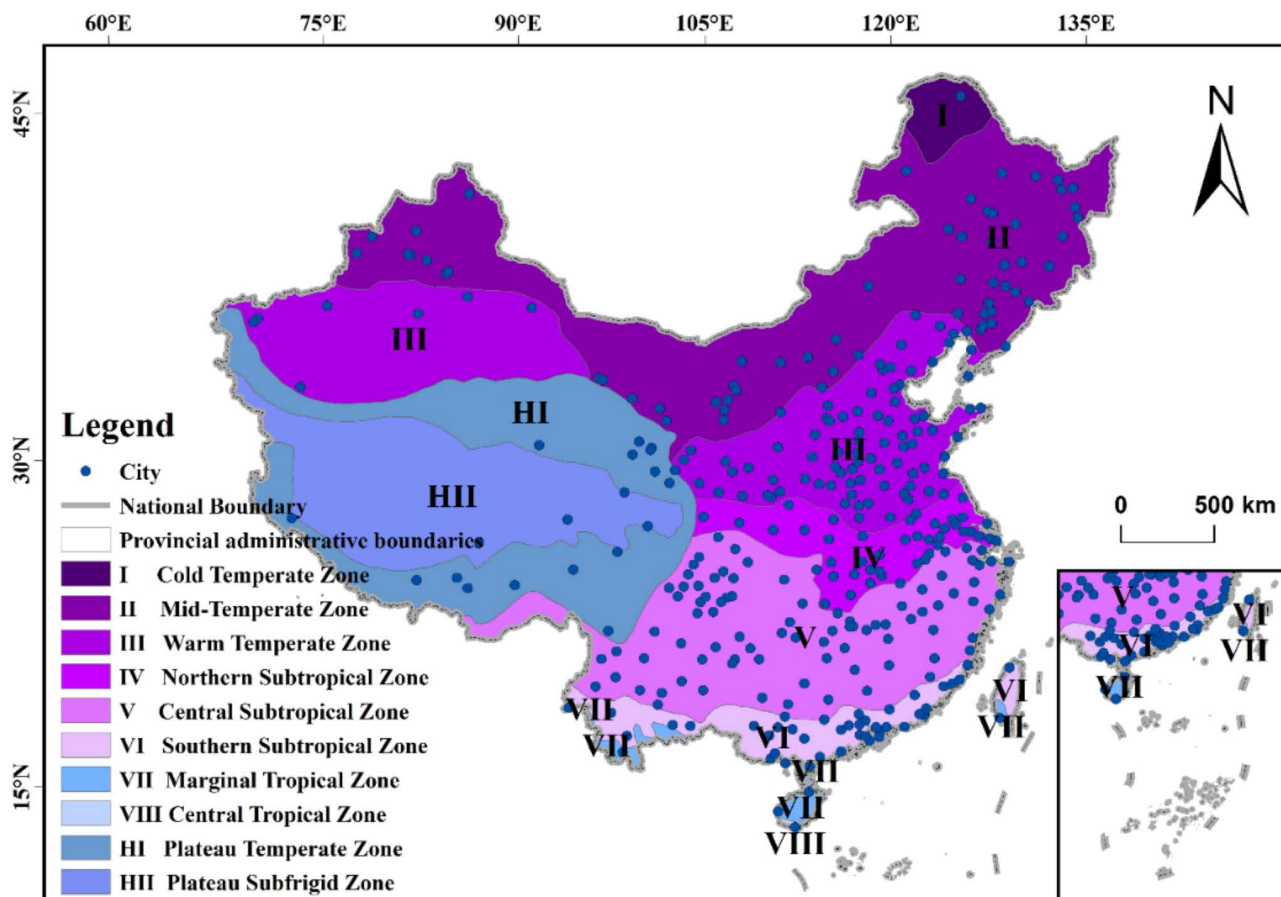


Fig 7. Map of China's climate zoning. The administrative division data in GeoJSON format is sourced from the National Geospatial Information Public Service Platform (Tianditu), with the website: <https://cloudcenter.tianditu.gov.cn/administrativeDivision>. The map approval number is GS (2024) 0650. The data coordinates are in GCS_WGS_1984.

<https://doi.org/10.1371/journal.pone.0334535.g007>

TNO₂CC exhibited an overall upward trend (Fig 10a). The relationship between Pressure and TNO₂CC followed a wave-shaped pattern, rapidly escalating TNO₂CC response magnitude when Pressure reached approximately 95,000 Pa. A pronounced positive correlation was observed in high-Pressure regions (> 95,000 Pa), indicating that elevated pressure consistently coincided with increased TNO₂CC levels. The response of TNO₂CC to AOD can be divided into two stages (Fig 10b). In the lower AOD range, TNO₂CC sensitivity to AOD increased rapidly. Beyond an AOD value of approximately 330, the response gradually stabilized. TNO₂CC participates in photochemical reactions in the atmosphere with other compounds, forming secondary aerosols. These secondary aerosols increase PM concentration, thereby elevating AOD values. The influence of Lai on TNO₂CC exhibited a nonlinear trend, initially declining and then stabilizing (Fig 10c). Lower Lai values correspond to higher TNO₂CC, while higher Lai values correspond to lower concentrations. ET exhibited a negative correlation with TNO₂CC (Fig 10d). As ET increases, TNO₂CC decreases, following an overall nonlinear trend of an initial sharp decline followed by gradual stabilization. Additionally, Dew also showed a negative correlation with TNO₂CC overall, characterized by a brief increase followed by a continuous decrease, dividing the response into two phases (Fig 10e). Below approximately 263 K (low Dew range), Dew and TNO₂CC exhibited a positive correlation. Above 263 K (high Dew range), the relationship shifts from positive to negative correlation, with TNO₂CC gradually decreasing as Dew increases.

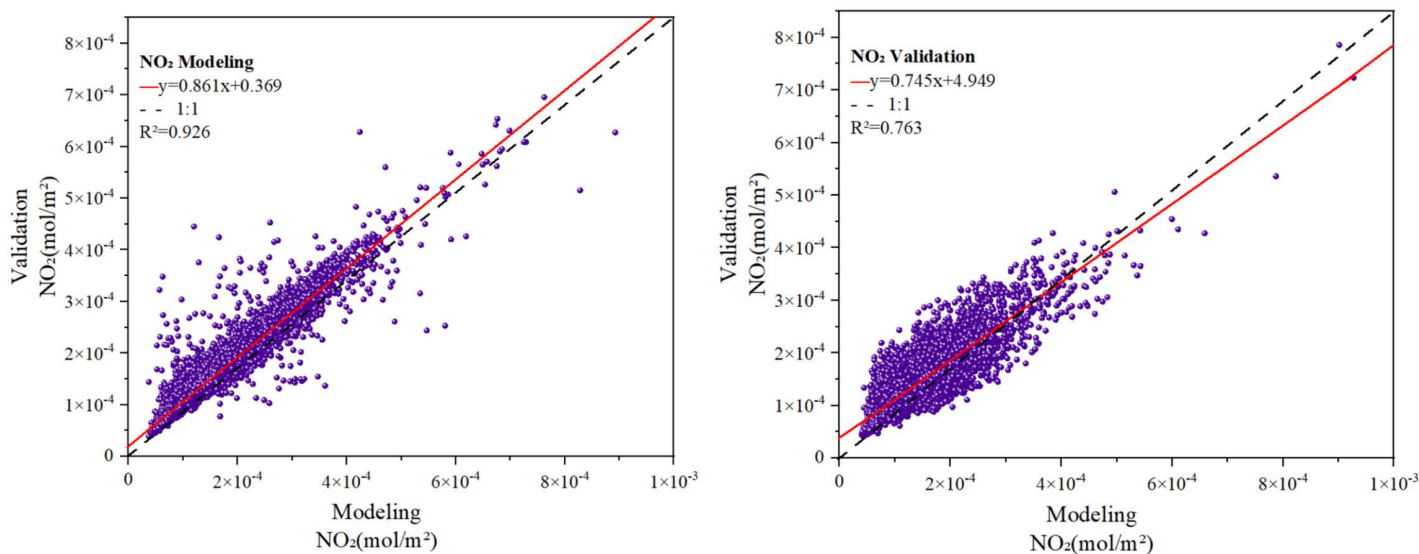


Fig 8. Performance of the Random Forest (RF) model for nitrogen dioxide concentration. (a): modeling, (b): validation.

<https://doi.org/10.1371/journal.pone.0334535.g008>

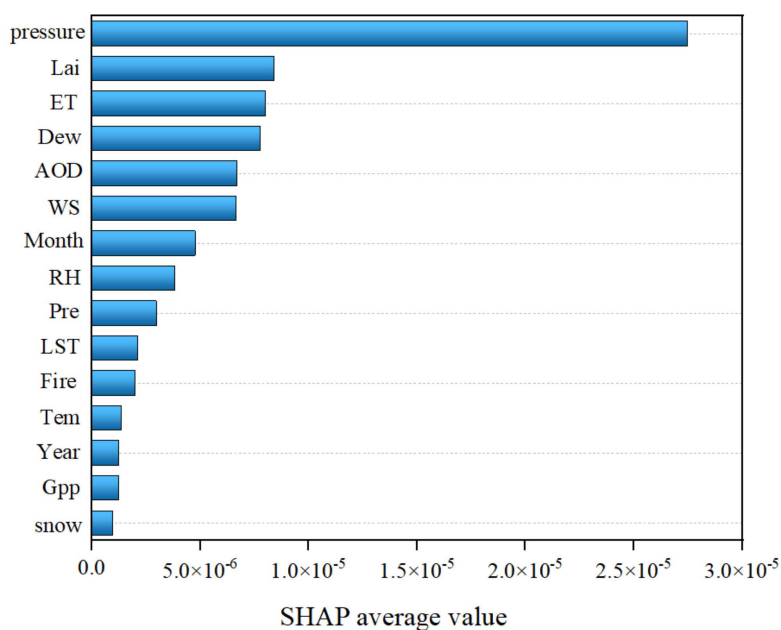


Fig 9. Importance ranking of the affecting factors based on RF-SHAP model.

<https://doi.org/10.1371/journal.pone.0334535.g009>

Discussion

A previous study applied the geographically weighted regression model to identify the the determinants of $PM_{2.5}$ concentration and explore variations in atmospheric pollutants [45]. Other research has employed principal component analysis, concluding that effective pollution control requires coordinated management of major pollutants, such as PM_{10} , $PM_{2.5}$, and O_3 [46]. The RF model offers several advantages, including low sensitivity to parameters [47], strong robustness

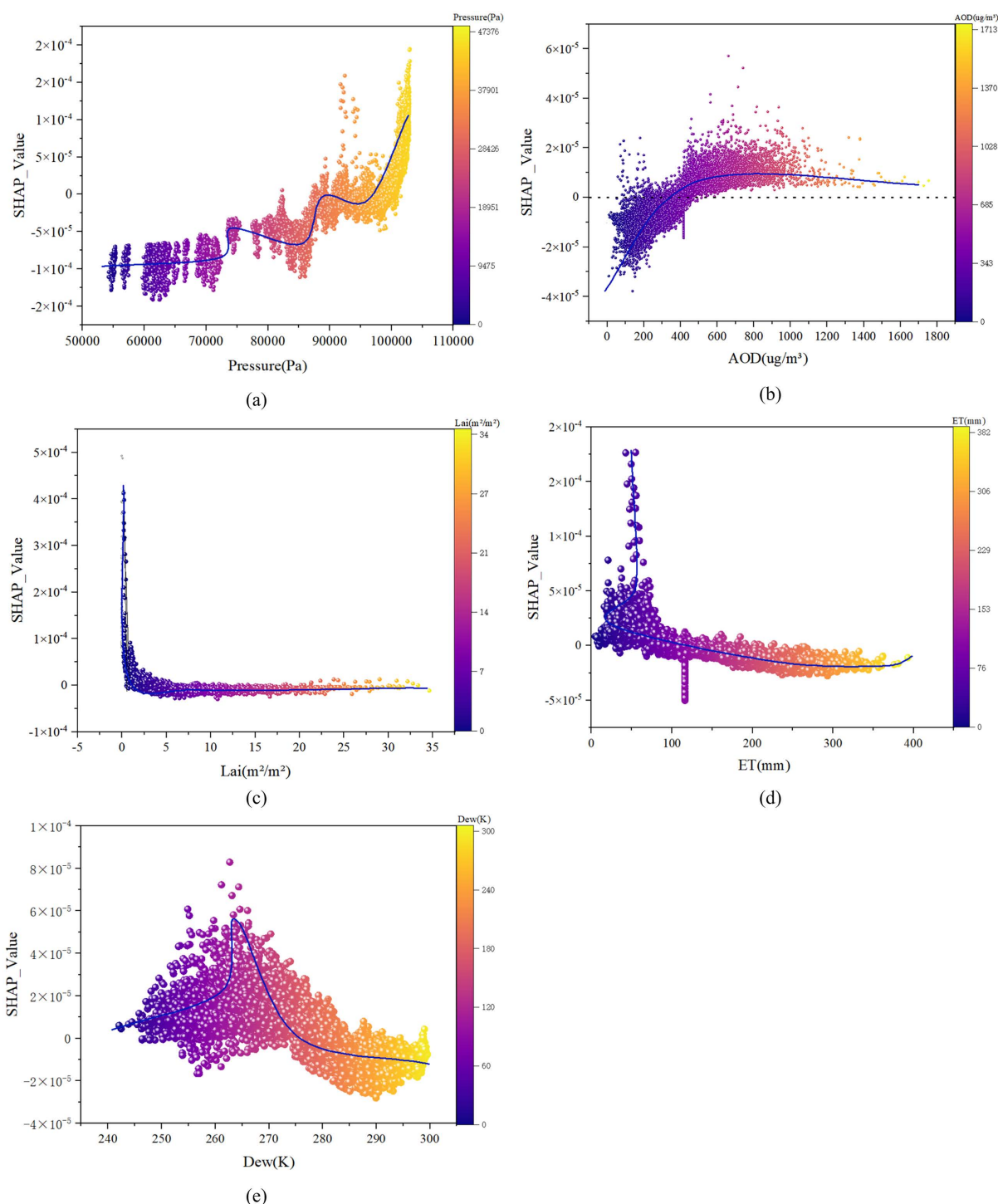


Fig 10. Influence intensity of Pressure(a), AOD(b), Lai(c), ET(d), Dew(e) on nitrogen dioxide concentration. (a): Pressure: atmospheric pressure; (b): AOD: aerosol optical depth; (c) Lai: Leaf Area Index; (d) ET: evapotranspiration; (e): Dew: dew point temperature. The blue line represents the trend of the impact of key influencing factors on TNO₂CC variation. SHAP values indicate the absolute effect size of features on the target variable, where the magnitude directly reflects each feature's marginal contribution to TNO₂CC variations. A SHAP value > 0 indicates a positive contribution, confirming that the factor's influence on TNO₂CC change increases as the SHAP value rises. Conversely, a SHAP value < 0 indicates a negative contribution, where a decreasing SHAP value corresponds to a stronger negative impact on TNO₂CC changes. Data on other influencing factors, except for the top five, can be found in [S1 File](#).

<https://doi.org/10.1371/journal.pone.0334535.g010>

against missing values, and improved computational efficiency through optimized variable selection [48]. The PSO algorithm exhibits excellent optimization performance [49]. This is particularly true in the field of environmental monitoring—for instance, in aspects such as pollutant concentration monitoring and water quality monitoring—where it demonstrates excellent performance [50]. This study employed an optimized RF model to identify key natural factors influencing TNO₂CC, providing theoretical foundations and data-driven insights for atmospheric pollution control.

This study analyzed correlation coefficients between natural variables to gain deeper insight in-to the factors and the mechanisms influencing TNO₂CC (Fig 11), the sample points comprised monthly average data from 346 cities spanning 2019–2023. Analysis of 15 natural variables determined the correlation analysis between variable pairs. The results indicated strong correlations (coefficients: > 0.65) between Lai and ET, GPP, Dew, Pre, LST, and Temp, whereas weak correlations (coefficients: > 0.65) were observed between AOD, Year, and other natural factors [51]. These findings suggest that the combined effects of vegetation and atmospheric processes mainly influence TNO₂CC changes.

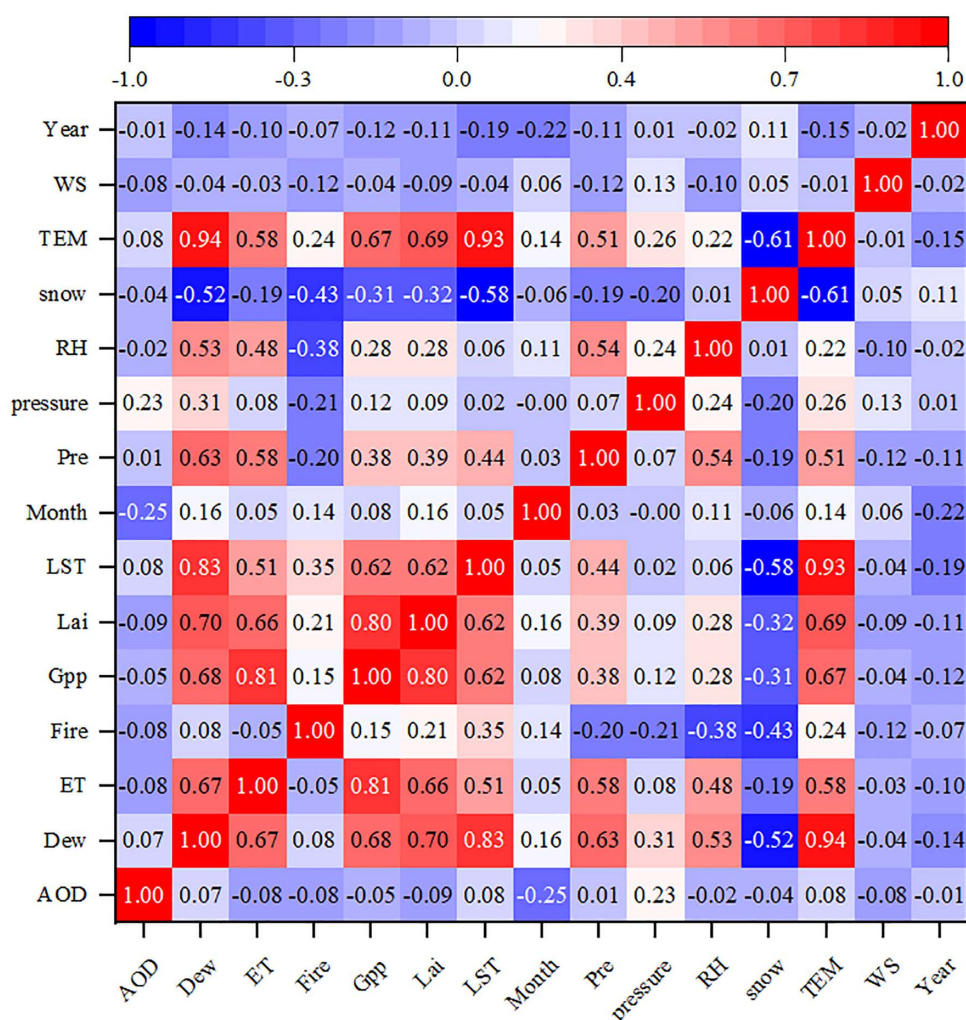


Fig 11. Thermodynamic diagram of NO₂ influencing factors. AOD (aerosol optical depth), Dew (dew point temperature), ET (evapotranspiration), Fire (fire activity), GPP (gross primary productivity), Lai (Leaf Area Index), LST (land surface temperature), Month(month), Pre (precipitation), Pressure (atmospheric pressure), RH (relative humidity), Snow (snow cover), TEMP (air temperature), WS (wind speed), Year (year).

<https://doi.org/10.1371/journal.pone.0334535.g011>

The PSO algorithm was applied to optimize the RF model's hyperparameters and evaluate its robustness. Model performance was confirmed by calculating the R^2 using a linear regression model, indicating reasonable fitting accuracy (Fig 8). The built-in feature importance ranking function of the RF model was employed to prioritize 15 natural factors, including Pressure, Lai, Dew, ET, RH, AOD, Fire, GPP, Temp, Year, Month, Pre, LST, WS, and Snow, across the studied cities. Combined with the SHAP interpretability algorithm (Fig 9). The analysis identified Pressure, AOD, Lai, ET, and Dew as the top five factors influencing TNO_2CC variations, emphasizing the significant roles of pressure and vegetation in urban NO_2 pollutant dynamics. This finding contrasts with Wu et al who re-reported Temp and Pre as dominant natural drivers of NO_2 variability in the Shaanxi-Gansu-Ningxia region [52]. The discrepancy possibly arises from the region's unique topographic and climatic characteristics, highlighting the spatial heterogeneity of NO_2 pollution mechanisms.

TNO_2CC exhibited strong seasonal variability, peaking in winter and declining in summer (Fig 4). This pattern is attributed to elevated pressure and lower temperatures in winter, which reduce the planetary boundary layer height and hinder pollutant dispersion [53]. Under such stable air conditions, NO_2 air pollutant accumulates [54]. Moreover, winter is characterized by static and stable weather conditions with low WS, leading to increased concentrations, whereas summer conditions promote better diffusion [55]. Although the annual average value in autumn is higher than that in summer and spring, the interannual difference is the smallest, as shown in Table 2. This phenomenon may occur because the meteorological conditions (such as WS, Temp and Pre) in autumn are more stable between years, so even if the TNO_2CC is high, its interannual variation may be small. As shown in Fig 3, the interannual trend from 2019 to 2023 displays a wave-like pattern in annual average TNO_2CC . This pattern aligns with Zhang et al, who reported similar TNO_2CC temporal variability in China, possibly influenced by prolonged high-Pressure systems during an extreme cold wave in December 2021 [56]. Notably, 2022 recorded the lowest annual TNO_2CC within five years, followed by a rebound in 2023. Meteorological reports attribute the decline in 2022 to weaker cold air activity compared to previous years [57], collectively reducing emissions. In contrast, stronger cold air processes in 2023 may explain the renewed increase in NO_2 levels. Based on the analysis of climate zones, the impact of meteorological factors on the changes in TNO_2CC has also been confirmed. The warm temperate zone features distinct four seasons, moderate precipitation, slightly cold winters and hot summers. In winter, static and stable weather conditions and temperature inversion phenomena often occur, making it difficult for near-surface pollutants (such as NO_2 and $\text{PM}_{2.5}$) to diffuse [58]. In summer, photochemical reactions and precipitation can remove part of the NO_2 in the atmosphere and reduce its concentration [59].

The model identified LAI, ET, and Dew as three of the top five natural influencing factors, all of which are vegetation-related variables. Correlation analysis underscores vegetation's role as a key regulator of TNO_2CC . LAI, defined as the total one-sided green leaf area per unit of ground surface [60], enhances photosynthetic efficiency, thereby reducing atmospheric TNO_2CC through direct uptake and by facilitating wet deposition during Pre events [61]. Seasonally, higher LAI values in summer correspond to lower TNO_2CC levels due to increased vegetation absorption and ET-driven rainfall. Spatially, southern China's dense vegetation exhibits stronger NO_2 removal capacity compared to northern regions, highlighting the importance of urban greening and vegetation coverage in shaping TNO_2CC spatiotemporal patterns [62]. In addition to meteorological and vegetation factors, fire activity also has a certain impact on changes in TNO_2CC . However, according to the ranking results, the effect is relatively limited. Therefore, this factor will only cause a slight change in TNO_2CC under special circumstances (such as urban fires) [63].

However, it is worth noting that this study focuses on the temporal-spatial distribution of TNO_2CC and the research on its natural influencing factors. Nevertheless, it must be acknowledged that natural factors play a regulatory role in the influencing process, while human factors are the fundamental source [64]. This is an aspect that needs improvement in our future research—we should fully consider the impacts of other social factors. Nowadays, emission inventory data has become well-developed. Examples include the U.S. National Emission Inventory (NEI), the UK National Atmospheric Emission Inventory (NAEI), and China's Multi-resolution Emission Inventory for China (MEIC). Because of this, more researchers are using these emission inventories to analyze where pollutants come from [65–68]. In future work, we can

use these inventories to study the spatial distribution of air pollution sources. In terms of data validation, China's ground-based air quality monitoring stations are already quite comprehensive, and the temporal resolution of the measured data is also very high. Some researchers have already used ground monitoring data to verify the usability of S5P data from other countries [69–72]. Next, we can conduct ground-based validation research on China's S5P data to ensure the reliability of the data. Our goal is to provide a more scientific theoretical basis for air pollution control.

Conclusions

This study analyzed the interannual and seasonal spatiotemporal distribution characteristics of TNO₂CC in 346 major Chinese cities from 2019 to 2023, using satellite remote sensing data, statistical methods, and local spatial autocorrelation analysis. The key findings are summarized as follows:

- (1) TNO₂CC exhibited significant temporal heterogeneity across the five years. Annual trends demonstrated an initial rise, followed by a decline and rebound. Seasonally, concentration remained consistently higher in winter and lower in summer.
- (2) TNO₂CC across 346 major Chinese cities displayed notable spatial clustering. High-concentration hotspots were mainly distributed in urban areas of the North China Plain, Yangtze River Delta, and northeastern China. In contrast, low-concentration zones were concentrated in northwestern, southwestern, southern coastal, and Inner Mongolia regions. Additionally, monsoon climates in eastern China contribute to increased summer Pre, which dilutes TNO₂CC levels, while centralized coal heating in northern regions during winters exacerbates TNO₂CC accumulation, resulting in pronounced seasonal contrasts.

An RF model incorporating 15 natural influencing factors was optimized through iterative hyperparameter tuning, achieving high predictive accuracy with training and validation datasets closely aligned along the 1:1 fit line. Feature importance ranking using the SHAP algorithm identified Pressure, AOD, Lai, ET, and Dew as the top five drivers of TNO₂CC variability. SHAP analysis revealed the following patterns: Pressure exhibited a positive correlation with TNO₂CC, following a wave-shaped upward trend. AOD transitioned from negative to positive effects on TNO₂CC at thresholds of approximately 330. Lai and ET displayed negative correlations, with TNO₂CC decreasing initially and stabilizing as these factors increased. Dew exerted a greater influence on TNO₂CC near 263 K. Three different temperature measurements, Dew, LST, Temp. They themselves have a strong correlation. This shows that TNO₂CC is strongly influenced by temperature.

Supporting information

S1 File. Results of random forest parameter tuning. This text presents the optimal parameters obtained after tuning the random forest using the particle swarm optimization algorithm.

(PDF)

S2 File. RF model parameter tuning code. This document contains the code for the parameter tuning process of the Particle Swarm Optimization (PSO) algorithm.

(DOCX)

S3 File. SHAP explanation algorithm code. This document contains the code for the SHAP explanation algorithm, which is used to analyze the response degree of various influencing factors to NO₂ concentration.

(DOCX)

S4 File. Raw data.

(ZIP)

Author contributions

Conceptualization: Feifei Cheng, Qing Sun, Heming Yang.

Data curation: Feifei Cheng, Qing Sun.

Formal analysis: Feifei Cheng, Qing Sun.

Funding acquisition: Heming Yang.

Methodology: Feifei Cheng, Qing Sun, Jiaqi Zhang, Jiahe Liu.

Supervision: Heming Yang.

Visualization: Feifei Cheng, Qing Sun, Jiaqi Zhang, Jiahe Liu, Wenwen Lü, Fapeng Shen.

Writing – original draft: Feifei Cheng, Qing Sun.

Writing – review & editing: Feifei Cheng, Qing Sun.

References

1. Wu Z, Tian Y, Li M, Wang B, Quan Y, Liu J. Prediction of air pollutant concentrations based on the long short-term memory neural network. *J Hazard Mater*. 2024;465:133099. <https://doi.org/10.1016/j.jhazmat.2023.133099> PMID: [38237434](https://pubmed.ncbi.nlm.nih.gov/38237434/)
2. Accarino G, Lorenzetti S, Aloisio G. Assessing correlations between short-term exposure to atmospheric pollutants and COVID-19 spread in all Italian territorial areas. *Environ Pollut*. 2021;268(Pt A):115714. <https://doi.org/10.1016/j.envpol.2020.115714> PMID: [33120339](https://pubmed.ncbi.nlm.nih.gov/33120339/)
3. Rahaman S, Tu X, Ahmad K, Qadeer A. A real-time assessment of hazardous atmospheric pollutants across cities in China and India. *J Hazard Mater*. 2024;479:135711. <https://doi.org/10.1016/j.jhazmat.2024.135711> PMID: [39255663](https://pubmed.ncbi.nlm.nih.gov/39255663/)
4. Ojeda-Castillo V, Murillo-Tovar MA, Hernández-Mena L, Saldarriaga-Noreña H, Vargas-Amado ME, Herrera-López EJ, et al. Tropospheric NO₂: anthropogenic influence, global trends, satellite data, and machine learning application. *Remote Sens*. 2024;17(1):49. <https://doi.org/10.3390/rs17010049>
5. Huangfu P, Atkinson R. Long-term exposure to NO₂ and O₃ and all-cause and respiratory mortality: A systematic review and meta-analysis. *Environ Int*. 2020;144:105998. <https://doi.org/10.1016/j.envint.2020.105998> PMID: [33032072](https://pubmed.ncbi.nlm.nih.gov/33032072/)
6. Rani B, Singh U, Chuhan AK, Sharma D, Maheshwari R. Photochemical smog pollution and its mitigation measures. *J Adv Res*. 2011;2:28–33.
7. Lim C-H, Ryu J, Choi Y, Jeon SW, Lee W-K. Understanding global PM_{2.5} concentrations and their drivers in recent decades (1998–2016). *Environ Int*. 2020;144:106011. <https://doi.org/10.1016/j.envint.2020.106011> PMID: [32795749](https://pubmed.ncbi.nlm.nih.gov/32795749/)
8. Voiculescu M, Constantin D-E, Condurache-Bota S, Călmuc V, Roșu A, Dragomir Bălănică CM. Role of meteorological parameters in the diurnal and seasonal variation of NO₂ in a romanian urban environment. *Int J Environ Res Public Health*. 2020;17(17):6228. <https://doi.org/10.3390/ijerph17176228> PMID: [32867209](https://pubmed.ncbi.nlm.nih.gov/32867209/)
9. Kusters MSW, Granés L, Petricola S, Tiemeier H, Muetzel RL, Guxens M. Exposure to residential air pollution and the development of functional connectivity of brain networks throughout adolescence. *Environ Int*. 2025;196:109245. <https://doi.org/10.1016/j.envint.2024.109245> PMID: [39848092](https://pubmed.ncbi.nlm.nih.gov/39848092/)
10. Coker ES, Cavalli L, Fabrizi E, Guastella G, Lippo E, Parisi ML, et al. The effects of air pollution on COVID-19 related mortality in Northern Italy. *Environ Resour Econ (Dordr)*. 2020;76(4):611–34. <https://doi.org/10.1007/s10640-020-00486-1> PMID: [32836855](https://pubmed.ncbi.nlm.nih.gov/32836855/)
11. Diener A, Mudu P. How can vegetation protect us from air pollution? A critical review on green spaces' mitigation abilities for air-borne particles from a public health perspective - with implications for urban planning. *Sci Total Environ*. 2021;796:148605. <https://doi.org/10.1016/j.scitoenv.2021.148605> PMID: [34271387](https://pubmed.ncbi.nlm.nih.gov/34271387/)
12. Lee Y-G, Lee P-H, Choi S-M, An M-H, Jang A-S. Effects of air pollutants on airway diseases. *Int J Environ Res Public Health*. 2021;18(18):9905. <https://doi.org/10.3390/ijerph18189905> PMID: [34574829](https://pubmed.ncbi.nlm.nih.gov/34574829/)
13. Yang R, Zhong C. Analysis on spatio-temporal evolution and influencing factors of air quality index (AQI) in China. *Toxics*. 2022;10(12):712. <https://doi.org/10.3390/toxics10120712> PMID: [36548545](https://pubmed.ncbi.nlm.nih.gov/36548545/)
14. Müller I, Erbertseder T, Taubenböck H. Tropospheric NO₂: explorative analyses of spatial variability and impact factors. *Remote Sensing of Environment*. 2022;270:112839. <https://doi.org/10.1016/j.rse.2021.112839>
15. Han J, Liu H, Zhu H, Xiong H. Kill two birds with one stone: a multi-view multi-adversarial learning approach for joint air quality and weather prediction. *IEEE Trans Knowl Data Eng*. 2023;35(11):11515–28. <https://doi.org/10.1109/tkde.2023.3236423>
16. Khajavi H, Rastgoo A. Predicting the carbon dioxide emission caused by road transport using a Random Forest (RF) model combined by Meta-Heuristic Algorithms. *Sustainable Cities and Society*. 2023;93:104503. <https://doi.org/10.1016/j.scs.2023.104503>

17. Forster PM, Smith CJ, Walsh T, Lamb WF, Lamboll R, Hauser M, et al. Indicators of Global Climate Change 2022: annual update of large-scale indicators of the state of the climate system and human influence. *Earth Syst Sci Data*. 2023;15(6):2295–327. <https://doi.org/10.5194/essd-15-2295-2023>
18. Gupta VK, Gupta A, Kumar D, Sardana A. Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model. *Big Data Min Anal*. 2021;4(2):116–23. <https://doi.org/10.26599/bdma.2020.9020016>
19. Wang T, Chen MP, Zhou MP, Li J, Wu SJ, Luo M. Pollution characteristics and influence factors of air pollutants in Wuxi city. *Environ Pollut & Control*. 2015;37:74–8. <https://doi.org/10.15985/j.cnki.1001-3865.2015.12.015>
20. Liu YJ, Li Y, Miao SG. Spatial-temporal characteristics and meteorological factors analysis of air pollution in Fang Shan District of Beijing. *Met&EnvSci*. 2018;41(4):60–9. <https://doi.org/10.16765/j.cnki.1673-7148.2018.04.009>
21. Kong FB, Cheng D, Zhang Q. Spatiotemporal Changes and Influencing Factors of Dust Fall in Gansu Province During the 13th Five-Year Plan Period. *Environ Monit China*. 2022;38:25–31. <https://doi.org/10.19316/j.issn.1002-6002.2022.03.04>
22. Li J, Chen H, Li Z, Wang P, Fan X, He W, et al. Analysis of low-level temperature inversions and their effects on aerosols in the lower atmosphere. *Adv Atmos Sci*. 2019;36(11):1235–50. <https://doi.org/10.1007/s00376-019-9018-9>
23. You T, Wu R, Huang G, Fan G. Regional meteorological patterns for heavy pollution events in Beijing. *J Meteorol Res*. 2017;31(3):597–611. <https://doi.org/10.1007/s13351-017-6143-1>
24. Islam ARMT, Hasanuzzaman M, Azad MAK, Salam R, Toshi FZ, Khan MSI, et al. Effect of meteorological factors on COVID-19 cases in Bangladesh. *Environ Dev Sustain*. 2021;23(6):9139–62. <https://doi.org/10.1007/s10668-020-01016-1> PMID: 33052194
25. Yin Z, Wang H, Chen H. Understanding severe winter haze events in the North China Plain in 2014: roles of climate anomalies. *Atmos Chem Phys*. 2017;17(3):1641–51. <https://doi.org/10.5194/acp-17-1641-2017>
26. Sangkham S, Thongtip S, Vongruang P. Influence of air pollution and meteorological factors on the spread of COVID-19 in the Bangkok metropolitan region and air quality during the outbreak. *Environ Res*. 2021;197:111104. <https://doi.org/10.1016/j.envres.2021.111104> PMID: 33798521
27. Chen ZQ. Air quality, influence factors and control countermeasure in Quanzhou, southeastern coast of China. *J Earth Environ*. 2019;10:201–9. <https://doi.org/10.7515/JEE182075>
28. Tui Y, Qiu J, Wang J, Fang C. Analysis of spatio-temporal variation characteristics of main air pollutants in Shijiazhuang city. *Sustainability*. 2021;13(2):941. <https://doi.org/10.3390/su13020941>
29. Guo L-C, Zhang Y, Lin H, Zeng W, Liu T, Xiao J, et al. The washout effects of rainfall on atmospheric particulate pollution in two Chinese cities. *Environ Pollut*. 2016;215:195–202. <https://doi.org/10.1016/j.envpol.2016.05.003> PMID: 27203467
30. Guo WW, Chen YJ, Liu G, Song KS, Tao BX. Analysis on the characteristics and influencing factors of air quality of urban agglomeration in the middle reaches of the Yangtze River in 2016 to 2019. *Ecol Environ Sci*. 2020;29:2034–44. <https://doi.org/10.16258/j.cnki.1674-5906.2020.10.014>
31. Tsui T, Derumigny A, Peck D, van Timmeren A, Wandl A. Spatial clustering of waste reuse in a circular economy: a spatial autocorrelation analysis on locations of waste reuse in the Netherlands using global and local Moran's I. *Front Built Environ*. 2022;8. <https://doi.org/10.3389/fbuil.2022.954642>
32. Anselin L. Local indicators of spatial association—LISA. *Geograph Anal*. 1995;27(2):93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
33. Anselin L, Sridharan S, Gholston S. Using exploratory spatial data analysis to leverage social indicator databases: the discovery of interesting patterns. *Soc Indic Res*. 2006;82(2):287–309. <https://doi.org/10.1007/s11205-006-9034-x>
34. Littenberg TB, Cornish NJ. Prototype global analysis of LISA data with multiple source types. *Phys Rev D*. 2023;107(6). <https://doi.org/10.1103/physrevd.107.063004>
35. Grömping U. Variable importance assessment in regression: linear regression versus random forest. *Am Statist*. 2009;63(4):308–19. <https://doi.org/10.1198/tast.2009.08199>
36. Zhang Z-H, Chen N, Zhu B, Tao H-T, Cheng H-R. Source analysis of ambient PM_{2.5} in Wuhan city based on random forest model. *Huan Jing Ke Xue*. 2022;43(3):1151–8. <https://doi.org/10.13227/j.hjxx.202108051> PMID: 35258179
37. Huang L, Zhu Y, Zhai H, Xue S, Zhu T, Shao Y, et al. Recommendations on benchmarks for numerical air quality model applications in China – Part 1: PM_{2.5} and chemical species. *Atmos Chem Phys*. 2021;21(4):2725–43. <https://doi.org/10.5194/acp-21-2725-2021>
38. Wan C, Xu H, Luo W, Ma J, Li Z. Estimation of regional PM_{2.5} concentration in China based on fine-mode aerosol optical thickness (AOD_f) and study of influencing factors. *Atmospheric Environment*. 2025;344:121026. <https://doi.org/10.1016/j.atmosenv.2025.121026>
39. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International conference on neural information processing systems*, 2017. <https://doi.org/10.48550/arXiv.1705.07874>
40. Tan ZH, Ma S, Han D, Gao D, Yan W. FY-4A cloud base height estimation method based on random forest algorithm. *Infrared Millim Waves*. 2019;38:381–8. <https://doi.org/10.11972/j.issn.1001-9014.2019.03.020>
41. Pakdehi M, Ahmadisharaf E, Azimi P, Yan Z, Keshavarz Z, Caballero C, et al. Modeling the latent impacts of extreme floods on indoor mold spores in residential buildings: application of machine learning algorithms. *Environ Int*. 2025;196:109319. <https://doi.org/10.1016/j.envint.2025.109319> PMID: 39946930
42. Zhang X-S, Nie D-W, Chen Z-Z, Wang R-Z, Su J. Spatiotemporal evolution characteristics and influencing factors of carbon emissions from construction industry in Western China. *Huan Jing Ke Xue*. 2025;46(9):5475–89. <https://doi.org/10.13227/j.hjxx.202408170> PMID: 40962742

43. Wang S, Peng H, Hu Q, Jiang M. Analysis of runoff generation driving factors based on hydrological model and interpretable machine learning method. *J Hydrol Reg Stud*. 2022;42:101139. <https://doi.org/10.1016/j.ejrh.2022.101139>
44. Rabiei-Dastjerdi H, Mohammadi S, Saber M, Amini S, McArdle G. Spatiotemporal analysis of NO₂ production using TROPOMI time-series images and google earth engine in a middle eastern country. *Remote Sensing*. 2022;14(7):1725. <https://doi.org/10.3390/rs14071725>
45. Li MH, Que X, Liu JF, Su SQ, Ding XT. Spatial heterogeneity analysis of PM_{2.5} influencing factors in Chinese cities. *J Chifeng Univ (Natural Science Edition)*. 2021;37:37–41. <https://doi.org/10.3969/j.issn.1673-260X.2021.01.012>
46. Bu XB, Fang ZL, Feng Q, Liao C, Wang HZ. Comprehensive evaluation of air quality based on principal component analysis—take 21 cities in Sichuan province as an example. *Sichuan Environ*. 2023;42:51–6. <https://doi.org/10.14034/j.cnki.schj.2023.03.009>
47. Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Machine Learning*. 2000;40(2):139–57. <https://doi.org/10.1023/a:1007607513941>
48. Yang X-T, Kang P, Wang A-Y, Zang Z-L, Liu L. Prediction of ozone pollution in sichuan basin based on random forest model. *Huan Jing Ke Xue*. 2024;45(5):2507–15. <https://doi.org/10.13227/j.hjxx.202304226> PMID: 38629516
49. Shami TM, El-Saleh AA, Alswaitti M, Al-Tashi Q, Summakieh MA, Mirjalili S. Particle swarm optimization: a comprehensive survey. *IEEE Access*. 2022;10:10031–61. <https://doi.org/10.1109/access.2022.3142859>
50. Gad AG. Particle swarm optimization algorithm and its applications: a systematic review. *Arch Computat Methods Eng*. 2022;29(5):2531–61. <https://doi.org/10.1007/s11831-021-09694-4>
51. Dong J, Cai X, Tian L, Chen F, Xu Q, Li T, et al. Satellite-based estimates of daily NO₂ exposure in urban agglomerations of China and application to spatio-temporal characteristics of hotspots. *Atmosph Environ*. 2023;293:119453. <https://doi.org/10.1016/j.atmosenv.2022.119453>
52. Wu YR, Wang H, Wang MJ. Spatiotemporal variation and impact factors of tropospheric NO₂ column density in Shanxi-Gansu-Ningxia region based on OMI satellite data. *J Atmosph Environ Optics*. 2023;:553–68. <https://doi.org/10.3969/j.issn.1673-6141.2023.06.005>
53. Feng JJ, Shen JF, Liang RZ, Mo CH. Analysis of relationship of meteorological elements and PM10 in Guangzhou. *Environ Monit China*. 2009;25:78–82.
54. Wang H, Chen XQ, Yu YJ, Chen BB, Sui P. Preliminary analysis of distribution characteristics of PM_{2.5}, PM_{2.5}/PM₁₀ and their relationship with meteorological conditions in Fuzhou. *J Trop Meteorol*. 2014;30:387–91.
55. Jiang L, Chen Y, Zhou H, He S. NO_x emissions in China: temporal variations, spatial patterns and reduction potentials. *Atmos Pollut Res*. 2020;11(9):1473–80. <https://doi.org/10.1016/j.apr.2020.06.003>
56. Liu ZK, Yuan WJ. Diagnostic analysis of a cold wave weather process in December 2021. *CCRL*. 2023;12(01):18–30. <https://doi.org/10.12677/ccrl.2023.121003>
57. Shi L-J, Suo N, Ma B-J, Tang T, Wang S, Ma X-Y, et al. Evolution of PM_{2.5} and O₃ pollution in the qinhuangdao city during sustained air quality improvement from 2018 to 2022. *Huan Jing Ke Xue*. 2025;46(4):2115–24. <https://doi.org/10.13227/j.hjxx.202402150> PMID: 40230121
58. Zhan C, Xie M, Fang D, Wang T, Wu Z, Lu H, et al. Synoptic weather patterns and their impacts on regional particle pollution in the city cluster of the Sichuan Basin, China. *Atmos Environ*. 2019;208:34–47. <https://doi.org/10.1016/j.atmosenv.2019.03.033>
59. Zhang X, Ding X, Talifu D, Wang X, Abulizi A, Maihemuti M, et al. Humidity and PM_{2.5} composition determine atmospheric light extinction in the arid region of northwest China. *J Environ Sci (China)*. 2021;100:279–86. <https://doi.org/10.1016/j.jes.2020.07.007> PMID: 33279040
60. Zhou WW, Qian CM, Shu QT, Qiu S, Huang JJ, Yu JG. Estimation of Pinus densata in Northwest Yunnan based on UAV and Sentinel-2 data. *J Southwest Forestry Univ (Nat Sci)*. 2024;44:141–9. <https://doi.org/10.11929/j.swfu.202310009>
61. Wang Y, Teng ZY, Zhang XL, Che YH, Sun GY. Research progress on the effects of atmospheric nitrogen dioxide on plant growth and metabolism. *Ying Yong Sheng Tai Xue Bao*. 2019;30(1):316–24. <https://doi.org/10.13287/j.1001-9332.201901.028> PMID: 30907555
62. Wang L. Analysis of purification effect of landscape vegetation on urban air pollution. *ESM*. 2024;49:162–6. <https://doi.org/10.3969/j.issn.1673-1212.2024.07.035>
63. Warneke C, Schwarz JP, Dibb J, Kalashnikova O, Frost G, Al-Saad J, et al. Fire influence on regional to global environments and air quality (FIREX-AQ). *JGR Atmospheres*. 2023;128(2). <https://doi.org/10.1029/2022jd037758>
64. Huang C, Chen CH, Li L. Anthropogenic air pollutant emission characteristics in the Yangtze River Delta region, China. *Acta Scientiae Circumstantiae*. 2011;31(9):1858–71. <https://doi.org/10.13671/j.hjxxb.2011.09.013>
65. Li M, Liu H, Geng G, Hong C, Liu F, Song Y, et al. Anthropogenic emission inventories in China: a review. *Nat Sci Rev*. 2017;4(6):834–66. <https://doi.org/10.1093/nsr/nwx150>
66. Arioli MS, D'Agosto M de A, Amaral FG, Cybis HBB. The evolution of city-scale GHG emissions inventory methods: a systematic review. *Environ Impact Assess Rev*. 2020;80:106316. <https://doi.org/10.1016/j.eiar.2019.106316>
67. Garg A, Shukla PR, Kapshe M. The sectoral trends of multigas emissions inventory of India. *Atmos Environ*. 2006;40(24):4608–20. <https://doi.org/10.1016/j.atmosenv.2006.03.045>
68. Reff A, Bhawe PV, Simon H, Pace TG, Pouliot GA, Mobley JD, et al. Emissions inventory of PM_{2.5} trace elements across the United States. *Environ Sci Technol*. 2009;43(15):5790–6. <https://doi.org/10.1021/es802930x> PMID: 19731678

69. Zheng H, Wang Z. Analysis of air pollution influencing factors of PM_{2.5} secondary particles by random forest. *IOP Conf Ser: Earth Environ Sci*. 2021;804(4):042065. <https://doi.org/10.1088/1755-1315/804/4/042065>
70. Verhoelst T, Compennolle S, Pinardi G, Lambert J-C, Eskes HJ, Eichmann K-U, et al. Ground-based validation of the Copernicus Sentinel-5P TROPOMI NO₂ measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks. *Atmos Meas Tech*. 2021;14(1):481–510. <https://doi.org/10.5194/amt-14-481-2021>
71. Lange K, Richter A, Schönhardt A, Meier AC, Bösch T, Seyler A, et al. Validation of Sentinel-5P TROPOMI tropospheric NO₂ products by comparison with NO₂ measurements from airborne imaging DOAS, ground-based stationary DOAS, and mobile car DOAS measurements during the S5P-VAL-DE-Ruhr campaign. *Atmos Meas Tech*. 2023;16(5):1357–89. <https://doi.org/10.5194/amt-16-1357-2023>
72. Cofano A, Cigna F, Santamaria Amato L, Siciliani de Cumis M, Tapete D. Exploiting Sentinel-5P TROPOMI and ground sensor data for the detection of volcanic SO₂ plumes and activity in 2018-2021 at Stromboli, Italy. *Sensors (Basel)*. 2021;21(21):6991. <https://doi.org/10.3390/s21216991> PMID: [34770296](https://pubmed.ncbi.nlm.nih.gov/34770296/)