

## Article

# TROPOMI-Referenced Reconstruction and Model Interpretation of Long-Term City-Scale NO<sub>2</sub> Column Density in East Asia Using Machine Learning

Jiaqi Zhang <sup>1,†</sup>, Qing Sun <sup>2,†</sup> , Heming Yang <sup>1,\*</sup>, Yanbiao Xi <sup>3</sup>, Feifei Cheng <sup>4</sup> and Jie Feng <sup>1</sup>

<sup>1</sup> Geography and Tourism, Jilin Normal University, Siping 136000, China; zhangjiaqi@mails.jlnu.edu.cn (J.Z.); 15500181231@163.com (J.F.)

<sup>2</sup> College of Forestry and Grassland Science, Jilin Agricultural University, Changchun 130118, China; sunqing@jlau.edu.cn

<sup>3</sup> Laboratory of Black Soils Conservation and Utilization, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130118, China; xiyanbiao111@163.com

<sup>4</sup> Chaoyang No. 3 Senior High School, Chaoyang 122000, China; yuxinwen@mails.jlnu.edu.cn

\* Correspondence: wosyhm@jlnu.edu.cn

† These authors contributed equally to this work.

## Abstract

Reliable long-term city-scale nitrogen dioxide (NO<sub>2</sub>) records are essential for evaluating urban air quality change, but the short observation period of TROPOMI limits long-term applications. This study developed a TROPOMI-referenced machine learning framework to reconstruct annual tropospheric NO<sub>2</sub> column density for 436 cities in China, Japan, South Korea, North Korea, and Mongolia from 2000 to 2022 using 18 annual predictors comprising five natural environmental variables, six meteorological variables, and seven sectoral anthropogenic NO<sub>x</sub> emission variables. To reduce spatial leakage, the 436 city polygons were assigned to a regular 5° × 5° grid using the largest equal-area polygon intersection fraction. The 61 occupied, non-overlapping blocks were allocated deterministically to five folds, with all 2019–2022 observations from each city retained in its assigned block. Model performance was calculated from the concatenated predictions for the five held-out block sets. Among nine models, random forest (RF) achieved the best independent test performance ( $R^2 = 0.885$ ; RMSE =  $1.968 \times 10^{-5}$  mol m<sup>-2</sup>). During 2005–2022, annual city-level RF–OMI correlations ranged from 0.850 to 0.945, and their normalized regional annual series were strongly correlated ( $r = 0.877$ ). Against ground observations, RF better represented intercity differences in China (mean annual  $r = 0.791$  versus 0.735 for OMI), whereas OMI performed better at the Japanese city scale (0.857 versus 0.810 for RF); nevertheless, RF closely reproduced the Japanese national annual trend ( $r = 0.989$ ). The East Asian mean increased significantly during 2000–2011 (Theil–Sen slope =  $+0.095 \times 10^{-5}$  mol m<sup>-2</sup> yr<sup>-1</sup>), declined significantly during 2011–2018 ( $-0.140 \times 10^{-5}$  mol m<sup>-2</sup> yr<sup>-1</sup>), and remained nonsignificantly negative during 2018–2022 ( $-0.060 \times 10^{-5}$  mol m<sup>-2</sup> yr<sup>-1</sup>). China peaked in 2011, Japan and North Korea showed significant long-term decreases, South Korea showed a significant overall decline, and Mongolia had no significant full-period trend. SHAP analysis showed that industrial combustion emissions, surface pressure, and road emissions had the highest global mean absolute SHAP values, with nonlinear and direction-dependent associations with RF predictions. The resulting dataset supports regional and national long-term NO<sub>2</sub> assessment, while country-specific and city-scale uncertainties should be considered in local applications.



Received: 2 July 2026

Revised: 31 August 2026

Accepted: 7 September 2026

Published: 11 September 2026

**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

**Keywords:** tropospheric NO<sub>2</sub> column density; long-term reconstruction; random forest; OMI validation; trend analysis; East Asia

## 1. Introduction

As a typical toxic atmospheric trace pollutant, nitrogen dioxide (NO<sub>2</sub>) concentration is a core indicator for assessing regional air pollution severity. Investigations into its spatiotemporal distribution characteristics and driving factors have attracted widespread attention in air pollution research [1]. NO<sub>2</sub> is not only a key precursor for the formation of acid rain, acid fog, and photochemical smog, but it can also be converted into nitrates through physical and chemical reactions, accelerating the rise in PM<sub>2.5</sub> concentration, and becoming an important trigger for smog pollution in autumn and winter [2]; the atmospheric transformation process of nitrogen oxides (NO<sub>x</sub>) is also the main source of secondary particles in PM<sub>2.5</sub>, further aggravating the degree of regional air pollution [3]. From the perspective of ecological and health impacts, high concentrations of NO<sub>2</sub> inhaled by the human body can trigger respiratory diseases, cardiovascular diseases, and other health problems, and its destructive effect on hemoglobin can cause hypoxia in the body, even having adverse effects on human brain development. Therefore, in order to accurately grasp the long-term regional impact of NO<sub>2</sub>, and scientifically control NO<sub>2</sub> and related atmospheric pollutants, it is urgent to carry out long-term monitoring and research on atmospheric NO<sub>2</sub> concentration with high spatial resolution [4].

Currently, the mainstream methods for NO<sub>2</sub> pollution research are mainly divided into three categories: ground-based observation data analysis, three-dimensional chemical transport models (CTM) simulation, and remote sensing estimation [5]. Although ground-based observations can obtain high-precision fixed-point data, they are limited by the sparse distribution of stations and the inherent defects of point-source monitoring, making it impossible to fully characterize the spatial distribution characteristics of NO<sub>2</sub> pollution and difficult to achieve comprehensive monitoring of regional NO<sub>2</sub> exposure [6]. CTM can integrate meteorological and emission inventory data to achieve NO<sub>2</sub> concentration simulation at regional and even global scales, as shown by He et al. (2007) who discovered an abrupt increase in the vertical column density of nitrogen dioxide in central and eastern China after 2000 through CTM [7], and UNO et al. (2007) used model simulations of tropospheric nitrogen dioxide in industrialized regions of North America and Europe to have the same magnitude and pattern characteristics as the GOME inversion results [8], but the simulation results of CTM are easily affected by uncertainties in the meteorological field, emission inventory, and parameterization schemes, and there are certain discrepancies with ground-based observation data [9]. Satellite remote sensing technology, with its advantages of large-scale and long-time series observations, has become a key means to compensate for the shortcomings of ground-based observations and model simulations [10]. Among them, the Sentinel-5p satellite equipped with tropospheric monitoring instruments (TROPOMI) has been in use since 2018 [11], providing stable global NO<sub>2</sub> column concentration data with the highest spatial resolution of 5.5 km × 3.5 km, significantly superior to that of previous sensors such as GOME-2 and OMI in terms of observation accuracy and timeliness, providing a high-quality data source for NO<sub>2</sub> pollution monitoring [12]. However, the time limitation of TROPOMI satellite observations starting in 2018 leads to the absence of long-time series data from 2000 to 2017, making it impossible to support a comprehensive analysis of the long-term evolution law of NO<sub>2</sub> in East Asia; at the same time, urban NO<sub>2</sub> mainly originates from mobile source emissions and industrial combustion, with a large gradient of diurnal concentration changes and significant spatiotemporal heterogeneity. The

fixed overflight time of satellites and limited spatial resolution are prone to homogenize these fine features, making it difficult to meet the needs of precise monitoring at the urban scale [13].

Rapid advances in machine learning techniques, covering both conventional machine learning and deep learning frameworks, offer a feasible approach to address the aforementioned research issues. These methods can accurately capture the nonlinear change characteristics of pollutants by mining the complex statistical relationships between multi-source data, while improving the spatiotemporal resolution and ensuring the accuracy of estimation, yielding significantly superior results to traditional statistical and physical models [14]. Existing research has confirmed that machine learning models perform outstandingly in NO<sub>2</sub> concentration estimation: Kuhn et al. (2024) [15] developed a deep learning model named NitroNet—a subclass of machine learning algorithms—which takes the simulated profiles from WRF-Chem as training samples and TROPOMI column concentrations along with auxiliary variables such as meteorological and emission data as inputs. By adopting strategies of data screening and bias correction, this model enables the accurate prediction of tropospheric NO<sub>2</sub> vertical profiles across multiple global regions. In summer, the correlation coefficients of the model predictions with satellite, ground-based and MAX-DOAS observations reach up to 0.95, 0.75 and 0.99, respectively, demonstrating a higher accuracy than that of traditional chemical transport models [15]. Li et al. (2025) [16] constructed a long-term fused dataset of OMI and TROPOMI satellite data based on the spatial zoning and cumulative distribution function (CDF) fusion strategy. They further conducted comparative modeling by integrating ensemble models including RF, XGBoost and Light Gradient Boosting Machine (LGBM), and optimized the LGBM algorithm to achieve high-precision daily estimation of near-surface NO<sub>2</sub> at a 1 km resolution over mainland China during 2014–2020. The results of 10-fold cross-validation show that the R<sup>2</sup> reaches 0.85 with RMSE of 7.51 µg/m<sup>3</sup>, which can clearly depict the spatial pattern of “higher concentrations in the east and lower in the west” and the seasonal evolution characteristics of the pollutant [16]. These studies show that machine learning models can fully integrate multi-source data such as satellite remote sensing, meteorology, and socioeconomic data, and have an irreplaceable advantage in solving spatiotemporal discontinuity and improving estimation accuracy, especially providing a feasible path for filling the time series gap of TROPOMI data and reconstructing the long-term NO<sub>2</sub> concentration distribution [17–19].

Based on these considerations, this study focuses on 436 cities in five East Asian countries, including China, Japan, South Korea, North Korea, and Mongolia, and aims to construct a long-term annual NO<sub>2</sub> column density dataset at the urban scale. TROPOMI NO<sub>2</sub> observations were used as the target variable for model development, while MODIS-derived natural environmental variables, ERA5-Land meteorological factors and annual EDGAR NO<sub>x</sub> emission inventory data were integrated as multi-source auxiliary predictors. In particular, the use of annual EDGAR NO<sub>x</sub> emissions allows the interannual variation in anthropogenic emissions to be explicitly represented, thereby improving the ability to characterize long-term changes in human emission intensity.

Nine machine learning and statistical learning models, including BP, CNN, ELM, PLS, LSTM, RBF, SVM, XGBoost, and RF, were compared using five-fold spatial block cross-validation to identify the model that most reliably generalized to held-out geographic blocks within the sampled East Asian domain. Considering the temporal availability of stable TROPOMI observations and the incompleteness of the 2018 observation record, complete annual TROPOMI observations from 2019 to 2022 were used as the target data for model development and evaluation. The selected RF model was subsequently applied uniformly to the corresponding annual auxiliary predictors for each year from 2000 to

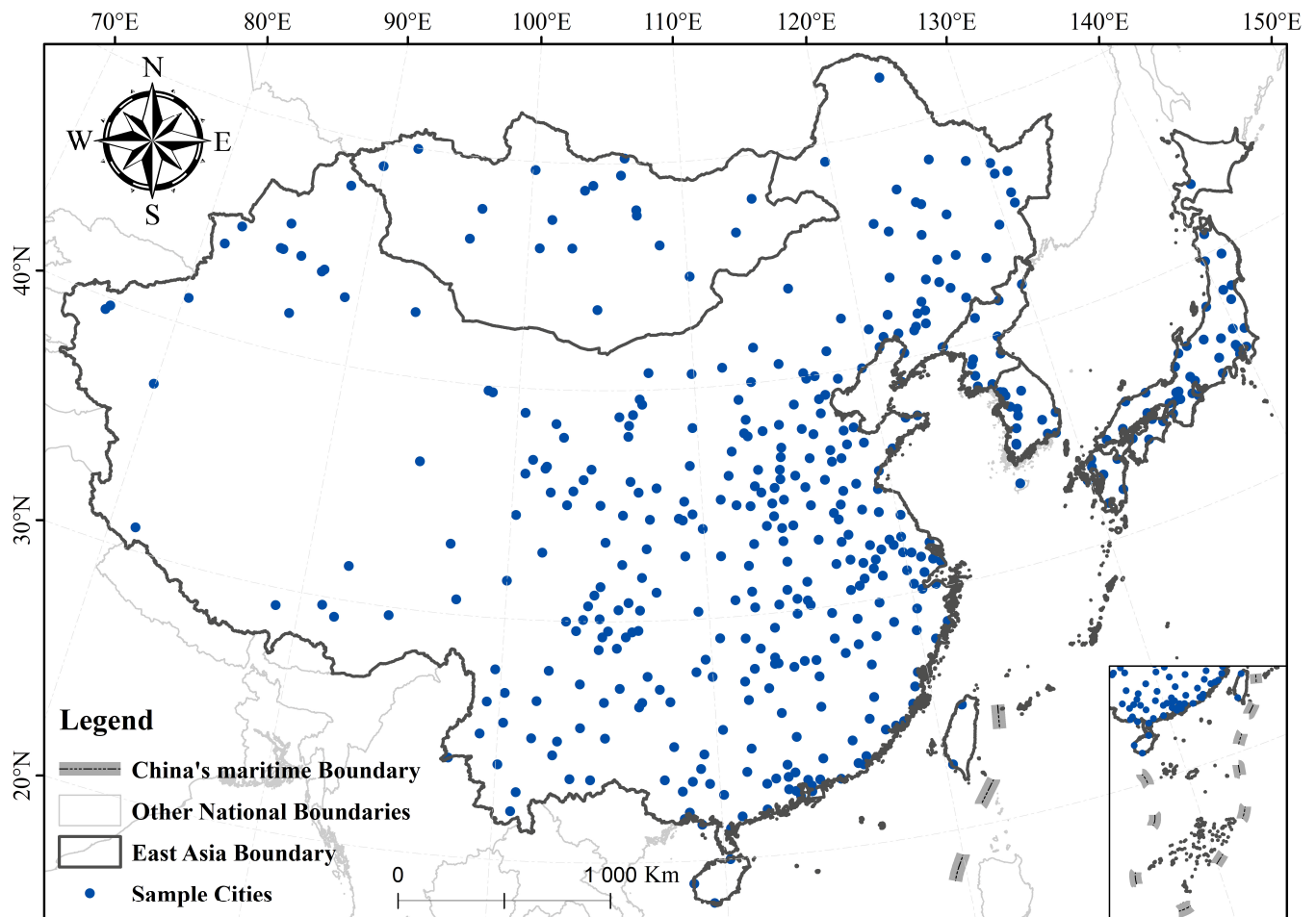
2022, thereby generating a continuous and methodologically consistent RF-retrieved NO<sub>2</sub> column density dataset. This framework aims to extend the advantages of TROPOMI's high-spatial-resolution observations to the historical period and support the construction of a continuous annual NO<sub>2</sub> dataset for East Asian cities.

To evaluate the reliability and applicability of the constructed dataset, the OMI NO<sub>2</sub> product and ground-based NO<sub>2</sub> observations from China and Japan are used as independent validation references. The OMI product provides a long-term satellite-based comparison, while ground observations from more than 340 Chinese cities and long-term Japanese monitoring stations provide independent evidence for assessing the consistency of the annual NO<sub>2</sub> dataset at the city scale. On this basis, this study further investigates the spatiotemporal evolution of urban NO<sub>2</sub> column density in East Asia to identify the predictors most strongly associated with variation in the RF-reconstructed NO<sub>2</sub> column density. The results are expected to provide a long-term city-scale NO<sub>2</sub> dataset and methodological support for regional air pollution assessment, long-term atmospheric environmental monitoring, and cross-border pollution control in East Asia. From a sustainability perspective, the constructed long-term city-scale NO<sub>2</sub> dataset can provide scientific evidence for evaluating urban air pollution trajectories, identifying sectors whose emission predictors are strongly associated with the reconstructed NO<sub>2</sub> patterns, and supporting sustainable transportation, industrial emission mitigation, and regional environmental governance in East Asia.

## 2. Materials and Methods

### 2.1. Study Area

East Asia is located between 4° N and 53° N, 73° E and 150° E, covering China, Japan, South Korea, North Korea, and Mongolia, with a total area of about 12.5 million square kilometers. Its unique natural geographical climate and distinct population, economy, and emission source characteristics jointly shape the distribution, transport, and transformation characteristics of nitrogen dioxide (NO<sub>2</sub>) in the region [3]. The western part of the region is highland and mountainous, while the eastern and southern parts are plains and hills, with many islands, reefs, and peninsulas along the coast, providing a key underlying surface for the accumulation of NO<sub>2</sub>. The climate is diverse, with different meteorological conditions such as monsoon areas and continental climate zones, which directly regulate the dilution, diffusion, and chemical transformation of NO<sub>2</sub>. The total population of the region exceeds 1.6 billion, with China, Japan, and South Korea as core economies. The level of industrialization and urbanization is high, and energy industry activities are intensive. North Korea and Mongolia have lagging economic development. There are significant differences in population distribution and energy industry structures among countries, and each country has its own emission sources with different emphases—the main sources in China are industry and mobile sources, Japan and South Korea focus on transportation and intensive industrial sources, North Korea mainly relies on coal-fired industry and agricultural sources, and Mongolia mainly relies on transportation and civil coal combustion. The differentiated emission structures further cause differences in the basic pattern of NO<sub>2</sub> concentration in the region [20]. The locations of the 436 urban areas included in this study are illustrated in Figure 1.



**Figure 1.** Locations of the 436 retained urban areas in five East Asian countries.

## 2.2. Data and Processing

### 2.2.1. Satellite NO<sub>2</sub> Products

TROPOMI NO<sub>2</sub> column density was used as the target variable for machine learning model development and validation. It was derived from the Sentinel-5P/TROPOMI offline NO<sub>2</sub> product and represents the tropospheric NO<sub>2</sub> vertical column density. Because the TROPOMI observations in 2018 did not cover a complete year, complete annual TROPOMI NO<sub>2</sub> observations from 2019 to 2022 were used as the target variable to establish the statistical relationship between tropospheric NO<sub>2</sub> column density and multi-source auxiliary predictors and to support model training, parameter optimization, and performance evaluation. The selected RF model was subsequently applied uniformly to the corresponding long-term auxiliary predictors for each year from 2000 to 2022, producing city-scale tropospheric NO<sub>2</sub> column density retrievals for the entire study period. These retrievals were used to construct a continuous and methodologically consistent TROPOMI-referenced RF-derived NO<sub>2</sub> column density dataset [21].

OMI NO<sub>2</sub> column density was used as an independent satellite reference dataset for validation and comparison. It was derived from the OMI/Aura Level-3 daily gridded NO<sub>2</sub> product and represents the cloud-screened tropospheric NO<sub>2</sub> column amount. Considering the long temporal coverage of OMI observations, the OMI NO<sub>2</sub> product was used to compare with the annual NO<sub>2</sub> dataset developed in this study and ground-based NO<sub>2</sub> observations. The daily OMI observations were aggregated to the annual scale to maintain temporal consistency with the annual NO<sub>2</sub> dataset and ground station records.

OMI and TROPOMI differ mainly in their temporal coverage and spatial resolution. OMI provides a long-term NO<sub>2</sub> record dating back to 2004, but its relatively coarse spatial resolution may smooth urban-scale spatial heterogeneity. In contrast, TROPOMI has provided observations since 2018 with a finer spatial resolution, allowing a more detailed characterization of intra-urban NO<sub>2</sub> variations and emission hotspots. Therefore, OMI was mainly used for long-term comparison, whereas TROPOMI was used as the target variable for model development during 2019–2022.

### 2.2.2. Natural Environmental and Meteorological Data

Natural environmental factors were mainly derived from MODIS remote sensing products, including aerosol optical depth, leaf area index, land surface temperature, normalized difference vegetation index, and snow cover extent. These variables were used to characterize atmospheric aerosol loading, vegetation growth conditions, surface thermal properties, ecological cover characteristics, and snow-related surface attributes, respectively. All MODIS variables were uniformly preprocessed on the Google Earth Engine platform, composited at the annual scale, and spatially averaged at the city scale to obtain annual natural environmental characteristics for each city.

Meteorological factors were obtained from ERA5-Land reanalysis data, including air temperature, precipitation, wind speed, surface pressure, dew point temperature, and relative humidity. These variables were used to describe the regulatory effects of meteorological conditions on NO<sub>2</sub> concentrations, including atmospheric diffusion, wet deposition removal, boundary layer stability, humidity-related chemical transformation, and regional transport processes. Monthly meteorological data were further aggregated to the annual scale and extracted at the city scale to ensure consistency with other annual auxiliary variables in both temporal and spatial dimensions.

### 2.2.3. Anthropogenic Emission Inventory Data

Anthropogenic emission factors were derived from the EDGAR v8.1 global air pollutant emission inventory. Since emission inventories generally provide NO<sub>x</sub> emissions rather than NO<sub>2</sub> emissions alone, and NO<sub>x</sub>, including NO and NO<sub>2</sub>, is the direct precursor pool controlling the formation of and variation in tropospheric NO<sub>2</sub> in urban environments, NO<sub>x</sub> emissions were used as emission-related explanatory variables in the NO<sub>2</sub> inversion model. Annual gridded NO<sub>x</sub> emission data were processed by year, clipped to the study area covering the five East Asian countries, and further aggregated to the city scale to characterize anthropogenic emission intensity in different cities.

To more reasonably represent the major anthropogenic sources of urban NO<sub>2</sub> and reduce variable redundancy and multicollinearity caused by overly detailed sectoral classifications, the original sector-specific EDGAR NO<sub>x</sub> emissions were aggregated into seven emission factors: road transportation emissions (NO<sub>x\_road</sub>), power generation emissions (NO<sub>x\_power</sub>), manufacturing combustion emissions (NO<sub>x\_ind\_combustion</sub>), residential and commercial building energy emissions (NO<sub>x\_residential</sub>), shipping and non-road transportation emissions (NO<sub>x\_shipping\_nonroad</sub>), industrial process- and fuel-related emissions (NO<sub>x\_ind\_process</sub>), and agricultural emissions (NO<sub>x\_agriculture</sub>). Specifically, NO<sub>x\_shipping\_nonroad</sub> was calculated as the pixel-wise sum of TNR\_Ship and TNR\_Other; NO<sub>x\_ind\_process</sub> was calculated as the pixel-wise sum of REF\_TRF, PRO\_FFF, NMM, CHE, IRO, NFE, and FOO\_PAP; and NO<sub>x\_agriculture</sub> was calculated as the sum of AGS and AWB.

#### 2.2.4. Ground-Based Validation Data

Ground-based NO<sub>2</sub> observations from China and Japan were used as independent validation data. The Chinese ground observations were obtained from the National Urban Air Quality Real-time Publishing Platform operated by the China National Environmental Monitoring Centre. These data were aggregated to the annual city scale and used to evaluate the consistency between the annual NO<sub>2</sub> dataset developed in this study and ground-level NO<sub>2</sub> variations in Chinese cities.

Japanese ground-based NO<sub>2</sub> observations were obtained from the Continuous Air Pollution Monitoring Data provided by the Environmental Information Platform of the National Institute for Environmental Studies. These data provide long-term monitoring records of air pollution conditions in Japan and were aggregated to the annual city scale to evaluate the regional applicability and uncertainty of the annual NO<sub>2</sub> dataset in Japan.

Considering that satellite NO<sub>2</sub> column density and ground-level NO<sub>2</sub> concentration represent different atmospheric quantities and have different units, the ground-based validation focused on correlation-based consistency rather than direct absolute error comparison. Pearson and Spearman correlation coefficients were used to compare the agreement of the annual NO<sub>2</sub> dataset and the OMI product with ground-based observations at the city-year scale.

#### 2.3. Data Preprocessing and Temporal Applicability Domain Assessment

All multi-source variables were harmonized at the city-year level using the fixed city polygons in the city boundary dataset derived from the Global Artificial Impervious Area (GAIA) dataset [22]. During boundary data quality control, two erroneous duplicate city records (codes 35768 and 123582), originating from the source boundary dataset and spatially duplicating the city polygon retained under code 123623, were removed, reducing the initial 438 records to 436 unique cities used in the subsequent analyses. City centroids were calculated solely to provide representative point locations for cartographic visualization of the discrete city-level spatial distribution; they were not used for spatial data extraction or for the spatial block cross-validation procedure.

No source product was explicitly resampled or exported to a common raster grid before city-scale aggregation. For each data source, annual summaries were generated using an operation consistent with the variable definition, with annual means used for concentration-like or state variables and annual sums used for precipitation. City-level values were subsequently obtained by averaging valid pixels within each city polygon. In Google Earth Engine, polygon reductions were performed using `ee.Reducer.mean()` at a nominal analysis scale of 1000 m, with `tileScale = 16`, `maxPixels = 1 × 1013`, and `bestEffort = false`. The specified scale defined the working scale of the zonal reduction rather than representing an explicit preprocessing step that resampled and exported all source products to a unified raster grid. When no valid source observations were available for a given city-year, the corresponding value remained missing; no unmasking, zero filling, or temporal interpolation was applied. Records with missing TROPOMI target values were excluded from model fitting, whereas missing observations in validation analyses were handled by pairwise deletion. Detailed product identifiers, bands, unit transformations, quality control rules, and annual aggregation procedures are provided in Table 1.

**Table 1.** Data sources, variable definitions, and actual harmonization procedures.

| Variable(s)             | Product, Band, or Source                                                                                     | Native Scale                                     | Conversion/Output Unit                                                                   | Annual and City-Scale Processing                                  | Quality Control, Coverage, and Missing Values                                    |
|-------------------------|--------------------------------------------------------------------------------------------------------------|--------------------------------------------------|------------------------------------------------------------------------------------------|-------------------------------------------------------------------|----------------------------------------------------------------------------------|
| TROPOMI NO <sub>2</sub> | COPERNICUS/S5P/OFFL/L <sup>3</sup> _NO <sub>2</sub> ;<br>tropospheric_NO <sub>2</sub> _column_number_density | Daily L <sup>3</sup> ; 0.01<br>degrees (~1.1 km) | mol·m <sup>-2</sup>                                                                      | Annual pixel mean;<br>city-polygon mean<br>(scale = 1000 m)       | GEE source QA ≥ 0.75; no<br>extra script mask;<br>unavailable years blank        |
| OMI NO <sub>2</sub>     | OMI/AuraOMNO <sub>2</sub> d/v003;<br>ColumnAmountNO <sub>2</sub> TropCloudScreened                           | Daily L <sup>3</sup> ;<br>0.25 degrees           | Native:<br>molecules·cm <sup>-2</sup> →<br>unified mol·m <sup>-2</sup><br>(see footnote) | Valid daily pixels →<br>annual mean → mean<br>within city polygon | Cloud-screened field;<br>masked/no-data<br>observations omitted<br>pairwise      |
| AOD                     | MODIS/061/MCD19A2_GRANULES;<br>Optical_Depth_047                                                             | Daily; 1 km                                      | ×0.001; dimensionless                                                                    | Annual mean → mean<br>within city polygon                         | No extra QA-bit filter;<br>masked pixels omitted;<br>empty years blank           |
| LAI                     | MODIS/061/MCD15A3H; Lai                                                                                      | 4-day; 500 m                                     | ×0.1; m <sup>2</sup> m <sup>-2</sup>                                                     | Annual mean → mean<br>within city polygon                         | No extra QA-bit filter;<br>generally available from<br>2002; earlier years blank |
| LST                     | MODIS/061/MOD11A2; LST_Day_1km                                                                               | 8-day; 1 km                                      | ×0.02–273.15;<br>degrees C                                                               | Annual mean → mean<br>within city polygon                         | No extra QA-bit filter;<br>masked pixels omitted;<br>empty years blank           |
| NDVI                    | MODIS/061/MOD13Q1; NDVI                                                                                      | 16-day; 250 m                                    | ×0.0001; dimensionless                                                                   | Annual mean → mean<br>within city polygon                         | No extra QA-bit filter;<br>masked pixels omitted;<br>empty years blank           |
| SCE                     | MODIS/061/MOD10A1; NDSI_Snow_Cover                                                                           | Daily; 500 m                                     | None; 0–100 (%)                                                                          | Annual mean → mean<br>within city polygon                         | No extra QA-bit filter;<br>masked pixels omitted;<br>empty years blank           |
| T; DPT                  | ECMWF/ERA5_LAND/MONTHLY_AGGR;<br>temperature_2m; dewpoint_temperature_2m                                     | Monthly;<br>~0.1 degrees                         | K – 273.15; degrees C                                                                    | Monthly pixel mean →<br>annual mean; city<br>mean at 1000 m       | Reanalysis coverage<br>2000–2022; non-finite<br>values recorded as missing       |
| Ppt                     | ECMWF/ERA5_LAND/MONTHLY_AGGR;<br>total_precipitation_sum                                                     | Monthly;<br>~0.1 degrees                         | ×1000; mm yr <sup>-1</sup>                                                               | Monthly sum → city<br>mean at 1000 m                              | Reanalysis coverage<br>2000–2022; non-finite<br>values recorded as missing       |

Table 1. Cont.

| Variable(s)                               | Product, Band, or Source                                                                                                                   | Native Scale                      | Conversion/Output Unit                                           | Annual and City-Scale Processing                                        | Quality Control, Coverage, and Missing Values                                                             |
|-------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|------------------------------------------------------------------|-------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| WS                                        | ECMWF/ERA5_LAND/MONTHLY_AGGR;<br>u/v_component_of_wind_10m                                                                                 | Monthly;<br>~0.1 degrees          | $\sqrt{u^2 + v^2}$ ; m s <sup>-1</sup>                           | Monthly $\sqrt{u^2 + v^2}$<br>-> annual mean; city<br>mean at 1000 m    | Reanalysis coverage<br>2000–2022; non-finite<br>values recorded as missing                                |
| P                                         | ECMWF/ERA5_LAND/MONTHLY_AGGR;<br>surface_pressure                                                                                          | Monthly;<br>~0.1 degrees          | divide by 100; hPa                                               | Monthly pixel mean -><br>annual mean; city<br>mean at 1000 m            | Reanalysis coverage<br>2000–2022; non-finite<br>values recorded as missing                                |
| RH                                        | Derived from ERA5-Land annual T and DPT                                                                                                    | Annual city level                 | Magnus equation; %                                               | Calculated after annual<br>T and DPT city means                         | Missing when T or DPT<br>was unavailable                                                                  |
| Seven EDGAR<br>NO <sub>x</sub> predictors | EDGAR v8.1 *_emi_nc.zip; TRO, ENE, IND, RCO,<br>TNR_Ship + TNR_Other, REF_TRF + PRO_FFF +<br>NMM + CHE + IRO + NFE + FOO_PAP,<br>AGS + AWB | Annual; 0.1 degrees;<br>1970–2022 | tonnes substance per<br>0.1-degree<br>grid cell yr <sup>-1</sup> | Component sectors<br>summed pixelwise -><br>annual city-polygon<br>mean | Complete annual<br>inventory; NO <sub>x</sub> _total<br>excluded from main<br>model to limit collinearity |
| China ground<br>NO <sub>2</sub>           | China National Environmental Monitoring Centre                                                                                             | Monitoring stations               | micrograms m <sup>-3</sup>                                       | Valid observations -><br>station annual mean -><br>city mean            | Provider-valid records<br>only; missing<br>station-years excluded<br>pairwise                             |
| Japan ground<br>NO <sub>2</sub>           | NIES Continuous Air Pollution Monitoring Data                                                                                              | Monitoring stations               | ppb or ppm as<br>reported                                        | Valid observations -><br>station annual mean -><br>city mean            | Provider-valid records<br>only; missing<br>station-years excluded<br>pairwise                             |

Note: OMI NO<sub>2</sub> native unit is molecules·cm<sup>-2</sup>. To convert to mol·m<sup>-2</sup> for consistency with TROPOMI, both the molecular-to-molar conversion (divide by Avogadro's constant,  $N_A = 6.022 \times 10^{23}$  NA = 6.022 × 10<sup>23</sup>) and the area unit conversion (cm<sup>-2</sup> → m<sup>-2</sup>, multiply by 104,104) are applied. The combined scaling factor is 10<sup>4</sup>/6.022 × 10<sup>23</sup>.

The EDGAR v8.1 Air Pollutants inventory was processed as an annual, sector-resolved input rather than as a static emission proxy. Its annual 0.1-degree NO<sub>x</sub> fields were aggregated pixel by pixel into road transport, power generation, industrial combustion, residential-commercial combustion, shipping and non-road transport, industrial processes, and agriculture before city-polygon means were calculated. Thus, the value assigned to each city-year was derived from the corresponding annual and sector-specific EDGAR layer, allowing the predictors to represent long-term changes in the magnitude and sectoral composition of anthropogenic NO<sub>x</sub> emissions. Because the EDGAR source files are expressed as tonnes substance grid-cell-1 yr<sup>−1</sup>, the resulting city variable is the mean of intersecting grid cell values; it is not interpreted as a city total emission or converted to an emission density.

To address the temporal extrapolation concern associated with reconstructing 2000–2018 from a model trained on 2019–2022, we conducted a predictor domain assessment before historical reconstruction. The assessment was restricted to the same 18 predictors used in the final model. All non-predictor metadata and the NO<sub>2</sub> target variable were excluded from the predictor domain assessment. For each predictor, we compared effective sample size, missingness, descriptive statistics, and the standardized mean difference (SMD) between the historical and training reference periods. We also quantified the proportion of non-missing historical observations falling within the training period minimum–maximum range and within the more conservative training period 1st–99th percentile range. Missing values were retained and reported separately rather than being classified as out-of-range observations. At the city-year level, we additionally recorded the number and proportion of valid predictors outside each training period range; annual and country-specific summaries were generated in parallel.

To examine temporal stability within the study period rather than relying only on pooled historical-versus-training distributions, we additionally calculated year-specific means, medians, and standardized mean differences for all 18 predictors from 2000 to 2022, using the pooled 2019–2022 data as the training reference. Trends in the annual mean and median series were assessed using the Mann–Kendall test and Theil–Sen slope, and potential abrupt changes were evaluated using the Pettitt test. Benjamini–Hochberg false discovery rate correction was applied across all 18 predictors within each combination of annual statistic, sensitivity scenario, and test family. All analyses were repeated after excluding 2020, 2021, and both years. Because TROPOMI observations were unavailable before 2019, longer-term relationship stability was assessed indirectly using OMI NO<sub>2</sub> for 2005–2022. For each predictor and year, we calculated the cross-city Spearman correlation with OMI NO<sub>2</sub> and applied the same temporal tests to the Fisher-z-transformed annual correlations. Predictor–TROPOMI correlations for 2019–2022 were reported descriptively. The main text focuses on the five predictors with the highest global mean absolute SHAP values, while complete results for all 18 predictors are provided in the Supplementary Materials.

$$\text{SMD} = \frac{\text{mean}_{\text{historical}} - \text{mean}_{\text{training}}}{\sqrt{\frac{\text{variance}_{\text{historical}} + \text{variance}_{\text{training}}}{2}}} \quad (1)$$

For concise presentation in the main text, the five predictors were prespecified on the basis of the verified global mean absolute SHAP ranking, rather than selected according to their overlap results. As shown in Table 2, their absolute SMDs were all below 0.10 (0.001–0.091), indicating very small mean differences between the historical and training reference periods. The historical coverage within the training period minimum–maximum range was 99.6–99.9%, while coverage within the stricter 1st–99th percentile range was 98.2–99.0%.

**Table 2.** Temporal predictor domain assessment for the five SHAP-prioritized predictors.

| SHAP Rank | Predictor                       | SMD   | Historical Observations<br>Within the Training<br>Min–Max Range (%) | Historical Observations<br>Within the Training 1st–99th<br>Percentile Range (%) |
|-----------|---------------------------------|-------|---------------------------------------------------------------------|---------------------------------------------------------------------------------|
| 1         | NO <sub>x</sub> _ind_combustion | 0.008 | 99.7                                                                | 99.0                                                                            |
| 2         | P                               | 0.001 | 99.9                                                                | 98.5                                                                            |
| 3         | NO <sub>x</sub> _road           | 0.091 | 99.6                                                                | 98.2                                                                            |
| 4         | NO <sub>x</sub> _ind_process    | 0.002 | 99.9                                                                | 98.2                                                                            |
| 5         | NO <sub>x</sub> _agriculture    | 0.005 | 99.8                                                                | 99.0                                                                            |

Note: The historical reconstruction period covers 2000–2018 ( $n = 8284$  city–year records), and the training reference period covers 2019–2022 ( $n = 1744$ ).  $SMD = (\text{mean\_historical} - \text{mean\_training}) / \sqrt{[(\text{variance\_historical} + \text{variance\_training})/2]}$ . Coverage rates were calculated only among non-missing historical observations; missing values were not classified as out of range. The five variables were prespecified using the verified global mean absolute SHAP ranking. Complete 18-predictor results are provided in the Supplementary Materials.

Year-specific diagnostics provided a more stringent assessment of temporal stability. Across the five SHAP-prioritized predictors, the maximum annual |SMD| relative to the pooled 2019–2022 reference ranged from 0.009 to 0.288, and the minimum annual coverage within the training period 1st–99th percentile range remained at least 96.8%. Pettitt tests identified FDR-significant level shifts in NO<sub>x</sub>\_ind\_combustion (2005,  $q = 0.046$ ), NO<sub>x</sub>\_ind\_process (2006,  $q = 0.024$ ), and NO<sub>x</sub>\_agriculture (2005,  $q = 0.044$ ), indicating that predictor levels evolved over time rather than remaining static. Nevertheless, annual cross-city associations with OMI NO<sub>2</sub> remained positive throughout 2005–2022 (Spearman’s  $\rho = 0.438$ – $0.711$  across the five predictors), and no FDR-significant abrupt change point was detected in any of the five association series (all Pettitt  $q \geq 0.142$ ). Four predictors showed no monotonic change in association strength, whereas NO<sub>x</sub>\_ind\_process exhibited a small gradual strengthening (Fisher- $z$  Theil–Sen slope =  $0.004 \text{ yr}^{-1}$ ; Mann–Kendall  $q = 0.005$ ). Excluding 2020, 2021, or both years did not materially alter these conclusions. Complete 18-predictor results and the year-specific diagnostics are provided in Supplementary Materials.

Across all 18 final predictors, 12 had |SMD| < 0.10 and 15 had |SMD| < 0.20. Sixteen predictors had at least 99% of non-missing historical observations within the training period minimum–maximum range, and 17 predictors had at least 95% within the training period 1st–99th percentile range. The complete 18-predictor distributions, annual and country-specific summaries, city–year extrapolation flags, and strict core-range sensitivity analyses are provided in Supplementary Tables and Figures.

These checks show that the dynamic emission, meteorological, and land surface predictors retained substantial overlap between the historical and training periods, including for the most influential emission-related predictors. The consistently positive OMI associations and the absence of FDR-significant abrupt relationship change points among the five SHAP-prioritized predictors provide additional evidence that their broad associations with NO<sub>2</sub> were maintained during 2005–2022. However, the detected changes in predictor levels and the gradual strengthening of the NO<sub>x</sub>\_ind\_process association show that neither the predictors nor their bivariate relationships were completely static.

#### 2.4. Machine Learning Models

In this study, urban-scale NO<sub>2</sub> column density across five East Asian countries was used as the target variable. The final predictor set comprised 18 annual variables: five natural environmental predictors (AOD, LAI, LST, NDVI, and SCE), six meteorological predictors (T, Ppt, WS, P, DPT, and RH), and seven sectoral anthropogenic NO<sub>x</sub> emission predictors (NO<sub>x</sub>\_road, NO<sub>x</sub>\_power, NO<sub>x</sub>\_ind\_combustion, NO<sub>x</sub>\_residential,

NO<sub>x</sub>\_shipping\_nonroad, NO<sub>x</sub>\_ind\_process, and NO<sub>x</sub>\_agriculture). Based on these predictors, nine machine learning and statistical learning models, including PLS, BP, CNN, ELM, LSTM, RBF, SVM, XGBoost, and RF, were constructed to perform remote sensing inversion of NO<sub>2</sub> column concentration. The model construction, parameter optimization, and performance evaluation followed standard machine learning procedures [23]. The key parameters of these nine models are summarized in Table 3.

**Table 3.** Core information of nine machine learning models.

| Name | Core Principles and Features                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Key Parameters                                                                                                                                                                                                                                                                                                                                                                                   |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| BP   | Conventional neural network model. Implementation of parameter iterative optimization for multi-layer fully connected neural networks based on gradient descent method, calculating output loss through forward propagation, completing gradient backpropagation and weight update through backpropagation, as the basic training framework for multi-layer neural networks.                                                                                                                                                                                            | <ol style="list-style-type: none"> <li>1. Training epochs = 1000</li> <li>2. Error goal threshold = <math>1 \times 10^{-6}</math></li> <li>3. Learning rate (lr) = 0.01</li> </ol>                                                                                                                                                                                                               |
| CNN  | Deep learning model. Feedforward neural network, through the hierarchical structure of convolutional layers, pooling layers, and fully connected layers, relies on parameter sharing and local receptive fields to automatically extract features of grid-like data and compress redundant information.                                                                                                                                                                                                                                                                 | <ol style="list-style-type: none"> <li>1. Optimizer: SGDM</li> <li>2. MiniBatchSize = 100</li> <li>3. MaxEpochs = 1200</li> <li>4. InitialLearnRate = 0.01</li> <li>5. LearnRateSchedule = piecewise</li> <li>6. LearnRateDropFactor = 0.1</li> <li>7. LearnRateDropPeriod = 800</li> <li>8. Shuffle = every-epoch</li> <li>9. Plots = training-progress</li> <li>10. Verbose = false</li> </ol> |
| ELM  | Conventional neural network model. Single hidden layer feedforward neural network, randomly initialized input layer—hidden layer weights and biases, no iterative training of hidden layer weights, directly solve hidden layer—output layer optimal weights through Moore–Penrose generalized inverse.                                                                                                                                                                                                                                                                 | <ol style="list-style-type: none"> <li>1. Number of hidden layer nodes (num_hiddens) = 50</li> <li>2. Hidden activation function (activate_model) = 'sig' (Sigmoid)</li> </ol>                                                                                                                                                                                                                   |
| PLS  | Statistical learning model. Multivariate regression method based on latent variable extraction. The original predictor matrix and response variable are projected into a low-dimensional component space, and the latent variables are extracted by maximizing the covariance between independent variables and the dependent variable. This method can reduce the influence of multicollinearity among predictors and is suitable for modeling NO <sub>2</sub> column concentrations driven by multiple correlated natural, meteorological, and anthropogenic factors. | <ol style="list-style-type: none"> <li>1. Number of latent variables (k) = 5</li> <li>2. Scaling method: standardized normalization</li> <li>3. Cross-validation folds: 10-fold</li> <li>4. Convergence threshold: <math>1 \times 10^{-6}</math></li> <li>5. Maximum iteration count: 1000</li> </ol>                                                                                            |

Table 3. Cont.

| Name    | Core Principles and Features                                                                                                                                                                                                                                                                                                     | Key Parameters                                                                                                                                                                                                                                                                                 |
|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| RBF     | Conventional neural network model. Feedforward neural network with radial basis function as the activation function, maps the input space nonlinearly to a high-dimensional feature space through hidden layers, and uses linear fitting in the output layer to approximate complex functions.                                   | Radial basis spread coefficient (rbf_spread) = 100                                                                                                                                                                                                                                             |
| SVM     | Kernel-based conventional machine learning model implemented for regression. It maps the predictors into a high-dimensional feature space through a kernel function and estimates a continuous response using an epsilon-insensitive loss while controlling model complexity.                                                    | <ol style="list-style-type: none"> <li>1. Kernel type: RBF (-t 2)</li> <li>2. Penalty factor (c) = 4.0</li> <li>3. RBF kernel gamma parameter (g) = 0.8</li> <li>4. Task mode: epsilon-SVR (-s 3)</li> <li>5. Epsilon loss threshold (-p) = 0.01</li> </ol>                                    |
| XGBoost | Ensemble machine learning model. Optimized distributed gradient boosting tree model, based on improved gradient boosting decision tree (GBDT), introduces second-order derivatives and regularization terms, adopts weighted histogram and cache optimization to enhance training efficiency, and supports parallel computation. | <ol style="list-style-type: none"> <li>1. Number of base decision trees (num_trees) = 500</li> <li>2. Objective loss function = reg:linear (linear regression task)</li> <li>3. Maximum tree depth (max_depth) = 1</li> </ol>                                                                  |
| RF      | Ensemble machine learning model. Integrative learning algorithm, based on Bootstrap resampling to construct multiple independent decision trees, outputting results by voting method (classification)/mean method (regression), and reducing the overfitting risk of a single tree by randomly selecting features and samples.   | <ol style="list-style-type: none"> <li>1. Number of base learners (trees) = 100</li> <li>2. Minimum leaf sample size (leaf) = 5</li> <li>3. Task method = regression</li> <li>4. OOBPrediction = on</li> <li>5. OOBPredictorImportance = on</li> </ol>                                         |
| LSTM    | Deep learning model. Improved RNN, introducing memory units and gating mechanisms (input/forget/output gates), precisely controlling the retention, input, and output of information through gates, solving the long-term dependency problem of traditional RNN.                                                                 | <ol style="list-style-type: none"> <li>1. Optimizer: Adam</li> <li>2. MaxEpochs = 1500</li> <li>3. InitialLearnRate = 0.01</li> <li>4. LearnRateSchedule = piecewise</li> <li>5. LearnRateDropFactor = 0.1</li> <li>6. LearnRateDropPeriod = 1200</li> <li>7. Shuffle = every-epoch</li> </ol> |

To avoid spatiotemporal information leakage that could arise from randomly splitting individual city–year records, We evaluated model generalization using five-fold spatial-block cross-validation. Fixed regular grids of  $4^\circ \times 4^\circ$ ,  $5^\circ \times 5^\circ$ , and  $6^\circ \times 6^\circ$ , anchored to integer longitude–latitude multiples in EPSG:4326, were compared using geometric and block-size diagnostics; the  $5^\circ \times 5^\circ$  grid was retained as the primary design, providing 112 complete grid cells and 61 occupied blocks. Each city polygon was assigned uniquely to the block containing the largest fraction of its area, with intersection areas calculated in the equal-area CRS EPSG:6933. Assignment did not use city centroids, city codes, NO<sub>2</sub> values, city counts, or manual adjustment. Occupied blocks were ordered deterministically from southwest to northeast and allocated cyclically to Fold 1–Fold 5. All 2019–2022 records from a city remained in its assigned block. The five held-out sets contained 13, 12, 12, 12, and 12 blocks and 102, 88, 102, 84, and 60 cities, respectively. Within every fold, training and test block IDs were disjoint and their positive-area intersection was 0 m<sup>2</sup>. Each occupied block—and therefore every city—served as test data exactly once. Detailed city assignments to the training and test sets are provided in Supplementary Materials.

We did not adopt leave-one-country-out validation as our primary validation scheme, primarily because the city sample sizes among the five countries were highly imbalanced, with the vast majority of cities located in China. If China were held out entirely, the training set would contain only about 90 cities, substantially reducing the sample size and causing marked differences in spatial coverage, emission levels, and environmental conditions between training and test data. Conversely, holding out a country with relatively few cities—such as North Korea, Mongolia, or South Korea—would result in a very small test set, making the evaluation metrics overly sensitive to a few individual cities and limiting their stability and comparability. Therefore, our design is intended to assess the model's ability to generalise to previously unseen cities within East Asia, rather than to extrapolate to an entirely unseen country; corresponding conclusions have been duly qualified in the text.

### 2.5. Inversion Accuracy Verification

Model comparison was based on the five-fold spatial block cross-validation design described above. In each fold, models were fitted using cities in the training blocks and evaluated exclusively on cities in the held-out blocks. All preprocessing steps and parameter selection were performed using the training blocks only, and the held-out blocks were not used during model fitting or parameter selection. Predictions from the five held-out block sets were concatenated across folds before  $R^2$ , RMSE, MAE, and MAPE were calculated. Because each spatial block served as the held-out set in exactly one fold, each city-year observation contributed exactly once to these pooled out-of-fold test metrics. Fold-specific training predictions were likewise concatenated to provide descriptive in-sample fitting diagnostics; because training observations recur across folds, these pooled training metrics were used only to characterize model fitting and not as evidence of independent spatial generalization. After model selection, the final RF model was refitted using all available 2019–2022 city-year samples for historical reconstruction [5]. The definitions of  $R^2$ , RMSE, MAE, and MAPE are as follows:

The  $R^2$  statistic is an important indicator of model fit, calculated as the ratio of the sum of squares of regression to the total sum of squares. The value reflects the relative contribution of regression. The specific calculation formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

RMSE is the square root of the mean of the squared differences between the predicted values and the true values, with a magnitude comparable to the true values; this index is used to measure the degree of deviation between the predicted values and the true values and is more sensitive to outliers in the data. Its calculation formula is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (3)$$

MAE is a core indicator for measuring the deviation between predicted values and actual values, obtained by calculating the arithmetic mean of the absolute errors of all samples, and is less sensitive to outliers than Mean Squared Error. Different from MSE, MAE can truly characterize the average prediction offset of the model; lower MAE values correspond to superior model prediction accuracy, and the results have good interpretability. The calculation formula is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (4)$$

Among them,  $x_i$  is the  $i$ -th measured value,  $y_i$  is the corresponding model estimate,  $\bar{y}$  is the mean of the model estimates, and  $n$  is the number of samples.

MAPE quantifies the mean absolute prediction error relative to the observed value and was calculated as:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

Here,  $y_i$  and  $\hat{y}_i$  denote the observed and predicted NO<sub>2</sub> column densities for sample  $i$ , respectively;  $\bar{y}$  denotes the mean of the observed values used in the  $R^2$  calculation; and  $n$  denotes the number of valid paired samples.

All observed TROPOMI NO<sub>2</sub> values included in the metric calculations were strictly positive; therefore, no zero denominators occurred. No epsilon adjustment or near-zero exclusion threshold was applied, and all positive observations were retained. Because MAPE can assign relatively large errors to observations with low concentrations, it was interpreted together with RMSE and MAE.

In addition, to assess temporal generalization across the available TROPOMI observation period, leave-one-year-out validation was performed for RF. In each iteration, the model was trained on three years and evaluated on the omitted year, with the procedure repeated for 2019, 2020, 2021, and 2022. Year-specific training and omitted-year testing metrics are provided in Supplementary Materials.

Considering that satellite-derived NO<sub>2</sub> column density and ground-level NO<sub>2</sub> concentration represent different atmospheric quantities and have different units, absolute error metrics were not directly used to evaluate their differences. Instead, Pearson correlation coefficient and Spearman rank correlation coefficient were used to assess the consistency between different NO<sub>2</sub> datasets and ground-based observations. Pearson correlation coefficient was used to measure the linear association between the two variables, while Spearman rank correlation coefficient was used to evaluate the consistency of their rank-order relationship among different cities or years.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

$$\rho = \frac{\sum_{i=1}^n (R(x_i) - \bar{R}_x)(R(y_i) - \bar{R}_y)}{\sqrt{\sum_{i=1}^n (R(x_i) - \bar{R}_x)^2} \sqrt{\sum_{i=1}^n (R(y_i) - \bar{R}_y)^2}} \quad (7)$$

where  $x_i$  represents the ground-based NO<sub>2</sub> observation for the  $i$ -th city-year sample,  $y_i$  represents the corresponding satellite-derived or reconstructed NO<sub>2</sub> value,  $(\bar{x})$  and  $(\bar{y})$  are the mean values of the two datasets, and  $(n)$  is the number of matched samples.  $(R(x_i))$  and  $(R(y_i))$  represent the ranks of  $x_i$  and  $y_i$ , respectively.

## 2.6. External Validation and Statistical Analysis of Temporal Trends

External validation was performed at the city-year scale using the OMI tropospheric NO<sub>2</sub> product and ground-based NO<sub>2</sub> observations from China and Japan. The RF product was matched with OMI for all 436 common cities during 2005–2022. RF and OMI were also matched separately with available ground observations in each country. Missing observations were handled by pairwise deletion, without zero filling or temporal interpolation. Because satellite NO<sub>2</sub> column density and ground-level NO<sub>2</sub> concentration represent different atmospheric quantities and have different units, these comparisons were intended to evaluate spatial and temporal consistency rather than absolute agreement.

For each year, the Pearson correlation coefficient was calculated across matched cities to quantify linear consistency in intercity spatial differences. Spearman's rank correlation coefficient was used where rank-based consistency was required, including the comparison

of city-level trend slopes. Uncertainty in each annual Pearson correlation was estimated using 1000 city-level bootstrap resamples, and percentile-based 95% confidence intervals were reported. To evaluate temporal agreement within individual cities, city-specific Pearson correlation coefficients were calculated between annual ground observations and the corresponding RF estimates and OMI observations using identical city–year pairs. Only cities with at least five valid paired years and nonconstant time series were included. Spearman rank correlations were calculated as a sensitivity analysis. Complete city-level results are provided in Supplementary Materials. For regional and national temporal comparisons, annual city means were calculated first. The 95% confidence interval of each annual mean was expressed as the mean  $\pm$  1.96 standard errors.

Because the RF, OMI, and ground series had different units and value ranges, each annual series was independently transformed to the interval [0, 1] using min–max normalization:

$$z_t = [x_t - \min(x)] / [\max(x) - \min(x)]. \quad (8)$$

Temporal agreement was evaluated using Pearson's  $r$ , normalized root-mean-square error (RMSEn), and the Theil–Sen slope. The Theil–Sen estimator, defined as the median of all pairwise annual slopes, was used because it is less sensitive to outlying years than ordinary least-squares regression. The significance of monotonic trends was evaluated using the two-sided Mann–Kendall test, with  $p < 0.05$  considered statistically significant.

For the 2000–2022 RF series, change points were examined using segmented trend analysis. Candidate segmentations were compared using the Bayesian information criterion, subject to a minimum of five annual observations per segment. Theil–Sen slopes and Mann–Kendall  $p$ -values were then calculated for the selected segments, in addition to the full-period Mann–Kendall test. This combination allowed the identification of potential multi-stage changes while providing a robust assessment of both period-specific and full-period temporal trends.

To examine whether the observations from 2020 to 2021 materially affected the reported long-term trends, we repeated the East Asian and country-level analyses after excluding 2020 and 2021, while retaining 2022. The sensitivity dataset therefore contained 21 annual observations covering 2000–2019 and 2022. The same Mann–Kendall test, Theil–Sen slope estimation, Benjamini–Hochberg adjustment, segmented regression, and Pettitt test used in the main analysis were repeated. The country-specific results are provided in Supplementary Materials. This analysis was used only to evaluate the robustness of the long-term trend conclusions and was not interpreted as an estimate of pandemic-related effects.

## 2.7. SHAP-Based Interpretation of the RF Model

To improve the transparency of the RF model, SHapley Additive exPlanations (SHAP) were used to quantify the contribution of each predictor to the model output [24]. SHAP is a post hoc feature attribution framework that decomposes a model prediction into predictor-specific contributions relative to a reference value. In this study, SHAP was used to characterize global predictor importance and the direction of predictor associations with RF predictions. These attribution values describe the behavior of the fitted model and should not be interpreted as independent causal effects of the predictors on atmospheric NO<sub>2</sub>.

The core calculation formula of SHAP is as follows:

$$\phi_i = \sum_{S \subseteq \{1, \dots, N\} \setminus \{i\}} \frac{|S|! (N - |S| - 1)!}{N!} (f(S \cup \{i\}) - f(S)) \quad (9)$$

where  $S$  is the feature subset excluding feature  $i$ ;  $|S|$  is the number of features in the subset;  $N$  is the total number of features;  $f(S \cup \{i\})$  and  $f(S)$  represent the model retrieval values with and without feature  $i$ , respectively.

In this study, the 18 final model predictors were ranked according to the mean absolute SHAP values calculated from the model evaluation samples. A larger mean absolute SHAP value indicates stronger reliance of the fitted RF model on that predictor and should not be interpreted as an independent causal effect on atmospheric  $\text{NO}_2$ .

In addition to the mean absolute SHAP values, a SHAP beeswarm plot was used to characterize the direction and distribution of predictor contributions across all samples. Positive SHAP values indicate that a predictor increases the modeled  $\text{NO}_2$  column concentration, whereas negative values indicate a decreasing contribution. The color gradient represents the magnitude of each predictor. Calendar year was not explicitly included as a predictor because it has no direct physical meaning; instead, temporal information was represented by the annual variations in the emission, meteorological, and environmental predictors.

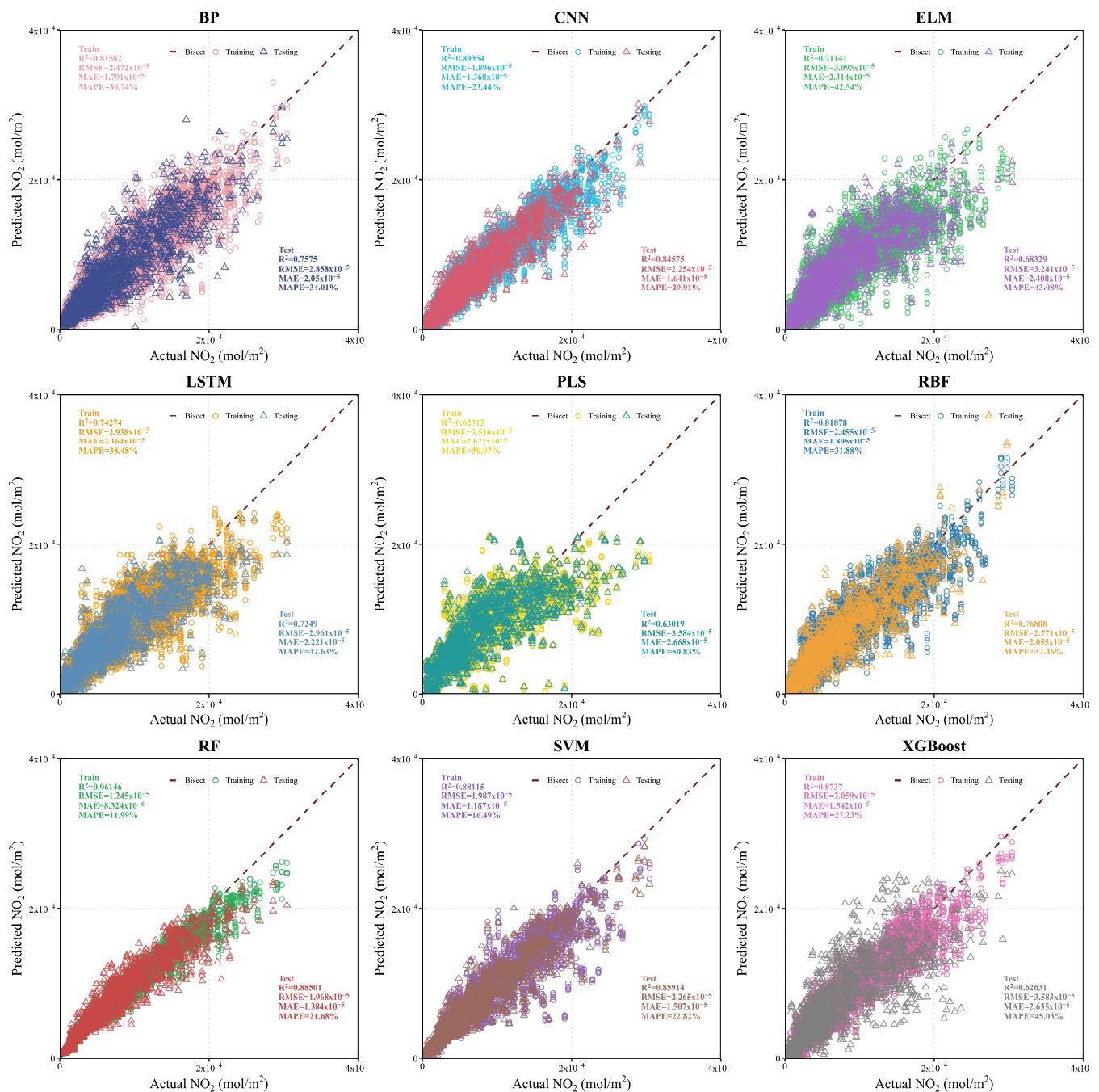
SHAP dependence plots were constructed for the five predictors with the largest global mean absolute SHAP values:  $\text{NO}_x_{\text{ind\_combustion}}$ ,  $P$ ,  $\text{NO}_x_{\text{road}}$ ,  $\text{NO}_x_{\text{ind\_process}}$ , and  $\text{NO}_x_{\text{agriculture}}$ . Each point represents one independent test city–year. Because the four sectoral emission predictors were strongly right-skewed, their horizontal axes were displayed as  $\log_{10}(x + 1)$ , whereas surface pressure was retained in its original hPa scale. For visualization, medians and interquartile ranges of SHAP values were overlaid for 20 equal-frequency bins. Spearman's  $\rho$  was also calculated between each raw predictor and its SHAP value to summarize monotonic association. These plots describe conditional contributions within the fitted RF model and are not partial-dependence estimates or causal effects.

Ranking stability was additionally assessed using 10,000 grouped-city bootstrap resamples of the independent test set, with city as the resampling unit and all four annual records retained together. Complete procedures and results are provided in the Supplementary Materials. This analysis evaluates sample–resampling stability of the existing fixed-model SHAP explanations rather than stability across refitted RF models.

### 3. Results

#### 3.1. $\text{NO}_2$ Inversion Accuracy of the Nine Models

To quantitatively evaluate the inversion performance and spatial generalization ability of the nine models, including BP, CNN, ELM, PLS, LSTM, RBF, RF, SVM, and XGBoost, four accuracy indicators were calculated: the coefficient of determination ( $R^2$ ), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Model performance was assessed using polygon-based, non-overlapping block-level five-fold spatial cross-validation. All 436 cities were assigned uniquely to 61 occupied  $5^\circ \times 5^\circ$  grid blocks according to maximum polygon intersection area, and entire blocks—not individual cities—were held out in each fold, with zero positive-area overlap between training and testing blocks. Training set indicators therefore describe model fitting, whereas the pooled held-out-block testing indicators provide the primary basis for evaluating out-of-block spatial generalization. The detailed metrics and observed-versus-predicted comparisons are presented in Table 4 and Figure 2.



**Figure 2.** Training and test results of  $\text{NO}_2$  retrieval models based on different machine learning algorithms.

The results show clear differences among the nine models in  $\text{NO}_2$  column concentration inversion. Based primarily on the spatially held-out testing results, RF delivered the strongest overall performance, followed by SVM and CNN. RBF and BP formed a middle-performing group, LSTM and ELM showed weaker predictive accuracy, and PLS and XGBoost exhibited the poorest overall spatial generalization. Thus, the updated results support a clearer performance hierarchy rather than treating all non-leading models as a single group.

**Table 4.** Training and testing results of the nine NO<sub>2</sub> retrieval models under five-fold spatial cross-validation.

| Model   | Training Set   |                              |                             |             | Testing Set    |                              |                             |             |
|---------|----------------|------------------------------|-----------------------------|-------------|----------------|------------------------------|-----------------------------|-------------|
|         | R <sup>2</sup> | RMSE<br>( $\times 10^{-5}$ ) | MAE<br>( $\times 10^{-5}$ ) | MAPE<br>(%) | R <sup>2</sup> | RMSE<br>( $\times 10^{-5}$ ) | MAE<br>( $\times 10^{-5}$ ) | MAPE<br>(%) |
| BP      | 0.81582        | 2.472                        | 1.791                       | 30.74       | 0.75750        | 2.858                        | 2.050                       | 34.01       |
| SVM     | 0.88115        | 1.987                        | 1.187                       | 16.49       | 0.85914        | 2.265                        | 1.507                       | 22.82       |
| LSTM    | 0.74274        | 2.938                        | 2.164                       | 38.48       | 0.72490        | 2.961                        | 2.221                       | 42.63       |
| RBF     | 0.81878        | 2.455                        | 1.805                       | 31.88       | 0.76808        | 2.771                        | 2.055                       | 37.46       |
| RF      | 0.96146        | 1.245                        | 0.8324                      | 11.99       | 0.88501        | 1.968                        | 1.384                       | 21.68       |
| XGBoost | 0.87370        | 2.059                        | 1.542                       | 27.23       | 0.62631        | 3.583                        | 2.635                       | 45.03       |
| ELM     | 0.71141        | 3.095                        | 2.314                       | 42.54       | 0.68329        | 3.241                        | 2.408                       | 43.08       |
| PLS     | 0.62315        | 3.516                        | 2.627                       | 50.07       | 0.63019        | 3.584                        | 2.668                       | 50.83       |
| CNN     | 0.89354        | 1.896                        | 1.368                       | 23.44       | 0.84575        | 2.254                        | 1.641                       | 29.91       |

Among all models, RF achieved the best overall performance. In the training set, RF obtained the highest R<sup>2</sup> value of 0.96146, with an RMSE of  $1.245 \times 10^{-5}$  mol m<sup>-2</sup>, an MAE of  $8.324 \times 10^{-6}$  mol m<sup>-2</sup>, and a MAPE of 11.99%. In the spatially held-out testing set, RF again achieved the highest R<sup>2</sup> (0.88501) and the lowest RMSE ( $1.968 \times 10^{-5}$  mol m<sup>-2</sup>), MAE ( $1.384 \times 10^{-5}$  mol m<sup>-2</sup>), and MAPE (21.68%) among all nine models. These results indicate that RF combined strong fitting capacity with the most accurate out-of-block predictions under the true spatial validation design.

SVM ranked second overall and showed particularly stable spatial generalization. Its R<sup>2</sup> values were 0.88115 in the training set and 0.85914 in the testing set, corresponding to a relatively small decrease of 0.02201. The testing set RMSE, MAE, and MAPE were  $2.265 \times 10^{-5}$  mol m<sup>-2</sup>,  $1.507 \times 10^{-5}$  mol m<sup>-2</sup>, and 22.82%, respectively. Although these values did not surpass RF, SVM provided the closest overall alternative, and its testing set MAPE was only 1.14 percentage points higher than that of RF.

CNN also presented strong predictive performance, with training and testing set R<sup>2</sup> values of 0.89354 and 0.84575, respectively. Its testing set RMSE was  $2.254 \times 10^{-5}$  mol m<sup>-2</sup>, slightly lower than that of SVM and second only to RF, while its testing set MAE and MAPE were  $1.641 \times 10^{-5}$  mol m<sup>-2</sup> and 29.91%. Accordingly, CNN captured the overall variation in NO<sub>2</sub> column concentrations well, but its larger relative error than RF and SVM places it third in the updated overall assessment.

In comparison, BP, RBF, LSTM, ELM, PLS, and XGBoost showed less satisfactory performance on spatially held-out blocks. Their testing set R<sup>2</sup> values ranged from 0.62631 to 0.76808, RMSE values ranged from  $2.771 \times 10^{-5}$  to  $3.584 \times 10^{-5}$  mol m<sup>-2</sup>, MAE values ranged from  $2.050 \times 10^{-5}$  to  $2.668 \times 10^{-5}$  mol m<sup>-2</sup>, and MAPE values ranged from 34.01% to 50.83%. RBF and BP retained moderate predictive ability, whereas LSTM, ELM, PLS, and XGBoost showed progressively greater limitations in accuracy or error control when transferred to unseen spatial blocks.

Specifically, RBF achieved the highest testing set R<sup>2</sup> within this group (0.76808), whereas BP produced slightly lower absolute and relative errors (MAE =  $2.050 \times 10^{-5}$  mol m<sup>-2</sup>; MAPE = 34.01%). LSTM showed a small training-to-testing R<sup>2</sup> change but only moderate testing accuracy (R<sup>2</sup> = 0.72490; MAPE = 42.63%), and ELM and PLS remained comparatively weak, with PLS recording the highest testing set MAPE (50.83%). XGBoost displayed the clearest evidence of overfitting: its R<sup>2</sup> decreased from 0.87370 in training to 0.62631 in testing, while RMSE rose from  $2.059 \times 10^{-5}$  to  $3.583 \times 10^{-5}$  mol m<sup>-2</sup> and MAPE increased from 27.23% to 45.03%. This pronounced deteri-

oration shows that strong in-sample fitting did not translate into reliable spatial generalization.

The leave-one-year-out analysis showed stable omitted-year performance across the four TROPOMI years:  $R^2$  ranged from 0.85622 to 0.87421, RMSE from  $1.958 \times 10^{-5}$  to  $2.155 \times 10^{-5} \text{ mol m}^{-2}$ , MAE from  $1.415 \times 10^{-5}$  to  $1.463 \times 10^{-5} \text{ mol m}^{-2}$ , and MAPE from 20.51% to 22.01%. No marked deterioration occurred when 2020 or 2021 served as the held-out year, indicating that the unusual conditions during these years did not strongly alter model performance. These results support short-term temporal generalization across the 2019–2022 TROPOMI observation period. Detailed annual metrics are reported in Supplementary Materials.

The RF–TROPOMI comparison was recalculated for 1744 matched city–year pairs (436 cities  $\times$  four years; Table 5), with bias defined as RF minus TROPOMI. Across all cities, the signed relative biases were 1.87%, 5.24%,  $-7.90\%$ , and  $2.44\%$  in 2019, 2020, 2021, and 2022, respectively, while the corresponding RMSE values were 0.972, 0.979, 1.569, and  $0.869 \times 10^{-5} \text{ mol m}^{-2}$ . To evaluate country-level heterogeneity, China (346 cities) and Japan (46 cities)—the only country strata containing at least 30 cities—were reported separately by year. China showed relative biases of 2.03%, 5.24%,  $-8.29\%$ , and  $2.74\%$ , whereas Japan showed 2.01%, 7.04%,  $-7.99\%$ , and  $-0.08\%$ . Both major country strata therefore reproduced the negative-bias pattern in 2021, although the magnitude and RMSE differed between countries. Pooled 2019–2022 estimates for Mongolia (18 cities), South Korea (15 cities), and North Korea (11 cities) are presented as exploratory because of their smaller sample sizes. These results show that a near-zero pooled regional bias can mask compensating temporal and country-specific errors.

**Table 5.** Overall annual and country-stratified bias metrics between the RF estimates and TROPOMI observations during 2019–2022.

| Scope         | Period    | Cities | N   | Mean Bias | Median Bias | RMSE  | Relative Bias (%) |
|---------------|-----------|--------|-----|-----------|-------------|-------|-------------------|
| All cities    | 2019      | 436    | 436 | 0.150     | 0.215       | 0.972 | 1.87              |
| All cities    | 2020      | 436    | 436 | 0.408     | 0.400       | 0.979 | 5.24              |
| All cities    | 2021      | 436    | 436 | $-0.748$  | $-0.281$    | 1.569 | $-7.90$           |
| All cities    | 2022      | 436    | 436 | 0.192     | 0.268       | 0.869 | 2.44              |
| China         | 2019      | 346    | 346 | 0.173     | 0.255       | 0.978 | 2.03              |
| China         | 2020      | 346    | 346 | 0.437     | 0.463       | 1.036 | 5.24              |
| China         | 2021      | 346    | 346 | $-0.849$  | $-0.405$    | 1.691 | $-8.29$           |
| China         | 2022      | 346    | 346 | 0.227     | 0.332       | 0.915 | 2.74              |
| Japan         | 2019      | 46     | 46  | 0.119     | 0.162       | 0.599 | 2.01              |
| Japan         | 2020      | 46     | 46  | 0.384     | 0.358       | 0.545 | 7.04              |
| Japan         | 2021      | 46     | 46  | $-0.524$  | $-0.229$    | 1.003 | $-7.99$           |
| Japan         | 2022      | 46     | 46  | $-0.005$  | 0.094       | 0.640 | $-0.08$           |
| Mongolia *    | 2019–2022 | 18     | 72  | 0.057     | 0.061       | 0.284 | 4.92              |
| South Korea * | 2019–2022 | 15     | 60  | $-0.352$  | $-0.086$    | 1.435 | $-2.46$           |
| North Korea * | 2019–2022 | 11     | 44  | 0.525     | 0.439       | 0.751 | 14.09             |

Note: Mean bias, median bias, and RMSE are expressed in  $10^{-5} \text{ mol m}^{-2}$ . Bias was defined as RF minus TROPOMI, and relative bias was calculated as  $\text{sum}(\text{RF} - \text{TROPOMI}) / \text{sum}(\text{TROPOMI}) \times 100\%$ . “Cities” denotes the number of distinct cities and N denotes city–year pairs. All-city, China, and Japan rows are reported annually. An asterisk indicates an exploratory estimate: Mongolia, South Korea, and North Korea are pooled over 2019–2022 because each country contained fewer than 30 retained cities.

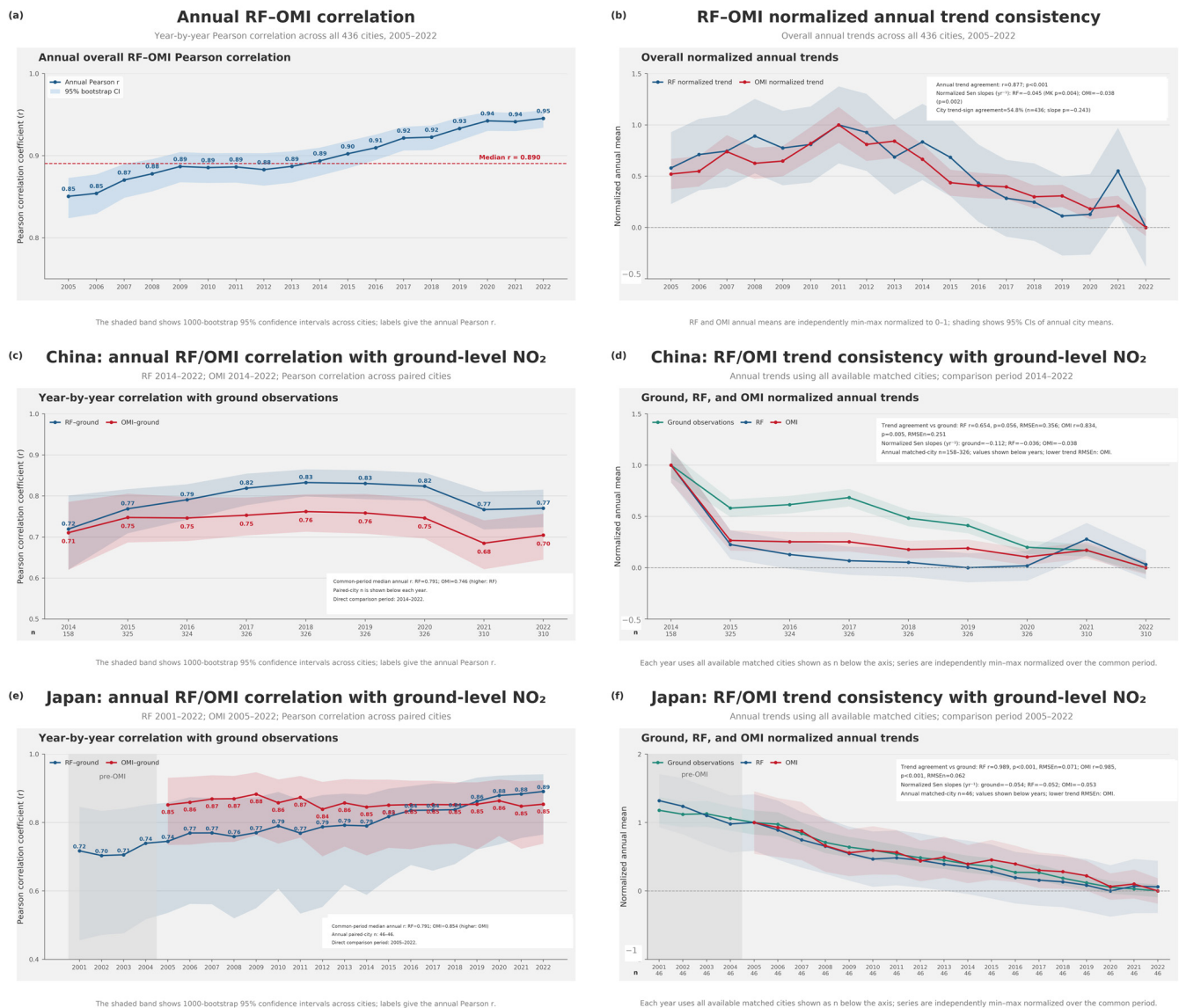
In summary, RF was identified as the optimal model for NO<sub>2</sub> column concentration inversion under the true block-level spatial validation design. It achieved the highest testing set R<sup>2</sup> and the lowest testing set RMSE, MAE, and MAPE, demonstrating the strongest out-of-block generalization among the nine models. SVM was the most competitive alternative, while CNN ranked third because its relative-error control was weaker despite a low RMSE. The marked training-to-testing deterioration of XGBoost further highlights the importance of spatially blocked validation for avoiding overly optimistic model assessment. Therefore, the RF inversion results were retained as the basis for subsequent spatiotemporal pattern analysis and driving-mechanism interpretation.

### 3.2. Ground-Based Validation and Comparison with OMI

To further evaluate the reliability of the annual NO<sub>2</sub> dataset developed in this study, ground-based NO<sub>2</sub> observations from China and Japan were used as independent reference data. Because satellite-derived NO<sub>2</sub> column density and ground-level NO<sub>2</sub> concentration represent different atmospheric quantities and are expressed in different units, their absolute differences are not directly comparable. Therefore, Pearson correlation coefficients were used to evaluate linear consistency in relative intercity variations. The annual NO<sub>2</sub> dataset reconstructed using the random forest model (hereafter, the RF dataset) and the OMI NO<sub>2</sub> product were separately matched with ground observations at the city–year scale, and annual correlations were calculated with pairwise deletion. Temporal consistency was evaluated after independent min–max normalization of the annual mean series using Pearson *r*, normalized root-mean-square error (RMSEn), Theil–Sen slopes, and two-sided Mann–Kendall tests (Figure 3). The RF dataset for all years, including 2019–2022, consists of model-retrieved values generated using a consistent random forest retrieval framework; it does not contain a splice of TROPOMI observations and historical reconstructions.

The RF dataset and OMI product showed strong annual spatial agreement during 2005–2022. All 436 common cities were available in each year, and the annual Pearson correlations ranged from 0.850 to 0.945 (mean = 0.900; median = 0.890), increasing from 0.850 in 2005 to 0.945 in 2022. After independent normalization, the correlation between the RF and OMI annual mean series was 0.877 ( $p < 0.001$ ), with an RMSEn of 0.152. Their normalized Theil–Sen slopes were  $-0.045$  and  $-0.038 \text{ yr}^{-1}$ , respectively, and both declines were significant (Mann–Kendall  $p < 0.01$ ). At the city level, however, RF and OMI trend directions agreed in only 54.8% of cities, and the city-specific slopes were negatively rank-correlated (Spearman  $\rho = -0.243$ ,  $p < 0.001$ ). Thus, the strong annual cross-sectional and aggregate temporal agreement did not translate into consistent local trend magnitudes or rankings.

For Chinese cities, annual RF–ground correlations ranged from 0.719 to 0.833 during 2014–2022 (mean = 0.791; median = 0.791), whereas OMI–ground correlations ranged from 0.685 to 0.762 (mean = 0.735; median = 0.746). RF outperformed OMI in the annual cross-sectional comparison in all nine validation years. The available matched samples comprised 158 cities in 2014, 324–326 cities during 2015–2020, and 310 cities in 2021–2022. The normalized annual-mean comparison produced a different result: RF versus ground yielded  $r = 0.654$  ( $p = 0.056$ ) and RMSEn = 0.356, whereas OMI versus ground yielded  $r = 0.834$  ( $p = 0.005$ ) and RMSEn = 0.251. The normalized Theil–Sen slopes of the ground observations, RF, and OMI were  $-0.112$ ,  $-0.036$ , and  $-0.038 \text{ yr}^{-1}$ , respectively. RF therefore represented annual relative differences among Chinese cities more effectively, while OMI tracked the national-scale annual variation more closely; RF captured the declining direction but underestimated its ground-observed magnitude.



**Figure 3.** Ground-based validation of the reconstructed NO<sub>2</sub> dataset and OMI using city-year NO<sub>2</sub> observations in China (2014–2022) and Japan (2001–2022; common RF–OMI period 2005–2022).

For Japanese cities, RF–ground correlations during the common RF–OMI period of 2005–2022 ranged from 0.745 to 0.891 (mean = 0.810; median = 0.791), while OMI–ground correlations ranged from 0.839 to 0.883 (mean = 0.857; median = 0.854). OMI was higher in 14 of the 18 common years, whereas RF was slightly higher from 2019 to 2022. Before OMI became available, RF–ground correlations during 2001–2004 ranged from 0.703 to 0.739. The normalized annual mean series of both satellite products closely followed the ground series: RF yielded  $r = 0.989$  and  $\text{RMSEn} = 0.071$ , and OMI yielded  $r = 0.985$  and  $\text{RMSEn} = 0.062$  (both  $p < 0.001$ ). The normalized Theil–Sen slopes of the ground observations, RF, and OMI were  $-0.054$ ,  $-0.052$ , and  $-0.053 \text{ yr}^{-1}$ , respectively. Accordingly, OMI generally reproduced annual intercity differences more strongly, but both products closely reconstructed Japan’s national long-term decline.

City-specific temporal correlations provided a complementary within-city assessment. In China during 2014–2022, RF–ground correlation exceeded OMI–ground correlation in 93 of 326 cities (28.5%), and the median city-specific correlations were 0.277 for RF and 0.612 for OMI. In Japan during 2005–2022, RF performed better in 23 of 46 cities (50.0%), while the median city-specific correlations remained high for both RF (0.885) and OMI

(0.925). These calculations used identical available city–year pairs for the two products and required at least five valid annual observations per city.

Overall, the updated ground-based validation confirms country- and scale-dependent performance. RF better represented annual cross-sectional intercity differences in China, whereas OMI more closely tracked within-city and national temporal variation. In Japan, OMI generally performed better for annual intercity differences, while both products showed high within-city temporal agreement and closely reproduced the national decline. These results support the use of the RF dataset primarily for regional and national-scale spatiotemporal assessment, while city-specific trend magnitudes and rankings should be interpreted cautiously because local agreement with OMI and ground observations remains limited [25].

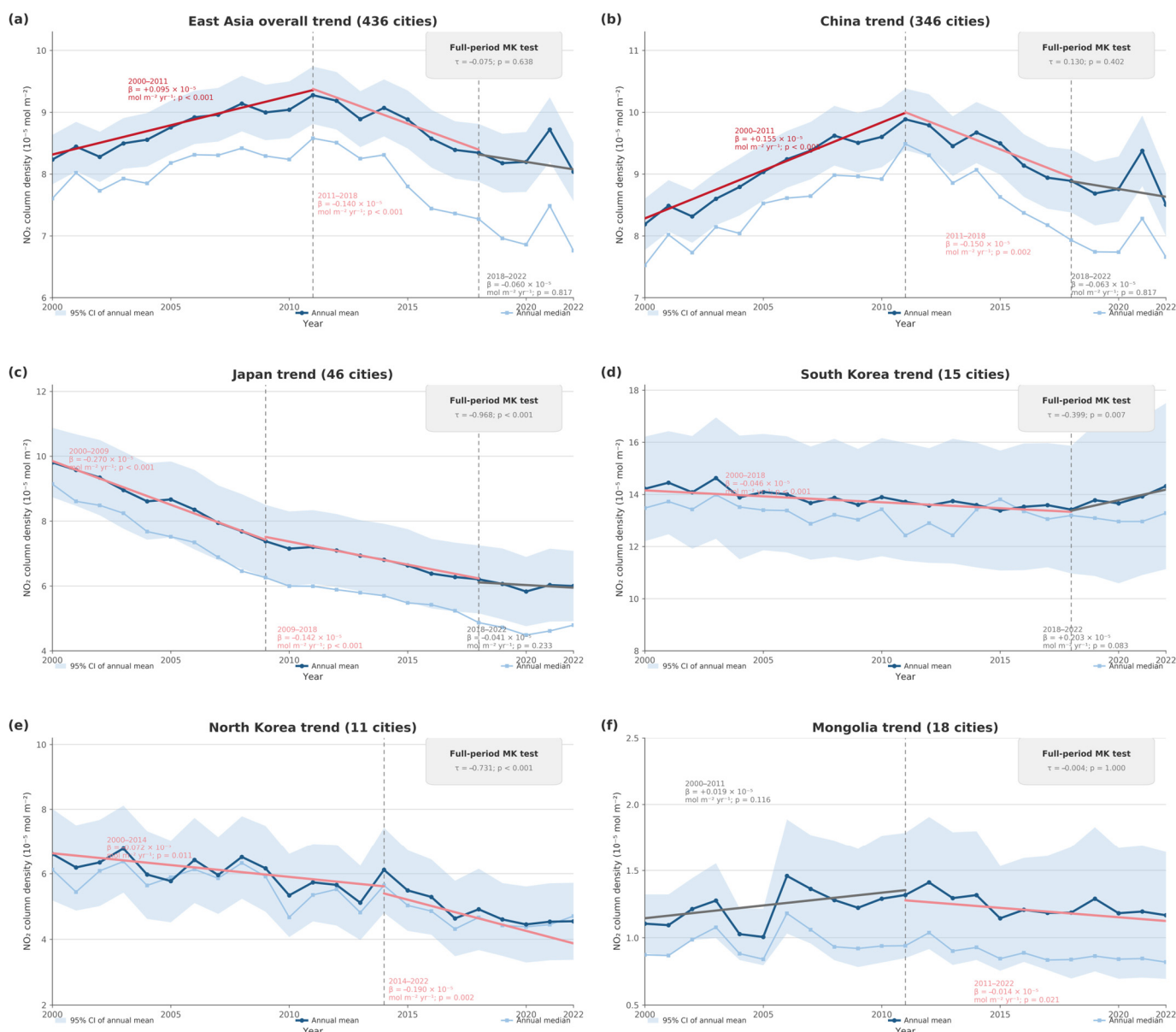
### 3.3. Spatiotemporal Distribution Characteristics of NO<sub>2</sub> Column Density

The temporal evolution of city-level NO<sub>2</sub> column density showed clear interannual variability and substantial differences among the five East Asian countries (Figure 4). At the regional scale, the annual mean increased from  $8.23 \times 10^{-5} \text{ mol m}^{-2}$  in 2000 to  $8.76 \times 10^{-5} \text{ mol m}^{-2}$  in 2005 and  $9.04 \times 10^{-5} \text{ mol m}^{-2}$  in 2010, reaching a maximum of  $9.28 \times 10^{-5} \text{ mol m}^{-2}$  in 2011. The corresponding median increased from  $7.60 \times 10^{-5} \text{ mol m}^{-2}$  in 2000 to  $8.24 \times 10^{-5} \text{ mol m}^{-2}$  in 2010 and  $8.58 \times 10^{-5} \text{ mol m}^{-2}$  in 2011. Thereafter, both measures generally declined: the mean and median were  $8.88 \times 10^{-5}$  and  $7.80 \times 10^{-5} \text{ mol m}^{-2}$  in 2015,  $8.20 \times 10^{-5}$  and  $6.86 \times 10^{-5} \text{ mol m}^{-2}$  in 2020, and  $8.04 \times 10^{-5}$  and  $6.76 \times 10^{-5} \text{ mol m}^{-2}$  in 2022, respectively.

BIC-selected segmented analysis identified breakpoints in 2011 and 2018 for the East Asian mean series. The regional trend increased significantly during 2000–2011 ( $\beta = +0.095 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p < 0.001$ ), decreased significantly during 2011–2018 ( $\beta = -0.140 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p < 0.001$ ), and remained nonsignificantly negative during 2018–2022 ( $\beta = -0.060 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p = 0.817$ ). Because these phases offset one another, the full-period Mann–Kendall test indicated no significant monotonic trend over 2000–2022 ( $\tau = -0.075$ ,  $p = 0.638$ ).

China, which accounted for 346 of the 436 cities, exhibited a temporal pattern similar to the regional series. Its mean increased from  $8.19 \times 10^{-5} \text{ mol m}^{-2}$  in 2000 to a maximum of  $9.89 \times 10^{-5} \text{ mol m}^{-2}$  in 2011 and then declined to  $8.50 \times 10^{-5} \text{ mol m}^{-2}$  in 2022; the corresponding medians were  $7.53 \times 10^{-5}$ ,  $9.49 \times 10^{-5}$ , and  $7.66 \times 10^{-5} \text{ mol m}^{-2}$ . The 2000–2011 increase was significant ( $\beta = +0.155 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p < 0.001$ ), as was the 2011–2018 decrease ( $\beta = -0.150 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p = 0.002$ ). The 2018–2022 slope remained negative but was not significant ( $\beta = -0.063 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p = 0.817$ ). The full-period trend was likewise nonsignificant ( $\tau = 0.130$ ,  $p = 0.402$ ).

Japan showed the strongest long-term decline among the five countries, although the rate of decline weakened and the 2018–2022 trend was not statistically significant. Its mean decreased from  $9.81 \times 10^{-5} \text{ mol m}^{-2}$  in 2000 to  $8.67 \times 10^{-5} \text{ mol m}^{-2}$  in 2005,  $7.15 \times 10^{-5} \text{ mol m}^{-2}$  in 2010, and  $6.00 \times 10^{-5} \text{ mol m}^{-2}$  in 2022; the median declined from  $9.14 \times 10^{-5}$  to  $4.79 \times 10^{-5} \text{ mol m}^{-2}$  over the full period. Significant reductions occurred during 2000–2009 ( $\beta = -0.270 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p < 0.001$ ) and 2009–2018 ( $\beta = -0.142 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p < 0.001$ ), followed by a smaller, nonsignificant decline during 2018–2022 ( $\beta = -0.041 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p = 0.233$ ). The full-period Mann–Kendall trend was strongly negative ( $\tau = -0.968$ ,  $p < 0.001$ ).



**Figure 4.** Annual trends in urban NO<sub>2</sub> column density for East Asia and the five countries from 2000 to 2022. Lines show annual means and medians, shaded bands show 95% confidence intervals of the mean, and colored segments show BIC-selected Theil-Sen trends.

South Korea retained the highest absolute NO<sub>2</sub> column density among the five countries. Its mean was  $14.21 \times 10^{-5}$  mol m<sup>-2</sup> in 2000, declined to  $13.38 \times 10^{-5}$  mol m<sup>-2</sup> in 2015, and increased to  $14.32 \times 10^{-5}$  mol m<sup>-2</sup> in 2022. BIC selected a 2018 breakpoint: the 2000–2018 trend was significantly negative ( $\beta = -0.046 \times 10^{-5}$  mol m<sup>-2</sup> yr<sup>-1</sup>,  $p < 0.001$ ), whereas the 2018–2022 increase was not significant at the 0.05 level ( $\beta = +0.203 \times 10^{-5}$  mol m<sup>-2</sup> yr<sup>-1</sup>,  $p = 0.083$ ). Despite the recent rebound, the full-period Mann–Kendall test indicated a significant overall decline ( $\tau = -0.399$ ,  $p = 0.007$ ).

North Korea also showed a significant long-term decrease. Its mean declined from  $6.60 \times 10^{-5}$  mol m<sup>-2</sup> in 2000 to  $5.34 \times 10^{-5}$  mol m<sup>-2</sup> in 2010 and  $4.55 \times 10^{-5}$  mol m<sup>-2</sup> in 2022. The decline was significant during both BIC-selected phases:  $\beta = -0.072 \times 10^{-5}$  mol m<sup>-2</sup> yr<sup>-1</sup> for 2000–2014 ( $p = 0.011$ ) and  $\beta = -0.190 \times 10^{-5}$  mol m<sup>-2</sup> yr<sup>-1</sup> for 2014–2022 ( $p = 0.002$ ). The full-period test likewise confirmed a strong downward trend ( $\tau = -0.731$ ,  $p < 0.001$ ).

Mongolia had the lowest NO<sub>2</sub> column density and substantial year-to-year variability. Its mean increased from  $1.10 \times 10^{-5} \text{ mol m}^{-2}$  in 2000 to a maximum of  $1.46 \times 10^{-5} \text{ mol m}^{-2}$  in 2006, and was  $1.17 \times 10^{-5} \text{ mol m}^{-2}$  in 2022. The 2000–2011 increase was not statistically significant ( $\beta = +0.019 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p = 0.116$ ), whereas the 2011–2022 segment showed a modest significant decline ( $\beta = -0.014 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ,  $p = 0.021$ ). Over the full period, however, there was no monotonic trend ( $\tau = -0.004$ ,  $p = 1.000$ ).

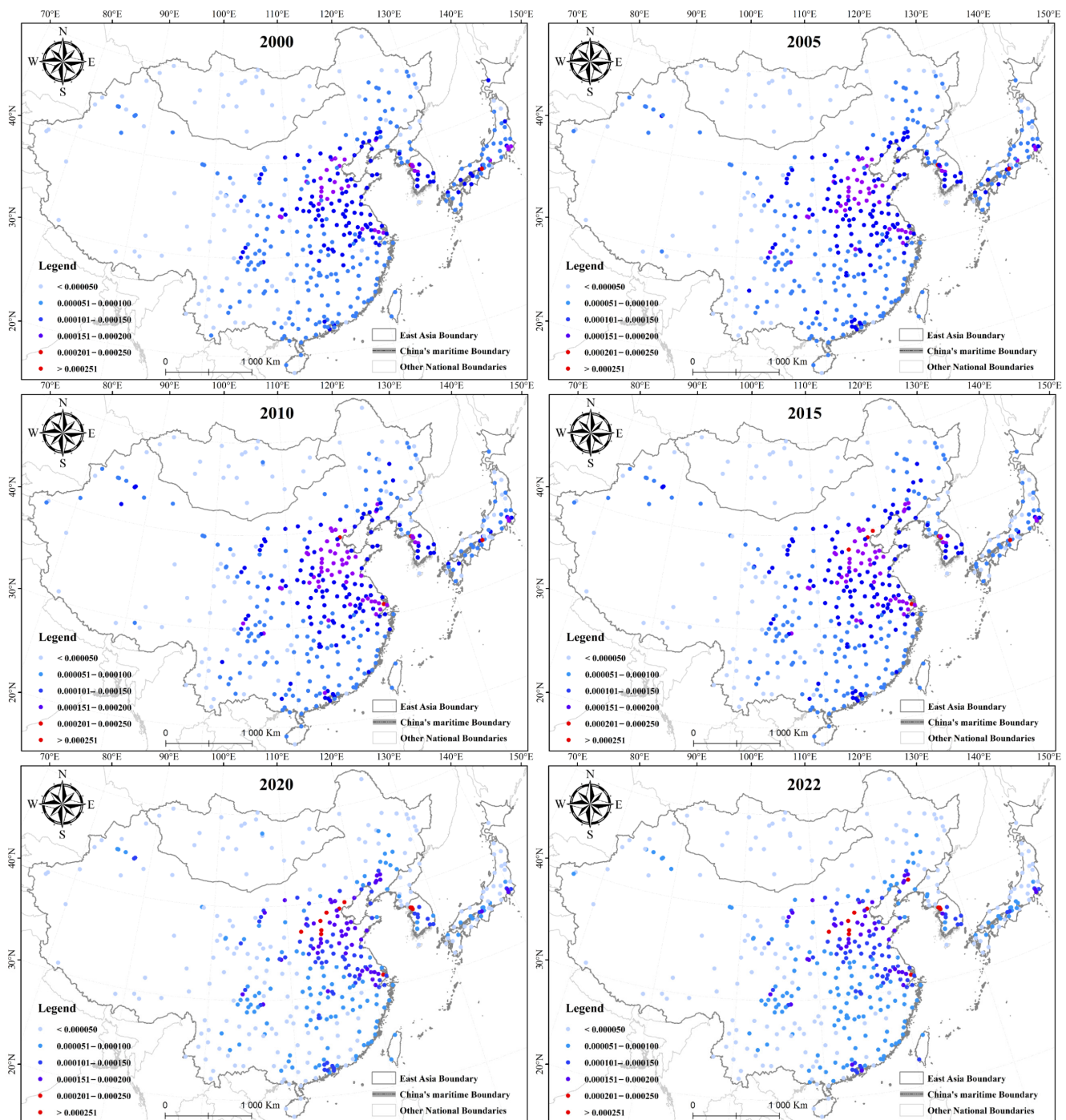
At the regional scale and for most country-years, the annual mean exceeded the corresponding median, consistent with right-skewed city-level NO<sub>2</sub> distributions. The mean–median differences and the widths of the 95% confidence intervals varied markedly among countries and over time, indicating persistent spatial heterogeneity in urban NO<sub>2</sub> column density throughout the study period [26].

Excluding 2020 and 2021 did not change the principal trend conclusions. For East Asia, the Theil–Sen slope changed from  $-0.060 \times 10^{-6} \text{ mol m}^{-2} \text{ yr}^{-1}$  for the complete series to  $+0.028 \times 10^{-6} \text{ mol m}^{-2} \text{ yr}^{-1}$  after exclusion, but both remained close to zero and nonsignificant after Benjamini–Hochberg adjustment ( $p = 0.770$  and  $0.976$ , respectively). Japan, South Korea, and North Korea retained significant negative trends, whereas China and Mongolia remained nonsignificant. Thus, the regional conclusions were not driven by the two pandemic-affected years; complete slope, significance, breakpoint, and Pettitt test results are provided in Supplementary Materials.

The spatial distribution shown at the discrete city centroid locations further confirmed the uneven pattern of NO<sub>2</sub> column density across the five East Asian countries (Figure 5). Figure 5 displays city-level estimates at their corresponding centroid locations rather than a continuously interpolated surface; intervening blank areas therefore indicate areas without represented city-level estimates, rather than zero pollution or a gradual spatial transition. High values were concentrated mainly in densely urbanized and industrialized areas, especially eastern and northern China, including the North China Plain, Beijing–Tianjin–Hebei, the Shandong Peninsula, the Yangtze River Delta, and parts of the eastern coastal urban belt. In contrast, western China, Mongolia, and other sparsely populated inland areas were dominated by lower values.

From 2000 to 2015, the number of cities exceeding  $1.5 \times 10^{-4} \text{ mol m}^{-2}$  increased from 31 in 2000 to 47 in 2005 and 61 in 2010, before reaching 56 in 2015. The corresponding counts were 61 in 2020 and 56 in 2022, showing that a substantial set of high-value urban sites persisted. At the lower end of the distribution, cities below  $5.0 \times 10^{-5} \text{ mol m}^{-2}$  numbered 101 in 2000, 95 in 2005, 101 in 2010, and 107 in 2015, then increased sharply to 150 in 2020 and 152 in 2022. The recent period therefore combined a wider prevalence of low-value cities with continued localized high-value clusters.

Although the prevalence of low-value cities increased after 2015, the spatial change was uneven, and persistent high-value clusters remained visible in 2020 and 2022 in parts of the North China Plain, the Shandong Peninsula, and several major urban agglomerations in eastern China, with additional localized high values in metropolitan areas of South Korea and Japan. Overall, East Asian cities retained a pronounced spatial contrast: higher NO<sub>2</sub> column densities were concentrated in economically developed coastal and industrial regions, whereas lower values predominated in inland and sparsely populated areas.

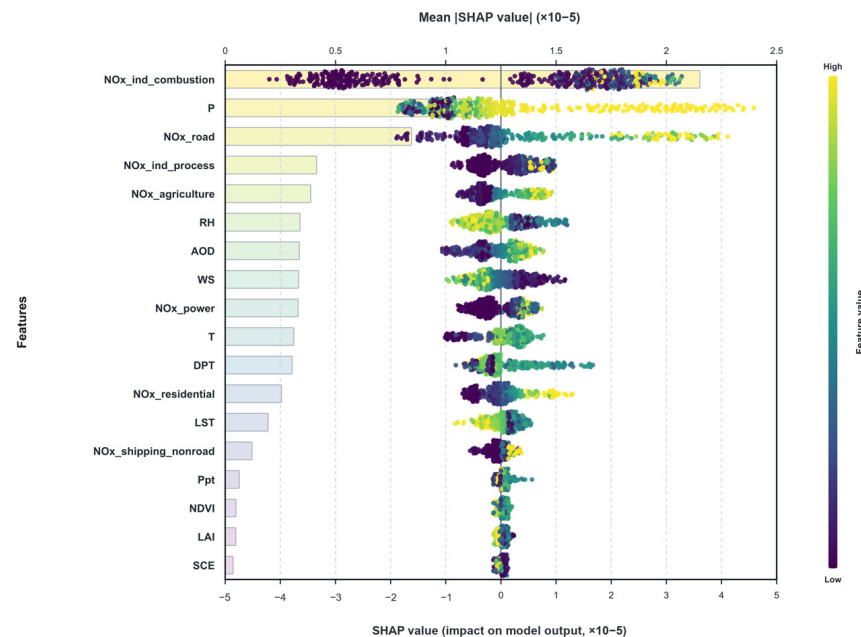


**Figure 5.** Spatial Distribution of  $\text{NO}_2$  in Five East Asian Countries from 2000 to 2022. The symbols represent city centroids used to display the corresponding city-level  $\text{NO}_2$  estimates. The map does not represent a continuously interpolated spatial surface; the white/gap areas between city centroids indicate areas without represented city-level estimates rather than zero or intermediate  $\text{NO}_2$  values.

### 3.4. SHAP-Based Interpretation of Predictor Associations

Figure 6 presents the global importance ranking and SHAP distributions of the predictors in the RF model. The horizontal bars represent the mean absolute SHAP value of each variable; a longer bar indicates a larger global mean absolute contribution to RF predictions. Each point in the beeswarm represents one sample, and its horizontal position is the SHAP value. A positive SHAP value indicates that the variable increases the predicted  $\text{NO}_2$  column density, whereas a negative SHAP value indicates that it decreases the prediction. Point colors from purple to yellow represent feature values from low to

high, so the beeswarm shows the direction of association between predictor values and the RF model output at different feature value levels. These directions represent conditional contributions to the RF model output and should not be interpreted as independent causal effects on atmospheric NO<sub>2</sub>.



**Figure 6.** SHAP-Based Feature Importance Analysis.

From the overall ranking structure, significant differences were observed in the contribution degrees of different influencing factors, showing an obvious hierarchical distribution of importance. The mean absolute SHAP importance order was NO<sub>x</sub>\_ind\_combustion, P, NO<sub>x</sub>\_road, NO<sub>x</sub>\_ind\_process, NO<sub>x</sub>\_agriculture, RH, AOD, WS, NO<sub>x</sub>\_power, T, DPT, NO<sub>x</sub>\_residential, LST, NO<sub>x</sub>\_shipping\_nonroad, Ppt, NDVI, LAI, and SCE. NO<sub>x</sub>\_ind\_combustion, P, and NO<sub>x</sub>\_road had the highest global mean absolute SHAP values and therefore contributed most strongly to variation in the RF model predictions.

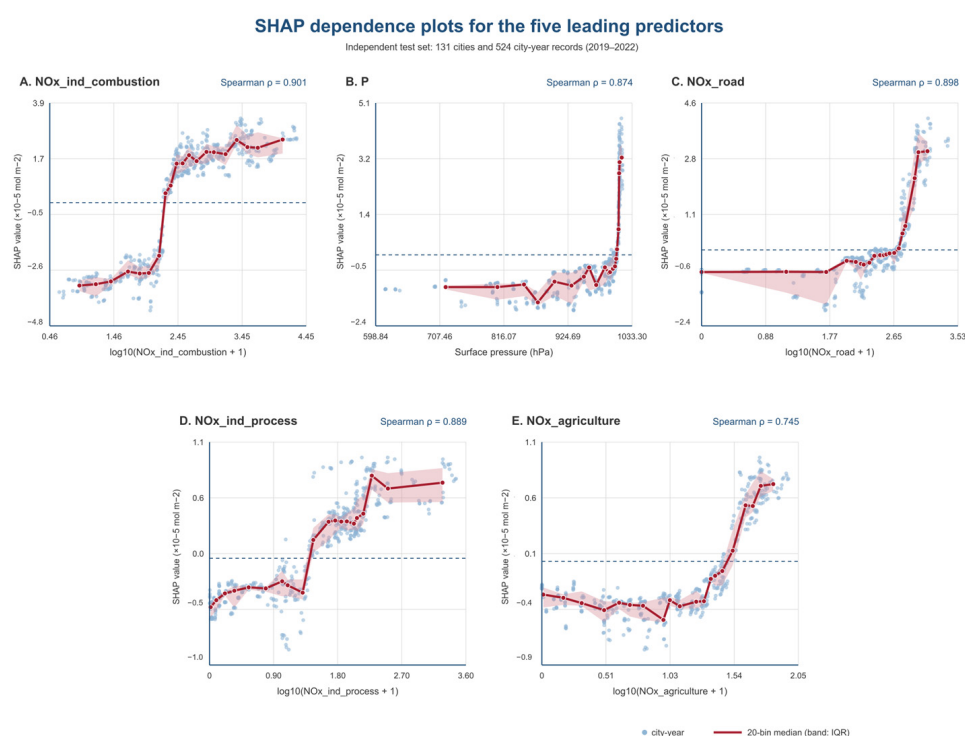
Among the environmental and meteorological factors, the contributions of individual variables to the model reconstruction differed substantially. Their importance order was P, RH, AOD, WS, T, DPT, LST, Ppt, NDVI, LAI, and SCE. The beeswarm indicates that high values of P, AOD, and T were generally associated with positive SHAP contributions, whereas high values of RH and WS were generally associated with negative contributions. High LST values were also predominantly negative, while DPT showed a non-monotonic relationship. In contrast, Ppt, NDVI, LAI, and SCE had relatively lower SHAP values and ranked lower in the overall importance order.

Among anthropogenic NO<sub>x</sub> emission sources, the overall contributions of different sectors also varied considerably. Their importance order was NO<sub>x</sub>\_ind\_combustion, NO<sub>x</sub>\_road, NO<sub>x</sub>\_ind\_process, NO<sub>x</sub>\_agriculture, NO<sub>x</sub>\_power, NO<sub>x</sub>\_residential, and NO<sub>x</sub>\_shipping\_nonroad. These emission variables generally showed positive SHAP contributions at high values and negative contributions at low values, indicating that higher values of these emission predictors were generally associated with higher RF-predicted NO<sub>2</sub> column density. NO<sub>x</sub>\_ind\_combustion exhibited the widest range of SHAP values, whereas NO<sub>x</sub>\_shipping\_nonroad had a weaker overall effect.

From the overall ranking pattern, both anthropogenic NO<sub>x</sub> emission sources and environmental-meteorological factors contributed to the RF-based NO<sub>2</sub> reconstruction model, but the intensity of their effects varied substantially among different variables.

$\text{NO}_x\text{_{ind\_combustion}}$ ,  $P$ , and  $\text{NO}_x\text{_{road}}$  were the most important predictors, and the beeswarm further supplied positive and negative directional information that is not captured by the mean absolute SHAP ranking.

The SHAP dependence plots showed positive but nonlinear associations between predictor values and their conditional contributions to the RF estimates (Figure 7). Spearman's rho between the raw predictor and its SHAP value was 0.901 for  $\text{NO}_x\text{_{ind\_combustion}}$ , 0.874 for  $P$ , 0.898 for  $\text{NO}_x\text{_{road}}$ , 0.889 for  $\text{NO}_x\text{_{ind\_process}}$ , and 0.745 for  $\text{NO}_x\text{_{agriculture}}$ . The binned medians generally transitioned from negative contributions at lower predictor values to positive contributions at higher values, although the transition ranges differed among predictors. These patterns represent conditional model associations rather than causal effects.



**Figure 7.** The SHAP dependence plots for the five leading predictors.

A 10,000-replicate grouped-city bootstrap further indicated high ranking stability (median Spearman's rho = 0.986; the original top-five set was recovered in 96.97% of resamples), with complete results provided in Supplementary Materials.

## 4. Discussion

### 4.1. Advantage of Using RF

Based on the updated model comparison results, RF, SVM, and CNN all showed relatively strong performance in  $\text{NO}_2$  column density inversion, but RF achieved the strongest overall spatial generalization, followed by SVM and CNN. Compared with SVM, RF achieved a higher testing set  $R^2$  (0.88501 versus 0.85914), a lower RMSE ( $1.968 \times 10^{-5}$  versus  $2.265 \times 10^{-5} \text{ mol m}^{-2}$ ), a lower MAE ( $1.384 \times 10^{-5}$  versus  $1.507 \times 10^{-5} \text{ mol m}^{-2}$ ), and a lower MAPE (21.68% versus 22.82%). Compared with CNN, RF also achieved a higher testing set  $R^2$  (0.88501 versus 0.84575), together with a lower RMSE ( $1.968 \times 10^{-5}$  versus  $2.254 \times 10^{-5} \text{ mol m}^{-2}$ ), MAE ( $1.384 \times 10^{-5}$  versus  $1.641 \times 10^{-5} \text{ mol m}^{-2}$ ), and MAPE (21.68% versus 29.91%). Although CNN had a slightly lower RMSE than SVM ( $2.254 \times 10^{-5}$  versus  $2.265 \times 10^{-5} \text{ mol m}^{-2}$ ), SVM ranked second overall because it achieved a higher  $R^2$  and lower MAE and MAPE. These results indicate that RF explained a larger proportion of

NO<sub>2</sub> variability while maintaining the most stable overall error control on spatially held-out blocks. Therefore, RF was more suitable as the basis for subsequent spatiotemporal analysis and predictor interpretation.

From the perspective of model structure, the advantages of RF are closely related to its ensemble learning framework. RF constructs multiple decision trees through bootstrap resampling and random feature selection, and the final prediction is obtained by averaging the outputs of all trees. This structure enables RF to capture nonlinear relationships, threshold effects, and local response patterns between NO<sub>2</sub> column concentration and explanatory variables. In contrast, SVM relies on kernel functions to map predictors into a high-dimensional feature space and constructs a relatively global regression relationship. Although SVM performs well in nonlinear regression, its performance is sensitive to kernel selection and parameter settings. CNN extracts features through convolutional operations and has strong advantages in regular grid or image-like spatial data. However, in this study, the predictors were mainly organized as city-level annual statistical variables rather than continuous image patches or high-frequency spatial grids, which may limit the full use of CNN's spatial feature extraction capability.

RF also showed strong adaptability to the multi-source predictor system used in this study. The 18 predictors were derived from MODIS remote sensing products, ERA5-Land meteorological reanalysis data, and annual sectoral EDGAR v8.1 NO<sub>x</sub> emission inventories. These predictors differ in physical meaning, spatial resolution, statistical distribution, and units and may exhibit nonlinear interactions and multicollinearity [27]. RF does not require strict assumptions about the distribution of input variables and is relatively robust to outliers and collinearity. During tree splitting, RF can automatically identify important variable thresholds and high-order interactions, making it suitable for modeling the complex coupling effects of natural environmental conditions, meteorological regulation, and anthropogenic emissions on NO<sub>2</sub> column concentration [28]. By comparison, SVM is more dependent on data scaling and parameter optimization, while CNN requires a more structured spatial input to fully exploit its architectural advantages.

In terms of the requirements of remote sensing inversion, the purpose of this study was not only to obtain high prediction accuracy, but also to generate reliable city-level NO<sub>2</sub> reconstruction results for spatiotemporal pattern analysis and mechanism interpretation. The selected model therefore needed to preserve inter-city differences, represent pollution gradients, and provide a stable basis for subsequent SHAP-based driving factor analysis. RF is well suited to these requirements because its tree-based structure can reflect heterogeneous local relationships and nonlinear responses under different environmental and emission conditions. Although CNN and SVM also achieved good inversion accuracy, RF provided a more balanced combination of prediction accuracy, robustness, and interpretability [19].

It should be noted that the testing set MAPE of RF was 21.68%, which was relatively high compared with its R<sup>2</sup>, RMSE, and MAE values. This phenomenon is mainly related to the physical characteristics of NO<sub>2</sub> column concentration inversion and the intrinsic sensitivity of MAPE. First, MAPE uses the observed value as the denominator; therefore, when TROPOMI NO<sub>2</sub> column concentrations are low, even small absolute deviations can generate large relative errors, thereby increasing the overall MAPE. Second, the EDGAR NO<sub>x</sub> emission inventory represents surface NO<sub>x</sub> emission intensity, whereas TROPOMI NO<sub>2</sub> column concentration reflects an atmospheric state after emission, transport, chemical transformation, vertical mixing, and satellite retrieval processes. Therefore, there is no simple one-to-one correspondence between NO<sub>x</sub> emissions and satellite-observed NO<sub>2</sub> columns. In addition, city-scale inversion involves scale conversion and spatial matching among emission grids, meteorological reanalysis grids, satellite pixels, and ad-

ministrative boundaries, which further increases prediction uncertainty. NO<sub>2</sub> also has a relatively short atmospheric lifetime and strong spatial heterogeneity, and its concentration is jointly affected by boundary layer height, wind fields, radiation, cloud conditions, aerosols, and photochemical processes. Thus, variables such as NO<sub>x</sub> emissions, wind speed, precipitation, and temperature cannot fully explain all spatiotemporal variations in NO<sub>2</sub> column concentration.

Therefore, the relatively high MAPE does not indicate model failure, but rather reflects the combined effects of low-concentration sample sensitivity, the nonlinear emission–concentration relationship, and scale mismatch among multi-source datasets. Considering that RF achieved the highest pooled testing set R<sup>2</sup> and the lowest pooled testing set RMSE, MAE, and MAPE among all nine models, it retained the strongest overall out-of-block predictive performance and the most stable error control. For this reason, RF was selected as the optimal model for NO<sub>2</sub> column density inversion and was used as the basis for subsequent spatiotemporal distribution analysis and model interpretation.

#### 4.2. Interpretation of Predictors Associated with Reconstructed NO<sub>2</sub>

##### 4.2.1. Associations of Anthropogenic NO<sub>x</sub> Emission Predictors with RF Predictions

Anthropogenic NO<sub>x</sub> emission factors were one of the variable categories with relatively high contributions in the RF model, indicating that the RF model relied substantially on anthropogenic emission predictors when reconstructing the spatiotemporal variation in city-level NO<sub>2</sub> column density. NO<sub>x\_ind\_combustion</sub>, NO<sub>x\_road</sub>, NO<sub>x\_ind\_process</sub>, and NO<sub>x\_agriculture</sub> were the four most prominent anthropogenic emission variables. NO<sub>x\_power</sub> and NO<sub>x\_residential</sub> showed moderate importance, whereas NO<sub>x\_shipping\_nonroad</sub> had a relatively low overall contribution. The beeswarm further shows that these anthropogenic emission variables generally produced positive SHAP contributions at high values and negative contributions at low values.

NO<sub>x\_road</sub> represents NO<sub>x</sub> emissions from road transportation. In terms of the direction of association with the RF output, high NO<sub>x\_road</sub> values were mainly distributed in the positive SHAP region, whereas low values were mainly distributed in the negative SHAP region. This indicates that stronger road transport emissions generally increased the model-predicted NO<sub>2</sub> column density, whereas lower emissions generally reduced the model output. Motor vehicles are one of the important emission sources of urban NO<sub>2</sub>. Regions with denser transportation networks, larger vehicle fleets, and higher traffic volumes usually have stronger vehicle exhaust emissions, leading to higher NO<sub>2</sub> column densities [29,30].

NO<sub>x\_agriculture</sub> also showed a high contribution in the SHAP ranking. This variable mainly represents agriculture-related NO<sub>x</sub> emissions, including agricultural soil emissions and open burning of agricultural waste. In terms of the direction of influence, high NO<sub>x\_agriculture</sub> values were mainly associated with positive SHAP contributions, whereas low values were mainly associated with negative contributions, indicating that stronger agriculture-related emissions generally increased the predicted NO<sub>2</sub> column density. Agricultural activities can affect NO<sub>2</sub> column density through several pathways [31]. On the one hand, nitrogen fertilizer application can promote nitrification and denitrification processes in soils, thereby producing reactive nitrogen emissions. On the other hand, straw burning and agricultural waste combustion can directly release NO<sub>x</sub> into the atmosphere [32]. In addition, agricultural areas are often distributed around urban fringes and rural–urban transition zones, and their emissions may influence city-level NO<sub>2</sub> columns through regional transport and background pollution. The relatively high SHAP importance of NO<sub>x\_agriculture</sub> indicates that the RF model used this predictor when reconstructing

regional NO<sub>2</sub> patterns. This association does not by itself establish a causal contribution of agricultural emissions to the reconstructed urban values [33].

NO<sub>x</sub>\_ind\_combustion represents NO<sub>x</sub> emissions from manufacturing combustion. NO<sub>x</sub>\_ind\_combustion ranked first among all variables. Its high-value samples were mainly located in the positive SHAP region, whereas low-value samples formed a broad negative SHAP distribution. Thus, higher manufacturing combustion emissions generally increased the model prediction substantially, while lower emissions generally reduced it. A small number of low-value samples still occurred in the positive SHAP region, suggesting that the effect may be moderated by interactions with other variables. Industrial fuel combustion is an important source of NO<sub>x</sub>, especially in areas with intensive manufacturing activity, concentrated heavy industry, and high energy consumption. Cities or urban agglomerations with clustered industrial facilities usually have stronger combustion-related emissions, which may increase tropospheric NO<sub>2</sub> column density [34]. The high mean absolute SHAP value of NO<sub>x</sub>\_ind\_combustion indicates that manufacturing combustion emissions were an important predictor in the RF model, particularly for distinguishing cities with different reconstructed NO<sub>2</sub> levels. This model importance should not be interpreted as a causal effect estimate [35].

NO<sub>x</sub>\_residential mainly reflects emissions from household energy consumption, commercial activities, heating, and small-scale combustion processes, whereas NO<sub>x</sub>\_power represents NO<sub>x</sub> emissions from fossil fuel combustion during power generation and energy supply [36–38]. Both variables occupied moderate positions in the mean absolute SHAP ranking, with lower overall contributions than manufacturing combustion, road transportation, industrial processes, and agriculture-related emissions. In terms of direction, high values of both variables were mainly associated with positive SHAP contributions, whereas low values were mainly associated with negative contributions. Thus, higher residential energy and power sector emissions generally increased the model-predicted NO<sub>2</sub> column density. Their effects on urban NO<sub>2</sub> may be jointly regulated by population density, energy consumption structure, heating demand, the spatial location of power plants, emission control measures, and atmospheric transport conditions. Therefore, they acted as secondary contributors in the overall model [39].

NO<sub>x</sub>\_shipping\_nonroad mainly involves emissions from shipping, port activities, construction machinery, and other non-road mobile sources, while NO<sub>x</sub>\_ind\_process mainly includes emissions from refinery transformation, chemical industries, iron and steel production, non-ferrous metals, non-metallic minerals, and food and paper industries [40,41]. NO<sub>x</sub>\_shipping\_nonroad had a relatively low overall contribution, whereas NO<sub>x</sub>\_ind\_process ranked fourth among all variables and was an important anthropogenic emission factor. In terms of direction, both variables generally showed positive SHAP contributions at high values and negative contributions at low values. However, NO<sub>x</sub>\_shipping\_nonroad had a narrower SHAP range, whereas NO<sub>x</sub>\_ind\_process showed larger positive and negative effects and greater overall importance. These emission sources may be spatially localized and can be important in specific port cities, industrial cities, or resource-based cities [42].

#### 4.2.2. Environmental and Meteorological Predictors in the RF Model

Natural environmental and meteorological factors mainly played a regulatory role in the diffusion, transformation, and accumulation processes of NO<sub>2</sub>. Unlike anthropogenic NO<sub>x</sub> emission factors, these variables do not directly represent emission sources, but they can influence NO<sub>2</sub> column density by affecting atmospheric stability, photochemical reactions, humidity conditions, aerosol loading, surface thermal properties, and pollutant transport. The environmental and meteorological importance order was P, RH, AOD, WS,

T, DPT, LST, Ppt, NDVI, LAI, and SCE. P was the most important variable in this category; RH, AOD, and WS formed the second tier, followed by T and DPT. Ppt, NDVI, LAI, and SCE had relatively low overall contributions.

T represents air temperature. T ranked fifth among the environmental and meteorological variables rather than first. In terms of direction, high T values generally tended toward positive SHAP contributions, whereas low values tended toward negative contributions, indicating that higher temperatures increased the model-predicted NO<sub>2</sub> column density in most samples. Temperature can influence NO<sub>2</sub> through multiple atmospheric processes, including photochemical reaction rates, boundary layer development, vertical mixing, and the conversion between NO<sub>x</sub> and NO<sub>2</sub> [43]. Therefore, the SHAP importance of T shows that temperature provided useful predictive information to the RF model. The observed SHAP pattern is consistent with several plausible atmospheric mechanisms, but it does not quantify a causal temperature effect [44].

P represents atmospheric pressure. P ranked second among all variables and first among the environmental and meteorological variables. In terms of direction, high P values extended mainly toward positive SHAP values, whereas low P values were mainly associated with negative SHAP contributions. This indicates that higher surface pressure values were generally associated with higher RF-predicted NO<sub>2</sub> column density within the fitted model [45]. Under relatively stable high-pressure conditions, pollutants are more likely to accumulate near the surface and in urban areas, whereas stronger vertical motion and atmospheric disturbance can enhance pollutant dispersion. The high SHAP importance of P indicates that the RF model relied strongly on surface pressure when differentiating city-year samples. However, surface pressure may also encode elevation, regional geography, and persistent climatic differences among cities. Its SHAP importance should therefore not be interpreted as evidence that pressure itself causally regulates the observed NO<sub>2</sub> pattern [46].

DPT and RH both represent atmospheric moisture conditions. In terms of direction, high RH values were mainly associated with negative SHAP contributions, whereas low values were more often associated with positive contributions, indicating that higher relative humidity generally reduced the model-predicted NO<sub>2</sub> column density. DPT showed a non-monotonic relationship: intermediate values formed a distinct positive SHAP tail, whereas lower and higher values were more often close to zero or in the negative SHAP region. DPT therefore cannot be interpreted as having a simple unidirectional positive or negative effect. Unlike Ppt, which mainly reflects precipitation events, DPT and RH can more continuously characterize water vapor content and humidity conditions in the atmosphere. These variables have physically plausible relationships with NO<sub>2</sub> chemistry and atmospheric moisture conditions. However, the SHAP results only demonstrate conditional associations within the RF model and do not establish these chemical pathways as the causes of the reconstructed pattern [47]. Higher humidity may promote the conversion of NO<sub>2</sub> into nitrate or other secondary components, while also altering the optical and chemical properties of the atmosphere. The mean absolute SHAP values of both variables were higher than that of Ppt, indicating that continuous atmospheric moisture conditions provided stronger model-explanatory information than intermittent precipitation processes [48].

AOD represents aerosol optical depth and reflects atmospheric aerosol loading. AOD ranked third among the environmental and meteorological variables. Its high-value samples were mainly located in the positive SHAP region, whereas low-value samples were mainly located in the negative SHAP region, indicating that higher AOD values were generally associated with higher RF-predicted NO<sub>2</sub> column density. AOD is often closely associated with regional pollution processes. Although AOD does not directly represent

NO<sub>2</sub> emissions, it can indicate particulate pollution and atmospheric conditions that are favorable for pollutant accumulation [49]. In addition, aerosols may affect solar radiation, photolysis rates, and boundary layer processes, thereby indirectly regulating NO<sub>2</sub> column density. Its relatively high SHAP contribution suggests that aerosol-related atmospheric conditions provided important information for NO<sub>2</sub> reconstruction.

LST reflects surface thermal conditions and may be related to land surface heating, urban heat island effects, and boundary layer exchange, while WS directly affects the horizontal transport and dilution of pollutants [50]. WS ranked fourth among the environmental and meteorological variables and was clearly more important than LST. In terms of direction, high WS values were mainly associated with negative SHAP contributions, whereas low values were mainly associated with positive contributions, indicating higher wind-speed values were generally associated with lower RF predictions and weak winds generally increased it. High LST values were also predominantly associated with negative SHAP contributions, whereas low values were more often close to zero or slightly positive; however, the overall influence of LST was smaller than that of WS. Higher wind speed usually promotes pollutant dispersion, whereas weak wind conditions are more favorable for local accumulation [51].

Ppt represents precipitation; NDVI and LAI mainly reflect vegetation cover and leaf area conditions, which may affect NO<sub>2</sub> through dry deposition and land cover differences; and SCE reflects snow cover conditions that may influence surface reflectance and winter atmospheric processes. In terms of direction, Ppt showed a weak nonlinear relationship: some intermediate values produced small positive SHAP contributions, whereas high values were generally close to zero or slightly negative. High values of NDVI and LAI were more often associated with small negative SHAP contributions. SCE points were tightly concentrated around zero and did not show a stable, clear unidirectional effect. Because the SHAP ranges of these variables were small, their directional patterns should be interpreted cautiously. NDVI, Ppt, LAI, and SCE had relatively low mean absolute SHAP values and ranked near the bottom of the importance order, indicating that their overall contributions to annual city-level NO<sub>2</sub> reconstruction were weaker than those of the major meteorological and aerosol-related factors [52]. All directions described above represent conditional contributions after accounting for other variables and model interactions, rather than independent causal effects.

#### 4.3. Uncertainty Analysis

Although the RF reconstruction showed strong spatial validation performance and broad agreement with OMI and ground observations, several sources of uncertainty remain. First, reconstruction for 2000–2018 relies on model relationships learned during the 2019–2022 TROPOMI period and therefore involves temporal transfer. However, the predictors used in this study were not static: they incorporated annually varying meteorological, land surface, and sector-specific anthropogenic NO<sub>x</sub> emission information. In particular, the annual EDGAR inventory can, to a certain extent, represent long-term changes in both emission magnitude and source sector composition over the past two decades, thereby avoiding the unrealistic assumption of temporally fixed anthropogenic emissions. The predictor domain assessment showed substantial distributional overlap between the historical and training periods, while the OMI-based analysis indicated broadly consistent annual spatial associations between the major predictors and NO<sub>2</sub>. These results support the historical reconstruction, but they cannot demonstrate that the full nonlinear predictor–NO<sub>2</sub> relationship remained invariant throughout 2000–2018. Earlier estimates should therefore be interpreted as model-based reconstructions using dynamic annual predictors rather than direct observations.

Second, the validation results indicate scale- and region-dependent uncertainty. RF and OMI showed generally good agreement at regional and national scales, whereas city-level trend agreement was weaker, indicating that regional performance cannot be directly generalized to every individual city. Differences in native spatial resolution, annual city-scale aggregation, and the physical distinction between satellite column densities and surface concentrations further contribute to this uncertainty. In addition, ground-based validation was concentrated in China and Japan, leaving South Korea, North Korea, and Mongolia less directly constrained by surface observations. The reconstructed dataset is therefore more appropriate for long-term regional and national comparisons and broad city-scale spatial patterns than for precise interpretation of trend magnitudes or isolated anomalies in individual cities.

Third, uncertainty is also associated with the SHAP interpretation. The 10,000-replicate grouped-city bootstrap indicated high ranking stability for the fixed RF model, with a median Spearman rank correlation of 0.986 and recovery of the original top-five predictors in 96.97% of resamples. However, this procedure assessed the sensitivity of the existing SHAP explanations to resampling of test cities and did not retrain the RF model in each bootstrap replicate. It therefore does not quantify uncertainty arising from model refitting. Moreover, when correlated predictors contain overlapping information, SHAP contributions may be redistributed among them, and predictors with similar importance may change rank across samples. The SHAP results should therefore be interpreted primarily as robust overall patterns of model reliance and conditional association rather than as evidence for exact rank ordering or independent causal effects.

## 5. Conclusions

This study developed a continuous annual RF-based tropospheric NO<sub>2</sub> column density product for 436 cities in five East Asian countries from 2000 to 2022 by integrating TROPOMI observations with 18 annual environmental, meteorological, and sectoral NO<sub>x</sub> emission predictors. Model performance was evaluated using five-fold spatial block cross-validation. The 436 city polygons were assigned uniquely to 61 occupied, non-overlapping 5° × 5° spatial blocks, and predictions from the five held-out block sets were concatenated before the testing metrics were calculated. Among the nine models, RF provided the best pooled held-out-block performance, with  $R^2 = 0.88501$ ,  $RMSE = 1.968 \times 10^{-5} \text{ mol m}^{-2}$ ,  $MAE = 1.384 \times 10^{-5} \text{ mol m}^{-2}$ , and  $MAPE = 21.68\%$ . Leave-one-year-out validation showed stable performance across 2019–2022, while the predictor domain and OMI-based annual association assessments indicated broad distributional overlap and predominantly stable association directions. Together, these findings support, but do not prove, the temporal applicability of the reconstruction; the detected predictor-level shifts, gradual association changes, and lack of OMI observations before 2005 warrant continued caution when interpreting the earliest historical years.

The complete 2000–2022 series was generated using a uniform RF framework rather than by splicing historical estimates with satellite observations. During 2005–2022, RF and OMI exhibited strong annual spatial consistency ( $r = 0.850\text{--}0.945$ ) and strongly correlated normalized regional annual series ( $r = 0.877$ ). Ground validation demonstrated that performance varied by country and scale. RF better represented intercity differences in China than OMI, whereas OMI performed better for intercity differences in Japan. Conversely, RF reproduced the Japanese national annual trend very closely, while it underestimated the magnitude of the ground-observed national decline in China. The 54.8% agreement in city-level RF and OMI trend directions further indicates that regional reliability does not eliminate local uncertainty.

The annual series revealed distinct nonlinear trajectories across East Asia. The regional mean increased significantly during 2000–2011 (Theil–Sen slope =  $+0.095 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ), decreased significantly during 2011–2018 ( $-0.140 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ), and remained nonsignificantly negative during 2018–2022 ( $-0.060 \times 10^{-5} \text{ mol m}^{-2} \text{ yr}^{-1}$ ); consequently, the full-period monotonic trend was not significant. China increased significantly to a maximum in 2011 and then declined significantly. Japan and North Korea showed persistent significant decreases, and South Korea showed a significant full-period decline despite a nonsignificant increase during 2018–2022. Mongolia exhibited no significant full-period trend. These results demonstrate that the long-term evolution of urban NO<sub>2</sub> in East Asia cannot be adequately characterized using a single linear trend and should instead be interpreted according to country-specific and stage-specific changes.

SHAP analysis identified industrial combustion emissions, surface pressure, and road emissions as the three most influential predictors, followed by industrial process and agricultural emissions, relative humidity, aerosol optical depth, and wind speed. The beeswarm distribution further showed that predictor contributions were nonlinear and direction dependent. These SHAP values describe conditional contributions to RF predictions after accounting for other predictors and interactions; they should not be interpreted as independent causal effects. Overall, the dataset is suitable for evaluating broad long-term NO<sub>2</sub> patterns, national differences, and urban pollution distributions in East Asia. Its use for individual-city trend magnitude, short-term anomalies, or causal policy assessment should account for temporal transfer uncertainty, scale mismatch, uneven ground monitoring coverage, and the fact that the 2000–2018 uncertainty bounds are transferred from the 2019–2022 calibration period rather than directly validated historically.

Overall, the long-term annual city-scale NO<sub>2</sub> column density dataset developed in this study compensates, to some extent, for the limitation caused by the relatively short observation record of the high-precision TROPOMI satellite and provides a continuous data basis for long-term NO<sub>2</sub> variation analysis in urban areas of East Asia. This study not only demonstrates the applicability of machine learning methods for large-scale and long-term NO<sub>2</sub> column density reconstruction, but also characterizes the spatiotemporal evolution patterns of reconstructed urban NO<sub>2</sub> and identifies the predictors with the highest global importance in the RF model in East Asia. The dataset and technical framework can provide data support and methodological reference for historical reconstruction of satellite-based trace gases, regional air pollution assessment, identification of key pollution cities, and collaborative management of transboundary atmospheric pollution.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su18189349/s1>, Figure S1: Annual standardized mean differences relative to the 2019–2022 reference for the five leading predictors; Figure S2: Annual trajectories of the five leading predictors; Figure S3: Annual cross-city associations between the five leading predictors and OMI NO<sub>2</sub>; Figure S4: Sensitivity of standardized Theil–Sen slopes to the exclusion of 2020 and 2021; Figure S5: Polygon-based  $5^\circ \times 5^\circ$  spatial blocks used for five-fold spatial cross-validation; Table S1: Predictor temporal stability analysis for 2000–2022; Table S2: Five-fold spatial cross-validation statistics; Table S3: RF leave-one-year-out validation metrics; Supplementary Data S1: City-year NO<sub>2</sub> data and predictors for 2000–2022; Supplementary Data S2a: Spatial CV fold 1 training data; Supplementary Data S2b: Spatial CV fold 1 test data; Supplementary Data S2c: Spatial CV fold 2 training data; Supplementary Data S2d: Spatial CV fold 2 test data; Supplementary Data S2e: Spatial CV fold 3 training data; Supplementary Data S2f: Spatial CV fold 3 test data; Supplementary Data S2g: Spatial CV fold 4 training data; Supplementary Data S2h: Spatial CV fold 4 test data; Supplementary Data S2i: Spatial CV fold 5 training data; Supplementary Data S2j: Spatial CV fold 5 test data; Supplementary Code S1a: BP model; Supplementary Code S1b: CNN model; Supplementary Code S1c: ELM model; Supplementary Code S1d: LSTM model; Supplementary Code S1e: PLS model;

Supplementary Code S1f: RBF model; Supplementary Code S1g: RF model; Supplementary Code S1h: SVM model; Supplementary Code S1i: XGBoost model; Supplementary Note S1: Spatial validation design; Supplementary Note S2: Quality-control report for the spatial cross-validation data split.

**Author Contributions:** Conceptualization, J.Z. and Q.S.; methodology, J.Z. and Q.S.; validation, Q.S.; formal analysis, J.Z. and F.C.; data curation, J.F. and Y.X.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z., Q.S. and H.Y.; visualization, J.Z. and F.C.; supervision, Y.X.; project administration, H.Y.; funding acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the China National Postdoctoral Program for Innovative Talents [BX20250085], and National Natural Science Foundation of China [42501130].

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data generated and used in the main analyses of this study, together with the associated reproducibility materials and modeling code, are publicly available in Mendeley Data: Yang, Heming (2026), “TROPOMI-referenced Reconstruction and Model Interpretation of Long-Term City-Scale NO<sub>2</sub> Column Density in East Asia Using Machine Learning”, Version 2, DOI: 10.17632/7b74kkkwhz.2. All Supplementary Materials cited throughout the manuscript, including extended validation, sensitivity analyses, uncertainty analyses, and related supplementary results, are also publicly available with the study materials.

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

## References

1. Xu, J.; Lindqvist, H.; Liu, Q.; Wang, K.; Wang, L. Estimating the spatial and temporal variability of the ground-level NO<sub>2</sub> concentration in China during 2005–2019 based on satellite remote sensing. *Atmos. Pollut. Res.* **2021**, *12*, 57–67. [\[CrossRef\]](#)
2. Ojeda-Castillo, V.; Murillo-Tovar, M.A.; Hernández Mena, L.; Saldarriaga-Noreña, H.; Vargas-Amado, M.E.; Herrera-López, E.J.; Díaz, J. Tropospheric NO<sub>2</sub>: Anthropogenic Influence, Global Trends, Satellite Data, and Machine Learning Application. *Remote Sens.* **2025**, *17*, 49. [\[CrossRef\]](#)
3. Park, J.; Hong, H.; Lee, H.; Kim, S.-W.; Kim, J.; Van Roozendaal, M.; Fayt, C.; Ahn, M.-H.; Jacob, D.; Seo, S.; et al. Tropospheric nitrogen dioxide levels vary diurnally in Asian cities. *Commun. Earth Environ.* **2025**, *6*, 389. [\[CrossRef\]](#)
4. He, S.; Dong, H.; Zhang, Z.; Yuan, Y. An Ensemble Model-Based Estimation of Nitrogen Dioxide in a Southeastern Coastal Region of China. *Remote Sens.* **2022**, *14*, 2807. [\[CrossRef\]](#)
5. Li, M.; Wu, Y.; Bao, Y.; Liu, B.; Petropoulos, G.P. Near-Surface NO<sub>2</sub> Concentration Estimation by Random Forest Modeling and Sentinel-5P and Ancillary Data. *Remote Sens.* **2022**, *14*, 3612. [\[CrossRef\]](#)
6. Musollari, S.; Pseftogkas, A.; Koukouli, M.-E.; Manders, A.; Segers, A.; Garane, K.; Balis, D. The Spatiotemporal Variability of Ozone and Nitrogen Dioxide in the Po Valley Using In Situ Measurements and Model Simulations. *Remote Sens.* **2025**, *17*, 1794. [\[CrossRef\]](#)
7. He, Y.; Uno, I.; Wang, Z.; Ohara, T.; Sugimoto, N.; Shimizu, A.; Richter, A.; Burrows, J.P. Variations of the increasing trend of tropospheric NO<sub>2</sub> over central east China during the past decade. *Atmos. Environ.* **2007**, *41*, 4865–4876. [\[CrossRef\]](#)
8. Uno, I.; He, Y.; Ohara, T.; Yamaji, K.; Kurokawa, J.-I.; Katayama, M.; Wang, Z.; Noguchi, K.; Hayashida, S.; Richter, A.; et al. Systematic Analysis of Interannual and Seasonal Variations of Model simulated Tropospheric NO<sub>2</sub> in Asia and comparison with GOME-satellite data. *Atmos. Chem. Phys.* **2007**, *7*, 1671–1681. [\[CrossRef\]](#)
9. Kong, L.; Tang, X.; Zhu, J.; Wang, Z.; Fu, J.S.; Wang, X.; Itahashi, S.; Yamaji, K.; Nagashima, T.; Lee, H.-J.; et al. Evaluation and uncertainty investigation of the NO<sub>2</sub>, CO and NH<sub>3</sub> modeling over China under the framework of MICS-Asia III. *Atmos. Chem. Phys.* **2020**, *20*, 181–202. [\[CrossRef\]](#)
10. Zhang, Y.; Li, Z.; Chen, Y.; de Leeuw, G.; Zhang, C.; Xie, Y.; Li, K. Improved inversion of aerosol components in the atmospheric column from remote sensing data. *Atmos. Chem. Phys.* **2020**, *20*, 12795–12811. [\[CrossRef\]](#)
11. Hu, K.; Feng, X.; Zhang, Q.; Shao, P.; Liu, Z.; Xu, Y.; Wang, S.; Wang, Y.; Wang, H.; Di, L.; et al. Review of Satellite Remote Sensing of Carbon Dioxide Inversion and Assimilation. *Remote Sens.* **2024**, *16*, 3394. [\[CrossRef\]](#)
12. Rabiei-Dastjerdi, H.; Mohammadi, S.; Saber, M.; Amini, S.; McArdle, G. Spatiotemporal Analysis of NO<sub>2</sub> Production Using TROPOMI Time-Series Images and Google Earth Engine in a Middle Eastern Country. *Remote Sens.* **2022**, *14*, 1725. [\[CrossRef\]](#)

13. Cooper, M.J.; Martin, R.V.; McLinden, C.A.; Brook, J.R. Inferring ground-level nitrogen dioxide concentrations at fine spatial resolution applied to the TROPOMI satellite instrument. *Environ. Res. Lett.* **2020**, *15*, 104013. [\[CrossRef\]](#)
14. Tawabini, B. Using machine learning algorithms to study the relationship between meteorological conditions and air quality parameters. *Sci. Rep.* **2026**, *16*, 10392. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Kuhn, L.; Beirle, S.; Osipov, S.; Pozzer, A.; Wagner, T. NitroNet—A machine learning model for the prediction of tropospheric NO<sub>2</sub> profiles from TROPOMI observations. *Atmos. Meas. Tech.* **2024**, *17*, 6485–6516. [\[CrossRef\]](#)
16. Li, Z.; Dong, H.; He, S.; Huang, H. Quantitative reconstruction of long-term spatiotemporal patterns of high-resolution ground-level NO<sub>2</sub> concentrations in mainland China using fusion techniques and a machine learning framework. *Environ. Int.* **2025**, *202*, 109672. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Wei, C.; Zhao, C.; Hu, Y.; Tian, Y. Predicting the Concentration Levels of PM<sub>2.5</sub> and O<sub>3</sub> for Highly Urbanized Areas Based on Machine Learning Models. *Sustainability* **2025**, *17*, 9211. [\[CrossRef\]](#)
18. Ravindiran, G.; Hayder, G.; Kanagarathinam, K.; Alagumalai, A.; Sonne, C. Air quality prediction by machine learning models: A predictive study on the indian coastal city of Visakhapatnam. *Chemosphere* **2023**, *338*, 139518. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Abdelmalek, M.M.; Mahmoud, H.; Shokry, H. Prognosis of air quality index and air pollution using machine learning techniques. *Sci. Rep.* **2025**, *15*, 25890. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Hu, M.; Chen, Y.; Yuan, D.; Yu, R.; Lu, X.; Fung, J.C.H.; Chen, W.; Huang, Y.; Lau, A.K. Estimation and spatiotemporal analysis of NO<sub>2</sub> pollution in East Asia during 2001–2016. *J. Geophys. Res. Atmos.* **2020**, *127*, e2021JD035129. [\[CrossRef\]](#)
21. Li, T.; Wang, Y.; Yuan, Q. Remote Sensing Estimation of Regional NO<sub>2</sub> via Space-Time Neural Networks. *Remote Sens.* **2020**, *12*, 2514. [\[CrossRef\]](#)
22. Gong, P.; Li, X.; Wang, J.; Bai, Y.; Chen, B.; Hu, T.; Liu, X.; Xu, B.; Yang, J.; Zhang, W.; et al. Annual Maps of Global Artificial Impervious Area (GAIA) between 1985 and 2018. *Remote Sens. Environ.* **2020**, *236*, 111510. [\[CrossRef\]](#)
23. Hagenauer, J.; Omrani, H.; Helbich, M. Assessing the performance of 38 machine learning models: The case of land consumption rates in Bavaria, Germany. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 1399–1419. [\[CrossRef\]](#)
24. Radjabaycolle, J.E.T.; Wattimena, E.M.C.; Pattiradjawane, V.E. Improving Air Quality Forecasts with LSTM and SHAP Explainability: A Case Study in Jakarta. *J. Embed. Syst. Secur. Intell. Syst.* **2025**, *6*, 337–347. [\[CrossRef\]](#)
25. Dhongade, S.; Tiwaskar, S.; Koturwar, A.; Motiramani, R.; Zarekar, D. Air Predictive Modeling for Air Quality: A Comparative Study of Machine Learning and Deep Learning Techniques. In Proceedings of the 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 24–25 February 2024; pp. 1–7. [\[CrossRef\]](#)
26. Cui, Y.; Zhang, W.; Bao, H.; Wang, C.; Cai, W.; Yu, J.; Streets, D.G. Streets, Spatiotemporal dynamics of nitrogen dioxide pollution and urban development: Satellite observations over China, 2005–2016. *Resour. Conserv. Recycl.* **2019**, *142*, 59–68. [\[CrossRef\]](#)
27. Fahim, M.A.S.; Visockienė, J.S. An Assessment of the Multi-Input Spatiotemporal RF–XGBoost Hybrid Framework for PM<sub>10</sub> Estimation in Lithuania. *Sustainability* **2026**, *18*, 2022. [\[CrossRef\]](#)
28. Faye, D.; Lguensat, R.; Kaly, F.; Sudmant, A.; Gaye, A.T.; Kalisa, E. Machine Learning for Air Quality Forecasting: Insights from Five Provinces of Rwanda. *Sci. Afr.* **2025**, *30*, e02959. [\[CrossRef\]](#)
29. Kerr, G.H.; Meyer, M.; Goldberg, D.L.; Miller, J.; Anenberg, S.C. Air Pollution Impacts from Warehousing in the United States Uncovered with Satellite Data. *Nat. Commun.* **2024**, *15*, 6006. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Goldberg, D.L.; Tao, M.; Kerr, G.H.; Ma, S.; Tong, D.Q.; Fiore, A.M.; Dickens, A.F.; Adelman, Z.E.; Anenberg, S.C. Evaluating the Spatial Patterns of U.S. Urban NO<sub>x</sub> Emissions Using TROPOMI NO<sub>2</sub>. *Remote Sens. Environ.* **2024**, *300*, 113917. [\[CrossRef\]](#)
31. Lin, X.; van der, A.R.; de Laat, J.; Huijnen, V.; Mijling, B.; Ding, J.; Eskes, H.; Douros, J.; Liu, M.; Zhang, X.; et al. European Soil NO<sub>x</sub> Emissions Derived from Satellite NO<sub>2</sub> Observations. *J. Geophys. Res. Atmos.* **2024**, *129*, e2024JD041492. [\[CrossRef\]](#)
32. Sha, T.; Ma, X.; Jia, H.; Tian, R.; Chang, Y.; Cao, F.; Zhang, Y. Impacts of Soil NO<sub>x</sub> Emission on O<sub>3</sub> Air Quality in Rural California. *Environ. Sci. Technol.* **2021**, *55*, 7113–7122. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Pan, S.Y.; He, K.H.; Lin, K.T.; Fan, C.; Chang, C.-T. Addressing nitrogenous gases from croplands toward low-emission agriculture. *npj Clim. Atmos. Sci.* **2022**, *5*, 43. [\[CrossRef\]](#)
34. Liu, F.; Beirle, S.; Joiner, J.; Choi, S.; Tao, Z.; Knowland, K.E.; Smith, S.J.; Tong, D.Q.; Ma, S.; Fasnacht, Z.T.; et al. High-Resolution Mapping of Nitrogen Oxide Emissions in Large US Cities from TROPOMI Retrievals of Tropospheric Nitrogen Dioxide Columns. *Atmos. Chem. Phys.* **2024**, *24*, 3717–3728. [\[CrossRef\]](#)
35. Badia, A.; Segura Barrero, R.; Ventura, S.; Guevara, M.; Peñuelas, J.; Villaba, G. Effect of land use changes on air quality: impacts of urbanization, urban vegetation, and agriculture. *npj Urban Sustain.* **2025**, *5*, 113. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Kuhlmann, G.; Koene, E.F.M.; Schooling, C.N.; Palmer, P.I.; López, Ò.C.; Guevara, M. Temporal Variability of NO<sub>x</sub> Emissions from Power Plants: A Comparison of Satellite- and Inventory-Based Estimates. *Atmos. Chem. Phys.* **2026**, *26*, 4405–4421. [\[CrossRef\]](#)
37. Varon, D.J.; Jervis, D.; Pandey, S.; Gallardo, S.L.; Balasus, N.; Yang, L.H.; Jacob, D.J. Quantifying NO<sub>x</sub> Point Sources with Landsat and Sentinel-2 Satellite Observations of NO<sub>2</sub> Plumes. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2317077121. [\[CrossRef\]](#) [\[PubMed\]](#)

38. Kashtan, Y.; Nicholson, M.; Finnegan, C.; Ouyang, Z.; Lebel, E.D.; Michanowicz, D.R.; Shonkoff, S.B.C.; Jackson, R.B. Nitrogen Dioxide Exposure, Health Outcomes, and Associated Demographic Disparities Due to Gas and Propane Combustion by U.S. Stoves. *Sci. Adv.* **2024**, *10*, eadm8680. [[CrossRef](#)] [[PubMed](#)]
39. Liu, Z.; Li, Y.; Law, A.; Tan, J.Y.K.; Chua, W.H.; Zhu, Y.; Feng, C.-C.; Luo, W. Association between NO<sub>2</sub> and human mobility: a two-year spatiotemporal study during the COVID-19 pandemic in Southeast Asia. *Ann. GIS* **2024**, *30*, 475–492. [[CrossRef](#)]
40. Luo, Z.; He, T.; Yi, W.; Zhao, J.; Zhang, Z.; Wang, Y.; Liu, H.; He, K. Advancing Shipping NO<sub>x</sub> Pollution Estimation through a Satellite-Based Approach. *PNAS Nexus* **2024**, *3*, pgad430. [[CrossRef](#)] [[PubMed](#)]
41. Latsch, M.; Richter, A.; Burrows, J.P.; Bösch, H. Improved Detection of Global NO<sub>2</sub> Signals from Shipping in Sentinel-5P TROPOMI Data. *Atmos. Meas. Tech.* **2025**, *18*, 4373–4395. [[CrossRef](#)]
42. Chen, X.; Feng, J.X. Health effects of built environment based on a comparison of walkability and air pollution: A case study of Nanjing City. *Prog. Geogr.* **2019**, *38*, 296–304. [[CrossRef](#)]
43. Ahmad, N.; Lin, C.; Lau, A.K.H.; Kim, J.; Zhang, T.; Yu, F.; Li, C.; Li, Y.; Fung, J.C.H.; Lao, X.Q. Estimation of Ground-Level NO<sub>2</sub> and Its Spatiotemporal Variations in China Using GEMS Measurements and a Nested Machine Learning Model. *Atmos. Chem. Phys.* **2024**, *24*, 9645–9665. [[CrossRef](#)]
44. Bai, X.; Wang, Y.; Gui, L.; Tao, M.; Zeng, M. Comparing the Influences on NO<sub>2</sub> Changes in Terms of Inter-Annual and Seasonal Variations in Different Regions of China: Meteorological and Anthropogenic Contributions. *Remote Sens.* **2025**, *17*, 121. [[CrossRef](#)]
45. Hegglin, M.I.; Yuan, Y. Impact of Weather Patterns and Meteorological Factors on PM<sub>2.5</sub> and O<sub>3</sub> Pollution in China: Insights from Multi-Year Observations. *Atmos. Chem. Phys.* **2024**, *24*, 6539–6564. [[CrossRef](#)]
46. Liu, X.; Yi, G.; Zhou, X.; Zhang, T.; Lan, Y.; Yu, D.; Wen, B.; Hu, J. Atmospheric NO<sub>2</sub> Distribution Characteristics and Influencing Factors in Yangtze River Economic Belt: Analysis of the NO<sub>2</sub> Product of TROPOMI/Sentinel-5P. *Atmosphere* **2021**, *12*, 1142. [[CrossRef](#)]
47. Tan, S.P.; Piri, M. Modeling the Solubility of Nitrogen Dioxide in Water Using Perturbed-Chain Statistical Associating Fluid Theory. *Ind. Eng. Chem. Res.* **2013**, *52*, 16032–16043. [[CrossRef](#)]
48. Yavuz, V. Variations in Air Pollutant Concentrations on Dry and Wet Days with Varying Precipitation Intensity. *Atmosphere* **2024**, *15*, 896. [[CrossRef](#)]
49. Masoom, A.; Kazadzis, S.; Valeri, M.; Raptis, I.-P.; Brizzi, G.; Papachristopoulou, K.; Barnaba, F.; Casadio, S.; Kreuter, A.; Niro, F. Assessment of the Impact of NO<sub>2</sub> Contribution on Aerosol-Optical-Depth Measurements at Several Sites Worldwide. *Atmos. Meas. Tech.* **2024**, *17*, 5525–5545. [[CrossRef](#)]
50. Sharma, M.; Kumar, P.; Choudhary, M.P.; Mathur, A.K. Integrated geospatial and statistical analysis of urban air pollutants (PM<sub>10</sub>, SO<sub>2</sub> and NO<sub>2</sub>) and their relationships with land surface temperature over a semi-arid environment. *Environ. Monit. Assess.* **2025**, *197*, 981. [[CrossRef](#)] [[PubMed](#)]
51. Li, D.; Liu, M.; Han, H.; Wang, J. Nonlinear Impacts of Air Pollutants and Meteorological Factors on PM<sub>2.5</sub>: An Interpretable GT-iFormer Model with SHAP Analysis. *Atmosphere* **2026**, *17*, 266. [[CrossRef](#)]
52. Dai, A.; Liu, C.; Ji, Y.; Sheng, Q.; Zhu, Z. Effect of Different Plant Communities on NO<sub>2</sub> in an Urban Road Greenbelt in Nanjing, China. *Sci. Rep.* **2023**, *13*, 3353. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.