remote sensing

MDPI

Article

Machine Learning-Based Estimation of Surface NO_2 Concentrations over China: A Comparative Analysis of Geostationary (GEMS) and Polar-Orbiting (TROPOMI) Satellite Data

Yijin Ma [1], Yi Wang [1,*], Jun Wang [2,3], Minghui Tao [1], Jhoon Kim [4], Chenyang Wu [5] and Shanshan Zhang [1]

[1] Hubei Key Laboratory of Regional Ecology and Environmental Change, School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China; 1202310756@cug.edu.cn (Y.M.); taomh@cug.edu.cn (M.T.); 1202310793@cug.edu.cn (S.Z.)
[2] Department of Chemical and Biochemical Engineering, The University of Iowa, Iowa City, IA 52242, USA; jun-wang-1@uiowa.edu
[3] Center for Global and Regional Environmental Research, The University of Iowa, Iowa City, IA 52242, USA
[4] Department of Atmospheric Sciences, Yonsei University, Seoul 03722, Republic of Korea; jkim2@yonsei.ac.kr
[5] School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China; m202373036@hust.edu.cn
* Correspondence: wangyi34@cug.edu.cn; Tel.: +86-027-67883727

Highlights

What are the main findings?

- The CatBoost model performed best, with GEMS data yielding higher accuracy (R^2 = 0.842) than TROPOMI data (R^2 = 0.765).
- GEMS's high temporal resolution provided a much larger training dataset, which was the key factor for its superior model performance.

What are the implications of the main findings?

- Geostationary satellite data (like GEMS) offers a critical advantage for high-resolution air quality monitoring via machine learning due to its frequent sampling.
- GEMS enables the reconstruction of detailed diurnal pollution patterns and near-real-time tracking of emission events, providing valuable insights for dynamic air quality management.

Abstract

High-accuracy spatiotemporal monitoring of surface nitrogen dioxide (NO_2) concentrations is essential for air quality management. This study evaluates machine learning-based estimates of near-surface NO_2 concentrations using data from the geostationary GEMS instrument and the polar-orbiting TROPOMI over China in 2022. Four tree-based models—Random Forest, XGBoost, CatBoost, and LightGBM—were trained by integrating satellite vertical-column densities with multi-source meteorological and ancillary data. Results show that CatBoost achieved the highest accuracy, with an R^2 of 0.842 for GEMS and 0.765 for TROPOMI, alongside the lowest RMSE and MAE. Models trained on GEMS data consistently outperformed TROPOMI-based models across all metrics. This advantage is primarily attributed to the substantially larger training sample size enabled by GEMS's high temporal resolution, as confirmed through a controlled experiment with consistent sample sizes which isolated the effect of data volume. Spatially, GEMS

Academic Editors: Qian Liu and Manzhu Yu

Received: 15 January 2026
Revised: 7 February 2026
Accepted: 13 February 2026
Published: 15 February 2026

estimates captured sharper concentration gradients and localized emission hotspots, while TROPOMI produced smoother fields. Temporally, only GEMS allowed the reconstruction of detailed diurnal patterns and near-real-time pollution episode tracking. This study confirms the significant added value of geostationary satellite data for high-frequency air quality monitoring and analysis when combined with machine learning.

Keywords: TROPOMI; GEMS; machine learning; surface NO_2 concentrations

1. Introduction

Nitrogen dioxide (NO_2) is an irritating brownish-red trace gas [1] that poses significant threat to human health and the environment. Short-term exposure can cause lung irritation, edema, and impaired immunity [2], while long-term exposure increases risks of respiratory [3], cardiovascular, and neurological disorders [4,5]; macular degeneration [6]; and adverse fetal development [7]. As a key precursor, NO_2 participates in chemical reactions in the atmosphere to form ozone and nitrate aerosols [8], with O_3 causing ~84,000 annual premature deaths in China and nitrate being included in $PM_{2.5}$'s ~2.71 million annual premature deaths [9,10]. Notably, surface NO_2 exhibits strong spatiotemporal variability due to complex anthropogenic (power generation, industry, transport, and residential) and natural (soil, wildfires, and lightening) emission sources, coupled with a short lifetime of ~2 to ~5 h in summer [11–13] and ~12 to ~24 h in winter [14]. Thus, high-spatiotemporal-resolution monitoring of surface NO_2 concentrations is crucial for public health protection and atmospheric environmental regulation.

China National Environmental Monitoring Centre (CNEMC) stations can accurately monitor surface NO_2 concentrations on an hourly basis, but these stations are sparse and unevenly distributed, with more in the East and fewer in the west, and are mainly concentrated in urban or suburban areas, leaving a significant portion of China unmonitored. In contrast to the in situ observations, satellite tropospheric NO_2 monitoring has the advantages of low cost, wide coverage and good continuity, hence effectively overcoming the shortcomings of the ground-based monitoring network [15]. The development of satellite remote sensing technology has made it possible to monitor spatial and temporal variations in NO_2. While early sensors like the Global Ozone Monitoring Experiment (GOME) (global coverage: ~3 days; resolution: 320 × 40 km) and the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) (6 days; 60 × 30 km) provided foundational data, subsequent instruments such as the Ozone Monitoring Instrument (OMI) (1 day; 13 × 14 km) and GOME-2 (1–3 days; 40 × 80 km) progressively enhanced resolution and revisit time. Currently, the TROPOspheric Monitoring Instrument (TROPOMI) sensor aboard the polar satellite Sentinel-5 Precursor achieves a revolutionary combination of daily global coverage and 3.5 × 5.5 km spatial resolution. This unprecedented capability makes it the premier tool for retrieving surface NO_2 concentrations at scales relevant to urban and point-source emissions [16].

As satellite remote sensing retrieves the tropospheric NO_2 vertical-column density (VCD) rather than surface NO_2 concentrations, methods for converting the vertical loadings to surface concentrations have undergone a transition from the atmospheric chemical transport model (CTM)-based method [17–21] to the empirical statistical method [22–25] to the machine learning method [26,27]. The CTM-based method simulates tropospheric NO_2 vertical profiles to derive a scaling factor between model-simulated near-surface concentrations and column concentrations, which is then combined with satellite data for conversion [28]. However, it suffers from high computational costs, reliance on sparse ground validation data, and sensitivity to outdated emission inventories. Empirical

statistical methods estimate near-surface NO_2 concentrations via linear regression models linking measured values to factors like meteorology and land use [29], but fail to capture complex NO_2 generation/diffusion mechanisms or multivariate dependencies. With the advancement of Artificial Intelligence (AI), machine learning has become the dominant approach due to its robust nonlinear-relationship handling and high prediction accuracy [30]. Up to now, many researchers have combined TROPOMI tropospheric NO_2 VCD retrievals with a variety of machine learning models to estimate surface NO_2 concentrations, such as Random Forest (RF) [31], XGBoost [32], deep convolutional neural network (DeepCNN) [33], spatiotemporal neural network (SNN) [34], coupled deep learning model [30], multi-task transformer model [35], etc., all of which have demonstrated varying degrees of excellent performance, and these machine learning models are generally superior to all empirical methods developed to date [36]. In order to select a high-performance model, the researchers compared different models. Kang et al. used TROPOMI NO_2 data to compare four types of machine learning, namely, Support Vector Regression (SVR), Random Forests (RF), Extreme Gradient Boosting (XGB), and Lightweight Gradient Boosting Machine (LGBM), with Multivariate Linear Regression (MLR) as the underlying statistical method [37]. Researchers have compared the effectiveness of TROPOMI with other polar-orbiting satellite products for estimating surface NO_2 concentrations, e.g., with the GOME-2 L2 NO_2 and OMI standard product (OMNO2d) [38], and with the POMINO-TROPOMI NO_2 [39]. Data from different sources have also been fused and combined to leverage multiple strengths for estimation, such as fusing OMI and GOME-2 data using different transit times [40].

Despite advances in estimating surface NO_2 concentrations from polar-orbiting satellite tropospheric NO_2 VCDs, their low temporal resolution prevents capturing diurnal variations. This gap is addressed by the geostationary GEO-KOMPSAT-2B (GK-2B) satellite mission that carries the Geostationary Environment Monitoring Spectrometer (GEMS) sensor to observe and deliver daytime hourly data across East Asia. Ideal for tracking fast-changing atmospheric conditions, it enables hourly near-surface NO_2 retrieval. Current research on estimating surface NO_2 using GEMS tropospheric NO_2 VCD retrievals focuses on model improvement and optimization. Ahmad et al. introduced the nested machine learning Model Mixing Height (NMH) as a key parameter to improve near-surface NO_2 concentration estimation accuracy by predicting NMH [41]. Advanced algorithms include CatBoost [42] and the multi-output Random Forest (MORF) model [43], both of which are widely used for surface pollutant gas estimation, capturing complex nonlinear relationships and spatiotemporal variability with high R^2 values and low Root Mean Square Errors (RMSE). High spatiotemporal resolution enables refined surface NO_2 spatiotemporal analysis, allowing scholars to capture sudden pollution events and the diurnal variation processes [43,44]. This facilitates a comprehensive understanding of pollutant spatiotemporal distribution, emission source characteristics, and NO_2 concentration variation patterns across different periods. Additionally, uncertainty quantification is emphasized for robust regional and conditional NO_2 reliability assessments, typically via pixel-wise concentration distribution prediction and generalization ability evaluation using spatial cross-validation [45].

While machine learning has become a dominant and extensively applied approach for estimating surface NO_2 from both polar-orbiting and geostationary satellite NO_2 VCDs, a critical and underexplored question remains regarding the comparative performance of polar-orbiting versus geostationary satellite observing systems within the same machine learning framework. Specifically, this includes two key questions: (1) Compared with TROPOMI, what advantages does the larger data volume of GEMS (a geostationary satellite) provide for training machine learning models to estimate near-surface NO_2 concentrations? (2) What differences exist in retrieved near-surface NO_2 concentrations

between the two satellite datasets when using co-located sampling points? To address these gaps, this study conducts a systematic comparative analysis using an identical machine learning framework. We train four machine learning models—Random Forest (RF) and three mainstream gradient boosting models (XGBoost, CatBoost, LightGBM (LGBM))—to estimate surface NO_2 concentrations over China in 2022. We compare the performance of the four models to select the optimal one, then further evaluate and compare the accuracy and capability of the two satellite datasets in estimating surface NO_2 concentrations. This study aims to investigate: (1) performance differences between models trained on the two datasets; (2) whether GEMS's high temporal resolution enhances estimation accuracy; and (3) the synergistic effects of optimal models and data sources. Section 2 outlines satellite, meteorological, ground-based monitoring and auxiliary data, along with their preprocessing and feature engineering. Section 3 presents the comparative results on model performance and retrieved spatiotemporal patterns. Section 4 discusses the implications and broader context of the findings, and Section 5 summarizes the key conclusions and future research directions.

2. Materials and Methods

2.1. Data

We selected satellite data (TROPOMI [Airbus Defence and Space Netherlands, Leiden, The Netherlands] and GEMS [Korean Aerospace Research Institute (KARI), Daejeon, South Korea] NO_2 tropospheric VCDs), meteorological data (surface pressure, total precipitation, 10 m U wind component, 10 m V wind component, 2 m temperature, boundary layer height, relative humidity at 1000 hPa pressure), ground monitoring station NO_2 data (CNEMC hourly observations) and auxiliary data (Normalized Difference Vegetation Index (NDVI), population density, Digital Elevation Model (DEM), road kernel density) in 2022. Satellite data were quality-controlled, auxiliary data were processed according to Table S1, and all datasets were resampled to a 0.25° × 0.25° spatial resolution using bilinear interpolation.

2.1.1. TROPOMI NO_2

The TROPOMI sensor on board the Copernicus Sentinel-5P satellite was launched by the European Space Agency in 2017 and is designed to monitor atmospheric composition at high spatial resolution. Its initial spatial resolution of 7 km × 3.5 km was improved to 5.5 km × 3.5 km after August 2019 [46]. TROPOMI has a swath width of approximately 2600 km, hence providing global observational coverage with a full Earth scan each day with an equatorial crossing time of 13:30. TROPOMI observes in the spectral range between 270 and 775 nm and between 2305 and 2385 nm, including channels critical for detecting NO_2.

In this study, we utilize tropospheric NO_2 vertical-column density from NASA's TROPOMI Level-2 products for the year 2022 across China. The retrieval algorithm employs a modified Differential Optical Absorption Spectroscopy (DOAS) method to derive total NO_2 Slant Column Densities (SCDs) using solar-backscatter radiation in the spectral range between 405 and 465 nm. These total SCDs are then corrected by subtracting stratospheric NO_2 contributions, yielding tropospheric NO_2 SCDs. Finally, the tropospheric SCDs are converted into vertical-column densities (VCDs) using an Air Mass Factor (AMF). The AMF calculation incorporates scattering weights influenced by surface albedo and aerosol parameters, along with assumptions on NO_2 vertical profiles [47]. Compared to Pandora ground-based NO_2 tropospheric VCDs, TROPOMI NO_2 tropospheric VCDs show a slight underestimation, with a Pearson correlation coefficient of r = 0.89 and

slopes of 0.79 [48]. These products are filtered to ensure data quality (quality assurance (QA) value > 0.75 and effective cloud area fraction < 0.3).

2.1.2. GEMS NO_2

The GEMS sensor on board the geostationary satellite GEO-KOMPSAT-2B was launched by the National Institute for Environmental Studies in 2020. It is dedicated to hourly monitoring of daytime air quality in East Asia, boasting a swath coverage of 5000 km that spans 75°E–145°E and 5°S–45°N. Equipped with 1000 spectral channels across the 300–500 nm wavelength range, the GEMS provides a spatial resolution of 7 km × 8 km via its specialized Step-and-Stare scanning technique [49]. The significant advantage of GEMS is the ability to collect data six to ten times in the daytime, hence providing a more complete picture of the intraday variation in NO_2 in China.

In this study, we utilize the GEMS Level-2 products (GK2_GEMS_L2_NO2) for the year 2022 across China. The GEMS NO_2 retrieval algorithm utilizes calibrated radiance and irradiance data from GEMS L1C to calculate the SCD of NO_2 using DOAS in the 435 nm to 450 nm band. Subsequently, the SCD is converted into VCD by employing scattering weight LookUp Tables (LUTs) and shape factors derived from the GEOS-Chem chemical transport model. Finally, a Stratosphere–Troposphere Separation (STS) technique is applied to distinguish between stratospheric and tropospheric NO_2, providing tropospheric NO_2 VCDs. The GEMS L2 and Pandora ground-based tropospheric NO_2 VCDs have a Pearson correlation coefficient of $r = 0.76$, with a slope of 1.28, implying that the GEMS L2 VCDs have an overestimation [48]. These products are filtered to ensure data quality ($0 \leq$ VCDs $< 10^{20}$ molecules·m^{-2}, algorithm quality flag ≤ 112, final algorithm ≤ 1, cloud fraction < 0.3) [50].

2.1.3. Meteorological Data

Meteorological fields significantly modulate surface NO_2 concentrations through multiple pathways, including vertical mixing, horizontal transport, chemical reaction, and deposition [51]. Vertical mixing intensity, governed primarily by boundary layer height (BLH), regulates the dilution capacity of pollutants within the lower troposphere [52]. Horizontal transport dynamics, driven by wind speed and direction, modulate the advection and dispersion of emission plumes [53]. Temperature and relative humidity collectively influence chemical reaction rates, affecting photochemical NO_2 production and secondary aerosol formation. Precipitation scavenging provides a critical removal pathway, particularly efficient for particulate matter through wet deposition mechanisms [54]. In this study, meteorological data were obtained from ECMWF's ERA5 hourly reanalysis dataset for 2022 over China, with the 0.25° × 0.25° spatial resolution and hourly temporal resolution [55].

2.1.4. Ground Monitoring Station NO_2 Data

To ensure accurate and reliable NO_2 monitoring, the CNEMC established a nationwide high-resolution ground monitoring network and regularly releases corresponding datasets. The CNEMC NO_2 dataset primarily utilizes Chemiluminescence Detection (CLD) to generate surface-level hourly measurements, capturing spatiotemporal distribution characteristics of regional NO_2 pollution, diurnal variation patterns, regional transport influences, and pollution emission intensities. This dataset thereby provides high-resolution observational support for air quality assessment, pollution source tracking, and atmospheric chemistry research [56]. In this study, we obtained hourly NO_2 data from CNEMC ground monitoring stations. The network comprised 2024 stations until 13 February 2022, expanding to 2026 stations thereafter (Figure S1). All available station data for their respective operational periods were used. Within each 0.25° grid cell, data were temporally

averaged from stations reporting at the same UTC hour. These gridded averages were then spatially co-located with resampled satellite data using a 15 min temporal matching window.

2.1.5. Auxiliary Data

The Normalized Difference Vegetation Index (NDVI), as a proxy for vegetation coverage and physiological activity, indirectly modulates surface NO_2 concentrations through two primary pathways: enhanced dry deposition velocity via stomatal uptake by plant canopies, which directly removes NO_2 from the atmosphere [27]; and alteration of boundary layer dynamics through changes in surface roughness and evapotranspiration, thereby influencing pollutant dispersion and transport patterns [57]. In this study, we utilized MOD13A3 Version 6 product, which provides monthly composite NDVI at a 1 km spatial resolution. This dataset is derived from measurements by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument [Raytheon Santa Barbara Remote Sensing, Goleta, CA, USA] aboard the Terra satellite. The data were obtained from the EarthData dataset official website [58].

Population density, defined as midyear population divided by land area in square kilometers, can directly or indirectly affect NO_2 concentrations by influencing factors such as transport, industry, energy use, and the urban heat island effect. Population density (pop) is obtained from LandScan, with a spatial resolution of 1 km × 1 km, and a yearly temporal resolution [59].

Topography can influence factors such as atmospheric dispersion, pollutant transport and local meteorological conditions. In this study, DEM data are from the National Tibetan Plateau Data Center with a spatial resolution of 30 m [60].

Transport is an important anthropogenic source of NO_x emissions. Higher road grades mean higher traffic volumes and more NO_x emissions. Road data was acquired from OpenStreetMap [61], and processed by weighted kernel density analysis (see Supplementary Note S1 for a description of the analysis methodology) based on road classes, gridded to a spatial resolution of 0.25° × 0.25° and a yearly temporal resolution.

2.2. Machine Learning Models

Machine learning has shown excellent performance in dealing with multifactor nonlinear relationships [62]. Random Forest (RF) is the basic integrated learning model, while XGBoost, CatBoost, and LightGBM (LGBM) are the mainstream gradient-boosted decision tree (GBDT) models. Each of the four models has its own unique strengths to cope with different types of data challenges. In this study, we will conduct a comparative exploration of these four models applied to the estimation of surface NO_2 concentration in China, and provide an optimal machine learning solution for NO_2 pollution inversion through systematic evaluation.

2.2.1. Model Introduction

Random Forest is an integrated learning model that improves the accuracy and stability of prediction by constructing multiple decision trees and summarizing the results. In the regression task, the average of the predicted values of all the sub-decision trees is taken as the output, and this way of constructing endows RF with a powerful noise processing capability. Even if there is noise interference in the data, the combined results of multiple sub-trees can eliminate the noise influence to a certain extent. Moreover, Random Forest is robust to missing feature values in prediction data. Since each tree uses a random subset of features, predictions can often be made effectively using only the trees that did not rely on the missing data, complemented by the ensemble's averaging mechanism. In addition, RF can also provide feature importance scores to help identify key influencing

factors [63]. However, the process of constructing a large number of sub-decision trees is extremely computationally intensive, and when facing large-scale datasets, the training time of RF will increase dramatically, and its efficiency will be significantly reduced.

XGBoost is an optimized version of Gradient Boosting Decision Tree (GBDT). Its core improvement lies in the processing of the loss function. It introduces a regular term on the basis of the traditional loss function, which effectively prevents model overfitting and enhances the generalization ability of the model by constraining the model complexity. At the same time, XGBoost supports parallel computing, which can make full use of the computational resources of multi-core processors and greatly improve the training speed [64]. XGBoost excels in efficient optimization and flexibility, and is especially good at dealing with complex datasets. However, the training process of XGBoost is relatively complex, and parameter tuning is tedious and time-consuming. Moreover, when dealing with large-scale datasets, XGBoost needs to store a large number of intermediate computation results, resulting in high memory consumption.

CatBoost is an efficient machine learning algorithm based on Gradient Boosting Decision Trees (GBDT), with the core features of native support for categorical features without preprocessing and avoiding the data leakage problem of traditional mean coding through ordered target encoding. It uses Ordered Boosting to reduce overfitting through sample alignment and rolling gradient computation, while using a symmetric, balanced tree structure to accelerate prediction and enhance generalization [65]. CatBoost has significant advantages in handling classification prediction tasks and provides automatic hyperparameter-tuning tools to improve practicability and ease of use. However, its algorithm design and the automatic parameter-tuning process have complexity, and the training speed of CatBoost is slow, which may be limited in scenarios with high requirements on training time.

LightGBM is a histogram-based splitting method that simplifies the traditional decision tree-splitting process. It discretizes the continuous feature values into different buckets of the histogram, calculates the statistical information of each bucket, and only needs to consider the boundaries of the buckets when splitting the nodes, which greatly reduces the amount of computation. At the same time, LightGBM reduces the amount of data needed to compute the gradient by techniques such as Gradient One-Sided Sampling (GOSS), optimizing memory usage and training time [66]. This results in a significant increase in training speed and memory efficiency, but on extremely class-imbalanced datasets, its histogram-based splitting tends to favor the majority class, which may result in the minority class samples being ignored and inferior prediction performance to other models.

2.2.2. Hyperparameter Optimization

In order to balance model optimality and computational cost, this study adopts a differentiated optimization strategy for the hyperparameter spatial characteristics of different models. For the Random Forest (RF) and XGBoost models, due to the high degree of parameter discretization, the RandomisedSearchCV is used to perform a 5-fold cross-validated stochastic search, which obtains a global sub-optimal solution with limited computational resources [67]. For CatBoost and LightGBM (LGBM) models with continuous dominant parameters, the Optuna framework is used to achieve Bayesian optimization, and the Tree-structured Parzen Estimator (TPE) algorithm is used to dynamically adjust the search direction according to the historical information to efficiently explore the high-dimensional space [68–70] (Note S2 in the Supplement for details). The specific hyperparameters optimized for each model (e.g., n_estimators, max_depth, learning_rate) along with their search ranges are provided in Table S3.

2.2.3. Model Performance Evaluation Metrics

In order to comprehensively and accurately measure the performance of the employed machine learning models (RF, XGBoost, CatBoost, and LGBM) in predicting surface NO_2 concentrations, we selected four key metrics for evaluation: the coefficient of determination (R^2), the Nash–Sutcliffe efficiency coefficient (NSE), the Mean Absolute Error (MAE), and the RMSE. Note S2 provides a detailed description of the metrics and their calculation methods. Through the comprehensive analysis of these indicators, we are able to comprehensively and objectively judge the performance of different models in the task of surface NO_2 concentration prediction, which provides a strong basis for model screening and optimization.

2.2.4. Feature Importance

Quantifying the extent to which each feature contributes to model predictions can reveal the relative importance of different factors in NO_2 concentration predictions and is an effective way to gain a deeper understanding of the factors that influence surface NO_2 concentrations. In this study, feature importance was calculated using model-specific intrinsic metrics with standardized postprocessing to enable cross-model comparisons. For the RF, feature importance was derived from Gini impurity reduction [63]. For XGBoost and LGBM, gain-based importance was adopted, which measures the total information gain from splits involving each feature during tree construction [64]. For CatBoost, the "PredictionValuesChange" metric was used, reflecting the average absolute change in model predictions when a specific feature is excluded [65]. To ensure comparability across models, raw importance scores (with model-specific scales) were normalized to a percentage scale (summing to 100) after extraction [71]. Finally, the normalized feature importance rankings were visualized to facilitate interpretation across all models.

2.3. Flowchart

The estimation of surface NO_2 concentrations using machine learning methods can be divided into 3 main parts, which are data preprocessing (collection and integration), model training and validation, and comparison and mapping (Figure 1). In the data preprocessing section, we collected data, initially processed the data, resampled, matched the data by time and location at the same grid point (Figure S2), and deleted samples that were missing values, obtaining samples of 89,932 and 949,212 for the TROPOMI-based dataset ($Group_T$) and GEMS-based dataset ($Group_G$), respectively (the descriptive feature statistics are shown in Tables S5 and S6). NO_2 concentration measurements from ground stations were used as the target, and Tropospheric column concentrations of TROPOMI/GEMS NO_2 (NO2_TROPOMI/NO2_GEMS), month (mon), day of a month (day), hour of a day (hour), 2 m temperature (t2m), surface pressure (sp), 10 m U wind component (u10), 10 m V wind component (v10), relative humidity (rh), total precipitation (tp), boundary layer height (blh), population (pop), DEM, NDVI and road were used as features. In the model training and validation section, both datasets were randomly split into a training set (80% of samples) and a test set (20% of samples), with the ground station data as targets and other data as features input into four models (Random Forest (RF), XGBoost, CatBoost, and LightGBM (LGBM)), to optimize each model, and then compare the training effect by R^2 and RMSE derived from cross-check validation. In the estimation section, the trained models are used to estimate the surface NO_2 concentration in China in 2022, comparing the estimation effects of the four models and two satellite datasets.

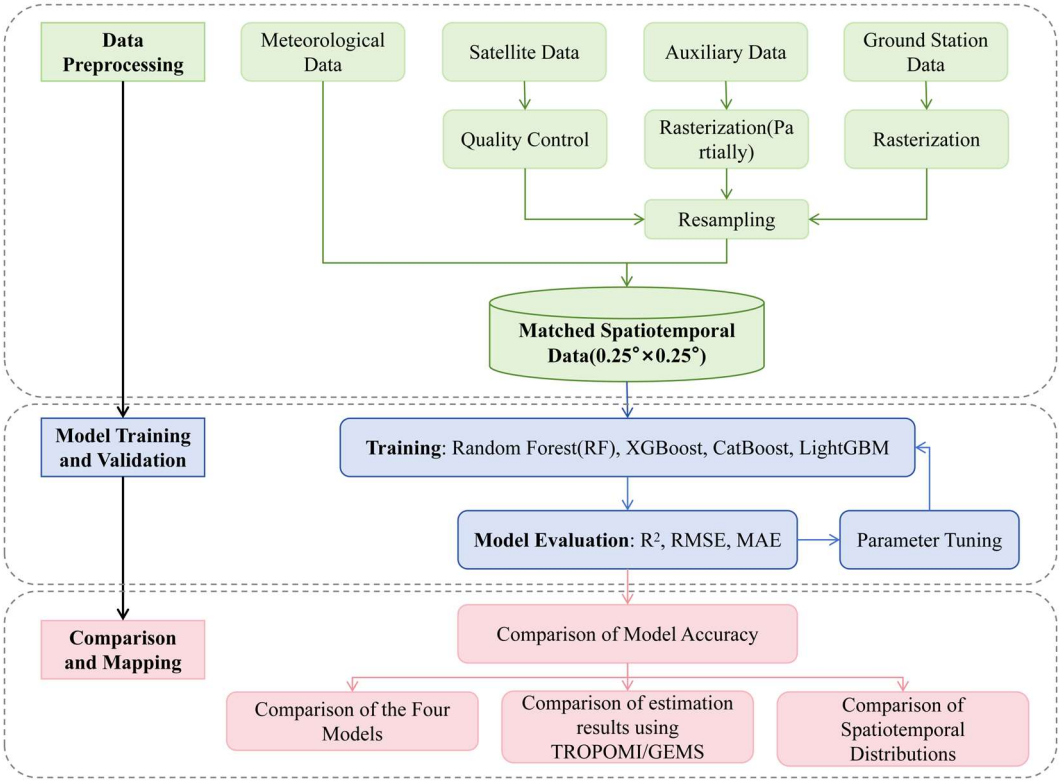

Figure 1. Flowchart of the methodology, which consists of data preprocessing, model training and validation, and comparison and mapping. Arrows indicate the workflow direction, and dashed boxes represent key processing steps.

2.4. Comparative Experimental Design with Identical Sampling Points

The experiments were designed to investigate the causes of the difference in model training results between the Group$_T$ and Group$_G$: using the intersection of the Group$_T$ and Group$_G$ datasets, i.e., selecting samples with the same time and location as the training set of the CatBoost model, and controlling the sample size consistently in order to investigate the effect of different satellite NO_2 data on the model results. The ALL$_{TROPOMI}$ and ALL$_{GEMS}$ datasets' sample sizes were 89,932 and 949,212, respectively, and taking the intersection resulted in 73,908 samples for this supplemental experiment (Table 1). In the experiment, the training set and the test set are 80% and 20% of the dataset (the random seed is uniformly set to 42), respectively.

Table 1. Experimental design for investigating the causes of the difference in model training results between the Group$_T$ and Group$_G$.

Name	Usage Data	Sample Number
ALL$_{TROPOMI}$	All matching TROPOMI dataset samples	89,932
ALL$_{GEMS}$	All matching GEMS dataset samples	949,212
INT$_{TROPOMI}$	Matching TROPOMI dataset samples (intersection of ALL$_{TROPOMI}$ and ALL$_{GEMS}$ dataset sample points)	73,908
INT$_{GEMS}$	Matching GEMS dataset samples (intersection of ALL$_{TROPOMI}$ and ALL$_{GEMS}$ dataset sample points)	73,908

3. Results

3.1. Model Evaluation and Analysis

3.1.1. Model Evaluation of Four Models with Original Datasets

Figures 2 and 3 illustrate the evaluation of surface NO_2 estimation and corresponding feature importance of the four models for TROPOMI and GEMS, respectively. Analysis of training set performance revealed that all models achieved exceptionally high accuracy on their respective training data. For instance, the CatBoost model attained an R^2 of 0.998 and 0.997 on the TROPOMI and GEMS training sets, respectively, with correspondingly low RMSE values. Similar near-perfect fits were observed for RF, XGBoost, and LGBM on the training data (Figures 2a,d,g,j and 3a,d,g,j). This indicates that all models possess sufficient capacity to learn the complex relationships within the training data.

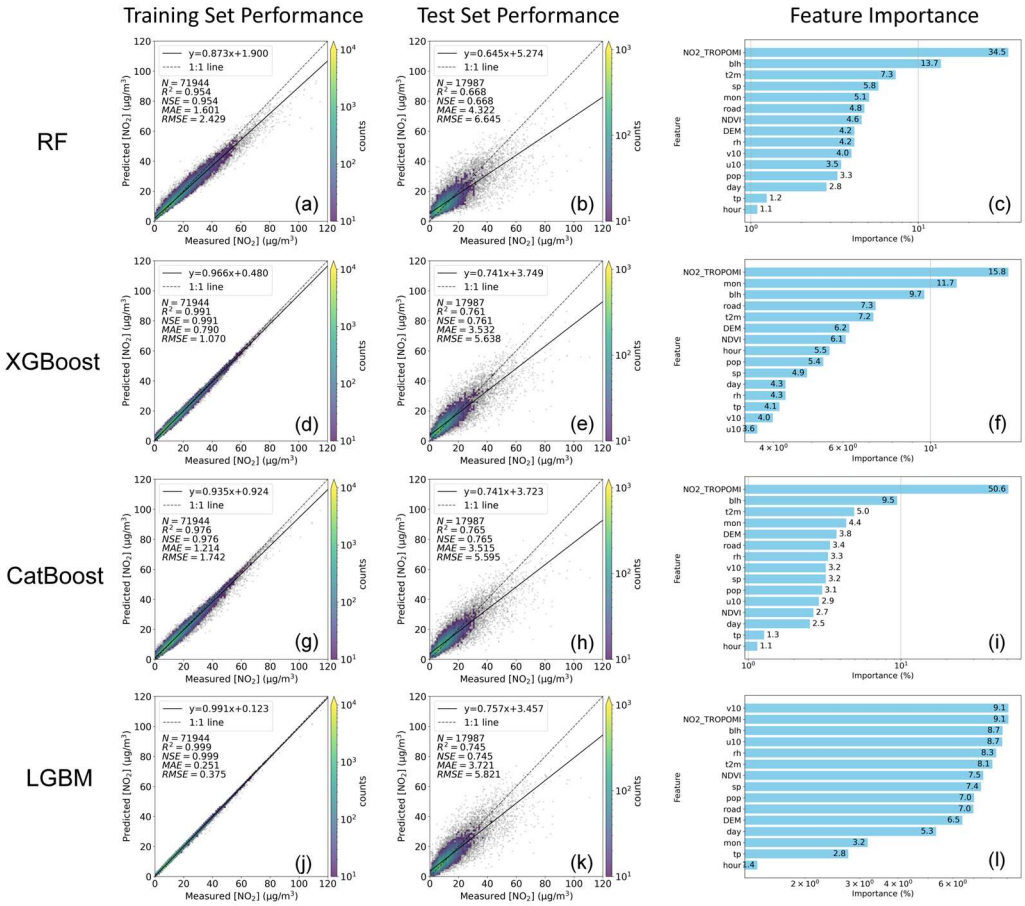

Figure 2. Evaluation results of the four models with Group$_T$ data used. From left to right: Training set performance, test set performance and feature importance (the meaning of the variables can be referred to Table S1). From top to bottom: RF, XGBoost, CatBoost and LGBM. Each subfigure is described as follows: (**a**) RF training set performance; (**b**) RF test set performance; (**c**) RF feature importance; (**d**) XGBoost training set performance; (**e**) XGBoost test set performance; (**f**) XGBoost feature importance; (**g**) CatBoost training set performance; (**h**) CatBoost test set performance; (**i**) CatBoost feature importance; (**j**) LGBM training set performance; (**k**) LGBM test set performance; (**l**) LGBM feature importance.

The critical comparison, however, lies in the generalization to unseen data. The RF model has the lowest R^2 for measured and predicted NO_2 in both the TROPOMI and GEMS groups ($R^2_{TROPOMI}$ = 0.668, R^2_{GEMS} = 0.723), whereas the CatBoost model has the highest R^2 among all models ($R^2_{TROPOMI}$ = 0.765, R^2_{GEMS} = 0.842), with the XGBoost model being the next highest, followed by the LGBM model (Figures 2b,e,h,k and 3b,e,h,k). In terms of

RMSE, the RF model exhibits the highest values ($RMSE_{TROPOMI}$ = 6.645 μg·m^{-3}, $RMSE_{GEMS}$ = 6.044 μg·m^{-3}), followed by the LGBM model ($RMSE_{TROPOMI}$ = 5.860 μg·m^{-3}, $RMSE_{GEMS}$ = 4.994 μg·m^{-3}). The XGBoost model achieves a lower RMSE than LGBM ($RMSE_{TROPOMI}$ = 5.638 μg·m^{-3}, $RMSE_{GEMS}$ = 4.878 μg·m^{-3}), and the CatBoost model yields the lowest RMSE ($RMSE_{TROPOMI}$ = 5.595, $RMSE_{GEMS}$ = 4.567). For MAE, the performance order is consistent with RMSE. The RF model has the highest MAE ($MAE_{TROPOMI}$ = 4.322 μg·m^{-3}, MAE_{GEMS} = 3.868 μg·m^{-3}), followed by the LGBM model ($MAE_{TROPOMI}$ = 3.761 μg·m^{-3}, MAE_{GEMS} = 3.276 μg·m^{-3}). The XGBoost model shows a lower MAE than LGBM ($MAE_{TROPOMI}$ = 3.532 μg·m^{-3}, MAE_{GEMS} = 3.166 μg·m^{-3}), and the CatBoost model has the lowest MAE ($MAE_{TROPOMI}$ = 3.515 μg·m^{-3}, MAE_{GEMS} = 2.890 μg·m^{-3}) (Figures 2b,e,h,k and 3b,e,h,k). Comparison of the feature importance (FI) of the four models reveals that each model has a different focus during the training process, with the CatBoost model having the highest feature importance for the satellite NO_2 data (NO2_TROPOMI/NO2_GEMS), which reaches 50.6% in the $Group_T$ and 32.2% in the $Group_G$, followed by the RF model ($FI_{TROPOMI}$ = 34.5%, FI_{GEMS} = 27.3%) and XGBoost model ($FI_{TROPOMI}$ = 15.8%, FI_{GEMS} = 13.7%), with the LGBM model ($FI_{TROPOMI}$ = 9.1%, FI_{GEMS} = 9.1%) ranking last (Figures 2c,f,i,l and 3c,f,i,l). Therefore, among the four models, the CatBoost model makes more efficient use of core input characteristics during the training process and has the highest estimation accuracy and outstanding performance in all aspects.

According to the test set accuracy (R^2, RMSE, and MAE), training with GEMS NO_2 satellite data results in a better-performing model for the test set compared to using TROPOMI NO_2 satellite data. The direct comparison between $Group_G$ and $Group_T$ across all models and metrics is clearly shown in Figures 2 and 3. In terms of test set R^2, $Group_G$ outperforms $Group_T$ by 0.055 (RF), 0.058 (XGBoost), 0.077 (CatBoost), and 0.069 (LGBM), respectively. In terms of test set RMSE, $Group_G$ outperforms $Group_T$ by 0.601 μg·m^{-3} (RF), 0.760 μg·m^{-3} (XGBoost), 1.028 μg·m^{-3} (CatBoost) and 0.866 μg·m^{-3} (LGBM), respectively. In terms of test set MAE, $Group_G$ outperforms $Group_T$ by 0.454 μg·m^{-3} (RF), 0.366 μg·m^{-3} (XGBoost), 0.625 μg·m^{-3} (CatBoost), 0.485 μg·m^{-3} (LGBM), respectively. The performance advantage of the Catboost model with GEMS data is particularly prominent, showing the strongest contrast.

3.1.2. Model Evaluation for Comparative Experiments with Identical Sampling Points

Figure 4 shows that the $INT_{TROPOMI}$ performs slightly better than the INT_{GEMS} in the test set when the sample size control is the same ($R^2_{TROPOMI}$ = 0.757, R^2_{GEMS} = 0.747; $RMSE_{TROPOMI}$ = 5.141 μg·m^{-3}, $RMSE_{GEMS}$ = 5.239 μg·m^{-3}; $MAE_{TROPOMI}$ = 3.440 μg·m^{-3}, MAE_{GEMS} = 3.499 μg·m^{-3}). Compared to the test set results obtained from the original datasets (Figures 2h and 3h), the accuracy of the test set for this experiment is reduced in both cases, with the R^2 decreasing by 0.008 for the TROPOMI group and 0.095 for the GEMS group. It was found that the test set of $INT_{TROPOMI}$ outperformed INT_{GEMS} in this experiment, contrary to the results of the previous experiment. This may be attributed to the higher spatial resolution of TROPOMI NO_2 compared to GEMS NO_2, and the advantage of the high spatial resolution of TROPOMI satellites is demonstrated when the sample sizes of the two groups are controlled in the same way. Meanwhile, we find that the accuracy of the GEMS group decreases more than that of the TROPOMI group, which is most likely due to the significant reduction in the sample size. In other words, due to the high temporal resolution of GEMS, a large number of samples can be increased to improve the effectiveness of the model.

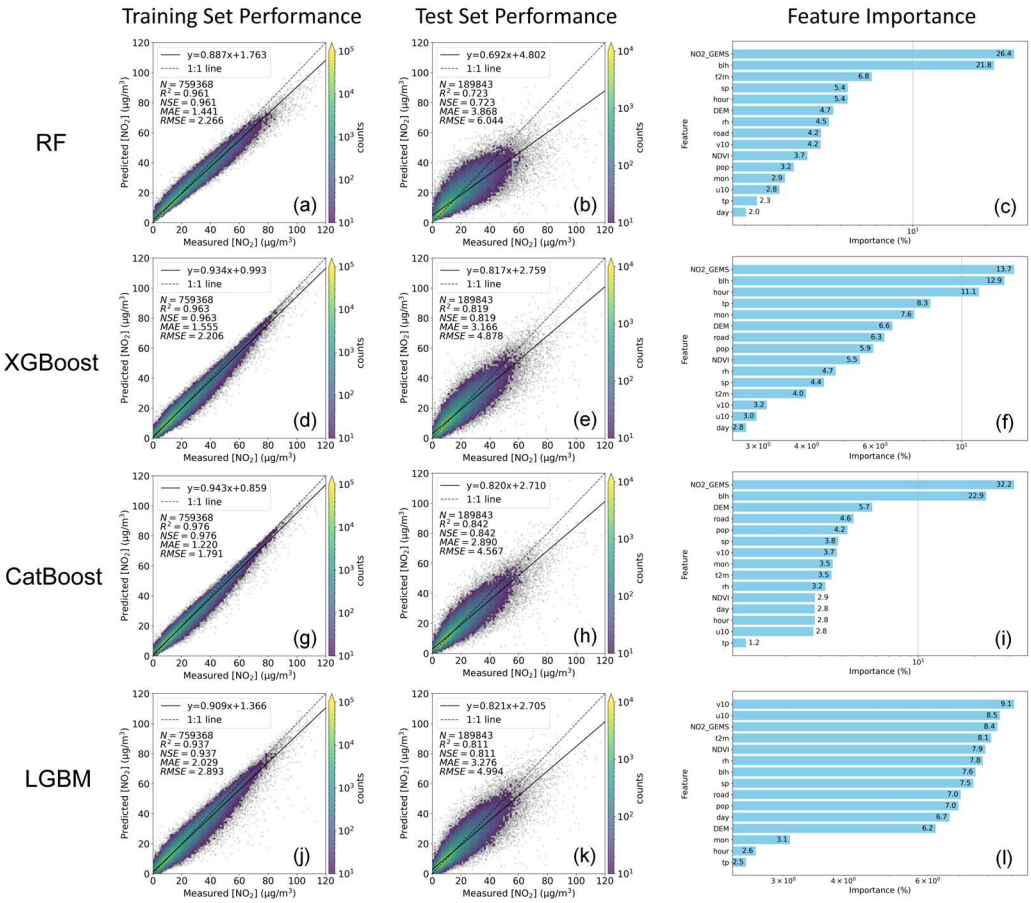

Figure 3. Evaluation results of the four models with GroupG data used. From left to right: Training set performance, test set performance and feature importance (the meaning of the variables can be referred to Table S1). From top to bottom: RF, XGBoost, CatBoost and LGBM. Each subfigure is described as follows: (**a**) RF training set performance; (**b**) RF test set perfor-mance; (**c**) RF feature importance; (**d**) XGBoost training set performance; (**e**) XGBoost test set performance; (**f**) XGBoost feature importance; (**g**) CatBoost training set performance; (**h**) CatBoost test set performance; (**i**) CatBoost feature importance; (**j**) LGBM training set performance; (**k**) LGBM test set performance; (**l**) LGBM feature importance.

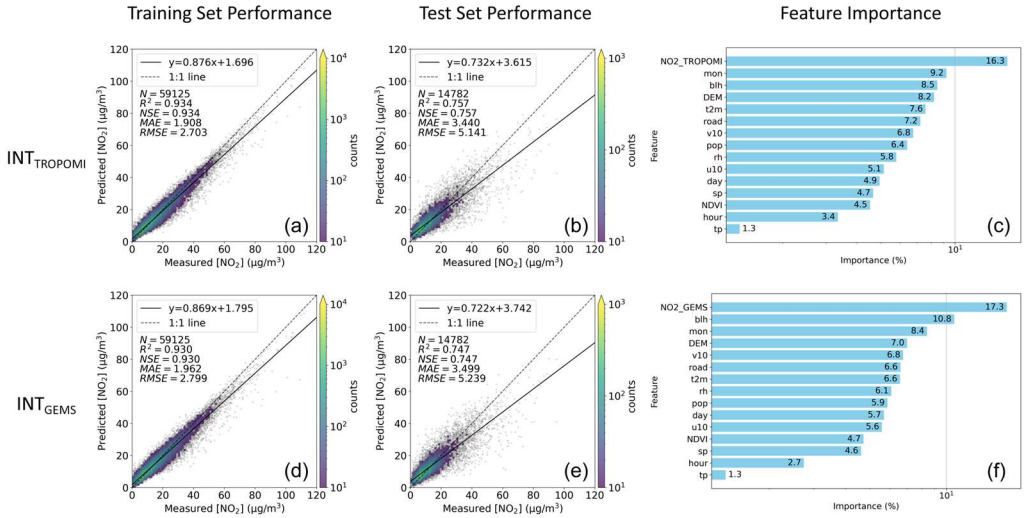

Figure 4. Evaluation results of the CatBoost models with INTTROPOMI and INTGEMS data used. From left to right: Training set performance, test set performance and feature importance (the meaning of the variables can be referred to Table S1). From top to bottom: INTTROPOMI and INTGEMS. Each

subfigure is described as follows: (**a**) $INT_{TROPOMI}$ training set performance; (**b**) $INT_{TROPOMI}$ test set performance; (**c**) $INT_{TROPOMI}$ feature importance; (**d**) INT_{GEMS} training set performance; (**e**) INT_{GEMS} test set performance; (**f**) INT_{GEMS} feature importance.

3.1.3. Analysis of Performance Discrepancy and Dominant Factors

The significant performance gap between models trained on GEMS and TROPOMI data (Section 3.1.1), along with the reversed outcome under controlled sample sizes (Section 3.1.2), stems from two key factors.

First, though TROPOMI NO_2 VCDs have a higher Pearson correlation with ground data (0.65 vs. 0.52 for GEMS; see Supplementary Figure S3 for scatterplots), the CatBoost model for $ALL_{TROPOMI}$ shows a smaller slope (0.741) and larger intercept (3.723 $\mu g \cdot m^{-3}$), leading to more severe spatial gradient compression and greater accuracy loss. In contrast, the GEMS-based model has a steeper slope and smaller intercept, making this bias milder and less impactful on estimation results.

Second, the larger sample size provided by GEMS is fundamentally more beneficial for model training. This is evidenced by a stronger correlation between per-site sample count and model performance (R^2) for GEMS (r = 0.38) than for TROPOMI (r = 0.23), as shown in Figure S4. The key advantage lies in data representativeness. GEMS's 949,212 samples (vs. TROPOMI's 89,932), collected at multiple daytime hours, inherently capture a wider spectrum of diurnal meteorological conditions (e.g., varying boundary layer height, temperature, solar radiation) and pollution states (from morning peaks to afternoon lows). This diversity during training enables tree-based models like CatBoost to learn more robust and generalizable rules for predicting NO_2 under various scenarios, thereby improving stability and accuracy on unseen data. In contrast, TROPOMI's single daily overpass results in a smaller and less temporally diverse dataset. This not only provides fewer examples for the model to learn from but also leads to uneven spatial sampling (Figure S4a). Consequently, the model trained on TROPOMI data may be less adept at predicting concentrations during underrepresented times or for extreme pollution events that are not well-captured in its limited sample.

The advantage of GEMS's massive sample size (driven by high temporal resolution) outweighs TROPOMI's superior single-sample quality. While TROPOMI has a slight edge under equal sample sizes, GEMS's large, diverse dataset better meets the needs of tree-based models like CatBoost, making temporal-resolution-induced sample volume more critical for surface NO_2 estimation.

3.2. Retrieval Results

3.2.1. Spatial Variation Comparison

Figure 5 shows spatial distributions of surface NO_2 concentrations over China during four seasons, which are retrieved via the CatBoost model based on TROPOMI and GEMS NO_2 VCDs at co-located sampling points, along with a direct comparison of the two datasets for each season. The results reveal that the two products share consistent seasonal cycles—concentrations are highest in winter and lowest in summer in both datasets (Figure 5a,b,e,f,i,j,m,n). The key distinctions lie in spatial gradients and regional details. Estimates based on TROPOMI (TROPOMI Intersect) generally show smoother and more extensive areas of moderate-to-high concentrations, particularly in eastern China (Figure 5a,e,i,m). In contrast, estimates based on GEMS (GEMS Intersect) often exhibit sharper gradients, with high concentrations more concentrated around major urban agglomerations and industrial zones, while showing lower background values in surrounding and remote areas (e.g., Northwest China) (Figure 5b,f,j,n). This indicates that estimates based on GEMS may exhibit sharper concentration gradients, with high values being more concentrated around major emission sources. In contrast, estimates based on TROPOMI

present smoother spatial fields with more regionally extended areas of moderate concentrations. Additionally, comparing GEMS Intersect (concentrations at TROPOMI overpass time) with GEMS Daytime (full daytime average at the same sites) indicates that full daytime average concentrations are lower across all seasons, while the basic spatial distribution pattern remains unchanged (Figure 5b,c,f,g,j,k,n,o). The spatial pattern of these differences (Figure 5d,h,l,p) displays distinct seasonal patterns. In spring, positive values dominate the southeastern region, whereas most of the western area is characterized by negative values. Summer exhibits the smallest magnitude of differences, with the central region showing positive values and the northwestern region negative values. During autumn, the northern region falls within the positive zone, in contrast to the negative values observed in the western and southern regions. The most pronounced disparities occur in winter, with the northern and southwestern regions marked by negative values and positive values sporadically distributed across the northeastern, northwestern, and parts of the southern regions.

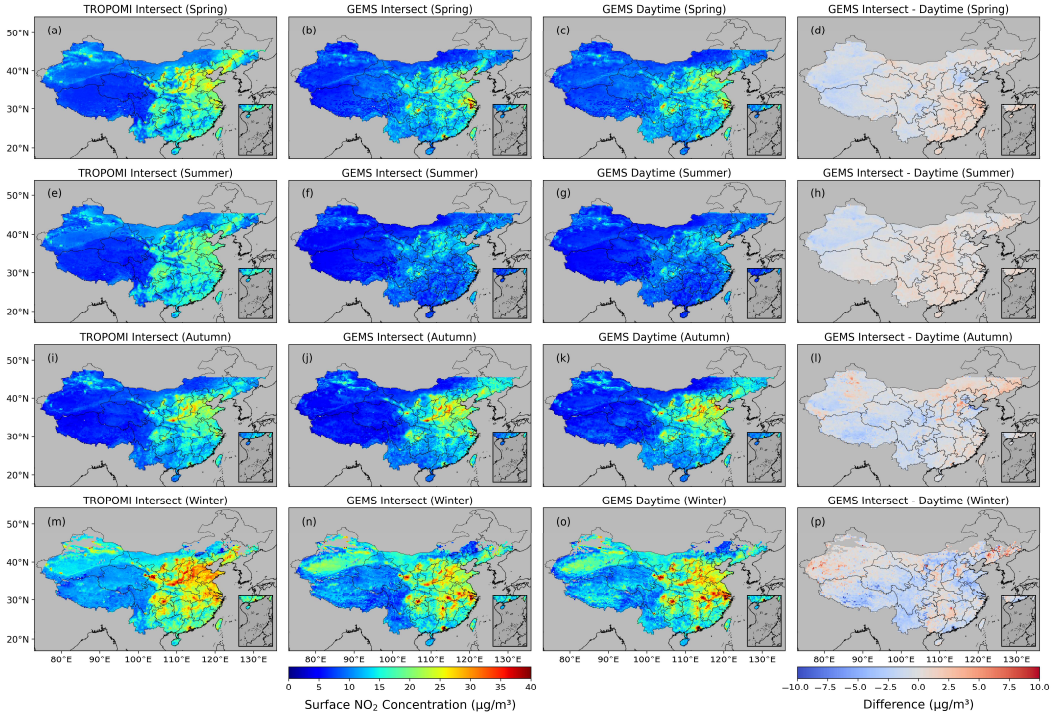

Figure 5. Spatial distributions of surface NO_2 concentrations over China: seasonal patterns and dataset comparisons. The first column (**a**,**e**,**i**,**m**) shows TROPOMI Intersect (surface NO_2 concentrations derived from TROPOMI NO_2 VCDs at locations co-located with GEMS observations). The second column (**b**,**f**,**j**,**n**) shows GEMS Intersect (surface NO_2 concentrations derived from GEMS NO_2 VCDs at locations co-located with TROPOMI observations). The third column (**c**,**g**,**k**,**o**) shows GEMS Daytime (pixel-level daily averages computed only when valid values exist, under GEMS Intersect spatiotemporal matching constraints). The fourth column (**d**,**h**,**l**,**p**) shows GEMS Intersect − Daytime (GEMS Intersect minus GEMS Daytime). Rows correspond to spring (**a**–**d**), summer (**e**–**h**), autumn (**i**–**l**), and winter (**m**–**p**).

Figure 6 shows the complete spatial coverage of surface NO_2 concentrations derived from TROPOMI ($ALL_{TROPOMI}$) and GEMS (ALL_{GEMS}) NO_2 VCDs across four seasons using the CatBoost model, complementing the sample-based spatial patterns in Figure 5. Confirming the pattern in Figure 5, both $ALL_{TROPOMI}$ and ALL_{GEMS} show that high-concentration areas expand significantly and their intensity increases substantially in winter, reaching the highest annual values. In spring, northern China and the Yangtze River Delta, Pearl River Delta, and Sichuan–Chongqing urban agglomerations show higher NO_2

concentrations in ALL$_{TROPOMI}$ (Figure 6a). In contrast, ALL$_{GEMS}$ shows lower NO_2 concentrations across most of eastern China (Figure 6e). In summer, ALL$_{TROPOMI}$ shows an overall decrease, with a pronounced reduction in previous high-concentration areas (Figure 6b). ALL$_{GEMS}$ maintains relatively high concentrations in northern China and southern urban agglomerations, but the spatial extents of these highs are smaller than in ALL$_{TROPOMI}$ (Figure 6f). In autumn, high-concentration areas expand in both datasets, with peak values primarily occurring in northern China (Figure 6c,g). In winter, regional contrasts are more pronounced in ALL$_{GEMS}$. ALL$_{TROPOMI}$ shows a continuous belt of high concentrations across North China (Figure 6d). ALL$_{GEMS}$ shows more scattered high-concentration areas focused on major cities and industrial zones (Figure 6h). For the entire year, ALL$_{TROPOMI}$ provides complete spatial coverage (Figure 6a–d). ALL$_{GEMS}$ lacks annual data in the Northeast and Northwest regions due to its geostationary field of regard (Figure 6e–h). These spatial differences stem partly from fundamental variations in sampling strategies: GEMS captures multiple daytime observations daily, whereas TROPOMI provides only one observation per day. This leads to differences in the spatial representativeness and the averaging of diurnal pollution patterns captured by each dataset.

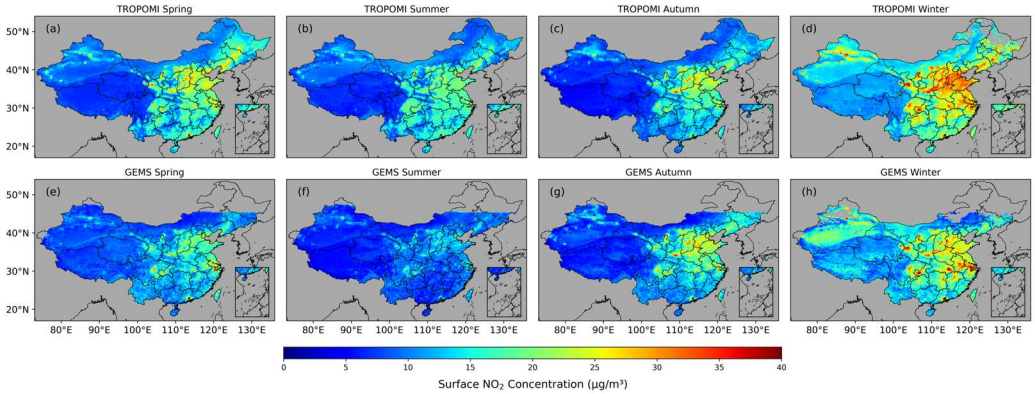

Figure 6. Surface NO_2 concentrations derived from TROPOMI NO_2 VCDs in spring (**a**), summer (**b**), fall (**c**), and winter (**d**). The bottom row panels are the same as the top row panels, but for GEMS. Each subfigure is described as follows: (**a**) TROPOMI spring; (**b**) TROPOMI summer; (**c**) TROPOMI autumn; (**d**) TROPOMI winter; (**e**) GEMS spring; (**f**) GEMS summer; (**g**) GEMS autumn; (**h**) GEMS winter.

3.2.2. Temporal Variation Comparison

Figure 7 presents the monthly temporal variations in surface NO_2 concentrations, which are derived from TROPOMI and GEMS datasets and retrieved using the CatBoost model. Panel (a) directly compares the average concentrations derived from TROPOMI and GEMS at co-located sampling points, while Panel (b) focuses on the region south of 45°N—a densely populated and industrialized area of interest. Both panels confirm that the two datasets follow a consistent seasonal cycle: concentrations peak in winter and reach their trough in summer, which aligns with the seasonal characteristics of anthropogenic emissions and atmospheric chemical reactivity. Notably, TROPOMI-based estimates are systematically higher than GEMS-based ones across most months, with the discrepancy being most pronounced during the low-concentration summer period, reflecting the inherent differences in satellite sampling strategies and retrieval biases.

Supplementary figures (Figures S5–S8) further illustrate the temporal variations in NO_2 concentrations. Intra-monthly variations in GEMS-based estimates (ALL$_{GEMS}$) are more pronounced, with a larger range between extreme values compared to TROPOMI-based estimates (ALL$_{TROPOMI}$). The differences between the two datasets are smaller in summer than in winter, and comparisons at the same sampling points show that GEMS-

based estimates have a wider range of extreme values. The discrepancies between the two datasets are smaller in summer than in winter. Comparisons at co-located sampling points further confirm that GEMS-based estimates display a wider range of extreme values and reveal more localized high-concentration clusters, whereas TROPOMI-based estimates tend to produce smoother spatial fields. More details are provided in Note S3 of the Supplementary Materials.

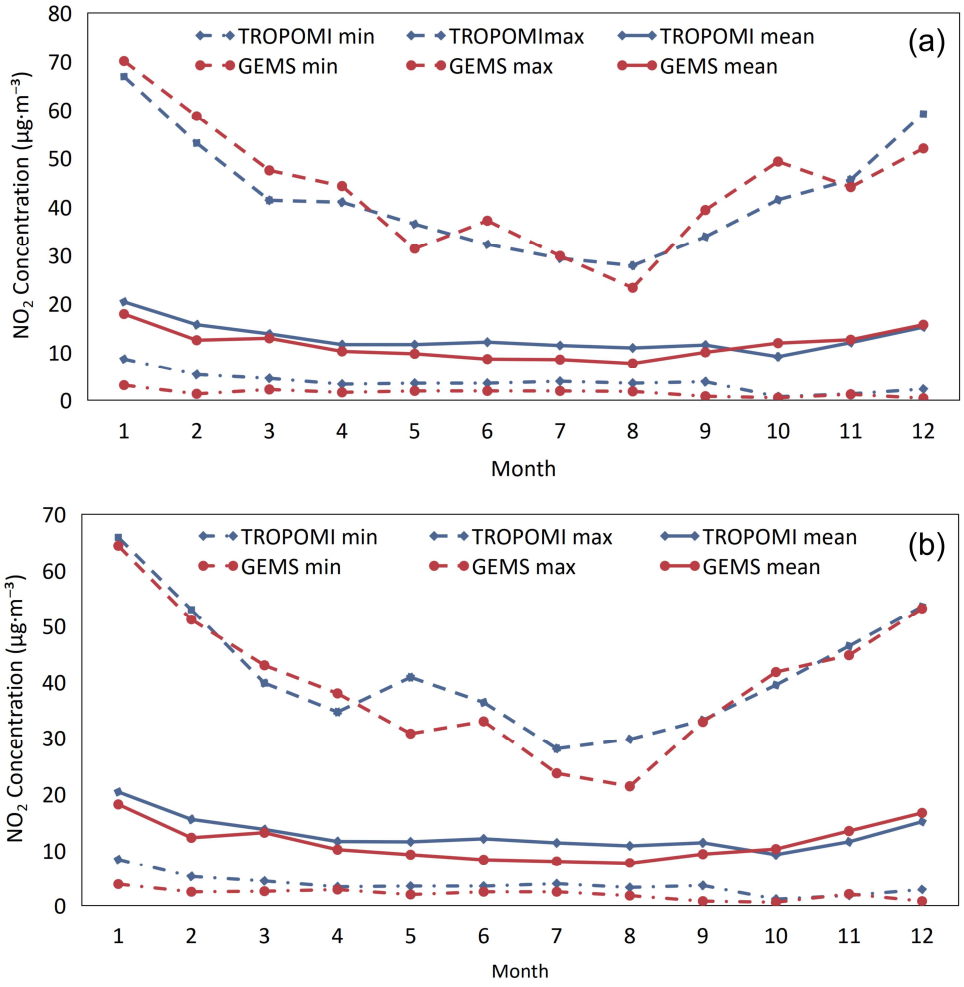

Figure 7. (**a**) shows a line graph of monthly average surface NO_2 concentrations derived from TROPOMI and GEMS at the co-located sampling points. (**b**) shows a line graph of monthly average surface NO_2 concentrations south of 45°N derived from TROPOMI and GEMS.

3.2.3. Intraday Variations

Figures 8 and 9 show the hourly variations in average surface NO_2 concentrations retrieved by the CatBoost model in January (winter) and July (summer). January exhibits fewer valid observations than July at 7:00, 8:00, 15:00, and 16:00 Beijing Time (BJT). By comparison, the overall NO_2 concentration is significantly higher in the winter than in the summer. In January, only limited data are available at 9:00 BJT in the southern part of Northeast China (Figure 8a). At 10:00 BJT, the NO_2 concentration reaches its peak, as shown in the figure (Figure 8b), with extensive areas of high NO_2 concentrations in East China. By 11:00 BJT, the concentration exhibits a decreasing trend (Figure 8c): the Yangtze River Delta region and the southern Gansu Province have the highest relative NO_2 concentrations, followed by the North China Plain, Sichuan–Chongqing region, Pearl River Delta and other areas. Overall, NO_2 concentrations declined significantly between 10:00 and 14:00 BJT (Figure 8d–f). For July, data availability is more comprehensive than in

January, with continuous data coverage from 7:00 to 16:00 BJT. Notably, NO_2 concentrations across China remained at low levels during this period (Figure 9). At 7:00 BJT, NO_2 concentrations exhibit minimal variability (Figure 9a). From 8:00 BJT onward, concentrations continue to decrease until 16:00 BJT, when a slight rebound is observed (Figure 9b–j). High NO_2 concentrations are mainly concentrated in North and East China—regions characterized by intensive industrial activities and heavy traffic—while low concentrations are distributed in western inland regions, including Tibet and Qinghai.

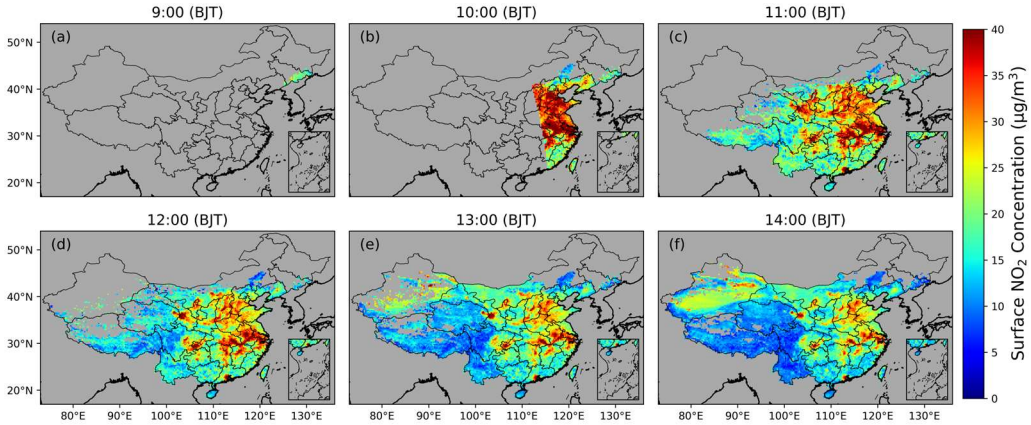

Figure 8. Hourly surface NO_2 concentrations of ALL$_{GEMS}$ in winter (take January as an example). BJT means Beijing Time, and BJT = UTC + 8 h. Each subfigure is described as follows: (**a**) surface NO_2 concentrations at 9:00 BJT; (**b**) surface NO_2 concentrations at 10:00 BJT; (**c**) surface NO_2 concentrations at 11:00 BJT; (**d**) surface NO_2 concentrations at 12:00 BJT; (**e**) surface NO_2 concentrations at 13:00 BJT; (**f**) surface NO_2 concentrations at 14:00 BJT.

3.2.4. Pollution Process Tracking

To demonstrate the utility of GEMS data in near-real-time pollution monitoring, we selected the 11 March 2022 NO_2 pollution event in the Yangtze River delta as a case study (Figure 10). This region was chosen because it represents a major emission hotspot with a dense ground monitoring network for robust validation, and its complex meteorology often leads to clear, dynamically evolving pollution episodes that are ideal for showcasing the high-temporal-resolution capabilities of geostationary observations. Our retrieved near-surface NO_2 concentrations show good consistency with ground station observations, verifying the robustness of our model in capturing NO_2 variations throughout this event. Spatially, the NO_2 pollution was concentrated in southern Jiangsu Province, Shanghai and the northern border of Zhejiang Province. Temporally, the pollution changed as follows: At 9:00 BJT, the maximum NO_2 concentration exceeded 90 $\mu g \cdot m^{-3}$ (Figure 10(a-1)), rising to over 100 $\mu g \cdot m^{-3}$ at 10:00 BJT (Figure 10(a-1)); it then decreased to about 60 $\mu g \cdot m^{-3}$ by 14:00 BJT (Figure 10(f-1)), before rebounding slightly to over 70 $\mu g \cdot m^{-3}$ after 15:00 BJT (Figure 10(g-1)). The mean NO_2 concentration decreased sharply between 9:00 and 14:00 BJT (Figure 10(a-1–f-1)), followed by a gradual decline thereafter. Notably, the entire event was accompanied by a continuous shrinkage of the NO_2 pollution plume, with its northern boundary moving southward, which is partly due to the southward migration of the wind field convergence around 32°N (Figure 10(a-1–d-1)). Furthermore, GEMS data effectively captures the seamless migration of the pollution core from southern Jiangsu to eastern Shanghai, unveiling the spatial trajectory within gaps between ground stations. It also resolves the continuous concentration gradient (from 90 to 70 $\mu g \cdot m^{-3}$) across the site-sparse border region between southern Jiangsu and northern Zhejiang.

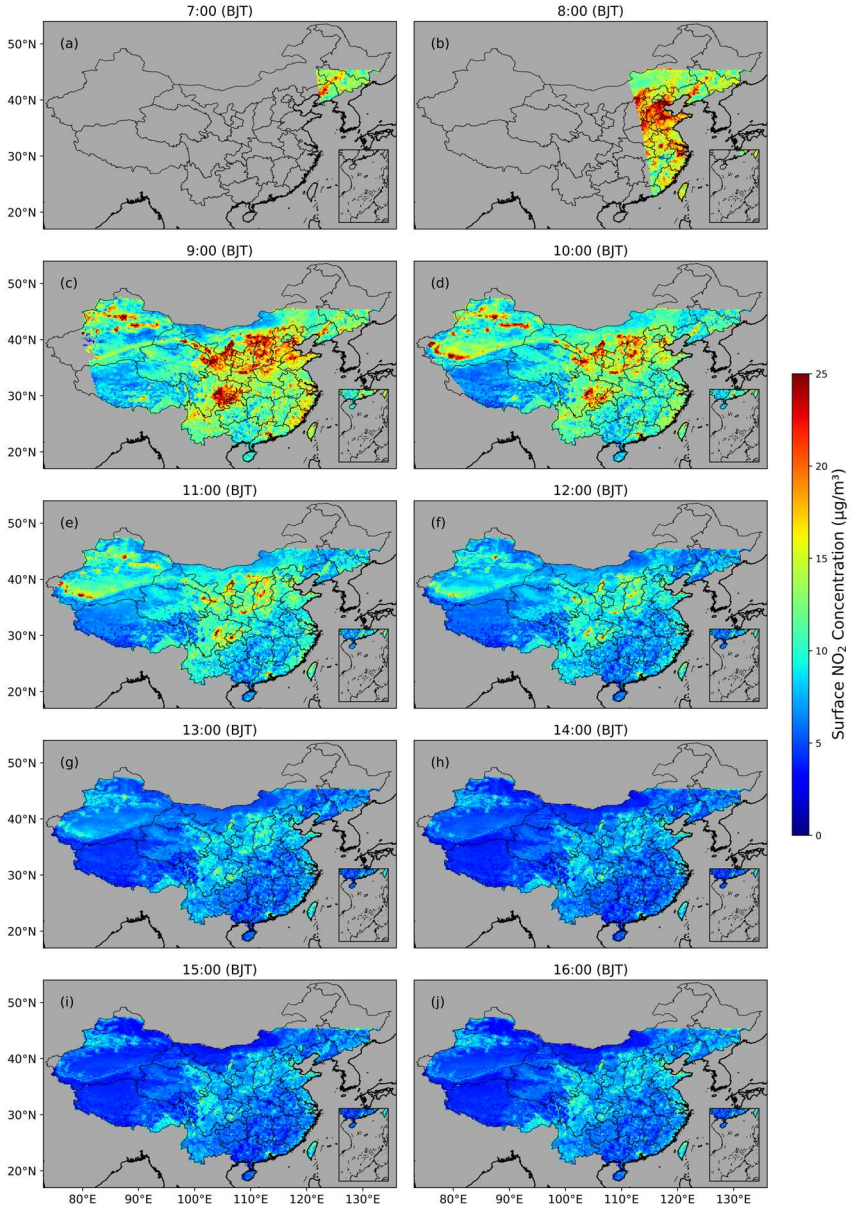

Figure 9. Hourly surface NO_2 concentrations of ALL$_{GEMS}$ in the summer (take July as an example). BJT means Beijing Time, and BJT = UTC + 8 h. Each subfigure is described as follows: (**a**) surface NO_2 concentrations at 7:00 BJT; (**b**) surface NO_2 concentrations at 8:00 BJT; (**c**) surface NO_2 concentrations at 9:00 BJT; (**d**) surface NO_2 concentrations at 10:00 BJT; (**e**) surface NO_2 concentrations at 11:00 BJT; (**f**) surface NO_2 concentrations at 12:00 BJT; (**g**) surface NO_2 concentrations at 13:00 BJT; (**h**) surface NO_2 concentrations at 14:00 BJT; (**i**) surface NO_2 concentrations at 15:00 BJT; (**j**) surface NO_2 concentrations at 16:00 BJT.

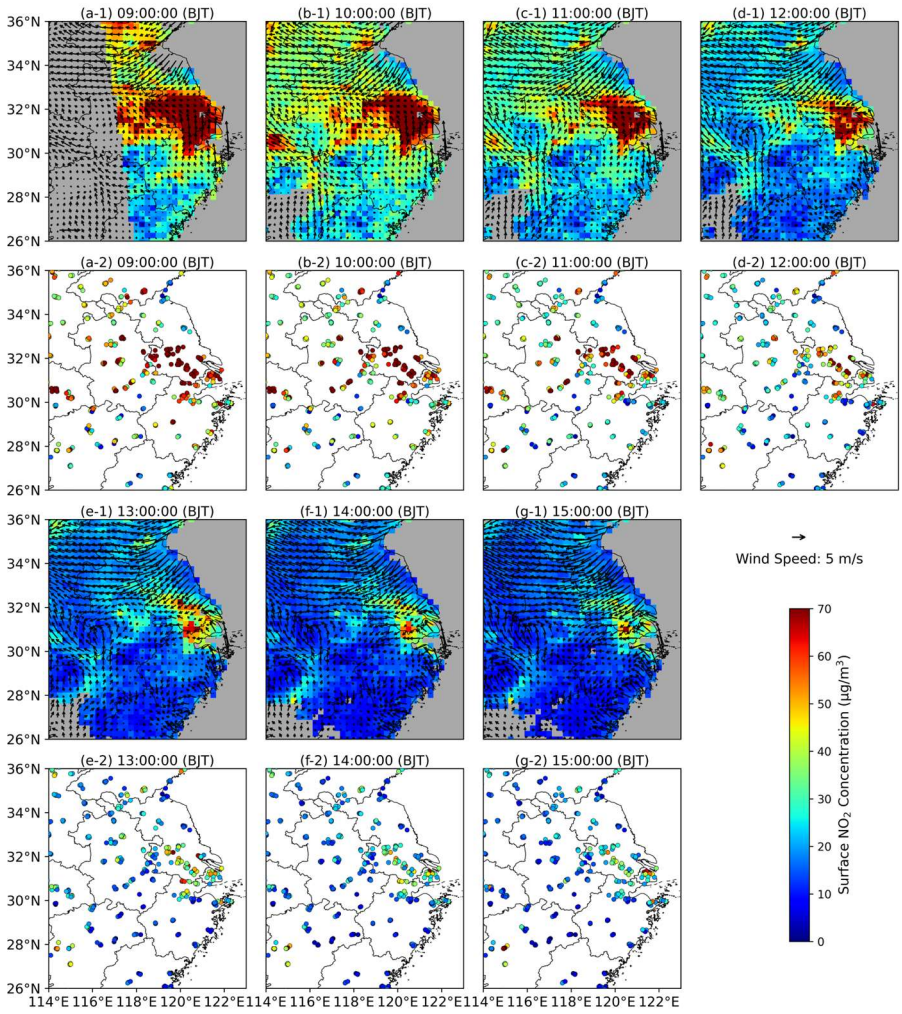

Figure 10. Variations in surface NO_2 concentrations during pollution in the Yangtze River Delta on 11 March 2022. Rows 1 and 3 (**a-1**–**g-1**) show our retrieval results, while rows 2 and 4 (**a-2**–**g-2**) show the measurements from the ground station. Overlaid vector arrows represent the contemporaneous surface wind field (10 m wind speed and direction).

4. Discussion

When designing our machine learning models, we evaluated the inclusion of explicit spatial coordinates as input features but ultimately excluded them. Preliminary tests showed that while adding static geographic labels such as longitude and latitude could lead to a modest improvement in statistical metrics, for example, an approximately 1–2% increase in R^2, it also introduced spatially unrealistic artifacts like striped patterns in the estimated concentration maps. This issue aligns with problems noted in prior studies [72]. More fundamentally, since latitude and longitude are static locational markers, they do not represent the dynamic physical and chemical processes governing surface NO_2 variability. Their use could therefore introduce a location-dependent bias that lacks physical interpretability. Future work may integrate spatial context more appropriately by incorporating physically grounded variables such as wind fields or by adopting transport-aware modeling frameworks.

It is noteworthy that all models achieved exceptionally high accuracy on the training set (e.g., CatBoost training $R^2 \approx 0.998$), which confirms the strong intrinsic capacity of the tree-based ensemble models, particularly the gradient boosting frameworks, to capture the complex nonlinear relationships between satellite VCDs and surface NO_2 concentrations. The reasonable degree of performance decline observed on the independent test set

(e.g., CatBoost test $R^2 \approx 0.765$) is precisely the key indicator that the models have successfully learned generalizable patterns rather than memorizing noise or idiosyncrasies in the training data. The final test set performance achieved (R^2: 0.765–0.842) is already superior to levels commonly reported in similar studies [32,43,45] and is fully sufficient for high-accuracy retrieval. More importantly, the consistent ranking of all four models (CatBoost > XGBoost > LGBM > RF) across both training and test sets underscores the robustness of our evaluation, providing a solid foundation for the subsequent spatial–temporal analysis and pollution process tracking. These spatiotemporal patterns are consistent with previous satellite-based findings regarding NO_2 distributions over China [27,38], their seasonal cycles [14,73], and diurnal variations captured by high-frequency observations [43,44].

It is important to address the systematic underestimation of high-concentration values by the models. All models, including the best-performing CatBoost, exhibit a tendency to underestimate high concentrations. This may stem from multiple factors: increased uncertainty in satellite retrievals under heavily polluted conditions, the inherent scarcity of extreme samples in training data, and potential nonlinear responses of ground-based monitoring instruments at very high concentrations. To address this issue, future research could explore modeling techniques specifically optimized for extreme values, such as quantile regression [74], cost-sensitive learning [75], or attempts to integrate physical constraints from chemical transport models [76], thereby improving prediction reliability during pollution events.

5. Conclusions

This study conducted a comprehensive comparison of machine learning-based surface NO_2 concentration estimation using tropospheric column data from the polar-orbiting TROPOMI and the geostationary GEMS instruments. Four tree-based models—Random Forest, XGBoost, CatBoost, and LightGBM—were trained and evaluated with multi-source data across China in 2022. The primary objectives were to assess the performance differences between models trained on the two satellite datasets, examine whether GEMS's high temporal resolution enhances estimation accuracy, and explore the synergistic effects of optimal models and data sources.

Among the four machine learning models evaluated, the results demonstrate that the CatBoost model achieved the highest estimation accuracy, with an R^2 of 0.765 for TROPOMI and 0.842 for GEMS, along with the lowest errors ($RMSE_{TROPOMI}$: 5.595 $\mu g \cdot m^{-3}$, $RMSE_{GEMS}$: 4.567 $\mu g \cdot m^{-3}$; $MAE_{TROPOMI}$: 3.515 $\mu g \cdot m^{-3}$, MAE_{GEMS}: 2.890 $\mu g \cdot m^{-3}$). The feature importance analysis further revealed that CatBoost made the most effective use of satellite NO_2 data as a core input, with normalized importance scores reaching 50.6% for TROPOMI and 32.2% for GEMS. In contrast, Random Forest showed the lowest accuracy, while XGBoost and LightGBM delivered intermediate results. These findings highlight CatBoost's superior capability in handling the nonlinear relationships and feature interactions inherent in the estimation task.

Models trained on GEMS data consistently outperformed those using TROPOMI across all evaluated metrics, which is largely attributable to the substantially larger training sample size (949,212 for GEMS vs. 89,932 for TROPOMI) enabled by GEMS's high temporal resolution. The controlled experiment with identical sample sizes showed that TROPOMI provides a slight advantage under equal data volume ($R^2_{TROPOMI} = 0.757$, $R^2_{GEMS} = 0.747$). However, in practical applications, the benefits of GEMS's frequent sampling—leading to a more representative and temporally diverse dataset—far outweighed the marginal gains from TROPOMI's spatial precision. This underscores the critical role of sample volume in enhancing model generalization, stability, and accuracy in surface NO_2 estimation.

Spatially, estimates derived from GEMS exhibited sharper gradients and more localized emission patterns, effectively highlighting urban and industrial hotspots. In contrast, TROPOMI-based estimates tended to produce smoother and more regionally extended concentration fields, potentially blending localized peaks with background levels. Temporally, both datasets captured consistent seasonal cycles (higher in winter, lower in summer). However, GEMS captures multiple daytime observations daily, enabling better representation of diurnal pollution dynamics, while TROPOMI's single daily overpass may average out short-term variations.

The high temporal resolution of GEMS uniquely enabled the reconstruction of detailed diurnal variation patterns in surface NO_2. Hourly analysis showed clear morning peaks and afternoon declines, with stronger diurnal cycles in winter than in summer. Notably, GEMS data allowed for near-real-time tracking of pollution episode evolution, as demonstrated in a case study over the Yangtze River Delta on 11 March 2022. The model successfully captured the dynamic changes in NO_2 concentrations throughout the daytime, aligning closely with ground-based observations. This capability highlights the significant added value of geostationary data for high-frequency air quality monitoring, process analysis, and timely pollution event management.

This research confirms the strong potential of geostationary satellite data for high-spatiotemporal-resolution air quality monitoring. Building on the findings presented here, several meaningful directions for future work can be outlined, which also address certain aspects that could be further developed: (1) extending the analysis to multi-year time series to better capture interannual variability and consolidate model robustness; (2) integrating GEMS observations with other satellite datasets to complement its spatial coverage, especially in regions such as Northwest and Northeast China; and (3) incorporating atmospheric chemical transport modeling into the machine learning framework to enhance interpretability and adaptivity under changing pollution regimes.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18040614/s1, Table S1: Summary of the datasets used in this study; Note S1: Weighted kernel density analysis; Table S2: Road types and corresponding classes and weights; Note S2: Detailed machine learning model specifications and implementation; Table S3: Model optimization frameworks and hyperparameter search spaces; Figure S1: Distribution of monitoring stations in China in 2022; Table S4: Model performance evaluation metrics; Figure S2: Matching of ground station data, satellite data, meteorological data and auxiliary data; Table S5: Descriptive characteristic statistics of Group$_T$ data; Table S6: Descriptive characteristic statistics of Group$_G$ data; Figure S3: Scatterplots of satellite-derived tropospheric NO_2 vertical-column densities (VCDs) against co-located ground-level NO_2 concentrations for (a) TROPOMI and (b) GEMS; Figure S4: Spatial distributions of training set sample sizes and test set correlation coefficients (r), with correlation analyses between sample size and model performance for surface NO_2 concentration retrievals; Figure S5: Monthly surface NO_2 concentrations derived from TROPOMI NO_2 VCDs; Figure S6: Monthly surface NO_2 concentrations derived from GEMS NO_2 VCDs; Figure S7: Monthly surface NO_2 concentrations derived from TROPOMI NO_2 VCDs at locations co-located with GEMS observations; Figure S8: Monthly surface NO_2 concentrations derived from GEMS NO_2 VCDs at locations co-located with TROPOMI observations; Note S3: Detailed temporal variability of surface NO_2 concentrations (Corresponding to Figures S5–S8).

Author Contributions: Conceptualization, Y.M. and Y.W.; methodology, Y.M. and Y.W.; software, Y.M. and C.W.; validation, Y.M. and Y.W.; formal analysis, Y.M.; investigation, Y.M. and Y.W.; resources, Y.W. and J.W.; data curation, Y.M.; writing—original draft preparation, Y.M.; writing—review and editing, Y.M., Y.W., J. W., J.K. and M.T.; visualization, Y.M., C.W. and S.Z.; supervision, Y.W.; project administration, Y.W.; funding acquisition, Y.W. All authors have read and agreed to the published version of the manuscript.

https://doi.org/10.3390/rs18040614

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 42201409).

Data Availability Statement: TROPOMI NO_2 data can be obtained from ERATHDATA GESDISC data portal (https://disc.gsfc.nasa.gov/, last access: 30 November 2025). The GEMS NO_2 data can be obtained from (https://nesc.nier.go.kr/, last access: 30 September 2025). The meteorological data can be obtained from the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis of the global climate (ERA5) from the Copernicus Climate Change Service website (https://cds.climate.copernicus.eu/, last access: 30 December 2025). The ground monitoring station NO_2 data can be obtained from the Urban Air Quality from China National Environmental Monitoring Centre (CNEMC) data portal (https://air.cnemc.cn:18007/, last access: 12 January 2026). The NASA-MOD13A3 NDVI data can be obtained from the ERATHDATA MODIS data portal (https://www.earthdata.nasa.gov/, last access: 27 May 2024). The population data can be obtained from the LandScan data portal (https://landscan.ornl.gov/, last access: 27 May 2024). The DEM can be obtained from the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/, last access: 27 May 2024). The road data can be obtained from OpenStreetMap (https://www.openstreetmap.org/, last access: 27 May 2024).

Acknowledgments: Jun Wang's contribution is made possible by in-kind support through the Lichtenberger Family Chair professorship in Chemical Engineering at the University of Iowa.

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

References

1. Change, I.P.O.C. Climate change 2007: The physical science basis. *Agenda* **2007**, *6*, 333.
2. Fu, J.; Xiao, L. Five cases of nitrogen dioxide toxic pulmonary edema reported. *Chin. J. Occup. Health Occup. Dis.* **1991**, *9*, 40–41.
3. Xia, X.; Meng, X.; Liu, C.; Guo, Y.; Li, X.; Niu, Y. Associations of long-term nitrogen dioxide exposure with a wide spectrum of diseases: A prospective cohort study of 0.5 million Chinese adults. *Lancet Public Health* **2024**, *9*, e1047–e1058.
4. de Bont, J.; Jaganathan, S.; Dahlquist, M.; Persson, A.; Stafoggia, M.; Ljungman, P. Ambient air pollution and cardiovascular diseases: An umbrella review of systematic reviews and meta-analyses. *J. Intern. Med.* **2022**, *291*, 779–800.
5. Wang, K.; Yuan, Y.; Wang, Q.; Yang, Z.; Zhan, Y.; Wang, Y.; Wang, F.; Zhang, Y. Incident risk and burden of cardiovascular diseases attributable to long-term NO_2 exposure in Chinese adults. *Environ. Int.* **2023**, *178*, 11.
6. Chang, K.; Hsu, P.; Lin, C.; Lin, C.; Juo, S.H.; Liang, C. Traffic-related air pollutants increase the risk for age-related macular degeneration. *J. Investig. Med.* **2019**, *67*, 1076–1081.
7. Wang, W.; Zhong, C.; Huang, L.; Zhou, X.; Chen, R.; Wu, J.; Li, X.; Xiong, T.; Liu, C.; Xiao, M.; et al. Prenatal NO_2 exposure and ultrasound measures of foetal growth: A prospective cohort study in Wuhan, China. *Occup. Environ. Med.* **2017**, *74*, 204–210.
8. Li, K.; Jacob, D.J.; Liao, H.; Shen, L.; Zhang, Q.; Bates, K.H. Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 422–427.
9. Lelieveld, J.; Klingmueller, K.; Pozzer, A.; Burnett, R.T.; Haines, A.; Ramanathan, V. Effects of fossil fuel and total anthropogenic emission removal on public health and climate. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 7192–7197.
10. Zhou, M.; Wang, H.; Zeng, X.; Yin, P.; Zhu, J.; Chen, W.; Li, X.; Wang, L.; Wang, L.; Liu, Y.; et al. Mortality, morbidity, and risk factors in China and its provinces, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2019**, *394*, 1145–1158.
11. Beirle, S.; Boersma, K.F.; Platt, U.; Lawrence, M.G.; Wagner, T. Megacity Emissions and Lifetimes of Nitrogen Oxides Probed from Space. *Science* **2011**, *333*, 1737–1739.
12. de Foy, B.; Wilkins, J.L.; Lu, Z.; Streets, D.G.; Duncan, B.N. Model evaluation of methods for estimating surface emissions and chemical lifetimes from satellite data. *Atmos. Environ.* **2014**, *98*, 66–77.
13. Laughner, J.L.; Cohen, R.C. Direct observation of changing NO_x lifetime in North American cities. *Science* **2019**, *366*, 723.
14. Shah, V.; Jacob, D.J.; Li, K.; Silvern, R.F.; Zhai, S.; Liu, M.; Lin, J.; Zhang, Q. Effect of changing NO_x lifetime on the seasonality and long-term trends of satellite-observed tropospheric NO_2 columns over China. *Atmos. Chem. Phys.* **2020**, *20*, 1483–1495.
15. Liu, W. Development Opportunities for Atmospheric Environmental Optical Monitoring Technology under the "Dual Carbon" Goals. *Acta Opt. Sin.* **2022**, *42*, 10–25.

16. Goldberg, D.L.; Anenberg, S.C.; Kerr, G.H.; Mohegh, A.; Lu, Z.; Streets, D.G. TROPOMI NO_2 in the United States: A Detailed Look at the Annual Averages, Weekly Cycles, Effects of Temperature, and Correlation With Surface NO_2 Concentrations. *Earth Future* **2021**, *9*, 16.
17. Cooper, M.J.; Martin, R.; 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*, 12.
18. Grajales, J.F.; Baquero-Bernal, A. Inference of surface concentrations of nitrogen dioxide (NO_2) in Colombia from tropospheric columns of the ozone measurement instrument (OMI). *Atmósfera* **2014**, *27*, 193–214.
19. Gu, J.; Chen, L.; Yu, C.; Li, S.; Tao, J.; Fan, M.; Xiong, X.; Wang, Z.; Shang, H.; Su, L. Ground-Level NO_2 Concentrations over China Inferred from the Satellite OMI and CMAQ Model Simulations. *Remote Sens.* **2017**, *9*, 17.
20. Wang, Y.; Wang, J.; Xu, X.; Henze, D.K.; Qu, Z.; Yang, K. Inverse modeling of SO_2 and NO_x emissions over China using multi-sensor satellite data – Part 1: Formulation and sensitivity analysis. *Atmos. Chem. Phys.* **2020**, *20*, 6631–6650.
21. Wang, Y.; Wang, J.; Zhou, M.; Henze, D.K.; Ge, C.; Wang, W. Inverse modeling of SO_2 and NO_x emissions over China using multi-sensor satellite data: 2. Downscaling techniques for air quality analysis and forecasts. *Atmos. Chem. Phys.* **2019**, *20*, 6631–6650.
22. Bechle, M.J.; Millet, D.B.; Marshall, J.D. National Spatiotemporal Exposure Surface for NO_2: Monthly Scaling of a Satellite-Derived Land-Use Regression, 2000–2010. *Environ. Sci. Technol.* **2015**, *49*, 12297–12305.
23. Qin, K.; Rao, L.; Xu, J.; Bai, Y.; Zou, J.; Hao, N.; Li, S.; Yu, C. Estimating Ground Level NO_2 Concentrations over Central-Eastern China Using a Satellite-Based Geographically and Temporally Weighted Regression Model. *Remote Sens.* **2017**, *9*, 20.
24. Robinson, D.P.; Lloyd, C.D.; McKinley, J.M. Increasing the accuracy of nitrogen dioxide (NO_2) pollution mapping using geographically weighted regression (GWR) and geostatistics. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 374–383.
25. Young, M.T.; Bechle, M.J.; Sampson, P.D.; Szpiro, A.A.; Marshall, J.D.; Sheppard, L.; Kaufman, J.D. Satellite-Based NO_2 and Model Validation in a National Prediction Model Based on Universal Kriging and Land-Use Regression. *Environ. Sci. Technol.* **2016**, *50*, 3686–3694.
26. Qin, K.; Han, X.; Li, D.; Xu, J.; Loyola, D.; Xue, Y.; Zhou, X.; Li, D.; Zhang, K.; Yuan, L. Satellite-based estimation of surface NO_2 concentrations over east-central China: A comparison of POMINO and OMNO2d data. *Atmos. Environ.* **2020**, *224*, 10.
27. Zhan, Y.; Luo, Y.; Deng, X.; Zhang, K.; Zhang, M.; Grieneisen, M.L.; Di, B. Satellite-based estimates of daily NO_2 exposure in China using hybrid random forest and spatiotemporal kriging model. *Environ. Sci. Technol.* **2018**, *52*, 4180–4189.
28. Lamsal, L.N.; Martin, R.V.; van Donkelaar, A.; Steinbacher, M.; Celarier, E.A.; Bucsela, E.; Dunlea, E.J.; Pinto, J.P. Ground-level nitrogen dioxide concentrations inferred from the satellite -borne Ozone Monitoring Instrument. *J. Geophys. Res. Atmos.* **2008**, *113*, 15.
29. Kim, D.; Lee, H.; Hong, H.; Choi, W.; Lee, Y.G.; Park, J. Estimation of Surface NO_2 Volume Mixing Ratio in Four Metropolitan Cities in Korea Using Multiple Regression Models with OMI and AIRS Data. *Remote Sens.* **2017**, *9*, 16.
30. Ghahremanloo, M.; Lops, Y.; Choi, Y.; Mousavinezhad, S.; Jung, J. A Coupled Deep Learning Model for Estimating Surface NO_2 Levels From Remote Sensing Data: 15-Year Study Over the Contiguous United States. *J. Geophys. Res. Atmos.* **2023**, *128*, 18.
31. Becerra-Rondon, A.; Ducati, J.; Haag, R. Satellite-based estimation of NO_2 concentrations using a machine-learning model: A case study on Rio Grande do Sul, Brazil. *Atmósfera* **2023**, *37*, 175–190.
32. Deng, F.; Chen, Y.; Liu, W.; Li, L.; Chen, X.; Tiwari, P.; Qin, K. Satellite-Based Estimation of Near-Surface NO_2 Concentration in Cloudy and Rainy Areas. *Remote Sens.* **2024**, *16*, 14.
33. Ghahremanloo, M.; Lops, Y.; Choi, Y.; Yeganeh, B. Deep Learning Estimation of Daily Ground -Level NO_2 Concentrations From Remote Sensing Data. *J. Geophys. Res. Atmos.* **2021**, *126*, 18.
34. Zhang, C.; Liu, C.; Li, B.; Zhao, F.; Zhao, C. Spatiotemporal neural network for estimating surface NO_2 concentrations over north China and their human health impact. *Environ. Pollut.* **2022**, *307*, 8.
35. Zeng, Q.; Wang, L.; Zhu, H.; Liu, S.; Wang, C.; Chen, L.; Tao, J. Estimating Daily Concentrations of Near-Surface CO, NO_2, and O_3 Simultaneously Over China Based on Spatiotemporal Multi-Task Transformer Model. *Atmos. Environ.* **2024**, *316*, 16.
36. Siddique, M.A.; Naseer, E.; Usama, M.; Basit, A. Estimation of Surface-Level NO_2 Using Satellite Remote Sensing and Machine Learning: A review. *IEEE Geosci. Remote Sens. Mag.* **2024**, *12*, 8–34.
37. Kang, Y.; Choi, H.; Im, J.; Park, S.; Shin, M.; Song, C.; Kim, S. Estimation of surface-level NO_2 and O_3 concentrations using TROPOMI data and machine learning over East Asia. *Environ. Pollut.* **2021**, *288*, 14.
38. Li, D. Optimization and Integrated Mapping of Near-Surface NO_2 Estimation Supported by GOME-2B/OMI/TROPOMI Remote Sensing Data. *Master's Thesis*, China University of Mining and Technology, Xuzhou, China, 2021.

39. Long, S. Estimation of Near-Surface NO_2 Concentrations in China Based on Machine Learning. *Master's Thesis*, Nanjing University of Information Science & Technology, Nanjing, China, 2022.
40. Li, D.; Qin, K.; Cohen, J.B.; He, Q.; Wang, S.; Li, D.; Zhou, X.; Ling, X.; Xue, Y. Combing GOME-2B and OMI Satellite Data to Estimate Near-Surface NO_2 of Mainland China. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 10269–10277.
41. Ahmad, N.; Lin, C.; Lau, A.K.H.; Kim, J.; Yu, F.; Li, C.; Li, Y.; Fung, J.C.H.; Lao, X.Q. Improving Ground-Level NO_2 Estimation in China Using GEMS Measurements and a Nested Machine Learning. *Egusphere*. **2024**, *2024*, 1–26.
42. Ding, Y.; Li, S.; Xing, J.; Li, X.; Ma, X.; Song, G.; Teng, M.; Yang, J.; Dong, J.; Meng, S. Retrieving hourly seamless PM2.5 concentration across China with physically informed spatiotemporal connection. *Remote Sens. Environ.* **2024**, *301*, 16.
43. Yang, Q.; Kim, J.; Cho, Y.; Lee, W.; Lee, D.; Yuan, Q.; Wang, F.; Zhou, C.; Zhang, X.; Xiao, X.; et al. A synchronized estimation of hourly surface concentrations of six criteria air pollutants with GEMS data. *npj Clim. Atmos. Sci.* **2023**, *6*, 9.
44. 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_2 and its spatiotemporal variations in China using GEMS measurements and a nested machine learning model. *Atmos. Chem. Phys.* **2024**, *24*, 9645–9665.
45. He, Q.; Qin, K.; Cohen, J.B.; Li, D.; Kim, J. Quantifying Uncertainty in ML-Derived Atmosphere Remote Sensing: Hourly Surface NO_2 Estimation With GEMS. *Geophys. Res. Lett.* **2024**, *51*, 13.
46. Ludewig, A.; Kleipool, Q.; Bartstra, R.; Landzaat, R.; Leloux, J.; Loots, E.; Meijering, P.; van der Plas, E.; Rozemeijer, N.; Vonk, F.; et al. In-flight calibration results of the TROPOMI payload on board the Sentinel-5 Precursor satellite. *Atmos. Meas. Tech.* **2020**, *13*, 3561–3580.
47. Lamsal, L.N.; Krotkov, N.A.; Vasilkov, A.; Marchenko, S.; Qin, W.; Yang, E.; Fasnacht, Z.; Joiner, J.; Choi, S.; Haffner, D.; et al. OMI/aura nitrogen dioxide standard product with improved surface and cloud treatments. *Atmos. Meas. Tech.* **2021**, *14*, 455–479.
48. Lange, K.; Richter, A.; Boesch, T.; Zilker, B.; Latsch, M.; Behrens, L.K.; Okafor, C.M.; Boesch, H.; Burrows, J.P.; Merlaud, A.; et al. Validation of GEMS tropospheric NO_2 columns and their diurnal variation with ground-based DOAS measurements. *Atmos. Meas. Tech.* **2024**, *17*, 6315–6344.
49. Kim, J.; Jeong, U.; Ahn, M.; Kim, J.H.; Park, R.J.; Lee, H. New era of air quality monitoring from space: Geostationary Environment Monitoring Spectrometer (GEMS). *Bull. Amer. Meteorol. Soc.* **2020**, *101*, E1–E22.
50. Lee, H.; Kim, D.; Hong, H.; Chang, L.S. *Geostationary Environment Monitoring Spectrometer (GEMS) User's Guide-Nitrogen Dioxide*; The National Institute of Environmental Research: Incheon, Republic of Korea, 2020; Volume 2024.
51. Lin, C.; Lau, A.K.H.; Fung, J.C.H.; Song, Y.; Li, Y.; Tao, M.; Lu, X.; Ma, J.; Lao, X.Q. Removing the effects of meteorological factors on changes in nitrogen dioxide and ozone concentrations in China from 2013 to 2020. *Sci. Total Environ.* **2021**, *793*, 11.
52. Dieudonne, E.; Ravetta, F.; Pelon, J.; Goutail, F.; Pommereau, J. Linking NO_2 surface concentration and integrated content in the urban developed atmospheric boundary layer. *Geophys. Res. Lett.* **2013**, *40*, 1247–1251.
53. Oluleye, A. Satellite Observation of Spatio-temporal Variations in Nitrogen Dioxide over West Africa and Implications for Regional Air Quality. *J. Health Pollut.* **2021**, *11*, 16.
54. Li, R.; Wang, Z.; Cui, L.; Fu, H.; Zhang, L.; Kong, L.; Chen, W.; Chen, J. Air pollution characteristics in China during 2015–2016: Spatiotemporal variations and key meteorological factors. *Sci. Total Environ.* **2019**, *648*, 902–915.
55. Hersbach, H.; Bell, B.; Berrisford, P.; Biavati, G. *ERA5 Hourly Data on Pressure Levels from 1940 to Present*; Copernicus Climate Change Service (C3S): Bonn, Germany; Climate Data Store (CDS): Bologna, Italy, 2023.
56. Fan, C.; Li, Z.; Li, Y.; Dong, J.; van der A, R.; de Leeuw, G. Variability of NO_2 concentrations over China and effect on air quality derived from satellite and ground -based observations. *Atmos. Chem. Phys.* **2021**, *21*, 7723–7748.
57. Wei, J.; Li, Z.; Wang, J.; Gupta, P.; Cribb, M. Ground-level gaseous pollutants (NO_2, SO_2, and CO) in China: Daily seamless mapping and spatiotemporal variations. *Atmos. Chem. Phys.* **2023**, *23*, 1511–1532.
58. Didan, K. *MOD13A3 MODIS/Terra Vegetation Indices Monthly L3 Global 1 km SIN Grid V006*; NASA EOSDIS Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2015; Volume 2024.
59. Laboratory, O.R.N. *LandScan Global Population Database*; Oak Ridge National Laboratory: Oak Ridge, TN, USA, 2023.
60. Center, N.T.P.D. *Tibetan Plateau 30m DEM Dataset*; National Tibetan Plateau Data Center: Beijing, China, 2023.
61. OpenStreetMap Contributors. OpenStreetMap; Available online: https://www.openstreetmap.org (accessed on 1 December 2025).
62. Li, Y.; Qin, K.; Li, D.; Fan, W. Estimation of ground-level ozone concentration based on GBRT. *Chiana Environ. Sci.* **2020**, *40*, 997–1007.
63. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

64. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the Kdd'16 22Nd Acm Sigkdd International Conference On Knowledge Discovery and Data Minning, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
65. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (Nips 2018), Montréal, QC, Canada, 3–8 December 2018; Volume 31, p. 11.
66. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 9.
67. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
68. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25Th Acm Sigkdd International Conference On Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631.
69. Bergstra, J.; Yamins, D.; Cox, D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. *Int. Conf. Mach. Learn. (PMLR)* **2013**, *28*, 115–123.
70. Falkner, S.; Klein, A.; Hutter, F. BOHB: Robust and efficient hyperparameter optimization at scale. *Int. Conf. Mach. Learn.* **2018**, *80*, 1437–1446.
71. Fisher, A.; Rudin, C.; Dominici, F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 81.
72. Zhou, W.Y.; Qin, K.; He, Q.; Wang, L.Y.; Luo, J.H.; Xie, W.L. Comparison and Optimization of Ground-Level NO_2 Concentration Estimation in China Based on TROPOMI and OMI. *Acta Opt. Sin.* **2024**, *44*, 12.
73. Long, S.; Wei, X.; Zhang, F.; Zhang, R.; Xu, J.; Wu, K.; Li, Q.; Li, W. Estimating daily ground-level NO_2 concentrations over China based on TROPOMI observations and machine learning approach. *Atmos. Environ.* **2022**, *289*, 12.
74. Vasseur, S.P.; Aznarte, J.L. Comparing quantile regression methods for probabilistic forecasting of NO_2 pollution levels. *Sci. Rep.* **2021**, *11*, 8.
75. Lu, W.; Wang, D. Ground-level ozone prediction by support vector machine approach with a cost-sensitive classification scheme. *Sci. Total Environ.* **2008**, *395*, 109–116.
76. Mariel, D.F.; Ralph, A.K.; Limbacher, J.A.; Appel, K.W.; James, A.M. Constraining chemical transport PM2.5 modeling outputs using surface monitor measurements and satellite retrievals: Application over the san joaquin valley. *Atmos. Chem. Phys.* **2018**, *18*, 12891–12913.