

Ground-Level NO₂ Surveillance from Space Across China for High Resolution Using Interpretable Spatiotemporally Weighted Artificial Intelligence

Jing Wei,*^{III} Song Liu,^{III} Zhanqing Li,*^{IV} Cheng Liu, Kai Qin, Xiong Liu, Rachel T. Pinker, Russell R. Dickerson, Jintai Lin, K. F. Boersma, Lin Sun, Runze Li, Wenhao Xue, Yuanzheng Cui, Chengxin Zhang, and Jun Wang*



Cite This: <https://doi.org/10.1021/acs.est.2c03834>



Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information

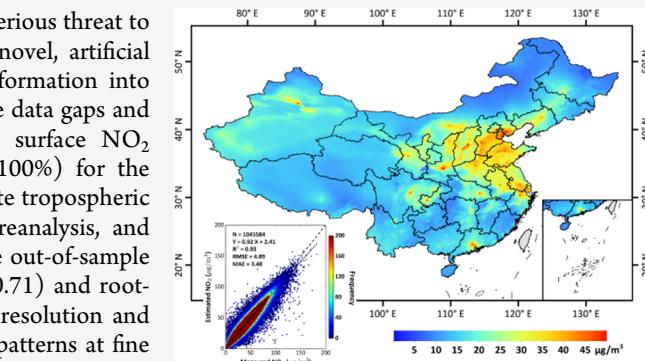
ABSTRACT: Nitrogen dioxide (NO₂) at the ground level poses a serious threat to environmental quality and public health. This study developed a novel, artificial intelligence approach by integrating spatiotemporally weighted information into the missing extra-trees and deep forest models to first fill the satellite data gaps and increase data availability by 49% and then derive daily 1 km surface NO₂ concentrations over mainland China with full spatial coverage (100%) for the period 2019–2020 by combining surface NO₂ measurements, satellite tropospheric NO₂ columns derived from TROPOMI and OMI, atmospheric reanalysis, and model simulations. Our daily surface NO₂ estimates have an average out-of-sample (out-of-city) cross-validation coefficient of determination of 0.93 (0.71) and root-mean-square error of 4.89 (9.95) $\mu\text{g}/\text{m}^3$. The daily seamless high-resolution and high-quality dataset “ChinaHighNO₂” allows us to examine spatial patterns at fine scales such as the urban–rural contrast. We observed systematic large differences between urban and rural areas (28% on average) in surface NO₂, especially in provincial capitals. Strong holiday effects were found, with average declines of 22 and 14% during the Spring Festival and the National Day in China, respectively. Unlike North America and Europe, there is little difference between weekdays and weekends (within $\pm 1 \mu\text{g}/\text{m}^3$). During the COVID-19 pandemic, surface NO₂ concentrations decreased considerably and then gradually returned to normal levels around the 72nd day after the Lunar New Year in China, which is about 3 weeks longer than the tropospheric NO₂ column, implying that the former can better represent the changes in NO_x emissions.

KEYWORDS: surface NO₂, air pollution, big data, artificial intelligence, COVID-19

1. INTRODUCTION

Nitrogen dioxide (NO₂) is one of the most important trace gases in the atmosphere, greatly impacting the ecological environment and air quality.^{1–4} It is a major pollutant near the ground and can be inhaled, posing a health threat.^{5,6} Nitrogen oxides (NO_x = NO₂ + NO) are PM_{2.5} precursors in haze and lead to the formation of surface ozone.^{7,8} NO_x comes from diverse and complex sources including fossil-fuel-fired power plants, automobile exhaust, industrial activities, biofuel, and resident cooking; the major natural sources are wildfires, soil, and lightning.⁹

Surface NO₂ levels are an important measure of air quality/pollution. Because NO_x has a short atmospheric lifetime, it is challenging to accurately quantify surface NO₂ concentrations with good spatial coverage based on sparse ground-based monitoring stations, especially in developing countries with dense populations, like China. Satellite remote sensing allows monitoring of global NO₂ distributions and variations such as with the global ozone monitoring experiment (GOME)



instrument¹⁰ and the ozone monitoring instrument (OMI).^{11,12} However, satellites can provide only the total or tropospheric NO₂ columns. Attempts have been made to convert satellite tropospheric NO₂ retrievals to ground-level concentrations using different chemical transport¹³ and statistical models. Qin et al. adopted the geographically and temporally weighted regression (GTWR) and extremely randomized trees (ERT) models to derive daily surface NO₂ from OMI tropospheric NO₂ products and meteorological data in central–eastern China.^{14,15} The daily ambient NO₂ exposure was also estimated in China from OMI tropospheric NO₂ products and environmental data by combining the

Received: May 28, 2022

Revised: June 13, 2022

Accepted: June 14, 2022

universal kriging or the K-means approaches with land-use regression (LUR)¹⁶ and different machine-learning models (e.g., random forest or RF).^{17–19} Li and Wu applied the deep learning of full residual deep networks (FSDN) to obtain the outdoor NO₂ concentrations in China by incorporating OMI tropospheric NO₂ and other covariates.²⁰

These surface NO₂ estimates were generated from OMI on the Aura satellite with spatial resolutions, 13 km × 24 km at nadir and up to ~26 km × 160 km at the edge of the swath, which cannot resolve high variations of surface NO₂ at medium or small (urban) scales.^{21,22} The Tropospheric Monitoring Instrument (TROPOMI) is onboard the copernicus sentinel-5 precursor (SSP) satellite, launched on 13 October 2017, and provides higher-spatial-resolution tropospheric NO₂ retrievals (e.g., 3.5 km × 5.5 km after August 2019). So far, only a handful of studies have employed the TROPOMI data to estimate near-surface high-resolution NO₂ concentration-adopted statistical regression (e.g., GTWR) and machine-learning [e.g., light gradient boosting machine (LightGBM) and extreme gradient boosting (XGBoost)] models.^{23–25}

Most previous studies have simply directly applied traditional methods that ignored the spatiotemporal heterogeneity of air pollution, especially machine learning that works via the single-pixel-based processing mode.²⁶ This leads to an inhomogeneous distribution of air pollution with the non-smooth transition or the discontinuity of air pollution mass in the space (and possibly yielding “salt and pepper” white noises in the data). In addition, the weaker signals of trace gases bring greater difficulties in both satellite gap filling and ground-level air pollutant estimation. It is thus necessary to consider additional factors such as model simulations and emission inventories to develop more accurate methods to improve accuracy. Here, a novel framework integrating machine and deep-learning models by involving spatiotemporally weighted characteristics of air pollution is developed to fill missing values in satellite tropospheric NO₂ products and derive daily seamless 1 km resolution ground-level NO₂ concentrations from rich big data across mainland China. Last, the application and fidelity of the dataset are demonstrated by analyzing the spatial distributions of surface NO₂ across China and their variations during statutory holidays and the COVID-19 pandemic.

2. MATERIALS AND METHODS

2.1. Big Data. Data used in our study include ground-based in situ observations, satellite remote sensing products, atmospheric reanalysis, and model simulation. Hourly ground-level NO₂ concentrations (unit: $\mu\text{g}/\text{m}^3$) measured at ~1630 monitoring stations across mainland China from 1 January 2019 to 31 December 2020 were collected from the Chinese Ministry of Environment and Ecology. Daily means were calculated from hourly observations that had undergone additional quality control measures,²⁷ i.e., filtering out invalid values and outliers caused by suspected instrument malfunction (details provided in Supporting Text 1).

Daily TROPOMI tropospheric NO₂ products (unit: $1\text{e}15 \text{ mol}/\text{cm}^2$) at a high spatial resolution of 1 km in China generated using a new algorithm and downscaled following the area-weighted method were employed.²⁸ This approach has significantly reduced the uncertainty compared to official products, especially in highly polluted areas that had been severely underestimated.^{29–31} Here, the recommended good-

quality (quality assurance value > 0.75) tropospheric vertical column NO₂ retrievals were selected. In addition, new daily OMI tropospheric NO₂ products ($0.25 \times 0.25^\circ$) reconstructed in mainland China³² were employed. NO₂ simulations ($0.75 \times 0.75^\circ$) every 1 h modeled at tropospheric and ground levels were calculated from the CAMS global reanalysis, and monthly anthropogenic NO_x emissions ($0.1 \times 0.1^\circ$) were obtained from the CAMS global emission inventory.³³

The ERA5 hourly atmospheric reanalysis product contains eight meteorological fields, i.e., boundary layer height (BLH), 2 m temperature (TEM), evaporation (ET), precipitation (PRE), relative humidity (RH), 10 m u-component of wind (WU), 10 m v-component of wind (WV), and surface pressure (SP), collected from the ERA5 hourly atmospheric reanalysis products.^{34,35} Population-related data include the 1 km landScan population distribution (POD) and 1 km visible infrared imaging radiometer suite nighttime light (NTL) products. Land-surface-related data include the moderate resolution imaging spectroradiometer (MODIS) land-use type (500 m) and normalized difference vegetation index (NDVI, 1 km) products and the shuttle radar topography mission (SRTM) digital elevation model (DEM, 90 m) products. Table S1 summarizes the detailed information on big data used in our study, and all of the auxiliary data were regridded to a uniform spatial resolution of $0.01 \times 0.01^\circ$ ($\approx 1 \text{ km} \times 1 \text{ km}$).

2.2. Methodology. We combine machine- and deep-learning models that consider the spatiotemporal heterogeneity of air pollution. The model framework includes a spatiotemporally weighted missing extra-trees (SWMET) model for filling the satellite tropospheric NO₂ gaps and a spatiotemporally weighted deep forest (SWDF) model to estimate surface NO₂.

2.2.1. Tropospheric NO₂ Gap Filling. Ubiquitous clouds in optical remote sensing images prevent tropospheric below-cloud NO₂ information from being detected due to shielding. An efficient machine-learning model, named missing extra-trees (MET),³⁶ was adopted to fill satellite data gaps. Differing from traditional methods (e.g., inverse distance weight or kriging), MET is a nonparametric spatial interpolation method that works like the missing forest³⁷ but with stronger randomizations, which can impute a dataset with missing values in multiple variables using an iterative way. It belongs to tree-based ensemble learning, which has a strong antinoise capability and is insensitive to multivariate collinearity. Here, spatiotemporal autocorrelations in satellite tropospheric NO₂ retrievals were considered in the MET model, leading to a new spatiotemporally weighted extra-trees (SWMET) model. This model was used to impute missing values in OMI and TROPOMI tropospheric NO₂ retrievals in sequence through two iterations, respectively, together with other spatiotemporally continuous auxiliary variables with potential influence (details provided in Supporting Text 2).

2.2.2. Ground-Level NO₂ Estimation. Considering the more complex and weaker relationships with tropospheric signals, deep learning (more flexible with a higher capability) was employed to estimate ground-level NO₂ concentrations. Deep forest (DF), developed by Zhou and Feng in 2017³⁸ and recently updated in 2021, was adopted. Based on the idea of deep neural networks that stack neural networks, DF stacks multilayer RFs and completely random tree forests in a cascade to obtain better feature representation and learning performance. DF can handle data of different scales without setting

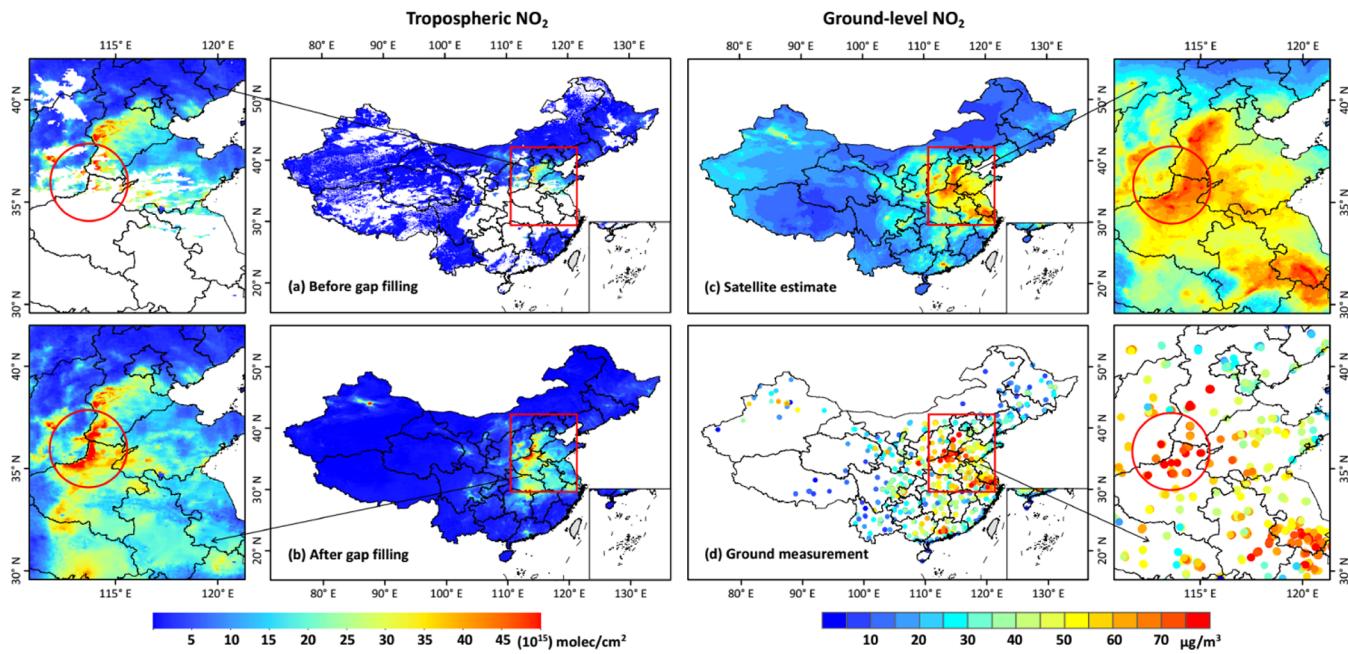


Figure 1. National and regional (zoomed-in subplots) spatial distributions of (a) original and (b) gap-filled TROPOMI tropospheric NO₂ columns (mol/cm², bottom-left legend), and (c) our model-derived and (d) ground-measured surface NO₂ concentrations (µg/m³, bottom-right legend) on 28 January 2019 in China. Red circles in the subplots outline areas of heavy pollution.

super parameters and has more competitive performance and a better physical interpretation than other “black box” deep-learning models. We also introduce a new SWDF model that incorporates the spatiotemporally weighted information into the DF model to construct a robust tropospheric-surface NO₂ conversion model involving all potential influencing factors (details provided in Supporting Text 3). Here, the model was trained and built each year separately for ground-level NO₂ estimations in China.

2.2.3. Spatiotemporal Weight Information. Temporally, air pollution can have strong seasonal cycles, and concentrations vary on the daily as well as synoptic scales, which may be similar on adjacent days, but differences increase as the time interval increases. In addition, such short-lived species also vary significantly spatially, where the farther away the two points are, the more different the polluted level is. Such differences in time and space are often not equal. Therefore, in this study, different from our previous studies that only considered the equal effects,^{39,40} spatiotemporally weighted information at different spatial points was updated and incorporated in the original artificial intelligence models to better distinguish spatiotemporal differences in air pollutants and improve their estimates.

The spatial term (P_s) is represented by the latitude (Lat) and longitude (Lon) information of one point, and the square root of inverse Haversine great-circle distances⁴¹ $\left(\frac{1}{\sqrt{D_p}}\right)$ from the point to the corners in eight directions, i.e., top-left (D_{p1}), top-middle (D_{p2}), top-right (D_{p3}), right-middle (D_{p4}), bottom-right (D_{p5}), bottom-middle (D_{p6}), bottom-left (D_{p7}), and left-middle (D_{p8}), and the center (D_{p9}) of the study area (eq 1). The temporal term (P_t) is represented by the day of the year (DOY), and the inverse distances $\left(\frac{1}{D_t}\right)$ from 1 day to the middle day of four seasons, i.e., spring equinox (21 March, D_{t1}), summer solstice (21 June, D_{t2}), autumn equinox (22

September, D_{t3}), and winter solstice (22 December, D_{t4}) (eq 2).

$$P_s = \left(\text{Lon}, \text{Lat}, \frac{1}{\sqrt{D_{p1}}}, \frac{1}{\sqrt{D_{p2}}}, \frac{1}{\sqrt{D_{p3}}}, \frac{1}{\sqrt{D_{p4}}}, \frac{1}{\sqrt{D_{p5}}}, \frac{1}{\sqrt{D_{p6}}}, \frac{1}{\sqrt{D_{p7}}}, \frac{1}{\sqrt{D_{p8}}}, \frac{1}{\sqrt{D_{p9}}} \right) \quad (1)$$

$$P_t = \left(\text{DOY}, \frac{1}{D_{t1}}, \frac{1}{D_{t2}}, \frac{1}{D_{t3}}, \frac{1}{D_{t4}} \right) \quad (2)$$

2.2.4. Validation Method. The widely used out-of-sample 10-fold cross-validation (10 CV) method, where data samples are randomly divided into 10 folds, saving one fold for testing, was selected to evaluate the overall accuracy of surface NO₂ estimates.⁴² In addition, considering the use of information from neighboring stations, a new independent out-of-city 10 CV approach, where cities are randomly divided into 10 folds, saving one fold for testing, was employed to assess the model’s predictive ability in space, i.e., surface NO₂ predictions at locations where ground-based measurements are not available.

3. RESULTS AND DISCUSSION

3.1. Model Performance. **3.1.1. Tropospheric and Surface NO₂ Results.** The OMI tropospheric NO₂ product has nearly complete spatial coverage (average = 87%) after preliminary data fusion by integrating global ozone monitoring experiment-2B (GOME-2B) NO₂ information using a reconstructed framework (Figure S1a).³² In contrast, TROPOMI tropospheric NO₂ products have a large number of missing retrievals with an average spatial coverage of only about 51%, especially in the pluvial areas of southern China (Figure S1b). Applying the SWMET model, daily 1 km tropospheric NO₂ maps with a spatial coverage increasing to

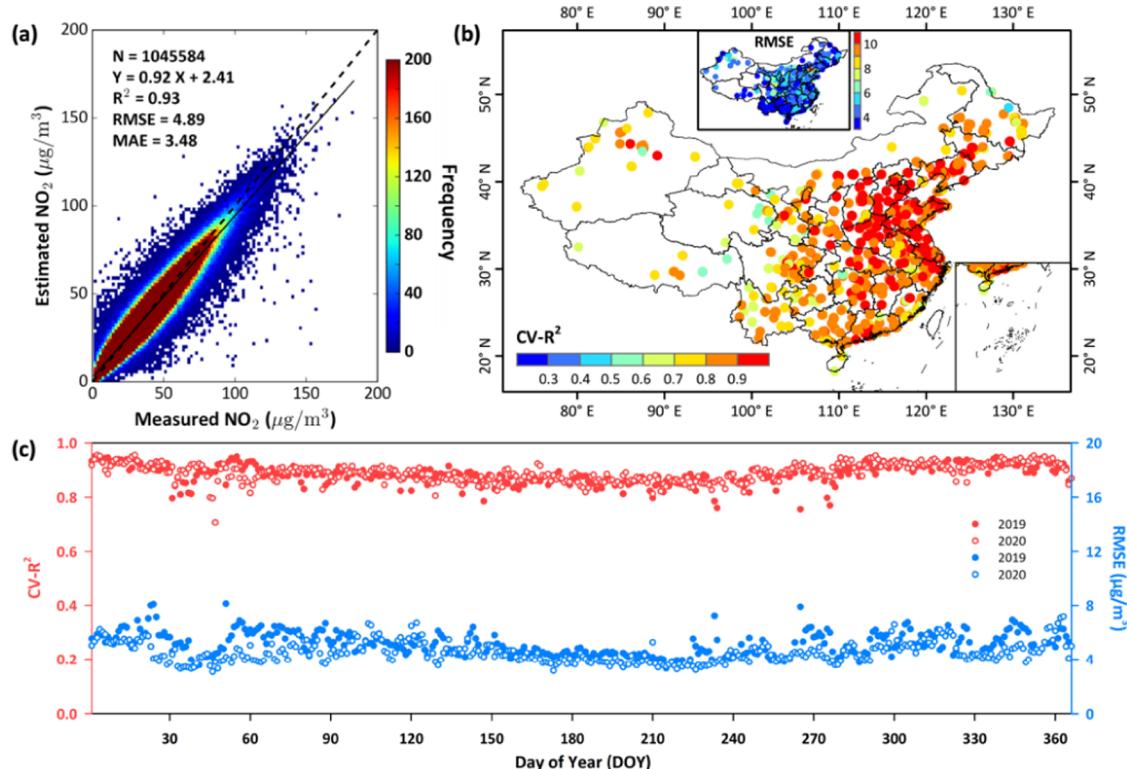


Figure 2. Out-of-sample cross-validation of daily ground-level NO₂ estimates ($\mu\text{g}/\text{m}^3$) (a) over the whole of China, (b) at each monitoring station, and (c) for each day of 2019 (filled dots) and 2020 (unfilled dots).

100% covering mainland China were generated. In general, tropospheric NO₂ information in missing areas can be well reconstructed (Figure 1a,b) even under highly polluted conditions (areas outlined by red circles).

Figure 1c shows a typical example of the spatial distribution of our model-derived ground-level NO₂ concentrations on an individual day, i.e., 28 January 2019. By filling satellite gaps, our ChinaHighNO₂ dataset can provide surface NO₂ information at any location throughout the country. Compared with ground measurements, our gap-filled dataset can also capture the spatial distributions of surface NO₂ in lightly polluted areas, e.g., northeast and southwest China (Figure 1d). More importantly, our predictions are highly consistent with observations both in spatial patterns and magnitudes over severely NO₂-polluted areas in eastern China, especially the North China Plain, where a large number of missing values existed in satellite tropospheric data (Figure 1a). This further illustrates the superior performance of our gap-filling method under polluted conditions.

3.1.2. Validation of NO₂ Estimates and Predictions. Gap-filled tropospheric NO₂ data compare well with multiaxis differential optical absorption spectroscopy (MAX-DOAS) measurements (only a slightly decreased correlation of 0.72 from 0.85 compared to nongap-filled retrievals) at the individual Xuzhou site (34.22°N , 117.14°E). However, limited by the available number of ground monitors, the same independent cross-validation method adopted in previous studies^{20,24} was selected to further validate gap-filling data. The results illustrate that our tropospheric NO₂ predictions are reliable, with the average coefficient of determination (R^2) values ranging from 0.89 to 0.96 and root-mean-square error (RMSE) values ranging from 0.46×10^{15} to $1.51 \times 10^{15} \text{ mol}/\text{cm}^2$.

For ground-level NO₂ data, we first evaluate the estimation accuracy of the developed SWDF model based on the out-of-sample CV approach at different spatiotemporal scales in China (Figure 2). Daily surface NO₂ estimates (number of estimates, $N = 1,045,584$) are highly consistent with ground measurements (CV-R² = 0.93), showing low uncertainties, with an average RMSE of $4.89 \mu\text{g}/\text{m}^3$ and mean absolute error (MAE) of $3.48 \mu\text{g}/\text{m}^3$ during 2019–2020 over the whole of China (Figure 2a). However, our surface NO₂ estimates tend to be biased low but not much (slope from linear regression = 0.92). This is mainly because the smaller number of data samples in the case of heavy pollution can affect the model training. Also, satellite tropospheric NO₂ columns under heavily polluted conditions are easily underestimated. Our model works well (e.g., CV-R² = 0.91–0.94, RMSE = 5.2–5.5 $\mu\text{g}/\text{m}^3$) in three typical urban agglomerations (Table S2). It also performs well at the individual-site scale, especially in East China (e.g., CV-R² > 0.9, RMSE < 4 $\mu\text{g}/\text{m}^3$). In general, approximately 83 and 85% of the monitoring stations across mainland China have high CV-R² values greater than 0.8 and low RMSE values of less than 6 $\mu\text{g}/\text{m}^3$, respectively (Figure 2b). The model is also relatively stable and less affected by time changes, and it can well estimate the time series of surface NO₂ concentrations on most days (Figure 2c).

We next examine the predictive ability of our model according to the out-of-city CV approach at varying spatiotemporal scales in China (Figure S2). The model can well predict daily NO₂ concentrations at locations in China where there are no ground-based measurements (i.e., CV-R² = 0.71, RMSE = $9.95 \mu\text{g}/\text{m}^3$, and MAE = $7.4 \mu\text{g}/\text{m}^3$; Figure S2a). It also shows a strong predictive ability at regional scales (e.g., CV-R² = 0.7–0.77, RMSE = 9.5 – $11.1 \mu\text{g}/\text{m}^3$; Table S2). Predictions are also reliable with small uncertainties at ~84%

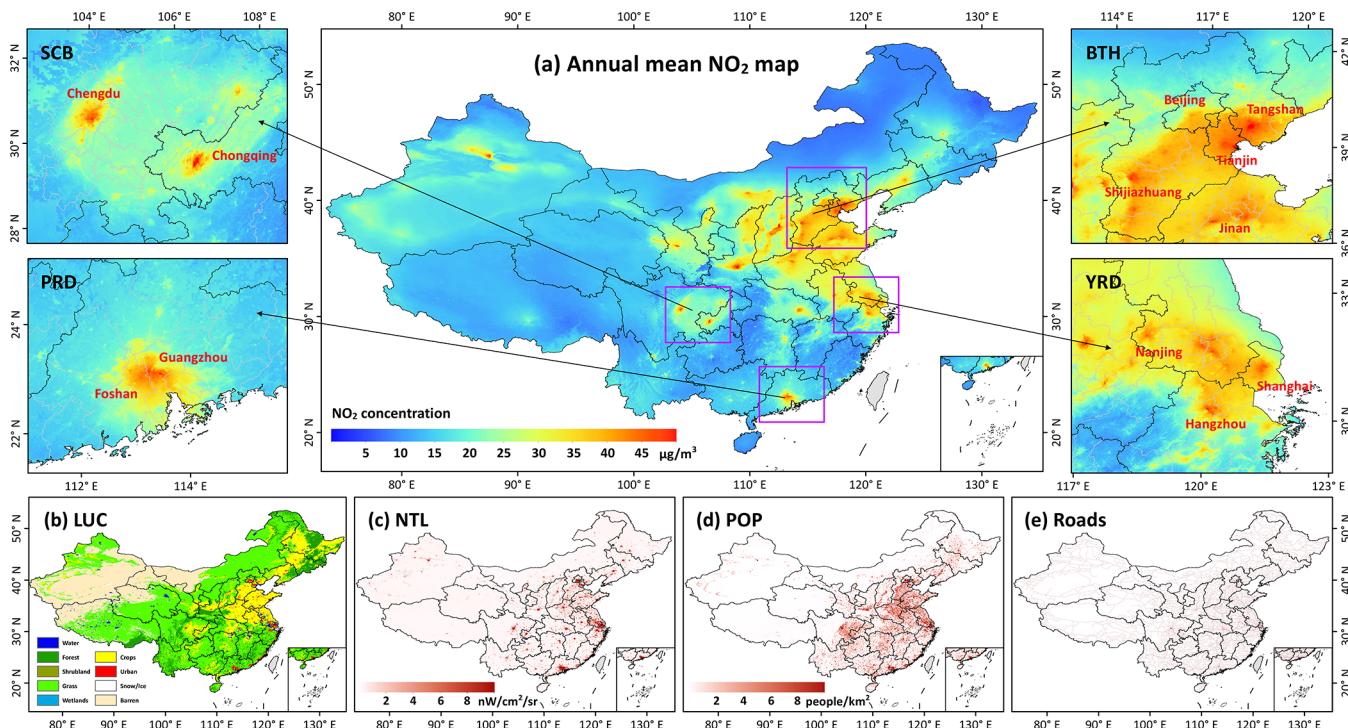


Figure 3. Spatial distributions of annual mean (a) national and regional (zoomed-in subplots) ground-level NO₂ concentrations ($\mu\text{g}/\text{m}^3$), (b) land-use cover (LUC), (c) nighttime lights (NTL), (d) population (POP), and (e) roads in China. Regions shown in panel (a) are the Sichuan Basin (SCB), Beijing–Tianjin–Hebei (BTH), the Pearl River Delta (PRD), and the Yangtze River Delta (YRD).

of the stations (i.e., $\text{CV-}R^2 > 0.5$, $\text{RMSE} < 12 \mu\text{g}/\text{m}^3$), especially those located in eastern China where the observation network has a high station density (Figure S2b). The model can capture daily variations of surface NO₂ in areas without available surface observations on most days in the years considered (Figure S2c). Compared to the estimation accuracy, the decline in predictive accuracy is mainly due to huge differences in the level of economic development among Chinese cities, especially in the eastern and western regions, which is closely related to NO_x emissions.

3.2. Spatiotemporal Characteristics of Surface NO₂.

3.2.1. ChinaHighNO₂ Dataset. Using the SWMET model, we have generated a daily, full-coverage (100%), high-resolution (1 km), and high-quality ground-level NO₂ dataset across mainland China (i.e., ChinaHighNO₂) for the years 2019 and 2020, one of the series of ChinaHighAirPollutants (CHAP) dataset. Monthly and annual surface NO₂ datasets are also synthesized by averaging daily data. These temporal surface NO₂ composites ($N = 35,496$, and 2958, respectively) agree well with ground observations ($R^2 = 0.95$ and 0.96, respectively), with low uncertainties (e.g., $\text{RMSE} = 3.11$ and $2.35 \mu\text{g}/\text{m}^3$, respectively) in China. These results further illustrate that our ChinaHighNO₂ dataset can be used to study spatiotemporal variations in surface NO₂ exposure across China.

Figure 3 shows the spatial distributions of national and regional annual mean surface NO₂ concentrations during 2019–2020, and their potential influential factors across China. The annual mean NO₂ concentration was $17.5 \pm 6.5 \mu\text{g}/\text{m}^3$ in China. In general, urban and cropland areas (red and yellow areas, respectively, in Figure 3b) corresponded to high surface NO₂ concentrations, where high pollution levels (i.e., annual NO₂ > $40 \mu\text{g}/\text{m}^3$) were mainly observed in central and eastern China regions with developed economies and

concentrated populations (Figure 3c,d), e.g., the Beijing–Tianjin–Hebei (BTH) urban agglomeration and the Yangtze River Delta (YRD). In northwest China, the distribution of surface NO₂ concentrations generally followed the distributions of populations and roads, suggesting that the sources of NO₂ were mainly from anthropogenic and transportation emissions (Figure 3d,e). By contrast, southwest and northeast China had low NO₂-pollution levels due to limited human activities in those regions. However, there are still positive biases in surface NO₂ estimates in the vast cleaner areas (e.g., the Taklimakan Desert and Tibet), mainly because oxidized nitrogen species are present, leading to uncertainties in surface measurements.^{43,44} It is also worth noting that for any machine-learning (ML), especially deep-learning (DL), method, having an adequate number of ground-based stations to provide training data is essential. Less than a certain number, the ML would not have enough samples to “learn”, which is especially the case in western China where the stations are very few, and they do not represent the vast area of uninhabited land. Fortunately, this issue does not affect our conclusions that are chiefly concerned with air quality in more populated regions in eastern China.

The ChinaHighNO₂ dataset can provide detailed surface NO₂ information at finer city levels due to its high 1 km spatial resolution. Tianjin, Tangshan, and Langfang are the three cities with the most severe surface NO₂ exposure. Most of the top 30 polluted cities are located in traditional heavy industrial areas in China, e.g., the BTH region and Shandong and Henan provinces. Large amounts of pollution gases are emitted from manufacture plants (Figure S3a).⁴⁵ In general, surface NO₂ levels are positively correlated with the level of urban economic development and population density, as proxied by the logarithm of NTL ($R = 0.64$, $p < 0.001$)²¹ and the population number ($R = 0.41$, $p < 0.001$), respectively (Figure S3b). We

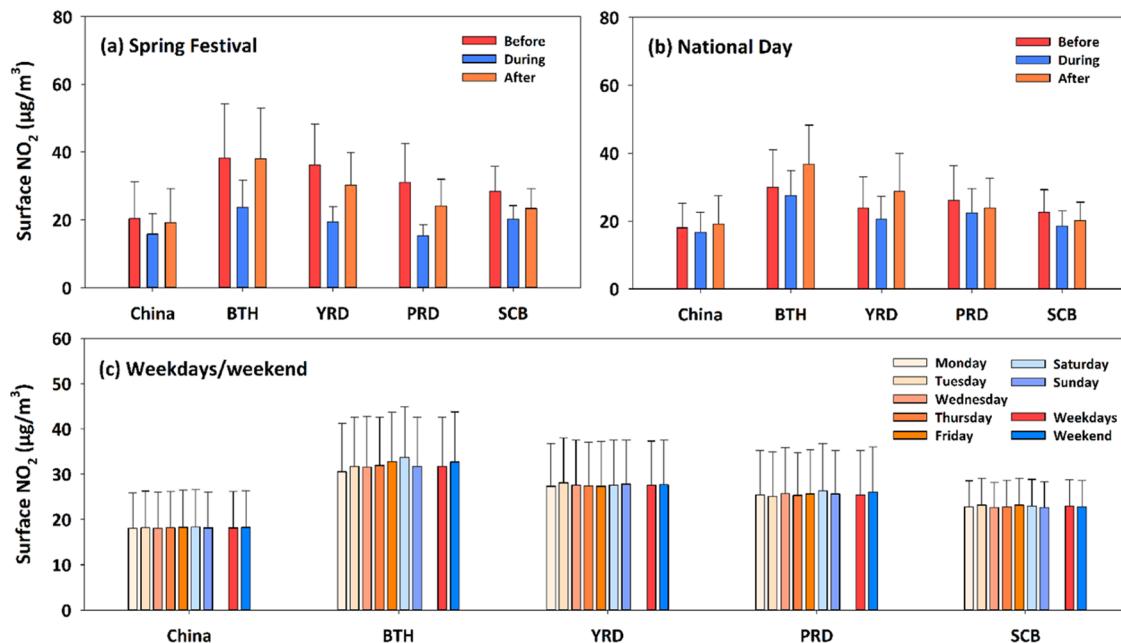


Figure 4. Comparison of average surface NO_2 concentrations ($\mu\text{g}/\text{m}^3$) before, during, and after (a) Spring Festival and (b) National Day holidays, and (c) during weekdays and the weekend in China and four typical regions. BTH, YRD, PRD, and SCB stand for Beijing–Tianjin–Hebei, Yangtze River Delta, Pearl River Delta, and Sichuan Basin, respectively.

then calculated surface NO_2 concentrations between urban and suburban areas, segmented from harmonized NTL using the stepwise-partitioning framework,⁴⁶ at all cities in China. Large urban–rural differences are seen (average relative difference = 28%), especially for the provincial capitals, accounting for more than one-third of the top 30 prefecture-level cities (Figure S4). Such differences are closely associated with the size of the urban area compared with the size of the rural area and the distribution and density of populations. For example, Urumqi and Xi'an in western China have small core urban areas located with the highest population density (Figure 3d), leading to greater urban/rural NO_2 differences.

Surface NO_2 concentrations varied significantly on a seasonal basis (Figure S5), where the highest surface NO_2 occurred in winter, with an average value of $20.9 \pm 8.9 \mu\text{g}/\text{m}^3$, especially in the BTH region ($\sim 39.2 \pm 13.0 \mu\text{g}/\text{m}^3$). The next highest surface NO_2 concentrations occurred in autumn (average = $19.3 \pm 7.4 \mu\text{g}/\text{m}^3$) followed by spring (average = $16.2 \pm 6.2 \mu\text{g}/\text{m}^3$). The peak in winter results from the combustion of coal and fossil fuels for heating in northern China, emitting a large amount of NO_x . NO_2 concentrations were lowest in summer (average = $13.8 \pm 4.1 \mu\text{g}/\text{m}^3$), especially in southern China. This was mainly due to faster chemical loss via the $\text{OH} + \text{NO}_2 + \text{M}$ reaction. Summertime and low latitudes generally have higher levels of solar radiation and OH , shortening the NO_2 lifetime.

3.2.2. Holiday and Weekly Effects. The effects of holidays in China, i.e., the Spring Festival (5–11 February 2019) and National Day (1–7 October 2019), on surface NO_2 concentrations are first investigated (Figures S6 and S7). In general, the spatial patterns over time of our model-derived surface NO_2 concentrations agree well with ground observations. Before the Spring Festival, surface NO_2 first remained high until February 1 due to normal conditions in most areas in China, especially in northern China, dominated by heavy industry. During the holiday, surface NO_2 concentrations declined rapidly in eastern China because most factories

gradually closed down over time, greatly reducing anthropogenic emissions. The lowest level of NO_2 concentration in the country occurred around 7 February. Toward the end of the holiday, factories began to reopen, and surface NO_2 levels gradually picked up. After the lantern festival (19 February), surface NO_2 concentrations returned to their normal levels.

Surface NO_2 concentrations during the National Day holiday had a similar temporal trend, i.e., high before the holiday, gradually decreasing over time, reaching a minimum value around 4 October during the holiday, then gradually increasing, returning to their normal levels on 9 October. These two general rounds of change and recovery of surface NO_2 for the two festivals lasted for about 4 and 2 weeks, respectively, which was closely related to anthropogenic emissions. However, significant decreases in NO_2 concentration before the festivals (e.g., 30 January to 1 February, 25–27 September) or the sudden fluctuations over a short period (e.g., 17–21 February, 9–13 October) may have likely been due to changes in meteorological conditions.

On the national and regional levels, surface NO_2 concentrations during the Spring Festival and National Day holidays were much lower by 22 and 14% than those before and after these holidays in China, respectively. In the BTH and YRD regions, in particular, maximum relative differences reached above 60 and 39%, respectively, illustrating strong holiday effects on surface NO_2 concentrations (Figure 4a,b). Surface NO_2 concentrations showed a slowly increasing trend on weekdays, decreasing by ~6% on Sundays in BTH. In the other three regions and over the whole of China, surface NO_2 concentrations stayed at a relatively stable level, with small, irregular changes (Figure 4c). Overall, differences in surface NO_2 concentrations in China and in each region between weekdays and the weekend were small, within $\pm 1 \mu\text{g}/\text{m}^3$. This differs from other parts of the world, such as Europe and the United States, where surface NO_2 is much higher on weekdays than on weekends.^{47–49} This is mainly attributed to differences in economic production activities, e.g., factories in China

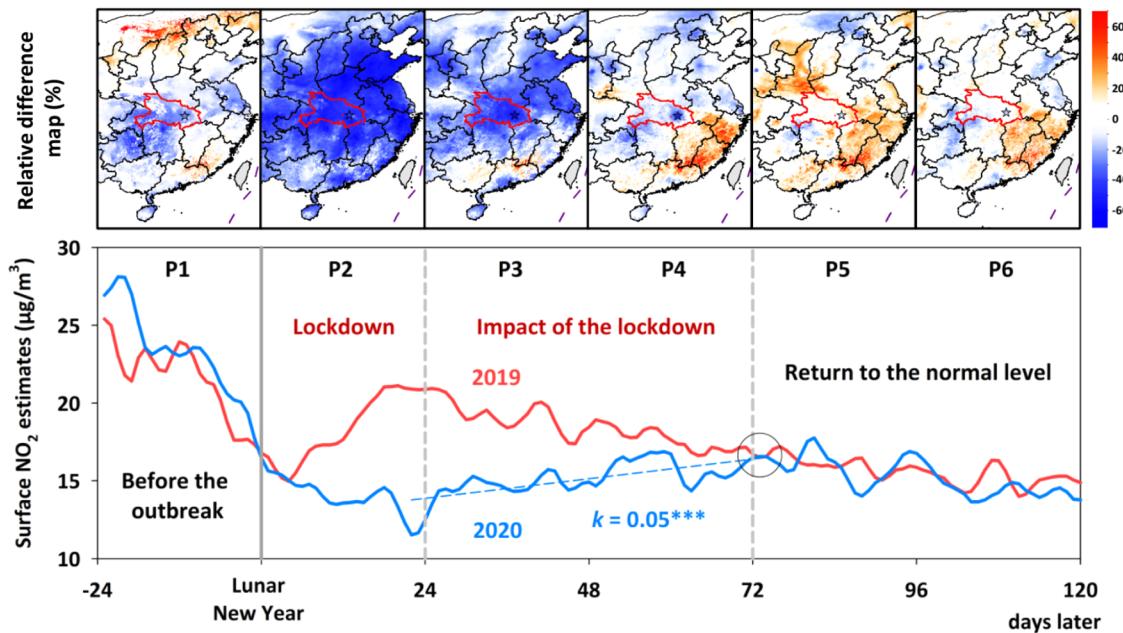


Figure 5. Time series of the 3-day moving average of daily surface NO₂ concentrations ($\mu\text{g}/\text{m}^3$) in China (bottom panel) and the relative difference (%) in surface NO₂ concentrations ($\mu\text{g}/\text{m}^3$) between 2020 and 2019 (top panel) during six periods (i.e., P1–P6) in eastern China before and after the Lunar New Year. The red border and star in the top panel indicate Hubei Province and Wuhan City, respectively. The gray circle in the bottom panel highlights when the surface-measured NO₂ concentration from 2020 reached the 2019 historical level. The dashed blue line shows the linear trend for observations during the period experiencing the impact of the lockdown in 2020. The slope (k) is given, and the three asterisks indicate $p < 0.001$.

operate on a continuous schedule, leading to continuous emissions.

Although similar results can be concluded from ground measurements on holidays (Figure S8a,b), larger amplitudes are observed in surface NO₂ changes before and after the holidays (2.8–3.7 times larger, especially over the whole of China and SCB). The main reason is that stations are unevenly distributed and mainly located in cities, reflecting NO_x emissions in urban areas. By contrast, satellite tropospheric NO₂ data capture weaker or even opposite holiday effects (Figure S9a,b). Similar weak weekday/weekend differences are observed from all data sources. With respect to NO₂ changes through the week, our results are more consistent with ground measurements, while tropospheric NO₂ column changes tend to be irregular (Figures S8c and S9c). The full coverage of our dataset makes up for the spatial and data heterogeneities inherent to surface observations and more accurately describes changes in surface NO₂ in the region.

3.2.3. Surface NO₂ Changes During the COVID-19 Pandemic. Surface NO₂ variations related to the COVID-19 epidemic,^{50–52} that broke out in Wuhan, Hubei province, China, are investigated. Figure 5 shows the time series of daily ground-level NO₂ concentrations and relative changes between 2020 and 2019 during six periods (denoted as P1–P6) before and after the Lunar New Year in eastern China. Before the outbreak (P1), surface NO₂ concentrations changed little in most areas of eastern China but began to decrease in Hubei province and surrounding areas. During the lockdown (P2), surface NO₂ concentrations dropped sharply, with relative changes greater than 60% over nearly all of eastern China. This was mainly due to the cessation of industrial production and human activities, significantly reducing NO_x emissions.^{53,54} Previous studies based on ground-based measurements and satellite estimates have also reported this phenomenon.^{51,55}

The lockdown had a strong and sustained impact on surface NO₂ concentrations (relative changes > 30%) until the 48th day after the Lunar New Year (P3). After P3, the impact gradually decreased in most areas (P4) due to increasing anthropogenic emissions as the epidemic was gradually controlled and cities lifted bans.⁵³ During P5 and P6, surface NO₂ concentrations gradually recovered to their historical levels, with relative changes within $\pm 20\%$ in most areas of East China, including severe epidemic areas (e.g., Hubei province) and even slightly higher in some low-risk areas, e.g., southeast China, indicating that human life had returned to normal. Note that there were also some significant hotspots, e.g., in Henan and Fujian provinces, in different periods. These areas are mainly forested and densely vegetated areas with little or no population (Figure 3b,d) and very low surface NO₂ levels, where small changes will lead to large relative differences.

Figure S10 compares variations in tropospheric NO₂ columns and ground NO₂ measurements during the COVID-19 pandemic. A similar temporal trend was found, namely, that NO₂ significantly decreased during the lockdown,⁵⁶ reaching a minimum value around the 22nd day after the Lunar New Year and rising steadily over time after the lockdown ended. However, noticeable differences were seen in the timing of the NO₂ recovery. Tropospheric NO₂ in 2020 first intersected with the historical level of 2019 (highlighted by the black circle in Figure S10b) on the 52nd day after the Lunar New Year (16 March 2020), with both time series generally following the same trend during P5 and P6. For surface-measured NO₂ concentrations, this intersection with the historical level of 2019 (highlighted by the black circle in Figure S10a) occurred around the 72nd day after the Lunar New Year (5 April 2020), also seen in our modeled results (Figure 5). The duration of the impact of the lockdown on the tropospheric NO₂ column was about 1.7 times shorter than

that on surface NO₂. Large differences were also found in the speed of NO₂ recovery after lockdown. The slope of the trend for observations (Figure S10a) was the steepest ($\sim 0.2 \mu\text{g}/\text{m}^3/\text{day}$, $p < 0.001$), four times that of ours (Figure 5), while that for the tropospheric NO₂ column (Figure S10b) was the least steep ($\sim 0.02 \mu\text{g}/\text{m}^3/\text{day}$, $p < 0.001$). This is mainly due to the difference in spatial representation, i.e., observations represent the change in NO₂ in urban areas, while the tropospheric NO₂ column represents the whole troposphere. Such a difference suggests that surface NO₂ can better represent changes in NO_x emissions at the local scale compared to the tropospheric NO₂ column. The potential reasons may be that NO₂ is directly related to the emission strength as a result of the relatively short lifetime (hours near the surface). Therefore, surface NO₂ is more sensitive to anthropogenic emissions, e.g., fuel combustion, urban automobile exhaust, and industrial production, often occurring at the ground level. By contrast, tropospheric NO₂ contains more information coming from complex sources that are smeared over large areas and can be further affected by convection or diffusion like transport, chemistry, and atmospheric lightning.

3.2.4. Surface NO₂-Pollution Exposure Risk. The unique advantage of the full coverage of our ChinaHighNO₂ dataset is the ability to assess the daily surface NO₂ exposure risk, i.e., exceedance of the national air quality standard (i.e., daily NO₂ = $80 \mu\text{g}/\text{m}^3$), at every location across China in 2019 and 2020, separately (Figure S11). In 2019, most areas of China met the acceptable ambient NO₂ standard year-round, except for key urban agglomerations, i.e., BTH, YRD, and the Pearl River Delta. In particular, core urban areas of the capitals and megalopolises in the main provinces (e.g., Tianjin, Shijiazhuang, Jinan, Taiyuan, Xi'an, Wuhan, Shanghai, and Foshan) had a high exposure risk, i.e., the percentage of days in 2019 not meeting the acceptable ambient NO₂ standard exceeded 10%. A similar spatial pattern was observed in 2020, the area with an ambient NO₂ exposure risk expanding into the YRD. Overall, in 2020, the probability of NO₂-pollution occurrence declined across mainland China, especially in three typical urban agglomerations, mainly due to the impact of the epidemic.

3.3. Discussion. **3.3.1. Advantage of the New Model.** Unlike traditional “black box” deep-learning models, the tree-based DF can be more physically interpretable, allowing the assessment of the importance of each input variable to model construction (Figure S12). Tropospheric NO₂ is dominant, with the highest importance score (>31%), followed by the modeled surface NO₂ (importance score = 15%). Spatial and temporal terms account for 14 and 10%, respectively, highlighting the importance of spatiotemporal information to air pollution modeling. Meteorological conditions also play key roles, especially RH and TEM, with a cumulative importance score near 20%. Variables related to the surface and population also have important impacts, contributing ~5% each.

3.3.2. Comparison of Different NO₂ Products. Only a few model studies developed focused on near-surface NO₂ in China have performed gap filling. Our model performs better than others,^{24,25} or is compatible,²⁰ in terms of overall accuracy (Table S3). Our results should be superior to traditional interpolation methods (e.g., inverse distance weighting and kriging), especially in areas with a complex terrain with rapidly varying land cover and topography.^{S2} Our gap filling is smoother with less noise compared to FSDN-filling results (e.g., Figure 8 in Li and Wu, 2021)²⁰ that neglect tropospheric

model simulations, especially spatiotemporal autocorrelations of air pollution.

OMI tropospheric NO₂ products were first used to derive ground-level NO₂ concentrations at coarse spatial resolutions from the original $\sim 0.25^\circ$ ^{15,17,18,57} to direct resampling of $\sim 0.125^\circ$ ^{16,19} or $\sim 0.1^\circ$ ¹⁴ (Table S4). Later, the resolutions of surface NO₂ estimates improved to $\sim 0.05^\circ$ ^{24,58} and $\sim 0.025^\circ$ ²⁵ using the newly launched TROPOMI satellite. Our study further improved the spatial resolution to 1 km using TROPOMI NO₂ retrievals via area-weighted downscaling, about 2.5–25 times higher than previous studies. Note that another study also generated 1 km surface NO₂ data but relied on the much coarser OMI NO₂ information as the main input.²⁰ In terms of overall accuracy, our SWDF model is superior to traditional statistical regression models (e.g., GTWR,¹⁴ LUR,¹⁶ UK&SBM,¹⁹ and GTWR-SK²⁵) and popular machine- and deep-learning models (e.g., RF-K, RF-SK,¹⁸ FSDN,²⁰ LightGBM,²⁴ BME,⁵⁷ and XGBoost^{58,59}), improving the CV-R² by 9–55% and reducing the RMSE by 23–64%. The 1 km ground-level NO₂ dataset represents a substantial improvement over existing surface NO₂ data in China, potentially valuable for nitrogen-cycle and health-related studies, especially in urban areas.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.est.2c03834>.

Additional quality control (Test S1); tropospheric NO₂ gap filling (Test S2); ground-level NO₂ estimation (Test S3); spatial coverage of available daily (a) OMI and (b) TROPOMI tropospheric NO₂ retrievals across China (Figure S1); and summary of the data sources used in this study (Table S1) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Jing Wei – Department of Chemical and Biochemical Engineering, Iowa Technology Institute, Center for Global and Regional Environmental Research, University of Iowa, Iowa City, Iowa 52242, United States; Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, Maryland 20742, United States; orcid.org/0000-0002-8803-7056; Email: weijing_rs@163.com

Zhanqing Li – Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, Maryland 20742, United States; orcid.org/0000-0001-6737-382X; Email: zli@atmos.umd.edu

Jun Wang – Department of Chemical and Biochemical Engineering, Iowa Technology Institute, Center for Global and Regional Environmental Research, University of Iowa, Iowa City, Iowa 52242, United States; orcid.org/0000-7334-0490; Email: jun-wang-1@uiowa.edu

Authors

Song Liu – School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

Cheng Liu – Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of

China, Hefei 230026, China;  orcid.org/0000-0002-3759-9219

Kai Qin – School of Environment and Geoinformatics, China University of Mining and Technology, Xuzhou 221116, China

Xiong Liu – Atomic and Molecular Physics Division, Center for Astrophysics | Harvard and Smithsonian, Cambridge, Massachusetts 02138, United States;  orcid.org/0000-0003-2939-574X

Rachel T. Pinker – Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, Maryland 20742, United States

Russell R. Dickerson – Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, Maryland 20742, United States

Jintai Lin – Laboratory for Climate and Ocean-Atmosphere Studies, Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871, China;  orcid.org/0000-0002-2362-2940

K. F. Boersma – Satellite Observations Department, Royal Netherlands Meteorological Institute, De Bilt 3731GA, the Netherlands; Meteorology and Air Quality Group, Wageningen University, Wageningen 6708PB, the Netherlands

Lin Sun – College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China

Runze Li – Department of Civil and Environmental Engineering, University of California, Irvine, California 92697, United States

Wenhao Xue – School of Economics, Qingdao University, Qingdao 266071, China

Yuanzheng Cui – College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China

Chengxin Zhang – Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230026, China

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.est.2c03834>

Author Contributions

J.W. and S.L. contributed equally to this study.

Notes

The authors declare no competing financial interest. The ChinaHighNO₂ dataset is available at <https://doi.org/10.5281/zenodo.4571660>, and the ChinHighAirPollutants (CHAP) dataset is available at <https://weijing-rs.github.io/product.html>.

ACKNOWLEDGMENTS

This work was supported by the internal funds (to J. Wang's research group) and the High Performance Computing facility in the University of Iowa. J.Wei and Z.Li were also partially supported by the NASA grant (80NSSC21K1980) and J. Wang received partial support from the NASA grant (80NSSC19K0950). K.B. was supported by the EU FP7 Project Quality Assurance for Essential Climate Variables (QA4ECV, grant no. 607405) and the Dutch Ministry of Infrastructure and the Environment (AVES-oculuS project). R.D. acknowledges the NSF support.

REFERENCES

- (1) Han, X. L.; Naehler, L. P. A review of traffic-related air pollution exposure assessment studies in the developing world. *Environ. Int.* **2006**, *32*, 106–120.
- (2) Lamsal, L. N.; Martin, R. V.; Parrish, D. D.; Krotkov, N. A. Scaling Relationship for NO₂ Pollution and Urban Population Size: A Satellite Perspective. *Environ. Sci. Technol.* **2013**, *47*, 7855–7861.
- (3) De Craemer, S.; Vercauteren, J.; Fierens, F.; Lefebvre, W.; Meysman, F. J. R. Using Large-Scale NO₂ Data from Citizen Science for Air-Quality Compliance and Policy Support. *Environ. Sci. Technol.* **2020**, *54*, 11070–11078.
- (4) Crutzen, P. A discussion of the chemistry of some minor constituents in the stratosphere and troposphere. *Pure Appl. Geophys.* **1973**, *106–108*, 1385–1399.
- (5) Chiusolo, M.; Cadum, E.; Stafoggia, M.; Galassi, C.; Berti, G.; Faustini, A.; Bisanti, L.; Vigotti, M. A.; Dassi, M. P.; Cerniglio, A.; Mallone, S.; Pacelli, B.; Minerba, S.; Simonato, L.; Forastiere, F.; EpiAir Collaborative Group. Short-Term Effects of Nitrogen Dioxide on Mortality and Susceptibility Factors in 10 Italian Cities: The EpiAir Study. *Environ. Health Perspect.* **2011**, *119*, 1233–1238.
- (6) McConnell, R.; Islam, T.; Shankardass, K.; Jerrett, M.; Lurmann, F.; Gilliland, F.; Gauderman, J.; Avol, E.; Kunzli, N.; Yao, L.; Peters, J.; Berhane, K. Childhood Incident Asthma and Traffic-Related Air Pollution at Home and School. *Environ. Health Perspect.* **2010**, *118*, 1021–1026.
- (7) 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. U.S.A.* **2019**, *116*, 422–427.
- (8) Qu, Z.; Jacob, D. J.; Silvern, R. F.; Shah, V.; Campbell, P. C.; Valin, L. C.; Murray, L. T. US COVID-19 Shutdown Demonstrates Importance of Background NO₂ in Inferring NO_x Emissions From Satellite NO₂ Observations. *Geophys. Res. Lett.* **2021**, *48*, No. e2021GL092783.
- (9) Lin, J. T. Satellite constraint for emissions of nitrogen oxides from anthropogenic, lightning and soil sources over East China on a high-resolution grid. *Atmos. Chem. Phys.* **2012**, *12*, 2881–2898.
- (10) Burrows, J. P.; Weber, M.; Buchwitz, M.; Rozanov, V.; Ladstätter-Weißenmayer, A.; Richter, A.; DeBeek, R.; Hoogen, R.; Bramstedt, K.; Eichmann, K.-U.; Eisinger, M.; Perner, D. The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results. *J. Atmos. Sci.* **1999**, *56*, 151–175.
- (11) Boersma, K. F.; Eskes, H. J.; Dirksen, R. J.; van der A, R. J.; Veefkind, J. P.; Stammes, P.; Huijnen, V.; Kleipool, Q. L.; Sneep, M.; Claas, J. An improved tropospheric NO₂ column retrieval algorithm for the Ozone Monitoring Instrument. *Atmos. Meas. Tech.* **2011**, *4*, 1905–1928.
- (12) Wang, Y.; Wang, J. Tropospheric SO₂ and NO₂ in 2012–2018: Contrasting views of two sensors (OMI and OMPS) from space. *Atmos. Environ.* **2020**, *223*, No. 117214.
- (13) 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.* **2008**, *113*, No. D16308.
- (14) Qin, K.; Rao, L.; Xu, J.; Bai, Y.; Zou, J.; Hao, N.; Li, S.; Yu, C. Estimating Ground Level NO₂ Concentrations over Central-Eastern China Using a Satellite-Based Geographically and Temporally Weighted Regression Model. *Remote Sens.* **2017**, *9*, 950.
- (15) Qin, K.; Han, X.; Li, D. H.; Xu, J.; Loyola, D.; Xue, Y.; Zhou, X. R.; Li, D.; Zhang, K. F.; Yuan, L. M. Satellite-based estimation of surface NO₂ concentrations over east-central China: A comparison of POMINO and OMNO2d data. *Atmos. Environ.* **2020**, *224*, No. 117322.
- (16) Xu, H.; Bechle, M. J.; Wang, M.; Szapiro, A. A.; Vedal, S.; Bai, Y. Q.; Marshall, J. D. National PM_{2.5} and NO₂ exposure models for China based on land use regression, satellite measurements, and universal kriging. *Sci. Total Environ.* **2019**, *655*, 423–433.
- (17) Dou, X.; Liao, C.; Wang, H.; Huang, Y.; Tu, Y.; Huang, X.; Peng, Y.; Zhu, B.; Tan, J.; Deng, Z.; Wu, N.; Sun, T.; Ke, P.; Liu, Z.

- Estimates of daily ground-level NO₂ concentrations in China based on Random Forest model integrated K-means. *Adv. Appl. Energy* **2021**, *2*, No. 100017.
- (18) Zhan, Y.; Luo, Y. Z.; Deng, X. F.; Zhang, K. S.; Zhang, M. H.; Grieneisen, M. L.; Di, B. F. Satellite-Based Estimates of Daily NO₂ Exposure in China Using Hybrid Random Forest and Spatiotemporal Kriging Model. *Environ. Sci. Technol.* **2018**, *52*, 4180–4189.
- (19) Chen, Z. Y.; Zhang, R.; Zhang, T. H.; Ou, C. Q.; Guo, Y. M. A kriging-calibrated machine learning method for estimating daily ground-level NO₂ in mainland China. *Sci. Total Environ.* **2019**, *690*, 556–564.
- (20) Li, L. F.; Wu, J. J. Spatiotemporal estimation of satellite-borne and ground-level NO₂ using full residual deep networks. *Remote Sens. Environ.* **2021**, *254*, No. 112257.
- (21) Wang, Y.; Wang, J.; Zhou, M.; Henze, D. K.; Ge, C.; Wang, W. Inverse modeling of SO₂ and NO_x emissions over China using multisensor satellite data - Part 2: Downscaling techniques for air quality analysis and forecasts. *Atmos. Chem. Phys.* **2020**, *20*, 6651–6670.
- (22) Kong, H.; Lin, J.; Chen, L.; Zhang, Y.; Yan, Y.; Liu, M.; Ni, R.; Liu, Z.; Weng, H. Considerable Unaccounted Local Sources of NO_x Emissions in China Revealed from Satellite. *Environ. Sci. Technol.* **2022**, *56*, 7131–7142.
- (23) 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*, No. 104013.
- (24) Wang, Y.; Yuan, Q. Q.; Li, T. W.; Zhu, L. Y.; Zhang, L. P. Estimating daily full-coverage near surface O₃, CO, and NO₂ concentrations at a high spatial resolution over China based on SSP-TROPOMI and GEOS-FP. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 311–325.
- (25) Wu, S. S.; Huang, B.; Wang, J. H.; He, L. J.; Wang, Z. Y.; Yan, Z.; Lao, X. Q.; Zhang, F.; Liu, R. Y.; Du, Z. H. Spatiotemporal mapping and assessment of daily ground NO₂ concentrations in China using high-resolution TROPOMI retrievals. *Environ. Pollut.* **2021**, *273*, No. 116456.
- (26) Wei, J.; Huang, W.; Li, Z. Q.; Sun, L.; Zhu, X. L.; Yuan, Q. Q.; Liu, L.; Cribb, M. Cloud detection for Landsat imagery by combining the random forest and superpixels extracted via energy-driven sampling segmentation approaches. *Remote Sens. Environ.* **2020**, *248*, No. 112005.
- (27) Wei, J.; Huang, W.; Li, Z. Q.; Xue, W. H.; Peng, Y. R.; Sun, L.; Cribb, M. Estimating 1-km-resolution PM_{2.5} concentrations across China using the space-time random forest approach. *Remote Sens. Environ.* **2019**, *231*, No. 111221.
- (28) Zhang, C. X.; Liu, C.; Chan, K. L.; Hu, Q. H.; Liu, H. R.; Li, B.; Xing, C. Z.; Tan, W.; Zhou, H. J.; Si, F. Q.; Liu, J. G. First observation of tropospheric nitrogen dioxide from the Environmental Trace Gases Monitoring Instrument onboard the GaoFen-5 satellite. *Light: Sci. Appl.* **2020**, *9*, No. 66.
- (29) Liu, M.; Lin, J.; Kong, H.; Boersma, K. F.; Eskes, H.; Kanaya, Y.; He, Q.; Tian, X.; Qin, K.; Xie, P. H.; Spurr, R. A new TROPOMI product for tropospheric NO₂ columns over East Asia with explicit aerosol corrections. *Atmos. Meas. Tech.* **2020**, *13*, 4247–4259.
- (30) Tan, W.; Liu, C.; Wang, S.; Xing, C.; Su, W.; Zhang, C.; Xia, C.; Liu, H.; Cai, Z.; Liu, J. Tropospheric NO₂, SO₂, and HCHO over the East China Sea, using ship-based MAX-DOAS observations and comparison with OMI and OMPS satellite data. *Atmos. Chem. Phys.* **2018**, *18*, 15387–15402.
- (31) van Geffen, J.; Boersma, K. F.; Eskes, H.; Sneep, M.; ter Linden, M.; Zara, M.; Veefkind, J. P. SSP TROPOMI NO₂ slant column retrieval: method, stability, uncertainties and comparisons with OMI. *Atmos. Meas. Tech.* **2020**, *13*, 1315–1335.
- (32) He, Q.; Qin, K.; Cohen, J. B.; Loyola, D.; Li, D.; Shi, J. C.; Xue, Y. Spatially and temporally coherent reconstruction of tropospheric NO₂ over China combining OMI and GOME-2B measurements. *Environ. Res. Lett.* **2020**, *15*, No. 125011.
- (33) Crippa, M.; Guizzardi, D.; Muntean, M.; Schaaf, E.; Dentener, F.; van Aardenne, J. A.; Monni, S.; Doering, U.; Olivier, J. G. J.; Pagliari, V.; Janssens-Maenhout, G. Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2. *Earth Syst. Sci. Data* **2018**, *10*, 1987–2013.
- (34) Muñoz-Sabater, J.; Dutra, E.; Agusti-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; Martens, B.; Miralles, D. G.; Piles, M.; Rodriguez-Fernandez, N. J.; Zsoter, E.; Buontempo, C.; Thepaut, J. N. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383.
- (35) Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horanyi, A.; Munoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; Simmons, A.; Soci, C.; Abdalla, S.; Abellán, X.; Balsamo, G.; Bechtold, P.; Biavati, G.; Bidlot, J.; Bonavita, M.; De Chiara, G.; Dahlgren, P.; Dee, D.; Diamantakis, M.; Dragani, R.; Flemming, J.; Forbes, R.; Fuentes, M.; Geer, A.; Haimberger, L.; Healy, S.; Hogan, R. J.; Holm, E.; Janiskova, M.; Keeley, S.; Laloyaux, P.; Lopez, P.; Lupu, C.; Radnoti, G.; de Rosnay, P.; Rozum, I.; Vamborg, F.; Villaume, S.; Thepaut, J. N. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049.
- (36) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42.
- (37) Stekhoven, D. J.; Bühlmann, P. MissForest-non-parametric missing value imputation for mixed-type data. *Bioinformatics* **2012**, *28*, 112–118.
- (38) Zhu, Z. H.; Feng, J. Deep Forest: Towards an Alternative to Deep Neural Networks *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2021, pp 3553–3559.
- (39) Wei, J.; Li, Z.; Xue, W.; Sun, L.; Fan, T.; Liu, L.; Su, T.; Cribb, M. The ChinaHighPM10 dataset: generation, validation, and spatiotemporal variations from 2015 to 2019 across China. *Environ. Int.* **2021**, *146*, No. 106290.
- (40) Wei, J.; Li, Z. Q.; Lyapustin, A.; Sun, L.; Peng, Y. R.; Xue, W. H.; Su, T. N.; Cribb, M. Reconstructing 1-km-resolution high-quality PM_{2.5} data records from 2000 to 2018 in China: spatiotemporal variations and policy implications. *Remote Sens. Environ.* **2021**, *252*, No. 112136.
- (41) Wei, J.; Li, Z.; Cribb, M.; Huang, W.; Xulanterne, W.; Sun, L.; Guo, J.; Peng, Y.; Li, J.; Lyapustin, A.; Liu, L. Improved 1 km resolution PM_{2.5} estimates across China using enhanced space-time extremely randomized trees. *Atmos. Chem. Phys.* **2020**, *20*, 3273–3289.
- (42) Rodriguez, J. D.; Perez, A.; Lozano, J. A. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 569–575.
- (43) Liu, M. Y.; Lin, J. T.; Wang, Y. C.; Sun, Y.; Zheng, B.; Shao, J. Y.; Chen, L. L.; Zheng, Y. X.; Chen, J. X.; Fu, T. M.; Yan, Y. Y.; Zhang, Q.; Wu, Z. H. Spatiotemporal variability of NO₂ and PM_{2.5} over Eastern China: observational and model analyses with a novel statistical method. *Atmos. Chem. Phys.* **2018**, *18*, 12933–12952.
- (44) Dickerson, R. R.; Anderson, D. C.; Ren, X. R. On the use of data from commercial NO_x analyzers for air pollution studies. *Atmos. Environ.* **2019**, *214*, No. 116873.
- (45) Benish, S. E.; He, H.; Ren, X.; Roberts, S. J.; Salawitch, R. J.; Li, Z.; Wang, F.; Wang, Y.; Zhang, F.; Shao, M.; Lu, S.; Dickerson, R. R. Measurement report: Aircraft observations of ozone, nitrogen oxides, and volatile organic compounds over Hebei Province, China. *Atmos. Chem. Phys.* **2020**, *20*, 14523–14545.
- (46) Zhao, M.; Cheng, C.; Zhou, Y.; Li, X.; Shen, S.; Song, C. A global dataset of annual urban extents (1992–2020) from harmonized nighttime lights. *Earth Syst. Sci. Data* **2022**, *14*, 517–534.
- (47) Boersma, K. F.; Jacob, D. J.; Trainic, M.; Rudich, Y.; DeSmedt, I.; Dirksen, R.; Eskes, H. J. Validation of urban NO₂ concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities. *Atmos. Chem. Phys.* **2009**, *9*, 3867–3879.
- (48) Streets, D. G.; Canty, T.; Carmichael, G. R.; de Foy, B.; Dickerson, R. R.; Duncan, B. N.; Edwards, D. P.; Haynes, J. A.

Henze, D. K.; Houyoux, M. R.; Jacobi, D. J.; Krotkov, N. A.; Lamsal, L. N.; Liu, Y.; Lu, Z. F.; Martini, R. V.; Pfister, G. G.; Pinder, R. W.; Salawitch, R. J.; Wecht, K. J. Emissions estimation from satellite retrievals: A review of current capability. *Atmos. Environ.* **2013**, *77*, 1011–1042.

(49) Krotkov, N. A.; McLinden, C. A.; Li, C.; Lamsal, L. N.; Celarier, E. A.; Marchenko, S. V.; Swartz, W. H. Aura OMI observations of regional SO₂ and NO₂ pollution changes from 2005 to 2015. *Atmos. Chem. Phys.* **2016**, *16*, 4605–4629.

(50) Zu, Z. Y.; Jiang, M. D.; Xu, P. P.; Chen, W.; Ni, Q. Q.; Lu, G. M.; Zhang, L. J. Coronavirus Disease 2019 (COVID-19) A Perspective from China. *Radiology* **2020**, *296*, E15–E25.

(51) Su, T. N.; Li, Z. Q.; Zheng, Y. T.; Luan, Q. Z.; Guo, J. P. Abnormally Shallow Boundary Layer Associated With Severe Air Pollution During the COVID-19 Lockdown in China. *Geophys. Res. Lett.* **2020**, *47*, No. e2020GL090041.

(52) Wei, J.; Li, Z.; Li, K.; Dickerson, R. R.; Pinker, R. T.; Wang, J.; Liu, X.; Sun, L.; Xue, W.; Cribb, M. Full-coverage mapping and spatiotemporal variations of ground-level ozone (O₃) pollution from 2013 to 2020 across China. *Remote Sens. Environ.* **2022**, *270*, No. 112775.

(53) Ding, J.; van der A, R. J.; Eskes, H. J.; Mijling, B.; Stavrakou, T.; van Geffen, J. H. G. M.; Veefkind, J. P. NO_x Emissions Reduction and Rebound in China Due to the COVID-19 Crisis. *Geophys. Res. Lett.* **2020**, *47*, No. e2020GL089912.

(54) Feng, S. Z.; Jiang, F.; Wang, H. M.; Wang, H. K.; Ju, W. M.; Shen, Y.; Zheng, Y. H.; Wu, Z.; Ding, A. J. NO_x Emission Changes Over China During the COVID-19 Epidemic Inferred From Surface NO₂ Observations. *Geophys. Res. Lett.* **2020**, *47*, No. e2020GL090080.

(55) Cooper, M. J.; Martin, R. V.; Hammer, M. S.; Levelt, P. F.; Veefkind, P.; Lamsal, L. N.; Krotkov, N. A.; Brook, J. R.; McLinden, C. A. Global fine-scale changes in ambient NO₂ during COVID-19 lockdowns. *Nature* **2022**, *601*, 380–387.

(56) Liu, F.; Page, A.; Strode, S. A.; Yoshida, Y.; Choi, S.; Zheng, B.; Lamsal, L. N.; Li, C.; Krotkov, N. A.; Eskes, H.; van der A, R.; Veefkind, P.; Levelt, P. F.; Hauser, O. P.; Joiner, J. Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak of COVID-19. *Sci. Adv.* **2020**, *6*, No. eabc2992.

(57) Jiang, Q. T.; Christakos, G. Space-time mapping of ground-level PM_{2.5} and NO₂ concentrations in heavily polluted northern China during winter using the Bayesian maximum entropy technique with satellite data. *Air Qual. Atmos. Health* **2018**, *11*, 23–33.

(58) Liu, J. J. Mapping high resolution national daily NO₂ exposure across mainland China using an ensemble algorithm. *Environ. Pollut.* **2021**, *279*, No. 116932.

(59) Chi, Y. L.; Fan, M.; Zhao, C. A. F.; Yang, Y. K.; Fan, H.; Yang, X. C. A.; Yang, J.; Tao, J. H. Machine learning-based estimation of ground-level NO₂ concentrations over China. *Sci. Total Environ.* **2022**, *807*, No. 150721.