



# Assessment of NO<sub>2</sub> population exposure from 2005 to 2020 in China

Zhongyu Huang<sup>1,2</sup> · Xiankang Xu<sup>1,2</sup> · Mingguo Ma<sup>1,2</sup> · Jingwei Shen<sup>1,2</sup>

Received: 28 February 2022 / Accepted: 8 June 2022 / Published online: 17 June 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Nitrogen dioxide (NO<sub>2</sub>) is a major air pollutant with serious environmental and human health impacts. A random forest model was developed to estimate ground-level NO<sub>2</sub> concentrations in China at a monthly time scale based on ground-level observed NO<sub>2</sub> concentrations, tropospheric NO<sub>2</sub> column concentration data from the Ozone Monitoring Instrument (OMI), and meteorological covariates (the MAE, RMSE, and  $R^2$  of the model were 4.16 µg/m<sup>3</sup>, 5.79 µg/m<sup>3</sup>, and 0.79, respectively, and the MAE, RMSE, and  $R^2$  of the cross-validation were 4.3 µg/m<sup>3</sup>, 5.82 µg/m<sup>3</sup>, and 0.77, respectively). On this basis, this article analyzed the spatial and temporal variation in NO<sub>2</sub> population exposure in China from 2005 to 2020, which effectively filled the gap in the long-term NO<sub>2</sub> population exposure assessment in China. NO<sub>2</sub> population exposure over China has significant spatial aggregation, with high values mainly distributed in large urban clusters in the north, east, south, and provincial capitals in the west. The NO<sub>2</sub> population exposure in China shows a continuous increasing trend before 2012 and a continuous decreasing trend after 2012. The change in NO<sub>2</sub> population exposure in western and southern cities is more influenced by population density compared to northern cities. NO<sub>2</sub> pollution in China has substantially improved from 2013 to 2020, but Urumqi, Lanzhou, and Chengdu still maintain high NO<sub>2</sub> population exposure. In these cities, the Environmental Protection Agency (EPA) could reduce NO<sub>2</sub> population exposure through more monitoring instruments and limiting factory emissions.

**Keywords** Ground-level NO<sub>2</sub> concentration · Random forest model · Long-term NO<sub>2</sub> population exposure assessment · Trend analysis · Spatial and temporal variation

## Introduction

Nitrogen dioxide (NO<sub>2</sub>) plays an important role in free radical chemistry and in photochemical processes in the troposphere and stratosphere (Crutzen 1979) and can generate ozone and fine particulate matter through complex physicochemical processes (Bidleman 1988; Odum et al. 1996; Pankow 1987). The products of these complex processes, as well as NO<sub>2</sub> itself, can have a profound impact on the global environment (Altshuller and Bufalini 1971; Atkinson 2000). In addition to its environmental impact,

NO<sub>2</sub> can also enter the human body and diffuse through the alveoli and pulmonary capillaries to all organs of the respiratory system. The health effects of NO<sub>2</sub> have been studied by many researchers, including the classification of NO<sub>2</sub> toxicity (Anyanwu 1999), the complex association with various diseases (Achakulwisut et al. 2019; Hu et al. 2020; Li et al. 2020; Niu et al. 2021; Zhang et al. 2021), and premature death caused by NO<sub>2</sub> (Chen et al. 2018; Crouse et al. 2015; He et al. 2020; Hu et al. 2021; Jerrett et al. 2013; Liu et al. 2017; Nie et al. 2021). These studies demonstrate that NO<sub>2</sub> has important effects on human health. Therefore, it is essential to monitor the NO<sub>2</sub> concentration and NO<sub>2</sub> concentration trend.

Currently, the main NO<sub>2</sub> concentration monitoring approach includes ground station and satellite remote sensing monitoring. Station monitoring has high accuracy but a small monitoring range, and there is a high uncertainty in assessing the pollution level over a large area, especially for areas far from ground stations (Boersma et al. 2008). In contrast, the near real-time continuous, large-scale area characteristics of remote sensing monitoring largely compensate

Responsible Editor: Marcus Schulz

✉ Jingwei Shen  
sjwgis@swu.edu.cn

<sup>1</sup> Chongqing Jinpo Mountain Karst Ecosystem National Observation and Research Station, School of Geographical Sciences, Southwest University, Chongqing 400715, China

<sup>2</sup> Chongqing Engineering Research Center for Remote Sensing Big Data Application, School of Geographical Sciences, Southwest University, Chongqing 400715, China

for the shortcomings of station monitoring (Fishman et al. 2008; Martin 2008), providing a reliable way to measure NO<sub>2</sub> atmospheric concentrations (Bechle et al. 2013; Cheng et al. 2019; Ialongo et al. 2020; Krotkov et al. 2016; Penn and Holloway 2020).

Satellite monitoring can obtain NO<sub>2</sub> atmospheric column concentrations, but ground-level NO<sub>2</sub> concentrations are more relevant to the environment and human health. Therefore, many researchers have tried to establish a mathematical model between the NO<sub>2</sub> atmospheric column and ground-level NO<sub>2</sub> concentrations and then use the NO<sub>2</sub> atmospheric column concentration to retrieve the ground-level NO<sub>2</sub> concentration (Araki et al. 2018; Gu et al. 2017; Larkin et al. 2017; Liu 2021; Xu et al. 2019; Zhan et al. 2018; Wong et al. 2021). Larkin et al. (2017) attempted to develop a land use regression model for estimating global NO<sub>2</sub> concentrations, but the accuracy of the model differed significantly in different regions and had limited applicability. The accuracy of the retrieval model for a single country or region has been improved compared to the global model (Araki et al. 2018; He et al. 2019; Silibello et al. 2021). Many researchers have constructed retrieval models of ground-level NO<sub>2</sub> concentrations in China, which have different spatial scales, such as regional and national scales, as well as different temporal scales, such as daily, monthly, and annual concentrations (Chi et al. 2021, 2022; Liu 2021; Qin et al. 2020, 2017; Wu et al. 2021; Xu et al. 2019). However, the majority of these studies retrieved ground-level NO<sub>2</sub> concentrations from 1 year or a certain number of years, without long-term research on ground-level NO<sub>2</sub> concentrations. Additionally, they mostly focus on the changes in NO<sub>2</sub> concentrations, with fewer studies involving the assessment of NO<sub>2</sub> population exposure.

Traditional pollutant exposure assessments generally interpolate air quality station data spatially to represent regional pollutant concentrations (Fridell et al. 2014), and the study areas are mostly at urban or small regional scales (Fenech and Aquilina 2021; Ramacher and Karl 2020). The estimation of ground-level NO<sub>2</sub> concentrations using satellite remote sensing data provides important support for large-scale NO<sub>2</sub> population exposure assessments. Silibello et al. (2021) used chemical transport models and machine learning to analyze NO<sub>2</sub> population exposure in Italy from 2013 to 2015, and a combination of both models reduced NO<sub>2</sub> concentration underestimation, which provides data support for environmental epidemiological studies. Zhan et al. (2018) conducted a study on NO<sub>2</sub> population exposure in China from 2013 to 2016. The study found that approximately a quarter of the population was exposed to NO<sub>2</sub> pollution and that urbanization exacerbated NO<sub>2</sub> pollution. Qin et al. (2017) studied NO<sub>2</sub> exposure levels in China from 2013 to 2014 and found that NO<sub>2</sub> exposure levels were significantly higher in densely populated areas than in other areas. Previous studies often used

residential addresses to replace population distribution at small regional scales (Fridell et al. 2014; Im et al. 2018), while census data for large regions are often discontinuous in time, which becomes a major limitation for large-scale population exposure assessments (Jerrett et al. 2005). However, with the release of multiple population time series products, this limitation is well remedied, and the products can support NO<sub>2</sub> population exposure assessments over long periods of time (Tatem 2017).

In recent years, China has undergone rapid industrialization and urbanization, but the air pollution problems associated with the development process are also very serious (Li and Zhang 2014). NO<sub>2</sub> is one of the major air pollutants in China and has a significant impact on human health and the environment, so it is essential to research the variation in ground-level NO<sub>2</sub> concentration and population exposure at the national scale (Xu et al. 2019; Yang et al. 2017). Currently, most studies related to NO<sub>2</sub> population exposure are for small areas or short periods, and there is a lack of studies on NO<sub>2</sub> population exposure in China over long periods of time. Therefore, this study aimed to conduct a long time series study on NO<sub>2</sub> population exposure in China. First, we used ground-based monitoring data, tropospheric NO<sub>2</sub> column concentration data from the Ozone Monitoring Instrument (OMI), and meteorological data to build a random forest model for estimating ground-level NO<sub>2</sub> concentrations and then assess NO<sub>2</sub> population exposure in China from 2005 to 2020 to analyze the trend and persistence of population exposure. This will fill the gap of long time series NO<sub>2</sub> population exposure assessments in China.

The remainder of this paper is organized as follows. The “**Study area and data sources**” section describes the study area and dataset. The “**Method**” section introduces the random forest model and statistical methods. The “**Result**” section introduces the results of ground-level NO<sub>2</sub> concentrations from the random forest, analyzes the changes in ground-level NO<sub>2</sub> concentrations and population exposure over multiple years, and discusses the trends in NO<sub>2</sub> population exposure and the persistence of changes. The “**Discussion**” section discusses the causes of variation in NO<sub>2</sub> concentrations and population exposure, comparing multiple models, and the “**Conclusion**” section summarizes the main findings.

## Study area and data sources

### Study area

In this study, the land area of China was taken as the study area, with a latitude range of 7 ~ 53°N and a longitude range of 72 ~ 136°E. The terrain of this region is

high in the west and low in the east. China is rich in land surface types, including basins, mountains, hills, plains, and plateaus. Additionally, China contains five climatic zones: cold temperate, middle temperate, warm temperate, subtropical, and tropical, with diverse climate types and geographical environments. In recent decades, China's industrialization has accelerated, and its economy has developed rapidly, which has led to more serious environmental problems. Figure 1a shows the distribution of major cities in China, and Fig. 1b shows the distribution of ground-level air quality monitoring stations in China in 2020.

### Ground NO<sub>2</sub> data

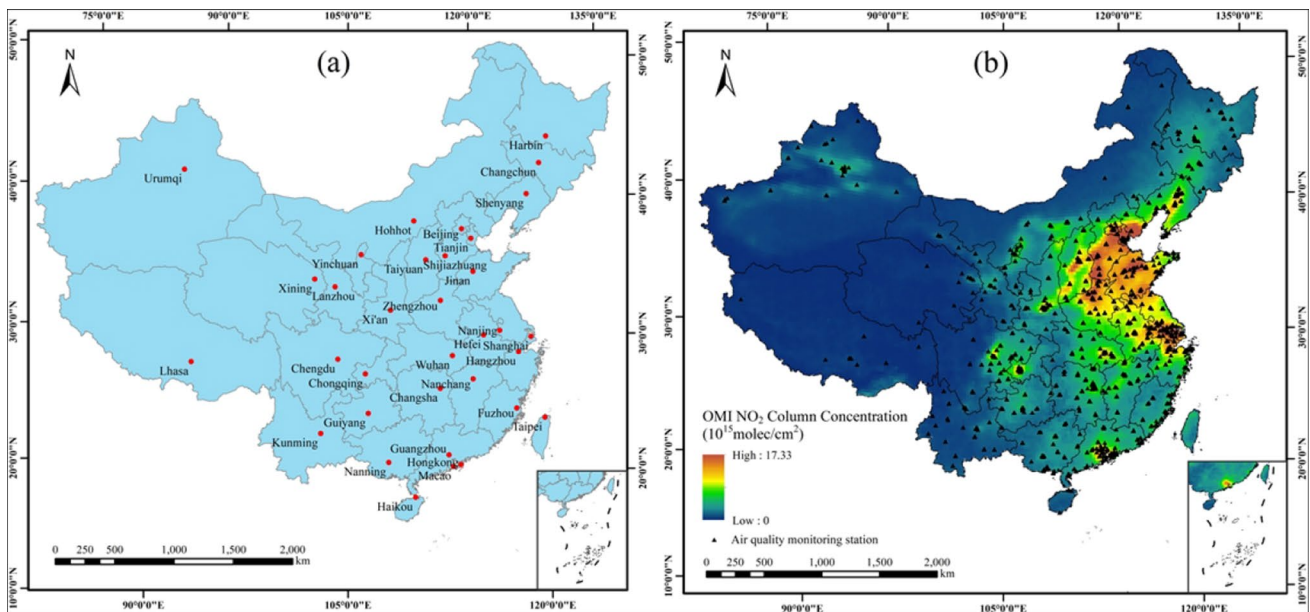
China has gradually established a national air quality monitoring network, and by 2015, the number and coverage of monitoring stations had increased significantly. In this study, hour-by-hour ground-level NO<sub>2</sub> concentration data from January 1, 2015, to December 31, 2020, were selected from the China National Environmental Monitoring Center (CNEMC, <http://106.37.208.233:20035/>). We filtered the stations with at least 80% valid values throughout the year from all stations as input to the model, and finally, approximately 1450 stations passed the filtration. Since the crossing time of the OMI was approximately 13:45 min local time, the average value from 13:00 to 14:00 for each station was selected as the daily measurement.

### OMI NO<sub>2</sub> data

The NO<sub>2</sub> tropospheric column concentration data used in this study were obtained from OMI, which is carried out on the Aura satellite of the Earth Observing System (EOS) and obtains information by observing the backscattered radiation from the Earth's atmosphere and the Earth's surface. OMI can pass wavelengths in the range of 270–500 nm, with an orbital swath width of 2600 km and a spatial resolution of 13 km × 24 km. The product used in this study is the OMI OMNO2d NO<sub>2</sub> cloud-screened tropospheric column concentration level 3 product, which is obtained by quality control on the basis of level 2 product and generates the NO<sub>2</sub> tropospheric column concentrations by area weighting to produce gridded data with a spatial resolution of 0.25° × 0.25°. The production criteria for the cloud-screened column concentration product are zenith angle < 85°, surface reflectivity < 30%, cloud cover < 30%, and 10 < cross-orbit position < 50.

### Meteorological data

The meteorological data used in the study were obtained from the fifth-generation European Centre for Medium-range Weather Forecasts atmospheric reanalysis product (ERA5) of the European Centre for Medium-Range Weather Forecasts (Hersbach 2016). We chose eight meteorological estimates with moderate resolution (0.125° or 0.25°), including the atmospheric boundary layer height (BLH), relative humidity (RH), 2 m atmospheric temperature (TEM),



**Fig. 1** Overview of the study area. **a** The distribution of major cities in China. **b** The annual OMI NO<sub>2</sub> column concentration and the distribution of NO<sub>2</sub> monitoring stations in 2020

u-components and v-components of the 10 m wind, surface pressure (SP), total precipitation (TP), evaporation (ET), and wind speed (WS) and wind direction (WD) was calculated from the u-components and v-components of the 10 m wind.

## Population data

The WorldPop dataset was developed by the WorldPop project (<https://www.worldpop.org>), which provides annual gridded population data for the period 2000 to 2020, and this study used the global population dataset for 2005 to 2020 with a spatial resolution of 1 km. The WorldPop dataset uses a random forest model to reallocate population numbers to the grid. The input variables for this model are the most recent official census data and a spatial auxiliary dataset. The spatial auxiliary dataset includes settlement locations and ranges, satellite nighttime lighting data, land cover data, and road and building maps. The estimated grid population is then finally adjusted to form the final dataset based on the UN Population Division's total national estimates for the target year (Tatem 2017).

## Method

### Data integration

The spatial resolution of the datasets used in the study differed, so the spatial resolution of the ERA5 reanalysis product was chosen to represent all data in this study. The OMI data were interpolated to this resolution using bilinear interpolation, and all ground station measurements contained in a single grid were averaged as the ground NO<sub>2</sub> concentration of the grid. ERA5 data were sampled to be simultaneous to the daily satellite passage time. WorldPop data were calculated as the sum of the population in the 0.125° × 0.125° grid by partition statistics. Finally, the monthly average of all data was calculated as the input to the model.

### Random forest model

The random forest (RF) model is a machine learning theory proposed by Breiman (2001). The basic idea of the algorithm is to construct a certain number of decision trees and combine them according to certain criteria to generate a random forest. Due to the existence of a multilayer random process, the random forest can generate hundreds or even thousands of decision trees randomly and ensure that the decision trees constructed each time may be different due to randomness, which can be used to simulate multiple nonlinear relationships to form complex models.

Random forest regression first randomly selects the sample data by a put-back method to generate K random training

sets, and the unselected part of the data forms the test sample set. For each training set, a fixed number of  $n$  ( $n < p$ ) variables are randomly selected from  $p$  variables as branching nodes of the classification tree to build a regression tree, and each training set generates a corresponding regression tree. The model finally obtains the predicted values by taking the mean of the regression trees.

### Model accuracy evaluation

We evaluated the performance of the random forest model by using mean absolute error (MAE), root mean square error (RMSE), and R-Square ( $R^2$ ). MAE is the mean absolute error and ranges from 0 to positive infinity; the smaller the value is, the smaller the error. RMSE is similar to MAE in that the smaller the value is, the higher the accuracy of the model prediction.  $R^2$  has a value range between 0 and 1; the closer the value is to 1, the better the model fit is. MAE, RMSE, and  $R^2$  are calculated by Formula (1–3), where  $\hat{y}_i$  is the estimated value of the  $i$ -th sample of the model,  $y_i$  is the true value of the  $i$ -th sample,  $\bar{y}$  is the mean of the samples, and  $n$  is the total number of samples.

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

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

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

### Population exposure assessment

NO<sub>2</sub> population exposure was obtained by weighting the ground-level NO<sub>2</sub> concentration and the population. Since the WorldPop dataset calculates the annual population of China, we estimated the annual NO<sub>2</sub> population exposure in China at the prefecture-level city scale. The annual NO<sub>2</sub> population exposure level can be calculated by Formula (4).

$$E_j = \frac{\sum_{i=1}^n (Pop_{ij} \times NO_{2ij})}{\sum_{i=1}^n Pop_{ij}} \quad (4)$$

where  $E_j$  is the NO<sub>2</sub> exposure of a city in year  $j$ ,  $Pop_{ij}$  and  $NO_{2ij}$  are the population and NO<sub>2</sub> concentration of the  $i$ -th

grid in a given year  $j$ , respectively, and  $n$  is the number of all grids in the city.

### Theil-Sen trend analysis

Theil-Sen median trend analysis is able to capture the temporal trend of each grid. Therefore, the results are able to reflect the multiyear trend of NO<sub>2</sub> population exposure. In addition, the method does not require the sample to obey a certain distribution, which makes it highly resistant to data errors (Sen 1968). The trend is calculated by Formula (5).

$$S_R = \text{Median}\left(\frac{E_j - E_i}{j - i}\right) \quad 2005 \leq i < j \leq 2020 \tag{5}$$

where  $S_R$  is the slope of the fit,  $E_i$  is the NO<sub>2</sub> population exposure in year  $i$ , and  $E_j$  is the NO<sub>2</sub> population exposure in year  $j$ . When  $S_R > 0$ , it indicates an increasing trend of the NO<sub>2</sub> population exposure level, and vice versa, a decreasing trend.

### Mann–Kendall test method

The Mann–Kendall test is a nonparametric test to determine the significance of changes in a given variable (Kendall 1955) and is calculated as follows: for the time series  $\{E_i\}$ ,  $i = 2005, 2006, \dots, 2020$ , the  $Z$  statistic is defined as Formula (6).

$$Z = \begin{cases} \frac{S-1}{\sqrt{s(S)}} & S > 0 \\ 0 & S = 0 \\ \frac{S+1}{\sqrt{s(S)}} & S < 0 \end{cases} \tag{6}$$

$$S = \sum_{j=1}^{n-1} \sum_{i=j+1}^n \text{sgn}(E_j - E_i) \tag{7}$$

$$\text{sgn}(E_j - E_i) = \begin{cases} 1 & E_j - E_i > 0 \\ 0 & E_j - E_i = 0 \\ -1 & E_j - E_i < 0 \end{cases} \tag{8}$$

$$s(S) = \frac{n(n-1)(2n+5)}{18} \tag{9}$$

In Formulas (7–9),  $E_i$  and  $E_j$  are the NO<sub>2</sub> population exposure levels in year  $i$  and year  $j$ ,  $n$  represents the length of the time series, and  $\text{sgn}$  is the sign function. In this paper, we determine the significance of the trend of NO<sub>2</sub> population exposure change at the 95% confidence level and then grade the  $Z$  value results into highly significant change ( $|Z| > 2.58$ ), significant change ( $1.96 < |Z| < 2.58$ ), weakly

significant change ( $1.65 < |Z| < 1.96$ ), and no significant change ( $|Z| < 1.65$ ).

### Hurst index analysis

The Hurst index can quantitatively describe the persistence of variables over a time series (HURST 2013); here, we used the Hurst index to analyze the persistence characteristics of NO<sub>2</sub> population exposure. The Hurst index is calculated by Formulas (10–13): for the time series  $\{E_i\}$ ,  $i = 1, 2, \dots, n$ , and for any positive integer  $\tau \geq 1$ , there is the sequence:

$$\bar{E}_\tau = \frac{1}{\tau} \sum_{i=1}^{\tau} E_i \quad \tau = 1, 2, \dots, n \tag{10}$$

$$X_{(i,\tau)} = \sum_{i=1}^{\tau} (E_i - \bar{E}_\tau) \quad 1 \leq i \leq \tau \tag{11}$$

$$E_\tau = \max X_{(i,\tau)} - \min X_{(i,\tau)} \quad \tau = 1, 2, \dots, n \tag{12}$$

$$S_\tau = \sqrt{\frac{1}{\tau} \sum_{i=1}^{\tau} (E_i - E_\tau)^2} \quad \tau = 1, 2, \dots, n \tag{13}$$

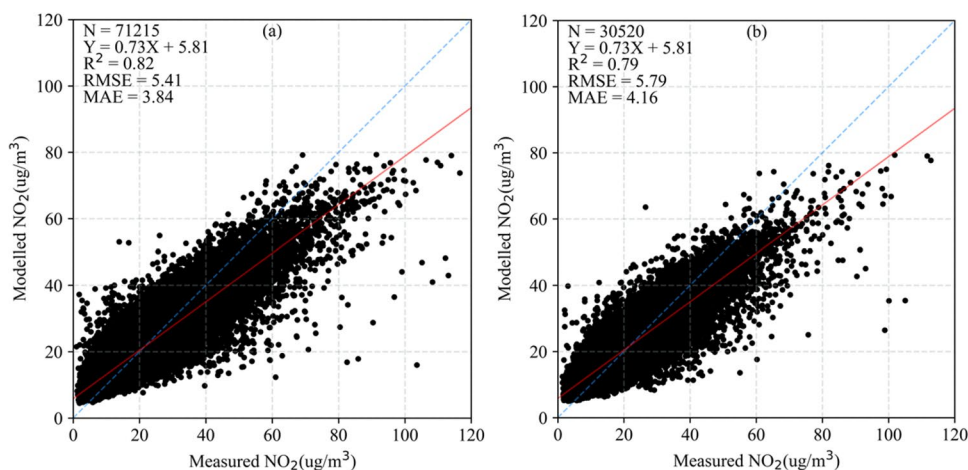
For the standard deviation  $S_\tau$  and range  $E_\tau$ , if  $E_\tau/S_\tau \propto \tau^H$ , then the time series is said to have the Hurst phenomenon, and  $H$  is the Hurst index. When  $0.5 < H < 1$ , the NO<sub>2</sub> population exposure is persistent, and vice versa, it is nonpersistent.

## Result

### Model accuracy

We constructed a random forest model to predict ground-level NO<sub>2</sub> concentrations by combining satellite, meteorological, population, and ground station data. In the model parameters, `n_estimators` was 100, `max_depth` was 30, `max_features` was 4, `min_samples_split` was 15, and `min_samples_leaf` was 15. In the model construction process, 70% of the data were randomly selected as the training set for model training, and 30% of the data were used as the test set to evaluate the accuracy of the model. Figure 2 shows the correlation between the model-simulated and measured NO<sub>2</sub> concentrations in the training and test datasets. There was a significant correlation between the model-simulated concentration and the measured concentration. The MAE, RMSE, and  $R^2$  of the model test dataset were 4.16  $\mu\text{g}/\text{m}^3$ , 5.79  $\mu\text{g}/\text{m}^3$ , and 0.79, respectively, which were less different from the accuracy of the training dataset. Additionally, the  $R^2$  of the model was greater than 0.75 in both the training and test datasets, indicating that the model performs well in

**Fig. 2** Correlation between the model-simulated  $\text{NO}_2$  and measured  $\text{NO}_2$  concentrations. **a** Training dataset. **b** Test dataset



simulating ground-level  $\text{NO}_2$  concentrations. In addition, we evaluated the model by using fivefold cross-validation, and the MAE, RMSE, and  $R^2$  of the model cross-validation were  $4.3 \mu\text{g}/\text{m}^3$ ,  $5.82 \mu\text{g}/\text{m}^3$ , and 0.77, respectively. The cross-validation results indicated that the random forest model has no overfitting phenomenon. Compared with the validation results in the model test dataset, the cross-validation  $R^2$  decreased by 0.02, RMSE increased by  $0.03 \mu\text{g}/\text{m}^3$ , and MAE increased by  $0.14 \mu\text{g}/\text{m}^3$ . The cross-validation results were basically consistent with the validation results in the model test dataset, which proved that the model was stable and reliable. Therefore, the simulated ground-level  $\text{NO}_2$  concentration from the random forest model can be used to analyze the spatial and temporal variations in  $\text{NO}_2$  concentration and population exposure in China.

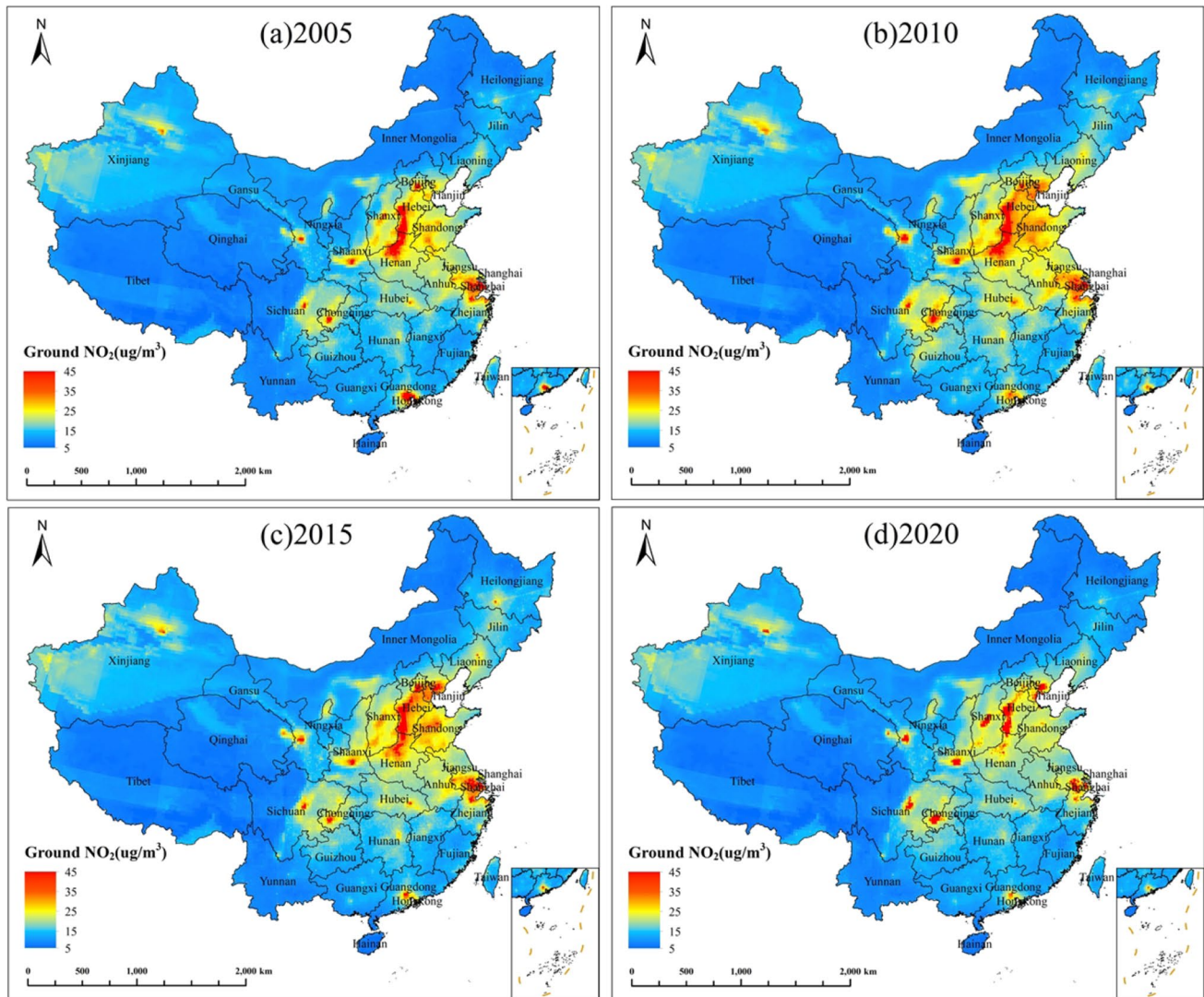
### Temporal and spatial changes in ground-level $\text{NO}_2$ concentrations

We studied the spatial and temporal variations in annual ground-level  $\text{NO}_2$  concentrations in China. The annual  $\text{NO}_2$  concentration was calculated from the monthly  $\text{NO}_2$  concentration predicted by the model. Figure 3 shows the distribution of the national annual  $\text{NO}_2$  concentration in 2005, 2010, 2015, and 2020, and  $\text{NO}_2$  showed spatial aggregation features.  $\text{NO}_2$  pollution was most serious in northern China, not only due to the high  $\text{NO}_2$  concentration but also due to the large area of high  $\text{NO}_2$  concentration, covering seven provinces and municipalities directly under the central government, including Henan, Hebei, Shandong, Beijing, and Tianjin. Within the region,  $\text{NO}_2$  pollution was concentrated in the south-central part of Hebei Province and the northern part of Henan Province. High  $\text{NO}_2$  concentrations in other regions were mostly found in cities with developed regions and their surrounding areas. For example, high  $\text{NO}_2$  concentrations in the southwest were located in the western part of Chongqing and Chengdu, and those in the northwest were

located in Lanzhou and Xi'an. These are both provincial capitals or the main urban areas of municipalities directly under the central government. The situation in southern China was very similar to that in the west. High  $\text{NO}_2$  concentrations were concentrated in large cities such as Shenzhen, Guangzhou, and other cities in Guangdong Province. In contrast, the areas of high  $\text{NO}_2$  concentrations in eastern China were more dispersed. These areas included Shanghai, the southern part of Jiangsu Province, and the northern part of Zhejiang Province. This may be because the urbanization and industrialization levels of cities in eastern China differed less, causing  $\text{NO}_2$  pollution levels to be relatively similar.

Temporally, the distribution of  $\text{NO}_2$  concentrations in China showed a trend of first increasing and then decreasing. From 2005 to 2010, the annual  $\text{NO}_2$  concentration increased. Compared to 2005, the increasing trend of  $\text{NO}_2$  concentration in eastern and northern China was obvious in 2010, such as Henan, Hebei, Shandong, Jiangsu, and Zhejiang Provinces. Furthermore,  $\text{NO}_2$  pollution increased in some cities in the west, such as Chengdu, Chongqing, and Lanzhou. From 2010 to 2015, the change in  $\text{NO}_2$  concentration was slight, and there was a degree of decrease in  $\text{NO}_2$  concentration in western China. The change in other regions was not obvious. From 2015 to 2020, the  $\text{NO}_2$  concentration in China declined significantly. In 2020, the national  $\text{NO}_2$  concentrations decreased to low levels, and the extent of  $\text{NO}_2$  pollution also decreased significantly, especially in the northern provinces of China, such as Henan and Shandong. Guangzhou and its surrounding cities contained concentrated areas of  $\text{NO}_2$  pollution in southern China, but the change in  $\text{NO}_2$  concentration in this region was different from the overall change trend, which was always in a decreasing trend from 2005 to 2020.

Because the  $\text{NO}_2$  concentration differs significantly in different seasons, we selected the maximum and mean  $\text{NO}_2$  concentrations in different seasons to analyze the  $\text{NO}_2$  concentration variations in China. The  $\text{NO}_2$  concentration was



**Fig. 3** The annual  $\text{NO}_2$  concentration in China over multiple years. **a** 2005; **b** 2010; **c** 2015; **d** 2020

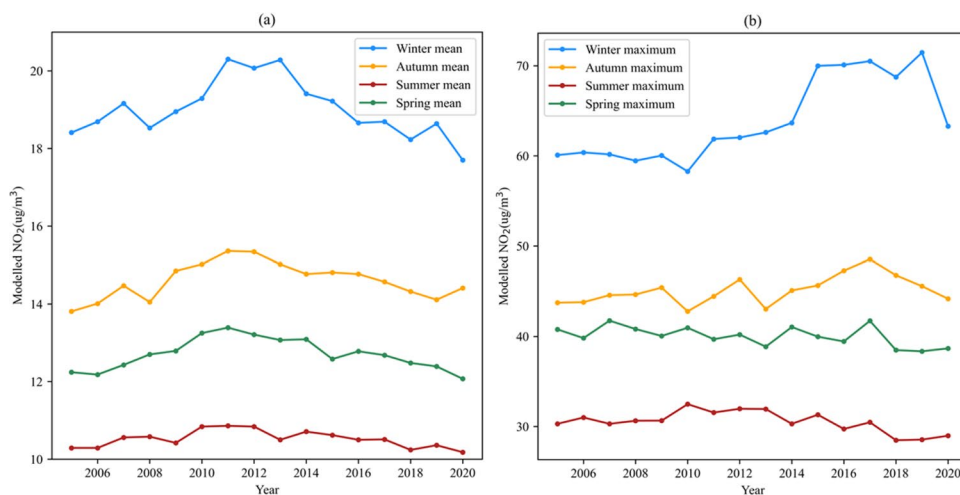
low in spring and summer, while it was higher in autumn and winter, as shown in Fig. 4. The lowest  $\text{NO}_2$  concentration throughout the year occurs during the summer due to higher temperatures, which were conducive to the decomposition of  $\text{NO}_2$ . Central heating in winter produces a large amount of air pollutants, including  $\text{NO}_2$ , resulting in winter being the most serious season for  $\text{NO}_2$  pollution. The  $\text{NO}_2$  concentration showed a trend of increasing and then decreasing from the mean concentration change. The average  $\text{NO}_2$  concentration in each season gradually increased from 2005 to 2012 and was in the decreasing stage between 2013 and 2020. The maximum  $\text{NO}_2$  concentration reflects the serious areas of  $\text{NO}_2$  pollution. The maximum  $\text{NO}_2$  concentrations in spring and summer in 2005–2020 gradually declined, while autumn had a fluctuating upwards trend. The maximum  $\text{NO}_2$  concentration in winter has a degree of increase from 2005 to 2019 and a significant decrease in 2020. This change was

likely related to the COVID-19 epidemic. In addition, the maximum concentrations in autumn and winter were both higher than  $40 \mu\text{g}/\text{m}^3$ , indicating that  $\text{NO}_2$  pollution was still serious in some regions in autumn and winter. Therefore, corresponding environmental protection policies need to be formulated for provinces with serious pollution in China, such as Henan and Hebei. Additionally, although  $\text{NO}_2$  pollution was very serious in some regions, it can be revealed from the variations in seasonal mean  $\text{NO}_2$  concentrations that  $\text{NO}_2$  concentrations in most regions of China were at a low level and  $\text{NO}_2$  pollution was concentrated in a small number of regions.

### Ground-level $\text{NO}_2$ exposure assessment

We assessed  $\text{NO}_2$  population exposure at the prefecture-level city scale. Figure 5 shows the  $\text{NO}_2$  population exposure in

**Fig. 4** Seasonal variations in NO<sub>2</sub> concentration. **a** Seasonal mean. **b** Seasonal maximum



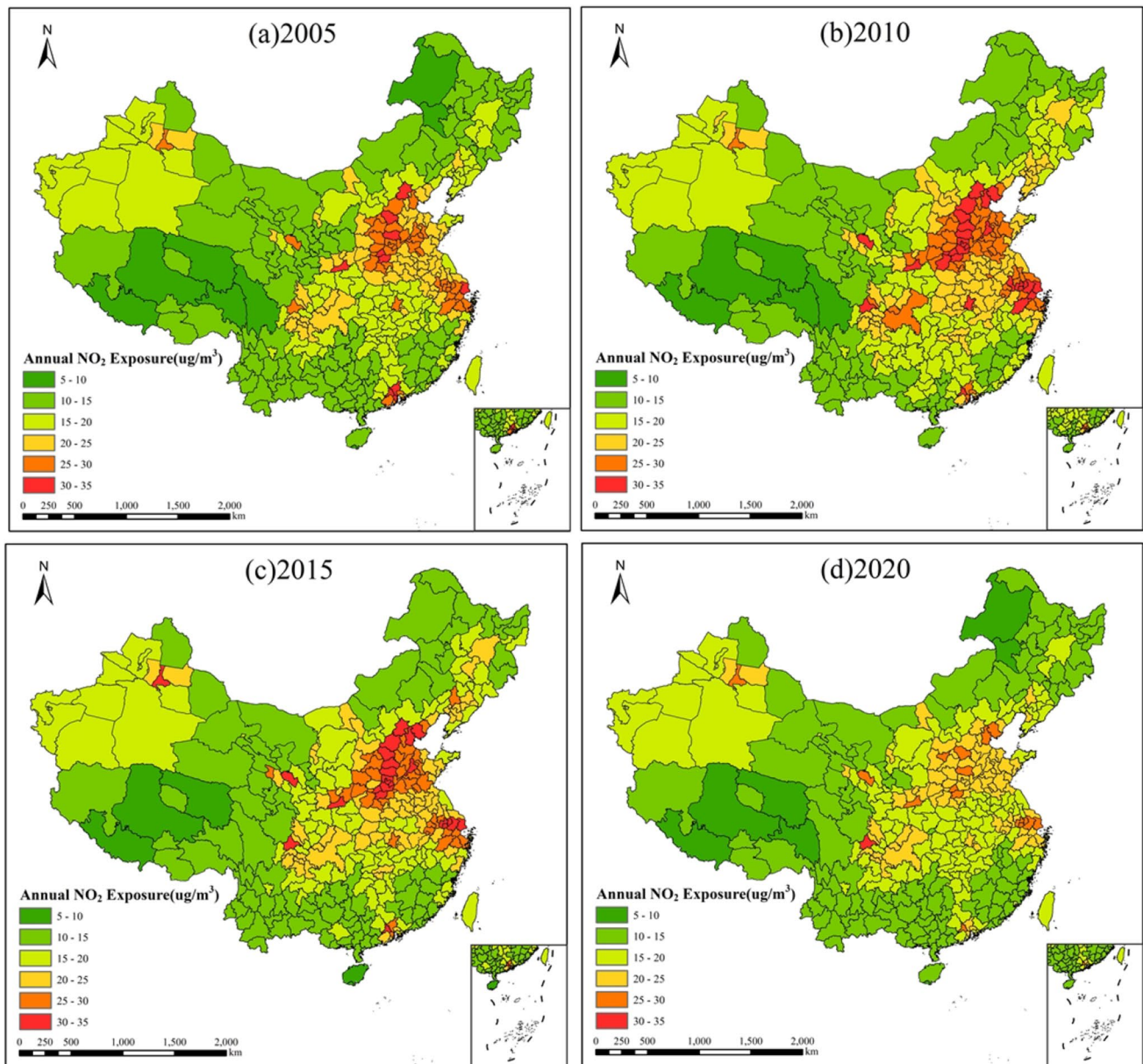
2005, 2010, 2015, and 2020. The spatial distribution of NO<sub>2</sub> population exposure was similar to the distribution of NO<sub>2</sub> concentration, with both having obvious spatial aggregation. The areas of high NO<sub>2</sub> population exposure in the northwest were centred on Lanzhou, Xi'an, and Urumqi. The areas of high NO<sub>2</sub> exposure in the southwest were centred on Chengdu and Chongqing, and the NO<sub>2</sub> population exposure of Chengdu was significantly higher than that of Chongqing. The central region had a relatively low NO<sub>2</sub> population exposure, and only Wuhan had a significantly higher NO<sub>2</sub> population exposure than the other cities. In the eastern region, there were several cities with high NO<sub>2</sub> population exposure, such as Hangzhou, Nanjing, and Shanghai. Additionally, the surrounding cities also had high NO<sub>2</sub> population exposure, indicating that the NO<sub>2</sub> concentration in this region was high and that the population distribution was also concentrated. Northern China is the region with the highest NO<sub>2</sub> population exposure, with more cities at high NO<sub>2</sub> population exposure, including Beijing, Tianjin, Shijiazhuang, Zhengzhou, and Jinan. The distribution of cities with high NO<sub>2</sub> population exposure in Henan and Hebei was generally similar to the areas of high NO<sub>2</sub> concentrations. Shandong Province had a relatively high NO<sub>2</sub> population exposure due to its dense population, but the NO<sub>2</sub> concentration in Shandong was lower than that in Henan and Hebei.

In terms of time, the NO<sub>2</sub> population exposure in China showed a significant increasing trend from 2005 to 2010. The rising trend was most obvious in the eastern and northern cities, such as Shanghai and Hangzhou in the east and Shijiazhuang, Zhengzhou, Beijing, and Jinan in the north. These cities were mostly located in large urban agglomerations, and the surrounding cities were also very densely populated. Therefore, in these two regions, cities with high NO<sub>2</sub> population exposure tend to be distributed in clusters. In the northwest region, the population distribution was relatively concentrated due to the smaller population, so high

NO<sub>2</sub> population exposure was usually in large cities, such as Lanzhou and Urumqi, while the NO<sub>2</sub> population exposure in the surrounding cities of these cities was relatively low. The two major cities in the southwestern region, Chongqing and Chengdu, also increased to some degree, but the NO<sub>2</sub> population exposure in the surrounding cities did not change significantly. The NO<sub>2</sub> population exposure in the southern region did not increase noticeably, and some cities even decreased to a certain extent. The NO<sub>2</sub> population exposure of some cities in central and southwest China decreased to a certain extent from 2010 to 2015, particularly in two cities, Wuhan and Chongqing. The NO<sub>2</sub> population exposure increased in Urumqi. The remaining cities in the country showed no significant change. The NO<sub>2</sub> population exposure decreased significantly in almost all cities from 2015 to 2020, and NO<sub>2</sub> pollution improved substantially in 2020. However, the NO<sub>2</sub> population exposure in Chengdu was still higher than 30 µg/m<sup>3</sup>, and the NO<sub>2</sub> population exposure in Urumqi did not change significantly, indicating that NO<sub>2</sub> pollution was still serious in some cities in western China.

There were 33 cities with NO<sub>2</sub> population exposure greater than 30 µg/m<sup>3</sup> in 2012, which was the largest number of cities in the period 2005–2020. Therefore, we chose 2012 as the dividing year to study the NO<sub>2</sub> population exposure trends in both periods. We calculated the NO<sub>2</sub> population exposure trends based on Theil-Sen trend analysis and then analyzed the significance of the trends based on the Mann–Kendall test. Figure 6a shows the result of the M–K trend test for NO<sub>2</sub> population exposure in each city from 2005 to 2012. During this period, NO<sub>2</sub> population exposure significantly increased in the majority of Chinese cities. Some areas, such as Qinghai, Tibet, northern Gansu, western Sichuan, and western Yunnan, did not change significantly. These areas are mainly sparsely populated areas and have comparatively lower NO<sub>2</sub> concentrations. In addition, a few densely populated cities had no significant changes



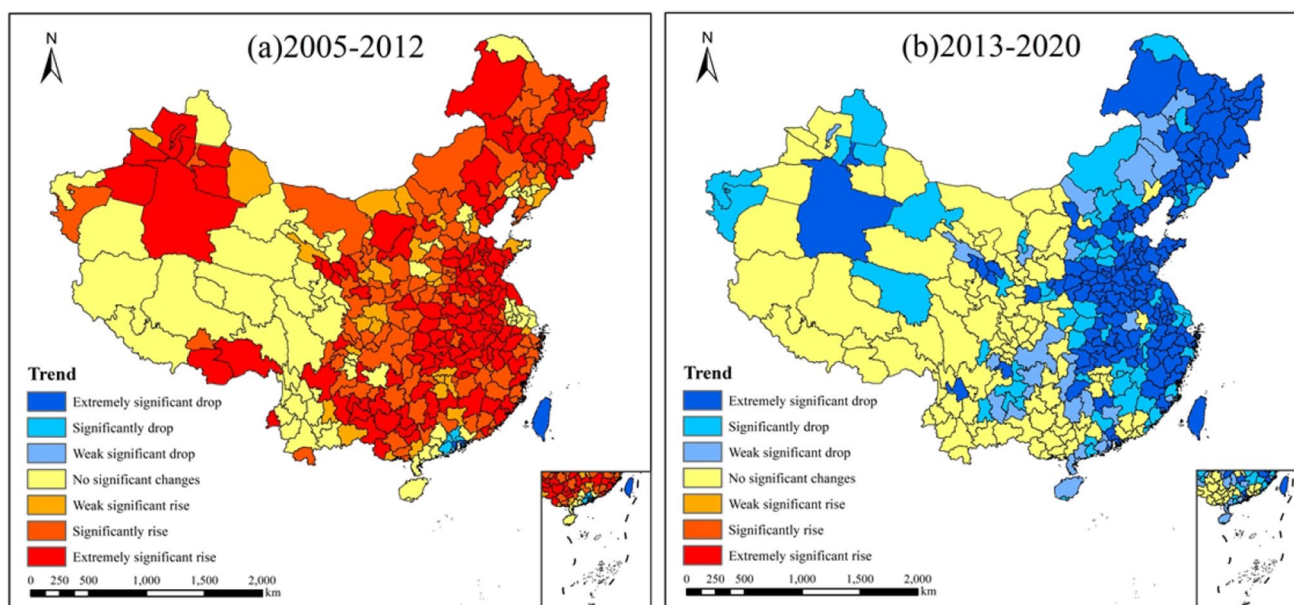


**Fig. 5** NO<sub>2</sub> population exposure in China for multiple years. **a** 2005; **b** 2010; **c** 2015; **d** 2020

in NO<sub>2</sub> population exposure, such as Beijing, Shanghai, and Suzhou. Some cities in Guangdong Province, such as Guangzhou, Dongguan, and Foshan, showed a decrease or even a significant decrease.

Figure 6b shows the trend of NO<sub>2</sub> population exposure from 2013 to 2020. This period was dominated by a significant decline in NO<sub>2</sub> population exposure, but the number of cities that experienced a decline was obviously less than the number of cities that rose in the previous period, and the decline was mostly concentrated in the central, eastern, and northern parts of the country. Among the regions showing a downwards trend, Wuhan, Nanchang, and Changsha in the central region are the centre, but

Hefei did not change significantly in this period. In the eastern region, Hangzhou and Shanghai were the centres. The north has the greatest number of cities with a significant downwards trend, including Beijing and Tianjin, the majority of cities in Henan and Shandong, and Harbin and Shenyang in the northeast. There was a clear difference in the west. The majority of cities in the west are dominated by no significant changes or weak downwards trends. Among the major cities in this region, Lanzhou and Urumqi in the northwest had a significant downwards trend. Chongqing and Chengdu, the two central cities in the southwest, showed a weak downwards trend or no significant change. We found that cities with downwards

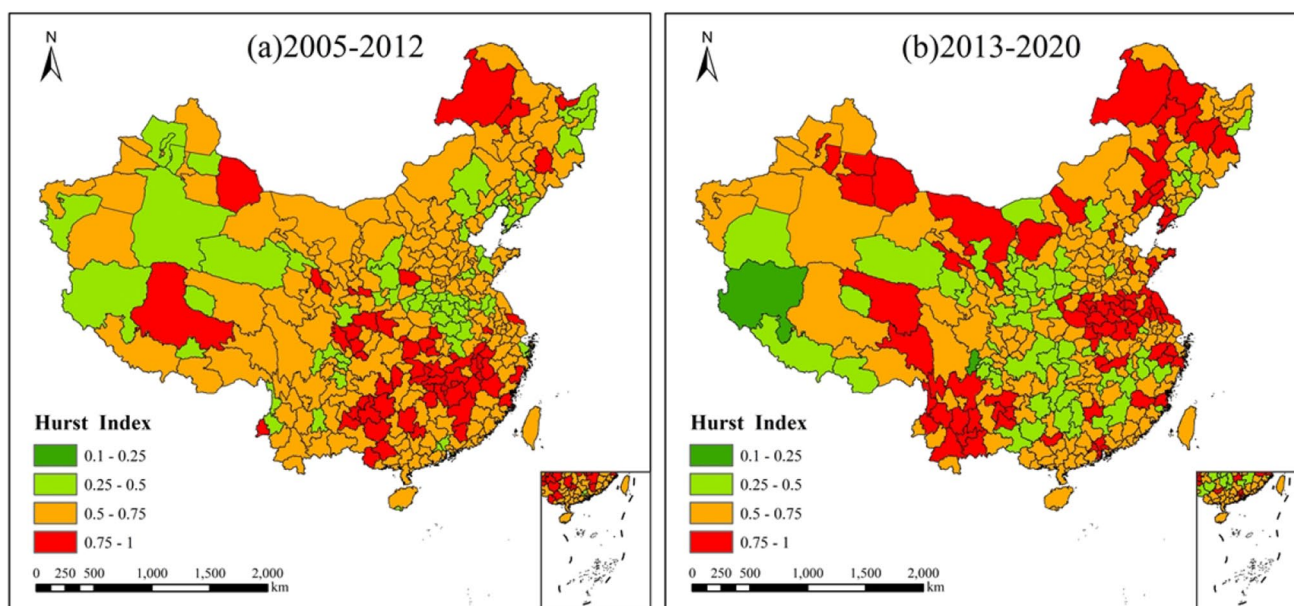


**Fig. 6** Trends in  $\text{NO}_2$  population exposure. **a** The trend from 2005 to 2012. **b** The trend from 2013 to 2020

trends were mostly concentrated in the north, while southern cities had fewer downwards trends.

The Hurst index measures the persistence of changes in  $\text{NO}_2$  population exposure. Figure 7a shows the Hurst index of  $\text{NO}_2$  population exposure from 2005 to 2012. The increase in  $\text{NO}_2$  population exposure was mostly persistent between 2005 and 2012, especially in the southern cities. Because  $\text{NO}_2$  concentrations were relatively low in southern China, the persistent increase in  $\text{NO}_2$  population exposure

indicated that the region has a strong population attraction and increasing population density that led to increasing  $\text{NO}_2$  population exposure. The north was also generally dominated by a continuous increase, but the growth of most cities in Henan was noncontinuous, and the  $\text{NO}_2$  concentrations of Henan showed an increasing trend during this period. Therefore, it may be that the population density in Henan Province decreased, resulting in a noncontinuous increase in  $\text{NO}_2$  population exposure. Figure 7b shows the Hurst



**Fig. 7** Hurst index of  $\text{NO}_2$  population exposure. **a** The index of 2005–2012. **b** The index of 2013–2020

index of NO<sub>2</sub> population exposure from 2013 to 2020. The declining trend in the north from 2013 to 2020 was mostly a continuous decline, and the western cities and a small number of southern cities exhibited noncontinuous changes. The decline in NO<sub>2</sub> population exposure in the north during this period was mostly due to policy factors, such as the enactment of strict environmental protection laws in 2015, which reduced nitrogen dioxide emissions. The western region was mostly nonpersistent in this period, which is consistent with the results of the lack of a significant trend above. In addition, there was also an impact of the COVID-19 epidemic during this period, which may explain the discontinuous changes in a small portion of the southern region.

## Discussion

In this study, we used a random forest model to retrieve ground-level NO<sub>2</sub> concentrations in China from 2005 to 2020 and analyzed the changes in NO<sub>2</sub> population exposure over the years. The accuracy of the model result was high; the MAE, RMSE, and  $R^2$  of the model were 4.16  $\mu\text{g}/\text{m}^3$ , 5.79  $\mu\text{g}/\text{m}^3$ , and 0.79, respectively, and the MAE, RMSE, and  $R^2$  of the model cross-validation were 4.3  $\mu\text{g}/\text{m}^3$ , 5.82  $\mu\text{g}/\text{m}^3$ , and 0.77, respectively. Apart from the random forest model, we compared different regression methods, such as the commonly used linear regression, backpropagation neural network (BPNN), and support vector machine (SVM) models. Table 1 shows the results of the multiple model comparison. The random forest model has the smallest error of all models, while the traditional linear model has the worst-fitting performance. The support vector machine model was the worst among deep learning models and has the longest computation time. The results of the comparison showed that the deep learning model has a clear advantage in large data simulations. All three deep learning methods perform better than the linear regression model. This indicates that deep learning regression is usually better than the traditional statistical model in the case of complex parameters and a large amount of data. Previous research has used a variety of models to estimate ground-level NO<sub>2</sub> concentrations, such as the extra tree model, geographically and temporally weighted regression model, community multi-scale air quality model, and land use regression model (Gu et al. 2017; Larkin et al. 2017; Qin et al. 2020, 2017). The

$R^2$  values for these models were between 0.51 and 0.7, and the RMSE values were all greater than 9  $\mu\text{g}/\text{m}^3$ . Compared to previous research, the random forest model used in this study significantly improved the accuracy of the simulated ground-level NO<sub>2</sub> concentrations, and the estimation results were more reliable. In addition, the random forest model is relatively simple to implement, with low computational overhead and strong interpretability of the model.

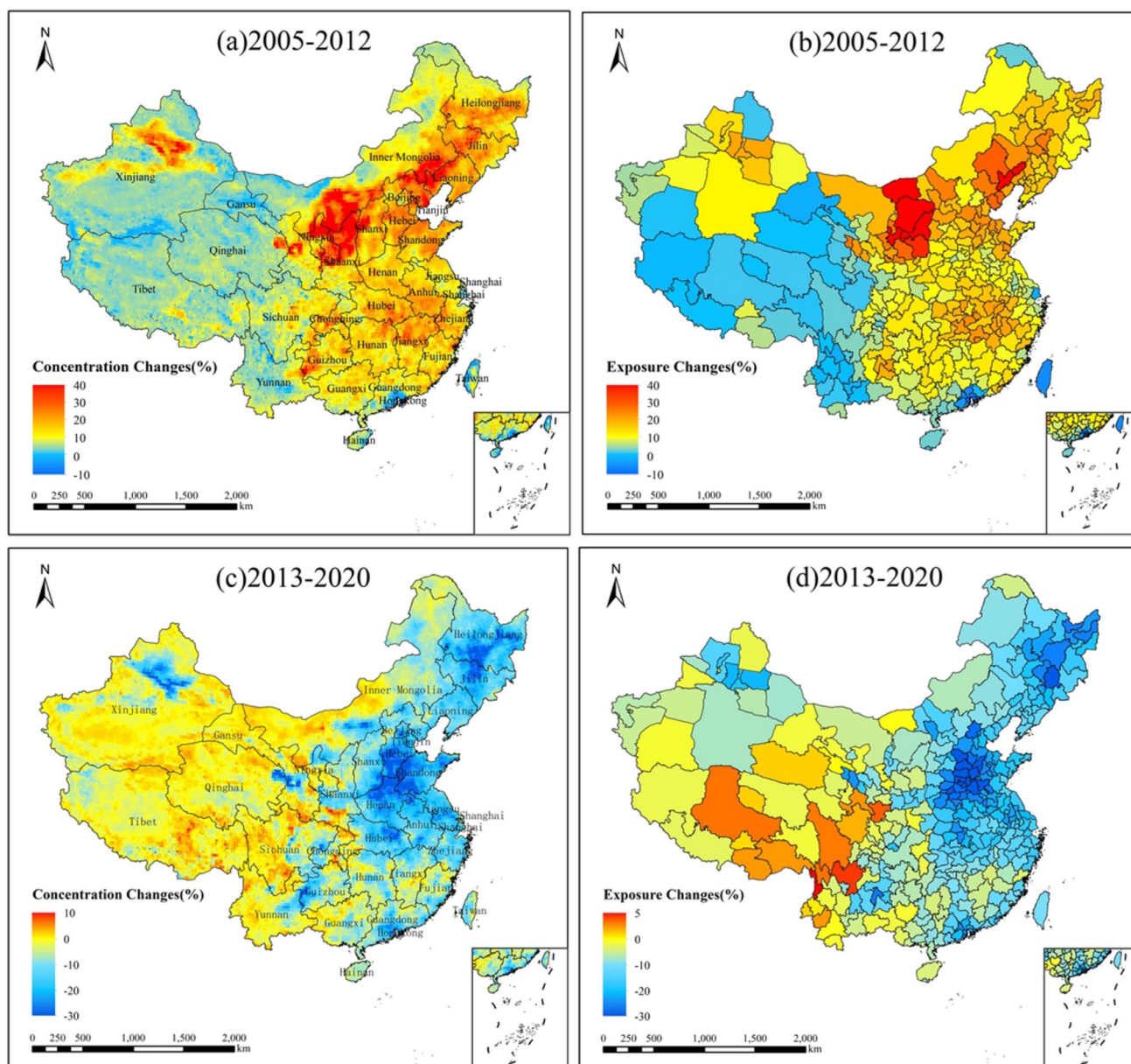
Figure 8 shows the amount of change in NO<sub>2</sub> concentration and NO<sub>2</sub> population exposure for 2005–2012 and 2013–2020, and we chose NO<sub>2</sub> concentration and NO<sub>2</sub> population exposure in 2005 and 2013 as the reference. In the first period, the changes in NO<sub>2</sub> concentration and NO<sub>2</sub> population exposure were generally similar in central and eastern China. Northern Shaanxi and south-central Inner Mongolia were the two regions with the highest increase in NO<sub>2</sub> concentration, and the increase in NO<sub>2</sub> population exposure in these two regions was also the highest in the country. Some differences in NO<sub>2</sub> concentration and NO<sub>2</sub> population exposure were found in the western region. For example, the rise in NO<sub>2</sub> population exposure in Urumqi and its surrounding cities was significantly lower than the rise in NO<sub>2</sub> concentration. Additionally, the NO<sub>2</sub> concentrations in some regions of Tibet, Qinghai, and Yunnan increased, but the NO<sub>2</sub> population exposure decreased, indicating that the population density in these regions was low and that the increase in NO<sub>2</sub> concentration did not directly cause an increase in NO<sub>2</sub> population exposure. The changes in NO<sub>2</sub> concentration and NO<sub>2</sub> population exposure were generally consistent in the second period.

The NO<sub>2</sub> concentrations and population exposure showed a significant increasing trend from 2005 to 2012, and the NO<sub>2</sub> population exposure persistently increased in most cities. During this period, industrial development in China was rapid. Industrial production increased from 7795.83 billion yuan in 2005 to 20,890.14 billion yuan in 2012, but the industries at this time were mostly rough industries, which seriously polluted the environment. Furthermore, the number of motor vehicles in this period rapidly increased from 43.29 to 120 million. Vehicle exhaust emissions are also a major source of NO<sub>2</sub>. Thus, the rapid growth of industry and vehicle ownership is responsible for the significant increase in NO<sub>2</sub> concentration and NO<sub>2</sub> population exposure during this period. Northern China, such as Henan and Hebei, has a large concentration of heavy industry and population, so the NO<sub>2</sub> population exposure is the highest in the country.

The NO<sub>2</sub> concentration and NO<sub>2</sub> population exposure showed a decreasing trend nationwide from 2013 to 2020. The decline over this period was mainly due to Chinese environmental protection policies, including the mandatory cleanup of coal and the extremely stringent environmental protection law enacted in 2015. These measures have significantly limited pollutant emissions and increased penalties

**Table 1** Comparison of multiple models

Model	MAE	RMSE	$R^2$
LR	6.09	8.37	0.55
RF	4.16	5.79	0.79
SVM	5.15	7.38	0.65
BPNN	4.91	6.54	0.73



**Fig. 8** Quantified changes in  $\text{NO}_2$  concentration and  $\text{NO}_2$  population exposure. **a**  $\text{NO}_2$  concentration changes from 2005 to 2012; **b**  $\text{NO}_2$  population exposure changes from 2005 to 2012; **c**  $\text{NO}_2$  concentra-

tion changes from 2013 to 2020; **d**  $\text{NO}_2$  population exposure changes from 2013 to 2020

for companies that violate the law on emissions. Meanwhile, the new regulations imposed strict requirements on government management and incorporated the effectiveness of pollution control into the government's performance evaluation. Chinese ecological and environmental departments issued a total of 224.8 billion yuan in ecological funds from 2016 to 2020 and environmental funds and initially established evaluation systems for air, water, and soil environmental protection. The  $\text{NO}_2$  concentrations in 2020 were significantly reduced compared to those in 2015. Although the COVID-19 epidemic also had some impact on the decrease

in  $\text{NO}_2$  concentrations, life in China largely normalized in the second half of the year, and the COVID-19 epidemic had a limited impact on the  $\text{NO}_2$  concentrations throughout the year.

## Conclusion

In this study, a random forest model based on OMI data and the ERA5 reanalysis product was constructed to retrieve the  $0.125^\circ \times 0.125^\circ$  ground-level  $\text{NO}_2$  concentrations in China

from 2005 to 2020. The MAE, RMSE, and  $R^2$  of the model were  $4.16 \mu\text{g}/\text{m}^3$ ,  $5.79 \mu\text{g}/\text{m}^3$ , and 0.79, respectively. The model results showed clear spatial aggregation of the  $\text{NO}_2$  concentration, which was consistent with  $\text{NO}_2$  population exposure. The average  $\text{NO}_2$  concentration in each season tended to increase and then decrease, which was consistent with the trend of the annual  $\text{NO}_2$  concentration. However, the maximum concentrations in autumn and winter still rose and were higher than the China environmental pollution standard, indicating that  $\text{NO}_2$  pollution did not improve significantly in some areas during autumn and winter.

The cities with high  $\text{NO}_2$  population exposure and areas with high  $\text{NO}_2$  concentrations basically overlap. Although the  $\text{NO}_2$  concentration was lower than that in Henan and Hebei Provinces, Shandong Province had a higher  $\text{NO}_2$  population exposure due to its dense population.  $\text{NO}_2$  population exposure increased significantly in most cities from 2005 to 2012, and most of the increase during this period was persistent. This result suggests that the increasing population density in southern China led to increased  $\text{NO}_2$  population exposure, as the  $\text{NO}_2$  concentration in southern China was relatively low. In contrast, the unsustainable increase in the  $\text{NO}_2$  population exposure in the northern cities was likely due to the outflow of the population. Most cities experienced a significant decrease in  $\text{NO}_2$  population exposure from 2013 to 2020, but the number of cities that experienced a decline was significantly less than the number of cities that rose in the previous period. The main reason for the significant upwards trend in both  $\text{NO}_2$  concentrations and  $\text{NO}_2$  population exposure from 2005 to 2012 was the rapid growth of industry and car ownership. The decline from 2013 to 2020 is mainly due to Chinese environmental protection policies.

By 2020, the southern cities still maintained low  $\text{NO}_2$  population exposure, and the eastern and northern cities significantly improved  $\text{NO}_2$  population exposure. However, the reduction in  $\text{NO}_2$  population exposure in the western region was not significant. Urumqi, Lanzhou, and Chengdu still maintained high  $\text{NO}_2$  population exposure, which indicated that the major cities in the western region require more attention. In these cities with high  $\text{NO}_2$  population exposure, the EPA could install more  $\text{NO}_2$  concentration monitoring instruments to broadcast real-time  $\text{NO}_2$  concentrations. People can avoid going to areas with high  $\text{NO}_2$  concentrations by broadcasting. For  $\text{NO}_2$  emission sources, the EPA could enforce factories to clean their emissions and encourage people to replace their fuel-powered vehicles with new energy electric vehicles.

There were also some shortcomings in this study, such as some studies suggesting that OMI data are somewhat underestimated in urban areas (Qin et al. 2020), which may lead to underestimation of ground-level  $\text{NO}_2$  concentrations in some regions. In addition, the uneven distribution of monitoring stations on the ground may also introduce errors into

the model. In a follow-up study, we intend to use multiple satellite datasets or introduce more geographic auxiliary elements, such as road data, to improve the accuracy of the model and then study the human health effects from prolonged exposure to high  $\text{NO}_2$  concentrations.

**Acknowledgements** We would like to thank the editors and anonymous referees for their constructive comments.

**Author contribution** Huang, ZY analyzed the data, constructed a random forest model to estimate ground-level  $\text{NO}_2$  concentrations in China, and wrote the manuscript. Shen, JW put forward the research objectives and ideas of the paper and was responsible for the planning and execution of research activities. Xu, XK was involved in the analysis of the data. Ma, MG offered some guidelines for the research.

**Funding** The research was supported by the National Natural Science Foundation of China (No. 41830648) and Chongqing Youth Innovative Talent Training Program (No. CY210208, CY220234).

**Data availability** The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

## References

- Achakulwisut P, Brauer M, Hystad P, Anenberg SC (2019) Global, national, and urban burdens of paediatric asthma incidence attributable to ambient  $\text{NO}_2$  pollution: estimates from global datasets. *Lancet Planet Health* 3(4):e166–e178
- Altshuller AP, Bufalini JJ (1971) Photochemical aspects of air pollution. *Rev Environ Sci Technol* 5(1):39–64
- Anyanwu E (1999) Complex interconvertibility of nitrogen oxides ( $\text{NO}_x$ ): impact on occupational and environmental health. *Rev Environ Health* 14(3):169–185
- Araki S, Shima M, Yamamoto K (2018) Spatiotemporal land use random forest model for estimating metropolitan  $\text{NO}_2$  exposure in Japan. *Sci Total Environ* 634:1269–1277
- Atkinson R (2000) Atmospheric chemistry of VOCs and  $\text{NO}_x$ . *Atmos Environ* 34(12–14):2063–2101
- Bechle MJ, Millet DB, Marshall JD (2013) Remote sensing of exposure to  $\text{NO}_2$ : satellite versus ground-based measurement in a large urban area. *Atmos Environ* 69:345–353
- Bidleman TF (1988) Atmospheric processes. *Environ Sci Technol* 22(4):361–367
- Boersma KF, Jacob DJ, Bucsela E, Perring A, Dirksen R, Yantosca R, ... Cohen R (2008) Validation of OMI tropospheric  $\text{NO}_2$  observations during INTEX-B and application to constrain  $\text{NO}_x$  emissions over the eastern United States and Mexico. *Atmos Environ* 42(19):4480–4497
- Breiman L (2001) Random forests. *Machine Learning* 45(1):5–32

- Chen R, Yin P, Meng X, Wang L, Liu C, Niu Y, ... Qi J (2018) Associations between ambient nitrogen dioxide and daily cause-specific mortality: evidence from 272 Chinese cities. *Epidemiology* 29(4):482–489
- Cheng L, Tao J, Valks P, Yu C, Liu S, Wang Y, ... Chen L (2019) NO<sub>2</sub> retrieval from the environmental trace gases monitoring instrument (EMI): preliminary results and intercomparison with OMI and TROPOMI. *Remote Sensing* 11(24):3017
- Chi Y, Fan M, Zhao C, Sun L, Yang Y, Yang X, Tao J (2021) Ground-level NO<sub>2</sub> concentration estimation based on OMI tropospheric NO<sub>2</sub> and its spatiotemporal characteristics in typical regions of China. *Atmos Res* 264:105821
- Chi Y, Fan M, Zhao C, Yang Y, Fan H, Yang X, ... Tao J (2022) Machine learning-based estimation of ground-level NO<sub>2</sub> concentrations over China. *Sci Total Environ* 807:150721
- Crouse DL, Peters PA, Hystad P, Brook JR, van Donkelaar A, Martin RV, Pope CA III (2015) Ambient PM<sub>2.5</sub>, O<sub>3</sub>, and NO<sub>2</sub> exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC). *Environ Health Perspect* 123(11):1180–1186
- Crutzen PJ (1979) The role of NO and NO<sub>2</sub> in the chemistry of the troposphere and stratosphere. *Annu Rev Earth Planet Sci* 7(1):443–472
- Fenech S, Aquilina NJ (2021) Estimation of the NO<sub>2</sub> population exposure in the Northern Harbour district of Malta. *Atmos Environ* 244:117918
- Fishman J, Bowman KW, Burrows JP, Richter A, Chance KV, Edwards DP, ... Ziemke JR (2008) Remote sensing of tropospheric pollution from space. *Bull Am Meteor Soc* 89(6):805–822
- Fridell E, Haeger-Eugensson M, Moldanova J, Forsberg B, Sjöberg K (2014) A modelling study of the impact on air quality and health due to the emissions from E85 and petrol fuelled cars in Sweden. *Atmos Environ* 82:1–8
- Gu J, Chen L, Yu C, Li S, Tao J, Fan M, ... Su L (2017) Ground-level NO<sub>2</sub> concentrations over China inferred from the satellite OMI and CMAQ model simulations. *Remote Sensing* 9(6):519
- He B, Heal MR, Humstad KH, Yan L, Zhang Q, Reis S (2019) A hybrid model approach for estimating health burden from NO<sub>2</sub> in megacities in China: a case study in Guangzhou. *Environ Res Lett* 14(12):124019
- He MZ, Kinney PL, Li T, Chen C, Sun Q, Ban J, ... Kioumourtzoglou M-A (2020) Short-and intermediate-term exposure to NO<sub>2</sub> and mortality: a multi-county analysis in China. *Environ Pollut* 261:114165
- Hersbach H (2016) The ERA5 atmospheric reanalysis. Paper presented at the AGU fall meeting abstracts
- Hu Y, Liu C, Chen R, Kan H, Zhou M, Zhao B (2021) Associations between total mortality and personal exposure to outdoor-originated NO<sub>2</sub> in 271 Chinese cities. *Atmos Environ* 246:118170
- Hu Y, Yao M, Liu Y, Zhao B (2020) Personal exposure to ambient PM<sub>2.5</sub>, PM<sub>10</sub>, O<sub>3</sub>, NO<sub>2</sub>, and SO<sub>2</sub> for different populations in 31 Chinese provinces. *Environment International* 144:106018
- Hurst H (2013) Closure to long-term storage capacity of reservoirs by HE Hurst. *Trans Am Soc Civ Eng* 116:804–808
- Ialongo I, Virta H, Eskes H, Hovila J, Douros J (2020) Comparison of TROPOMI/Sentinel-5 Precursor NO<sub>2</sub> observations with ground-based measurements in Helsinki. *Atmospheric Measurement Techniques* 13(1):205–218
- Im U, Brandt J, Geels C, Hansen KM, Christensen JH, Andersen MS, ... Balzarini A (2018) Assessment and economic valuation of air pollution impacts on human health over Europe and the United States as calculated by a multi-model ensemble in the framework of AQMEII3. *Atmos Chem Phys* 18(8):5967–5989
- Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahuvaroglu T, ... Giovis C (2005) A review and evaluation of intraurban air pollution exposure models. *J Exposure Sci Environ Epidemiol* 15(2):185–204
- Jerrett M, Burnett RT, Beckerman BS, Turner MC, Krewski D, Thurston G, ... Shi Y (2013) Spatial analysis of air pollution and mortality in California. *Am J Respir Crit Care Med* 188(5):593–599
- Kendall MG (1955) Rank correlation methods. *Br J Psychol* 25(1):86–91
- Krotkov NA, McLinden CA, Li C, Lamsal LN, Celarier EA, Marchenko SV, ... Duncan BN (2016) Aura OMI observations of regional SO<sub>2</sub> and NO<sub>2</sub> pollution changes from 2005 to 2015. *Atmos Chem Phys* 16(7):4605–4629
- Larkin A, Geddes JA, Martin RV, Xiao Q, Liu Y, Marshall JD, ... Hystad P (2017) Global land use regression model for nitrogen dioxide air pollution. *Environ Sci Technol* 51(12):6957–6964
- Li J, Huang J, Wang Y, Yin P, Wang L, Liu Y, ... Li G (2020) Years of life lost from ischaemic and haemorrhagic stroke related to ambient nitrogen dioxide exposure: a multicity study in China. *Ecotoxicol Environ Saf* 203:111018
- Li M, Zhang L (2014) Haze in China: current and future challenges. *Environ Pollut* 189:85–86
- Liu J (2021) Mapping high resolution national daily NO<sub>2</sub> exposure across mainland China using an ensemble algorithm. *Environ Pollut* 279:116932
- Liu M, Huang Y, Ma Z, Jin Z, Liu X, Wang H, ... Bi J (2017) Spatial and temporal trends in the mortality burden of air pollution in China: 2004–2012. *Environ Int* 98:75–81
- Martin RV (2008) Satellite remote sensing of surface air quality. *Atmos Environ* 42(34):7823–7843
- Nie D, Shen F, Wang J, Ma X, Li Z, Ge P, ... Chen M (2021) Changes of air quality and its associated health and economic burden in 31 provincial capital cities in China during COVID-19 pandemic. *Atmos Res* 249:105328
- Niu Z, Liu F, Yu H, Wu S, Xiang H (2021) Association between exposure to ambient air pollution and hospital admission, incidence, and mortality of stroke: an updated systematic review and meta-analysis of more than 23 million participants. *Environ Health Prev Med* 26(1):1–14
- Odum JR, Hoffmann T, Bowman F, Collins D, Flagan RC, Seinfeld JH (1996) Gas/particle partitioning and secondary organic aerosol yields. *Environ Sci Technol* 30(8):2580–2585
- Pankow JF (1987) Review and comparative analysis of the theories on partitioning between the gas and aerosol particulate phases in the atmosphere. *Atmospheric Environment* (1967) 21(11):2275–2283
- Penn E, Holloway T (2020) Evaluating current satellite capability to observe diurnal change in nitrogen oxides in preparation for geostationary satellite missions. *Environ Res Lett* 15(3):034038
- Qin K, Han X, Li D, Xu J, Loyola D, Xue Y, ... Yuan L (2020) Satellite-based estimation of surface NO<sub>2</sub> concentrations over east-central China: a comparison of POMINO and OMNO2d data. *Atmos Environ* 224:117322
- Qin K, Rao L, Xu J, Bai Y, Zou J, Hao N, ... Yu C (2017) Estimating ground level NO<sub>2</sub> concentrations over Central-Eastern China using a satellite-based geographically and temporally weighted regression model. *Remote Sensing* 9(9):950
- Ramacher MOP, Karl M (2020) Integrating modes of transport in a dynamic modelling approach to evaluate population exposure to ambient NO<sub>2</sub> and PM<sub>2.5</sub> pollution in urban areas. *Int J Environ Res Public Health* 17(6):2099
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. *J Am Stat Assoc* 63(324):1379–1389
- Silibello C, Carlino G, Stafoggia M, Gariazzo C, Finardi S, Pepe N, ... Viegi G (2021) Spatial-temporal prediction of ambient nitrogen dioxide and ozone levels over Italy using a Random Forest model for population exposure assessment. *Air Qual Atmos Health* 14(6):817–829
- Tatem AJ (2017) WorldPop, open data for spatial demography. *Scientific Data* 4(1):1–4

- Wong P-Y, Su H-J, Lee H-Y, Chen Y-C, Hsiao Y-P, Huang J-W, ... Spengler JD (2021) Using land-use machine learning models to estimate daily NO<sub>2</sub> concentration variations in Taiwan. *J Clean Prod* 317:128411
- Wu S, Huang B, Wang J, He L, Wang Z, Yan Z, ... Du Z (2021) Spatiotemporal mapping and assessment of daily ground NO<sub>2</sub> concentrations in China using high-resolution TROPOMI retrievals. *Environ Pollut* 273:116456
- Xu H, Bechle MJ, Wang M, Szpiro AA, Vedal S, Bai Y, Marshall JD (2019) National PM<sub>2.5</sub> and NO<sub>2</sub> exposure models for China based on land use regression, satellite measurements, and universal kriging. *Sci Total Environ* 655:423–433
- Yang X, Zheng Y, Geng G, Liu H, Man H, Lv Z, ... de Hoogh K (2017) Development of PM<sub>2.5</sub> and NO<sub>2</sub> models in a LUR framework incorporating satellite remote sensing and air quality model data in Pearl River Delta region China. *Environ Pollut* 226:143–153
- Zhan Y, Luo Y, Deng X, Zhang K, Zhang M, Grieneisen ML, Di B (2018) Satellite-based estimates of daily NO<sub>2</sub> exposure in China using hybrid random forest and spatiotemporal kriging model. *Environ Sci Technol* 52(7):4180–4189
- Zhang Q, Liu C, Wang Y, Gong J, Wang G, Ge W, ... Kan H (2021) Associations of long-term exposure to ambient nitrogen dioxide with indicators of diabetes and dyslipidemia in China: a nationwide analysis. *Chemosphere* 269:128724

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.