Environmental Research Letters

CrossMark

LETTER

A hybrid model approach for estimating health burden from NO$_2$ in megacities in China: a case study in Guangzhou

Baihuiqian He[1,2] , Mathew R Heal[1,6] , Kamilla H Humstad[3] , Liu Yan[4] , Qiang Zhang[4] and Stefan Reis[2,5,6]

[1] School of Chemistry, University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh, EH9 3FJ, United Kingdom
[2] NERC Centre for Ecology & Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, United Kingdom
[3] School of GeoSciences, University of Edinburgh, Institute of Geography, Drummond Street, Edinburgh, EH8 9XP, United Kingdom
[4] Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, People's Republic of China
[5] University of Exeter Medical School, European Centre for Environment and Health, Knowledge Spa, Truro, TR1 3HD, United Kingdom
[6] Author whom to any correspondence should be addressed.

E-mail: srei@ceh.ac.uk and M.Heal@ed.ac.uk

Abstract

Background: Nitrogen dioxide (NO$_2$) poses substantial public health risks in large cities globally. Concentrations of NO$_2$ shows high spatial variation, yet intra-urban measurements of NO$_2$ in Chinese cities are sparse. The size of Chinese cities and shortage of some datasets is challenging for high spatial resolution modelling. The aim here was to combine advantages of dispersion and land-use regression (LUR) modelling to simulate population exposure to NO$_2$ at high spatial resolution for health burden calculations, in the example megacity of Guangzhou. *Methods:* Ambient concentrations of NO$_2$ simulated by the ADMS-Urban dispersion model at 83 'virtual' monitoring sites, selected to span both the range of NO$_2$ concentration and weighting by population density, were used to develop a LUR model of 2017 annual-mean NO$_2$ across Guangzhou at 25 m $\times$ 25 m spatial resolution. *Results:* The LUR model was validated against both the 83 virtual sites (adj R^2: 0.96, RMSE: 5.48 μg m^{-3}; LOOCV R^2: 0.96, RMSE: 5.64 μg m^{-3}) and, independently, against available observations ($n = 11$, $R^{2:}$ 0.63, RMSE: 18.0 μg m^{-3}). The modelled population-weighted long-term average concentration of NO$_2$ across Guangzhou was 52.5 μg m^{-3}, which contributes an estimated 7270 (6960–7620) attributable deaths. Reducing concentrations in exceedance of the China air quality standard/WHO air quality guideline of 40 μg m^{-3} would reduce NO$_2$-attributable deaths by 1900 (1820–1980). *Conclusions:* We demonstrate a general hybrid modelling method that can be employed in other cities in China to model ambient NO$_2$ concentration at high spatial resolution for health burden estimation and epidemiological study. By running the dispersion model with alternative mitigation policies, new LUR models can be constructed to quantify policy effectiveness on NO$_2$ population health burden.

1. Introduction

It is expected that by 2025 roughly 70% of the global population will live in urban areas (United Nations 2014). China is at the forefront of this trend of rapid economic development and urbanisation, which has led to the emergence of megacities (>10 million people) and megacity clusters and concomitant problems of poor air quality (Chan and Yao 2008, Wang

and Hao 2012, Cohen *et al* 2017, Sun and Zhou 2017, Airvisual 2019).

Nitrogen dioxide (NO$_2$) is a key urban air pollutant with well-documented public health impacts. Exposure to ambient NO$_2$ is associated with increased mortality (WHO 2013, Faustini *et al* 2014, Crouse *et al* 2015). It is also linked to respiratory symptoms such as bronchitis, asthma and reduced lung-function growth in children, and to cardiovascular and Alzheimer's

disease (Brunekreef and Holgate 2002, Weinmayr *et al* 2010, Chiusolo *et al* 2011, Luo *et al* 2016, Yan *et al* 2016, Anenberg *et al* 2018).

Ambient concentrations of NO_2 in many Chinese cities exceed the air quality standard (Ministry of Environmental Protection 2018a). The predominant sources of NO_2 are road traffic, other combustion processes in domestic, commercial and industrial settings, including cooking, and (where relevant) shipping (Fu *et al* 2017, Liu *et al* 2017, Ding *et al* 2018, Zheng *et al* 2018a). Due to the widespread sources and relatively short lifetime of NO_2, its concentrations are strongly spatially varying (Beirle *et al* 2011, Cyrys *et al* 2012, Gurung *et al* 2017). The China National Environmental Monitoring Centre has implemented a nationwide monitoring network for routine measurements of ambient NO_2 (and other pollutants), but despite the increase in the number of monitoring sites across China in recent years, air quality monitoring networks cannot provide the highly spatially resolved concentration fields needed to accurately evaluate the population health burden of exposure to NO_2. Measurements also cannot provide the data needed to evaluate future mitigation scenarios that address the challenges of reducing population exposure to NO_2.

Instead, modelling approaches are needed. Urban-scale air pollution models fall into two basic categories: dispersion models (Visscher 2013), and land-use regression (LUR) models (Briggs *et al* 1997, Jerrett *et al* 2005). A dispersion model strives to accurately simulate the fundamental physical–chemical processes in the atmosphere from emission source to all selected receptor locations, whilst a LUR model establishes significant predictor variables for concentrations at known locations and uses the statistical relationship to estimate concentrations at all other locations. However, modelling NO_2 concentrations in China is challenging as the detailed input data required for modelling that exist for Western cities are much more limited for Chinese cities (He *et al* 2018). Chinese cities are large in both population number and geographical area compared with cities in Europe and North America (United Nations 2017), and are also characterised by high population density: an average of 5000 per km^2 in urban areas with a population >500 000, which is almost double that for EU cities (2900 per km^2) and triple that for North American cities (1600 per km^2) (Demographia 2019). The lack of a full suite of input data is a particular challenge for dispersion models, as is the computational demand of running such a model over the size of a typical Chinese city. LUR models are more flexible in terms of data requirement but can generally only be used to estimate long-term-average pollutant concentration fields, and they lack predictive capability. The lack of measurement data and difficulty in modelling limit epidemiological studies and health impact assessments in China (Cao *et al* 2011, Gu *et al* 2017, Sun and Zhou 2017, Huang *et al* 2018, Li *et al* 2018, Lin *et al* 2018, Ji *et al* 2019).

In this work, we demonstrate a practical approach to simulating annual-average concentrations of NO_2 in Chinese megacities in which we combine the advantages of both dispersion and LUR modelling (Mölter *et al* 2010). We use an urban dispersion model to simulate annual-average NO_2 concentrations at strategically chosen 'virtual' sites. This network of virtual measurements is then used to develop an LUR model to estimate NO_2 concentrations over the whole region at 25 m × 25 m resolution, from which mortality health burdens can be calculated. A particular advantage of this approach is that it can be used to derive spatially-resolved LUR models for future scenarios that are rooted in process-based dispersion model simulations of those scenarios. Here we demonstrate our methodology for the example Chinese megacity of Guangzhou, which is the third largest city in China. We use our hybrid model to estimate the scale of long-term health burden currently existing in Guangzhou due to the spatially varying NO_2 concentrations.

2. Method

2.1. Study design
The city of Guangzhou is located on the north side of the Pearl River Delta (figure 1). It has a total area of 7434 km^2 and a population of 14 million, divided into six districts (Conghua, Guangzhou city centre, Huaxian, Nansha, Panyu, and Zengcheng) (table 1). Concentrations of NO_2 are measured at 11 sites (figure 1). The workflow for this study is shown in figure 2. The ADMS-Urban dispersion model v4.1 (CERC 2017) was used to derive NO_2 concentrations at 83 additional strategically-selected receptors locations across Guangzhou, which were then used to develop an LUR model for NO_2 concentrations across the whole city.

2.2. The ADMS-urban model
The ADMS-Urban dispersion model v4.1 (Carruthers *et al* 1994, CERC 2017) is a widely-used (McHugh *et al* 1997, Carruthers *et al* 2000, Di Sabatino *et al* 2007, Righi *et al* 2009) quasi-Gaussian model for simulating dispersion in the atmosphere of continuous releases from a range of explicit sources in an urban area. The NO_x–NO_2 chemistry is simulated using the 8-reaction set of Venkatram *et al* (1994) that includes reactions with ozone and hydrocarbons. This General Reaction Set has been thoroughly evaluated against measurements (see e.g. Tonnesen and Jeffries 1994, Venkatram *et al* 1994, Chaney *et al* 2011, Carruthers *et al* 2017). Surface deposition is included in the model. The sources of data used for the ADMS-urban dispersion modelling of NO_2 concentrations at monitor locations and at the virtual sites in this study are given in supplementary information (SI) table S2. Gridded emissions of NO_x and VOCs for the year 2016 were obtained from the Multi-resolution Emission Inventory for China (Zheng *et al* 2018a) and downscaled to

Figure 1. (A) The location of Guangzhou in China. (B), (C) The locations of the 11 NO$_2$ monitoring sites and 83 virtual receptor sites for the dispersion modelling, which were stratified into the six districts of Guangzhou indicated by the boundaries. The labels on the maps are the Guangzhou codes for the monitoring sites; additional site location description is provided in supplementary information table S1, available online at stacks.iop.org/ERL/14/124019/mmedia. Map tiles by Stamen Design (2019), under CC BY 3.0. Data by OpenStreetMap, under ODbL.
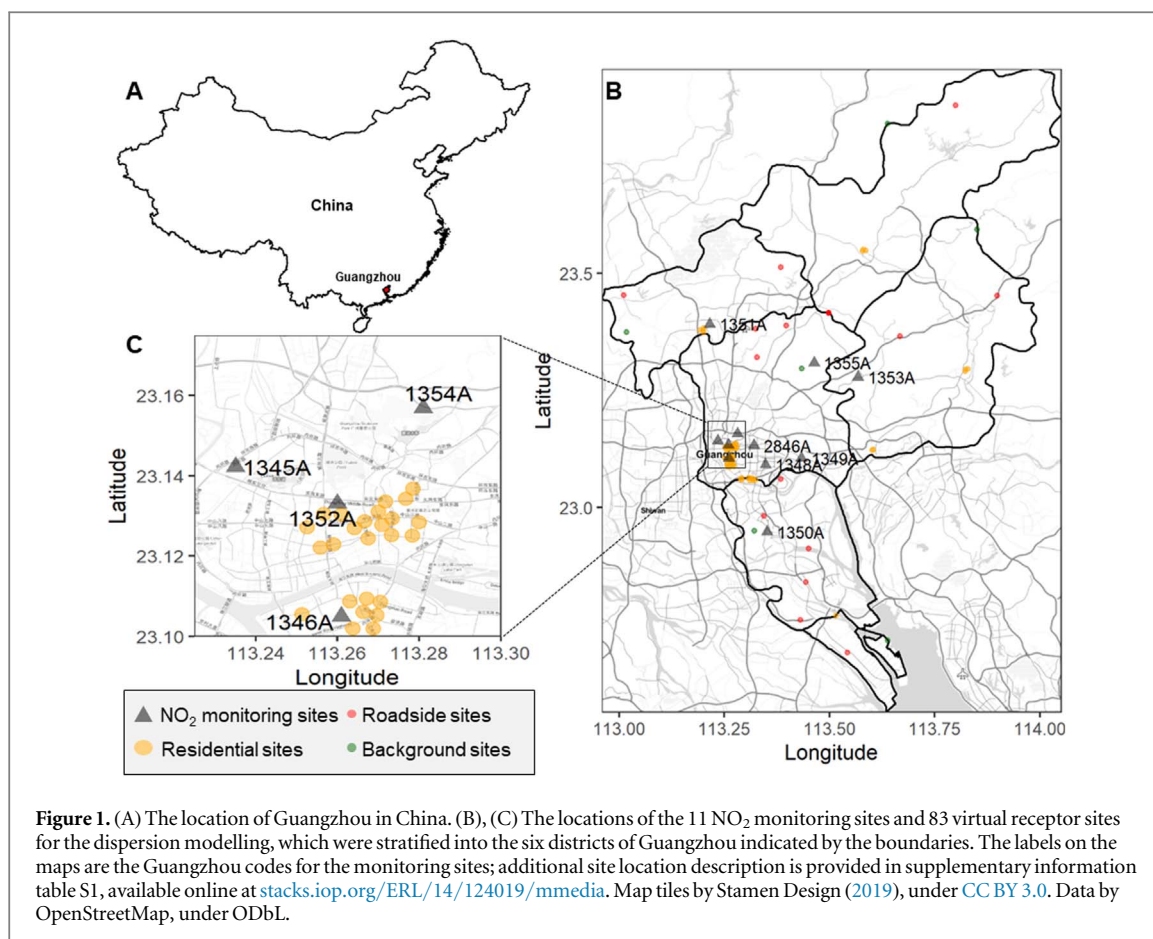
Table 1. Population of each district, the number of residential and roadside sites selected and the number residential sites also satisfying the definition of roadside site.

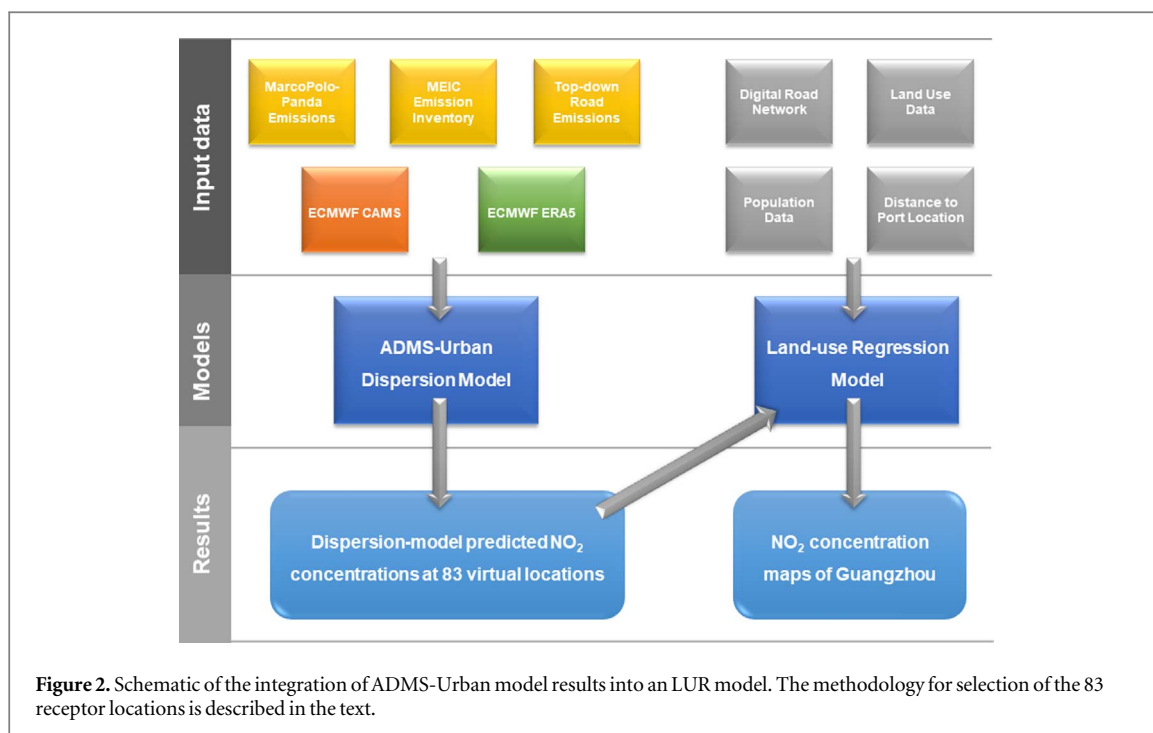| District	Population	No. of residential sites	No. of residential sites also fitting roadside definition	No. of roadside sites
Conghua	596 816	3	0	1
Guangzhou city centre	7909 399	37	1	6
Huaxian	916 916	4	1	2
Nansha	110 923	1	0	1
Panyu	1986 223	10	0	5
Zengcheng	1120 186	5	0	2

4 km × 4 km horizon resolution using various spatial proxies (Geng *et al* 2017, Zheng *et al* 2018b). Shipping emissions were taken from MarcoPolo-Panda for the year 2014 (MarcoPolo-Panda 2017). A map of the total NO$_x$ emissions is shown in SI figure S1. Motorway, trunk, primary, secondary, and tertiary roads, as defined by OpenStreetMap, were explicitly modelled as road sources (OpenStreetMap 2018). Emissions from roads were calculated by assigning total on-road traffic emissions to the explicitly modelled roads according to road type and total length of each road, as described in SI section A.

Wind speed and direction at 10 m above the ground, and cloud cover were obtained from ECMWF-ERA5 (ECMWF 2019). Background concentrations of NO$_2$, NO$_x$, and O$_3$ were obtained from the Copernicus Atmosphere Monitoring Service

(CAMS) (ECMWF 2019). As the prevailing wind direction in Guangzhou is from the northeast (SI figure S2), the background site for pollutant concentrations for the model was chosen to be a rural location to the northeast of the model domain (SI figure S3). The model was evaluated using the NO$_2$ concentrations at the 11 monitoring sites shown in figure 1.

2.3. Selection of the ADMS-urban virtual receptor sites

As the main purpose of modelling NO$_2$ concentrations is to estimate NO$_2$ where people live, the virtual sites were primarily selected to be proportionally located across the residential areas of Guangzhou. The six districts have a large range in population (table 1) so site selection was first stratified across the six districts

Figure 2. Schematic of the integration of ADMS-Urban model results into an LUR model. The methodology for selection of the 83 receptor locations is described in the text.

so as to ensure that sites in the virtual network were located across the full domain area. Centroids of each cell of population data (100 m × 100 m) (Worldpop 2015) were extracted as potential virtual receptor sites. Three types of virtual sites were defined for each district so as also to span the range in anticipated NO$_2$ concentration:

- *Background site:* centroid of the population cell with greatest distance to the nearest road.

- *Roadside sites:* centroids of the population cells with a distance to nearest road of no more than 5 m.

- *Residential sites:* centroids of the population cells with the highest population density.

In each district, residential sites were selected first, with number of residential sties proportional to the relative proportion of the total Guangzhou population in that district (the specific number is given in table 1). One residential site in two districts also fitted the definition of roadside site (see table 1), but these were retained under the category of residential. All other sites that satisfied the definition of roadside site (17 across the six districts, table 1) were selected as receptor locations within the roadside site category. A background site was selected in each district. An additional overarching criterion was a minimum distance of 300 m between any pair of sites in each district to ensure that all sites were distributed across the range of localities in the study area. The locations of the 83 sites are shown in figure 1. ADMS-Urban was used to simulate annual-mean NO$_2$ concentrations at these locations at a height of 1.5 m.

2.4. LUR model predictor variables

Sources of data used to develop the LUR model are given in SI table S2. Potential predictor variables (SI table S5) for the LUR model were chosen based on expectation that emission sources, dispersion, and physical geography may contribute to NO$_2$ concentration variation in urban areas (ESCAPE 2008). The inclusion of different buffer sizes allows for potential different influences of a predictor variables over different distances from the receptor (Su *et al* 2009, Beelen *et al* 2013).

Road lengths in a buffer had to be used as surrogates for traffic flow and fleet composition, as neither of these data are publicly available for Guangzhou. The variables of number of people and artificial (i.e. non-natural) area within a buffer do not directly affect NO$_2$ concentration, but they are indirectly related to road transport emissions and to domestic, industrial and electrical generation emissions which contribute to total NO$_x$ emission in Guangzhou (SI figure S4). Distance to nearest port and its derivatives were included as non-buffer variables to account for the impacts of shipping, Guangzhou is close to the South China Sea, and is a major port, with ship activities significantly contributing to total NO$_x$ emissions in the southern part of the model domain (Liu *et al* 2016, Johansson *et al* 2017) (SI figure S4). A maximum distance of 35 000 m was specified for the distance to port variable as the dispersion modelling at the virtual sites indicated that NO$_2$ concentration decreased to urban background level at this distance from nearest port as shown in SI figure S5.

Green, i.e. not built-up, areas are assumed to mitigate NO$_2$ concentration since they are not sources of NO$_2$ and also act as areas over which NO$_2$

concentrations disperse and dilute (and deposit), hence the *a priori* direction of effect of the green area variable is negative (Jim and Chen 2008). Coordinates and altitude were included to reflect physical geography influence and wind speed and direction proxies.

2.5. LUR model development and validation
A stepwise multiple regression approach was used to select the potential predictor variables to maximise the adjusted percentage explained variance (R^2) and minimise the root mean square error (RMSE) (ESCAPE 2008). First, all predictor variables were individually regressed against the dependent variable data (annual-mean NO_2 concentrations modelled by ADMS-urban at the 83 receptor sites). The predictor variable which explained the most NO_2 concentration variation (highest R^2) and with a coefficient in the *a priori*-defined direction (SI table S5) formed the initial model. The univariate linear regression process was repeated with the remaining variables. A variable with the highest increase in adjusted R^2 was added into the initial model if it met the following criteria: (1) the increase in adjusted R^2 was greater than 0.01; (2) the coefficients of this variable and the variables already in the model were in accord with the *a priori*-defined direction. This process was repeated until no further variable satisfied the criteria. In the final step, variables with *p*-value greater than 0.1 were removed from the model starting from the variable with highest *p*-value.

Tests were then performed to check multicollinearity and influential observations. Multicollinearity in the variables was checked using Variance Inflation Factor (VIF). Variables with VIF values greater than three were removed from the model. Extreme values or many zero values in a variable data can skew the final model and this can be indicated by Cook's distance >1. In this study, no observation was removed. The model was validated using both the measurement and virtual site concentration data.

2.6. Health burden calculation
Calculation of total premature deaths from concentrations of NO_2 followed the methodology described by Walton *et al* (2015). We used the association with all-cause mortality of 2.45% (95% CI: 2.34%, 2.58%) per 10 μg m^{-3} elevation of NO_2 reported by Zhang *et al* (2011). This health risk coefficient was used since it was derived from data in Shenyang, a province in China. The number of deaths in 2017 in Guangzhou was 60 900 (Guangzhou Statistics Bureau 2018). The population data (100 m × 100 m) was resampled to the resolution of the concentration map (25 m × 25 m). The population-weighted average concentration (*E*) for NO_2 across the whole of Guangzhou was calculated as follows

$$E = \frac{1}{Pop} \sum_i C_i Pop_i \qquad (1)$$

where C_i and Pop_i are the concentration and the number of people in each cell *i* of the concentration map.

The attributable deaths from exposure to ambient NO_2 in Guangzhou was calculated by multiplying the attributable fraction by number of all-cause deaths (equations (2)–(4)), where *RR* refers to relative risk

$$RR = 1.0245^{(E/10)}, \qquad (2)$$

$$AF = (RR - 1)/RR, \qquad (3)$$

Attributable death = the number of deaths × AF.
$$\qquad (4)$$

3. Results

3.1. ADMS-urban model validation
The ADMS-Urban model showed good statistical performance against the annual-mean NO_2 concentrations at the 11 monitoring sites in Guangzhou (figure 3(A)). The model explains 72% of the spatial variation in the measured NO_2, with an RMSE of 17.7 μg m^{-3}, albeit with a tendency to overpredict NO_2 concentrations (MB = 11.2 μg m^{-3}, NMB = 0.22). A likely explanation for the general overprediction is that the precise height of each monitoring site is unknown. The only information provided to us is that they range in elevation from 2 to 20 m, which is consistent with anecdotal information that pollutant monitoring sites in Chinese cities are often situated on tops of buildings, and thus at higher elevation (where NO_2 concentrations would be lower) than the receptor height used in the model. The greatest model overprediction at site 1354A (figure 3(A)) is likely explained by the fact that this site is located in a park in the city centre and is surrounded by dense vegetation whose effects on NO_2 concentration (through dispersion, deposition, and temperature reduction) are not sufficiently captured by the model. As the distance to nearest road for all 11 monitoring sites is greater than 5 m, the observed data can not verify the model for roadside locations where highest NO_2 concentrations are anticipated, but at the other end of the concentration range the model predicted well the background site 1355A which is located in an area of natural land cover (Maofengshan). It is also important to recognise that the observational data have uncertainty. Measurement errors have been reported previously in China but more recent literature indicates that data is more reliable since 2013 (Yuyu *et al* 2012, Ghanem and Zhang 2014, Stoerk 2016). The monitoring specification HJ654-2013 (Ministry of Environmental Protection 2018b) also specifies detailed calibration procedures for each instrument such as zero and scale noises, error of indication, zero and span drifts tests. However, overestimation of NO_2 concentrations using instruments based on the chemiluminescence method (TEI Model 42i from Thermo Fisher Scientific
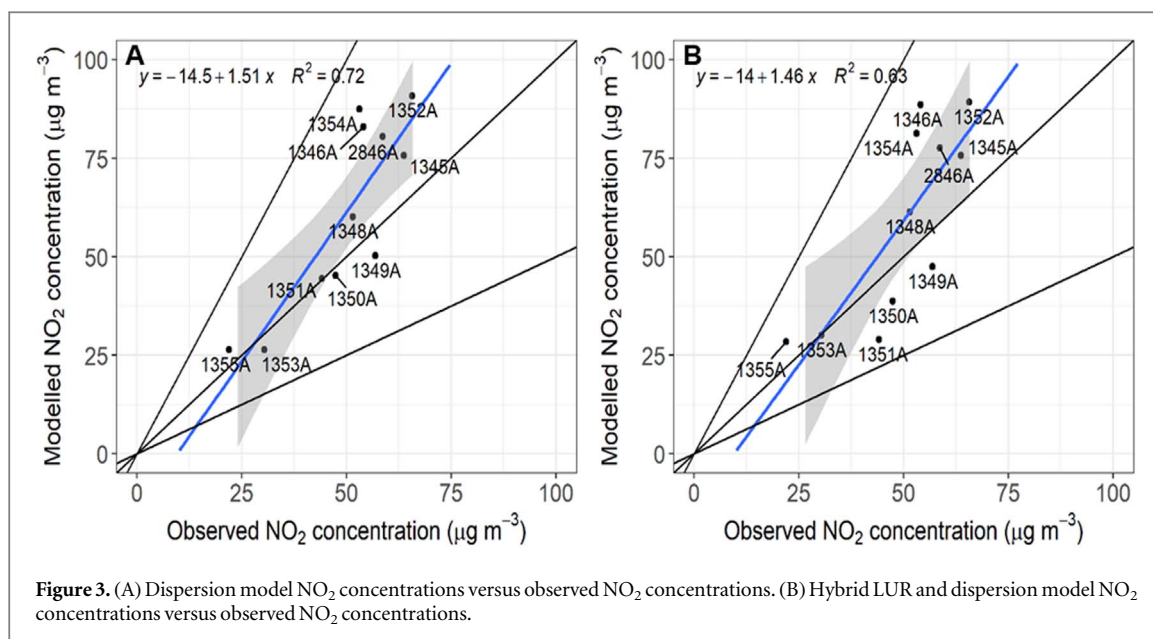
Figure 3. (A) Dispersion model NO_2 concentrations versus observed NO_2 concentrations. (B) Hybrid LUR and dispersion model NO_2 concentrations versus observed NO_2 concentrations.

Table 2. The final LUR model variables and their coefficients and standard errors.

Variable	Definition	Coefficient	Standard error
Intercept	/	33.09	1.77
pop_5000	Number of people within a 5000 m buffer	2.875×10^{-5}	1.087×10^{-6}
sq_dis_port	Squared distance to the nearest port	-9.543×10^{-9}	1.832×10^{-9}
all_len_25	Length of all type of roads within a 25 m buffer	6.362×10^{-2}	1.261×10^{-2}

Table 3. LUR model performance evaluated by training (R^2 and RMSE), LOOCV (R^2 and RMSE), and observation (R^2 and RMSE).

Evaluation method	R^2	RMSE ($\mu g\,m^{-3}$)
Using 83 virtual sites	0.96	5.48
LOOCV using 83 virtual sites	0.96	5.64
Using measurement at 11 monitoring sites	0.63	18.0

Inc., USA) ranged from 6% to 280% depending on the composition of the oxidation products of NO_x has been reported (Xu *et al* 2013). It has not been feasible to robustly quantify uncertainty for the NO_2 measurements in Guangzhou within the scope of this study, as documentation of the processes applied for data quality control and assurance (QA/QC) and the operating principles at each monitoring site is not publicly available.

3.2. Hybrid LUR model and validation

The final LUR model contained three variables (table 2): *pop_5000* (the number of people within a 5000 m buffer) *sq_dis_port* (the squared distance to the nearest port), and *all_len_25* (the length of all type of roads within a 25 m buffer). The three variables respectively represent the influence of people (indirectly related to residential, industrial, and traffic emission, and inversely related to green space), shipping emissions, and traffic emissions. Table 3

summarizes the results of hybrid LUR model evaluation against both the 83 virtual sites and the 11 monitoring sites and figure 3(B) shows the scatter plot for the LUR model predictions at the 11 monitoring sites.

Against the 83 virtual sites, LOOCV evaluation shows the hybrid LUR model explains 96% of the spatial variation of NO_2 concentration with an RMSE of $5.64\ \mu g\,m^{-3}$. The NO_2 concentrations at the monitoring site provide an additional evaluation of the hybrid LUR model independent of the NO_2 data used to construct the model (table 3 and figure 3(B)). For this plot, $R^2 = 0.63$ and MB $= 9.08\ \mu g\ m^{-3}$, which corresponds to a NMB of 18.2%. These statistical uncertainties seem reasonable for intra-urban modelling/ mapping of a domain of this size and with the paucity of detailed emissions data to support modelling. Figure 3(B) shows that the hybrid model slightly overestimates NO_2 concentrations, which as discussed earlier is likely the consequence of the model predicting NO_2 concentration at 1.5 m height, which is lower than the height of most monitoring sites (2–20 m). Closer inspection of figure 3(B) shows that the model generally over-predicts more at locations with higher NO_2 concentrations which can be rationalised because these are the locations most affected by local traffic emissions for which there is insufficiently detailed input data. As noted above, it is also important to recognise that there is uncertainty in the observed values.

Figure 4. The final hybrid dispersion-LUR model of NO$_2$ concentrations in Guangzhou. The most polluted district is Guangzhou city centre. NO$_2$ concentrations are higher at locations around the road links, industrial areas, and close to the ports (in the southern part of the model domain).

3.3. Concentration map and health burden calculation

Figure 4 shows the spatial distribution of annual-mean NO$_2$ concentration across Guangzhou in 2017. Modelled concentrations vary between 21.5 μg m^{-3} in the most rural areas in the north of the domain to 99.7 μg m^{-3} in the most polluted district of Guangzhou city centre. The map indicates that, as expected, concentrations of NO$_2$ are highest on road links, particularly those with heavy traffic as shown on the expanded map of the city centre. The model NO$_2$ concentration show high values at the boundary of Nansha, which is likely due to shipping (SI figure S4). The population-weighted annual-mean concentration of NO$_2$ across Guangzhou is 52.5 μg m^{-3}. Even this population-weighted concentration exceeds the Chinese air quality standard and WHO guideline of 40 μg m^{-3} (WHO 2006, Ministry of Environmental Protection 2012), but figure 4 indicates that, in large areas of Guangzhou, concentrations of NO$_2$ are substantially in excess of this value.

The modelled number of deaths attributable to long-term exposure to NO$_2$ in Guangzhou in 2017 was 7270 (6960–7620; range based only on the 95% CI for the health risk coefficient). For comparison, the total number of deaths from road-traffic accidents in Guangzhou was 847 in 2015 (Guangzhou Municipal Public Security Bureau 2016). Approximately 60% of people in Guangzhou experienced NO$_2$ concentration above the China air quality standard/WHO guideline of 40 μg m^{-3}. The air quality plan (2016–2025) published by the Guangzhou government to combat air pollution problems sets the target of meeting the NO$_2$

air quality standard of 40 μg m^{-3} by 2020 (Guangzhou Government 2017). Our model predicts that if all areas with NO$_2$ concentrations greater than 40 μg m^{-3} were reduced to the air quality standard, the number of lives saved would be 1900 (1820–1980) compared to the 2017 estimate.

4. Discussion

Compared to LUR, dispersion modelling has the advantage of grounding simulations of pollutant concentrations for different scenarios in a process-based simulation of the controlling processes; but the approach requires detailed input data on emissions (e.g. emission inventories), meteorological parameters, and background concentrations, which are currently not fully available/accessible in most cities in China. Aleksankina *et al* (2018, 2019) have investigated model uncertainties in relation to emission input data in detail, highlighting that atmospheric chemistry transport models show relatively robust responses to changes in emission input data. LUR is an efficient modelling approach, but in areas with limited monitoring sites such as Guangzhou, with only 11 monitoring sites, the selected variables may overfit the model and hence cause bias in health-effect estimates (Basagaña *et al* 2012). LUR is also limited to modelling measured air pollutants and cannot predict air pollutant concentrations under potential future emissions scenarios. However, despite the disadvantages mentioned above, LUR is relatively cheap and easy to implement; and once input datasets are assembled,

LUR models can be readily adapted to new 'measurement' data without needing to consider complex physicochemical processes.

The challenge of high spatial resolution air pollution modelling has driven integrated data and modelling approaches (Johnson *et al* 2010, Beevers *et al* 2012, Hao *et al* 2016, Yang *et al* 2017, Hood *et al* 2018, Xue *et al* 2019). For megacities of large area like Guangzhou, without additional monitoring data, our solution for simulation of an NO_2 concentration map has been to derive a hybrid modelling approach that uses a virtual network of receptor locations for dispersion and LUR modelling. This approach gives flexibility to choose the input sites (such as weighting towards residential population (around 60% of the receptors are in Guangzhou city centre) and number of sites (83 receptor sites). It also takes less time to generate a high-resolution map over a large domain compared with dispersion modelling only. In this hybrid approach, dispersion modelling is used to conduct scenario analysis and LUR is used for interpolating the spatial trends. This approach can be easily used to derive concentration maps to calculate health burdens (including for future scenarios) or for epidemiological studies.

There are of course uncertainties, as for any modelling approach. Even though we have sought to play to the strengths of the available data for Guangzhou, data accessibility and quality are generally an issue for China (He *et al* 2018). For example, detailed industrial emissions as point sources for the dispersion model are not accessible in Guangzhou. Similarly, detailed traffic flow and speed are not available, hence road length was used as a proxy in the LUR model. Also daily vehicle NO_x emissions on different road types in Beijing had to be used to estimate emissions on different roads in Guangzhou. Since detailed land-use data are not available, satellite observation derived data Globeland30 were used (Globeland30 2010). It is extremely difficult to get access to monitoring data in Guangzhou; several websites provide real time data (Aqicn 2019, CNEMC 2019), but historical data are not available for download from official sites and the detailed location and environment of the monitoring sites are unclear. State-controlled sites were used in this study, but no information about monitoring site heights was available other than they range in elevation from 2 to 20 m. The quality of monitoring data is unknown. ADMS requires concentrations of NO and NO_2 background concentrations but only NO_2 was reported at each monitoring site, therefore ECMWF-CAMS re-analysis data was used for the boundary concentration. No roadside monitoring sites were available, so neither the dispersion nor the LUR model simulations could be validated at highest NO_2 concentrations. Since the LUR model used ADMS-Urban modelled concentrations as input, the hybrid model inevitably transmits any bias in the dispersion modelling to the final model.

In terms of the health burden calculation, detailed census data is not available and UN modelled population data (Worldpop 2015) were used instead. A health risk coefficient from a single-pollutant model was used which may overestimate the actual effects due to NO_2 (COMEAP 2018).

5. Conclusions

Given the current challenges related to data availability, accessibility and quality in Chinese megacities, and their large spatial area, a single model struggles to provide high spatio-temporal resolution air pollution maps for the whole city. In this work, a combined dispersion and LUR model approach to simulate annual average NO_2 for health burden assessments was demonstrated for the megacity of Guangzhou. Ambient concentrations of NO_2 simulated by the ADMS-Urban dispersion model at 83 'virtual' monitoring sites, selected to span both the range of NO_2 concentration and weighting by population density, were used to develop a highly spatial resolved LUR model. To our knowledge, this is the first attempt to apply this approach to maximise the advantages of both modelling approaches and to overcome data shortcomings in relation to data availability in China. This method can be employed in other cities in China, or globally, to model ambient NO_2 concentration at high spatial resolution to investigate the effectiveness of potential mitigation policies. The dispersion model simulates the effect of different emissions scenarios on concentrations at the network of receptor locations, from which LUR modelled maps are constructed to estimate future concentrations and health burdens across the whole domain.

Data availability statement

Any data that support the findings of this study are openly available at DOI (https://datashare.is.ed.ac.uk/handle/10283/3420).

ORCID iDs

Baihuiqian He ⓘ https://orcid.org/0000-0003-1994-4151

Mathew R Heal ⓘ https://orcid.org/0000-0001-5539-7293

Kamilla H Humstad ⓘ https://orcid.org/0000-0003-1017-2191

Liu Yan ⓘ https://orcid.org/0000-0001-9101-5168

Stefan Reis ⓘ https://orcid.org/0000-0003-2428-8320

References

Airvisual 2019 World Most Polluted Countries in 2018—$PM_{2.5}$ Ranking | AirVisual [WWW Document](https://airvisual.com/world-most-polluted-countries)(Accessed: 9 April 2019)

Aleksankina K, Heal M R, Dore A J, Van Oijen M and Reis S 2018 Global sensitivity and uncertainty analysis of an atmospheric chemistry transport model: the FRAME model (version 9.15.0) as a case study *Geosci Model Dev* **11** 1653–64

Aleksankina K, Reis S, Vieno M and Heal M R 2019 Advanced methods for uncertainty assessment and global sensitivity analysis of a Eulerian atmospheric chemistry transport model *Atmos. Chem. Phys.* **19** 2881–98

Anenberg S C *et al* 2018 Estimates of the global burden of ambient $PM_{2.5}$, Ozone, and NO_2 on asthma incidence and emergency room visits *Environ. Health Perspect.* **126** 107004

Aqicn 2019 中国空气污染: 实时空气质量指数地图 (China Air Pollution: Real-time Air Quality Index Map) [WWW Document](https://aqicn.org/map/china/cn/#@g/39.5901/88.9673/5z) (Accessed: 23 April 2019)

Basagaña X *et al* 2012 Effect of the number of measurement sites on land use regression models in estimating local air pollution *Atmos. Environ.* **54** 634–42

Beelen R *et al* 2013 Development of NO_2 and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe—the ESCAPE project *Atmos. Environ.* **72** 10–23

Beevers S D, Kitwiroon N, Williams M L and Carslaw D C 2012 One way coupling of CMAQ and a road source dispersion model for fine scale air pollution predictions *Atmos. Environ.* **59** 47–58

Beirle S, Boersma K F, Platt U, Lawrence M G and Wagner T 2011 Megacity emissions and lifetimes of nitrogen oxides probed from space *Science* **333** 1737–9

Briggs D J, Collins S, Elliott P, Fischer P, Kingham S, Lebret E, Pryl K, Van Reeuwijk H, Smallbone K and Van Der Veen A 1997 Mapping urban air pollution using GIS: a regression-based approach *Int. J. Geogr. Inf. Sci.* **11** 699–718

Brunekreef B and Holgate 2002 Air pollution and health *Lancet* **360** 1233–42

Cao J, Yang C, Li J, Chen R, Chen B, Gu D and Kan H 2011 Association between long-term exposure to outdoor air pollution and mortality in China: a cohort study *J. Hazard. Mater.* **186** 1594–600

Carruthers D J, Edmunds H A, Lester A E, McHugh C A and Singles R J 2000 Use and validation of ADMS-Urban in contrasting urban and industrial locations *Int. J. Environ. Pollut.* **14** 364

Carruthers D J, Holroyd R J, Hunt J C R, Weng W S, Robins A G, Apsley D D, Thomson D J and Smith F B 1994 UK-ADMS: a new approach to modelling dispersion in the earth's atmospheric boundary layer *J. Wind Eng. Ind. Aerod.* **52** 139–53

Carruthers D J, Stocker J R, Ellis A, Seaton M D and Smith S E 2017 Evaluation of an explicit NO_x chemistry method in AERMOD *J. Air Waste Manag. Assoc.* **67** 702–12

CERC 2017 ADMS-Urban User Guide [WWW Document](https://cerc.co.uk/environmental-software/assets/data/doc_userguides/CERC_ADMS-Urban4.1.1_User_Guide.pdf) (Accessed: 21 May 2019)

Chan C K and Yao X 2008 Air pollution in mega cities in China *Atmos. Environ.* **42** 1–42

Chaney A M, Cryer D J, Nicholl E J and Seakins P W 2011 NO and NO_2 interconversion downwind of two different line sources in suburban environments *Atmos. Environ.* **45** 5863–71

Chiusolo M *et al* 2011 Short-term effects of nitrogen dioxide on mortality and susceptibility factors in 10 Italian Cities: the Epiair study *Environ. Health Perspect.* **119** 1233–8

CNEMC 2019 实时数据-中国环境监测总站 (Real-time data-China National Environmental Monitoring Center) [WWW Document](http://cnemc.cn/sssj/)(Accessed: 23 April 2019)

Cohen A J *et al* 2017 Estimates and 25 year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015 *Lancet* **389** 1907–18

COMEAP 2018 Associations of Long-Term Average Nitrogen Dioxide with Mortality, UK Department of Health Committee on the Medical Effects of Air Pollutants. PHE report no. 2018238 (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/734799/COMEAP_NO2_Report.pdf)

Crouse D L *et al* 2015 Ambient PM2.5, O_3, and NO_2 exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC) *Environ. Health Perspect.* **123** 1180–6

Cyrys J *et al* 2012 Variation of NO_2 and NO_x concentrations between and within 36 European study areas: Results from the ESCAPE study *Atmos. Environ.* **62** 374–90

Demographia 2019 World Urban Areas [WWW Document](http://demographia.com/db-worldua.pdf) (Accessed: 9 April 2019)

Di Sabatino S, Buccolieri R, Pulvirenti B and Britter R 2007 Simulations of pollutant dispersion within idealised urban-type geometries with CFD and integral models *Atmos. Environ.* **41** 8316–29

Ding J, van der A R J, Mijling B, Jalkanen J-P, Johansson L and Levelt P F 2018 Maritime NO_x emissions over Chinese seas derived from satellite observations *Geophys. Res. Lett.* **45** 2031–7

ECMWF 2019 ECMWF | Public Datasets [WWW Document] (Accessed: 30 March 2019) (https://doi.org/10.24381/cds.adbb2d47)

ESCAPE 2008 ESCAPE Study Manual [WWW Document] (http://escapeproject.eu/manuals/ESCAPE-Study-manual_x007E_final.pdf) (Accessed: 21 May 2019)

Faustini A, Rapp R and Forastiere F 2014 Nitrogen dioxide and mortality: review and meta-analysis of long-term studies *Eur. Respir. J.* **44** 744–53

Fu M, Liu H, Jin X and He K 2017 National- to port-level inventories of shipping emissions in China *Environ. Res. Lett.* **12** 114024

Geng G, Zhang Q, Martin R V, Lin J, Huo H, Zheng B, Wang S and He K 2017 Impact of spatial proxies on the representation of bottom-up emission inventories: a satellite-based analysis *Atmos. Chem. Phys.* **17** 4131–45

Ghanem D and Zhang J 2014 Effortless perfection:' do Chinese cities manipulate air pollution data? *J. Environ. Econ. Manag.* **68** 203–25

Globeland30 2010 GLC30 Information Service [WWW Document](http://globallandcover.com/GLC30Download/index.aspx) (Accessed: 19 May 2019)

Gu Y, Lin H, Liu T, Xiao J, Zeng W, Li Z, Lv X and Ma W 2017 The interaction between ambient PM_{10} and NO_2 on Mortality in Guangzhou, China *Int. J. Environ. Res. Public. Health* **14** 1381

Guangzhou Government 2017 Environmental Air Quality Standardization Plan of Guangzhou City (2016–2025) [WWW Document]. Website Guangzhou Gov (http://gz.gov.cn/gzgov/s2811/201712/57727a1d77354f5dbc22bb5831aa7d93.shtml) (Accessed: 25 September 2018)

Guangzhou Municipal Public Security Bureau 2016 The Number of Deaths in Guangzhou Traffic Accidents has Dropped for 13 Consecutive Years [WWW Document] (http://gzjd.gov.cn/gzjd/gaxw_ztbd_ctgc/201602/67b06e82303f4f1bb3e6f7225f187552.shtml) (Accessed: 21 May 2019)

Guangzhou Statistics Bureau 2018 Guangzhou Statistics Bureau [WWW Document] (http://gzstats.gov.cn/gzstats/tjgb_qtgb/201802/cf533209a9cc46d08da1f6070a65067e.shtml) (Accessed: 15 April 2019)

Gurung A, Levy J I and Bell M L 2017 Modeling the intraurban variation in nitrogen dioxide in urban areas in Kathmandu Valley, Nepal *Environ. Res.* **155** 42–8

Hao W, Reis S, Lin C and Heal M R 2016 Effect of monitoring network design on land use regression models for estimating residential NO_2 concentration *Atmos. Environ.* **149** 24–33

He B, Heal M R and Reis S 2018 Land-use regression modelling of intra-urban air pollution variation in China: current status and future needs *Atmosphere* **9** 134

Hood C, MacKenzie I, Stocker J, Johnson K, Carruthers D, Vieno M and Doherty R 2018 Air quality simulations for London using a coupled regional-to-local modelling system *Atmos. Chem. Phys.* **18** 11221–45

Huang J, Pan X, Guo X and Li G 2018 Health impact of China's air pollution prevention and control action plan: an analysis of national air quality monitoring and mortality data *Lancet. Planet. Health.* **2** 313–23

Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, Morrison J and Giovis C 2005 A review and evaluation of intraurban air pollution exposure models *J. Expo. Anal. Environ. Epidemiol.* **15** 185–204

Ji X *et al* 2019 Nitrogen dioxide air pollution and preterm birth in Shanghai, China *Environ. Res.* **169** 79–85

Jim C Y and Chen W Y 2008 Assessing the ecosystem service of air pollutant removal by urban trees in Guangzhou (China) *J. Environ. Manage.* **88** 665–76

Johansson L, Jalkanen J-P and Kukkonen J 2017 Global assessment of shipping emissions in 2015 on a high spatial and temporal resolution *Atmos. Environ.* **167** 403–15

Johnson M, Isakov V, Touma J S, Mukerjee S and Özkaynak H 2010 Evaluation of land-use regression models used to predict air quality concentrations in an urban area *Atmos. Environ.* **44** 3660–8

Li W, Cao Y, Li R, Ma X, Chen J, Wu Z and Xu Q 2018 The spatial variation in the effects of air pollution on cardiovascular mortality in Beijing, China *J. Expo. Sci. Environ. Epidemiol.* **28** 297–304

Lin H *et al* 2018 Ambient particulate matter air pollution associated with acute respiratory distress syndrome in Guangzhou, China *J. Expo. Sci. Environ. Epidemiol.* **28** 392–9

Liu F, Beirle S, Zhang Q, van der A R J, Zheng B, Tong D and He K 2017 NO_x emission trends over Chinese cities estimated from OMI observations during 2005 to 2015 *Atmos. Chem. Phys.* **17** 9261–75

Liu H, Fu M, Jin X, Shang Y, Shindell D, Faluvegi G, Shindell C and He K 2016 Health and climate impacts of ocean-going vessels in East Asia *Nat. Clim. Change* **6** 1037–41

Luo K, Li R, Li W, Wang Z, Ma X, Zhang R, Fang X, Wu Z, Cao Y and Xu Q 2016 Acute effects of nitrogen dioxide on cardiovascular mortality in beijing: an exploration of spatial heterogeneity and the district-specific predictors *Sci. Rep.* **6** 38328

MarcoPolo-Panda 2017 MarcoPolo Emission Inventory | MarcoPolo—Panda (http://marcopolo-panda.eu/products/toolbox/emission-data/marcopolo-emission-inventory/) (Accessed: 30 March 2019)

McHugh C A, Carruthers D J and Edmunds H A 1997 ADMS–Urban: an air quality management system for traffic, domestic and industrial pollution *Int. J. Environ. Pollut.* **8** 666–74

Ministry of Environmental Protection 2012 Ambient Air Quality Standards [WWW Document] (http://210.72.1.216:8080/gzaqi/Document/gjzlbz.pdf) (Accessed: 21 May 2019)

Ministry of Environmental Protection 2018a Report on the State of the Environment in China [WWW Document] (http://english.mee.gov.cn/Resources/Reports/soe/) (Accessed: 9 April 2019)

Ministry of Environmental Protection 2018b 环境空气气态污染物 (SO_2, NO_2, O_3, CO)连续自动监测系统技术要求及检测方法 (Specifications and Test Procedures for Ambient Air Quality Continuous Automated Monitoring System for SO_2, NO_2, O_3 and CO) [WWW Document] (http://kjs.mee.gov.cn/hjbhbz/bzwb/jcffbz/201308/t20130802_256853.shtml) (Accessed: 13 October 2019)

Mölter A, Lindley S, de Vocht F, Simpson A and Agius R 2010 Modelling air pollution for epidemiologic research: I. A novel approach combining land use regression and air dispersion *Sci. Total Environ.* **408** 5862–9

OpenStreetMap 2018 OpenStreetMap [WWW Document] OpenStreetMap (https://openstreetmap.org/) (Accessed: 19 May 2019)

Righi S, Lucialli P and Pollini E 2009 Statistical and diagnostic evaluation of the ADMS-Urban model compared with an urban air quality monitoring network *Atmos. Environ.* **43** 3850–7

Stamen 2019 Stamen Maps [WWW Document] (http://maps.stamen.com/) (Accessed: 18 April 2019)

Stoerk T 2016 Statistical corruption in Beijing's air quality data has likely ended in 2012 *Atmos. Environ.* **127** 365–71

Sun J and Zhou T 2017 Health risk assessment of China's main air pollutants *BMC Public Health* **17** 212

Su J G, Jerrett M and Beckerman B 2009 A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures *Sci. Total Environ.* **407** 3890–8

Tonnesen S and Jeffries H E 1994 Inhibition of odd oxygen production in the carbon bond four and generic reaction set mechanisms *Atmos. Environ.* **28** 1339–49

United Nations 2014 2014 Revision of the World Urbanization Prospects [WWW Document] (https://esa.un.org/unpd/wup/publications/files/wup2014-highlights.pdf) (Accessed: 9 April 2019)

United Nations 2017 The World's Cities in 2018 [WWW Document] (https://un.org/en/events/citiesday/assets/pdf/the_worlds_cities_in_2018_data_booklet.pdf) (Accessed: 25 January 2018)

Venkatram A, Karamchandani P, Pai P and Goldstein R 1994 The development and application of a simplified ozone modeling system (SOMS) *Atmos. Environ.* **28** 3665–78

Visscher A D 2013 *Air Dispersion Modeling: Foundations and Applications* (New York: Wiley)

Walton H, Dajnak D, Beevers S, Williams M, Watkiss P and Hunt A 2015 Understanding the Health Impacts of Air Pollution in London [WWW Document] (https://london.gov.uk/sites/default/files/HIAinLondon_KingsReport_14072015_final_0.pdf) (Accessed 21 May 2019)

Wang S and Hao J 2012 Air quality management in China: issues, challenges, and options *J. Environ. Sci. China* **24** 2–13

Weinmayr G, Romeo E, De Sario M, Weiland S K and Forastiere F 2010 Short-term effects of PM_{10} and NO_2 on respiratory health among children with asthma or asthma-like symptoms: a systematic review and meta-analysis *Environ. Health Perspect.* **118** 449–57

WHO 2006 *Air Quality Guidelines: Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide.* (Copenhagen: World Health Organization)

WHO 2013 Health Risks of Air Pollution in Europe—HRAPIE Project, Recommendations for Concentration–Response Functions for Cost–Benefit Analysis of Particulate Matter, Ozone and Nitrogen Dioxide [WWW Document]

(http://euro.who.int/__data/assets/pdf_file/0006/238956/Health_risks_air_pollution_HRAPIE_project.pdf) (Accessed: 21 May 2019)

Worldpop 2015 China 100m Population (https://doi.org/10.5258/SOTON/WP00645)

Xu Z, Wang T, Xue L K, Louie P K K, Luk C W Y, Gao J, Wang S L, Chai F H and Wang W X 2013 Evaluating the uncertainties of thermal catalytic conversion in measuring atmospheric nitrogen dioxide at four differently polluted sites in China *Atmos. Environ.* **76** 221–6

Xue T, Zheng Y, Tong D, Zheng B, Li X, Zhu T and Zhang Q 2019 Spatiotemporal continuous estimates of $PM_{2.5}$ concentrations in China, 2000–2016: a machine learning method with inputs from satellites, chemical transport model, and ground observations *Environ. Int.* **123** 345–57

Yan W, Yun Y, Ku T, Li G and Sang N 2016 NO_2 inhalation promotes Alzheimer's disease-like progression: cyclooxygenase-2-derived prostaglandin E2 modulation and monoacylglycerol lipase inhibition-targeted medication *Sci. Rep.* **6** 22429

Yang X, Zheng Y, Geng G, Liu H, Man H, Lv Z, He K and de Hoogh K 2017 Development of $PM_{2.5}$ and NO_2 models in a LUR framework incorporating satellite remote sensing and air quality model data in Pearl River Delta region, China *Environ. Pollut.* **226** 143–53

Yuyu C, Zhe J G, Naresh K and Guang S 2012 Gaming in air pollution data? Lessons from China *BE J. Econ. Anal. Policy* **13** 1–43

Zhang P *et al* 2011 Long-term exposure to ambient air pollution and mortality due to cardiovascular disease and cerebrovascular disease in Shenyang, China *PLoS One* **6** e20827

Zheng B *et al* 2018a Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions *Atmos. Chem. Phys.* **18** 14095–111

Zheng B, Zhang Q, Davis S J, Ciais P, Hong C P, Li M, Liu F, Tong D, Li H Y and He K B 2018b Infrastructure shapes differences in the carbon intensity of Chinese cities *Environ. Sci. Technol.* **52** 6032–41