Atmospheric Pollution Research 12 (2021) 101096

Contents lists available at ScienceDirect

Atmospheric Pollution Research

journal homepage: www.elsevier.com/locate/apr

Development of spatiotemporal land use regression models for $PM_{2.5}$ and NO_2 in Chongqing, China, and exposure assessment for the CLIMB study

Alexander Harper [a], Philip N. Baker [a,b], Yinyin Xia [c], Tao Kuang [c], Hua Zhang [b], Yingxin Chen [d], Ting-Li Han [b], John Gulliver [e,*]

[a] Department of Health Sciences, University of Leicester, Leicester, UK
[b] Department of Obstetrics and Gynaecology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, 400016, China
[c] School of Public Health and Management, Chongqing Medical University, Chongqing, 400016, China
[d] Centre for Environmental Health and Sustainability, University of Leicester, Leicester, UK
[e] Centre for Environmental Health and Sustainability & School of Geography, Geology and the Environment, University of Leicester, Leicester, UK

A R T I C L E I N F O

Keywords:
Air pollution
Generalised additive modelling
Land use regression
Exposure assessment
CLIMB study

A B S T R A C T

Limited research has been conducted in Asia on the association of maternal exposure to ambient air pollution and the increased risk of adverse pregnancy outcomes such as low birth weight and preterm birth. The aim of this study was to develop spatiotemporal land use regression (LUR) models for fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO_2) in Chongqing, China, and to use the models to estimate $PM_{2.5}$ and NO_2 exposure for the participants in a randomized trial of complex lipid supplementation (the Complex Lipids In Mothers and Babies (CLIMB) study), before and during pregnancy. Spatiotemporal generalised additive models were developed for 2015–2016 on a daily basis incorporating measurement data from 16 sites, temporal variables on meteorology, and spatial variables produced using a geographical information system. Hold-out validation (HOV) was performed using daily and monthly averaged measurements for 2017 at 17 sites with 4 of the sites in different locations to 2015–16. The $PM_{2.5}$ spatiotemporal model had good overall predictive ability (daily HOV correlation (COR)-$R^2 = 0.75$ and HOV mean-squared-error (MSE)-$R^2 = 0.69$; monthly HOV COR-$R^2 = 0.87$ and HOV MSE-$R^2 = 0.76$). The NO_2 spatiotemporal model estimates had moderate-to-good correlation with measurements (daily HOV COR-$R^2 = 0.44$; monthly HOV COR-$R^2 = 0.65$), but estimates were subject to bias (daily HOV MSE-$R^2 = 0.24$; monthly HOV MSE-$R^2 = -0.02$). On this basis, we recommend that $PM_{2.5}$ models are used for predicting absolute exposure and NO_2 models are used for relative ranking of exposures.

1. Introduction

Particulate matter less than 2.5 μm in diameter ($PM_{2.5}$) and nitrogen dioxide (NO_2) exposure during pregnancy increases the risk of adverse pregnancy outcomes such as low birth weight (Pedersen et al., 2013) and small for gestational age (Stieb et al., 2016). Exposure to $PM_{2.5}$ is particularly harmful since its small size allows for deposition deep in the lungs where it exerts systemic damage either indirectly, due to respiratory-mediated release of inflammatory markers (Suwa et al., 2002), or directly through diffusion into the bloodstream and deposition around the body (Nemmar et al., 2002). The relationship of maternal air pollution exposure and adverse pregnancy outcomes has been firmly established by major studies conducted in North America and Europe (Brauer et al., 2008; Pedersen et al., 2013a; Stieb et al., 2016). However, relatively little research has been conducted in Asia (Li et al., 2020). To conduct epidemiological studies, methods for individual exposure estimation are required.

In the absence of detailed emissions inventories for dispersion modelling, land use regression (LUR) models are commonly used in epidemiological studies for air pollutant exposure estimation (Hoek et al., 2008). LUR models are trained and validated against measured pollutant concentrations using variables generated from a geographic information system (GIS), such as distance to nearest source, road network density, land use, terrain and population density, to predict the

Peer review under responsibility of Turkish National Committee for Air Pollution Research and Control.

* Corresponding author. Centre for Environmental Health and Sustainability & School of Geography, Geology and the Environment, University of Leicester, Leicester, LE1 7RH, UK.

E-mail address: jg435@leicester.ac.uk (J. Gulliver).

https://doi.org/10.1016/j.apr.2021.101096
Received 5 March 2021; Received in revised form 12 May 2021; Accepted 26 May 2021
Available online 6 June 2021

concentrations at residential locations. Models of annual average concentrations can be developed using widely available spatial data and can capture the fine scale variability in air pollution concentrations (De Hoogh et al., 2014). For use in pregnancy cohort studies, where sub-annual exposure periods are required, LUR models have been developed to account for both the spatial and temporal variation in pollutant concentrations (Barratt et al., 2018; Shi et al., 2018; Xu et al., 2019; Zhang et al., 2020). In China, various techniques have been used to develop LUR models of differing temporal resolution. Tian et al. (2019) and Kong and Tian (2020), for example, developed separate LUR models for different seasons. Xu et al. (2019) used temporal trends in air pollution concentration measurements to develop a model capable of predicting two-week averages. Models with daily temporal resolution have also been developed. Barratt et al. (2018), for example, used air quality monitoring campaigns to produce dynamic, three-dimensional LUR models which considered horizontal and vertical variation in air pollution concentration and population movements. This included the use of monitoring reference sites designed to measure background concentration to strengthen the temporal model component. In the absence of reference sites, studies have used temporal variables describing meteorology and have incorporated satellite data including Aerosol Optical Density (AOD) for $PM_{2.5}$ and NO_2 column data (Anand and Monks, 2017; Shi et al., 2018). Whilst satellite data has been shown to improve LUR model performance, its use is restricted by cloud coverage which reduces its suitability for use in the development of spatiotemporal models for regions commonly covered by cloud. In Europe, Chemical Transport Models (CTMs) have been used in LUR models and shown to improve model performance (De Hoogh et al., 2016).

Since the relationships of temporal variables and pollutant concentrations are often complex and non-linear, more sophisticated modelling techniques are often required. Generalised additive models (GAMs), for example, allow handling of non-linear relationships using smooth terms. Dimakopoulou et al. (2018), used smooth terms to model the relationship between spatial and temporal variables with air pollution concentrations (Xu et al., 2019). The additive nature of GAMs means that the effect of each variable on the predicted outcome is independent of other variables, which allows for interpretation of the impact of each variable on the explained variability of measured concentrations.

The Complex Lipids in Mothers and Babies (CLIMB) study recruited a total of 1500 women in from two hospitals in Chongqing, China. Complex lipids are important constituents of the central nervous system, and studies have shown that supplementation during pregnancy may improve offspring cognitive outcomes (Vickers et al., 2009; Gurnida et al., 2012). The primary aim of the CLIMB study was to study the effects of supplementation of complex lipids in pregnancy on maternal ganglioside status and offspring cognitive outcomes. The data collected also presents an opportunity to study the relationship between air pollution exposure and a range of pregnancy outcomes, including pregnancy complications and fetal biometry, and offspring outcomes such as infant cognitive development, growth and general health in a population of pregnant women (Huang et al., 2017; Norris et al., 2019). To enable such research to be conducted, this study aimed to develop and validate granular spatiotemporal LUR models for $PM_{2.5}$ and NO_2 in Chongqing, China.

2. Methods

2.1. Study area

The study area focussed on the urban centre of the Chinese municipality of Chongqing (Fig. 1). With a population of approximately 6.52 million and a land area of 5472 km^2, the study area has a high population density of approximately 1191 people/km^2 (Chongqing Municipal Bureau of Statistics, 2015). The city is characterised by high-rise buildings within a valley where the Yangtze and Jialing rivers converge,

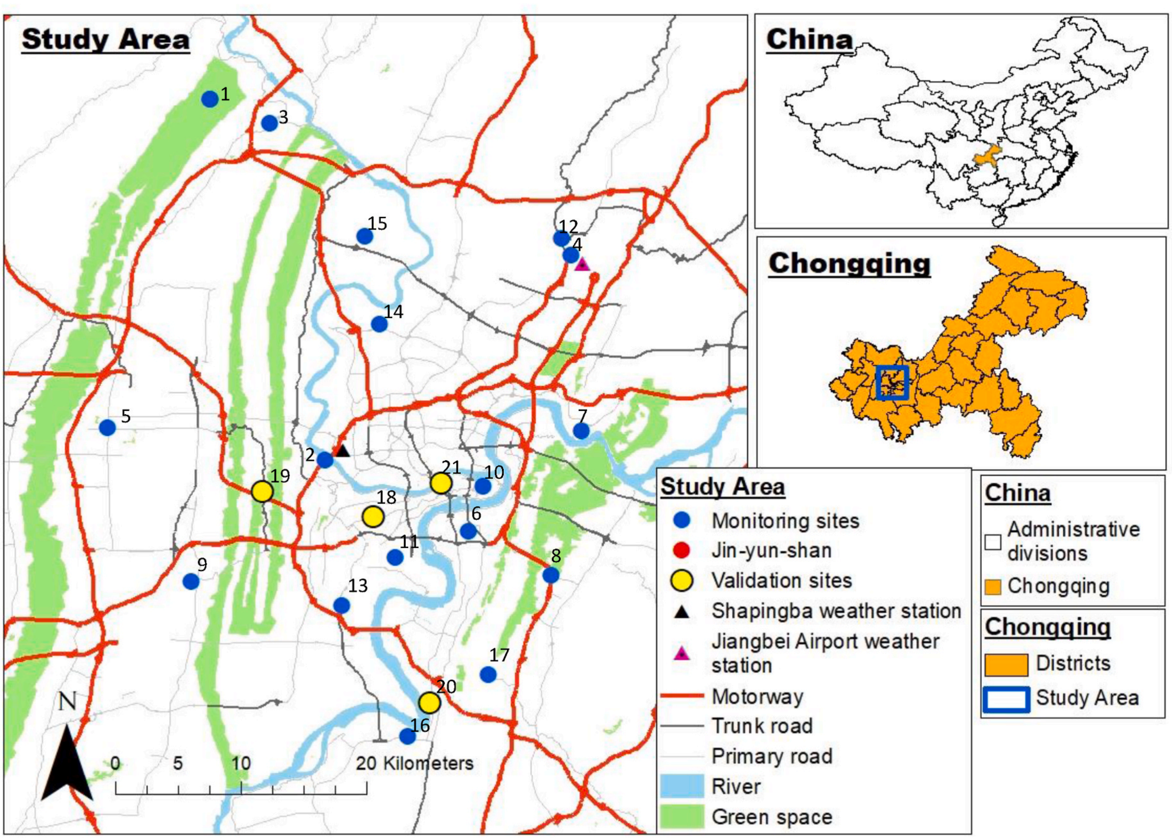

Fig. 1. Study area and location of monitoring sites (OpenStreetMap contributors, 2015; https://data.nextgis.com/en/region/CN-50/).

and the surrounding rural area is bounded to the west by mountains. Chongqing is considered one of the four 'furnace cities' of China due to its long and hot summers (Xinhua, 2012). In contrast, winters tend to be mild with low wind speeds and temperature inversions (Liao et al., 2018). Rain falls all year round, although summer is considered the rainy season (Wang et al., 2020). Chongqing experiences one of the lowest levels of sunshine in China, with only 1178.7 sunshine hours in 2018 (Wong, 2020).

2.2. Air pollution measurement data

Hourly $PM_{2.5}$ and NO_2 measurements taken from January 2, 2015 (data was not available for the January 1, 2015) to December 31, 2017 were obtained from the Chongqing Ecological Protection Bureau (2020). Monitoring sites were given a site ID number and classified as traffic sites if they were <100 m from a major road (motorway, trunk road, primary road), and background sites if they were >100 m away from a major road (Table S1, supporting information). In 2015, measurements were taken at a total of 17 sites, of which 6 are traffic sites and 11 are background sites, and 14 are in urban areas and 2 in suburban areas. Site 1 is in a rural forested area at relatively high altitude and is not representative of residential locations, therefore it was excluded from model development. In October and November 2016, Sites 2, 11 and 17 were discontinued and activity at the urban traffic sites 18 and 20, and suburban background site 19 started. On September 30, 2017, activity at Site 10 ceased and activity at urban background site 21 started. Values reported as $PM_{2.5}$ and NO_2 concentrations of 0 were considered as missing data.

2.3. Model development

2.3.1. Model formulation

The spatiotemporal models took the following form:

$$Y_{ij} = \alpha + \sum(\beta X_i) + \sum(f(Z_j)) + f(Long_i) + f(Lat_i) + \varepsilon_{ij}$$

where Y_{ij} represents the predicted daily average pollutant concentrations at location i and on day j. $\alpha + \sum(\beta X_i)$ represents the spatial component of the models, where α is the intercept and β is the coefficient of each GIS predictor variable X. $\sum(f(Z_j))$ represents the temporal term, where $f(Z_j)$ is the smooth function of temporal variables Z. $f(Long_i)$ and $f(Lat_i)$ are the smooth terms of longitude and latitude, respectively, and ε is the random error term. The approach therefore combines linear regression models for the spatial terms and generalised additive models (GAM) for the temporal terms and location.

2.3.2. Spatial predictor variables

Spatial variables were calculated for each site using the monitoring site coordinates and GIS data in ArcMap 10.8 software with a map projection of WGS 72BE South China Sea Lambert Conformal Conic. Variables were generated to reflect air pollution sources and sinks. The following groups of variables were generated: road network, land use, topography, vegetation, and population density. Road network variables included distance to nearest road and lengths of roads within circular buffers of radii 25 m, 50 m, 100 m, 300 m, 500 m and 1000 m to reflect local influences of nearby traffic and the influence of road density on the urban background concentration. The different road types used are summarised in Table S2. Land use and vegetation variables were derived using circular buffers of radii 300 m, 500 m, 1000 m and 5000 m. The vegetation variables were based on the Normalised Difference Vegetation Index (NDVI) developed from Moderate Resolution Imaging Spectroradiometer (MODIS) images in which cloud coverage did not negatively impact the data quality over the study area. Images taken on two days (July 12, 2015 and August 29, 2015) met these criteria and were merged to produce average NDVI. Population density at the point locations of monitoring sites was extracted from a 1 km grid. The average population density within circular buffers was not used since the population density data had a relatively low spatial resolution of 1 km. Site elevation was extracted from a digital elevation model and distance to centre (using site 10 as the centre; see Fig. 1) was calculated for each site. A description of the potential spatial variables can be found in Table S3. The GIS data used were obtained from the following sources:

1. Road network data was from OpenStreetMap (OpenStreetMap contributors, 2015; https://data.nextgis.com/en/region/CN-50/).
2. Land use variables were derived from layers with a 100 m resolution in which the values of each pixel represented the fraction of each land use type within the pixel area. Data was provided by the Copernicus Global Land Service (Buchhorn et al., 2019).
3. Vegetation variables were based on the NDVI derived from images taken in 2015 from the MODIS MOD13Q1 v006 dataset which has a temporal resolution of 16 days and spatial resolution of 250 m (Didan, 2015).
4. A 1 km × 1 km population density grid was downloaded from the Chinese Academy of Sciences website (Xu, 2017).
5. Site elevation was extracted from the Global Digital Surface Model, which had a resolution of 30 m and was developed using data from the Advanced Land Observing Satellite (Tadono et al., 2016).

2.3.3. Spatial model component

The spatial models took the form:

$$Y_i = \alpha + \sum \beta X + \varepsilon$$

where Y is the measured annual air pollution concentration at location i, α is the intercept, β is the coefficient to predictor variables X and ε is the random error term.

The spatial component of the models was developed using a stepwise multiple linear regression approach which related the spatial variables to the daily average air pollutant concentration measurements taken at sites 2 to 17 in 2015. Measurements taken in 2016 were not used due to the closure of three monitoring sites. Variable selection involved use of a similar algorithm to those employed by previous studies (Abernethy et al., 2013; Habermann et al., 2015). *A priori* assumptions regarding the direction of correlation between each variable and annual average air pollutant concentrations were made based on prior knowledge of air pollution sources and sinks, and variables filtered as follows:

1. Variables for which the correlation did not match the *a priori* assumption were excluded.
2. Remaining variables were ranked in each sub-group by the size of the correlation coefficient.
3. Within each sub-group, variables which were strongly correlated ($r > 0.6$) with the highest ranked variable were excluded.

Developed models ensured that the direction of the effect of each variable was consistent with the *a priori* assumption and that variables were statistically significant ($p < 0.05$ for entry and $p < 0.1$ in the final model) and had a variance inflation factor of less than 3. Models were also checked for outliers using Cook's Distance.

2.3.4. Temporal predictor variables

Meteorological variables were used to account for the influence of weather on the change in air pollution concentration over time (i.e. day-to-day, season). Measurement data from the Shapingba weather station (Fig. 1) was downloaded from the National Centers for Environmental Information website (National Centers for Environmental Information, 2020), and data from the Jiangbei Airport weather station was downloaded from the Reliable Prognosis 5 website (Raspisaniye Pogodi Ltd, 2020). Since the location of the Shapingba weather station was more central than the Jiangbei station, measurements from the Shapingba station were used to generate the variables of daily average temperature, amount of rainfall in the past 24 and 48 h and a binary variable for rainfall events. On days where data from the Shapingba weather station was missing, data from the Jiangbei station were used. Measurements taken at the Jiangbei Airport station were used to generate the following daily average variables: relative humidity, horizontal visibility, wind direction and wind speed. The daily average wind direction was calculated using the Openair R package (Carslaw and Ropkins, 2012). The temporal variables generated are summarised in Table S4.

2.3.5. Temporal model component

The temporal component took the form:

$$Y_j = \sum (f(Z_j)) + \varepsilon_j$$

where Y is the predicted daily residuals on day j, $f(Z_j)$ are smooth functions of the temporal variables Z and ε is the random error term.

The temporal component of models was developed using the mgcv package in R (Wood, 2011). The temporal GAMs were fitted to the residuals from spatial component at each monitoring site on a daily basis, calculated by subtracting the predicted annual average concentration from the observed daily average concentrations measured in 2015 and 2016. Fig. S1-S5 presents descriptive statistics and the univariate relationships between temporal variables and daily average pollutant concentrations, which were used to identify terms to include in the GAMs. Identified variables were included in the temporal $PM_{2.5}$ and NO_2 GAMs if they were significant ($p < 0.05$).

The optimal number of basis dimensions was determined through graphical analysis of GAMs to ensure that modelled relationships did not overfit the measurement data and key aspects of relationships were not missed. The smoothing parameter values were estimated using the Restricted Maximum Likelihood method (Wood, 2011).

2.3.6. Remaining spatial autocorrelation

To account for the remaining spatial autocorrelation, the smooth terms of longitude and latitude were fitted to spatiotemporal residuals which were calculated by subtracting the sum of the spatial temporal predictions from the measured daily average concentrations in 2015 and 2016. The terms took the form:

$$S_i = f(Long_i) + f(Lat_i)$$

where S_i is the spatiotemporal residual at location i, and $f(Long_i)$ and $f(Lat_i)$ represent smooth terms of longitude and latitude, respectively. The term was included in the spatiotemporal models if it was statistically significant ($p < 0.05$).

2.4. Model performance and validation

Spatial predictions were assessed through comparison of the measured and predicted annual average pollutant concentration in 2015 and via leave-one-out-cross-validation (LOOCV). The LOOCV involved the removal of the annual average concentration at each site in turn and refitting the model to the annual average concentrations at the remaining sites whilst maintaining the same variables.

Measured daily, weekly and monthly average pollutant concentrations from training sites in 2017, including four newly opened sites (Sites 18–21), were used for hold-out validation (HOV). The performance of spatiotemporal models was assessed through analysis of measured against predicted pollutant concentrations to produce four groups of performance statistics: COR-R^2 (square of Pearson's correlation coefficient), MSE-R^2 (mean-square-error-based-R^2; i.e. 1- (mean square error/variance of observations)), RMSE (root-mean-square-error), and NRMSE (normalised root mean square error; i.e. RMSE/interquartile range (IQR) of the observations). In addition to COR-R^2, some studies have used MSE-R^2 as a composite measure of correlation and bias to provide a more stringent test of model prediction performance (Gulliver et al., 2016; Knibbs et al., 2018; Wang et al., 2012).

A summary of the data used in model development and validation is provided in Table 1.

2.5. Exposure assessment of CLIMB study participants

The study utilised data collected on pregnant women enrolled in the CLIMB study, a three armed randomized controlled trial of a complex milk lipid supplement during pregnancy that was first established in Chongqing, China from September 2015 (Huang et al., 2017; Norris et al., 2019). Women were recruited in their first trimester of pregnancy from the First Affiliated Hospital of Chongqing Medical University and Chongqing Health Centre for Women and Children (CHCWC) in China. Recruitment was completed in June 2017. Eligible gravidas were aged 20–40 years and had a singleton pregnancy. Women with a history of premature delivery before 32 weeks of gestation, known milk allergy or aversion, or lactose intolerance were excluded. Women who withdrew from the study, whose pregnancies were terminated, who miscarried, or were lost to follow up (n = 40), were excluded from the analysis.

The developed spatiotemporal models were used to estimate the $PM_{2.5}$ and NO_2 exposures of 1183 of the 1500 participants recruited in the CLIMB study (Huang et al., 2017; Norris et al., 2019) for whom the detailed residential addresses during pregnancy were known. Using the R software (Wood, 2011), the average exposures in the 1st, 2nd and 3rd trimesters and throughout pregnancy were predicted, as well as during the 90 days prior to conception, since this has been reported to be a critical exposure period (Robledo et al., 2015).

3. Results

3.1. Measured concentrations

Fig. 2 presents the variability in the annual average NO_2 and $PM_{2.5}$ concentrations in 2015, the year used for development of the spatial component. During this period, the annual average NO_2 and $PM_{2.5}$ concentrations were 44.9 μg/m^3 (SD = 8.4 μg/m^3) and 55.9 μg/m^3 (SD = 2.8 μg/m^3), respectively. There was greater variation in the annual average NO_2 concentration (range = 30.6 μg/m^3) than in that of $PM_{2.5}$ concentration (range = 11.1 μg/m^3). Site 10, an urban centre background site, recorded the highest annual average NO_2 (63.4 μg/m^3) and $PM_{2.5}$ concentrations (63.6 μg/m^3). Aside from Site 1, which was considered a spatial outlier and was excluded from model development, Site 9, an urban background site (Airport site), reported the lowest annual average NO_2 concentration (32.8 μg/m^3) and Site 5, an urban traffic site, reported the lowest annual average $PM_{2.5}$ concentration

Table 1
Summary of data used for model development and validation.

	Site ID[a]	Measurement time period[b]
Spatial component	2 to 17	January 2, 2015–December 31, 2015
Temporal component	3 to 10 and 12 to 16	January 2, 2015–December 31, 2016
	2	January 2, 2015–October 22, 2016
	11	January 2, 2015 – November 5, 2016
	17	January 2, 2015–November 13, 2016
Hold-out validation	3 to 9 and 12 to 16	January 1, 2017–December 31, 2017
	10	January 1, 2017–September 30, 2017
	18 to 20	November 21, 2016–December 31, 2017
	21	September 30, 2017–December 31, 2017

[a] The ID of sites at which measurements used in the corresponding stage of model development and validation were taken.

[b] The time period over which the measurements in the corresponding stage of model development and validation were taken.

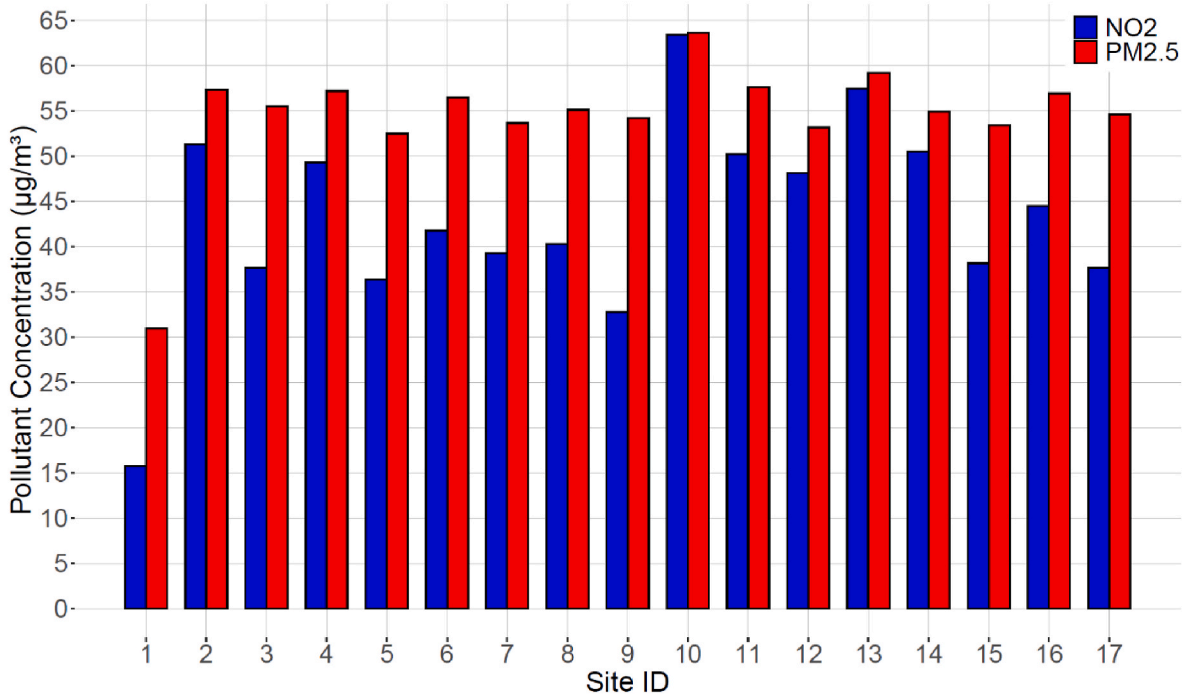

Fig. 2. Variation in measured annual averages of NO_2 and $PM_{2.5}$ concentration by site in 2015.

(52.5 μg/m^3). Fig. S6 presents the variability in the annual average NO_2 and $PM_{2.5}$ concentrations in 2016 and 2017 which both include site location changes, hence the spatial model was developed using sites only from 2015.

The temporal model was developed using daily average measurements from 2015 to 2016. Daily average measurements in 2015 and 2016 are presented graphically in Fig. S2 and S3. Average monthly concentrations are summarised in Table S5. Across the winter months of January to March and October to December in both years, $PM_{2.5}$ and NO_2 concentrations were generally higher and showed greater variation (in winter 2015, average $PM_{2.5}$ = 70.5 μg/m^3 (SD = 42.9), NO_2 = 47.8 μg/m^3 (SD = 14.1 μg/m^3; in winter 2016, average $PM_{2.5}$ = 61.3 μg/m^3 (SD = 30.7), NO_2 = 50.6 μg/m^3 (SD = 12.7)) compared to summer months of April to September. In summer 2015, average $PM_{2.5}$ = 41.7 μg/m^3 (SD = 15.9 μg/m^3), NO_2 = 42.3 μg/m^3 (SD = 42.2 μg/m^3); in summer 2016, average $PM_{2.5}$ = 46.8 μg/m^3 (SD = 20.0 μg/m^3), NO_2 = 42.1 μg/m^3 (SD = 11.0 μg/m^3). Air pollutant concentration did not vary significantly by day of the week (Fig. S1). In 2015, 2016 and 2017, the percentages of missing $PM_{2.5}$ measurements were 1.9%, 2.8% and 1.4%, and those of NO_2 measurements were 2.5%, 2.9% and 1.4%, respectively.

3.2. Model variables

Variables included in the spatiotemporal models are summarised in Table 2. In the NO_2 and $PM_{2.5}$ spatial models, the length of Highways within a 500 m circular buffer (Highway_500) variable was included and was the most important variable in both models (partial R^2 = 0.54 in the $PM_{2.5}$ model, partial R^2 = 0.31 in the NO_2 model). In addition, the average NDVI within a 5000 m circular buffer (NDVI_5000) was included in the NO_2 model, and the fraction of land within a 5000 m circular buffer which was rural (Rural_5000) was included in the $PM_{2.5}$ model. For the $PM_{2.5}$ model, the partial R^2 values were as follows: 0.54 for Highway_500, 0.33 for Rural_5000. For the NO_2 model, the partial R^2

Table 2
Summary of the spatial and temporal components of the $PM_{2.5}$ and NO_2 models.

Linear terms	$PM_{2.5}$ B	SE B	β	p	Partial R^2	ΔR^2	NO_2 B	SE B	β	p	Partial R^2	ΔR^2
Spatial intercept	56.0	1.6	–	<0.01	–	–	59.5	11.9	–	<0.01	–	–
Temporal intercept	3.12	1.07	–	<0.01		–	–	–	–	–	–	–
Rural 5000	−0.049	0.020	−0.48	0.02	0.33	–	–	–	–	–	–	–
HW 500	7.5e-4	1.9e-4	0.61	<0.01	0.54	–	1.8e-3	7.7e-4	0.49	0.03	0.31	
NDVI 5000	–	–	–	–	–	–	−3.8e-3	1.8e-3	−0.43	0.05	0.26	
Rain in 48 h	−6.66	1.50	–	<0.01	–	0.01	–	–	–	–	–	–
Smooth terms	k	edf		p	ΔR^2		k	edf		p	ΔR^2	
Month	3	1.98		<0.01	0.04		5	3.67		<0.01	0.03	
Temperature	5	3.79		<0.01	0.01		9	5.00		<0.01	0.05	
Horizontal visibility	9	5.72		<0.01	0.31		9	1.00		<0.01	0.07	
Relative humidity	9	5.53		<0.01	0.11		9	3.51		<0.01	0.13	
Wind speed	–	–		–	–		9	2.03		<0.01	0.07	

B, unstandardized coefficient; SE B, standard error of coefficients (B); β, standardized coefficient; p, probability value for significance; Partial R^2, coefficient of partial determination; ΔR^2, change in R^2 when the variable was removed; Rural 5000, fractional cover of rural land within a 5000 m circular buffer; HW 500, length of Highways within a 500 m circular buffer; NDVI 5000, average Normalised Difference Vegetation Index within a 5000 m circular buffer; k, number of basis dimensions; edf, effective degrees of freedom.

values were as follows: 0.31 for Highway_500, 0.26 for NDVI_5000. The multiple linear regression assumptions were assessed graphically (Fig. S7 and S8). Whilst the NO_2 model met all assumptions tested for, the $PM_{2.5}$ model failed to meet the assumption of normality at extreme low and high concentrations, and the measured annual average concentration at Site 10 acted as an influential observation (Cook's D = 1.6). Modelled annual average $PM_{2.5}$ and NO_2 concentrations surfaces for the study area are presented in Fig. 3.

Smooth terms for month, temperature, horizontal visibility and relative humidity variables were selected for inclusion in both $PM_{2.5}$ and

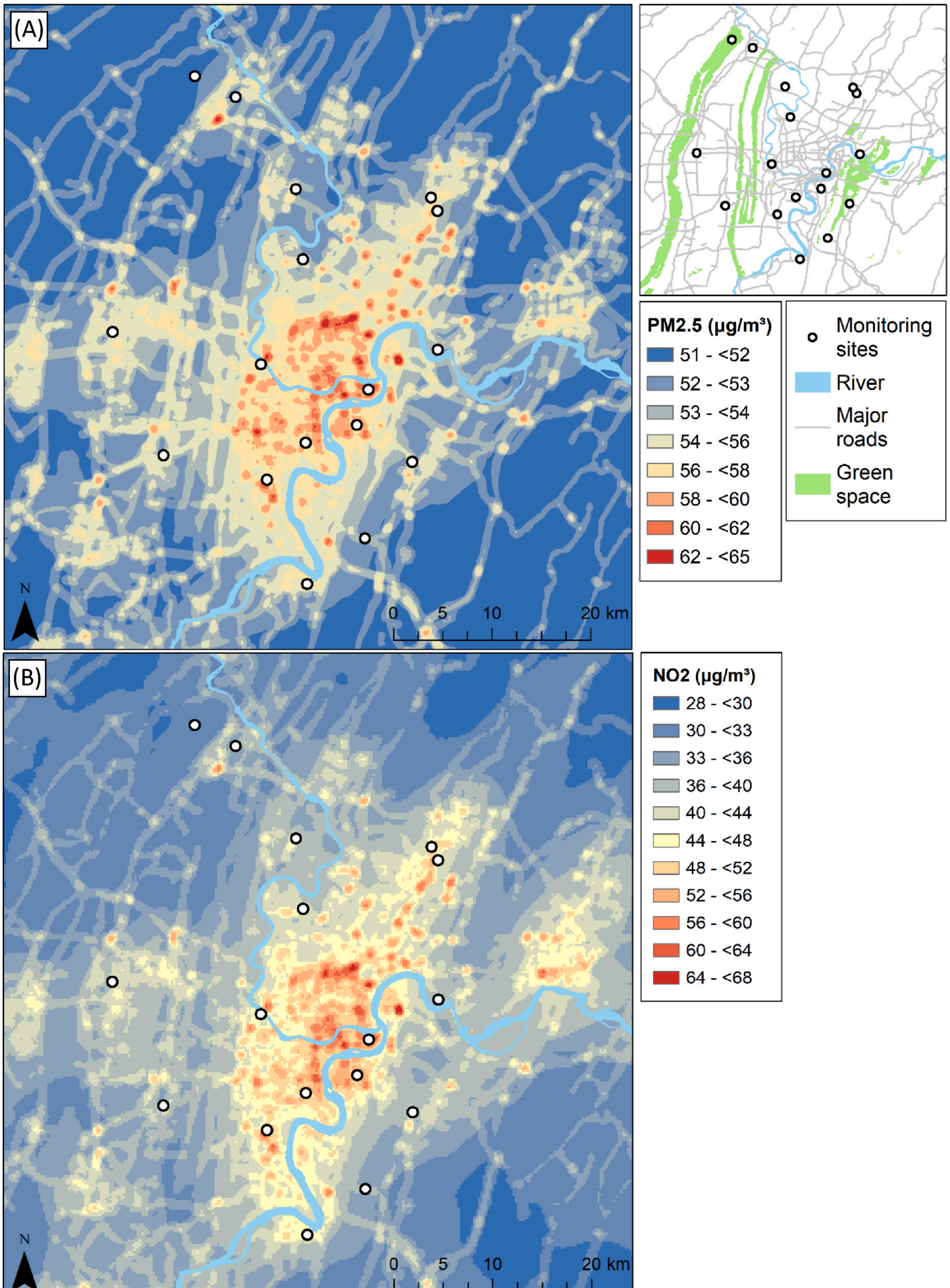

Fig. 3. Modelled annual average $PM_{2.5}$ (A) and NO_2 (B) concentration in 2015.

NO_2 temporal GAMs. In addition, a linear term of the binary variable describing the presence or absence of rainfall within 48 h was included in the $PM_{2.5}$ GAM and a smooth term of the wind speed variable was included in the NO_2 GAM. In the $PM_{2.5}$ GAM, the number of basis dimensions (*k*) specified for temperature and month smooth terms were 5 and 3, respectively. In the NO_2 GAM, k was 5 for the month smooth term. For other included variables, the default k was used (k = 9). The most important variable in the $PM_{2.5}$ GAM was visibility (R^2 change = 0.31) with low visibility related to increased $PM_{2.5}$ concentrations. The most important variable in the NO_2 GAM was humidity (R^2 change = 0.13), with high humidity related to decreased NO_2 concentrations. In both GAMs, pollutant concentrations were highest in winter and lowest in summer. The partial effect plots for the $PM_{2.5}$ and NO_2 GAMs are presented in Fig. S9 and S10. The GAM assumptions were tested graphically (Fig. S11 and S12). Both the $PM_{2.5}$ and NO_2 GAMs met the assumption of homogeneity, however, at extreme low and high concentrations the assumption of normality was not met. The longitude and latitude smooth terms, which accounted for remaining spatial autocorrelation, were only included in the NO_2 spatiotemporal model.

3.3. Model evaluation

For the $PM_{2.5}$ spatial component, model development yielded 0.78 for COR-R^2, 0.79 for MSE-R^2, 1.2 μg/m^3 for RMSE and 0.39 for NRMSE. For the NO_2 spatial component, model development yielded 0.71 for COR-R^2, 0.73 for MSE-R^2, 4.4 μg/m^3 for RMSE and 0.36 for NRMSE. For the $PM_{2.5}$ spatial component, the LOOCV values were 0.60 for COR-R^2, 0.62 for MSE-R^2, 1.7 μg/m^3 for RMSE and 0.54 for NRMSE. For the NO_2 spatial component, the LOOCV values were 0.58 for COR-R^2, 0.60 for MSE-R^2, 5.3 μg/m^3 for RMSE and 0.44 for NRMSE. Performance and validation statistics are summarised in Table S6.

Table 3 shows summary statistics combining all sites for daily and monthly COR-R^2, MSE-R^2, RMSE and NRMSE of the $PM_{2.5}$ and NO_2 spatiotemporal LUR models. The values for individual sites are presented in Table S7.

For spatiotemporal model development (i.e. using measurements from 2015 to 2016), the average COR-R^2 of the $PM_{2.5}$ was 0.72 and for NO_2 was 0.39. HOV yielded similar results, whereby the $PM_{2.5}$ model had good performance (HOV daily COR-R^2 = 0.75; monthly COR-R^2 = 0.87) and the NO_2 model had moderate performance (HOV daily COR-R^2 = 0.44; monthly CO-R^2 = 0.65). The average HOV MSE-R^2 of the $PM_{2.5}$ model (HOV daily MSE-R^2 = 0.69; monthly MSE-R^2 = 0.76) was substantially greater than that of the NO_2 model (HOV daily MSE-R^2 = 0.24; monthly MSE-R^2 = −0.02) and the NO_2 spatiotemporal model predictions were subject to greater bias than those of the $PM_{2.5}$ spatiotemporal model.

3.4. Exposure assessment of CLIMB study participants

Fig. 4 shows the variation in the estimated exposure to $PM_{2.5}$ and NO_2 during the 90 days prior to pregnancy, 1st, 2nd and 3rd trimesters and during the whole of pregnancy. Table S9 presents the summary statistics. The estimated $PM_{2.5}$ and NO_2 exposures of CLIMB study participants over the whole pregnancy were 57.4 μg/m^3 (SD = 4.0 μg/m^3) and 50.5 μg/m^3 (SD = 5.1 μg/m^3), respectively. The estimated $PM_{2.5}$ exposures during the 90 days prior to conception (mean = 52.9 μg/m^3; SD = 11.0 μg/m^3), 1st (mean = 52.0 μg/m^3; SD = 11.0 μg/m^3), 2nd (mean = 58.6 μg/m^3; SD = 12.2 μg/m^3) and 3rd (mean = 61.8 μg/m^3; SD = 16.0 μg/m^3) trimesters showed greater variation than the corresponding NO_2 exposures during the 90 days prior to conception (mean = 49.6 μg/m^3; SD = 6.3 μg/m^3), 1st (mean = 48.8 μg/m^3; SD = 6.3 μg/m^3), 2nd (mean = 51.0 μg/m^3; SD = 6.2 μg/m^3) and 3rd (51.8 μg/m^3; SD = 6.8 μg/m^3) trimesters.

4. Discussion

Spatiotemporal LUR models for estimating daily and monthly average $PM_{2.5}$ and NO_2 concentration estimates were developed for Chongqing, China. The $PM_{2.5}$ spatiotemporal model had good performance when providing concentration estimates in absolute terms. Whilst the ability of the NO_2 spatiotemporal model to provide concentration estimates in absolute terms was substantially weaker than for $PM_{2.5}$, the ability of the model to provide a relative ranking for NO_2 was moderate.

The developed spatiotemporal LUR models were used to estimate the $PM_{2.5}$ and NO_2 exposure during time periods before and during the pregnancies of 1183 women who lived in Chongqing and were recruited as part of the CLIMB study (Huang et al., 2017; Norris et al., 2019). Exposure estimates highlighted that models could capture contrasts in space and time during different pregnancy exposure periods, although greater variation in average $PM_{2.5}$ exposure compared to NO_2 was observed.

4.1. Performance of the spatiotemporal land use regression models

From measurements taken from the network of fixed-site monitors, there was greater spatial variation in concentrations of NO_2 than $PM_{2.5}$ and, conversely, greater temporal variation in concentrations of $PM_{2.5}$ than NO_2. Since the spatial component of the models was based on annual average concentrations from 16 measurement sites and the temporal component was developed from data covering 730 days (the measured daily average concentrations from 2015 to 2016), the models had a comparatively weak spatial basis and strong temporal basis. Hence, we yielded more accurate predictions of $PM_{2.5}$ concentrations than NO_2 concentrations overall. NO_2 models performed well in HOV in

Table 3
Performance of spatiotemporal land use regression models.

	COR-R^2		MSE-R^2		RMSE (μg/m^3)		NRMSE	
	Daily	Monthly	Daily	Monthly	Daily	Monthly	Daily	Monthly
$PM_{2.5}$								
2015–2016								
	0.72 (0.67–0.77)	0.84 (0.77–0.89)	0.71 (0.66–0.76)	0.83 (0.78–0.88)	17.5 (15.3–20.1)	8.7 (7.2–10.5)	0.32 (0.28–0.37)	0.16 (0.13–0.19)
HOV								
	0.75 (0.68–0.80)	0.87 (0.74–0.97)	0.69 (0.53–0.75)	0.76 (0.47–0.91)	17.8 (16.2–20.9)	10.7 (7.3–14.2)	0.54 (0.48–0.68)	0.36 (0.25–0.55)
NO_2								
2015–2016								
	0.39 (0.24–0.50)	0.61 (0.29–0.86)	0.31 (0.00–0.49)	0.33 (−0.45–0.78)	12.8 (9.2–16.3)	5.9 (3.0–11.1)	0.29 (0.23–0.42)	0.13 (0.08–0.26)
HOV								
	0.44 (0.30–0.69)	0.65 (0.35–0.88)	0.24 (−0.17–0.59)	−0.02 (−1.39–0.66)	13.4 (9.6–17.7)	8.4 (4.1–13.2)	0.62 (0.49–0.76)	0.68 (0.35–1.12)

COR-R^2, MSE-R^2, RMSE, and NRMSE are presented as the mean, with minimum and maximum in parentheses, of the COR-R^2, MSE-R^2, RMSE, and NRMSE values achieved across sites (Table S7 and S8). Abbreviations: COR-R^2, Pearson's correlation coefficient, squared; MSE-R^2, mean-square-error-based-R^2; RMSE, root mean square error; NRMSE, normalised root mean square error (normalised by the inter-quartile range (IQR) of measurements); HOV, hold-out validation.

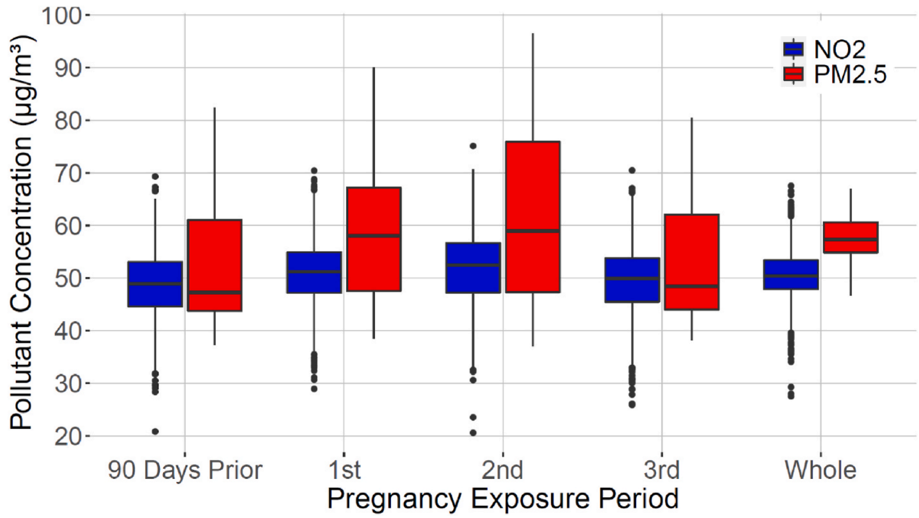

Fig. 4. Exposure assessment of CLIMB study participants.

terms of correlation (COR-R^2) and poor-to-good depending on the site and poorly overall in terms of MSE-R^2. It is not completely clear why NO_2 model performance was stronger at some sites compared to others, but there was a tendency for MSE-R^2 to be lower at sites close to highways which suggests weaknesses in the variables accounting for fluctuations in traffic. The overall performance suggests that using the NO_2 spatiotemporal model to provide relative rankings of concentration would be more appropriate than using the model to provide concentration estimates in absolute terms. Models for $PM_{2.5}$ performed well overall and may be applied to predict absolute values of exposures.

We used a two-stage approach to model development: 1) the spatial variables were selected using a multiple linear regression model and 2) the residuals from the spatial model were included along with meteorological variables and month of year in a spatiotemporal GAM. One option was to include both the spatial and temporal variables in the GAM, which would have allowed the two groups of variables to interact. However, this approach was seen in experimentation to supress the ability of the models to capture spatial contrasts in concentrations for both the NO_2 and $PM_{2.5}$ models. Therefore, the two-stage approach which was adopted in this paper was favoured.

4.2. Variables included in the spatiotemporal land use regression models

The models incorporated spatial variables describing the road network, land use and vegetation. Land use and vegetation variables represented the influence of areas of greenspace in Chongqing where the urban background concentration was lower due to the absence of air pollution sources. The length of highways (motorways, trunk roads, primary roads, secondary roads and tertiary roads) within a 500 m circular buffer, which was included in both $PM_{2.5}$ and NO_2 models, accounted for the influence of traffic on air pollution concentrations. In areas of high road density there were likely more high-rise buildings and narrower streets. Therefore, the highways variable may have also acted as a proxy for the effect of street canyons where air pollution dispersion from traffic is reduced resulting in higher pollutant concentrations at the road-side (Park et al., 2004).

The temporal GAMs incorporated meteorological variables and a variable to describe month which modelled the seasonal change in pollutant concentration. In both the $PM_{2.5}$ and NO_2 temporal GAMs, inclusion of horizontal visibility greatly increased GAM performance and was likely to have accounted for the level of smog. In the NO_2 GAM, wind speed was included to account for reduced NO_2 concentrations associated with higher wind speed (Zhang et al., 2015). In the $PM_{2.5}$ GAM, the binary rainfall variable accounted for the wet deposition of particulate matter which reduces concentration (Guo et al., 2016). An explanation for the observed relationships between pollutant concentrations with daily average relative humidity and temperature in the GAMs was unclear. Both temperature and humidity demonstrated seasonal variation (whereby temperature tended to be lower in winter and higher in summer, and humidity tended to be higher in winter and lower in summer). Therefore, interpretation of the relationships between pollutant concentration with daily average temperature and relative humidity was complicated by concurvity (the nonparametric analogue of multicollinearity; Ramsey et al., 2003) between the daily average relative humidity and temperature variables with the month variable. The same meteorological variables included in the spatiotemporal models have also been included in other $PM_{2.5}$ and NO_2 spatiotemporal models (Shi et al., 2018; Zhang et al., 2020), which highlights their relevance in different settings. Overall, weather conditions had a greater influence on the variability in $PM_{2.5}$ concentrations than the variability in NO_2 concentrations.

Anand and Monks (2017) and Shi et al. (2018) incorporated satellite derived data including column density (OMI) for NO2 and AOD for PM2.5 and found that these variables improved LUR model performance. Chongqing has the second lowest number of sunshine hours for cities in China (Wong, 2020) and subsequently satellite derived data was missing on many days due to cloud cover. Furthermore, we could not establish a daily dataset for NO_2 from OMI. CTMs have also been shown to improve LUR model performance but at the time of this study there were no suitable CTMs for Chongqing. Limitations with these data sources meant that only meteorological variables were incorporated into the temporal component of the LUR models.

4.3. Comparison with other land use regression models developed for regions in China

Wu et al. (2015) developed a spatial $PM_{2.5}$ LUR model for Chongqing with similar variables to represent the main $PM_{2.5}$ sources and sinks, but with different buffer sizes (primary road length within a 1000 m buffer, cropland within a 500 m buffer), and also elevation. Wu et al. (2015) reported higher model R^2 (0.84) than the spatial component of the $PM_{2.5}$ spatiotemporal model developed in the present study. However, their inclusion of measurements taken at Site 1, excluded from model development in the present study, may have been an outlier and resulted in inflation of model performance. This conjecture is supported by Wu et al. (2015) including an elevation variable which likely reflected the influence of Site 1 (altitude of 724 m) as it is 296 m higher in altitude than the next highest site and there is a comparatively low level of variability in elevation between the other sites. Direct comparison of the performance of the spatial component of our model and Wu et al. (2015) is not possible as they are based on measured concentrations from different years.

The spatiotemporal models developed for Wuhan, and subsequently used in birth cohort studies, were based on measurements taken at 10 to 20 monitoring sites (Kang et al., 2020; Liao et al., 2019; Liu et al., 2019; Song et al., 2019). Five spatial variables were incorporated into a $PM_{2.5}$ spatiotemporal model (Kang et al., 2020; Liao et al., 2019; Liu et al., 2019) and six spatial variables were incorporated into the NO_2 spatiotemporal model (Song et al., 2019). Models developed in the current study included only two spatial variables based on prior rules for variable selection (see section 2.3.3). We have provided a parsimonious model given the level of support (i.e. relatively low number of monitoring sites). There may be other sources of $PM_{2.5}$ and NO_2, such as discrete areas of industry, that we were unable to represent due to limitations with the data available to produce predictor variables. Nonetheless, were able to differentiate exposures for the study population in terms of proximity to sources using major roads and sinks using rural areas and vegetation.

Although the Wuhan $PM_{2.5}$ spatiotemporal model reported the same model R^2 as that achieved by the $PM_{2.5}$ model developed in the present study (average daily $R^2 = 0.72$), lack of model parsimony may have increased the risk of the Wuhan models overfitting the data which would have inflated the model R^2 value (Gillespie et al., 2016; Hawkins, 2004). Since the risk of overfitting increases in models based on measurements taken at a relatively low number of sites (Gillespie et al., 2016), in the present study development of a parsimonious model was considered preferable.

Spatiotemporal LUR models developed for regions in China using measurements taken at a limited number of sites typically used cross-validation techniques to validate models. Basagaña et al. (2012) found that model R^2 and cross-validation R^2 were inflated in spatial LUR models developed using measurements taken at a limited number of monitoring sites and suggested that hold-out validation (HOV) provided a better indicator of model performance, but is not possible for studies with a low number of sites (i.e. ~20). We believe that there have been no studies which have investigated the effect of the number of monitoring sites on performance of spatiotemporal LUR models. In the present study, the HOV MSE-R^2 values of both $PM_{2.5}$ and NO_2 spatiotemporal models were similar to the model development MSE-R^2 values, with the exception of monthly models for NO_2 but we recognise the weakness of our study is the relatively lower number of measurement sites which may have limited our capacity to fully capture spatiotemporal contrasts in concentrations. In the absence of detailed emissions inventories to apply dispersion modelling, approaches that rely on statistical methods, such as LUR, would benefit from increasing the number of sites in routine monitoring networks or undertaking supplementary measurements at other locations. The latter however has implications for the affordability of these types of study.

Barratt et al. (2018) demonstrated that reference sites, which provide a good measure of background air pollutant concentration, can be used to strengthen the temporal component of models. In Chongqing, one option was to use Site 1 as a reference site to represent background concentration levels on a daily basis. The use of a reference site would however have prevented predictions being made on days where measurement data was missing and would have risked the model becoming non-functional for adaptation to years outside of the study period if that site closed.

4.4. Strengths and limitations

This study had two main strengths. Firstly, the use of HOV provided a more robust test of performance of spatiotemporal models than the cross-validation techniques used in other studies which have developed spatiotemporal LUR models in China. Secondly, since the variables included in the spatiotemporal LUR models are likely to be freely available for most urban areas, our models could be transferred and locally calibrated for use in other regions.

The study faced several limitations which provide direction for future research. In addition to the low number of monitoring sites mentioned above, we lacked data on traffic intensity of the road network and were therefore limited to creating variables based on road length or distance to road. Chongqing is characterised by high-rise buildings, yet we did not have data on individual building geometry with heights which limited our ability to statistically represent the effects of building density on dispersion (i.e. street canyons) and concentration variability. Previous studies have used potential predictor variables describing building height and street configuration such as aspect-ratio. Barratt et al. (2018) developed a three-dimensional LUR model for Hong-Kong and reported that the addition of street configuration and building height to reflect pollutant concentration in street canyons improved model performance. However, in their preferred LUR models, only the nitric oxide model included a building height variable and the other models incorporated variables similar to the potential predictor variables included in our study. Other studies have reported increases in model R^2 with inclusion of urban morphology variables between 2% and 13% (Eeftens et al., 2013; Shi et al., 2016; Su et al., 2008; Tang et al., 2013). Su et al. (2008) used measurements taken at 15 monitoring sites, which was similar to the number present in Chongqing, and they reported an increase in NO_2 model R^2 of 0.11 with inclusion of variables on building height, suggesting that inclusion of such variables in the present study would have improved model performance. Other studies used measurement data taken at a much greater number of monitoring sites than were available in Chongqing ($\geq$100 for Barratt et al. (2018) and Eeftens et al. (2013) and >40 for Tang et al. (2013)) and Shi et al. (2016) used data provided by a mobile measurement campaign, which reduces their relevance to the present study. In the future, with increased LiDAR mapping coverage, it may be possible to obtain data regarding building heights in Chongqing for inclusion in LUR models.

Meteorological data was only available from two weather stations and it was unclear whether the weather experienced at these two stations would be experienced throughout the study area. Finally, some missing air pollution measurements may have negatively impacted LUR model development and performance. However, since the amount of missing data was small (<3%), the effect on model performance was likely to be minimal.

5. Conclusions

Air pollution exposure models are now available to study pregnancy outcomes in the CLIMB study in Chongqing. We recommend that $PM_{2.5}$ models are used for predicting absolute exposure whereas NO_2 models are used for relative ranking of exposures. Considering the data collected in the CLIMB study, the exposure estimates made could be used in future research studying the effects of maternal air pollution exposure on adverse pregnancy outcomes as well as infant cognitive development and health.

Credit author statement

Alexander Harper - Conceptualization; Data curation; Formal analysis; Methodology; Validation; Roles/Writing - original draft; Writing - review & editing. **Philip N. Baker** - Conceptualization; Funding acquisition; Writing - review & editing. **Yinyin Xia** - Writing - review & editing. **Tao Kuang** - Data curation; Writing - review & editing. **Hua Zhang** - Writing - review & editing. **Yingxin Chen** - Writing - review & editing. **Ting-Li Han** - Writing - review & editing. **John Gulliver** - Conceptualization; Formal analysis; Methodology; Validation; Roles/Writing - original draft; Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research was supported by funding from the Wolfson Foundation as part of the Wolfson Intercalated Award. We would like to thank everyone involved in the CLIMB study for providing the locations and dates of pregnancy of the pregnant women cohort. The CLIMB study was supported financially by the New Zealand Primary Growth Partnership postfarm gate dairy programme which is funded by Fonterra Co-operative Group Ltd, New Zealand and the New Zealand Ministry for Primary Industries.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apr.2021.101096.

References

Abernethy, R.C., Allen, R.W., McKendry, I.G., Brauer, M., 2013. A land use regression model for ultrafine particles in Vancouver, Canada. Environ. Sci. Technol. 47 (10), 5217–5225.

Anand, J.S., Monks, P.S., 2017. Estimating daily surface NO_2 concentrations from satellite data - a case study over Hong Kong using land use regression models. Atmos. Chem. Phys. 17 (13), 8211–8230.

Barratt, B., Lee, M., Wong, P., Tang, R., Tsui, T.H., Cheng, W., Yang, Y., Lai, P.C., Tian, L., Thach, T.Q., Allen, R., Brauer, M., 2018. A Dynamic Three-Dimensional Air Pollution Exposure Model for Hong Kong. Health Effects Institute Research Report, p. 194.

Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., Foraster, M., Nazelle, A., Nieuwenhuijsen, M., Vila, J., Künzli, N., 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. Atmos. Environ. 54, 634–642.

Brauer, M., Lencar, C., Tamburic, L., Koehoorn, M., Demers, P., Karr, C., 2008. A cohort study of traffic-related air pollution impacts on birth outcomes. Environ. Health Perspect. 116 (5), 680–686.

Buchhorn, M., Smets, B., Bertels, L., Lesiv, M., Tsendbazar, N.-E., Herold, M., Fritz, S., 2019. Copernicus Global Land Service: Land Cover 100m: Epoch 2015. Globe'. Retrieved from: http://doi.org/10.5281/zenodo.3243509. (Accessed 20 October 2019).

Carslaw, D.C., Ropkins, K., 2012. Openair — an R package for air quality data analysis. Environ. Model. Software 27–28, 52–61.

Chongqing Ecological Protection Bureau, 2020. Atmospheric Environmental Quality. Available at: http://sthjj.cq.gov.cn/hjzl_249/. (Accessed 20 September 2019).

Chongqing Municipal Bureau of Statistics, 2015. Chongqing Statistical Yearbook 2015. China statistics press, 2016.

De Hoogh, K., Korek, M., Vienneau, D., Keuken, M., Kukkonen, J., Nieuwenhuijsen, M., Badaloni, C., Beelen, R., Bolignano, A., Cesaroni, G., Pradas, M., Cyrys, J., Douros, J., Eeftens, M., Forastiere, F., Forsberg, B., Fuks, K., Gehring, U., Gryparis, A., Gulliver, J., Hansell, A., Hoffmann, B., Johansson, C., Jonkers, S., Kangas, L., Katsouyanni, K., Künzli, N., Lanki, T., Memmesheimer, M., Moussiopoulos, N., Modig, L., Pershagen, G., Probst-Hensch, N., Schindler, C., Schikowski, T., Sugiri, D., Teixidó, O., Tsai, M., Yli-Tuomi, T., Brunekreef, B., Hoek, G., Bellander, T., 2014. Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies. Environ. Int. 73, 382–392.

De Hoogh, K., Gulliver, J., van Donkelaar, A., Martin, R.V., Marshall, J.D., Bechle, M.J., Cesaroni, G., Pradas, M.C., Dedele, A., Eeftens, M., Forsberg, B., Galassi, C., Heinrich, J., Hoffmann, B., Jacquemin, B., Katsouyanni, K., Korek, M., Künzli, N., Lindley, S.J., Lepeule, J., Meleux, F., de Nazelle, A., Nieuwenhuijsen, M., Nystad, W., Raaschou-Nielsen, O., Peters, A., Peuch, V.-H., Rouil, L., Udvardy, O., Slama, R., Stempfelet, M., Stephanou, E.G., Tsai, M.Y., Yli-Tuomi, T., Weinmayr, G., Brunekreef, B., Vienneau, D., Hoek, G., 2016. Development of West-European $PM_{2.5}$ and NO_2 land use regression models incorporating satellite-derived and chemical transport modelling data. Environ. Res. 151, 1–10.

Didan, K., 2015. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. Retrieved from: https://doi.org/10.5067/MODIS/MOD13Q1.006. (Accessed 30 May 2020).

Dimakopoulou, K., Samoli, E., Katsouyanni, K., 2018. Spatio-temporal land use regression modelling of ozone levels in Athens, Greece. Glob. NEST J. 22 (1), 85–94.

Eeftens, M., Beekhuizen, J., Beelen, R., Wang, M., Vermeulen, R., Brunekreef, B., Huss, A., Hoek, G., 2013. Quantifying urban street configuration for improvements in air pollution models. Atmos. Environ. 72 (2013), 1–9.

Gillespie, J., Beverland, I.J., Hamilton, S., Padmanabhan, S., 2016. Development, evaluation, and comparison of land use regression modeling methods to estimate residential exposure to nitrogen dioxide in a cohort study. Environ. Sci. Technol. 50 (20), 11085–11093.

Gulliver, J., de Hoogh, K., Hoek, G., Vienneau, D., Fecht, D., Hansell, A., 2016. Back-extrapolated and year-specific NO_2 land use regression models for Great Britain - do they yield different exposure assessment? Environ. Int. 92–93 (2016), 202–209.

Guo, L.C., Zhang, Y., Lin, H., Zeng, W., Liu, T., Xiao, J., Rutherford, S., You, J., Ma, W., 2016. The washout effects of rainfall on atmospheric particulate pollution in two Chinese cities. Environ. Pollut. 215, 195–202.

Gurnida, D.A., Rowan, A.M., Idjradinata, P., Muchtadi, D., Sekarwana, N., 2012. Association of complex lipids containing gangliosides with cognitive development of 6-month-old infants'. Early Hum. Dev. 88 (8), 595–601.

Habermann, M., Billger, M., Haeger-Eugensson, M., 2015. Land use regression as method to model air pollution. Prev. Res. Gothenburg/Sweden', Procedia Eng. 115, 21–28.

Hawkins, D.M., 2004. The problem of overfitting. J. Chem. Inf. Comput. Sci. 44 (1), 1–12.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 42 (33), 7561–7578.

Huang, S., Mo, T.-T., Norris, T., Sun, S., Zhang, T., Han, T.-L., Rowan, A., Xia, Y.-Y., Zhang, H., Qi, H.-B., Baker, P.N., 2017. The CLIMB (Complex Lipids in Mothers and Babies) study: protocol for a multicentre, three-group, parallel randomised controlled trial to investigate the effect of supplementation of complex lipids in pregnancy, on maternal ganglioside status and subsequent cognitive outcomes in the offspring. BMJ Open 7 (10), e016637. https://doi.org/10.1136/bmjopen-2017-016637.

Kang, J., Liao, J., Xu, S., Xia, W., Li, Y., Chen, S., Lu, B., 2020. Associations of exposure to fine particulate matter during pregnancy with maternal blood glucose levels and gestational diabetes mellitus: potential effect modification by ABO blood group. Ecotoxicol. Environ. Saf. 198, 110673.

Knibbs, L.D., Coorey, C.P., Bechle, M.J., Marshall, J.D., Hewson, M.G., Jalaludin, B., Morgan, G.G., Barnett, A.G., 2018. Long-term nitrogen dioxide exposure assessment using back-extrapolation of satellite-based land-use regression models for Australia. Environ. Res. 163 (2018), 16–25.

Kong, L., Tian, G., 2020. Assessment of the spatio-temporal pattern of PM2.5 and its driving factors using a land use regression model in Beijing, China. Environ. Monit. Assess. 192 (2), 95.

Li, C., Yang, M., Zhu, Z., Sun, S., Zhang, Q., Cao, J., Ding, R., 2020. Maternal Exposure to Air Pollution and the Risk of Low Birth Weight: A Meta-Analysis of Cohort Studies, vol. 190. Environmental Research, p. 109970.

Liao, J., Li, Y., Wang, X., Zhang, B., Xia, W., Peng, Y., Zhang, W., Cao, Z., Zhang, Y., Liang, S., Hu, K., Xu, S., 2019. Prenatal exposure to fine particulate matter, maternal hemoglobin concentration, and fetal growth during early pregnancy: associations and mediation effects analysis. Environ. Res. 173, 366–372.

Liao, T., Gui, K., Jiang, W., Wang, S., Wang, B., Zeng, Z., Che, H., Wang, Y., Sun, Y., 2018. Air stagnation and its impact on air quality during winter in Sichuan and Chongqing, southwestern China, 635. Science of The Total Environment, pp. 576–585.

Liu, H., Liao, J., Jiang, Y., Zhang, B., Yu, H., Kang, J., Hu, C., Li, Y., Xu, S., 2019. Maternal exposure to fine particulate matter and the risk of fetal distress. Ecotoxicol. Environ. Saf. 170, 253–258.

National Centers for Environmental Information, 2020. Climate Data Online. Available at: https://www.ncdc.noaa.gov/cdo-web/. (Accessed 14 March 2020).

Nemmar, A., Hoet, P.H., Vanquickenborne, B., Dinsdale, D., Thomeer, M., Hoylaerts, M. F., Vanbilloen, H., Mortelmans, L., Nemery, B., 2002. Passage of inhaled particles into the blood circulation in humans. Circulation 105 (4), 411–414.

Norris, T., Souza, R., Xia, Y., Zhang, T., Rowan, A., Gallier, S., Zhang, H., Qi, H., Baker, P., 2019. Effect of supplementation of complex milk lipids in pregnancy on fetal growth: results from the Complex Lipids in Mothers and Babies (CLIMB) randomized controlled trial. J. Matern. Fetal Neonatal Med. 1–10. Ahead-of-print.

OpenStreetMap contributors, 2015. Planet Dump Retrieved from. Available at: https://planet.openstreetmap.org https://data.nextgis.com/en/. (Accessed 10 October 2019).

Park, S.-K., Kim, S.-D., Lee, H., 2004. Dispersion characteristics of vehicle emission in an urban street canyon. Sci. Total Environ. 323 (1), 263–271.

Pedersen, M., Giorgis-Allemand, L., Bernard, C., Aguilera, I., Andersen, A., Ballester, F., Beelen, R., Chatzi, L., Cirach, M., Danileviciute, A., Dedele, A., Eijsden, M., Estarlich, M., Fernández-Somoano, A., Fernández, M., Forastiere, F., Gehring, U., Grazuleviciene, R., Gruzieva, O., Heude, B., Hoek, G., Hoogh, K., van den Hooven, E., Håberg, S., Jaddoe, V., Klümper, C., Korek, M., Krämer, U., Lerchundi, A., Lepeule, J., Nafstad, P., Nystad, W., Patelarou, E., Porta, D., Postma, D., Raaschou-Nielsen, O., Rudnai, P., Sunyer, J., Stephanou, E., Sørensen, M., Thiering, E., Tuffnell, D., Varró, M., Vrijkotte, T., Wijga, A., Wilhelm, M., Wright, J., Nieuwenhuijsen, M., Pershagen, G., Brunekreef, B., Kogevinas, M., Slama, R., 2013. Ambient air pollution and low birthweight: a European cohort study (ESCAPE). Lanc. Respirat. Med. 1 (9), 695–704.

Ramsey, T.O., Burnett, R.T., Kreski, D., 2003. The effect of concurvity in generalized additive models linking mortality to ambient particulate matter. Epidemiology 14 (1), 18–23.

Raspisaniye Pogodi Ltd, 2020. Weather Archive in Chongqing/Jiulongpo (Airport), METAR. Available at: https://rp5.ru/Weather_archive_in_Chongqing,_Jiulongpo_(airport),_METAR. (Accessed 10 October 2019).

Robledo, C.A., Mendola, P., Yeung, E., Männistö, T., Sundaram, R., Liu, D., Ying, Q., Sherman, S., Grantz, K.L., 2015. Preconception and early pregnancy air pollution exposures and risk of gestational diabetes mellitus. Environ. Res. 137, 316–322.

Shi, Y., Ho, H.C., Xu, Y., Ng, E., 2018. Improving satellite aerosol optical depth-$PM_{2.5}$ correlations using land use regression with microscale geographic predictors in a high-density urban context. Atmos. Environ. 190, 23–34.

Shi, Y., Lau, K.K.-L., Ng, E., 2016. Developing street-level $PM_{2.5}$ and PM_{10} land use regression models in high-density Hong Kong with urban morphological factors. Environ. Sci. Technol. 50 (15), 8178–8187.

Song, L., Zhang, B., Liu, B., Wu, M., Zhang, L., Wang, L., Xu, S., Cao, Z., Wang, Y., 2019. Effects of maternal exposure to ambient air pollution on newborn telomere length. Environ. Int. 128, 254–260.

Stieb, D.M., Chen, L., Beckerman, B., Jerrett, M., Crouse, D., Omariba, D., Peters, P., van Donkelaar, A., Martin, R., Burnett, R., Gilbert, N., Tjepkema, M., Liu, S., Dugandzic, R., 2016. A national study of the association between traffic-related air pollution and adverse pregnancy outcomes in Canada, 1999–2008. Environ. Res. 148, 513–526.

Su, J.G., Brauer, M., Buzzelli, M., 2008. Estimating urban morphometry at the neighborhood scale for improvement in modeling long-term average air pollution concentrations. Atmos. Environ. 42 (34), 7784–7893.

Suwa, T., Hogg, J.C., Quinlan, K.B., Ohgami, A., Vincent, R., van Eeden, S.F., 2002. Particulate air pollution induces progression of atherosclerosis. J. Am. Coll. Cardiol. 39 (6), 935–942.

Tadono, T., Nagai, H., Ishida, H., Oda, F., Naito, S., Minakawa, K., Iwamoto, H., 2016. Generation of the 30 m-mesh global digital surface model by ALOS Prism. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. 41 (B4), 157–162.

Tang, R., Blangiardo, M., Gulliver, J., 2013. Using building heights and street configuration to enhance intraurban PM_{10}, NO_X, and NO_2 land use regression models. Environ. Sci. Technol. 47, 11643–11650.

Tian, Y., Yao, X., Chen, L., 2019. Analysis of spatial and seasonal distributions of air pollutants by incorporating urban morphological characteristics. Comput. Environ. Urban Syst. 75, 35–48.

Vickers, M.H., Guan, J., Gustavsson, M., Krägeloh, C.U., Breier, B.H., Davison, M., Fong, B., Norris, C., McJarrow, P., Hodgkinson, S.C., 2009. Supplementation with a mixture of complex lipids derived from milk to growing rats results in improvements in parameters related to growth and cognition. Nutr. Res. 29 (6), 426–435.

Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., Brunekeef, B., 2012. Systematic evaluation of land use regression models for NO_2. Environ. Sci. Technol. 46 (8), 4481–4489.

Wang, Y., Xu, C., Ren, J., Zhao, Y., Li, Y., Wang, L., Yao, S., 2020. The long-term effects of meteorological parameters on pertussis infections in Chongqing, China, 2004–2018. Sci. Rep. 10, 17235.

Wong, S., 2020. Climate in China: Annual Average Sunshine Hours in Major Cities 2018. Available at: https://www.statista.com/statistics/282497/china-annual-average-sunshine-hours-in-major-cities/. (Accessed 16 July 2020).

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. Roy. Stat. Soc. B 73 (1), 3–36.

Wu, J.-S., Liao, X., Peng, J., Huang, X.-L., 2015. [Simulation and influencing factors of spatial distribution of PM2.5 concentrations in Chongqing]. Huanjing Kexue 36 (3), 759–767.

Xinhua, 2012. Meteorologists Debunk China's 'Four Furnaces. China Daily Europe, 23 August 2012. Available at: http://europe.chinadaily.com.cn/china/2012-08/23/content_15701995.htm. (Accessed 20 July 2020).

Xu, J., Yang, W., Han, B., Wang, M., Wang, Z., Zhao, Z., Bai, Z., Vedal, S., 2019. An advanced spatio-temporal model for particulate matter and gaseous pollutants in Beijing, 211. Atmospheric Environment, China, pp. 120–127.

Xu, X., 2017. China's Population Spatial Distribution Kilometer Grid. Available at: http://www.resdc.cn/DOI/DOI.aspx?DOIid=32. (Accessed 15 January 2020).

Zhang, H., Xu, T., Zong, Y., Tang, H., Liu, X., Wang, Y., 2015. Influence of meteorological conditions on pollutant dispersion in street canyon. Procedia Eng. 121, 899–905.

Zhang, X., Just, A.C., Hsu, H.-H.L., Kloog, I., Woody, M., Mi, Z., Rush, J., Georgopoulos, P., Wright, R.O., Stroustrup, A., 2020. A hybrid approach to predict daily NO_2 concentrations at city block scale. Sci. Total Environ. 143279.