Environment International 116 (2018) 300–307

Contents lists available at ScienceDirect

Environment International

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

Spatiotemporal modeling of $PM_{2.5}$ concentrations at the national scale combining land use regression and Bayesian maximum entropy in China

Li Chen[a], Shuang Gao[a], Hui Zhang[a], Yanling Sun[a], Zhenxing Ma[a], Sverre Vedal[b], Jian Mao[a,*], Zhipeng Bai[a,c,**]

[a] School of Geographic and Environmental Sciences, Tianjin Normal University, Tianjin 300387, China
[b] Department of Environmental and Occupational Health Sciences, University of Washington School of Public Health, 4225 Roosevelt Way Ave NE, Suite 100, Seattle, WA 98105, USA
[c] State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China

ARTICLE INFO

Keywords:
$PM_{2.5}$ pollution
Land use regression
Bayesian maximum entropy
China
Spatio-temporal model

ABSTRACT

Concentrations of particulate matter with aerodynamic diameter < 2.5 μm ($PM_{2.5}$) are relatively high in China. Estimation of $PM_{2.5}$ exposure is complex because $PM_{2.5}$ exhibits complex spatiotemporal patterns. To improve the validity of exposure predictions, several methods have been developed and applied worldwide. A hybrid approach combining a land use regression (LUR) model and Bayesian Maximum Entropy (BME) interpolation of the LUR space-time residuals were developed to estimate the $PM_{2.5}$ concentrations on a national scale in China. This hybrid model could potentially provide more valid predictions than a commonly-used LUR model. The LUR/BME model had good performance characteristics, with $R^2 = 0.82$ and root mean square error (RMSE) of 4.6 μg/m³. Prediction errors of the LUR/BME model were reduced by incorporating soft data accounting for data uncertainty, with the R^2 increasing by 6%. The performance of LUR/BME is better than OK/BME. The LUR/BME model is the most accurate fine spatial scale $PM_{2.5}$ model developed to date for China.

1. Introduction

China has experienced rapid economic growth and urbanization which has been accompanied by increased emissions of particulate matter with aerodynamic diameter < 2.5 μm ($PM_{2.5}$). In 2014 the Chinese government launched a nationwide monitoring program for $PM_{2.5}$ in an effort to obtain good data on the spatial distribution of $PM_{2.5}$ nationally. The number of monitoring sites increased from > 900 in 2014 to > 1400 in 2016 (http://106.37.208.233:20035/). In 2016, $PM_{2.5}$ was measured in 338 cities in China (http://datacenter.mep.gov.cn/). The annual average $PM_{2.5}$ concentration averaged over these cities was 47 μg/m³ (range 12–158) in 2016, which was 6% less than that measured in 2015. There are two annual average $PM_{2.5}$ concentration standard classes for the China National Ambient Air Quality Standard (No. GB3095-2012): the Class I standard at 15 μg/m³ and the Class II standard at 35 μg/m³. The annual average $PM_{2.5}$ concentration standards of the World Health Organization (WHO), the European Union (EU) and the United States Environmental Protection Agency (USEPA) are 10 μg/m³, 25 μg/m³ and 12 μg/m³, respectively. Most cities had relatively high concentrations, with only 1.2% of these cities having concentrations of 15 μg/m³ or less, 26.9% between 16 and 35 μg/m³, and 71.9% above 35 μg/m³. The annual average $PM_{2.5}$ concentrations of the Beijing-Tianjin-Hebei, Yangtze River Delta and Pearl River Delta regions were 71 μg/m³, 46 μg/m³, and 32 μg/m³, respectively. This national air quality monitoring network provides an opportunity to estimate accurate $PM_{2.5}$ concentrations across China. Based on a recent source apportionment performed by the Ministry of Environment (http://dqhj.mep.gov.cn/dqhjzl/dqklwyjx/), in recent years the major sources of $PM_{2.5}$ in China are traffic emissions, coal combustion, industrial emissions and resuspended dust. Concern over $PM_{2.5}$ pollution is becoming more prevalent in China due to its harmful effects on human health and the environment. Numerous studies have shown that exposure to $PM_{2.5}$ pollution is associated with respiratory and cardiovascular diseases and increased mortality (Krewski et al., 2009; Burnett et al., 2014). It has been estimated that there were 1.23 million premature deaths and 25 million disability-adjusted life years (DALYs) lost in 2010 in China attributed to high concentrations of outdoor $PM_{2.5}$ (Lim et al., 2012). In addition, estimated health effects and related economic losses incurred by $PM_{2.5}$ pollution are high in China (Chen et al., 2017; Lim et al., 2012).

* Correspondence to: J. Mao, College of Urban and Environmental Science, Tianjin Normal University, 393 Binshuixi Rd., Tianjin 300387, China.
** Correspondence to: Z. Bai, Chinese Research Academy of Environmental Science, Beijing 100012, China.
E-mail addresses: mao_jian1018@163.com (J. Mao), baizp@craes.org.cn (Z. Bai).

https://doi.org/10.1016/j.envint.2018.03.047
Received 10 December 2017; Received in revised form 31 March 2018; Accepted 31 March 2018
0160-4120/

Epidemiological studies of the health effects of air pollution need accurate exposure estimates. There are several methods used to estimate air pollution concentrations across space and time, including monitoring data with spatial interpolation, chemical transport models, dispersion models, remote sensing image retrieval and land use regression (LUR) models. Air quality monitoring networks can provide air pollutant concentration at discrete points, but the points have limited spatial coverage. Interpolation methods such as kriging are often unable to capture small-scale spatial variability and produce overly smoothed concentration surfaces (Zou et al., 2015). Chemical transport models (CTM) such as Community Modeling and Analysis System (CMAQ) can simulate air pollutant concentration over large spatial domains with fine temporal resolution based on emission inventories and meteorological conditions and an advanced understanding of chemical processes, but the spatial resolution of the simulated results is relatively coarse (Zhang et al., 2012; Menut and Bessagnet, 2010). Dispersion models such as the Atmospheric Dispersion Model System (ADMS) can be used to predict air pollutants concentrations using information on meteorology, sources and emissions (De et al., 2014; Wilson and Zawar-Reza, 2006; DäDelä and Miå, 2015). Remote sensing techniques can simulate the air pollutant distributions over an area based on the relationship between air pollutant concentration and aerosol optical depth (AOD), but missing data in time as well as limited spatial resolution are limitations (Beckerman et al., 2013). LUR models have become a commonly used method to assess air pollutants exposures in epidemiologic researches in recent years. LUR models use a linear regression framework to predict air pollutants levels based on spatial predictors that include land use, traffic, population, and meteorological data. While LUR models can provide predictions at fine spatial resolution, most LUR models have been developed over relatively small spatial domains (Briggs et al., 1997; Wu et al., 2015; Chen et al., 2010a; Chen et al., 2010b; Chen et al., 2012; Meng et al., 2015; Meng et al., 2016). A few LUR models have been developed for large regions like Europe with relatively good results (Vienneau et al., 2013; De et al., 2016), but they are not appropriate for making short-term concentration estimates. No single method is capable of describing both the spatial and temporal variability with relatively high resolution. There has therefore been interest in combining or fusing these methods (De et al., 2017; Stafoggia et al., 2017; Yang et al., 2017b; Di et al., 2016; Akita et al., 2014; Beckerman et al., 2013; Reyes and Serre, 2014).

Bayesian maximum entropy (BME) is a geostatistical method for analyzing spatial/temporal data based on epistemics developed by Christakos (Christakos, 1990). Inputs to a BME analysis consist of general knowledge (G-K) describing generalizable characteristics, and the site specific knowledge (S-K) that includes hard data (corresponding to measurements) and soft data (having uncertainty characterized by a probability density function (PDF) which can be non-Gaussian). Two advantages of BME are: 1) simulating spatial/temporal variances; and 2) accounting for missing data by means of a nonlinear formulation of the PDF at each spatiotemporal point. One difference between a BME model and a CTM model is that a BME model is a geostatistical method and does not require emission inventory or meteorological data. Such data are not easy to acquire at a high spatial resolution in China at present. Therefore, our preferred approach was to initially fuse LUR with BME, and in future work to combine other inputs with the LUR/BME model as other data become available.

The goal of this study is to develop an estimation method combining LUR and BME that can be used to accurately estimate the spatial and temporal distribution of $PM_{2.5}$ concentrations across China for application in health effects analyses.

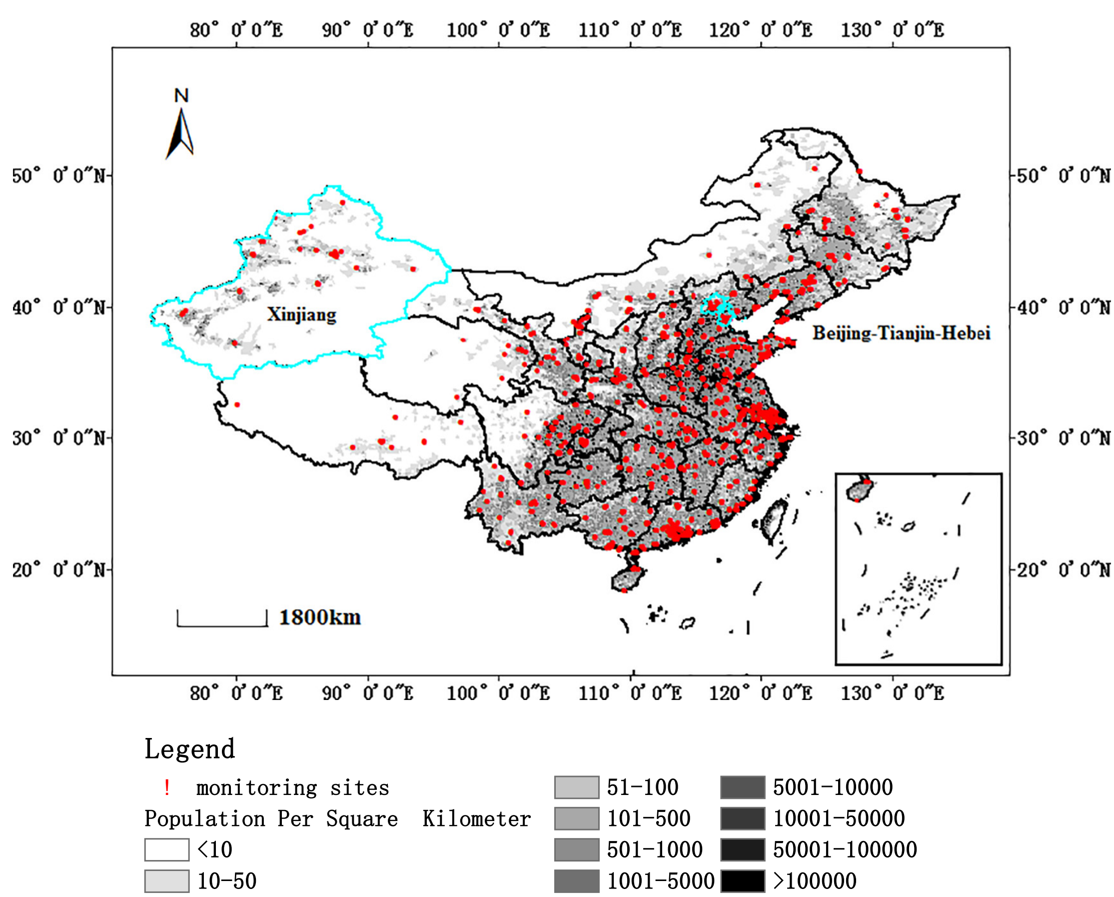

Fig. 1. Geographic locations of the $PM_{2.5}$ monitoring sites in 2016 and the spatial distribution of the population in 2010.