Atmospheric Environment 173 (2018) 295–305

Contents lists available at ScienceDirect

Atmospheric Environment

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

High-resolution spatiotemporal mapping of $PM_{2.5}$ concentrations at Mainland China using a combined BME-GWR technique

Lu Xiao[a], Yichao Lang[a], George Christakos[a,b,*]

[a] Institute of Islands and Coastal Ecosystems, Ocean College, Zhejiang University, Zhoushan, China
[b] Department of Geography, San Diego State University, San Diego, CA, USA

A R T I C L E I N F O

Keywords:
Particulate matter
Space-time analysis
Remote sensing
Aerosol optical depth
Bayesian Maximum Entropy
Geographically weighted regression
Meteorological fields

A B S T R A C T

With rapid economic development, industrialization and urbanization, the ambient air $PM_{2.5}$ has become a major pollutant linked to respiratory, heart and lung diseases. In China, $PM_{2.5}$ pollution constitutes an extreme environmental and social problem of widespread public concern. In this work we estimate ground-level $PM_{2.5}$ from satellite-derived aerosol optical depth (AOD), topography data, meteorological data, and pollutant emission using an integrative technique. In particular, Geographically Weighted Regression (GWR) analysis was combined with Bayesian Maximum Entropy (BME) theory to assess the spatiotemporal characteristics of $PM_{2.5}$ exposure in a large region of China and generate informative $PM_{2.5}$ space-time predictions (estimates). It was found that, due to its integrative character, the combined BME-GWR method offers certain improvements in the space-time prediction of $PM_{2.5}$ concentrations over China compared to previous techniques. The combined BME-GWR technique generated realistic maps of space-time $PM_{2.5}$ distribution, and its performance was superior to that of seven previous studies of satellite-derived $PM_{2.5}$ concentrations in China in terms of prediction accuracy. The purely spatial GWR model can only be used at a fixed time, whereas the integrative BME-GWR approach accounts for cross space-time dependencies and can predict $PM_{2.5}$ concentrations in the composite space-time domain. The 10-fold results of BME-GWR modeling ($R^2 = 0.883$, RMSE $= 11.39\ \mu g/m^3$) demonstrated a high level of space-time $PM_{2.5}$ prediction (estimation) accuracy over China, revealing a definite trend of severe $PM_{2.5}$ levels from the northern coast toward inland China (Nov 2015–Feb 2016). Future work should focus on the addition of higher resolution AOD data, developing better satellite-based prediction models, and related air pollutants for space-time $PM_{2.5}$ prediction purposes.

1. Introduction

China's atmospheric pollution is very serious. The main pollution sources are fuel combustion, human activities and other natural processes (e.g., dust). Ambient air pollutants that have a significant impact on human health and the environment are composed of complex materials, including particulate matter, ozone, sulfur dioxide and nitrogen dioxide. Fine particle particulate matter, $PM_{2.5}$ (aerodynamic diameters smaller than 2.5 μm), is the most problematic among these pollutants as regards public health. $PM_{2.5}$ particles can enter into the alveoli, subsequently being retained in the lung parenchyma (Christakos and Hristopulos, 1998; Dockery, 2009), thus causing severe heart disease, cardiovascular diseases, respiratory diseases and even lung cancer (Brook et al., 2010; Dhondt et al., 2011; Hoek et al., 2013). With rapid economic development, $PM_{2.5}$ pollution in China has become an extreme environmental and social problem having an important impact on the human body, the environment and the climate (Song et al., 2014; Peng et al., 2016).

Relatively small $PM_{2.5}$ datasets are available in China. Ground-based measurements are considered as the most reliable way of collecting $PM_{2.5}$ concentrations (Tao et al., 2017). Accordingly, most pollutant concentration information has been obtained from ground-level monitoring sites, a fact that has many limitations. Among them is the considerable information bias, the limited credibility of the exposure response results (Liu et al., 2007), and the sparsity and uneven distribution of the monitoring stations (Gupta and Christopher, 2008). These limitations affect the geographical and demographic range of a study, so that it is usually impossible to determine the temporal and spatial variation of $PM_{2.5}$ concentrations over large geographic areas (You et al., 2016b), For example, $PM_{2.5}$ was not included in China's national monitoring system until 2013 (Chu et al., 2016). So far, domestic and foreign scholars have carried out $PM_{2.5}$ variation analysis and generated ground level $PM_{2.5}$ concentration estimates using a variety of statistical models. Christakos and Serre (2000) and Christakos

* Corresponding author. Institute of Islands and Coastal Ecosystems, Ocean College, Zhejiang University, Zhoushan, China.
E-mail address: gchristakos@zju.edu.cn (G. Christakos).

https://doi.org/10.1016/j.atmosenv.2017.10.062
Received 11 June 2017; Received in revised form 19 October 2017; Accepted 29 October 2017
Available online 21 November 2017

et al. (2001) have used Bayesian maximum entropy (BME) theory to represent and predict spatiotemporal particulate matter distributions in North Carolina and California, USA. Wang and Christopher (2003) used linear regression models, whereas Liu et al. (2004) proposed a Chemical Transport Model (CTM). Later, Reid et al. (2015) and Donkelaar et al. (2011) also used the CTM. Lee et al. (2011) developed the day-specific Mixed-Effect Model (MEM), Lee et al. (2012) used a space-time geostatistical kriging model to estimate long-term ambient $PM_{2.5}$ exposure in U.S.A. Liu et al. (2009) and Kloog et al. (2011) proposed a two-stage generalized additive model (GAM), and Ma et al. (2014) and (Xiao et al., 2017) used Geographically Weighted Regression (GWR) tools. Hence, remote sensing techniques, spatial and temporal modeling, and statistical prediction theory have been individually or in combination employed in the quantitative assessment of air pollution and environmental health (Kim et al., 2015; Xiao et al., 2017).

In view of the above considerations, the objective of the present work is to introduce and validate in terms of real data a new satellite-based technique of composite space-time modeling and estimation of $PM_{2.5}$ concentrations in China during a four-month period: this is the combined (or integrative) Bayesian Maximum Entropy-Geographically Weighted Regression (BME-GWR) method. Advantages of the combined BME-GWR method include its rigorous consideration of the physical cross-space-time dependencies of pollutant distribution, the generation of pollutant predictions in a realistic space-time domain rather than separately, the inclusion of more general cases (non-linear and non-Gaussian predictors), and its ability to jointly incorporate different environmental predictors, including topography data (elevation) and meteorological data (wind speed, precipitation, temperature, relative humidity and pressure), as well as pollutant emission indicators (such as NO_2, CO, land use, population, and road network information).

2. Data and method

2.1. Study area

The present study focuses on the entire China, with the exception of the Xinjiang, Tibet, Qinghai, Mongolia and Heilongjiang provinces (Fig. 1), which were not included in the study because the $PM_{2.5}$ monitoring sites in these areas are sparse (while these provinces cover a total area of 5.33 million Km^2, they have only 89 monitoring sites, and this monitoring limitation would seriously affect pollutant estimation accuracy in provinces with serious pollution problems). On the other hand, the study area covers about 4.18 million Km^2 that include 1408 monitoring sites and 93% of the total population of China. For data processing and mapping purposes, the study area is covered with a grid consisting of 357,997 grid cells of $3 \times 3Km^2$ size (Xiao et al., 2017).

2.2. Data

2.2.1. Ground-level pollutant measurements

The 24*h*-averaged $PM_{2.5}$, NO_2, and CO concentrations at nationally-referenced monitoring stations during the period November 1, 2015 to February 29, 2016 were downloaded from the China Environmental Monitoring Center (CEMC, http://106.37.208.233:20035/). The observed $PM_{2.5}$ concentrations, which served as the dependent variable of the pollutant space-time prediction (estimation) techniques used in this work, include 1408 monitoring sites (Fig. 1) with a total of 3009 observations in the study area, and, also, 43 monitoring sites that were evenly distributed in adjacent to the study area provinces to avoid any edge-effects. $PM_{2.5}$, NO_2, and CO concentrations less than 2 $\mu g/m^3$ (5.6% of total records) were discarded since they are below the established detection limit (EPA, 2008). Also, stations where data were available during less than 15 days per month were removed, according to China National Ambient Air Quality Standards (CNAAQS). All 3009 observations were distributed during four months, i.e., there exist 953, 707, 529 and 820 observations during the months of November, December, January and February, respectively. Daily $PM_{2.5}$, NO_2, and CO data were used to calculate the monthly average pollutant concentration at each site, and the monthly averages were obtained using the R programming language (R version 3.3.2, https://www.r-project.org/). Notice that most of the $PM_{2.5}$ monitoring sites are clustered in urban areas (rural areas have little coverage in China).

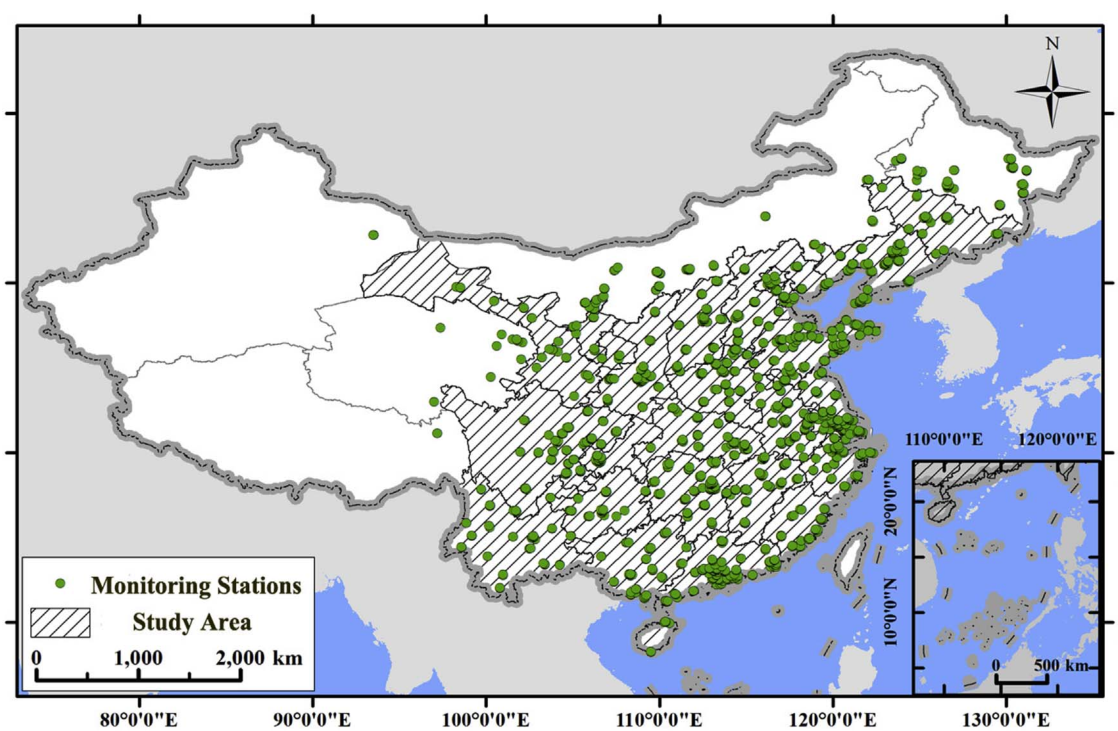

Fig. 1. Study area. The green dots represent the 1408 $PM_{2.5}$ monitoring sites within the study area and the 43 sites in neighboring provinces. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)