Environmental Pollution 240 (2018) 319–329

Contents lists available at ScienceDirect

Environmental Pollution

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

Space-time $PM_{2.5}$ mapping in the severe haze region of Jing-Jin-Ji (China) using a synthetic approach[☆]

Junyu He [a, b], George Christakos [a, b, *]

[a] Ocean College, Zhejiang University, Zhoushan, China
[b] Department of Geography, San Diego State University, San Diego, CA, USA

ARTICLE INFO

Article history:
Received 27 February 2018
Received in revised form
4 April 2018
Accepted 21 April 2018

Keywords:
BME
Dimensionality reduction
$PM_{2.5}$
Remote sensing
Uncertainty
Mapping

ABSTRACT

Long- and short-term exposure to $PM_{2.5}$ is of great concern in China due to its adverse population health effects. Characteristic of the severity of the situation in China is that in the Jing-Jin-Ji region considered in this work a total of 2725 excess deaths have been attributed to short-term $PM_{2.5}$ exposure during the period January 10–31, 2013. Technically, the processing of large space-time $PM_{2.5}$ datasets and the mapping of the space-time distribution of $PM_{2.5}$ concentrations often constitute high-cost projects. To address this situation, we propose a synthetic modeling framework based on the integration of (*a*) the Bayesian maximum entropy method that assimilates auxiliary information from land-use regression and artificial neural network (ANN) model outputs based on $PM_{2.5}$ monitoring, satellite remote sensing data, land use and geographical records, with (*b*) a space-time projection technique that transforms the $PM_{2.5}$ concentration values from the original spatiotemporal domain onto a spatial domain that moves along the direction of the $PM_{2.5}$ velocity spread. An interesting methodological feature of the synthetic approach is that its components (methods or models) are complementary, i.e., one component can compensate for the occasional limitations of another component. Insight is gained in terms of a $PM_{2.5}$ case study covering the severe haze Jing-Jin-Ji region during October 1–31, 2015. The proposed synthetic approach explicitly accounted for physical space-time dependencies of the $PM_{2.5}$ distribution. Moreover, the assimilation of auxiliary information and the dimensionality reduction achieved by the synthetic approach produced rather impressive results: It generated $PM_{2.5}$ concentration maps with low estimation uncertainty (even at counties and villages far away from the monitoring stations, whereas during the haze periods the uncertainty reduction was over 50% compared to standard $PM_{2.5}$ mapping techniques); and it also proved to be computationally very efficient (the reduction in computational time was over 20% compared to standard mapping techniques).

1. Introduction

Numerous studies have shown that chronic and acute exposure to ambient $PM_{2.5}$ can increase the risk of cardiovascular and respiratory morbidity and mortality (e.g., Zhang et al., 2013; Franklin et al., 2007; Lepeule et al., 2012). Specifically, ambient $PM_{2.5}$ pollution influence large populations in China, contributing to about 1.6 million deaths per year (Rohde and Muller, 2015). Hence, the adequate assessment of the space-time distribution of $PM_{2.5}$ concentrations is a high priority for China's environmental administrators and public health decision makers seeking to control air pollution and its adverse health effects. As a result, various quantitative methods have been used to estimate pollutant concentrations across space and time (Christakos and Kolovos, 1999; Christakos and Serre, 2000; Yu et al., 2009; Akita et al., 2012, 2014; Ma et al. 2014, 2016; Reyes and Serre, 2014; Fang et al., 2016). Since 2013 the Chinese Ministry of Environmental Protection has established a systematic ground $PM_{2.5}$ monitoring network across the entire country in order to better understand the $PM_{2.5}$ pollutant distribution. Currently, 1500 monitoring stations exist, but most of them are located in cities, resulting in station clustering. The non-uniform distribution of monitoring stations seriously limits the quality and effectiveness of $PM_{2.5}$ exposure assessment studies.

Recently, land-use regression (LUR) modeling has proven to be a useful tool of air pollution assessment in China (Chen et al. 2010a,

☆ This paper has been recommended for acceptance by Haidong Kan.

* Corresponding author. Hai Ke Building 421, No.1 Zheda Road, Dinghai District, Zhoushan, Zhejiang Province, China.
E-mail address: gchristakos@zju.edu.cn (G. Christakos).

https://doi.org/10.1016/j.envpol.2018.04.092
0269-7491/

2010b; Wu et al., 2015; Liu et al., 2016). By considering local characteristics (land use or land cover, meteorological factors, elevation, traffic etc.), LUR often offers a reasonable pollution assessment approach based on relevant explanatory variables at the annual scale. Further, it was found that, because of its large spatial and temporal coverage, satellite remote sensing could improve $PM_{2.5}$ estimation using sparse monitoring stations. For example, aerosol optical depth (AOD) derived from satellite remote sensing has been introduced in $PM_{2.5}$ estimation, and the results showed a significantly improvement of LUR modeling (Mao et al., 2012).

Unlike LUR, artificial neural network (ANN) is a non-linear model with the advantages of self-learning, self-organizing and self-adaption (Feng and Hong, 2008). The ANN has been proved less biased and more accurate in the context of PM_{10} estimation than multiple regression models (Chaloulakou et al., 2003). In that study it was shown that the non-linear models were more appropriate than the linear ones in cases of complicated process-based real-world prediction. Moreover, ANN models have been used to predict $PM_{2.5}$ concentrations (Yao et al., 2012; Zhang et al., 2015, Di et al., 2016). Specifically, Yao et al. (2012) and Di et al. (2016) took into consideration satellite remote sensing data as predictors, including AOD (or AOT, aerosol optical thickness) and meteorological factors. However, it should be stressed that these and similar studies rarely discussed model prediction uncertainty across the space-time domain of interest, i.e., neither parameter uncertainty (due to model input parameters) nor conceptual uncertainty (due to the model itself) has been considered.

On the other hand, Bayesian maximum entropy (BME, Christakos, 2000) is a space-time stochastic approach that includes a set of powerful quantitative tools to represent, predict and map physical attributes at unsampled spatial locations and time instants under conditions of in-situ uncertainty. BME is based on a very general theory that considers non-Gaussian distributions and involves non-linear space-time estimators; also, it can integrate a variety of knowledge bases, including physical laws, scientific theories, theoretical models and empirical relationships (Christakos and Li, 1998; Christakos and Serre, 2000; Beckerman et al., 2013). These features distinguish BME from many mainstream techniques that are based on restrictive assumptions (Gaussian distributions, linear estimators etc.). BME has been successfully applied in many air quality, environmental exposure and public health assessment studies (e.g., Yu et al., 2010; Messier et al., 2012; Akita et al., 2012; Reyes and Serre, 2014; Yang and Christakos, 2015).

Interestingly, combinations of the LUR and BME methods have been proposed in the literature that improved the space-time estimation of $PM_{2.5}$ concentrations (e.g., Yu et al., 2011; Beckerman et al., 2013; Reyes and Serre, 2014). In particular, these studies have found that, in view of the auxiliary information provided by LUR, the joint BME-LUR framework produced more accurate estimates of $PM_{2.5}$ concentrations than previous techniques. However, as far as we know, the combination of ANN with BME has not been previously considered in the literature, which is why it is one of the objectives of the present study (see below).

The $PM_{2.5}$ distribution in the atmosphere is a physical attribute that occurs in a composite space-time domain under conditions of in-situ uncertainty. The BME approach represents the $PM_{2.5}$ distribution as a spatiotemporal random field (the term "spatiotemporal" accounts for the lawful space-time $PM_{2.5}$ pattern and "random" for the relevant in-situ uncertainties), and then provides a rigorous quantitative framework to describe important characteristics of space-time pollutant variation, such as $PM_{2.5}$ covariance functions. It is a matter of physical intuition that to obtain the $PM_{2.5}$ covariance, composite space-time pollutant dependencies and interactions need to be considered rather than separate spatial and temporal covariances (Christakos and Serre, 2000). However, as the space-time variability of a physical attribute can be very complex, the data-based calculation of a space-time $PM_{2.5}$ covariance is practically inefficient for large datasets (Wikle and Cressie, 1999). Therefore, the development of quantitative notions and techniques with the dual objective of reducing the computational cost (when dealing with large datasets) while accounting for physical space-time dependencies is a major issue of current air pollution research and development. In an effort to address this issue, Xie et al. (2001) suggested three computationally cost-effective schemes: keeping the number of data points as small as possible, calculating covariance functions as fast as possible, and calculating realistic covariance functions using as few data-pairs as possible. Unfortunately, in addition to the obvious concern that relying on a small number of data points may lead to unrealistic results, none of the above three schemes are practically useful when confronted with a case-specific dataset and a given software library. Christakos et al. (2016) proposed another way to confront these difficulties in practice based on the spatiotemporal projection (STP) technique. It was found that by means of a domain dimensionality reduction (DDR) process that transformed the original spatiotemporal domain of the attribute of interest (in our case this would be the pollutant distribution) into a lower-dimensionality domain, the STP method led to considerable savings in computational cost while producing space-time maps of similar accuracy as those produced by high computational cost techniques.

In view of the above considerations, the present work proposes a novel synthetic approach based on the integration of the ANN, BME and STP techniques. The synthetic approach was applied in practice to assess the $PM_{2.5}$ pollution situation in the seriously polluted Jing-Jin-Ji region of northern China during Oct 1–31, 2015 (i.e., accurate $PM_{2.5}$ estimates and informative maps were generated that cover the entire region and time period of interest). The synthetic approach combines the strengths and advantages of its individual components: (*a*) BME can assimilate different types of core knowledge and auxiliary information from land-use regression and ANN outputs (using $PM_{2.5}$ monitoring, satellite remote sensing data, land-use and geographical records); (*b*) as a non-linear model, the ANN performs better than linear models; and (*c*) based on the DDR transformation, the STP offers considerable modeling benefits (theoretical and computational). As a result, the synthetic approach can account for composite space-time $PM_{2.5}$ dependencies, decrease mapping uncertainty, and decrease computational cost.

2. Materials and methods

2.1. Datasets

Study Area and Monitoring $PM_{2.5}$ Data. The study area is the Jing-Jin-Ji region located in northern China. This area includes Beijing, Tianjin and Hebei province with a combined territory of approximately 218 thousand km^2 and a population of 110 *million* people. Daily mean $PM_{2.5}$ concentrations from 99 monitoring stations in 13 cities were collected during the period Oct. 1st through Oct. 31st, 2015 (Fig. 1). These 99 stations are regarded as hard data points (in this work, "hard" are termed the measurements that are uncertainty-free) and, accordingly, the measured daily mean $PM_{2.5}$ concentrations are treated as hard data. The tapered element oscillating microbalance (TEOM) and the beta-attenuation monitoring (BAM) instruments were employed to measure $PM_{2.5}$ concentrations. The mean $PM_{2.5}$ concentration and the coefficient of variation were found to be $60.58\ \mu g/m^3$ and 1.05, respectively, indicating a large variability of the monitored $PM_{2.5}$ data during Oct. 2015 (Table S1).