Environmental Research 158 (2017) 54–60

Contents lists available at ScienceDirect

Environmental Research

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

Evaluation of a data fusion approach to estimate daily $PM_{2.5}$ levels in North China

Fengchao Liang[a,c], Meng Gao[b], Qingyang Xiao[c], Gregory R. Carmichael[b], Xiaochuan Pan[a,*], Yang Liu[c,*]

[a] Department of Occupational and Environmental Health, School of Public Health, Peking University, Beijing 100191, China
[b] Center for Global and Regional Environmental Research, the University of Iowa, Iowa City, IA 52242, USA
[c] Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA

ARTICLE INFO

Keywords:
$PM_{2.5}$
KED
WRF-Chem
Data fusion
Spatiotemporal model

ABSTRACT

$PM_{2.5}$ air pollution has been a growing concern worldwide. Previous studies have conducted several techniques to estimate $PM_{2.5}$ exposure spatiotemporally in China, but all these have limitations. This study was to develop a data fusion approach and compare it with kriging and Chemistry Module. Two techniques were applied to create daily spatial cover of $PM_{2.5}$ in grid cells with a resolution of 10 km in North China in 2013, respectively, which was kriging with an external drift (KED) and Weather Research and Forecast Model with Chemistry Module (WRF-Chem). A data fusion technique was developed by fusing $PM_{2.5}$ concentration predicted by KED and WRF-Chem, accounting for the distance from the central of grid cell to the nearest ground observations and daily spatial correlations between WRF-Chem and observations. Model performances were evaluated by comparing them with ground observations and the spatial prediction errors. KED and data fusion performed better at monitoring sites with a daily model R^2 of 0.95 and 0.94, respectively and $PM_{2.5}$ was overestimated by WRF-Chem ($R^2=0.51$). KED and data fusion performed better around the ground monitors, WRF-Chem performed relative worse with high prediction errors in the central of study domain. In our study, both KED and data fusion technique provided highly accurate $PM_{2.5}$. Current monitoring network in North China was dense enough to provide a reliable $PM_{2.5}$ prediction by interpolation technique.

1. Introduction[1]

The North China region, home to 350 million people and the Beijing Metropolitan Area, is the cultural and political center of China and one of its economic hubs. For the past three decades, rapid urbanization, population growth and expansion of industrial land use have severely affected the air quality in this region, specifically fine particulate matter ($PM_{2.5}$). $PM_{2.5}$ levels in most cities in the North China remain above both the WHO guideline and China's National Ambient Air Quality Standard (NAAQS, http://kjs.mep.gov.cn/) (Hu et al., 2014; Zhang and Cao, 2015). Numerous studies worldwide have linked exposure to $PM_{2.5}$ to various adverse health effects, including respiratory and cardiovascular mortality and morbidity, birth outcomes, as well as diabetes (Brook et al., 2013; Gauderman et al., 2004; Lepeule et al., 2012; Miller et al., 2007; Stieb et al., 2016).

Previously, most studies in China have typically characterized the exposures to $PM_{2.5}$ using ground observations (Zhang and Cao, 2015). Since 2013, China has invested heavily in the development of a national regulatory monitoring network, with now over 1500 monitors in operation nationwide. Similar to the network operated by the U.S. Environmental Protection Agency, the coverage of China's regulatory network is mostly concentrated in urban areas. Since major emission sources such as coal-fired power plants and biomass burning are located in rural areas, this network alone is insufficient to fully characterize the spatial and temporal variability of $PM_{2.5}$ levels.

For the past 15 years, various techniques have been developed to fill the data gaps left by ground monitors. At one end of this spectrum of techniques, geostatistical interpolation methods such as inverse distance weighting (IDW), spline interpolation and kriging have been applied to obtain $PM_{2.5}$ concentration surfaces using ground

* Corresponding authors.
E-mail addresses: liangfengchao126@126.com (F. Liang), meng-gao@uiowa.edu (M. Gao), qingyang.xiao@emory.edu (Q. Xiao), gregory-carmichael@uiowa.edu (G.R. Carmichael), xcpan@bjmu.edu.cn (X. Pan), yang.liu@emory.edu (Y. Liu).

[1] $PM_{2.5}$, fine particulate matter; NAAQS, National Ambient Air Quality Standard; IDW, inverse distance weighting; OK, Ordinary kriging; CTMs, chemical transport models; CMAQ, Community Multiscale Air Quality; WRF-Chem, Weather Research and Forecast Model with Chemistry Module; KED, kriging with external drift; DOY, day-of-year; MOSAIC, Model for Simulating Aerosol Interactions and Chemistry; CV, cross-validation; OBS, observation.

http://dx.doi.org/10.1016/j.envres.2017.06.001
Received 18 December 2016; Received in revised form 15 April 2017; Accepted 1 June 2017
Available online 03 July 2017
0013-9351/

observations (Li and Liu, 2014; Li et al., 2016; Ramos et al., 2016; Sampson et al., 2013). For example, a study in Korea used multiple spatial interpolation methods to estimate air pollutant exposure, including average values from all monitors, nearest neighbor, IDW and Ordinary kriging (OK), which were all based on 13 ground monitors. It showed that OK provided the most accurate estimated exposures (Son et al., 2010). Since kriging and similar techniques are heavily dependent on ground data support, it works poorly in regions or during the time periods with few monitors. In China, very little $PM_{2.5}$ data exist before 2013 that can be used for spatial interpolation. At the other end of this spectrum, techniques independent from ground observations such as atmospheric chemical transport models (CTMs) can predict $PM_{2.5}$ concentrations with complete spatial and temporal coverage (Gao et al., 2016b; Hu et al., 2016). A well-known limitation of CTMs is that the errors in pollutant emissions inventory and meteorological fields can introduce substantial errors in model predictions (Hogrefe et al., 2015; Zhang et al., 2016). To take advantage of the high accuracy of kriging and full coverage of CMT predictions, a fusion approach has been proposed that OK fused with Community Multiscale Air Quality (CMAQ) model to predict 12 pollutants in Georgia, United States (Friberg et al., 2016; Puttaswamy et al., 2014).

Given that data fusion technique for kriging and CTMs has not been applied in China, our objective is to develop such a fusion approach and evaluate the accuracy of spatiotemporally resolved ambient $PM_{2.5}$ concentration predicted by kriging, Weather Research and Forecast Model with Chemistry Module (WRF-Chem) and data fusion techniques in North China.

2. Data and methods

2.1. Study domain, modeling grid and ground monitoring data

Our study domain covers Beijing, Tianjin, Hebei, Shanxi, and Shandong province as well as part of Liaoning, Inner Mongolia, Shaanxi, Henan, Anhui, and Jiangsu province (Fig. 1). The air mass over the densely populated and highly industrialized low lands is isolated by the Inner Mongolian Plateau to the north, the Taihang Mountains to the west with elevations above 1500 m, and the Bohai Sea to the east. Strong local emissions and poor dispersion conditions are major contributors to the severe $PM_{2.5}$ pollution levels (Gao et al., 2016b; Li and Liu, 2011). Daily mean $PM_{2.5}$ concentrations in 2013 from 410 monitors were obtained from China's Ministry of Environmental Protection (http://english.mep.gov.cn/) and local Environmental Protection Bureaus in the study domain, of which 362 are operated by the national government and 48 by local governments; 365 monitors are located in our study domain and the others were used to make reliable predictions at the border areas. Daily mean number (N) of operating monitors in this work was 283, with a range of 59–349. We created a 10 km resolution modeling grid in this domain in order to compare different $PM_{2.5}$ modeling results, a total of 13,326 grid cells contributed in daily prediction.

2.2. Kriging model

Kriging is a linear weighted combination of observed values that uses spatial autocorrelation among data to determine the weights with or without taking account for an external trend or trend model. Under suitable assumptions on the priors, kriging gives the best linear unbiased prediction of the intermediate values (Cressie, 1990). We developed a kriging with external drift (KED) model to spatially interpolate daily $PM_{2.5}$ concentration observations to the entire modeling grid. KED is a particular case of universal kriging, where the prediction for nonstationary processes is performed by taking into account both local trends of the target variable and external drift (a spatial trend) when minimizing the estimation variance (Chauvet and Galli, 1982; Hengl et al., 2003). Previous studies have showed that elevation is a

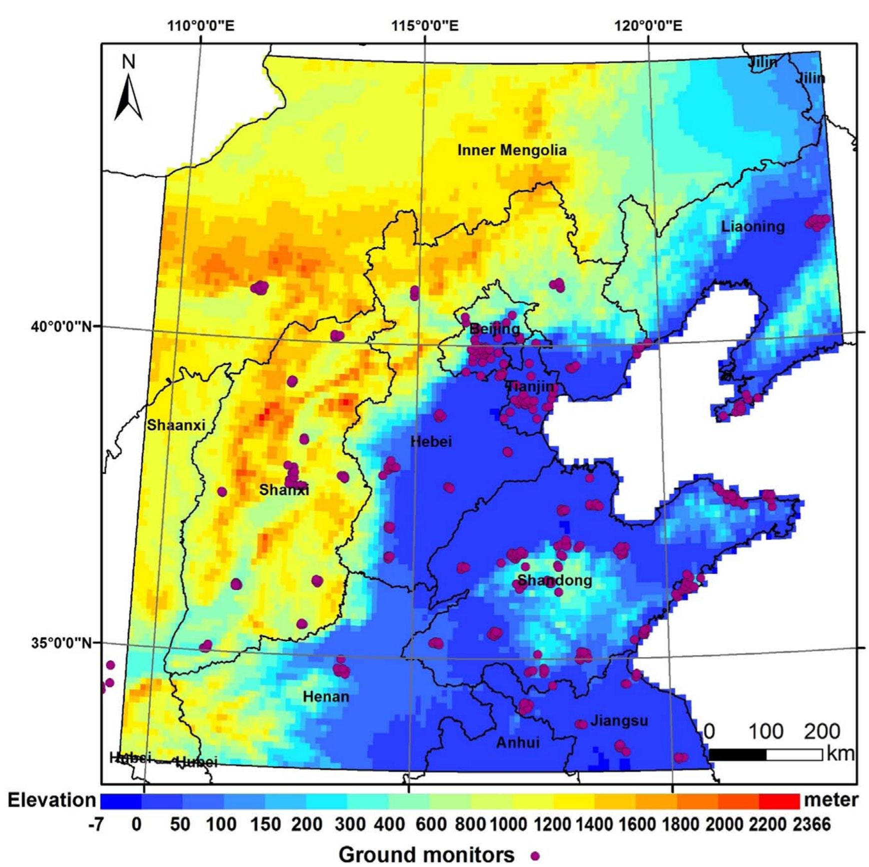

Fig. 1. Study domain and distribution of ground monitors in North China.