Atmospheric Environment 134 (2016) 51–60

Contents lists available at ScienceDirect

Atmospheric Environment

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

Application of plume analysis to build land use regression models from mobile sampling to improve model transferability

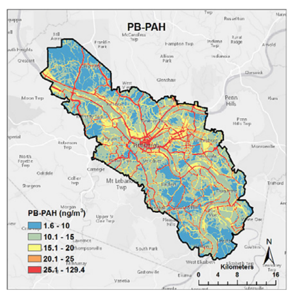
Yi Tan, Timothy R. Dallmann, Allen L. Robinson, Albert A. Presto*

Center for Atmospheric Particle Studies, Carnegie Mellon University, Pittsburgh, PA, 15213, United States

HIGHLIGHTS

- We observed spatial gradients of polycyclic aromatic hydrocarbons (PB-PAH) and BC.
- BC and PB-PAH variability is driven by plumes from high emitting vehicles.
- Two-layer models (plume + background) were developed to describe spatial patterns.
- The model plume layer is transferable to an independent holdout dataset.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 5 January 2016
Received in revised form
11 March 2016
Accepted 12 March 2016
Available online 15 March 2016

Keywords:
Air pollution
Black carbon
Exposure assessment
Geographic information systems
Land use regression

ABSTRACT

Mobile monitoring of traffic-related air pollutants was conducted in Pittsburgh, PA. The data show substantial spatial variability of particle-bound polycyclic aromatic hydrocarbons (PB-PAH) and black carbon (BC). This variability is driven in large part by pollutant plumes from high emitting vehicles (HEVs). These plumes contribute a disproportionately large fraction of the near-road exposures of PB-PAH and BC. We developed novel statistical models to describe the spatial patterns of PB-PAH and BC exposures. The models consist of two layers: a plume layer to describe the contributions of high emitting vehicles using a near-roadway kernel, and an urban-background layer that predicts the spatial pattern of other sources using land use regression. This approach leverages unique information content of highly time resolved mobile monitoring data and provides insight into source contributions. The two-layer model describes 76% of observed PB-PAH variation and 61% of BC variation. On average, HEVs contribute at least 32% of outdoor PB-PAH and 14% of BC. The transferability of the models was examined using measurements from 36 hold-out validation sites. The plume layer performed well at validation sites, but the background layer showed little transferability due to the large difference in land use between the city and outer suburbs.

1. Introduction

Exposure to traffic related air pollutants is linked with adverse health effects including childhood cancer, respiratory, and cardiovascular diseases (Brugge et al., 2007; Heck et al., 2013). The spatial variability of traffic related pollutants, such as black carbon (BC) and particle bound polycyclic aromatic hydrocarbons (PB-PAH), is substantial in urban areas (Clougherty et al., 2013; Tan et al., 2014a). However, the large spatial variation of traffic related pollutants cannot be characterized by sparse monitoring systems such as the U.S. EPA Air Quality System (AQS) that are designed to

* Corresponding author.
E-mail address: apresto@andrew.cmu.edu (A.A. Presto).

http://dx.doi.org/10.1016/j.atmosenv.2016.03.032

monitor regional compliance with the National Ambient Air Quality Standards. In many cases markers of traffic related emissions such as BC or PB-PAH are not widely monitored.

Pollutant mapping studies aim to measure pollutants at high spatial density using either saturation sampling or mobile monitoring. Saturation sampling studies collect integrated samples over multiple weeks in different seasons, but there are still large uncertainties in reproducing annual mean concentrations (Tan et al., 2014b). Nevertheless, distributed sampling data is frequently used to build land use regression (LUR) models that predict pollutant spatial patterns (Zhang et al., 2014, 2015; Wang et al., 2013; Kheirbek et al., 2012; Jedynska et al., 2014; Clougherty et al., 2008). Mobile monitoring is even more uncertain in reproducing annual mean concentrations because it collects data for shorter durations than distributed sampling. However, the ability of mobile monitoring to capture pollutant spatial patterns is comparable with saturation sampling (Tan et al., 2014b), and mobile monitoring data have been used to build LUR models (Larson et al., 2009; Patton et al., 2015; Saraswat et al., 2013). While most mobile sampling platforms are equipped with high time resolution (~1 s–1 min) instrumentation, many LUR models built from mobile sampling use highly averaged data that loses much of the informational value inherent in high time resolution measurements.

Highly time resolved data collected in mobile monitoring studies can provide important information on pollutant sources. For example, a mobile monitoring study in the Los Angeles area quantified particle emissions from the Los Angeles International Airport (Hudda et al., 2014). High time resolution data can be used to estimate emission factors for on-road traffic (Hudda et al., 2013), including analysis of individual vehicle plumes (Dallmann et al., 2011, 2012; Canagaratna et al., 2004). Mobile sampling can also identify pollutant hotspots not captured by stationary monitoring (Brantley et al., 2014). In our recent mobile monitoring campaign in the Pittsburgh region, Tan et al. analyzed pollutant plumes from high emitting vehicles (HEVs), most of which were diesel trucks and buses, to partially resolve the sources of particle bound polycyclic aromatic hydrocarbons (PB-PAH) and black carbon (BC) (Tan et al., 2014a). HEVs contributed up to 70% of the on-road PB-PAH and 30% of BC, with significant spatial variability that showed strong linear correlation between the contribution of HEVs and the Average Daily Truck Traffic (ADTT) counts (Tan et al., 2014a).

LUR are statistical relationships between land-use variables and pollutant concentrations. Land-use variables typically include traffic, zoning (e.g., industrial or residential), and elevation (Hoek et al., 2008). Some variables in typical LUR models may be indicative of pollutant sources. For example, traffic variables are related with vehicle emissions, and variables associated with industrial land use may indicate that a particular pollutant is emitted from point sources. However, LUR is not a rigorous method to apportion pollutant sources, and the regression coefficients in LUR models do not necessarily represent the contributions of specific sources (Hoek et al., 2008). LUR models may also include variables that lack obvious physical interpretability or are not directly related to sources. These limitations of LUR models limit their ability to directly attribute observed pollutant concentrations to specific sources and to predict potential changes in air quality due to mitigation strategies. Additionally, LUR and other statistical models often suffer from poor transferability. Models built for a specific city, or even a portion of a city, typically are not applicable outside of that region (Patton et al., 2015; Poplawski et al., 2009). Poor transferability is most likely a consequence of the purely statistically based, rather than physically based, representation of pollutant spatial patterns in LUR models.

An alternative method to predict spatial distribution of pollutant concentrations is the distance-weighted kernel algorithm (Loibl and Orthofer, 2001; Vienneau et al., 2009; Pratt et al., 2014; Gulliver and Briggs, 2011). This method explicitly links emissions to pollutant concentrations based on the proximity to sources and expected dispersion patterns. Compared to other geospatial approaches, the kernel method better represents the transport of pollutants away from sources by assuming a smooth fall-off near sources rather than the sharp cutoff created by using fixed-distance buffers, as recently demonstrated by Pratt et al (Pratt et al., 2014). The distance-weighted kernel therefore offers the possibility of improving model transferability. The impact of potential changes in emission sources (e.g., reduction in high emitting diesel trucks) can also be readily estimated.

In this manuscript, we develop two types of spatial models based on mobile sampling data collected in Pittsburgh, PA. The first model is a traditional LUR. The second is a novel two-layer model that leverages the unique attributes of highly time resolved data to predict the spatial patterns of PB-PAH and BC with insight into source contributions. The plume layer of the two-layer model uses a previously published relationship between HEV plumes and ADTT reported by Tan et al. (Tan et al., 2014a) and a distance-weighted kernel algorithm to predict near-road contributions of HEVs. The background layer predicts the spatial variability of the non-plume background using LUR. We assess model transferability using a separate holdout dataset, and compare the performance of the two-layer model to the traditional LUR model.

2. Methods

2.1. Air pollution dataset

This paper analyzes data that were collected using the Carnegie Mellon University mobile laboratory, which is equipped with real-time instruments to measure black carbon (BC; Magee Scientific AE31 Aethalometer), air toxics (e.g., benzene and toluene), PB-PAH (EcoChem PAS2000), NOx, SO_2, O_3, and CH_4. The mobile monitoring campaign was conducted in two phases, and Table 1 summarizes all the data.

Phase I of this study and the mobile laboratory have been described in detail previously (Tan et al., 2014a). Briefly, the Phase I monitoring domain included the city of Pittsburgh and its immediate suburbs (Fig. S1). The monitoring was conducted during the 2011–2012 winter (Nov 2011–Feb 2012) and the 2012 summer (Jun 2012–Aug 2012). A total of 42 sites were selected using random sampling stratified by elevation (valley or upland) and traffic volume (high or low traffic). Eight sites were valley sites with low traffic, 11 sites were valley sites with high traffic, 13 sites were upland sites with low traffic, and 10 sites were upland sites with high traffic. Monitoring sites included different neighborhoods within the city, suburban sites, and locations near major pollution sources.

The mobile laboratory was driven along a prescribed driving route at each site. While some applications of mobile monitoring sampled specified intersections in a cloverleaf pattern, such as Larson et al. (Larson et al., 2009), the roadway network in Pittsburgh was not always conducive to this strategy. Instead, each sampling site is defined as the centroid of a driving route consisting of local major and minor roadways. Points along the driving route were within 250 m of the sampling site, and were within the same stratum (e.g., valley and low traffic). The mobile laboratory was typically driven ~5 mph below the posted speed limit (25 mph for most roads). We avoided high-speed highway driving, and avoided following specific vehicles, such as diesel trucks and buses, that could have high emissions that might skew estimates of site average concentrations. Mobile measurements were performed in three periods in both seasons to cover different times of day: