Science of the Total Environment 634 (2018) 77–86

Contents lists available at ScienceDirect

Science of the Total Environment

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

Development of land-use regression models for fine particles and black carbon in peri-urban South India

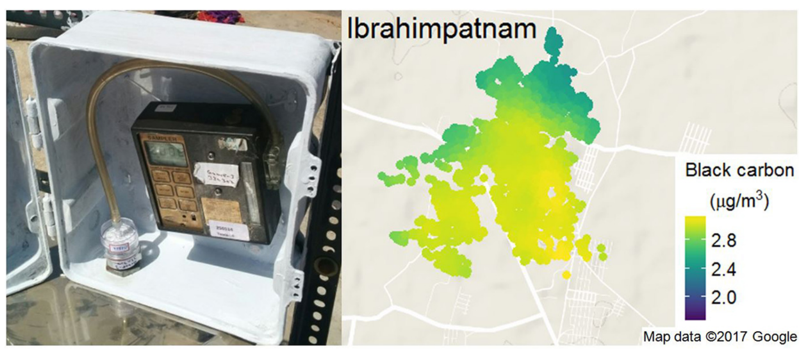
Margaux Sanchez [a,b,c,*], Albert Ambros [a,b,c], Carles Milà [a,b,c], Maëlle Salmon [a,b,c], Kalpana Balakrishnan [d], Sankar Sambandam [d], V. Sreekanth [e], Julian D. Marshall [e], Cathryn Tonne [a,b,c]

[a] Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain
[b] Universitat Pompeu Fabra (UPF), Barcelona, Spain
[c] CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain
[d] Department of Environmental Health Engineering, Sri Ramachandra University (SRU), Chennai, India
[e] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, United States

HIGHLIGHTS

- We developed LUR models for $PM_{2.5}$ and black carbon in peri-urban South India.
- We derived predictors from local built-environment survey and satellite imagery.
- $PM_{2.5}$ and black carbon models reached 58% and 79% of explained variability.
- Data from local built-environment survey were relevant for black carbon model.
- We observed more spatial variability than typical values in other study areas.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 7 December 2017
Received in revised form 21 March 2018
Accepted 24 March 2018
Available online xxxx

Editor: P. Kassomenos

Keywords:
Black carbon
Exposure assessment
India
Land-use regression
Particulate matter
Peri-urban area

ABSTRACT

Land-use regression (LUR) has been used to model local spatial variability of particulate matter in cities of high-income countries. Performance of LUR models is unknown in less urbanized areas of low-/middle-income countries (LMICs) experiencing complex sources of ambient air pollution and which typically have limited land use data. To address these concerns, we developed LUR models using satellite imagery (e.g., vegetation, urbanicity) and manually-collected data from a comprehensive built-environment survey (e.g., roads, industries, non-residential places) for a peri-urban area outside Hyderabad, India. As part of the CHAI (Cardiovascular Health effects of Air pollution in Telangana, India) project, concentrations of fine particulate matter ($PM_{2.5}$) and black carbon were measured over two seasons at 23 sites. Annual mean (sd) was 34.1 (3.2) μg/m^3 for $PM_{2.5}$ and 2.7 (0.5) μg/m^3 for black carbon. The LUR model for annual black carbon explained 78% of total variance and included both local-scale (energy supply places) and regional-scale (roads) predictors. Explained variance was 58% for annual $PM_{2.5}$ and the included predictors were only regional (urbanicity, vegetation). During leave-one-out cross-validation and cross-holdout validation, only the black carbon model showed consistent performance. The LUR model for black carbon explained a substantial proportion of the spatial variability that could not be captured by simpler interpolation technique (ordinary kriging). This is the first study to develop a LUR model for ambient concentrations of $PM_{2.5}$ and black carbon in a non-urban area of LMICs, supporting the applicability of the LUR approach in such settings. Our results provide insights on the added value of manually-collected built-environment data to improve the performance of LUR models in settings with limited data availability. For

* Corresponding author at: Barcelona Institute for Global Health (ISGlobal), Doctor Aiguader, 88, 08003 Barcelona, Spain.
E-mail address: margaux.sanchez@isglobal.org. (M. Sanchez).

https://doi.org/10.1016/j.scitotenv.2018.03.308

both pollutants, LUR models predicted substantial within-village variability, an important feature for future epidemiological studies.

1. Introduction

Air pollution is a leading risk factor for mortality and morbidity worldwide. The World Health Organization estimated that 3 million deaths were attributable to ambient air pollution in 2012; 87% of these occurred in low- and middle-income countries (LMICs) (World Health Organization, 2016). Still, high-income countries (HICs) remain the focus of much of the current literature investigating the health effects of air pollution (Tonne et al., 2017). Epidemiological evidence based on populations in HICs may not apply to LMIC populations because of differences in exposure ranges, air pollution sources, age distributions, and baseline health status. Thus, there is a critical need for population-level estimates of long-term exposure that can be used for epidemiological purposes in LMIC settings (Ma et al., 2017), where the majority of air pollution and pollution-related health effects occur.

Land-use regression (LUR) is a modeling method widely used in epidemiological studies to estimate particulate matter and black carbon concentrations at fine spatial scale in urban areas of North America and Europe (Eeftens et al., 2012, 2016; Hoek et al., 2008; Montagne et al., 2015; van Nunen et al., 2017; Weichenthal et al., 2016; Wolf et al., 2017; Zhang et al., 2014). There have been some attempts to apply the methodology in China and India but this literature is scarce and is limited to urban areas (Huang et al., 2017; Saraswat et al., 2013; Wu et al., 2015, 2017). LUR models have shown reasonable performance in urban areas where much of the locally-emitted particulate matter is generated by road traffic (Karagulian et al., 2015), but conditions in peri-urban or rural areas in LMICs, such as India, may differ (e.g. higher contribution of domestic fuel use and open burning) (Paliwal et al., 2016). Previous studies have demonstrated the limited transferability of LUR prediction models to areas other than those in which they were developed (Patton et al., 2015; Wang et al., 2014). In Bangalore, India, increasing concentrations of $PM_{2.5}$ were observed with increasing proximity to roads in a middle-income neighborhood, a spatial pattern similar to what would be observed in HIC (Both et al., 2011). However, this was not the pattern observed in a low-income neighborhood, which was attributed by the authors to solid fuel use (Both et al., 2011), highlighting the complexity of spatial patterns of air pollution in LMICs. It is unknown how well LUR prediction models perform in less urbanized area of LMICs that showed different contribution of particulate matter sources. Challenges to the development of LUR models in these settings include limited availability of geographic information systems (GIS) data, sources of particulate matter emissions not well correlated with existing land use data, and lack of routine monitoring data.

To address the need for air pollution exposure assessment in LMICs, we developed LUR models to predict spatial variation of $PM_{2.5}$ and black carbon in a peri-urban area of South India in which local emission sources include household solid fuel use, local industries, and motor vehicles. We here advance the science of LUR modeling by coupling land-use data derived from satellite imagery and data derived from a built-environment survey, a data collection approach not previously employed in LUR, allowing us to overcome the limited access to GIS data often found in LMICs. To help future exposure assessment, we evaluated the added value of LUR modeling predictions as compared to ordinary kriging, an interpolation technique that does not require any additional geographic data. We contribute to the exposure assessment literature by studying a peri-urban environment in a LMIC for which there is scarce literature regarding the spatial variability of ambient air pollution and no literature on LUR as an approach to predict such variability.

2. Methods

2.1. Study area

The study area consists of 28 rural and peri-urban villages in the Southeast of Hyderabad, covering 8.2 km^2 in a 22 km × 35 km (i.e., 770 km^2) region (Fig. 1). The area between villages was not of interest as it includes mostly crops and agricultural lands with no or few inhabitants. All villages were included in the existing APCAPS (Andhra Pradesh Children and Parents' Study) cohort and CHAI (Cardiovascular Health effects of Air pollution in Telangana, India) project (Kinra et al., 2014; Tonne et al., 2017). Villages vary in terms of surface, population size, socioeconomic status, urbanization, and primary cooking fuel.

2.2. Sampling campaign ($PM_{2.5}$ and black carbon)

We identified 23 households located in 16 different villages as fixed sites for sampling. We selected sites to maximize contrasts in several variables expected to correlate with particulate matter: distance to primary roads, distance to the city of Hyderabad, distance to industry, 500-meter buffer household density, and village-level solid fuel use (Fig. 1). This was done to avoid extrapolation of the LUR models to values of predictors for which we had no measurements. The households located in the 12 villages that were not directly monitored were similar to the households located in the other 16 villages in terms of distance to primary roads, distance to the city of Hyderabad, household density in a 50-meter buffer, and village-level proportion of solid fuel use.

We measured 24-hour integrated gravimetric $PM_{2.5}$ concentrations at these 23 locations for a total of 21 days in two sessions: 11 nonconsecutive days during post-monsoon season (Sep–Oct 2015) and 10 nonconsecutive days during summer season (Mar–Apr 2016). Sampling was done every other day. All sites were sampled the same days. Monitors included pumps (model 224-PCMTX8, SKC Ltd., Dorset, UK) that drew air through a cyclone separator (cut point: 2.5 μm) attached to a cassette containing 37-mm filter (Emfab filter, Pallflex®). Monitors were placed on the households' rooftops. Filters were left on site and collected daily at noon. Filters were weighed pre- and post-exposure using the TAPHE (Tamil Nadu Air Pollution and Health Effects) study protocol which follows RTI (Research Triangle Institute) guidelines (Balakrishnan et al., 2015). Daily $PM_{2.5}$ concentrations were derived from filter mass after correction for mass accumulated on blank filters (season-specific correction using median blank weight). Of the 483 sampled filters (21 days at 23 sites), 13 experienced device malfunction (running time < 75% of the expected sampling duration or pump airflow < 20% of the expected value) and 5 showed unexplained weighing errors for $PM_{2.5}$ (post-weight < pre-weight). We imputed all missing values (4% of the data) using a linear combination of date and $PM_{2.5}$ measurements from an ambient background monitoring site (see below). The overall performance of the imputation model was fair (adjusted-R^2 = 0.61) and predicted and measured values correlated positively ($R_{Spearman}$ = 0.74).

Daily black carbon concentrations were derived from optical attenuation (880 nm) of the mass collected on the filters, using a Magee OT21 Sootscan Optical Transmissometer (Magee Scientific, Berkeley, California, USA). Negative concentrations were obtained for 41 filters (8% of the sample). We assumed these were due to concentrations below the lower end of our standard curve and we imputed them with the seasonal 5th percentile concentration: 0.21 μg/m^3 in summer season and 1.31 μg/m^3 in post-monsoon season. Days with device malfunction or