Science of the Total Environment 639 (2018) 40–48

Contents lists available at ScienceDirect

Science of the Total Environment

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

Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters

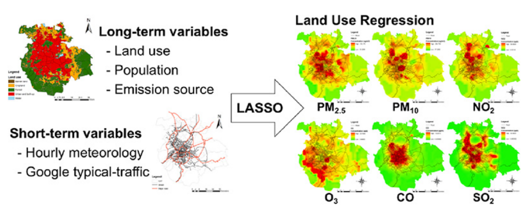
Yeongkwon Son [a,b,1], Álvaro R. Osornio-Vargas [c], Marie S. O'Neill [d], Perry Hystad [e], José L. Texcalac-Sangrador [f], Pamela Ohman-Strickland [g], Qingyu Meng [a,*], Stephan Schwander [a,b,*]

[a] Department of Environmental and Occupational Health, School of Public Health, Rutgers University, Piscataway, NJ, USA
[b] Office of Global Public Health Affairs, School of Public Health, Rutgers University, Piscataway, NJ, USA
[c] Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada
[d] Departments of Environmental Health Sciences and Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA
[e] College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA
[f] Environmental Health Department, National Institute of Public Health, Cuernavaca, Morelos, Mexico
[g] Department of Biostatistics, School of Public Health, Rutgers University, Piscataway, NJ, USA

HIGHLIGHTS

- Air pollution exposures in Mexico City result in adverse health outcomes.
- Land use regression models for six air pollutants were developed.
- Hourly meteorology and traffic data facilitated finer timescale simulation.
- A new regression method improved model performance.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 1 March 2018
Received in revised form 7 May 2018
Accepted 11 May 2018
Available online xxxx

Editor: P. Kassomenos

Keywords:
Land use regression
Air pollution
Exposure
LASSO
Mexico City

ABSTRACT

The Mexico City Metropolitan Area (MCMA) is one of the largest and most populated urban environments in the world and experiences high air pollution levels. To develop models that estimate pollutant concentrations at fine spatiotemporal scales and provide improved air pollution exposure assessments for health studies in Mexico City. We developed finer spatiotemporal land use regression (LUR) models for $PM_{2.5}$, PM_{10}, O_3, NO_2, CO and SO_2 using mixed effect models with the Least Absolute Shrinkage and Selection Operator (LASSO). Hourly traffic density was included as a temporal variable besides meteorological and holiday variables. Models of hourly, daily, monthly, 6-monthly and annual averages were developed and evaluated using traditional and novel indices. The developed spatiotemporal LUR models yielded predicted concentrations with good spatial and temporal agreements with measured pollutant levels except for the hourly $PM_{2.5}$, PM_{10} and SO_2. Most of the LUR models met performance goals based on the standardized indices. LUR models with temporal scales greater than one hour were successfully developed using mixed effect models with LASSO and showed superior model performance compared to earlier LUR models, especially for time scales of a day or longer. The newly developed LUR models will be further refined with ongoing Mexico City air pollution sampling campaigns to improve personal exposure assessments.

* Corresponding authors at: School of Public Health, Rutgers University, 683 Hoes Lane West, Piscataway, NJ08854, USA.
E-mail addresses: yeongkwon.son@dri.edu, (Y. Son), osornio@ualberta.ca, (Á.R. Osornio-Vargas), marieo@umich.edu, (M.S. O'Neill), Perry.Hystad@oregonstate.edu, (P. Hystad), jtexcalac@insp.mx, (J.L. Texcalac-Sangrador), ohmanpa@sph.rutgers.edu, (P. Ohman-Strickland), mengqi@sph.rutgers.edu, (Q. Meng), schwansk@sph.rutgers.edu (S. Schwander).
[1] Present address: Division of Atmospheric Sciences, Desert Research Institute, Reno, NV, USA

https://doi.org/10.1016/j.scitotenv.2018.05.144
0048-9697/© 2018 Published by Elsevier B.V.

1. Introduction

The Mexico City Metropolitan Area (MCMA) is one of the largest and most populated urban environments in the world and experiences high air pollution levels (Molina et al., 2010). A large body of published studies link air pollution exposure in the MCMA to various adverse health outcomes, including inflammation and/or DNA damage (Alfaro-Moreno et al., 2002; Osornio-Vargas et al., 2003), respiratory diseases including asthma (Rojas-Martinez et al., 2007), cardiovascular disease (Shields et al., 2013), suppression of innate antibacterial immunity (Rivas-Santiago et al., 2015) and overall mortality (Bell et al., 2008; O'Neill et al., 2004). Most of these health studies have relied on ambient air pollution data from fixed monitoring stations to estimate exposure. A major disadvantage of this methodological approach is that it may not capture spatial variability in exposure due to local sources, urban topography and local meteorological variables (Jerrett et al., 2005).

Several other methods, such as geostatistical methods and land use regression (LUR), have therefore been proposed to model and account for the within-city spatial distribution of air pollution concentrations in health studies (Brauer et al., 2003; Hoek et al., 2008; Rivera-González et al., 2015; Ryan and LeMasters, 2007). On one hand, geostatistical methods, including inverse distance weighting and ordinary Kriging, have been evaluated as exposure assessment methods in the MCMA (Rivera-González et al., 2015), but these methods cannot deal with local emissions (Jerrett et al., 2005). On the other hand, LUR analyses have not been used widely for air pollution research in the MCMA. Currently available LUR methods developed for the MCMA air pollution study have identified limitations associating with data availability and extrapolation from seasonal variables to annual simulations (Just et al., 2015; Sangrador et al., 2008).

The current effort to improve LUR models in the MCMA with fine spatial and temporal exposure analysis methods was made to advance an ongoing study of the relationship between ambient air pollution exposure and human host immune cell functions among participants in a Mexico City-based study known as "MexAir". Based on data from our lab (Rivas-Santiago et al., 2015; Sarkar et al., 2012), we hypothesize that exposure to poor air quality in the MCMA may increase susceptibility to infection with *Mycobacterium tuberculosis* (*M.tb*) and risk of reactivation tuberculosis (TB).

In order to develop a LUR model with fine spatial and temporal scale, two key variables need to be considered. First, traffic data have been identified as an important predictor to improve spatial variability (Beelen et al., 2013; Hoek et al., 2001). However, there are no publicly obtainable traffic data in most cities (Hoek et al., 2008). Indeed, traffic information were not included in the previous MCMA LUR study due to the data availability (Just et al., 2015). Second, meteorological variables might improve the performance of LUR models with short temporal resolution (Ainslie et al., 2008; Just et al., 2015; Liu et al., 2015). However, most LUR models performed to date outside of the MCMA, have focused on long-term air pollutant exposures, and assessments of annual, seasonal and monthly averages of $PM_{2.5}$ (Henderson et al., 2007; Johnson et al., 2013), PM_{10} (Hart et al., 2009; Liu et al., 2015), NO_2 (Arain et al., 2007; Beelen et al., 2013; Dons et al., 2014; Hart et al., 2009; Henderson et al., 2007; Johnson et al., 2013; Liu et al., 2015; Wang et al., 2013; Wheeler et al., 2008), and SO_2 (Wheeler et al., 2008). A limited number of earlier LUR models that predicted shorter-term pollutant levels used temporal calibration approaches (e.g., dummy variables) (Dons et al., 2013; Johnson et al., 2013) but did not consider time-varying meteorological covariates that are useful in the optimization and refinement of LUR models (Ainslie et al., 2008; Liu et al., 2015).

The aim of the current study was to introduce 'hourly meteorological variables' and an 'hourly traffic density variable' into the LUR model to allow modeling of air quality on a finer temporal resolution scale and integrate the temporal variations found throughout the three major weather seasons in the MCMA: wet, cold-dry, warm-dry (de Foy et al., 2006; Manzano-León et al., 2016).

2. Material and methods

2.1. Study areas and polygonal regions

The MCMA includes the states of Mexico and Hidalgo, and the Federal District (Distrito Federal). The current study area (Regions I, II, III, and IV) covered a central part of the MCMA with a total area of 4238 km^2 (Fig. 1). LUR models were developed to cover the whole MCMA study area with different time scales, and then evaluated using whole and separate air quality monitoring data based on polygonal regions. Thiessen polygons were generated using the locations of the 28 automatic air quality monitoring stations of the Red Automática de Monitoreo Atmosférico (RAMA) in the MCMA. Thiessen polygons were then combined into four regions based on air quality distribution, pollutant emission sources and urban population density. The four regions are defined as: Region I (central MCMA region with large population density; the MexAir study municipalities Iztapalapa and Iztacalco are included in this zone), Region II (north western MCMA region with medium population density and high air pollution levels), Region III (north eastern MCMA region with medium air pollution levels) and IV (southern MCMA region with highest O_3 levels and low levels of other air pollutants).

2.2. Data collection

Data sets for regression analyses are described in the Supplementary material (Table S1). In brief, ambient air pollution data from 2011 to 2014 was downloaded from the RAMA (SIMAT, 2018). The air quality monitoring stations operate beta-attenuation monitors, UV photometric ambient ozone analyzers, chemiluminescence NO-NO_2-NO_x analyzers, UV Fluorescence Sulfur Dioxide analyzers, and Infrared CO analyzers to measure particulate matter ($PM_{2.5}$, PM_{10}), O_3, NO_2, SO_2, and CO, respectively. Outliers (i.e., any pollutant values lower than 5% and higher than 95%) were not excluded in our study as the pollutant data sets had been cleaned and verified by the Mexican government and were therefore considered accurate.

Hourly meteorological data for the same time period was obtained from the meteorology and solar radiation monitoring network (Red de Meteorología y Radiación Solar; REDMET) and the atmospheric deposition monitoring network (Red de Depósito Atmosférico, REDDA), which is operated by the Mexican Ministry of Environment (Secretaría del Medio Ambiente y Recursos Naturales; SEMARNAT) (SIMAT, 2018). Data from a total of 28 RAMA stations, 21 REDMET stations and 16 REDDA stations covering the entire Federal District area and part of Hidalgo and Mexico states are included in the current LUR study (Tables S2 and S3).

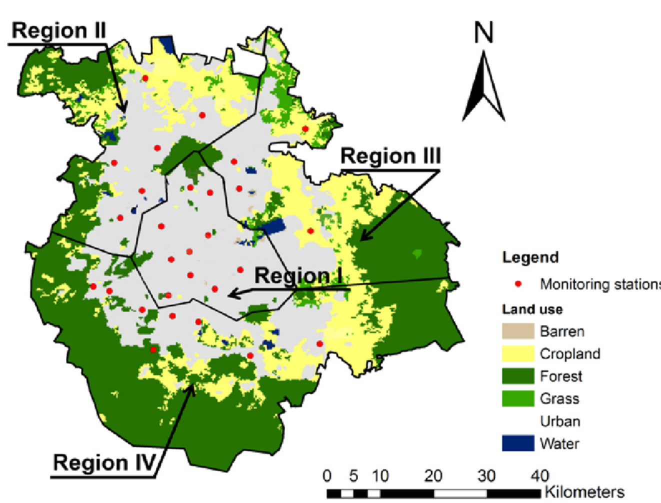


Fig. 1. The study area and polygonal regions.