Science of the Total Environment 488–489 (2014) 343–353

Contents lists available at ScienceDirect

Science of the Total Environment

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

Land use regression models to estimate the annual and seasonal spatial variability of sulfur dioxide and particulate matter in Tehran, Iran

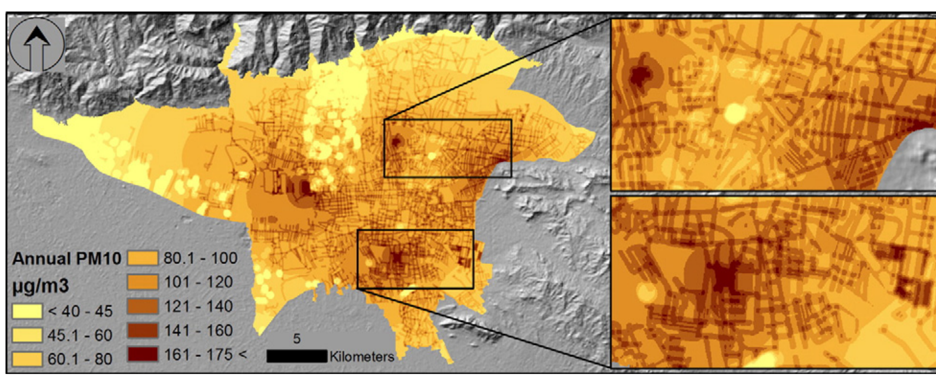
Hassan Amini [a,b,*], Seyed Mahmood Taghavi-Shahri [c,d,**], Sarah B. Henderson [e,f,***], Kazem Naddafi [b,g], Ramin Nabizadeh [b], Masud Yunesian [b,g,****]

[a] Kurdistan Environmental Health Research Center, Kurdistan University of Medical Sciences, Sanandaj, Iran
[b] Department of Environmental Health Engineering, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
[c] Research Center for Environmental Pollutants, Qom University of Medical Sciences, Qom, Iran
[d] Department of Epidemiology and Biostatistics, School of Public Health, Isfahan University of Medical Sciences, Isfahan, Iran
[e] Environmental Health Services, British Columbia Centre for Disease Control, 655 West 12th Avenue, Vancouver, BC V5Z 4R4, Canada
[f] School of Population and Public Health, The University of British Columbia, 2206 East Mall, Vancouver, BC V5T 1Z3, Canada
[g] Center for Air Pollution Research (CAPR), Institute for Environmental Research (IER), Tehran University of Medical Sciences, Tehran, Iran

HIGHLIGHTS

- Application of LUR in novel context of Tehran, Iran, where air quality is poor
- A novel variable selection method for LUR model building
- Several new predictive variables and variable types were developed and explored.
- LUR models were developed for annual, cooler and warmer seasons for SO_2 and PM_{10}.
- The entire population lived in areas where the WHO Air Quality Guidelines were exceeded.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 25 November 2013
Received in revised form 5 April 2014
Accepted 24 April 2014
Available online 16 May 2014

Editor: Lidia Morawska

ABSTRACT

The Middle Eastern city of Tehran, Iran has poor air quality compared with cities of similar size in Europe and North America. Spatial annual and seasonal patterns of SO_2 and PM_{10} concentrations were estimated using land use regression (LUR) methods applied to data from 21 air quality monitoring stations. A systematic algorithm for LUR model building was developed to select variables based on (1) consistency with *a priori* assumptions about the assumed directions of the effects, (2) a *p*-value of <0.1 for each predictor, (3) improvements to the leave-one-out cross-validation (LOOCV) R^2, (4) a multicollinearity index called the variance inflation factor, and (5) a grouped (leave-25%-out) cross-validation (GCV) for final model. In addition, several new predictive

Abbreviations: AQCC, Air Quality Control Company; DOE, Department of Environment; GIS, Geographic Information System; GCV, grouped (leave-25%-out) cross-validation; LOOCV, leave-one-out cross-validation; LUR, land use regression; MAR, missing at random; PM_{10}, particulate matter with an aerodynamic diameter of 10 µm or less; PPVs, potentially predictive variables; RMSE, root mean square error; SO_2, sulfur dioxide; VIF, variance inflation factor; WHO, World Health Organization.

* Correspondence to: H. Amini, Kurdistan Environmental Health Research Center, Kurdistan University of Medical Sciences, Sanandaj, Iran.
** Correspondence to: S.M. Taghavi-Shahri, Research Center for Environmental Pollutants, Qom University of Medical Sciences, Qom, Iran.
*** Correspondence to: S.B. Henderson, Environmental Health Services, British Columbia Centre for Disease Control, 655 West 12th Avenue, Vancouver, BC V5Z 4R4, Canada.
**** Correspondence to: M. Yunesian, Center for Air Pollution Research (CAPR), Institute for Environmental Research (IER), Tehran University of Medical Sciences, Tehran PO Box: 14155-6446, Iran. Tel.: +98 21 8898 7379; fax: +98 21 8163 3623.
E-mail address: yunesian@tums.ac.ir (M. Yunesian).

http://dx.doi.org/10.1016/j.scitotenv.2014.04.106

Keywords:
Air pollution exposure modeling
Geographic Information Systems (GIS)
Land use regression (LUR)
Particulate matter
Sulfur dioxide
Tehran

variables and variable types were explored. The annual mean concentrations of SO_2 and PM_{10} across the stations were 38 ppb and 100.8 μg/m^3, respectively. The R^2 values ranged from 0.69 to 0.84 for SO_2 models and from 0.62 to 0.67 for PM_{10} models. The LOOCV and GCV R^2 values ranged, respectively, from 0.40 to 0.56 and 0.40 to 0.50 for the SO_2 models; they were 0.48 to 0.57 and 0.50 to 0.55, respectively, for the PM_{10} models. There were clear differences between the SO_2 and PM_{10} models, but the warmer and cooler season models were consistent with the annual models for both pollutants. Although there was limited similarity between the SO_2 and PM_{10} predictive variables, measures of street density and proximity to airport or air cargo facilities were consistent across both pollutants. In 2010, the entire population of Tehran lived in areas where the World Health Organization guidelines for 24-hour mean SO_2 (7 ppb) and annual average PM_{10} (20 μg/m^3) were exceeded.

1. Introduction

Foundational research in air pollution epidemiology demonstrated that cities with more air pollution have higher morbidity and mortality rates than those with less air pollution (Krewski et al., 2000; Miller et al., 2007). More recent work has demonstrated that more polluted areas within cities have higher morbidity and mortality rates than less polluted areas (Hankey et al., 2012; Schwela, 2011). Such studies require methods to assess within-city variability in exposure to different air pollutants. In small studies it is possible to make measurements for individual subjects, but population-based studies require methods that can be simply and objectively applied to thousands, hundreds of thousands, or even millions of subjects (Hoek et al., 2008). Many different approaches have been used. In areas with air quality monitoring networks, it has been assumed that pollutant concentrations from the nearest measurement station are representative of exposure in the area (Jerrett et al., 2004). However, these stations are generally situated to capture overall background concentrations (Kanaroglou et al., 2005), and not the very small-scale variability reported by several studies (Henderson et al., 2007).

Land use regression (LUR) is a geospatial technique that has been used to model within-city spatial and spatiotemporal variability for multiple air pollutants. There is no standard way of conducting LUR, but detailed descriptions of different approaches are published elsewhere (Hoek et al., 2008). In brief, a pollutant is measured at multiple sites around a city. Under ideal circumstances, the sites are specifically selected to optimize the spatial variability in pollutant concentrations, but some studies have leveraged data from pre-existing monitoring networks (Gulliver et al., 2011a; Kashima et al., 2009; Stedman et al., 1997). Physical and geographic characteristics that might be associated with the pollutant concentrations are then measured around each site using a Geographic Information System (GIS). These potentially predictive variables typically describe site location, surrounding land use, population density, point sources, and traffic patterns. Multiple linear regression is used to correlate the measurements with the most predictive variables, and the resulting equation can be used to estimate pollutant concentrations anywhere that all of the predictors can be measured.

The LUR approach was first developed in Europe (Briggs et al., 1997) and was first widely applied in both Europe and North America (Hoek et al., 2008). More recently, LUR has been extended to cities in Asia (Allen et al., 2011; Chen et al., 2010a,b; Kashima et al., 2009; Saraswat et al., 2013), South America (Habermann and Gouveia, 2012; Sangrador et al., 2008), and Africa (Dionisio et al., 2010). Here we describe the application of LUR in a large Middle Eastern city. Tehran, the capital of Iran, experiences high air pollution concentrations due mainly to motorized traffic and industrial sources. The traffic consists of cars, motorbikes, trucks, and buses. There are four regional bus terminals in the city, namely West, South, East, and Park-Savar Beyhaghi (Figure S1 in the supplemental information). Approximately 60,000 (<1%) and 185,000 (2.2%) persons live within 500 and 1000 m of these four bus terminals, respectively. There is also a wide variety of heavy and light industries across the city, including dairy production and manufacturing of detergents, inks, metals, machinery, paving materials, roofing materials, plastics, and pharmaceuticals (Figure S2 in the supplemental information). Because Iran has large natural gas reservoirs, the majority of people use natural gas for cooking and heating (Amini et al., 2014). Air pollutant concentrations in Tehran exceed those reported for Europe and North America, where most LUR models have been developed (Vlachogianni et al., 2011; Wheeler et al., 2008; Yanosky et al., 2008). Although there have been a few studies on the acute health effects of air pollution in Tehran (Kheirandish-Gozal et al., 2014; Qorbani et al., 2012), studies on the chronic health effects have been limited. Results of work reported here will provide the foundation for the next generation of air pollution epidemiology and science based policy making in Tehran, and Iran in general. Furthermore, we add to the LUR literature by presenting a novel variable selection method for model building, by comparing models for the warmer and cooler seasons, and by exploring several new potentially predictive variables and variable types.

2. Materials and methods

2.1. Study area

Tehran is located at the foot of the Alborz Mountains, south of the Caspian Sea (Fig. 1). The annual mean daily temperature is 18.5 °C, with highs around 40 °C in July and lows around −10 °C in January. The average annual precipitation is 150 millimeters (mm), with the maximum in January (40 mm) and the minimum in September (0 mm). The weather is typically sunny, with an annual average of 2800 h of bright sunshine and a mean cloud cover of 30%. The prevailing winds blow from west and north (Figure S3, supplemental information). The average elevation is 1200 m above sea level, and the city covers a large area. Tehran is the most populous city in Iran, with 8.2 million urban residents, and a daytime population of more than 10 million people due to diurnal migration from the outlying areas. The estimation domain for our LUR models was 613 km^2 (Amini et al., 2013; Leili et al., 2008).

2.2. Air pollution data

Hourly SO_2 and PM_{10} concentrations for the 2010 calendar year were obtained from 23 air quality monitoring stations administered by two government agencies (Fig. 1). Of the stations, 16 belonged to the Air Quality Control Company (AQCC), and 7 belonged to the Department of Environment (DOE). Both the AQCC and DOE monitoring stations used UV fluorescence analyzers (model AF22M of Environment SA, France) to measure SO_2, and beta-radiation attenuation instruments or beta-gauge monitors (model MP 101M of Environment SA, France; FH 62 IN, FAG Kugelfischer, Germany, and APDA-351E of Horiba, Japan) to measure PM_{10}.

Complete datasets would have contained 8760 measured values (24 h/day × 365 days in 2010), but 33% and 43% of the SO_2 and PM_{10} values were missing, respectively (Figure S4, supplemental information). The Amelia program was used for imputation of the missing data (Honaker et al., 2011). The program uses a new expectation–maximization algorithm with bootstrapping to impute missing values and return a complete dataset. We provided the program with all