Atmospheric Environment 79 (2013) 421–427

Contents lists available at SciVerse ScienceDirect

Atmospheric Environment

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

Estimation of sulfur dioxide air pollution concentrations with a spatial autoregressive model

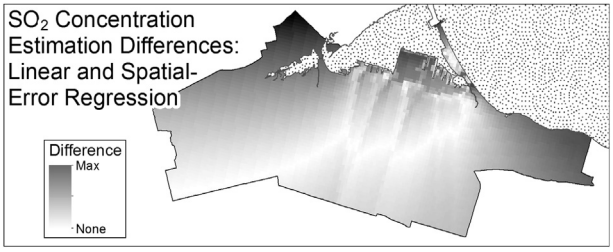
Pavlos S. Kanaroglou [a], Matthew D. Adams [a,*], Patrick F. De Luca [a], Denis Corr [b,c], Nazmul Sohel [a]

[a] School of Geography & Earth Sciences, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada
[b] Corr Research[1], Hamilton, ON, Canada
[c] Rotek Environmental Inc., 43 Keefer Court, Unit 104, Hamilton, ON, Canada

HIGHLIGHTS

- Few land use regression models have evaluated spatial effects prompting this study.
- Application of OLS regression violated the independence of errors assumption.
- Spatial error model removed spatial autocorrelation of residuals.
- Locations' direction to source and wind direction alignment is a strong predictor.
- Mobile monitoring data are applicable for land use regression.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 7 February 2013
Received in revised form
3 July 2013
Accepted 5 July 2013

Keywords:
Air pollution
Sulfur dioxide
Mobile monitoring
Land use regression
Spatial error model
Hamilton
Ontario
Canada

ABSTRACT

In this paper, we develop a land-use regression model for sulfur dioxide air pollution concentrations. We make use of mobile monitoring data collected in Hamilton, Ontario, Canada, between 2005 and 2010 inclusive. The observed SO_2 concentrations are regressed against a comprehensive set of land use and transportation variables. Land use and transportation variables are assessed as the amount of each characteristic within buffers of 50, 100, 200, 400, 800, and 1600 m around pollution observation locations. In the first instance of regression modeling, we apply ordinary least-squares regression. The OLS model R^2 for training data was 0.38, and an R^2 of 0.3 for a 50% held out cross-validation data set. The residuals are spatially correlated with the OLS model as determined with Moran's I. We thus applied a simultaneous autoregressive model, specifically the spatial error model. The resulting model slightly improved fit as determined by a pseudo $R^2 = 0.4$, improved log-likelihood, and reduced MSE, RMSE, and MAE. The spatial error model residuals were not spatially auto-correlated, resulting in a valid model. SAR modeling is a natural extension to OLS regression models and solves the issue of spatial autocorrelation in model residuals with a one-stage model.

1. Introduction

Natural and anthropogenic activities – examples include volcanic eruptions and industrial emissions – often result in sulfur dioxide concentrations exceeding the World Health Organization guideline of 20 μg m^{-3} [7.6 ppbv @ 25 C°] (24-h mean) (WHO,

* Corresponding author. Tel.: +1 905 525 9140x26949; fax: +1 905 546 0463.
E-mail address: adamsmd@mcmaster.ca (M.D. Adams).
[1] www.corr-research.com.

1352-2310/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.atmosenv.2013.07.014

2006). Epidemiological studies reveal that exposure to elevated SO_2 concentrations can lead to adverse health effects including: all cause respiratory and cardiovascular mortality (Medley et al., 2002), risks for non-accidental mortality (Park et al., 2010), and cardiac hospitalization (65 years or older) (Fung et al., 2005). For a review of air pollution health effects including SO_2, see Bascom (1996). Therefore, accurate and reliable estimates are crucial for identifying the populations at risk of elevated exposure. However, traditional fixed-location monitoring stations are often few in number and their use is inappropriate for assigning individual level air pollution exposure, due to spatial variation of SO_2 at the intra-urban scale (Adams et al., 2012; Chattopadhyay et al., 2010).

Land use regression (LUR) modeling is a powerful tool for assigning the spatial distribution of air pollutant concentrations. LUR utilizes measurements of air pollutant concentrations monitored throughout the study area, along with geographic variables representing the land use types surrounding the locations where the air pollution concentrations were obtained. The underlying theory is that surrounding land use types play a major role in generating activities that act as emission sources, and they can be used to model the concentrations. The goal is to develop a model capable of predicting pollutant concentrations at unmonitored sites (Jerrett et al., 2005; Hoek et al., 2008; Adamkiewicz et al., 2010; Allen et al., 2011). Once the model is developed it can be applied to estimate air pollution concentrations at locations within the study area. Thus far, LUR models have been developed for a number of environmental pollutants including: NO, NO_2, particulate matter, soot, VOCs, elemental carbon, and SO_2, see Hoek et al. (2008) for a review. Although some research has mentioned or considered LUR model developments for SO_2, the literature concerning this pollutant is rather limited (Atari et al., 2008; Wheeler et al., 2008; Gulliver et al., 2011).

Often the evaluation of LUR model residuals for spatial autocorrelation is not undertaken, though this spatial correlation would violate the independence errors assumption for these models. Few studies have applied methods to deal with the spatial correlation of residuals. Recently, Li et al. (2012) applied a two-stage approach, first a generalized additive model followed with cokriging of the spatial residuals. They found this approach produced better-fit models compared to universal kriging, multiple linear LUR, and GAM with and without a spatial smoothing term. Spatial autocorrelation accounted for about 20% of the variance. Mercer et al. (2011) utilized two approaches both involving kriging to handle the spatial structure in the model residuals, both performing better than non-spatial LUR models. The use of kriging to handle the spatial autocorrelation in the model residuals is one approach; however it does not directly handle the spatial process in a one-stage model, which we are interested in.

This study aims to present the use of a spatial autoregressive model (spatial error model) to produce valid models when spatial structure occurs in the error terms of ordinary least squares regression models. We apply this to sulfur dioxide concentrations measured by mobile monitoring to model the long-term average concentration between 2005 and 2010 for Hamilton, Ontario, Canada.

2. Methods

2.1. Study area

Hamilton, Ontario, Canada is situated at the western tip of Lake Ontario (43.3°N, 79.9°W) with a population of 520 000 (Statistics Canada, 2012). Hamilton features an upper and lower city separated by the Niagara Escarpment (~90 m). The selection of Hamilton is appropriate for this study, as SO_2 is a pollutant of concern, with concentrations of considerable spatial variability over the city (Adams et al., 2012). Air pollution concerns have led to multiple air pollution studies being conducted in Hamilton (Adams et al., 2012; Wallace et al., 2010, 2009; Sahsuvaroglu et al., 2006; Jerrett et al., 2001; Finkelstein et al., 2004, 2003; Buzzelli et al., 2003). SO_2 sources are mainly located in the harbour/industrial region of the city. The government run air quality monitoring network has three monitors located directly within Hamilton, and one more slightly beyond Hamilton's north-west boundary. These monitors cover a small spatial extent, and do not capture the entire spatial complexity of the intra-urban air pollution (Adams et al., 2012).

2.2. Air pollution data

Mobile monitoring techniques collected SO_2 concentration data with a modified van outfitted with air pollution monitoring equipment and roof mounted air intakes (3 m above ground). Collection campaigns occurred in non-peak traffic hours (10:00 AM–4:00 PM inclusive) across 62 days between 2005 and 2010. We analyzed the SO_2 concentrations at a centrally located stationary monitor in our study site, operated by the provincial government. We averaged the concentrations for each hour over the entire study period and we observed that the surveyed hours (10:00 AM–4:00 PM) include the peak concentration times. City-wide coverage scans occurred during surveys under different meteorological conditions. A Monitor Labs™ 8850 SO_2 analyzer (range 0–100 ppm) analyzed SO_2 concentrations, calibrated with an ESA Model VE-3M SO_2 calibrator. The monitor has a minimum detection limit of 1 ppb, and a precision of 5 ppb. SO_2 concentrations were log-normal distributed, thus the log-transformed concentration was used in the analysis. Data reduction was applied because the data-logger's time resolution of 1 s resulted in many redundant observations. The monitoring system uses a two-minute rolling average of air intake. Detailed methods of data collection are found in Wallace et al. (2009). Mobile monitoring data locations are shown in Fig. 1. We selected one observation every one minute to be included in the analysis. The mobile monitoring locations are available by day in a kml file in Appendix A.

Mobile monitoring data were adjusted to minimize temporal variability at each site, using data from a centrally located stationary monitoring site operated by the Ontario Ministry of the Environment. The temporal variability reduction (TVR) adjustment approach followed: $MS_t = MO_t^*(\log(SL + e)/\log(SD_t + e))$, where MS_t = Standardized mobile value at time t; MO_{nt} = Original mobile value at time t; SL = Stationary Monitor long-term mean (6-Year); SD_t = Hourly value from the stationary monitor that time t falls in. This is similar to an approach applied in Hoek et al. (2002). This approach is designed to minimize the variability in concentrations that is due to temporal variability. We examined the central monitor and averaged by year and month the concentrations it obtained. No year to year trend occurred in the data. Monthly, the average was similar across all months, with a range of 1.3 ppb excluding April, May and June, having elevated concentrations above the mean of the other months (3.1, 4.1, and 5.6 ppb above respectively). Thus, TVR adjustment is necessary to remove the temporal variability in the data set.

Mobile data following the derivation of all independent variables and temporal adjustment were averaged within 50 m grid cells, this was done because many of the points were collocated. We chose to average within 50 m based on the accuracy that can be obtained from the GPS unit during collection. We feel the choice is appropriate as it is smaller, or in most cases much smaller, than the minimum buffer sizes used in other LUR studies of SO_2 (Atari et al., 2008; Wheeler et al., 2008; Gulliver et al., 2011).