Transportation Research Part D 43 (2016) 181–191

Contents lists available at ScienceDirect

Transportation Research Part D

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

A hybrid land use regression/line-source dispersion model for predicting intra-urban NO_2

Drew R. Michanowicz [a,*], Jessie L.C. Shmool [a], Leah Cambal [a], Brett J. Tunno [a], Sara Gillooly [a], Megan J. Olson Hunt [b], Sheila Tripathy [a], Kyra Naumoff Shields [c], Jane E. Clougherty [a]

[a] University of Pittsburgh Graduate School of Public Health, Department of Environmental and Occupational Health, Pittsburgh, PA, United States
[b] University of Wisconsin–Green Bay, Department of Natural and Applied Sciences (Statistics), Green Bay, WI, United States
[c] Colorado State University, Center for Energy Development and Global Health, Fort Collins, CO, United States

ARTICLE INFO

Article history:

Keywords:
Air pollution
Land use regression
Nitrogen dioxide
Dispersion
Caline3
Traffic
Near-road exposure

ABSTRACT

Land use regression (LUR) has evolved as a standard tool for estimating intra-urban variation in long-term air pollution exposures. LUR, however, generally relies on observed spatial relationships between distributed sources and pollution measures, rather than meteorological or physio-chemical information. Incorporating traffic-meteorological interaction information into LURs via dispersion models may improve accuracy and physical interpretability. To examine whether deterministic dispersion output improves LUR predictions for nitrogen dioxide (NO_2), we incorporated hourly Caline3QHCR dispersion information into existing winter-time regression models originally designed to disentangle effects of multiple sources (e.g., legacy industry, vehicle traffic) and concentration modifiers (e.g., elevation) across Pittsburgh, PA. Caline3QHCR output improved overall cross-validated R^2 values by 0.10, 0.03, and 0.05 for the weekday, full-week, and merged years models, respectively. The addition of Caline3 output as an independent covariate effectively displaced two road-specific predictors, one built environment predictor and one meteorological predictor across two winter-season models. The incorporation of study-specific dispersion principles may corroborate overall LUR model interpretability and could be applied similarly to preexisting LUR models that lack strong source-concentration relationships. Model improvements were observed at sites with relatively higher NO_2 concentrations that exhibited higher traffic volumes and were in closer proximity to roadways (< 300 m), which may have an important bearing towards better characterizing exposures in both near-roadway and high-traffic environments.

Introduction

Land use regression (LUR) has emerged as a standard tool for intra-urban air pollution exposure assessment in recent years (Brauer et al., 2003; Briggs et al., 1997; Clougherty et al., 2013; Jerrett et al., 2005). LUR, however, offers limited capability to incorporate source–meteorology interaction information, thereby producing estimates based on empirical relationships, rather than a theoretical–physical basis (Jerrett et al., 2005; Su et al., 2008; Wilton et al., 2010). Thus, there is now growing interest in incorporating study-specific air dispersion information into LUR, in the hopes of improving accuracy,

* Corresponding author at: The Landmark Center, 401 Park Drive, Room 415-M West, P.O. Box 15677, Boston, MA 02215, United States. Tel.: +1 814 934 5624.
E-mail address: Arm73@pitt.edu (D.R. Michanowicz).

http://dx.doi.org/10.1016/j.trd.2015.12.007
1361-9209/

calibration, and interpretability of such models (Gulliver and Briggs, 2011; Lindström et al., 2013; Mölter et al., 2010; Wilton et al., 2010; Zwack et al., 2011).

LUR quantifies statistical relationships between measured pollution concentrations and emission source indicators to estimate concentrations at non-sampled locations (Hoek et al., 2008). Significant traffic-source indicators have included total length of roadway (Henderson et al., 2007), distance from nearest roadway (Gilbert et al., 2005) and traffic count density (Ross et al., 2006) within various radial buffer distances. The statistical relationships derived from these metrics in LUR are based on observed values and statistical principles, and generally fail to account for short-term interactions between sources and atmospheric conditions (Wilton et al., 2010). Moreover, traffic-related pollution can lead to complex spatiotemporal patterns in air pollution, necessitating dedicated near-roadway sampling (Gulliver and Briggs, 2011; Mölter et al., 2010) beyond the data obtained from fixed-site monitors (Jerrett et al., 2005), and refined spatial analysis.

To further refine small-scale (e.g., intra-urban) spatial concentration gradients, techniques to combine spatially-scalable models to better capture near-source variability have been employed (e.g., localized traffic demand modeling for emissions factor estimation) (Cook et al., 2008; Isakov et al., 2007; Kinnee et al., 2004). Temporal variance has been included in similar predictive models by including meteorological covariates (e.g., mean wind speed or direction) (Arain et al., 2007; Clougherty et al., 2009; Jerrett et al., 2007; Su et al., 2008), or by weighting source-concentration relationships by predominant wind direction (Clougherty et al., 2009; Mavko et al., 2008; Van den Hooven et al., 2012). Ainslie et al. (2008) and Su et al. (2008) attempted to capture atmospheric dispersion using a source-area concentration grid of distributed emissions under varying atmospheric conditions. Likewise, Wilton (2011) incorporated meteorologically-varying covariates as volume sources in a CALPUFF Lagrangian puff model (Scire et al., 1990). Zwack et al. (2011) jointly utilized regression analyses and the Quick Urban & Industrial Complex (QUIC) dispersion model to estimate source characterization and emissions factors of ultrafine particulates from a major roadway in Brooklyn, NY. Some authors have implemented hybrid LUR-dispersion models, finding marginal improvements in overall model fit (Lindström et al., 2013; Wilton et al., 2010). It has been less explored, however, whether similar model performance differs as a function of source characteristics (e.g., traffic volume and roadway proximity) across sites and time.

Ideally, estimation of ground-level air pollutant concentrations should include emissions characteristics, meteorologically-related dispersion, transformation and removal processes (Bekhor and Broday, 2013), along with a means of model validation (Chang and Hanna, 2004). Of the many types of models employed, Gaussian-type plume dispersion models are the most widely developed and utilized regulatory atmospheric dispersion models (Ristic et al., 2014). Gaussian dispersion models have been employed extensively in regulatory air quality management, and to a lesser degree in human exposure assessments (Jerrett et al., 2005; Johnson et al., 2010; Marshall et al., 2008; Mölter et al., 2010; Nafstad et al., 2003; Nyberg et al., 2000; Van den Hooven et al., 2012). Gaussian dispersion models may be used to simulate transport of pollutants deterministically, as a function of source characteristics (e.g., location, strength, size) and temporally-varying meteorological conditions (e.g., wind speed, direction, atmospheric stability) (Briant et al., 2013; Chang and Hanna, 2004). Therefore, standard LUR could be corroborated by incorporating physically-based estimates derived from study-specific source–meteorology interaction information via dispersion model output (Jerrett et al., 2005; Su et al., 2008; Wilton et al., 2010).

In this study, we aimed to improve winter-time prediction of ground-level NO_2 across Pittsburgh, PA, USA, by incorporating the Caline3QHCR line-source (roadway) dispersion model (Benson, 1992; Eckhoff and Braverman, 1995) output as an independent covariate into pre-constructed LUR models. Our multi-pollutant spatial saturation study was designed to disentangle impacts of multiple pollution sources (e.g., legacy industry, vehicle traffic), and to assess potential modifiers of source-concentration relationships (e.g., elevation) across an urban-to-suburban landscape (Shmool et al., 2014). We utilized two successive years of winter-season-only NO_2 measurements. We evaluated improvements in model fit by adding Caline3 predictions as an additional term to three pre-constructed LUR models, and observing changes in regression coefficients and covariate significance. Specifically, we tested: (1) Caline's effectiveness given diurnal traffic variability in a weekday-only (year 1) vs. full-week (year 2) LUR model; (2) whether Caline's improvements in fitting accuracy differed across sampling intervals by including modeled predictions in a combined-years LUR model (year 1 + year 2) and (3) Caline's effect on LUR predictions as a function of traffic density and distance from roadway, in order to assess whether hybrid LUR-dispersion may better explain near-roadway variability, relative to LUR alone.

Methods

NO_2 measurements for Pittsburgh

NO_2 measurements were sampled across two successive winter seasons from early January through late March of 2012 and 2013. Year 1 was comprised of six successive 5-day (Monday through Friday) sampling sessions, which are hereafter referred to as the *weekday measures* and *weekday model*. Year 2 was comprised of six successive 7-day (Monday through Sunday) sampling sessions, referred to hereafter as the *full-week measures* and *full-week model*. We employed a spatial saturation design to characterize intra-urban variability in multiple air pollutants (e.g., $PM_{2.5}$, NO_2, O_3, SO_2) across the greater Pittsburgh, PA metropolitan area, systematically allocating sampling sites across complex topography and emission source regimes, as detailed in Shmool et al. (2014) and Tunno et al. (2015).