Atmospheric Environment 171 (2017) 1–16

Contents lists available at ScienceDirect

Atmospheric Environment

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

Review article

A systematic review of land use regression models for volatile organic compounds

Heresh Amini[a,b,*], Masud Yunesian[c], Vahid Hosseini[d], Christian Schindler[a,b], Sarah B. Henderson[e,f], Nino Künzli[a,b]

[a] *Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland*
[b] *University of Basel, Basel, Switzerland*
[c] *Center for Air Pollution Research (CAPR), Institute for Environmental Research (IER), Tehran University of Medical Sciences, Tehran, Iran*
[d] *Mechanical Engineering Department, Sharif University of Technology, Tehran, Iran*
[e] *Environmental Health Services, British Columbia Centre for Disease Control, Vancouver, Canada*
[f] *School of Population and Public Health, University of British Columbia, Vancouver, Canada*

GRAPHICAL ABSTRACT

ARTICLE INFO

Keywords:
Aliphatic compounds
Aromatic alkylbenzenes
Benzene
LUR models
VOC

ABSTRACT

Various aspects of land use regression (LUR) models for volatile organic compounds (VOCs) were systematically reviewed. Sixteen studies were identified published between 2002 and 2017. Of these, six were conducted in Canada, five in the USA, two in Spain, and one each in Germany, Italy, and Iran. They were developed for 14 different individual VOCs or groupings: benzene; toluene; ethylbenzene; *m*-xylene; *p*-xylene; (*m*/*p*)-xylene; *o*-xylene; total BTEX; 1,3-butadiene; formaldehyde; *n*-hexane; total hydro carbons; styrene; and acrolein. The models were based on measurements ranging from 22 sites in El Paso (USA) to 179 sites in Tehran (Iran). Only four studies in Rome (Italy), Sabadell (Spain), Tehran, and Windsor (Canada) met the Cocheo's criterion of having at least one passive sampler per 3.4 km^2 of study area. The range of R^2 values across all models was from 0.26 for 1,3-butadiene in Dallas (USA) to 0.93 for benzene in El Paso. The average R^2 values among two or more studies of the same VOCs were as follows: benzene (0.70); toluene (0.60); ethylbenzene (0.66); (*m*/*p*)-xylene (0.65); *o*-xylene (0.61); total BTEX (0.66); 1,3-butadiene (0.46); and formaldehyde (0.56). The common spatial predictors of studied VOC concentrations were dominated by traffic-related variables, but they also included proximity to ports in the USA, number of chimneys in Canada, altitude in Spain, northern latitudes in Italy, and proximity to sewage treatment plants and to gas filling stores in Iran. For the traffic-related variables, the review suggests that large buffers, up to 5,000 m, should be considered in large cities. Although most studies reported logical directions of association for predictors, some reported inconsistent results. Some studies included log-transformed predictors while others divided one variable by another. Only six studies provided the p-values of predictors. Future work may incorporate chemistry-transport models, satellite observations, meteorological

* Corresponding author. Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland.
E-mail address: Hassan.Amini@unibas.ch (H. Amini).

http://dx.doi.org/10.1016/j.atmosenv.2017.10.010
Received 12 June 2017; Received in revised form 2 October 2017; Accepted 6 October 2017
Available online 07 October 2017

variables, particularly temperature, consider specific sources of aromatic vs aliphatic compounds, or may develop hybrid models. Currently, only one national model has been developed for Canada, and there are no global LUR models for VOCs. Overall, studies from outside North America and Europe are critically needed to describe the wide range of exposures experienced by different populations.

1. Introduction

Approximately 8% of global deaths and 4% of global disability-adjusted life-years (DALYs) were attributed to ambient air pollution in 2015, making it one of the leading modifiable risk factors for the Global Burden of Disease (GBD) worldwide (Cohen et al., 2017). These estimates were based on exposure to particulate matter ≤ 2.5 µm ($PM_{2.5}$) and ozone (O_3) (Brauer et al., 2016), for ischemic heart disease (IHD), cerebrovascular disease (ischemic stroke and hemorrhagic stroke), lung cancer, chronic obstructive pulmonary disease (COPD), and lower respiratory infections (LRI) (Cohen et al., 2017). The GBD estimates would be considerably higher if more air pollutants and health outcomes where included in the analyses (Amini et al., 2014a). For example, pollutants such as nitrogen dioxide (NO_2) and air toxics have been independently associated with increased risk of morbidity and mortality from the health outcomes already considered (Cesaroni et al., 2013; Filippini et al., 2015; Samoli et al., 2006; Thomas et al., 2014). On the other hand, health outcomes such as leukemia and other cancers (Lavigne et al., 2017; Weichenthal et al., 2017), neurodegenerative diseases (Chen et al., 2017; Heydarpour et al., 2014), and many others have been associated with the air pollutants (Bakian et al., 2015; Künzli et al., 2000; Nhung et al., 2017; Thurston et al., 2017; West et al., 2016).

The GBD estimates are partly influenced by the estimated relative risks extracted from cohort studies on long-term exposure to air pollution (GBD 2015 Risk Factors Collaborators, 2016). The most commonly used exposure assessment method in health outcomes analysis has been land use regression (LUR) (Hoek et al., 2013). Since its introduction in Europe (Briggs et al., 1997), the approach has been extensively used over the last 20 years to estimate the spatial variability in a wide range of pollutants, mainly in high-income countries (HICs) (Beelen et al., 2013, 2014; Dirgawati et al., 2016; Eeftens et al., 2012; Henderson et al., 2007; Raaschou-Nielsen et al., 2013). However, there have also been some studies in low- and middle-income countries (LMICs) (Amini et al., 2014b, 2016; Gurung et al., 2017; Lee et al., 2017; Yang et al., 2017).

Generally, LUR models have been applied to map concentrations of particulate matter and nitrogen oxides and, to a lesser extent, other pollutants, such as volatile organic compounds (VOCs) (Hoek et al., 2008). There are hundreds of VOCs species in ambient air, but they are all characterized by a low boiling point and ready transformation to the gaseous phase (Monks et al., 2009). Important groups of VOCs include: aliphatic alkanes, such as hexane; aromatic alkylbenzenes, such as benzene, toluene, ethylbenzene, and xylenes (BTEX); halogenated hydrocarbons, such as tetrachloromethane; and terpenes, such as carene (Baldasano et al., 1998). To date, most LUR models of VOCs have focused on the aromatic alkylbenzenes, which are ubiquitous in fossil fuels and combustion products (Monks et al., 2009).

One of the largest studies to use LUR for long-term exposure assessment in research on the development of chronic disease is the European Study of Cohorts for Air Pollution Effects (ESCAPE) (Beelen et al., 2013; Eeftens et al., 2012). However, studies such as ESCAPE and the GBD have not used LUR to model VOCs, despite their carcinogenicity (International Agency for Research on Cancer (IARC), 2016; Lipfert, 2017). We see three possible reasons for this. First, most studies in HICs have reported very low concentrations of VOCs in the ambient air (Guerreiro et al., 2014). Second, both $PM_{2.5}$ and O_3 have been consistently useful for predicting the range of health outcomes included in large studies (Brauer et al., 2016; Cohen et al., 2017). Third, study of VOCs has not been a priority for funding agencies, likely due to the first and second points. However, some studies have reported that ambient VOC concentrations are very high within large LMICs cities where large numbers of people are exposed (Amini et al., 2017a; Hoque et al., 2008; Matysik et al., 2010).

There is no evidence of a safe threshold for carcinogenic VOCs (e.g., benzene), and they may contribute to a considerable burden of disease even at concentrations below the current standards (Beelen et al., 2014; Künzli et al., 2015; Kutlar Joss et al., 2017). As such, the exclusion of VOCs from burden of disease studies likely leads to underestimation of the deaths and DALYs attributable to a wide range of cancers, particularly in densely populated and highly exposed LMICs. Based on the evidence to date, one key difference between VOCs and pollutants such as $PM_{2.5}$ or O_3 is the distributions of their ambient concentrations in HICs and LMICs. The burden of disease associated with $PM_{2.5}$ and O_3 also tends to be high in LMICs, but the vast literature from HICs can be used to inform LMICs calculations because there is considerable overlap between the distributions across contexts (Brauer et al., 2016). When it comes to VOCs, however, there appears to be little overlap in the distributions, which is consistent with the nature of pollutants that disperse quickly to the atmosphere. This results in a dearth of literature from highly-studied areas that can be used to inform calculations for highly-exposed areas.

Given (1) the potential burden of disease attributable to VOCs and (2) the limited information about population exposures to VOCs, we have conducted a systematic review of the published LUR models. Many different components of LUR modeling, such as monitoring data, geographic predictors, model development, and validation have been reviewed elsewhere (Hoek et al., 2008; Jerrett et al., 2005; Ryan and LeMasters, 2007), but no other study has focused on VOCs. In this article, we aim to provide a systematic review of LUR models developed for VOCs over the last two decades to summarize different elements of this literature, including: the geographic distribution of published studies; the VOCs modeled; the number of measurement sites used for VOC modeling; air quality data collection methods; VOC pollutant-specific predictor variables; and model evaluation. Finally, we discuss the knowledge gaps and future directions of research in this area.

2. Materials and methods

We searched 12 databases in *The Web of Science*: *Web of Science™ Core Collection*; *Medline®*; *BIOSIS Citation Index℠*; *BIOSIS Previews®*; *Current Contents Connected®*; *Data Citation Index℠*; *Derwent Innovations Index℠*; *Inspect®*; *KCI-Korean Journal Database; Russian Science Citation Index; SciELO Citation Index; and Zoological Record®*. The search phrases and keywords were: Land use regression; LUR; volatile organic compound; VOC; BTEX; benzene; toluene; and xylene. There was no restriction on timespan or language, and the final search was done on August 25, 2017. Only original research articles were retained for the analyses. Studies were only included if they were related to air pollution and they had developed an LUR model for at least one VOC. Studies were excluded if they had only applied VOC estimates from an LUR model described in another publication, or if no full-text was available (e.g. a conference abstract). The references of identified LUR articles on VOCs were also checked to identify any relevant articles missed by the search.

The full-text of all identified articles was reviewed and the following data were extracted: the year of publication; citation; study location (country, city/area, and population size); modeled VOCs; years of