Spatial modeling of ambient concentrations of volatile organic compounds in Montreal, Canada

Sara Zapata-Marin [iD][a,*], Alexandra M. Schmidt[a,b], Dan Crouse[c], Vikki Ho[d,e], France Labrèche[f], Eric Lavigne[g,h], Marie-Élise Parent[e,i], Mark S. Goldberg[b,j,k,l]

Background: Volatile organic compounds (VOCs) are components of the complex mixture of air pollutants within cities and can cause various adverse health effects. Therefore, it is necessary to understand their spatial distribution for exposure assessment in epidemiological studies.
Objectives: The objective was to model measured concentrations of five VOCs within the city of Montreal, Canada, developing spatial prediction models that can be used in health studies.
Methods: We measured concentrations using 3M 3500 Organic Vapor Monitors, over 2-week periods, for three monitoring campaigns between 2005 and 2006 in over 130 locations in the city. Using GC/MSD (Gas Chromatography/Mass Selective Detector), we measured concentrations of benzene, n-decane, ethylbenzene, hexane, and trimethylbenzene. We fitted four different models that combine land-use regression and geostatistical methods to account for the potential spatial structure that remains after accounting for the land-use variables. The fitted models also accounted for possible variations in the concentration of air pollutants across campaigns.
Results: The highest concentrations for all VOCs were found in December with hexane being the most abundant followed by ethylbenzene. We obtained predicted surfaces for the VOCs for the three campaigns and mean surfaces across campaigns. We found higher concentrations of some VOCs along highways and in the Eastern part of Montreal, which is a highly industrialized area.
Conclusions: Each of the fitted models captured the spatial and across-campaigns variability for each VOC, and we found that different VOCs required different model structures.

Keywords: Air pollution; Monitoring; Spatial statistics; Volatile organic compounds

Introduction

Volatile organic compounds (VOCs) are organic compounds that have high vapor pressures (~10 Pa) at room temperature (25°C).[1] Acute and chronic exposures to these chemical compounds can cause adverse health effects such as irritation of the eyes and upper respiratory tract, and effects on the central nervous system (e.g., loss of coordination), as well as being toxic to the liver and kidneys.[2,3] Furthermore, benzene, trichloroethylene, and vinyl chloride are accepted carcinogens[3] and ethylbenzene, carbon tetrachloride, 1,2-dichloroethane, chloroform, and other trihalomethanes have been identified as possible carcinogens.

In Canada, it has been estimated that 2.3 Mt of VOCs were released in 2005, a 21% (600 Kt) decrease from 1990.[4] Oil and gas industries were the main source of VOC emissions (29.6% of total emissions; 650 kt), followed by transportation and mobile equipment (27.4%; 630 kt), and paints and solvents (19.13%; 440 kt).[4]

Spatial distribution and seasonality of concentrations of VOCs in urban areas

Governments use annual emission inventories of greenhouse gases and air pollutants to identify sources of atmospheric pollutants. Given restrictions related to data, time, staff, funding, and the lack of a systematic assessment, it has been shown that emission inventories often underestimate concentrations of air pollutants.[5–7] Although most emission inventories have a 1° by 1° spatial resolution and a temporal resolution of one year, this

What this study adds

It is critical to study the spatial distribution of volatile organic compounds (VOCs) to understand the population health risks associated with their exposure. Bayesian hierarchical models were used to account for the spatial and across-campaigns variation of benzene, n-decane, ethylbenzene, trimethylbenzene, and hexane concentration in Montreal for three monitoring campaigns. Higher concentrations were found during December with hexane being the most abundant VOC. The predicted surfaces also showed higher concentrations along some of the major highways and in the Eastern part of the island where refineries were operating at the time of the study.

level of spatial resolution is not sufficient to characterize the variability of concentrations of some pollutants within cities, as some studies have shown that for pollutants related to traffic, such as nitrogen dioxide (NO_2), intraurban variability exceeds inter-city variability.[8–10] Furthermore, recent studies found differences in the distribution of sources and in the concentrations of ozone precursors, especially VOCs, between urban, suburban, rural, and industrial areas.[11–14]

Within cities, traffic is a significant source of many air pollutants such as nitrogen dioxide, carbon dioxide, particulate matter, black carbon, ultrafine particles, and VOCs.[15] More specifically, in downtown areas, vehicle exhaust is the most abundant source of airborne pollutants.[16]

In urban areas, VOCs are emitted by different sources such as traffic, industrial processes, construction, waste, and solvents.[16] Some VOCs related to traffic include benzene, toluene, ethylbenzene, and xylene (BTEX), hexane, methylcyclopentane, 2-methylhexane, 2,2,4-trimethylpentane, trans-2-pentene, ethyltoluene, 1,2,4-trimethylbenzene and 1,2,4-trichlorobenzene, and many others.[17] Moreover, alkanes associated with transport-related emissions, such as hexane, are highly correlated with carbon monoxide.[18]

The concentration and mixture of traffic-related pollutants can change across space and time depending on a combination of different factors such as the proximity to roads or other point sources, traffic volume, the composition of the vehicle fleet, and the presence of other pollutants.[19] Depending on pollution levels, individual susceptibility, the population levels of activity, acute and past exposures can lead to diverse adverse health effects immediately or after years of exposure.[19]

Given the adverse health effects of these pollutants, in this study, we focus on five VOCs commonly found in urban areas: benzene, n-decane, ethylbenzene, hexane, and 1,2,3-trimethylbenzene. Benzene, ethylbenzene, and trimethylbenzene are three aromatic hydrocarbons found in fossil fuels and urban air masses, predominantly emitted by vehicle exhausts, fuel evaporation, and spoilage.[20] N-decane and hexane, alkane hydrocarbons, are found in fossil fuels and solvents that become airborne by evaporation or combustion.[21]

Land-use regression (LUR) methods have been extensively used to estimate the spatial distribution of air pollutants such as nitrogen oxides (NOx), NO_2, particulate matter 2.5 ($PM_{2.5}$), and VOCs[10,22,23] and to study their relationship with environmental factors in urban settings. Although there is no consensus on the monitoring process[22] (e.g., number of monitoring sites, monitoring period, the distance between sites), a study on NO_2 found that studies using LUR should be based on a large number of sites (>80) for better performance.[24]

Previous studies have shown that, contrary to ozone and $PM_{2.5}$, there is little overlap in the distributions of VOCs between high-income countries and low- and middle-income countries, given how quickly VOCs disperse into the atmosphere.[22] However, within urban areas, LUR studies for $PM_{2.5}$, ozone, and VOCs usually involve the same set of predictors, such as traffic-related variables, land use, and other unique features of the study area.[22]

Objectives

To support two population-based case-control studies of postmenopausal breast cancer[25,26] and one prostate cancer case-control study[27] that we conducted in the mid-1990s and early 2000s in Montreal, Quebec, we conducted a dense monitoring program of NO_2[14] and selected VOCs to link these to the residential addresses of participants in these studies.

The main objective of the present study was to determine the spatial distribution of ambient concentrations of the five selected VOCs from our monitoring campaign conducted in 2005 and 2006, using a combination of land-use regression and geostatistical methods.

Additionally, we obtained predicted surfaces by interpolating concentrations at locations where measurements were not made, while accounting for local variations in concentrations so that in our case-control studies, we can link the predictions with residential addresses of participants,[25–27] and hence estimate risks associated with these exposures. The implications of using these predictions in health-related studies as proxies of environmental exposures are beyond the scope of the present article.

Materials and methods

The greater Montreal area is the second most populated city in Canada, with a population in 2016 of over 4 million inhabitants.[28] From 1981 to 2010, the mean daily temperature in April was around 6.4°C (temperature range, 1.2° to 11.6°C), 20.1°C in August (temperature range, 14.8°C to 25.3°C), and −5.4°C in December (temperature range, −9.3°C to −1.4°C).[29] The average annual concentration of NO_2 in 2018 was of 10.4 parts per billion (ppb), while the 3-year average from 2016 to 2018 was 7.4 µg/m³ for fine particulate matter and 57 ppb for ozone.[30]

The east end of Montreal is of particular interest as it is an industrial area with refineries and various other heavy industries. Between 2005 and 2006, refineries included Shell Canada Montreal East Refinery (closed in 2010), the Petro Canada Montreal Refinery (now Suncor), and petrochemical plants like Parachem Petrochemical and Petromont (closed in 2008).[31]

Data collection

The location of the samplers was chosen using a population-weighted location-allocation model that placed 133 samplers in areas likely to have high spatial variability of traffic-related pollution and in areas with high population densities.[14] In addition, we added about 20 samplers to capture concentrations in residential areas that were under-represented by the initial allocation scheme. The minimum distance between any two neighboring samplers was approximately 100 m and the maximum distance was just over 3 km. The samplers were deployed in three monitoring campaigns: December 2005 ("cold" weather), April 2006 ("temperate" weather), and August 2006 ("hot" weather).

In addition to the Ogawa samplers that measured concentrations of NO_2,[14] we co-located passive 3M 3500 Organic Vapor Monitors (3M Company, Saint Paul, MN). After a 2-week uninterrupted sampling period, we retrieved each monitor, snapped the shipping cap onto the monitor, ensuring that the two-port plugs were sealed firmly, and then recorded the date and time. We placed the sampler in the shipping container, closed the container with its plastic lid, and then sealed it immediately with Teflon tape. These were then shipped to a commercial laboratory that conducted all of the analyses (Airzone, Mississauga, ON[27]).

Samples were extracted with 2 ml of solvent (carbon disulfide) and concentrations using GC/MSD (Gas Chromatography/ Mass Selective Detector) were estimated using NIOSH (National Institute for Occupational Safety and Health) methods 1,003, 1,500, and 1,501 with a detection limit of 0.2 µg/m³. We had

three field blanks per sampling survey, and all sample results were corrected with the blanks, deuterated internal standard, and recovery. The multipoint calibration curve had a $R^2 > 0.999$ and the detection limits, based on the U.S. Federal Register Code of Federal Regulations (CFRs) 40 method, are shown in Table SM-1 (http://links.lww.com/EE/A199).

As part of the fieldwork, we included additional monitors in each sampling period in randomly selected monitoring sites to determine the measurement campaign's reliability. We had 30, 8, and 9 duplicates in the December, April, and August campaigns, respectively, and 4 triplicates in the December campaign. To assess agreement between these duplicate monitors, we computed differences between duplicate measurements to produce Bland-Altman plots, and we also computed the intraclass correlation coefficient (ICC) and associated 95% confidence intervals using a multilevel procedure. ICC values close to 1 imply a high similarity between duplicated measurements.

Volatile organic compounds data

We analyzed concentrations of five exhaust-related VOCs (n-decane, hexane, ethylbenzene, benzene, and 1,2,4-trimethylbenzene) for three, continuous 2-week monitoring campaigns in December 2005, April 2006, and August 2006. There were 133 monitoring locations for the December and April campaigns and 131 for the August campaign.

Land-use variables

Potential predictors of the different VOCs were obtained using circular buffers at 50-, 100-, 200-, 500-, and 1,000-m radii around each monitoring location (Table SM-3; http://links.lww.com/EE/A199). Land-use variables were available as a proportion of the area of each buffer covered by each specific variable. The available land-use variables for each VOC were obtained from DMTI Spatial Inc.[32] (Markham, Ontario, Canada) and included buildings, open areas, residential, industrial, commercial, waterbody, parks and recreational, governmental and institutional, commercial, and roads land-use, which are common predictors for local variability of urban air pollution.[14,33,34] Average and total NOx and total daily traffic volume were obtained from VISSIM (Verkehr In Städten – SIMulationsmodell),[35] a traffic simulation software, and MOVES (MOtor Vehicle Emission Simulator),[36] an emission modeling system for mobile sources. Population density for 2016 was based on Canadian census data.[28] Finally, the easting and northing coordinates were also included in the mean structure of the models.

After obtaining the spatial variables at the different buffer sizes, we selected the appropriate variables and buffer sizes for each VOC in each campaign by using the procedure of least absolute shrinkage and selection operator (LASSO).[37] This regression analysis method shrinks the coefficients towards zero to reduce the set of covariates used in the model. The goal is to minimize the sum of squared errors with a bound on the absolute values of the coefficients. We included the variables selected using LASSO for each campaign so that we would have the same set of variables for all three campaigns.

Because our main goal was to predict concentrations where measurements were not made rather than to find associations between the concentrations and the land-use variables, we excluded variables that were highly correlated (>0.99) with each other (Tables SM-3–10; http://links.lww.com/EE/A199). For example, average and total NOx were highly correlated, and we included only average NOx.

Statistical Analysis

For each VOC, we fitted four different regression models (see below). These models considered possible variations between

monitoring campaigns and a possible spatial structure between the monitoring locations after accounting for the land-use variables. To select the best model among the fitted ones for each VOC, the Watanabe-Akaike information criterion (WAIC) was used. WAIC is a measure of the predictive accuracy that allows us to measure the performance of a model and to compare multiple models accounting for both goodness of fit and model complexity.[38] Smaller values indicate the optimal model among the fitted ones.

Model description

To obtain predicted surfaces over a region given the information collected at a set of monitoring stations, we used a combination of land-use regression and geostatistical methods to obtain the predicted surfaces. This model considered possible variations across campaigns and a possible latent (unobservable) spatial structure after accounting for the land-use variables. This latent structure is such that neighboring locations will tend to have similar adjustment after accounting for the available explanatory variables.[39]

Specifically, for all our models, let $Y_j(\mathbf{s})$ be the natural logarithmic concentration of a VOC at location s, and campaign j where $j = 1, 2, 3$ for the December, April, and August campaigns, respectively. The base model is defined as follows:

$$Y_j(\mathbf{s}) = Z_j' \alpha_j + \mathbf{X}'(\mathbf{s}) \beta_j + \omega_j(\mathbf{s}) + \epsilon_j(\mathbf{s}) \quad (1)$$

where $\mathbf{X}(\mathbf{s})$ is a q-dimensional vector containing the land-use predictors, an intercept, and the standardized Universal Transverse Mercator (UTM) coordinates at each monitoring site; β_j is a vector of coefficients associated with the land-use variables for each campaign; Z_j represents each campaign, such that $Z_1 = 0$, and α_j is the coefficient associated with this indicator variable. Finally, the latent spatial structure ω_j helps to accommodate for a possible residual spatial structure left after accounting for the land-use variables and the campaign fixed effect Z_j. The spatial residual $\omega_j(\cdot)$ follows a zero mean Gaussian process with variance σ_j^2 and an exponential correlation function. Then $\boldsymbol{\omega_j} = (\omega_j(\mathbf{s}_1), \cdots, \omega_j(\mathbf{s}_n))'$ follows, a priori, a multivariate normal distribution with mean 0 and an exponential covariance matrix $\boldsymbol{\Sigma_j} = \sigma_j^2 exp(-\mathbf{d}/\phi_j)$ where σ_j^2 is the partial sill at each campaign j, d is the Euclidean distance matrix, and ϕ_j is a parameter that controls how fast the spatial correlation among $\omega_j(\cdot)$ decays to zero; and the measurement error $\epsilon_j(\mathbf{s})$ follows a zero mean normal distribution with variance τ^2 (nugget effect).

We fitted four models that are particular cases of the general structure in equation (1). For model 1, we let $\boldsymbol{\beta_j} = \boldsymbol{\beta}$ meaning the effect of the land-use variables is the same across campaigns, in model 2, we also let $\boldsymbol{\beta_j} = \boldsymbol{\beta}$, and $\omega_j(\mathbf{s}) = 0$, so that the spatial variation is fully captured by the land-use variables. In model 3, we let $\boldsymbol{\alpha} = 0$, so that all the variability across campaigns is captured by β_j, and $\omega_j(\mathbf{s})$, and finally in model 4, we let $\boldsymbol{\alpha} = 0$ and $\boldsymbol{\omega_j}(\mathbf{s}) = 0$, meaning there is no additional spatial structure after accounting for the land-use variables.

Under the Bayesian paradigm, model specification is complete after assigning a prior distribution to the parameter vector. In our case, all parameters were assumed to be independent a priori. We assigned a normal distribution having a mean of zero and with large variances for $\boldsymbol{\alpha}$ and for $\boldsymbol{\beta}$ in models 1 and 2, as this reflects our prior ignorance about the association between the land-use variables and the concentrations of VOCs; for σ^2 and τ^2, we assigned an inverse gamma prior with mean fixed at 1 and an infinite variance; for the spatial range $\boldsymbol{\varphi}$ we assigned an exponential prior with mean equal to the practical range (assuming the correlation between sites decreases to 0.05 at half of the maximum observed distance); and for model 4, the regression coefficients $\boldsymbol{\beta_j}$ follow a normal prior with mean $\boldsymbol{\gamma}$, and variance ψ^2 where $\boldsymbol{\gamma} \sim N(0, 10)$, and ψ^2 follows an inverse gamma distribution with mean equals to 1 and an infinite variance. The parameter vector for each model is defined as $\theta_1 = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau, \boldsymbol{\sigma}, \boldsymbol{\varphi}\}$ for model 1, $\theta_2 = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau\}$ for model 2, $\theta_3 = \{\boldsymbol{\beta}, \tau, \boldsymbol{\sigma}, \boldsymbol{\varphi}\}$ for model 3, and $\theta_3 = \{\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \boldsymbol{\sigma}, \psi\}$ for model 4.

We used Markov Chain Monte Carlo methods[40] to obtain samples from the posterior distribution. The statistical analysis was conducted using the package Nimble in the software R (R Foundation for Statistical Computing, Vienna, Austria; version 3.4.5).[41] The code for fitting these models is available at https://github.com/SaraZM/VOCs.

After fitting all four models, we chose the appropriate model for each VOC using the minimum value of the WAIC.

Results

Comparison of duplicate monitors

Bland-Altman plots of the analyses of agreement of duplicate monitors are shown in the Supplement (Figure SM-1; http://links.lww.com/EE/A199 and Table SM-2; http://links.lww.com/EE/A199), where we observed an average difference across observations close to zero, except for n-decane and hexane where we found an average absolute difference of about 0.15 and 0.28 µg/m³, respectively. For most pollutants, we found in each campaign ICCs close to unity, indicating almost perfect agreement in measured concentrations, but an anomalously low ICC for hexane in the December campaign and for 1,2,4-trimethylbenzene in the August campaign.

Volatile organic compounds

For the December 2005 campaign, there were eight measurements below the detection limit for n-decane, and in the April and August 2006 campaigns, there were three and two measurements below the detection limit for n-decane and hexane, respectively. Given there are only few measurements below the limit of detection these measurements were excluded from the analysis. Additionally, only the locations that were present in all three campaigns were analyzed. Therefore, the analysis included 127 monitoring locations for hexane, 121 for n-decane, and 129 for ethylbenzene, 1,2,4-trimethylbenzene, and benzene.

The concentration of benzene ranged from 0.18 to 5.27 µg/m³, n-decane ranged between 0.13 and 5.25 µg/m³, ethylbenzene ranged between 0.59 and 27.47 µg/m³, hexane ranged between 0.40 and 32.03 µg/m³, and 1,2,4-trimethylbenzene between

0.36 and 2.22 µg/m³ (Table 1). The largest variability for all the VOCs was found in December, except for ethylbenzene, that had a greater variability in August. Hexane had the highest concentrations across campaigns, followed by ethylbenzene.

Model comparison and diagnostics

To choose the best model for each VOC, among the proposed ones, we used the minimum WAIC (Table 2). For benzene and n-decane, the selected model was one with spatial structure and an indicator variable to capture the variability across campaigns (model 1). For ethylbenzene, hexane, and 1,2,4-trimethylbenzene, the selected model had different coefficients for the land-use variables across campaigns and did not have a component for spatial structure (model 4).

We also computed the observed versus the predicted values for each VOC, and we did not identify important outliers (Figure 1). Although not shown here, we found no important patterns in the residuals for each campaign after accounting for the covariates and spatial structure.

For benzene and n-decane (Tables SM-4–7; http://links.lww.com/EE/A199), we obtained different intercepts for each campaign, with August having the lowest values. For benzene (Tables SM-4 and 5; http://links.lww.com/EE/A199), the spatial variance and practical range also changed across campaigns, suggesting that the spatial structure was different for each of the campaigns, with the highest spatial variance in the August campaign and similar values for the December and April campaigns. In the case of n-decane (Tables SM-6 and 7; http://links.lww.com/EE/A199), we obtained similar values for the spatial variance across campaigns but different values for the practical range, with a posterior mean of 8.82 km and 1.53 km for the April and December campaigns respectively.

For ethylbenzene, hexane, and 1,2,4-trimethylbenzene (Tables SM-8–10; http://links.lww.com/EE/A199), some of the coefficients associated with land-use variables differed in magnitude and direction across campaigns. This might be due to seasonal effects that can affect the relationship between the land-use variables and the pollutants levels.

To estimate the concentrations of each VOC at unsampled locations across Montreal, we obtained samples from

Table 1.

Selected moments of the distributions of benzene, n-decane, ethylbenzene, hexane, and 1,2,4-trimethylbenzene levels (in µg/m³) across three sampling campaigns in Montreal, between 2005 and 2006

Descriptive statistics	Benzene				n-Decane			
	December	April	August	Average	December	April	August	Average
Mean	1.35	1.30	0.56	1.07	2.08	2.05	0.95	1.69
Median	1.28	1.18	0.49	0.98	1.92	1.96	0.88	1.61
SD	0.61	0.60	0.31	0.44	0.73	0.66	0.44	0.42
Minimum	0.40	0.68	0.18	0.54	0.25	0.89	0.13	1.01
Maximum	4.72	5.27	2.51	3.35	4.27	5.25	3.31	3.27

Descriptive statistics	Ethylbenzene				Hexane			
	December	April	August	Average	December	April	August	Average
Mean	3.63	2.77	2.03	2.81	14.25	6.35	1.57	7.39
Median	3.23	2.65	1.82	2.62	13.78	5.83	1.42	7.07
SD	2.15	0.92	2.30	1.62	5.21	3.80	0.83	2.11
Minimum	1.12	1.22	0.59	1.30	2.24	2.32	0.40	3.32
Maximum	23.70	8.90	27.47	20.02	30.77	32.03	5.56	16.79

Descriptive statistics	1,2,4-Trimethylbenzene			
	December	April	August	Average
Mean	1.14	0.98	1.00	1.04
Median	1.10	0.98	0.97	0.99
SD	0.38	0.22	0.26	0.23
Minimum	0.36	0.53	0.53	0.63
Maximum	2.16	1.85	2.22	1.85

WAIC of the fitted models for each VOC

Model	Benzene	n-Decane	Ethylbenzene	Hexane	1,2,4-Trimethylbenzene
Model 1	**119.64**	**331.18**	235.37	448.28	−19.80
Model 2	190.47	363.36	269.78	417.39	−13.58
Model 3	151.23	382.25	260.87	496.53	32.20
Model 4	187.05	348.87	**219.44**	**415.02**	**−63.02**

Bold values (minimum WAIC) identify the selected models.

the resultant posterior predictive distributions of the different VOCs at unobserved locations of interest (e.g., residences of participants in the three case-control studies), and we computed means, medians, and SDs at each unobserved location, as described in Section 3 of the Supplementary Material (http://links.lww.com/EE/A199).

Figure 2 shows the predicted surfaces on a natural logarithmic scale obtained for a grid with a 0.25 by 0.25 km cell size. We also obtained the SD of the posterior predictive distribution showing higher uncertainty in areas where no monitors were located (Figures SM-2–6; http://links.lww.com/EE/A199).

The predicted surfaces for all the analyzed VOCs resulted in higher concentrations across the island for the December campaign, especially in the north part of the island. The predicted surfaces for benzene showed higher concentrations in an area in the northeast part of the island with December and April showing similar levels. The mean predicted surface also showed the highest values at the northeast part of Montreal. For benzene and n-decane, we also found high concentrations during the April campaign. For n-decane and 1,2,4-trimethylbenzene, we also found higher concentrations in the central part of the island and along some of the most important highways in Montreal. The predicted surface of hexane shows the highest concentrations for the December campaign with small spatial variability across campaigns (Figure SM-7; http://links.lww.com/EE/A199).

Discussion

We fitted spatial regression models to data from three dense sampling campaigns in Montreal. Land-use variables were used to determine the predicted concentrations of five selected VOCs related to traffic. For each VOC, we selected the model that explained the most variability in the data, accounting for the complexity of the model, and found reasonably good fits to the data (Figure 1). From these models, we then predicted the concentration of each VOC across the island at a resolution of 0.25 by 0.25 km cell grid size (Figure 2).

We found the highest concentrations for all five VOCs during the December campaign, followed by April and August. Meteorological conditions such as anticyclones in colder weather, leading to stagnant meteorological conditions, might have facilitated the accumulation of these air pollutants during the December campaign. Higher concentrations during winter have also been found in other cities for ozone and different VOCs.[11,12,42,43] Additionally, higher levels of benzene and hexane were observed during this period compared with levels in other Canadian cities.[44,45]

We also found that the spatial distribution of the different VOCs changed by season; therefore, using one season to represent the annual exposure is not recommended, as shown in a previous study.[45]

There is face validity to our results. Specifically, for benzene, we found that areas located in the northeast part of Montreal where the highest benzene levels were predicted correspond to an area with oil refineries operating at the time the study was conducted. Additionally, we found for n-decane that the highest levels were predicted for sections of a major highway (Autoroute 40), especially in the North. As well, for 1,2,4-trimethylbenzene,

the predicted areas with the highest levels corresponded to multiple sections of several highways (Autoroute 40, Autoroute 136, and Autoroute 15).

Given the nature of these VOCs and their emission sources, these findings are consistent with previous studies on NO₂, ozone, and VOCs in other Canadian cities and low- and middle-income countries.[14,44,46–48] These studies have found a similar spatial pattern with higher concentrations of air pollutants along road networks, in high traffic areas (e.g., downtown areas), and near industrial sources. Previous studies on the spatial distribution of NO₂ in Montreal have also found higher concentrations along some of the major highways.[14,46] However, due to the multiple emission sources of VOCs, in addition to major roads, we also found higher concentrations in industrialized areas.

Additionally, the majority of the LUR models for intraurban variation of VOCs are concentrated in high-income countries, particularly in Europe and North America.[22] Furthermore, low- and middle-income countries can be more affected by air pollution than high-income countries.[22] Even though the mean concentration of VOCs can vary a lot across different cities, given the nature of the five VOCs analyzed in this study, higher values are consistently found in high traffic and industrialized areas.[11,16,44]

Strengths and limitations

We used 3M passive monitors because of ease in installing on fixed city poles at 10-feet heights, they did not require electricity or pumps, if stolen they would not be costly, and were thought to be sufficiently accurate and precise. Passive samplers have been shown to be reliable in measuring VOCs over extended periods of time.[49] Indeed, we found values of the intraclass correlation coefficient close to unity for duplicate samplers. Furthermore, an analysis of 3M organic passive dosimeters outdoors using a sampling duration of 72 hours was comparable to automated continuous gas chromatography measurements.[50] Other methods could have been used that could have led to more accurate estimates, such as passivated Summa canisters and flame and photoionization detectors, but they are not suitable for remote sites without electricity, their operation is difficult in cold weather, they require knowledge of the proportions of concentrations of the different VOCs, and these methods are expensive.

The present study sampled a considerably larger number of sites than some of the previous studies.[8,11–13,43–45,48,51] Additionally, both the spatial and campaign variability was accounted for by the model instead of averaging the data across campaigns. This allowed us to determine if the associations between each of the VOCs and explanatory variables changed across campaigns. From a statistical point of view, when one averages the observations across campaigns and then fits the models, one is implicitly assuming that these associations do not vary across campaigns. We believe this is a strong assumption and must be verified. By obtaining the predicted surface for several VOCs, we not only facilitate the comparison of pollutant levels across campaigns but also across VOCs.

As the main goal of the present study was prediction, the results presented here should not be used to identify the relationship between land-use variables and VOC concentrations. To meet this goal, one needs to adjust the method for variable selection by testing for collinearity and confounding and possibly changing the set of predictors for each campaign. Additionally, it would also be valuable to account for meteorological covariates such as temperature, wind speed, humidity, and atmospheric pressure.

Some previous reviews on LUR methods highlighted the importance of using local characteristics and sources specific to each VOC (e.g., dry cleaners, gas stations, etc.) to improve the predictions of their spatial distribution.[22] For example, a previous study found that area sources were correlated with the
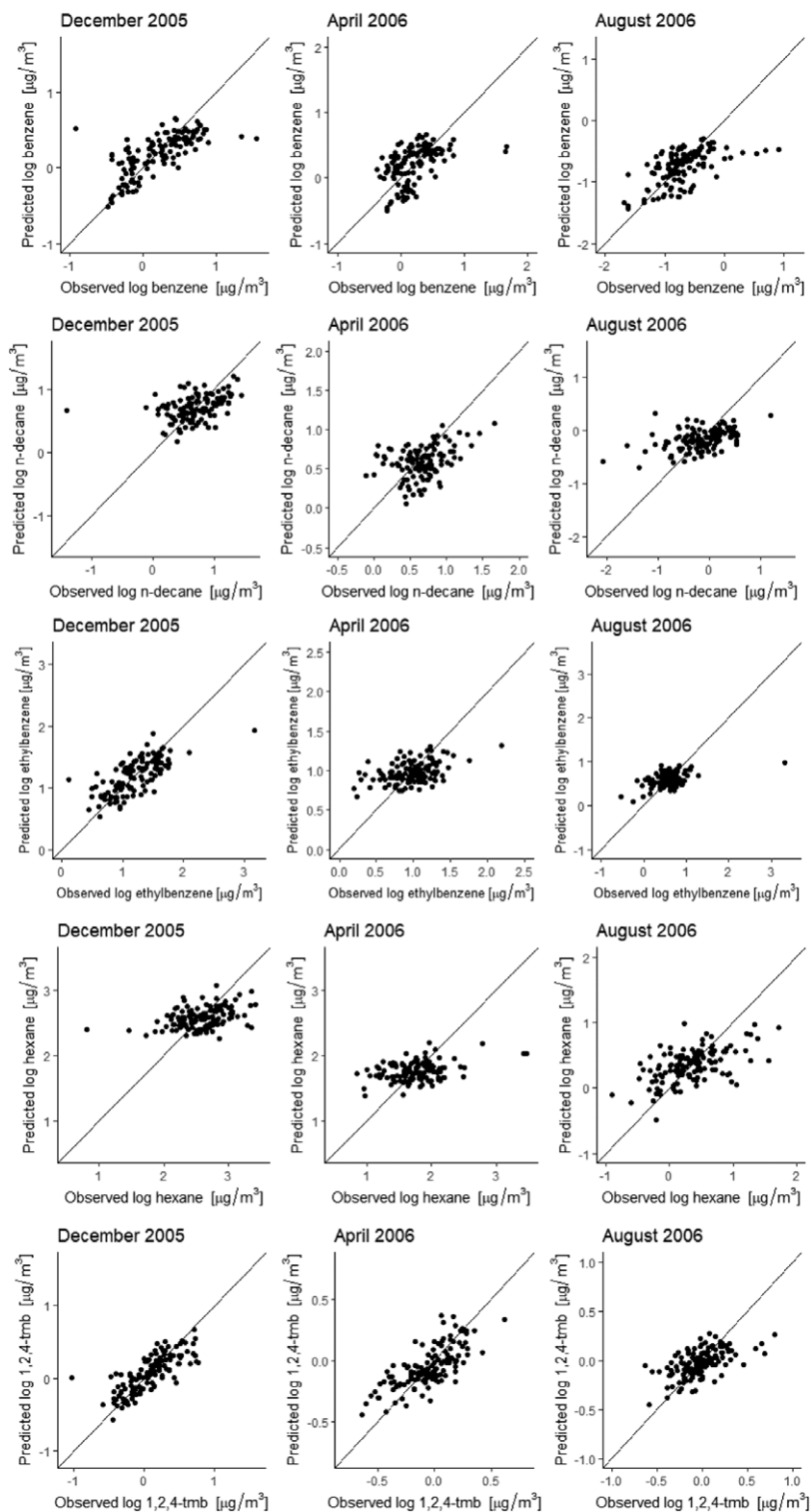
Figure 1. Scatter plots of the observed versus fitted values for benzene, n-decane, ethylbenzene, hexane, and 1,2,4-trimethylbenzene using the selected models (Table SM-3; http://links.lww.com/EE/A199). The straight line represents perfect prediction.

concentrations of VOCs at small spatial scales.[52] Furthermore, there might also be variations between local VOC sources in rural and urban areas. For example, a study found higher levels of benzene, toluene, ethylbenzene, and xylene in rural areas compared with cities due to different point sources such as domestic coal combustion during winter.[53]

We did not make use of point sources because in Canada, it is based on industry-reported data, and these emissions from point sources are known to underestimate emissions[54] and, to our knowledge, there is no emission inventory specific to the Montreal area that includes all relevant sources.

Finally, this study shares the same limitation of other LUR methods[22] namely that the results are case- and area-specific; therefore, the results found here are only valid for the studied area.
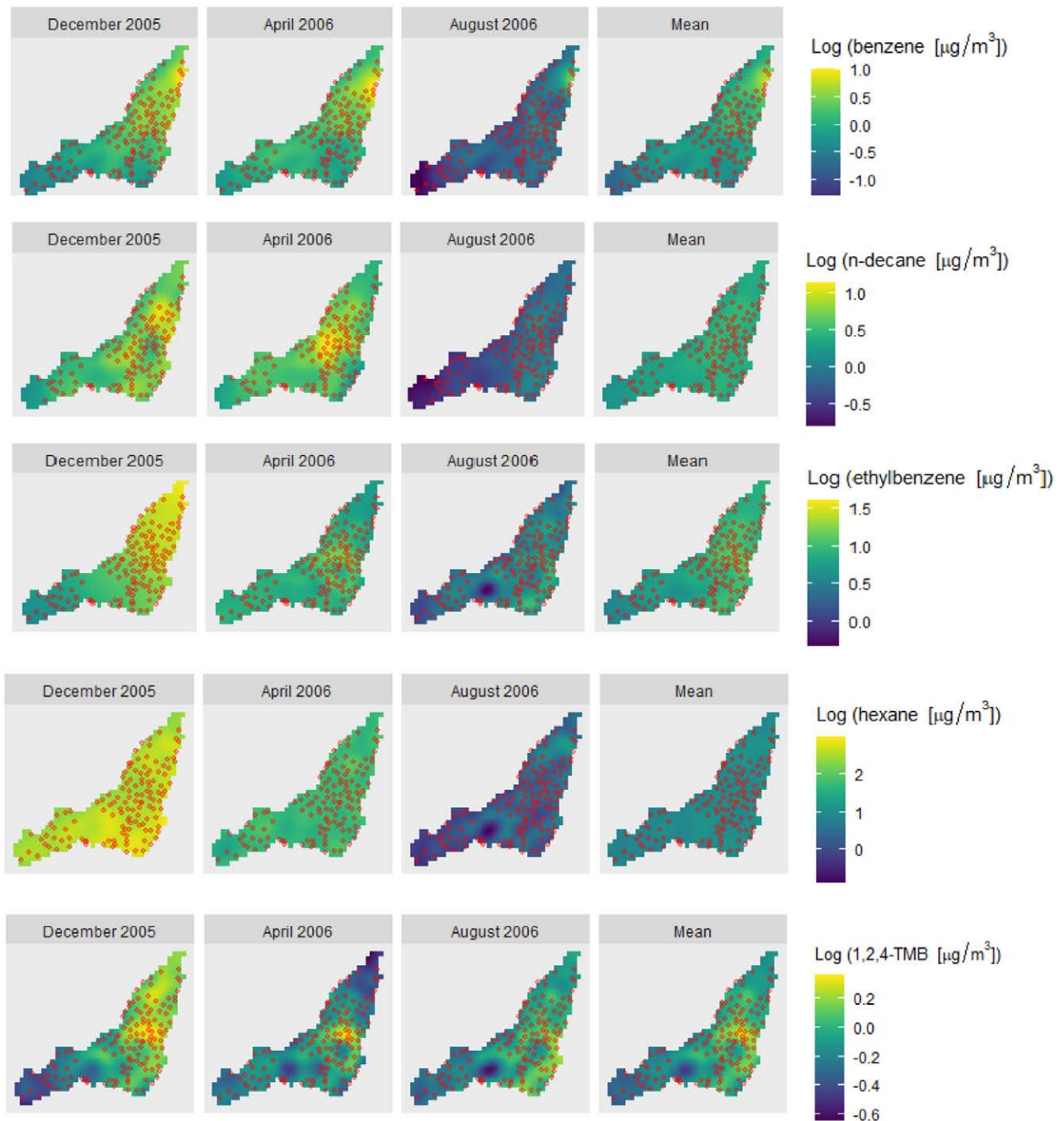
Figure 2. Posterior mean of the predicted surfaces in the log scale for benzene, n-decane, ethylbenzene, hexane, and 1,2,4-trimethylbenzene concentration at each campaign. Red solid circles represent the locations of the monitors.

Conclusions

In the present study, we obtained predicted surfaces showing the spatial variability of each of the five VOCs. We found higher concentrations of VOCs in the east and central part of Montreal with higher concentrations during the winter campaign.

We proposed four models, each of which accounted for spatial and campaign variability. The model that fitted the best, according to WAIC, for benzene and n-decane accounted for the spatial structure after adjusting for the land-use variables, and for seasonal variability through the intercept. For hexane, ethylbenzene, and 1,2,4-trimethylbenzene, land-use covariates alone accounted for the spatial variability, and the campaign variability was accounted for through the coefficients associated with the land-use variables.

Inference was performed under the Bayesian framework; therefore, it was straightforward to obtain summaries of the predictive posterior distribution such that spatial interpolation to unobserved locations of interest naturally accounts for the uncertainty in the estimation of the unknowns in the model (Section 3 of the Supplementary Material; http://links.lww.com/EE/A199).

The proposed models are flexible to adjust for any set of land-use variables or air pollutants concentrations, and the methods are easily reproducible. The predicted surfaces obtained here,

and the spatial interpolation methods used in this study, can help estimate the air pollutant levels at residential addresses of participants for health studies.

References

1. Williams J, Koppmann R. Volatile organic compounds in the atmosphere: an overview. In: Koppmann R, ed. Volatile Organic Compounds in the Atmosphere. John Wiley & Sons, Ltd. 2007:1–32.
2. United States Environmental Protection Agency. *Volatile Organic Compounds' Impact on Indoor Air Quality*. Available at: https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality. Accessed 16 August 2021.
3. Centre international de recherche sur le cancer. *IARC Working Group on the Evaluation of Carcinogenic Risks to Humans*. Benzene. Lyon, France; 2018. Available at: https://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans/Benzene-2018.
4. Environment and Climate Change Canada. *Canada's Air Pollutant Emissions Inventory Report*. 2020. Available at: https://publications.gc.ca/collections/collection_2022/eccc/En81-30-2020-eng.pdf.
5. Gurney KR, Liang J, Roest G, et al. Under-reporting of greenhouse gas emissions in U.S. cities. Nat Commun. 2021;12:553.
6. Ren X, Salmon OE, Hansford JR, et al. Methane emissions from the Baltimore-Washington area based on airborne observations: comparison to emissions inventories. J Geophys Res Atmos. 2018;123:8869–8882.
7. Turner AJ, Jacob DJ, Benmergui J, et al. A large increase in U.S. methane emissions over the past decade inferred from satellite data and surface observations. Geophys Res Lett. 2016;43:2218–2224.
8. Briggs DJ, Collins S, Elliott P, et al. Mapping urban air pollution using GIS: a regression-based approach. Int J Geogr Inf Sci. 1997;11:699–718.
9. Yifang Zhu WCH, Seongheon K, Shen S, et al. Study of ultrafine particles near a major highway with heavy-duty diesel traffic. Atmos Environ. 2002;36:12.
10. Jerrett M, Arain A, Kanaroglou P, et al. A review and evaluation of intraurban air pollution exposure models. J Expo Anal Environ Epidemiol. 2005;15:185–204.
11. Bozkurt Z, Üzmez OO, Döğeroğlu T, et al. Atmospheric concentrations of SO₂, NO₂, ozone and VOCs in Düzce, Turkey using passive air samplers: sources, spatial and seasonal variations and health risk estimation. Atmos Pollut Res. 2018;9:1146–1156.
12. Kumar A, Singh D, Kumar K, et al. Distribution of VOCs in urban and rural atmospheres of subtropical India: temporal variation, source attribution, ratios, OFP and risk assessment. Sci Total Environ. 2018;613–614:492–501.
13. Li B, Ho SSH, Qu L, et al. Temporal and spatial discrepancies of VOCs in an industrial-dominant city in China during summertime. Chemosphere. 2021;264:128536.
14. Crouse DL, Goldberg MS, Ross NA. A prediction-based approach to modelling temporal and spatial variability of traffic-related air pollution in Montreal, Canada. Atmos Environ. 2009;43:5075–5084.
15. Health Canada. 2022. Traffic-Related Air Pollution: Asthma, Allergies, and Lung Function. Available at: https://www.canada.ca/en/health-canada/services/publications/healthy-living/traffic-related-air-pollution-asthma-allergies-lung-function.html Accessed July 12, 2022.
16. Tsai J-H, Lu Y-T, Chung I, et al. Traffic-related airborne VOC profiles variation on road sites and residential area within a microscale in urban area in Southern Taiwan. Atmosphere. 2020;11:1015.
17. Yang H-H, Dhital NB, Wang Y-F, et al. Effects of short-duration vehicular traffic control on volatile organic compounds in roadside atmosphere. Atmos Pollut Res. 2020;11:419–428.
18. Baker AK, Beyersdorf AJ, Doezema LA, et al. Measurements of nonmethane hydrocarbons in 28 United States cities. Atmos Environ. 2008;42:170–182.
19. Krzyżanowski M, Kuna-Dibbert B, Schneider J. Health Effects of Transport-related Air Pollution. WHO Regional Office Europe; 2005.
20. Reimann S, Lewis AC. Anthropogenic VOCs. In: Koppmann R, ed. Volatile Organic Compounds in the Atmosphere. John Wiley & Sons, Ltd. 2007:33–81.
21. Agency for Toxic Substances and Disease Registry (ATSDR). Toxicological Profile for n-Hexane. U.S. Department of Health and Human Services, Public Health Service; 1999.
22. Amini H, Yunesian M, Hosseini V, et al. A systematic review of land use regression models for volatile organic compounds. Atmos Environ. 2017;171:1–16.
23. Hoek G, Beelen R, de Hoogh K, et al. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ. 2008;42:7561–7578.
24. Basagaña X, Rivera M, Aguilera I, et al. Effect of the number of measurement sites on land use regression models in estimating local air pollution. Atmos Environ. 2012;54:634–642.
25. Crouse DL, Goldberg MS, Ross NA, et al. Postmenopausal breast cancer is associated with exposure to traffic-related air pollution in Montreal, Canada: a case control study. Environ Health Perspect. 2010;118:1578–1583.
26. Goldberg MS, Labrèche F, Weichenthal S, et al. The association between the incidence of postmenopausal breast cancer and concentrations at street-level of nitrogen dioxide and ultrafine particles. Environ Res. 2017;158:7–15.
27. Parent M-E, Goldberg MS, Crouse DL, et al. Traffic-related air pollution and prostate cancer risk: a case–control study in Montreal, Canada. Occup Environ Med. 2013;70:511–518.
28. Statistics Canada. 2017. Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001. Ottawa. Released November 29, 2017. Available at: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E. Accessed 30 July 2021.
29. Government of Canada. *Canadian Climate Normals 1981-2010 Station Data*. 2021. Available at: https://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?searchType=stnProv&lstProvince=QC&txtCentralLatMin=0&txtCentralLatSec=0&txtCentralLongMin=0&txtCentralLongSec=0&stnID=5415&dispBack=0. Accessed 30 July 2021.
30. Réseau de surveillance de la qualité de l'air (RSQA). Environmental Assessment Report 2018 Air Quality Montreal. Ville de Montréal, Service de l'environnement, Division de la planification et du suivi environnemental. 2018.
31. Boulet D, Melançon S. Environmental Assessment Report Air Quality Montreal. Ville de Montréal, Service des infrastructures, du transport et de l'environnement, Direction de l'environnement, Division de la planification et du suivi environnemental, RSQA 2012;8. Available at: https://numerique.banq.qc.ca/patrimoine/details/52327/3289851?docref=f6IyJPE5DWnEegaISPAySg&docsearchtext=air%20quality%20montreal%202012
32. DMTI Spatial. *CanMap® GIS Data for GIS Mapping Software*. 2021. Available at: https://www.dmtispatial.com/canmap/. Accessed 19 August 2021.
33. Xie X, Semanjski I, Gautama S, et al. A review of urban air pollution monitoring and exposure assessment methods. ISPRS Int J Geo-Inf. 2017;6:389.
34. Ramos Y, Requia WJ, St-Onge B, et al. Spatial modeling of daily concentrations of ground-level ozone in Montreal, Canada: a comparison of geostatistical approaches. Environ Res. 2018;166:487–496.
35. PTV Vissim. *Traffic Simulation Software*. 2021. Available at: https://www.ptvgroup.com/en/solutions/products/ptv-vissim/. Accessed 19 August 2021.
36. United States Environmental Protection Agency. *MOVES and Other Mobile Source Emissions Models*. 2021. Available at: http://www.epa.gov/moves. Accessed 19 August 2021.
37. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: With Applications in R. Springer; 2013.
38. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. Stat Comput. 2014;24:997–1016.

39. Banerjee S, Gelfand AE, Carlin BP. Hierarchical Modeling and Analysis for Spatial Data. Monographs on Statistics & Applied Probability; 135. 2nd ed. Chapman & Hall/CRC Press; 2015.

40. Gamerman D, Lopes HF. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Texts in Statistical Science Series; 68. 2nd ed. Taylor & Francis; 2006.

41. de Valpine P, Paciorek C, Turek D, et al. NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling. 2021, R package version 0.12.2. Available at: https://cran.r-project.org/package=nimble.

42. Gu P, Dallmann TR, Li HZ, et al. Quantifying urban spatial variations of anthropogenic VOC concentrations and source contributions with a mobile sampling platform. Int J Environ Res Public Health. 2019;16:1632.

43. Yang Y, Liu X, Zheng J, et al. Characteristics of one-year observation of VOCs, NOx, and O3 at an urban site in Wuhan, China. J Environ Sci (China). 2019;79:297–310.

44. Su JG, Jerrett M, Beckerman B, et al. A land use regression model for predicting ambient volatile organic compound concentrations in Toronto, Canada. Atmos Environ. 2010;44:3529–3537.

45. Wheeler AJ, Smith-Doiron M, Xu X, et al. Intra-urban variability of air pollution in Windsor, Ontario--measurement and modeling for human exposure assessment. Environ Res. 2008;106:7–16.

46. Deville Cavellin L, Weichenthal S, Tack R, et al. Investigating the use of portable air pollution sensors to capture the spatial variability of traffic-related air pollution. Environ Sci Technol. 2016;50:313–320.

47. Amini H, Hosseini V, Schindler C, et al. Spatiotemporal description of BTEX volatile organic compounds in a Middle Eastern megacity: Tehran Study of Exposure Prediction for Environmental Health Research (Tehran SEPEHR). Environ Pollut. 2017;226:219–229.

48. Atari DO, Luginaah IN. Assessing the distribution of volatile organic compounds using land use regression in Sarnia, "Chemical Valley", Ontario, Canada. Environ Health. 2009;8:1–14.

49. Shields HC, Weschler CJ. Analysis of ambient concentrations of organic vapors with a passive sampler. JAPCA. 1987;37:1039–1045.

50. Stock TH, Morandi MT, Afshar M, et al. Evaluation of the use of diffusive air samplers for determining temporal and spatial variation of volatile organic compounds in the ambient air of urban communities. J Air Waste Manag Assoc. 2008;58:1303–1310.

51. Oiamo TH, Johnson M, Tang K, et al. Assessing traffic and industrial contributions to ambient nitrogen dioxide and volatile organic compounds in a low pollution urban environment. Sci Total Environ. 2015;529:149–157.

52. Lu T, Lansing J, Zhang W, et al. Land use regression models for 60 volatile organic compounds: comparing Google Point of Interest (POI) and city permit data. Sci Total Environ. 2019;677:131–141.

53. Liu K, Zhang C, Cheng Y, et al. Serious BTEX pollution in rural area of the North China Plain during winter season. J Environ Sci. 2015;30:186–190.

54. Walker TR. Effectiveness of the national pollutant release inventory as a policy tool to curb atmospheric industrial emissions in Canada. Pollutants. 2022;2:289–305.