Available online at www.sciencedirect.com
ScienceDirect
Atmospheric Environment 41 (2007) 7778–7787
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

Determinants of airborne benzene concentrations in rural areas of western Canada

Igor Burstyn[a,b,*], Xiaoqing (Isabelle) You[a],
Nicola Cherry[b], Ambikaipakan Senthilselvan[a]

[a]*Department of Public Health Sciences, The University of Alberta, Edmonton, Alta., Canada*
[b]*Community and Occupational Medicine Program, Department of Medicine, Faculty of Medicine and Dentistry, The University of Alberta, 13-103E Clinical Sciences Building, Edmonton, Alta., Canada T6G 2G3*

Received 2 November 2006; received in revised form 1 June 2007; accepted 9 June 2007

Abstract

This study estimated the level and determinants of airborne benzene concentrations in rural western Canada. A multi-site, multi-month unbalanced two-factorial design was used to collect air samples at 1206 fixed sites across a geographic area associated with primary oil and gas industry in Canadian provinces of Alberta, north-eastern British Columbia, and central and southern Saskatchewan from April 2001 to December 2002. Benzene concentrations integrated over 1 calendar month were determined using passive organic vapour monitors. Linear mixed effects models were applied to identify the determinants of airborne benzene concentrations, in particular the proximity to oil and gas facilities. The observed geometric mean of benzene concentrations was 158 ng m^{-3}, with large geometric standard deviation: 4.9. Benzene concentrations showed a seasonal variation with maxima in winter and minima in summer. Emissions from oil well (within 2 km) and compressor influenced monthly airborne benzene concentrations. However, in our study, being located in the general area of a gas plant seems to be the most important in determining monthly airborne benzene concentrations. These findings support the need for investigation of the impact of oil and gas industry on quality of rural air.

Keywords: Air pollution; Oil and gas industry; Source identification; Statistical modelling

---

1. Introduction

Benzene is a volatile organic compound (VOC) that has been classified as carcinogenic to humans by the International Agency for Research on Cancer (IARC, 1987). Environmental exposure to benzene is an important public health concern since low-level exposure is widespread in general population (Parra et al., 2006; Rappaport and Kupper, 2004) and there is considerable uncertainty about the existence of a threshold for toxic effects over long-term exposure (Austin et al., 1988; Glass et al., 2003). Benzene is ubiquitous, since it is a natural constituent of crude oil and natural gas, as well as of a wide variety of manufactured chemicals and fuels.

---

*Corresponding author. Community and Occupational Medicine Program, Department of Medicine, Faculty of Medicine and Dentistry, The University of Alberta, 13-103E Clinical Sciences Building, Edmonton, Alta., Canada T6G 2G3.
Tel.: +1 780 492 3240; fax: +1 780 492 9677.
E-mail address: iburstyn@ualberta.ca (I. Burstyn).

1352-2310/$ - see front matter
doi:10.1016/j.atmosenv.2007.06.011

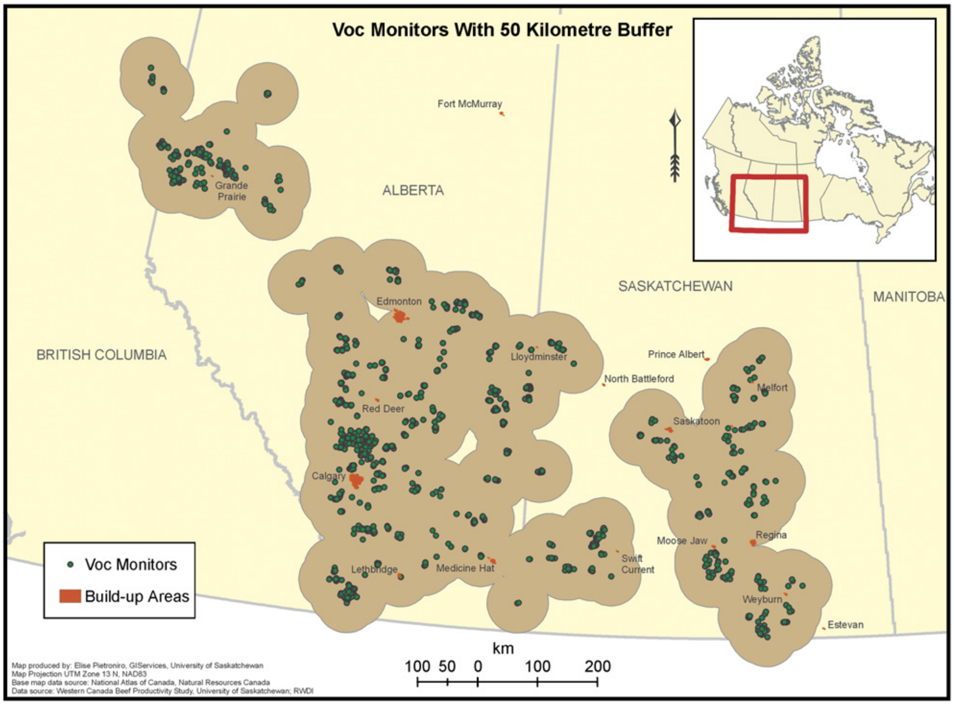
The upstream oil and gas industry plays a very important economic role in western Canada. This industry comprises all infrastructure used to find, produce, gather, treat/process and transport natural gas, liquefied petroleum gas, condensate, crude oil, heavy oil and crude bitumen to market. The benzene emission sources range from small fugitive leaks on valves and fittings to large point sources such as flaring, process vents and occasional well blowouts and pipeline release. The impact of emissions from upstream oil and gas industry on animal health raises substantial concerns in western Canada among beef cattle producers, because their pastures and primary oil and gas facilities are scattered across the rural areas and often overlap (Scott et al., 2003a). In response, the Western Interprovincial Scientific Studies Association (WISSA) initiated a study to evaluate the impact of exposure to emissions from oil and natural gas facilities on animal health; the focus of epidemiological study that produced exposure data was on health effects of low levels of exposure (WISSA, 2006). WISSA oversaw the design, funding and implementation of the overall project, including the collection and analysis of exposure information but only provided funding for the data preparation and statistical work that led to results presented in this manuscript. As a part of the study, exposure measurements for VOC were collected at fixed locations throughout the Canadian provinces of Alberta, north-eastern British Columbia, and central and southern Saskatchewan over a period from April 2001 to December 2002. We previously studied the effects of oil and gas infrastructure on measured mixtures of VOC (You et al., 2007); this paper specifically focuses on benzene, which was not adequately represented by measures of mixtures of VOC in our previous work.

The objective of this article is to determine how the proximity to oil and gas infrastructure and other anthropogenic sources of benzene may influence environmental concentrations of benzene in the air.

2. Methods and materials

2.1. Sampling strategy

Air samplers were located wherever cattle from the study herds were managed or pastured from April 2001 (205 herds) to June 2002 (203 herds), and—for a subset of 50 herds—to December 2002. The geographical distribution of sampling sites is shown in Fig. 1. Sampling devices were located to account for all ‘management groups’ within herds and therefore the numbers of monitoring locations



Map produced by: Elise Pietroniro, GIServices, University of Saskatchewan
Map Projection UTM Zone 13 N, NAD83
Base map data source: National Atlas of Canada, Natural Resources Canada
Data source: Western Canada Beef Productivity Study, University of Saskatchewan; RWDI

Fig. 1. Spatial distribution of benzene monitoring stations.

per herd differ. In a random sample of locations (~10%, ~120 locations, 1472 pairs in total), replicate measurements were collected for each month of the sampling campaign. All samples (including replicates) were collected and analysed using the same methods, described in detail below. Replicate monitors were located on the same fixed stand, in very close proximity to each other.

All monitors were set 1.5–1.8 m above the ground, at sites chosen according to the following criteria: (1) adjacent to areas within the pasture where study cattle spent most of their time; (2) away from minor sources of exposure, such as roads, farm equipment operated by internal combustion engines, >100 m from fuel and farm equipment storage areas, and local oil field equipment; (3) >10 m from roadways and other areas where vehicles were to be expected (e.g., gates where vehicles stopped and idled); (4) outside the immediate area of local oil and gas facilities, to avoid "worst-case" sampling; (5) at least 20 m from the nearest tree canopy, as defined by drip line; (6) away from buildings, hay storage and other objects that may obstruct airflow; (7) in flat terrain (i.e., not at the tops or bottoms of hills).

2.2. Sampling devices and chemical analyses

Airborne vapours were collected by exposing 3 M OVM 3500 badges for 1-month periods. Benzene was extracted from exposed and blank sampling media with 1.5 ml of pre-cleaned carbon disulphide. The extraction solvent was previously spiked with deuterated standard at nominal concentrations of 2 ng l^{-1} to assess extraction efficiency and quality of analyses. The extracts were transferred by glass Pasteur pipettes to auto-sampler amber glass vials for quantification. Benzene concentration were determined by injecting (split mode, 16:1) 2 μl aliquots of extracts, standard solutions and carbon disulphide into the GC-MSD (HP 5890 Series II GC; 5972 A MSD; HP7673 auto-injector; Hewlett Packard Co. Inc., Palo Alto, CA) operated in the selected ion monitoring mode (3 ions/compound). The GC was equipped with a DB624 column (60 m long, 0.32 mm I.D., 1.8 mm film thickness, J&W Scientific, Folsom, CA). The temperature program for resolution of the target compounds was: 4 min hold at 40 °C; ramp at 10 °C $\min^{-1}$ to 220 °C; ramp at 20 °C $\min^{-1}$ to 250 °C and hold for 2 min. For identification and quantification, one quantification (Q) and two confirmation (C) ions were monitored. Generally, a retention time window of ±2.5%, and a tolerance of ±25% for abundance ratios of Q and C ions were the confirmation criteria, and the external standard method was used to determine the concentration of quantification ions in the samples. Extract volumes, corrections for benzene in extracts, and the extraction recovery efficiency (95%) were used in quantification calculations, as was the airborne sampling rate (38.6 ml $\min^{-1}$). Deviations from standard application of the vapour monitors, including airborne sampling rate, were examined under experimental conditions and it was observed that sampling rates were insensitive to temperature, relative humidity and air velocity (Fellin et al., 1989). *Sampling time* can adversely affect measurement of OVM if capacity of sampling media is exceeded or sampled material re-volatilizes. Since overall concentrations measured in the study were well below occupational exposure levels for which OVM was designed, capacity should not affect performance of OVM in this study. Re-volatilization was evaluated by Airzone One Ltd. in experimentally generated atmospheres over a wide range of environmental conditions: temperature (−5 to 30 °C), relative humidity (10–95%), wind (stagnant to >3.6 m s^{-1}) and competing compounds (VOCs in concentrations of 1–200 μg m^{-3}). It was observed that there was little evidence of re-volatilization under these conditions. The same set of chamber tests (results of which were not available to the authors), revealed that 38.6 ml $\min^{-1}$ *sampling rate* is appropriate for benzene under these conditions and this seems to be the most appropriate method for selecting sampling rate (as compared to using sampling rate provided by 3 M for monitoring occupational, rather than environmental exposures). For example, in one set of chamber tests with 22 pairs of 3 M OVMs and active samplers with charcoal tubes, the correlation between benzene concentrations determined by active and passive samplers was 0.99 (Phil Fellin, personal communications, 2007).

None of measurements of benzene concentration were censored for the limit of quantification, and actual instrument readings were used in calculating air concentration. Replicate field samples were used for quality assurance and control (same as repeated measurements at a given site and month in statistical modelling). Sampling and analytical procedures were developed and implemented by AirZone One Inc. (Mississauga, Ont., Canada), a commercial laboratory contracted by WISSA to analyse VOC samples.

2.3. *Documenting determinants of benzene concentrations*

Field technicians contracted by RWDI Air Inc., Calgary, Canada, documented nearby potential sources of emissions around sampling sites. Presence of minor sources located within 2 km of the pastures and major sources located within 5 km of the pastures were documented by visual observations. Out of a total of 1241 sites, 432 had no industrial source documented; 87 had no information about nearby potential sources, but other information such as site terrain and tree canopy were documented; 35 had no data. These 35 sites and associated measurements were excluded from modelling determinants of benzene concentrations. The 87 sites that had missing information about nearby potential sources were coded as if there were no industrial sources of benzene. From free text entries of field technicians, we classified the nearby potential industrial sources into 10 categories. Classification of sources and their description by field technicians (verbatim) were: well (well, wells, gas well, oil well, wellhead, drilling rig, drilling well, pump, pump jack and pump tank), plant (gas plant, gas processing plant, plant), battery (battery, battery tank, batteries), compressor (compressor), flare (flaring, flare stack, flare), other oil and gas facilities (heater, pipeline riser, gas tank, diesel tank(s), diesel, purple clear gas, gas risers, gas valve, pipeline, tank(s), tank farm), X'mas tree (clarification by the authors: vertical arrangement of valves and controls), gas operations, many oil activity, oil field, pump station, highway (highway(s)), road (road, gravel road, railway (just used once or twice a week)), other industries (power sub-station, industrial park, town, city, energy corporation, power transformer, shop, welding shop) and non-industrial source (acreage(s), forest reserve, yard, farmyard, power line, electric line, irrigation pump, lease site, pit, slough, water well, hay field, rittle range, river, salt lick, silage operation, water pipeline, farm(s), farmstead, homestead, corral(s), dugout, barn(s), farm erwip, feed lot, home, loading corrals, quarters, school, Sask Tel tower, bale stack, grain bins, grain terminal, gravel pile, gun club, hat stack, hay bales, microwave tower(s), shacks, Telus tower, tin shed).

Data on location and type of oil and gas facilities in 2001 and 2002 were supplied by the provincial regulatory (government) agencies. Coordinates of monitoring stations and oil and gas facilities were used to estimate distances between each putative source and monitoring station. Effects of sources on measured concentrations were considered for two concentric distance classes around monitoring stations: 0–2 and 2–50 km. The following types of sources were considered separately: oil wells, gas wells, bitumen wells, batteries and gas plants. Only facilities classified as 'active' by the regulatory agencies in a given year were considered in analysis.

For each type of source, distance class and year, we developed a measure of proximity of monitoring stations to facilities in oil and gas industry in the following manner. The weighted sum of distances of similar sources in a given distance class were computed as sum of 1/(distance)$^{2/3}$, as in the dispersion model of Strosher (1996), which assumes that concentration of air pollutants in the air is directly proportional to emission rate and 1/(distance)$^{2/3}$. Only those types of sources that could be identified in all provinces were used in subsequent analysis (except for bitumen wells, which were assumed to exist exclusively in Alberta). Thus, proximity (P_{ijkl}) of kth monitoring station $(1,\ldots,K)$ at lth time $(1,\ldots,L)$ to s sources $(1,\ldots,S_{ijkl})$ of ith type $(1,\ldots,I)$ within jth distance class $(1,\ldots,J)$ can be quantified as

$$P_{ijkl} = \sum\nolimits_{\text{all}\ s\,(ijkl)} [D_{s(ijkl)}^{-(2/3)}], \tag{1}$$

where D is distance in kilometres (km) from the source to monitoring station, and each sum is specific to a fixed combination of source type, distance class, location and year of air quality measurement. If a value of proximity could not be determined, as when there was no bitumen wells within 2 km of a monitoring station, P_{ijkl} was set to zero.

2.4. *Identifying predictors of benzene concentrations*

The following initial mixed effects model was used with fixed seasonal effect and random effects of month/time, sampling location, and replicates at some locations and months:

$$Y_{jm(hi)} = \mu_y + \delta\,\text{Season}_m + \tau_h + \lambda_i + \varepsilon_{j(hi)}. \tag{2}$$

$Y_{jm(hi)} = \log_e$(benzene concentration) the jth replicate at the ith sampling site in the hth consecutive month (not month nested in a year) and mth season; $\mu_y =$ the true unknown mean of $\log_e$(benzene concentration); $\delta =$ the regression coefficient (fixed effect) for the dummy variable 'season' (winter: November–April or summer: May–October); τ_h = the random effects of the *h*th month; λ_i = the random effects of *i*th sampling site; $\varepsilon_{j(hi)}$ = the random effect of the *j*th replicate nested in the *i*th sampling site and the *h*th month. It was assumed that $\tau_h \sim N\ (0, \sigma_m^2)$, $\lambda_i \sim N\ (0, \sigma_s^2)$, and $\varepsilon_{j(hi)} \sim N\ (0, \sigma_r^2)$, and that these are mutually independent. The ratio of 97.5th to 2.5th percentiles of estimated frequency distribution ($R_{0.95}$) of the benzene concentrations between repeats, locations, and months were estimated by $\exp(3.92\sigma_r)$, $\exp(3.92\sigma_s)$, and $\exp(3.92\sigma_m)$, respectively (Rappaport, 1991).

The empirical models of the determinants of benzene concentration were constructed in several steps. First, we added each fixed effect to the initial mixed effects model (Eq. (2)) one at a time. We chose proximity scores of oil wells, gas wells, bitumen wells, batteries, and gas plants from the regulatory data, and flare, other oil and gas facilities, other industry, highway, road, and vegetation from reports of field technicians as potential determinants of benzene concentrations. We considered only predictor variables that may be reasonably expected to be associated with benzene concentrations. In order to use functions of proximity scores as fixed effects in statistical models, we log-transformed them. Furthermore, a constant value of 1 was added to all proximity scores (P_{ijkl}) before log-transformation to avoid the problem with logarithm of 0 being undefined. (For the sake of simplicity, we will refer to $\log_e(1 + P_{ijkl})$ as 'proximity' in describing statistical model.) We evaluated statistical significance ($p \leqslant 0.05$) of effect estimates in these models.

Next, we evaluated final mixed effects models of the following general form:

$$Y_{jm(hi)} = \mu_y + \sum\nolimits_{\text{all } w} \beta_w X_{whi} + \delta\,\text{Season}_m + \tau_h + \lambda_i + \varepsilon_{j\,(hi)}, \tag{3}$$

where β_w represents estimates of fixed effect of *w*th (1, …, *W*) determinant of benzene concentration, and X_{whi} represents values of fixed effects of the determinant on *h*th consecutive month (not month nested in a year) at *i*th site; $Y_{jm(hi)}$, μ_y, δ, τ_h, λ_i, $\varepsilon_{j(hi)}$ have the same meaning as in Eq. (2).

Only variables reflecting proximity to sources that were significant on the first step (i.e., after being added to Eq. (2)) were considered further. Two types of statistical models were constructed, focusing separately on nearby and distant sources of emissions. The first modelling strategy focused on identification of the effects of sources that were located close to monitoring stations. It used proximity scores in <2 km distance classes and sources reported by field technicians. The second modelling strategy aimed to identify effect of sources that were further away from monitoring stations and therefore considered only proximity scores from 2 to 50 km distance classes. In building these models, correlations among predictor variables were examined as indicators of co-linearity. Pearson correlation coefficient and phi correlation coefficients were computed, for continuous and dichotomous predictors, respectively. If two variables had correlation of at least 0.7, only the one that produced greater improvement in overall model fit (as judged by Akaike Information Criterion (AIC)) was included in the final model. The predictors which were statistically significant were retained in the final models. Significant proximal and distant predictors from the two parallel modelling strategies were evaluated together in the final model and only significant predictors retained. Assumptions of mixed effects model were examined through plots of the residuals. All statistical analyses were implemented in SAS software (SAS Institute Inc., Cary, NC).

3. Results

The number of air samplers deployed ranged from 145 to 1031 per month and there were 117–947 sampling sites per month. A total of 11,399 air samples (13% of them were replicates) were collected. Wells (including oil well, gas well) were the most common type of oil and gas facility reported by field technicians in the vicinity of monitoring stations, followed by gas plant, flare and battery (Table 1). For 24 measurements information on sources recorded by field technicians was missing (Table 1), consequently analyses were restricted to the complete data set of 11,375 measurements.

Air concentrations of benzene were described reasonably well by lognormal frequency distribution. The geometric mean of the distribution of benzene concentrations was $158\,\text{ng}\,\text{m}^{-3}$, somewhat smaller than the arithmetic mean ($288\,\text{ng}\,\text{m}^{-3}$). The observed concentrations were quite variable, ranging from <0.5 to $9036\,\text{ng}\,\text{m}^{-3}$; geometric standard deviation was 4.93. Only 5.3% of measurements were non-detectable ($<0.5\,\text{ng}\,\text{m}^{-3}$). Among replicates, coefficient of variation was 101%.

Table 1
Frequencies of potential determinants of concentration and observations (samples)

Source type (present/absent)	Count of sites	%	Count of samples	%
No industrial sources documented			4695[a]	41.19
No industrial source	432	34.81		
Source information missing	87	7.01		
No data	35[b]	2.82	24	0.2
Battery	87	7.01	1079	9.47
Compressor	56	4.51	609	5.34
Flare	110	8.86	1350	11.84
Highway	74	5.96	831	7.29
Other industry	26	2.1	282	2.47
Other oil and gas facilities	57	4.59	667	5.85
Plant	124	9.99	1313	11.52
Road	218	17.57	1841	16.15
Well	311	25.06	3182	27.91

Total documented sites = 1241, total observations = 11,399.

[a] The number of samples measured at sites with no industrial sources and at sites with source information missing.

[b] Of the 35 sites, 6 sites had a total of 24 measurements, but no data about sites, so these 24 observations were excluded from regression modelling and other 29 sites did not have any measurements.

Table 2
Variance components of the initial benzene model with season as the only fixed effect ($N = 11{,}375$)

Variance component (s^2)	Estimate	Standard error	p-Value	$R_{0.95}$[a]
Between locations	0.293	0.021	<0.0001	8
Month-to-month (within a season)	0.372	0.121	0.0011	11
Between repeats[b]	1.453	0.020	<0.0001	113

[a] $R_{0.95}$ represents the fold range, includes 95% of the values and $R_{0.95} = \exp(3.92S)$.

[b] Repeats that were collected at the same location in the same month.

The dominant source of random variation in benzene levels was due to discordance among repeated measurements at the same location and on the same month (Table 2). There was on average a factor of 100 differences among the replicates. Benzene concentrations were significantly higher in winter/cold months than warmer seasons (estimated fixed effect = 1.23, standard error (S.E.) = 0.27, $p<0.0001$; from the initial model in Table 2). After we corrected random effects for seasonal variation, month-to-month differences of approximately one order of magnitude remained. The spatial component of overall variability was the smallest, but still indicated residual difference of a factor of eight among locations in the same season and month.

Both distant (2–50 km) and nearby (<2 km) sources appear to affect measured benzene concentrations (Table 3). None of the predictor variables in <2 km distance class appeared to be appreciably correlated among themselves and all significant ones from Table 3 were considered in multivariate model. The proximity of oil wells, batteries and compressors appears to affect benzene concentrations in the air when only nearby facilities were considered (Table 4). The following estimates of random effects were obtained for the model that only considered nearby sources: between-location variance of 0.270 (S.E. = 0.020), month-to-month variance of 0.372 (S.E. = 0.118) and between-repeat variance of 1.453 (S.E. = 0.021).

Among variables retained from step 1 for analysis of impact of distant sources, we observed correlation suggestive of co-linearity among proximities scores in 2–50 km distance class for batteries with oil wells ($r = 0.8$), gas plants ($r = 0.8$) and ‘all large facilities’ ($r = 0.9$); ‘all large facilities’ with oil wells ($r = 0.7$) and gas plants ($r = 0.9$). All large facilities, a less specific variable compared to ‘gas plants’ was

Table 3
Initial analyses of determinants of natural logarithms of airborne benzene concentration (in $\mu g\,m^{-3}$) one at a time using mixed effects model[a] ($N = 11{,}375$)

Potential predicator	Estimate	Standard error	p-Value
Logarithmic transformations of the proximity scores in units of $\log_e(1+\sum[\text{distance(km)}^{-2/3}])$, see Eq. (1)			
Oil well <2 km	0.800	0.174	<0.0001
Oil well 2–50 km	0.348	0.028	<0.0001
Gas well <2 km	0.619	0.429	0.149
Gas well 2–50 km	0.208	0.035	<0.0001
Bitumen well 2–50 km[b]	0.267	0.351	0.447
Battery <2 km	7.559	1.195	<0.0001
Battery 2–50 km	1.217	0.073	<0.0001
Gas plant <2 km	12.710	10.067	0.207
Gas plant 2–50 km	22.632	1.246	<0.0001
All large facilities[c] <2 km	9.927	3.979	0.013
All large facilities[c] 2–50 km	6.629	0.325	<0.0001
Present/absent			
Flare	0.197	0.067	0.0036
Highway	0.089	0.083	0.2841
Other oil and gas industries	0.168	0.094	0.0717
Other industries	−0.158	0.140	0.2607
Road	−0.097	0.054	0.074
Compressor	0.397	0.095	<0.0001

[a] Each determinant was added one at a time to the initial mixed effects model described by Eq (2).

[b] There were no bitumen wells within 2 km of the studied locations.

[c] All large facilities include gas plants.

Table 4
Mixed-effects models of natural logarithms of benzene concentrations (in $\mu g\,m^{-3}$) that considered nearby and distance sources separately ($N = 11{,}375$)[a]

Model	Effect	Estimate	Standard error	p-Value
Nearby sources	Intercept	−2.262	0.178	<0.0001
	Winter[b]	1.229	0.270	<0.0001
	Oil well <2 km[c]	0.391	0.190	0.0401
	Battery <2 km[c]	5.925	1.333	<0.0001
	Compressor (yes/no)	0.339	0.093	0.0003
Distant sources	Intercept	−1.281	0.204	<0.0001
	Winter[b]	1.233	0.267	<0.0001
	Gas plant 2–50 km[c]	22.632	1.246	<0.0001

[a] Results from the multivariate mixed effects model described by Eq. (3).

[b] Winter, a dichotomous variable indicating 1 for Winter and 0 for summer, was defined as the period from November to April, while summer was defined as the period from May to October.

[c] Logarithmic transformations of the proximity scores in units of $\log_e(1+\sum[\text{distance(km)}^{-2/3}])$, see Eq. (1).

excluded from further consideration. For the 2–50 km distance class, batteries, oil wells and gas plants were offered one at a time to determine which one of them produced greater improvement in model fit; the proximity score for gas wells was initially offered in all models (but see below). Inclusion of proximity score for gas plants in 2–50 km distance class resulted in the lowest value of AIC (37524.8) compared to models with oil wells (AIC = 37675.4) and batteries (AIC = 37565.8), indicating that that proximity score for gas plants fit the data better. Fixed effect estimate for the proximity score for gas well (2–50 km distance class) among the three alternative model exhibited erratic behaviour: it was not significant in two of the models (including the one with the proximity score for the gas plants) and changed direction of effect among models; it was consequently excluded from the final model for 2–50 km distance class (Table 4). The model of the impact of distant sources had between-location variance of 0.174 (S.E. = 0.016), month-to-month variance of 0.362 (S.E. = 0.118) and between-repeat variance of 1.460 (S.E. = 0.021).

In deriving a final model of benzene concentration we combined predictors in Table 4 from the two models. Fixed effect of proximity to batteries in <2 km distance class lost statistical significance (fixed effect = 0.750, S.E. = 1.226, $p = 0.5$). Therefore, we excluded it from the final model (AIC = 37511.5) presented in Table 5. The following estimates of random effects were obtained for the final model: between-location variance of 0.171 (S.E. = 0.015), month-to-month variance of 0.362 (S.E. = 0.118) and between-repeat variance of 1.460 (S.E. = 0.021). As expected, predictor variables describing oil and gas infrastructure variables only contributed to explaining the spatial variability of benzene concentrations, reducing unexplained between-location variance from 0.293 to 0.171, i.e., explaining 42% of spatial variability; the variance of other two random effect estimates remained unchanged. It would appear that presence of gas plants in the general area of monitoring stations was the most important explanatory variable for benzene concentrations: it had the highest fixed effect estimate and it alone accounted for much of the between-location variance explained by the final model. Residual plots did not indicate severe violations of the assumptions underlying the models.

Table 5
Final mixed-effects model of natural logarithms of benzene concentrations (in $\mu g\,m^{-3}$), ($N = 11{,}375$)[a]

Effect	Estimate	Standard error	p-Value
Intercept	−1.298	0.204	<0.0001
Winter[b]	1.235	0.267	<0.0001
Oil well <2 km[c]	0.494	0.148	0.0009
Gas plant 2–50 km[c]	21.875	1.255	<0.0001
Compressor (yes/no)	0.187	0.082	0.02

[a] Results from the multivariate mixed effects model described by Eq. (3).

[b] Winter, a dichotomous variable indicating 1 for winter and 0 for summer, was defined as the period from November to April, while summer was defined as the period from May to October.

[c] Logarithmic transformations of the proximity scores in units of $\log_e(1+\sum[\text{distance(km)}^{-2/3}])$, see Eq. (1).

4. Discussion

We observed that even though environmental concentrations of benzene in the rural air in western Canada are both low and variable, they can be linked to emissions from oil and gas facilities that are common in the studied areas. Gas plants appear to be among the most prominent sources, exerting influence over at least a 50 km range. However, much of the variability in observed benzene concentrations remained unexplained, raising the possibility that we did not capture all important sources of benzene emissions among the potential determinants and/or that the determinants of exposure were assessed with large errors.

In a related paper arising from the same project (You et al., 2007), individual VOC were combined through factor analysis, with benzene concentrations amalgamated into a factor score represented by one of the factors. Other VOC that were represented by that factor were, in order of importance: ethylbenzene, xylenes, toluene, trimethylbenzenes, hexane, cumene, and benzene. The extent to which factor score represented benzene was poor: factor loading (correlation of factor score with benzene) of 0.52, the lowest of the group of VOC that was considered to be characteristic of that factor. Previously, we identified the presence of oil and gas wells, batteries, highways and roads, 'various' industries and winter season as being associated with the factor score that included benzene. These results are similar, though not identical, to those presented here as significant determinants of benzene concentration: presence of compressors seems to be only important for predicting benzene levels and not for other VOC in these data; in contrast to our previous analysis, the presence of batteries does not appear to be a strong predictor of benzene concentration, but here the proximity to gas plants emerged as an important correlate of elevated benzene levels.

It should be noted that it was estimated that industrial sources contributed only 6% of total benzene emissions in Canada in 1991, while the major source of benzene pollution was light-duty vehicles (67% of 22,000 tonnes) (Solomon et al., 1993). Nonetheless, it is not surprising to find that facilities that are associated with extraction and gathering of crude oil, a natural source of benzene, would be linked to elevated benzene levels. The most plausible mechanism for this is low-intensity fugitive emissions that do not register as process upsets. It should be noted that a decrease in benzene concentrations near industrial sites (including refineries) has been observed between 1989 and 1993 (Dann and Wang, 1995).

Greater benzene concentrations during winter months in rural areas in Canada have been previously reported, suggesting that the 24-h benzene concentrations are a factor of 4–8 higher in January and February then in June and July (Dann and Wang, 1995), which is similar to a three- to four-fold difference in monthly benzene concentrations between cold and warm seasons in our study. This seasonal trend is typically attributed to greater abundance in summer of OH radical that degrades benzene (US Environmental Protection Agency, 1993).

Observed benzene concentrations are similar to those reported at three rural monitoring sites in Canada between 1988 and 1993: 100–2000 ng m^{-3} with median of 500 ng m^{-3} (Dann and Wang, 1995). Two of these sites were within 50 km of urban areas and this may explain overall elevated benzene concentrations relative to geometric mean observed in our data. For monitoring stations located with 3 km of refineries, benzene concentrations between approximately 500 and 3000 ng m^{-3} were reported in down-wind 24-h samples (Dann and Wang, 1995), which is in agreement with our results that suggest an influence of oil and gas facilities on benzene concentrations. Our study did not capture information on wind direction relative to studied sources. Overall, short duration of sampling in the Environment Canada study (Dann and Wang, 1995) (24 h) makes it difficult to compare to our results, obtained from month-long samples. For comparison of the concentrations observed in this study with regulator standards, we computer location-specific arithmetic annual means of benzene concentrations ($n = 2068$): mean of means: 253 ng m^{-3}, median of means: 206 ng m^{-3}, standard deviation of means: 225 ng m^{-3}, minimum: 5 ng m^{-3}, maximum: 5 326 ng m^{-3}.

Our work suffers from numerous limitations. As previously mentioned, we lack data on many potential determinants of benzene concentrations, such as climatic information, prevailing wind direction, precipitation and topography around the monitoring sites. Sampling strategy aimed to reduce the influence of some of these factors, by specifying in what terrain monitoring stations should be placed. Some of the features of the locations of monitoring stations would have been captured by random 'location' effect in the models. Likewise, some of the features of climate and wind patterns would have been captured by random month effects and fixed effect of season. This would reduce the likelihood of confounding of the effects of measures of proximity to sources of emission. In addition, better characterization of strength of sources of emissions may also have help to explain spatiotemporal variability in the observed concentrations. In constructing measures of proximity to sources, the split point (2 km) and maxima (50 km) were selected based on the consultations with WISSA Science Advisory Panel. There are neither strict rules nor empirical data that would guide such selection and this in fact may be one of the sources of uncertainty in our results. However, dispersion models employed to study SO_2 emission from point sources in oil and gas industry in western Canada show no decay in concentrations within 1 km and reduce to background level at about 70 km (Scott et al., 2003b). Thus, our choice of distance classes in somewhat inconsistent with previous work (albeit with a different pollutant). We only had information on whether a potential source was active in a given calendar period, but it is possible that some industrial facilities shut down and then restarted within the study period. Our models would not reflect this and therefore this source of error in our data remains uncontrolled.

Environmental measurements used in modelling were imprecise, with between-repeat variance dwarfing spatiotemporal variability. This emphasises the need to correct partitioning of variability through mixed effects models we employed. Errors in the dependent variable should not bias estimates of associations with determinants of exposure, but would adversely affect precision of predicted concentrations if our models are directly applied to exposure assessment.

This project is the only large-scale study into determinants of environmental benzene concentrations in rural area with diverse and numerous primary oil and gas industry. Our results suggest that there is a detectable impact of primary oil and gas industry on quality of rural air. Given the recognized toxicity of benzene, any possible human exposures due to modifiable sources, as appears to be the case when oil and gas infrastructure intermingles with farmland, should be carefully investigated with a view towards conducting rigorous risk assessment.

Acknowledgements

The study would not have been possible without funding from the Western Interprovincial Scientific Studies Association, RWDI Air Inc. that collected air samples, and AirZone One Ltd. that conducted laboratory analyses. Phil Fellin of AirZone One Ltd. was most helpful in clarifying the details of sampling and analytical methodology for this publication. Guidance from an independent Scientific Advisory Board (Drs. Tee Guidotti, Ole Nielsen, Kiros Berhane, Beverly S. Cohen, Bruce Hunter, Bill Lasley, Wayne Martin, Carl Ribble, Peter Thorne, David Tollerud, Hanspeter Witchi) was invaluable and we are thankful to them for overseeing data collection and survey design which occurred before we became involved in the project. We also wish to thank Alberta Energy and Utilities Board, Oil and Gas Commission of British Columbia and Petroleum Development Branch of Government of Saskatchewan who provide regulatory data on location and type of oil and gas facilities. Elise Pietroniro and Cheryl Waldner at the University of Saskatchewan earned our gratitude for computing distances between air monitoring stations and oil and gas facilities and creating the map of monitoring sites. Cheryl Waldner is also the principal investigator of the epidemiological (animal health) project for which air quality measurements were collected. Finally, we thank Jianduan Liang and Narmatha Thanigasalam at the University of Alberta who created SAS program that calculated proximity scores.

References

Austin, H., Delzell, E., Cole, P., 1988. Benzene and leukemia. A review of the literature and a risk assessment. American Journal of Epidemiology 127, 419–439.

Dann, T.F., Wang, D.K., 1995. Ambient air benzene concentrations in Canada (1989–1993): seasonal and day of week trends, and source influences. Journal of Air and Waste Management Association 45, 695–702.

Fellin, P., Otson, R., Ernst, D.L., 1989. A versatile system for evaluation of organic vapour monitoring methods. In: Proceedings of the Eight World Clean Air Congress, vol. 3. The Hague, The Netherlands, September 1989, pp. 675–680.

Glass, D.C., Gray, C.N., Jolley, D.J., Gibbons, C., Sim, M.R., Fritschi, L., Adams, G.G., Bisby, J.A., Manuell, R., 2003. Leukemia risk associated with low-level benzene exposure. Epidemiology 14, 569–577.

IARC, 1987. Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs, vols. 1–42. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Suppl. 7. International Agency for Research on Cancer, Lyon.

Parra, M.A., Gonzalez, L., Elustondo, D., Garrigo, J., Bermejo, R., Santamaria, J.M., 2006. Spatial and temporal trends of volatile organic compounds (VOC) in a rural area of northern Spain. The Science of the Total Environment 370, 157–167.

Rappaport, S.M., 1991. Assessment of long-term exposures to toxic substances in air. Annals of Occupational Hygiene 35, 61–121.

Rappaport, S.M., Kupper, L.L., 2004. Variability of environmental exposures to volatile organic compounds. Journal of Exposure Analysis and Environmental Epidemiology 14, 92–107.

Scott, H.M., Soskolne, C.L., Martin, S.W., Shoukri, M.M., Lissemore, K.D., Coppock, R.W., Guidotti, T.L., 2003a. Air emissions from sour-gas processing plants and dairy-cattle reproduction in Alberta, Canada. Preventive Veterinary Medicine 57, 69–95.

Scott, H.M., Soskolne, C.L., Wayne, M.S., Ellehoj, E.A., Coppock, R.W., Guidotti, T.L., Lissemore, K.D., 2003b. Comparison of two atmospheric-dispersion models to assess farm-site exposure to sour-gas processing-plant emissions. Preventive Veterinary Medicine 57, 15–34.

Solomon, P., Davis, C., Tharby, R., 1993. Programs and Strategies to Minimize and Prevent Benzene Releases from Industrial, Commercial and Transportation Sources in Canada (Report IP-132). Environment Canada, Ottawa, Ont.

Strosher, M.T., 1996. Investigations of flare gas emissions in Alberta. Final Report to: Environment Canada Conservation and Protection, the Alberta Energy and Utilities Board and the Canadian Association of Petroleum Producers. Alberta Research Council, Calgary, Alta.

US Environmental Protection Agency, 1993. Motor Vehicle-Related Air Toxics Study (EPA 420-R-93-005). EPA Office of Mobile Sources, Emission Planning and Strategies Division, Ann Arbor, MI.

WISSA, 2006. Western Canada Study of Animal Health Effects Associated with Exposure to Emissions from Oil and Natural Gas Field Facilities: Research Appendices. Calgary, Alta., Canada.

You, X., Senthilselvan, A., Cherry, N., Kim, H.-M., Burstyn, I., 2007. Determinants of airborne concentrations of volatile organic compounds in rural areas of Western Canada. Journal of Exposure Science and Environmental Epidemiology (published online Feb 25, 2007; http://dx.doi.org/10.1038/sj.jea.7500556).