Modeling the Shape of the Dependency of Airborne Benzene Concentration in the Air on Distance to Primary Oil and Gas Facilities

Irina Dinu[*,1], Yan Chen[1] and Igor Burstyn[2,3]

[1]*Department of Public Health Sciences, School of Public Health, University of Alberta, Edmonton, Alberta, Canada*

[2]*Community and Occupational Medicine Program, Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada*

[3]*Departement of Environmental and Occupational Health, School of Public Health, Drexel University, Philadelphia, Pennsylvania, USA*

Abstract: The level and determinants of airborne concentrations were estimated by collecting air samples at 1206 fixed sites across a geographic area associated with primary oil and gas industry in the rural western Canada, in the provinces of Alberta, north-east British Columbia, and central and southern Sasketchewan from April 2001 to December 2002. Benzene concentrations integrated over one calendar month were determined using passive organic vapor monitors. Previous work applied linear mixed effects models to identify the determinants of airborne benzene concentrations, in particular the proximity to oil and gas facilities. We present results of a more flexible model using cubic splines to accommodate nonlinearities in the effects of determinants of airborne benzene concentrations, as well as time. Benzene concentrations exhibited monotonically increasing time trends for the months from July through December, and monotonically decreasing time trends corresponding to the months from December to July. We illustrated here how cubic splines can be used to identify complex relations between proximity to point sources of air pollution and observed extent of contamination, during the study period, and identified batteries as an important source of benzene emissions that was missed in previous analysis of the same data. These findings contribute to better understanding how positioning oil and gas facilities impacts air quality.

Keywords: Air pollution, oil and gas industry, source identification, cubic splines.

INTRODUCTION

Benzene is a colorless, flammable liquid that easily volatilizes. It is produced from both coal and petroleum sources and is naturally present in crude oil. Based on classification schemes to evaluate exposures that may be carcinogenic, the International Agency for Research on Cancer [1], the Environmental Protection Agency [2], and The National Toxicology Program [3] rated benzene in their highest carcinogenic class. Low level exposure to benzene is widespread in the general population, making it an important public health concern [4,5]. Also, there is considerable uncertainty on a threshold for toxic effects over prolonged exposure [6,7].

The impact of emissions from oil and gas industries on animal health is a major concern in western Canada, among beef cattle producers, as their pastures and primary oil and gas facilities often overlap [8]. A study to evaluate the impact of exposure to emissions from oil and natural gas facilities on animal health was initiated by the Western Interprovincial Scientific Studies Association (WISSA). The epidemiological study focus was on health effects of low levels of exposure [9]. WISSA oversaw the design, funding

and implementation of the project, including collection and analysis of exposure information [10-12]. WISSA also provided funding for some of the data manipulation presented in this manuscript.

The goal of this paper is to model benzene concentration as a function of distance to sources of emissions around sampling sites, adjusting for monthly effects and repeated sampling at the same location and month, in the context of the WISSA study introduced above. Exposure to benzene is of interest because in the WISSA study it was observed that exposures to benzene in excess of 0.236 $\mu g/m^3$ (month-long average) were associated with a small increase in time-to-pregnancy (a measure of sub-fertility) in cattle [13]; it must be noted that relevance of these findings to human health in the affected region is currently unclear. Consequently, there is some interest in refining our understanding of factors that influence exposure to benzene in this context. Previous work [14], assumed that the logarithm of benzene concentration depends linearly on the function of the proximity scores of monitoring stations to sources of emissions [15]. As a result, the concentration of benzene depended on the distance to the source by a function that was forced to resemble exponential decay, which seems reasonable especially for those sources emitting benzene at ground level. However, for those sources which are not emitters at ground level (e.g. from a stack or a pipe on an elevated structure) the benzene concentration peaks at a certain distance, and then decreases as we get further away from the source. In such cases, the assumptions

*Address correspondence to this author at the Department of Public Health Sciences, School of Public Health, University of Alberta, Edmonton, Alberta, Canada; Tel: (780) 492-8336; Fax: (780) 492-0364; E-mail: idinu@ualberta.ca

that the logarithm of benzene concentration depends linearly on the distances from sources, and that the benzene concentration follows an exponential decay function are no longer appropriate. Failure to model nonlinearities in the effect of a covariate may result in missing its effect on the response [16]. We propose to use cubic splines [17] to relax this assumption about the shape of the relationship, resulting in a more flexible form of the dependence of benzene concentration on distances from sources. Furthermore, previous work [14] modeled the effect of time by using a binary summer/winter type of variable. We use cubic splines to model time through a continuous function, allowing more flexible control for this covariate and thereby more effectively adjusting for its influence on the main association of interest: impact of the proximity to oil and gas infrastructure on the extent of air pollution by benzene.

MATERIALS AND METHODS

Study Design and Methods

This large scale project gives a unique setting for studying determinants of environmental benzene concentrations in rural area, with a variety of facilities of primary oil and gas industries. Detailed descriptions of the sampling strategy, sampling devices and chemical analyses, and determinants of benzene concentration are provided elsewhere [10,14]. Briefly, replicated benzene concentrations were available from April 2001 to December 2002, at various monitoring stations scattered throughout oil and gas producing region of western Canada that also overlapped with cattle ranching, please see Fig. (**1**) for a map of the area (in the figure, VOC sands for a panel of "volatile organic compounds" that include benzene). Replicate measurements were collected for each month of the sampling campaign, in a random sample of 10% of the locations. The same collection and analysis methods were used for all samples and replicates. All monitors were set 1.5 to 1.8 meters above the ground, at sites following several criteria. Each site was away from minor sources of exposure, such as roads, farm equipment operated by internal combustion engines, more than 100 meters from fuel and farm equipment storage areas, and local oil field equipment; more than 10 meters from roadways and other areas where vehicles were to be expected; outside the immediate area of local oil and gas facilities, to avoid "worst-case" sampling; at least 20 meters from the nearest tree canopy, as defined by drip line; away from buildings, hay storage, and other objects that may obstruct air flow; and in flat terrain. Airborne vapors were collected by exposing 3M Organic Vapor Monitor 3500 badges for one month periods. Benzene was extracted with carbon disulphide and analyzed by gas chromatography. The extraction recovery efficiency was 95%; the airborne sampling rate estimated to be 38.6ml/min. Deviations from standard application of the vapor 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 [18]. Little re-volatilization was observed in experimentally generated atmospheres over a wide range of environmental conditions. None of measurements of benzene concentration were censored for the limit of quantification, and actual instrument readings were used in calculating air concentration. Sampling and analytical procedures were developed and implemented by AirZone One Inc. (Mississauga, Ont., Canada), a commercial laboratory contracted by WISSA to analyze benzene samples. Distances between each source and monitoring station were estimated based on coordinates of monitoring stations and oil and gas facilities supplied by the provincial regulatory (government) agencies. Air benzene concentrations were measured at 1,200 sampling locations, from April 2001 to December 2002, for a total of 11,375 observations. Following [14], we studied the following sources as potential determinants of benzene concentrations: batteries, oil wells, gas wells, bitumen wells, other wells, and compressors and/or gas plants.

Statistical Analysis

Cubic splines have been found to have a good ability to fit sharply curving shapes over the entire range of the predictor variable [19]. We describe here our approach in modeling the dependency of benzene concentration on time and distances from various source types. More specifically, for a given monitoring location l we consider only sources situated within 100 km, and summarize across sources by taking the inverse distance of the *closest* source of a specific source of type i to that monitoring location (note that in our previous work summary function of distance to all sources of a given type was examined), i.e.:

$$D_{il} = \begin{cases} (\min_{\substack{\text{source } s \\ \text{of type } i}} D_{isl})^{-1}, & \text{if at least one source of type } i \text{ is} \\ & \text{within 100km of location } l \\ 0, & \text{otherwise} \end{cases}$$

For a given source of type i, we assume that the benzene concentration at a given monitoring location l, follows a cubic spline function with k knots, given by:
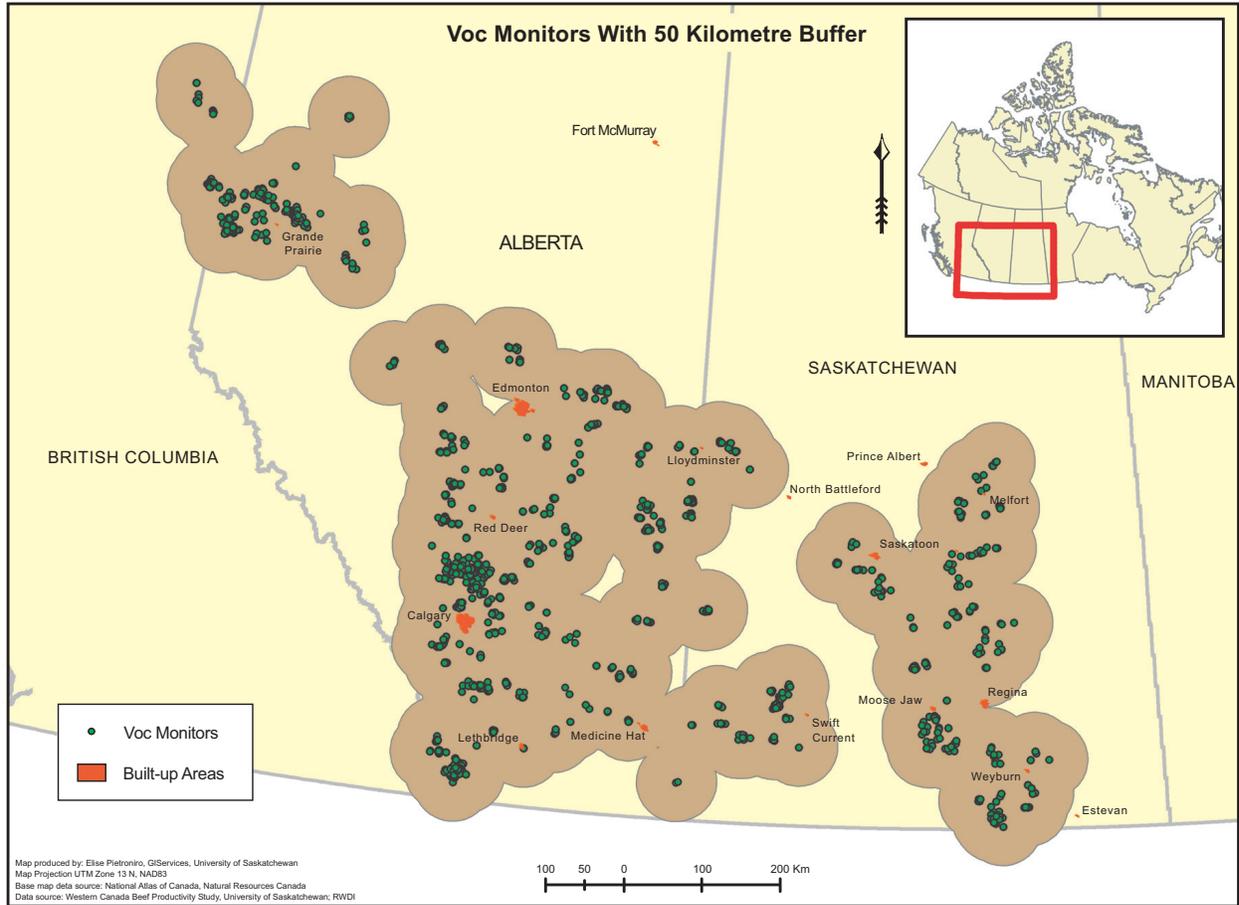
$$f_i(D_{il}) = \beta_i^0 + \beta_i^1 D_{il}^1 + ... + \beta_i^{k-1} D_{il}^{k-1},$$

where $\beta_i^0, \beta_i^1, ..., \beta_i^{k-1}$ represent the k coefficient parameters corresponding to the basis giving the cubic spline function, $D_{il}^1 = D_{il}$, and for $r=1,...,k-2$,

$$D_{il}^{r+1} = (D_{il} - t_r)_+^3 - (D_{il} - t_{k-1})_+^3 (t_k - t_r)/(t_k - t_{k-1}) \\ + (D_{il} - t_k)_+^3 (t_{k-1} - t_r)/(t_k - t_{k-1})$$

Here we denote by $t_1, ..., t_k$, the k knots and follow the recommendations given in [14] when choosing the number and position of knots. Previous work [20] showed that the location of knots in a cubic spline model is not very crucial and that the fit depends much more on the choice than the number of knots (k). For a large sample size ($n \geq 100$), it is recommended to use five knots. Therefore, we chose five knots located at the inverse of 0.5 km, 2 km, 10 km, 20 km, and 50 km to construct our cubic splines for the various source types.

To model the time effect over consecutive months starting at April 2001, and up to December 2002, we can use a cubic spline with five knots, corresponding to June 2001, October 2001.

Fig. (1). Map of the area under study.

February 2002, June 2002 and October 2002. Let X_h denote the time corresponding to the measurement taken at the h-th consecutive month. The cubic spline function to model time effect is given by:

$$f_{time}(X_h) = \gamma_0 + \gamma_1 X_h^1 + ... + \gamma_{k-1} X_h^{k-1}$$

where $\gamma_0, \gamma_1, ... \gamma_{k-1}$ represent the k coefficient parameters corresponding to the basis giving the cubic spline function, $X_h^1 = X_h$, and for $r=1,...,k-2$,

$$X_h^{r+1} = (X_h - t_r)_+^3 - (X_h - t_{k-1})_+^3 (t_k - t_r)/(t_k - t_{k-1})$$
$$+ (X_h - t_k)_+^3 (t_{k-1} - t_r)/(t_k - t_{k-1})$$

Overall, we propose the following model for benzene concentration:

$$Y_{jlh} = \sum_{type\ i} f_i(D_{il}) + f_{time}(X_h) + \tau_h + \lambda_l + \varepsilon_{j(hl)} ,$$

where Y_{jlh} is the logarithm of benzene concentration of the j-th replicate at the l-th sampling site, h-th consecutive month, τ_h is the random effect of the h-th month, λ_l is the random effect of the l-th location, and $\varepsilon_{j(hl)}$ is the random effect of the j-th replicate nested in the l-th location and the h-th

month. As in the previous work, we assumed that $\tau_h \sim N(0, \sigma_1^2)$, $\lambda_l \sim N(0, \sigma_2^2)$, and $\varepsilon_{j(hl)} \sim N(0, \sigma_3^2)$, and that these are mutually independent.

To select a final model, we first fit regression models consisting of terms corresponding to individual sources, and calculate the likelihood-ratio test (LRT), which is the change in deviance from the null model. Then, we carried out a forward selection process, adding terms to this model, while preserving the group structure of each term, rather than adding one coefficient at a time [16]. For multiple regression models, the LRT is the change in deviance when that term is added to the model *via* the forward selection procedure. The predictors were screened for multicolinearities using Pearson correlation coefficient. Plots of residuals were used to examine the assumptions of mixed effects models. Statistical analyses were performed using SAS software, version 9.1 (SAS Institute Inc., Cary, NC), and R package, version 2.5.1 [21].

RESULTS

The distributions of distances of nearest source to monitoring stations, for each nearest source type within 100 km, are presented in Table **1**. It appears that there is a fairly good coverage from 0 km to 100 km, for all the source types,

Table 1.　Distributions of Distances (in Kilometers) Between Monitoring Stations and the Nearest Source of a Given Type

Source Type		N	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
Batteries		9265	0.06	1.45	2.79	6.77	99.72
Compressors and/or Gas Plants		10101	0.17	2.96	5.44	9.34	98.06
Wells	Bitumen	1414	9.71	34.35	58.55	91.75	99.32
	Gas	10041	0.03	0.65	1.36	3.16	98.04
	Oil	10232	0.02	1.06	2.92	6.53	98.88
	Other	8936	0.03	0.86	1.51	2.70	88.93

except for bitumen wells. There were no bitumen wells within 9.7 km of any monitoring stations, and approximately 75% of the sources are located at distances larger than 34.4 km of monitoring stations. Given this lack of data at closer distances, it was unlikely we would be able to get reasonable predictions of the impact of bitumen wells on air concentration of benzene. Also, previous analysis did not find bitumen well to be a significant predictor of benzene concentrations [14]. Based on these considerations, we excluded this variable from the model-building process. After screening for multicolinearities, we did not exclude any of the variables, with 0.35 being the largest correlation coefficient.

Likelihood-ratio tests in Table **2** indicate that individual sources are important in explaining benzene concentrations, with the largest effect corresponding to time. The final model selected by the forward procedure consists of time (LRT X^2=2777.2, p-value<0.0001), oil wells (LRT X^2=151.1, p-value<0.0001) and batteries (LRT X^2=26.8, p-value<0.0001) (Table **2**).

We displayed in Figs. (**2**, **3**) the predicted benzene concentrations curves versus distance, for each source selected in the final benzene concentration model *via* forward selection procedure, and for each of the following months: April 2001, July 2001, July 2002, April 2002, December 2002, December 2001 (this order follows the magnitude of the predicted benzene concentrations). More precisely, the horizontal axis in Figs. (**2**, **3**) represents distance away from oil wells only, and

batteries only, respectively. The predicted benzene concentrations curves versus distance away from both oil wells and batteries (e.g. both one oil well and one battery 10 km away from the monitoring station) are displayed in Fig. (**4**), for the same months as described above. As there was no variation in the functions at larger distances, we chose to present the distance axis up to 50 km. Regarding time trends, the benzene concentration increased from April 2001 to December 2001, decreased to a minimum in July 2002, and then increased again up to a maximum in December 2002 (Figs. **2-4**). The benzene concentrations were a factor of 3 to 8 higher in December than July, for years of 2002 and 2001, respectively. The magnitudes of the effects of oil wells and batteries on benzene concentration are comparable: for example, for the month of December 2001, a maximum concentration of 5.5 µg/m^3 was predicted for oil well approximately 340 m away and 5.2 µg/m^3 for battery approximately 800 m away. For oil wells, our findings are not supportive of a peak in the benzene concentration at a certain distance. However, for batteries, the fitted cubic spline function indicated a peak at around 800 m, and then a decrease as we get further away from the source. Fig. (**4**) indicates that maximum concentrations were predicted at approximately 800 m away from both oil wells and batteries: for example, in the month of December 2002, a maximum concentration of 7 µg/m^3 was predicted when both an oil well and a battery are at approximately 800 m away. These predicted concentrations cover the entire range of observed benzene concentrations, from <0.0005 (5.3% of observation that are below the limit of

Table 2.　Results for Individual Source and Multiple Regression Benzene Concentration Models *via* Forward Selection. The LRT is the Likelihood Ratio Test, which is Change in Deviance from the Null Model, for the Individual Source Regression Models, and the Change in Deviance when that Term is Added to the Model *via* Forward Selection Procedure, for the Multiple Regression Models

| Effect	DF*	Individual Source Regression Model		Multiple Regression Model *Via* Forward Selection	
		LRT[§]	P-Value	LRT[§§]	P-value
Time	4	2777.2	<0.0001	2777.2	<0.0001
Wells oil	4	137.9	<0.0001	151.1	<0.0001
Batteries	4	108.3	<0.0001	26.8	<0.0001
Compressors and/or gas plants	4	94.9	<0.0001	8.3	0.08
Wells other	4	64.1	<0.0001	7.5	0.11
Wells gas	4	59.2	<0.0001	6.2	0.18

*DF is the degrees of freedom.
[§]For individual source models, the LRT is the likelihood-ratio test when that term is added to the null model.
[§§]For multiple regression model, the LRT is the likelihood ratio test when that term is added to the model *via* the forward selection procedure.

detection) to 9.036 μg/m^3 and indicate that our model has a good fit and predictive power (NB: geometric mean of the observed month-long average benzene concentration was 1.58 μg/m^3 with geometric standard deviation of 4.93). These findings indicate that the concentrations of benzene at which health effects in cattle were observed [13] can be explained by proximity to oil and gas infrastructure.
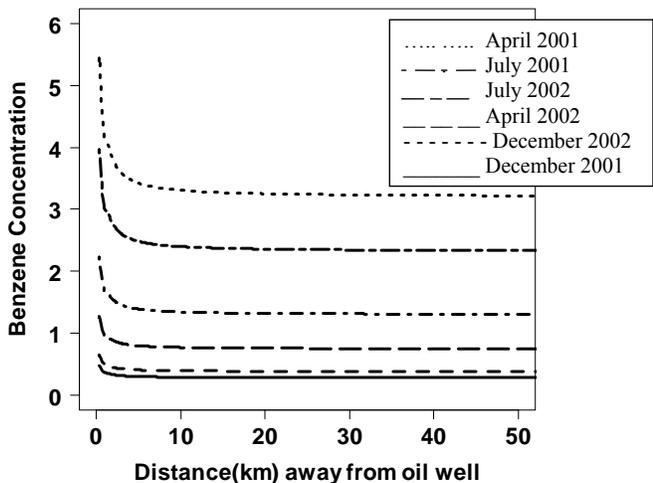
Fig. (2). Predicted benzene concentrations (μg/m^3) curves versus distance, for oil wells only, and for each of the following months: April 2001, July 2001, July 2002, April 2002, December 2002, December 2001 (this order follows the magnitude of the predicted benzene concentrations); horizontal axis represents distance away from oil wells only.
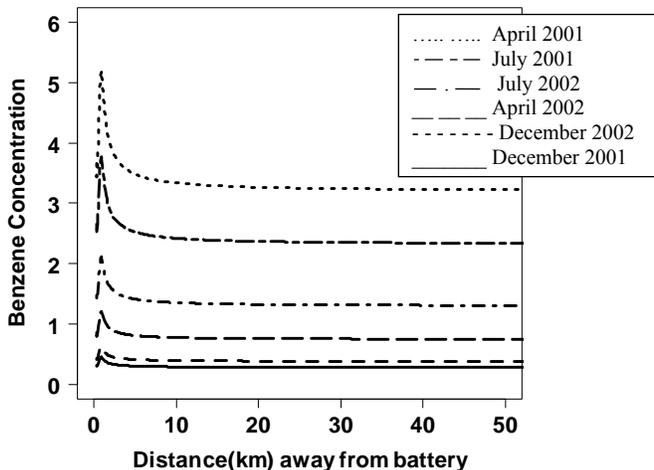
Fig. (3). Predicted benzene concentrations (μg/m^3) curves versus distance, for batteries only, and for each of the following months: April 2001, July 2001, July 2002, April 2002, December 2002, December 2001 (this order follows the magnitude of the predicted benzene concentrations); the horizontal axis represents distance away from batteries only.
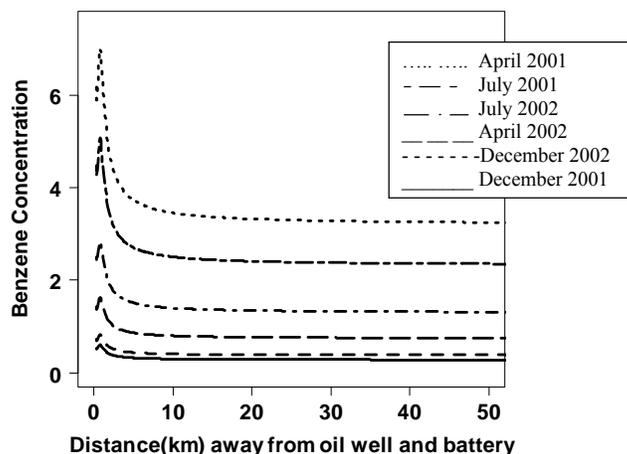
Fig. (4). Predicted benzene concentrations (μg/m^3) curves versus distance, for oil wells and batteries, and for each of the following months: April 2001, July 2001, July 2002, April 2002, December 2002, December 2001 (this order follows the magnitude of the predicted benzene concentrations); the horizontal axis represents distance away from oil wells and batteries.

Examination of the plots of the standard error bands around the fitted cubic spline functions corresponding to log scale of benzene concentrations (plots not shown) indicated increasing variability of the fitted curves within approximately 800m from the source, with a magnitude of the standard error band around the fitted logarithm of benzene concentration curve of 0.3 at approximately 800m away from the source), followed by a decrease and stabilization as we get further away from the source, with a magnitude of the standard error band around the fitted logarithm of benzene concentration curve of 0.1 at approximately 800m away from the source. Examination of residuals did not indicate severe violations of the assumptions underlying the models (plots not shown).

The following estimates of random effects were obtained for the final model: between-location variance of 1.29 (standard error (SE)=0.11), month-to-month variance of 0.004 (SE=0.0005), and between-repeat variance of 1.59 (SE=0.02). An increase of 87% and 8% in the magnitudes of these estimates for between-location and between-repeat variance, respectively, compared to those reported in [14], may be attributed to the bias-variance tradeoff, i.e., a more flexible model such as the one we used based on the cubic splines is expected to improve upon distortion of the shape of the association (bias) at the expense of increasing the variance. The fact that the magnitude of month-to-month variance in [14] was 100 times larger may be explained by the fact that they used a winter/summer type of variable, and we used consecutive months.

DISCUSSION

This large scale study of determinants of environmental benzene concentrations in rural area is ideal for the use of cubic splines: its large size allows for enough knots to

capture the fluctuations in benzene concentrations; the cubic splines are flexible enough to fit sharp shapes. We were able to construct a novel benzene exposure model that identified oil wells and batteries as important source of emission, after allowing for time trend and random variation among repeats. Previous work [14] indicated oil wells as important, with an estimate of 0.49 on the logarithm of benzene concentration scale. The size of the nonlinear effect in our model was fairly similar, 0.53 on the logarithm of benzene concentration scale. The proximity to batteries shows a significant nonlinear effect, which was not previously detected by the linear model [14]. Modeling nonlinear effects of distances from sources is important, as it may have an impact on assessing significance of candidate variables in predicting benzene concentrations. It is worthwhile to mention that previous analysis of WISSA data suggested an influence of batteries on benzene concentration, although not in a straightforward way. More precisely, You *et al.* [22] found out that proximity to batteries (within 2 km) was most influential in determining monthly airborne concentrations of components of a factor summarizing concentrations of all compounds similar to benzene among 26 volatile organic compounds candidates, based on principal component analysis.

We were able to fit a reasonable dependency function of benzene concentration on time, indicating monotonically increasing trends for the months from July through December, and monotonically decreasing trends corresponding to the months from December to July. Our findings that the benzene concentrations are higher in the month of December than July are in line with previous work [23], reporting that 24-hour benzene concentrations are a factor of 4 to 8 higher in January and February than in June and July, and also with a report that a 3-4 fold difference in monthly benzene concentrations between cold (November to April) and warm (May to October) seasons (in the same dataset as examined here) [14]. The greater abundance in summer of OH-radical that degrades benzene is usually used to explain such seasonal trend [24].

There are, however, limitations to our modeling approach. We summarized over sources of certain types, by taking minimum distances to monitoring stations. The validity of this approach relies to a large extent on soundness of the assumption that the closest source impacts the benzene concentrations is the most potent/influential one. Another limitation is that we overcome the lack of sources of certain types within 100 km of some of the monitoring stations by approximating their inverse distances to those monitoring stations with zero. We note that a similar approach was taken in [14], when the proximity scores were set to zero for those monitoring stations that did not have any sources of a certain type in a fixed distance. Another limitation is that the present work is based on the assumption that the distances from sources to monitoring stations were determined with negligible error. Previous work [25] showed that if the error was appreciable, then the shape of the association of interest can be distorted and power to detect associations diminished. Although this was shown under different parameterization of separation distance, a similar effect may be expected using our proposed model [26].

We have no data on wind speed and direction, but Alberta does not have strong prevailing winds and our prior work with direction seminvariograms did not reveal any general trends for concentrations [22]. Therefore it is unlikely that wind patterns, especially for month-long averages, are an important source of variance. But this of course can be studied further, especially in assessment of each specific facility, in which case matters of local topology and weather pattern will also have to be examined.

A variety of models and relevant variables to be included have been proposed to study the relationships between airborne concentrations and distance from the source [27-29]. Overall, our work indicates how cubic splines can be used to identify complex relations between proximity to point sources of air pollution and observed extent of contamination, and identified batteries as an important source of benzene emissions that was missed in previous analysis of the same data. This last finding contributes to better understanding how positioning of oil and gas facilities impacts air quality and may be of considerable impact on assessment and management of risk to health, especially if the findings that long-term exposure to benzene above 0.236 μg/m^3 affect cattle fertility [13] is found to be relevant to human health.

ACKNOWLEDGEMENTS

REFERENCES

[1]   IARC. Overall Evaluations of Carcinogenicity: An updating of IARC monographs. IARC monographs on the evaluation of carcinogenic risks to humans. Lyon: International Agency for Research on Cancer 1987; (Suppl. 7): pp. 1-42.

[2]   US Environmental Protection Agency. Integrated risk information system (IRIS) on benzene. National Center for Environmental Assessment. Washington, DC: Office of Research and Development 2002.

[3]   Report on Carcinogens, Eleventh Edition; 2005. U.S. department of health and human services, public health service, national toxicology program; (http://ntp.niehs.nih.gov/ntp/roc/toc11.html); January 31, 2005 Benzene profile can be found at (http://ntp.niehs.nih.gov/ntp/roc/eleventh/profiles/s019benz.pdf).

[4]   Parra MA, Gonzalez L, Elustondo D, Garrigo J, Bermejo R, Santamaria JM. Spatial and temporal trends of volatile organic compounds (VOC) in a rural area of northern Spain. Sci Total Environ 2006; 370: 157-67.

[5]   Rappaport SM, Kupper LL. Variability of environmental exposures to volatile organic compounds. J Exposure Anal Environ Epidemiol 2004; 14: 92-107.

[6]   Austin H, Delzell E, Cole P. Benzene and leukemia. A review of the literature and a risk assessment. Am J Epidemiol 1988; 127: 419-39

[7]   Glass DC, Gray CN, Jolley DJ, *et al.* Leukemia risk associated with low-level benzene exposure. Epidemiology 2003; 14: 569-77.

[8]   Scott HM, Soskolne CL, Martin SW, *et al.* Air emissions from sour-gas processing plants and dairy-cattle reproduction in Alberta, Canada. Prev Vet Med 2003a; 57: 69-95.

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

[10] Waldner CL. Western Canada Study of Animal Health Effects Associated With Exposure to Emissions From Oil and Natural Gas Field Facilities. Study Design and Data Collection III. Methods of Assessing Animal Exposure to Contaminants from the Oil and Gas Industry. Arch Environ Occup Health 2008; 63(4): 201-19.

[11] Waldner CL. Western Canada Study of Animal Health Effects Associated With Exposure to Emissions From Oil and Natural Gas Field Facilities. Study Design and Data Collection II. Location of Study Herds Relative to the Oil and Gas Industry in Western Canada. Arch Environ Occup Health 2008; 63(4):187-99.

[12] Waldner CL. Western Canada study of animal health effects associated with exposure to emissions from oil and natural gas field facilities. Study design and data collection I. Herd performance records and management. Arch Environ Occup Health 2008; 63(4):167-84.

[13] Waldner CL, Stryhn H. Risk of nonpregnancy, risk of disposal for pregnant cows, and duration of the calving interval in cow-calf herds exposed to the oil and gas industry in Western Canada. Arch Environ Occup Health 2008; 63(4): 241-61.

[14] Burstyn I, You X, Cherry N, Senthilselvan A. Determinants of airborne benzene concentrations in rural areas of western Canada. Atmos Environ 2007; 41: 7778-87.

[15] Strosher MT. 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, Atlanta 1996.

[16] Hastie T, Tibshirani R, Friedmann J. The Elements of Statistical Learning. Springer: USA 2001; 122-4.

[17] Harrell FE, Regression Modeling Strategies with applications to linear models, logistic regression, and survival analysis, Springer Verlag: New York, 2001; 19-24.

[18] Fellin P, Otson R, Ernst DL. A versatile system for evaluation of organic vapour monitoring methods. In: Proceedings of the Eight World Clean Air Congress. The Hague: The Netherlands, 1989; 3: 675-80.

[19] Stone CJ, Koo CY. Additive splines in statistics. In Proceedings of the Statistical Computing Section, ASA, 1985; 45-8.

[20] Stone CJ. Comment: Generalized Additive Models. Stat Sci 1986; 1:312-4.

[21] R Development Core Team, R: A language and environment for statistical computing. ISBN 3-900051-07-0, R Foundation for Statistical Computing, Vienna, Austria 2006.

[22] You XI, Senthilselvan A, Cherry NM, Kim HG, Burstyn I. Determinants of airborne concentrations of volatile organic compounds in rural areas of Western Canada. J Expo Sci Environ Epidemiol 2008; 18(2):117-28.

[23] Dann TF, Wang DK. Ambient air benzene concentrations in Canada (1989-1993): seasonal and day of week trends, and source influences. J Air Waste Manage Assoc 1995; 45: 695-702.

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

[25] Burstyn I. Impact of measurement error on quantifying the importance of proximity to point sources of air pollution. J Expo Sci Environ Epidemiol 2008; in press.

[26] Wang N, Davidian M. A note on covariate measurement error in nonlinear mixed effects models. Biometrika 1996; 83: 801-812.

[27] Turner DB. Atmospheric Dispersion Estimates. CRC Press: USA 1994.

[28] Araina MA, Blaira R, Finkelsteina N. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. Atmos Environ 2007; 41: 3453-64.

[29] Hoek G, Beelen R, Hoogh K. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ 2008; 42: 7561-78.