

Elsevier Editorial System(tm) for Science of the Total Environment  
Manuscript Draft

Manuscript Number: STOTEN-D-08-02063R1

Title: Spatial analysis and land use regression of VOCs and NO<sub>2</sub> from school-based urban air monitoring in Detroit/Dearborn, USA

Article Type: Full Length Article

Section/Category: Environmental Sources and Environmental Changes

Keywords: Air pollution; GIS; Spatial analysis; Land use regression (LUR); Urban air quality; Traffic.

Corresponding Author: Dr Shaibal Mukerjee, PhD

Corresponding Author's Institution: U.S. Environmental Protection Agency

First Author: Shaibal Mukerjee, PhD

Order of Authors: Shaibal Mukerjee, PhD; Luther A Smith, PhD; Mary M Johnson, PhD; Lucas M Neas, ScD; Casson A Stallings, PhD

Abstract: Passive ambient air sampling for nitrogen dioxide (NO<sub>2</sub>) and volatile organic compounds (VOCs) was conducted at 25 school and two compliance sites in Detroit and Dearborn, Michigan, USA during the summer of 2005. Geographic Information System (GIS) data were calculated at each of 116 schools. The 25 selected schools were monitored to assess and model intra-urban gradients of air pollutants to evaluate impact of traffic and urban emissions on pollutant levels. Schools were chosen to be statistically representative of urban land use variables such as distance to major roadways, traffic intensity around the schools, distance to nearest point sources, population density, and distance to nearest border crossing. Two approaches were used to investigate spatial variability. First, Kruskal-Wallis analyses and pairwise comparisons on data from the schools examined coarse spatial differences based on city section and distance from heavily trafficked roads. Secondly, spatial variation on a finer scale and as a response to multiple

factors was evaluated through land use regression (LUR) models via multiple linear regression. For weeklong exposures, VOCs did not exhibit spatial variability by city section or distance from major roads; NO<sub>2</sub> was significantly elevated in a section dominated by traffic and industrial influence versus a residential section. Somewhat in contrast to coarse spatial analyses, LUR results revealed spatial gradients in NO<sub>2</sub> and selected VOCs across the area. The process used to select spatially representative sites for air sampling and the results of coarse and fine spatial variability of air pollutants provide insights that may guide future air quality studies in assessing intra-urban gradients.

Response to Reviewers: RESPONSES TO STOTEN REVIEWER COMMENTS  
(REPLIES IN BOLD TEXT IMMEDIATELY BELOW REVIEWER COMMENT)

NOTE: Redlined/deleted (track changes) version of revised manuscript should be used in evaluation of responses below.

Reviewers' comments:

Reviewer #1:

The study positions itself as addressing 3 different issues: (1) evaluating impact of traffic and urban emissions on respiratory effects in children, (2) developing spatial models at the coarse and fine scale (3) providing an insight into site selection.

None of these goals is adequately addressed by the paper. Since the focus is not clear, the paper is confusing at many points.

RESPONSE: While point 1 was a motivating factor, this paper does not address health effects at all. This is planned for a later paper. With respect to point 2, we would not characterize our coarse scale analysis as a modeling effort. We have reviewed the phrasing of the paper and do not feel it is misleading in this regard.

Referencing in the paper is poor. For example:

(1) pg 2, line 21: These and other results have influenced enactment of recent legislation in California. When I read the CA legislation, it did not mention air pollution - it mentioned noise, hazardous spills and other things.

RESPONSE: The California Act in question states that a governing school board is prohibited from siting new schools within 500 feet from the “closest traffic lane of a freeway or other busy traffic corridor...” as it relates to “hazardous substances.” Hazardous substance is referred to by the Act as defined by California Health and Safety Code Section 25316; this definition includes air pollutants. The introductory language of the Act explicitly mentions as a concern the effect of such pollutants as benzene and 1,3-butadiene on children’s health problems including asthma. As the paper indicates, we took this same viewpoint in the Detroit study.

(2) Pg 3, line 8: Based on these and other monitoring approaches the US EPA conducted air monitoring studies in late 1999. Reading this I assumed they were referring to the citations in the paragraph above it. All those references are 2000 or later. Very clairvoyant of the US EPA!

RESPONSE: Owing to the size of the manuscript, we were attempting to limit references to the most relevant that reference earlier LUR papers. The seminal LUR reference (Briggs et al., 1997) has been added in the preceding paragraph. Phrasing has been revised to remove ambiguity (see first paragraph of Introduction.)

(3) In the same line, there is no citation for the EPA El Paso study (it is referenced later).

RESPONSE: Smith et al. reference now added at the end of the sentence in question (second paragraph of Introduction).

(4) Surprisingly, no reference to a study by the same author (Johnson MM primary author) to Evaluation of Land-Use Regression Models in Detroit Michigan, Epidemiology 19:6 (2008)

This looks and reads like a paper that has been very hastily put together.

RESPONSE: The reference in question was an abstract for a poster, not a peer-reviewed journal article. We focused our references on peer reviewed journal articles and reports as well as textbooks.

Key methodology questions are not in the paper: how many passive samplers were deployed at each of the 25 schools?

RESPONSE: Phrasing has been revised to indicate that duplicate passive samplers were located at the compliance sites (first paragraph of Section 2.3).

Were samplers deployed in pairs?

RESPONSE: See comment above.

How well did the paired samplers match?

RESPONSE: We did not want to repeat information from another paper in this manuscript.

Coefficients of variation are shown for the duplicates in current Table 6 (former Table 2) to give the reader a sense of duplicate variability versus spatial variability.

What is the effective least count of the samplers? While a reference to a paper giving details is fine (pg 8, line 3), these details need to be in this paper, as the reliability of the measured data affects the reliability of the statistical model the paper aims to present in this paper.

RESPONSE: The current paper is already lengthy as it is. We disagree that these details need to be repeated here as well as in the reference cited.

Statistical analysis is shoddy.

(1) No motivation is given for the Enumeration Area (EA)-based coarse spatial analysis. The analysis itself is not interpreted. It is left at some variables being significant if an inequality given on pg 10 line 1 holds!

RESPONSE: We agree that the motivation for EAs was not present in the original version. First paragraph of Section 2.4 has been revised to indicate why EAs were considered as part of the coarse spatial analyses. However, the remainder of this comment is incorrect. It is true that the inequality cited on current first paragraph of page 11 is used to declare significance. We chose to display the inequality in the Methods section because we modified the standard version of Dunn's test and we wished to be explicit about exactly what was done. As stated in the paper, this was done to improve the power of Dunn's test and appropriate references are provided. However, we did not simply leave it at this. Section 3.2 in the Results section provides a complete description of the outcome of these comparisons. Furthermore, the first paragraph of the Discussion and Conclusions section (Section 4) interprets the outcome of the coarse spatial analysis.

(2) In the LUR model (i) some variables are log-transformed. Which variables were transformed and why is not noted. While log transformation are not uncommon, the reason needs to be noted, and hopefully tied to a physical process or interpretation.

RESPONSE: Log transforms were utilized based on graphical inspection of the data. While this was stated in the original version, the first paragraph of Section 2.5 has been reworded for clarity. Contrary to the comment, Table 4 explicitly shows all variables that were log transformed in the LURs. For brevity, we left this to this table but we have added a reference to it (first paragraph of Section 2.5).

(ii) Why was MN\_BIG\_DIST even considered as one of the explanatory variables? The explanatory variables were picked a priori, but no adequate reason is given for picking this variable.

RESPONSE: We stated in second paragraph of Methods section 2.1 and paragraph 7 of Discussion and Conclusions section 4 that distance from large manganese and PM sources were considered because of their potential use in the health study.

(iii) When the predictors were not significant because of high variance, the observations were weighted with the inverse of Cook's D - pg 11, line 2. There has to be theoretical reasoning for adding weights - not just a statistical argument for fitting the model better, specially when the weighting adds collinearity (pg 13, line 14)

RESPONSE: As the paper indicates, the weighting procedure was employed to de-emphasize the effect of specific sites on the individual regressions. The reason why these sites do not conform as well to the behavior exhibited at the other sites is unknown and may provide fodder for future research. However, the fact that this reason is not known does not preclude the use of a weighting scheme to obtain a more reliable and interpretable result from the regression analysis. It is incorrect to state that a statistical technique must await a physical interpretation before it may be applied in an analysis. Indeed, scientific research is often conducted to explain such anomalies uncovered by analyses such as this.

(iv) A note in Table 4 (pg 29) notes that the R<sup>2</sup> values reported are for the original scale, not for the log-transformed scale the model is reported in. At this point, there we are left with a model which cannot be interpreted, and hence cannot be applied, defeating the purpose of model development.

RESPONSE: The reviewer is incorrect. It is precisely for interpretability that we calculated the R<sup>2</sup> values on the original scale. This indicates how reliable the model estimates will be when applied to measurement data. We are not alone in this opinion. We have added references along this line at the point where mentioned in the text (first paragraph of Section 3.3).

For these reasons, the analysis as presented is not worthy of publication.

RESPONSE: As the above responses indicate, we disagree with this statement.

However, the data is interesting and the paper raises issues (not followed up on) that are worth pursuing.

(1) One of the goals mentioned in the paper is an insight into site selection. The authors of this paper have an ad hoc approach based on maximizing the range of the independent variables, while controlling for collinearity. This approach could possibly be refined using principal component

analysis. A comparison between this (refined) approach and Kanaroglou et al's [Atmospheric Environment 39 (2005) 2399-2409] location-allocation algorithm - based on optimizing the range of the pollutant and the exposure - would be interesting to the LUR community.

RESPONSE: We point out that our method involves more than just covering the range of independent variables. Multivariate considerations are also taken into account. The reviewer's suggestion of comparing different approaches has merit for future research but is beyond the scope of this paper. With respect to location-allocation, we previously referenced Sahsuvaroglu et al. and Henderson et al. In considering the reviewer's comments, we have added the Kanaroglou et al. reference. We note that part of this future research comparing different approaches would need to deal with the fact that these other papers take a regional perspective in developing their initial model as opposed to the much more geographically compact Detroit/Dearborn focus we have in this paper.

(2) The transferability of LUR models - across space and time - is another interesting topic. A discussion of transferability of LUR models for Detroit, using this data as the data to be fitted, is also an interesting exercise.

RESPONSE: Again, this is a useful topic for future research that we plan to pursue but is beyond the scope of this paper. In general, transference of an LUR model developed in one situation for another application should be approached with caution. For example, in the El Paso study referenced in the paper Smith et al., elevation was used as an explanatory variable but is irrelevant for Detroit. While similar traffic, land use, and other variables can be considered, the pre-analysis approach we present can be utilized to select an appropriate set of relevant explanatory variables for the individual case under study.

(3) The authors could consider alternate modeling techniques to fit the data, for example, spatial interpolation methods (Beleen et al, Mapping of background air pollution at a fine spatial scale across the European Union, Science of the Total Environment, 2009, in press).

REPOSNE: This is beyond the scope of this paper and most LUR studies in specific cities. The approach presented in Beleen at al. is being applied on a continent-wide basis where considerably more monitors are available. The kriging utilized in the Beleen at al. reference could not have been supported with the available monitoring data from Detroit.

Reviewer #2: This is an interesting piece of research, the selection of air sampling sites and the sequential development of land use regression models. The topic is relevant to STOTEN.

However, there are some concerns, making it very difficult to critique the validity of the conclusions. I have listed below things that have caught my attention.

Major comments/suggestions:

1. The method used to select the eight variables and the 25 sampling sites was not well justified, thus it is nearly impossible to replicate the method in another area based on the information provided.

The selection of the eight variables from a pool of 45 was somewhat ambiguous or arbitrary to me. The authors should provide a quantitative procedure on how to select those variables. Similarly, the method to select 25 schools from 116 was presented rather vaguely: "schools were selected to reflect the range of combinations of group numbers across all the potential predictors", please describe quantitatively the selection criteria.

RESPONSE: We understand the reviewer's desire for a more quantitative procedure and this may be pursued in future work. However, the key point is that for the predictive equations to be valid across the entire geographic region, combinations of the predictor variables at the chosen sites must span the mathematical space covered by the predictor variables. Otherwise, the predictive equations developed cannot be reliably extended over the whole area for which predictions are desired. The difficulty arises from the fact that a variety of variables are to be considered simultaneously. To retain maximum flexibility, we felt that the goals of the project were better served by the procedure described in the paper. To address the reviewer's concerns in this area, we have added a new Table 1 and moved the Supplementary Table S5 to become Table 2. These present summary statistics and correlation structure of the chosen variables. In addition, by revising the text we have tried to clarify the objective of spanning the mathematical space with our choice of sites. We have also added to the text a note that the procedure does not lead to a unique choice of a set of variables or sites.

The authors mentioned that 25 schools were chosen from a pool of 116. It is not clear to me whether 25 is a predetermined number due to resource limitations, or only 25 schools in the area met the site selection criteria. Similarly, 4 or 5 schools were chosen from each EA. It is not clear to me whether those numbers were predetermined based on the size of each EA or 4/5 schools in the each EA met the site selection criteria. Also, did you decide that you wanted eight variables out of 45 before hand, or only eight met the selection criteria?

RESPONSE: The number of schools to select was based on resource limitations. Text has been added to indicate this (first paragraph of Section 2.2). Equal distribution of school numbers within

EAs was done with an eye to the later health analysis while maintaining the ancillary variable distribution; the paper has been revised to indicate this (second paragraph of Section 2.2). The 8 variables were not decided a priori but resulted from the procedures described in the paper.

2. Air monitoring. You mentioned that "weeklong" sampling was conducted during a 6-week period of "July 18 - August 30, 2005"; it is not clear which sites in which week were monitored. If all 25 sites were monitored for 6 weeks, please clarify how to get one concentration per site per pollutant; alternatively, if each week only a few sites were monitored, followed by a different group of sites the following week, and so on and so forth, please justify the use of rotating monitoring data in a span of 6 weeks.

RESPONSE: We indicated in second paragraph of Section 2.3 that concurrent monitoring was performed at the schools. We revised the first sentence of that paragraph to make it clearer that concurrent monitoring was done at all sites for six weeks.

Page 7, L12-14. I am not sure whether correlation between those variables is the property that one wants to persevere in the site selection process, or the distributions ought to be persevered, but distributions cannot be assessed directly using correlation analysis.

RESPONSE: We have clarified in the text that both the correlation structure and the distributions are preserved between the chosen and unmonitored sites. Indeed as indicated above, we have expanded our description of spanning the mathematical space (Section 2.2).

Page 8, L10-11. "Concurrent monitoring was conducted at local neighborhood schools to reflect children's exposures in the immediate community." It is not clear to me: 1) whether this sentence refers to this study or the study of (Morishita et al., 2006). 2) whether the "local neighborhood schools" are the 25 schools or a different set of schools. Suggest rephrasing this sentence and maybe start a new paragraph.

RESPONSE: As suggested by the reviewer, a new paragraph was made in terms of the sampling description. Text has been clarified (new third paragraph of Section 2.3).

Page 8, L16. "using a portable, calibrated VOC monitor accurate to ppm levels; no contamination was detected.". Please provide the make and model of the VOC monitor, and describe quantitatively the meaning of "no contamination was detected".

RESPONSE: Description of the VOC monitor and the levels it detected are now included (new third paragraph of Section 2.3).

It not clear to me how to obtain EA-level concentrations using 4/5 sites in each EA. The power of the Dunn's test could be low due to small sample sizes (4 or 5)

RESPONSE: The Dunn's test utilized the 6-week average concentration at each site within the EA. Text has been added to indicate this (first paragraph of Section 2.4). The reviewer is correct that small sample size inhibits the power of the test. As we noted in the paper, as noted in the paper we modified Dunn's test to increase its power.

3. The method used to develop LUR models is overly simplified or flawed.

My major concern is the departure of the LUR model development method presented in this paper from the rest of the models in this field. Generally, a multiple linear regression is conducted by screening all potential independent variables, or predictors, using one-on-one correlation analysis where variables with a very low  $r^2$  value or a high p-value were eliminated; the remaining variables are then used in a stepwise procedure (manually or using SAS script) where the final model is chosen, in which all variables were statistically significant (e.g.  $p < 0.05$ ) and presented the best model fit, in addition to high  $r^2$  and other diagnostic analysis as the authors presented.

RESPONSE: As the paper notes, we did effectively initiate the process as indicated by the reviewer but we did it graphically instead of specifically using one-on-one correlations. However, we chose to present these as plots in the supplemental data (Figs. S1) rather than lengthen the paper. The reviewer is correct that stepwise regression is often applied in a multiple linear regression setting. However, we were concerned with applying it within the context of weighted regression. We chose to present the complete regression results in Table 4 which clearly indicates which variables were significant predictors and coupled this with a discussion of the collinearity resulting from the weighted regression. Note we have added what was the supplemental collinearity material in the text. (see third paragraph of Section 3.3 and new Table 7). The fact that we have chosen a slightly different approach to that commonly used does not invalidate the approach. We believe there is room in the LUR tool kit for multiple options. For example, a generalized additive models approach was used in El Paso (Smith et al., 2006 reference in paper).

Specifically, in this paper,

1) only eight variables from a pool of 45 were considered in LUR models, leading to a) few correlation coefficients being significant (page 10, L18), b) some variables "acting as surrogates for variables not considered" (page 17, L19), and c) some other issues listed below.

RESPONSE: We are not sure that the points listed here are the result of using only 8 variables as predictors. Using additional variables from the original 45, or a different set of variables, does not

guarantee a different outcome to the analysis. In fact, the collinearity problem might have been worse. a) We assume the reviewer meant regression coefficients, not correlation coefficients. This particular reference is to the original unweighted regression, not the weighted regression reported as the outcome in this paper. b) We only suggested this as part of our discussion of the results, not as a conclusion of the analysis.

2) A weighting scheme was used which cause collinearity (page 13, L14), which might also be the one of the reasons of low R<sup>2</sup> of original modeled and predicted concentrations in some models. RESPONSE: Collinearity does not cause low R<sup>2</sup> values. It does interfere with the interpretation of the regression coefficients, as was noted in the original paper and has been expanded upon in the revised version (third paragraph of Section 3.3).

3) Eight out of nine final models (Table 4) have, as many as five, variables with statically insignificant coefficients (i.e. coefficient was not statically different from zero), which warrants justification. Also I am wondering if the FULL models were significant or not.

RESPONSE: Each FULL model reported was significant. Again, we feel that collinearity may have been the problem here and this led us to report the entire suite of variables used in each regression in Table 4 and shifted the bulk of collinearity discussion from supplementary data to the text.

4) Eight out of nine final models (Table 4) have some coefficients with signs that point to a causal relationship that is unexpected. For example, a possible sign of coefficients to "Distance to a border crossing" or "distance to a road with X amount of traffic" suggests that the farther away you are from those places, the higher the concentrations. Similarly, a negative sign of coefficient to "traffic density" indicates that higher traffic density is associated with lower concentrations. Six of those eight models have two or three predictors with questionable signs, I am wondering if this "double incorrect" makes the models "correct". The inclusion of those predictors could partially explain less reliable predictions (page 14, L11-16).

RESPONSE: We noticed this as well and, in light of this and another reviewer's comments, we have expanded our discussion of the effect of collinearity on the regression coefficients. In particular, the specific cases cited in this point are noted in the new Table 7.

In summary, I am not convinced that the screening of LUR predictors prior to the air sampling was appropriate. In other words, the authors should be encouraged to explore other variables in the LUR process, which may or may not eliminate issues listed in 2) to 4) above.

RESPONSE: See our responses above.

4. The authors may want to make the presentation of the results more effective.

Suggest including general statistics of all concentration data or the selected species, the method detection limits, and the number of valid samples, in a table or a box plot.

RESPONSE: General statistics for all schools now added to Table 5 (was Table 3) which includes method detection limits.

The authors may want to rearrange the materials in the Results section such that the results of this study are presented first, followed by comparison to other studies, instead of a comparison without results (e.g. page 11, L14-18).

RESPONSE: As suggested, summary statistics of study results presented first before comparisons. See redlined text in first paragraph of Results section 3 on p. 12. Also, Section 3.1 has been re-titled as Concentrations. We did this to provide an overall context.

Page 14, L6-9, suggest tabulating all indexes (e.g. RMSR and ME) of all modes in a table.

RESPONSE: We feel this level of detail would be uninviting to readers and prefer to leave the discussion of the cross-validation as it is.

Table 2. I didn't find any numbers about CV between weeks for temporal variability (page 12, L8-10).

RESPONSE: Temporal CVs have been added to Table 6.

Discussion and conclusions. The reviewer found that this section is a bit hard to follow due to a frequent switch between the two pollutants (NO<sub>2</sub> and VOCs) and between the two scales (coarse and LUR). For example, the reason of little variability in VOCs was "due to pervasive mobile source effect" in one paragraph (page 15, L16-17), but "due to the fact that winds were generally coming from all directions during each week of the study" in another (page 16, L2-3), then the limitations of the coarse scale approach (page 16).

RESPONSE: We thank the reviewer for pointing this out. We merged the two paragraphs in question and consolidated the discussion (first paragraph of Section 4).

Some conclusions in the paper are not supported by appropriate analyses and seem to be almost arbitrary. e.g. page 15, L14-17: "Analysis ..found little variability of VOCs. This suggested a pervasive mobile source effect." and page 17, L3-8.

RESPONSE: With respect to the first comment, phrasing has been revised to indicate that this is a possible explanation, not a conclusion. With respect to the comment about p. 17 (now p. 18, third paragraph), it is well known from regression analysis that extrapolation beyond the range of the independent variable space is inherently risky. The discussion on these lines simply reiterates this point within the context of LUR modeling.

Table 4 is a bit hard to follow, suggest the following format:

| Dependent variable and Model R2(%) | Parameter | Estimate | p-value |
|------------------------------------|-----------|----------|---------|
| Y1                                 | Intercept | 1        | 0.001   |
| R2=0.80                            | X1        | 0.1      | 0.002   |
|                                    | X2        | 0.3      | 0.04    |
| Y2 (60)                            | Intercept | 1        | 0.001   |
| R2=0.60                            | X1        | 2        | 0.00    |
|                                    | X3        | 3        | 0.03    |

RESPONSE: We disagree that the suggested format for Table 4 would be easier to follow. For reasons indicated above, we wish to report the full regression model that was attempted in each case.

Fig 2. The LUR model predicted surfaces should be provided with all schools marked. The discussion (page 15, L-10) should be based on these LUR maps, instead of values at the schools. The reason is that one should establish the credibility of the models (i.e. predicted surfaces are smooth and free from irregularity or unexpected features) first before using the models to predict the concentrations at the schools.

RESPONSE: We disagree with this comment. We feel that the use of Figs. 2a-c and the relevant discussion are appropriate. The paper already provides considerable discussion with respect to the credibility of the models. In particular, the regression coefficient estimates, cross-validation results, regression diagnostics including residual diagnostics and AIC and BIC results, and comparison of predicted versus measured values are already presented.

Editorial comments/suggestions:

The term "mathematical space" was rather confusing; suggest replacing it with better words, maybe "statistical space" or "statistical property".

RESPONSE: The correct term is “mathematical space” and we have elaborated on this in the revised version (first paragraph of Section 2.2).

Suggest using international units, since most journals would not allow other units. For the same reasons, please delete scales in Miles in Figs 1 and 2. Units of the eight variables (pages 5&6) should be provided.

RESPONSE: Scale was provided in km and miles for ease of reference by local community groups and US and international investigators. We have adopted the reviewer’s suggestions and have provided the units for the eight variables in the text (see last paragraph of Section 2.1). Complete descriptions of the ancillary variables are provided in Tables S1 to S4.

Significant numbers: suggest one decimal, i.e. 12.3, for NO<sub>2</sub> concentrations throughout the text and in all tables.

RESPONSE: We prefer to leave the tables as they are.

There are a bit too many sentences in ( ). Suggest limiting its use to minima.

RESPONSE: Sentences and phrases in parenthesis have been minimized throughout revised manuscript.

## Abstract

The flow is a bit hard to follow. Suggest following the default structure: Objectives, Methods (site selection and monitoring methods), Results, and Conclusions.

RESPONSE: Some text has been added and some phrases deleted to enhance the flow of the Abstract.

## Introduction

I don't think that the study of Luginaah et al. (2006) was conducted using school sites. Maybe this reference could be moved to page 3, L5. Several recent studies have also included LUR modeling of SO<sub>2</sub>.

RESPONSE: Luginaah et al. (2006) stated that for their passive NO<sub>2</sub> monitoring to develop LURs, “...(t)he monitoring site selections were based primarily upon proximity to elementary schools, as well as ensuring inclusion of all types of land uses such as road networks, industry, and residential settings...”

A recent study on LURs for SO<sub>2</sub> (Wheeler et al., 2008) is now included.

## Method

Section 2.1 is a bit hard to follow. The authors may want to rearrange the materials in order to improve the flow. For example, start with the types of variables that you wanted to include (page 5), move on to the data sources (page 4), then how to select 8 from 45.

RESPONSE: We have added a sentence, as the reviewer suggested, indicating the general types of variables originally considered (second paragraph of Section 2.1).

Page 5, L8-10. Suggest clarifying: 1) how many border-crossing points were considered, 2) whether the distance to a crossing was used as a surrogate of Canadian emissions.

RESPONSE: Added text that two border crossings were considered, Ambassador Bridge and Detroit-Windsor Tunnel (second paragraph of Section 2.1). We stated in that same paragraph that "No point source or traffic data for neighboring Windsor, Ontario were available". But the distance to border crossing was not used as a surrogate for Canadian emissions; the text has in the same paragraph been revised to clarify this.

Page 5, L16, I think the authors meant "used in site selection"; since LUR method was presented in Section 2.5.

RESPONSE: Text has been revised to say "...to be used in the LUR models..." (see second paragraph of p. 6).

Page 10, L10-12. Please clarify how to decide which variables to log-transform.

RESPONSE: Revised the text to make it clearer what we did in first paragraph of Section 2.5.

## Results

Page 11, subheading of 3.1, could read "Concentrations"

RESPONSE: Subheading so revised.

Page 11, L12-14. The sentence is a bit confusing.

RESPONSE: We thank the reviewer for pointing this out. The text has been revised (first paragraph of Section 3.1).

Page 11, L18. Suggest including a few recent studies in the reference list.

RESPONSE: An additional, recent, reference has been added (first paragraph of Section 3.1).

Other editorial suggestions:

| Page                                                                                                       | Line  | Suggestion                                                                                                                                                                                                                           |
|------------------------------------------------------------------------------------------------------------|-------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1                                                                                                          | Title | Spatial analysis and land use regression of VOCs and NO <sub>2</sub> from school-based urban air monitoring in Detroit/Dearborn, USA                                                                                                 |
| RESPONSE: Title revised and emphasized Detroit and Dearborn, MI, USA in the abstract.                      |       |                                                                                                                                                                                                                                      |
| 2                                                                                                          | 4     | NO <sub>2</sub> and selected VOCs across the area                                                                                                                                                                                    |
| RESPONSE: Done.                                                                                            |       |                                                                                                                                                                                                                                      |
| 3                                                                                                          | 2     | predict pollutant concentrations                                                                                                                                                                                                     |
| RESPONSE: Done.                                                                                            |       |                                                                                                                                                                                                                                      |
| 416                                                                                                        | 914   | LUR models                                                                                                                                                                                                                           |
| RESPONSE: Done.                                                                                            |       |                                                                                                                                                                                                                                      |
| 4                                                                                                          | 20-22 | distance from school to nearest point source of VOCs and PM (emission data from the EPA 1999 National Emission Inventory database) as well as manganese (Mn) (emission data from the Michigan 2002 Toxic Release Inventory database) |
| RESPONSE: Revised as suggested but minimizing the parentheses.                                             |       |                                                                                                                                                                                                                                      |
| 5                                                                                                          | 2-3   | 2006; Hoek et al., 2008); these variables were also considered in this study.                                                                                                                                                        |
| RESPONSE: Done.                                                                                            |       |                                                                                                                                                                                                                                      |
| 5                                                                                                          | 17    | Pearson correlation coefficient                                                                                                                                                                                                      |
| RESPONSE: Done.                                                                                            |       |                                                                                                                                                                                                                                      |
| 5                                                                                                          | 23    | Suggest deleting "large" since the emission amount was specified.                                                                                                                                                                    |
| RESPONSE: Retained since emphasizing what the large source is designated in terms of range pounds emitted. |       |                                                                                                                                                                                                                                      |
| 6                                                                                                          | 1, 2  | Suggest deleting "large" since the emission amount was specified.                                                                                                                                                                    |
| RESPONSE: Retained large to be clear that smaller sources did exist.                                       |       |                                                                                                                                                                                                                                      |
| 6                                                                                                          | 5     | In this study, schools were.                                                                                                                                                                                                         |
| RESPONSE: Revised first paragraph of Section 2.2 for reasons above.                                        |       |                                                                                                                                                                                                                                      |

6                      7                      ensure that reliable predictions of something.

RESPONSE: Revised first paragraph of Section 2.2 for reasons above.

7                      3-4                      for each of the eight variables

RESPONSE: Done.

7                      18-20                      Passive samplers were deployed outdoors at the 25 selected schools and at two compliance sites operated by the Michigan Department of Environmental Quality (MDEQ), as shown in Fig. 1.

RESPONSE: Done.

8                      4-5                      Weeklong integrated sampling or Weeklong sampling interval

RESPONSE: Done.

8                      21                      These sampling methods have been validated

RESPONSE: Done.

9                      2-4                      Seems out of place, suggest deleting.

RESPONSE: Done.

9                      11                      Monitored concentrations of NO<sub>2</sub>, 1,3-butadiene, BTEX species, and styrene were.

RESPONSE: Done.

9                      16-17                      The Kruskal-Wallis test (SAS, 2004b) was used to

RESPONSE: Sentence revised but NPAR1WAY procedure retained. See first paragraph of Section 2.4.

10                                      equation and 1st paragraph, suggest deleting and referencing if possible.

RESPONSE: We feel it should be retained because we modified Dunn's test.

10                      19                      coefficients were statistically significant ( $p < 0.05$ ). The same, i.e. ( $p < 0.05$ ), applies to other places.

RESPONSE: Done.

11                      3-4                      A discussion of both transformations and weighting in a regression context can be found in Carroll and Ruppert (1988).

RESPONSE: Done.

11                                      subheading of 3.1, could read "Concentrations"

RESPONSE: Done.

11                      12-14                      The sentence is confusing.

RESPONSE: Revised as noted above.

11                      18                      Suggest including more recent studies in the reference list.

RESPONSE: Revised as noted above.

11                      19-21                      Seems to be out place, suggest deleting.

RESPONSE: Would like to retain since this indicates other pollutants measured. Revised to state these pollutants were not analyzed further since below detection (end of first paragraph, Section 3.1).

12                      2                       $p = 0.27$ ), could be due to slightly lower ambient levels (Table X).

RESPONSE: Text revised as suggested (second paragraph of Section 3.1).

12                      6-7                      coefficients of variation (CV) were calculated

RESPONSE: Done.

12                      18                      largest length of freeways

RESPONSE: Done.

13                      5                      Seems awkward, please rephrase: "figured into one of the potential predictors and in recognition of the California school siting legislation".

RESPONSE: Phrasing revised (last paragraph of Section 3.2).

13                      10                      Suggest deleting.

RESPONSE: Done (first paragraph, Section 3.3).

13                      11                      Table 4 presents the final LUR models.

RESPONSE: Done.

14                      2                      Cross-validation (Cressie, 1993),

RESPONSE: Done.

14                      10-11                      this was attributable to large discrepancies between measurements and predictions at one, two, or three sites.

RESPONSE: Done.

14                      21                      NO<sub>2</sub> was predicted well at the East 7 Mile site.

RESPONSE: Done.

15                      12                      School-based ambient air monitoring was successfully conducted in Detroit and.

RESPONSE: Done.

16                      10                      "has a large number of .", or "has facilities that emit large amounts of ."

RESPONSE: Done.

16                      19                      coarse level testing, Sentence seems awkward, please rephrase.

RESPONSE: Sentence has been rephrased, fourth paragraph of Section 4.

16                      21-22                      the prediction of pollutant levels across the entire area by considering multiple variables simultaneously, without being restricted by any discretization.

RESPONSE: Done.

17                      13-14                      why all GA-level mean values were statistically the same, expect one pair.

RESPONSE: Revised, paragraph 6 of Section 4.

17                      23                      Suggest moving to Results.

RESPONSE: Sentence put in Methods section (paragraph 3, Section 2.1).

18                      9-11                      Seems awkward, please rephrase.

RESPONSE: Sentence revised, last sentence of text.

Table 1                      Caption                      Group rankings by explanatory variables for 25 monitored schools.

RESPONSE: Changed "by" to "at"; new Table 3.

Table 3                      Caption                      Median values of NO<sub>2</sub> and selected VOCs (all in ppbV) in each enumeration area

RESPONSE: Current Table 5 caption revised. Table also includes summary statistics for all schools.

Table 5                      Caption                      Comprising of measurements and LUR predictions of NO<sub>2</sub> and selected VOCs (ppbV) at Dearborn and E7 Mile MDEQ sites.

RESPONSE: Done, current Table 8 caption.

Figure 1                      Caption                      Locations of schools and compliance monitoring sites: Site 10 at East 7 Mile and Site 23 at Dearborn.

RESPONSE: Revised.

Tables                      footnotes                      Table 2, delete 2 Table 3, delete last two sentences

RESPONSE: Current Table 5 (was 3) footnote revised.

Reviewer #3: This is an impressive paper that documents the performed research very well.

However, I have some minor comments for the authors:

1.                      Line 14, pg1 -Abstract should indicated that Michigan is in the USA

RESPONSE: Done.

2.                      Line 21, pg2 - "These and other results." influenced the law enacted in 2003, except that the articles referred to in "these" (Singer et al., 2004, Luginaah et al., 2006) are dated after the law. Please reword the beginning of this sentence to tie the information together better

RESPONSE: Text has been revised, first paragraph of Introduction.

3. Line 8, pg3 - Similar to #2 above, ". these and other monitoring approaches", the dates do not align. Please reword.

RESPONSE: Sentence in question has been deleted; second paragraph of Introduction.

4. Line 1, pg4 - please describe or suggest what other urban influences are referred to here.

RESPONSE: Added industrial emissions and population (last paragraph of Introduction).

5. Line 18, pg 4 - why was this classification of traffic volumes used? Were other ranges tested?

RESPONSE: All the potential predictors are listed in the supplementary material. Increments at 10,000 was based on a combination of convenience, prior experience, and the California legislation.

6. Line 10, p5 - no source data WAS available (But this could be a stylistic comment)

RESPONSE: Moved the sentence but left it as "were"; second paragraph, Section 2.1.

7. Line 17, p5 - Please discuss some of the correlations between selected and non-selected variables within the same group. I could not locate this, but it would be of interest.

RESPONSE: The revised paper has been lengthened and we don't feel adding this would be worth the additional length.

8. Line 18, p5 - What does "reasonable" amount of variability mean?

RESPONSE: All ancillary variables had a coefficient of variation > 30% which has been added to the text. We describe variability in Table 1. Text so revised (fourth paragraph, Section 2.1).

9. Line 4, p8 -

a. What was the summer season chosen? If during summer, the highest spatial gradients are observed, what is the implication of what this means for the distribution of pollution during the rest of the year? Or for an annual summary? The issue of seasonality is a common theme for all the LURs and has to be addressed in this paper.

RESPONSE: Budgetary constraints restricted monitoring to one season. The reasons for choosing the summer season were noted in second paragraph of Section 2.3. Higher concentrations will

likely be more suitable for the planned health analyses. We appreciate the reviewer's concern about seasonality. Unfortunately, we could only obtain one season of monitoring and cannot support a seasonal analysis in this paper.

b. Why was only one week chosen?

RESPONSE: The paper has been revised to indicate that weeklong sampling was conducted for the 6 weeks from mid-July to August (second paragraph of Section 2.3). The weeklong sampling period was chosen to better represent the chronic exposure that will be studied in the health analyses.

10. Table 1, p 25 - Each school was ranked by the 8 variables. Does it make sense to provide an average ranking across each school? It would add some meaning to the table.to the table.

RESPONSE: Averaging the rankings across each school would not be appropriate since the variables are of different types. Moreover, it is a school's individual rankings on the different variables that jointly are of importance in site selection relative to other schools.

**Spatial analysis and land use regression of VOCs and NO<sub>2</sub> from school-based urban air monitoring in Detroit/Dearborn, USA**

Shaibal Mukerjee<sup>a,\*</sup>, Luther A. Smith<sup>b</sup>, Mary M. Johnson<sup>c</sup>, Lucas M. Neas<sup>c</sup>, Casson A. Stallings<sup>b</sup>

<sup>a</sup> National Exposure Research Laboratory, U.S. Environmental Protection Agency (E205-02), Research Triangle Park, NC, 27711, USA

<sup>b</sup> Alion Science and Technology, Inc., Research Triangle Park, NC, 27709, USA

<sup>c</sup> National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, 27711, USA

\* Corresponding author. Tel: +1 919 541 1865; fax: +1 919 541 4787

E-mail: [mukerjee.shaibal@epa.gov](mailto:mukerjee.shaibal@epa.gov)

**ABSTRACT**

Passive ambient air sampling for nitrogen dioxide (NO<sub>2</sub>) and volatile organic compounds (VOCs) was conducted at 25 school and two compliance sites in Detroit and Dearborn, Michigan, USA during the summer of 2005. Geographic Information System (GIS) data were calculated at each of 116 schools. The 25 selected schools were monitored to assess and model intra-urban gradients of air pollutants to evaluate impact of traffic and urban emissions on pollutant levels. Schools were chosen to be statistically representative of urban land use variables such as distance to major roadways, traffic intensity around the schools, distance to nearest point sources, population density, and distance to nearest border crossing. Two approaches were used to investigate spatial variability. First, Kruskal-Wallis analyses and pairwise comparisons on data from the schools examined coarse spatial differences based on city section and distance from heavily trafficked roads. Secondly, spatial variation on a finer scale and as a response to multiple factors was evaluated through land use regression (LUR) models

via multiple linear regression. For weeklong exposures, VOCs did not exhibit spatial variability by city section or distance from major roads; NO<sub>2</sub> was significantly elevated in a section dominated by traffic and industrial influence versus a residential section. Somewhat in contrast to coarse spatial analyses, LUR results revealed spatial gradients in NO<sub>2</sub> and selected VOCs across the area. The process used to select spatially representative sites for air sampling and the results of coarse and fine spatial variability of air pollutants provide insights that may guide future air quality studies in assessing intra-urban gradients.

*Keywords:* Air pollution; GIS; Spatial analysis; Land use regression (LUR); Urban air quality; Traffic.

## **1. Introduction**

Recent studies indicate spatial differences exist for gaseous and particulate air pollutants within urban areas and that these intra-urban gradients can occur near major roadways (Brauer et al., 2003; Zhu et al., 2004; Sahsuvaroglu et al., 2006; Henderson et al., 2007). These studies suggest that compliance monitors, typically sited at no more than a few locations in urban areas, may be limited in assessing spatial variability of air pollution for respiratory health studies in those areas. Another approach has been deployment of passive samplers and other field portable monitors at various locations in urban areas to assess intra-urban variability of air pollutants; pollutants studied have included nitrogen dioxide (NO<sub>2</sub>), particulate matter, and volatile organic compounds (VOCs) (Ware et al., 1993; Spicer et al., 1996; Janssen et al., 2001). Similar sites have been deployed at schools to assess spatial influence and/or respiratory health effects associated with near roadway emissions such as NO<sub>2</sub>, sulfur dioxide, and VOCs (Singer et al., 2004; Luginaah et al., 2006; Wheeler et al., 2008). In addition, children's health concerns resulting from air pollutants have influenced enactment of recent legislation in California

1 regarding siting of new schools within 152.4 m (500 feet) of roads with traffic > 100,000  
2 vehicles per day in urban areas (California State Senate Legislation, 2003). Urban network  
3 studies have also used air pollution data in combination with traffic, emissions inventories, and  
4 demographic data to develop spatial models, such as land use regression (LUR) models, that  
5 predict pollutant concentrations at unmonitored locations in the study area to estimate individual  
6 exposures from such ambient sources for spatially based epidemiologic assessments (Briggs et  
7 al., 1997; Brauer et al., 2003; Sahsuvaroglu et al., 2006; Jerrett et al., 2005; Hoek et al., 2008).  
8 This modeling approach has been advocated as being cost effective for assessment of long term  
9 health effects of ambient air pollution (Künzli and Tager, 2000).

10 The U.S. Environmental Protection Agency (EPA) conducted air monitoring studies in  
11 late 1999 at elementary schools in El Paso, Texas and subsequently developed LUR models to  
12 assess intra-urban variability of air pollutants (Smith et al., 2006). Passive air monitors were  
13 deployed to measure ambient levels of VOCs and NO<sub>2</sub>, and LUR models were developed.  
14 Modeled pollutant concentrations were used to assess spatial differences in respiratory health  
15 effects among children attending El Paso schools. School sites for monitoring were selected  
16 based on field observations and sampling convenience. Traffic variables, elevation, population  
17 density, distance from border crossing, and distance to a major oil facility were common  
18 explanatory variables in the regression analyses for VOCs and NO<sub>2</sub>. The observed gradient of  
19 pollutant levels indicated that BTEX species (benzene, toluene, ethylbenzene, o-xylene, and m,p-  
20 xylene) and NO<sub>2</sub> levels were significantly higher in the central section of El Paso, located in a  
21 valley, compared with eastern and western sections which were at higher elevation (Smith et al.,  
22 2006).

23 A similar approach was used in the study presented here. The EPA monitored weeklong

concentrations of VOCs and NO<sub>2</sub> at elementary schools in Detroit and Dearborn, Michigan, USA. This study assessed spatial gradients of NO<sub>2</sub> and VOCs as functions of traffic and other urban influences such as industrial emissions and population density. The approach is presented in terms of how potential explanatory variables were chosen and school sites selected for monitoring to capture air pollution variability across the Detroit/Dearborn area. Note that the procedure described below will not necessarily lead to a unique set of explanatory variables or monitor locations. Overall spatial analyses on a coarse level are presented by comparing city sections and school distances from major roadways. Finer scale variability and the influence of different variables on pollutant levels are assessed through use of LUR models. Estimates from the LUR models will be applied in a children's respiratory health study; the children's health study results are to be presented elsewhere..

## **2. Methods**

### *2.1. Sources and selection of ancillary variables*

Spatially representative school sites were selected (Section 2.2) and LUR models developed (Section 2.5) based on traffic and other urban land-use variables. These geographic information system (GIS) variables were generated using ArcView 3 and 9 (ESRI, Redlands, CA) with statistical analyses implemented in SAS version 9.1 (SAS, 2004a; 2004b).

Initial identification of ancillary variables was based, in part, on previous spatial assessments done in El Paso (Smith et al., 2006) and other urban areas (Sahsuvaroglu et al., 2006; Ross et al., 2006). The general types of ancillary variables were distance to roadways, traffic intensity, population density, and distance to point sources including border crossings. Data sources for variables were: 1) school data, including location, from the National Center of Education Statistics (NCES); 2) traffic volumes from the Southeast Michigan Council of

1 Governments Travel Demand Forecast Model for 2000 based on fixed radii from the school and  
2 distance of school from roads classified into traffic volumes of 10,000 vehicles per day  
3 increments up to 90,000 and more; 3) 2000 U.S. Census data; 4) school distance to nearest point  
4 source of VOCs and PM using emission data from the EPA 1999 National Emission Inventory  
5 database as well as distance to nearest manganese (Mn) point source using emission data from  
6 the Michigan 2002 Toxic Release Inventory database; and 5) school distance to nearest U.S.-  
7 Canada border crossing. Spatial studies of urban areas, including Detroit, have used explanatory  
8 variables such as distance from roadways and population density (Brauer et al, 2003;  
9 Sahsuvaroglu et al., 2006; Henderson et al., 2007; Ross et al., 2006; Jerrett et al., 2005, Wu et al.,  
10 2006; Hoek et al., 2008); these variables were also considered here. Radii used for population  
11 and housing unit density were 125, 250, 500, and 1000 m buffers; traffic intensity within fixed  
12 radii included these buffers and 1500 and 2000 m. Distances to major VOC and PM point  
13 sources were also considered since the Detroit/Dearborn area is impacted by automobile and  
14 other heavy industries (Hammond et al., 2008). (No point source or traffic data for neighboring  
15 Windsor, Ontario were available.) Distance to Mn and PM<sub>2.5</sub> sources were considered because  
16 of their potential use in the subsequent health study. Proximity of schools to nearest  
17 international border crossing (Ambassador Bridge or Detroit-Windsor Tunnel) was used since  
18 these crossings have the largest traffic volumes between the U.S. and Canada. School locations  
19 identified by the NCES database were checked with a global positioning system. Ancillary  
20 variables in categories 2 to 5 listed above generated 45 variables presented in Tables S1 to S4 of  
21 the supplementary data.

22 Explanatory variables were selected from the 45 variables by performing separate correlation  
23 analyses for the variable groups, these being distance to road, traffic intensity, housing unit and

population density, and distance from point sources. The following eight variables were selected as potential explanatory variables: distance to the nearest road with traffic volume of at least 50,000 vehicles per day (DIST\_50KP), distance to the nearest road with traffic volume of at least 90,000 vehicles per day (DIST\_90KP), traffic intensity within 1000 m (INT1000), population density within 500 m (P\_DEN500), distance to the nearest large VOC source emitting 100-1500 lbs/year (VOC\_BIG\_DIST), distance to the nearest large PM<sub>2.5</sub> source emitting > 100 lbs/year (PM25\_BIG\_DIST), distance to the nearest large Mn source emitting 1000-2800 lbs/year (MN\_BIG\_DIST), and distance to the nearest border crossing (BRDR\_DIST). Units for all distances are in m, traffic intensity in vehicles per day/km, and population density is in persons per km<sup>2</sup>. Table 1 presents summary statistics for these variables, and Table 2 indicates the correlation structure among them. Both these tables provide a breakdown of these statistics by monitored and unmonitored schools. No dominant wind direction occurred during the weeklong sampling periods and, therefore, wind direction was not considered in the LUR modeling.

The selected variables exhibited a reasonable amount of variability (coefficient of variation, CV > 30%) across the population of elementary schools. Within these groups, the selected variables to be used in the LUR models were weakly correlated (Pearson correlation coefficient  $\leq 0.7$ ) with each other (see Table 2). However, they often were strongly correlated (Pearson correlation coefficient > 0.7) with some non-selected variables within the same group.

## *2.2. Selection of schools*

One of the goals of this effort was to develop predictive equations for the pollutants across the entire Detroit-Dearborn study area. To ensure that this could be done reliably, it was required that the schools selected as passive monitoring sites span the mathematical space determined by the predictor variables (described in the previous paragraph). Note that the

1 mathematical space is established by the variables' ranges, variabilities, and their overall  
2 correlation structure. Resource limitations dictated that 25 schools could be selected for air  
3 monitoring. Therefore, it was necessary to choose monitoring sites jointly with respect to each  
4 other and with respect to the values of the ancillary variables. All local elementary schools from  
5 pre-kindergarten to Grade 6 in the Detroit and Dearborn Public School systems were candidates  
6 for passive ambient air monitoring (Fig. 1).

7 Distribution of school sites with respect to enumeration area (EA) was used in site  
8 selection since EA was initially considered for selecting participants in the health study. EAs  
9 were defined as a common census tract grouping based on the first two digits of the 2000 U.S.  
10 Census tract number. Near equal numbers of schools were selected per EA to provide distinct  
11 demographic characteristics for the planned health study while maintaining a reasonable  
12 distribution of ancillary variables. Fig. 1 shows the five EAs, 50 to 54, for Detroit; Dearborn  
13 was part of EA 57.

14 School monitoring locations were chosen as follows. The 116 schools were ranked based  
15 on the values of their ancillary variables, with separate rankings for each of the 8 potential  
16 explanatory variables. Each of the 8 rankings was divided into 13 groups of nine; the groups  
17 were numbered from 1 (nine lowest ranked) to 13 (nine highest ranked). For example, the school  
18 denoted as School 9 was near a DIST\_50KP road and, thus, was in group 1 for DIST\_50KP; this  
19 school also had a high INT1000 vehicle count and was in group 12 for INT1000 (Table 3).  
20 Within EAs, schools were selected to reflect the range of combinations of group numbers across  
21 all the potential predictors, rather than simply spacing sites across the Detroit/Dearborn  
22 geographic area. The 25 chosen schools are shown in Fig. 1.

23 For the selected schools, Table 3 indicates by EA each school's group numbers for each

of the eight variables. Note the variety of combinations of group numbers. For example within EA 50, School 9 is in: lower groupings for DIST\_50KP, VOC\_BIG\_DIST, and Mn\_BIG\_DIST; intermediate groupings for DIST\_90KP, POP\_DEN500, and BRDR\_DIST; and high groups for INT1000, and PM25\_BIG\_DIST. In contrast, in the same EA, School 13 is low on INT1000 and POP\_DEN500 and high on the remaining variables.

Four schools were chosen from each EA, except for EA 50 which had five (Fig. 1). Examination of the distribution of the individual ancillary variables (Table 1) showed that the selected and unmonitored schools had comparable ranges for each variable. Also, correlations among the ancillary variables for the same paired variables were similar for the 25 schools chosen versus remaining schools (Table 2). In addition, Table 3 shows that the chosen schools encompass a variety of high, medium, and low combinations of ancillary variables. Finally, an eight dimensional cluster analysis confirmed that the chosen schools were distributed across the various clusters constructed from the total 116 schools. Thus, the collective evidence from Tables 1 to 3 and the cluster analysis indicated that the procedure described above covered the mathematical space spanned by the ancillary variables. This objective process, coupled with actual site visits to confirm feasibility, ensured that the subsequent spatial analysis of the ambient data collected would be based on a representative sample of school sites for Detroit and Dearborn.

### *2.3. Air monitoring*

Passive samplers were deployed outdoors at the 25 selected schools and at two compliance sites operated by the Michigan Department of Environmental Quality (MDEQ), as shown in Fig. 1. Monitoring was done at the MDEQ compliance sites to evaluate LUR model predictions and to compare VOCs and NO<sub>2</sub> measurements with corresponding reference method

1 measurements reported in the EPA Air Quality System (AQS) database. In addition, duplicate  
2 passive samplers were collocated at the compliance sites to evaluate passive sampler precision  
3 (see Mukerjee et al., 2009). Compliance sites were AQS Site 261630033 near the River Rouge  
4 Industrial Complex in Dearborn and AQS Site 261630019 in the East 7 Mile Road area of  
5 northeast Detroit. Further details on monitoring conducted at the compliance sites are discussed  
6 elsewhere (Mukerjee et al., 2009).

7 Ambient monitoring was conducted concurrently at all sites for six weeks from July 18 -  
8 August 30, 2005. Weeklong integrated sampling was chosen to represent chronic ambient  
9 exposures. Evaluation of meteorological data indicated summers in Detroit were characterized  
10 by a greater proportion of stagnation persistence of low wind speeds and stable air masses,  
11 conditions perceived as associated with development of spatial gradients in air pollutants  
12 reflective of local-scale emissions. Previous monitoring studies also suggested that higher  
13 concentrations of ambient particulate air pollution occur in Detroit during the summer (Morishita  
14 et al., 2006).

15 For the present study, concurrent monitoring was conducted at the 25 chosen local  
16 neighborhood schools to reflect children's exposures in the immediate community. Samplers  
17 were placed in shelters and suspended 1.5 to 2 m in height at gas-line cages, fenced courtyards,  
18 or chain link cages designed to minimize vandalism. Shelters were custom-designed, 2.8 liter  
19 size stainless steel bowls to protect samplers from wind and precipitation. All sites were checked  
20 prior to sampler deployment and subsequent to sampler retrieval for immediate VOC influences  
21 using a portable, calibrated MiniRAE 2000 VOC monitor (RAE Systems Inc., Sunnyvale, CA,  
22 USA) accurate to 0.1 ppm levels; no contamination was detected at that level. All sites were  
23 photographed from all directions.

1 Air monitoring was conducted using Ogawa Model 3300 passive samplers for NO<sub>2</sub>  
2 (Ogawa & Co., Pompano Beach, FL, USA) and thermal desorption diffusion tubes packed with  
3 40/60 mesh size, unwashed Carbopack X adsorbent for VOCs (Supelco, Inc., Bellefonte, PA,  
4 USA). These sampling methods have been validated (Yu et al., 2008; Cox, 2003) and were  
5 further evaluated in lab and field studies in Detroit (Mukerjee et al., 2009). Ogawa samplers  
6 were analyzed using ion chromatography. VOC analysis was performed using thermal  
7 desorption followed by GC/MS. Monitoring and chemical analysis procedures for these  
8 samplers are detailed elsewhere (McClenny et al., 2005; 2006; Mukerjee et al., 2004; 2009).

9 Twenty-five VOC species were measured including BTEX species and 1,3-butadiene.  
10 These compounds were used as markers for transportation, refineries and aromatic emissions  
11 (Ware et al., 1993; Fujita, 2001; Buzcu and Fraser, 2006). Process dominated compounds such  
12 as styrene (Ware et al., 1993; Fujita, 2001) were also measured. Nitrogen dioxide was measured  
13 as an indicator of mobile and stationary combustion sources.

#### 14 *2.4. Coarse spatial distinctions*

15 In this study, comparison of the monitored schools' ambient data by their respective EA  
16 and major road distance category DIST\_90KP was performed to assess coarse urban gradients of  
17 air pollution. In addition, it is envisioned that the planned health analyses may be conducted, in  
18 part, by EA. Monitored concentrations of NO<sub>2</sub>, 1,3-butadiene, BTEX species, and styrene were  
19 used for this purpose; see Coarse-Scale Comparisons in the Results section for these analyses.  
20 To avoid data distribution assumptions, nonparametric statistical tests were used in these  
21 comparisons. The Kruskal-Wallis test, using the NPAR1WAY procedure in SAS (SAS, 2004b),  
22 was used to determine whether overall differences existed among the EAs. Subsequent pairwise  
23 comparisons of the EAs and road distance categories were made using a modification of Dunn's

multiple comparison procedure to control false positives within the stated significance level (Dunn, 1964); the comparisons utilized the six week average concentrations at the individual sites. To improve statistical power, the test was modified (Hochberg and Tamhane, 1987) using a critical point value from the studentized range distribution. For the two-sided comparisons for EA and road distance category effect, Dunn's test indicates statistical significance according to whether the following inequality holds:

$$|\bar{R}_i - \bar{R}_{i'}| > \left( \frac{Q_{k,\infty}^{(\alpha)}}{\sqrt{2}} \right) \left\{ \frac{N(N+1)}{12} \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right) \right\}^{\frac{1}{2}}$$

where k reflects the number of EA or road distance category pairwise comparisons being done for each chemical,  $\bar{R}_i$  is the mean rank for the  $i^{\text{th}}$  EA or road distance category,  $n_i$  is the number of observations in the  $i^{\text{th}}$  EA or road distance category, N is the total number of observations over all EAs or road distance categories, and  $Q_{k,\infty}^{(\alpha)}$  is the  $\alpha$ -level critical point from the studentized range distribution. The value of  $Q_{k,\infty}^{(\alpha)}$  was taken from the tabulated critical points of the studentized range tables given by Harter (1960) using  $\alpha=0.05$  and  $k=5$  for the EA comparisons and  $k=3$  for the road distance category effects; all tests were two-sided.

## 2.5. LUR models

The monitored values were plotted against the ancillary variables selected as potential predictors. These plots suggested the use of multiple linear regression, after a logarithmic transformation of the pollutant concentrations. In some cases, the explanatory variables were also logarithmically transformed (see Table 4). Example scatterplots for log transformed  $\text{NO}_2$  are displayed in Figs. S1a-g of the supplementary material.

Residual analyses after the initial regression attempts indicated that the linear regression approach was reasonable in terms of the posited model using the REG procedure in SAS.

1 However, for each pollutant, individual schools had large differences between their measured  
2 and predicted values. These generated large enough variance estimates that very few, if any,  
3 regression coefficients were statistically significant ( $p < 0.05$ ). These departures from the  
4 general pattern were further investigated. However, no clear pattern was evident. For example,  
5 from pollutant to pollutant, it was not always the same school that was high or low. Review of  
6 field comments, lab notes, and photographic documentation at each site revealed no reason for  
7 these departures.

8 To de-emphasize the effect of these sites with large departures, the regressions were  
9 repeated with observations weighted by the inverse of the Cook's D influence statistic (Cook,  
10 1977). A discussion of both transformations and weighting in a regression context can be found  
11 in Carroll and Ruppert (1988). This approach better satisfied the basic regression assumptions  
12 and allowed important predictors to be identified via the significance of their regression  
13 coefficients.

14 Beyond residual analysis, regressions were evaluated for collinearity. Also, cross-  
15 validation was conducted on each regression. The two compliance sites were not utilized in  
16 developing the predictive equations, and regression performance was assessed by comparing the  
17 values measured at these two sites against the predicted values.

### 18 **3. Results**

#### 19 *3.1. Concentrations*

20 Table 5 shows summary statistics of the air pollutants collected at the schools overall and  
21 by EA. Data from passive sampling at the compliance sites was found to be in general  
22 agreement with the continuous monitoring data (Mukerjee et al., 2009). This provided assurance  
23 that the passive data were representative of air quality conditions. Comparison of data from the

El Paso study referenced earlier (Smith et al., 2006) found that median pollutant concentrations in El Paso were comparable to or higher than Detroit, with the exception of o-xylene. The levels for BTEX species in Detroit were also similar or lower in comparison to historical levels measured in other U.S. cities (Singh et al., 1985; Edgerton et al., 1989; Smith et al., 2007). All data were above method detection limits (Table 5). Although the passive samplers were capable of measuring carbon tetrachloride, trichloroethylene, and chlorobenzene, these pollutants were found to be below detection in nearly every case and were not analyzed further.

Pearson correlations of total BTEX suggested the VOC passive data generally tracked well with 7-day integrated canister samples at the Dearborn compliance site ( $r=0.98$ ,  $p<0.01$ ) but less so in comparison to auto-GC measurements at the East 7 Mile compliance site ( $r=0.62$ ,  $p=0.27$ ), which might be due to low ambient levels. The NO<sub>2</sub> passive samplers were also comparable to reference methods ( $r=0.81$ ,  $p=0.05$ ). The BTEX and NO<sub>2</sub> pollutant data from the samplers used here were above method detection limits. Field method evaluations and comparisons of the samplers used in this study with detection limits are detailed elsewhere (Mukerjee et al., 2009).

Based on a similar approach (Spicer et al., 1996), CV using the MEANS procedure in SAS were calculated for: 1) duplicate passive samplers to represent measurement variability, 2) between school sites to represent spatial variability, and 3) between weeks to represent temporal variability. For NO<sub>2</sub> and BTEX species, duplicate CVs were  $\leq 4\%$ , while spatial and temporal CVs were  $\geq 16\%$  (Table 6). The latter reflects the spatial and temporal variation across school sites and weeks and indicates that the variation of pollutant concentrations is attributable to spatial and temporal differences rather than measurement precision.

### *3.2. Coarse-Scale Comparisons*

The following pairwise comparisons were of interest: EA 52 (southwestern Detroit) versus all others, EA 57 (Dearborn) versus all others, and EA 51 (southeastern Detroit) versus EA 54 (western Detroit). EA 52 was compared with the other EAs since this area contained the largest length of freeways and high-volume city roads and was generally downwind of the River Rouge/Zug Island industrial area; EA 57 represented Dearborn for comparison with the other areas in Detroit. In addition, EA 51 was compared to EA 54 since EA 51 also contained heavy industry while EA 54 is generally a residential area.

Table 5 presents, by EA, summary statistics for the pollutants. An overall difference ( $p < 0.05$ ) among EAs was only found for  $\text{NO}_2$ . Pairwise comparisons of EAs using the modified Dunn's test found that  $\text{NO}_2$  levels were significantly higher in EA 52 relative to EA 54 at the 5% level (Table 5); no other significant differences were found.

Additional pairwise comparisons were performed using school distance from road segments with traffic volumes  $\geq 90,000$  vehicles per day. This traffic volume was used in one of the potential predictors and is close to the volume cited in the California school siting legislation. The road distance categories used in these comparisons were 0 – 1.6 km, 1.6 - 4 km, and  $> 4$  km. The modified Dunn's test revealed no differences at the 5% level between the distance categories for any pollutant.

### 3.3. LUR Modeling

Table 4 presents the final LUR models. The explanatory power of these LUR models ranged from quite good for  $\text{NO}_2$  ( $R^2=82\%$ ) to poor for toluene ( $R^2=31\%$ ).  $R^2$  in Table 4 reflects the original scale, not the log-transformed scale (Kvålseth, 1985; Scott and Wild, 1991). This is done for ease in interpretation of LUR predictive power in the field environment.

Unfortunately, the weighting procedure introduced collinearity. Each of the regressions

1 except for toluene and *o*-xylene suffered from this. Collinearity hinders assessment of the  
2 relative importance of individual predictors and possibly the sign of the regression coefficients  
3 because variables may be mathematically related to some extent. Because of this, the LUR  
4 regression coefficients reported in Table 4 must be interpreted cautiously.

5       Using the criteria described by Kleinbaum et al. (1998), Table 7 indicates the sets of  
6 variables within each regression that are subject to having the variables “play off” against each  
7 other in terms of their regression coefficients. For example, Table 7 reports that for the NO<sub>2</sub>  
8 regression a set of five predictors (including the intercept) are involved in collinearity. Each of  
9 the five was reported as being a statistically significant ( $p < 0.05$ ) predictor. What this implies is  
10 that collectively this group of variables is important for predicting NO<sub>2</sub>, but because of the  
11 collinearity issue, the relative importance of the individual predictors is uncertain.

12       Notwithstanding the influence of collinearity, it is interesting that MN\_BIG\_DIST and  
13 PM<sub>25</sub>\_BIG\_DIST seem to be important predictors for several pollutants. To reiterate, Mn and  
14 PM<sub>2.5</sub> sources were originally considered for their utility in the planned children’s health study.  
15 Their appearance in the final equations was investigated by deleting them and recomputing the  
16 regressions. However, residual analyses and both the Akaike (1969) and Bayesian information  
17 criteria indicators (Sawa, 1978) suggested that retaining these predictors was advisable.

18       Cross-validation (Cressie, 1993), whereby monitored schools were left out one at a time  
19 and the model re-estimated, was employed to evaluate the regressions; the resulting predicted  
20 value was then checked against the monitored value not used in the re-estimation. Satisfactory  
21 performance of the models was obtained for benzene, toluene, *o*-xylene, and the BTEX sum.  
22 Good performance was found for 1,3-butadiene, styrene, and NO<sub>2</sub>, though these latter three  
23 compounds had higher root mean squares of approximately 4 to 6 than the desired value of 1.

For ethylbenzene and *m,p*-xylene, cross-validation results were not as good, with mean errors on the order of 3, as opposed to 0, and root mean squares near 16 and 17. For each compound where prediction variability of root mean square error was elevated, this was attributable to large discrepancies between measurements and LUR predictions at one, two, or three sites. NO<sub>2</sub> appeared to be less reliably predicted at sites which were simultaneously close to roads carrying traffic volumes greater than both 50,000 vehicles and 90,000 vehicles per day and were also in an area of heavy traffic intensity. Ethylbenzene and *m,p*-xylene were poorly predicted in locations with combinations of being near roads carrying 50,000 or more vehicles per day, in areas of high traffic intensity, and far from large PM<sub>2.5</sub> sources.

Performance of the regressions was also assessed by comparing the predictions for the LUR effort to measured values from compliance sites (Table 8). Generally, the predictions and measurements agreed well. Of the 18 comparisons in Table 8, 15 show a difference of less than 20%, and 7 predictions are within 10% of the measured value. The worst performance was a discrepancy of 31% for toluene at the Dearborn site. NO<sub>2</sub> was predicted well at the East 7 Mile site.

Figs. 2a-c display the pollutant levels predicted by the LUR models for NO<sub>2</sub>, benzene, and ethylbenzene, respectively for all Detroit/Dearborn local elementary schools. The blank area in the figures represent the cities of Highland Park and Hamtramck, referenced here as Highland Park; see Fig. 1. Fig. 2a shows generally higher predicted NO<sub>2</sub> levels in the south-central section of the Detroit/Dearborn area below Highland Park. For benzene, Fig. 2b indicates lower predicted values in the western portion of the area, higher levels around and to the north and east of Highland Park, and a pocket of higher values in EA 52. Styrene and 1,3-butadiene displayed similar patterns. Fig. 2c indicates that ethylbenzene was similar to benzene in that predictions

were lower to the west and relatively elevated around Highland Park; however, ethylbenzene did not display the group of higher values in the south-central area. The other BTEX species were similar to ethylbenzene.

#### **4. Discussion and conclusions**

School-based ambient air monitoring was successfully conducted in Detroit and Dearborn where sites were statistically chosen based on GIS data to be spatially representative of air pollution explanatory variables throughout the airshed. Analysis on two different coarse levels, geographic sections of the area and categories reflecting distances to heavily trafficked roads, found little variability of VOCs. Spatial uniformity of VOCs also occurred in Columbus, OH based on samples collected for  $\leq 3$  hour integrals (Spicer et al., 1996).  $\text{NO}_2$  exhibited coarse spatial difference between the traffic and industrial-dominated area of southwestern Detroit (EA 52) versus the more residential western section (EA 54). Qualitatively, these results may indicate a mobile source influence on  $\text{NO}_2$  due to the greater length of major road segments in southwestern Detroit and that this industrial section is more impacted than the western section. For distance from very high volume road segments, coarse spatial analysis found no significant differences for  $\text{NO}_2$ . The lack of coarse spatial differences, particularly for VOCs, may have been due to the fact that winds were generally coming from all directions during each week of the study and a pervasive mobile source effect.

LUR modeling revealed spatial gradients for all pollutants. Generally, pollutants were lower in the western part and, except  $\text{NO}_2$ , higher near and to the north and east of Highland Park.  $\text{NO}_2$ , benzene, styrene and 1,3-butadiene also had elevated levels in the south-central portion of the Detroit-Dearborn area. Coarse spatial analyses did find significantly higher  $\text{NO}_2$  levels in the south-central section versus the western portion. This difference is reflected in the LUR

1 predictions.

2 Although the Detroit/Dearborn area has facilities that emit large amounts of VOC and PM<sub>2.5</sub>,  
3 such as the River Rouge area (Hammond et al., 2008), the coarse analyses found no significant  
4 differences in VOC levels between EAs. A recent study of VOCs along commuting routes near  
5 and away from industrial areas in Detroit suggested mobile source emissions dominate other  
6 sources (Batterman et al., 2002). However, LURs obtained in this study indicated that mobile  
7 and point sources may play roles in determining ambient VOC concentrations.

8 Results between coarse analyses and LUR modeling with respect to spatial variability may  
9 have arisen for more than one reason. For coarse analyses, the basic question was whether  
10 certain specific, well-defined city sections differed between themselves for the observed  
11 pollutant levels. Furthermore, though modified to improve its power, the Dunn's test used for  
12 the coarse-scale analysis sacrifices power to guard against false positives.. LUR modeling had as  
13 its goal the prediction of pollutant levels across the entire area by considering multiple variables  
14 simultaneously, without being restricted by any discretization, whereas coarse analyses were  
15 restricted to pollutant levels relative to a single categorized variable. Thus, LUR is more  
16 amenable to detecting spatial gradients across the entire area, while the coarse analyses are  
17 applicable in assessing section-to-section differences.

18 Utility of LUR regressions is very much dependent on the "pre-analysis" described above,  
19 whereby both monitoring locations and potential predictors are determined beforehand. Recall  
20 that this procedure does not necessarily lead to a unique set of explanatory variables or monitor  
21 locations. Incorporating statistical considerations of the pre-analysis into the LUR approach  
22 provides greater confidence that the appropriate mathematical space has been identified and  
23 covered by the monitoring locations. Therefore, there is higher confidence in the use of the

1 predictive equations to determine pollutant levels across the entire study area. Similar, though  
2 not identical, pre-analyses like population-weighted location-allocation models have been  
3 emphasized in other LUR studies (Sahsuvaroglu et al., 2006; Henderson et al., 2007).

4 Since an EA occupies a large fraction of the area, each can contain a range of pollutant  
5 levels. Pollutant gradients seen in Figs. 2a-c suggest that pollutant levels do not change abruptly  
6 across the area. This combination may explain why all EA levels were statistically the same,  
7 except one pair.

8 Since MN\_BIG\_DIST and PM25\_BIG\_DIST were considered for utility in the health study,  
9 their appearance in some predictive equations is puzzling. While their relative importance as  
10 predictors is somewhat uncertain due to collinearity, their use in these predictions appears  
11 justified based on analyses of models excluding these variables. MN\_BIG\_DIST and  
12 PM25\_BIG\_DIST may be acting as surrogates for variables not considered in the models. For  
13 example, large manganese and PM<sub>2.5</sub> sources considered here might also be smaller emitters of  
14 NO<sub>2</sub> or VOCs. Only large VOC sources could enter the predictive equations, and  
15 MN\_BIG\_DIST and PM25\_BIG\_DIST may be substituting for distances to lower level VOC or  
16 NO<sub>2</sub> emitters.

17 Previous research has found spatial gradients for these pollutants under similar sampling  
18 intervals at El Paso schools (Smith et al., 2006). There, the center section exhibited significantly  
19 higher VOC and NO<sub>2</sub> levels. El Paso's proximity to Ciudad Juárez, Mexico and the area's  
20 complex terrain were likely reasons that coarse-level spatial gradients were found, versus Detroit  
21 with flat terrain and smaller cross-border influences.

22 The results presented here demonstrated the utility of coarse scale analyses for examining  
23 differences between specific subsets of a geographic area and LUR modeling for finer scale

descriptions of an entire metropolitan area. This study has emphasized the importance of statistical pre-analysis in selecting explanatory variables and site locations and of diagnostic analyses in determining final LUR models.

#### **Acknowledgements and disclaimer**

We thank Hunter Daughtrey, Karen Oliver, Herb Jacumin, Laura Liao, Chris Fortune, Mike Wheeler, and Dennis Williams, all from Alion Science and Technology, Felicia Venable and Mathew Sam of the Detroit Public Schools, Don Ball of Dearborn Public Schools, and principals and staff. We also thank Mary Ann Heindorf, Ann Chevalier, Deborah Sherrod, Amy Robinson, and Craig Fitzner for access to MDEQ sites and data and Ron Williams and Alan Vette of EPA for manuscript review. Finally, we thank Ann Williams, Gina Andrews, Paul Killough of EPA, and John Archer of Eastern Research Group for assistance. The U.S. Environmental Protection Agency through its Office of Research and Development funded and managed the research described here under contract EP-D-05-065 to Alion. The paper has been subjected to Agency review and approved for publication.

#### **Supplementary data**

Tables of ancillary variables considered for use in LUR models and scatterplots of log transformed NO<sub>2</sub> by predictor variables. This material is available via the Internet at <http://www.sciencedirect.com>.

#### **REFERENCES**

- Akaike, H., 1969. Fitting autoregressive models for prediction. *Ann. I. Stat. Math.* 21, 243-247.
- Batterman, S.A., Peng, C.Y., Braun, J., 2002. Levels and composition of volatile organic compounds on commuting routes in Detroit, Michigan. *Atmos. Environ.* 36, 6015–6030.
- Brauer, M., Hoek, G., van Vliet, P., Meliefste, K., Fischer, P., Gehring, U., Heinrich, J., Cyrys,

J., Bellander, T., Lewne, M., Brunekreef, B., 2003. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology* 14 228-239.

Briggs, D.J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebre, E., Pryl, K., Van Reeuwijk, H., Smallbone, K., Van Der Veen, A., 1997. Mapping air pollution using GIS: a regression approach. *Int. J. Geogr. Inf. Sci.* 11 699-718.

Buzcu, B., Fraser, M.P., 2006. Source identification and apportionment of volatile organic compounds in Houston, TX. *Atmos. Environ.* 40, 2385-2400.

California State Senate Legislation, October 3, 2003. Senate Bill 352, Chapter 668, An act to amend Section 17213 of the Education Code, and to amend Section 21151.8 of the Public Resources Code, relating to public schools. (Internet version [http://info.sen.ca.gov/pub/03-04/bill/sen/sb\\_0351-0400/sb\\_352\\_bill\\_20031003\\_chaptered.pdf](http://info.sen.ca.gov/pub/03-04/bill/sen/sb_0351-0400/sb_352_bill_20031003_chaptered.pdf))

Carroll, R.J., Ruppert, D., 1988. Transformation and weighting in regression. Chapman and Hall, New York, pp. 115-208.

Cook, R.D., 1977. Detection of influential observations in linear regression. *Technometrics* 19, 15-18.

Cox, R.M., 2003. The use of passive sampling to monitor forest exposure to O<sub>3</sub>, NO<sub>2</sub> and SO<sub>2</sub>: a review and some case studies. *Environ. Pollut.* 126, 301–311.

Cressie, N.A.C., 1993. Statistics for spatial data, revised edition. Wiley, New York, pp. 101-104.

Dunn, O.J., 1964. Multiple comparisons using rank sums. *Technometrics* 6, 241-252.

Edgerton, S.A., Holdren, M.W., Smith, D.L., Shah, J.J., 1989. Inter-urban comparison of ambient volatile organic compound concentrations in U.S. cities. *J. Air Waste Manage.*

Assoc. 39, 729-732.

Fujita, E.M., 2001. Hydrocarbon source apportionment for the 1996 Paso del Norte Ozone Study. *Sci. Total Environ.* 276, 171-184.

Hammond, D., Dvonch, J.T., Keeler, G.J., Parker, E.A., Kamal, A.S., Barres, J.A., Yip, F.Y., Brakefield-Caldwell, W., 2008. Sources of ambient fine particulate matter at two community sites in Detroit, Michigan. *Atmos. Environ.* 42, 720-732.

Harter, H.L., 1960. Tables of range and studentized range. *Ann. Math. Stat.* 31, 1122-1147.

Henderson, S.B., Beckerman, B., Jerrett, M., Brauer, M., 2007. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ. Sci. Technol.* 41, 2422-2428.

Hochberg, Y., Tamhane, A.C., 1987. Multiple comparison procedures. Wiley, New York, pp. 244-246.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* 42, 7561-7578.

Janssen, N.A.H., van Vliet, P.H.N., Aarts, F., Harssema, H., Brunekreef, B., 2001. Assessment of exposure to traffic related air pollution of children attending schools near motorways. *Atmos. Environ.* 35, 3875-3884.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuvaroglu, T., Morrison J., Giovis, C., 2005b. A review and evaluation of intraurban air pollution exposure models. *J. Exp. Anal. Environ. Epidemiol.* 15, 185-204.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., Nizam, A., 1988. Applied regression analysis and other multivariable methods. Duxbury Press, Pacific Grove, CA, pp. 237-248.

1 Künzli, N., Tager, I.B., 2000. Long-term health effects of particulate and other ambient air  
 2 pollution: Research can progress faster if we want it to. *Environ. Health Persp.* 108, 915-  
 3 918.  
 4 Kvalseth, T.A., 1985. Cautionary rule about  $R^2$ . *Am. Statist.* 39, 279-285.  
 5 Luginaah, I., Xu, X., Fung, K.Y., Grgicak-Mannion, A., Wintermute, J., Wheeler, A., Brook, J.,  
 6 2006. Establishing the spatial variability of ambient nitrogen dioxide in Windsor,  
 7 Ontario. *Int. J. Environ. Stud.* 63, 487-500.  
 8 McClenny, W.A., Oliver, K.D., Jacumin, H.H., Daughtrey, E.H., Whitaker, D.A., 2005. 24 h  
 9 diffusive sampling of toxic VOCs in air onto Carbopack X solid adsorbent followed by  
 10 thermal desorption/GC/MS analysis—laboratory studies. *J. Environ. Monit.* 7, 248-256.  
 11 McClenny, W.A., Jacumin, H.H., Oliver, K.D., Daughtrey, E.H., Whitaker, D.A., 2006.  
 12 Comparison of 24 h averaged VOC monitoring results for residential indoor and outdoor  
 13 air using Carbopack X-filled diffusive samplers and active sampling—a pilot study. *J.*  
 14 *Environ. Monit.* 8, 263-269.  
 15 Morishita, M., Keeler, G.J., Wagner, J.G., Harkema, J.R., 2006. Source identification of ambient  
 16 PM<sub>2.5</sub> during summer inhalation exposure studies in Detroit, MI. *Atmos. Environ.* 40,  
 17 3823-3834.  
 18 Mukerjee, S., Oliver, K.D., Seila, R.L., Jacumin, H.H., Croghan, C., Daughtrey, E.H., Neas,  
 19 L.M., Smith, L.A., 2009. Field comparison of passive air samplers with reference  
 20 monitors for ambient volatile organic compounds and nitrogen dioxide under week-long  
 21 integrals. *J. Environ. Monit.* 11, 220-227.  
 22 Mukerjee S., Smith L.A., Norris G.A., Morandi M.T., Gonzales M., Noble C.A., Neas L.M., and  
 23 Özkaynak A.H. 2004. Field method comparison between passive air samplers and

continuous monitors for volatile organic compounds and NO<sub>2</sub> in El Paso, Texas, USA. J. Air Waste Manage. Assoc. 54, 307-319.

Ross, Z, English, P.B., Scalf, R., Gunier, R., Smorodinsky, S., Wall, S., Jerrett, M., 2006. Nitrogen dioxide prediction in Southern California using land use regression modeling: potential for environmental health analyses. J. Exp. Sci. Environ. Epidemiol. 16, 106-114.

Sahsuvaroglu, T., Arain, A., Kanaroglou, P., Finkelstein, N., Newbold, B., Jerrett, M., Beckerman, B., Brook, J., Finkelstein, M., Gilbert, N.L., 2006. A land use regression model for predicting ambient concentrations of nitrogen dioxide in Hamilton, Ontario, Canada. J. Air Waste Manage. Assoc. 56, 1059-1069.

SAS, 2004a. Base SAS® 9.1.3 procedures guide. SAS Institute Inc., Cary, NC, pp. 524-550.

SAS, 2004b. SAS/STAT® 9.1 user's guide. SAS Institute Inc., Cary, NC, pp. 1291-1320, 3143-3162.

Sawa, T., 1978. Information criteria for discriminating among alternative regression models. Econometrica 46, 1273-1282.

Scott, A., Wild, C., 1991. Transformations and R<sup>2</sup>. Am. Statist. 45, 127-129.

Singer, B.C., Hodgson, A.T., Hotchi, T., Kim, J.J., 2004. Passive measurement of nitrogen oxides to assess traffic-related pollutant exposure for the East Bay Children's Respiratory Health Study. Atmos. Environ. 38, 393-403.

Singh, H.B., Salas, L.J., Cantrell, B.K., Redmond, R.M., 1985. Distribution of aromatic hydrocarbons in the ambient air. Atmos. Environ. 19, 1911-1919.

Smith, L., Mukerjee, S., Gonzales, M., Stallings, C., Neas, L., Norris, G., Özkaynak, H., 2006. Use of GIS and ancillary variables to predict volatile organic compound and nitrogen

dioxide levels at unmonitored locations. *Atmos. Environ.* 40, 3773–3787.

Smith, L.A., Stock, T.H., Chung, K.C., Mukerjee, S., Liao, X.L., Stallings, C., Afshar, M., 2007. Spatial analysis of volatile organic compounds from a community-based air toxics monitoring network in Deer Park, Texas, USA. *Environ. Monit. Assess.* 128, 369–379.

Spicer, C.W., Buxton, B.E., Holdren, M.W., Smith, D.L., Kelly, T.J., Rust, S.W., Pate, A.D., Sverdrup, G.M., Chuang, J.C., 1996 Variability of hazardous air pollutants in an urban area. *Atmos. Environ.* 1996, 30 3443-3456.

Ware, J.H., Spengler, J.D., Neas, L.M., Samet, J.M., Wagner, G.R., Coultas, D., Özkaynak, H., Schwab, M., 1993. Respiratory and irritant health effects of ambient volatile organic compounds: the Kanawha County Health Study. *Am. J. Epidemiol.* 137, 1287-1301.

Wheeler, A.J., Smith-Doiron, M., Xu, X., Gilbert, N.L., Brook, J.R., 2008. Intra-urban variability of air pollution in Windsor, Ontario—Measurement and modeling for human exposure assessment. *Environ. Res.* 106, 7-16.

Wu, Y.C., Batterman, S.A., 2006. Proximity of schools in Detroit, Michigan to automobile and truck traffic. *Journal of Exposure Science and Environmental Epidemiology* 16, 457-470.

Yu, C.H., Morandi, M.T., Weisel, C.P., 2008. Passive dosimeters for nitrogen dioxide in personal/indoor air sampling: A review. *J. Expo. Sci. Environ. Epid.* 18, 441-451.

Zhu, Y., Hinds, W.C., Shen, S., Sioutas, C. 2004. Seasonal trends of concentration and size distribution of ultrafine particles near major highways in Los Angeles. *Aerosol Sci. Technol.* 2004, 38 5-13.

Table 1

Summary statistics of ancillary variables for monitored and unmonitored schools

| Ancillary variable          | Monitored schools (n=25) |         |         | Unmonitored schools (n=91) |        |         |
|-----------------------------|--------------------------|---------|---------|----------------------------|--------|---------|
|                             | Minimum                  | Median  | Maximum | Minimum                    | Median | Maximum |
| DIST_50KP <sup>a</sup>      | 0                        | 806     | 4,518   | 100                        | 1,389  | 5,400   |
| DIST_90KP <sup>b</sup>      | 500                      | 4,000   | 8,905   | 500                        | 4,046  | 10,655  |
| INT1000 <sup>c</sup>        | 24,711                   | 118,398 | 249,513 | 17,872                     | 64,728 | 493,302 |
| POP_DEN500 <sup>d</sup>     | 1,319                    | 3,039   | 4,659   | 950                        | 3,216  | 6,339   |
| VOC_BIG_DIST <sup>e</sup>   | 1,476                    | 3,940   | 11,229  | 1,237                      | 4,478  | 13,124  |
| PM2.5_BIG_DIST <sup>f</sup> | 2,600                    | 10,781  | 21,570  | 2,154                      | 12,368 | 23,278  |
| Mn_BIG_DIST <sup>g</sup>    | 2,941                    | 7,901   | 17,016  | 1,910                      | 8,670  | 19,413  |
| BRDR_DIST <sup>h</sup>      | 1,208                    | 9,306   | 19,482  | 943                        | 11,416 | 20,461  |

<sup>a</sup> DIST\_50KP: distance to nearest road with traffic volume  $\geq 50,000$  vehicles per day (m)

<sup>b</sup> DIST\_90KP: distance to nearest road with traffic volume  $\geq 90,000$  vehicles per day (m)

<sup>c</sup> INT1000: traffic intensity within 1000 m of location (vehicles per day / km)

<sup>d</sup> POP\_DEN500: population within a 500 m radius of location

<sup>e</sup> VOC\_BIG\_DIST: distance (m) to nearest large (45-680 kg/year) VOC emission source

<sup>f</sup> PM2.5\_BIG\_DIST: distance (m) to nearest large ( $>45$  kg/year) PM2.5 emission source

<sup>g</sup> Mn\_BIG\_DIST: distance (m) to nearest large (454-1270 kg/year) Manganese emission source

<sup>h</sup> BRDR\_DIST: distance to nearest border crossing (m)

**TABLE 2 - Pearson correlations between explanatory variables considered for LURs<sup>a</sup>**

|                    | DIST_<br>50KP | DIST_<br>90KP | INT<br>1000 | POP_DEN<br>500 | VOC_BIG_<br>DIST | PM2.5_BIG_<br>DIST | Mn_BIG_<br>DIST | BRDR_<br>DIST |
|--------------------|---------------|---------------|-------------|----------------|------------------|--------------------|-----------------|---------------|
| DIST_50KP          | 1             | 0.38          | -0.66       | -0.02          | 0.07             | 0.04               | 0.53            | 0.45          |
| DIST_90KP          | 0.26          | 1             | -0.28       | 0.22           | 0.06             | 0.31               | 0.18            | 0.21          |
| INT1000            | -0.56         | -0.24         | 1           | -0.28          | -0.14            | 0.04               | -0.30           | -0.19         |
| POP_DEN500         | -0.10         | 0.12          | -0.17       | 1              | 0.13             | 0.08               | -0.03           | -0.15         |
| VOC_BIG_<br>DIST   | -0.15         | -0.08         | 0.26        | 0.02           | 1                | 0.27               | 0.66            | 0.59          |
| PM2.5_BIG_<br>DIST | 0.31          | -0.02         | -0.10       | 0.07           | 0.13             | 1                  | -0.09           | 0.44          |
| Mn_BIG_<br>DIST    | 0.07          | 0.09          | 0.16        | -0.05          | 0.70             | -0.27              | 1               | 0.68          |
| BRDR_DIST          | 0.29          | 0.25          | 0.04        | -0.02          | 0.56             | 0.40               | 0.59            | 1             |

<sup>a</sup> Schools with passive sampling data (n=25) appear in the upper triangular portion of the matrix (i.e., above the diagonal of 1s); correlations within the group of unmonitored schools (n=91) appear in the lower triangular portio

1 **Table 3 - Group rankings for explanatory variables at monitored schools (see Selection of Schools Section)**

| School number            | Group DIST_50KP                   | Group DIST_90KP | Group INT1000            | Group POP_DEN500                    | Group VOC_BIG_DIST      | Group PM2.5_BIG_DIST            | Group Mn_BIG_DIST            | Group BRDR_DIST              |
|--------------------------|-----------------------------------|-----------------|--------------------------|-------------------------------------|-------------------------|---------------------------------|------------------------------|------------------------------|
| Enumeration area (EA) 50 |                                   |                 |                          |                                     |                         |                                 |                              |                              |
| 3                        | 3                                 | 1               | 11                       | 1 (lowest population density group) | 6                       | 9                               | 2                            | 7                            |
| 9                        | 1 (near road of specified volume) | 6               | 12 (high traffic volume) | 6                                   | 4                       | 11                              | 3                            | 5                            |
| 11                       | 6                                 | 12              | 8                        | 13                                  | 8                       | 13                              | 6                            | 8                            |
| 12                       | 1                                 | 12              | 12                       | 12                                  | 6                       | 12                              | 6                            | 7                            |
| 13                       | 10                                | 13              | 1                        | 3                                   | 9                       | 13                              | 9                            | 9                            |
| EA 51                    |                                   |                 |                          |                                     |                         |                                 |                              |                              |
| 4                        | 2                                 | 1               | 11                       | 2                                   | 3                       | 8                               | 1 (close to large Mn source) | 4                            |
| 6                        | 5                                 | 7               | 9                        | 12                                  | 8                       | 5                               | 6                            | 1 (close to border crossing) |
| 7                        | 11                                | 11              | 8                        | 3                                   | 1 (close to VOC source) | 11                              | 7                            | 5                            |
| 8                        | 4                                 | 4               | 9                        | 5                                   | 2                       | 11                              | 2                            | 4                            |
| EA 52                    |                                   |                 |                          |                                     |                         |                                 |                              |                              |
| 24                       | 2                                 | 13              | 9                        | 4                                   | 1                       | 1 (close to large PM2.5 source) | 3                            | 2                            |
| 25                       | 6                                 | 7               | 1                        | 11                                  | 4                       | 2                               | 6                            | 1                            |
| 26                       | 2                                 | 7               | 12                       | 11                                  | 5                       | 2                               | 6                            | 1                            |
| 27                       | 3                                 | 1               | 10                       | 3                                   | 6                       | 4                               | 5                            | 1                            |
| EA 53                    |                                   |                 |                          |                                     |                         |                                 |                              |                              |

|                  |    |    |    |    |    |    |    |    |
|------------------|----|----|----|----|----|----|----|----|
| 1                | 5  | 7  | 9  | 7  | 10 | 10 | 10 | 12 |
| 2                | 9  | 9  | 4  | 12 | 8  | 8  | 8  | 10 |
| 5                | 1  | 2  | 9  | 13 | 4  | 6  | 2  | 3  |
| 14               | 5  | 8  | 6  | 11 | 11 | 6  | 6  | 5  |
| EA 54            |    |    |    |    |    |    |    |    |
| 15               | 10 | 11 | 7  | 9  | 13 | 11 | 13 | 13 |
| 16               | 1  | 3  | 13 | 1  | 12 | 6  | 13 | 12 |
| 17               | 9  | 3  | 1  | 10 | 10 | 4  | 12 | 11 |
| 19               | 7  | 2  | 4  | 6  | 7  | 2  | 10 | 6  |
| EA 57 (Dearborn) |    |    |    |    |    |    |    |    |
| 18               | 13 | 11 | 7  | 4  | 3  | 4  | 12 | 12 |
| 20               | 10 | 4  | 7  | 10 | 5  | 1  | 9  | 5  |
| 21               | 7  | 6  | 7  | 1  | 4  | 1  | 9  | 6  |
| 22               | 3  | 9  | 12 | 4  | 1  | 1  | 8  | 4  |

1  
2

1 **Table 4 - LUR models developed for Detroit-Dearborn passive monitoring**

| Response                  | Predictive equation                                                                                                                                                                                                  | R <sup>2</sup> (%) |
|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| <b>Log(NO2)</b>           | <b>4.43</b> – <b>0.05 Log(DIST_50KP)</b> + <b>0.02 Log(DIST_90KP)</b> + 0.01 Log(INT1000) + 0.02 * Log(VOC_BIG_DIST) – <b>0.10 * Log(PM25_BIG_DIST)</b> – <b>0.08 Log(MN_BIG_DIST)</b> – <b>1.5 E-5 * BRDR_DIST</b>  | 82                 |
| <b>Log(Benzene)</b>       | <b>6.23</b> – <b>5.04 E-5 * DIST_50KP</b> – 3.4 E-8 * INT1000 + <b>5.99 E-5 * P_DEN500</b> – <b>3.8 E-5 * MN_BIG_DIST</b> + <b>1.23 E-5 * BRDR_DIST</b>                                                              | 43                 |
| <b>Log(Toluene)</b>       | <b>7.42</b> – <b>4.17 E-5 * MN_BIG_DIST</b> + <b>1.92 E-5 * BRDR_DIST</b>                                                                                                                                            | 31                 |
| <b>Log(Ethylbenzene)</b>  | <b>10.04</b> – 0.04 * Log(DIST_50KP) – 2.91 E-6 * DIST_90KP – 0.05 * Log(INT1000) – 8.45 E-6 * VOC_BIG_DIST – <b>0.46* Log(MN_BIG_DIST)</b> + <b>1.79 E-5 * BRDR_DIST</b>                                            | 63                 |
| <b>Log(m,p-xylene)</b>    | <b>8.43</b> – <b>1 E-4 * DIST_50KP</b> – <b>2.81 E-5 * DIST_90KP</b> – <b>1.51 E-6 * INT1000</b> – 0.13 * Log(VOC_BIG_DIST) + <b>0.16* Log(PM25_BIG_DIST)</b> – <b>0.22* Log(MN_BIG_DIST)</b> – 6.02 E-6 * BRDR_DIST | 55                 |
| <b>Log(o-xylene)</b>      | <b>6</b> – 0.04 * Log(DIST_50KP) – <b>2.34 E-5 * DIST_90KP</b> + <b>0.24 * Log(P_DEN500)</b> – 1.34 E-5 * VOC_BIG_DIST + <b>0.09 * Log(PM25_BIG_DIST)</b> – <b>0.34 * Log(MN_BIG_DIST)</b> + 5.42 E-6 * BRDR_DIST    | 60                 |
| <b>Log(BTEX_sum)</b>      | <b>10.78</b> – 4.52 E-5 * DIST_50KP – <b>0.30 * Log(MN_BIG_DIST)</b>                                                                                                                                                 | 40                 |
| <b>Log(styrene)</b>       | <b>3.75</b> + 2.82 E-6 * DIST_50KP – <b>2.47 E-5 * DIST_90KP</b> + 2.88 E-7 * INT1000 + 2.53 E-5 * P_DEN500 – <b>2.03 E-5 * MN_BIG_DIST</b> – 3.45 E-6 * BRDR_DIST                                                   | 43                 |
| <b>Log(1,3-butadiene)</b> | <b>4.56</b> – 0.05 * Log(DIST_50KP) + 4.1 E-8 * INT1000 + <b>1 E-4 * P_DEN500</b> + 0.01 * Log(VOC_BIG_DIST) – 2.82 E-6 * PM25_BIG_DIST – <b>2.95 E-5 * MN_BIG_DIST</b> + 1.61 E-6 * BRDR_DIST                       | 43                 |

2 Notes: Bold indicates regression coefficients significant at the 5% level. Log is the natural logarithm. R<sup>2</sup> is reported for the original  
3 scale, not the log-transformed scale.

**Table 5 - Median values of NO<sub>2</sub> and selected VOCs (all in ppbV) for all schools and each enumeration area (EA)<sup>a</sup>**

| Pollutant       | MDL <sup>b</sup> | All schools<br>(n=25)   | EA                      |                         |                                |                         |                                |                         |
|-----------------|------------------|-------------------------|-------------------------|-------------------------|--------------------------------|-------------------------|--------------------------------|-------------------------|
|                 |                  |                         | 50<br>(n=5)             | 51<br>(n=4)             | 52<br>(n=4)                    | 53<br>(n=4)             | 54<br>(n=4)                    | 57<br>(n=4)             |
| NO <sub>2</sub> | 1.00             | 15.63<br>(11.49, 23.59) | 14.21<br>(11.49, 18.58) | 16.00<br>(12.56, 16.20) | <b>21.15</b><br>(18.50, 22.20) | 15.96<br>(13.65, 23.59) | <b>13.58</b><br>(12.98, 15.08) | 15.52<br>(13.80, 15.98) |
| 1,3-butadiene   | 0.04             | 0.07<br>(0.05, 0.13)    | 0.08<br>(0.05, 0.10)    | 0.07<br>(0.05, 0.10)    | 0.09<br>(0.06, 0.13)           | 0.08<br>(0.07, 0.11)    | 0.06<br>(0.05, 0.07)           | 0.06<br>(0.05, 0.07)    |
| Benzene         | 0.02             | 0.46<br>(0.34, 0.70)    | 0.48<br>(0.39, 0.61)    | 0.47<br>(0.34, 0.65)    | 0.52<br>(0.44, 0.70)           | 0.52<br>(0.46, 0.62)    | 0.40<br>(0.34, 0.45)           | 0.39<br>(0.35, 0.45)    |
| Toluene         | 0.02             | 1.40<br>(0.98, 1.99)    | 1.54<br>(1.23, 1.98)    | 1.50<br>(1.00, 1.99)    | 1.45<br>(1.20, 1.85)           | 1.56<br>(1.40, 1.75)    | 1.26<br>(1.09, 1.42)           | 1.11<br>(0.98, 1.39)    |
| Ethylbenzene    | 0.02             | 0.18<br>(0.12, 0.36)    | 0.20<br>(0.16, 0.25)    | 0.22<br>(0.12, 0.36)    | 0.20<br>(0.16, 0.25)           | 0.20<br>(0.17, 0.31)    | 0.17<br>(0.14, 0.19)           | 0.15<br>(0.13, 0.18)    |
| m,p-xylene      | 0.04             | 0.59<br>(0.36, 1.23)    | 0.63<br>(0.49, 0.79)    | 0.73<br>(0.36, 1.23)    | 0.64<br>(0.52, 0.83)           | 0.65<br>(0.56, 1.04)    | 0.54<br>(0.41, 0.62)           | 0.46<br>(0.40, 0.59)    |
| o-xylene        | 0.02             | 0.20<br>(0.12, 0.34)    | 0.20<br>(0.16, 0.28)    | 0.25<br>(0.12, 0.31)    | 0.21<br>(0.17, 0.28)           | 0.22<br>(0.19, 0.34)    | 0.18<br>(0.15, 0.21)           | 0.16<br>(0.13, 0.20)    |
| Total BTEX      |                  | 2.81<br>(1.95, 4.11)    | 3.06<br>(2.43, 3.90)    | 3.39<br>(1.95, 4.11)    | 3.04<br>(2.49, 3.90)           | 3.15<br>(2.79, 4.06)    | 2.56<br>(2.13, 2.87)           | 2.26<br>(2.02, 2.81)    |
| Styrene         | 0.02             | 0.03<br>(0.03, 0.06)    | 0.03<br>(0.03, 0.04)    | 0.04<br>(0.03, 0.06)    | 0.04<br>(0.04, 0.05)           | 0.04<br>(0.03, 0.05)    | 0.03<br>(0.03, 0.04)           | 0.03<br>(0.03, 0.04)    |

<sup>a</sup> Minimum and maximum values in parentheses. Values in bold indicate significant difference of pollutant concentrations between pairs of EAs. See Fig. 1 for EA locations.

<sup>b</sup> Method detection limit

**Table 6 - Coefficients of variation (CV) for duplicate samplers at compliance monitoring stations and CV over all school sites**

| Pollutant          | East 7 Mile               | Dearborn                  | School sites            |                          |
|--------------------|---------------------------|---------------------------|-------------------------|--------------------------|
|                    | Duplicate CV <sup>1</sup> | Duplicate CV <sup>1</sup> | Spatial CV <sup>2</sup> | Temporal CV <sup>3</sup> |
| NO <sub>2</sub>    | 2                         | 3                         | 20                      | 16                       |
| 1,3-butadiene      | 9                         | 11                        | 26                      | 31                       |
| Benzene            | 2                         | 4                         | 20                      | 24                       |
| Toluene            | 1                         | 2                         | 20                      | 25                       |
| Ethylbenzene       | 1                         | 4                         | 28                      | 25                       |
| <i>o</i> -xylene   | 1                         | 3                         | 27                      | 26                       |
| <i>m,p</i> -xylene | 1                         | 4                         | 32                      | 26                       |
| Styrene            | 4                         | 6                         | 21                      | 19                       |

<sup>1</sup> CV calculated from 6 pairs. Summarization presents the mean of CVs by week.

<sup>2</sup> CV calculated from mean pollutant values from the 25 school sites.

<sup>3</sup> Temporal CV calculated by site. Tabulated values are the average CV across sites.

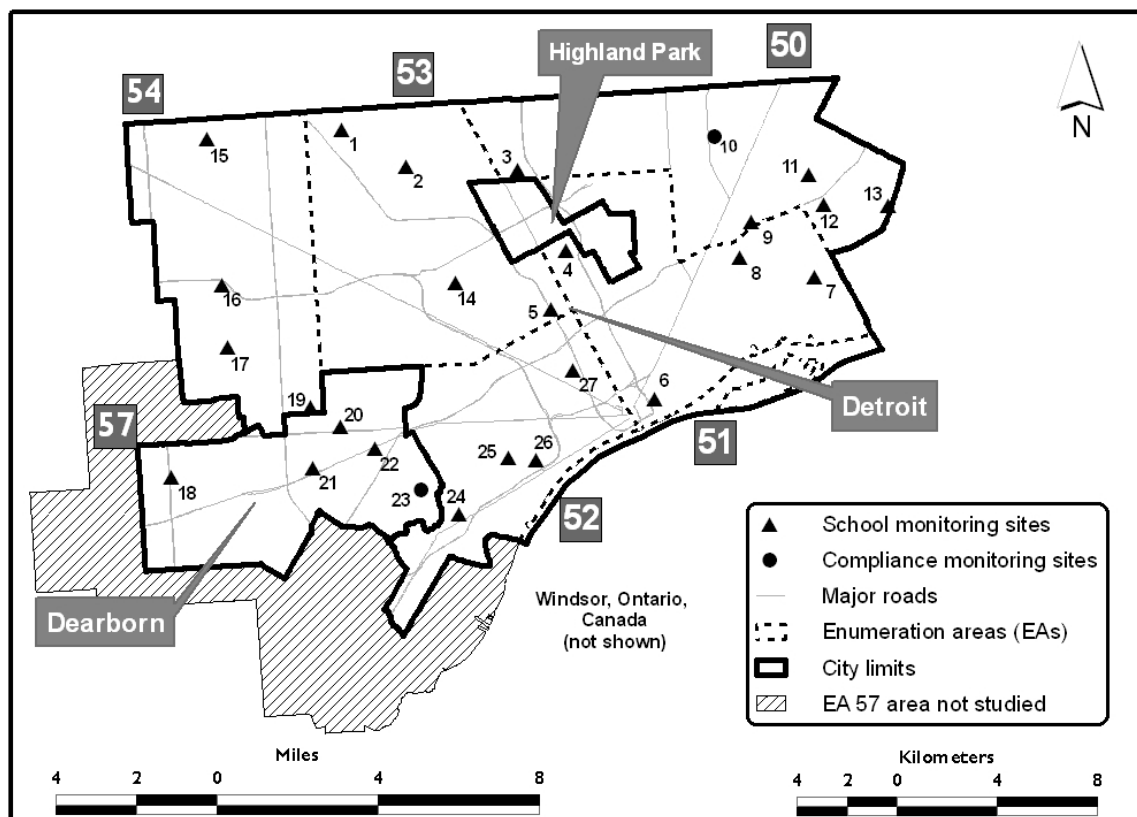
**Table 7 - Variables associated with problematic regression coefficients due to collinearity**

| Response compound  | Comingled variable sets                                                 |
|--------------------|-------------------------------------------------------------------------|
| NO <sub>2</sub>    | (int, DIST_50KP, PM25_BIG_DIST, MN_BIG_DIST, BRDR_DIST)                 |
| benzene            | (int, DIST_50KP, INT1000, P_DEN500, BRDR_DIST)                          |
| ethylbenzene       | (DIST_50KP, MN_BIG_DIST)<br>(int, INT1000)<br>(VOC_BIG_DIST, BRDR_DIST) |
| <i>m,p</i> -xylene | (VOC_BIG_DIST, PM25_BIG_DIST)<br>(int, BRDR_DIST)                       |
| BTEX sum           | (int, MN_BIG_DIST)                                                      |
| stryrene           | (int, INT1000, P_DEN500)<br>(MN_BIG_DIST, BRDR_DIST)                    |
| 1,3-butadiene      | (int, VOC_BIG_DIST, BRDR_DIST)<br>(PM25_BIG_DIST, MN_BIG_DIST)          |

**Note: “int” refers to the regression intercept.**

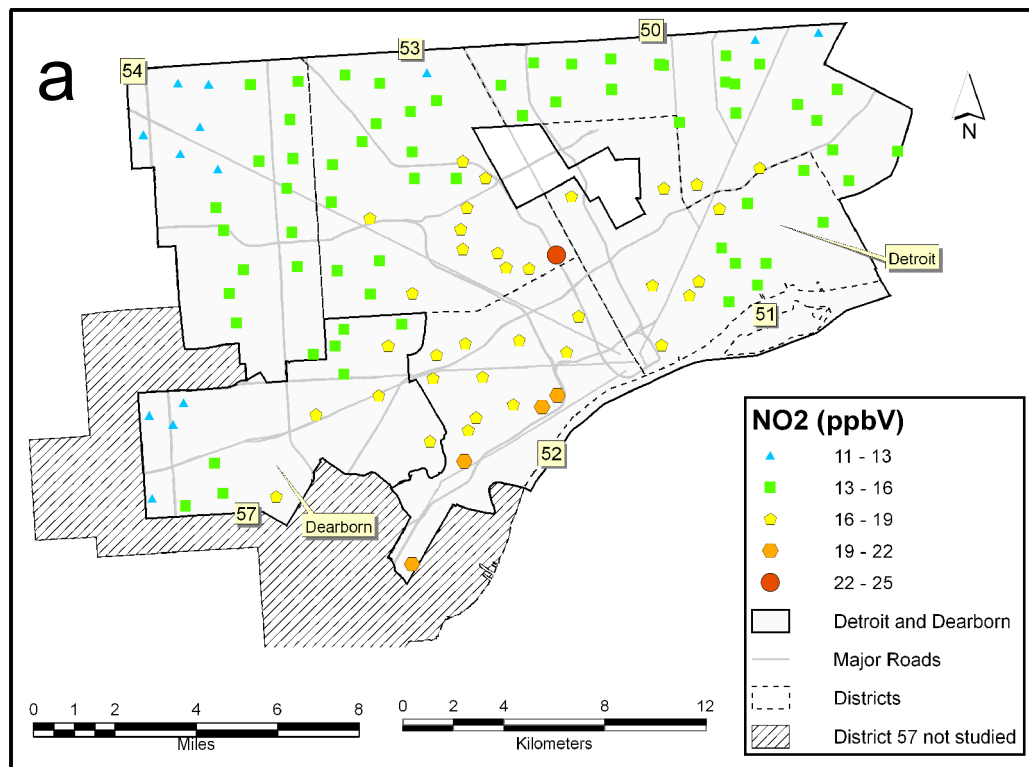
**Table 8 – Comparison of measurements and LUR predictions of NO<sub>2</sub> and selected VOCs (ppbV) at Dearborn and E7 Mile MDEQ sites. (Pollutant concentrations and differences are rounded to the nearest integer or nearest whole per cent. Concentrations in ppbV for VOCs and NO<sub>2</sub>.)**

| Site     | Pollutant          | Measured | LUR Predicted | Difference | % difference |
|----------|--------------------|----------|---------------|------------|--------------|
| E7 Mile  | NO <sub>2</sub>    | 14       | 14            | 0          | 0            |
|          | benzene            | 0.46     | 0.48          | 0.02       | 4            |
|          | toluene            | 1.42     | 1.66          | 0.24       | 16           |
|          | ethylbenzene       | 0.19     | 0.24          | 0.05       | 24           |
|          | <i>m,p</i> -xylene | 0.60     | 0.59          | -0.02      | -3           |
|          | <i>o</i> -xylene   | 0.21     | 0.23          | 0.02       | 11           |
|          | BTEX sum           | 2.88     | 3.30          | 0.42       | 14           |
|          | styrene            | 0.03     | 0.03          | 0.002      | 7            |
|          | 1,3-butadiene      | 0.08     | 0.08          | 0.003      | 3            |
|          |                    |          |               |            |              |
| Dearborn | NO <sub>2</sub>    | 22       | 18            | -4         | -17          |
|          | benzene            | 0.39     | 0.42          | 0.03       | 8            |
|          | toluene            | 1.10     | 1.45          | 0.34       | 31           |
|          | ethylbenzene       | 0.20     | 0.22          | 0.02       | 12           |
|          | <i>m,p</i> -xylene | 0.66     | 0.68          | 0.02       | 3            |
|          | <i>o</i> -xylene   | 0.18     | 0.16          | -0.02      | -12          |
|          | BTEX sum           | 2.53     | 3.04          | 0.51       | 20           |
|          | styrene            | 0.04     | 0.03          | -0.005     | -13          |
|          | 1,3-butadiene      | 0.06     | 0.07          | 0.01       | 13           |

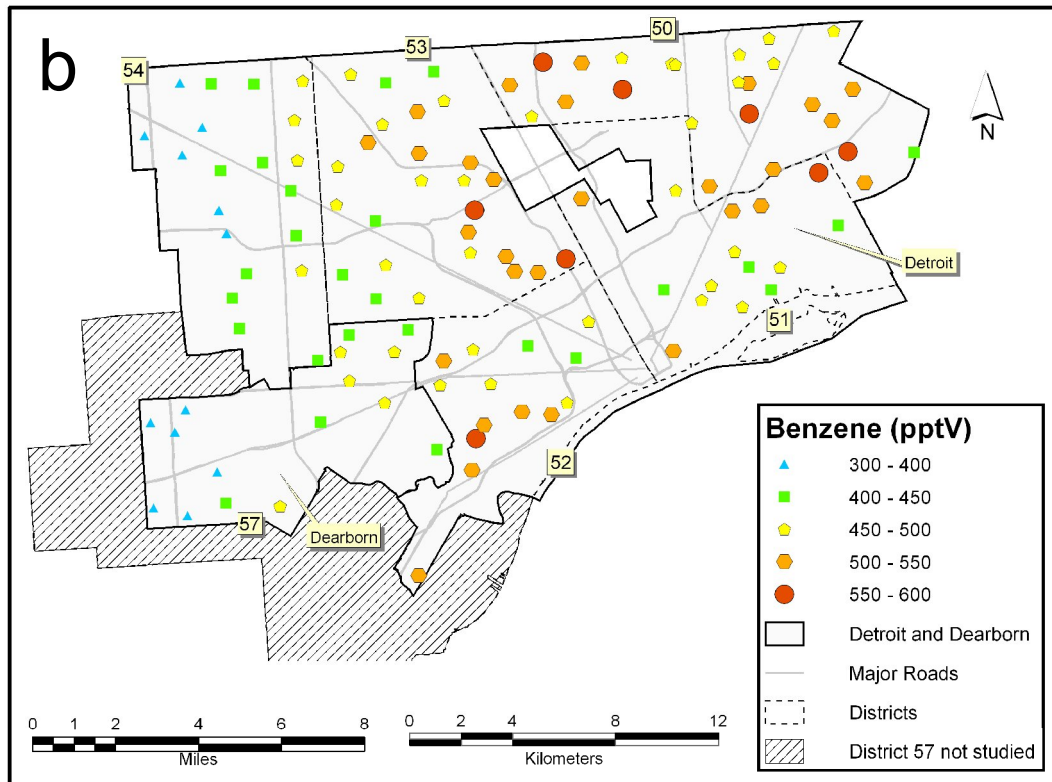


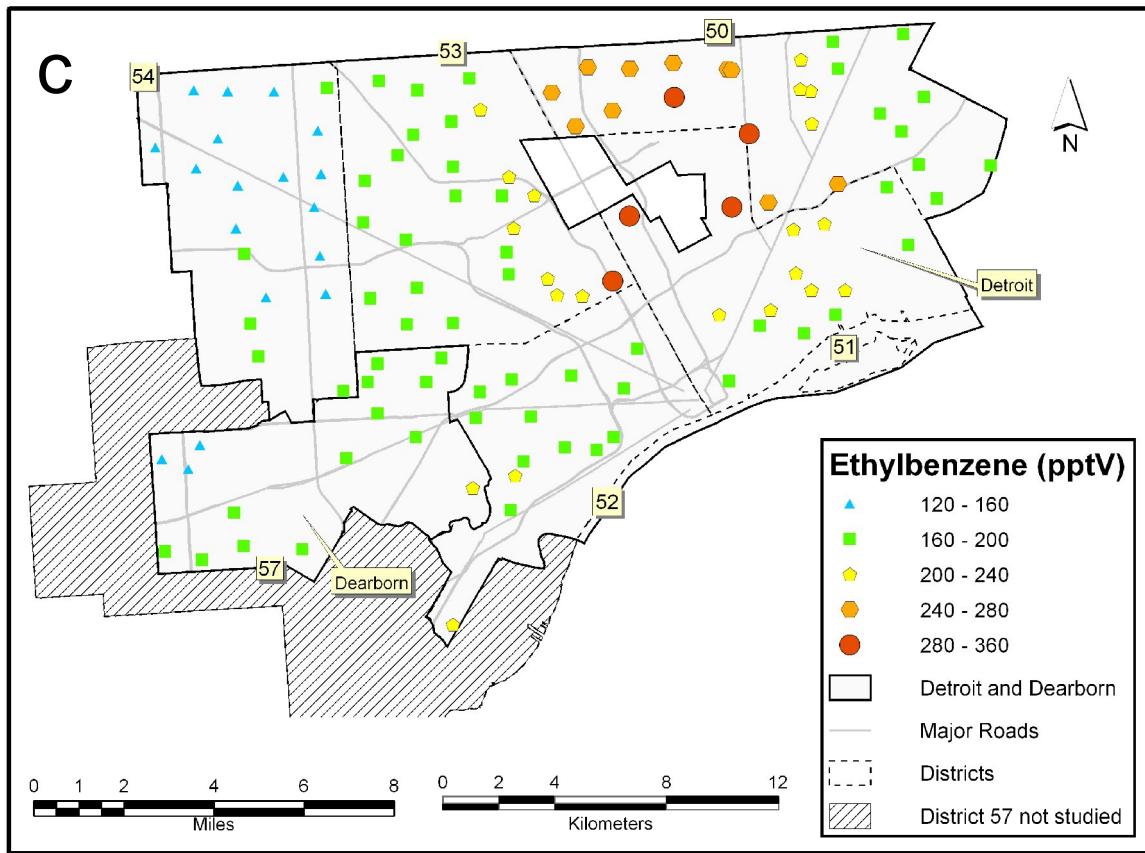
**Fig. 1 - Locations of schools and compliance monitoring sites<sup>a</sup> in Detroit and Dearborn, with enumeration areas.**

<sup>a</sup> All sites were numbered in this study. The two compliance sites, Site 10 (East 7 Mile) and Site 23 (Dearborn), used in this study were established by the State of Michigan for regulatory purposes.



**Fig. 2 - Land use regression predicted ambient air pollution levels at all Detroit and Dearborn local elementary schools for: (a) NO<sub>2</sub>; (b) benzene; (c) ethylbenzene.**





1   **Spatial analysis and land use regression of VOCs and NO<sub>2</sub> from school-based urban air**  
2   **monitoring in Detroit/Dearborn, USA sites**  
3   Shaibal Mukerjee<sup>a,\*</sup>, Luther A. Smith<sup>b</sup>, Mary M. Johnson<sup>c</sup>, Lucas M. Neas<sup>c</sup>, Casson A. Stallings<sup>b</sup>  
4   <sup>a</sup> National Exposure Research Laboratory, U.S. Environmental Protection Agency (E205-02),  
5   Research Triangle Park, NC, 27711, USA  
6   <sup>b</sup> Alion Science and Technology, Inc., Research Triangle Park, NC, 27709, USA  
7   <sup>c</sup> National Health and Environmental Effects Research Laboratory, U.S. Environmental  
8   Protection Agency, Research Triangle Park, NC, 27711, USA  
9   \* Corresponding author. Tel: +1 919 541 1865; fax: +1 919 541 4787  
10   E-mail: [mukerjee.shaibal@epa.gov](mailto:mukerjee.shaibal@epa.gov)

11   **ABSTRACT**  
12   Passive ambient air sampling for nitrogen dioxide (NO<sub>2</sub>) and volatile organic compounds  
13   (VOCs) was conducted at 25 school and two compliance sites in Detroit and Dearborn,  
14   Michigan, USA during the summer of 2005. Geographic Information System (GIS) data were  
15   calculated at each of 116 schools. The 25 selected schools were monitored to assess and model  
16   intra-urban gradients of air pollutants to evaluate impact of traffic and urban emissions on  
17   ~~pollutant levels-respiratory effects in children~~. Schools were chosen to be statistically  
18   representative of urban land use variables such as distance to major roadways, traffic intensity  
19   around the schools, distance to nearest point sources, population density, and distance to nearest  
20   border crossing; ~~these variables were used to develop land-use regression (LUR) models~~. Two  
21   approaches were used to investigate spatial variability. First, Kruskal-Wallis analyses and  
22   pairwise comparisons on data from the schools examined coarse spatial differences based on city  
23   section and distance from heavily trafficked roads. Secondly, spatial variation on a finer scale

and as a response to multiple factors was evaluated through land use regression (LUR) models via multiple linear regression. For weeklong exposures, VOCs did not exhibit spatial variability by city section or distance from major roads; NO<sub>2</sub> was significantly elevated in a section dominated by traffic and industrial influence versus a residential section. Somewhat in contrast to coarse spatial analyses, LUR results revealed spatial gradients in NO<sub>2</sub> and selected VOCs ~~and NO<sub>2</sub>~~ across the area. The process used to select spatially representative sites for air sampling and the results of coarse and fine spatial variability of air pollutants provide insights that may guide future air quality studies in assessing intra-urban gradients.

*Keywords:* Air pollution; GIS; Spatial analysis; Land use regression (LUR); Urban air quality; Traffic.

## **1. Introduction**

Recent studies indicate spatial differences exist for gaseous and particulate air pollutants within urban areas and that these intra-urban gradients can occur near major roadways (Brauer et al., 2003; Zhu et al., 2004; Sahsuvaroglu et al., 2006; Henderson et al., 2007). These studies suggest that compliance monitors, typically sited at no more than a few locations in urban areas, may be limited in assessing spatial variability of air pollution for respiratory health studies in those areas. Another approach has been deployment of passive samplers and other field portable monitors at various locations in urban areas to assess intra-urban variability of air pollutants; pollutants studied have included nitrogen dioxide (NO<sub>2</sub>), particulate matter, and volatile organic compounds (VOCs) (Ware et al., 1993; Spicer et al., 1996; Janssen et al., 2001). Similar sites have been deployed at schools to assess spatial influence and/or respiratory health effects associated with near roadway emissions such as NO<sub>2</sub>, sulfur dioxide, and VOCs (Singer et al., 2004; Luginaah et al., 2006; Wheeler et al., 2008). In addition, children's health concerns

1 | ~~resulting from air pollutants. These and other results~~ have influenced enactment of recent  
2 | legislation in California regarding siting of new schools within 152.4 m (500 feet) of roads with  
3 | traffic > 100,000 vehicles per day in urban areas (California State Senate Legislation, 2003).  
4 | Urban network studies have also used air pollution data in combination with traffic, emissions  
5 | inventories, and demographic data to develop spatial models, such as land use regression (LUR)  
6 | models, that predict pollutant concentrations at unmonitored locations in the study area to  
7 | estimate individual exposures from such ambient sources for spatially based epidemiologic  
8 | assessments (Briggs et al., 1997; Brauer et al., 2003; Sahsuvaroglu et al., 2006; Jerrett et al.,  
9 | 2005; Hoek et al., 2008). This modeling approach has been advocated as being cost effective for  
10 | assessment of long term health effects of ambient air pollution (Künzli and Tager, 2000).  
11 | ~~Based on these and other monitoring approaches to examine within city variability, the~~  
12 | U.S. Environmental Protection Agency (EPA) conducted air monitoring studies in late 1999 at  
13 | elementary schools in ~~the U.S.-Mexico border city of~~ El Paso, Texas and subsequently developed  
14 | LUR models to assess intra-urban variability of air pollutants (Smith et al., 2006). Passive air  
15 | monitors were deployed to measure ambient levels of VOCs and NO<sub>2</sub>, and LUR models were  
16 | developed. Modeled pollutant concentrations were used to assess spatial differences in  
17 | respiratory health effects among children attending El Paso schools. School sites for monitoring  
18 | were selected based on field observations and sampling convenience. Traffic variables,  
19 | elevation, population density, distance from border crossing, and distance to a major oil facility  
20 | were common explanatory variables in the regression analyses for VOCs and NO<sub>2</sub>. The  
21 | observed gradient of pollutant levels indicated that BTEX species (benzene, toluene,  
22 | ethylbenzene, o-xylene, and m,p-xylene) and NO<sub>2</sub> levels were significantly higher in the central  
23 | section of El Paso, located in a valley, compared with eastern and western sections which were at

higher elevation (Smith et al., 2006).

A similar approach was used in the study presented here. The EPA monitored weeklong concentrations of VOCs and NO<sub>2</sub> at elementary schools in Detroit and Dearborn, Michigan, USA. This study assessed spatial gradients of NO<sub>2</sub> and VOCs as functions of traffic and other urban influences such as industrial emissions and population density. The approach is presented in terms of how potential explanatory variables were chosen and school sites selected for monitoring to capture air pollution variability across the Detroit/Dearborn area. Note that the procedure described below will not necessarily lead to a unique set of explanatory variables or monitor locations. Overall spatial analyses on a coarse level are presented by comparing city sections and school distances from major roadways. Finer scale variability and the influence of different variables on pollutant levels are assessed through use of LUR models. Estimates from the LUR models will be applied in a children's respiratory health study; the children's health study results are to be ~~(not presented elsewhere.)~~.

## 2. Methods

### 2.1. Sources and selection of ancillary variables

Spatially representative school sites were selected (Section 2.2) and LUR modelss developed (Section 2.5) based on traffic and other urban land-use variables. These geographic information system (GIS) variables were generated using ArcView 3 and 9 (ESRI, Redlands, CA) with statistical analyses implemented in SAS version 9.1 (SAS, 2004a; 2004b).

Initial identification of ancillary variables was based, in part, on previous spatial assessments done in El Paso (Smith et al., 2006) and other urban areas (Sahsuvaroglu et al., 2006; Ross et al., 2006). The general types of ancillary variables were distance to roadways, traffic intensity, population density, and distance to point sources including border crossings.

1 Data sources for variables were: 1) school data, ~~(including location,)~~ from the National Center  
2 of Education Statistics (NCES); 2) traffic volumes from the Southeast Michigan Council of  
3 Governments Travel Demand Forecast Model for 2000 based on fixed radii from the school and  
4 distance of school from roads classified into traffic volumes of 10,000 vehicles per day  
5 increments up to 90,000 and more; 3) 2000 U.S. Census data; 4)- school distance to nearest point  
6 source of VOCs and PM using emission data from the EPA 1999 National Emission Inventory  
7 database as well as distance to nearest manganese (Mn) point source using emission data from  
8 the Michigan 2002 Toxic Release Inventory database~~distance of school from point source~~  
9 ~~emissions data for VOCs and PM from the EPA 1999 National Emission Inventory database and~~  
10 ~~manganese (Mn) from the Michigan 2002 Toxic Release Inventory database;~~ and 5) school  
11 distance to nearest U.S.-Canada border crossing. Spatial studies of urban areas, ~~(including~~  
12 ~~Detroit,)~~ have used explanatory variables such as distance from roadways and population density  
13 (Brauer et al, 2003; Sahsuvaroglu et al., 2006; Henderson et al., 2007; Ross et al., 2006; Jerrett et  
14 al., 2005; Wu et al., 2006; Hoek et al., 2008); ~~and,~~ these variables were also considered here.  
15 Radii used for population and housing unit density were 125, 250, 500, and 1000 m buffers;  
16 traffic intensity within fixed radii included these buffers and 1500 and 2000 m. Distances to  
17 major VOC and PM point sources were also considered since the Detroit/Dearborn area is  
18 impacted by automobile and other heavy industries (Hammond et al., 2008). (No point source or  
19 traffic data for neighboring Windsor, Ontario were available.) Distance to Mn and PM<sub>2.5</sub> sources  
20 were considered because of their potential use in the subsequent health study. Proximity of  
21 schools to nearest international border crossing (Ambassador Bridge or Detroit-Windsor Tunnel)  
22 was used since these crossings have ing the largest traffic volumes between the U.S. and Canada,  
23 ~~and no source data for neighboring Windsor, Ontario were available.~~ School locations identified

by the NCES database were checked with a global positioning system. Ancillary variables in categories 2 to 5 listed above generated 45 variables presented in ~~(Tables S1 to S4 of the supplementary data).~~

Explanatory variables were selected from the 45 variables by performing separate correlation analyses for the variable groups these being ~~(distance to road, traffic intensity, housing unit and population density, and distance from point sources).~~ The following eight variables were selected as potential explanatory variables: distance to the nearest road with traffic volume of at least 50,000 vehicles per day (DIST\_50KP), distance to the nearest road with traffic volume of at least 90,000 vehicles per day (DIST\_90KP), traffic intensity within 1000 m (INT1000), population density within 500 m (P\_DEN500), distance to the nearest large VOC source emitting 100-1500 lbs/year (VOC\_BIG\_DIST), distance to the nearest large PM<sub>2.5</sub> source emitting > 100 lbs/year (PM25\_BIG\_DIST), distance to the nearest large Mn source emitting 1000-2800 lbs/year (MN\_BIG\_DIST), and distance to the nearest border crossing (BRDR\_DIST). Units for all distances are in m, traffic intensity in vehicles per day/km, and population density is in persons per km<sup>2</sup>. Table 1 presents summary statistics for these variables, and Table 2 indicates the correlation structure among them. Both these tables provide a breakdown of these statistics by monitored and unmonitored schools. No dominant wind direction occurred during the weeklong sampling periods and, therefore, wind direction was not considered in the LUR modeling.

The selected variables exhibited a reasonable amount of variability (coefficient of variation, CV > 30%) across the population of elementary schools. Within these groups, the selected variables to be used in the LUR models were weakly correlated (Pearson correlation coefficient ≤ 0.7) with each other (see Table 2). However, though they often were strongly correlated (Pearson correlation coefficient > 0.7) with some non-selected variables within the same group.

Formatted: Indent: First line: 0.25"

~~Also, the selected variables exhibited a reasonable amount of variability across the population of elementary schools. The following eight variables were selected as potential explanatory variables: distance to the nearest road with traffic volume of at least 50,000 vehicles per day (DIST\_50KP), distance to the nearest road with traffic volume of at least 90,000 vehicles per day (DIST\_90KP), traffic intensity within 1000 m (INT1000), population density within 500 m (P\_DEN500), distance to the nearest large VOC source emitting 100-1500 lbs/year (VOC\_BIG\_DIST), distance to the nearest large PM<sub>2.5</sub> source emitting > 100 lbs/year (PM25\_BIG\_DIST), distance to the nearest large Mn source emitting 1000-2800 lbs/year (MN\_BIG\_DIST), and distance to the nearest border crossing (BRDR\_DIST).~~

## 2.2. Selection of schools

~~Unlike El Paso~~One of the goals of this effort was to develop predictive equations for the pollutants across the entire Detroit-Dearborn study area. To ensure that this could be done reliably, it was required that the schools selected as passive monitoring sites span the mathematical space determined by the predictor variables (described in the previous paragraph). Note that the mathematical space is established by the variables' ranges, variabilities, and their overall correlation structure. Resource limitations dictated that 25 schools could be selected for air monitoring. Therefore, it was necessary to choose monitoring sites jointly with respect to each other and with respect to the values of the ancillary variables.~~schools were selected as passive ambient air monitoring sites so that they spanned the mathematical space determined by the eight explanatory variables; this was done to ensure that reliable predictions could be generated across the entire area. Therefore, it was necessary to select monitoring sites jointly both with respect to each other and the values of the ancillary variables.~~ All local elementary schools from ~~(pre-kindergarten to Grade 6)~~ in the Detroit and Dearborn Public School systems

1 were candidates for passive ambient air monitoring (Fig. 1).

2 Distribution of school sites with respect to enumeration area (EA) was used in site  
3 selection since EA was initially considered for selecting participants in the health study. EAs  
4 were defined as a common census tract grouping based on the first two digits of the 2000 U.S.  
5 Census tract number. Near equal numbers of schools were selected per EA to provide distinct  
6 demographic characteristics for the planned health study while maintaining a reasonable  
7 distribution of ancillary variables. Fig. 1 shows the five EAs, 50 to 54, for Detroit ~~(50 to 54)~~;  
8 Dearborn was part of EA 57.

9 School monitoring locations were chosen as follows. The 116 schools were ranked based  
10 on the values of their ancillary variables, with separate rankings for each of the 8 potential  
11 explanatory variables. Each of the 8 rankings was divided into 13 groups of nine; the groups  
12 were numbered from 1 (nine lowest ranked) to 13 (nine highest ranked). For example, the school  
13 denoted as School 9 was near a DIST\_50KP road and, thus, was in group 1 for DIST\_50KP; this  
14 school also had a high INT1000 vehicle count and was in group 12 for INT1000 (Table ~~43~~).  
15 Within EAs, schools were selected to reflect the range of combinations of group numbers across  
16 all the potential predictors, rather than simply spacing sites across the Detroit/Dearborn  
17 geographic area. The 25 chosen schools are shown in ~~From a pool of 116 schools, 25 were~~  
18 ~~selected~~ (Fig. 1).

19 For the selected schools, Table ~~34~~ indicates ~~(by EA)~~ each school's group numbers for  
20 each of the eight variables. Note the variety of combinations of group numbers. For example  
21 within EA 50, School 9 is in: lower groupings for DIST\_50KP, VOC\_BIG\_DIST, and  
22 Mn\_BIG\_DIST; intermediate groupings for DIST\_90KP, POP\_DEN500, and BRDR\_DIST; and  
23 high groups for INT1000, and PM25\_BIG\_DIST. In contrast, in the same EA, School 13 is low

on INT1000 and POP\_DEN500 and high on the remaining variables.

Four schools were chosen from each EA, except for EA 50 which had five (Fig. 1).

Examination ~~both with respect to~~ the distribution of the individual ancillary variables (Table 1) showed that the selected and unmonitored schools had comparable ranges for each variable. Also, correlations among the ancillary variables for the same paired variables were similar for the 25 schools chosen versus remaining schools (Table 2). In addition, Table 3 shows that the chosen schools encompass a variety of high, medium, and low combinations of ancillary variables. ~~and the results of a~~ Finally, an eight dimensional cluster analysis confirmed that the chosen schools were distributed across the various clusters constructed from the total 116 schools. Thus, the collective evidence from Tables 1 to 3 and the cluster analysis indicated that the procedure described above covered the mathematical space spanned by the ancillary variables. ~~Correlations among the ancillary variables for the same paired variables were similar for the 25 schools chosen versus remaining schools (Table S5).~~ This objective process, coupled with actual site visits to confirm feasibility, ensured that the subsequent spatial analysis of the ambient data collected would be based on a representative sample of school sites for Detroit and Dearborn.

### 2.3. Air monitoring

Passive samplers were deployed outdoors at the 25 ~~spatially representative~~ selected schools and at two compliance sites operated by the Michigan Department of Environmental Quality (MDEQ), as shown in Fig. 1). Monitoring was done at the MDEQ compliance sites to evaluate LUR model predictions and to compare VOCs and NO<sub>2</sub> measurements with corresponding reference method measurements reported in the EPA Air Quality System (AQS) database. In addition, ~~D~~uplicate passive samplers were collocated at the compliance sites to

1 ~~evaluate passive sampler precision (see Mukerjee et al., 2009) to compare VOCs and NO<sub>2</sub>~~  
2 ~~measurements with corresponding reference method measurements reported in the EPA Air~~  
3 ~~Quality System (AQS) database.~~ Compliance sites were AQS Site 261630033 near the River  
4 Rouge Industrial Complex in Dearborn (~~AQS Site 261630033~~) and AQS Site 261630019 in the  
5 East 7 Mile Road area of northeast Detroit (~~AQS Site 261630019~~). Further details on monitoring  
6 conducted at the compliance sites are discussed elsewhere (Mukerjee et al., 2008<sup>9</sup>).

7 Ambient monitoring was conducted concurrently at all sites for six weeks from July 18 -  
8 August 30, 2005. Weeklong integrated sampling ~~integrals were~~ chosen to represent chronic  
9 ambient exposures. Evaluation of meteorological data indicated summers in Detroit were  
10 characterized by a greater proportion of stagnation persistence ~~of~~ (low wind speeds and stable  
11 air masses), conditions perceived as associated with development of spatial gradients in air  
12 pollutants reflective of local-scale emissions. Previous monitoring studies also suggested that  
13 higher concentrations of ambient particulate air pollution occur in Detroit during the summer  
14 (Morishita et al., 2006).

15 ~~For the present study,~~ concurrent monitoring was conducted at the 25 chosen local  
16 neighborhood schools to reflect children's exposures in the immediate community. Samplers  
17 were placed in shelters and suspended 1.5 to 2 m in height at gas-line cages, fenced courtyards,  
18 or chain link cages designed to minimize vandalism. Shelters were custom-designed, 2.8 liter  
19 size stainless steel bowls to protect samplers from wind and precipitation. All sites were checked  
20 prior to sampler deployment and subsequent to sampler retrieval for immediate VOC influences  
21 using a portable, calibrated MiniRAE 2000 VOC monitor (RAE Systems Inc., Sunnyvale, CA,  
22 USA) accurate to 0.1 ppm levels; no contamination was detected at that level. All sites were  
23 photographed from all directions.

Air monitoring was conducted using Ogawa Model 3300 passive samplers for NO<sub>2</sub> (Ogawa & Co., Pompano Beach, FL, USA) and thermal desorption diffusion tubes packed with 40/60 mesh size, unwashed Carbopack X adsorbent for VOCs (Supelco, Inc., Bellefonte, PA, USA). These sampling methods have been validated (Yu et al., 2008; Cox, 2003) and were further evaluated in lab and field studies in Detroit (Mukerjee et al., 2008). Ogawa samplers were analyzed using ion chromatography. VOC analysis was performed using thermal desorption followed by GC/MS. Monitoring and chemical analysis procedures for these samplers are detailed elsewhere (McClenny et al., 2005; 2006; Mukerjee et al., 2004; 2008).

~~Other Detroit exposure studies using these passive samplers under varying time integrals are discussed elsewhere (Johnson et al., 2008; Williams et al., 2008).~~

Twenty-five VOC species were measured including BTEX species and 1,3-butadiene. These compounds were used as markers for transportation, refineries and aromatic emissions (Ware et al., 1993; Fujita, 2001; Buzcu and Fraser, 2006). Process dominated compounds such as styrene (Ware et al., 1993; Fujita, 2001) were also measured. Nitrogen dioxide was measured as an indicator of mobile and stationary combustion sources.

#### 2.4. Coarse spatial distinctions

~~In this study, comparison of the monitored schools' ambient data by their respective EA and major road distance category DIST\_90KP was performed to assess coarse urban gradients of air pollution. In addition, it is envisioned that the planned health analyses may be conducted, in part, by EA. Monitoring data/Monitored concentrations (of NO<sub>2</sub>, 1,3-butadiene, BTEX species, and styrene) were used for this purpose; investigated for coarse-scale differences across the area. Selected pairwise comparisons between EAs were conducted. Also, the DIST\_90KP explanatory variable was divided into three categories, and pairwise comparisons made between~~

1 ~~these categories~~ (see Coarse-Scale Comparisons in the Results section for these analyses). To  
2 avoid data distribution assumptions, nonparametric statistical tests were used in these  
3 comparisons. The Kruskal-Wallis test, using the (SAS procedure NPAR1WAY procedure in  
4 SAS (SAS, 2004b), ~~was~~ was used to determine whether overall differences existed among the EAs.  
5 Subsequent pairwise comparisons of the EAs and road distance categories were made using a  
6 modification of Dunn's multiple comparison procedure to control false positives within the  
7 stated significance level (Dunn, 1964); the comparisons utilized the six week average  
8 concentrations at the individual sites. To improve statistical power, the test was modified  
9 (Hochberg and Tamhane, 1987) using a critical point value from the studentized range  
10 distribution. For the two-sided comparisons for EA and road distance category effect, Dunn's  
11 test indicates statistical significance ~~if~~ according to whether the following inequality holds:

$$12 \quad \left| \bar{R}_i - \bar{R}_{i'} \right| > \left( \frac{Q_{k,\infty}^{(\alpha)}}{\sqrt{2}} \right) \left\{ \frac{N(N+1)}{12} \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right) \right\}^{\frac{1}{2}}$$

13 where k reflects the number of EA or road distance category pairwise comparisons being done  
14 for each chemical,  $\bar{R}_i$  is the mean rank for the  $i^{\text{th}}$  EA or road distance category,  $n_i$  is the number  
15 of observations in the  $i^{\text{th}}$  EA or road distance category, N is the total number of observations over  
16 all EAs or road distance categories, and  $Q_{k,\infty}^{(\alpha)}$  is the  $\alpha$ -level critical point from the studentized  
17 range distribution. The value of  $Q_{k,\infty}^{(\alpha)}$  was taken from the tabulated critical points of the  
18 studentized range tables given by Harter (1960) using  $\alpha=0.05$  and  $k=5$  for the EA comparisons  
19 and  $k=3$  for the road distance category effects; all tests were two-sided.

## 20 2.5. LUR models

21 The monitored values were plotted against the ancillary variables selected as potential  
22 predictors. These plots suggested the use of multiple linear regression, after a logarithmic

1 ~~transformation of the pollutant concentrations. Plotting of the monitored values against the~~  
2 ~~potential predictor variables suggested the use of multiple linear regression, after a logarithmic~~  
3 ~~transformation of the pollutant concentrations.~~ In some cases, the explanatory variables were  
4 also logarithmically transformed ([see Table 4](#)). Example scatterplots for log transformed NO<sub>2</sub>  
5 are displayed in [Figs. S1a-g](#) of the supplementary material (~~Figs. S1a-g~~).

6 Residual analyses after the initial regression attempts indicated that the linear regression  
7 approach was reasonable in terms of the posited model ~~using the (SAS procedure REG procedure~~  
8 ~~in SAS)~~. However, for each pollutant, individual schools had large differences between their  
9 measured and predicted values. These generated large enough variance estimates that very few,  
10 if any, regression coefficients were statistically significant (~~p < 0.05~~) (~~5% level~~). These  
11 departures from the general pattern were further investigated. However, no clear pattern was  
12 evident. For example, from pollutant to pollutant, it was not always the same school that was  
13 high or low. Review of field comments, lab notes, and photographic documentation at each site  
14 revealed no reason for these departures.

15 To de-emphasize the effect of these sites with large departures, the regressions were  
16 repeated with observations weighted by the inverse of the Cook's D influence statistic (Cook,  
17 1977). ~~(A discussion of both transformations and weighting in a regression context can be found~~  
18 ~~in Carroll and Ruppert (1988). See Carroll and Ruppert (1988) for a discussion of both~~  
19 ~~transformations and weighting in a regression context.)~~ This approach better satisfied the basic  
20 regression assumptions and allowed important predictors to be identified via the significance of  
21 their regression coefficients.

22 Beyond residual analysis, regressions were evaluated for collinearity. Also, cross-  
23 validation was conducted on each regression. The two compliance sites were not utilized in

developing the predictive equations, and regression performance was assessed by comparing the values measured at these two sites against the predicted values.

### 3. Results

#### 3.1. Preliminary Evaluation Concentrations

Table 5 shows summary statistics of the air pollutants collected at the schools overall and by EA. Data from passive sampling at the compliance sites was found to be in general agreement with the continuous monitoring data (Mukerjee et al., 2009). This comparison of air monitoring data at the school sites with data collected through passive monitoring at the compliance sites found them to be in general agreement and provided assurance that the passive data were representative of air quality conditions. Comparison of data from the El Paso study referenced earlier (Smith et al., 2006) found that median pollutant concentrations in El Paso were comparable to or higher than Detroit, with the exception of o-xylene. The levels for BTEX species in Detroit were also similar or lower in comparison to historical levels measured in other U.S. cities (Singh et al., 1985; Edgerton et al., 1989; Smith et al., 2007). All data were above method detection limits (Table 5). Although the passive samplers were capable of measuring carbon tetrachloride, trichloroethylene, and chlorobenzene, these pollutants were found to be below detection in nearly every case and were not analyzed further.

Pearson correlations of total BTEX suggested the VOC passive data generally tracked well with 7-day integrated canister samples at the Dearborn compliance site ( $r=0.98$ ,  $p<0.01$ ) but less so in comparison to auto-GC measurements at the East 7 Mile compliance site ( $r=0.62$ ,  $p=0.27$ ), which might be due to low ambient levels. The NO<sub>2</sub> passive samplers were also comparable to reference methods ( $r=0.81$ ,  $p=0.05$ ). The BTEX and NO<sub>2</sub> pollutant data from the samplers used here were above method detection limits. Field method evaluations and

comparisons of the samplers used in this study with detection limits are detailed elsewhere (Mukerjee et al., 2009).

Based on a similar approach (Spicer et al., 1996), ~~coefficients of variation~~ (CV using the MEANS procedure in SAS; SAS procedure MEANS) were calculated for: 1) duplicate passive samplers to represent measurement variability, 2) between school sites to represent spatial variability, and 3) between weeks to represent temporal variability. For NO<sub>2</sub> and BTEX species, duplicate CVs were  $\leq 4\%$ , while spatial and temporal CVs were  $\geq 16\%$  (Table 62). The latter reflects the spatial and temporal variation across school sites and weeks and indicates that the variation of pollutant concentrations is attributable to spatial and temporal differences rather than measurement precision.

### 3.2. Coarse-Scale Comparisons

The following pairwise comparisons were of interest: EA 52 (southwestern Detroit) versus all others, EA 57 (Dearborn) versus all others, and EA 51 (southeastern Detroit) versus EA 54 (western Detroit). EA 52 was compared with the other EAs since this area contained the largest ~~mileage-length~~ of freeways and high-volume city roads and was generally downwind of the River Rouge/Zug Island industrial area; EA 57 represented Dearborn for comparison with the other areas in Detroit. In addition, EA 51 was compared to EA 54 since EA 51 also contained heavy industry while EA 54 is generally a residential area.

Table 35 presents, by EA, summary statistics for the pollutants. An overall difference ( $p < 0.05$ ) among EAs was only found for NO<sub>2</sub>. Pairwise comparisons of EAs using the modified Dunn's test found that NO<sub>2</sub> levels were significantly ~~(5% level)~~ higher in EA 52 relative to EA 54 at the 5% level (Table 53); no other significant differences were found.

Additional pairwise comparisons were performed using school distance from road

segments with traffic volumes  $\geq 90,000$  vehicles per day. This traffic volume was used ~~insinee~~  
~~it figured into~~ one of the potential predictors and is close to the volume cited in recognition of  
the California school siting legislation. The road distance categories used in these comparisons  
were 0 – 1.6 km, 1.6 - 4 km, and > 4 km. The modified Dunn's test revealed no differences at  
the (5% level) between the distance categories for any pollutant.

### 3.3. LUR Modeling

~~The application of LUR allowed an assessment of the variability of the different~~  
~~pollutants on a finer, continuous scale.~~ Table 44 presents the ~~predictive equations~~ final LUR  
models. The explanatory power of these LUR models ranged from quite good for NO<sub>2</sub>  
(R<sup>2</sup>=82%) to poor for toluene (R<sup>2</sup>=31%). ~~(R<sup>2</sup> in Table 4-4 reflects the original scale, not the log-~~  
~~transformed scale (Kvålseth, 1985; Scott and Wild, 1991). This is done for ease in interpretation~~  
~~of LUR predictive power in the field environment.)~~

Unfortunately, the weighting procedure introduced collinearity. Each of the regressions  
except for toluene and *o*-xylene suffered from this. Collinearity hinders assessment of the  
relative importance of individual predictors and possibly the sign of the regression coefficients  
because since-variables may be mathematically related to some extent. Because of this, the LUR  
regression coefficients reported in Table 44 must be interpreted cautiously ~~(see supplemental~~  
~~discussion and Table S6 on collinearity).~~

Using the criteria described by Kleinbaum et al. (1998), Table 7 indicates the sets of  
variables within each regression that are subject to having the variables “play off” against each  
other in terms of their regression coefficients. For example, Table 7 reports that for the NO<sub>2</sub>  
regression a set of five predictors (including the intercept) are involved in collinearity. Each of  
the five was reported as being a statistically significant (p < 0.05) predictor. What this implies is

that collectively this group of variables is important for predicting NO<sub>2</sub>, but because of the collinearity issue, the relative importance of the individual predictors is uncertain.

Notwithstanding the influence of collinearity, it is interesting that MN\_BIG\_DIST and PM25\_BIG\_DIST seem to be important predictors for several pollutants. To reiterate, Mn and PM<sub>2.5</sub> sources were originally considered for their utility in the planned children's health study. Their appearance in the final equations was investigated by deleting them and recomputing the regressions. However, residual analyses and both the Akaike (1969) and Bayesian information criteria indicators (Sawa, 1978) suggested that retaining these predictors was advisable.

Cross-validation (~~see~~ Cressie, 1993), whereby monitored schools were left out one at a time and the model re-estimated, was employed to evaluate the regressions; the resulting predicted value was then checked against the monitored value not used in the re-estimation. Satisfactory performance of the models was obtained for benzene, toluene, *o*-xylene, and the BTEX sum. Good performance was found for 1,3-butadiene, styrene, and NO<sub>2</sub>, though these latter three compounds had higher root mean squares ~~of~~ (approximately 4 to 6) than the desired value of 1. For ethylbenzene and *m,p*-xylene, cross-validation results were not as good, with mean errors on the order of 3 ~~-(as opposed to 0.)~~ and root mean squares near 16 and 17. For each compound where prediction variability ~~of~~ (root mean square error) was elevated, this was attributable to large discrepancies between measurements and LUR predictions ~~predictions being missed~~ at one, two, or three sites. NO<sub>2</sub> appeared to be less reliably predicted at sites which were simultaneously close to roads carrying traffic volumes greater than both 50,000 vehicles and 90,000 vehicles per day and were also in an area of heavy traffic intensity. Ethylbenzene and *m,p*-xylene were poorly predicted in locations with combinations of being near roads carrying 50,000 or more vehicles per day, in areas of high traffic intensity, and far from large PM<sub>2.5</sub>

sources.

Performance of the regressions was also assessed by comparing the predictions for the LUR effort to measured values from compliance sites (Table 85). Generally, the predictions and measurements agreed well. Of the 18 comparisons in Table 85, 15 show a difference of less than 20%, and 7 predictions are within 10% of the measured value. The worst performance was a discrepancy of 31% for toluene at the Dearborn site. NO<sub>2</sub> was predicted ~~perfectly (within rounding well)~~ at the East 7 Mile site.

Figs. 2a-c display the pollutant levels predicted by the LUR models for NO<sub>2</sub>, benzene, and ethylbenzene, respectively for all Detroit/Dearborn local elementary schools. ~~(The blank area in the figures represent the cities of Highland Park and Hamtramck, referenced here as Highland Park; see Fig. 1.)~~ Fig. 2a shows generally higher predicted NO<sub>2</sub> levels in the south-central section of the Detroit/Dearborn area below Highland Park. For benzene, Fig. 2b indicates lower predicted values in the western portion of the area, higher levels around and to the north and east of Highland Park, and a pocket of higher values in EA 52. Styrene and 1,3-butadiene displayed similar patterns. Fig. 2c indicates that ethylbenzene was similar to benzene in that predictions were lower to the west and relatively elevated around Highland Park; however, ethylbenzene did not display the group of higher values in the south-central area. The other BTEX species were similar to ethylbenzene.

#### 4. Discussion and conclusions

School-based ambient air monitoring ~~sites-were~~was successfully ~~deployed-conducted~~ in Detroit and Dearborn where sites were statistically chosen based on ~~land-use-(GIS)~~ data to be spatially representative of air pollution explanatory variables throughout the airshed. Analysis on two different coarse levels, geographic sections of the area and categories reflecting distances

to heavily trafficked roads, found little variability of VOCs. ~~This suggested a pervasive mobile source effect.~~ Spatial uniformity of VOCs also occurred in Columbus, OH based on samples collected for  $\leq 3$  hour integrals (Spicer et al., 1996).

NO<sub>2</sub> exhibited coarse spatial difference between the traffic and industrial-dominated area of southwestern Detroit (EA 52) versus the more residential western section (EA 54).

Qualitatively, these results may indicate a mobile source influence on NO<sub>2</sub> due to the greater length of major road segments in southwestern Detroit and that this industrial section is more impacted than the western section. For distance from very high volume road segments, coarse spatial analysis found no significant differences for NO<sub>2</sub>. The lack of coarse spatial differences, particularly for VOCs, may have been due to the fact that winds were generally coming from all directions during each week of the study and a pervasive mobile source effect.

LUR modeling revealed spatial gradients for all pollutants. Generally, pollutants were lower in the western part and, except NO<sub>2</sub>, higher near and to the north and east of Highland Park. NO<sub>2</sub>, benzene, styrene and 1,3-butadiene also had elevated levels in the south-central portion of the Detroit-Dearborn area. ~~Coarse spatial analyses did find significantly higher NO<sub>2</sub> levels in the south-central section versus the western portion. This difference is reflected in the LUR predictions.~~

Although the Detroit/Dearborn area has facilities that emit large amounts of large VOC and PM<sub>2.5</sub> ~~emitters~~, such as the River Rouge area (Hammond et al., 2008), the coarse analyses found no significant differences in VOC levels between EAs. A recent study of VOCs along commuting routes near and away from industrial areas in Detroit suggested mobile source emissions dominate other sources (Batterman et al., 2002). However, LURs obtained in this study indicated that mobile and point sources may play roles in determining ambient VOC

1 concentrations.

2 Results between coarse analyses and LUR modeling with respect to spatial variability may  
3 have arisen for more than one reason. For coarse analyses, the basic question was whether  
4 certain specific, well-defined ~~groups-city sections~~ differed between themselves for the observed  
5 pollutant levels. Furthermore, ~~though modified to improve its power, the Dunn's test used for~~  
6 ~~the coarse-scale analysis sacrifices power~~ ~~testing was conducted~~ to guard against false  
7 positives, ~~though the modification to Dunn's test did serve to boost statistical power~~. LUR  
8 modeling had as its goal the prediction of pollutant levels across the entire area ~~by without being~~  
9 ~~restricted by any discretization and~~ considering ~~ing~~ multiple variables simultaneously, ~~without~~  
10 ~~being restricted by any discretization~~, whereas coarse analyses were restricted to pollutant levels  
11 relative to a single categorized variable. Thus, LUR is more amenable to detecting spatial  
12 gradients across the entire area, while the coarse analyses are applicable in assessing  
13 ~~neighborhood-to-neighborhood~~ ~~section-to-section~~ differences.

14 Utility of LUR regressions is very much dependent on the “pre-analysis” described above,  
15 whereby both monitoring locations and potential predictors are determined beforehand. ~~Recall~~  
16 ~~that this procedure does not necessarily lead to a unique set of explanatory variables or monitor~~  
17 ~~locations~~. Incorporating statistical considerations of the pre-analysis into the LUR approach  
18 provides greater confidence that the appropriate mathematical space has been identified and  
19 covered by the monitoring locations. Therefore, there is higher confidence in the use of the  
20 predictive equations to determine pollutant levels across the entire study area. Similar, though  
21 not identical, pre-analyses ~~like population-weighted location-allocation models have been~~ ~~were~~  
22 emphasized in other LUR studies (Sahsuaroglu et al., 2006; Henderson et al., 2007).

23 Since an EA occupies a large fraction of the area, each can contain a range of pollutant

1 levels. Pollutant gradients seen in Figs. 2a-c suggest that pollutant levels do not change abruptly  
2 across the area. This combination may explain why all EA levels were statistically the same,  
3 except one pair~~coarse-level testing gave only one significant test.~~

4 Since MN\_BIG\_DIST and PM25\_BIG\_DIST were considered for utility in the health study,  
5 their appearance in some predictive equations is puzzling. While their relative importance as  
6 predictors is somewhat uncertain due to collinearity, their use in these predictions appears  
7 justified based on analyses of models excluding these variables. MN\_BIG\_DIST and  
8 PM25\_BIG\_DIST may be acting as surrogates for variables not considered in the models. For  
9 example, large manganese and PM<sub>2.5</sub> sources considered here might also be smaller emitters of  
10 NO<sub>2</sub> or VOCs. Only large VOC sources could enter the predictive equations, and  
11 MN\_BIG\_DIST and PM25\_BIG\_DIST may be substituting for distances to lower level VOC or  
12 NO<sub>2</sub> emitters.~~No dominant wind direction occurred during the weeklong sampling periods and,~~  
13 ~~therefore, wind direction was not considered in the LUR modeling.~~

14 Previous research has found spatial gradients for these pollutants under similar sampling  
15 intervals at El Paso schools (Smith et al., 2006). There, the center section exhibited significantly  
16 higher VOC and NO<sub>2</sub> levels. El Paso's proximity to Ciudad Juárez, Mexico and the area's  
17 complex terrain were likely reasons that coarse-level spatial gradients were found, versus Detroit  
18 with flat terrain and smaller cross-border influences.

19 The results presented here demonstrated the utility of coarse scale analyses for examining  
20 differences between specific subsets of a geographic area and LUR modeling for finer scale  
21 descriptions of an entire metropolitan area. This study has ~~also~~ emphasized the importance of  
22 ~~the~~ statistical pre-analysis in selecting explanatory variables and site selectionlocations and of the  
23 ~~importance of~~ diagnostic analyses in ~~the~~ determining ~~of~~ final LUR models.

1   **Acknowledgements and disclaimer**

2   We thank Hunter Daughtrey, Karen Oliver, Herb Jacumin, Laura Liao, Chris Fortune, Mike  
3   Wheeler, and Dennis Williams, all from Alion Science and Technology, Felicia Venable and  
4   Mathew Sam of the Detroit Public Schools, Don Ball of Dearborn Public Schools, and principals  
5   and staff. We also thank Mary Ann Heindorf, Ann Chevalier, Deborah Sherrod, Amy Robinson,  
6   and Craig Fitzner for access to MDEQ sites and data and Ron Williams and Alan Vette of EPA  
7   for manuscript review. Finally, we thank Ann Williams, Gina Andrews, Paul Killough of EPA,  
8   and John Archer of Eastern Research Group for assistance. The U.S. Environmental Protection  
9   Agency through its Office of Research and Development funded and managed the research  
10   described here under contract EP-D-05-065 to Alion. The paper has been subjected to Agency  
11   review and approved for publication.

12   **Supplementary data**

13   Tables of ancillary variables considered for use in LUR models, ~~Pearson correlations between~~  
14   ~~explanatory variables considered for LURs, LUR collinearity discussion and table, and~~  
15   scatterplots of log transformed NO<sub>2</sub> by predictor variables. This material is available via the  
16   Internet at <http://www.sciencedirect.com>.

17   **REFERENCES**

18   Akaike, H., 1969. Fitting autoregressive models for prediction. Ann. I. Stat. Math. 21, 243-247.  
19   Batterman, S.A., Peng, C.Y., Braun, J., 2002. Levels and composition of volatile organic  
20   compounds on commuting routes in Detroit, Michigan. Atmos. Environ. 36, 6015–6030.  
21   Brauer, M., Hoek, G., van Vliet, P., Meliefste, K., Fischer, P., Gehring, U., Heinrich, J., Cyrus,  
22   J., Bellander, T., Lewne, M., Brunekreef, B., 2003. Estimating long-term average  
23   particulate air pollution concentrations: application of traffic indicators and geographic

information systems. Epidemiology 14 228-239.

Briggs, D.J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebre, E., Pryl, K., Van Reeuwijk, H., Smallbone, K., Van Der Veen, A., 1997. Mapping air pollution using GIS: a regression approach. Int. J. Geogr. Inf. Sci. 11 699-718.

Buzcu, B., Fraser, M.P., 2006. Source identification and apportionment of volatile organic compounds in Houston, TX. Atmos. Environ. 40, 2385-2400.

California State Senate Legislation, October 3, 2003. Senate Bill 352, Chapter 668, An act to amend Section 17213 of the Education Code, and to amend Section 21151.8 of the Public Resources Code, relating to public schools. (Internet version  
[http://info.sen.ca.gov/pub/03-04/bill/sen/sb\\_0351-0400/sb\\_352\\_bill\\_20031003\\_chaptered.pdf](http://info.sen.ca.gov/pub/03-04/bill/sen/sb_0351-0400/sb_352_bill_20031003_chaptered.pdf))

Carroll, R.J., Ruppert, D., 1988. Transformation and weighting in regression. Chapman and Hall, New York, pp. 115-208.

Cook, R.D., 1977. Detection of influential observations in linear regression. Technometrics 19, 15-18.

Cox, R.M., 2003. The use of passive sampling to monitor forest exposure to O<sub>3</sub>, NO<sub>2</sub> and SO<sub>2</sub>: a review and some case studies. Environ. Pollut. 126, 301-311.

Cressie, N.A.C., 1993. Statistics for spatial data, revised edition. Wiley, New York, pp. 101-104.

Dunn, O.J., 1964. Multiple comparisons using rank sums. Technometrics 6, 241-252.

Edgerton, S.A., Holdren, M.W., Smith, D.L., Shah, J.J., 1989. Inter-urban comparison of ambient volatile organic compound concentrations in U.S. cities. J. Air Waste Manage. Assoc. 39, 729-732.

Fujita, E.M., 2001. Hydrocarbon source apportionment for the 1996 Paso del Norte Ozone

Study. Sci. Total Environ. 276, 171-184.

Hammond, D., Dvonch, J.T., Keeler, G.J., Parker, E.A., Kamal, A.S., Barres, J.A., Yip, F.Y.,  
Brakefield-Caldwell, W., 2008. Sources of ambient fine particulate matter at two  
community sites in Detroit, Michigan. Atmos. Environ. 42, 720-732.

Harter, H.L., 1960. Tables of range and studentized range. Ann. Math. Stat. 31, 1122-1147.

Henderson, S.B., Beckerman, B., Jerrett, M., Brauer, M., 2007. Application of land use  
regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine  
particulate matter. Environ. Sci. Technol. 41, 2422-2428.

Hochberg, Y., Tamhane, A.C., 1987. Multiple comparison procedures. Wiley, New York, pp.  
244-246.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A  
review of land-use regression models to assess spatial variation of outdoor air pollution.  
Atmos. Environ. 42, 7561-7578.

Janssen, N.A.H., van Vliet, P.H.N., Aarts, F., Harssema, H., Brunekreef, B., 2001. Assessment  
of exposure to traffic related air pollution of children attending schools near motorways.  
Atmos. Environ. 35, 3875-3884.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison  
J., Giovis, C., 2005b. A review and evaluation of intraurban air pollution exposure  
models. J. Exp. Anal. Environ. Epidemiol. 15, 185-204.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., Nizam, A., 1988. Applied regression analysis and  
other multivariable methods. Duxbury Press, Pacific Grove, CA, pp. 237-248.~~Johnson,  
M., Hudgens, E., Williams, R., Andrews, G., Neas, L., Gallagher, J., Özkaynak, H., 2008.~~

~~A participant-based approach to indoor/outdoor air monitoring in community health studies. J. Exp. Sci. Environ. Epidemiol. in press (doi:10.1038/jes.2008.39).~~

Künzli, N., Tager, I.B., 2000. Long-term health effects of particulate and other ambient air pollution: Research can progress faster if we want it to. *Environ. Health Persp.* 108, 915-918.

Kvålseth, T.A., 1985. Cautionary rule about  $R^2$ . *Am. Statist.* 39, 279-285.

Luginaah, I., Xu, X., Fung, K.Y., Grgicak-Mannion, A., Wintermute, J., Wheeler, A., Brook, J., 2006. Establishing the spatial variability of ambient nitrogen dioxide in Windsor, Ontario. *Int. J. Environ. Stud.* 63, 487-500.

McClenny, W.A., Oliver, K.D., Jacumin, H.H., Daughtrey, E.H., Whitaker, D.A., 2005. 24 h diffusive sampling of toxic VOCs in air onto Carbopack X solid adsorbent followed by thermal desorption/GC/MS analysis—laboratory studies. *J. Environ. Monit.* 7, 248-256.

McClenny, W.A., Jacumin, H.H., Oliver, K.D., Daughtrey, E.H., Whitaker, D.A., 2006. Comparison of 24 h averaged VOC monitoring results for residential indoor and outdoor air using Carbopack X-filled diffusive samplers and active sampling—a pilot study. *J. Environ. Monit.* 8, 263-269.

Morishita, M., Keeler, G.J., Wagner, J.G., Harkema, J.R., 2006. Source identification of ambient PM<sub>2.5</sub> during summer inhalation exposure studies in Detroit, MI. *Atmos. Environ.* 40, 3823-3834.

Mukerjee, S., Oliver, K.D., Seila, R.L., Jacumin, H.H., Croghan, C., Daughtrey, E.H., Neas, L.M., Smith, L.A., 2008. Field comparison of passive air samplers with reference monitors for ambient volatile organic compounds and nitrogen dioxide under week-long integrals. *J. Environ. Monit.* ~~in press (DOI: 10.1039/b809588d)~~ 11, 220-227.

1 Mukerjee S., Smith L.A., Norris G.A., Morandi M.T., Gonzales M., Noble C.A., Neas L.M., and  
2 Özkaynak A.H. 2004. Field method comparison between passive air samplers and  
3 continuous monitors for volatile organic compounds and NO<sub>2</sub> in El Paso, Texas, USA. J.  
4 Air Waste Manage. Assoc. 54, 307-319.

5 Ross, Z, English, P.B., Scalf, R., Gunier, R., Smorodinsky, S., Wall, S., Jerrett, M., 2006.  
6 Nitrogen dioxide prediction in Southern California using land use regression modeling:  
7 potential for environmental health analyses. J. Exp. Sci. Environ. Epidemiol. 16, 106-  
8 114.

9 Sahsuvaroglu, T., Arain, A., Kanaroglou, P., Finkelstein, N., Newbold, B., Jerrett, M.,  
10 Beckerman, B., Brook, J., Finkelstein, M., Gilbert, N.L., 2006. A land use regression  
11 model for predicting ambient concentrations of nitrogen dioxide in Hamilton, Ontario,  
12 Canada. J. Air Waste Manage. Assoc. 56, 1059-1069.

13 SAS, 2004a. Base SAS® 9.1.3 procedures guide. SAS Institute Inc., Cary, NC, pp. 524-550.  
14 SAS, 2004b. SAS/STAT® 9.1 user's guide. SAS Institute Inc., Cary, NC, pp. 1291-1320, 3143-  
15 3162.

16 Sawa, T., 1978. Information criteria for discriminating among alternative regression models.  
17 Econometrica 46, 1273-1282.

18 [Scott, A., Wild, C., 1991. Transformations and R<sup>2</sup>. Am. Statist. 45, 127-129.](#)

19 Singer, B.C., Hodgson, A.T., Hotchi, T., Kim, J.J., 2004. Passive measurement of nitrogen  
20 oxides to assess traffic-related pollutant exposure for the East Bay Children's Respiratory  
21 Health Study. Atmos. Environ. 38, 393-403.

22 Singh, H.B., Salas, L.J., Cantrell, B.K., Redmond, R.M., 1985. Distribution of aromatic  
23 hydrocarbons in the ambient air. Atmos. Environ. 19, 1911-1919.

- 1 Smith, L., Mukerjee, S., Gonzales, M., Stallings, C., Neas, L., Norris, G., Özkaynak, H., 2006.
- 2 Use of GIS and ancillary variables to predict volatile organic compound and nitrogen
- 3 dioxide levels at unmonitored locations. *Atmos. Environ.* 40, 3773–3787.
- 4 [Smith, L.A., Stock, T.H., Chung, K.C., Mukerjee, S., Liao, X.L., Stallings, C., Afshar, M., 2007.](#)
- 5 [Spatial analysis of volatile organic compounds from a community-based air toxics](#)
- 6 [monitoring network in Deer Park, Texas, USA. \*Environ. Monit. Assess.\* 128, 369–379.](#)
- 7 Spicer, C.W., Buxton, B.E., Holdren, M.W., Smith, D.L., Kelly, T.J., Rust, S.W., Pate, A.D.,
- 8 Sverdrup, G.M., Chuang, J.C., 1996 Variability of hazardous air pollutants in an urban
- 9 area. *Atmos. Environ.* 1996, 30 3443-3456.
- 10 Ware, J.H., Spengler, J.D., Neas, L.M., Samet, J.M., Wagner, G.R., Coultas, D., Özkaynak, H.,
- 11 Schwab, M., 1993. Respiratory and irritant health effects of ambient volatile organic
- 12 compounds: the Kanawha County Health Study. *Am. J. Epidemiol.* 137, 1287-1301.
- 13 [Wheeler, A.J., Smith-Doiron, M., Xu, X., Gilbert, N.L., Brook, J.R., 2008. Intra-urban variability](#)
- 14 [of air pollution in Windsor, Ontario—Measurement and modeling for human exposure](#)
- 15 [assessment. \*Environ. Res.\* 106, 7-16.](#)
- 16 ~~Williams, R., Rea, A., Vette, A., Croghan C., Whitaker, D., Wilson, H., Stevens, C., McDow, S.,~~
- 17 ~~Fortmann, R., Sheldon, L., Wilson, H., Thornburg, J., Phillips, M., Lawless, P., Rhodes,~~
- 18 ~~C., Daughtrey, H. 2008. The design and field implementation of the Detroit Exposure~~
- 19 ~~and Aerosol Research Study. *J. Exp. Sci. Environ. Epidemiol.* in press~~
- 20 ~~(doi:101038/jes.2008.61).~~
- 21 Wu, Y.C., Batterman, S.A., 2006. Proximity of schools in Detroit, Michigan to automobile and
- 22 truck traffic. *Journal of Exposure Science and Environmental Epidemiology* 16, 457-470.
- 23 Yu, C.H., Morandi, M.T., Weisel, C.P., 2008. Passive dosimeters for nitrogen dioxide in

- 1           personal/indoor air sampling: A review. J. Expo. Sci. Environ. Epid. 18, 441-451.
- 2    Zhu, Y., Hinds, W.C., Shen, S., Sioutas, C. 2004. Seasonal trends of concentration and size
- 3           distribution of ultrafine particles near major highways in Los Angeles. Aerosol Sci.
- 4    Technol. 2004, 38 5-13.
- 5    |

Field Code Changed

Table 1

Summary statistics of ancillary variables for monitored and unmonitored schools

| Ancillary variable                | Monitored schools (n=25) |                |                | Unmonitored schools (n=91) |               |                |
|-----------------------------------|--------------------------|----------------|----------------|----------------------------|---------------|----------------|
|                                   | Minimum                  | Median         | Maximum        | Minimum                    | Median        | Maximum        |
| <u>DIST_50KP<sup>a</sup></u>      | <u>0</u>                 | <u>806</u>     | <u>4,518</u>   | <u>100</u>                 | <u>1,389</u>  | <u>5,400</u>   |
| <u>DIST_90KP<sup>b</sup></u>      | <u>500</u>               | <u>4,000</u>   | <u>8,905</u>   | <u>500</u>                 | <u>4,046</u>  | <u>10,655</u>  |
| <u>INT1000<sup>c</sup></u>        | <u>24,711</u>            | <u>118,398</u> | <u>249,513</u> | <u>17,872</u>              | <u>64,728</u> | <u>493,302</u> |
| <u>POP_DEN500<sup>d</sup></u>     | <u>1,319</u>             | <u>3,039</u>   | <u>4,659</u>   | <u>950</u>                 | <u>3,216</u>  | <u>6,339</u>   |
| <u>VOC_BIG_DIST<sup>e</sup></u>   | <u>1,476</u>             | <u>3,940</u>   | <u>11,229</u>  | <u>1,237</u>               | <u>4,478</u>  | <u>13,124</u>  |
| <u>PM2.5_BIG_DIST<sup>f</sup></u> | <u>2,600</u>             | <u>10,781</u>  | <u>21,570</u>  | <u>2,154</u>               | <u>12,368</u> | <u>23,278</u>  |
| <u>Mn_BIG_DIST<sup>g</sup></u>    | <u>2,941</u>             | <u>7,901</u>   | <u>17,016</u>  | <u>1,910</u>               | <u>8,670</u>  | <u>19,413</u>  |
| <u>BRDR_DIST<sup>h</sup></u>      | <u>1,208</u>             | <u>9,306</u>   | <u>19,482</u>  | <u>943</u>                 | <u>11,416</u> | <u>20,461</u>  |

<sup>a</sup> DIST\_50KP: distance to nearest road with traffic volume  $\geq 50,000$  vehicles per day (m)<sup>b</sup> DIST\_90KP: distance to nearest road with traffic volume  $\geq 90,000$  vehicles per day (m)<sup>c</sup> INT1000: traffic intensity within 1000 m of location (vehicles per day / km)<sup>d</sup> POP\_DEN500: population within a 500 m radius of location<sup>e</sup> VOC\_BIG\_DIST: distance (m) to nearest large (45-680 kg/year) VOC emission source<sup>f</sup> PM2.5\_BIG\_DIST: distance (m) to nearest large ( $>45$  kg/year) PM2.5 emission source<sup>g</sup> Mn\_BIG\_DIST: distance (m) to nearest large (454-1270 kg/year) Manganese emission source<sup>h</sup> BRDR\_DIST: distance to nearest border crossing (m)

**TABLE 2 - Pearson correlations between explanatory variables considered for LURs<sup>a</sup>**

|                       | <u>DIST_50KP</u> | <u>DIST_90KP</u> | <u>INT_1000</u> | <u>POP_DEN_500</u> | <u>VOC_BIG_DIST</u> | <u>PM2.5_BIG_DIST</u> | <u>Mn_BIG_DIST</u> | <u>BRDR_DIST</u> |
|-----------------------|------------------|------------------|-----------------|--------------------|---------------------|-----------------------|--------------------|------------------|
| <u>DIST_50KP</u>      | 1                | 0.38             | -0.66           | -0.02              | 0.07                | 0.04                  | 0.53               | 0.45             |
| <u>DIST_90KP</u>      | 0.26             | 1                | -0.28           | 0.22               | 0.06                | 0.31                  | 0.18               | 0.21             |
| <u>INT1000</u>        | -0.56            | -0.24            | 1               | -0.28              | -0.14               | 0.04                  | -0.30              | -0.19            |
| <u>POP_DEN500</u>     | -0.10            | 0.12             | -0.17           | 1                  | 0.13                | 0.08                  | -0.03              | -0.15            |
| <u>VOC_BIG_DIST</u>   | -0.15            | -0.08            | 0.26            | 0.02               | 1                   | 0.27                  | 0.66               | 0.59             |
| <u>PM2.5_BIG_DIST</u> | 0.31             | -0.02            | -0.10           | 0.07               | 0.13                | 1                     | -0.09              | 0.44             |
| <u>Mn_BIG_DIST</u>    | 0.07             | 0.09             | 0.16            | -0.05              | 0.70                | -0.27                 | 1                  | 0.68             |
| <u>BRDR_DIST</u>      | 0.29             | 0.25             | 0.04            | -0.02              | 0.56                | 0.40                  | 0.59               | 1                |

<sup>a</sup> Schools with passive sampling data (n=25) appear in the upper triangular portion of the matrix (i.e., above the diagonal of 1s); correlations within the group of unmonitored schools (n=91) appear in the lower triangular portion

1 | **Table 31 - Group rankings for explanatory variables atby monitored schools (see Selection of Schools Section)**

| School number            | Group DIST_50KP                   | Group DIST_90KP | Group INT1000            | Group POP_DEN500                    | Group VOC_BIG_DIST      | Group PM2.5_BIG_DIST            | Group Mn_BIG_DIST            | Group BRDR_DIST              |
|--------------------------|-----------------------------------|-----------------|--------------------------|-------------------------------------|-------------------------|---------------------------------|------------------------------|------------------------------|
| Enumeration area (EA) 50 |                                   |                 |                          |                                     |                         |                                 |                              |                              |
| 3                        | 3                                 | 1               | 11                       | 1 (lowest population density group) | 6                       | 9                               | 2                            | 7                            |
| 9                        | 1 (near road of specified volume) | 6               | 12 (high traffic volume) | 6                                   | 4                       | 11                              | 3                            | 5                            |
| 11                       | 6                                 | 12              | 8                        | 13                                  | 8                       | 13                              | 6                            | 8                            |
| 12                       | 1                                 | 12              | 12                       | 12                                  | 6                       | 12                              | 6                            | 7                            |
| 13                       | 10                                | 13              | 1                        | 3                                   | 9                       | 13                              | 9                            | 9                            |
| EA 51                    |                                   |                 |                          |                                     |                         |                                 |                              |                              |
| 4                        | 2                                 | 1               | 11                       | 2                                   | 3                       | 8                               | 1 (close to large Mn source) | 4                            |
| 6                        | 5                                 | 7               | 9                        | 12                                  | 8                       | 5                               | 6                            | 1 (close to border crossing) |
| 7                        | 11                                | 11              | 8                        | 3                                   | 1 (close to VOC source) | 11                              | 7                            | 5                            |
| 8                        | 4                                 | 4               | 9                        | 5                                   | 2                       | 11                              | 2                            | 4                            |
| EA 52                    |                                   |                 |                          |                                     |                         |                                 |                              |                              |
| 24                       | 2                                 | 13              | 9                        | 4                                   | 1                       | 1 (close to large PM2.5 source) | 3                            | 2                            |
| 25                       | 6                                 | 7               | 1                        | 11                                  | 4                       | 2                               | 6                            | 1                            |
| 26                       | 2                                 | 7               | 12                       | 11                                  | 5                       | 2                               | 6                            | 1                            |
| 27                       | 3                                 | 1               | 10                       | 3                                   | 6                       | 4                               | 5                            | 1                            |
| EA 53                    |                                   |                 |                          |                                     |                         |                                 |                              |                              |

|                  |    |    |    |    |    |    |    |    |
|------------------|----|----|----|----|----|----|----|----|
| 1                | 5  | 7  | 9  | 7  | 10 | 10 | 10 | 12 |
| 2                | 9  | 9  | 4  | 12 | 8  | 8  | 8  | 10 |
| 5                | 1  | 2  | 9  | 13 | 4  | 6  | 2  | 3  |
| 14               | 5  | 8  | 6  | 11 | 11 | 6  | 6  | 5  |
| EA 54            |    |    |    |    |    |    |    |    |
| 15               | 10 | 11 | 7  | 9  | 13 | 11 | 13 | 13 |
| 16               | 1  | 3  | 13 | 1  | 12 | 6  | 13 | 12 |
| 17               | 9  | 3  | 1  | 10 | 10 | 4  | 12 | 11 |
| 19               | 7  | 2  | 4  | 6  | 7  | 2  | 10 | 6  |
| EA 57 (Dearborn) |    |    |    |    |    |    |    |    |
| 18               | 13 | 11 | 7  | 4  | 3  | 4  | 12 | 12 |
| 20               | 10 | 4  | 7  | 10 | 5  | 1  | 9  | 5  |
| 21               | 7  | 6  | 7  | 1  | 4  | 1  | 9  | 6  |
| 22               | 3  | 9  | 12 | 4  | 1  | 1  | 8  | 4  |

1  
2

|   |                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                      |                          |
|---|------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| 1 | <b>Table 4 - LUR models developed for Detroit-Dearborn passive monitoring</b>                                                                        |                                                                                                                                                                                                                                                                                                                                      |                          |
|   | <b>Response</b>                                                                                                                                      | <b>Predictive equation</b>                                                                                                                                                                                                                                                                                                           | <b>R<sup>2</sup> (%)</b> |
|   | <b>Log(NO<sub>2</sub>)</b>                                                                                                                           | <b><math>4.43 - 0.05 \text{ Log}(\text{DIST}_{50\text{KP}}) + 0.02 \text{ Log}(\text{DIST}_{90\text{KP}}) + 0.01 \text{ Log}(\text{INT1000}) + 0.02 * \text{Log}(\text{VOC\_BIG\_DIST}) - 0.10 * \text{Log}(\text{PM}_{25\_BIG\_DIST}) - 0.08 \text{ Log}(\text{MN\_BIG\_DIST}) - 1.5 \text{ E-}5 * \text{BRDR\_DIST}</math></b>     | <b>82</b>                |
|   | <b>Log(Benzene)</b>                                                                                                                                  | <b><math>6.23 - 5.04 \text{ E-}5 * \text{DIST}_{50\text{KP}} - 3.4 \text{ E-}8 * \text{INT1000} + 5.99 \text{ E-}5 * \text{P\_DEN500} - 3.8 \text{ E-}5 * \text{MN\_BIG\_DIST} + 1.23 \text{ E-}5 * \text{BRDR\_DIST}</math></b>                                                                                                     | <b>43</b>                |
|   | <b>Log(Toluene)</b>                                                                                                                                  | <b><math>7.42 - 4.17 \text{ E-}5 * \text{MN\_BIG\_DIST} + 1.92 \text{ E-}5 * \text{BRDR\_DIST}</math></b>                                                                                                                                                                                                                            | <b>31</b>                |
|   | <b>Log(Ethylbenzene)</b>                                                                                                                             | <b><math>10.04 - 0.04 * \text{Log}(\text{DIST}_{50\text{KP}}) - 2.91 \text{ E-}6 * \text{DIST}_{90\text{KP}} - 0.05 * \text{Log}(\text{INT1000}) - 8.45 \text{ E-}6 * \text{VOC\_BIG\_DIST} - 0.46 * \text{Log}(\text{MN\_BIG\_DIST}) + 1.79 \text{ E-}5 * \text{BRDR\_DIST}</math></b>                                              | <b>63</b>                |
|   | <b>Log(m,p-xylene)</b>                                                                                                                               | <b><math>8.43 - 1 \text{ E-}4 * \text{DIST}_{50\text{KP}} - 2.81 \text{ E-}5 * \text{DIST}_{90\text{KP}} - 1.51 \text{ E-}6 * \text{INT1000} - 0.13 * \text{Log}(\text{VOC\_BIG\_DIST}) + 0.16 * \text{Log}(\text{PM}_{25\_BIG\_DIST}) - 0.22 * \text{Log}(\text{MN\_BIG\_DIST}) - 6.02 \text{ E-}6 * \text{BRDR\_DIST}</math></b>   | <b>55</b>                |
|   | <b>Log(o-xylene)</b>                                                                                                                                 | <b><math>6 - 0.04 * \text{Log}(\text{DIST}_{50\text{KP}}) - 2.34 \text{ E-}5 * \text{DIST}_{90\text{KP}} + 0.24 * \text{Log}(\text{P\_DEN500}) - 1.34 \text{ E-}5 * \text{VOC\_BIG\_DIST} + 0.09 * \text{Log}(\text{PM}_{25\_BIG\_DIST}) - 0.34 * \text{Log}(\text{MN\_BIG\_DIST}) + 5.42 \text{ E-}6 * \text{BRDR\_DIST}</math></b> | <b>60</b>                |
|   | <b>Log(BTEX sum)</b>                                                                                                                                 | <b><math>10.78 - 4.52 \text{ E-}5 * \text{DIST}_{50\text{KP}} - 0.30 * \text{Log}(\text{MN\_BIG\_DIST})</math></b>                                                                                                                                                                                                                   | <b>40</b>                |
|   | <b>Log(styrene)</b>                                                                                                                                  | <b><math>3.75 + 2.82 \text{ E-}6 * \text{DIST}_{50\text{KP}} - 2.47 \text{ E-}5 * \text{DIST}_{90\text{KP}} + 2.88 \text{ E-}7 * \text{INT1000} + 2.53 \text{ E-}5 * \text{P\_DEN500} - 2.03 \text{ E-}5 * \text{MN\_BIG\_DIST} - 3.45 \text{ E-}6 * \text{BRDR\_DIST}</math></b>                                                    | <b>43</b>                |
|   | <b>Log(1,3-butadiene)</b>                                                                                                                            | <b><math>4.56 - 0.05 * \text{Log}(\text{DIST}_{50\text{KP}}) + 4.1 \text{ E-}8 * \text{INT1000} + 1 \text{ E-}4 * \text{P\_DEN500} + 0.01 * \text{Log}(\text{VOC\_BIG\_DIST}) - 2.82 \text{ E-}6 * \text{PM}_{25\_BIG\_DIST} - 2.95 \text{ E-}5 * \text{MN\_BIG\_DIST} + 1.61 \text{ E-}6 * \text{BRDR\_DIST}</math></b>             | <b>43</b>                |
| 2 | Notes: Bold indicates regression coefficients significant at the 5% level. Log is the natural logarithm. R <sup>2</sup> is reported for the original |                                                                                                                                                                                                                                                                                                                                      |                          |
| 3 | scale, not the log-transformed scale.                                                                                                                |                                                                                                                                                                                                                                                                                                                                      |                          |

**Table 2—Coefficients of variation (CV) for duplicate samplers at compliance monitoring stations and CV over all school sites**

| Pollutant          | East 7 Mile               | Dearborn                  | School sites            |
|--------------------|---------------------------|---------------------------|-------------------------|
|                    | Duplicate CV <sup>1</sup> | Duplicate CV <sup>1</sup> | Spatial CV <sup>2</sup> |
| NO <sub>2</sub>    | 2                         | 3                         | 20                      |
| 1,3-butadiene      | 9                         | 11                        | 26                      |
| Benzene            | 2                         | 4                         | 20                      |
| Toluene            | 1                         | 2                         | 20                      |
| Ethylbenzene       | 1                         | 4                         | 28                      |
| <i>o</i> -xylene   | 1                         | 3                         | 27                      |
| <i>m,p</i> -xylene | 1                         | 4                         | 32                      |
| Styrene            | 4                         | 6                         | 21                      |

<sup>1</sup> Coefficients of variation calculated from 6 pairs. Summarization presents the mean of CVs by week.

<sup>2</sup> Coefficients of variation calculated from mean pollutant values from the 25 school sites.

**Table 53 - Median values of NO<sub>2</sub> and selected VOCs (all in ppbV) for all schools and each monitored schools as a function of enumeration area (EA)<sup>a</sup>**

| Pollutant       | MDL <sup>b</sup> | All schools<br>(n=25)          | EA                      |                         |                                |                         |                                |                         |
|-----------------|------------------|--------------------------------|-------------------------|-------------------------|--------------------------------|-------------------------|--------------------------------|-------------------------|
|                 |                  |                                | 50<br>(n=5)             | 51<br>(n=4)             | 52<br>(n=4)                    | 53<br>(n=4)             | 54<br>(n=4)                    | 57<br>(n=4)             |
| NO <sub>2</sub> | <u>1.00</u>      | <u>15.63</u><br>(11.49, 23.59) | 14.21<br>(11.49, 18.58) | 16.00<br>(12.56, 16.20) | <b>21.15</b><br>(18.50, 22.20) | 15.96<br>(13.65, 23.59) | <b>13.58</b><br>(12.98, 15.08) | 15.52<br>(13.80, 15.98) |
| 1,3-butadiene   | <u>0.04</u>      | <u>0.07</u><br>(0.05, 0.13)    | 0.08<br>(0.05, 0.10)    | 0.07<br>(0.05, 0.10)    | 0.09<br>(0.06, 0.13)           | 0.08<br>(0.07, 0.11)    | 0.06<br>(0.05, 0.07)           | 0.06<br>(0.05, 0.07)    |
| Benzene         | <u>0.02</u>      | <u>0.46</u><br>(0.34, 0.70)    | 0.48<br>(0.39, 0.61)    | 0.47<br>(0.34, 0.65)    | 0.52<br>(0.44, 0.70)           | 0.52<br>(0.46, 0.62)    | 0.40<br>(0.34, 0.45)           | 0.39<br>(0.35, 0.45)    |
| Toluene         | <u>0.02</u>      | <u>1.40</u><br>(0.98, 1.99)    | 1.54<br>(1.23, 1.98)    | 1.50<br>(1.00, 1.99)    | 1.45<br>(1.20, 1.85)           | 1.56<br>(1.40, 1.75)    | 1.26<br>(1.09, 1.42)           | 1.11<br>(0.98, 1.39)    |
| Ethylbenzene    | <u>0.02</u>      | <u>0.18</u><br>(0.12, 0.36)    | 0.20<br>(0.16, 0.25)    | 0.22<br>(0.12, 0.36)    | 0.20<br>(0.16, 0.25)           | 0.20<br>(0.17, 0.31)    | 0.17<br>(0.14, 0.19)           | 0.15<br>(0.13, 0.18)    |
| m,p-xylene      | <u>0.04</u>      | <u>0.59</u><br>(0.36, 1.23)    | 0.63<br>(0.49, 0.79)    | 0.73<br>(0.36, 1.23)    | 0.64<br>(0.52, 0.83)           | 0.65<br>(0.56, 1.04)    | 0.54<br>(0.41, 0.62)           | 0.46<br>(0.40, 0.59)    |
| o-xylene        | <u>0.02</u>      | <u>0.20</u><br>(0.12, 0.34)    | 0.20<br>(0.16, 0.28)    | 0.25<br>(0.12, 0.31)    | 0.21<br>(0.17, 0.28)           | 0.22<br>(0.19, 0.34)    | 0.18<br>(0.15, 0.21)           | 0.16<br>(0.13, 0.20)    |
| Total BTEX      |                  | <u>2.81</u><br>(1.95, 4.11)    | 3.06<br>(2.43, 3.90)    | 3.39<br>(1.95, 4.11)    | 3.04<br>(2.49, 3.90)           | 3.15<br>(2.79, 4.06)    | 2.56<br>(2.13, 2.87)           | 2.26<br>(2.02, 2.81)    |
| Styrene         | <u>0.02</u>      | <u>0.03</u><br>(0.03, 0.06)    | 0.03<br>(0.03, 0.04)    | 0.04<br>(0.03, 0.06)    | 0.04<br>(0.04, 0.05)           | 0.04<br>(0.03, 0.05)    | 0.03<br>(0.03, 0.04)           | 0.03<br>(0.03, 0.04)    |

<sup>a</sup> Minimum and maximum values in parentheses. Values in bold indicate significant difference of ~~mean-ranked~~ pollutant concentrations between pairs of EAs; ~~only NO<sub>2</sub> indicated significantly elevated concentration between EA52 and EA54.~~ See Fig. 1 for EA locations.

<sup>b</sup> Method detection limit

Formatted: Superscript

Formatted: Superscript

Formatted Table

Formatted: Superscript

Formatted: Left

Formatted: Left

Formatted: Left

Formatted: Left

Formatted: Left

Formatted: Left

Formatted: Left

Formatted: Left

Formatted: Left

Formatted: Left

**Table 6 - Coefficients of variation (CV) for duplicate samplers at compliance monitoring stations and CV over all school sites**

| Pollutant             | East 7 Mile               | Dearborn                  | School sites            |                          |
|-----------------------|---------------------------|---------------------------|-------------------------|--------------------------|
|                       | Duplicate CV <sup>1</sup> | Duplicate CV <sup>1</sup> | Spatial CV <sup>2</sup> | Temporal CV <sup>3</sup> |
| <u>NO<sub>2</sub></u> | <u>2</u>                  | <u>3</u>                  | <u>20</u>               | <u>16</u>                |
| <u>1,3-butadiene</u>  | <u>9</u>                  | <u>11</u>                 | <u>26</u>               | <u>31</u>                |
| <u>Benzene</u>        | <u>2</u>                  | <u>4</u>                  | <u>20</u>               | <u>24</u>                |
| <u>Toluene</u>        | <u>1</u>                  | <u>2</u>                  | <u>20</u>               | <u>25</u>                |
| <u>Ethylbenzene</u>   | <u>1</u>                  | <u>4</u>                  | <u>28</u>               | <u>25</u>                |
| <u>o-xylene</u>       | <u>1</u>                  | <u>3</u>                  | <u>27</u>               | <u>26</u>                |
| <u>m,p-xylene</u>     | <u>1</u>                  | <u>4</u>                  | <u>32</u>               | <u>26</u>                |
| <u>Styrene</u>        | <u>4</u>                  | <u>6</u>                  | <u>21</u>               | <u>19</u>                |

<sup>1</sup> CV calculated from 6 pairs. Summarization presents the mean of CVs by week.

<sup>2</sup> CV calculated from mean pollutant values from the 25 school sites.

<sup>3</sup> Temporal CV calculated by site. Tabulated values are the average CV across sites.

Formatted: Centered

Formatted: Superscript

**Table 4 – LUR models developed for Detroit-Dearborn passive monitoring**

| Response                   | Predictive equation                                                                                                                                                                                                                                                                                                   | R <sup>2</sup> (%) |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| <b>Log(NO<sub>2</sub>)</b> | <del>4.43</del> <b>0.05</b> <del>Log(DIST_50KP)</del> <b>+ 0.02</b> <del>Log(DIST_90KP)</del> <b>+ 0.01</b> <del>Log(INT1000)</del> <b>+</b><br><del>-0.02 * Log(VOC_BIG_DIST)</del> <b>-0.10</b> <del>* Log(PM25_BIG_DIST)</del> <b>-0.08</b> <del>Log(MN_BIG_DIST)</del><br><del>-1.5 E-5 * BRDR_DIST</del>         | 82                 |
| <b>Log(Benzene)</b>        | <del>6.23</del> <b>-5.04 E-5</b> <del>* DIST_50KP</del> <b>-3.4 E-8</b> <del>* INT1000</del> <b>+ 5.99 E-5</b> <del>* P_DEN500</del><br><del>-3.8 E-5 * MN_BIG_DIST</del> <b>+ 1.23 E-5</b> <del>* BRDR_DIST</del>                                                                                                    | 43                 |
| <b>Log(Toluene)</b>        | <del>7.42</del> <b>-4.17 E-5</b> <del>* MN_BIG_DIST</del> <b>+ 1.92 E-5</b> <del>* BRDR_DIST</del>                                                                                                                                                                                                                    | 34                 |
| <b>Log(Ethylbenzene)</b>   | <del>10.04</del> <b>-0.04</b> <del>* Log(DIST_50KP)</del> <b>-2.91 E-6</b> <del>* DIST_90KP</del> <b>-0.05</b> <del>* Log(INT1000)</del><br><del>-8.45 E-6 * VOC_BIG_DIST</del> <b>-0.46</b> <del>* Log(MN_BIG_DIST)</del> <b>+ 1.79 E-5</b> <del>* BRDR_DIST</del>                                                   | 63                 |
| <b>Log(m,p-xylene)</b>     | <del>8.43</del> <b>-1 E-4</b> <del>* DIST_50KP</del> <b>-2.81 E-5</b> <del>* DIST_90KP</del> <b>-1.51 E-6</b> <del>* INT1000</del><br><del>0.13 * Log(VOC_BIG_DIST)</del> <b>+ 0.16</b> <del>* Log(PM25_BIG_DIST)</del> <b>-0.22</b> <del>* Log(MN_BIG_DIST)</del><br><del>-6.02 E-6 * BRDR_DIST</del>                | 55                 |
| <b>Log(o-xylene)</b>       | <del>6</del> <b>-0.04</b> <del>* Log(DIST_50KP)</del> <b>-2.34 E-5</b> <del>* DIST_90KP</del> <b>+ 0.24</b> <del>* Log(P_DEN500)</del><br><del>-1.34 E-5 * VOC_BIG_DIST</del> <b>+ 0.09</b> <del>* Log(PM25_BIG_DIST)</del> <b>-0.34</b> <del>* Log(MN_BIG_DIST)</del> <b>+</b><br><del>-5.42 E-6 * BRDR_DIST</del>   | 60                 |
| <b>Log(BTEX_sum)</b>       | <del>10.78</del> <b>-4.52 E-5</b> <del>* DIST_50KP</del> <b>-0.30</b> <del>* Log(MN_BIG_DIST)</del>                                                                                                                                                                                                                   | 40                 |
| <b>Log(styrene)</b>        | <del>3.75</del> <b>+ 2.82 E-6</b> <del>* DIST_50KP</del> <b>-2.47 E-5</b> <del>* DIST_90KP</del> <b>+ 2.88 E-7</b> <del>* INT1000</del> <b>+</b><br><del>-2.53 E-5 * P_DEN500</del> <b>-2.03 E-5</b> <del>* MN_BIG_DIST</del> <b>-3.45 E-6</b> <del>* BRDR_DIST</del>                                                 | 43                 |
| <b>Log(1,3-butadiene)</b>  | <del>4.56</del> <b>-0.05</b> <del>* Log(DIST_50KP)</del> <b>+ 4.1 E-8</b> <del>* INT1000</del> <b>+ 1 E-4</b> <del>* P_DEN500</del><br><b>+ 0.01</b> <del>* Log(VOC_BIG_DIST)</del> <b>-2.82 E-6</b> <del>* PM25_BIG_DIST</del><br><b>-2.95 E-5</b> <del>* MN_BIG_DIST</del> <b>+ 1.61 E-6</b> <del>* BRDR_DIST</del> | 43                 |

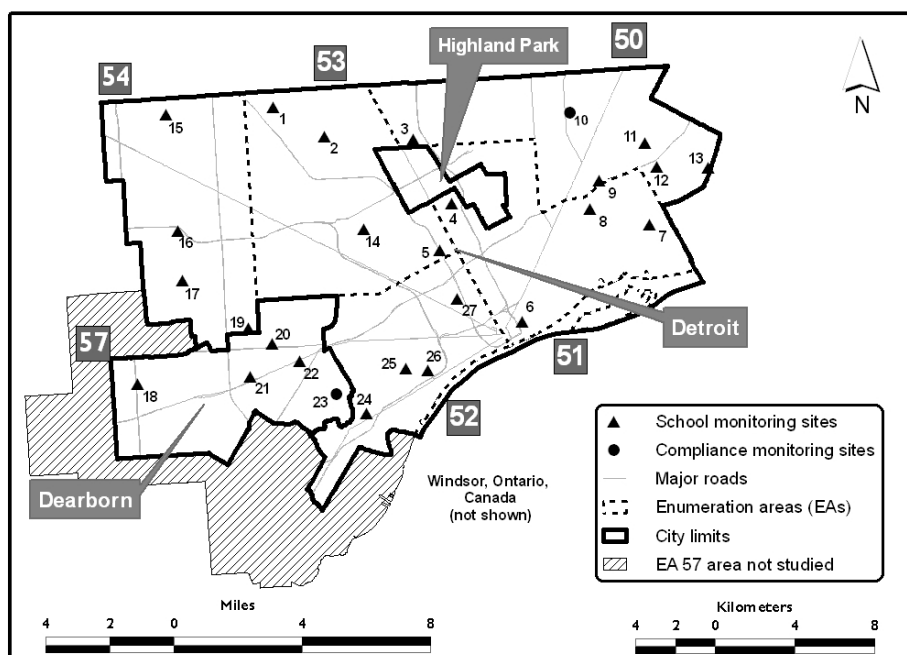
Notes: Bold indicates regression coefficients significant at the 5% level. Log is the natural logarithm. R<sup>2</sup> is reported for the original scale, not the log-transformed scale. **Table 7 - Variables associated with problematic regression coefficients due to collinearity**

| <u>Response compound</u> | <u>Comingled variable sets</u>                                                               |
|--------------------------|----------------------------------------------------------------------------------------------|
| <u>NO<sub>2</sub></u>    | <u>(int, DIST_50KP, PM25_BIG_DIST, MN_BIG_DIST, BRDR_DIST)</u>                               |
| <u>benzene</u>           | <u>(int, DIST_50KP, INT1000, P_DEN500, BRDR_DIST)</u>                                        |
| <u>ethylbenzene</u>      | <u>(DIST_50KP, MN_BIG_DIST)</u><br><u>(int, INT1000)</u><br><u>(VOC_BIG_DIST, BRDR_DIST)</u> |
| <u><i>m,p</i>-xylene</u> | <u>(VOC_BIG_DIST, PM25_BIG_DIST)</u><br><u>(int, BRDR_DIST)</u>                              |
| <u>BTEX sum</u>          | <u>(int, MN_BIG_DIST)</u>                                                                    |
| <u>stryrene</u>          | <u>(int, INT1000, P_DEN500)</u><br><u>(MN_BIG_DIST, BRDR_DIST)</u>                           |
| <u>1,3-butadiene</u>     | <u>(int, VOC_BIG_DIST, BRDR_DIST)</u><br><u>(PM25_BIG_DIST, MN_BIG_DIST)</u>                 |

**Note: “int” refers to the regression intercept.**

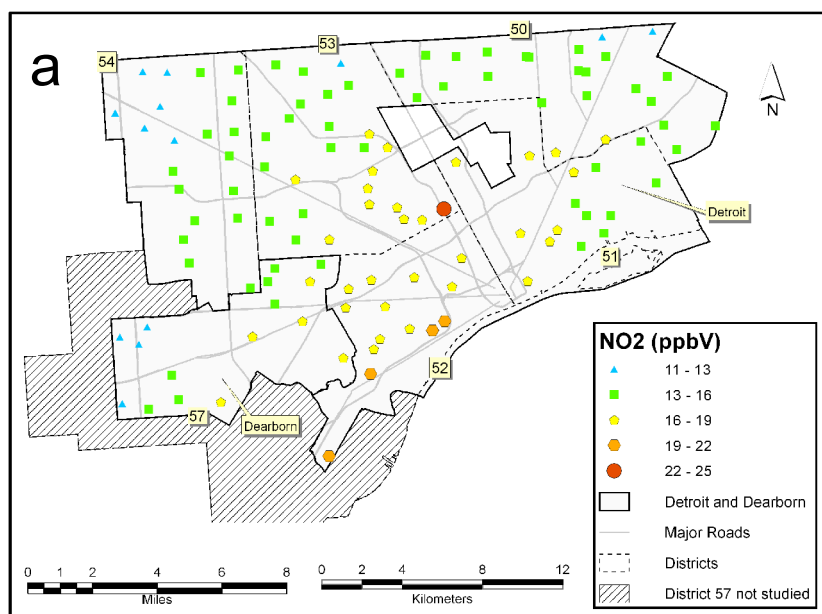
**Table 85 – Comparison of measurements and LUR predictions of NO<sub>2</sub> and selected VOCs (ppbV) at Dearborn and E7 Mile MDEQ sites. ~~LUR predictions at Dearborn and E7 Mile MDEQ sites.~~ (Pollutant concentrations and differences are rounded to the nearest integer or nearest whole per cent. Concentrations in ppbV for VOCs and NO<sub>2</sub>.)**

| Site     | Pollutant          | Measured | LUR Predicted | Difference | % difference |
|----------|--------------------|----------|---------------|------------|--------------|
| E7 Mile  | NO <sub>2</sub>    | 14       | 14            | 0          | 0            |
|          | benzene            | 0.46     | 0.48          | 0.02       | 4            |
|          | toluene            | 1.42     | 1.66          | 0.24       | 16           |
|          | ethylbenzene       | 0.19     | 0.24          | 0.05       | 24           |
|          | <i>m,p</i> -xylene | 0.60     | 0.59          | -0.02      | -3           |
|          | <i>o</i> -xylene   | 0.21     | 0.23          | 0.02       | 11           |
|          | BTEX sum           | 2.88     | 3.30          | 0.42       | 14           |
|          | styrene            | 0.03     | 0.03          | 0.002      | 7            |
|          | 1,3-butadiene      | 0.08     | 0.08          | 0.003      | 3            |
|          |                    |          |               |            |              |
| Dearborn | NO <sub>2</sub>    | 22       | 18            | -4         | -17          |
|          | benzene            | 0.39     | 0.42          | 0.03       | 8            |
|          | toluene            | 1.10     | 1.45          | 0.34       | 31           |
|          | ethylbenzene       | 0.20     | 0.22          | 0.02       | 12           |
|          | <i>m,p</i> -xylene | 0.66     | 0.68          | 0.02       | 3            |
|          | <i>o</i> -xylene   | 0.18     | 0.16          | -0.02      | -12          |
|          | BTEX sum           | 2.53     | 3.04          | 0.51       | 20           |
|          | styrene            | 0.04     | 0.03          | -0.005     | -13          |
|          | 1,3-butadiene      | 0.06     | 0.07          | 0.01       | 13           |

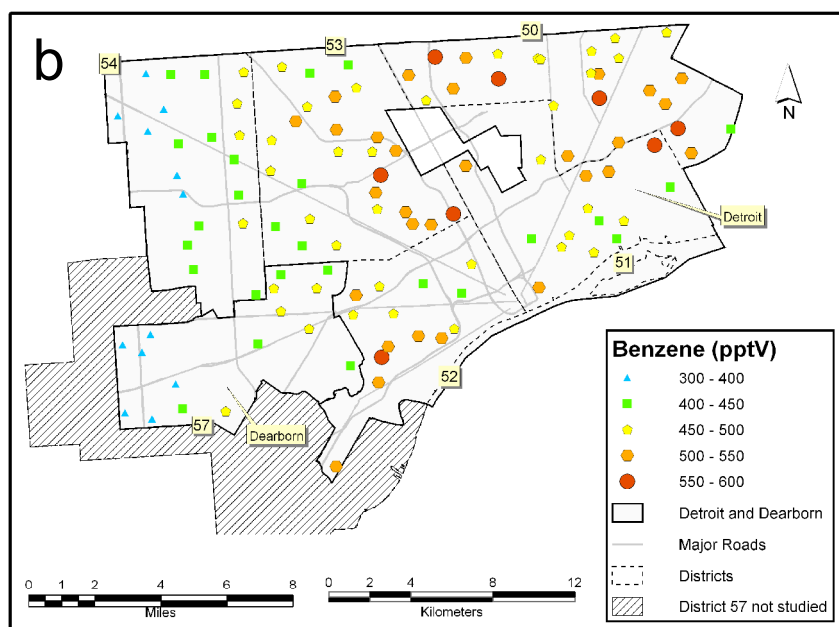


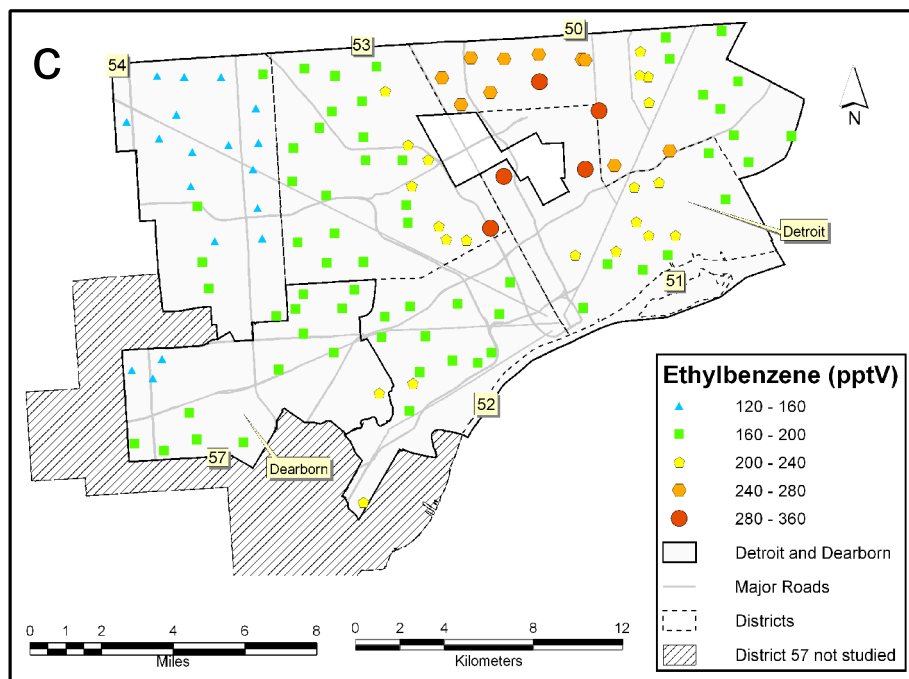
**Fig. 1 - Locations of school<sub>s</sub> and compliance monitoring sites<sup>a</sup> in Detroit and Dearborn, ~~used for this study~~ with enumeration areas ~~delineated and numbered~~.**

<sup>a</sup> All sites were numbered in this study. The two compliance sites, Site 10 (East 7 Mile) and Site 23 (Dearborn), used in this study were established by the State of Michigan for regulatory purposes.



**Fig. 2 - Land use regression predicted ambient air pollution levels at all Detroit and Dearborn local elementary schools for: (a) NO<sub>2</sub>; (b) benzene; (c) ethylbenzene.**





## Supplementary Material

[Click here to download Supplementary Material: Supplementary\\_data\\_\(STOTEN\)\[2\].doc](#)

**RESPONSES TO STOTEN REVIEWER COMMENTS**

(REPLIES IN **BOLD TEXT** IMMEDIATELY BELOW REVIEWER COMMENT)

**NOTE: Redlined/deleted (track changes) version of revised manuscript should be used in evaluation of responses below.**

Reviewers' comments:

Reviewer #1:

The study positions itself as addressing 3 different issues: (1) evaluating impact of traffic and urban emissions on respiratory effects in children, (2) developing spatial models at the coarse and fine scale (3) providing an insight into site selection.

None of these goals is adequately addressed by the paper. Since the focus is not clear, the paper is confusing at many points.

**RESPONSE: While point 1 was a motivating factor, this paper does not address health effects at all. This is planned for a later paper. With respect to point 2, we would not characterize our coarse scale analysis as a modeling effort. We have reviewed the phrasing of the paper and do not feel it is misleading in this regard.**

Referencing in the paper is poor. For example:

(1) pg 2, line 21: These and other results have influenced enactment of recent legislation in California. When I read the CA legislation, it did not mention air pollution - it mentioned noise, hazardous spills and other things.

**RESPONSE: The California Act in question states that a governing school board is prohibited from siting new schools within 500 feet from the "closest traffic lane of a freeway or other busy traffic corridor..." as it relates to "hazardous substances." Hazardous substance is referred to by the Act as defined by California Health and Safety Code Section 25316; this definition includes air pollutants. The introductory language of the Act explicitly mentions as a concern the effect of such pollutants as benzene and 1,3-butadiene on children's health problems including asthma. As the paper indicates, we took this same viewpoint in the Detroit study.**

(2) Pg 3, line 8: Based on these and other monitoring approaches the US EPA conducted air monitoring studies in late 1999. Reading this I assumed they were referring to the citations in the paragraph above it. All those references are 2000 or later. Very clairvoyant of the US EPA!

**RESPONSE: Owing to the size of the manuscript, we were attempting to limit references to the most relevant that reference earlier LUR papers. The seminal LUR reference (Briggs et al., 1997) has been added in the preceding paragraph. Phrasing has been revised to remove ambiguity (see first paragraph of Introduction.)**

(3) In the same line, there is no citation for the EPA El Paso study (it is referenced later).

**RESPONSE: Smith et al. reference now added at the end of the sentence in question (second paragraph of Introduction).**

(4) Surprisingly, no reference to a study by the same author (Johnson MM primary author) to Evaluation of Land-Use Regression Models in Detroit Michigan, Epidemiology 19:6 (2008)

This looks and reads like a paper that has been very hastily put together.

**RESPONSE:** The reference in question was an abstract for a poster, not a peer-reviewed journal article. We focused our references on peer reviewed journal articles and reports as well as textbooks.

Key methodology questions are not in the paper: how many passive samplers were deployed at each of the 25 schools?

**RESPONSE:** Phrasing has been revised to indicate that duplicate passive samplers were located at the compliance sites (first paragraph of Section 2.3).

Were samplers deployed in pairs?

**RESPONSE:** See comment above.

How well did the paired samplers match?

**RESPONSE:** We did not want to repeat information from another paper in this manuscript. Coefficients of variation are shown for the duplicates in current Table 6 (former Table 2) to give the reader a sense of duplicate variability versus spatial variability.

What is the effective least count of the samplers? While a reference to a paper giving details is fine (pg 8, line 3), these details need to be in this paper, as the reliability of the measured data affects the reliability of the statistical model the paper aims to present in this paper.

**RESPONSE:** The current paper is already lengthy as it is. We disagree that these details need to be repeated here as well as in the reference cited.

Statistical analysis is shoddy.

(1) No motivation is given for the Enumeration Area (EA)-based coarse spatial analysis. The analysis itself is not interpreted. It is left at some variables being significant if an inequality given on pg 10 line 1 holds!

**RESPONSE:** We agree that the motivation for EAs was not present in the original version. First paragraph of Section 2.4 has been revised to indicate why EAs were considered as part of the coarse spatial analyses. However, the remainder of this comment is incorrect. It is true that the inequality cited on current first paragraph of page 11 is used to declare significance. We chose to display the inequality in the Methods section because we modified the standard version of Dunn's test and we wished to be explicit about exactly what was done. As stated in the paper, this was done to improve the power of Dunn's test and appropriate references are provided. However, we did not simply leave it at this. Section 3.2 in the Results section provides a complete description of the outcome of these comparisons. Furthermore, the first paragraph of the Discussion and Conclusions section (Section 4) interprets the outcome of the coarse spatial analysis.

(2) In the LUR model (i) some variables are log-transformed. Which variables were transformed and why is not noted. While log transformation are not uncommon, the reason needs to be noted, and hopefully tied to a physical process or interpretation.

**RESPONSE:** Log transforms were utilized based on graphical inspection of the data. While this was stated in the original version, the first paragraph of Section 2.5 has been reworded for clarity. Contrary to the comment, Table 4 explicitly shows all variables that were log transformed in the LURs. For brevity, we left this to this table but we have added a reference to it (first paragraph of Section 2.5).

(ii) Why was MN\_BIG\_DIST even considered as one of the explanatory variables? The explanatory variables were picked a priori, but no adequate reason is given for picking this variable.

**RESPONSE:** We stated in second paragraph of Methods section 2.1 and paragraph 7 of Discussion and Conclusions section 4 that distance from large manganese and PM sources were considered because of their potential use in the health study.

(iii) When the predictors were not significant because of high variance, the observations were weighted with the inverse of Cook's D - pg 11, line 2. There has to be theoretical reasoning for adding weights - not just a statistical argument for fitting the model better, specially when the weighting adds collinearity (pg 13, line 14)

**RESPONSE:** As the paper indicates, the weighting procedure was employed to de-emphasize the effect of specific sites on the individual regressions. The reason why these sites do not conform as well to the behavior exhibited at the other sites is unknown and may provide fodder for future research. However, the fact that this reason is not known does not preclude the use of a weighting scheme to obtain a more reliable and interpretable result from the regression analysis. It is incorrect to state that a statistical technique must await a physical interpretation before it may be applied in an analysis. Indeed, scientific research is often conducted to explain such anomalies uncovered by analyses such as this.

(iv) A note in Table 4 (pg 29) notes that the R<sup>2</sup> values reported are for the original scale, not for the log-transformed scale the model is reported in. At this point, there we are left with a model which cannot be interpreted, and hence cannot be applied, defeating the purpose of model development.

**RESPONSE:** The reviewer is incorrect. It is precisely for interpretability that we calculated the R<sup>2</sup> values on the original scale. This indicates how reliable the model estimates will be when applied to measurement data. We are not alone in this opinion. We have added references along this line at the point where mentioned in the text (first paragraph of Section 3.3).

For these reasons, the analysis as presented is not worthy of publication.

**RESPONSE:** As the above responses indicate, we disagree with this statement.

However, the data is interesting and the paper raises issues (not followed up on) that are worth pursuing.

(1) One of the goals mentioned in the paper is an insight into site selection. The authors of this paper have an ad hoc approach based on maximizing the range of the independent variables, while controlling for collinearity. This approach could possibly be refined using principal component analysis. A comparison between this (refined) approach and Kanaroglou et al's [Atmospheric Environment 39 (2005) 2399-2409] location-allocation algorithm - based on optimizing the range of the pollutant and the exposure - would be interesting to the LUR community.

**RESPONSE:** We point out that our method involves more than just covering the range of independent variables. Multivariate considerations are also taken into account. The reviewer's suggestion of comparing different approaches has merit for future research but is beyond the scope of this paper. With respect to location-allocation, we previously referenced Sahsuvaroglu et al. and Henderson et al. In considering the reviewer's comments, we have added the Kanaroglou et al. reference. We note that part of this future research comparing different approaches would need to deal with the fact that these other papers take a regional perspective in developing their initial model as

opposed to the much more geographically compact Detroit/Dearborn focus we have in this paper.

(2) The transferability of LUR models - across space and time - is another interesting topic. A discussion of transferability of LUR models for Detroit, using this data as the data to be fitted, is also an interesting exercise.

**RESPONSE:** Again, this is a useful topic for future research that we plan to pursue but is beyond the scope of this paper. In general, transference of an LUR model developed in one situation for another application should be approached with caution. For example, in the El Paso study referenced in the paper Smith et al., elevation was used as an explanatory variable but is irrelevant for Detroit. While similar traffic, land use, and other variables can be considered, the pre-analysis approach we present can be utilized to select an appropriate set of relevant explanatory variables for the individual case under study.

(3) The authors could consider alternate modeling techniques to fit the data, for example, spatial interpolation methods (Beleen et al, Mapping of background air pollution at a fine spatial scale across the European Union, Science of the Total Environment, 2009, in press).

**REPONSE:** This is beyond the scope of this paper and most LUR studies in specific cities. The approach presented in Beleen et al. is being applied on a continent-wide basis where considerably more monitors are available. The kriging utilized in the Beleen et al. reference could not have been supported with the available monitoring data from Detroit.

Reviewer #2: This is an interesting piece of research, the selection of air sampling sites and the sequential development of land use regression models. The topic is relevant to STOTEN. However, there are some concerns, making it very difficult to critique the validity of the conclusions. I have listed below things that have caught my attention.

Major comments/suggestions:

1. The method used to select the eight variables and the 25 sampling sites was not well justified, thus it is nearly impossible to replicate the method in another area based on the information provided.

The selection of the eight variables from a pool of 45 was somewhat ambiguous or arbitrary to me. The authors should provide a quantitative procedure on how to select those variables. Similarly, the method to select 25 schools from 116 was presented rather vaguely: "schools were selected to reflect the range of combinations of group numbers across all the potential predictors", please describe quantitatively the selection criteria.

**RESPONSE: We understand the reviewer's desire for a more quantitative procedure and this may be pursued in future work. However, the key point is that for the predictive equations to be valid across the entire geographic region, combinations of the predictor variables at the chosen sites must span the mathematical space covered by the predictor variables. Otherwise, the predictive equations developed cannot be reliably extended over the whole area for which predictions are desired. The difficulty arises from the fact that a variety of variables are to be considered simultaneously. To retain maximum flexibility, we felt that the goals of the project were better served by the procedure described in the paper. To address the reviewer's concerns in this area, we have added a new Table 1 and moved the Supplementary Table S5 to become Table 2. These present summary statistics and correlation structure of the chosen variables. In addition, by revising the text we have tried to clarify the objective of spanning the mathematical space with our choice of sites. We have also added to the text a note that the procedure does not lead to a unique choice of a set of variables or sites.**

The authors mentioned that 25 schools were chosen from a pool of 116. It is not clear to me whether 25 is a predetermined number due to resource limitations, or only 25 schools in the area met the site selection criteria. Similarly, 4 or 5 schools were chosen from each EA. It is not clear to me whether those numbers were predetermined based on the size of each EA or 4/5 schools in the each EA met the site selection criteria. Also, did you decide that you wanted eight variables out of 45 before hand, or only eight met the selection criteria?

**RESPONSE: The number of schools to select was based on resource limitations. Text has been added to indicate this (first paragraph of Section 2.2). Equal distribution of school numbers within EAs was done with an eye to the later health analysis while maintaining the ancillary variable distribution; the paper has been revised to indicate this (second paragraph of Section 2.2). The 8 variables were not decided a priori but resulted from the procedures described in the paper.**

2. Air monitoring. You mentioned that "weeklong" sampling was conducted during a 6-week period of "July 18 - August 30, 2005"; it is not clear which sites in which week were monitored. If all 25 sites were monitored for 6 weeks, please clarify how to get one concentration per site per pollutant; alternatively, if each week only a few sites were monitored, followed by a different group of sites the following week, and so on and so forth, please justify the use of rotating monitoring data in a span of 6 weeks.

**RESPONSE:** We indicated in second paragraph of Section 2.3 that concurrent monitoring was performed at the schools. We revised the first sentence of that paragraph to make it clearer that concurrent monitoring was done at all sites for six weeks.

Page 7, L12-14. I am not sure whether correlation between those variables is the property that one wants to persevere in the site selection process, or the distributions ought to be persevered, but distributions cannot be assessed directly using correlation analysis.

**RESPONSE:** We have clarified in the text that both the correlation structure and the distributions are preserved between the chosen and unmonitored sites. Indeed as indicated above, we have expanded our description of spanning the mathematical space (Section 2.2).

Page 8, L10-11. "Concurrent monitoring was conducted at local neighborhood schools to reflect children's exposures in the immediate community." It is not clear to me: 1) whether this sentence refers to this study or the study of (Morishita et al., 2006). 2) whether the "local neighborhood schools" are the 25 schools or a different set of schools. Suggest rephrasing this sentence and maybe start a new paragraph.

**RESPONSE:** As suggested by the reviewer, a new paragraph was made in terms of the sampling description. Text has been clarified (new third paragraph of Section 2.3).

Page 8, L16. "using a portable, calibrated VOC monitor accurate to ppm levels; no contamination was detected.". Please provide the make and model of the VOC monitor, and describe quantitatively the meaning of "no contamination was detected".

**RESPONSE:** Description of the VOC monitor and the levels it detected are now included (new third paragraph of Section 2.3).

It not clear to me how to obtain EA-level concentrations using 4/5 sites in each EA. The power of the Dunn's test could be low due to small sample sizes (4 or 5)

**RESPONSE:** The Dunn's test utilized the 6-week average concentration at each site within the EA. Text has been added to indicate this (first paragraph of Section 2.4). The reviewer is correct that small sample size inhibits the power of the test. As we noted in the paper, as noted in the paper we modified Dunn's test to increase its power.

3. The method used to develop LUR models is overly simplified or flawed. My major concern is the departure of the LUR model development method presented in this paper from the rest of the models in this field. Generally, a multiple linear regression is conducted by screening all potential independent variables, or predictors, using one-on-one correlation analysis where variables with a very low  $r^2$  value or a high p-value were eliminated; the remaining variables are then used in a stepwise procedure (manually or using SAS script) where the final model is chosen, in which all variables were statistically significant (e.g.  $p < 0.05$ ) and presented the best model fit, in addition to high  $r^2$  and other diagnostic analysis as the authors presented.

**RESPONSE:** As the paper notes, we did effectively initiate the process as indicated by the reviewer but we did it graphically instead of specifically using one-on-one correlations. However, we chose to present these as plots in the supplemental data (Figs. S1) rather than lengthen the paper. The reviewer is correct that stepwise regression is often applied in a multiple linear regression setting. However, we were concerned with applying it

within the context of weighted regression. We chose to present the complete regression results in Table 4 which clearly indicates which variables were significant predictors and coupled this with a discussion of the collinearity resulting from the weighted regression. Note we have added what was the supplemental collinearity material in the text. (see third paragraph of Section 3.3 and new Table 7). The fact that we have chosen a slightly different approach to that commonly used does not invalidate the approach. We believe there is room in the LUR tool kit for multiple options. For example, a generalized additive models approach was used in El Paso (Smith et al., 2006 reference in paper).

Specifically, in this paper,

1) only eight variables from a pool of 45 were considered in LUR models, leading to a) few correlation coefficients being significant (page 10, L18), b) some variables "acting as surrogates for variables not considered" (page 17, L19), and c) some other issues listed below.

**RESPONSE:** We are not sure that the points listed here are the result of using only 8 variables as predictors. Using additional variables from the original 45, or a different set of variables, does not guarantee a different outcome to the analysis. In fact, the collinearity problem might have been worse. a) We assume the reviewer meant regression coefficients, not correlation coefficients. This particular reference is to the original unweighted regression, not the weighted regression reported as the outcome in this paper. b) We only suggested this as part of our discussion of the results, not as a conclusion of the analysis.

2) A weighting scheme was used which cause collinearity (page 13, L14), which might also be the one of the reasons of low R<sup>2</sup> of original modeled and predicted concentrations in some models.

**RESPONSE:** Collinearity does not cause low R<sup>2</sup> values. It does interfere with the interpretation of the regression coefficients, as was noted in the original paper and has been expanded upon in the revised version (third paragraph of Section 3.3).

3) Eight out of nine final models (Table 4) have, as many as five, variables with statically insignificant coefficients (i.e. coefficient was not statically different from zero), which warrants justification. Also I am wondering if the FULL models were significant or not.

**RESPONSE:** Each FULL model reported was significant. Again, we feel that collinearity may have been the problem here and this led us to report the entire suite of variables used in each regression in Table 4 and shifted the bulk of collinearity discussion from supplementary data to the text.

4) Eight out of nine final models (Table 4) have some coefficients with signs that point to a causal relationship that is unexpected. For example, a possible sign of coefficients to "Distance to a border crossing" or "distance to a road with X amount of traffic" suggests that the farther away you are from those places, the higher the concentrations. Similarly, a negative sign of coefficient to "traffic density" indicates that higher traffic density is associated with lower concentrations. Six of those eight models have two or three predictors with questionable signs, I am wondering if this "double incorrect" makes the models "correct". The inclusion of those predictors could partially explain less reliable predictions (page 14, L11-16).

**RESPONSE:** We noticed this as well and, in light of this and another reviewer's comments, we have expanded our discussion of the effect of

**collinearity on the regression coefficients. In particular, the specific cases cited in this point are noted in the new Table 7.**

In summary, I am not convinced that the screening of LUR predictors prior to the air sampling was appropriate. In other words, the authors should be encouraged to explore other variables in the LUR process, which may or may not eliminate issues listed in 2) to 4) above.

**RESPONSE: See our responses above.**

4. The authors may want to make the presentation of the results more effective.

Suggest including general statistics of all concentration data or the selected species, the method detection limits, and the number of valid samples, in a table or a box plot.

**RESPONSE: General statistics for all schools now added to Table 5 (was Table 3) which includes method detection limits.**

The authors may want to rearrange the materials in the Results section such that the results of this study are presented first, followed by comparison to other studies, instead of a comparison without results (e.g. page 11, L14-18).

**RESPONSE: As suggested, summary statistics of study results presented first before comparisons. See redlined text in first paragraph of Results section 3 on p. 12. Also, Section 3.1 has been re-titled as Concentrations. We did this to provide an overall context.**

Page 14, L6-9, suggest tabulating all indexes (e.g. RMSR and ME) of all modes in a table.

**RESPONSE: We feel this level of detail would be uninviting to readers and prefer to leave the discussion of the cross-validation as it is.**

Table 2. I didn't find any numbers about CV between weeks for temporal variability (page 12, L8-10).

**RESPONSE: Temporal CVs have been added to Table 6.**

Discussion and conclusions. The reviewer found that this section is a bit hard to follow due to a frequent switch between the two pollutants (NO<sub>2</sub> and VOCs) and between the two scales (coarse and LUR). For example, the reason of little variability in VOCs was "due to pervasive mobile source effect" in one paragraph (page 15, L16-17), but "due to the fact that winds were generally coming from all directions during each week of the study" in another (page 16, L2-3), then the limitations of the coarse scale approach (page 16).

**RESPONSE: We thank the reviewer for pointing this out. We merged the two paragraphs in question and consolidated the discussion (first paragraph of Section 4).**

Some conclusions in the paper are not supported by appropriate analyses and seem to be almost arbitrary. e.g. page 15, L14-17: "Analysis ..found little variability of VOCs. This suggested a pervasive mobile source effect." and page 17, L3-8.

**RESPONSE: With respect to the first comment, phrasing has been revised to indicate that this is a possible explanation, not a conclusion. With respect to the comment about p. 17 (now p. 18, third paragraph), it is well known from regression analysis that extrapolation beyond the range of the independent variable space is inherently risky. The discussion on these lines simply reiterates this point within the context of LUR modeling.**

Table 4 is a bit hard to follow, suggest the following format:

| Dependent variable and Model R2(%) | Parameter | Estimate | p-value |
|------------------------------------|-----------|----------|---------|
| Y1                                 | Intercept | 1        | 0.001   |
| R2=0.80                            | X1        | 0.1      | 0.002   |
|                                    | X2        |          |         |
| 0.3                                | 0.04      |          |         |
| Y2 (60)                            | Intercept | 1        | 0.001   |
| R2=0.60                            | X1        | 2        | 0.00    |
|                                    | X3        | 3        | 0.03    |

**RESPONSE:** We disagree that the suggested format for Table 4 would be easier to follow. For reasons indicated above, we wish to report the full regression model that was attempted in each case.

Fig 2. The LUR model predicted surfaces should be provided with all schools marked. The discussion (page 15, L-10) should be based on these LUR maps, instead of values at the schools. The reason is that one should establish the credibility of the models (i.e. predicted surfaces are smooth and free from irregularity or unexpected features) first before using the models to predict the concentrations at the schools.

**RESPONSE:** We disagree with this comment. We feel that the use of Figs. 2a-c and the relevant discussion are appropriate. The paper already provides considerable discussion with respect to the credibility of the models. In particular, the regression coefficient estimates, cross-validation results, regression diagnostics including residual diagnostics and AIC and BIC results, and comparison of predicted versus measured values are already presented.

Editorial comments/suggestions:

The term "mathematical space" was rather confusing; suggest replacing it with better words, maybe "statistical space" or "statistical property".

**RESPONSE:** The correct term is "mathematical space" and we have elaborated on this in the revised version (first paragraph of Section 2.2).

Suggest using international units, since most journals would not allow other units. For the same reasons, please delete scales in Miles in Figs 1 and 2. Units of the eight variables (pages 5&6) should be provided.

**RESPONSE:** Scale was provided in km and miles for ease of reference by local community groups and US and international investigators. We have adopted the reviewer's suggestions and have provided the units for the eight variables in the text (see last paragraph of Section 2.1). Complete descriptions of the ancillary variables are provided in Tables S1 to S4.

Significant numbers: suggest one decimal, i.e. 12.3, for NO2 concentrations throughout the text and in all tables.

**RESPONSE:** We prefer to leave the tables as they are.

There are a bit too many sentences in ( ). Suggest limiting its use to minima.

**RESPONSE:** Sentences and phrases in parenthesis have been minimized throughout revised manuscript.

Abstract

The flow is a bit hard to follow. Suggest following the default structure: Objectives, Methods (site selection and monitoring methods), Results, and Conclusions.

**RESPONSE:** Some text has been added and some phrases deleted to enhance the flow of the Abstract.

#### Introduction

I don't think that the study of Luginaah et al. (2006) was conducted using school sites. Maybe this reference could be moved to page 3, L5. Several recent studies have also included LUR modeling of SO<sub>2</sub>.

**RESPONSE:** Luginaah et al. (2006) stated that for their passive NO<sub>2</sub> monitoring to develop LURs, "...the monitoring site selections were based primarily upon proximity to elementary schools, as well as ensuring inclusion of all types of land uses such as road networks, industry, and residential settings..." A recent study on LURs for SO<sub>2</sub> (Wheeler et al., 2008) is now included.

#### Method

Section 2.1 is a bit hard to follow. The authors may want to rearrange the materials in order to improve the flow. For example, start with the types of variables that you wanted to include (page 5), move on to the data sources (page 4), then how to select 8 from 45.

**RESPONSE:** We have added a sentence, as the reviewer suggested, indicating the general types of variables originally considered (second paragraph of Section 2.1).

Page 5, L8-10. Suggest clarifying: 1) how many border-crossing points were considered, 2) whether the distance to a crossing was used as a surrogate of Canadian emissions.

**RESPONSE:** Added text that two border crossings were considered, Ambassador Bridge and Detroit-Windsor Tunnel (second paragraph of Section 2.1). We stated in that same paragraph that "No point source or traffic data for neighboring Windsor, Ontario were available". But the distance to border crossing was not used as a surrogate for Canadian emissions; the text has in the same paragraph been revised to clarify this.

Page 5, L16, I think the authors meant "used in site selection"; since LUR method was presented in Section 2.5.

**RESPONSE:** Text has been revised to say "...to be used in the LUR models..." (see second paragraph of p. 6).

Page 10, L10-12. Please clarify how to decide which variables to log-transform.

**RESPONSE:** Revised the text to make it clearer what we did in first paragraph of Section 2.5.

#### Results

Page 11, subheading of 3.1, could read "Concentrations"

**RSPONSE:** Subheading so revised.

Page 11, L12-14. The sentence is a bit confusing.

**RESPONSE:** We thank the reviewer for pointing this out. The text has been revised (first paragraph of Section 3.1).

Page 11, L18. Suggest including a few recent studies in the reference list.

**RESPONSE:** An additional, recent, reference has been added (first paragraph of Section 3.1).

Other editorial suggestions:

| Page | Line | Suggestion |
|------|------|------------|
|------|------|------------|

|                                                                                              |       |                                                                                                                                      |
|----------------------------------------------------------------------------------------------|-------|--------------------------------------------------------------------------------------------------------------------------------------|
| 1                                                                                            | Title | Spatial analysis and land use regression of VOCs and NO <sub>2</sub> from school-based urban air monitoring in Detroit/Dearborn, USA |
| <b>RESPONSE: Title revised and emphasized Detroit and Dearborn, MI, USA in the abstract.</b> |       |                                                                                                                                      |

|                        |   |                                                   |
|------------------------|---|---------------------------------------------------|
| 2                      | 4 | NO <sub>2</sub> and selected VOCs across the area |
| <b>RESPONSE: Done.</b> |   |                                                   |

|                        |   |                                  |
|------------------------|---|----------------------------------|
| 3                      | 2 | predict pollutant concentrations |
| <b>RESPONSE: Done.</b> |   |                                  |

|                        |     |            |
|------------------------|-----|------------|
| 416                    | 914 | LUR models |
| <b>RESPONSE: Done.</b> |     |            |

|                                                                       |       |                                                                                                                                                                                                                                      |
|-----------------------------------------------------------------------|-------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4                                                                     | 20-22 | distance from school to nearest point source of VOCs and PM (emission data from the EPA 1999 National Emission Inventory database) as well as manganese (Mn) (emission data from the Michigan 2002 Toxic Release Inventory database) |
| <b>RESPONSE: Revised as suggested but minimizing the parentheses.</b> |       |                                                                                                                                                                                                                                      |

|                        |     |                                                                               |
|------------------------|-----|-------------------------------------------------------------------------------|
| 5                      | 2-3 | 2006; Hoek et al., 2008); these variables were also considered in this study. |
| <b>RESPONSE: Done.</b> |     |                                                                               |

|                        |    |                                 |
|------------------------|----|---------------------------------|
| 5                      | 17 | Pearson correlation coefficient |
| <b>RESPONSE: Done.</b> |    |                                 |

|                                                                                                                   |    |                                                                   |
|-------------------------------------------------------------------------------------------------------------------|----|-------------------------------------------------------------------|
| 5                                                                                                                 | 23 | Suggest deleting "large" since the emission amount was specified. |
| <b>RESPONSE: Retained since emphasizing what the large source is designated in terms of range pounds emitted.</b> |    |                                                                   |

|                                                                             |      |                                                                   |
|-----------------------------------------------------------------------------|------|-------------------------------------------------------------------|
| 6                                                                           | 1, 2 | Suggest deleting "large" since the emission amount was specified. |
| <b>RESPONSE: Retained large to be clear that smaller sources did exist.</b> |      |                                                                   |

|                                                                            |   |                              |
|----------------------------------------------------------------------------|---|------------------------------|
| 6                                                                          | 5 | In this study, schools were. |
| <b>RESPONSE: Revised first paragraph of Section 2.2 for reasons above.</b> |   |                              |

|                                                                            |   |                                                |
|----------------------------------------------------------------------------|---|------------------------------------------------|
| 6                                                                          | 7 | ensure that reliable predictions of something. |
| <b>RESPONSE: Revised first paragraph of Section 2.2 for reasons above.</b> |   |                                                |

|                        |     |                                 |
|------------------------|-----|---------------------------------|
| 7                      | 3-4 | for each of the eight variables |
| <b>RESPONSE: Done.</b> |     |                                 |

|                        |       |                                                                                                                                                                                         |
|------------------------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 7                      | 18-20 | Passive samplers were deployed outdoors at the 25 selected schools and at two compliance sites operated by the Michigan Department of Environmental Quality (MDEQ), as shown in Fig. 1. |
| <b>RESPONSE: Done.</b> |       |                                                                                                                                                                                         |

|                        |     |                                                            |
|------------------------|-----|------------------------------------------------------------|
| 8                      | 4-5 | Weeklong integrated sampling or Weeklong sampling interval |
| <b>RESPONSE: Done.</b> |     |                                                            |

8                    21                    These sampling methods have been validated  
**RESPONSE: Done.**

9                    2-4                    Seems out of place, suggest deleting.  
**RESPONSE: Done.**

9                    11                    Monitored concentrations of NO<sub>2</sub>, 1,3-butadiene, BTEX species, and styrene were.  
**RESPONSE: Done.**

9                    16-17                    The Kruskal-Wallis test (SAS, 2004b) was used to  
**RESPONSE: Sentence revised but NPAR1WAY procedure retained. See first paragraph of Section 2.4.**

10                    equation and 1st paragraph, suggest deleting and referencing if possible.  
**RESPONSE: We feel it should be retained because we modified Dunn's test.**

10                    19                    coefficients were statistically significant (p<0.05). The same, i.e. (p<0.05), applies to other places.  
**RESPONSE: Done.**

11                    3-4                    A discussion of both transformations and weighting in a regression context can be found in Carroll and Ruppert (1988).  
**RESPONSE: Done.**

11                    subheading of 3.1, could read "Concentrations"  
**RESPONSE: Done.**

11                    12-14                    The sentence is confusing.  
**RESPONSE: Revised as noted above.**

11                    18                    Suggest including more recent studies in the reference list.  
**RESPONSE: Revised as noted above.**

11                    19-21                    Seems to be out place, suggest deleting.  
**RESPONSE: Would like to retain since this indicates other pollutants measured. Revised to state these pollutants were not analyzed further since below detection (end of first paragraph, Section 3.1).**

12                    2                    p=0.27), could be due to slightly lower ambient levels (Table X).  
**RESPONSE: Text revised as suggested (second paragraph of Section 3.1).**

12                    6-7                    coefficients of variation (CV) were calculated  
**RESPONSE: Done.**

12                    18                    largest length of freeways  
**RESPONSE: Done.**

13                    5                    Seems awkward, please rephrase: "figured into one of the potential predictors and in recognition of the California school siting legislation".  
**RESPONSE: Phrasing revised (last paragraph of Section 3.2).**

13                    10                    Suggest deleting.  
**RESPONSE: Done (first paragraph, Section 3.3).**

13                    11                    Table 4 presents the final LUR models.  
**RESPONSE: Done.**

14                    2                    Cross-validation (Cressie, 1993),  
**RESPONSE: Done.**

14                    10-11                    this was attributable to large discrepancies  
between measurements and predictions at one, two, or three sites.  
**RESPONSE: Done.**

14                    21                    NO2 was predicted well at the East 7 Mile site.  
**RESPONSE: Done.**

15                    12                    School-based ambient air monitoring was successfully  
conducted in Detroit and.  
**RESPONSE: Done.**

16                    10                    "has a large number of .", or "has facilities that  
emit large amounts of ."  
**RESPONSE: Done.**

16                    19                    coarse level testing, Sentence seems awkward, please  
rephrase.  
**RESPONSE: Sentence has been rephrased, fourth paragraph of Section 4.**

16                    21-22                    the prediction of pollutant levels across the  
entire area by considering multiple variables simultaneously, without being  
restricted by any discretization.  
**RESPONSE: Done.**

17                    13-14                    why all GA-level mean values were  
statistically the same, expect one pair.  
**RESPONSE: Revised, paragraph 6 of Section 4.**

17                    23                    Suggest moving to Results.  
**RESPONSE: Sentence put in Methods section (paragraph 3, Section 2.1).**

18                    9-11                    Seems awkward, please rephrase.  
**RESPONSE: Sentence revised, last sentence of text.**

Table 1                    Caption                    Group rankings by explanatory variables  
for 25 monitored schools.  
**RESPONSE: Changed "by" to "at"; new Table 3.**

Table 3                    Caption                    Median values of NO2 and selected VOCs  
(all in ppbV) in each enumeration area  
**RESPONSE: Current Table 5 caption revised. Table also includes summary  
statistics for all schools.**

Table 5                    Caption                    Comprising of measurements and LUR  
predictions of NO2 and selected VOCs (ppbV) at Dearborn and E7 Mile MDEQ  
sites.  
**RESPONSE: Done, current Table 8 caption.**

Figure 1                      Caption                      Locations of schools and compliance  
monitoring sites: Site 10 at East 7 Mile and Site 23 at Dearborn.

**RESPONSE: Revised.**

Tables                      footnotes                      Table 2, delete 2Table 3, delete last  
two sentences

**RESPONSE: Current Table 5 (was 3) footnote revised.**

Reviewer #3: This is an impressive paper that documents the performed research very well. However, I have some minor comments for the authors:

1. Line 14, pg1 -Abstract should indicated that Michigan is in the USA

**RESPONSE: Done.**

2. Line 21, pg2 - "These and other results." influenced the law enacted in 2003, except that the articles referred to in "these" (Singer et al., 2004, Luginaah et al., 2006) are dated after the law. Please reword the beginning of this sentence to tie the information together better

**RESPONSE: Text has been revised, first paragraph of Introduction.**

3. Line 8, pg3 - Similar to #2 above, ". these and other monitoring approaches", the dates do not align. Please reword.

**RESPONSE: Sentence in question has been deleted; second paragraph of Introduction.**

4. Line 1, pg4 - please describe or suggest what other urban influences are referred to here.

**RESPONSE: Added industrial emissions and population (last paragraph of Introduction).**

5. Line 18, pg 4 - why was this classification of traffic volumes used? Were other ranges tested?

**RESPONSE: All the potential predictors are listed in the supplementary material. Increments at 10,000 was based on a combination of convenience, prior experience, and the California legislation.**

6. Line 10, p5 - no source data WAS available (But this could be a stylistic comment)

**RESPONSE: Moved the sentence but left it as "were"; second paragraph, Section 2.1.**

7. Line 17, p5 - Please discuss some of the correlations between selected and non-selected variables within the same group. I could not locate this, but it would be of interest.

**RESPONSE: The revised paper has been lengthened and we don't feel adding this would be worth the additional length.**

8. Line 18, p5 - What does "reasonable" amount of variability mean?

**RESPONSE: All ancillary variables had a coefficient of variation > 30% which has been added to the text. We describe variability in Table 1. Text so revised (fourth paragraph, Section 2.1).**

9. Line 4, p8 -

a. What was the summer season chosen? If during summer, the highest spatial gradients are observed, what is the implication of what this means for the distribution of pollution during the rest of the year? Or for an annual summary? The issue of seasonality is a common theme for all the LURs and has to be addressed in this paper.

**RESPONSE: Budgetary constraints restricted monitoring to one season. The reasons for choosing the summer season were noted in second paragraph of Section 2.3. Higher concentrations will likely be more suitable for the planned health analyses. We appreciate the reviewer's concern about**

seasonality. Unfortunately, we could only obtain one season of monitoring and cannot support a seasonal analysis in this paper.

b. Why was only one week chosen?

**RESPONSE:** The paper has been revised to indicate that weeklong sampling was conducted for the 6 weeks from mid-July to August (second paragraph of Section 2.3). The weeklong sampling period was chosen to better represent the chronic exposure that will be studied in the health analyses.

10. Table 1, p 25 - Each school was ranked by the 8 variables. Does it make sense to provide an average ranking across each school? It would add some meaning to the table.to the table.

**RESPONSE:** Averaging the rankings across each school would not be appropriate since the variables are of different types. Moreover, it is a school's individual rankings on the different variables that jointly are of importance in site selection relative to other schools.