Statistical Analysis of Soil Lead Concentrations in Vernon, CA

prepared for:

Advanced GeoServices

West Chester, Pennsylvania

prepared by:

Dr. Mitchell J. Small and Dr. Stephen M. Rose

Carnegie Mellon University

Pittsburgh, Pennsylvania

August 12, 2015

Disclaimer

Dr. Mitchell J. Small and Dr. Stephen M. Rose are employees of Carnegie Mellon University (CMU). However, this study was conducted as an independent consulting project without the sponsorship or oversight of CMU. As such, the content contained within the report does not represent the official statements or views of CMU. Responsibility for the content of the report lies solely with Dr. Small and Dr. Rose.

Acknowledgment of Responsibility

The undersigned acknowledge they have conducted this study and concur with its presentation as found in this report.

Signature: ______________________ Date: August 12, 2015

Print Name: Mitchell J. Small

Signature: ______________________ Date: August 12, 2015

Print Name: Stephen M. Rose

Executive Summary

The purpose of this study is to analyze the spatial pattern of soil lead concentrations measured in the area extending outward from the Exide facility in Vernon, California, and to derive an estimate of the contribution to these concentrations made by the historic lead emissions from the facility. The analysis seeks to analyze and characterize the Exide lead contribution and compare Exide's contribution to lead concentrations resulting from urban lead sources and associated elevated ambient air concentrations historically present in cities such as Los Angeles (with both high industrial activity and high traffic density).

The study analyzes soil samples collected from 244 residential properties and 163 non-residential sites in and around Vernon. We apply three analysis methods to these data to estimate the contributions of the Exide and non-Exide factors to observed soil lead concentrations. The first analysis uses kriging to smoothly interpolate soil lead concentrations in order to graphically show larger spatial patterns. The second analysis fits a piecewise regression model to estimate soil lead concentration as a function of distance from the Exide facility. The third analysis fits a regression model to estimate soil lead concentration as a function of modeled ambient air concentration of lead dust from the Exide facility and site characteristics such as distance to the nearest freeway or arterial road and year of housing construction (a proxy for lead paint).

The two regression analyses give several results that are consistent with each other and with the kriging analysis. First, average soil lead concentrations tend to decrease with distance from the Exide facility and then, beyond a certain distance, level off and no longer continue to decrease. This distance represents a threshold at which the contribution from the Exide facility becomes statistically indistinguishable from the ambient urban background concentration. We estimate this distance to be as large as 1.2 km in the prevailing wind direction and as small as 0.3 km for some of the other directions. For comparison, the nearest houses in the Initial Assessment Areas (IAAs) are approximately 1.3 km from the Exide facility. In a separate analysis, we find that soil lead concentrations in the IAAs are statistically indistinguishable from concentrations in the Expanded Assessment Areas (EAAs). Taken together, these findings suggest that the Exide contribution to soil lead concentrations in these areas is not statistically distinguishable from the ambient urban background.

Second, we estimate the average soil lead concentration beyond the threshold distance, which we describe as the "ambient urban background" level, to range from 129 mg/kg for a sampling site not near old houses, freeways, and arterial roads to 242 mg/kg for a sampling site near a house built before 1940 and near an arterial road. An average value of 218 mg/kg is inferred using the results from all sites sampled beyond the threshold distance, given the mix of age and proximity to freeways and arterial roads of the sampling locations. These average urban background concentrations are somewhat higher than those estimated in other studies of the greater Los Angeles area and a section of Long Beach, California. However, this is consistent with the longer and more concentrated history of high traffic density and high industrial activity in the Vernon area compared to most portions of the other study areas, including a particular history of lead processing industrial plants in the vicinity of, but unrelated to, the Exide site.

Table of Contents

1 INTRODUCTION 1

1.1 .... BACKGROUND AND CONTEXT 1

1.2 .... OBJECTIVE OF THIS ANALYSIS 1

2 METHODS 2

3 IMPLEMENTATION AND RESULTS 3

3.1 .... KRIGING INTERPOLATION 3

3.2 .... REGRESSION ANALYSES OF SOIL LEAD CONCENTRATION VS. DISTANCE 4

3.3 .... STATISTICAL ANALYSIS OF CONTRIBUTORS TO SOIL LEAD CONCENTRATIONS 8

4 DISCUSSION 11

4.1 .... COMPARISON TO AMBIENT URBAN BACKGROUND FROM OTHER STUDIES 11

4.1.1 Expanded Assessment Areas (EAAs) 12

4.1.2 Long Beach 14

4.1.3 U.C. Irvine Study (Wu, et al.) 15

5 FINDINGS AND CONCLUSIONS 15

5.1 .... FINDINGS 15

5.2 .... CONCLUSIONS 16

6 REFERENCES 17

APPENDIX A. SOIL LEAD SAMPLING AND ANALYSIS 19

A.1 SOIL SAMPLES 19

A.1.1 Selection of included depth ranges 19

APPENDIX B. STATISTICAL METHODS AND ANALYSIS PROCEDURES 20

B.1 KRIGING INTERPOLATION 20

B.2 REGRESSION MODELS OF SOIL LEAD CONCENTRATION VS. DISTANCE 21

B.3 STATISTICAL MODEL OF CONTRIBUTORS TO SOIL LEAD CONCENTRATIONS 23

B.3.2 Ambient Air Concentrations of Atmospheric Dispersion Model (AERMOD) 23

B.3.3 Sensitivity of Results to Distances Used to Determine Proximity to Freeways and Arterial Roads 25

B.3.4 Proximity to Other Possible Industrial Lead Sources 26

B.3.5 Drip Zone, Downspout, and Other Areas of Enhanced-Deposition 27

B.3.6 Quantifying Uncertainty of Ambient Urban Background: An Approach Based on Measurement Error Determined from Duplicate Samples 27

B.3.7 Calculating Modeled Ambient Air Concentration Corresponding to Soil Lead Concentration Where Exide Contribution is Distinguishable from Background 29

APPENDIX C. HISTORIC LEAD INDUSTRIES AND HOT SPOTS **29**

APPENDIX D. AREAS USED TO ESTIMATE BACKGROUND FOR COMPARISON **32**

APPENDIX E. LARGER VERSIONS OF FIGURES ABOVE **34**

E.1 FIGURE 1 34
E.2 FIGURE 2 35
E.3 FIGURE 3 36
E.4 FIGURE 4 37
E.5 FIGURE 5 LEFT 38
E.6 FIGURE 5 CENTER 39
E.7 FIGURE 5 RIGHT 40
E.8 FIGURE 6 41
E.9 FIGURE 7 42

1 Introduction

1.1 Background and Context

Professor Mitchell Small and post-doctoral research associate Stephen Rose of Carnegie Mellon University were retained by Exide Technologies Inc. (Exide) to perform statistical analyses on soil lead data collected in the vicinity of its Vernon, California lead recycling facility. Exide has performed extensive soil sampling in the area surrounding its Vernon facility pursuant to agreements with the California Department of Toxic Substances Control (DTSC) that require Exide to establish the extent of impact from its facility to 80 mg/kg lead or background, whichever is higher. The November 2014 agreement required that the document *Strategies for Establishing and Using Background Estimates of Metals in Soils (DTSC, August, 2008)* be used as guidance for the purposes of establishing background. The techniques used in this study are consistent with this guidance document and have been used successfully by Dr. Small at other lead recycling facilities to separate contributions to soil lead levels that result from the deposition of lead from industrial site air emissions from those caused by other anthropogenic sources. (Small et al. 1995) The data used in this study were provided by Advanced GeoServices Corp. and ENVIRON International Corporation, both consultants to Exide.

Dr. Small and Dr. Rose were retained by Exide in September 2014. A meeting was held at DTSC's offices in Sacramento, CA on October 23, 2014 to introduce DTSC to the proposed methodology. Subsequent to that meeting, DTSC provided comments on the methodology in a letter dated November 24, 2014 that were incorporated into the work plan titled *Background Study Work Plan* (Work Plan) and dated December 23, 2014. DTSC provided comments on the Work Plan on February 26, 2015; Exide is providing responses to those comments under separate cover.

1.2 Objective of this Analysis

The purpose of this study is to analyze the spatial pattern of soil lead concentration measured in the area extending outward from the Exide facility in Vernon, California, and to derive an estimate of the contribution to these concentrations made by the historic lead emissions from the facility. The analysis seeks to characterize and distinguish the Exide lead contribution from that resulting from urban lead sources and associated elevated ambient air concentrations historically present in cities such as Los Angeles (with both high industrial activity and high traffic density). We refer to soil lead concentrations that result from this general, regional effect as "urban baseline (or ambient background) concentrations." The analysis also identifies sample and property-specific factors that lead to further enhancement and elevation of urban background lead concentrations, including proximity to highways or heavily travelled arterial roads, proximity to a house built before 1940 (and therefore subject to contributions from lead paint), location in a drainage "drip zone" (where deposited lead may accumulate and lead based paint effects are most pronounced), and proximity to other historic lead processing facilities. These latter, non-Exide factors are said to contribute to the "elevated urban background concentrations with attributable sources" in the study area.

2 Methods

The contributions of the Exide and non-Exide factors to observed soil lead concentrations are estimated using a set of graphical and quantitative statistical methods. The objective is to see whether similar findings are inferred from these methods regarding the spatial extent of discernible impact from the Exide site on the surrounding soil lead concentrations. The statistical methods include:

1) Developing a "kriged" map of soil lead concentrations in the study area. The map uses the observed spatial correlation structure of the soil lead measurements to interpolate between them, producing a smoothed estimate of soil lead variations across the mapped area. (Section 3.1)
2) Plotting and fitting least squares regression analyses of soil lead concentration vs. distance from the Exide site. These regressions yield an estimate of the distance from the site at which mean soil lead concentrations "level off" to an estimated urban background soil lead concentration for the surrounding area. The urban background soil lead concentration is also estimated as an output of the fitted regressions. (Section 3.2)
3) Implementing a statistical regression of soil lead concentration vs. modeled ambient air concentration and a set of 0-1 (where 0 indicates that the variable is not present and 1 indicates that the variable is present) indicator variables. These indicator variables include whether the property is near a freeway, near a major arterial road, includes a house built before 1940, or is near a lead processing facility (other than the Exide facility), paint manufacturing facility, or metal-working facility. The fitted regression provides estimates of the average urban background soil lead concentration, the average additional soil lead concentration attributable to each indicator variable, and the Exide contribution to soil lead concentrations associated with each value of the average annual ambient air lead concentration simulated by the atmospheric dispersion model for locations within the study area. (Section 3.3)

All three methods use a common dataset of property-averaged soil lead concentrations measured in the top 6 inches of the soil. Concentrations in this surface zone were very similar with depth (0-1 inches vs. 1-3 inches vs. 3 – 6 inches, see Appendix A.1.1). Soil samples collected in the "drip zone" are analyzed separately because samples collected within the drip zone may be affected by runoff from roofs containing suspended fine particulates and/or by the presence of lead based paint. Table 1 provides a summary of the 635 property averages that were analyzed, including the number of sample averages exhibiting each of the discrete designations for exposure to historic lead paint (built before 1940), and proximity to freeways, arterial roads, or historic industrial sites other than Exide.

Further details on the monitoring program are presented in Appendix A.

Table 1: Number of observations from Residential and Non-Residential areas in the Vernon Soil Lead Dataset used for the Statistical Analyses (values averaged over the 0-6 inch depth interval)

	Total	Built before 1940 [1]	Freeway < 200 ft	Arterial Road < 200 ft	Lead industry[2] < 1000 ft	Non-lead Metal industry[2] < 1000 ft	Paint mfg. [2] < 1000 ft
Residential	244	174	3	40	0	3	0
Residential Dripzone	228	162	3	38	0	2	0
Non-residential	163	0[3]	3	50	19	8	1
Total	635	336	9	136	19	13	1

1. The year of house construction was provided to Drs. Small and Rose by Advanced GeoServices Corp. based on data obtained from Zillow® real estate data.
2. The locations of historic non-Exide lead industries, non-lead metals industries or paint manufacturing facilities were provided to Drs. Small and Rose by Advanced GeoServices Corp. based on a review of Sanborn Fire Insurance maps.
3. Non-residential samples were taken from public areas, and therefore are not associated with houses or structures

3 Implementation and Results

3.1 Kriging Interpolation

We use kriging to smoothly interpolate measured soil lead concentrations in order to reveal larger-scale patterns. Kriging estimates the value of a variable (in this case, the soil lead concentration) at a given location as the weighted average of the measured values within a given distance of the location. Measured values within 500 meters were included in the estimate. Further details about the kriging procedure are given in Appendix B.1 .

The interpolated kriging map in Figure 1 shows a distinct area of elevated soil lead concentration at and near the Exide facility, with a notable dark red area (concentration greater than 400 mg/kg) extending out from the site. However, concentrations in this elevated zone fall rapidly to less than 320 mg/kg (the California soil screening value for industrial areas, shown in green), to light then dark blue areas indicative of concentrations less than 225 and 80 mg/kg, respectively. The locations of soil lead measurements (average concentration of yard area samples over a depth of 0-6 inches for residential properties or the average concentration over a depth of 0-6 inches at non-residential locations) are shown as white dots in Figure 1.

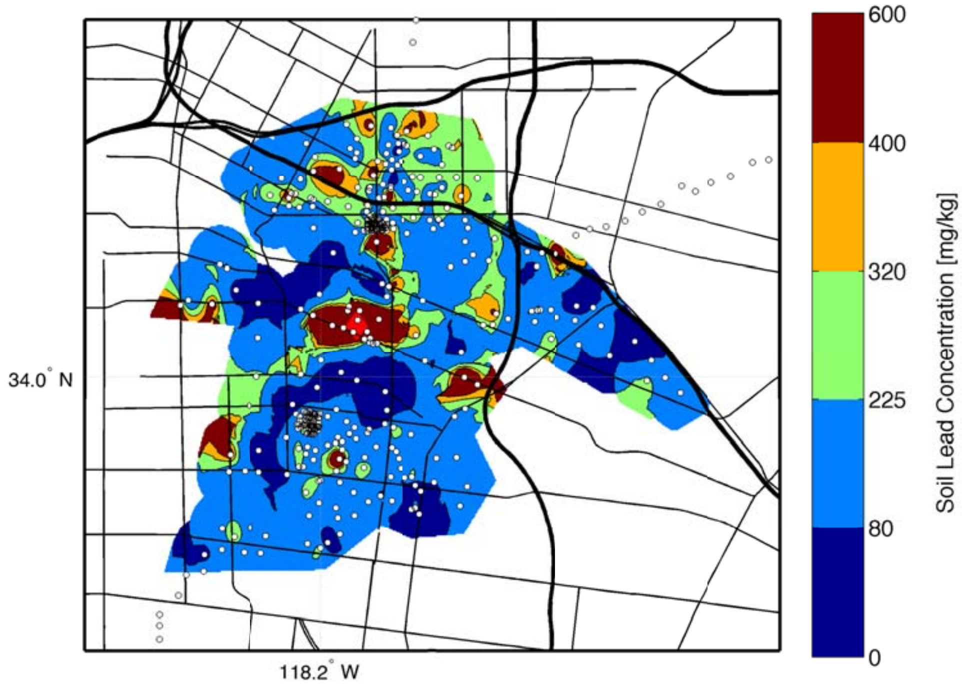


Figure 1: Map of property-average soil lead concentration, interpolated by kriging. Sampling locations are shown as white circles.

The kriging indicates that soil lead concentrations above the industrial soil screening value of 320 mg/kg surround the Exide facility and then quickly drop off to concentrations below that level before reaching the residential areas (represented by the higher density of sampling locations). A band of interpolated soil lead concentrations below 80 mg/kg (dark blue) lies between the facility and the residential area to the south and partially to the north. The soil lead concentrations within the residential areas are higher than the concentrations within these bands and more variable, with localized zones of soils with higher lead concentrations, indicating that sources other than Exide are contributing to soil lead concentrations in the residential area. In Appendix C. we overlay the kriged soil lead contour map with an aerial photograph showing historical industrial sites with lead or lead-associated activities. In some cases the red hot spots in Figure 1 that are displaced from the Exide facility can be associated with these historical sites. In some other cases, the hot spot results from a single soil lead measurement many times higher than all others on the property, suggesting the presence of a very specific local source. While some of these hot spots could be considered outliers and potentially excluded from the analysis, no measurements were excluded from the kriging or from the statistical analyses that follow.

3.2 Regression Analyses of Soil Lead Concentration vs. Distance

As a second level of analysis, we assess the soil lead concentration as a function of distance from the Exide facility using a number of alternative functions, beginning with a piecewise linear regression. Alternative logarithmic and exponential forms are then considered. All three regressions assume that soil lead concentrations decrease monotonically with distance from the facility until they reach a constant level that represents the ambient urban background concentration. The piecewise linear and the logarithmic regressions exhibit a distinct transition to the urban background concentration at a fitted distance from the site. The exponential equation exhibits a smooth transition to urban background concentrations with distance, so it requires some degree of subjective interpretation to determine at what point the transition has effectively occurred. We use these regression analyses to estimate the distance from the Exide facility at which soil lead concentrations reach the ambient urban background level and to estimate the ambient background concentration in the area in and around Vernon, CA. The equations for these regression analyses and additional details are given in Appendix B.2 .

The results, shown in Figure 2 and numerically summarized in Table 2, make it clear that all three analyses give very similar estimates for the distance from the site at which lead concentrations reach a constant value and for the constant value. It is clear from Figure 2 that there is a distinct relationship between lead concentration and distance, with a high level of statistical confidence in the estimated parameters of the fitted regressions (as indicated by the low p-values in Table B-1, Table B-2, and Table B-3). However, the fitted regressions do not explain a high amount of variation in the data. For example, the R^2 value of the piecewise linear regression is 0.25, which means that this regression explains 25% of the variation in the observed soil lead data. The fitted regressions and their coefficients are thus statistically significant at a high degree of confidence and indicative of a clear spatial pattern, but with a large amount of variation around this pattern. The regression analysis in Section 3.3, which uses the modeled air concentration and indicator variables, explains a larger amount of the variation in the data with a comparable degree of confidence.

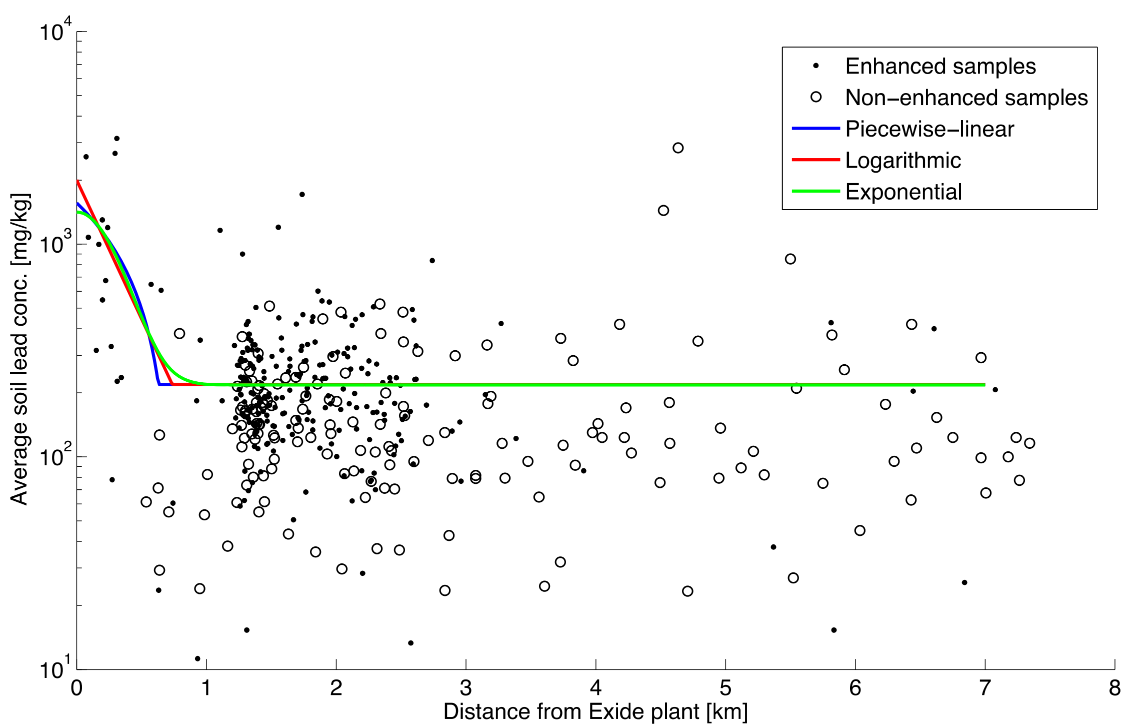


Figure 2: Soil lead concentration as a function of distance from the Exide facility. Three different fitted regression analyses are overlain on the empirical data. All three estimate similar values for the ambient background soil lead concentration and distances at which the concentrations fall to the ambient background level. Open circles represent data from "non-enhanced" sampling locations, which are not close to freeways, arterial roads, or houses built before 1940. Dots represent sampling sites that have one or more of those factors likely to enhance soil lead concentrations

The resulting distances at which the soil lead concentrations are predicted to become constant are very similar across the three considered models. Table 2 compares the results. For the exponential regression, we use the criteria that the predicted soil concentration falls 95% of the difference between the value at zero distance and the asymptotic (background) value. The three regression analyses yield similar estimates for the distance at which the urban background concentration for the area is effectively reached, ranging from 0.63 km for the linear regression to 0.74 km for the logarithmic regression. The result from the linear regression is plotted as a black dashed circle in Figure 3.

Table 2: Estimated distance to reach ambient background concentration and estimated ambient concentrations (standard errors in parentheses).

Fitted Regression	**Distance to reach ambient background**	**Ambient background concentration**
Linear	0.630 km (0.079)	218 mg/kg
Logarithmic	0.737 km (0.11)	219 mg/kg
Exponential	0.681 km[a] (0.11)	218 mg/kg

[a] distance where predicted soil concentration falls 95% from the value at zero distance and the estimated ambient background value

All three analyses also yield very similar estimates for the urban background concentration, shown here to be approximately 218 mg/kg.

In order to better account for prevailing wind directions and possible difference in wind speed along those directions, we split the sampled locations into six subsets by direction (sectors) and fit the piecewise linear regression function to the samples in each subset. Figure 3 shows the level-off distances for each of six radial sectors around the site. For each sector, a best estimate is shown (thin dark blue arc), plus and minus one standard error of the estimate (wider light blue arc). Property locations with sampling are also shown (black dots) to provide spatial perspective. The areas with the high densities of sampling to the north and south of the facility are the Northern and Southern Initial Assessment Areas where soil removal is presently taking place.

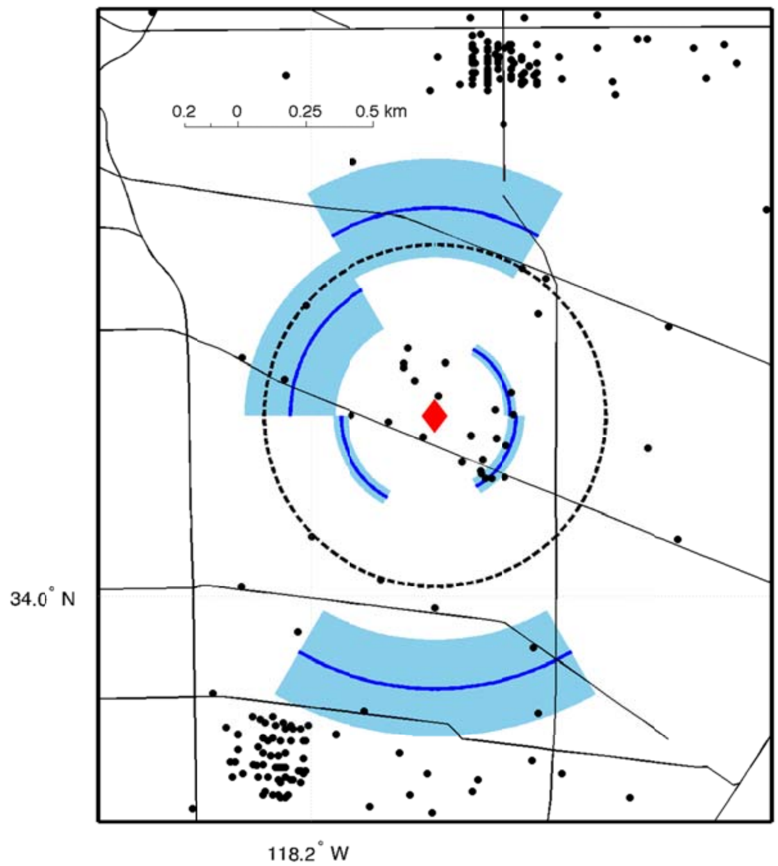


Figure 3: Distances to reach ambient background concentrations using the piecewise linear regression for directional subsets (sectors). Light blue bands represent the ±1 standard error confidence intervals. The clusters of black dots at the top and bottom are the Northern and Southern Initial Assessment Areas.

Table 3: Estimated distance to reach ambient background concentrations and estimated ambient concentrations for directional subsets shown in Figure 3 using the fitted piecewise linear function. The values in parentheses are the standard errors of the indicated estimates.

Sector	Distance to reach ambient background [km]	Ambient background concentration [mg/kg]
All directions	0.63 km (0.07)	218
330º-30º	0.77 km (0.18)	226
30º-90º	0.28 km (0.023)	227
90º-150º	0.30 km (0.031)	241
150º-210º	1.01 km (0.18)	205
210º-270º	0.35 km (0.028)	133
270º-330º	0.54 km (0.17)	217

As indicated in Table 3, the estimated distances needed to level off at the urban background value ranges from approximately 0.26 km (= 0.28 – 0.023) in the 30º - 90º sector to 1.19 km (= 1.01 + 0.18) in the 150º - 210º sector.

3.3 Statistical Analysis of Contributors to Soil Lead Concentrations

The statistical analyses presented above consider only the measured soil lead data. This has the advantage of limiting the number of assumptions that must be made when incorporating other information, especially when this information is uncertain in its completeness and representativeness. However, when such information is available, its use can help to better explain the contributions of the various factors associated with elevated soil lead, including Exide, other industrial facilities, historic leaded paint, and highway or road proximity. In this section, an atmospheric dispersion model for the Exide facility is used to predict ambient air concentrations (a surrogate for deposition), in conjunction with indicator variables for other sources, to fit a least-squares regression equation able to distinguish between the various sources contributing to the observed soil lead profile within and beyond the discernible influence of the site. The regression analysis estimates property-average soil lead concentration according to the following formula:

$$\begin{aligned} \textit{Soil conc.} = {} & b_0 + b_1 * \textit{Amb. air conc.} + b_2 * \textit{Freeway} + b_3 * \textit{Arterial road} \\ & + b_4 * \textit{Built before 1940} + b_5 * \textit{Drip zone} + \textit{Random error} \end{aligned} \quad (1)$$

Where b_0 is the intercept and $b_1 - b_5$ are the fitted regression coefficients for the following explanatory variables:

- **Ambient air concentration** of lead modeled using an atmospheric dispersion model described in Appendix B.3.2 .
- **Freeway**: a 0/1 variable that indicates whether a property is less than 200 ft from a freeway, described in Appendix B.3.3 .
- **Arterial road**: a 0/1 variable that indicates whether a property is less than 200 ft from an arterial road, described in Appendix B.3.3 .
- **Built before 1940**: a 0/1variable that indicates for whether a house was built before 1940, when lead paint was still used.
- **Drip zone**: a 0/1 variable that indicates whether samples were collected in a "drip zone" near the roof or downspout of a house, described in Appendix B.3.5 .

We tested four other explanatory variables but did not include them in the analysis. The variable for proximity to historic lead manufacturing industries is statistically significant, but we exclude it because it reduces the amount of variability explained by the model (R^2). The other variables are not statistically significant. These unused explanatory variables are:

- Proximity to historic lead manufacturing industries ($p = 0.022$)
- Proximity to historic paint manufacturing industries ($p = 0.26$)
- Proximity to historic metal working industries ($p = 0.79$)
- Ambient air concentration/Drip zone interaction term ($p = 0.80$)

We fit the regression analysis given in (1) to property-averaged soil lead concentrations from 0 – 6 inch soil depth (see Appendix A.1.1) using the explanatory variables included in (1). The equation is fitted using the "fitlm" function in the MATLAB Statistics Toolbox (Mathworks 2014), with the robust-fitting option that reduces the sensitivity to outliers. The resulting fitted parameter values are summarized in Table 4.

Table 4: Fitted coefficients of regression equation (1) for soil lead concentration.

		Estimate	Std. Err.	*t*-Stat.	*p*-Value
(intercept)		b_0 = +129.2 mg/kg	9.90	13.0	1.32e-34
Amb. air conc.	(*n*=635)	b_1 = +7.724 (mg/kg)/(10^{-3} µg/m^3)	0.42	18.4	5.84e-61
Freeway < 200 ft	(*n*=9)	b_2 = +173.8 mg/kg	50.4	3.45	6.04e-4
Arterial road < 200 ft	(*n*=136)	b_3 = +46.1 mg/kg	14.9	3.10	2.10e-3
Built before 1940	(*n*=336)	b_4 = +66.2 mg/kg	12.4	5.32	1.41e-7
Drip zone	(*n*=228)	b_5 = +70.3 mg/kg	12.9	5.44	7.81e-8

Number of observations: 635, Error degrees of freedom: 629
Root Mean Squared Error: 150 mg/kg
R-squared: 0.468, Adjusted R-Squared 0.463
F-statistic vs. constant model: 110, p-value = 1.08e-83

The fitted intercept of the regression analysis in Table 4 (b_0 = 129.2 +/- 9.9 mg/kg) provides an estimate for the average urban background concentration exclusive of factors that further enhance the soil lead levels. The fitted coefficients for each indicator variable (*Freeway*, *Arterial road*, *Built before 1940*, *Drip zone*) provide an estimate of the average additional soil lead concentration attributable to each. The average contribution from historic Exide lead deposition at each location is derived from the product of the fitted Exide slope coefficient (b_1) and the associated modeled Exide ambient air lead concentration (*Amb. Air conc.*). As such, the average soil lead contribution from historic Exide emissions for each specific location in the study area is estimated as the second term on the right hand side of equation (1).

As described in Appendix B.3.2 , the resulting estimate of the Exide contribution to soil lead does not depend on the emission rate used for the Exide facility in the atmospheric dispersion model, but only on the relative shape of the predicted spatial profile of the ambient air lead concentration. The estimated slope coefficient (b_1), and therefore the resulting Exide contribution to soil lead, are scaled by the regression analysis to match the observed soil lead data with minimal (sum of squares) error.

Given estimates for each of the Exide and non-Exide contributions to the observed soil lead concentrations in the study area, several decision rules may be considered for determining at what level (and therefore, in which geographic area) the estimated Exide contributions are high enough to make a statistically distinguishable contribution to the observed values. Two alternative approaches for determining the uncertainty of the ambient background concentration are described in Appendix B.3.6 . They lead to similar (though not identical) geographic areas for potential attribution of observed soil lead concentrations to the historic emissions from the facility, beyond the estimated (though highly uncertain) ambient urban background values of soil lead in the study area. Regardless of how the background uncertainty is calculated, we set the threshold for a statistically-distinguishable contribution beyond ambient background as twice the standard error of the background uncertainty. The contribution at or above 2σ (2 x standard deviation) would be sufficient to yield a total concentration (unenhanced background + Exide) that, with high confidence, would be distinguishable from the unenhanced ambient urban background concentration alone (i.e., an unenhanced ambient urban background distribution with zero Exide contribution). A 2σ soil lead contribution should thus allow us to reject the null hypothesis that this increment above the unenhanced urban ambient background occurred by random chance, in favor of the alternative hypothesis that it is attributable to the Exide soil lead contribution.

As described in Appendix B.3.6 , the method chosen to characterize the uncertainty in the background soil lead concentration (and determine σ) is based on the field and analytical error of soil lead concentration measurements. The value of σ is inferred from the duplicate sampling results collected as part of the QA/QC plan for the monitoring program. (Aelion et al. 2014) A lognormal (multiplicative) measurement error is assumed, so that the measurement error is characterized by a coefficient of variation that is then multiplied by the concentration at which it is assumed that measurements are to be taken (in this case, at the estimated non-enhanced ambient urban background soil lead concentration, 129.2 mg/kg) to obtain σ. An alternative approach was also considered based on the standard error of the estimate of the background concentration from the regression results in Table 4. However, this estimate is subject to change based on the magnitude of the sampling program – more samples lead to a smaller standard error in the regression model. The duplicate sample – measurement error approach is not subject to this artifact and was therefore chosen. As indicated in Table B-5, the measurement error approach yields an estimate of σ = 17.57 mg/kg (2σ = 35.1 mg/kg).

The geographic area where the Exide facility makes a statistically-distinguishable contribution to soil lead concentrations, shown in Figure 4, is calculated by finding the modeled ambient air concentration that yields a mean soil lead contribution equal to the 2σ measurement error, estimated to be 35.1 mg/kg. Given that a modeled ambient air concentration of 1 x 10^{-3} $\mu g/m^3$ corresponds to 7.724 mg/kg of soil concentration (b_1 from Table 4), a soil lead concentration of 35.14 mg/kg corresponds to a modeled ambient air concentration of (35.14/7.724 * 10^{-3}) = 4.55 x 10^{-3} $\mu g/m^3$, as illustrated in Figure B-4. The contour for that modeled ambient air concentration is plotted in green in Figure 4. Contours calculated using the modeled ambient air concentration coefficient plus or minus 1 standard error (7.724 $\pm$ 0.042) are also plotted in red and blue, respectively, to show the uncertainty. The distances of the outer contours from the Exide facility range from 0.3 km in the west-northwest to 1.3 km in the southwest.

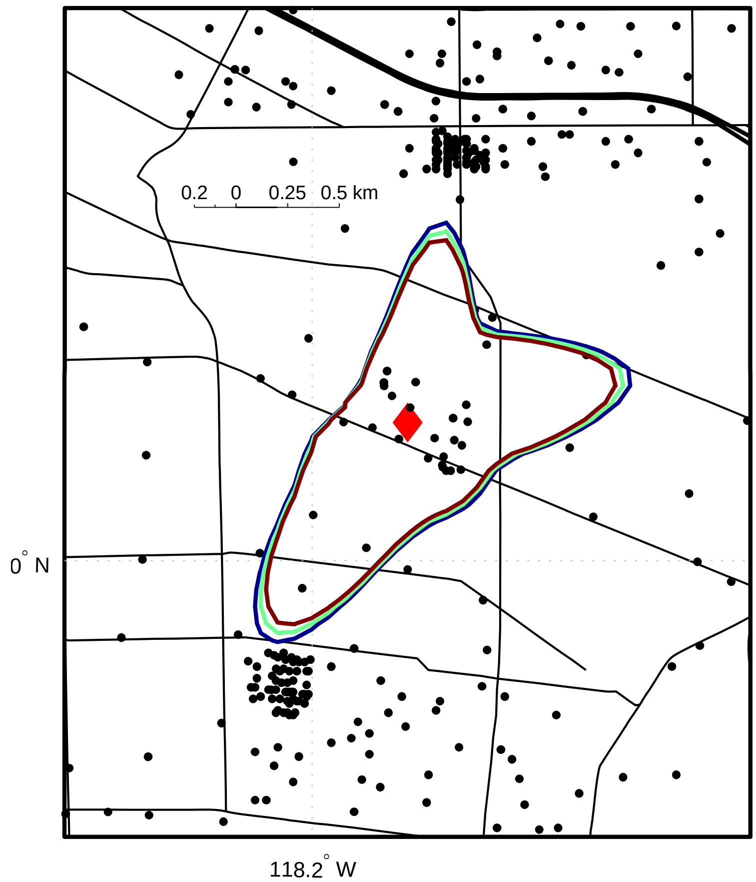


Figure 4: Contours of locations beyond which Exide's estimated contributions to soil lead concentrations are no longer statistically distinguishable from background concentrations based on two times the computed standard deviation of the soil lead measurement error at the estimated background concentration (=2*0.136*129.2 = 35.14 mg/kg; see Table 9). This results in a critical modeled ambient air concentration of 4.55 x 10^{-3} μg/m^3 from the Exide facility, plotted as the green line. Uncertainty in the fitted coefficient for modeled ambient concentrations yields the red and blue lines.

4 Discussion

4.1 Comparison to Ambient Urban Background from Other Studies

We consider three sets of soil lead measurements that can be used to estimate the ambient urban background soil lead concentrations for comparison with the assessment areas: the expanded assessment areas (EAA) adjacent to Vernon shown in Figure D-1, a residential area of Long Beach, CA near the I-405 freeway (shown in Figure D-3) where Exide collected samples as part of the initial sampling effort, and samples collected as part of a U.C. Irvine study (Wu et al. 2010) from sites scattered around the Los Angeles basin, shown in Figure D-2.

Each of these data sets has advantages and disadvantages. The EAAs are the most similar to the Initial Assessment Areas (IAAs) and thus meet the selection criteria for a background area in the Guidance. The Long Beach sampling areas are not affected by the Exide facility and have some characteristics similar to the assessment areas, but they contain significantly fewer houses likely to have lead paint based on the age of house construction. The U.C. Irvine study estimates average soil lead concentrations for residential and industrial properties across the entire Los Angeles basin, but their dataset contains very few samples in and around Vernon.

4.1.1 Expanded Assessment Areas (EAAs)

The EAAs are very similar to the IAAs in terms of characteristics that affect soil lead concentrations, as shown in Figure 5. Approximately equal fractions of houses in both areas were built before 1940, and the distributions of distances to freeways and arterial roads are very similar. These similarities make the EAAs well-suited for comparison with the IAAs as representative of ambient urban background conditions in the vicinity.

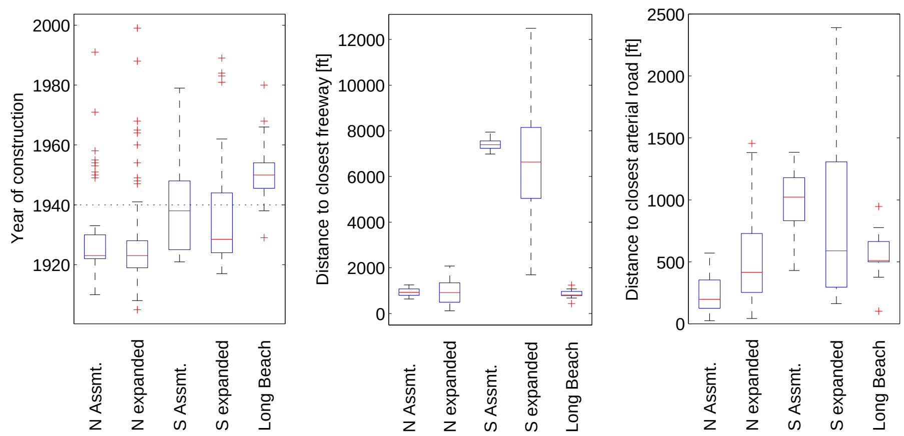


Figure 5: Boxplots comparing year of construction of homes (left), distance to the closest freeway (center), and distance to the closest arterial road (right) for each property.

Figure 6 plots the distributions of property-averaged soil lead concentrations for the IAAs, EAAs, and the Long Beach area (described in Section 4.1.2). Distributions for all properties are on the top and distributions for only properties built after 1940 are on the bottom. Vertical lines labeled "A" and "B" show the ambient urban background concentrations estimated in Sections 3.2 and 3.3, respectively. This figure shows that the distributions of property-averaged soil lead concentrations in the IAAs are nearly indistinguishable from the distributions of the corresponding EAAs, though the difference between the Northern IAA and the Northern EAA is larger for properties built after 1940.

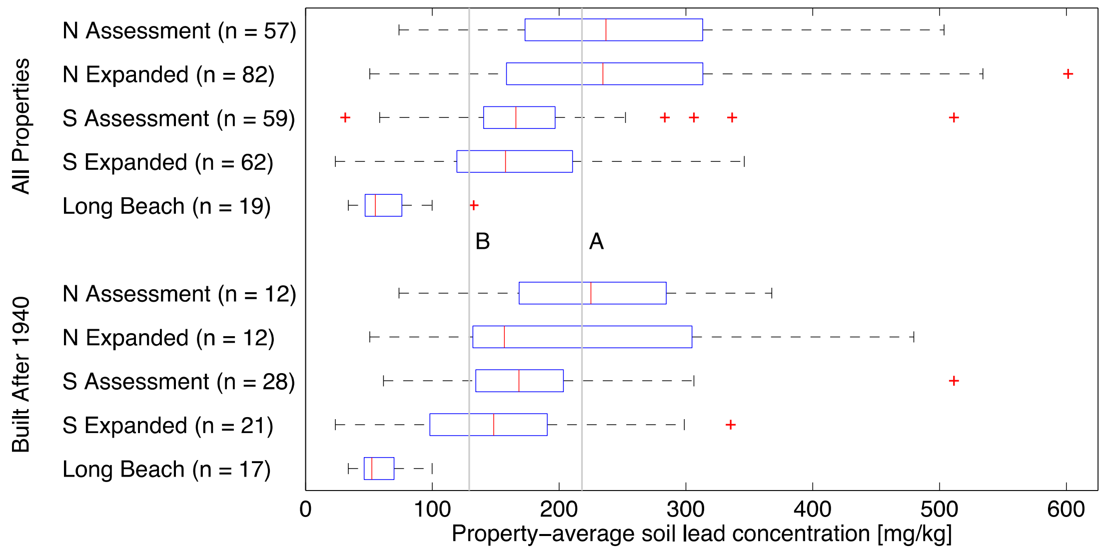


Figure 6: Boxplots comparing distributions of property-averaged soil lead concentrations in different areas. The top group includes all sampled properties; the bottom group includes only properties built after 1940. The vertical lines "A" and "B" correspond to the ambient urban background concentrations estimated in Sections 3.2 and 3.3, respectively.

The similar distributions of soil lead concentrations in IAAs and the corresponding EAAs suggest that emissions from the Exide facility have a small effect on soil lead concentrations in the IAAs. If the effects of the Exide facility were significant, we would expect lower average concentrations in the EAAs, which are farther from the facility. We refine this argument further by calculating the 95% upper confidence level of average soil lead concentrations for smaller groups of sampled properties at various distances from the Exide facility, shown in Figure 7. This figure, like the kriging map in Figure 1, shows that concentrations do not decrease monotonically with distance from the Exide facility and that the average concentrations in the Northern and Southern Assessment Areas are similar or lower than average concentrations farther from the facility. This is consistent with the models of soil concentration as a function of distance from the facility in Figure 2 and Figure 3, which show soil concentrations reaching ambient urban background levels less than 1 km from the facility. For comparison, the nearest residential properties in the IAAs and EAAs are approximately 1.3 km from the facility.

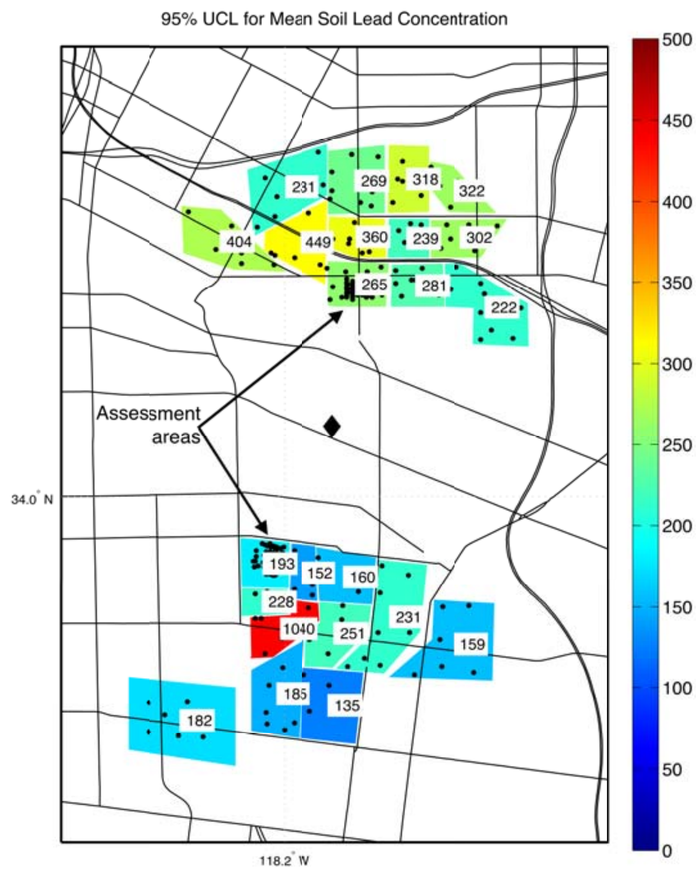


Figure 7: 95% upper confidence level for mean soil lead concentrations for small groups of sampled properties in the Initial Assessment Areas and Expanded Assessment Areas. Each group is labeled with the 95th UCL for the group mean.

4.1.2 Long Beach

The sampled properties in Long Beach are within approximately the same distance of a freeway as the northern assessment area properties are, as shown in Figure 5 (center). However, Figure 5 (left) shows that most of the Long Beach houses were built after 1940 (when lead paint was effectively phased out), but most houses in the assessment areas were built before 1940. Even if properties built before 1940 are excluded (as shown in the bottom grouping in Figure 6), there is still a significant difference between soil lead concentrations in the Long Beach sampled area and the assessment areas. This suggests there are differences between the Long Beach sampling area and the assessment areas we do not account for. A possible explanation is that the prevailing wind in Long Beach comes from the southwest (from offshore), so it travels less distance over highly urbanized and industrialized areas, and would blow emissions from the I-405 freeway *away* from the Long Beach sampling area. Also, the time period of high traffic density and high industrial activity is believed to extend further back into the first half of the 20th Century in the Vernon area compared to Long Beach, allowing the former to experience higher historic exposure to leaded gasoline, leaded paint, and lead-based industries.

4.1.3 U.C. Irvine Study (Wu, et al.)

The Wu et al. study samples soil lead concentrations from 144 residential properties in the greater Los Angeles area. (Wu et al. 2010) Collectively, the characteristics of those sampled sites are similar to the sites sampled in this Exide background study, though very few of the Wu, et al. samples are collected near Vernon, CA. Significant fractions of the Wu samples are collected within 300 meters of freeways (66) or arterial roads (35); our study collects 71 samples from within 300 m of freeways and 184 within 300 m of arterial roads (see B.3.3). The average year of construction of the nearest residential properties in the Wu study is 1945, which is similar to the Southern Assessment Area and the Southern Expanded Area in our study (see Figure 5).

We calculate consistently higher median soil lead concentrations in this study than the Wu et al. study, as shown in Table 5. Where necessary, we convert total lead concentrations to bioavailable lead concentrations (or vice versa) using the relationship fitted by Wu et al.:

$$bio\_Pb = -1.2 + 0.74\ total\_Pb$$

Table 5: Comparison of median soil lead concentrations.

	Total		**Bioavailable**	
	this study	Wu et al.	*this study*	Wu et al.
Freeway < 300m (n = –/32)	–[1]	153 mg/kg	–[1]	112 mg/kg
Arterial road < 300m (n = 50/66)	194 mg/kg	134 mg/kg	143 mg/kg	98 mg/kg
Other (n = 34/44)	157 mg/kg	56 mg/kg	115 mg/kg	40 mg/kg

[1] We cannot calculate median concentrations for sites near a freeway because there are no sites in this Exide study within 300 m of a freeway but *not* within 300 m of an arterial road.

These differences are likely statistically significant, in spite of the small sample sizes, because the magnitudes of the differences are large. We hypothesize that the medians in our study for the IAAs are higher than the medians calculated by Wu et al. because of the greater concentration of industry in Vernon and the predominant wind direction noted above. The median soil lead concentration of the Long Beach samples (see Figure 6) is similar to that for the non-freeway, non-arterial samples ("Other"), and those samples that are not consistently close to industrial areas.

5 Findings and Conclusions

5.1 Findings

Average soil lead concentrations tend to decrease with distance from the Exide facility and then, beyond a certain distance, do not decrease any farther. Although there are large variations in soil concentrations, even between adjacent properties, there are consistent patterns. We perform two different types of regression analysis in order to estimate where soil lead concentrations stop decreasing and "level off" at the ambient urban background concentrations. One group of regression equations, which relate soil concentration to distance from the Exide facility, estimate that soil lead concentrations level off between 0.28 and 1.01 km from the facility, depending on direction (Table 3). The other regression equation, which relates soil concentration to modeled ambient air concentrations of lead emitted by the Exide facility, estimates that the soil lead concentrations are statistically indistinguishable from the ambient urban background between 0.3 and 1.3 km from the Exide facility, depending on direction (Figure 4). For comparison, the nearest houses in the IAAs are approximately 1.3 km from the Exide facility.

We use the same regression models described above to estimate the ambient urban background concentration of soil lead in the study area. Using the fitted soil concentration vs. distance equations, we estimate the ambient urban background concentration as 218 mg/kg (Table 2). Using the soil concentration vs. modeled ambient air concentration regression, we estimate the ambient urban background as 129 mg/kg for a house built after 1940 that is not near a freeway or arterial road, 195 mg/kg for a more typical house built before 1940 that is not near a freeway or arterial road, and 242 mg/kg for a house built before 1940 that is near an arterial road (Table 4).

The ambient urban background concentrations of soil lead we estimate are higher than those estimated from two other sets of soil samples in the Los Angeles area. Samples from an area of Long Beach, CA indicate a median soil lead concentration of approximately 50 mg/kg (see Figure 6). However, the sampled Long Beach properties are upwind of a freeway and mostly built after 1940, whereas most of the sampled study area properties are downwind of one or more freeways and built before 1940 (when lead paint was effectively phased out). Wu et al. used samples collected from around the Los Angeles area to estimate a median soil lead concentration of 56 mg/kg for residential properties, whereas we estimated 157 mg/kg for the residential properties in the Vernon area (see Table 5). The area sampled in Long Beach is believed to have a shorter history of high industrial activity and traffic density compared to Vernon, and similar differences may also affect the comparisons between the study and the Wu et al. results.

The soil lead concentrations in the initial assessment areas (IAAs) are not statistically-significantly different from concentrations in the nearby expanded assessment areas (EAAs). The EAAs are farther away from the Exide facility, but they have similar characteristics to the IAAs (distributions of building age and distance to freeways and arterial roads). The finding of similar soil lead concentrations in the IAAs and the EAAs, but with both indicating significant sample to sample variability, suggests that the high concentration excursions exhibited by a subset of the samples in both the IAAs and the EAAs are unlikely to be associated with a single, broadly distributed source, but instead are due to the effects of multiple very localized sources. This finding is consistent with the inferences from our kriging and statistical analyses that indicate that the Exide impact is confined to an area within 1.3 km of the site, equal to or less than the distances to the IAAs.

5.2 Conclusions

This study analyzed soil samples collected from 244 residential properties and 163 non-residential sites in and around the Exide facility in Vernon, California. We applied three analysis methods to these data to estimate the contributions of the Exide and non-Exide factors to observed soil lead concentrations. The first analysis used kriging to smoothly interpolate soil lead concentrations in order to graphically show larger spatial patterns. The second analysis fitted a piecewise regression equation to estimate soil lead concentration as a function of distance from the Exide facility. The third analysis fitted a regression equation to estimate soil lead concentration as a function of modeled ambient air concentration of lead dust from the Exide facility and site characteristics such as the distance to the nearest freeway or arterial road, and the year of housing construction (a proxy for lead paint).

The two regression analyses give several results that are consistent with each other and with the kriging analysis. First, average soil lead concentrations tend to decrease with distance from the Exide facility and then, beyond a certain distance, level off and no longer continue to decrease. This distance represents a threshold at which the contribution from the Exide facility becomes statistically indistinguishable from the ambient urban background concentration. We estimate this distance to be as large as 1.3 km in the prevailing wind direction and as small as 0.3 km for some of the other directions. For comparison, the nearest houses in the Initial Assessment Areas (IAAs) are approximately 1.3 km from the Exide facility. In a separate analysis, we find that soil lead concentrations in the IAAs are statistically indistinguishable from concentrations in the Expanded Assessment Areas (EAAs). Taken together, these findings suggest that the Exide contribution to soil lead concentrations in these areas is not statistically distinguishable from the ambient urban background.

Second, we estimate the average soil lead concentration beyond the threshold distance, which we describe as the "ambient urban background" level, to range from 129 mg/kg for a sampling site not near old houses, freeways, and arterial roads, to 242 mg/kg for a sampling site near a house built before 1940 and near an arterial road. These average background urban concentrations are somewhat higher than those estimated in other studies of the greater Los Angeles area and for a section of Long Beach, California. However, this is consistent with differences in the predominant wind directions relative to freeways and other high traffic corridors in the Vernon vs. Long Beach sampling areas, and the longer and more concentrated history of high traffic density and high industrial activity in the Vernon area compared to most portions of the other study areas, including a particular history of lead processing industrial plants in the vicinity of, but unrelated to, the Exide site.

6 References

Aelion, C.M. et al., 2014. Temporal and spatial variation in residential soil metal concentrations: Implications for exposure assessments. *Environmental Pollution*, 185, pp.365–368.

Hafen, M.R. & Brinkmann, R., 1996. Analysis of lead in soils adjacent to an interstate highway in Tampa, Florida. *Environmental Geochemistry and Health*, 18(4), pp.171–179.

Johnson, D.E. et al., 1976. *Baseline levels of platinum and palladium in human tissue*, Research Triangle Park, NC: U.S. Environmental Protection Agency.

Lagerwer, J.V. & Specht, A.W., 1970. Contamination of Roadside Soil and Vegetation with Cadmium, Nickel, Lead, and Zinc. *Environmental Science and Technology*, 4(7), pp.583–586.

Lansdown, R. et al., 1986. The relationship between blood-lead concentrations, intelligence, attainment and behaviour in a school population: the second London study. *International Archives of Occupational and Environmental Health*, 57(3), pp.225–235.

Mathworks, 2014. MATLAB and Statistic Toolbox.

Motto, H.L. et al., 1970. Lead in soils and plants: its relation to traffic volume and proximity to highways. *Environmental Science and Technology*, 4(3), pp.231–237.

National Institute of Standards and Technology, 2015. NIST/SEMATECH e-Handbook of Statistical Methods. *itl.nist.gov*. Available at: http://www.itl.nist.gov/div898/handbook/eda/section3/eda3669.htm [Accessed May 26, 2015].

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30, pp.683–691.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. Available at: http://www.R-project.org/.

Rodríguez-Flores, M. & Rodríguez-Castellón, E., 1982. Lead and cadmium levels in soil and plants near highways and their correlation with traffic density. *Environmental Pollution Series B, Chemical and Physical*, 4(4), pp.281–290.

Small, M.J. et al., 1995. Source Attribution of Elevated Residential Soil Lead Near a Battery Recycling Site. *Environmental Science and Technology*, 29(4), pp.883–895.

Wu, J. et al., 2010. Spatial analysis of bioavailable soil lead concentrations in Los Angeles, California. *Environmental Research*, 110(4), pp.309–317.

Zupančič, N., 1999. Lead contamination in the roadside soils of Slovenia. *Environmental Geochemistry and Health*, 21(1), pp.37–50.

Appendices

Appendix A. Soil Lead Sampling and Analysis

A.1 Soil Samples

We analyze lead concentrations from soil samples collected at residential properties and public areas near Vernon, CA, and from residential properties in northern Long Beach, CA. The residential samples were collected by Advanced GeoServices at depths of 0-1 inches, 1-3 inches, 3-6 inches, 6-12 inches, and 12-18 inches according to procedures described in the "Revised Addendum To The November 15, 2013 Work Plan For Off-Site Soil Sampling Exide Technologies Vernon, California" (Paul Stratman, July 26, 2014). These residential samples are grouped into two Initial Assessment Areas (North and South) and two corresponding Expanded Assessment Areas (North and South), shown in Figure D-1, as well as a potential background area in Long Beach, shown in Figure D-3. The public-area samples (called the "Step-Out Samples") were collected by ENVIRON at depths of 0-1 inches, 1-3 inches, and 3-6 inches according to procedures described in the "Step-out Dust and Soil Sampling Report" (Russell Kemp, June 2014).

Approximately 60 samples were collected from each residential property and 3 samples were collected from each step-out location. In order to reduce the sample-to-sample variability and provide a more representative measure of exposure, we calculate the mean lead concentration at each property or sampling site using all samples from 0 – 6 inch depths. Section A.1.1 explains why we chose the 0 – 6 inch range. Table A-1**Error! Reference source not found.** gives summary statistics for the property-averaged soil lead concentrations.

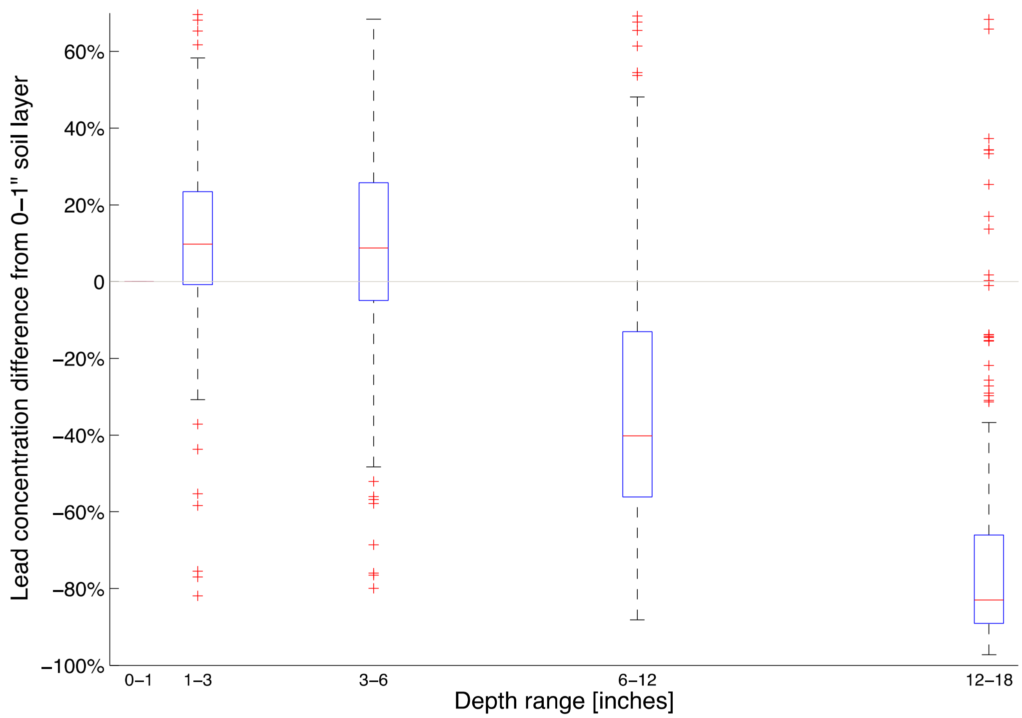
Table A-1: Summary statistics for soil lead concentrations (mg/kg) averaged over the 0-6 inch depth interval for each property/sampling site. These summary statistics exclude duplicate samples and samples from drip zones

	N	Min	Max	Mean	Median	Std. Dev.
Overall	423	11.2	3140	231.5	170.0	295.6
Northern IAA	52	73.7	898.5	257.6	237.0	126.9
Northern EAA	78	50.6	601.3	249.4	234.6	119.4
Southern IAA	58	31.4	511.3	172.9	165.9	72.0
Southern EAA	58	23.5	1711.6	198.4	157.6	215.5
Long Beach	19	33.7	132.6	132.6	63.5	25.4
Step Out	158	11.2	3140	267.9	123.3	443.9

A.1.1 Selection of included depth ranges

The analyses described in Appendix B. are based on the means of soil samples from depths between 0 and 6 inches at each property. We average only these samples from the top 6 inches of soil because the lead concentrations are highest in that range. This is consistent with the analysis plan described in the "Background Study Work Plan"; it was assumed that the 0 – 6 inch range was most likely to be affected by surface exposure and near-surface mixing.

Figure A-1 shows the distributions of lead concentrations in the various depth ranges, relative to concentrations in the top (0 – 1 inch) layer. For the median property, this figure shows that the mean lead concentration in samples from the 1 - 3 inch range and 3 – 6 inch range are approximately 10% higher than the mean concentration in the 0 – 1 inch range. Similarly, the mean concentration in the 6 – 12 inch range for the median property is approximately 45% lower than the 0 - 1 inch range and the 12 – 18 inch range is 85% lower than the 0 – 1 inch range. These results**Error! Reference source not found.** in Figure A-1 are based on mean concentrations from each depth range and each property, but the same pattern holds for the median, 10th percentile, and 90th percentile concentrations.



e to

property mean concentration

Appendix B. Statistical Methods and Analysis Procedures

B.1 Kriging Interpolation

We perform the kriging interpolation on $\log_{10}$-transformed property-average soil lead concentrations using the "gstat" package in R (Pebesma 2004; R Core Team 2015). We consider only the measured values within a 500-meter radius in our analysis. A larger radius would give more smoothing, while a smaller radius would reveal patterns that occur over smaller spatial scales (however, these small-scale variations may be due to property-specific factors, rather than the broader spatial pattern associated with urban and industrial source deposition). The weighting varies inversely with distance; i.e. measurements from closer sampling locations are weighted more heavily than more distant ones. The weighting is determined by fitting a model to the *semivariogram* in Figure B-1, which reflects the degree of correlation between sampling locations as a function of the distance between them (as the semivariance goes up, the correlation goes down).

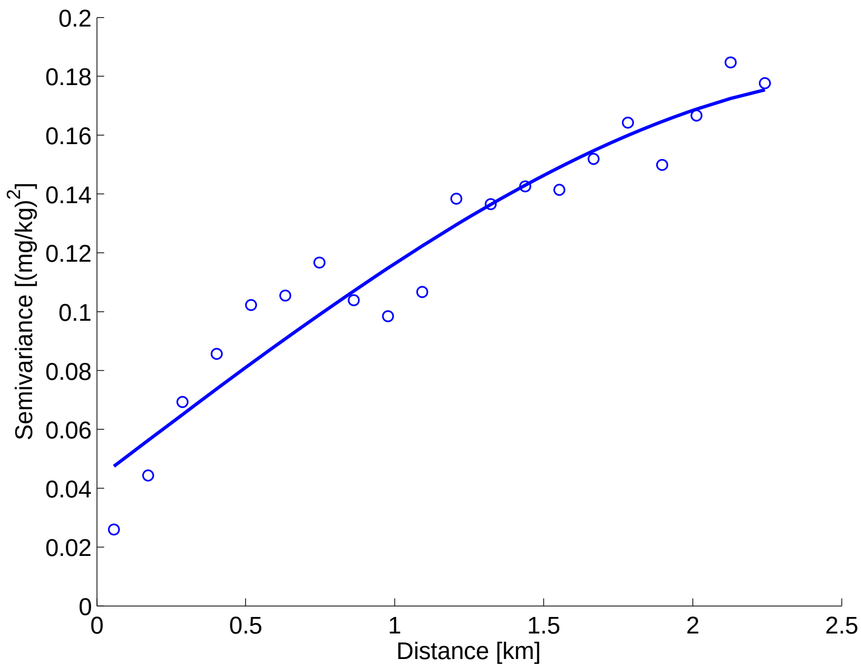


Figure B-1: Empirical semivariogram and fitted spherical model

We calculate the semivariogram only for separation distances less than 2.6 km, even though there are pairs of measurement locations farther apart, because the semivariogram is no longer monotonically increasing (i.e. the correlation is no longer decreasing) beyond that distance. This could result from a variety of factors, including the semi-regular grid of roads and highways in the area. Since measurements only within 500 m are used to calculate each interpolated soil lead concentration, truncating the semivariogram at 2.6 km has very little effect on the subsequent estimation. We fit a spherical model given in Equation (B–1) to the empirical semivariogram, where h is distance, g is the semivariance, a is the range parameter, c is the partial sill, and c_0 is the nugget parameter.

$$g(h) = \begin{cases} c_0 + c\left(1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right) & \text{if } h \leq a \\ c & \text{otherwise} \end{cases} \qquad \text{(B–1)}$$

The spherical model fitted to the $\log_{10}$ property-average soil lead concentrations has parameter values of $a = 0.0242$, $c = 0.1378$, and $c_0 = 0.043$. Figure B-1 shows the fitted model overlaid on the empirical data. A spherical model best fits the empirical semivariogram, though other common models did not significantly change the kriging results.

B.2 Regression Models of Soil Lead Concentration vs. Distance

We fit the models given in Equations (B–2), **Error! Reference source not found.**, and (B–4) to property-averaged lead concentrations from the top 6 inches of soil. Soil samples taken from drip zones are excluded because they are not expected to follow the same concentration vs. distance relationship.

For the piecewise linear model in Equation (B–2), the three fitted parameters of the model represent the soil lead concentration at zero distance from the Exide facility (β_1), the rate of decrease of soil lead with distance (β_2), and the distance at which soil lead concentrations reach a steady ambient urban background level (β_3). The ambient background concentration can be calculated as $\beta_1 + \beta_2 * \beta_3$. The fitted regression coefficients are given in Table B-1; all are statistically significant.

$$soil\ conc. = \beta_1 + \beta_2 * \min(distance, \beta_3) \quad \text{(B–2)}$$

Table B-1: Regression coefficients for piecewise linear model given in Equation (B–2).

	Estimate (std. err.)	t-Statistic	p Value
β_1	1560 (145)	10.8	6.24e-24
β_2	-2130 (445)	-4.79	2.29e-6
β_3	0.630 (0.0787)	8.01	1.20e-14

The logarithmic model given in Equation (B–3) is nearly the same as the linear model in Equation (B–2) but the dependent variable is the base-10 logarithm of soil lead concentration. The three fitted parameters of the model represent the $\log_{10}$ of soil lead concentration at zero distance from the Exide facility (β_1), the rate of decrease of the log of soil lead with distance (β_2), and the distance at which soil lead concentrations reach a steady ambient urban background level (β_3). The ambient background concentration can be calculated as $10^{\wedge}(\beta_1 + \beta_2 * \beta_3)$. The fitted regression coefficients are given in Table B-2; all are statistically significant.

$$\log_{10}(soil\ conc.) = \beta_1 + \beta_2 * \min(distance, \beta_3) \quad \text{(B–3)}$$

Table B-2: Regression coefficients for logarithmic model given in Equation (B–3)

	Estimate (std. err.)	t-Statistic	p Value
β_1	3.30 (0.0582)	56.7	1.59e-194
β_2	-1.31 (0.246)	-5.31	1.79e-7
β_3	0.737 (0.105)	7.01	1.02e-11

The exponential model given in Equation (B–4) models soil lead concentration as an exponential function of distance from the site. The three fitted parameters of the model represent the ambient background soil lead concentration (β_1), the difference between the soil concentration at zero distance and the ambient background concentration (β_2), and the rate at which concentrations decrease with distance from the site (β_3). The fitted regression coefficients are given in Table B-3; all are statistically significant.

$$soil\ conc. = \beta_1 + \beta_2 \exp(-\beta_3 * distance^2) \qquad \text{(B–4)}$$

Table B-3: Regression coefficients for exponential model given in Equation (B–4)

	Estimate (std. err.)	t-Statistic	p Value
β_1	218 (14.5)	15.0	6.91e-41
β_2	1200 (152)	7.88	3.09e-14
β_3	6.44 (2.02)	3.19	1.54e-3

B.3 Statistical Model of Contributors to Soil Lead Concentrations

The description of regression analysis in Section 3.3 omits details about procedures and assumptions in order to focus on the results of the analysis. This section documents those details. Section B.3.2 gives details about the atmospheric dispersion model that calculates ambient air concentrations of lead emitted from the Exide facility, which is an important explanatory variable in the regression model in (1). Section B.3.3 discusses how we selected the distance thresholds for sampling sites to be considered near freeways and arterial roads, and the sensitivity of our results to different threshold distances. Section B.3.4 discusses how we selected the distance thresholds for sampling sites to be considered near historic industrial sources of lead, and the sensitivity of our results to different threshold distances. Section B.3.5 explains the rationale for treating soil samples from "drip zones" separately from other samples from residential properties. Finally, Section B.3.6 describes the details of the method we use to calculate the uncertainty of the soil lead concentration where Exide's contribution becomes statistically indistinguishable from the ambient urban background.

B.3.2 Ambient Air Concentrations of Atmospheric Dispersion Model (AERMOD)

The atmospheric concentrations of lead from the Exide facility at surrounding locations was calculated by ENVIRON using the AERMOD model, version 12060, with wind and other meteorological input data for the two-year period 2006-2007 taken from the AQMD Central LA meteorological station. The resulting air concentrations are shown in Figure B-2. The 2-year period is within the range of time periods (1-5 years) recommended for regulatory air quality modeling[1]. A longer period (e.g., 5 years) is recommended when attempting to characterize the upper tail of an ambient air concentration distribution; a shorter period (e.g., 2 years) is believed sufficient when characterizing the average of the concentration profile (and the associated cumulative average deposition).

The ambient air concentration was computed by the AERMOD model without consideration of plume depletion by deposition. This is equivalent to assuming 100% resuspension of the deposited lead to the atmosphere for further transport and dispersion around (and downwind) of the site.

[1] See, for example, http://www.epa.gov/scram001/guidance/clarification/20130308_Met_Data_Clarification.pdf and http://oehha.ca.gov/air/hot_spots/SRP/Chapter%202%20.pdf , Section 2.8.6.2

Since the ambient air concentration is subsequently used in the regression model to compute the Exide contribution to the soil lead concentrations, this assumption is conservative in terms of the resulting spatial extent of the ambient air concentration profile from the site and its associated impact on the observed soil lead concentrations (in essence, we assume that the same lead particle remains where deposited *and* resuspends to contribute further downwind deposition). As the modeled ambient air concentrations provide an upper bound estimate on the extent to which the concentration contours extend away from the site, so too are the subsequent calculated mean soil lead concentrations reflective of a conservative, upper bound estimate of the site influence at increasing distances from the site.

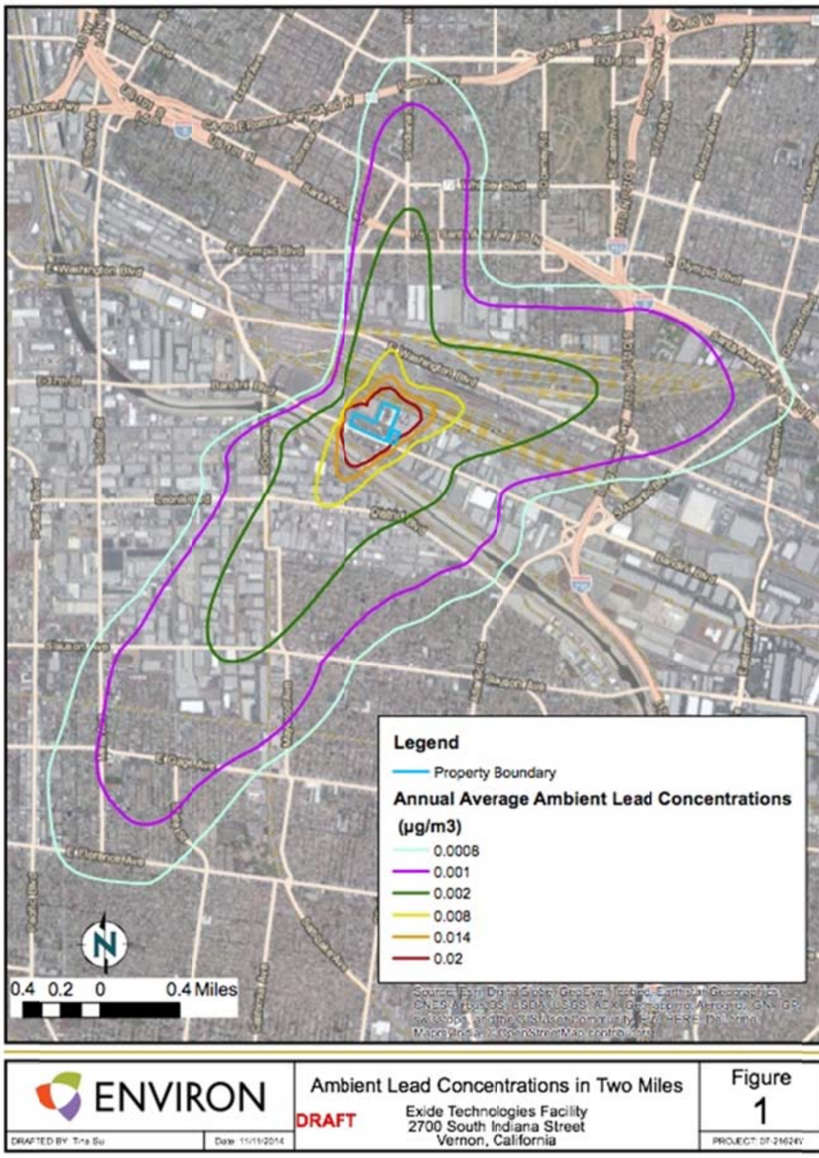


Figure B-2: Modeled ambient air concentrations of lead emitted by the Exide facility.

The AERMOD results were obtained using a constant emission rate from the site of 0.0279 lb/hr, reflective of observed conditions during 2010 and 2012. However, this choice of emission rate has no effect on the subsequent regression model. For the latter, all that matters is the relative shape of the ambient air concentration profile and the extent to which this shape explains a portion of the spatial variation in the observed soil lead data. If a different emission rate were used for the AERMOD simulation (e.g., 100 times larger, 2.79 lb/hr), the computed ambient air concentrations would be (100 times) larger, but the fitted regression coefficient for these ambient air concentrations would be (100 times) smaller. As noted below, it is the *product* of the regression coefficient and the modeled ambient air concentration (at a given location) that determines the Exide contribution to the mean soil lead concentration at that location, and this product is independent of the initial emission rate assumed for the AERMOD simulation.

B.3.3 Sensitivity of Results to Distances Used to Determine Proximity to Freeways and Arterial Roads

We set the threshold for proximity to freeways and arterial roads at 200 ft (61 m) based on previous studies. (Rodríguez-Flores & Rodríguez-Castellón 1982; Hafen & Brinkmann 1996; Motto et al. 1970; Lagerwer & Specht 1970) The Wu et al. study samples soil lead at distances of 0 - 10m, 150 - 160m, and 290 - 300m from freeways and roads, based on previous studies that found most of the lead (by mass) deposits within 150 m of a road. (Wu et al. 2010; Johnson et al. 1976; Lansdown et al. 1986; Zupančič 1999)

Increasing the threshold for distance to be considered close to freeways or arterial roads significantly increases the number of sampling sites that are considered close, as shown in Table B-4. The effect on the regression model is to reduce the estimated ambient urban background (intercept) soil concentration and reduce the estimated effect of being close to either a freeway or arterial road. However, these changes have a negligible effect on the region in which Exide's contribution to soil lead concentrations is distinguishable from background, as shown in Figure B-3. The innermost contour in Figure B-3 is calculated with a threshold of 200 ft (used for the results shown in Figure 4 above) and the outer two contours are calculated with thresholds of 150 m and 300m.

Table B-4: Number of sampling sites close to freeways and arterial roads, for a range of distance thresholds

	Freeway			Arterial Road		
	< 200 ft	< 150 m (492 ft)	< 300 m (984 ft)	< 200 ft	< 150 m (492 ft)	< 300 m (984 ft)
Residential	3	19	71	40	120	184
Residential Dripzone	3	18	64	38	113	172
Non-residential	3	5	19	50	86	120
Total	9	42	154	136	319	476

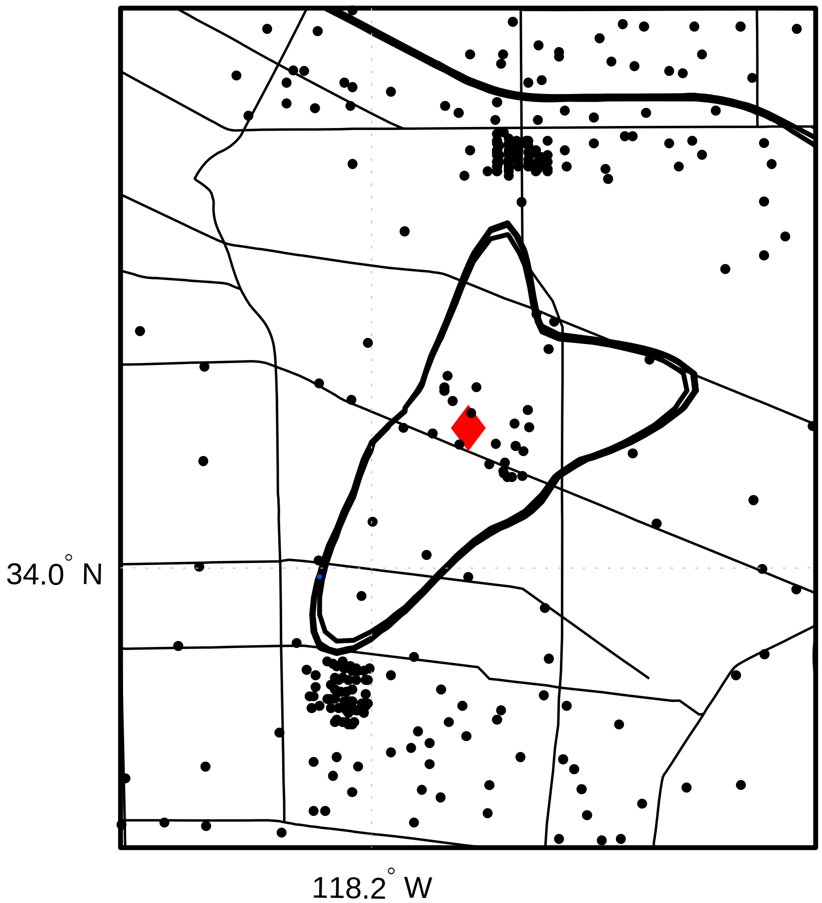


Figure B-3: Contours showing where the Exide contribution to soil lead is distinguishable from ambient urban background concentration, for proximity-to-highway thresholds ranging from 200 ft (innermost contour) to 984 ft (300 m; outermost contour).

B.3.4 Proximity to Other Possible Industrial Lead Sources

We set the threshold for proximity to historic lead industries, paint manufacturers, and non-lead metal processing industries at 1000 ft (305 m) in order to include enough samples to give statistically-significant model coefficients, but also represent a plausible distance for lead transport. A threshold of 1000 ft is approximately the same as the 300 m (984 ft) upper limit used in the Wu et al. study for the effects of highways. (Wu et al. 2010) The indicator variables for proximity to paint manufacturing and non-lead metal facilities are not statistically-significant predictors. The indicator variable for proximity to lead processing facilities is significant at a 2% level. Including it in the model slightly decreases the estimated ambient urban background soil lead concentration (intercept) from 129 to 127 mg/kg and decreases the contribution of modeled ambient airborne lead concentrations to soil lead (slope) from 7724 to 7325 (ug/m^3)/(mg/kg). However, both models (with and without this indicator variable) estimate approximately the same ambient air concentration where the Exide contribution to soil lead is distinguishable from the ambient urban background: 0.00047 ug/m^3 with vs. 0.00046 ug/m^3 without. Increasing the threshold distance increases the number of sampling sites that are considered close to industrial sites. However, the effect on the regression model is to make the indicator variable for proximity to lead-processing industries statistically insignificant.

B.3.5 Drip Zone, Downspout, and Other Areas of Enhanced-Deposition

The "Drip zone" variable in the regression model in (1) statistically controls for the higher soil lead concentrations in samples collected in a drip zone, near a downspout, or other area of enhanced lead deposition. For each property, the lead concentrations of drip zone samples are averaged separately from non-drip zone samples, which means that most of the sampled residential properties have two data points associated with them (see Table 1). The "Drip zone" indicator variable differentiates between those two data points.

Equation (1) models the effect of a drip zone as increasing soil lead concentration by a fixed amount above the concentration of non-drip zone samples at the same site. We test an alternate model, where the difference between drip zone and non-drip zone samples is proportional to the rate of lead deposition at that site by adding an Ambient Air Concentration/Drip Zone interaction term to the model. However, the interaction term is not statistically significant ($p = 0.80$), so we do not include it in the model.

B.3.6 Quantifying Uncertainty of Ambient Urban Background: An Approach Based on Measurement Error Determined from Duplicate Samples

In order to determine where the Exide contribution to soil lead is distinguishable from the ambient urban background concentration calculated by the regression model, we estimate the uncertainty of that background concentration based on measurement errors of duplicate soil samples. (Aelion et al. 2014) The method is described with accompanying spreadsheets at:

A New View of Statistics (Will G Hopkins, 2013): http://www.sportsci.org/resource/stats/index.html

In particular, see the section on Precision of Measurements, Calculations for Reliability: http://www.sportsci.org/resource/stats/relycalc.html

During the soil sampling campaign, duplicates of 210 samples were collected to assess measurement error. Each of these "Field Duplicate" samples was collected at the same location and depth as a regular sample, and analyzed for soil lead concentration separately. We compare the concentrations measured for the original and corresponding Field Duplicate samples to estimate the coefficient of variation (COV) of measurement uncertainty, which is the standard deviation divided by the mean. For a given pair of duplicate measurements $x_{i,1}$ and $x_{i,2}$, we assume each measurement is the product of the "true" value $x_{i,T}$ multiplied by a lognormally-distributed measurement error **r** because the errors are larger for larger measurements:

$$x_{i,1} = x_{i,T}\mathbf{r} \qquad \text{(B–5)}$$

We take the logarithm transform of both sides to get

$$\ln(x_{i,1}) = \ln(x_{i,T}) + \ln(\mathbf{r}) \qquad \text{(B–6)}$$

where the log measurement error ln(**r**) is normally distributed with mean = 0 and variance = σ^2. In this analysis ln( ) refers to the natural logarithm.

Then the difference between the two log-transformed duplicate measurements d_i is:

$$\begin{aligned} d_i &= \ln(x_{i,1}) - \ln(x_{i,2}) \\ &= (\ln(x_{i,T}) + \ln(r_{i,1})) - (\ln(x_{i,T}) + \ln(r_{i,2})) \\ &= \ln(r_{i,1}) - \ln(r_{i,2}) \end{aligned} \quad \text{(B–7)}$$

The log measurement error terms $\ln(r_{i,1})$ and $\ln(r_{i,2})$ are drawn from the same normal distribution, so their difference is drawn from a normal distribution with mean 0 and variance = $2\sigma^2$. This means that the variance we calculate from the difference **d** of the log-transformed measurements is two times the variance of the log-transformed measurement error ln(**r**).

The distribution of the corresponding un-transformed measurement error **r** is lognormal with a scale parameter $\sigma^2 = \text{var}(\mathbf{d})/2 = \text{var}(\ln(x_1) - \ln(x_2))/2$. Finally, we calculate the COV of measurement uncertainty using the following formula for the COV of a lognormal distribution with scale parameter σ^2 (National Institute of Standards and Technology 2015):

$$\text{COV} = \sqrt{\exp(\sigma^2) - 1} = \sqrt{\exp\left(\text{var}(\mathbf{d})/2\right) - 1} \quad \text{(B–8)}$$

For the 210 duplicate samples in the Vernon study, the COV of the duplicate measurement uncertainty is 13.6%. For the urban ambient soil lead concentration of 129.2 mg/kg estimated in Table 4, this gives a two-standard-deviation uncertainty bound of 35.14 mg/kg.

We also investigated another method to estimate the uncertainty of the background concentration calculated by the regression model: the standard error of the regression coefficients. The standard error of regression coefficients is a standard output of a fitted regression model. It is a function of the number of data points used to fit the regression model.

We choose to use the duplicate-measurement error uncertainty estimate because it is insensitive to the number of samples, unlike the standard error measure discussed above. Table B-5 shows a comparison of the measurement uncertainties calculated with these two methods.

Table B-5: Different measures of uncertainty around the estimated ambient urban background concentration of soil lead

	Coeff. of Variation (COV)	2σ uncertainty of ambient background soil conc.
Duplicate measurement uncertainty	0.136	35.1 mg/kg
Standard error of regression coefficients	0.077	19.8 mg/kg

B.3.7 Calculating Modeled Ambient Air Concentration Corresponding to Soil Lead Concentration Where Exide Contribution is Distinguishable from Background

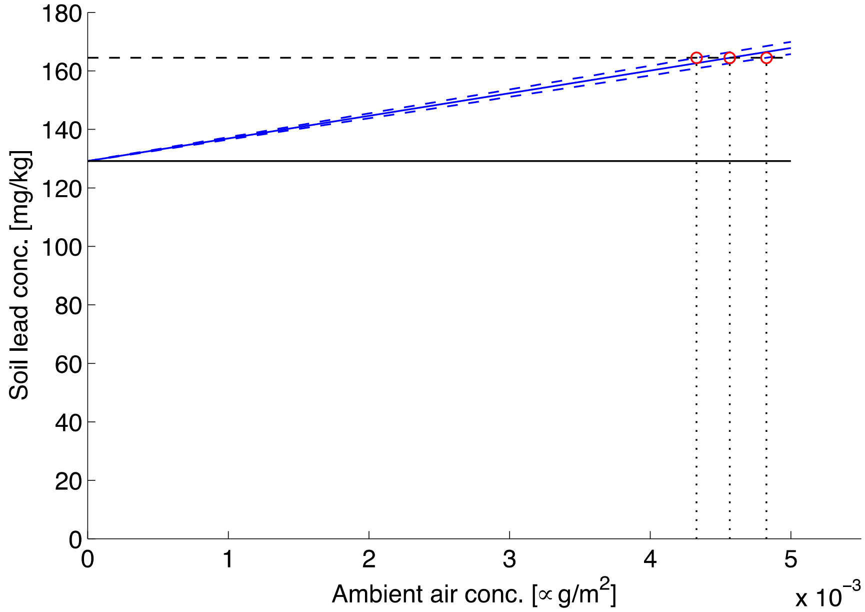


Figure B-4: Estimating the modeled ambient air concentrations of lead from the Exide facility that cause increases in the soil lead concentration that are statistically-distinguishable from uncertainty in the background concentration.

Appendix C. Historic Lead Industries and Hot Spots

In order to understand the variability of the kriged results within the residential area, simplified kriged contours were superimposed on a map that shows the historic industries within the sampling area as shown in Figure C-1. This comparison suggests that some of the isolated high concentration sample results are spatially associated with historic lead industries. There also appears to be higher soil lead associated with the confluence of the I-5, I-710 and CA-60 freeways in the north. These associations suggest that indicator variables of proximity to freeways and other highly trafficked roadways, proximity to historic industries and residential influences such as lead based paint should be examined further in the statistical analyses.

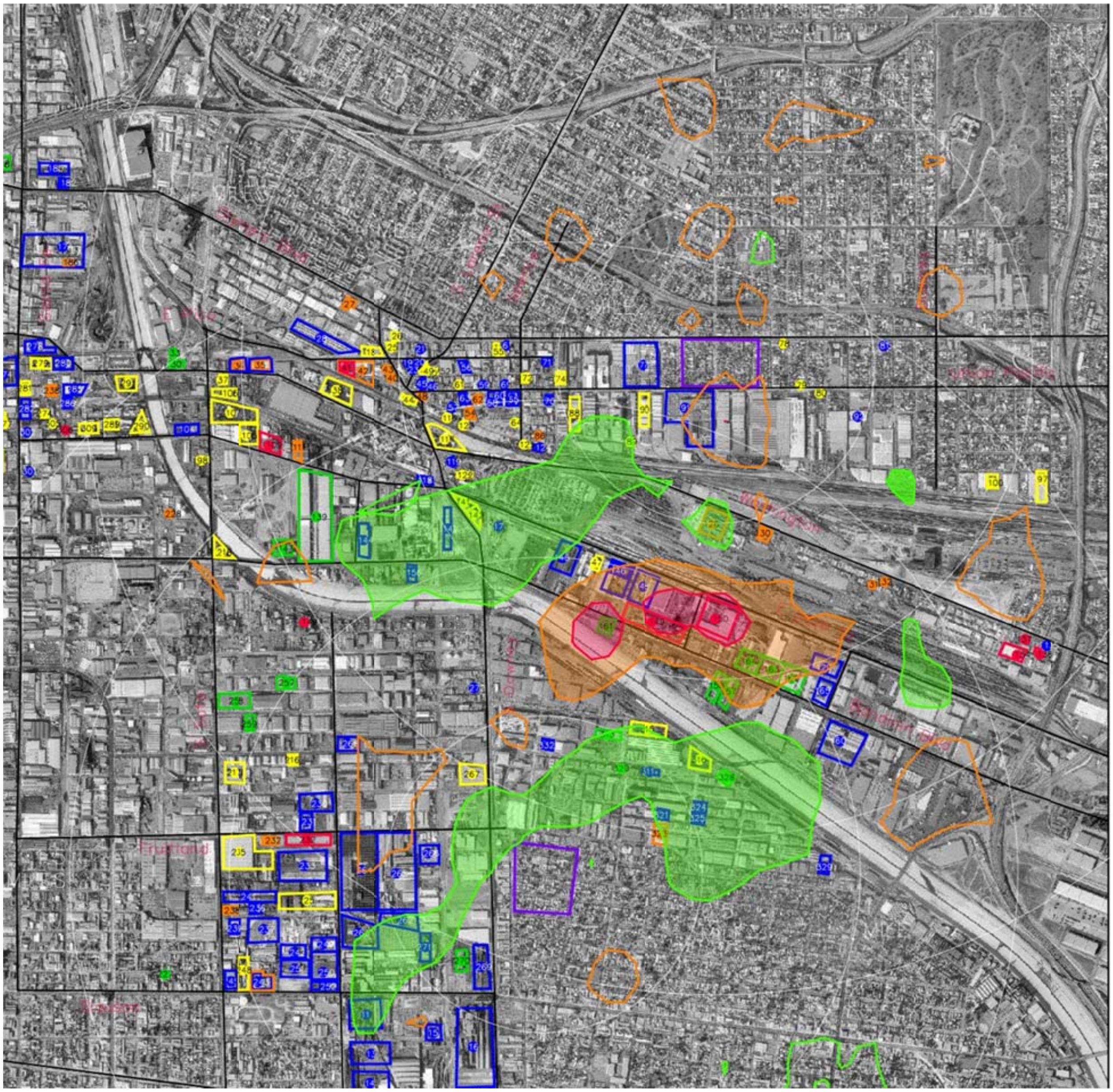


Figure C-1: Kriging contours in relation to locations of historic industrial facilities. Historic lead-processing facilities are outlined in red, metal (foundry/manufacturer) facilities in blue, metal (scrap/warehouse) in yellow, and paint manufacturing facilities in green.

Some of the “hot spots” on the kriging map are not clearly associated with historic industrial sites or freeways, but are associated with outlier soil samples. For example, a single sample at location ESA-36 (shown in Figure C-2) has a measured soil lead concentration of 45,600 mg/kg, which is 27 times higher than the next-highest measurement on that property. Excluding some of these clear outliers would likely improve the fit of the model, but we do not exclude any samples from our analysis.

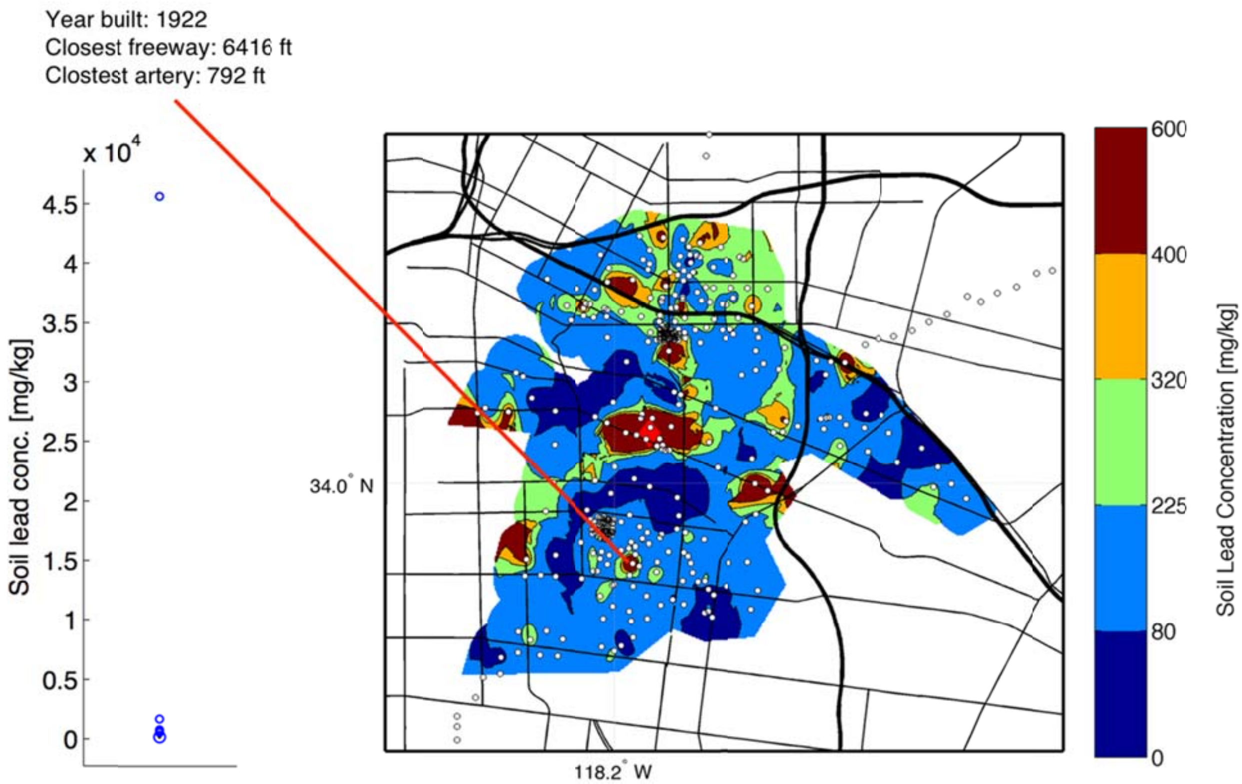


Figure C-2: Details of property ESA-36, which is at the center of a "hot spot". The boxplot on the left shows the distribution of soil lead concentrations sampled from that property.

Appendix D. Areas Used to Estimate Background for Comparison

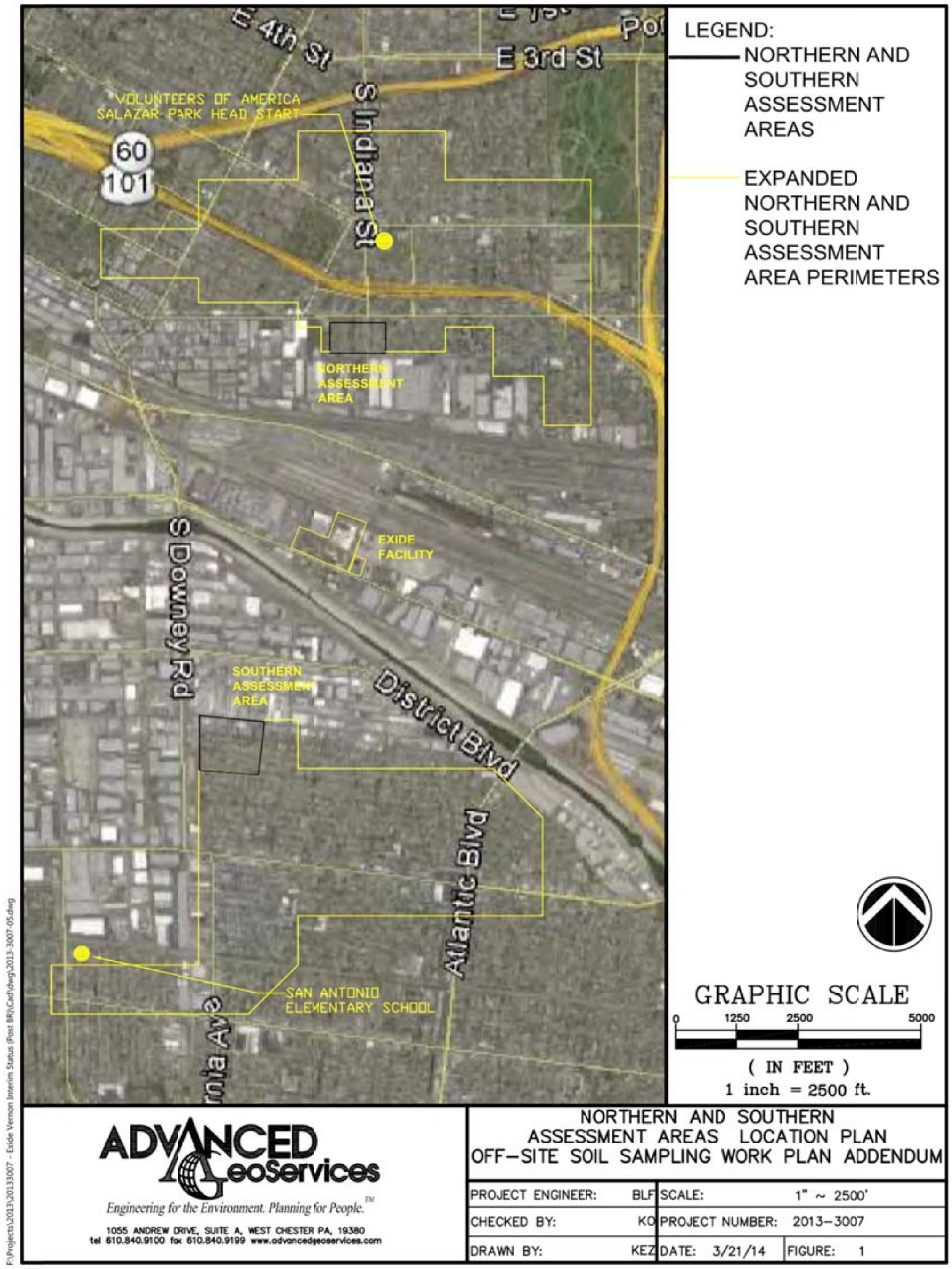
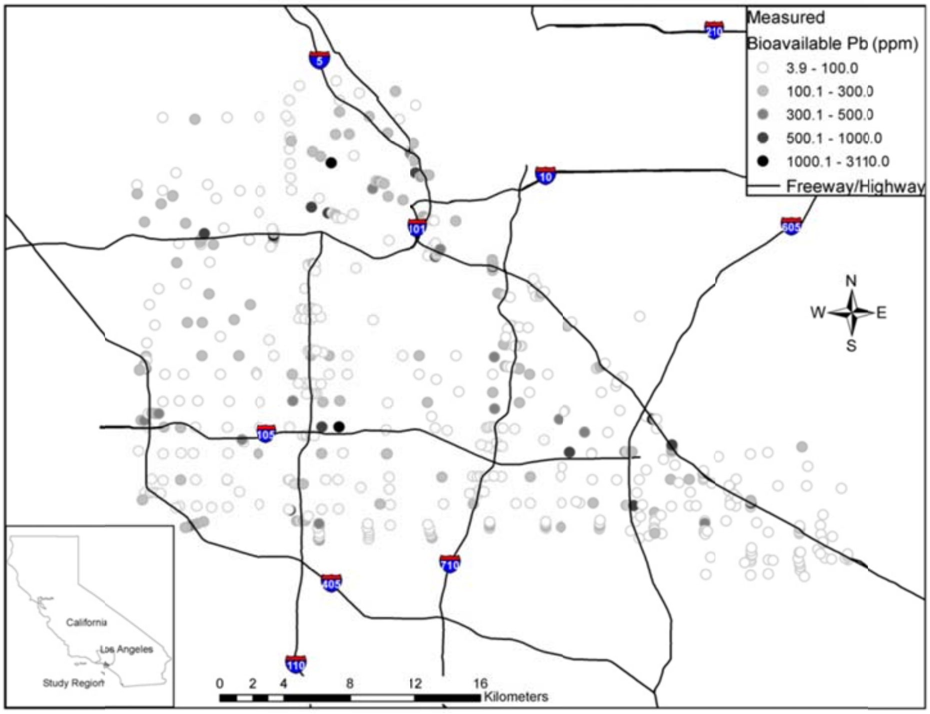
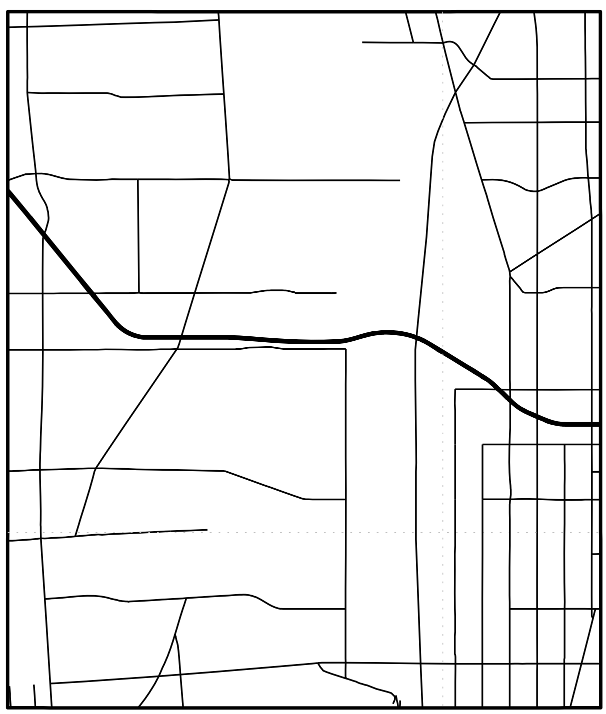


Figure D-1: Map of the Initial Assessment Areas (IAA) outlined in black and Expanded Assessment Areas (EAA) outlined in yellow.

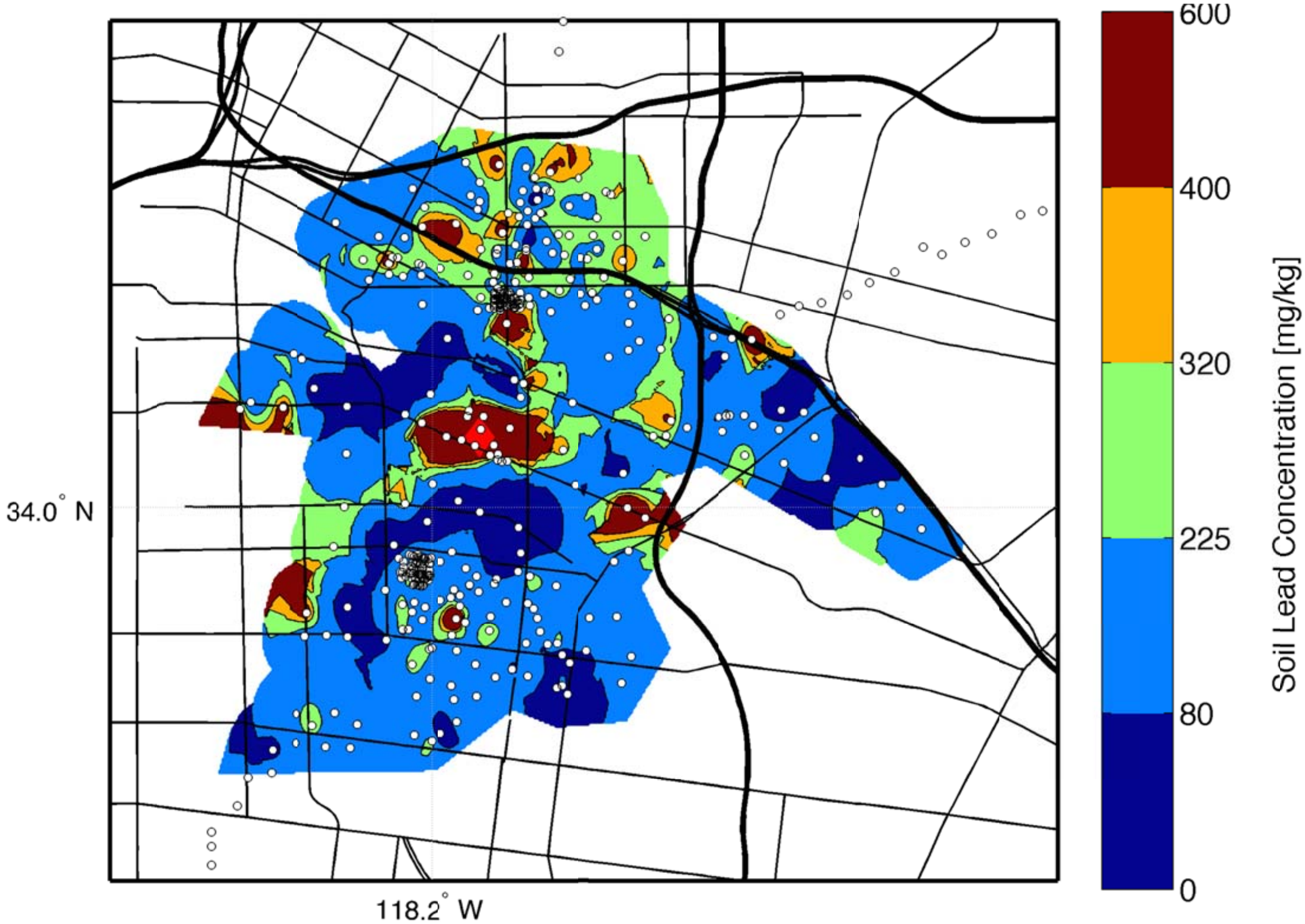

Measured
Bioavailable Pb (ppm)
3.9 - 100.0
100.1 - 300.0
300.1 - 500.0
500.1 - 1000.0
1000.1 - 3110.0
Freeway/Highway
N
W
E
S
California
Los Angeles
Study Region
0 2 4 8 12 16 Kilometers



Appendix E. Larger Versions of Figures Above

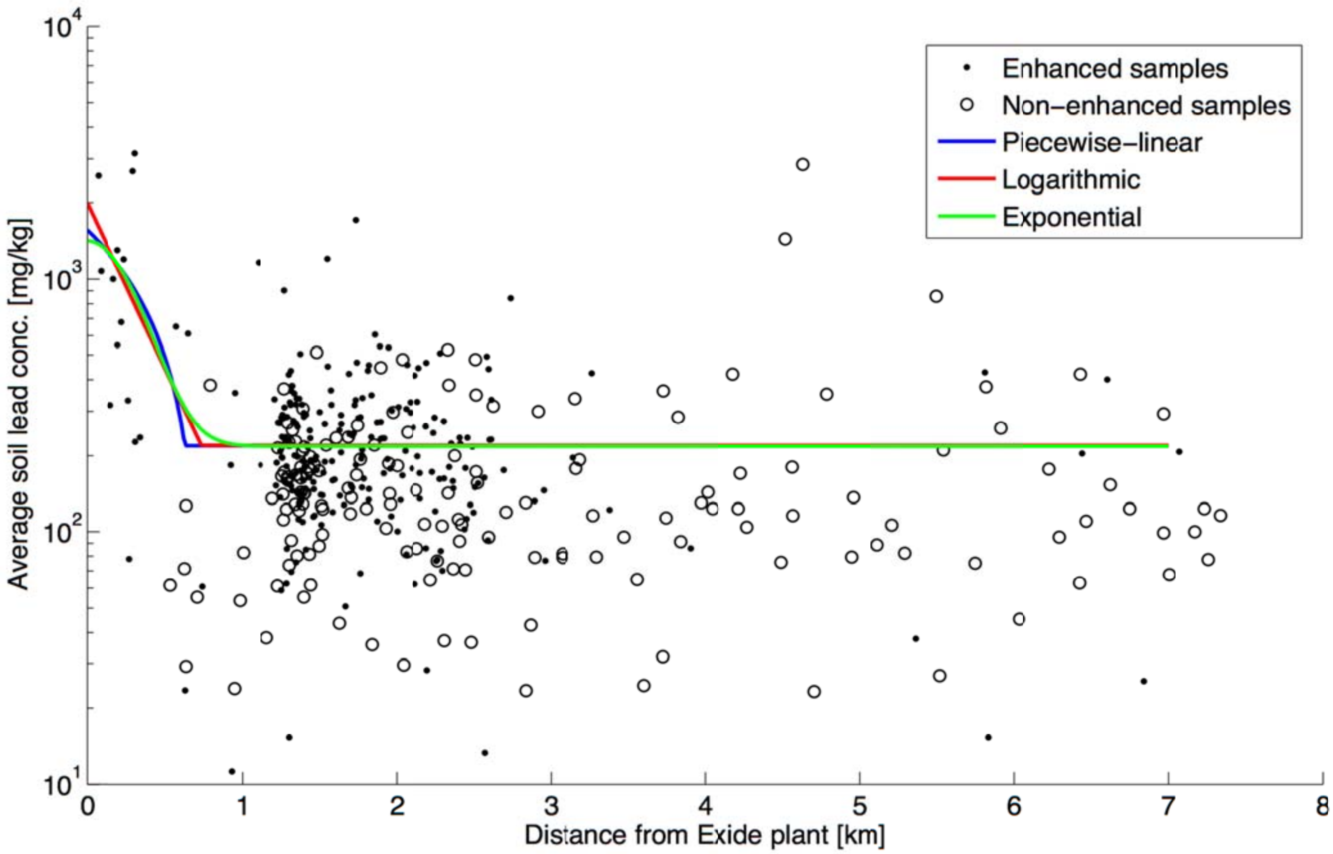
E.1 Figure 1

Map of property-average soil lead concentration, interpolated by kriging. Sampling locations are shown as white circles.



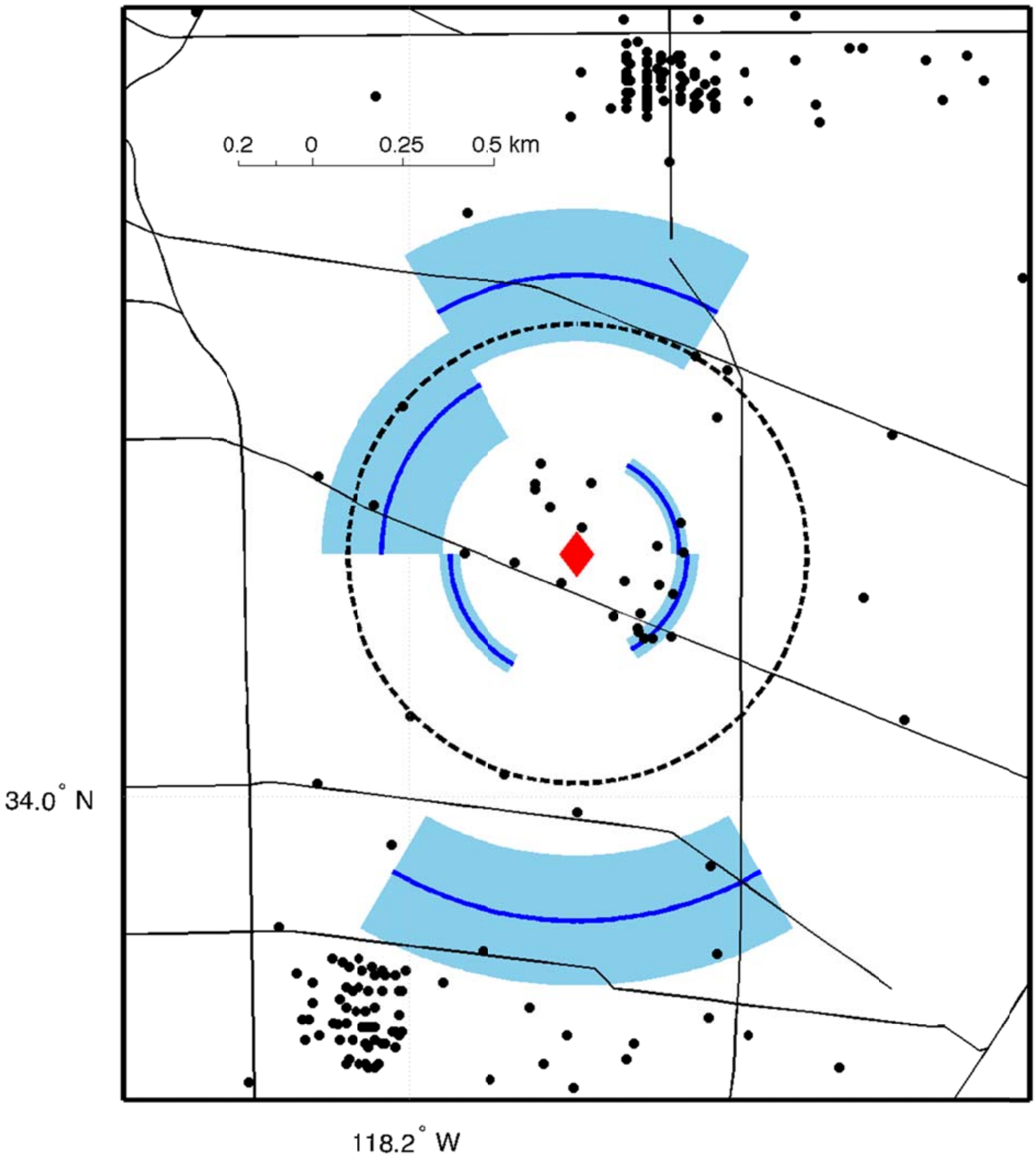
E.2 Figure 2

Soil lead concentration as a function of distance from the Exide facility. Three different fitted regression analyses are overlain on the empirical data. All three estimate similar values for the ambient background soil lead concentration and distances at which the concentrations fall to the ambient background level. Open circles represent data from "non-enhanced" sampling locations, which are not close to freeways, arterial roads, or houses built before 1940. Dots represent sampling sites that have one or more of those factors likely to enhance soil lead concentrations



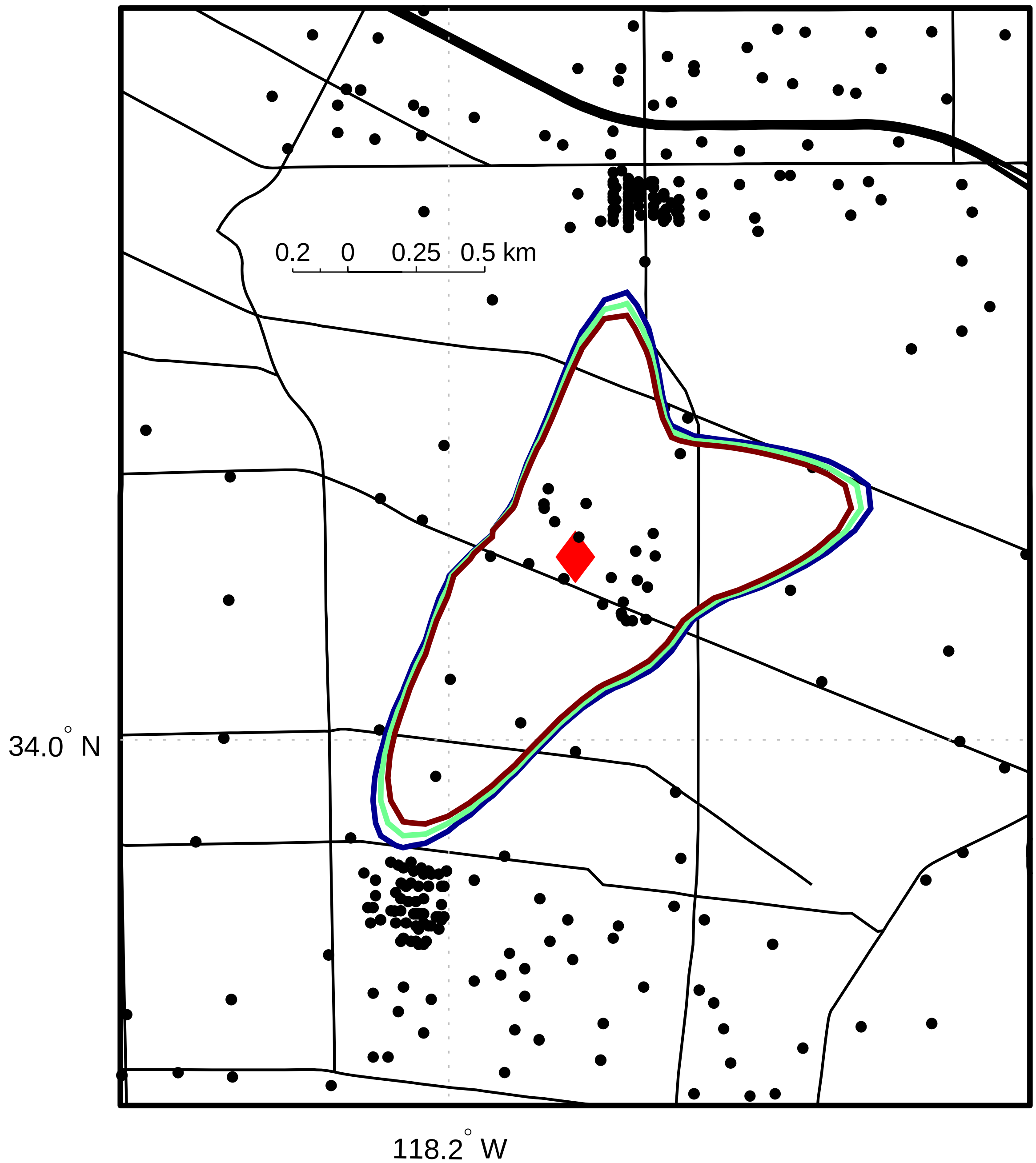
E.3 Figure 3

Distances to reach ambient background concentrations using the linear regression for directional subsets. Light blue bands represent the ±1 standard error confidence intervals. The clusters of black dots at the top and bottom at the northern and southern assessment areas.



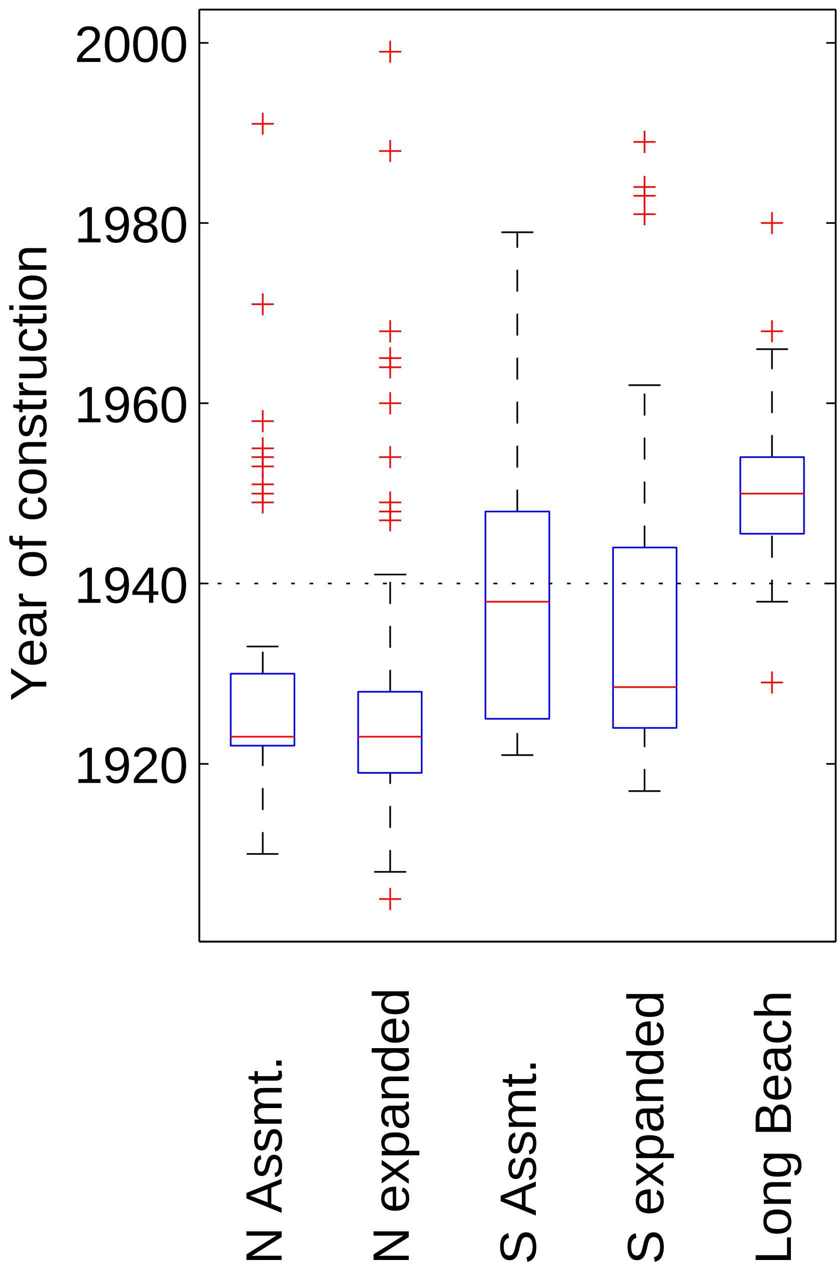
E.4 Figure 4

Contours of Exide's contribution to soil lead concentrations that are statistically distinguishable from background concentrations based on two times the computed standard deviation of the soil lead measurement error at the estimated background concentration (=2*0.136*129.2 = 35.14 mg/kg; see Table 9). This results in a critical modeled ambient air concentration of 4.55 x 10^{-3} μg/m^3 from the Exide facility, plotted as the green line. Uncertainty in the fitted coefficient for modeled ambient concentrations yields the red and blue lines.



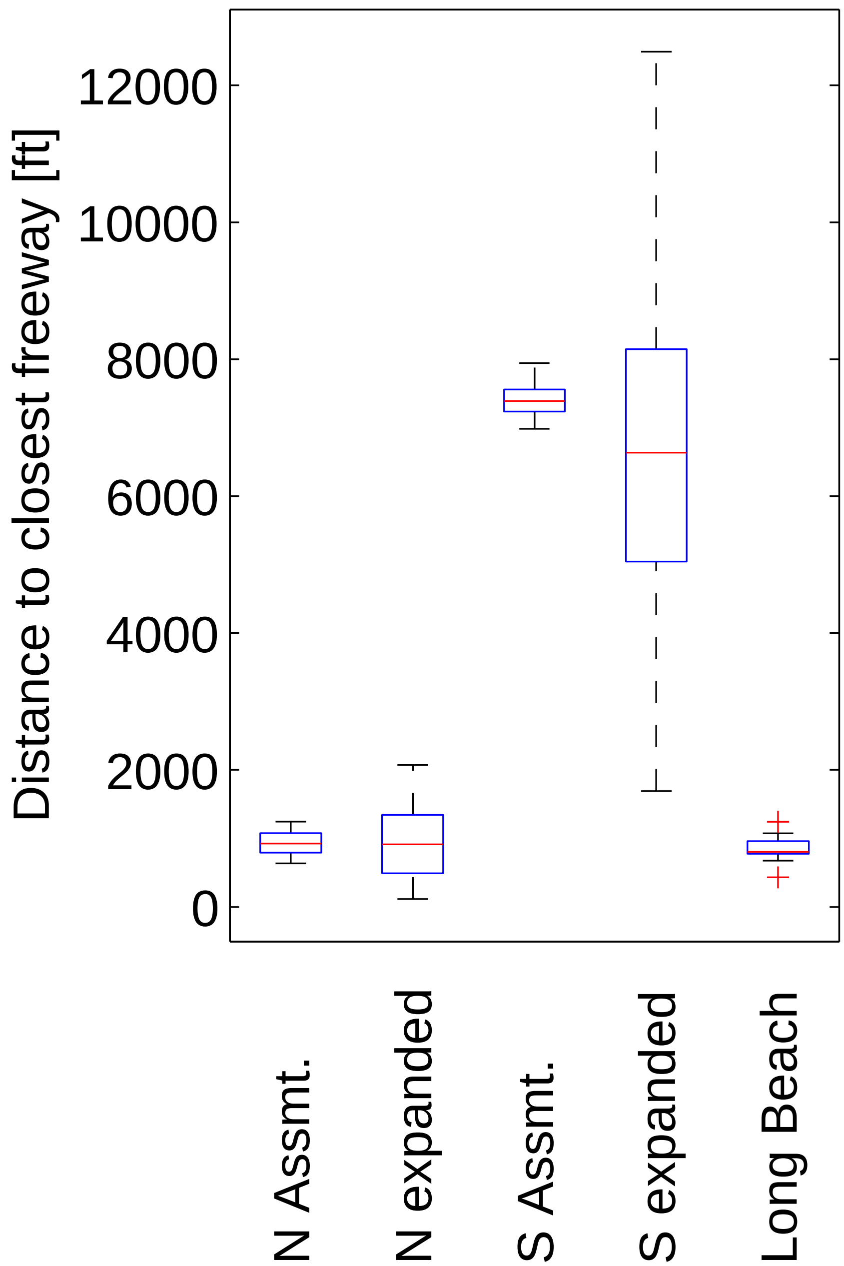
E.5 Figure 5 left

Boxplots comparing year of construction of homes for each property.



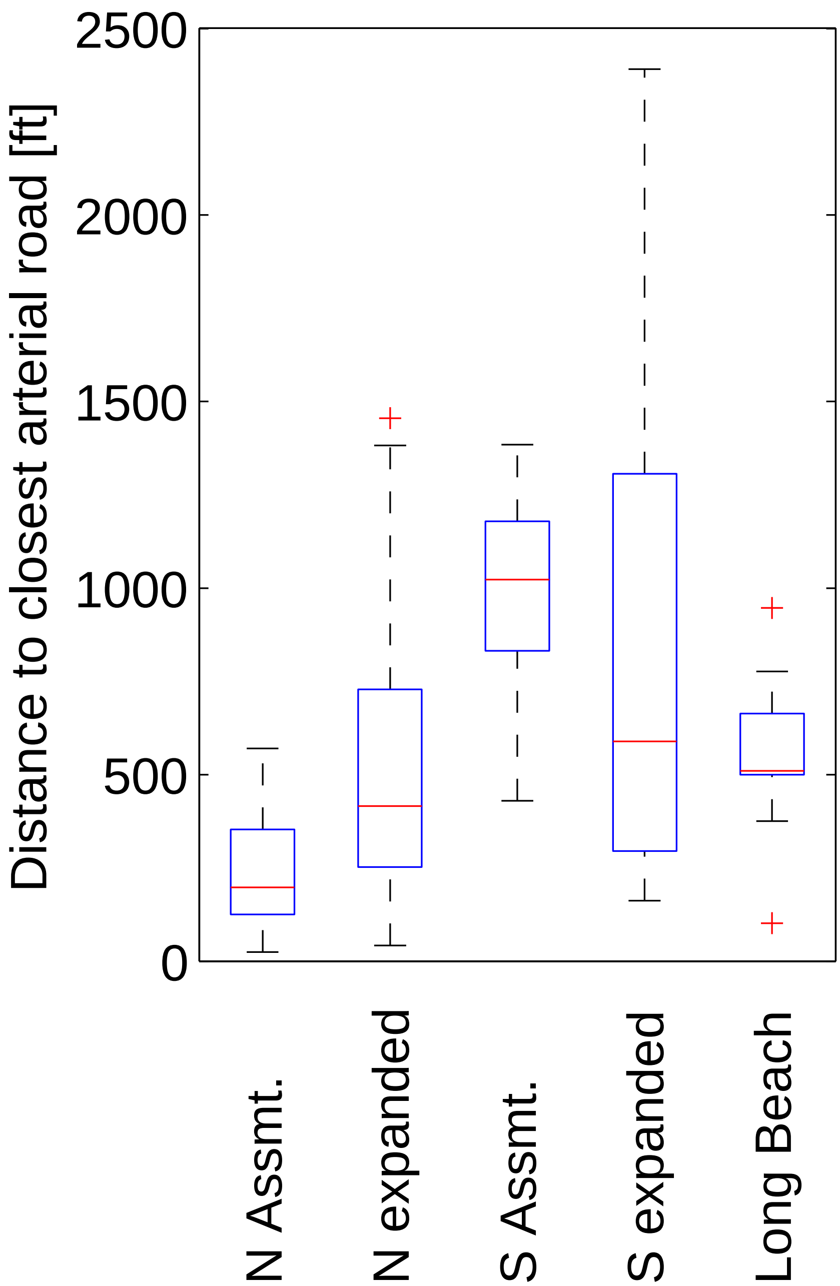
E.6 Figure 5 center

Boxplot comparing distance to the closest freeway for each property.



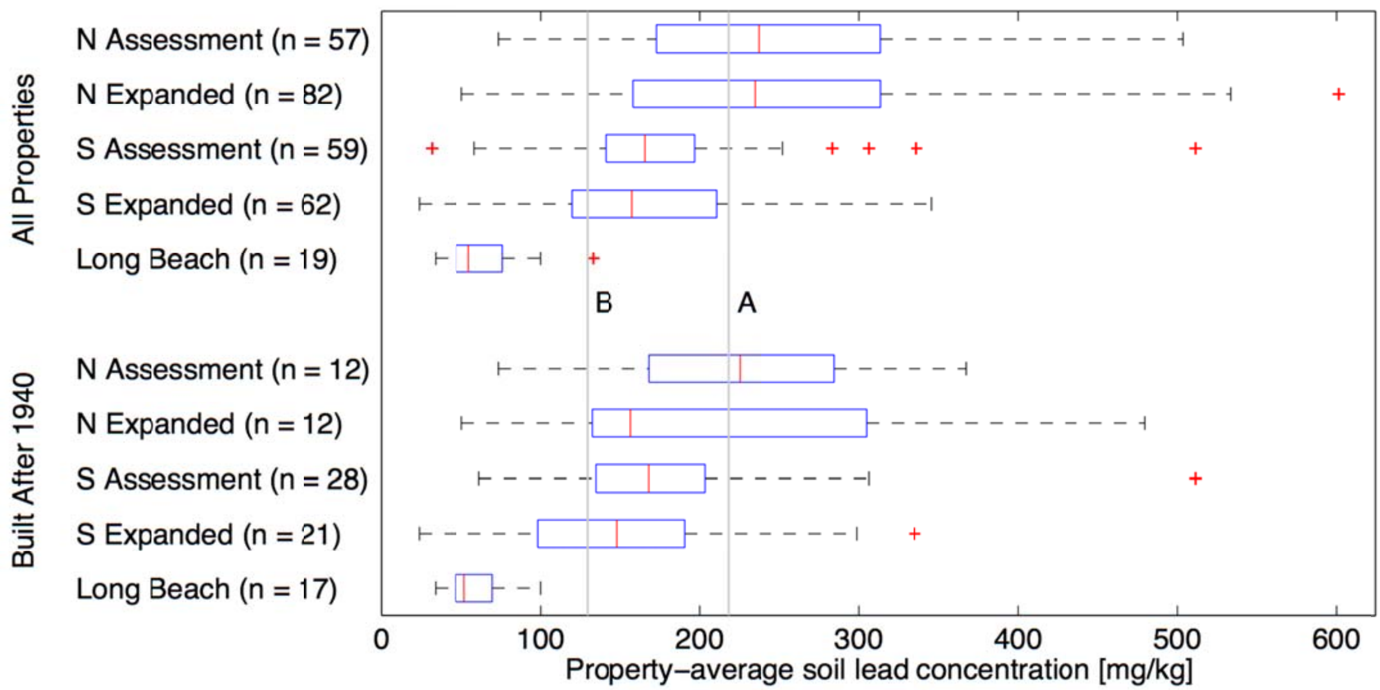
E.7 Figure 5 right

Boxplot comparing distance to the closest arterial road for each property.



E.8 Figure 6

Boxplots comparing distributions property-averaged soil lead concentrations in different areas. The top group includes all sampled properties; the bottom group includes only properties built after 1940. The vertical lines "A" and "B" correspond to the ambient urban background concentrations estimated in Sections 3.2 and 3.3, respectively.



E. 9 Figure 7

95% upper confidence level for mean soil lead concentrations for small groups of sampled properties in the assessment areas and expanded sampling areas. Each group is labeled with the 95^{th} UCL for the group mean.