Final Report: Statistical Modeling and Analysis Results for the Topsoil Lead Contamination Study (Quemetco Project)

Submitted to:

Prof. Shoumo Mitra
Department of Agriculture
Cal Poly Pomona

Russell Plumb
Masters Candidate
Cal Poly Pomona

Report Prepared By:

Scott M. Lesch
Principal Statistician, GEBJ Salinity Laboratory
Consulting Affiliate, Statistical Consulting Collaboratory
University of California: Riverside, CA 92521
(951) 369-4861

Daniel R. Jeske
Associate Professor, Department of Statistics
Director, Statistical Consulting Collaboratory
University of California: Riverside, CA 92521
(951) 827-3014

Javier Saurez
Ph.D. Student, Department of Statistics
University of California: Riverside, CA 92521

January 28, 2006

Table of Content

Executive Summary

This report summarizes the statistical modeling and analysis results associated with the Ca Poly Pomona Topsoil Lead Contamination study. The purpose of this report is to document both the implemented sampling design and all corresponding data modeling and inference techniques used during the subsequent statistical analyses.

The development of the sampling protocol, including both the initial recommended design and final implemented sampling strategy are discussed in Section 2. The initial Stratified Random sampling design was developed using a Neyman allocation scheme. After presenting this design to the client, a refined GIS analysis was performed and more accurate available sampling areas for each school were calculated. These calculations were used to revise the second-stage random sampling scheme. Additionally, two extra properties were added to the sampling design (one nursery located within 2 Km of the factory and one previously overlooked park) and 12 additional sampling locations were selected along the factory perimeter. After these refinements, the final sampling plan contained 361 sampling locations from 69 distinct non-factory properties (and the factory perimeter).

The basic univariate statistics that summarize the contamination data associated with the analyzed metals (for all 360 topsoil samples) are given in Section 3. A total of seven metal concentration measurements were made on each topsoil sample; the metals analyzed in this study include Arsenic (As), Cadmium (Cd), Chromium (Cr), Copper (Cu), Nickel (Ni), Lead (Pb), and Zinc (Zn). The univariate statistics summarize both the raw and natural log transformed metal data, where the transformed data is defined as $Y = \ln(X+1)$. The histograms and quantile plots of each log transformed metal data appear to be approximately symmetric (but in some cases also moderately heavy-tailed).

Section 4 presents the analysis of the sampling depth effect, based on the 43 sites were topsoil samples were acquired from two sampling depths.. Paired t-tests and sign-

rank tests are employed to determine what, if any, effect the sampling depth had on the observed metal concentration levels. Both sets of tests suggest that there was no sampling depth effect at the 0.05 level (i.e., the mean and/or median metal concentration levels did not change across sampling depths).

Two types of exploratory data analysis (EDA) plots for assessing the degree of spatial structure (present in the metal concentration data) are discussed in Section 5; quatile maps and robust variogram plots. The quantile maps suggest that a substantial amount of short-range, local variation is present in the metal concentration data. Additionally, both the quantile maps and variogram plots suggest that distinct property effects may also be present; i.e., samples gathers from within one property may be more similar (less variable) than samples gathered from different properties.

Section 6 introduces the idea of quantile indicator maps and describes the corresponding Chi-square tests of association that are derived from these maps. The corresponding Chi-square test results indicate that at the corrected 0.01 significance level, an excessive number of Pb samples near the factory exceed both the median and q90 cut-offs. Additionally, an excessive number of Cr and Ni samples exceed the q90 cut-off. These results imply that an abnormally high number of "hot" (i.e., contaminated) Cr, Ni, and Pb samples occur within close proximity (< 2 Km) to the factory location.

Section 7 presents the contamination by distance to factory (CD2F) plots. These plots display the natural log transformed contamination levels for each metal as a function of the distance (of each sample site) to the factory, along with a smoothed spline function fitted to the resulting contamination pattern. The CD2F plots for Cr, Ni, and Pb display fairly clear evidence of an increasing contamination trend towards the factory.

Finally, in Section 8 a mixed linear spline model is proposed for modeling the distance to factory effect, while simultaneously adjusting for secondary covariates that were hypothesized to also (possibly) influence the metal contamination levels. The fitted spline models are then used to estimate the *Baseline*, *Factory*, and *Proximity* effects. The

Baseline effect estimates the background log contamination level across the survey region (i.e., the background level not influenced by the factory), the *Factory* effect estimates the log contamination level within or immediately around the perimeter of the factory, and the *Proximity* effect quantifies the distance to factory contamination relationship. These results agree with the earlier test results presented in sections 6 and 7. More specifically, they confirm that (i) the factory perimeter samples appear to be highly contaminated with respect to the estimated baseline metal contamination levels observed throughout the sampling region (for all metals), and (ii) at least two (and possibly three) of the seven metals analyzed in this study (Cr, Ni, and Pb) exhibit significantly elevated contamination levels near the factory site.

1.0 Introduction

This report summarizes all of the primary statistical modeling and analysis results associated with the Ca Poly Pomona Topsoil Lead Contamination study. The purpose of this report is to document both the implemented sampling design and all corresponding data modeling and inference techniques used during the subsequent statistical analyses. Additionally, this report is designed to serve as a template for describing the sampling protocol and statistical analysis techniques in any future technical manuscripts developed by the client(s).

The remainder of this report is organized as follows. Section 2 describes the development of the sampling protocol, including both the initial recommended design and final implemented sampling strategy. Section 3 presents the basic univariate statistics that summarize the contamination data associated with the seven analyzed metals (for all 360 topsoil samples). Section 4 presents the analysis of the sampling depth effect, based on the 43 sites where topsoil samples were acquired from two sampling depths. Section 5 next describes the two types of exploratory data analysis (EDA) plots used for initially determining the degree of spatial structure present in the metal concentration data; i.e., the quatile maps and robust variogram plots. Section 6 then introduces the idea of quantile indicator maps and describes the corresponding Chi-square tests of association that are derived from these maps. Following this, section 7 presents the contamination by distance to factory (CD2F) plots, and section 8 presents the results for the formal mixed linear spline models (motivated by the CD2F plots). Note that the main confirmatory statistical results concerning the apparent factory contamination effect(s) are given in sections 6 and 8, respectively.

2.0 Sampling Protocol

It is well known that topsoil samples are very sensitive to historical near surface activities. For example, in highly industrialized areas it is not at all uncommon to find significant disturbances to the topsoil due to various (commercial, residential, or industrial construction) "cut-and-fill" activities. Thus, in order to collect reliable topsoil sample data for this study, the sampling locations were restricted to two specific types of well established, open-space areas: (i) public access parkland, and (ii) public or private school playgrounds.

Using a preliminary GIS analysis (performed by the client), 67 schools and public parklands were identified to be within 4.8 Km (3.0 miles) of the factory site. In addition to identifying the centroid location of each property (school or park), the approximate size of each property was also calculated and subsequently used in the initial sample allocation process. This initial protocol followed a two-stage sampling design. In the first stage the identified properties were divided into 3 strata based on their centroid distances to the factory; these strata were defined as follows:

Strata A:	within 0 – 1.6 Km (0-1 miles) of factory
Strata B:	within 1.6 – 3.2 Km (1-2 miles) of factory
Strata C:	within 3.2 – 4.8 Km (2-3 miles) of factory

Figure 2.1 shows an example of the circular stratum pattern used in the initial protocol. Based on prior research on sampling for trace metal concentrations in soil, we initially assumed that the sampling variances (for Pb) would be approximately $9\sigma^2$, $3\sigma^2$, and σ^2 for strata A, B, and C, respectively (Jackson et al., 1987). Using an initial target sample size of 300 sites, we used a Neyman allocation scheme to allocate the samples across these three stratum (Lohr, 1999). In the second stage of the sampling plan (i.e., within each stratum), we then employed a proportional allocation scheme to determine the number of samples chosen from each identified property (proportional to the size of each property). Note that during this initial analysis, 50% of the calculated area of each school property

was assumed to consist of playgrounds or school yards amenable to sampling (in contrast, 100% of the public park areas were assumed to be amenable to sampling).

After this initial sampling design was presented to the client, a refined GIS analysis was performed (again by the client). During this second stage GIS analysis, more accurate available sampling areas for each school were calculated and a simple random sampling scheme was employed to select random sampling positions within each identified property. Additionally, during this refined analysis two extra properties were added to the sampling design (one nursery located within 2 Km of the factory and one previously overlooked park), along with 12 additional sampling locations on the factory perimeter. Due to these refinements, the final sampling plan contained 361 sampling locations from 69 distinct non-factory properties (and the factory perimeter). Table 2.1 summarized the final number of properties and sample sites acquired within each strata; a complete listing of the property names and number of samples acquired at each property are given in the next section.

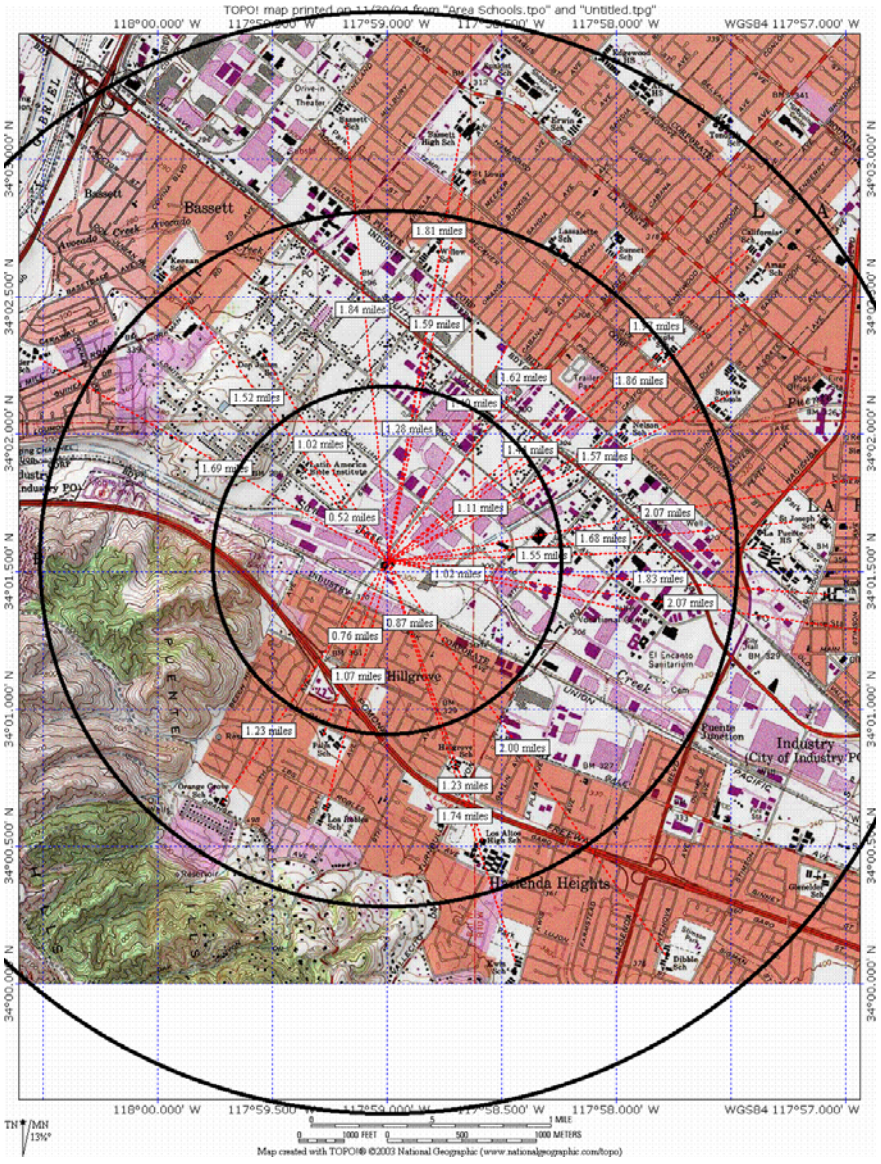
For the record, one topsoil sample that registered 0 concentration levels for all seven metals has been removed from the subsequent data analyses. This topsoil sample corresponds to sample site #3 on the Cedarlane Middle School property.

Table 2.1 Final number of properties and sample sites allocated within the three Strata defining the sampling region.

Strata	Number of Properties	Number of Sample Sites	Percent	Cumulative Frequency
A	6 **	44	12.2	44
B	29	158	43.8	202
C	35	159	44.0	361

(**): includes factory perimeter (12 samples)

Figure 2.1 Example of the circular stratum pattern used in the initial sampling protocol.



3.0 Basic Summary Statistics

In this study, 360 topsoil samples were acquired from 70 distinct property locations (including the perimeter of the factory site). Table 3.1 lists the number of samples associated with each property; note that the factory site is listed as "RSR Quemetco Inc.". Most properties contained between 3 to 8 sample sites each, but 52 distinct sample sites were collected from one very large park ("Industry Hills").

A total of seven metal concentration measurements were initially made on each topsoil sample and then these ICP measured concentration levels were converted into ppm units (for the subsequent statistical analyses). Each sample was analyzed for the following metals: Arsenic (As), Cadmium (Cd), Chromium (Cr), Copper (Cu), Nickel (Ni), Lead (Pb), and Zinc (Zn). Table 3.2 lists the basic univariate summary statistics for the ppm metal data. For the non-factory sites, these statistics include the mean, standard deviation, variance, skewness, and associated quantile estimates (minimum, 1%, 5%, 10%, median, 90%, 95%, 99%, and maximum) for each metal. For the 12 factory sites, the displayed statistics include the mean, standard deviation, variance, minimum, and maximum observed levels for each metal. Note that a cursory examination of Table 3.2 shows that the mean concentration level of each metal at the factory site tends to be 4 to 10 times higher than the surrounding area.

Table 3.3 lists the exact same set of univariate summary statistics for the natural log transformed metal data, where the transformation is defined as $Y = \ln(X+1)$. Thus, all statistics are expressed in ln(ppm) units in Table 3.3.

Table 3.4 lists the observed, log transformed metal correlation matrix for the 348 non-factory sample sites. This matrix shows the degree of correlation between each (log transformed) metal across the sampling region. The corresponding p-values are shown beneath each correlation estimate (these p-values can be used to assess the statistical significance of each correlation estimate, if desired). Interestingly, these correlation

estimates suggest that the seven metals are only weakly correlated with one another (across the sampling region).

Finally, univariate panel graphs for each metal are shown after the tables (Figures 3.1 through 3.7). Each panel graph contains two histogram charts and two quantile plots for a specific metal; these charts and plots show the distribution shape of the raw and natural log transformed data, respectively. All seven panel graphs suggest that the log transformed metal data is approximately symmetric (but in some cases also moderately heavy-tailed).

Table 3.1 Listing of property names and the number of samples acquired from within each property.

Property Name	# of Sample Sites	Percent	Cumulative Frequency	Cumulative Percent
Allen J. Martin Park	7	1.94	7	1.94
Amar High School	4	1.11	11	3.06
Amar School	3	0.83	14	3.89
Avacado Heights Park	6	1.67	20	5.56
B.F. Maxson Elementary School	1	0.28	21	5.83
Baldwin School	1	0.28	22	6.11
Bassett Park	8	2.22	30	8.33
Bassett Senior High School	13	3.61	43	11.94
Bixby School	2	0.56	45	12.50
California Elementary School	1	0.28	46	12.78
California School	1	0.28	47	13.06
Cedarlane Middle School	3	0.83	50	13.89
Charles T. Kranz Intermediate School	2	0.56	52	14.44
Del Valle School	4	1.11	56	15.56
Dibble School	7	1.94	63	17.50
Don julian Elementary School	3	0.83	66	18.33
Glen A. Wilson High School	9	2.50	75	20.83
Glenelder Elementary School	2	0.56	77	21.39
Hacienda La Puente Adult Education Schoo	6	1.67	83	23.06
Hillgrove School	8	2.22	91	25.28
Industry Hills	52	14.44	143	39.72
J.E. Van Wig Elementary School	1	0.28	144	40.00
Keenan School	4	1.11	148	41.11
Kwiss Elementary School	8	2.22	156	43.33

Table 3.1 continued...

Property Name	# of Sample Sites	Percent	Cumulative Frequency	Cumulative Percent
La Puente High School	13	3.61	169	46.94
La Puente Park	15	4.17	184	51.11
Lassalette Elementary School	4	1.11	188	52.22
Latin American Bible School	4	1.11	192	53.33
Los Altos Elementary School	2	0.56	194	53.89
Los Altos High School	13	3.61	207	57.50
Los Molinos Elementary School	2	0.56	209	58.06
Los Robles Elementary School	4	1.11	213	59.17
Los Robles Park	3	0.83	216	60.00
Manzanita Park	8	2.22	224	62.22
Mesa Robles Elementary School	4	1.11	228	63.33
Mountain View High School	5	1.39	233	64.72
Mountain View Park	2	0.56	235	65.28
Nelson School	4	1.11	239	66.39
Newton Middle School	3	0.83	242	67.22
Nueva Vista Continuation High School	2	0.56	244	67.78
Orange Grove Middle School	6	1.67	250	69.44
Orangewood Elementary School	1	0.28	251	69.72
Palm Elementary School	10	2.78	261	72.50
Palm Nursery	4	1.11	265	73.61
Puente Hills high School	2	0.56	267	74.17
RSR Quemetco Inc.	12	3.33	279	77.50
Rio Hondo Junior College	10	2.78	289	80.28
Saint Joseph School	8	2.22	297	82.50
Saint Louis School	2	0.56	299	83.06
San Angelo Park	2	0.56	301	83.61
Shadybend Elementary School	2	0.56	303	84.17
Sierra Vista Middle School	3	0.83	306	85.00
Sparks Elementary School	7	1.94	313	86.94
Stimson Park	3	0.83	316	87.78
Sunkist Elementary School	1	0.28	317	88.06
Sunset Elementary School	3	0.83	320	88.89
Temple Elementary School	5	1.39	325	90.28
Thomas Burton Park	3	0.83	328	91.11
Thomas E. Erwin Elementary School	1	0.28	329	91.39
Tonopah School	2	0.56	331	91.94
Torch Middle School	3	0.83	334	92.78
Tri Cities Acadamy	2	0.56	336	93.33
Valley Continuation HS	5	1.39	341	94.72
Valley Vocational Center	2	0.56	343	95.28
Wallen L. Andrews North Whittier Element	4	1.11	347	96.39
Walnut Creek Nature Park	1	0.28	348	96.67
Willwood School	1	0.28	349	96.94
Workman Elementary School	2	0.56	351	97.50
Workman High School	9	2.50	360	100.00

Table 3.2 Basic univariate summary statistics for As, Cd, Cr, Cu, Ni, Pb, and Zn: all statistics shown in ppm units.

Non-Factory Sites

Variable	Label	N	Mean	Std Dev	Variance	Skewness	Minimum
As	As [ppm]	348	5.04	4.23	17.91	3.99	0.00
Cd	Cd [ppm]	348	0.63	1.34	1.79	15.06	0.00
Cr	Cr [ppm]	348	19.85	12.83	164.49	3.29	1.30
Cu	Cu [ppm]	348	25.18	14.39	207.18	3.08	0.20
Ni	Ni [ppm]	348	14.54	8.81	77.57	3.36	1.50
Pb	Pb [ppm]	348	34.93	34.45	1186.58	3.05	2.35
Zn	Zn [ppm]	348	123.91	126.81	16079.95	6.95	8.00

Variable	Label	1st Pctl	5th Ptcl	10th Pctl	50th Pctl	90th Pctl
As	As [ppm]	0.70	1.68	2.10	3.77	9.21
Cd	Cd [ppm]	0.00	0.00	0.00	0.51	1.02
Cr	Cr [ppm]	4.30	9.80	11.28	17.00	29.37
Cu	Cu [ppm]	8.35	11.62	14.00	21.90	39.67
Ni	Ni [ppm]	5.07	6.60	7.96	12.69	20.53
Pb	Pb [ppm]	5.52	7.29	8.97	24.28	71.97
Zn	Zn [ppm]	33.40	42.53	47.14	87.13	244.00

Variable	Label	95th Pctl	99th Pctl	Maximum
As	As [ppm]	12.40	25.00	40.86
Cd	Cd [ppm]	1.48	2.80	23.71
Cr	Cr [ppm]	50.20	77.60	99.48
Cu	Cu [ppm]	54.00	73.60	130.08
Ni	Ni [ppm]	31.00	58.75	68.80
Pb	Pb [ppm]	107.15	198.50	277.45
Zn	Zn [ppm]	289.36	594.00	1737.53

Quemetco Factory Sites

Variable	Label	N	Mean	Std Dev	Variance	Minimum	Maximum
As	As [ppm]	12	21.09	14.57	212.35	6.25	49.90
Cd	Cd [ppm]	12	4.57	5.75	33.06	1.10	19.40
Cr	Cr [ppm]	12	80.98	51.48	2649.99	22.10	193.55
Cu	Cu [ppm]	12	157.41	128.18	16429.89	1.45	403.50
Ni	Ni [ppm]	12	63.37	31.88	1016.27	14.65	119.25
Pb	Pb [ppm]	12	1911.17	1941.45	3769223.65	284.00	5051.00
Zn	Zn [ppm]	12	1346.60	1301.18	1693058.14	167.00	5027.00

Table 3.3 Basic univariate summary statistics for the natural log transformed As, Cd, Cr, Cu, Ni, Pb, and Zn data: all statistics shown in ln(ppm +1) units.

Non-Factory Sites

Variable	Label	N	Mean	Std Dev	Variance	Skewness	Minimum
lnAs	ln(As+1)	348	1.656	0.497	0.247	0.700	0.000
lnCd	ln(Cd+1)	348	0.420	0.299	0.089	2.945	0.000
lnCr	ln(Cr+1)	348	2.920	0.454	0.206	0.534	0.833
lnCu	ln(Cu+1)	348	3.156	0.461	0.213	-0.345	0.182
lnNi	ln(Ni+1)	348	2.644	0.415	0.172	0.786	0.916
lnPb	ln(Pb+1)	348	3.283	0.742	0.551	0.359	1.209
lnZn	ln(Zn+1)	348	4.594	0.628	0.394	0.681	2.197

Variable	Label	1st Pctl	5th Ptcl	10th Pctl	50th Pctl	90th Pctl
lnAs	ln(As+1)	0.531	0.986	1.131	1.562	2.324
lnCd	ln(Cd+1)	0.000	0.000	0.000	0.409	0.703
lnCr	ln(Cr+1)	1.668	2.380	2.508	2.891	3.413
lnCu	ln(Cu+1)	2.235	2.535	2.708	3.131	3.706
lnNi	ln(Ni+1)	1.803	2.028	2.193	2.617	3.070
lnPb	ln(Pb+1)	1.875	2.116	2.299	3.230	4.290
lnZn	ln(Zn+1)	3.538	3.773	3.874	4.479	5.501

Variable	Label	95th Pctl	99th Pctl	Maximum
lnAs	ln(As+1)	2.595	3.258	3.734
lnCd	ln(Cd+1)	0.907	1.335	3.207
lnCr	ln(Cr+1)	3.936	4.364	4.610
lnCu	ln(Cu+1)	4.007	4.312	4.876
lnNi	ln(Ni+1)	3.466	4.090	4.246
lnPb	ln(Pb+1)	4.684	5.296	5.629
lnZn	ln(Zn+1)	5.671	6.389	7.461

Quemetco Factory Sites

Variable	Label	N	Mean	Std Dev	Variance	Minimum	Maximum
lnAs	ln(As+1)	12	2.890	0.675	0.456	1.981	3.930
lnCd	ln(Cd+1)	12	1.382	0.776	0.601	0.742	3.016
lnCr	ln(Cr+1)	12	4.198	0.711	0.505	3.140	5.271
lnCu	ln(Cu+1)	12	4.560	1.381	1.908	0.896	6.003
lnNi	ln(Ni+1)	12	4.018	0.616	0.380	2.750	4.790
lnPb	ln(Pb+1)	12	7.059	1.061	1.126	5.652	8.528
lnZn	ln(Zn+1)	12	6.818	0.969	0.938	5.124	8.523

Table 3.4 Calculated correlation matrix for the seven log transformed metals.

Ln[Metal+1] Transformed Data
Joint (7-metal) Correlation Structure
Non-Factory Sites

Pearson Correlation Coefficients, N = 348
Prob > |r| under H0: Rho=0

	lnAs	lnCd	lnCr	lnCu	lnNi	lnPb	lnZn
lnAs	1.00000	0.38086	0.12540	0.22097	0.33842	0.40267	0.33431
ln(As+1)		<.0001	0.0193	<.0001	<.0001	<.0001	<.0001
lnCd	0.38086	1.00000	-0.00175	0.10278	0.02427	0.20563	0.19315
ln(Cd+1)	<.0001		0.9740	0.0554	0.6519	0.0001	0.0003
lnCr	0.12540	-0.00175	1.00000	0.40564	0.42893	0.34987	0.30168
ln(Cr+1)	0.0193	0.9740		<.0001	<.0001	<.0001	<.0001
lnCu	0.22097	0.10278	0.40564	1.00000	0.28658	0.22581	0.24774
ln(Cu+1)	<.0001	0.0554	<.0001		<.0001	<.0001	<.0001
lnNi	0.33842	0.02427	0.42893	0.28658	1.00000	0.30287	0.28281
ln(Ni+1)	<.0001	0.6519	<.0001	<.0001		<.0001	<.0001
lnPb	0.40267	0.20563	0.34987	0.22581	0.30287	1.00000	0.52194
ln(Pb+1)	<.0001	0.0001	<.0001	<.0001	<.0001		<.0001
lnZn	0.33431	0.19315	0.30168	0.24774	0.28281	0.52194	1.00000
ln(Zn+1)	<.0001	0.0003	<.0001	<.0001	<.0001	<.0001	

Figure 3.1 Panel graph for As: histograms and quantile plots.

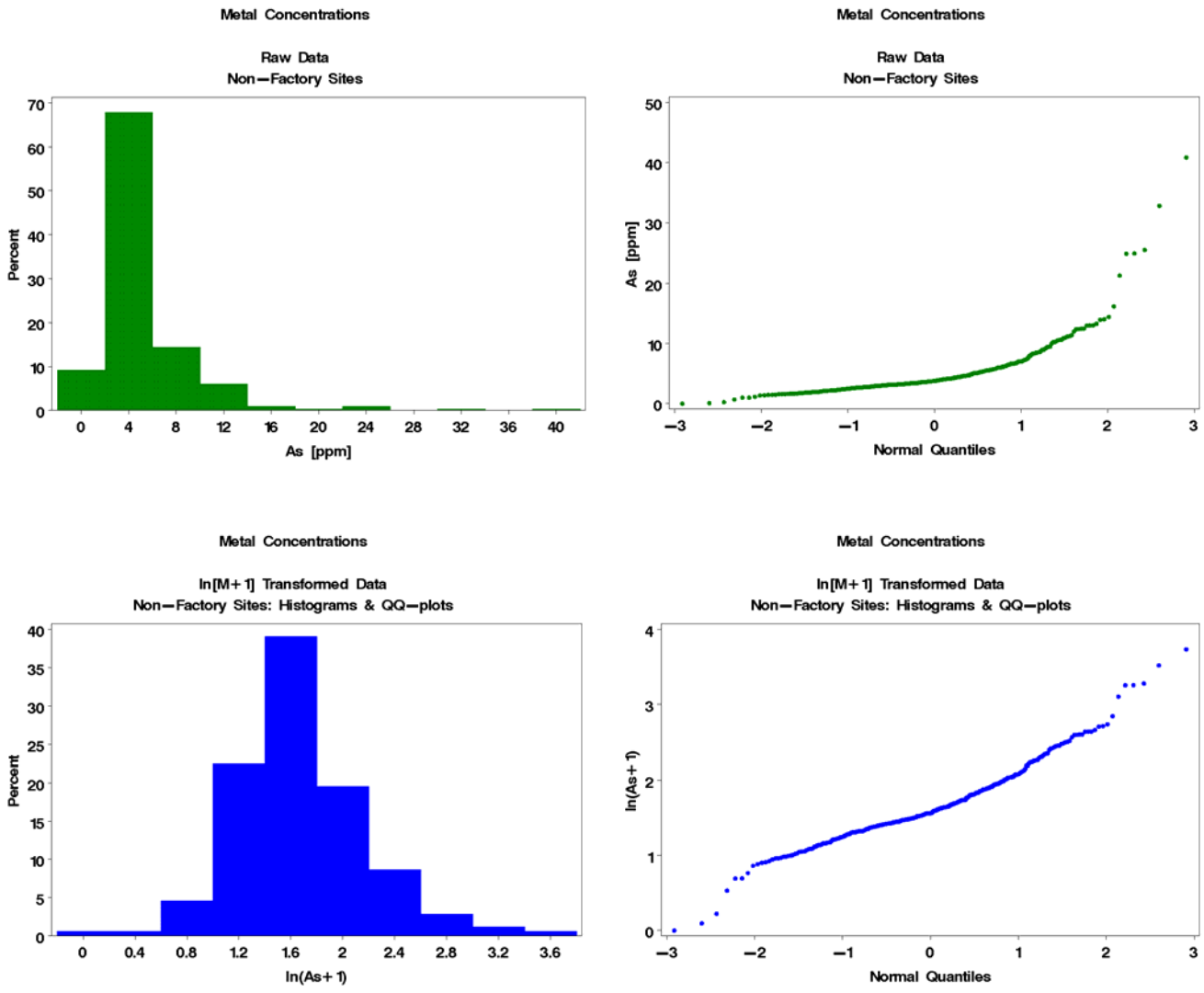


Figure 3.2 Panel graphs for Cd; histograms and quantile plots.

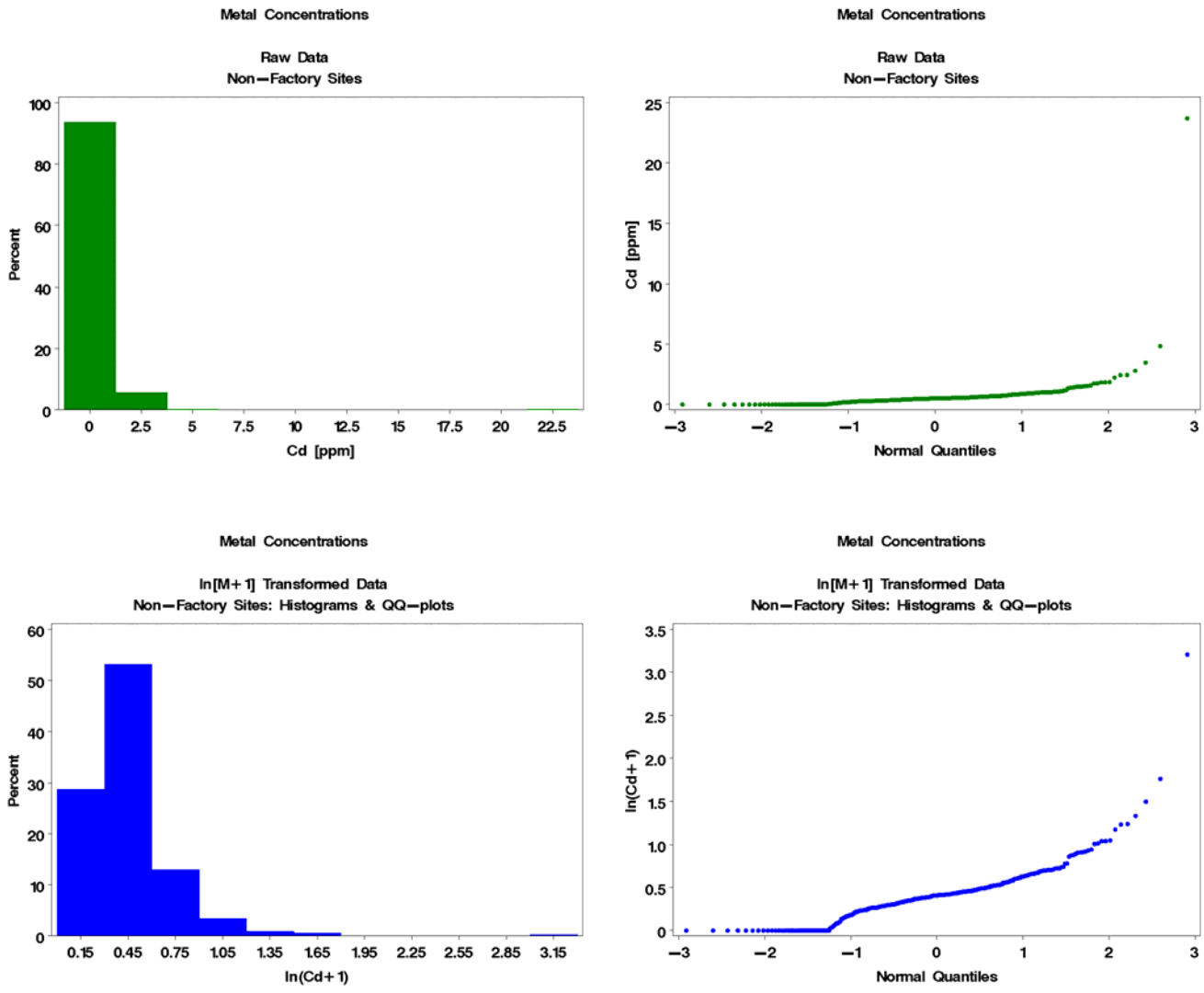


Figure 3.3 Panel graphs for Cr; histograms and quantile plots.

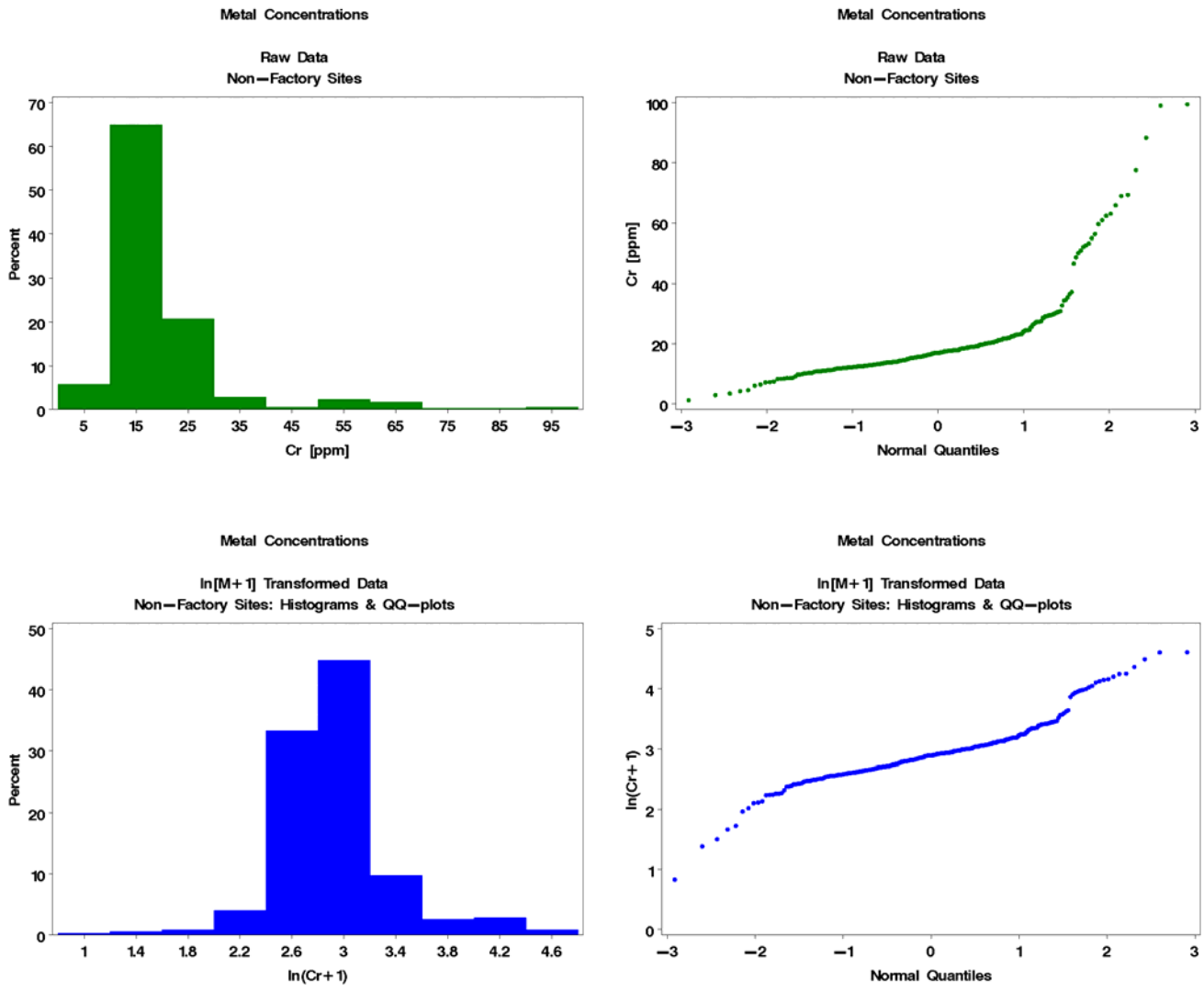


Figure 3.4 Panel graphs for Cu; histograms and quantile plots.

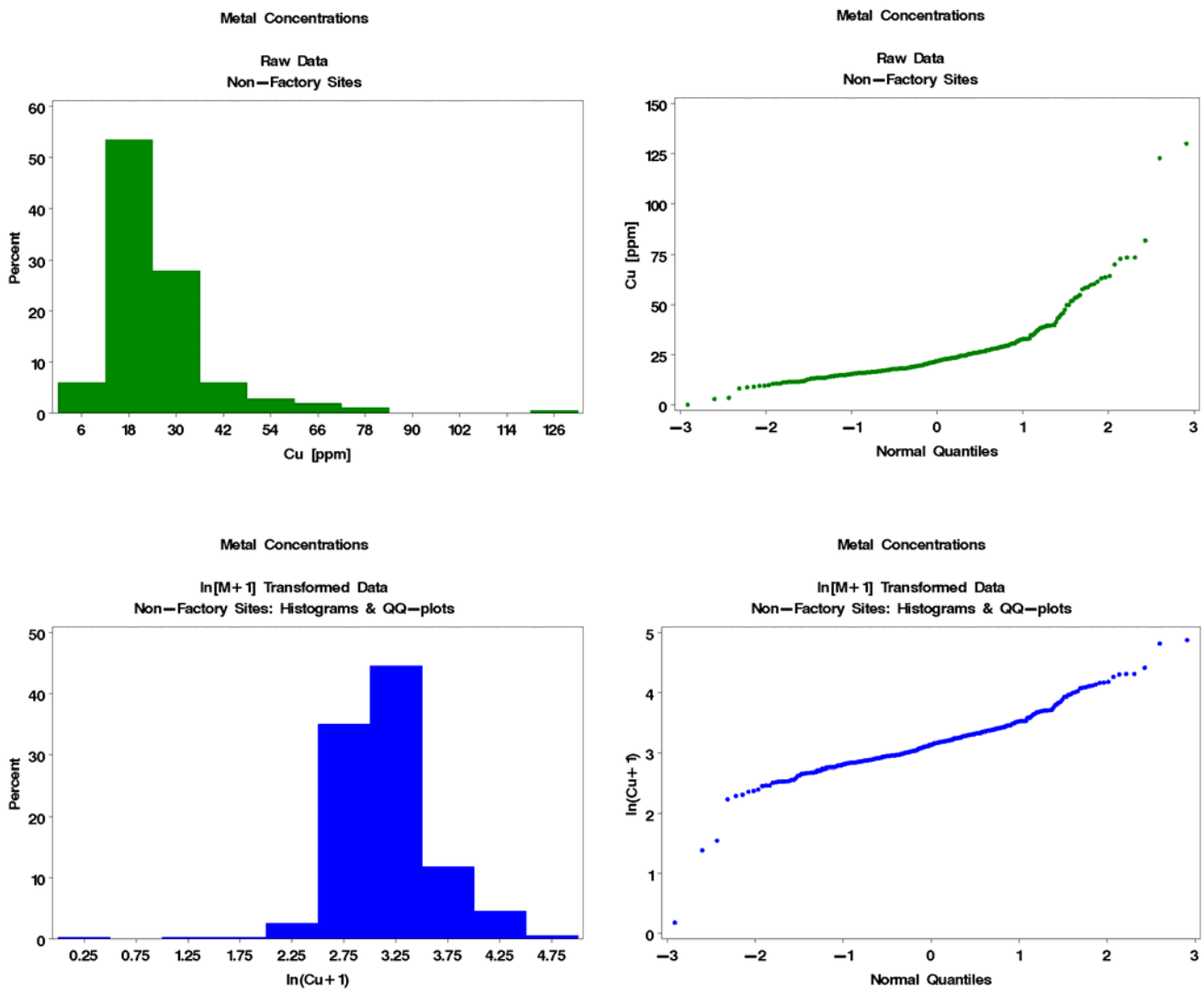


Figure 3.5 Panel graphs for Ni; histograms and quantile plots.

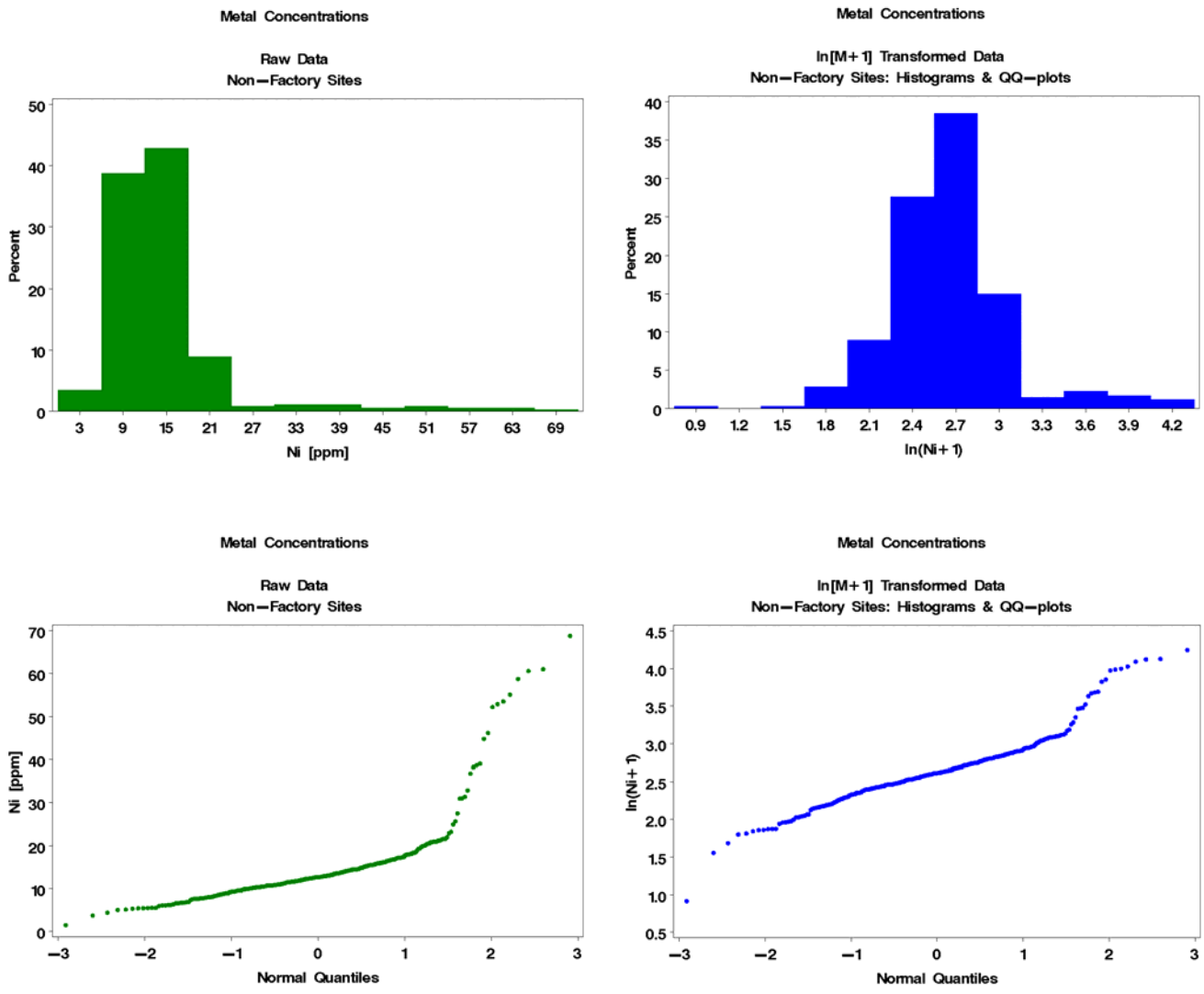


Figure 3.6 Panel graphs for Pb; histograms and quantile plots.

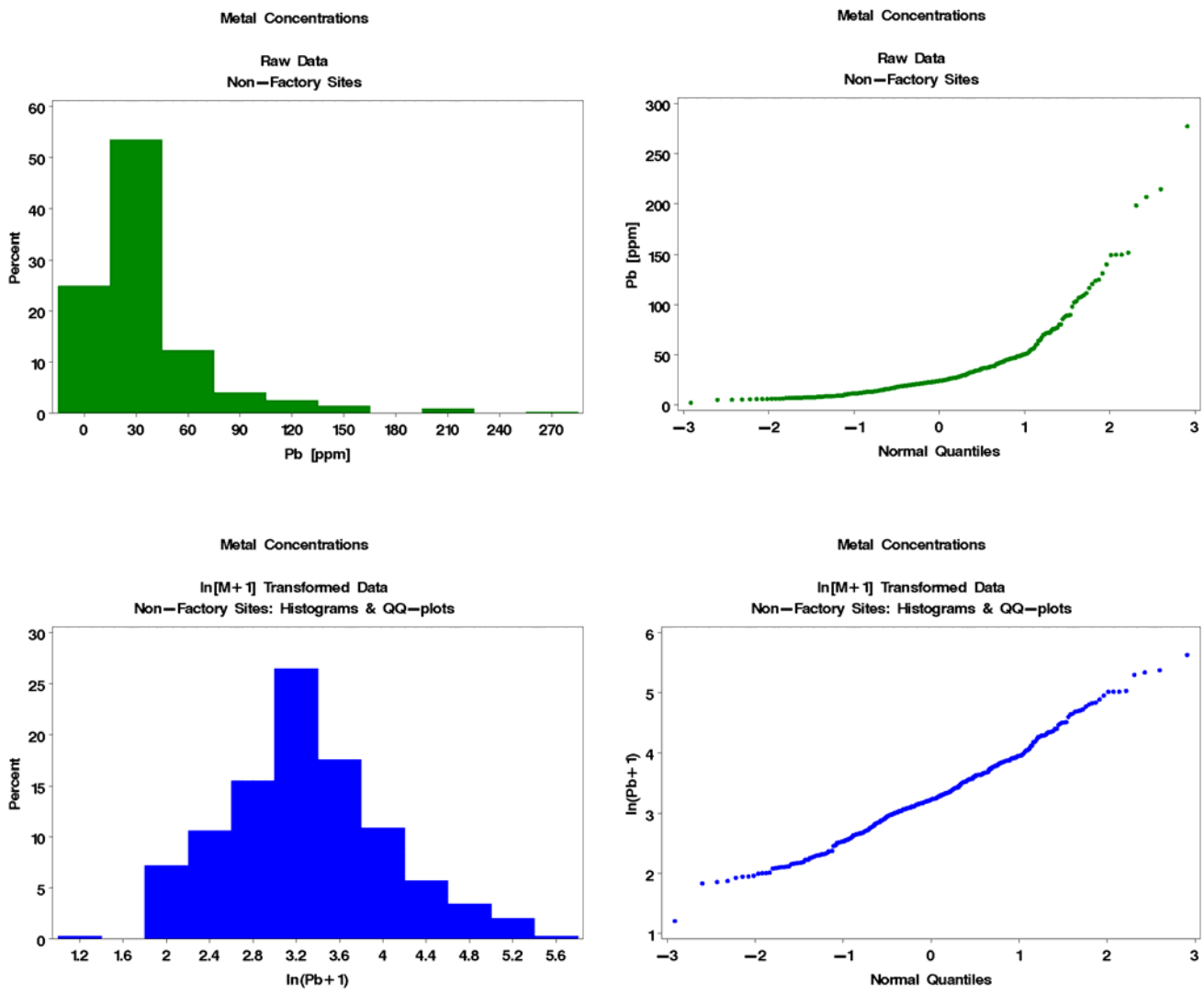
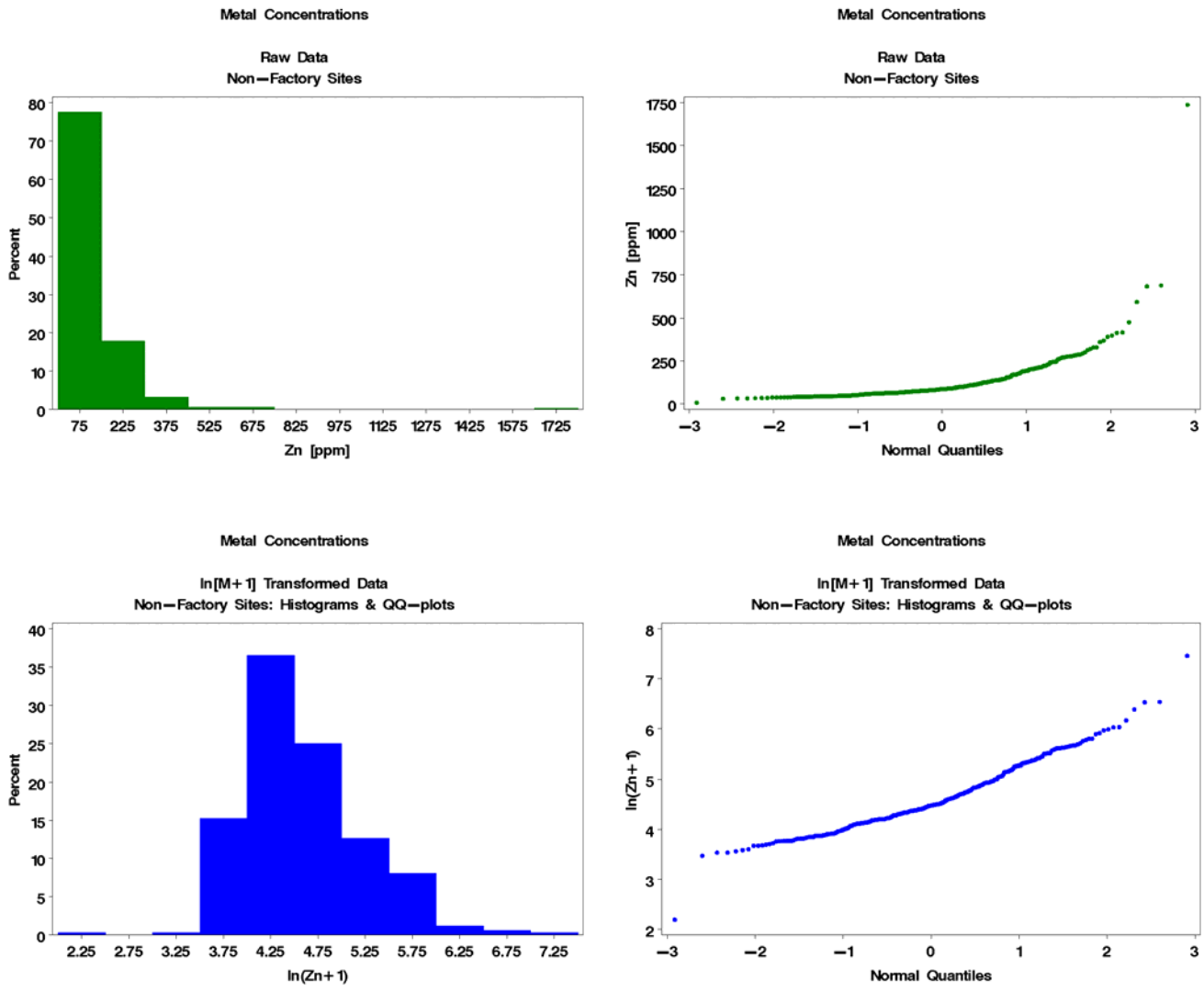


Figure 3.7 Panel graphs for Zn; histograms and quantile plots.



4.0 Analysis of the Sampling Depth Effect

A side issue of interest in this study was the effect of sampling depth on the measured metal concentration levels. To address this issue, soil samples from two distinct sampling depths (A:0-5 cm and B:5-10 cm) were acquired from six property locations visited during the first two rounds of sampling (i.e., five properties from strata A and one additional property from strata B). In all, 43 sites were sampled during this first round, yielding 43 locations with metal concentration levels for these two distinct depths.

Paired t-tests and sign-rank tests were employed to determine what, if any, effect the sampling depth had on the observed metal concentration levels (Hollander & Wolfe, 1999). The sign-rank tests were performed on the paired difference data for all seven metals; these test results are shown in Table 4.1. Only one of the seven tests produced a p-value < 0.05 (Zn). Furthermore, because seven simultaneous tests have been performed in this analysis, a Bonferonni corrected significance level of 0.05/7 = 0.0071 should be used to ensure an overall 0.05 type I error rate (Johnson & Wichern, 1988). Note that the observed p-value for the Zn difference (0.0362) is larger than 0.0071, suggesting that none of the sign-rank tests are statistically significant.

In addition to the sign-rank tests, paired t-tests were also performed on the six metals for which the ln(M+1) transformation induced approximate symmetry (i.e., all metals except Cd). These test results are shown in Table 4.2. Again, only one of the six tests produced a p-value below 0.05 (Cu: p=0.0367) and this p-value was larger than the Bonferonni corrected significance level of 0.05/6 = 0.0083. Additionally, a multivariate Hotelling's T test computed on all six metals was also non-significant (F=1.86, p=0.1141), further suggesting that none of the individual t-tests were statistically different from 0 (Press, 1981; Johnson & Wichern, 1988). These results, which agree with the sign-rank test results shown in Table 4.1, imply that the mean (and/or median) metal concentration levels were the same within each sampling depth.

In addition to the mean comparison tests, depth A versus B correlation estimates were calculated for each on the six ln[M+1] transformed metals. These correlation estimates and associated p-values are shown in Table 4.3. Surprisingly, both Cr and Ni exhibit no statistically significant correlation across depths. Additionally, As and Cu exhibit only moderate positive correlation; only two of the six metals (Pb and Zn) appear to be strongly correlated.

Table 4.1 Sign-rank test results for all seven metals (n=43 paired samples, depths A versus B).

Variable	Label	Mean	sr-test	p-value
As_A		10.39		
As_B		9.31		
d_rAS	As difference	1.08	116.5	0.1620
Cd_A		1.48		
Cd_B		2.62		
d_rCd	Cd difference	-1.13	-44.5	0.3196
Cr_A		53.24		
Cr_B		56.66		
d_rCr	Cr difference	-3.42	-4.0	0.9622
Cu_A		70.18		
Cu_B		83.76		
d_rCu	Cu difference	-13.58	-105.0	0.2086
Ni_A		40.71		
Ni_B		48.40		
d_rNi	Ni difference	-7.70	-72.5	0.3876
Pb_A		592.87		
Pb_B		739.38		
d_rPb	Pb difference	-146.51	63.0	0.4533
Zn_A		498.28		
Zn_B		456.43		
d_rZn	Zn difference	41.86	172.0	0.0362

Table 4.2 Univariate t-test results for the six log transformed metals (n=43 paired samples, depths A versus B).

Variable	Label	Mean	t-test	p-value
lnAs_A		2.063		
lnAs_B		1.849		
ld_rAS	ln(As+1) difference	0.214	1.506	0.1397
lnCr_A		3.681		
lnCr_B		3.626		
ld_rCr	ln(Cr+1) difference	0.055	0.248	0.8056
lnCu_A		3.616		
lnCu_B		4.041		
ld_rCu	ln(Cu+1) difference	-0.425	-2.158	0.0367
lnNi_A		3.445		
lnNi_B		3.630		
ld_rNi	ln(Ni+1) difference	-0.186	-1.043	0.3031
lnPb_A		4.904		
lnPb_B		4.842		
ld_rPb	ln(Pb+1) difference	0.061	0.343	0.7334
lnZn_A		5.407		
lnZn_B		5.039		
ld_rZn	ln(Zn+1) difference	0.368	1.862	0.0696

Table 4.3 Depth correlation estimates and p-values for the six log transformed metals (n=43 paired samples, depths A versus B).

Metal	Correlation (p-value)
ln[As+1]	0.4988 (p=0.0007)
ln[Cr+1]	0.0584 (p=0.7097)
ln[Cu+1]	0.3119 (p=0.0418)
ln[Ni+1]	0.0673 (p=0.6680)
ln[Pb+1]	0.8046 (p=0.0001)
ln[Zn+1]	0.6245 (p=0.0001)

5.0 Exploratory Spatial Data Analysis Plots

The degree of spatial structure apparent in the metal concentration data was assessed using two specific graphical techniques; quantile maps and (robust) variogram plots. A quantile map is simply a color-coded sampling map, where the symbols identify the sampling locations and their corresponding colors indicate the magnitude of the response variable. In this study, each metal concentration measurement was classified into one of 4 distinct quantile ranges: q0 – q25, q25 – q50, q50 – q75, and q75 – q100. Note that the 25%, 50%, and 75% percentile cut-off values for each metal are shown in Table 5.1.

Variogram plots are used extensively in geostatistical applications (Cressie, 1991; Wackernagel, 1998). Empirical variogram plots are used to show the degree of spatial similarity between data observations separated x-units apart. Usually (if the data are spatially correlated), the pattern (i.e., calucated variance) in the variogram plot initially increases as the separation distance increases, but then levels off (once the separation distance exceeds the maximum range of spatial correlation). Thus, in theory, variogram plots can be used to show the degree and range of spatial structure apparent in the data. In this study, robust variogram plots were constructed using the log transformed metal concentration data (Cressie, 1991).

The panel graphs shown in Figures 5.1 through 5.7 display the quantile and variogram plots for each individual metal, except for Figure 5.2 (which shows the quantile map only). Note that a meaningful variogram plot could not be constructed from the log transformed Cd data, due to the excessive number of zeros (non-detects).

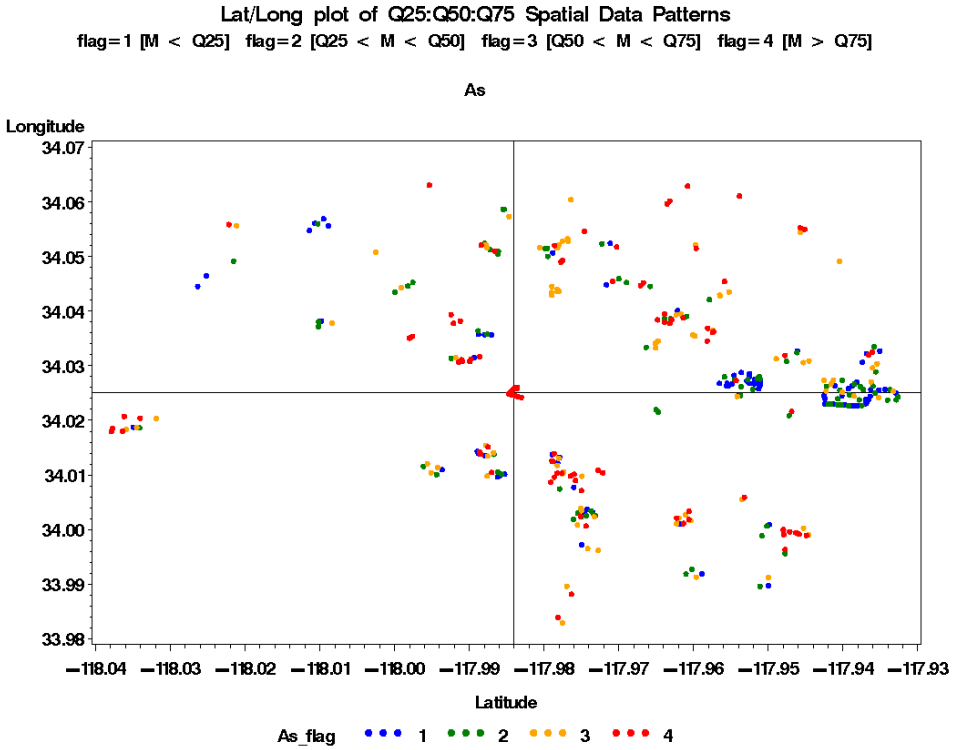
While the seven quantile maps suggest that there is a substantial amount of spatial variation in the metal concentration data over very short ranges, there is some hint that concentration levels of samples are correlated with their distance from the plant. Most of these maps also suggest that distinct property effects may also be present; i.e., samples gathers from within one property may be more similar (less variable) than samples

gathered from different properties. Perhaps not surprisingly, the majority of the variogram plots appear to be rather noisy and no consistent spatial structure appears to be present. Four of the six variogram plots (As, Cu, Pb, and Zn) suggest the presence of two distinct variance components, since the variograms appear to abruptly increase after the 0.5 Km distance. This effect could reflect a compound symmetric variance structure (i.e., within versus between property effects), or it may just be an artifact of the highly clustered sampling design. Overall, these variogram plots do not appear to be especially informative, except that they tend to confirm that the metal concentration data is indeed highly variable over very short distances.

Table 5.1 Percentile (q25, q50, q75) cut-off values used in the construction of the quantile maps (shown in Figures 5.1 through 5.7).

Variable	25th Pctl (ppm)	50th Pctl (ppm)	75th Pctl (ppm)
As [ppm]	2.94	3.77	5.70
Cd [ppm]	0.32	0.51	0.69
Cr [ppm]	13.38	17.00	20.85
Cu [ppm]	16.91	21.90	28.21
Ni [ppm]	10.39	12.69	15.83
Pb [ppm]	15.02	24.28	40.95
Zn [ppm]	64.30	87.13	139.81

Figure 5.1 Exploratory spatial plots (quantile map and variogram) for As.



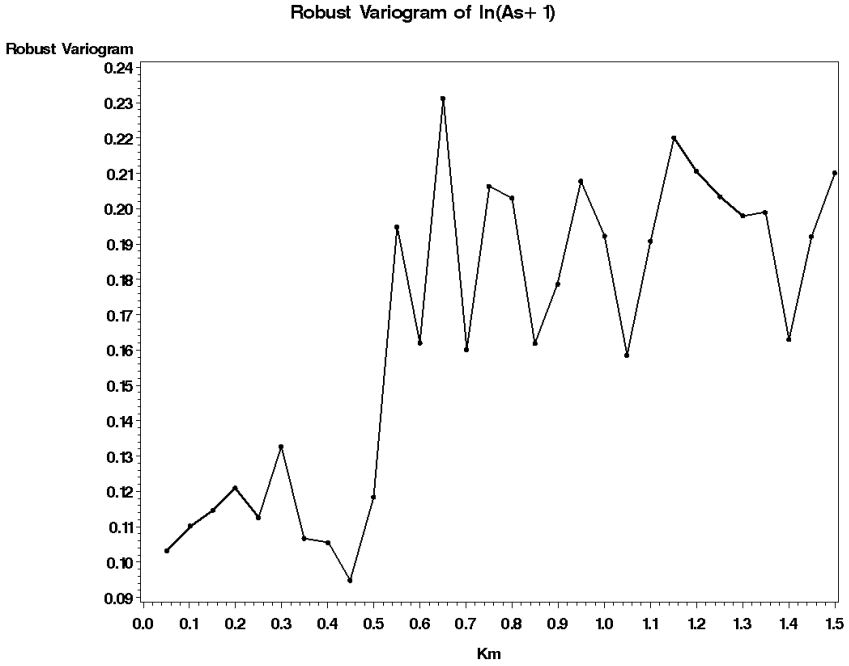


Figure 5.2 Exploratory spatial plots (quantile map only) for Cd.

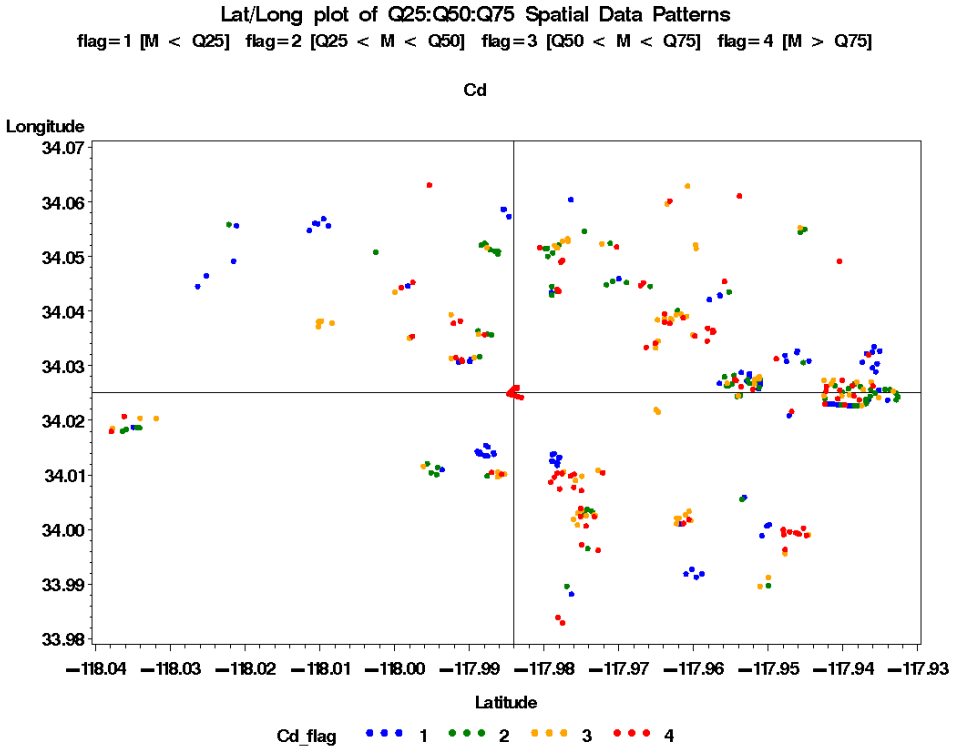
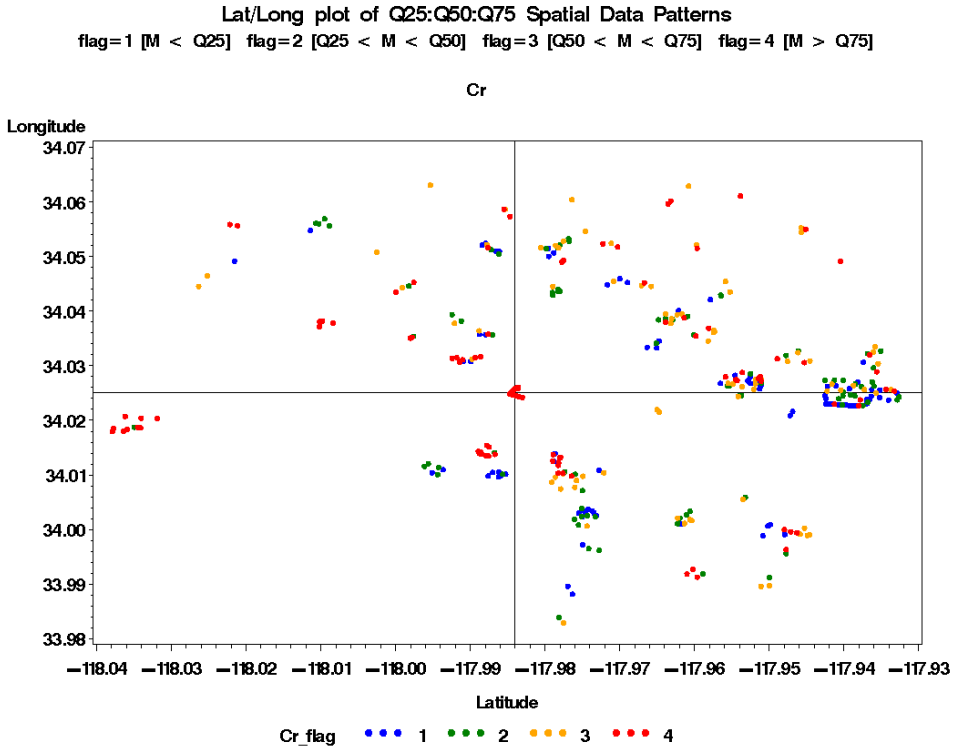


Figure 5.3 Exploratory spatial plots (quantile map and variogram) for Cr.



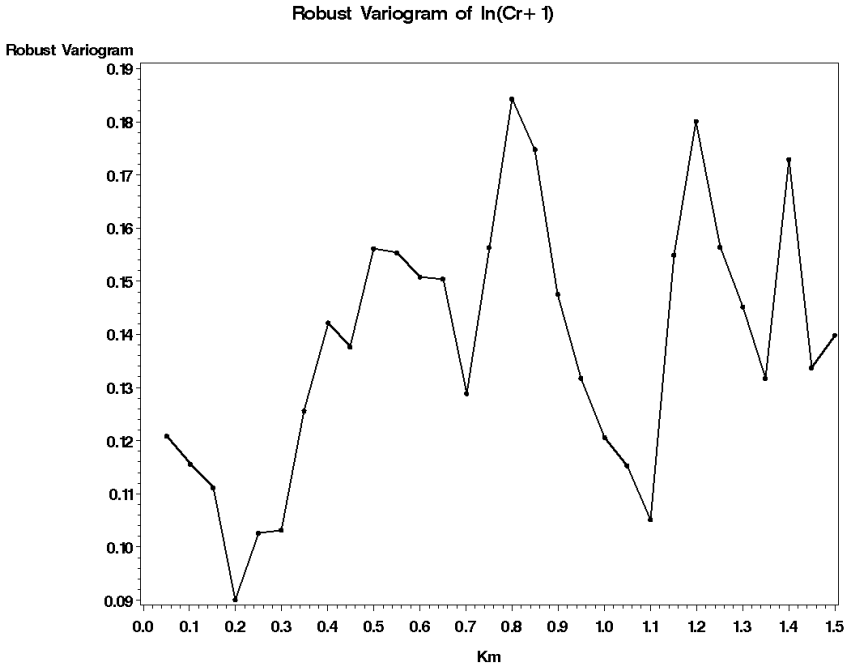
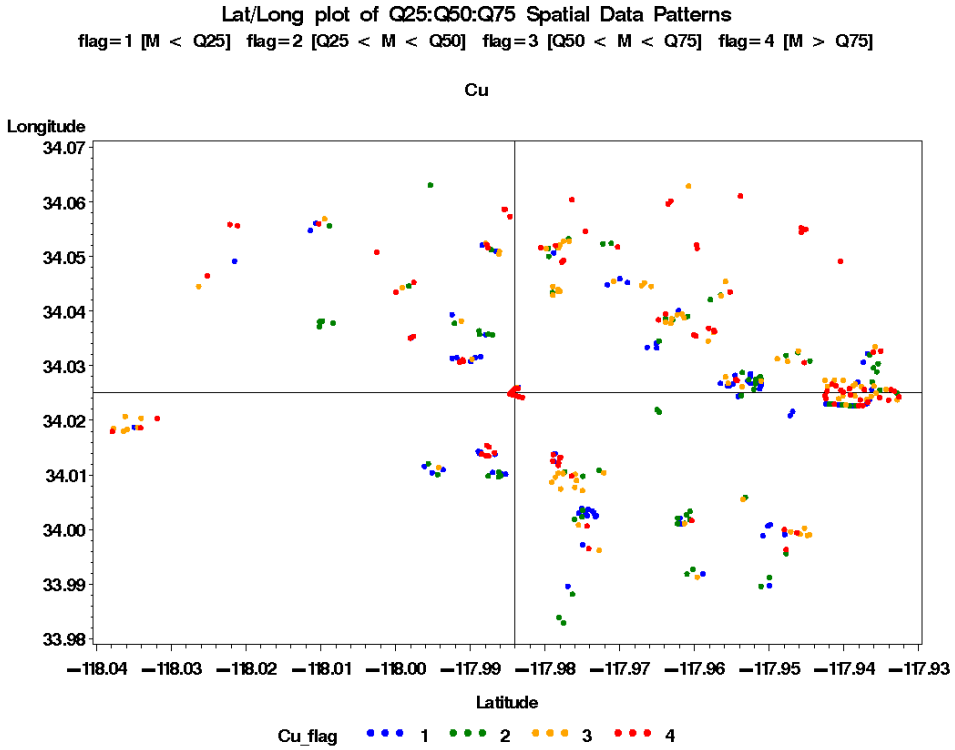


Figure 5.4 Exploratory spatial plots (quantile map and variogram) for Cu.



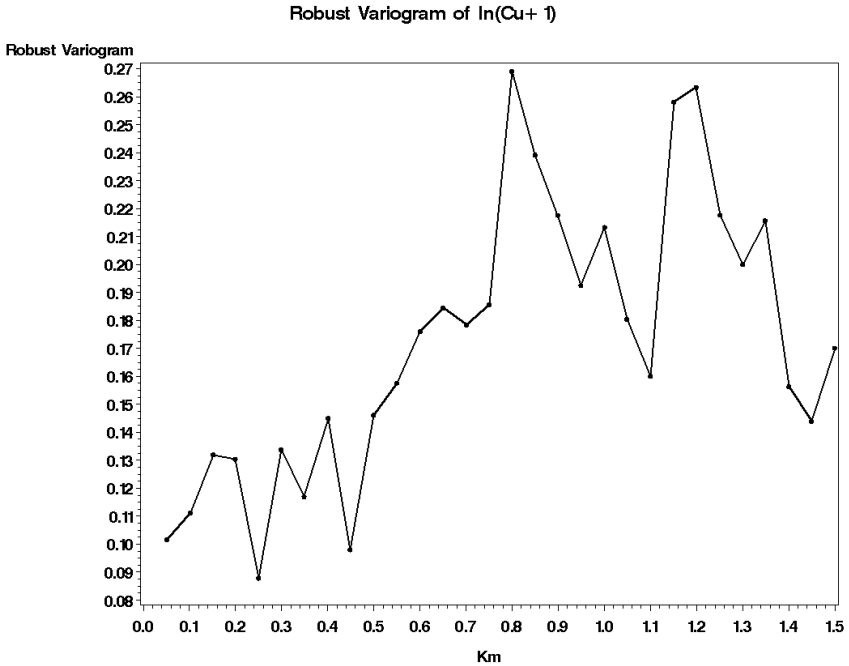
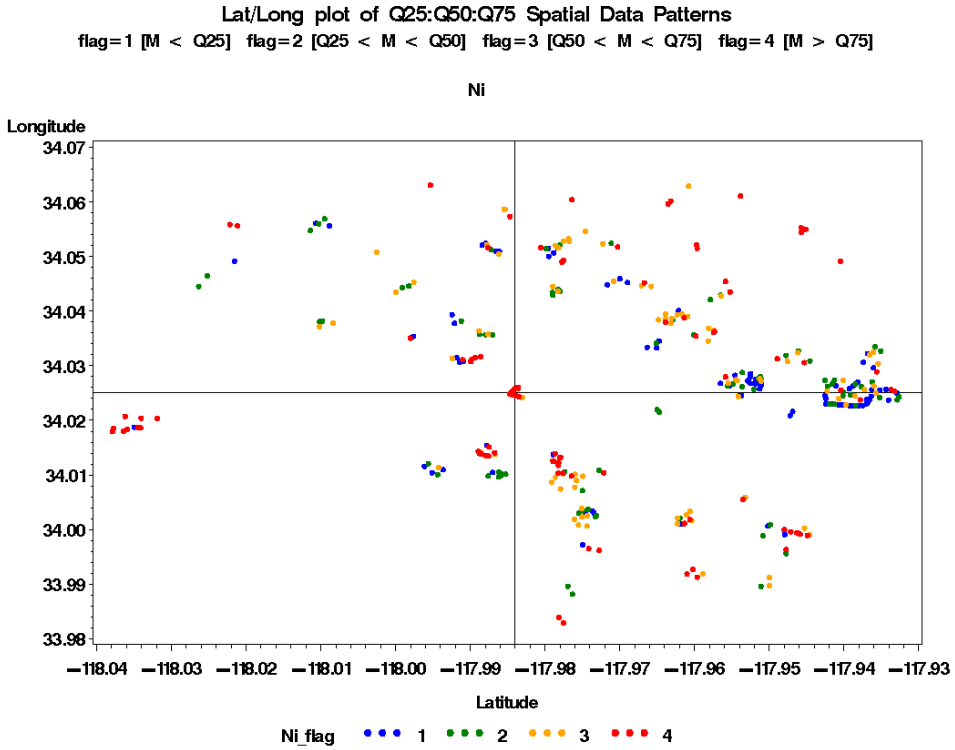


Figure 5.5 Exploratory spatial plots (quantile map and variogram) for Ni.



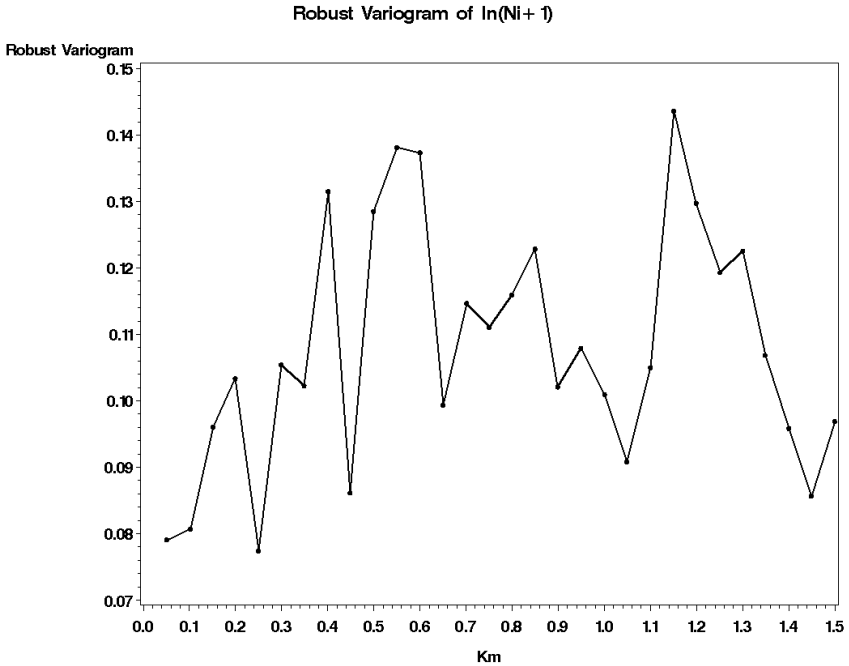
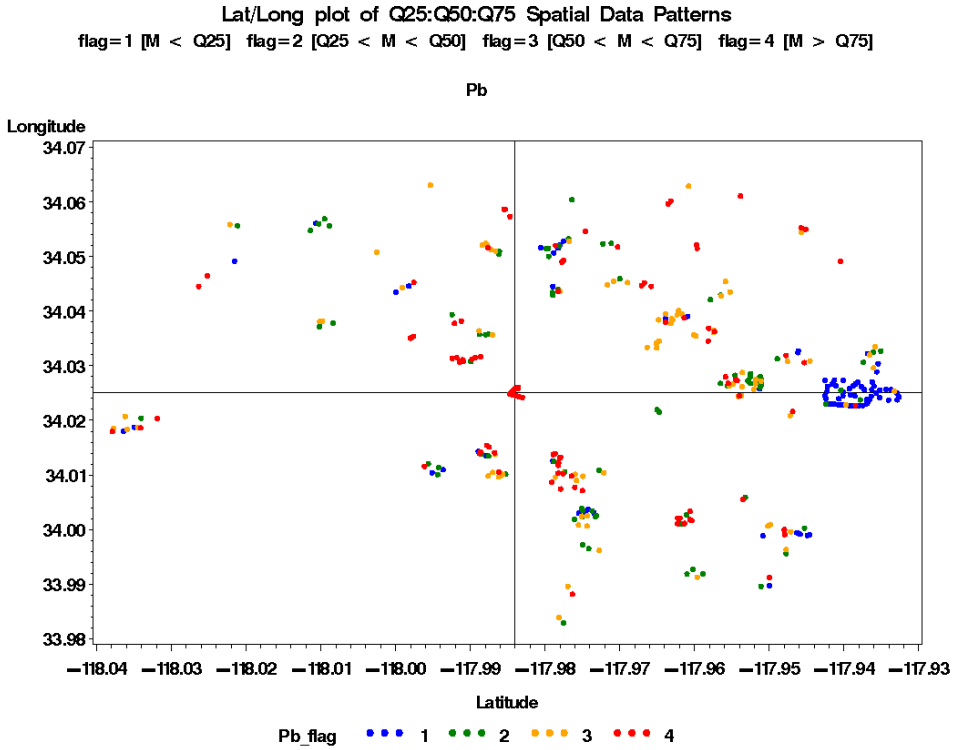


Figure 5.6 Exploratory spatial plots (quantile map and variogram) for Pb.



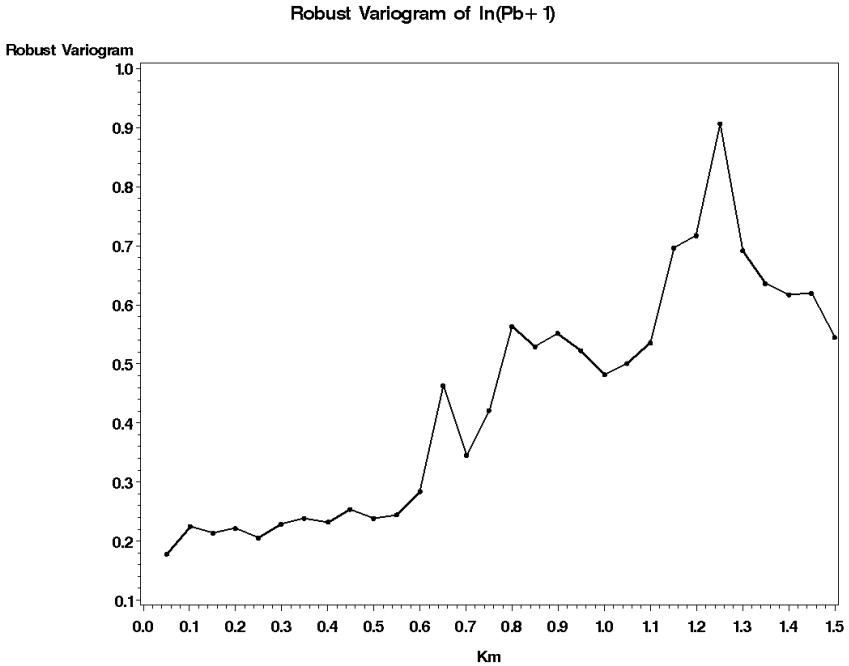
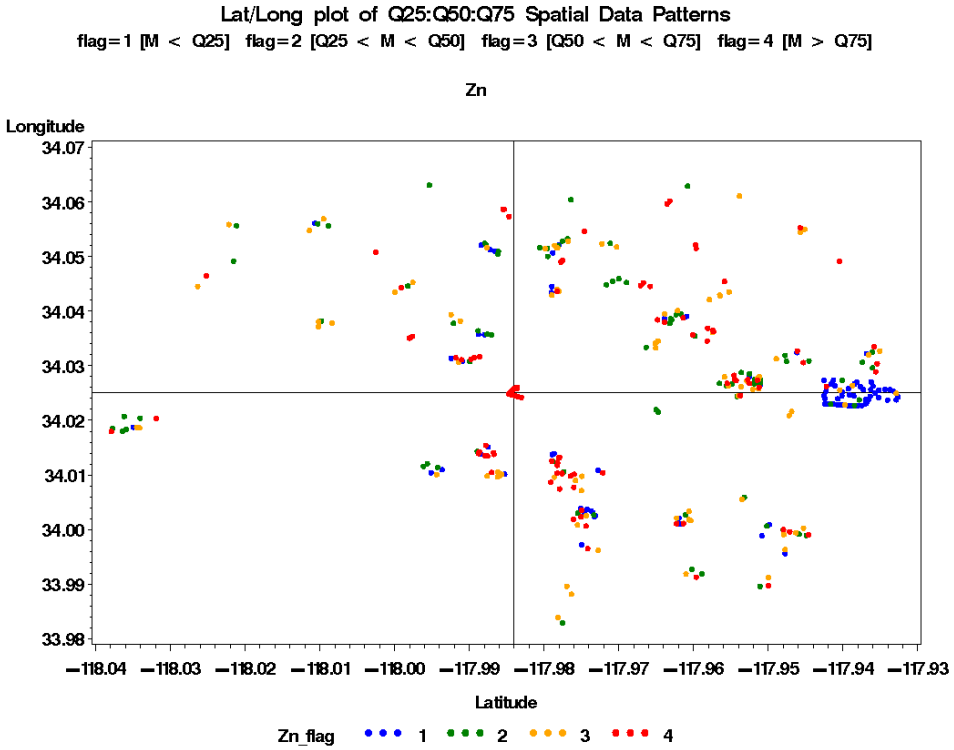
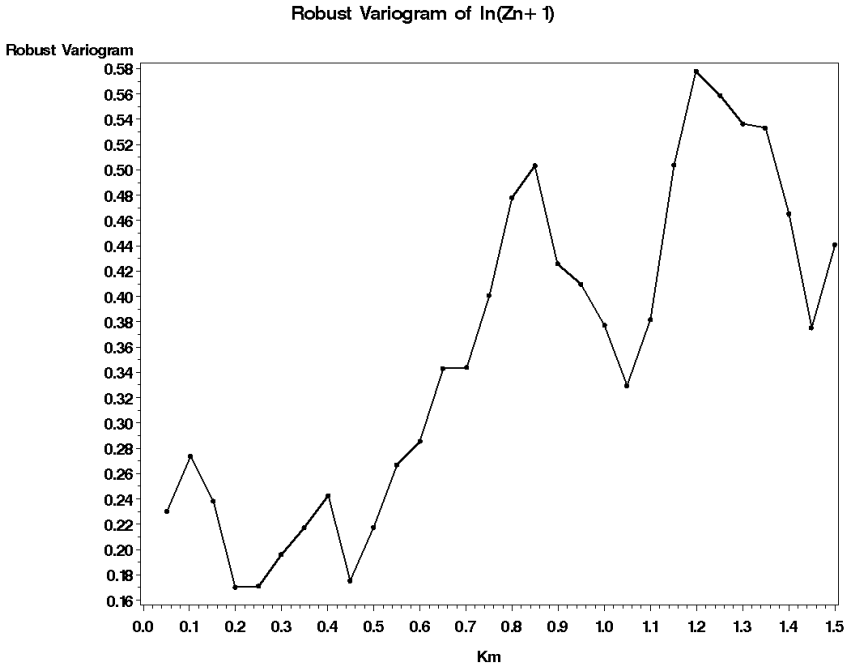


Figure 5.7 Exploratory spatial plots (quantile map and variogram) for Zn.





6.0 Quantile Indicator Maps and Tests of Association

A quantile indicator (QI) map can be used to detect non-random patterns in the magnitudes of spatially referenced sample data, specifically with respect to a hypothesized location of point-source contamination. The construction of a QI map is fairly simple; a cut-off quantile of interest is selected (such as the median, or 90% percentile, etc.) and all of the sampling locations that exceed this quantile are high-lighted. If a point-source contamination location is suspected, one simply inspects the QI map to see if an abnormally high number of high-lighted locations appear to occur within close proximity to the point-source.

The above idea can be translated into a formal hypothesis test as follows. First, define a region surrounding the point-source (usually a circle of constant radius) and classify the sampling locations accordingly (i.e., either contained or not contained within this region). Next, perform a Chi-square test of association on the resulting 2 by 2 table of Location versus Quantile score; note that significant positive association implies that sample sites closer to the point-source tend to exhibit a higher probability of exceeding the chosen cut-off level (Agresti, 2003).

Figures 6.1 through 6.7 show the QI maps for As, Cd, Cr, Cu, Ni, Pb, and Zn, respectively, using the 90% percentile as the cut-off quantile of interest. (Note that the factory sites are not included in any of these plots, or any of the following statistical analyses.) Four of these plots show evidence of non-random, elevated patterns around the factory location. Table 6.1 displays the Chi-square table and test results for Cr (Figure 5.3), assuming a 2 Km circular influence region around the factory. This likelihood ratio Chi-square test statistic of 31.9 is highly significant ($p < 0.0001$).

Table 6.2 summarizes 14 tests of association performed on these seven metals (using median [50% percentile] and 90% percentile cut-offs). The percentages of sites exhibiting contamination levels above these cut-offs are shown for both regions (within 2 Km versus > 2 Km), along with the individual p-values for each corresponding likelihood

ratio Chi-square test. The corresponding test results that are statistically significant at the Bonferonni corrected 0.05 and 0.01 significance levels are also high-lighted in Table 6.2. These results imply that at the corrected 0.01 significance level, an excessive number of Pb samples near the factory exceed both the median and q90 cut-offs and an excessive number of Cr and Ni samples exceed the q90 cut-off. Hence, an abnormally high number of "hot" (i.e., contaminated) Cr, Ni, and Pb samples appear to occur within close proximity (< 2 Km) to the factory location.

Table 6.1 Chi-square association table and likelihood test results for the q90 Cr sample pattern.

	Cr < q90 cut-off	Cr > q90 cut-off	# of Sample sites
Sample site > 2 Km from Factory	266 (95.0%)	14 (5.0%)	280
Sample site < 2 Km from Factory	47 (69.1%)	21 (30.9%)	68
# of Sample sites	313	35	348

Likelihood ratio Chi-square test score: 31.9 (p < 0.0001)

Table 6.2 Chi-square test results; all 7 metals using both median and q90 cut-offs.

	Median cut-off			q90 cut-off		
Metal	< 2 Km	> 2 Km	p-value	< 2 Km	> 2 Km	p-value
As	63.2%	46.8%	0.0145	14.7%	8.6%	0.1447
Cd	51.5%	49.6%	0.7869	7.4%	10.7%	0.3923
Cr	55.9%	48.2%	0.2562	30.9%	5.0%	0.0001 **
Cu	42.7%	52.1%	0.1595	20.6%	7.1%	0.0021 *
Ni	54.4%	48.9%	0.4170	29.4%	5.0%	0.0001 **
Pb	69.1%	45.4%	0.0004 **	27.9%	5.7%	0.0001 **
Zn	61.8%	47.1%	0.0299	14.7%	8.9%	0.1736

Notes: ** significant at Bonferonni corrected 0.01 level
* significant at Bonferonni corrected 0.05 level

Figure 6.1 QI-90 map for As.

Figure 6.2 QI-90 map for Cd.

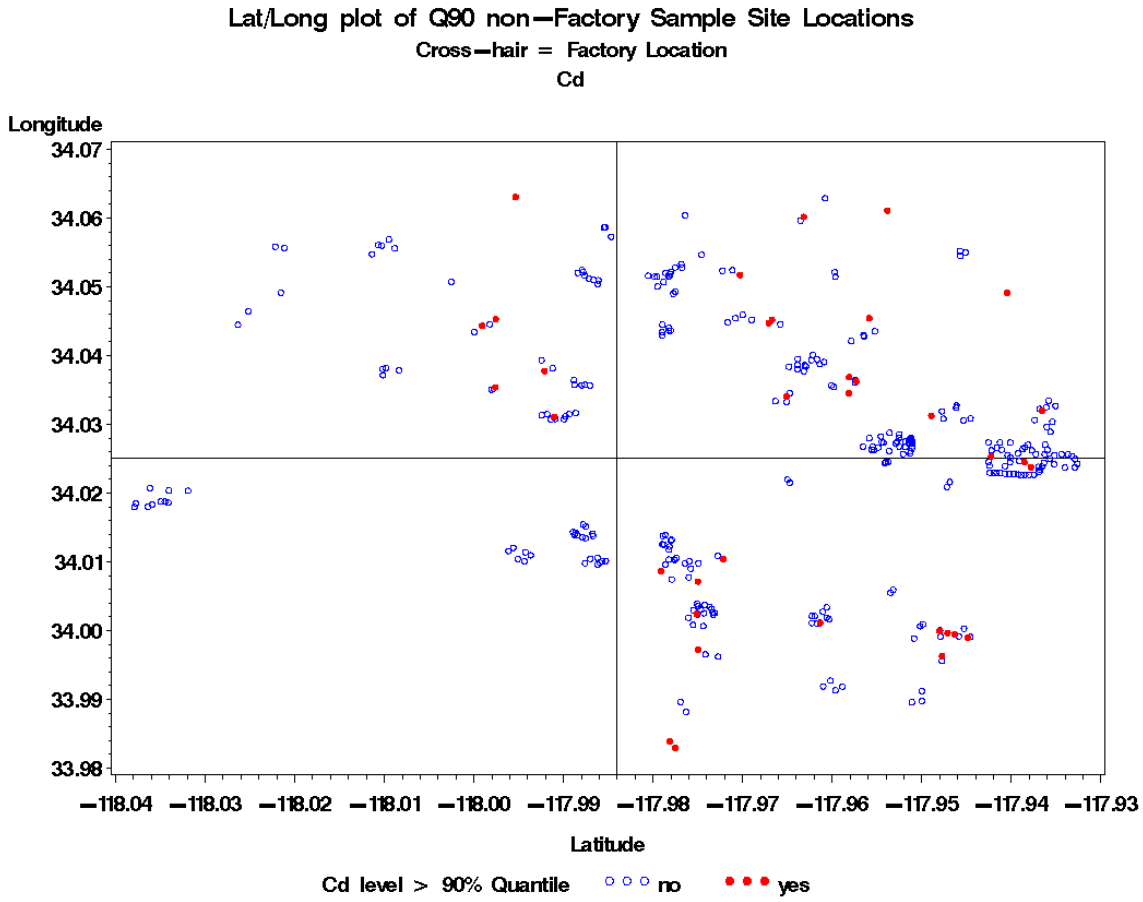


Figure 6.3 QI-90 map for Cr.

Figure 6.4 QI-90 map for Cu.

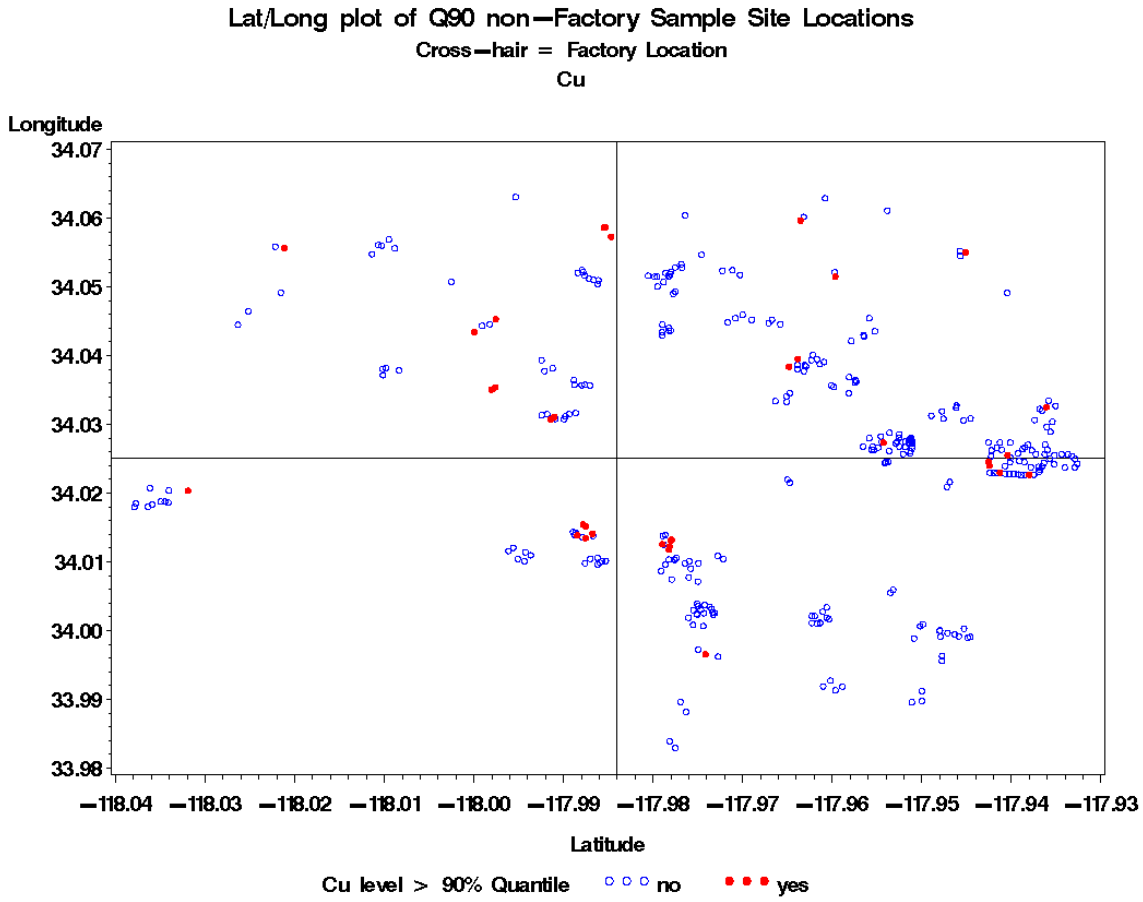


Figure 6.5 QI-90 map for Ni.

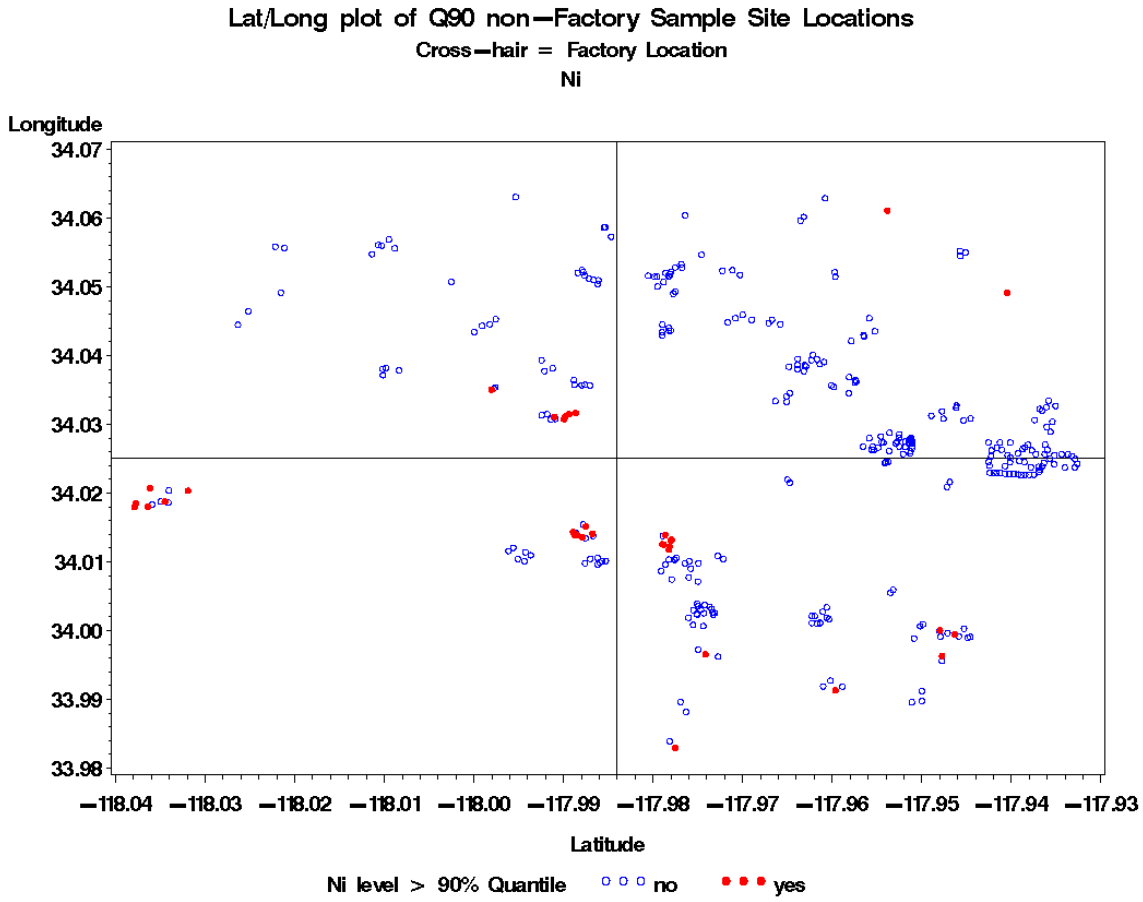
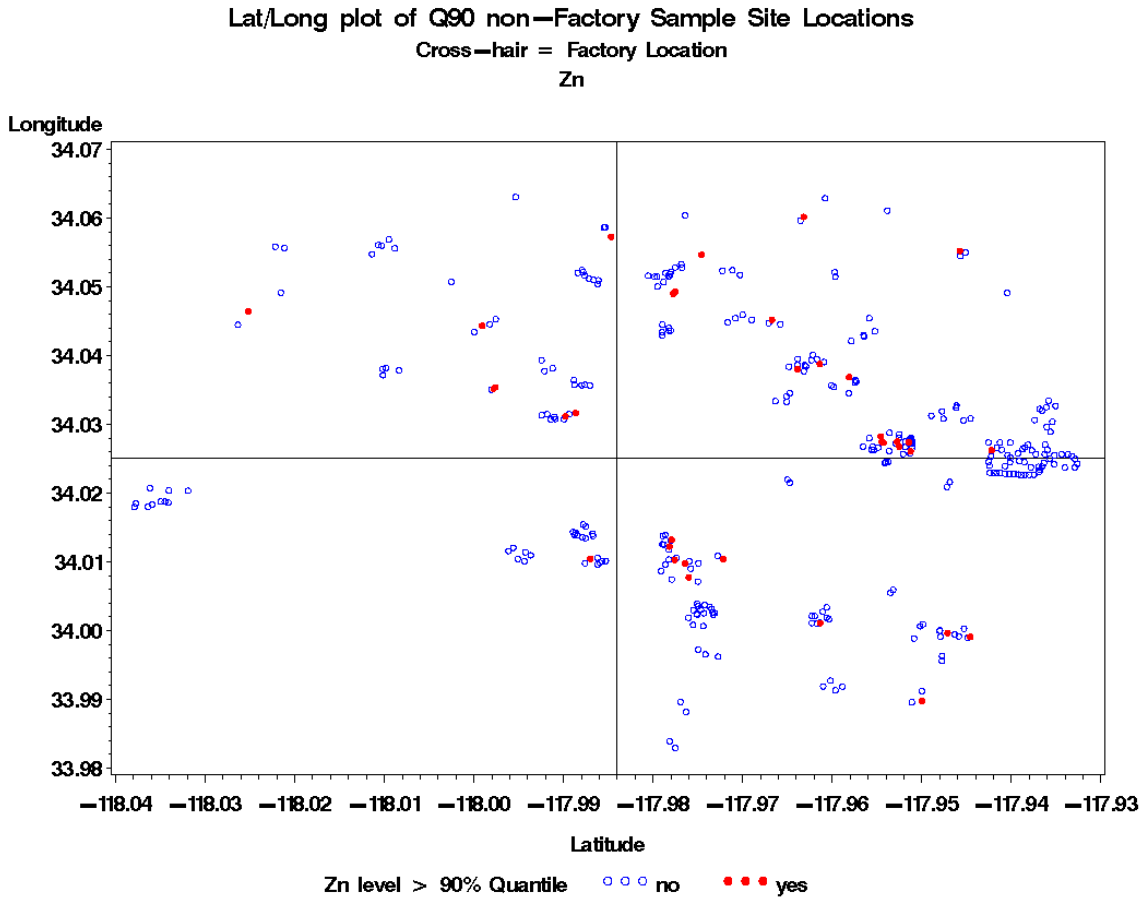


Figure 6.6 QI-90 map for Pb.

Figure 6.7 QI-90 map for Zn.



7.0 Contamination by Distance to Factory Plots

The primary goal of this study was to determine if (any of the seven) metal concentration levels tended to increase as the corresponding sampling locations got closer to the Quemetco factory site. In order to visually assess the degree of such trends, the natural log transformed contamination levels for each metal were plotted as a function of the distance (of each sample site) to the factory and then a smoothed spline was fitted to the resulting contamination pattern. These plots, referred to here as CD2F plots, are shown in Figure 7.1 through 7.7.

In each plot, the 348 non-factory sites are shown as open green circles and the 12 samples collected around the factory perimeter are shown as solid blue circles. The spline function displays the smoothed trend (if any) in the non-factory sites with respect to the distance from the factory effect; note that the factory perimeter sites were not used in the calibration of this spline function.

The CD2F plots associated with the metals Cr, Ni, and Pb display fairly clear evidence of an increasing contamination trend towards the factory. In each case, the increasing trend tends to occur on or about 2 Km away from the factory and appears to be approximately linear (on the log transformed scale). The CD2F plots for the remaining four metals (As, Cd, Cu, and Zn) display no evidence of any increasing trends. However, in all seven plots the contamination levels associated with the factory sites appear to be significantly elevated (as compared to the non-factory sites). Tables 7.1 shows the calculated log mean (and back-transformed median) factory contamination estimates for all seven metals, and contrasts these with the estimated means and medians for all samples greater than 2 Km away from the factory. These differences are substantial, and in all seven cases produced statistically significant t-tests and sign-rank tests (significant at the 0.0001 level).

Note that the shape of the Cr, Ni, and Pb CD2F spline plots suggest that a linear spline model might be used to model (and hence test for) the distance to factory effect for these three metals. This idea is developed in detail in the next section.

Table 7.1 Calculated mean and back-tansformed median contamination estimates for (i) factory perimeter samples, and (ii) all non-factory samples > 2 Km from the factory (for all seven metals).

Metal	location	Mean ln[M+1]	Std Error	Median	95% CI
As	Factory	2.890	0.136	17.99	(13.7, 23.6)
	> 2 Km	1.634	0.028	5.12	(4.8, 5.4)
Cd	Factory	1.382	0.095	3.98	(3.3, 4.8)
	> 2 Km	0.434	0.020	1.54	(1.5, 1.6)
Cr	Factory	4.198	0.102	66.57	(54.3, 81.6)
	> 2 Km	2.873	0.021	17.69	(17.0, 18.5)
Cu	Factory	4.560	0.131	95.59	(73.6, 124.2)
	> 2 Km	3.165	0.027	23.70	(22.4, 25.0)
Ni	Factory	4.018	0.091	55.57	(46.3, 66.7)
	> 2 Km	2.591	0.019	13.35	(12.9, 13.9)
Pb	Factory	7.059	0.202	1163.63	(776.4, 1743.9)
	> 2 Km	3.174	0.042	23.91	(22.0, 26.0)
Zn	Factory	6.818	0.177	914.43	(641.7, 1303.1)
	> 2 Km	4.556	0.037	95.24	(88.5, 102.5)

Figure 7.1 Contamination by distance to factory plot: ln[As+1] sample data.

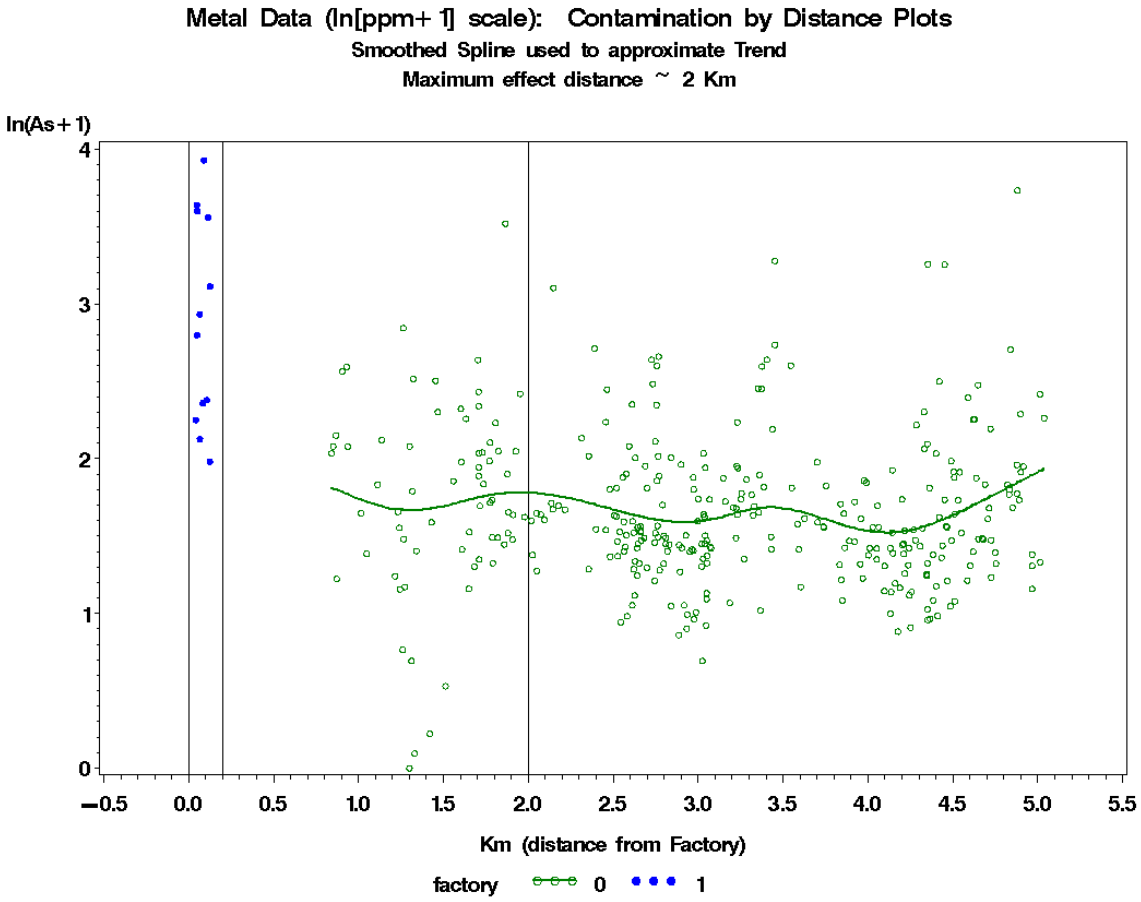


Figure 7.2 Contamination by distance to factory plot: ln[Cd+1] sample data.

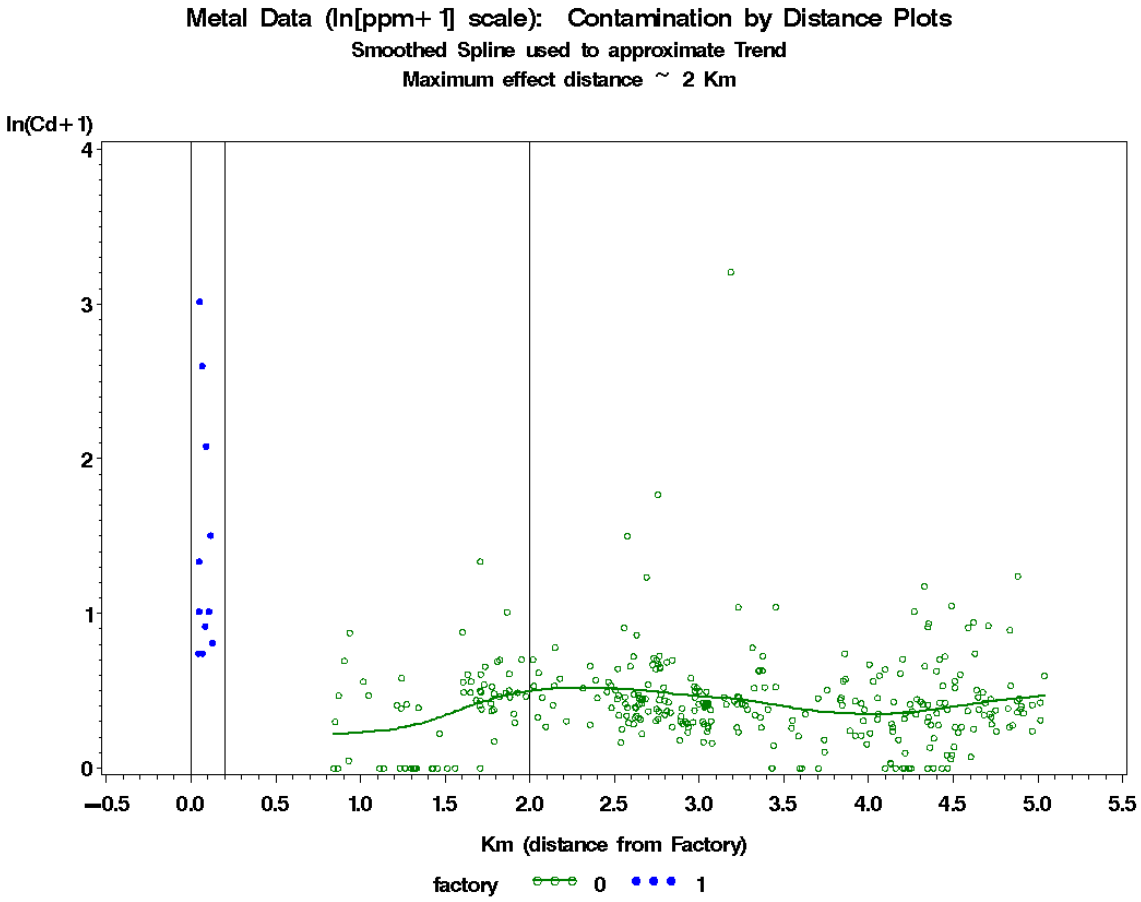


Figure 7.3 Contamination by distance to factory plot: ln[Cr+1] sample data.

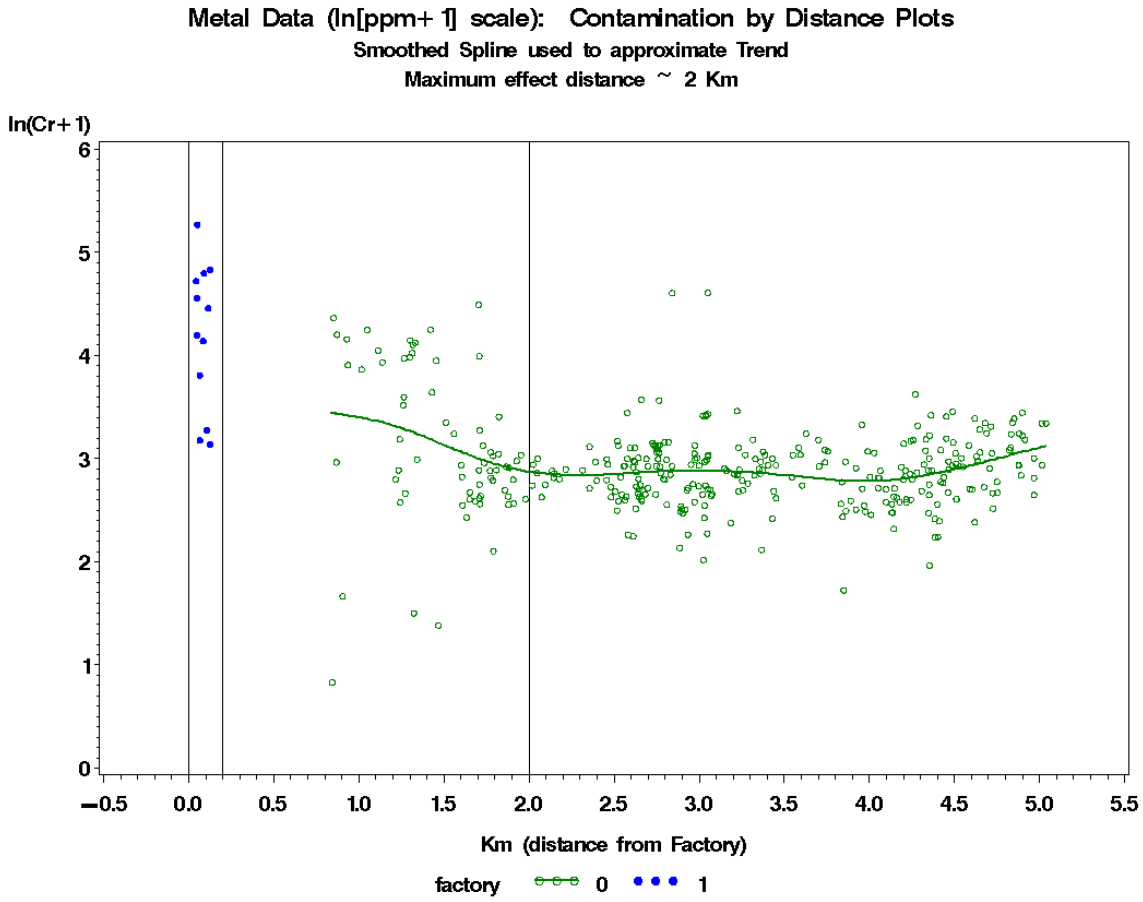


Figure 7.4 Contamination by distance to factory plot: ln[Cu+1] sample data.

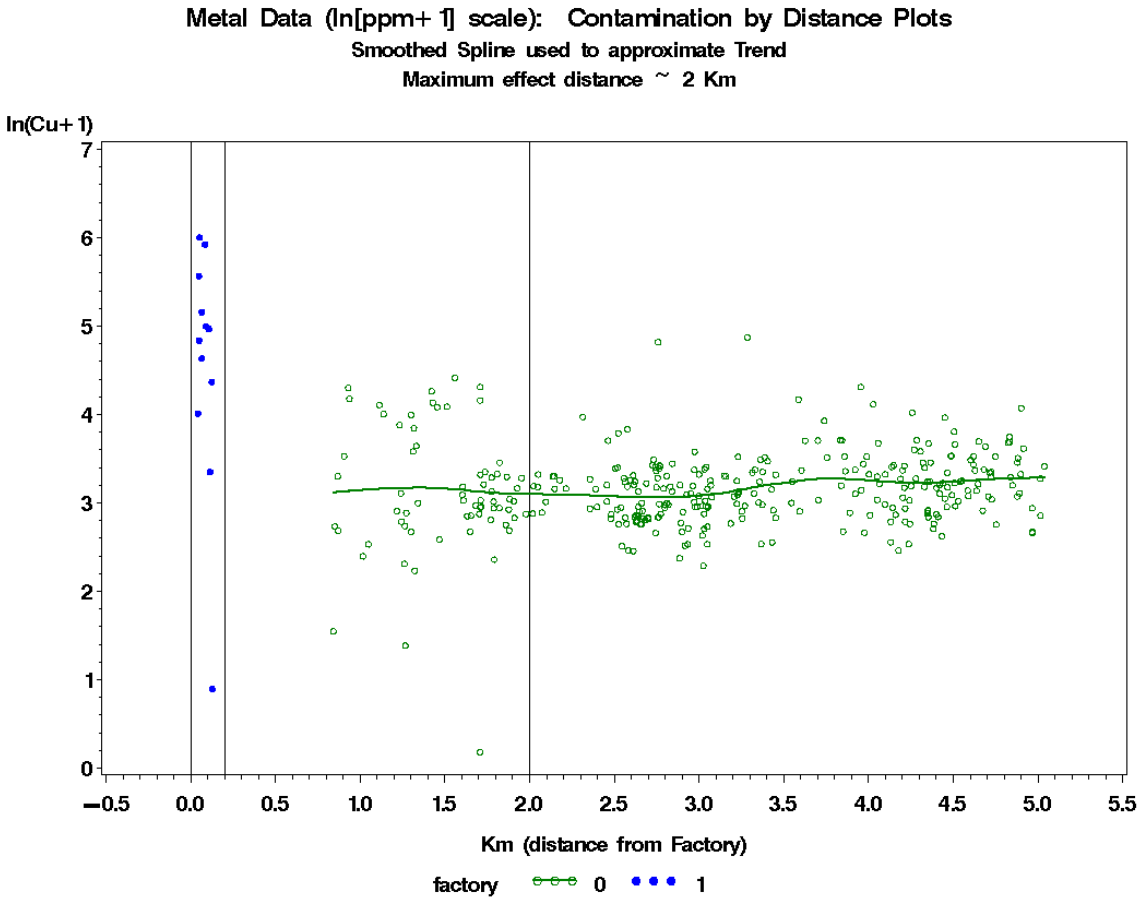


Figure 7.5 Contamination by distance to factory plot: ln[Ni+1] sample data.

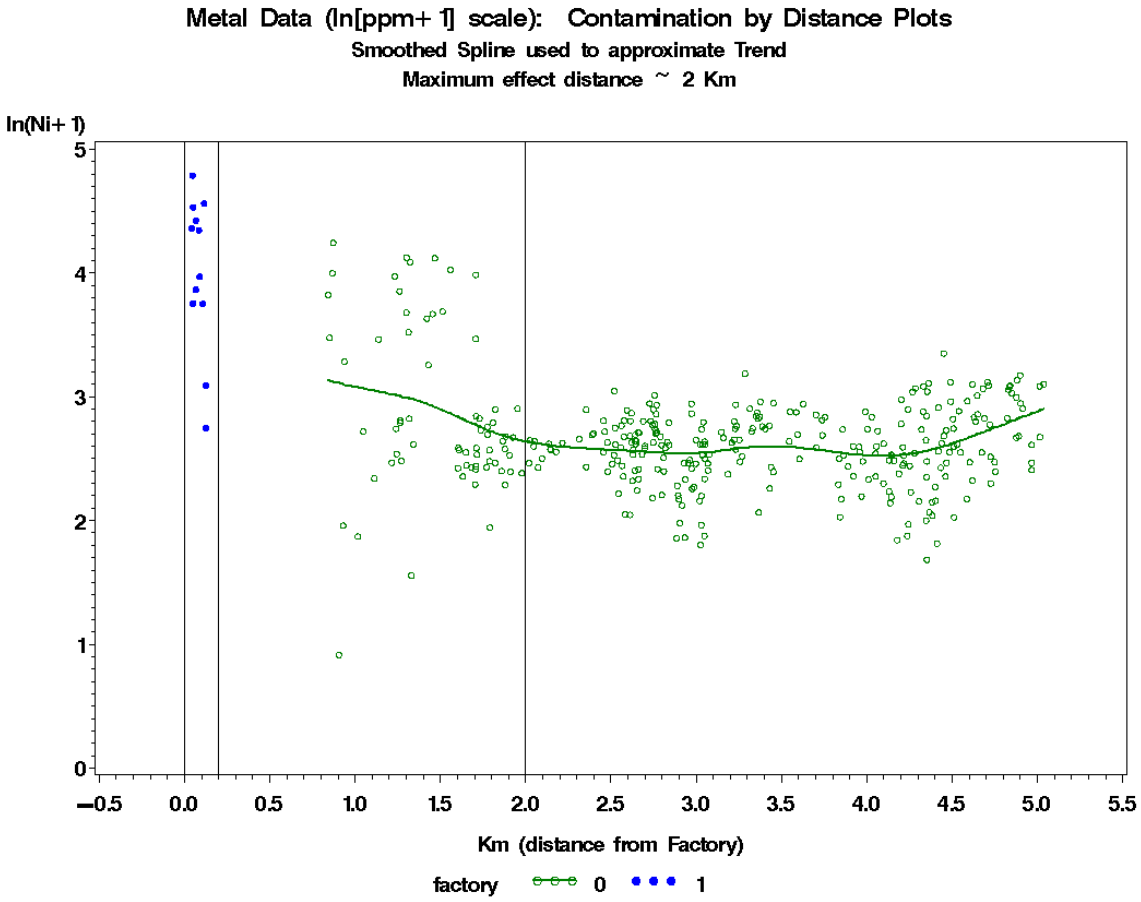


Figure 7.6 Contamination by distance to factory plot: ln[Pb+1] sample data.

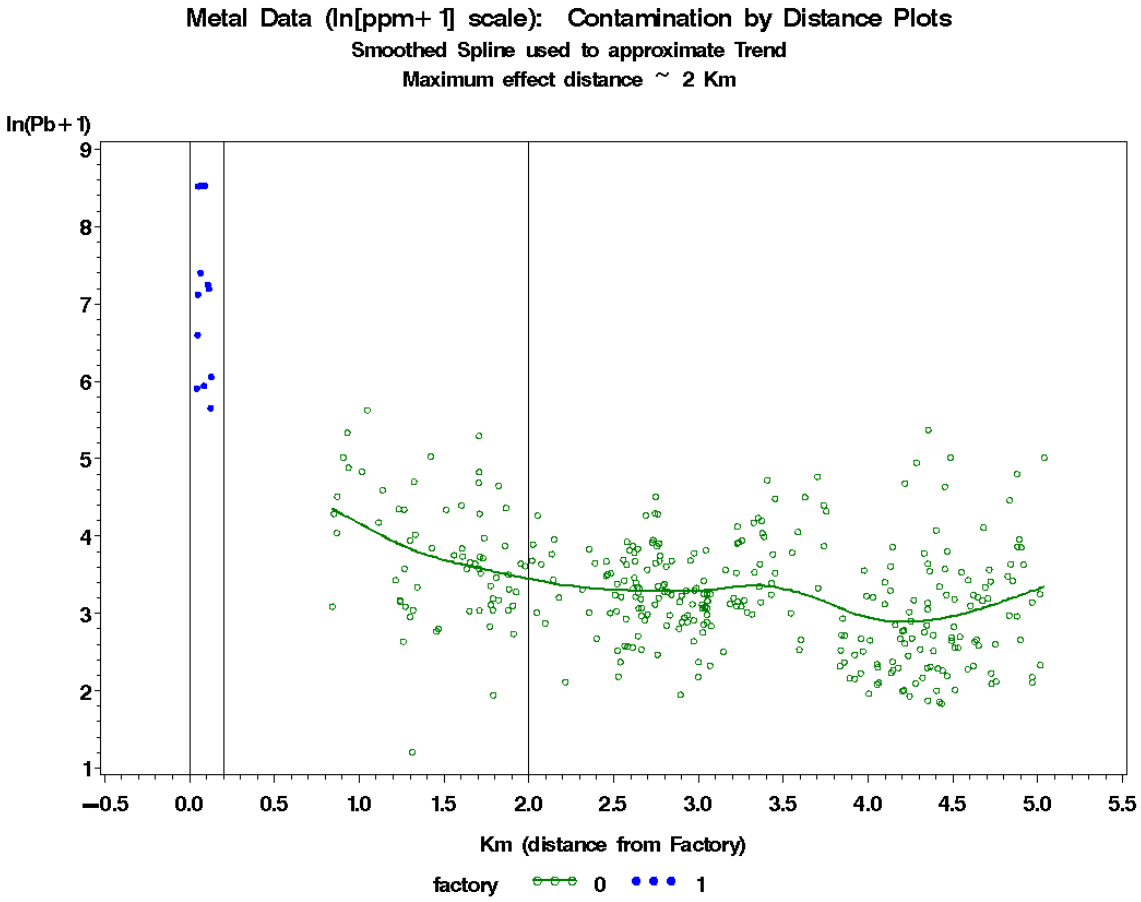
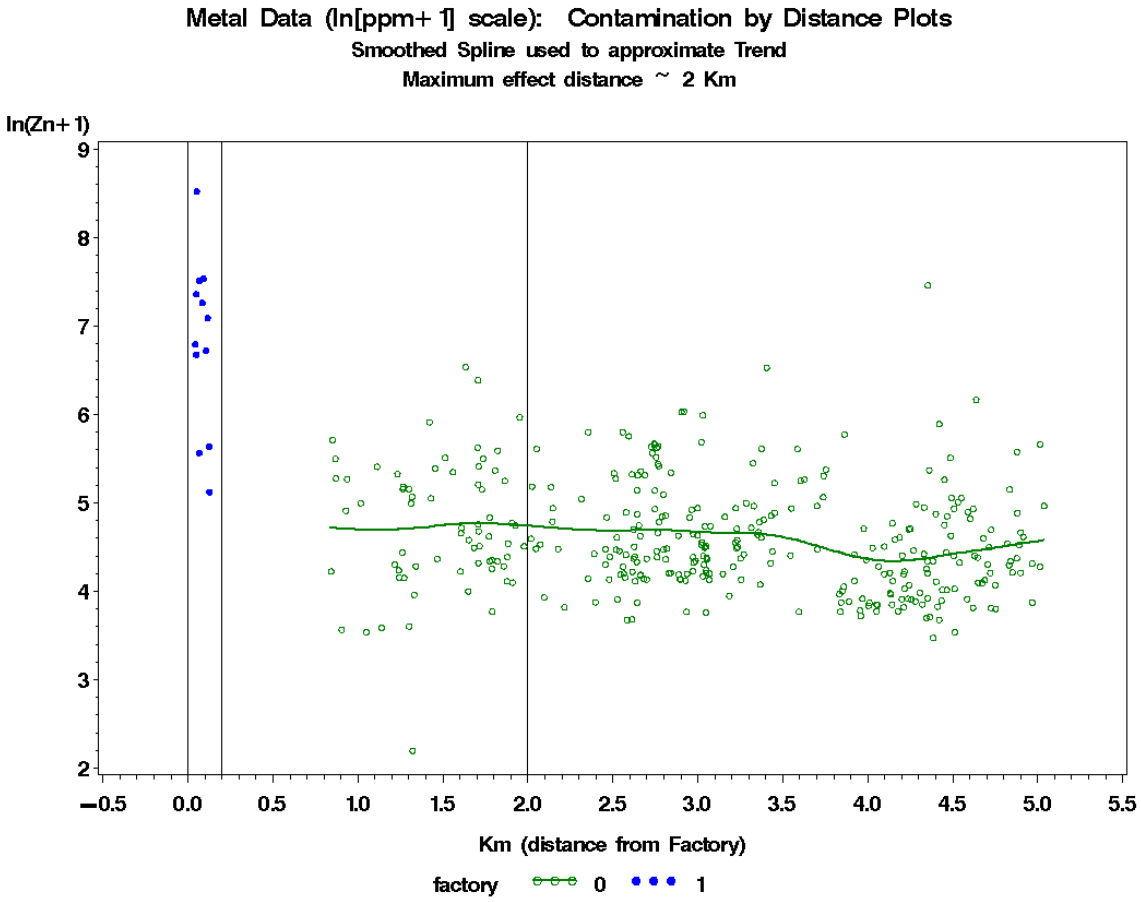


Figure 7.7 Contamination by distance to factory plot: ln[Zn+1] sample data.



8.0 Linear Spline Models

As alluded to in section 7, the CD2F plots suggest that a linear spline model might be used to test for distance to factory contamination effects for each metal examined in this study. As shown in Figure 8.1, a simple linear spline model suitable for describing such effects (and simultaneously adjusting for possible covariates) can be specified as follows:

$$y_i = \beta_0 + \beta_1(\min[d_i, T]) + \beta_2(f_i) + \{\text{extra covariates}\} + \eta_i$$

$$\text{for } f_i = \begin{cases} 1 & \text{if Factory site} \\ 0 & \text{otherwise} \end{cases} \qquad \text{Eqn [1]}$$

$$d_i = \text{distance of } i^{th} \text{ site to Factory}$$

where y represents the log transformed (ln[M+1]) metal contamination level and η represents a (possibly spatially dependent) random error component. In this linear spline model, the β_1 slope parameter captures the distance to factory contamination effect. If the average log contamination level in the soil increases as the distance to the factory decreases, then the β_1 slope parameter will be negative. Thus, a one-sided t-test can be performed to test for such a (distance to factory) effect.

This linear spline model is conditioned on an *a priori* specified threshold distance (T); no factory contamination effects are assumed to be present beyond this threshold distance. As such, assuming that all three parameters are statistically significant, the following three effects can be estimated from the final fitted equation:

Effect	*Estimate*	
Baseline	$\hat{\beta}_0 + \hat{\beta}_1(T)$	
Factory	$\hat{\beta}_0 + \hat{\beta}_2$	
Proximity	$\hat{\beta}_0 + \hat{\beta}_1(d_i)$	*for any* $d_i < T$

The *Baseline* effect estimates the background log contamination level across the survey region (i.e., the background level not influenced by the factory), the *Factory* effect estimates the log contamination level within or immediately around the perimeter of the factory, and the *Proximity* effect quantifies the distance to factory contamination relationship. Note that if the β_1 slope parameter is not found to be statistically significant, then we would conclude that there is no *Proximity* effect, etc.

In order to develop a realistic model, Eqn [1] should be modified to include any additional covariates that might be thought to influence the metal concentration levels (such as secondary proximity effects to other possible pollution sources, etc.). Additionally, the residual errors should be carefully examined for evidence of spatial correlation and an appropriately chosen error structure should be incorporated into the model to account for any detected correlation structure (Schabenberger & Gotway, 2005).

Based on the visual evidence seen in the CD2F plots, a threshold value of T = 2 Km was chosen for Eqn [1]. Additionally, in this study, information on two additional covariates were available at all sampling locations: (i) the altitude associated with the sampling position, and (ii) a 0/1 binary variable indicating if the sampling position was within 1 Km to any of the three major freeways (the 605, 60 and 10) running through the sampling region (1 if the sampling position is within 1 Km of any of the 3 freeways, 0 otherwise). Thus, both of these variables were initially incorporated into Eqn [1] as additional linear covariates, leading to a new regression model defined as:

$$y_i = \beta_0 + \beta_1(\min[d_i, 2]) + \beta_2(f_i) + \beta_3(alt_i) + \beta_4(fway_i) + \eta_i \qquad \text{Eqn [2]}$$

Next, following the standard approach for fitting mixed linear models, Eqn [2] was estimated under the assumption of different error structures (Davis, 2002; Schabenberger & Gotway, 2005). The following six error structures were used when estimating Eqn [2] (via maximum likelihood): (1) identical and independent [IID], (2) Isotropic spatial Exponential [SpExp], (3) Isotropic spatial Gaussian [SpGau], (4) Compound Symmetry [CS], (5) SpExp+CS, and (6) SpGau+CS. Note that the compound symmetry structure

was used to capture within versus between property effects (i.e., samples from the same property might exhibit more similarity than samples from different properties), even though this is not a "traditional" spatial error structure per se. Additionally, (5) and (6) represent composite spatial + compound symmetry error structures that are designed to adjust for both random property effects and spatially correlated residual errors, etc. All estimations were carried out using the MIXED procedure in SAS (SAS Institute Inc, 1999).

Table 8.1 displays the -2 log likelihood (-2LL) scores obtained under each of the six covariance structures for the six metals that could be examined using Eqn [2]. (The residual errors from the log transformed Cd data exhibited excessive asymmetry and non-Normality, violating the basic mixed linear modeling assumptions.) Somewhat surprisingly, the CS error structure produced the smallest -2LL score in all six regression models, typically much lower than either spatial error structure (both of which always included nugget effects). Additionally, in all six models both composite error structures converged to the corresponding CS structure (the spatial sill parameter converged to 0 in all cases). These -2LL scores imply that the CS error structure represented the most appropriate residual error assumption for Eqn [2], since the -2LL differences between the CS and IID structures were always statistically significant (at or below the 0.05 level). During this initial error structure analysis the freeway covariate parameter was also found to never be statistically significant. Thus, based upon these results, a final (revised) linear spline model was defined as:

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1(\min[d_i,2]) + \beta_2(f_i) + \beta_3(alt_i) + \theta_k + \varepsilon_{i(k)} \\
\theta_k &\sim iid\ \ N(0,\sigma_p^2) \\
\varepsilon_{i(k)} &\sim iid\ \ N(0,\sigma_s^2)
\end{aligned}
\qquad \text{Eqn [3]}
$$

Note that Eqn [3] represents a mixed linear spline model containing two error effects (random property effects, and random site within property effects).

Table 8.2 displays the β_1 parameter estimates, standard errors, t-tests and corresponding one-sided p-values obtained from the restricted maximum likelihood (REML) solutions to Eqn [3] for each metal. As shown in Table 8.2, the spline parameter estimates are highly significant for the Cr and Pb metals ($p < 0.005$) and significant below the individual 0.05 erro level for Ni as well. (Employing a simultaneous Bonferroni correction, the Cr and Pb parameter estimates can be declared to be jointly significant below the adjusted 0.01 level, while the Ni estimate is significant below the 0.1 level.) These test results suggest that the log contamination levels of three of the six metals appear to exhibit factory *Proximity* effects.

Figure 8.2 shows the observed ln[Pb+1] contamination data plotted as a function of distance to the factory site, with the predicted mixed linear spline equation superimposed on this data (adjusted to a reference altitude of 278.4 m above sea level). The increasing trend present in the log contamination data sampled within 2 Km of the factory site is apparent in this plot. Similar trends are also apparent in the Cr and Ni data, (see Figures 7.3 and 7.5, respectively).

Table 8.3 presents the final mixed linear spline model *Factory*, *Proximity*, and *Baseline* estimates for all six metals. Note that no Proximity effects are presented for As, Cu, or Zn, since the β_1 parameter estimates were clearly non-significant in these equations (for these three metals, Eqn [3] was re-estimated without the β_1 parameter in order to produce the final *Factory* and *Baseline* effect estimates). In addition to the mean log concentration estimates, back-transformed median (geometric mean) and approximate 95% confidence interval estimates are also shown in Table 8.3.

The median estimates shown in Table 8.3 can be used to quantify the degree of increased contamination on the factory perimeter with respect to the estimated baseline contamination level (throughout the sample region) for all six metals. For example, the Factory to Baseline ratios for As, Cr, Cu, Ni, Pb, and Zn are 3.1, 3.6, 4.1, 3.8, 37.4, and 8.2, respectively. Hence, topsoil samples around the factory perimeter exhibit about 3-4 times as much As, Cr, Cu, and Ni, 8 times as much Zn, and 37 times as much lead

concentration levels as the corresponding baseline levels. Likewise, the median proximity estimates associated with Cr, Pb, and Ni define the factory proximity effect. For Pb specifically, topsoil samples acquired with 0.5 Km and 1.0 Km of the factory will on average exhibit about 3 and 2.1 times higher concentration levels than the regional baseline level, respectively. Likewise, for Cr and Ni, the equivalent proximity ratios are 1.9 and 1.5 (Cr) and 1.6 and 1.3 (Ni), respectively.

Overall, these results agree with the earlier test results presented in sections 6 and 7. More specifically, they confirm that (i) the factory perimeter samples appear to be highly contaminated with respect to the estimated baseline metal contamination levels observed throughout the sampling region, and (ii) at least two (and quite possibly three) of the seven metals analyzed in this study (Cr, Ni, and Pb) exhibit significantly elevated contamination levels near the factory site.

Table 8.1 -2LL scores obtained for each of the six covariance structures used in conjunction with Eqn [2].

-2 Log Likelihood Scores (ML estimation)

Metal	IID	Spatial Exp	Spatial Gau	CS	Spatial Exp+CS	Spatial Gau+CS
As	493.2	464.6	466.4	453.8	453.8 *	453.8 *
Cr	422.4	418.8	418.5	416.7	416.7 *	416.7 *
Cu	540.5	536.0	535.4	531.1	531.1 *	531.1 *
Ni	325.6	304.1	302.6	275.9	275.9 *	275.9 *
Pb	736.4	662.8	667.6	653.6	653.6 *	653.6 *
Zn	662.9	650.1	650.2	636.4	636.4 *	636.5 *

Notes: (*) Hessian not positive definite, spatial sill parameter converged to 0.

Table 8.2 β_1 parameter estimates and t-tests for Eqn [3], using the CS error structure.

Metal	B1-spline Estimate	Std Error	t-score	one sided p-value
As	-0.0136	0.1753	-0.08	0.4692
Cr	-0.4195	0.1216	-3.45	0.0003 **
Cu	0.0677	0.1425	0.48	0.6826
Ni	-0.2985	0.1356	-2.20	0.0142 *
Pb	-0.7424	0.2461	-3.02	0.0014 **
Zn	0.0126	0.1862	0.07	0.5269

Notes: (*) significant at 0.1 Bonferroni corrected level
(**) significant at 0.01 Bonferonni corrected level

Table 8.3 Final mixed linear spline model predictions: *Baseline*, *Factory*, and *Proximity* effects (if applicable) for As, Cr, Cu, Ni, Pb, and Zn.

Metal	Location	Mean ln[M+1]	Std Error	Median (ppm)	95% CI (ppm)
As	Factory	2.864	0.323	17.53	(9.2, 33.5)
	Baseline	1.747	0.047	5.73	(5.2, 6.3)
Cu	Factory	4.550	0.221	94.59	(60.9, 147.0)
	Baseline	3.148	0.037	23.28	(21.6, 25.1)
Zn	Factory	6.779	0.312	878.84	(471.3, 1638.6)
	Baseline	4.672	0.049	106.94	(96.9, 118.0)
Cr	Factory	4.174	0.192	64.97	(44.3, 95.3)
	0.5 Km	3.517	0.173	33.69	(23.9, 47.6)
	1.0 Km	3.307	0.113	27.31	(21.8, 34.3)
	1.5 Km	3.098	0.058	22.14	(19.7, 24.9)
	Baseline	2.888	0.035	17.95	(16.8, 19.2)
Pb	Factory	7.023	0.474	1121.92	(434.8, 2895.2)
	0.5 Km	4.515	0.351	91.34	(45.3, 184.3)
	1.0 Km	4.143	0.232	63.02	(39.7, 100.1)
	1.5 Km	3.772	0.119	43.48	(34.3, 55.2)
	Baseline	3.401	0.072	30.00	(26.0, 34.6)
Ni	Factory	3.987	0.252	53.88	(32.5, 89.2)
	0.5 Km	3.099	0.193	22.17	(15.1, 32.6)
	1.0 Km	2.949	0.127	19.09	(14.8, 24.6)
	1.5 Km	2.800	0.065	16.45	(14.4, 18.7)
	Baseline	2.651	0.039	14.17	(13.1, 15.3)

Figure 8.1 Definition of the conceptual linear spline equation for modeling the distance to factory effect (if any) for each metal.

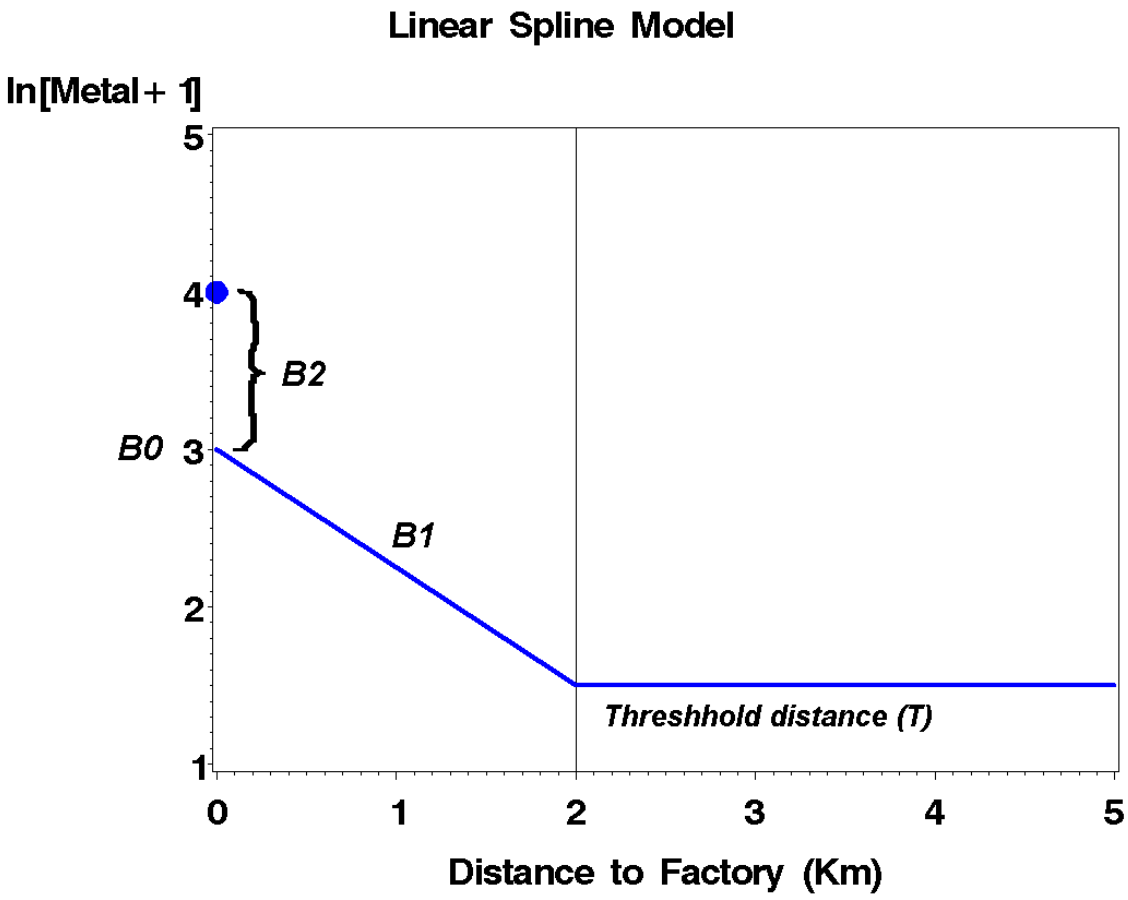
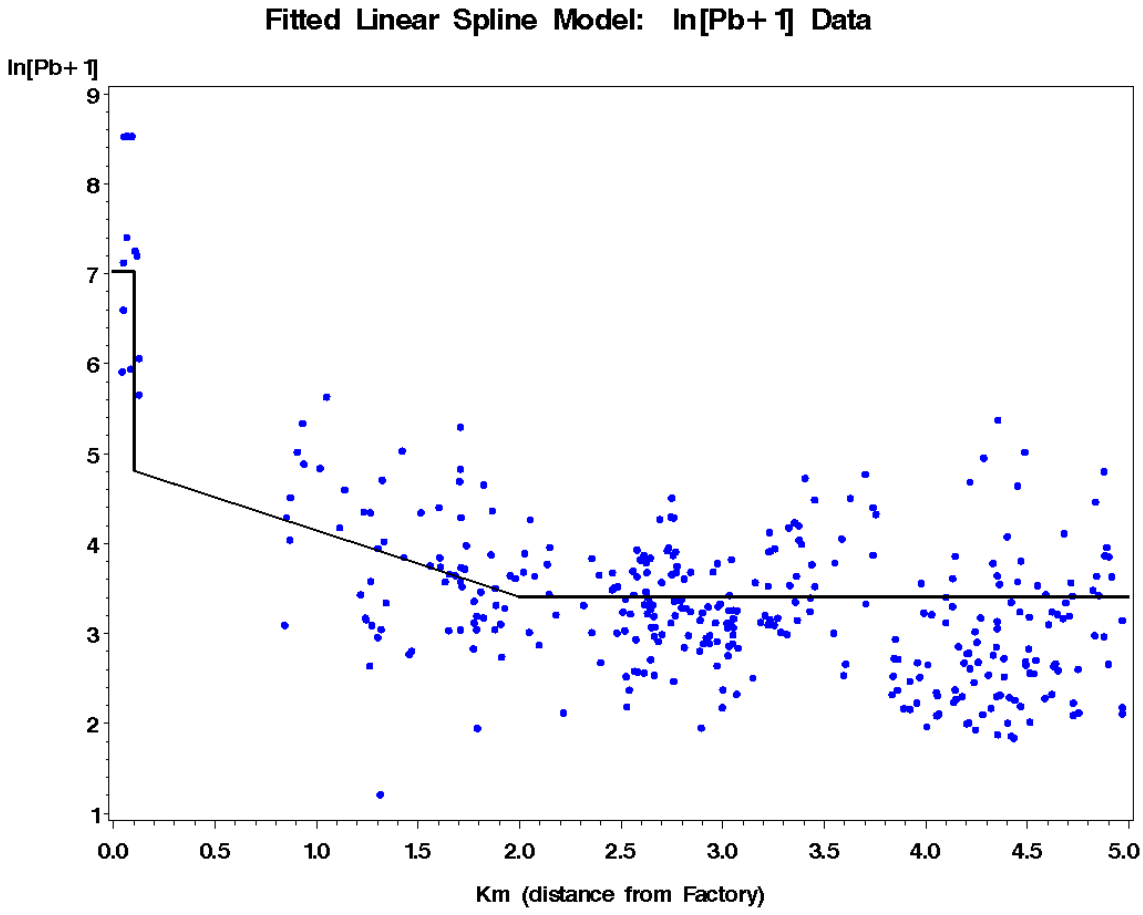


Figure 8.2 Observed ln[Pb+1] contamination data with the fitted mixed linear spline equation superimposed on the data.



9.0 References

Agresti, A. Categorical Data Analysis, 2nd Ed. John Wiley, New York, New York.

Cressie, N.A.C. 1991. Statistics for Spatial Data. John Wiley, New York, New York.

Davis, C.S. 2002. Statistical Methods for the Analysis of Repeated Measurements. Springer-Verlag, New York, New York.

Hollander, M., and Wolfe, D.A. 1999. Nonparametric Statistical Methods, 2nd Ed. John Wiley, New York, New York.

Jackson, K.W., Eastwood, I.W., and Wild, W.S. 1987. Stratified sampling protocol for monitoring trace metal concentrations in soil. Soil Science. 143: 436-443.

Johnson, R.A., and Wichern, D.W. 1988. Applied Multivariate Statistical Analysis, 2nd Ed. Prentice Hall, Englewood Cliffs, New Jersey.

Lohr, S.L. 1999. Sampling: Design and Analysis. Duxbury Press, Pacific Grove, CA.

Press, S.J. 1981. Applied Multivariate Analysis: using Bayesian and Frequentist Methods of Inference, 2nd Ed. R.E. Krieger Publishing Co., Malabar, Florida.

Schabenberger, O., and Gotway, C.A. 2005. Statistical Methods for Spatial Data Analysis. CRC Press, New York, New York.

Wackernagel, H. 1998. Multivariate Geostatistics, 2nd Ed. Springer-Verlag, Berlin, Germany.

Appendix: SAS Program Code

(for performing all data analyses described in Report)

Six SAS programs are shown in this Appendix. Each program performs a different set of statistical analyses, as shown in the Code Definition table below. Comments are also included in each set of source code, to help document the program flow. Note that all six programs use the following two data files:

Complete_metal_data.txt {metal concentration data file}
Final_data_cordfile.prn {lat/long coordinates, factory & freeway distances, etc.}

Code Definition Table

Program name	Purpose
Spatial_UniEDA_and_DepthAB.sas	Calculates all univariate distribution statistics and graphics (section 3), and analyzes the sampling depth effect (section 4).
Spatial_Variograms.sas	Produces the robust variogram plots (section 5).
Spatial_FC_tests_and_Qplots.sas	Produces both the quantile maps (section 5) and quantile indicator maps (section 6), and calculates the Chi-square association tests (section 6).
Spatial_D2Factory_plots.sas	Produces the CD2F plots (section 7).
Spatial_MLM_part_I.sas	Calculates the -2LL scores for the various covariance structures used in the linear spline models (section 8).
Spatial_MLM_part_II.sas	FIts the mixed linear spline models and produces all Baseline, Factory, and Proximity estimates (section 8).

```
/* sas file:  spatial_UniEDA_and_DepthAB.sas
   S.M. Lesch
   12/03/2005

   SAS code to perform univariate analyses of each metal, and to examine the correlation
   structure and relationships between A & B depth samples.
*/

/*  read in metal data from condensed data file...  */

goptions ftext='swissb' htitle=2 htext=2;

data metalA metalB;
infile "c:\UCR\Consult01\complete_metal_data.txt";
  input property $ site dpth $ As Cd Cr Cu Ni Pb Zn;
;
  lnAs = log(As+1);
  lnCd = log(Cd+1);
  lnCr = log(Cr+1);
  lnCu = log(Cu+1);
  lnNi = log(Ni+1);
  lnPb = log(Pb+1);
  lnZn = log(Zn+1);
  ;
  label As='As [ppm]' Cd='Cd [ppm]' Cr='Cr [ppm]' Cu='Cu [ppm]'
        Ni='Ni [ppm]' Pb='Pb [ppm]' Zn='Zn [ppm]'
               lnAs='ln(As+1)' lnCd='ln(Cd+1)' lnCr='ln(Cr+1)' lnCu='ln(Cu+1)'
               lnNi='ln(Ni+1)' lnPb='ln(Pb+1)' lnZn='ln(Zn+1)';
  ;
  factory = 0;
  if (property='Q') then factory = 1;
  ;
  if (property='63' and site=3) then delete;
  if (property='68') then delete;
  if (property='47' and site=3) then delete;
  ;
  if (dpth='A') then output metalA;
  if (dpth='B') then output metalB;
;
run;


title 'Frequency Counts: Properties & Sites w/in Properties';
proc freq data=metalA;
table property;
run;

title 'Metal Concentrations';
title2 ' ';

/*  Note: most statistical output for the 12 factory sites can be discarded (the
    sample size is not sufficient for creating histograms, qqplots, or reliable
    correlation estimates.  */

data statF statR;
set metalA;
if (factory=1) then output statF;
if (factory=0) then output statR;
run;

symbol1 i=none c=green v=dot;

title3 'Raw Data';
title4 'Non-Factory Sites';
proc means n mean std var skew min p1 p5 p10 p50 p90 p95 p99 max data=statR maxdec=2;
  var As Cd Cr Cu Ni Pb Zn;
run;

title4 'Factory Sites';
proc means n mean std var min max data=statF maxdec=2;
  var As Cd Cr Cu Ni Pb Zn;
```

```
run;

title4 'Non-Factory Sites';
proc univariate data=statR noprint;
  var As Cd Cr Cu Ni Pb Zn;
  histogram As Cd Cr Cu Ni Pb Zn / cfill=green;
  qqplot As Cd Cr Cu Ni Pb Zn;
run;

symbol1 i=none c=blue v=dot;

title3 'ln[M+1] Transformed Data';
title4 'Non-Factory Sites';
proc means n mean std var skew min p1 p5 p10 p50 p90 p95 p99 max data=statR maxdec=3;
  var lnAs lnCd lnCr lnCu lnNi lnPb lnZn;
run;

title4 'Factory Sites';
proc means n mean std var min max data=statF maxdec=3;
  var lnAs lnCd lnCr lnCu lnNi lnPb lnZn;
run;

title4 'Non-Factory Sites: Histograms & QQ-plots';
proc univariate data=statR noprint;
  var lnAs lnCd lnCr lnCu lnNi lnPb lnZn;
  histogram lnAs lnCd lnCr lnCu lnNi lnPb lnZn / cfill=blue;
  qqplot lnAs lnCd lnCr lnCu lnNi lnPb lnZn;
run;

title4 'Joint (7-metal) Correlation Structure';
proc corr data=statR;
  var lnAs lnCd lnCr lnCu lnNi lnPb lnZn;
run;


/*  Next, perform the A|B depth analyses.  First, we need to set up the A and B data
files.  */

proc sort data=metalA;  by property site;  run;
proc sort data=metalB;  by property site;  run;

data metalA;
set metalA;
  As_A = As;  Cd_A = Cd;  Cr_A = Cr;  Cu_A = Cu;  Ni_A = Ni;  Pb_A = Pb;  Zn_A = Zn;
  lnAs_A = lnAs;  lnCd_A = lnCd;  lnCr_A = lnCr;  lnCu_A = lnCu;  lnNi_A = lnNi;
  lnPb_A = lnPb;  lnZn_A = lnZn;
  keep property site As_A Cd_A Cr_A Cu_A Ni_A Pb_A Zn_A
       lnAs_A lnCd_A lnCr_A lnCu_A lnNi_A lnPb_A lnZn_A;
run;

data metalB;
set metalB;
  As_B = As;  Cd_B = Cd;  Cr_B = Cr;  Cu_B = Cu;  Ni_B = Ni;  Pb_B = Pb;  Zn_B = Zn;
  lnAs_B = lnAs;  lnCd_B = lnCd;  lnCr_B = lnCr;  lnCu_B = lnCu;  lnNi_B = lnNi;
  lnPb_B = lnPb;  lnZn_B = lnZn;
  keep property site As_B Cd_B Cr_B Cu_B Ni_B Pb_B Zn_B
       lnAs_B lnCd_B lnCr_B lnCu_B lnNi_B lnPb_B lnZn_B;
run;

/*  Now we will merge these two files together and compute the necessary difference
variables.  */

data metal2;
merge metalA metalB;
;
  by property site;
  if (As_B=.) then delete;

  /* compute raw differences for sign-rank tests */

  d_rAS = As_A - As_B;
```

```
  d_rCd = Cd_A - Cd_B;
  d_rCr = Cr_A - Cr_B;
  d_rCu = Cu_A - Cu_B;
  d_rNi = Ni_A - Ni_B;
  d_rPb = Pb_A - Pb_B;
  d_rZn = Zn_A - Zn_B;
  label d_rAs='As difference' d_rCd='Cd difference' d_rCr='Cr difference' d_rCu='Cu
difference'
        d_rNi='Ni difference' d_rPb='Pb difference' d_rZn='Zn difference';
  ;
  /* compute ln-transformed differences for t-tests & T2 test; note that Cd is excluded
     since there are too many non-detects (which result in transformed 0's */

  ld_rAS = lnAs_A - lnAs_B;
  ld_rCr = lnCr_A - lnCr_B;
  ld_rCu = lnCu_A - lnCu_B;
  ld_rNi = lnNi_A - lnNi_B;
  ld_rPb = lnPb_A - lnPb_B;
  ld_rZn = lnZn_A - lnZn_B;
  label ld_rAs='ln(As+1) difference' ld_rCr='ln(Cr+1) difference' ld_rCu='ln(Cu+1)
difference'
        ld_rNi='ln(Ni+1) difference' ld_rPb='ln(Pb+1) difference' ld_rZn='ln(Zn+1)
difference';

run;

/* check the validity of the Normality assumption for the ln-transformed differences &
   compute individual t-tests (note: sign-rank test results also available for this
   ln-transformed data)... */

title 'Normality Check on Differences of ln-transformed metal data';
title2 'Note: individual t-test results are also contained within this output';
proc univariate normal data=metal2;
  var ld_rAs ld_rCr ld_rCu ld_rNi ld_rPb ld_rZn;
  qqplot ld_rAs ld_rCr ld_rCu ld_rNi ld_rPb ld_rZn;
run;

/* perform sign-rank tests on raw difference data... */

title 'Univariate Output containing Sign-Rank test results (for raw difference data)';
proc univariate data=metal2;
  var d_rAs d_rCd d_rCr d_rCu d_rNi d_rPb d_rZn;
run;

/* perform the Hotellings T2 test on ln-transformed data using GLM procedure... */

title 'Hotellings T2 Test for Joint Hypothesis of ln(As,Cr,Cu,Ni,Pb,Zn) differences = 0';
proc glm data=metal2;
  model ld_rAs ld_rCr ld_rCu ld_rNi ld_rPb ld_rZn = / nouni;
  manova h=intercept;
run;

/* lastly, some extra code that calculates the correlation structures for the
   ln-transformed metal data across depths... */

title 'Within-Metal Correlation Structure (Depths 1 vs 2: ln-transformed data)';
title2 'Note: off-diagonal estimates can be ignored...';
proc corr data=metal2;
  var lnAs_A lnCr_A lnCu_A lnNi_A lnPb_A lnZn_A;
  with lnAs_B lnCr_B lnCu_B lnNi_B lnPb_B lnZn_B;
run;

title 'Raw Means & Differences';
proc means mean data=metal2 maxdec=2;
var As_A As_B d_rAs
    Cd_A Cd_B d_rCd
    Cr_A Cr_B d_rCr
    Cu_A Cu_B d_rCu
    Ni_A Ni_B d_rNi
    Pb_A Pb_B d_rPb
    Zn_A Zn_B d_rZn;
```

```
run;

title 'ln[M+1] Means & Differences'l;
proc means mean data=metal2 maxdec=3;
var lnAs_A lnAs_B ld_rAs
    lnCr_A lnCr_B ld_rCr
    lnCu_A lnCu_B ld_rCu
    lnNi_A lnNi_B ld_rNi
    lnPb_A lnPb_B ld_rPb
    lnZn_A lnZn_B ld_rZn;
run;
```

```
/* sas file:  spatial_Variograms.sas
   S.M. Lesch
   11/10/2005

   SAS code to create (robust) variogram plots of ln(Metal+1) transformed metal data
   (excluding Cd metal).
*/

/*  read in metal and coordinate data from condensed data files...  */

goptions ftext='swissb' htitle=1.7 htext=1.4;

data metal;
infile "c:\UCR\Consult01\complete_metal_data.txt";
  input property $ site dpth $ As Cd Cr Cu Ni Pb Zn;
;
lnAs = log(As+1);
lnCd = log(Cd+1);
lnCr = log(Cr+1);
lnCu = log(Cu+1);
lnNi = log(Ni+1);
lnPb = log(Pb+1);
lnZn = log(Zn+1);
;
oldsite = site;
if (dpth='B') then delete;
run;

data xylocs;
informat propname $40.;
infile "c:\UCR\Consult01\Final_data_cordfile.prn" truncover firstobs=6;
  input property $ code site lat long elevtn dQ d605 d60 d10 propname & ;
;
newsite = site;
run;

/*  create primary working file, and compute sampling distances to Factory...  */

proc sort data=metal;
by property site;
run;

proc sort data=xylocs;
by property site;
run;


data capoly;
```

```
merge metal xylocs;
by property site;
;
sx = 92.4*(lat+117.984);
sy = 111.0*(long-34.025);
;
deldist = ( sx**2 + sy**2 )**0.5;
olddist = 0.0003048*dQ;
f605dist = 0.0003048*d605;
f60dist = 0.0003048*d60;
f10dist = 0.0003048*d10;
;

label deldist='Km (distance from Factory)'
      lnAs='ln(As)' lnCd='ln(Cd)' lnCr='ln(Cr)' lnCu='ln(Cu)'
          lnNi='ln(Ni)' lnPb='ln(Pb)' lnZn='ln(Zn)'
      long='Longitude' lat='Latitude' olddist='Supplied by Russell';
;
if (property='63' and site=3) then delete;
if (property='68') then delete;
if (property='47' and site=3) then delete;
;
flag = 'no ';
if (deldist <= 2) then flag = 'yes';
label flag='w/in 2 Km of Factory';

run;

/*  IMPORTANT NOTE: a check on the supplied distance to factory values (versus the
computed values
    stored into the deldist variable) shows that about 10% of the recorded values appear
to be
    in error.  Thus, in this and all other SAS files, the SAS computed values are always
used.  */

title 'Cal Poly Study:  Merged Data Files';
title2 'QA/QC Check:  distance from Factory';
title3 'Computed versus Recorded';

symbol1 i=none c=black v=circle;

proc gplot data=capoly;
plot deldist*olddist;
run;

proc freq data=capoly;
table propname;
run;

/*  Create the robust variogram plots; note that these plots must be done one at a time
    (i.e., for each metal individually).  The empirical values used in the plots are also
    printed out for each metal.  */

symbol1 i=join c=black v=dot h=0.7 w=2;


title 'Robust Variogram Analysis: ln(Pb+1)';

proc variogram data=capoly outv=leadout;
compute lagd=0.05 maxlag=30 robust;
coordinates xc=sx yc=sy;
var lnPb;
run;

proc print data=leadout;
run;

data leadout;
set leadout;
label distance='Km';
if (COUNT > 30);
```

```
run;

title 'Robust Variogram of ln(Pb+1)';
proc gplot data=leadout;
plot rvario*distance;
run;


title 'Robust Variogram Analysis: ln(Ni+1)';

proc variogram data=capoly outv=Niout;
compute lagd=0.05 maxlag=30 robust;
coordinates xc=sx yc=sy;
var lnNi;
run;

proc print data=Niout;
run;

data Niout;
set Niout;
label distance='Km';
if (COUNT > 30);
run;

title 'Robust Variogram of ln(Ni+1)';
proc gplot data=Niout;
plot rvario*distance;
run;


title 'Robust Variogram Analysis: ln(Cr+1)';

proc variogram data=capoly outv=Crout;
compute lagd=0.05 maxlag=30 robust;
coordinates xc=sx yc=sy;
var lnCr;
run;

proc print data=Crout;
run;

data Crout;
set Crout;
label distance='Km';
if (COUNT > 30);
run;

title 'Robust Variogram of ln(Cr+1)';
proc gplot data=Crout;
plot rvario*distance;
run;


title 'Robust Variogram Analysis: ln(As+1)';

proc variogram data=capoly outv=Asout;
compute lagd=0.05 maxlag=30 robust;
coordinates xc=sx yc=sy;
var lnAs;
run;

proc print data=Asout;
run;

data Asout;
set Asout;
label distance='Km';
if (COUNT > 30);
run;
```

```
title 'Robust Variogram of ln(As+1)';
proc gplot data=Asout;
plot rvario*distance;
run;


title 'Robust Variogram Analysis: ln(Cu+1)';

proc variogram data=capoly outv=Cuout;
compute lagd=0.05 maxlag=30 robust;
coordinates xc=sx yc=sy;
var lnCu;
run;

proc print data=Cuout;
run;

data Cuout;
set Cuout;
label distance='Km';
if (COUNT > 30);
run;

title 'Robust Variogram of ln(Cu+1)';
proc gplot data=Cuout;
plot rvario*distance;
run;


title 'Robust Variogram Analysis: ln(Zn+1)';

proc variogram data=capoly outv=Znout;
compute lagd=0.05 maxlag=30 robust;
coordinates xc=sx yc=sy;
var lnZn;
run;

proc print data=Znout;
run;

data Znout;
set Znout;
label distance='Km';
if (COUNT > 30);
run;

title 'Robust Variogram of ln(Zn+1)';
proc gplot data=Znout;
plot rvario*distance;
run;

option pagesize=400;
title 'Variogram Data';

data capoly;
set capoly;
pcode = property;
if (pcode='48b') then pcode = '480';
if (pcode='M') then pcode = '101';
if (pcode='Q') then delete;
run;

proc print data=capoly noobs;
var pcode sx sy;
run;
```

```
/* sas file:  spatial_FC_tests_and_Qplots.sas
   S.M. Lesch
   11/15/2005

   Spatial Q90 quantile position plots for CaPoly metal data.  Associated Chi-square
tests
   for the Q50 and Q90 classifications also computed.  Additional factory contamination
tests (i.e.,
   a comparison of the factory levels versus sites > 2 Km away): both parametric (t-test)
and
   non-parametric (Kruskal-Wallis & Savage tests) performed.
*/

/*  read in metal and coordinate data from condensed data files...  */

goptions ftext='swissb' htitle=1.7 htext=1.4;

data metal;
infile "c:\UCR\Consult01\complete_metal_data.txt";
  input property $ site dpth $ As Cd Cr Cu Ni Pb Zn;
;
lnAs = log(As+1);
lnCd = log(Cd+1);
lnCr = log(Cr+1);
lnCu = log(Cu+1);
lnNi = log(Ni+1);
lnPb = log(Pb+1);
lnZn = log(Zn+1);
;
oldsite = site;
if (dpth='B') then delete;
run;

data xylocs;
informat propname $40.;
infile "c:\UCR\Consult01\Final_data_cordfile.prn" truncover firstobs=6;
  input property $ code site lat long elevtn dQ d605 d60 d10 propname & ;
;
newsite = site;
run;

/*  create primary working file, and compute sampling distances to Factory...  */

proc sort data=metal;
by property site;
run;

proc sort data=xylocs;
by property site;
run;

data capoly;
merge metal xylocs;
by property site;
;
mx = 92.4*(lat+117.984);
my = 111.0*(long-34.025);
;
deldist = ( mx**2 + my**2 )**0.5;
olddist = 0.0003048*dQ;
f605dist = 0.0003048*d605;
f60dist = 0.0003048*d60;
f10dist = 0.0003048*d10;
;
label deldist='Km (distance from Factory)'
      lnAs='ln(As+1)' lnCd='ln(Cd+1)' lnCr='ln(Cr+1)' lnCu='ln(Cu+1)'
         lnNi='ln(Ni+1)' lnPb='ln(Pb+1)' lnZn='ln(Zn+1)'
      long='Longitude' lat='Latitude';
;
if (property='63' and site=3) then delete;
if (property='68') then delete;
```

```
if (property='47' and site=3) then delete;
;
flag = 'no ';
if (deldist <= 2) then flag = 'yes';
label flag='w/in 2 Km of Factory';
;
factory = 0;
fzone = 2 - min(deldist,2);
f605buff = 0;
f60buff = 0;
f10buff = 0;
selevtn = (elevtn-322)/100 + 0.4357742;
;
if (property='Q') then do;
  factory = 1;
  fzone = 0;
end;
if (f605dist <= 1) then f605buff = 1;
if (f60dist <= 1) then f60buff = 1;
if (f10dist <= 1) then f10buff = 1;
freeway = 0;
if (f605buff=1 or f60buff=1 or f10buff=1) then freeway = 1;
;
samzone = 0;
if (fzone > 0) then samzone = 1;
if (property='Q') then samzone = 2;
run;

title 'Cal Poly Study:  Merged Data Files (data QA check)';

proc freq data=capoly;
table propname;
run;


/*  Set up a new data file: [basictest].  This file will be used to test for an
    elevated factory contamination level, versus sites > 2 Km from the factory.  */

data basictest;
set capoly;
if (samzone=1) then delete;
run;

title 'Basic t-test Results';
title2 'Factory vs > 2 Km away';
title3 'ANOVA Model Intercept represents Factory Estimate';
proc glm data=basictest;
  class samzone;
  model lnAs lnCd lnCr lnCu lnNi lnPb lnZn = samzone / solution;
  estimate '> 2 Km:  ' intercept 1 samzone 1 0;
run;

title 'Basic Nonparametric Results';
title2 'Factory vs > 2 Km away';
title3 'Kruskal-Wallis & Savage Tests';
proc npar1way data=basictest wilcoxon savage;
  class samzone;
  var lnAs lnCd lnCr lnCu lnNi lnPb lnZn;
run;

symbol1 i=none c=blue v=circle h=0.7;
symbol2 i=none c=red v=dot h=0.7;

/*  Set up a second new data file: [qq].  This file will be used to create a sample
location map and
    all 7 quantile maps.  Next, a reduced version of this file [qq2] which does not
contain the factory
    sites will be used to create all of the Q90 indicator maps and Chi-Sq tests.  */

data qq;
set capoly;
```

```
;
As_code50 = 'no ';
Cd_code50 = 'no ';
Cr_code50 = 'no ';
Cu_code50 = 'no ';
Ni_code50 = 'no ';
Pb_code50 = 'no ';
Zn_code50 = 'no ';
;
if (lnAs > 1.562) then As_code50 = 'yes';
if (lnCd > 0.409) then Cd_code50 = 'yes';
if (lnCr > 2.891) then Cr_code50 = 'yes';
if (lnCu > 3.131) then Cu_code50 = 'yes';
if (lnNi > 2.617) then Ni_code50 = 'yes';
if (lnPb > 3.230) then Pb_code50 = 'yes';
if (lnZn > 4.479) then Zn_code50 = 'yes';
;
label As_code50='As level > 50% Quantile'
      Cd_code50='Cd level > 50% Quantile'
      Cr_code50='Cr level > 50% Quantile'
      Cu_code50='Cu level > 50% Quantile'
      Ni_code50='Ni level > 50% Quantile'
      Pb_code50='Pb level > 50% Quantile'
      Zn_code50='Zn level > 50% Quantile';
;
As_code90 = 'no ';
Cd_code90 = 'no ';
Cr_code90 = 'no ';
Cu_code90 = 'no ';
Ni_code90 = 'no ';
Pb_code90 = 'no ';
Zn_code90 = 'no ';
;
if (lnAs > 2.324) then As_code90 = 'yes';
if (lnCd > 0.703) then Cd_code90 = 'yes';
if (lnCr > 3.413) then Cr_code90 = 'yes';
if (lnCu > 3.706) then Cu_code90 = 'yes';
if (lnNi > 3.070) then Ni_code90 = 'yes';
if (lnPb > 4.290) then Pb_code90 = 'yes';
if (lnZn > 5.501) then Zn_code90 = 'yes';
;
label As_code90='As level > 90% Quantile'
      Cd_code90='Cd level > 90% Quantile'
      Cr_code90='Cr level > 90% Quantile'
      Cu_code90='Cu level > 90% Quantile'
      Ni_code90='Ni level > 90% Quantile'
      Pb_code90='Pb level > 90% Quantile'
      Zn_code90='Zn level > 90% Quantile';
;
As_flag = 1;  if (lnAs > 1.372 ) then As_flag = 2;  if (lnAs > 1.562) then As_flag = 3;
if (lnAs > 1.902) then As_flag = 4;
Cd_flag = 1;  if (lnCd > 0.274 ) then Cd_flag = 2;  if (lnCd > 0.409) then Cd_flag = 3;
if (lnCd > 0.524) then Cd_flag = 4;
Cr_flag = 1;  if (lnCr > 2.666 ) then Cr_flag = 2;  if (lnCr > 2.891) then Cr_flag = 3;
if (lnCr > 3.084) then Cr_flag = 4;
Cu_flag = 1;  if (lnCu > 2.885 ) then Cu_flag = 2;  if (lnCu > 3.131) then Cu_flag = 3;
if (lnCu > 3.374) then Cu_flag = 4;
Ni_flag = 1;  if (lnNi > 2.432 ) then Ni_flag = 2;  if (lnNi > 2.617) then Ni_flag = 3;
if (lnNi > 2.823) then Ni_flag = 4;
Pb_flag = 1;  if (lnPb > 2.774 ) then Pb_flag = 2;  if (lnPb > 3.230) then Pb_flag = 3;
if (lnPb > 3.736) then Pb_flag = 4;
Zn_flag = 1;  if (lnZn > 4.179 ) then Zn_flag = 2;  if (lnZn > 4.479) then Zn_flag = 3;
if (lnZn > 4.947) then Zn_flag = 4;
;
run;


data qq2;
set qq;
if (samzone=2) then delete;
run;
```

```
title 'Lat/Long plot all Sample Site Locations';
proc gplot data=qq;
plot long*lat / href= -117.984 vref= 34.025;
run;


title 'Lat/Long plot of Q90 non-Factory Sample Site Locations';
title2 'Cross-hair = Factory Location';

title3 'As';
proc gplot data=qq2;
plot long*lat=As_code90 / href= -117.984 vref= 34.025;
run;
title3 'Cd';
proc gplot data=qq2;
plot long*lat=Cd_code90 / href= -117.984 vref= 34.025;
run;
title3 'Cr';
proc gplot data=qq2;
plot long*lat=Cr_code90 / href= -117.984 vref= 34.025;
run;
title3 'Cu';
proc gplot data=qq2;
plot long*lat=Cu_code90 / href= -117.984 vref= 34.025;
run;
title3 'Ni';
proc gplot data=qq2;
plot long*lat=Ni_code90 / href= -117.984 vref= 34.025;
run;
title3 'Pb';
proc gplot data=qq2;
plot long*lat=Pb_code90 / href= -117.984 vref= 34.025;
run;
title3 'Zn';
proc gplot data=qq2;
plot long*lat=Zn_code90 / href= -117.984 vref= 34.025;
run;

title 'Sampling Zone Influence Tests';
title2 'Zones 0 (> 2 Km) vs 1 (< 2 Km)';
title3 'Factory Sites Excluded';

proc freq data=qq2;
table samzone*As_code50 / chisq nocol ;
table samzone*As_code90 / chisq nocol ;
table samzone*Cd_code50 / chisq nocol ;
table samzone*Cd_code90 / chisq nocol ;
table samzone*Cr_code50 / chisq nocol ;
table samzone*Cr_code90 / chisq nocol ;
table samzone*Cu_code50 / chisq nocol ;
table samzone*Cu_code90 / chisq nocol ;
table samzone*Ni_code50 / chisq nocol ;
table samzone*Ni_code90 / chisq nocol ;
table samzone*Pb_code50 / chisq nocol ;
table samzone*Pb_code90 / chisq nocol ;
table samzone*Zn_code50 / chisq nocol ;
table samzone*Zn_code90 / chisq nocol ;
run;

symbol1 i=none c=blue v=dot h=0.7;
symbol2 i=none c=green v=dot h=0.7;
symbol3 i=none c=orange v=dot h=0.7;
symbol4 i=none c=red v=dot h=0.7;

title 'Lat/Long plot of Q25:Q50:Q75 Spatial Data Patterns';
title2 'flag=1 [M < Q25]  flag=2 [Q25 < M < Q50]  flag=3 [Q50 < M < Q75]  flag=4 [M >
Q75]';
title3 '   ';

title4 'As';
```

```
proc gplot data=qq;
plot long*lat=As_flag / href= -117.984 vref= 34.025;
run;
title4 'Cd';
proc gplot data=qq;
plot long*lat=Cd_flag / href= -117.984 vref= 34.025;
run;
title4 'Cr';
proc gplot data=qq;
plot long*lat=Cr_flag / href= -117.984 vref= 34.025;
run;
title4 'Cu';
proc gplot data=qq;
plot long*lat=Cu_flag / href= -117.984 vref= 34.025;
run;
title4 'Ni';
proc gplot data=qq;
plot long*lat=Ni_flag / href= -117.984 vref= 34.025;
run;
title4 'Pb';
proc gplot data=qq;
plot long*lat=Pb_flag / href= -117.984 vref= 34.025;
run;
title4 'Zn';
proc gplot data=qq;
plot long*lat=Zn_flag / href= -117.984 vref= 34.025;
run;
```

```
/* sas file:  spatial_D2Factory_plots.sas
   S.M. Lesch
   11/29/2005

   Contamination by Distance (to factory) plots: trend estimated using smoothing spline.
*/

/*  read in metal and coordinate data from condensed data files...  */

goptions ftext='swissb' htitle=1.7 htext=1.4;

data metal;
infile "c:\UCR\Consult01\complete_metal_data.txt";
  input property $ site dpth $ As Cd Cr Cu Ni Pb Zn;
;
lnAs = log(As+1);
lnCd = log(Cd+1);
lnCr = log(Cr+1);
lnCu = log(Cu+1);
lnNi = log(Ni+1);

lnPb = log(Pb+1);
lnZn = log(Zn+1);
;
oldsite = site;
if (dpth='B') then delete;
run;

data xylocs;
informat propname $40.;
```

```
infile "c:\UCR\Consult01\Final_data_cordfile.prn" truncover firstobs=6;
  input property $ code site lat long elevtn dQ d605 d60 d10 propname & ;
;
newsite = site;
run;

/*  create primary working file, and compute sampling distances to Factory...  */

proc sort data=metal;  by property site;  run;
proc sort data=xylocs;  by property site;  run;

data capoly;
merge metal xylocs;
by property site;
;
mx = 92.4*(lat+117.984);
my = 111.0*(long-34.025);
;
deldist = ( mx**2 + my**2 )**0.5;
olddist = 0.0003048*dQ;
f605dist = 0.0003048*d605;
f60dist = 0.0003048*d60;
f10dist = 0.0003048*d10;
;
label deldist='Km (distance from Factory)'
      lnAs='ln(As+1)' lnCd='ln(Cd+1)' lnCr='ln(Cr+1)' lnCu='ln(Cu+1)'
         lnNi='ln(Ni+1)' lnPb='ln(Pb+1)' lnZn='ln(Zn+1)'
      long='Longitude' lat='Latitude';
;
if (property='63' and site=3) then delete;
if (property='68') then delete;
if (property='47' and site=3) then delete;
;
flag = 'no ';
if (deldist <= 2) then flag = 'yes';
label flag='w/in 2 Km of Factory';
;
factory = 0;
if (property='Q') then factory = 1;
run;

title 'Cal Poly Study:  Merged Data Files';


proc freq data=capoly;
table propname;
run;


/*  make the contamination by distance to factory plots...  */

symbol1 i=sm50 c=green v=circle h=0.7 w=2;
symbol2 i=none c=blue v=dot h=0.7;

proc sort data=capoly;
by deldist;
run;

title 'Metal Data (ln[ppm+1] scale):  Contamination by Distance Plots';
title2 'Smoothed Spline used to approximate Trend';
title3 'Maximum effect distance ~ 2 Km';

proc gplot data=capoly;
plot (lnAs lnCd lnCr lnCu lnNi lnPb lnZn)*deldist=factory / href = 0 0.2 2 haxis = -0.5
to 5.5 by 0.5;
run;
```

```
/* sas file:  spatial_MLM_part_I.sas
   S.M. Lesch
   11/10/2005

   Preliminary MLM of Cal Poly metal data.  This code performs the covariance
identification process;
   i.e., the determination of the proper covariance structure for the proposed linear
spline model
   (used to determine if the proximity to factory effects the log transformed metal
contamination
   levels).  */

/*  read in metal and coordinate data from condensed data files...  */

goptions ftext='swissb' htitle=1.7 htext=1.4;

data metal;
infile "c:\UCR\Consult01\complete_metal_data.txt";
  input property $ site dpth $ As Cd Cr Cu Ni Pb Zn;
;
lnAs = log(As+1);
lnCd = log(Cd+1);
lnCr = log(Cr+1);
lnCu = log(Cu+1);
lnNi = log(Ni+1);
lnPb = log(Pb+1);
lnZn = log(Zn+1);
;
oldsite = site;
if (dpth='B') then delete;
run;

data xylocs;
informat propname $40.;
infile "c:\UCR\Consult01\Final_data_cordfile.prn" truncover firstobs=6;
  input property $ code site lat long elevtn dQ d605 d60 d10 propname & ;
;
newsite = site;
run;

/*  create primary working file, and compute sampling distances to Factory...  */

proc sort data=metal;  by property site;  run;
proc sort data=xylocs;  by property site;  run;

data capoly;
merge metal xylocs;
by property site;
;
mx = 92.4*(lat+117.984);    /* convert lat/long into Km */
my = 111.0*(long-34.025);
;
deldist = ( mx**2 + my**2 )**0.5;   /* re-calculate distance to factory */
olddist = 0.0003048*dQ;             /* re-scale feet to Km */
f605dist = 0.0003048*d605;
f60dist = 0.0003048*d60;
f10dist = 0.0003048*d10;
;
label deldist='Km (distance from Factory)'
      lnAs='ln(As)' lnCd='ln(Cd)' lnCr='ln(Cr)' lnCu='ln(Cu)'
          lnNi='ln(Ni)' lnPb='ln(Pb)' lnZn='ln(Zn)'
      long='Longitude' lat='Latitude';
;
if (property='63' and site=3) then delete;
if (property='68') then delete;
if (property='47' and site=3) then delete;
;
/* flag sites w/in 2 Km of factory */
flag = 'no ';
if (deldist <= 2) then flag = 'yes';
label flag='w/in 2 Km of Factory';
```

```
;
/* define factory indicator variable, proximity variable, freeway indicator variables,
   and normalized (centered and relative scaled) elevation variable */
factory = 0;
proximity = min(deldist,2);
f605buff = 0;
f60buff = 0;
f10buff = 0;
selevtn = (elevtn-322)/100 + 0.4357742;
;
if (property='Q') then do;
  factory = 1;
  proximity = 0;
end;
if (f605dist <= 1) then f605buff = 1;
if (f60dist <= 1) then f60buff = 1;
if (f10dist <= 1) then f10buff = 1;
;
/* use individual freeway indicator variables to create a single variable that
   indicates if a site is w/in 1 Km of any freeway */
freeway = 0;
if (f605buff=1 or f60buff=1 or f10buff=1) then freeway = 1;
;
run;

/*  Note: The MIXED procedure is run four times for each metal.  The first model
represents
    the IID error case.  The second model fits a compound symmetric (CS) error structure.
    The third model fits an isotropic spatial exponential (SpExp) error structure (with a
nugget
    effect), and the forth model fits a mixture of the CS and SpExp structures.  Note
that
    the 2nd and 3rd covariance structures are nested w/in the 4th structure.  Note also
that
    in all 6 cases the partial sill parameter in the SpExp structure converges to the
boundary (0)
    in the 4th model, suggesting that once the CS error structure is adopted, no further
    spatial structure is apparent in the MLM residuals.  */


title 'Mixed Modeling Analysis of ln[As]:  Part I';

title2 'IID Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnAs = factory proximity freeway selevtn / solution;
run;
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnAs = factory proximity freeway selevtn / solution;
  random property;
run;
title2 'Spatial Error Structure (Isotropic Exp w/Nugget)';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnAs = factory proximity freeway selevtn / solution;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.1 0.75 0.2;
run;
title2 'CS + Spatial Error Structure (Isotropic Exp w/Nugget)';
title3 'Compound Symmetry assumed to be additive & IID';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnAs = factory proximity freeway selevtn / solution;
  random property;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.05 0.1 0.75 0.2;
run;
```

```
title 'Mixed Modeling Analysis of ln[Cr]:  Part I';

title2 'IID Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnCr = factory proximity freeway selevtn / solution;
run;
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnCr = factory proximity freeway selevtn / solution;
  random property;
run;
title2 'Spatial Error Structure (Isotropic Exp w/Nugget)';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnCr = factory proximity freeway selevtn / solution;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.05 0.25 0.2;
run;
title2 'CS + Spatial Error Structure (Isotropic Exp w/Nugget)';
title3 'Compound Symmetry assumed to be additive & IID';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnCr = factory proximity freeway selevtn / solution;
  random property;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.05 0.05 0.25 0.2;
run;



title 'Mixed Modeling Analysis of ln[Cu]:  Part I';

title2 'IID Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnCu = factory proximity freeway selevtn / solution;
run;
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;

  model lnCu = factory proximity freeway selevtn / solution;
  random property;
run;
title2 'Spatial Error Structure (Isotropic Exp w/Nugget)';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnCu = factory proximity freeway selevtn / solution;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.1 0.5 0.2;
run;
title2 'CS + Spatial Error Structure (Isotropic Exp w/Nugget)';
title3 'Compound Symmetry assumed to be additive & IID';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnCu = factory proximity freeway selevtn / solution;
  random property;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.05 0.1 0.5 0.2;
run;



title 'Mixed Modeling Analysis of ln[Pb]:  Part I';

title2 'IID Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
```

```
  model lnPb = factory proximity freeway selevtn / solution;
run;
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnPb = factory proximity freeway selevtn / solution;
  random property;
run;
title2 'Spatial Error Structure (Isotropic Exp w/Nugget)';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnPb = factory proximity freeway selevtn / solution;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.3 0.5 0.2;
run;
title2 'CS + Spatial Error Structure (Isotropic Exp w/Nugget)';
title3 'Compound Symmetry assumed to be additive & IID';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnPb = factory proximity freeway selevtn / solution;
  random property;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.2 0.1 0.5 0.1;
run;



title 'Mixed Modeling Analysis of ln[Ni]:  Part I';

title2 'IID Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnNi = factory proximity freeway selevtn / solution;
run;
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnNi = factory proximity freeway selevtn / solution;
  random property;
run;
title2 'Spatial Error Structure (Isotropic Exp w/Nugget)';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnNi = factory proximity freeway selevtn / solution;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.09 0.5 0.09;
run;
title2 'CS + Spatial Error Structure (Isotropic Exp w/Nugget)';
title3 'Compound Symmetry assumed to be additive & IID';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnNi = factory proximity freeway selevtn / solution;
  random property;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.06 0.06 0.5 0.06;
run;



title 'Mixed Modeling Analysis of ln[Zn]:  Part I';

title2 'IID Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnZn = factory proximity freeway selevtn / solution;
run;
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnZn = factory proximity freeway selevtn / solution;
  random property;
```

```
run;
title2 'Spatial Error Structure (Isotropic Exp w/Nugget)';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnZn = factory proximity freeway selevtn / solution;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.1 0.5 0.3;
run;
title2 'CS + Spatial Error Structure (Isotropic Exp w/Nugget)';
title3 'Compound Symmetry assumed to be additive & IID';
proc mixed data=capoly covtest method=ml;
  class property;
  model lnZn = factory proximity freeway selevtn / solution;
  random property;
  repeated / subject=intercept local type=sp(exp)(mx my);
  parms 0.05 0.1 0.5 0.2;
run;
```

```
/* sas file:  spatial_MLM_part_II.sas
   S.M. Lesch
   11/29/2005

   Compound Symmetry MLM Analysis: model estimation and final predicted estimates
   for the 6 metals that can be analyzed.
*/

/*  read in metal and coordinate data from condensed data files...  */

goptions ftext='swissb' htitle=1.7 htext=1.4;

data metal;
infile "c:\UCR\Consult01\complete_metal_data.txt";
  input property $ site dpth $ As Cd Cr Cu Ni Pb Zn;
;
lnAs = log(As+1);
lnCd = log(Cd+1);
lnCr = log(Cr+1);
lnCu = log(Cu+1);
lnNi = log(Ni+1);
lnPb = log(Pb+1);
lnZn = log(Zn+1);
;
oldsite = site;
if (dpth='B') then delete;
run;

data xylocs;
informat propname $40.;
infile "c:\UCR\Consult01\Final_data_cordfile.prn" truncover firstobs=6;
  input property $ code site lat long elevtn dQ d605 d60 d10 propname & ;
;
newsite = site;
run;

/*  create primary working file, and compute sampling distances to Factory...  */

proc sort data=metal;  by property site;  run;
proc sort data=xylocs;  by property site;  run;

data capoly;
```

```
merge metal xylocs;
by property site;
;
mx = 92.4*(lat+117.984);    /* convert lat/long into Km */
my = 111.0*(long-34.025);
;
deldist = ( mx**2 + my**2 )**0.5;   /* re-calculate distance to factory */
olddist = 0.0003048*dQ;             /* re-scale feet to Km */
f605dist = 0.0003048*d605;
f60dist = 0.0003048*d60;
f10dist = 0.0003048*d10;
;
label deldist='Km (distance from Factory)'
      lnAs='ln(As)' lnCd='ln(Cd)' lnCr='ln(Cr)' lnCu='ln(Cu)'
         lnNi='ln(Ni)' lnPb='ln(Pb)' lnZn='ln(Zn)'
      long='Longitude' lat='Latitude';
;
if (property='63' and site=3) then delete;
if (property='68') then delete;
if (property='47' and site=3) then delete;
;
/* flag sites w/in 2 Km of factory */
flag = 'no ';
if (deldist <= 2) then flag = 'yes';
label flag='w/in 2 Km of Factory';
;
/* define factory indicator variable, proximity variable, freeway indicator variables,
   and normalized (centered and relative scaled) elevation variable */
factory = 0;
proximity = min(deldist,2);
f605buff = 0;
f60buff = 0;
f10buff = 0;
selevtn = (elevtn-322)/100 + 0.4357742;
;
if (property='Q') then do;
  factory = 1;
  proximity = 0;
end;
if (f605dist <= 1) then f605buff = 1;
if (f60dist <= 1) then f60buff = 1;
if (f10dist <= 1) then f10buff = 1;
;
/* use individual freeway indicator variables to create a single variable that
   indicates if a site is w/in 1 Km of any freeway */
freeway = 0;
if (f605buff=1 or f60buff=1 or f10buff=1) then freeway = 1;
;
run;

/*  Note: The MIXED procedure is run twice for each metal.  The first model includes the
    general freeway indicator variable, which never appears to be statistically
significant.
    The GLS residuals are also output using this first model (so QQ plots of the
residuals
    can be made).  The second model represents the reduced model; these models are used
to
    generate (i.e., estimate) the baseline, factory, and (if applicable) proximity
effects
    for each metal.  Note that these estimates are for the reference altitude of 278.5 ft
MSL.  */

symbol1 i=none c=black v=circle h=0.7;
symbol2 i=none c=blue v=dot h=0.7;


title 'Mixed Modeling Analysis of ln[As]:  Part II';
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnAs = factory proximity selevtn / solution outp=resAs;
```

```
  random property;
run;
title3 'Estimates of Interest';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnAs = factory selevtn / solution;
  estimate 'baseline: ' intercept 1;
  estimate 'factory:  ' intercept 1 factory 1;
  random property;
run;
title3 'Residual Diagnositics';
proc univariate normal data=resAs;
var Resid;
qqplot Resid;
run;


title 'Mixed Modeling Analysis of ln[Cr]:  Part II';
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnCr = factory proximity selevtn / solution outp=resCr;
  random property;
run;
title3 'Estimates of Interest';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnCr = factory proximity selevtn / solution;
  estimate 'baseline: ' intercept 1 proximity 2;
  estimate 'factory:  ' intercept 1 factory 1;
  estimate '0.5 Km away: ' intercept 1 proximity 0.5;
  estimate '1.0 Km away: ' intercept 1 proximity 1.0;
  estimate '1.5 Km away: ' intercept 1 proximity 1.5;
  random property;
run;
title3 'Residual Diagnositics';
proc univariate normal data=resCr;
var Resid;
qqplot Resid;
run;


title 'Mixed Modeling Analysis of ln[Cu]:  Part II';
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnCu = factory proximity selevtn / solution outp=resCu;
  random property;
run;
title3 'Estimates of Interest';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnCu = factory selevtn / solution;
  estimate 'baseline: ' intercept 1;
  estimate 'factory:  ' intercept 1 factory 1;
  random property;
run;
title3 'Residual Diagnositics';
proc univariate normal data=resCu;
var Resid;
qqplot Resid;
run;


title 'Mixed Modeling Analysis of ln[Pb]:  Part II';
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnPb = factory proximity selevtn / solution outp=resPb;
  random property;
run;
```

```
title3 'Estimates of Interest';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnPb = factory proximity selevtn / solution;
  estimate 'baseline: ' intercept 1 proximity 2;
  estimate 'factory:  ' intercept 1 factory 1;
  estimate '0.5 Km away: ' intercept 1 proximity 0.5;
  estimate '1.0 Km away: ' intercept 1 proximity 1.0;
  estimate '1.5 Km away: ' intercept 1 proximity 1.5;

  random property;
run;
title3 'Residual Diagnositics';
proc univariate normal data=resPb;
var Resid;
qqplot Resid;
run;


title 'Mixed Modeling Analysis of ln[Ni]:  Part II';
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnNi = factory proximity selevtn / solution outp=resNi;
  random property;
run;
title3 'Estimates of Interest';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnNi = factory proximity selevtn / solution;
  estimate 'baseline: ' intercept 1 proximity 2;
  estimate 'factory:  ' intercept 1 factory 1;
  estimate '0.5 Km away: ' intercept 1 proximity 0.5;
  estimate '1.0 Km away: ' intercept 1 proximity 1.0;
  estimate '1.5 Km away: ' intercept 1 proximity 1.5;
  random property;
run;
title3 'Residual Diagnositics';
proc univariate normal data=resNi;
var Resid;
qqplot Resid;
run;


title 'Mixed Modeling Analysis of ln[Zn]:  Part II';
title2 'Compound Symmetry Error Structure';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnZn = factory proximity selevtn / solution outp=resZn;
  random property;
run;
title3 'Estimates of Interest';
proc mixed data=capoly covtest method=reml;
  class property;
  model lnZn = factory selevtn / solution;
  estimate 'baseline: ' intercept 1;
  estimate 'factory:  ' intercept 1 factory 1;
  random property;
run;
title3 'Residual Diagnositics';
proc univariate normal data=resZn;
var Resid;
qqplot Resid;
run;
```