# Appendix B

Nickel Statistical Background Memo



16406 US Route 224 East Findlay, OH 45840

## Memorandum

Date: Thursday, October 27, 2005

To: Guy Gallello

CC:

From: John Carson, Ph.D.

RE: Results of Schenectady AD AOC-2 data analysis

As we discussed, I have analyzed the metals data from the excavations in AOC-2 of Schenectady Army Depot and the relevant background (BG) for the purpose of comparing mercury (Hg) and nickel (Ni) soil concentrations to BG.

#### Data

I started with the Excel files that you provided ('838360 AOC-2 Area B.xls', '838360 AOC-2 Area C.xls', '838360 AOC-2 Area F.xls' and 'AOC-2 Metal hits.xls'). Based on the field drawing you provided of the planned backfill area (Area B + part of Area C), I grouped those sample locations within the planned backfill area into an area B/C to facilitate the analysis. The data is listed in Attachment 3. I did not use the original data for locations that were resampled except for analysis of trace metals/ major metals relationships. For the comparisons of mean concentrations to BG, I used the data from duplicate sample pairs with weights of 0.5 for each member of the pair. Unduplicated samples were given a weight of one.

### Nondetects

There were no nondetects (ND) for Ni and two ND for Hg among the sample results used in this analysis. A number of J-qualified results were also reported at values less than the reporting limit (RL). The ND were given as <RL, where RL was the sample specific reporting limit. The laboratory RLs are almost always greater than or equal to the limit of quantitation (LOQ) which is a little more than three times the limit of detection (LOD). When J-qualified results are reported, their estimated concentrations lie between the LOD and the LOQ. In this case, ND are actually between zero and the LOD but are reported as if they are between zero and the LOQ. Unless the ND are assumed to be zero (as in Aitchison's method), reporting ND as <RL when they are really <LOD creates a high bias in statistical procedures. This problem is known as *informative* censoring (see Helsel, chapter 3). If the ND is replaced by substituting RL/2, then that result is biased by a factor greater than three, since RL ≥LOQ ≥3·LOD.



To reduce this high bias, when I know that an RL ≥LOQ is being used, I use RL/3 as the censoring point. This is applicable to any method of handling ND except for zero substitution. Since the mercury result for sample EX-B-3 was ND with a very high RL, 0.34, much higher than most of the reported data, I eliminated it from the analysis as it added almost no useful information. The other ND sample result (EX-C-10RE) was reported as '0.036 U' or < 0.036. I replaced the RL with RL/3, to approximate the LOD as discussed above, and used LOD/2 substitution in the analysis.

### **Computations**

I used the public domain software R (<a href="http://www.r-project.org/">http://www.r-project.org/</a>), a language and environment for statistical computing and graphics. It is based on the S language and environment, which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues.

### Comparison of Means

First, I used the nonparametric percentile bootstrap to compare the mean concentrations of Hg and Ni in the combined backfill areas B/C and in area F to the BG means. I used stratified bootstrap sampling with importance weights given by the sample weights mentioned above. I initially conducted four univariate tests, for Hg in area B/C, for Hg in area F, for Ni in area B/C and for Ni in area F, at the 5% significance level, by constructing 95% lower confidence limits (LCL) for the difference in means using the percentile method of constructing confidence intervals with 1000 bootstrap replications. If the LCL for an area/analyte is greater than 0, then that area/analyte fails the test and is deemed to (putatively) exceed BG.

Average Hg did not exceed BG in either area. Area C by itself significantly exceeded BG for Hg. However, over the planned backfill area B/C and in Area F, the sample mean Hg concentration was less than that for BG. Average Ni appeared to exceed BG in both in area F and in the combined backfill area B/C—more so in area B/C than in area F. The resulting 95% LCLs are given in Table 1. Because of the apparent failure for Ni, additional analysis was required.

Table 1: 95% LCLs for Mean Diff erence from BG

Analyte \ Area	Area B/C	Area F
Hg	-0.0292	-0.0287
Ni	10.75	0.838



### Predictive Analysis

Next I looked at the relation between Ni concentrations and the concentrations of the major constituent metals: aluminum (Al), calcium (Ca), iron (Fe), potassium (K) and magnesium (Mg). I found high correlation and strong linear relationships between Ni on one hand and Fe, K, Mg and the total of the major constituent metals on the other hand. The respective correlations with Ni were 0.92 for Fe, 0.90 for K, 0.73 for Mg and 0.81 for the total of major constituent metals. The full set of correlations is shown in Table 2.

	Hg	Ni	Al	Ca	Fe	K	Mg	Al+Ca+Fe+K+Mg
Hg	1.00	0.28	0.18	-0.11	0.25	0.24	0.12	0.11
Ni	0.28	1.00	0.55	0.26	0.92	0.90	0.73	0.81
Al	0.18	0.55	1.00	-0.25	0.63	0.68	0.25	0.39
Ca	-0.11	0.26	-0.25	1.00	0.13	0.26	0.64	0.71
Fe	0.25	0.92	0.63	0.13	1.00	0.84	0.63	0.77
K	0.24	0.90	0.68	0.26	0.84	1.00	0.72	0.80
Mg	0.12	0.73	0.25	0.64	0.63	0.72	1.00	0.88
Al+Ca+Fe+K+Mg	0.11	0.81	0.39	0.71	0.77	0.80	0.88	1.00

**Table 2: Correlations for Metals** 

Aluminum is a surrogate variable for clay content. The total of major constituent metals is a surrogate variable for the total inorganic content minus sand. The relationship between Fe and Ni appeared to be consistently the same in all the zones: BG, area B/C and area F. The relationship between the other metals and Ni did not appear to be as consistent by zone. This is illustrated in Figures 1 (Ni versus Fe) and 2 (Ni versus K).

To verify this, I fit various linear models (Rao, chapter 4) predicting Ni. For the models with Fe as the only continuous covariate, I tested the effect of zone on both the intercept and the slope of the regression using Analysis of Covariance (ANCOVA), but neither was statistically significant. For K, Mg and the total of the major constituent metals, there were linear relationships of varying strength with Ni, but they appeared to change by zone. In linear models with the other metals as predictors, Area was always a statistically significant factor. The strength and consistency of the Ni-Fe relationship across zones shows that Ni is geochemically strongly associated with Fe.

To explore this further, I included all major metals and area into a large ANCOVA model. I whittled down the list of predictors, refitting the model at each step until only statistically significant predictors remained. The final model included Fe, K, Al, Area and the interaction between Al and Area. It predicts 91% of the variation in the Ni concentrations (Adjusted R-squared: 0.914) versus 83% for Fe as the only predictor. The role of K and Al probably relates to



predicting the total silt and clay content of the soil. Ni concentration would be reduced as there is more sand or organics in the soil and increased as there is less.

The Ni levels appear higher in area B/C than in BG because the Fe levels are higher due to natural variation in the soil, as indicated by the boxplots in Figure 3, not because there is anthropogenic Ni contamination. In fact, after controlling for the concentrations of Fe, K and Al, the estimated effects for areas B, C and F area actually *lower* than the effect for BG. That means that based on modeling this data, for samples in each area having the same concentrations of Fe, K and Al, we would expect to see higher Ni concentrations in the samples from BG than in those from Areas B, C or F.

#### Conclusion

I conclude based on this analysis that there is no evidence that either Hg or Ni are elevated above BG in the studied areas: backfill area B/C and Area F. Please contact me if you have any questions about this analysis.

#### Carson, John, Ph.D.

Senior Statistician, Senior Technical Leader Shaw Environmental & Infrastructure 16406 US Rte. 224 East Findlay OH 45840 419-425-6156 direct 419-424-4939 fax www.shawgrp.com John.Carson@shawgrp.com

### References

Helsel, D. 2004. Nondetects and Data Analysis: Statistics for Censored Environmental Data. John Wiley. New York. 268 pp.

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rao, C R. 1973. *Linear Statistical Inference and Its Applications*. Second Edition. John Wiley. New York. 625 pp.



Figure 1: Nickel versus Iron Relationship

### AOC-2 Nickel versus Iron by Area

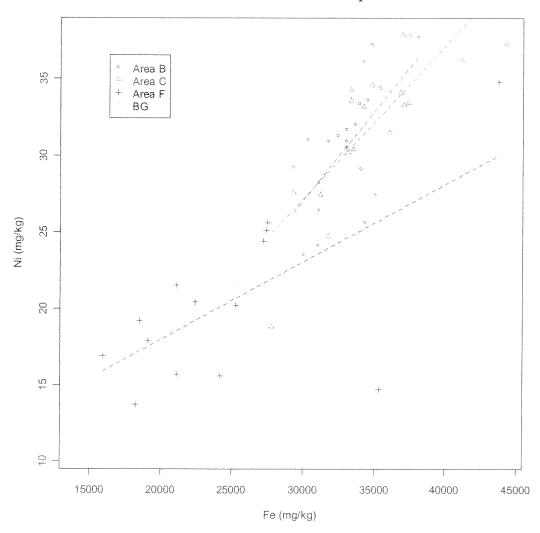


Figure 2: Nickel versus Potassium Relationship

### AOC-2 Nickel versus Potassium by Area

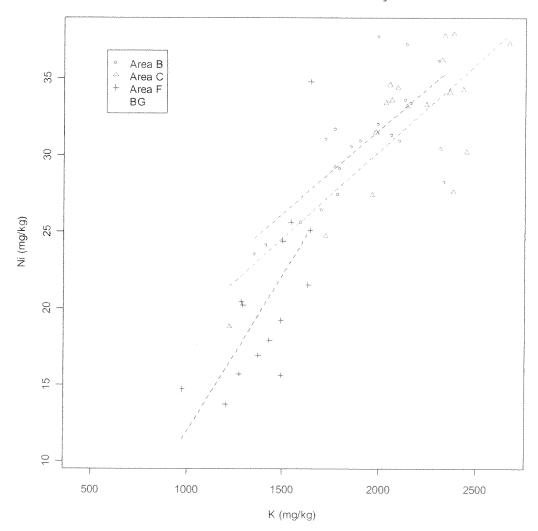
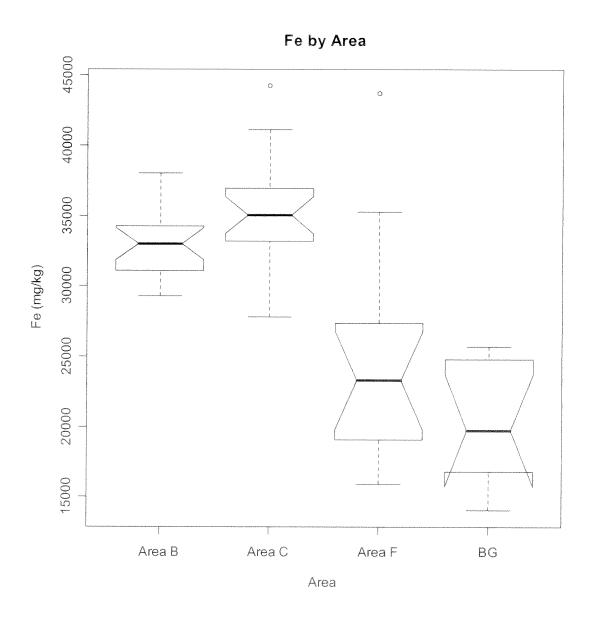


Figure 3: Boxplots of Iron Concentration by Area

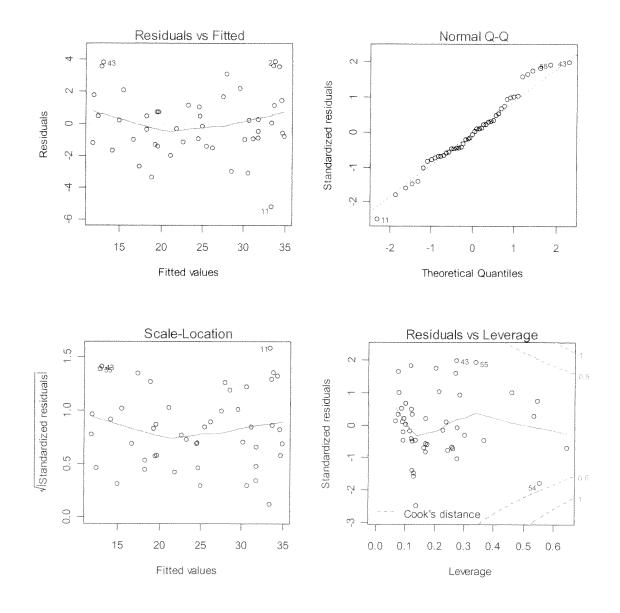




### **Attachment 1: Supporting Graphs**



Figure 4: Model Diagnostic Plots for Ni Predicted by Fe, K, Al and Area





### **Attachment 2: R Printouts**



### **Bootstrap Confidence Intervals for Ni**

```
Sken.AOC2.Ni.diff.bs$call <-
boot(data = Sken.AOC2.metals[Use, ], statistic = mean.wt.zone.Ni.diff,
    R = 1000, strata = backfill.area[Use], weights = Ni.wt[Use])</pre>
```

Note that the lower limit of a 90% two-sided confidence interval (CI) for the mean is also a 95% LCL for the mean.

#### Area B/C

### **Bootstrap Confidence Intervals for Hg**

```
Sken.AOC2.Hg.diff.bs$call <-
boot(data = Sken.AOC2.metals[Use, ], statistic = mean.wt.zone.Hg.diff,
    R = 1000, strata = backfill.area[Use], weights = Hg.wt[Use])

Area B/C
boot.ci(Sken.AOC2.Hg.diff.bs,index=1,conf=0.9,type="perc")
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates

Intervals :
Level    Percentile
90% (-0.0292, 0.0109)
Calculations and Intervals on Original Scale

Area F
boot.ci(Sken.AOC2.Hg.diff.bs,index=2,conf=0.9,type="perc")
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates</pre>
```



Intervals :
Level Percentile
90% (-0.0287, -0.0074 )
Calculations and Intervals on Original Scale

### Initial Linear Model with Fe and Area as Predictors

```
>summary(Sken.AOC2.Ni.lm)
Call:
lm(formula = Ni ~ Fe * Area, data = Sken.AOC2.metals, subset = Use)
Residuals:
    Min
            10 Median
                             30
-6.4720 -2.0780 0.5308 1.8158 6.1220
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.767e+00 1.112e+01 -0.519 0.60644
              1.103e-03 3.370e-04 3.274 0.00204 **
AreaArea C -1.536e+01 1.482e+01 -1.036 0.30569
AreaArea F 9.497e+00 1.152e+01 0.825 0.41395
AreaBG 8.601e+00 1.194e+01 0.720 0.47520
              8.601e+00 1.194e+01 0.720 0.47520
Fe:AreaArea C 4.408e-04 4.453e-04 0.990 0.32751
Fe:AreaArea F -3.891e-04 3.581e-04 -1.087 0.28297
Fe:AreaBG -4.083e-04 3.987e-04 -1.024 0.31131
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 2.958 on 45 degrees of freedom
Multiple R-Squared: 0.8671, Adjusted R-squared: 0.8464
F-statistic: 41.94 on 7 and 45 DF, p-value: < 2.2e-16
> Anova (Sken.AOC2.Ni.lm)
Anova Table (Type II tests)
Response: Ni
         Sum Sq Df F value
                               Pr(>F)
         667.44 1 76.2705 3.028e-11 ***
Area 32.60 3 1.2417 0.3058
Fe:Area 69.94 3 2.6642 0.0592
                                0.0592 .
Residuals 393.79 45
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Final Linear Model with Fe Only as Predictor

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.004e+00 1.787e+00 -1.122 0.267
            9.704e-04 6.096e-05 15.919 <2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 3.12 on 51 degrees of freedom
Multiple R-Squared: 0.8325, Adjusted R-squared: 0.8292
F-statistic: 253.4 on 1 and 51 DF, p-value: < 2.2e-16
> Anova (Sken.AOC2.Ni.lm3)
Anova Table (Type II tests)
Response: Ni
          Sum Sq Df F value
                              Pr(>F)
         2466.30 1 253.42 < 2.2e-16 ***
Residuals 496.33 51
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Initial Linear Model with Fe, K, Mg, Al, Area and All Interactions as Predictors

```
> summary(Sken.AOC2.Ni.Fe.K.Mg.Al.lm)
Call:
lm(formula = Ni ~ Fe + K * Area + Mq * Area + Al * Area + Area,
    data = Sken.AOC2.metals, subset = Use)
Residuals:
     Min
            10
                  Median
                            30
-4.68800 -0.95048 -0.09436 1.39682 3.67385
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)
                 -1.290e+01 9.551e+00 -1.350 0.1870
                 4.385e-04 9.312e-05 4.709 5.29e-05 ***
                 -3.311e-04 5.229e-03 -0.063
                                              0.9499
Area[T.Area C]
                -6.239e+01 4.772e+01
                                     -1.307
                                               0.2010
                                      1.273
Area[T.Area F]
                 1.898e+01 1.491e+01
                                              0.2127
Area[T.BG]
                  2.340e+01 1.003e+01
                                     2.334 0.0265 *
Mg
                 2.425e-03 1.194e-03 2.031 0.0512.
Al
                 7.729e-04 5.202e-04 1.486 0.1478
K:Area[T.Area C] -2.273e-02 2.570e-02 -0.884
                                              0.3836
K:Area[T.Area F] 5.660e-03 9.816e-03 0.577
                                              0.5685
K:Area[T.BG]
                 6.716e-03 5.744e-03
                                     1.169
                                              0.2515
Area[T.Area C]:Mg 6.606e-03 7.112e-03
                                      0.929
                                              0.3604
Area[T.Area F]:Mg -6.335e-04 2.409e-03 -0.263
                                              0.7944
Area[T.BG]:Mg -2.243e-03 1.226e-03 -1.829
                                              0.0773 .
Area[T.Area C]:Al 3.708e-03 3.042e-03 1.219 0.2324
Area[T.Area F]:Al -1.414e-03 6.136e-04 -2.305 0.0283 *
Area[T.BG]:Al -1.729e-03 6.690e-04 -2.584 0.0149 *
```



```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 2.234 on 30 degrees of freedom
Multiple R-Squared: 0.9453, Adjusted R-squared: 0.9162
F-statistic: 32.42 on 16 and 30 DF, p-value: 1.377e-14
> Anova (Sken.AOC2.Ni.Fe.K.Mg.Al.lm)
Anova Table (Type II tests)
Response: Ni
          Sum Sq Df F value
                             Pr(>F)
        110.715 1 22.1792 5.285e-05 ***
         30.236 1 6.0570 0.01982 *
Area
         44.340 3 2.9608 0.04800 *
          7.330 1 1.4683 0.23507
         15.953 1 3.1959
Al
                            0.08393 .
         13.090 3 0.8741
K:Area
                            0.46544
Area:Mg 26.332 3
Area:Al 50.229 3
                    1.7583
                             0.17639
                    3.3541
                            0.03184 *
Residuals 149.755 30
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Final Linear Model with Fe, K, Al, Area and Al:Area as Predictors

```
> summary(Sken.AOC2.Ni.Fe.K.Mg.Al.lm3)
lm(formula = Ni ~ Fe + K + Al * Area + Area, data = Sken.AOC2.metals,
    subset = Use)
Residuals:
   Min
           1Q Median
                          3 Q
                                  Max
-5.2428 -1.2051 -0.1643 1.1174 3.8753
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)
                -9.490e-01 6.557e+00 -0.145 0.8857
                                      6.087 4.79e-07 ***
                 4.898e-04 8.047e-05
Fe
                 8.520e-03 1.226e-03
                                      6.947 3.33e-08 ***
                 -4.591e-05 3.524e-04 -0.130 0.8971
Area[T.Area C]
              -1.608e+01 1.283e+01 -1.253 0.2179
Area[T.Area F]
                 1.093e+01 6.784e+00 1.611 0.1157
Area[T.BG]
                 1.245e+01 7.409e+00 1.680 0.1014
Al:Area[T.Area C] 7.663e-04 7.434e-04 1.031
                                              0.3093
Al:Area[T.Area F] -7.644e-04 4.121e-04 -1.855
                                              0.0716
              -1.232e-03 5.128e-04 -2.402
                                              0.0214 *
Al:Area[T.BG]
... ... ...
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note: See ANOVA table below where Area is a significant predictor. Area is coded as a factor whose first level is 'Area B'. The factor effects are fitted so that the effect for the first



# level is 0. So the other metals being equal, we estimate that Ni should be about 12 mg/kg higher on average in samples from Area B than in samples from BG.

```
Residual standard error: 2.259 on 37 degrees of freedom
Multiple R-Squared: 0.9311, Adjusted R-squared: 0.9143
F-statistic: 55.53 on 9 and 37 DF, p-value: < 2.2e-16
> Anova(Sken.AOC2.Ni.Fe.K.Mg.Al.lm3)
Anova Table (Type II tests)
Response: Ni
         Sum Sq Df F value
                             Pr(>F)
        189.062 1 37.0535 4.785e-07 ***
        246.278 1 48.2669 3.331e-08 ***
         96.431 1 18.8990 0.0001036 ***
Area
         45.269 3 2.9573 0.0448471 *
Al:Area 62.128 3 4.0588 0.0136914 *
Residuals 188.789 37
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Attachment 3: Data



Mg	6470	7980	8290	8070	5350	8440	5920	7610	8080	5730	8760	4880	8760	7800	9969	5670	8340	5400	5500	5100	7380	7860	4900	7200	0092	5360	6400	5810	0609	
2	1850	1990	1900	2060	1700	2100	1780	1720	1990	1770	2330 8	1350 4	2300 8	2160	2130	1770	2140	1590	1790	1410	2330	2670	1720 2	2360	2310	1960	2240 e	-	2320	
×	-	-		-	+	+		-		-	+				+	+		+	+	+-	-	+	<del> </del>		-			+		
ъ Б	33000	38000	33000	32400	31100	31700	35000	30300	33600	33000	31100	30000	34200	33900	34500	29300	34800	34300	34000	31000	37400	44200	31700	36800	33500	31200	37000	34200	41100	
Ca	4300	10100	16200	28100	2400	36700	922	20600	20300	2410	30600	1780	24900	24100	3100	1730	21900	2150	2510	2240	0269	24800	1560	7810	20100	2650	2840	2450	9430	
Al	16300	17300	15000	15000	16000	14600	17800	13300	15600	17000	15100	16400	15900	15000	17300	20100	15800	17500	17800	15000	18300	18000	18200	18900	17000	17500	19900	18500	18400	
Ni.wt	<b>7</b> (	-	house	quained	feeced			-	(manual		formet		-	0.5	period		0.5		Amend	yeard	0.5	fecand	-	Pared.		_			-	-
Ni.q																														-
Z	30.5	37.7	30.9	31.3	26.4	30.9	27.4	31	32	31.7	28.2	23.5	36.1	33.4	33.6	29.2	37.2	25.6	29.1	24.1	37.8	37.3	24.7	34.1	30.4	27.4	33.3	33.2	36.2	
Hg.wt	pond		0	_	-	-		-	quant	-		-	posed	0.5	-	henny	0.5	-	·		0.5	_		parent.	-	ferrent				_
Hg.q			5																						5					
Hg	0.22	0.05	NA	0.023	0.2	0.031	0.024	0.034	0.057	0.03	0.022	0.043	0.036	0.036	0.026	0.042	0.035	0.27	0.1	0.05	0.16	0.17	0.03	0.045	0.033	0.071	0.25	0.46	0.33	-
backfill.area		B/C	B/C	B/C		B/C	B/C	B/C	B/C	B/C	B/C	B/C	B/C	B/C		B/C	B/C	B/C	B/C	B/C	The state of the s		B/C	B/C					The second secon	
Backfill	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	-
SampleID	EX-B-1	EX-B-2	EX-B-3	EX-B-4	EX-B-5	EX-B-6	EX-B-7	EX-B-8	EX-B-9	EX-B-10	EX-B-11	EX-B-12	EX-B-13	EX-B-14	EX-B-15	EX-B-16	EX-B-DUP	EX-B-1RE	EX-B-5RE	EX-B-15RE	EX-C-1	EX-C-2	EX-C-3	EX-C-4	EX-C-5	EX-C-6	EX-C-7	EX-C-8	EX-C-9	
Area	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area B	Area C	Area C	Area C	Area C	Area C	Area C	Area C	Area C	Area C	
-	p(	7	60	4	S	9	7	∞	6	10	garand garand	12	13	14	15	16	17	<u>∞</u>	19	20	21	22			25	26	27	28	29	Contract of the Contract of th





	Area	SampleID	Backfill	backfill.area	Hg	Hg.q	Hg.wt	ź	Ni.q	Ni.wt	Al	Са	Fe	X	Mg
62	BG	HP16	FALSE	BG	0.039	1	П	9.01			8110	1520	14100	443	2150
63	BG	HP17	FALSE	BG	0.054			12.5	The state of the s	-	7950	1280	16900	539	2320
49	BG	HP18	FALSE	BG	0.04	janny		15.2		hwari	7080	46600	16900	503	5880