N00129.AR.000924
NSB NEW LONDON
5090.3a

TETRA TECH NUS, INC.
661 Andersen Drive ▪ Pittsburgh, Pennsylvania 15220-2745
(412) 921-7090 ▪ FAX (412) 921-4040 ▪ www.tetratech.com

PITT-03-2-074

April 1, 2002

Project Number 2863

Ms. Kymberlee Keckler
Remedial Project Manager
U.S. Environmental Protection Agency – Region 1
1 Congress St.
Suite 1100 (HBT)
Boston, Massachusetts 02114-2023

Mr. Mark Lewis
Connecticut Department of Environmental Protection
Bureau of Water Management
Permitting, Enforcement, & Remediation Division
79 Elm Street
Hartford, Connecticut 06106-5127

Subject: Final Year 3 Annual Groundwater Monitoring Report for DRMO
Naval Submarine Base-New London, Groton, Connecticut

Reference: CLEAN Contract Number N62467-94-D-0888
Contract Task Order Number 0816

Dear Ms. Keckler / Mr. Lewis:

On behalf of the U.S. Navy, Tetra Tech NUS, Inc. (TtNUS) is pleased to submit to the U.S. Environmental Protection Agency, Region 1 (EPA), 2 copies and to the Connecticut Department of Environmental Protection (CTDEP), 1 copy of the final version of the subject report. EPA comments dated December 5, 2001 and January 30, 2002 were addressed during preparation of this final report per the Navy's responses. The response-to-comment document is provided in Appendix D of the report.

If you have any questions regarding the report or the Groundwater Monitoring Program at the DRMO at Naval Submarine Base-New London, please contact Mr. Mark Evans of Engineering Field Activity Northeast at (610) 595-0567 (ext. 162) or me at (412) 921-8984.

Sincerely,

Corey A. Rich, P.E.
Project Manager

Attachment(s)

c: Mr. Mark Evans, EFANE (2 copies)
Mr. Richard Conant, NSB-NLON (2 copies)
Mr. Roger Boucher, EFANE (w/o enclosure)
Ms. Jennifer Hayes Stump, Gannett Fleming (1 copy)
Mr. John Trepanowski, TtNUS-KOP (1 copy)
CTO 816 – File Copy

GENERAL COMMENTS

Comment 1

The monitoring program adheres closely to the Groundwater Monitoring Plan, and appears to be moving forward in routine fashion. The monitoring data continue to support the conclusion that the soil removal and cap installation are successful in limiting transport of site contaminants to groundwater. The groundwater met all primary monitoring criteria in Rounds 9 through 12. Secondary criteria are exceeded for a few SVOCs present at relatively low levels, and for several inorganics. According to the Groundwater Monitoring Plan Decision Diagram, monitoring frequency can be reduced to twice a year when four consecutive rounds of representative monitoring have been performed with results below monitoring criteria.

Response

Agreed. The monitoring program is proceeding in routine fashion and there have only been a few compounds detected at concentrations that exceeded secondary criteria. The monitoring data support the conclusion that the soil removal and cap installation are successful in limiting transport of site contaminants to groundwater.

Figure 4-10 (i.e., Groundwater Monitoring Plan Decision Diagram) does indicate that the monitoring frequency can be reduced to twice a year when four consecutive rounds of representative monitoring have been performed with results below monitoring criteria. However, because 12 rounds of quarterly sampling have been completed at the DRMO site and the results have all been below the primary monitoring criteria, it is recommended in the report that the monitoring frequency be reduced from quarterly to annual, instead of biannual. The USEPA indicates their agreement with this recommendation in Comment No. 15. Therefore, no changes are recommended in response to this comment.

Comment 2

Statistics are used in the annual report to analyze whether compounds are present in the groundwater above upgradient levels. The Groundwater Monitoring Plan was finalized in February of 1998. Since that time, a number of detailed guidance documents for performing statistical comparisons using environmental data have been published by a variety of agencies. It appears that some of the statistical procedures currently used in this annual groundwater monitoring report do not fully comply with procedures recommended in these more current guidance documents. However, based on the Groundwater Monitoring Plan Decision Diagram, it does not appear that the errors identified in the statistics used in this monitoring program would change the recommendations presented in the annual report. Comments related to the statistics are provided in Attachment A to assist in developing statistical evaluation procedures that are in line with currently recommended guidance.

Response

Agreed. The Navy's recommendations for revised statistical methods are provided in the response for General Comment 3.

Comment 3

As described in §4.3.1.1, in the statistical analysis of the ground water data, non-detects have been replaced with one-half the sample quantitation limit. This section also states: "...The amount of data that are below the detection limit play an important role in selecting the statistical method of addressing the detection limit problem...." According to USEPA guidance (USEPA, 1992), parametric ANOVAs should not be used in the event nondetects exceed 50% of the data set. In addition, for analyses using the Wilcoxon Rank Sum Test, several environmental statistics guidance documents limit the percent of nondetects allowable in the test data sets to 50% (US Navy 1998) or even 40% (US DD, US DOE, USEPA, USNRC 2000).

REFERENCES:

US Department of Defense, US Department of Energy, US Environmental Protection Agency, and US Nuclear Regulatory Commission, 2000, *Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM)*, Revision 1 (includes replacement pages dated June 2001), NUREG-1575, EPA/402/R-97/016, DOE/EH-0624, August.

US Navy, 1998, *Procedural Guidance for Statistically Analyzing Environmental Background Data*, Naval Facilities Engineering Command, Southwest Division and Engineering Field Activities, West, September.

US Environmental Protection Agency, 1992, *Statistical Analysis of Ground Water Monitoring Data at RCRA Facilities*: Addendum to Interim Guidance

The data sets for a number of the parameters evaluated in this document contained more than 50% nondetects for either the upgradient wells, the down gradient wells or both. For example, for the constituents trans-1,2-dichloroethene, vinyl chloride, phenanthrene, arsenic, lead and silver the frequency of detection for each of these constituents in upgradient wells was zero out of twelve samples. Based on guidance provided in the aforementioned guidance documents, these constituents should not have been subject to statistical evaluation procedures using either ANOVA or Wilcoxon Rank-Sum tests. As such, the conclusions regarding the statistical comparisons presented in Section 5.1 (page 5-2, first full paragraph) are erroneous for the listed constituents. No statistical comparisons between the upgradient and downgradient wells can be performed using ANOVA or Wilcoxon Rank-Sum tests for these constituents. To ensure the validity of the statistical comparisons performed in the monitoring program, the Navy should confirm that the assumptions regarding the frequency of detection for each test are met for each constituent analyzed by either the ANOVA or the Wilcoxon Rank-Sum procedures.

One of the main assumptions in a parametric ANOVA is that the data sets being compared have homogeneity of variance. There is no mention of this in the text and no evidence that this assumption was checked in the tables. If homogeneity of variance cannot be demonstrated then ANOVAs are not the appropriate statistic to apply. Rather, the Wilcoxon Rank Sum Test should be used for comparison purposes. In this report, homogeneity of variance should be confirmed for data analyzed using the ANOVA techniques using the F-test for homogeneity of variance or a similar statistic. If homogeneity of variance cannot be confirmed then data sets should be tested with techniques for unequal variances.

Response to Original Comment (January 8, 2002)

Agreed. The statistical methodology used to evaluate the Year 3 data from the DRMO is the same method that has been used since the Year 1 analysis. This methodology failed to include:

- Homogeneity of variance between the upgradient and downgradient data sets as a requirement for performing the parametric ANOVA.
- A contingency for not performing ANOVA when nondetects exceed 50% of the data set.

Revisions are necessary to the statistical methodology performed on the analytical data. Because the conclusions of the Year 3 analysis would be unaffected by such changes, it is recommended that the following steps be incorporated into the statistical analysis performed for Year 4:

- Levene's Test of Homogeneity of Variance
- Two-Sample Test of Proportions

EPA Rebuttal (January 30, 2002)

General Comment 3 and Specific Comment 4: Although the recommended statistical steps are reasonable to address the comment, based on the reduction in sampling frequency and the elimination of the three deep wells, it is unclear how the Navy plans to use ANOVA techniques to compare upgradient wells to downgradient wells in year 4 and beyond. Two upgradient wells and five downgradient wells will be sampled one time each year. This number of data points is not sufficient to perform an ANOVA test. Please describe in greater detail how the comparison on this limited data will performed.

It should be noted that to date, the comparison of upgradient and downgradient wells has been something of an academic exercise since no primary monitoring criteria have been exceeded.

Response to Rebuttal

Agree with Clarification. Regarding the original comment about the inappropriateness of completing statistical comparison between upgradient and downgradient wells for the parameters trans-1,2-dichloroethene, vinyl chloride, phenanthrene, arsenic, lead and silver, the commentor is correct. To address this problem, new statistical evaluations were completed on the Year 3 data set following the new methodology provided in the response to the original comment. This change was necessary so that the conclusions and recommendations of the report were based on technically defensible statistical information. The results of the new evaluation (text and tables) are provided in the final report. As noted in the rebuttal, the comparison of upgradient and downgradient wells has been something of an academic exercise since no primary monitoring criteria have been exceeded. It should also be noted that the new statistical evaluation did not result in any changes to the conclusions or recommendations of the report.

Regarding the EPA's rebuttal, the commentor is correct in stating that the statistical methodologies suggested by the Navy for use in the future may not be suitable for evaluating analytical data from a single annual sampling event. However, the suggested methodologies would be appropriate for use under the other groundwater monitoring programs at the Area A Landfill and Goss Cove Landfill sites at Naval Submarine Base-New London, Groton, Connecticut. Quarterly sampling is still being conducted at those sites. The suggested methodologies will be incorporated, as appropriate, into the statistical evaluations completed for those monitoring programs.

The following methodology will be adopted for use for the DRMO groundwater monitoring program in the future.

> The Parametric ANOVA and the tests of it's assumptions (Shapiro-Wilk Test of Normality and Levene's Test of Homogeneity of Variance) will no longer be performed. Either a Wilcoxon Rank-Sum analysis (if 4 or more of the 7 concentrations are positive detections) or a Test of Proportions (if 3 or less of the 7 concentrations are positive detections) will be performed. Due to the small size of the data sets, these tests will have limited statistical power.
>
> Since the site has proven to be rather benign over the last three years, attention will be concentrated on determining if conditions are changing at the site by performing trend analysis. The Mann-Kendall Test (Gilbert, p. 208) will be used on the average concentrations for each COPC over all rounds to test for an upward trend. Those COPCs exhibiting an upward trend according to the test will be plotted as a function of time. Linear regression best-fit lines with confidence bands will also be plotted. If it becomes evident that the trend is not linear but rather follows another function the regression plot will be the function that best fits the shape of the data. If there are any COPCs that show trends indicating that the primary monitoring criteria is likely to be exceeded in the near future, additional methods may be used to confirm/verify the results.

Comment 4

As noted in previous reviews, it is noted that the secondary monitoring criterion for arsenic (0.14 µg/L, based human-health considerations for consumption of organisms), is well below the laboratory detection limits achieved (typically a few µg/L). Therefore, no meaningful comparisons can be made to this monitoring criterion. Nonetheless, this criterion appears to be rather conservative, particularly in view of the "background" value adopted (3.6 micrograms per liter); *i.e.*, "typical" NSBNL groundwater exceeds the secondary monitoring criterion.

Response to Original Comment (January 8, 2002)

Agreed. The secondary monitoring criteria for arsenic (0.14 µg/L) is below the laboratory detection limit for arsenic, and in general no meaningful comparisons can be made to this criterion. This

secondary criterion was agreed to by the Navy and regulators several years ago during development of the Groundwater Monitoring Program for the DRMO. Since the laboratory detection limit for arsenic is not low enough to allow for a meaningful comparison and the background concentration of arsenic determined for NSB-NLON is 1.92 µg/L, as presented in the Basewide Groundwater Operable Unit Remedial Investigation (TtNUS, 2001), the use of the secondary criterion should be reconsidered. It is recommended that the background concentration for arsenic be adopted as the secondary monitoring criterion for arsenic during future groundwater monitoring efforts at the DRMO. In addition, it is recommended that the background concentrations of the other inorganics of concern at the DRMO (i.e., barium, chromium, copper, lead, and zinc) also be adopted as the secondary monitoring criteria for future groundwater monitoring efforts at the DRMO.

It should be noted that the "background" concentration of 3.6 µg/L stated in the comment is not the background groundwater concentration presented in the subject report; however, it was erroneously presented as such in previous reports. The background concentration of arsenic determined for NSB-NLON is 1.92 µg/L, as presented in the Basewide Groundwater Operable Unit Remedial Investigation (TtNUS, 2001).

EPA Rebuttal (January 30, 2002)

General Comment 4: Regarding the comparison to arsenic secondary monitoring criterion, please expand the comparison table to include background values and discuss the limitations of the secondary monitoring criterion in the text.

Response to Rebuttal

Agree. The secondary criteria for inorganics were maintained and the analytical summary table (Table 4-1) was expanded by one column to include background concentrations for inorganics. In addition, the columns that include the criterion and background values were conditionally formatted, as appropriate, to indicate the criteria that was exceeded by the detected chemical concentrations. The second paragraph in Section 4.1 was revised to include a discussion on the limitations of the secondary monitoring criteria.

SPECIFIC COMMENTS

Comment 1: p. 1-1, §1.1

The text reads, "... to determine whether surface water protection have been attained ..."...Please check.

Response

The sentence will be corrected to read, "...to determine whether surface water protection criteria have been attained..." The basis for this sentence is that the groundwater at the DRMO is not used as a drinking water source and the only way receptors can come into contact with the groundwater is after it discharges to the Thames River (i.e., after it essentially becomes surface water).

Comment 2: p. 2-8, §2.3.5

The text states that the regression on the average arsenic concentrations performed after Year 1 of the monitoring program indicated a decline, and that this "... contradicted the results of the statistical evaluation." This contradiction is not apparent in the summary of results given here. The statistical evaluation established that arsenic concentrations were higher in downgradient wells than in upgradient wells. The trend analysis suggested a trend of declining concentrations in the downgradient area. Downgradient arsenic could have been declining, while still remaining higher than upgradient arsenic. Please clarify the argument for a contradiction in these analyses. In any case, it is agreed that the trend may not be meaningful, in view of the low confidence in the fit, as well as the possible influence of the reporting of "non-detects" and variable laboratory detection limits discussed previously.

Response to Original Comment (January 8, 2002)

Agreed. The paragraph in question will be revised as follows:

"A statistical evaluation of the data indicated that upgradient and downgradient concentrations of both organic and inorganic COCs were found to be similar except for arsenic. *The statistical evaluation established that arsenic concentrations were higher in downgradient wells than in upgradient wells.* The average arsenic concentrations for all downgradient monitoring wells for each round were plotted as a function of round to determine the trend of the concentrations. The regression line fit to the average arsenic concentrations showed a decreasing trend, which contradicted the results of the statistical evaluation. The confidence in the regression line fit to the data was low and it is likely the line did not represent the true trend in the data."

EPA Rebuttal (January 30, 2002)

Specific Comment 2: The suggested text does not address the comment. Trend analysis and the statistical analysis are not at odds with each other, as they are evaluating two independent issues. It is possible to have a statistically higher level of arsenic in downgradient wells and at the same time have a downward trend in these wells. While I agree that the trend analysis is somewhat inconclusive, this is not from a contradiction with the results of the statistical comparison. Although the proposed text is an improvement, please eliminate the phrase, "which contradicted the results of the statistical evaluation."

Response to Rebuttal

Agreed. The phrase was deleted.

Comment 3: p. 4-1, §4.1

Table 4-1 indicates that arsenic was detected at 6MW11S at 12.6 micrograms per liter in Round 12 (June 2001). This is a striking anomaly, in that arsenic has been non-detect at this location for the previous 11 rounds. Please check field records, chain-of-custody records, etc., to verify to the extent possible that the sample analyzed for this round was indeed collected from well 11S, rather than from 11D, which exhibits

nearly identical arsenic (12.7 micrograms per liter). (It is noted that other inorganic analytes appear to be distinctly different; e.g., iron was analyzed at 340 micrograms per liter for 11S and at 7890 micrograms per liter for 11D, arguing against a duplicate analysis.) Also, the lab should be asked to verify that there were no errors in entering data in the reporting tables.

Response to Original Comment (January 8, 2002)

Agree. The arsenic result for 6MW11S appears to be an anomaly; however, as discussed below, the cause of the anomaly is not apparent and there is no justification for changing the result. Therefore, the result will be maintained and a sentence will be added to the report that identifies the anomaly. Future monitoring results will help determine if this result is truly an anomaly.

Based on a review of the raw data, the result as reported by the laboratory of 12.6 µg/L is correct. All of the metals results for this sample were also compared to others in the same sample-delivery-group (SDG) and it does appear to be a different sample than any of the others that are in this SDG. The only thing noticeable from looking at the raw data relates to the fact that two replicates were done for the Inductively Coupled Plasma (ICP) analysis and the results were averaged. Some labs do more than two replicates but it is common practice to perform multiple replicate analyses for each analyte and report the average. The replicate analyses are done within seconds of each other and make for a more accurate result. In this case, the first replicate was 9.77 µg/L and the second one was 15.5 µg/L with a percent relative standard deviation (%RSD) of 32.086. The difference between the two replicate results can be due to several things such as sample matrix or instrumental problems. However, many of the other analytes in this sample had better %RSDs so it does not appear that there were any general instrument problems with this analysis.

The result could also be a false positive due to contamination from a variety of different sources. In addition, an instrumental baseline for arsenic might have fluctuated slightly from sample to sample or something present in the sample (either from the laboratory or the field) could change the sample matrix enough so that the replicates showed the poor reproducibility.

Another explanation could be due to the high levels of sodium that were present in the sample. Sodium is not an interfering analyte; however, it is possible that the high levels of sodium might have interfered with some of the analytes, producing false positives. If there is a large amount of any one analyte present, it could cause a "loading effect" on the plasma and interfere with the reading of other analytes. This could have happened in this case with arsenic. If this is the case, then the arsenic result could be reported from the 10X dilution for sodium, giving a result of 50 U.

<u>EPA Rebuttal (January 30, 2002)</u>

Specific Comment 3: It is not appropriate to calculate %RSD for two data points. Relative Percent Difference (RPD) would be more meaningful. In this case, the RPD between the two replicates is 45%. A high variation between replicates is frequently caused by improper peaking of the instrument or a broad-based interference in the vicinity of the analytical line of interest. In this instance the later is more likely since this level of arsenic is not expected. Improper peaking would explain a high variance between sample replicates but it would not account for the relatively high level of arsenic that was measured. A further review of the raw data associated with the sample to determine whether a matrix spike or analytical spike was

run on the sample could illuminate the issue. These would have been impacted by either of the data quality issues discussed above.

Response to Rebuttal

The initial result of 12.6 micrograms per liter for arsenic in sample DRMO-6MW11S-GW-12 as reported by the laboratory was verified by reviewing all of the associated raw data. There were no matrix spikes, post digestion spikes, sample duplicates, or serial dilutions performed by the laboratory on this sample to evaluate. The only information that could be evaluated was the raw data analysis for this sample. It is nearly impossible to determine without a doubt the cause of the anomaly based on the raw data. The scenarios presented in the initial response were possibilities that could have caused an erroneous result, in addition to those presented in Specific Comment 3.

The high variance between the two replicates for arsenic for this sample was not meant to explain the high level of arsenic that was measured for this sample. It was noted to present the possibility that there could have been some factors present that could have caused an erroneous result. All of the numbers mentioned in the original response came from the raw data. The laboratory's software associated with the analytical instrument calculates, from the two measured replicates, the average, standard deviation and %RSD. The average of the results is what is reported as the final value. These numbers, from the instrumental raw data analysis, are the only things that were available. Based on this limited information, there is no way to determine precisely what may have occurred at what stage to produce this high result for arsenic.

Based on a review of the available field data (sample log sheet, turbidity data, and chain-of-custody), it is not apparent that the anomalous result was the result of a field problem.

As stated previously in the original response, the result will be maintained and a sentence will be added to the report that identifies the anomaly. Future monitoring results will help determine if this result is truly an anomaly.

Comment 4: p. 4-4, §4.3.1, ¶ 2

The fourth sentence in this paragraph states: "Parametric ANOVA tests assume that the data are normally or lognormally distributed." Please expand the text here to include other important assumptions that parametric ANOVA methods make. For example, parametric methods assume that the data sets being compared not only have a normal or lognormal distribution, but that the data sets have the same distribution type and variance.

Response to Original Comment (January 8, 2002)

The suggested text was not included because the methodology was not testing the assumption of homogeneity of variance. The paragraph will be changed in future reports (i.e., Year 4 and subsequent reports) to the following:

"The Analysis of Variance (ANOVA) technique was the basic approach used to compare data from upgradient and downgradient monitoring well locations. The ANOVA technique is used to test whether there is statistically significant evidence of contamination. There are two types of ANOVA tests: parametric and non-parametric. The parametric ANOVA method makes two important assumptions: 1) the data are both normally (or both lognormally) distributed, and 2) the group variances are homogeneous. If either of these crucial assumptions to the parametric ANOVA are violated, a non-parametric test can be conducted using the ranks of the observations rather than the original observations themselves. These assumptions can be checked using the Shapiro-Wilk Test of Normality and Levene's test of Homogeneity of Variance, respectively. If the analysis of the data demonstrated that these assumptions, critical to the parametric ANOVA, were violated, non-parametric ANOVA techniques were conducted using the ranks of the observations rather than the observations themselves. The Wilcoxon Rank-Sum test (also known as the Mann-Whitney U test) was employed as the non-parametric ANOVA for comparing the downgradient results to the upgradient results."

EPA Rebuttal (January 30, 2002)

General Comment 3 and Specific Comment 4: Although the recommended statistical steps are reasonable to address the comment, based on the reduction in sampling frequency and the elimination of the three deep wells, it is unclear how the Navy plans to use ANOVA techniques to compare upgradient wells to downgradient wells in year 4 and beyond. Two upgradient wells and five downgradient wells will be sampled one time each year. This number of data points is not sufficient to perform an ANOVA test. Please describe in greater detail how the comparison on this limited data will performed.

It should be noted that to date, the comparison of upgradient and downgradient wells has been something of an academic exercise since no primary monitoring criteria have been exceeded.

Response to Rebuttal

Please refer to the response provided to the EPA's rebuttal to the original response to General Comment 3. The Year 3 data were evaluated using the new statistical approach. Extensive revisions were made to the text and tables associated with Section 4.3 of the report to address the comment. The text changes suggested in the original response to the comment were incorporated as appropriate.

Comment 5: p. 4-5, §4.3.2.1

The last sentence of the paragraph beginning "A 'W' statistic ..." states: "...If $W_{calc}<W_{test}$, then the null hypothesis is rejected, H_A is accepted, the data are not assumed to be normally distributed (or not lognormally distributed if log transformed data are tested)...." However, Tables 4-4 and 4-5 indicate that lognormality was selected as the distribution type if the test for normality failed regardless of whether W_{calc} was greater than or less than W_{test} for the transformed data. Please clarify this apparent discrepancy.

Response to Original Comment (January 8, 2002)

The paragraph will be changed to the following:

> "A 'W' statistic (W_{calc}) is computed for a data set and compared to a test statistic (W_{test}). If $W_{calc} \geq W_{test}$, then the null hypothesis is not rejected and the data are assumed to be normally distributed. If $W_{calc} < W_{test}$, then the null hypothesis is rejected, H_A is accepted, the data are not assumed to be normally distributed. Another 'W' statistic is computed for the log-transformed data set and compared to the test statistic as described above. If both the normal and lognormal W_{calc} are greater than or equal to the W_{test} then the underlying distribution is considered to be that one producing the highest W_{calc} value. If neither 'W' statistic is greater than or equal to the test statistic the underlying distribution is defaulted to lognormal. This is because "EPA's experience with environmental concentration data, and groundwater data in particular, suggests that the Lognormal distribution is generally more appropriate as a default statistical model than the Normal distribution... " (EPA, 1992)."

<u>EPA Rebuttal (January 30, 2002)</u>

Specific Comment 5: Please clarify the suggested text. The first sentence of the proposed text indicates that if Wcalc for the untransformed data is greater than or equal to Wtest, then the data are assumed to be normally distributed. However, the text goes on to say "...If both the normal and lognormal Wcalc are greater than or equal to the Wtest then the underlying distribution is considered to be that one producing the highest Wcalc value...." Please clarify this conflict.

<u>Response to Rebuttal</u>

The paragraph will be changed to the following:

"A 'W' Statistic (W_{calc}) will be computed and compared to the test statistic (W_{test}) for both the data set and the log transformed data set. If both the normal and lognormal W_{calc} are greater than or equal to the W_{test} then the underlying distribution is considered to be the one producing the highest W_{calc} value. If only one W_{calc} value exceeds the W_{test} value the underlying distribution is the one that exceeds the W_{test} value. If neither 'W' statistic is greater than or equal to the test statistic or if the W_{calcs} are equal for normal and lognormal, the underlying distribution is defaulted to lognormal. This is because "EPA's experience with environmental concentration data, and groundwater data in particular, suggests that the Lognormal distribution is generally more appropriate as a default statistical model than the Normal distribution..." (EPA, 1992).

Comment 6: p. 4-7, §4.3.2.2

The five equations at the bottom of the page are not in the correct order. Please move the equation for the grand mean to the end of the list.

Response to Original Comment (January 8, 2002)

Disagree. The equations are correct and their order is arbitrary. No changes are suggested in response to this comment.

EPA Rebuttal (January 30, 2002)

Specific Comment 6: Reordering the equations would be clearer because the equation for the grand mean uses variables that, as the text is written, have not yet been defined. However, since the technical conclusions of the document are not affected, the no change is required.

Response to Rebuttal

Agree. The equations will be reordered as requested.

Comment 7: p. 4-8, §4.3.2.2

The table at the bottom of the page appears to have some errors:
1. The first entry in the "Sum of Squares" column should read SS_{sample}.
2. The entries in the "Degrees of Freedom" column should just be "k-1", "N-k", and "N-1" respectively.
3. The "Mean Squares" column should read "$MS_{sample}=SS_{sample}/(k-1)$" in the first row and "$MS_{Error}=SS_{Error}/(N-k)$" in the second row.

Response

The typographical errors will be corrected. In the future, the table will appear as follows:

ONE-WAY PARAMETRIC ANOVA TABLE

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares	F
Between Locations	SS_{Well}	k-1	$MS_{Well}=SS_{Well}/(k-1)$	$F=MS_{Well}/MS_{Error}$
Error (within Locations)	SS_{Error}	N-k	$MS_{Error}=SS_{Error}/(N-k)$	
Total	SS_{Total}	N-1		

Comment 8: p. 4-9, §4.3.2.3, ¶ 2

The second sentence in this paragraph states: "However, parametric ANOVA methods make a key assumption; the results are normally (or lognormally) distributed." Please expand the text here to include other important assumptions that parametric ANOVA methods make. For example, parametric methods assume that the data sets being compared not only have a normal or lognormal distribution, but that the data sets have the same distribution type and variance.

Response

As explained in an earlier response, this wording was not included because the methodology was not testing the assumption of homogeneity of variance. This sentence will be expanded to the following:

> "However, parametric ANOVA methods make two key assumptions: 1) the data are both normally (or both lognormally) distributed, and 2) the group variances are homogeneous."

Comment 9: p. 4-11, §4.3.3

The text notes that the statistical analysis indicates that only TCE shows a significant change from upgradient to downgradient wells, and that TCE *decreases*. This is due to detections of TCE at 6MW6D, the upgradient, bedrock well. Although well below the primary and secondary monitoring criteria, this detection may bear on upgradient detections of chlorinated solvents (Area A Downstream Watercourses). It is appropriate that this well should remain in the monitoring program, as recommended on p. 5-3. Alternatively, at some point, it may be reasonable to move this well to long-term monitoring for another program.

Response

Comment noted.

Comment 10: p. 4-12, §4.3.3

In addressing the potential impact of the site on arsenic in groundwater, the Report notes that most of the elevated arsenic downgradient of the site is in the deep wells, and that the vertical gradients at these locations are upward, arguing against any influence of the site. This is a compelling argument; it is agreed that this, along with the observation that there is an intervening, silty, semi-confining layer, certainly suggests that the deep arsenic cannot reflect a site impact.

Response

Agreed. The data indicate that the significant concentrations of arsenic have been found in the deep downgradient wells and do not suggest an impact by the site. The deep downgradient wells have been recommended for elimination from the groundwater monitoring program to eliminate unnecessary monitoring.

Comment 11: p. 4-13, §4.3.3

The Report states that Year 3 data continue to show a difference in ORP between shallow upgradient wells and shallow downgradient wells, but also notes that significant differences in inorganics concentrations were not found. The Report concludes that "... the chemical data does not suggest that there are any anthropogenic sources influencing the mobility of metals." It is agreed that whatever redox impacts the site may exert on groundwater do not seem to be triggering significant impacts to metals. However, while impacts on the inorganics of potential concern are minimal, it is an overstatement to claim, "This data does not indicate any influence of landfill leachate on downgradient groundwater." There certainly seems to be a significant difference in ORP across the site, and, as noted in the text, in pH, as well. The ORP change is from moderately oxidizing conditions in shallow upgradient wells to strongly reducing conditions in shallow downgradient wells. This change may be due to interaction of groundwater with organic materials beneath the cap; it is also possible that this is simply a result of passing through the domain of organic-rich, estuarine sediments beneath the site, and would occur independent of the historic site use. In this regard, it is noted that 6MW1S, which is south of the cover, shows ORP comparable to the upgradient wells (e.g., 104.3 mV in Round 12, compared to 85.1 mV and 147.2 mV at 6MW9S and 6MW6S, respectively). This certainly leaves open the possibility that the low ORP observed in downgradient wells west of the cap reflects an influence of the site. It is also noted that the cap itself will tend to perpetuate these conditions, as it inhibits recharge of oxygenated water to the shallow aquifer. The text also notes that pH increases across the site (e.g., 5.55–6.03 in 6MW6S and 6MW9S, upgradient, compared to 6.99–7.86 in 6MW10S and 6MW11S, west of the cap; Table A-1), and states that this supports the conclusion that the groundwater does not show indications of landfill leachate. This conclusion is less convincing; landfill leachates do not consistently lower pH, as implied in the text. Another interpretation of the data is simply that water on the upgradient side of the site is recently infiltrated (showing, e.g., the effects of acid rain), and that it is buffered along the flow path beneath the cap.

Response

The comment provides several valid reasons for the distribution of ORP found in the groundwater at the DRMO and provides alternative interpretations of the data as compared to the report. This information will be incorporated into the text of the final report to provide a thorough interpretation of the data. However, it is important to note that the points made in the comment do not contradict the main conclusion of the report (i.e., the remaining landfill material at the DRMO is not enhancing the mobility of metals or increasing the migration of metals from the site) or suggest that further evaluation of ORP is necessary. The USEPA provides concurrence with this statement in Comment No. 15.

Comment 12: p. 4-14, §4.3.3

The text concludes, "The correlations for arsenic and barium are moderate at best and are not indicative of a significant cause/effect relationship between metals concentrations and the low ORP detected in downgradient wells. Therefore, the low ORP values appear to be related to natural conditions at the DRMO and not anthropogenic influences." The last statement does not follow from the first. The correlation between ORP and metals concentrations does not indicate anything with regard to the origin of the low ORP conditions in the downgradient wells. It is acknowledged that it is difficult to draw unequivocal conclusions with regard to the origin of the decreasing ORP across the site. However, the fact that 6MW1S, which is close to the

Thames shoreline, but south of the cap, shows ORP comparable to the upgradient shallow wells certainly suggests that the presence of buried organics (whether anthropogenic or natural) and the cutoff of recharge result in depletion of oxygen beneath the cap. The statistical analysis of correlations between ORP and metals cannot elucidate the relationship between the site and redox conditions. However, the analysis does seem to suggest that metals concentrations are not significantly impacted by the ORP change across the site. In the deep wells (e.g., 6MW11D), elevated arsenic does appear to be associated with reducing conditions (12.7 micrograms per liter As in Round 12, 7890 micrograms per liter iron, ORP -140.7 mV, DO 0.11 mg/L). Under the conditions observed in shallow downgradient wells, iron (and associated arsenic) would be expected to be mobile, if present in the system. Their relatively low levels in shallow, downgradient groundwater might suggest that they simply are not present in significant quantities in sorbed phases on the aquifer solids. (Note that iron is non-detect in the high-ORP wells (6MW01S, 6MW06S, and 6MW09S), and present at measurable concentrations, but still quite low, in the highly reducing, shallow, downgradient wells (6MW10S, 6MW11S). Arsenic is correspondingly non-detect in all of these wells, regardless of ORP, as reflected in the statistical analysis. An exception is 6MW11S, which showed an anomalous detection of As in Round 12 only. (Please see related Specific Comment.)

Response

The last sentence will be changed as follows:

> "Therefore, the low ORP values in the downgradient wells, whether caused by natural or anthropogenic influences, do not seem to be enhancing the mobility of metals or the migration of them from the site."

Comment 13: Table 4-1

The second page of this table does not have a column listing the chemical parameters. Please revise the table.

Response

Agreed. The table will be corrected and provided in the final version of the report.

Comment 14: Tables 4-4 & 4-5

The last column in both of these tables is labeled "95% UCL." In actuality, the numbers in this column appear to be the lesser of the maximum detected value and the 95% UCL. These numbers would typically be used as estimates of the exposure point concentrations (EPCs). Please change the column header to more appropriately identify the numbers listed. It is also unclear what these numbers are used for in the analysis.

Response to Original Comment (January 8, 2002)

Agree with clarification. The term exposure point concentration has been used for these values on other projects. Some reviewers have objected to this term because it has a risk assessment

connotation to it. If USEPA is comfortable with the nomenclature, it will be used in the future. These values are presented for comparison purposes only.

EPA Rebuttal (January 30, 2002)

Specific Comment 14: Please indicate on future tables that the number in the column is the lesser of the maximum detected value and the 95% UCL. This can be accomplished by adding a footnote to the last column that is currently entitled "95% UCL."

Response to Rebuttal

Agreed. A footnote will be added to the tables.

Comment 15: §5.2

Section 5.2 presents recommendations for changes in the monitoring program. The following are comments on these recommendations:

The Report recommends reduction in sampling frequency from quarterly to annual. This is well motivated. No primary monitoring criteria have been exceeded for any COC in the first three years of monitoring. The system as a whole appears to be changing very slowly, if at all, at this time. Annual monitoring is adequate to verify continued low levels of site COCs.

The Report recommends re-evaluation of the monitoring frequency after two more years of annual sampling. The Five-Year Review is an appropriate time to assess the extent and frequency of monitoring necessary for the ensuing period.

The Report recommends that deep monitoring wells 6MW2D, -10D, and -11D be dropped from the program. It is agreed that these wells have not shown evidence of site impacts, and are not likely to show such impacts. The wells are screened quite deep (between 77-87, 43.4-53.4, and 75-85 ft bgs, respectively), below what appears to be a semi-confining layer. Any contaminants related to the DRMO site (with the exception of DNAPL, for which there has been no evidence) would be expected to remain in shallow groundwater, due to the underlying silty confining layer and the close proximity of groundwater discharge to the Thames River. It is also agreed that 6MW6D, the bedrock well on the upgradient side of the DRMO site, should remain in the monitoring program. This well has shown historical low-level detections of chlorinated VOCs (5.5 — 10 micrograms per liter TCE and 3.2 — 7.0 micrograms per liter cis-1,2-DCE in Rounds 9 — 12), and may provide significant insight into chlorinated VOC issues upgradient of this area in the Area A Downstream Watercourses.

The Report recommends that analysis of inorganics and ORP be discontinued. It is agreed that a good effort has been made to investigate this question, and that it has yielded improved understanding of controls on mobility of inorganics at the site. The overall conceptual model for the predominant transport processes at the site is unlikely to be changed significantly by further evaluations of the sort already undertaken. This is particularly justified to the extent that inorganics have not proven to be an obstacle to meeting the primary monitoring criteria established for the DRMO. These criteria have been met consistently for the first 12 rounds of monitoring over the past three years.

Response

Comment noted. The Navy will continue to discuss the extent and frequency of the monitoring program at the DRMO with the USEPA and CTDEP. The recommendations provided in the Year 3 Annual Report will be implemented during the next (i.e., 14th) round of sampling at the DRMO. Because the 13th Round of sampling was completed in September 2001 and the monitoring frequency will change from quarterly to annual, the 14th Round of sampling will be conducted in September 2002.

Comment 16: Table A-1

It is noted that Table A-1, which summarizes field parameters from Rounds 9–12, as been added to the Report, and that it highlights suspect results (*e.g.*, a number of DO measurements in Round 10). This is a useful reference table, and acknowledges clearly the questionable DO data noted in a previous EPA review.

Original Response (January 8, 2002)

Comment noted. This table will be provided in subsequent annual groundwater monitoring reports.

Additional Response

Table A-1 was revised for the final report to provide a complete summary of water quality data collected for the groundwater monitoring program at DRMO. The table now includes water quality data collected during Rounds 1 through 12.