N00102.AR.002983
NSY PORTSMOUTH
5090.3a

TRANSMITTAL LETTER AND U S NAVY RESPONSES TO MAINE DEPARTMENT OF ENVIRONMENTAL PROTECTION COMMENTS ON THE DRAFT MEMORANDUM ON RE-EVALUATION OF FACILITY BACKGROUND REPORT NSY PORTSMOUTH ME
2/10/2014
TETRA TECH

TETRA TECH

PITT-02-14-026

February 10, 2014

Project Number 112G02100

Mr. Matthew Audet
USEPA, Region 1
5 Post Office Square
Suite 100
Mail Code OSRR07-3
Boston, Massachusetts 02109-3912

Mr. Iver McLeod
Maine Department of Environmental Protection
State House Station 17
Augusta, Maine 04333-0017

Reference: Contract No. N62470-08-D-1001 (CLEAN)
Contract Task Order No. WE13

Subject: Responses to Comments on Draft Memorandum on Re-Evaluation of Facility Background Report
Portsmouth Naval Shipyard (PNS), Kittery, Maine

Dear Mr. Audet/Mr. McLeod:

On behalf of the U.S. Navy, Tetra Tech is pleased to provide to U.S. Environmental Protection Agency Region I (USEPA) and Maine Department of Environmental Protection (MEDEP) 2 and 3 copies, respectively, of the responses to USEPA comments dated June 27, 2012 and July 10, 2012, MEDEP comments dated July 30, 2012, and MEDEP follow-up comments dated January 8, 2014 on the subject document. USEPA indicated no further comments in an email dated January 24, 2014. An electronic copy of the responses to comments is also being submitted via e-mail.

Comments are requested by **March 11, 2014**.

If you have any comments or questions, or if additional information is required, please contact Ms. Linda Cole at 757.341.2011.

For the Community Restoration Advisory Board (RAB) members; if you have any comments or questions on these issues, they can be provided to the Navy at a RAB meeting, by calling the Public Affairs office at 207.438.1140 or by writing to:

Portsmouth Naval Shipyard
Public Affairs Office
Attn: Danna Eddy
Portsmouth, NH 03804-5000

Sincerely,

Deborah J. Cohen, PE
Project Manager

DJC/clm
Enclosure

Tetra Tech, Inc.
661 Andersen Drive, Pittsburgh, PA 15220-2700
Tel 412.921.7090 Fax 412.921.4040 www.tetratech.com

Without Enclosure
Mr. Doug Bogen (e-mail)
Ms. Mary Marshall (e-mail)
Mr. Peter Britz (e-mail)
NH Fish & Game (D. Grout) (e-mail)
ME Dept. of Marine Resources (D. Nault) (e-mail)
Dr. Roger Wells (e-mail)
PNS Code 100PAO (e-mail)
Ms. Diana McNabb (e-mail)
Mr. Jack McKenna
Lisa Joy (e-mail)
Paul Dombrowski (e-mail)
NIRIS RDM (e-mail and hard copy)
NOAA (K. Finkelstein) (e-mail)
U.S. Fish and Wildlife (K. Munney)(e-mail)

Hard Copy and electronic copy via e-mail
NAVFAC MIDLANT. (Code OPTE3/L. Cole)
NAVFAC MIDLANT PWD ME (Code PRN4, M. Thyng)

RESPONSES TO USEPA COMMENTS DATED JUNE 27, 2012
DRAFT RE-EVALUATION OF FACILITY BACKGROUND REPORT TECHNICAL MEMORANDUM
PORTSMOUTH NAVAL SHIPYARD, KITTERY, MAINE

(The Navy provided draft responses to USEPA comments on April 29, 2013. USEPA's email dated January 24, 2014 indicated USEPA had no further comments.)

General Comments

1. **Comment:** The Navy is proposing to establish background concentrations for brackish groundwater based on a very limited data set of eight samples (four samples each collected from two wells). A database of only two wells is insufficient to establish background conditions. Similarly, the freshwater groundwater dataset is being comprised of only four wells, each having four rounds of samples collected. Navy's *Guidance for Environmental Background Analysis Volume III: Groundwater* states in Section 2.1.5: *"The background analysis and comparative statistical methods presented in this guidance document typically require chemical data representing at least 10 groundwater monitoring wells located in both impacted and unimpacted zones, and a series of semiannual or quarterly sampling events."* Also Section 1.6.2 states: *"Variations in chemical characteristics of groundwater over time must also be considered for background groundwater analysis. Temporal variations are typically due to seasonal fluctuations in aquifer recharge rates. Semiannual or more frequent sampling over several years is required to accurately characterize seasonal trends."*

   Response: The groundwater background data were collected in 1996 and 1997 prior to the release of the current Navy Guidance published in 2003. The purpose of this Technical Memorandum was to update the summary statistics characterizing the background data based on updated statistical procedures for the summary statistics since the Facility Background Report (Tetra Tech, May 2000) was prepared. The Navy does not plan to collect any additional background data. An uncertainty section will be added to the Technical Memorandum for the Re-Evaluation of Facility Background Report and the limited number of samples comprising the groundwater background data set will be discussed.

2. **Comment:** The Navy has provided no documentation to demonstrate that the two wells identified as brackish are statistically different from the four freshwater wells. Given the relative locations of brackish and freshwater wells, the distinction appears inappropriate. Please include appropriate justification for distinguishing between these groups of wells.

   Response: The purpose of the Technical Memorandum was to update the summary statistics presented in the Facility Background Report. In the 2000 Facility Background Report, salinity was the major factor used to distinguish between wells and summary statistics were calculated separately for each data set; therefore, the updated summary statistics were also calculated for each data set. Additionally, even though the data sets were not evaluated to determine if they were statistically different, concentrations of chemicals from the freshwater and brackish wells represent different populations and therefore it would not be appropriate to calculate statistics combining two separate populations.

3. **Comment:** The background groundwater dataset should also include geochemical data for temperature, redox potential, dissolved oxygen, and specific conductivity so that appropriate

comparisons can be made among background wells and with site wells when evaluating for COPCs.

Response: Field measurements (pH, specific conductance, temperature, turbidity, dissolved oxygen, salinity, and Eh) for groundwater samples were presented in Appendix B of the Facility Background Report. The data are included in Table 3-2 of the Groundwater Monitoring Summary Report (December, 1996 – November, 1997) (Tetra Tech, August 1999). A reference to these data will be provided.

Specific Comment

1. **Comment:** Table 2: There is a typographical error for the 95% UCL for pH; it should not be 74.

   Response: The 95 percent upper confidence limit (UCL) for pH is 7.4. Table 2 will be updated to show that the UCL is 7.4.

RESPONSES TO USEPA COMMENTS DATED JULY 10, 2012
DRAFT RE-EVALUATION OF FACILITY BACKGROUND REPORT TECHNICAL MEMORANDUM
PORTSMOUTH NAVAL SHIPYARD, KITTERY, MAINE

(USEPA comments dated July 10, 2012 are similar in nature to the June 27, 2012 comments. Please see the responses to USEPA comments dated June 27, 2012 for additional information for similar comments.)

1. **Comment:** Navy is proposing to establish background concentrations for brackish groundwater based on a very limited data set of eight samples (four samples each collected from two wells). A database of only two wells is insufficient to establish background conditions. Similarly, the freshwater groundwater dataset is also small being comprised of only four wells, each having four rounds of samples collected.

   Response: An uncertainty section will be added that will discuss uncertainty associated with the limited number of samples comprising the groundwater background data set.

2. **Comment:** Navy has provided no documentation to demonstrate that the two wells identified as brackish are statistically different from the four freshwater wells. Given the relative locations of brackish and freshwater wells, the distinction appears inappropriate. Please include appropriate statistical justification for distinguishing between these groups of wells.

   Response: Salinity was the major factor for distinguishing between the brackish and freshwater wells.

3. **Comment:** The background groundwater dataset should also include geochemical data for temperature, redox potential, dissolved oxygen, and specific conductivity so that appropriate comparisons can be made among background wells and with site wells when evaluating for COPCs.

   Response: Geochemical data were presented in Appendix B of the Facility Background Report, which was taken from the Groundwater Monitoring Summary Report. A reference to the geochemical data will be provided in the Technical Memorandum.

4. **Comment:** As a result of not excluding any potential outliers from the dataset plus the addition of location BGS-26, the calculated background concentrations proposed in the re-evaluation document have changed; the majority are somewhat greater in the re-evaluation document but some values are lower. The most notable differences are for pesticides and zinc for which the proposed background values increased significantly and for some PAHs for which the background values decreased significantly.

   Response: Comment noted.

5. **Comment: Table 2.** There is a typographical error for the 95% UCL for pH; it should not be 74.

   Response: The 95 percent UCL for pH will be updated to 7.4 in Table 2.

RESPONSES TO MEDEP COMMENTS DATED JULY 30, 2012 AND FOLLOW-UP COMMENTS DATED JANUARY 8, 2014
DRAFT RE-EVALUATION OF FACILITY BACKGROUND REPORT TECHNICAL MEMORANDUM
PORTSMOUTH NAVAL SHIPYARD, KITTERY, MAINE

(The Navy provided draft responses to MEDEP comments on April 29, 2013. MEDEP follow-up comments on the responses to comments 1, 3, 6, 9, and 12 and an additional comment were received on January 8, 2014. The following provides the July 30, 2012 comments, the April 29, 2013 Navy responses, and the January 2014 MEDEP follow-up comments and Navy responses.)

1. **July 30 2012 Comment:** Previously MEDEP has indicated the two biggest problems with the background data analyses are the improper use of non-detect values and the handling of statistical outliers. We discussed these issues during our June 29, 2012 conference call. The Navy indicated they would show the range of NDs on the report's histograms using different colors or markers to distinguish them. The Navy also stated they would provide detailed justification for including outliers in the dataset. Further discussion of these issues is included in the comments below. Several comments address more specific issues.

   April 29, 2013 Response: The methods mentioned by MEDEP for handling non-detects during the June 29, 2012 conference call still require that a numeric value be entered for these values. The detection limit was entered and noted as a non-detect in the statistical software used to calculate the summary statistics and graphical displays presented in the background re-evaluation. This includes calculations of UCLs, upper prediction limits (UPLs), and upper tolerance limits (UTLs) using USEPA software ProUCL version 4.1.00. The graphical displays in USEPA's Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities (2009) were constructed using the detection limit for non-detected concentrations. Please see the Navy's response to MEDEP Comment No. 9 for details on the changes the Navy made to the graphical displays. For further justification of including the outliers in the dataset please see the Navy's response to MEDEP Comment No. 12.

   January 8, 2014 Follow-up Comment: Response to Comment 1. MEDEP is satisfied with the identification of non-detects in the graphs and the use of ProUCL and the identification of non-detects and their detection levels. This should provide the most appropriate UCLs for each dataset. There are still some issues where there are too many non-detects to calculate a UCL or any other statistics. These should be clearly stated in the results and tables.

   February 10, 2014 Response: Uncertainty regarding non-detects will be noted in the technical memorandum as provided in the April 29, 2013 response to MEDEP Comment No. 11.

2. **July 30, 2012 Comment:** Please submit the data in the Maine EDD format with as many fields completed as possible.

   April 29, 2013 Response: The Navy's background dataset, available in NIRIS and provided in Appendix A of the Technical Memorandum, will be provided to MEDEP in the Maine EDD format. No additional data will be entered into the database for the Maine EDD data submission.

3. **July 30, 2012 Comment:** Introduction, third paragraph. The 2,3,7,8-TCDD TEQ and benzo(a)pyrene equivalent concentrations were calculated by 1) using only the detected

concentrations and 2) using detected concentrations and ½ the detection limit. Neither method is currently recommended for handling non-detects. We are concerned that given the large percentage of the PAH data with detection limits greater than most of the detected values, that there is no appropriate method for calculating equivalent concentrations. Please see references such as the EPA (2009), Hewett and Ganser (2007), and Helsel (2005, 2012) for acceptable ways to handle non-detects other than data substitution.

April 29, 2013 Response: The referenced documents do not discuss how to handle non-detect concentrations when calculating 2,3,7,8-TCDD toxicity equivalent quotients (TEQs) and benzo(a)pyrene equivalent concentrations or any other calculated chemical parameter. Additionally the calculations of 2,3,7,8-TCDD TEQ and benzo(a)pyrene equivalent concentrations for the background data should be conducted in the same manner as site data at PNS. It is common practice to use one-half the detection limit for 2,3,7,8-TCDD TEQ and benzo(a)pyrene equivalent concentrations and this procedure has been used for site data at PNS. Therefore, no change to the calculation of 2,3,7,8-TCDD TEQ and benzo(a)pyrene equivalent concentrations presented in the Technical Memorandum will be made.

To provide a better understanding of how handling of the detection limits may affect the calculation of 2,3,7,8-TCDD TEQs and benzo(a)pyrene equivalents for the facility background data set, the Navy also calculated UCLs, UPLs, and UTLs by summing the ProUCL outputs for UCLs, UPLs, and UTLs for the individual parameters multiplied by their corresponding toxicity equivalency factor (TEF). These were compared with the outputs from ProUCL for 2,3,7,8-TCDD TEQs and benzo(a)pyrene equivalents based on positive detections only and using one-half the detection limit for non-detects. In addition, benzo(a)pyrene equivalents were calculated using the February 2013 proposed Maine Remedial Action Guidelines (RAGs) background concentrations for the individual parameters to provide a comparison to the various facility background calculated benzo(a)pyrene equivalents. The tables with this information are provided at the end of the responses to comments. These tables show that there is little impact to the facility background concentrations using the two methods.

As shown on the table with the 2,3,7,8-TCDD TEQ calculations, there is little difference between the various calculations of TEQs based on the 95-percent UCLs. The lowest value (6.7 ng/kg) was based on using one-half the detection limit to represent non-detects and the calculation based on summation of the UCL for each parameter had the greatest value (9.2 ng/kg). The UPLs and UTLs are greater than the UCLs, but show the same pattern.

For benzo(a)pyrene equivalents, the various calculations for the facility background data are similar and all in the range of approximately 300 to 800 µg/kg, with the ProUCL output for the UCL being the lowest value. These are all less than equivalents based on the background numbers provided in the February 2013 Maine RAGs proposed revisions. The February 2013 Maine RAGs proposed revisions provide UPLs for rural developed areas, urban developed areas, and urban fill. The UPLs for benzo(a)pyrene ranged from 1,500 to 5,200 µg/kg, and benzo(a)pyrene equivalents that were calculated using the February 2013 UPLs for individual parameters ranged from approximately 2,000 to 13,500 µg/kg. The facility background benzo(a)pyrene equivalents (714 and 514 µg/kg) are also less than New England average urban concentrations of benzo(a)pyrene (1,323 µg/kg) (see the Navy's response to MEDEP Comment No. 12).

January 8, 2014 Follow-up Comment: RTC 2. "The referenced documents do not discuss how to handle non-detect concentrations when calculating 2,3,7,8-TCDD toxicity equivalent

quotients (TEQs) and benzo(a)pyrene equivalent concentrations or any other calculated chemical parameter."

Helsel discusses methods for calculating quantitative statistics, such as maximum likelihood estimation (MLE) and robust ROS. Although he does not specifically mention calculating TEQs, these methods provide ways of estimating missing data that are an improvement over arbitrary data substitution. Unfortunately, these methods require a substantial amount of data above the reporting limit in order to be used.

"It is common practice to use one-half the detection limit for 2,3,7,8-TCDD TEQ and benzo(a)pyrene equivalent concentrations and this procedure has been used for site data at PNS."

While we agree that the practice is common we do not believe it is a statistically appropriate practice. Likewise, continuing an inappropriate practice because it is standard practice is not prudent. Indeed, the Navy revised the background evaluation report in order to update statistical methods used in the evaluation. However, as discussed above, based on MEDEP's PAH background evaluation, the PAH background values the Navy has proposed are reasonable for an area with a long history of industrial activity.

February 10, 2014 Response: No change is necessary for the technical memorandum based on this follow-up comment. As provided in MEDEP's follow-up comment, the background values the Navy has proposed are acceptable. However, the following is provided to clarify that the Navy did not use an inappropriate practice to calculate TEQs for the facility background technical memorandum.

There are various methods that USEPA accepts to handle non-detections when calculating TEQs for individual samples, including using 0, ½ the detection limit, or full detection limit. Helsel (2009) proposes another method using the Kaplan-Meier (KM) statistics to calculate TEQs for individual samples. Use of the KM for calculating TEQs for individual samples has not specifically been approved by USEPA; however, USEPA's ProUCL incorporates KM statistical methods.

As stated in the April 29, 2013 response to MEDEP Comment No. 1, to better understand how handling of detection limits may affect the accuracy of calculation of 2,3,7,8-TCDD and benzo(a)pyrene TEQs for the facility background data set, the Navy compared the mean TEQs calculated in the technical memorandum to calculated TEQs using summations of ProUCL outputs for the individual chemicals. As shown in the April 29, 2013 response to MEDEP Comment No.1, the difference between the output of using ½ ND, positive detections only, and summation of ProUCL outputs for individual chemicals are considered insignificant for this project. Therefore, limitations of the method used for calculation of the TEQ mean for the background data set are perceived to have little effect on the outcome of the data analyses and the method used is considered appropriate and acceptable for this project.

4. **July 30, 2012 Comment:** Figure 1. BGS-22 is off the edge of the map. Please correct this.

   April 29, 2013 Response: Figure 1 will be updated to show BGS-22.

5. **July 30, 2012 Comment:** Table 2. In Table 2, how were averages of data with nondetects calculated? Note that it is inappropriate to use quantitation limits (or any other substitution of data) when calculating summary statistics. Please see EPA (2009), Hewett and Ganser (2007), and Helsel (2005, 2012) for alternative methods.

April 29, 2013 Response: One-half the detection limit was used for the average calculation. This was done to be consistent with how average data are presented in risk assessments for PNS. It is common practice to use one-half the detection limit for the average calculations presented in RAGs Table 3s. To be consistent with the presentation of site data and the background data the average calculations will not be changed. A footnote will be added to Tables 2, 4, and 5 indicating that one-half the detection limit was used in the calculation of the average concentrations. Note that UCLs, UPLs, and UTLs are outputs of ProUCL and were calculated in accordance with ProUCL requirements.

6. **July 30, 2012 Comment:** Table 2. The number of samples for metals, SVOC, VOC, PAH, and pesticide analyses is stated as 25, but we count 24 total samples if BGS-26 is a duplicate of BGS-10.

   April 29, 2013 Response: BGS-26 is not listed in the data base as a duplicate sample and therefore, was not treated as a duplicate. The data and the sample type for each location are presented in Table A.1 of Attachment A. Therefore, no changes are required to the number of samples listed in Table 2.

   January 8, 2014 Follow-up Comment: RTC 6. "BGS-26 is not listed in the data base as a duplicate sample and therefore, was not treated as a duplicate. The data and the sample type for each location are presented in Table A.1 of Attachment A. Therefore, no changes are required to the number of samples listed in Table 2."

   Based on the field sheets and figures from the time the data were collected, the database is incorrect. The field log clearly states BGS-26 is a field duplicate of BGS-10. MEDEP can forward a copy of this information if needed. Sample location maps from the era of the original study show there is no unique location for BGS-26. Following the convention used at PNSY the data for BGS-10 and BGS-26 need to be combined, and the database needs to be corrected for the final report.

   February 10, 2014 Response: In consideration that BGS-26 was not identified as a duplicate sample in previous data evaluations and that inclusion of BGS-26 as a separate sample has negligible impact on the evaluation of the facility background evaluation results, the Navy has elected not to make a change to the database at this time and will make a note of this in the technical memorandum. The following provides the justification.

   Duplicate samples for this project are typically averaged for calculation of dataset averages and other statistics in risk assessment. However, BGS-26 was not identified as a duplicate sample in the data base and therefore, was not averaged with BGS-10 as part of previous data evaluations or in this technical memorandum. Review of the data shows that the results for the two samples are similar (relative percent difference is generally less than 20 percent). The results for BGS-26 and BGS-10 are in the middle to lower end of the range of detections and are less than the 95 percent UCL. Therefore, including BGS-26 as an additional sample in the dataset as is presented in the technical memorandum tends toward a slightly lower average for the facility background dataset. This difference is slight because of the concentrations at BGS-26 and BGS-10 compared to the other 23 background samples. Updating the database and calculations to show BGS-26 as a duplicate would not have a significant impact on the output and usefulness of the background technical memorandum compared to the significant time and effort it would require to redo all of the calculations and data plotting.

7. **July 30, 2012 Comment:** In Table 2, the range of detection limits is incomplete. The range shown is only the range for data that have undetected results, which commonly are an order of magnitude greater than the detected values. Please record the detection limits for all the sample analyses.

   April 29, 2013 Response: The range of detection limits in Tables 2, 4, and 5 are the ranges of detection limits for samples with non-detected results. The titles of these columns will be changed to, "Range of Non-detected Results."

8. **July 30, 2012 Comment:** Table A1. The existence of detection limits that exceed detected values makes us question the usability of the data. It is impossible to quantitatively compare data with detection limits 10 times the values of other data. For example, consider the benzo(a)pyrene data where some of the undetected data have detection limits of 350-400 ppm and other data have detected values of 36-110 J ppm. Is 350 U greater or less than, say, 43 J? We have no way of knowing this, making it very problematic to assign a value to 350 U. Please provide a defensible method for handling data with such high detection limits.

   April 29, 2013 Response: Non-detected concentrations were handled in accordance with USEPA (2009) for the graphical displays and the technical guidance for ProUCL for calculating the 95-percent UCLs, UPLs, and UTLs. The Navy agrees that there is uncertainty with the detection limits of the PAH data. The uncertainty of these detection limits will be discussed in the added uncertainty section of the Technical Memorandum.

9. **July 30, 2012 Comment:** Statistical Methodology. The plotting of substituted data, such as quantitation limits, on graphs can show trends or patterns that have more to do with laboratory methods than environmental concerns. The substitution of values for non-detects will alter the plots depending on the value chosen and will not accurately represent the distribution of parameters in soil. In particular, the depiction of mercury, selenium, bis(2-ethylhexyl)phthalate, benzo(a)pyrene (and other PAH) data may be inaccurate due to the plotting of detection limits along with actual data. A probability plot of data where 22 out of 25 points are detection limits (as with the bis(2-ethylhexyl)phthalate data) will show the distribution of the detection limits not the environmental data. Please see EPA (2009) and Helsel (2005, 2012) for alternative plotting methods.

   April 29, 2013 Response: Chapter 9 of USEPA 2009 discusses the use of non-detects and graphical displays of data. Specifically Section 9.2 Box plots states, "It is important to plot the data as reported by the laboratory for non-detects or negative radionuclide data. Proxy values for non-detects should not be plotted since we want to see the distribution of the original data. Different symbols can be used to display non-detects, such as the open symbols described in Section 9.1. The mean will be biased high if using the RL of non-detects in the calculation, but the purpose of the box plot is to assess the distribution of the data, not quantifying a precise estimate of an unbiased mean. Displaying the frequency of detection (number of detected values/number of total sample) under the station name is also helpful. Unlike time series plots, box plots cannot use missing data, so missing data should be removed before producing a box plot." Also, the example box plot, Example 9-2 and Figure 9-2, non-detected data were used in the construction of the box plot. Section 9.3 Histograms states, "Also, use the data as reported by the laboratory for non-detects and eliminate any missing values, since histograms cannot include missing data." The example histograms, Figure 9-3, was constructed using non-detected data as well.

The graphical displays presented in the outlier assessment will be updated with the Frequency of Detection (FOD) added to the box plot, normal probability plot, histogram, and summary statistics. For the box plots solid squares will be added to the box plots displaying the detection limit of the non-detected results. The non-detected results were already plotted with a different symbol on the normal probability plot. For the histograms a solid square will be placed on the histogram denoting the detection limit of non-detected samples. The average concentration presented in the summary statistics will be removed. The non-detected results will continue to be used in the graphical displays as is consistent with the USEPA guidance.

January 8, 2014 Follow-up Comment: RTC 9. MEDEP appreciates the effort the Navy has made to make the non-detect data more obvious, such as plotting them with a different symbol. These changes will help clarify the limitations of a given comparison between site and background concentrations. However, these limitations are great in that creating a probability plot of mostly non-detect values will give you a good assessment of the distribution of the detection limits, but not the parameter of interest. A note should be added to such datasets that we do not have enough data to make valid statistical assessments and, hence, cannot provide an estimate of background concentrations for those parameters.

February 10, 2014 Response: The technical memorandum has been updated to note the limitations.

10. **July 30, 2012 Comment:** Statistical Methodology, last paragraph. Basic summary statistics were computed after assigning the non-detects ½ the detection limit. At the beginning of this section, it was stated that re-evaluation of the background data was based on updated statistical methods. It should be noted that current statistical practice for environmental data strongly discourages the use of replacing a fixed value for non-detects (e.g., USEPA, 2009; Singh et al, 2010; Helsel 2005) especially for calculating summary statistics. For data that have multiple detection limits, as is the case here, substitution of detection limits for non-detects results in the introduction of unrelated patterns into the data. It is highly recommended that other methods, such as maximum likelihood estimation (MLE), Kaplan-Meier nonparametric method, or regression on order statistics (ROS), be used instead of data substitution (Hewett and Ganser (2007); Helsel (2005, 2012)).

    April 29, 2013 Response: The referenced text states that the detection limit was used, not one-half the detection limit, for the statistical calculations. The 95-percent UCLs, UTLs, and UPLs were calculated using ProUCL, which incorporates the methods referenced for handling nondetect data. As stated in the Navy's response to MEDEP Comment No. 1, non-detects have to be assigned a numerical value. The referenced survival methods recommended by MEDEP have to have a value for the non-detects; the calculations do not ignore non-detected results. Therefore the detection limit reported by the laboratory was input into the software and the graphical displays. This procedure for handling non-detects is consistent with USEPA (2009) and in general the use of the methods referenced by MEDEP.

    One-half the detection limit was only used for the average calculations presented on Tables 2, 4 and 5. The average concentration presented on the graphical displays will be removed and the average concentration presented on Tables 2, 4, and 5 will not be changed as stated in the Navy's response to MEDEP Comment No. 5.

11. **July 30, 2012 Comment:** Statistical Methodology, last paragraph. The Navy states that summary statistics cannot be calculated for data where the majority of values are non-detect, therefore they only substituted the detection limit for non-detects if there were three or more detections. The case where there are three detections, however, is where the majority of values

are non-detect. The USEPA's 2006 Data Quality Assessment: Statistical Methods for Practitioners Guidance discourages the use of data substitution of non-detects for datasets with greater than 15% non-detects. It has been shown that even at non-detection percentages as low as 5-10%, substitution gives inconsistent results compared to other methods (Singh et al, 2006). The results from USEPA's ProUCL software in Appendix B warn that the results are likely to be invalid, yet they are presented in Table 2. MEDEP disagrees with this approach and recommends 10-15 measureable observations for the calculation of UCLs and summary statistics.

April 29, 2013 Response: As stated in the Navy's responses to MEDEP Comment Nos. 1 and 10, the methods referenced by MEDEP for handling non-detected concentrations still include the non-detected results. The detection limits are inputted for these results and they are noted as non-detects. The Navy does agree that there is uncertainty with calculations involving a high percentage of non-detects and the warning messages from ProUCL will be added as a footnote to Tables 2, 4, and 5. The uncertainty of these results will also be discussed in the uncertainty section that will be added to the Technical Memorandum. No decisions are being made based on the summary statistics presented in the Technical Memorandum. The methodology that the Navy used for the background comparisons presented in the Remedial Investigation Reports for Operable Units 7 and 9 will be used for statistical comparison of site and background data.

12. **July 30, 2012 Comment:** Soil, second paragraph. The Navy identifies at least 18 potential outliers based on statistical methods then dismisses them as outliers because they are outside of CERCLA sites. This raises the question of why outlier tests were performed if the conclusion is that no sample result could have been an outlier regardless of what proper statistical analyses show.

    The EPA (2002) defines outliers as, "Measurements (usually larger or smaller than other data values) that are not representative of the sample population from which they were drawn. Outliers distort statistics if used in any calculations." Therefore, the identification of outliers is important so that underlying assumptions in statistical analyses, i.e., all data come from the same population, are valid. Singh et al, 2006, recommended that all outliers be removed before calculating a UCL, UPL, or UTL. Although we do not agree with the automatic dismissal of outliers, we think that the reasons for retaining outliers in estimating background concentrations should be strongly supported.

    Outliers may be the result of errors in sampling, recording, or analysis. There is, however, no assessment of data quality in this report. Please evaluate the possibility that outliers are a result of error and provide a data quality assessment of all the data.

    The identification of outliers is not a simple, clear-cut process. The reasons for identifying outliers should have been established prior to evaluating background data. For example, outliers may be helpful for revealing errors in sampling, recording, or analysis and for identifying data belonging to another population (contaminated). On the other hand, outliers may be extreme values within a population (not uncommon in environmental data). The Unified Guidance (EPA, 2009) recommends assessing the possible causes for outliers, and not automatically removing or retaining them.

    Including outliers in the dataset solely because they were from locations outside CERCLA sites is unacceptable to MEDEP. As discussed during our June 29 conference call the Navy will need to provide additional justification for using these data in the background data set.

April 29, 2013 Response: USEPA's Data Quality Assessment: Statistical Methods for Practitioners (2006) states the following:

"Statistical outlier tests give the analyst probabilistic evidence that an extreme value does not "fit" with the distribution of the remainder of the data and is therefore a statistical outlier. These tests should only be used to identify data points that required further investigation. The tests alone cannot determine whether a statistical outliers should be discarded or corrected within a data set. This decision should be based on judgmental or scientific grounds. There are 5 steps involved in treating extreme values or outliers:

1. Identify extreme values that may be potential outliers.
2. Apply statistical test.
3. Scientifically review statistical outliers and decide on their disposition.
4. Conduct data analyses with and without statistical outliers; and
5. Document the process.

Potential outliers may be identified through the graphical representations of Chapter 2 (step 1 above). Graphs such as box and whisker plots, ranked data plot, normal probability plot, and time plot can all be used to identify observations that are much larger or smaller than the rest of the data. If potential outliers are identified, the next step is to apply one of the statistical tests described in the following sections.

If a data point is found to be an outlier, the analyst may either : 1) correct the data point; 2) discard the data point from the analysis; or 3) use the data point in the analysis. This decision should be based on scientific reasoning in addition to the results of the statistical test. For instance, data points containing transcription errors should be corrected, whereas data points collected while an instrument was malfunctioning may be discarded. Discarding an outlier from a data set should be done with extreme caution, particularly from environmental data sets, which often contain legitimate extreme values. If an outlier is discarded from the data set, all statistical analysis of the data should be applied to both the full and truncated data set so that the effect of discarding observations may be assessed. If scientific reasoning does not explain the outlier, it should be discarded from the data set."

The typical potential chemicals of concern (COPCs) for PNS IRP sites that have concentrations that may be similar to background are arsenic, lead, and PAHs. Therefore, outliers for these chemicals were evaluated further in comparison to available background data for other areas of the United States. Additional rationale is provided below for the soil outliers identified in the Re-Evaluation Technical Memorandum will be added to the outlier discussion in the Technical Memorandum.

One arsenic potential outlier and one lead potential outlier were identified, both at BGS-07. BGS-07 is located in a residential area (Shipyard Commander's Residence) and is away from any IRP Sites and is therefore not likely to represent contamination from a Navy site. Shacklett and Boerngen (1984) list the range of arsenic concentrations from eastern United States soil samples as <0.1 to 73 mg/kg. The potential arsenic outlier of 58.6 mg/kg is within this range; therefore, this concentration is likely representative of anthropogenic background and was retained in the background dataset.

The USEPA's Agency for Toxic Substances and Disease Reigstry (ATSDR, 2007) stated the following about typical anthropogenic lead soil concentrations:

"Anthropogenic sources of lead include the mining and smelting of ore, manufacture of lead-containing products, combustion of coal and oil, and waste incineration. Many anthropogenic sources of lead, most notably leaded gasoline, lead-based paint, lead solder in food cans, lead-arsenate pesticides, and shot and sinkers, have been eliminated or strictly regulated due to lead's persistence and toxicity. Because lead does not degrades, these former uses leave their legacy as higher concentrations of lead in the environment."

"The concentration of soil lead generally decreases as distance from contaminating sources increases. The estimated lead levels in the upper layer of soil beside roadways are typically 30–2,000 µg/g higher than natural levels, although these levels drop exponentially up to 25 m from the roadway (EPA 1986a)."

"In the state of Maine, soil samples taken from areas where the risk of lead contamination was considered high (within 1–2 feet of a foundation of a building >30 years old) indicated that 37% of the samples had high lead concentrations (>1,000 µg/g). In 44% of the private dwellings, high lead levels were found in the soil adjacent to the foundation; high levels were found in only 10% of the public locations (playgrounds, parks, etc.). In addition, the largest percentage (54%) of highly contaminated soil was found surrounding homes built prior to 1950; homes built after 1978 did not have any lead contamination in the soil (Krueger and Duguay 1989). Environmental health studies conducted near four NPL sites measured mean concentrations of lead in soil ranging from 317 to 529 mg/kg, and mean concentrations of lead in dust ranging from 206 to 469 mg/kg (Agency for Toxic Substances and Disease Registry 1995)."

Based on these anthropogenic lead concentrations the potential outlier of 1,100 mg/kg at BGS-07 was retained and concluded to represent anthropogenic background concentrations.

Three PAHs [benzo(a)pyrene, benzo(b)fluroanthene, and benzo(k)fluoranthene] and benzo(a)pyrene equivalents were identified as potential statistical outliers at BGS-14. BGS-14 is located in an undeveloped area of PNS away from any IRP sites and is therefore not likely to represent contamination from a Navy site. The following table shows the concentration of the outlier, the range and average concentrations found in New England Urban Soils (Bradley, 1994) and the February 2013 proposed revisions to Maine RAGs for Urban Developed Areas and Urban Fill. The benzo(a)pyrene, benzo(b)fluroanthene, and benzo(k)fluoranthene outlier concentrations were within the range and similar to the average concentrations reported in New England soil (Bradley, 1994). Except for the benzo(k)fluoranthene urban developed area, the outlier concentrations were less than the urban background values in the February 2013 proposed revisions to Maine RAGs. Therefore, it is likely that these samples represent anthropogenic concentrations and do not need to be removed from the background data set.

Parameter	**PNS Outlier concentration (µg/kg)**	**New England Urban Range (µg/kg)**	**New England Urban Average (µg/kg)**	**MEDEP Urban Developed Area/ Urban Fill Background (µg/kg)**
Benzo(a)pyrene	1100	40-13000	1323	1700/5200
Benzo(b)fluoranthene	1700	49-12000	1435	2000/6800
Benzo(k)fluoranthene	1100	43-25000	1681	760/12000

Dioxin/furan potential outliers were also evaluated; however, background values for dioxins/furans were not readily available. Instead, the outliers were multiplied by their TEFs and compared to the industrial USEPA soil regional screening level (RSL) for 2,3,7,8-TCDD of 18

ng/kg. None of the adjusted dioxin outlier concentrations exceed the screening criteria; therefore there is no reason to remove these concentrations from the background dataset.

Parameter	TEF	PNS Outlier Concentration (ng/kg)	Adjusted PNS Outlier Concentration (ng/kg)
1,2,3,4,6,7,8-HPCDD	0.01	826	8.26
1,2,3,4,6,7,8,9-OCDF	0.0001	129 94	0.0129 0.0094
1,2,3,4,7,8,9-HPDCF	0.01	3.4	0.034
2,3,7,8-TCDD TEQ	--	15.7	15.7

January 8, 2014 Follow-up Comment: RTC 12.

a) There are several references to background studies of a national or regional scale. Based on the direct impact of bedrock and soil types on metals in particular, and the effects of sample collection methods, statewide studies such as the recent USGS work cited in Maine's 2013 RAGs (p. 17, footnote: http://www.maine.gov/dep/ftp/RAGS-Background-Documents/Metals-and-PAH-Background-Study-2012/) are more appropriate, and rely on current sample and analytical methods. MEDEP will consider those studies as more relevant than older, less focused work, especially Shacklett and Boerngen (1984) which covers too broad an area to be relevant, especially as it concerns arsenic. We note that the study referenced in the Maine 2013 RAGs indicated the 90th percentile of arsenic in Maine ranged from approximately 13-16 ppm.

b) The urban fill background values listed with MEDEP's draft RAGs are not applicable to the background data, as the sample locations do not meet the criteria for urban fill listed in the document. Please remove them from the table.

c) MEDEP continues to object to inclusion of the outliers in the database. However, it appears that this is a situation in which MEDEP and the Navy will have to agree to disagree. In the future, we will endeavor to ensure that background comparisons are completed without allowing single outlier data points to drive the conclusions.

February 10, 2014 Response: a) No change is required based on this follow-up comment. It is noted that for the PNS facility background data set, the average is 14.5 mg/kg and the 95 percent UCL is 17.9 mg/kg. b) The table in the response to comment provides USEPA and Maine values for comparison, and the Navy believes that the urban fill values are valuable for this comparison, but agrees that the urban fill values will not be included in the technical memorandum outlier discussion. c) Comment noted. The Navy will continue to work with MEDEP to try to address MEDEP's concerns during future site-specific investigation and data evaluation.

13. **July 30, 2012 Comment:** Soil, 2nd paragraph. The reason given for this re-evaluation of background concentrations was to update statistical methods used in the previous background assessment report, but apparently the evaluation criteria for assessing outliers has also changed. Please explain what has changed between 2000 and now that would allow for a different conclusion on what outliers represent at this site?

April 29, 2013 Response: The purpose of the Technical Memorandum was to update the methodology for identifying outliers and computing UCLs, UPLs, and UTLs that were provided in the Facility Background Report to be consistent with the methodology that the Navy is currently using (e.g., methodology used for the RI Reports for Operable Units 7 and 9). In the Facility Background Report, outliers were identified as concentrations greater than two standard deviations exceeding the mean and were eliminated from the dataset without investigating the outliers. In the Technical Memorandum statistical hypothesis tests were used to identify potential outliers and any potential outliers were investigated for reasons they should be removed from the data sets as stated in USEPA's Data Quality Assessment: Statistical Methods for Practitioners (2006). The UCLs, UPLs, and UTLs presented in the Technical Memorandum were also re-calculated using the updated USEPA methodology for calculating UCLs, UPLs, and UTLs.

14. **July 30, 2012 Comment:** Soil, 4-Hydroxy-4-Methyl-2-Pentanone concentrations. We noticed that 4-Hydroxy-4-Methyl-2-Pentanone concentrations are unusually high and suspect that the lab has misidentified the compound. Please review the 4-Hydroxy-4-Methyl-2-Pentanone data and provide an explanation.

    April 29, 2013 Response: The Facility Background Development report listed 4-Hydroxy-4-Methyl-2-Pentanone as a tentatively identified compound. This result will be removed from the Table 2 and the outlier analysis. The following footnote will be added to 4-Hydroxy-4-Methyl-2-Pentanone in the database tables in Appendix, "tentatively identified compound not used in formal background comparisons or statistical evaluations."

15. **July 30, 2012 Comment:** Groundwater. Groundwater rarely has aluminum concentrations greater than 50 ppb, so four values greater than 800 ppb should raise a red flag. Why are these concentrations so high? Are the data valid? Were the metals samples filtered or is it possible that clay particles were included in the water sample bottles?

    April 29, 2013 Response: The data were validated. The four results referenced were from the four samples collected from the deep bedrock well at SI-01. The results for the total aluminum samples ranged from 996 to 1070 ug/L. The filtered results were less and ranged from 201 to 547 ug/L.

16. **July 30, 2012 Comment:** No spatial assessment of the concentrations was undertaken in this report. Evaluation of the distribution of potential contaminants is important in determining whether outliers indicate possible hotspots or randomly located high concentrations. For example, concentrations for most PAHs at BGS-14 were generally 1-2 orders of magnitude greater than most other locations, a strong indication that the location is contaminated well above typical background PAH values. We note that BGS-22 has PAH concentrations similar to BGS-14. However, BGS-22 is located at the edge of a parking lot where one would expect to see high PAH concentrations. BGS-14, however, is located in an undeveloped location on the Shipyard making such high concentrations suspect.

    April 29, 2013 Response: Background samples were collected from the various locations across the Shipyard and several were taken from locations off the Shipyard. The background samples were intended to represent concentrations across the area without impacts from IRP sites, and were not intended to represent pristine background. However, due to the sampling locations for the most part being relatively far apart a spatial analysis is not entirely appropriate. Please see the Navy's response to MEDEP Comment No. 12 on the additional information on the outlier evaluation. In addition, the February 2013 proposed revisions to Maine RAGs shows

that the concentrations of PAHs in BGS-14 are not dissimilar to the rural developed area background numbers.

Additional Comment from January 8, 2014: Comment No. 1. As we have discussed previously, the high PAH detection limits make comparison to site PAH concentrations difficult at best. Although resampling background soil locations for PAHs would be the most accurate way to address this problem another, though less preferable, approach is to compare the Navy's proposed values to MEDEP's PAH background evaluation. Based on this evaluation, the values being proposed are reasonable for an area with a long history of industrial and other activity.

February 10, 2014 Response: Comment noted. No change is required based on this comment.

2,3,7,8 TCDD TEQ CALCULATION

Parameter	TEF	UCL (ng/kg)	UCL * TEF (ng/kg)	UPL (ng/kg)	UPL * TEF (ng/kg)	UTL (ng/kg)	UTL * TEF (ng/kg)
1,2,3,4,6,7,8,9-OCDD	0.0003	1560	0.468	3230	0.969	5090	1.527
1,2,3,4,6,7,8,9-OCDF	0.0003	68	0.0204	85.4	0.02562	129	0.0387
1,2,3,4,6,7,8-HPCDD	0.01	251	2.51	191	1.91	455	4.55
1,2,3,4,6,7,8-HPCDF	0.01	28.1	0.281	51.7	0.517	78.3	0.783
1,2,3,4,7,8,9-HPCDF	1	1	1	2.2	2.2	4.4	4.4
1,2,3,4,7,8-HXCDD	0.1	1.3	0.13	2.8	0.28	3.5	0.35
1,2,3,4,7,8-HXCDF	0.1	5.1	0.51	16.7	1.67	28.3	2.83
1,2,3,6,7,8-HXCDD	0.1	3.2	0.32	9.1	0.91	15	1.5
1,2,3,6,7,8-HXCDF	0.1	2.7	0.27	7.9	0.79	12.4	1.24
1,2,3,7,8,9-HXCDD	0.1	4.3	0.43	12.3	1.23	20.4	2.04
1,2,3,7,8,9-HXCDF	0.1	0.76	0.076	2.3	0.23	3.8	0.38
1,2,3,7,8-PECDD	1	1.3	1.3	3.7	3.7	4.1	4.1
1,2,3,7,8-PECDF	0.03	1.2	0.036	2.8	0.084	3.8	0.114
2,3,4,6,7,8-HXCDF	0.1	4	0.4	7.5	0.75	9.2	0.92
2,3,4,7,8-PECDF	0.3	3.3	0.99	6.7	2.01	8.3	2.49
2,3,7,8-TCDD	1	not detected	---	not detected	---	not detected	---
2,3,7,8-TCDF	0.1	4.6	0.46	7.6	0.76	11.5	1.15

Calculated 2,3,7,8-TCDD TEQs	UCL	UPL	UTL
2,3,7,8 – TCDD TEQ based on summation of ProUCL output for each individual parameter	9.2 ng/kg	18 ng/kg	28.4 ng/kg
2,3,7,8 - TCDD TEQ ProUCL output for positive detections only	7.0 ng/kg	14.8 ng/kg	22.7 ng/kg
2,3,7,8 - TCDD TEQ ProUCL output for half the detection limit for non-detects	6.7 ng/kg	13.5 ng/kg	19.5 ng/kg

ProUCL output for UCL, UTL, and UPL provided on Table 2 of the Technical Memorandum.

CALCULATION OF BENZO(A)PYRENE EQUIVALENT VALUES

Parameter	PNS Facility Background Using ProUCL Output for BAP Equivalents [1]		TEF	PNS Facility Background Using ProUCL Outputs for Individual PAHs [2]						Maine Background Using Remedial Action Guideline UPL for Individual PAHs (February 2013 proposed numbers) [3]					
	Positive Detections Only (95% UCL)	Half Detection Limits for ND (95% UCL)		95% UCL (µg/kg)	UCL X TEF (µg/kg)	95% UPL (µg/kg)	UPL X TEF (µg/kg)	95% UTL (µg/kg)	UTL X TEF (µg/kg)	Rural Developed Area (µg/kg)	UPL X TEF (µg/kg)	Urban Developed Area (µg/kg)	UPL X TEF (µg/kg)	Urban Fill (µg/kg)	UPL X TEF (µg/kg)
Benzo(a)anthracene	NA	NA	0.1	174	17.4	307	30.7	449	44.9	860	86	1600	160	27000	2700
Benzo(a)pyrene	NA	NA	1	222	222	523	523	642	642	1500	1500	1700	1700	5200	5200
Benzo(b)fluoranthene	NA	NA	0.1	496	49.6	632	63.2	888	88.8	1300	130	2000	200	6800	680
Benzo(k)fluoranthene	NA	NA	0.01	224	2.24	402	4.02	632	6.32	690	6.9	760	7.6	12000	120
Chrysene	NA	NA	0.001	310	0.31	733	0.733	906	0.906	1000	1	2300	2.3	6400	6.4
Dibenzo(a,h)anthracene	NA	NA	1	Non-detect	--	Non-detect	--	Non-detect	--	320	320	230	230	4500	4500
Indeno(1,2,3-cd)pyrene	NA	NA	0.1	132	13.2	239	23.9	341	34.1	400	40	740	74	3300	330
Calculated Benzo(a)pyrene Equivalent on sample by sample basis	714 µg/kg	514 µg/kg	**Calculated Benzo(a)pyrene Equivalent (summation of individual PAHs*TEF)**	NA	304 µg/kg	NA	646 µg/kg	NA	817 µg/kg	Rural	2,084 µg/kg	Urban	2,374 µg/kg	Urban Fill	13,536 µg/kg

1. Benzo(a)pyrene Equivalents and Benzo(a)pyrene Equivalents - HalfND as provided on Table 2
2. Parameter ProUCL outputs provided on Table 2
3. Maine RAGs proposed revisions from February 2013 are provided for comparison.