N00102.AR.001317
NSY PORTSMOUTH
5090.3a

200 01

DEPARTMENT OF THE NAVY
PORTSMOUTH NAVAL SHIPYARD
PORTSMOUTH, N. H. 03804-5000

IN REPLY REFER TO:

December 4, 2003

MEMORANDUM

FOR THE MEMBERS OF THE RESTORATION ADVISORY BOARD (RAB), INSTALLATION RESTORATION PROGRAM, PORTSMOUTH NAVAL SHIPYARD, KITTERY, MAINE

RAB members are invited to participate in a technical meeting on Thursday, January 8, 2004. The purpose of the technical meeting is to resolve comments on the Navy's recommendations for resolution of selected items prior to the Rounds 1 through 7 Report for the Interim Offshore Monitoring Program. The meeting will begin at 9 a.m.

If you plan to attend this meeting, please contact Ms. Marty Raymond no later than December 30, 2003 to make arrangements to attend. She may be reached by email at Raymondm@mail.ports.navy.mil or by phone at 207-438-2536.

Sincerely,

Ken

Ken Plaisted
Navy Co-Chairman
Restoration Advisory Board

Distribution:

Doug Bogen
Michele Dionne
Phil McCarthy
Alan Davis
Carolyn Lepage
Jeff Clifford
Mary Marshall
Jack McKenna
James Horrigan
Peter Britz
Diana McNabb
Onil Roy
Roger Wells

EPA Region I (M. Audet)
MEDEP (I. McLeod)
NOAA (K. Finkelstein)
MEDMR (D. Card)
NHFG (C. McBane)
USFWS (K. Munney)
EFANE (F. Evans)
COMSUBGRU TWO (A. Stackpole)

DEPARTMENT OF THE NAVY
ENGINEERING FIELD ACTIVITY, NORTHEAST
NAVAL FACILITIES ENGINEERING COMMAND
10 INDUSTRIAL HIGHWAY
MAIL STOP, #82
LESTER, PA 19113-2090

IN REPLY REFER TO
5090
Code EV23/FE
February 23, 2004

Mr. Matthew Audet
U.S. Environmental Protection Agency, Region I
1 Congress Street
Suite 1100
Mail Code HBT
Boston, MA 02114-2023

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

Dear Mr. Audet/Mr. McLeod:

SUBJECT: OPERABLE UNIT 4 ROUND 1-7 TRENDING REPORT; INSTALLATION RESTORATION PROGRAM FOR PORTSMOUTH NAVAL SHIPYARD, KITTERY, ME

Enclosed are the Action Items from the January 8, 2004 Technical Meeting on the Operable Unit 4 Round 1-7 Trending Report. The minutes will be received by USEPA and MEDEP via overnight mail on March 1, 2004. In order to keep with our current schedule, the Navy requests comments on the following items, which will be submitted to USEPA and MEDEP on March 12, 2004, on or before April 2, 2004:

- Prepare a brief approach to describe what may initiate the need for conducting a power analysis as part of comparison of MS data to RS data
- Provide a memorandum describing how the TEQs will be calculated for the dioxin risk evaluation (i.e., with consideration of dioxin-like PCBs).
- Identify the data sets to represent the exposure units for the evaluation of potential human health and ecological risks.

Also, we are assuming a 30 day time period for other items in the enclosed Action Item list.

5090
Code EV23/FE
February 23, 2004

If additional information is required please contact Mr. Fred Evans at (610) 595-0567 extension 159.

Sincerely,

FREDERICK J. EVANS
Remedial Project Manager
By direction of the
Commanding Officer

Enclosure: 1. OU4 Round 1-7 Report Action Item List

Copy to:
Mr. Jeff Clifford
Mr. Jack McKenna
Ms. Carolyn Lepage
PNS (Code 106.3R)
TtNUS (D. Cohen)
NOAA (K. Finkelstein)
USFWS (K. Munney)
Mr. Onil Roy
Dr. Roger Wells
PNS Code 100PAO
COMSUBGRU TWO (A. Stackpole)
MEDMR (D. Card) (via email)
NHFG (Dr. C. McBane) (via email)
Mr. Doug Bogen(via email)
Mr. Peter Britz(via email)
Mr. Alan Davis(via email)
Ms. Michele Dionne(via email)
Mr. James Horrigan(via email)
Ms. Mary Marshall(via email)
Ms. Diana McNabb(via email)

OU4 ROUND 1-7 REPORT ACTION ITEM LIST

Discussion Item	Action Item	Agency Responsible	Due Date
Weighting of Monitoring Station Data	Determine if trending plots for MS should be distinguish between sampling locations within the MS (i.e., using color coding or different symbols).	Navy	Include as part of the draft report
Statistical Issues	Prepare a brief approach to describe what may initiate the need for conducting a power analysis as part of comparison of MS data to RS data	Navy	March 12, 2004
	Evaluate other methods to present data for understanding results (graphical presentation of distributions, etc.)	Navy	Include as part of the draft report
Dioxin Analysis Evaluation	Provide a memorandum describing how the TEQs will be calculated for the dioxin risk evaluation (i.e., with consideration of dioxin-like PCBs).	Navy	March 12, 2004
	Identify the data sets to represent the exposure units for the evaluation of potential human health and ecological risks.	Navy	March 12, 2004
Other Issues	Conduct metals analysis on a subset of the sediment samples using standard EPA methods – After reviewing the data, prepare memorandum evaluating the data and provide a recommendation for additional sample analysis/data evaluation if warranted.	Navy	April 30, 2004

TETRA TECH NUS, INC.
661 Andersen Drive • Pittsburgh, PA 15220
Tel 412.921.7090 • Fax 412.921.4040 • www.tetratech.com

PITT-02-4-045

February 27, 2004

Project Number 2833

Mr. Matthew Audet
Environmental Protection Agency
Region I (Mail Code: HBT)
1 Congress Street, Suite 1100
Boston, Massachusetts 02114-2023

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

Reference: Contract No. N62467-94-D-0888 (CLEAN)
Contract Task Order No. 815

Subject: Meeting Minutes from the January 8, 2004 OU4 Interim Offshore Monitoring Program Technical Meeting for the Rounds 1 - 7 Report
Portsmouth Naval Shipyard (PNS), Kittery, Maine

Dear Mr. Audet/Mr. McLeod:

On behalf of the U.S. Navy, Tetra Tech NUS, Inc. is pleased to provide to the U.S. Environmental Protection Agency Region I (USEPA) and to the Maine Department of Environmental Protection (MEDEP) 4 and 3 copies, respectively, of the meeting minutes from the January 8, 2004 OU4 Interim Offshore Monitoring Program technical meeting for the Rounds 1 - 7 Report.

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
Code 106.3R Bldg. 44
Attn: Marty Raymond
Portsmouth, NH 03804-5000

If you have any comments or questions, or if additional information is required, please contact Mr. Fred Evans at 610-595-0567 x 159.

Sincerely,

Aaron Bernhardt
Project Manager

AMB/kf
Enclosure

TETRA TECH NUS, INC.

Electronic Copy via E-mail
ME Dept. of Marine Resources (D. Card)
Mr. Doug Bogen
Ms. Michele Dionne
Ms. Mary Marshall
Mr. Peter Britz
Ms. Diane McNabb
Mr. Alan Davis
NH Fish & Game (C. McBane)
Mr. James Horrigan (SAPL)

Hard Copy
EFANE, (Code 1823/FE, F. Evans) (4 copies)
PNS (Code 106.3R, M. Raymond) (4 copies)
NOAA (K. Finkelstein)
US Fish & Wildlife Service (K. Munney)
Mr. Jack McKenna
Mr. Jeff Clifford
Ms. Carolyn Lepage
D. Cohen, TtNUS, Pittsburgh

Without Enclosure
Dr. Roger Wells
Mr. Onil Roy
PNS Code 100PAO
A. Lunsford, NEHC
COMSUBGRU TWO (A. Stackpole)

OU4 Interim Offshore Monitoring Program Technical Meeting
Recommendations for Resolution of Selected Items Prior to Rounds 1 - 7 Report
Comfort Inn, Portsmouth, NH
January 8, 2004

Meeting Participants:

Fred Evans (EFANE)
Amanda Kittelson (EFANE)
Jason Speicher (EFANE, via speaker phone)
Ken Plaisted (PNS Environmental)
Marty Raymond (PNS Environmental)
Matt Audet (USEPA)
Rick Sugatt (USEPA)
Ken Finkelstein (NOAA)
Ken Munney (USFWS)
Iver McLeod (MEDEP)
Deborah Rice (MEDEP)
Jim Horrigan (RAB member, SAPL)
Carolyn Lepage (TAG)
Dave Brown (consultant to Carolyn Lepage, via speaker phone)
Aaron Bernhardt (TtNUS)
Tom Johnston (TtNUS)
Debbie Cohen (TtNUS)
Todd Kushner (TtNUS, via speaker phone)

Handouts:

- Meeting Agenda
- Update on Interim Offshore Monitoring Program for Operable Unit 4 at Portsmouth Naval Shipyard
- Statistical Methodology for Interim Offshore Monitoring Program for Operable Unit 4 at Portsmouth Naval Shipyard
- SAPL Input for January 8, 2004 OU4 Technical Meeting

Introduction

Meeting participant introductions were conducted. Tom Johnston indicated he was participating as a facilitator for the meeting. Jason Speicher and Todd Kushner were participating via telephone. Dave Brown participated via telephone for a portion of the meeting (see minutes for when he was participating in the discussion)

Presentation and Questions on the OU4 Interim Offshore Monitoring Program

There are several regulatory personnel involved in Portsmouth projects who were not involved in the project when the OU4 Interim Offshore Monitoring Plan (TtNUS, October 1999) was developed or when the first four rounds of offshore data (Baseline Rounds) were evaluated and the results provided in the OU4 Interim Offshore Monitoring Baseline Report (TtNUS, July 2002). Therefore, the meeting began with a presentation providing a summary of the interim offshore monitoring program, the results provided in the Baseline Report, and the evaluation and decision-making activities that will be conducted as part of the upcoming Rounds 1 through 7 Report. Aaron Bernhardt provided the presentation, which is attached to these minutes. Questions raised during the presentation and the Navy's answers are summarized below:

- *Were biota samples only collected in the first two rounds?* No, mussel samples are collected every round from each sediment location that had mussels nearby. Juvenile lobster samples were collected in Rounds 1 through 5 but then lobster sampling was

discontinued after Round 5 based on the recommendations provided in the Baseline Report.

- *Did the Navy assume for the calculation of number of samples as part of the monitoring program development that the CV would be 0.5?* Yes. As part of the Baseline data evaluation, average CVs for the primary chemicals of concern (PAHs and metals) were generally 0.8 or lower.
- *Did the TEQ calculation in the Baseline Report included dioxins/furans and dioxin-like PCBs?* No, only the dioxin/furan congeners were included in the TEQ calculation. [See Decision Process for Determining Whether to Discontinue Dioxin Analysis for discussion related to TEQ calculation for the Rounds 1 through 7 data evaluation.]
- *What are the factors considered in the development of the confidence bands around the trend lines?* The variability of the data and the sample size go into determination of these bands.
- *Why is the 50 percent confidence band considered in one case and the 90 percent confidence band considered in the other case?* This is a more conservative approach. The trend lines predict concentration trends over the 5 years following the Round 7 sampling. Using the 90 percent confidence band for increasing trends results in a predicted exceedance of the OU4 Interim Remediation Goal (IRG) sooner than using the 50 percent confidence band. For decreasing trends, using the 50 percent confidence band results in a predicted decrease to less than IRG later than if a 90 percent confidence band is used.
- *What is the basis for looking at the next sampling round after 5 years?* Under CERCLA, the Navy is required to evaluate the IRP sites at least every 5 years. The technical basis for the 5-year interval in predicting when the trend line will cross the IRG is related to how rapidly the trend line is changing. Specifically, the evaluation of the trend for the offshore sediment concentrations predicts whether a significant change in concentrations will occur over the next 5 years (i.e., a decrease in concentration from above to below the IRG or increase in concentration from below to above the IRG). If the trend line is changing such that a cross-over of the IRG is predicted to occur in the next 5 years, then more sampling before the next 5 year sampling is necessary to better estimate when that change will occur. Therefore, this decision regarding the trend line is related to the need for additional sampling to understand the trend. A separate decision is made to determine the need for additional scrutiny. This decision, which also is made as part of the Rounds 1 through 7 data evaluation, is not based on the decision for more frequent sampling. [For example, the nickel concentration trend for MS04 is not changing rapidly and therefore additional sampling is not necessary to understand the trend. However, because of the elevated concentrations noted, the extent of nickel (and copper) concentrations in the sediment in the vicinity of MS04 and MS03 were investigated as part of the Site 32 RI investigation.]
- *Do the example trends provided at the end of the presentation include Round 7 data?* No, Round 7 data was not available at the time of the meeting. The laboratory had difficulty with analytical equipment so the Navy received the data late.
- *On the trend line, what is the correlation provided and how is the correlation being used for the trend line?* A correlation is not being used for the trend line. The plots show concentrations versus time. The trend line is a best estimate of where the mean concentration lies based on the data. The confidence band around the trend line takes into account the sample size and variability of the data around the trend line to estimate with a set confidence, a range that the actual mean would be expected to fall within.
- *Why not use random movement around the mean to evaluate the data?* The intent of the interim offshore monitoring program is to understand changes in concentrations over

time; therefore, we need to look at the trends and projected trends. Random distribution around a mean does not take into account the changes in concentrations over time. Quality control charts (used for some bioassays) are used when you expect to get the same number or results each time; therefore, these also are not applicable to the interim offshore monitoring program.

- *With all the data available for the offshore area, why can't the Navy make a decision now on what needs to be done at each monitoring station?* The monitoring program was designed to set out when decisions should be made regarding additional sampling or scrutiny at the monitoring stations to ensure that the Navy has sufficient data for understanding trends and nature and extent before beginning the OU4 Feasibility Study (FS). Although the Navy can make decisions earlier than the times set out in the interim offshore monitoring program (for example Site 32 sediment sampling conducted in 2003), the program was set up so that we have sufficient data to make decisions. Now that the Round 7 data has been collected, there adequate data to make decisions which will be included in the draft Rounds 1 through 7 Interim Offshore Monitoring Program report.

Based on the various questions and comments that were raised during the presentation discussion, Aaron Bernhardt provided a clarification on the purpose of the meeting. He indicated that one purpose of the meeting was to provide the newer technical people from USEPA and MEDEP on the Portsmouth project with a brief history regarding the development and activities to date as part of the interim offshore monitoring program. In addition, the Baseline Report listed a few issues that the Navy wanted to resolve before evaluating the data for the Rounds 1 through 7 Interim Offshore Monitoring Program report. The Navy provided its proposed resolution of the issues/methods to evaluate the data in a memorandum (dated September 23, 2003). The regulators provided comments and the Navy goal was to try to resolve those comments in the meeting or determine the necessary action items to resolve the comments. Fred Evans indicated that a separate technical meeting could be planned to go over the Navy's recommendations for each monitoring station once the draft Rounds 1 through 7 report is submitted but before comments are issued. The goal of that meeting would be to try to resolve/clarify some of the issues ahead of time and therefore receive less comments on the draft report. The Navy is not prepared at today's meeting to discuss the results of evaluation of the monitoring station data.

Iver McLeod indicated that MEDEP has questions on some of the evaluation methods planned, particularly related to the coefficient of variance (CV) and weighting of the data. Fred Evans indicated that these items are on the agenda for this meeting. The presentation provided some of the basis to support the discussions later on in the meeting related to the MEDEP's issues.

During a break in the meeting, Carolyn LePage and Fred Evans discussed the question of how OU3 would be linked with OU6 and OU4 (which was part of the discussion at the January 7, 2004 technical meeting). As this time the Navy would like to figure out how to incorporate the operable units offshore of OU3 (OU6 and portion of OU4) into the OU3 remedial monitoring program; however, discussion among the regulators and Navy will be necessary to determine how this should be done.

After the presentation and questions and answers, the meeting focused on the issues related to the evaluation of the Rounds 1 through 7 data.

Weighting of Monitoring Station Data

Issue: The issue raised is whether data at the three locations at a particular station should be weighted prior to averaging the sample results. If so, how should the weighting be conducted? Does the size of the habitat that each sample represents need to be considered in determining the average station concentration? Does the data need to be reviewed to determine whether the samples representing the different habitats may also indicate different sample populations based on different concentration patterns?

Discussion: Consideration of whether treating the habitats differently will affect the decision making for the interim offshore monitoring program. While understanding of the concentration patterns in the various habitats may affect the understanding of the nature and extent (which is important for the FS), for the interim offshore monitoring decisions, weighting of the data would not significantly change the conclusions for the monitoring program. TtNUS reviewed the data from the first four rounds of sampling and found that the samples within the station that generally had higher concentrations represented a smaller area of the station than the other samples and thus weighting based on size would tend to reduce the average concentrations and not significantly change the conclusions.

In the trending, the variability in the data is considered in developing the confidence bands. In addition, the proposed method for plotting the data shows the separate sample concentrations that contribute to the average concentration. Therefore, patterns of the data (i.e., if only samples from one location are greater than the IRG) may suggest that additional scrutiny of a specific habitat or area needs to be evaluated. TtNUS will determine whether the samples from the sample locations can be indicated with different symbols or colors on the trending plots[1].

Resolution: It was agreed that for calculating the average concentration, weighting of the data (based on habitat coverage) is not necessary. However, it is important to look at the data to see whether there may be different populations at each station that may impact the decision making and appropriateness of averaging the data for comparison to IRGs.

Coefficient of Variation Data

Issue: The original comment from the MEDEP related to the need to calculate the CV for the reference station locations.

Discussion: Aaron Bernhardt indicated that the reference station locations were determined based on representing the target habitats and not based on an estimation of the CV for reference data. There are 16 sediment samples per monitoring round in the reference data set. Deborah Rice clarified that her concern with the CV for the reference stations related to the apparent variability in the reference station data set. She believes that because of the high variability in the concentrations in the reference station data set, there may not be sufficient statistical power to tell whether the reference data set and the monitoring station data set are statistically different. Deborah Rice said the Navy needs to do a power analysis to support the Navy's conclusions that the monitoring station concentrations are similar to the reference station concentrations. Todd Kushner pointed out that typically a power analysis is used as part of project planning activities and there may be some qualifications to using the analysis after sampling. In addition, the need to do a power analysis should be considered on a case-by-case basis. The Navy indicated that it has done power analysis before at other Navy sites.

[1] Post meeting note: Different symbols will be used for each sampling location on the plot.

Resolution: Navy will present an approach to describe what may initiate the need for conducting a power analysis as part of comparison of site data to reference data. The Navy will get back to group with proposal. Also, the Navy will consider other methods for presenting the data to understand the results as part of the Round 1 through 7 Interim Offshore Monitoring Program report. Also, the term "statistically similar" is not correct and will be replaced with "not statistically different", where appropriate.

[Dave Brown joined the meeting during the discussion of the resolution. The following provides the discussion that relates to the CV resolution as well as the next meeting topic on statistical issues.]

In answer to a question of whether the Navy was changing the method of comparing site data to background data from what was proposed in the monitoring plan, the Navy indicated that only updates to the proposed methods are currently being considered to reflect updates in Navy and USEPA guidance. This will be discussed further as part of the statistical evaluation presentation.

Dave Brown indicated that his concern is that if we did not see something statistically significant that we would drop it without having the statistical power to support that action. This concern is consistent with the previous discussion related to the CV and need for evaluating the statistical power.

A discussion ensued about the effect of multiple analytes on the outcome of a statistical test when the analyte concentrations may be correlated. Tom Johnston indicated that this can be addressed through binomial probabilities and acknowledged that the binomial probabilities account for multiple analytes but tend to conservatively overestimate data set differences when analyte concentrations are correlated. [At this time, it is not known if binomial probabilities will need to be conducted as part of the Round 1 through 7 Interim Offshore Monitoring Program report.] Fred Evans indicated that the Navy looks at the data in different ways and does not only rely on the results of statistical evaluations.

Dave Brown indicated that with complicated projects the Navy needs to make the evaluations and results more understandable for the general public. Fred Evans agreed that communication with community is important so that people can understand what the Navy is doing and why. He indicated the Navy, USEPA, MEDEP, and SAPL, as the TAG recipient, all have a responsibility to communicate with the public on PNS's IR Program and suggested the groups work together on how to communicate the evaluations and results more understandable for the general public.

Several suggestions that Dave Brown made relate to evaluating the distribution of the data, graphically presenting the data to show distribution patterns, and conducting background comparisons even if the concentrations are below risk levels (to provide an understanding of how site data compares with background conditions). Todd Kushner indicated that we look at the distribution as part of our data evaluation. Dave Brown asked whether the Navy will do a T-test or similar test on the data. Todd Kushner indicated that TtNUS will conduct T-tests when comparing site data to reference data.

The conclusion of the discussion was that CVs for the reference area were not the primary issue of concern. The main cause of concern was the need for the Navy to conduct a power analysis in some instances to determine if the statistical conclusions are valid. It was also discussed that while the power analysis may have some value, the project team should understand that

post-hoc or retrospective power analysis is controversial because of the ease of misinterpretating the results.

[Jim Horrigan and Ken Plaisted left the meeting at the completion of this discussion and Dave Brown was not available on phone for remainder of meeting.]

OU4 Statistical Evaluation Presentation and Statistical Issues

Much of discussion on the CV was related to the statistical evaluation, so the meeting continued with the presentation on the statistical evaluation and issues related to the statistical treatment of the data. Tom Johnston provided the presentation and Todd Kushner provided answers to the specific statistical questions that were raised during the presentation. The presentation is attached to the minutes.

It was clarified that the normalization of the metals data to aluminum is not used for risk assessment purposes, but normalized data is used to understand nature and extent of the contamination. Therefore, both non-normalized and normalized data will be provided in the Rounds 1 through 7 report. For the statistical outlier analysis, data points are flagged that need to be evaluated further to determine whether there is an assignable reason for the outlier that suggests that the data point should not be included in the data set. If an assignable reason cannot be identified, then the data point is not removed from the data set. Aaron Bernhardt provided an example of where the reference results were very high in one sampling round and after further evaluation the Navy determined that there had been a mix up at the laboratory and the sample was actually from a different location. Therefore, the results for this sample were removed from the reference station data set.

Todd Kushner indicated how duplicate pairs are handled for the identification of the minimum and maximum, frequency of detection, and averaging. The text in the report describing the manipulation of the data needs to be clear on how the duplicate samples are handled.

When doing the statistical evaluation, a 50 percent non-detection rate in a dataset is used as a guide to determine what type of statistical analysis will be conducted. Todd Kushner indicated that the guidance typically considers a 40 to 50 percent non-detection rate as a guide. Aaron Bernhardt noted that because we are using NOAA analytical methods we typically have low detection limits so there are few parameters with a lot of non-detects and the low rate of detections may not be an issue.

During the presentation, Aaron Bernhardt noted that one of the presentation slides was incorrect. The slide showing "Parametric Satterthwaite t-Test" should indicate "HOV are dissimilar." This change is handmarked on the presentation handout attached to these minutes

Distribution shapes for the two data sets (lognormal versus normally distributed) that are being compared need to be the same to use the parametric tests. If the distribution shapes are different or undefined, nonparameteric tests are used. In answer to a question of whether TtNUS tries to determine the distribution of the data if the distribution is not normal or lognormal, it was indicated that very often the environmental data sets are lognormal. Typically in the environmental field the data set distribution is considered it to be undefined if the data set distribution is not lognormal or normal. Probability plots can be used to aid to understanding of the distribution of the data. Deborah Rice indicated that tissue data are often lognormal and mainly raised the question because parametric tests are typically more powerful than non-paramatric tests and trying to find the distribution of the data may allow for use of a parametric

test. Todd said there are many distributions that can be tested for, but typical environmental investigations only test for normal or lognormal distributions. Also Todd added that the data is evaluated qualitatively and quantitatively to understand the data in addition to evaluating the statistical results.

Decision Process for Determining Whether to Discontinue Dioxin Analysis

Discussion: The Navy is proposing a decision process to use to determine whether to discontinue dioxin analysis for the selected monitoring stations. Aaron Bernhardt reviewed the flow chart for the recommendation process, which was attached to the response to comments on the draft memorandum. He noted that one difference from before is that the screening level comparison comes before the background comparison to be consistent with Navy and USEPA guidance.

The decision process shows that if there are no exceedances of the screening level, then the risks are acceptable and dioxin analysis can be discontinued. First the maximum concentrations are compared to the screening levels. If there are exceedances of the screening level, the 95 percent UCL concentration is then compared to the screening level. Deborah Rice indicated that if there is a disparity between sample concentrations, then the 95 percent UCL may result in overlooking a potential hot spot or carrying through an area with low concentrations. This is particularly a concern if samples are included in the 95 percent UCL calculation that are not representative of the exposure unit (e.g., recreational exposure unit is the intertidal area only and should not include subtidal samples). The decision process will be revised to include a box before calculation of the 95 percent UCL to determine the samples within the appropriate exposure units (i.e., add "determine appropriate data set to represent exposure unit.")

Deborah Rice then brought up that the calculation of the TEQ values should include both dioxin/furans and the WHO congeners for the dioxin-like PCBs. She indicated that it was appropriate to include the dioxin-like PCBs because they have the same mechanism of aryl hydrocarbon (Ah)-receptor activation (once ingested). Rick Sugat, Ken Munney, and Deborah Rice indicated that this is the standard method for TEQ calculation that they have been involved with. Separate dioxin-like PCB TEQ value and dioxin/furan TEQ value could be calculated and then summed for a total TEQ value so that the relative contribution from the dioxin-like PCBs and dioxin/furans to the total TEQ value can be understood. It was suggested that understanding the pattern of the individual congeners as an indicator of source may be considered if the TEQ values suggest a potential concern.

The Navy will need to provide a proposal on how to calculate the TEQs. Jason Speicher noted that detection limits may present a concern for the TEQ calculation for dioxin-like PCBs because the detections limits are greater for those chemicals than they are for dioxins. The calculation of TEQ generally treats non-detections in two ways; TEQ values are calculated without considering the non-detections (i.e., using only the positive detections) or TEQ values are calculated using positive detections plus ½ the detection limit for non-detections. In this second method, detection limits could falsely elevate the TEQ value.

Other Issues

The Navy is considering changing analytical methods from the NOAA method to a USEPA method. In general, the NOAA method for metals is more difficult and therefore, fewer laboratories do the analysis. Also, the turn around time for the analysis is typically longer and

the costs are higher. Because the NOAA method includes complete digestion of the sediment, the metals concentrations for natural materials in the sediment particle (like aluminum) will be higher than if a USEPA method is used. For the more toxic metals, however, a significant difference between concentrations using the two methods is not expected. The Navy will propose a method for evaluating the differences between the two methods.

ROUND 1-7 REPORT ACTION ITEM LIST

Discussion Item	Action Item	Agency Responsible	Due Date
Weighting of Monitoring Station Data	Determine if trending plots for MS should be distinguish between sampling locations within the MS (i.e., using color coding or different symbols).	Navy	Include as part of the draft report
Statistical Issues	Prepare a brief approach to describe what may initiate the need for conducting a power analysis as part of comparison of MS data to RS data	Navy	March 12, 2004
	Evaluate other methods to present data for understanding results (graphical presentation of distributions, etc.)	Navy	Include as part of the draft report
Dioxin Analysis Evaluation	Provide a memorandum describing how the TEQs will be calculated for the dioxin risk evaluation (i.e., with consideration of dioxin-like PCBs).	Navy	March 12, 2004
	Identify the data sets to represent the exposure units for the evaluation of potential human health and ecological risks.	Navy	March 12, 2004
Other Issues	Conduct metals analysis on a subset of the sediment samples using standard EPA methods – After reviewing the data, prepare memorandum evaluating the data and provide a recommendation for additional sample analysis/data evaluation if warranted.	Navy	April 30, 2004

OU4 Interim Offshore Monitoring Program Meeting
January 8, 2004
Proposed Meeting Agenda

Location: Comfort Inn-Yokens
Portsmouth, NH

9:00 - 9:05	Introductions
9:05 - 10:00	OU4 Interim Offshore Monitoring Program Presentation
10:00 - 10:15	Questions on Monitoring Program
10:15 - 10:30	Weighting of Monitoring Station Data
10:30 - 10:45	Break
10:45 - 11:00	Weighting of Monitoring Station Data, Continued
11:00 - 11:30	Coefficient of Variance Data
11:30 - 12:00	Discontinuing Dioxin Analysis
12:00 - 12:45	Lunch
12:45 - 1:00	OU4 Statistical Evaluation Presentation
1:00 - 1:50	Statistical Issues
1:50 - 2:00	Break
2:00 - 3:00	Other Issues

Statistical Methodology for Interim Offshore Monitoring Program for Operable Unit 4 at Portsmouth Naval Shipyard

January 8, 2004

Tetra Tech NUS, Inc.

Data Preparation

- For any statistical analysis, use the average of duplicate samples.
- For any nondetect (ND), use ½ the detection limit.
- Normalize data for each sediment sample:
  - Organic Chemicals: $[\text{chemical}]_{sed}/[\text{TOC}]_{sed}$
  - Inorganic Chemicals: $[\text{chemical}]_{sed}/[\text{aluminum}]_{sed}$
- Normalize data for each mussel sample:
  - Organic Chemicals: $[\text{chemical}]_{mussel}$/% Lipids
  - Inorganic Chemicals: not normalized

Outlier Analysis

Analyze all data sets and subsets for statistical outliers by one of the following tests (outliers with an assignable cause will be omitted from the data set):

- Discordance Test
- Rosner's Test
- Walsh's Test
- Normal Probability Plots

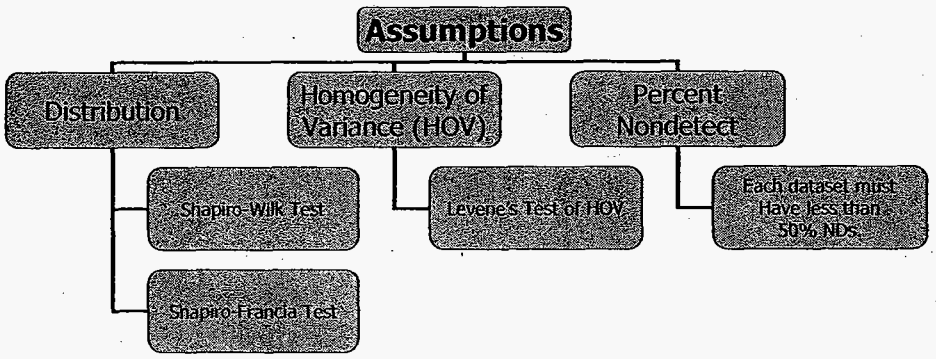
Descriptive Statistics

Compute descriptive statistics for all datasets.	
➢Frequency of Detection	➢Mean Concentration
➢Minimum Concentration	➢Maximum Concentration
➢Average of Positive Detections	➢Location of Maximum Detection
➢Range of NDs	➢Coefficient of Variation

Statistical Background Comparisons

- If all results = ND, no statistical analysis warranted
- If all reference station (RS) results = ND and monitoring stations (MS) results are detected, then [MS] concluded to exceed [RS]
- If all MS results = ND and RS results are detected, then [MS] is concluded not to exceed [RS]
- Assumptions will be tested to determine which comparison test will be used

Tests of Assumptions



Parametric ANOVA Tests

- Parametric ANOVA Student's *t*-Test:
  - Run if all tests of assumptions hold, i.e.:
    - Each dataset contains less than 50% NDs
    - Dataset distributions are equal
    - HOVs are similar
- Parametric Satterthwaite *t*-Test:
  - Run if (all conditions must hold):
    - Each dataset contains less than 50% NDs
    - Dataset distributions are equal
    - HOVs are similar

Nonparametric ANOVA Tests

- Wilcoxon Rank-Sum Test (WRS):
  - Run if:
    - Each dataset contains less than 50% NDs
    - Distributions are different
- Quantile Test:
  - Run in conjunction with WRS Test to ensure upper tail of MS data ≤RS data.
- Test of Proportions:
  - One of the datasets contains more than 50% NDs

Update on Interim Offshore Monitoring Program for Operable Unit 4 at Portsmouth Naval Shipyard

January 8, 2004

Tetra Tech NUS, Inc.

Objectives of Presentation

- Provide a general overview of the OU4 interim offshore monitoring program to date to orient meeting participants with the program.
- Present methodology for evaluating data in the Round 1-7 Report to assist in resolving meeting discussion items.

Monitoring Program

- Interim Offshore Monitoring Plan for OU4 (TtNUS, October 1999) provides the program components and decision process

- The program was developed, in accordance with the Interim Record of Decision (ROD) for OU4 to:
  - Provide monitoring before completion of the offshore Feasibility Study.
  - Provide the basis for any monitoring program that may be required as part of the final action.
  - Provide data to determine whether the interim remedial action objectives for OU4 are being met
  - Develop preliminary remediation goals (PRGs) for OU4

Baseline rounds, semi-annual sampling

- Round 1 (Year 0) – September 1999
- Round 2 (Year 0) – May 2000 (data for PRGs collected)
- Round 3 (Year 1) – August 2000
- Round 4 (Year 1) – May 2001

Subsequent rounds until final remedy

- Round 5 (Year 2) – August 2001
- Round 6 (Year 3) – August 2002
- Round 7 (Year 4) – August 2003
- Round 8 (Year 6) – Scheduled for August 2005
- Round 9 (Year 8) – Scheduled for August 2007
- Round 10 (Year 9) – Schedule for August 2008

Monitoring Program (cont.)

- Decisions to make
  - All decisions based on evaluations of representative chemicals of concern (COCs)
  - Baseline sampling
    - Most appropriate season for subsequent sampling rounds
      - Evaluated first four rounds of data
  - Every 5 Years
    - Need for increased sampling between 5 year evaluations
    - Need for additional scrutiny at particular stations
      - Additional evaluation of monitoring station data or collection of additional data

Sampling Program

- Sediment sampling primary measure
  - Identified in EERA as the primary reservoir for PNS-related chemicals
  - Risk in EERA was low to intermediate
- Mussel/lobster sampling secondary measures
  - Mussel data used to calculate BAFs/BSAFs

5

Sediment Samples (0 – 10 cm except as noted below)

- Metals
- PAHs (including alkylated PAHS)
- Pesticides
- PCBs (NOAA and WHO congeners)
- Dioxins (at MS 7-12 and all 4 reference stations)
- AVS/SEM
  - Acid volatile sulfides (0-2 cm and 0-10 cm)
  - Simultaneously extracted metals (0-10 cm)
  - Discontinued after 5 rounds
- Grain size
- Total organic carbon

Mussel/Juvenile Lobster Samples

- Metals
- PAHs (including alkylated PAHS)
- Pesticides
- PCBs (NOAA and WHO congeners)
- Dioxins (at MS 7-12 and all 4 reference stations)
- Percent lipids

NOAA analytical methods for consistency

Sampling Program (cont.)

- 14 Monitoring Stations around PNS (see Attachment)
  - ~100 ft. wide by ~300 ft. long (except MS11)
    - MS11 is larger to encompass Sites 6 and 29
  - 3 sediment sampling locations per station based on
    - USEPA DEFT software and variability of data
    - Providing adequate spatial coverage for habitats
      - Mussel beds
      - Eelgrass beds
      - Seep location
      - Salt marsh
      - Subtidal habitat
  - 1 to 3 mussel samples per station
  - 1 juvenile lobster sample per station
  - Targeted areas of contamination (i.e., near seeps)

Sampling Program (cont.)

- 4 Reference Stations in Piscataqua River (see Attachment)
  - Selected to provide data on regional chemical conc.
    - Away from immediate influence of PNS
    - Depositional areas with similar geotechemical characteristics to sediment near PNS
    - Represent contaminant exposure that estuarine biota would experience without PNS as a source and not near point source
    - Similar habitats to PNS
  - 4 sediment sample locations per station; based on habitats
    - Intertidal sediment by mussel bed and in salt marsh
    - Subtidal sediment, including within an eelgrass bed if present
  - 2 mussel samples (1 intertidal/1 subtidal)
  - 1 juvenile lobster sample

Preliminary Remediation Goals

- Sediment-based PRGs developed for list of limiting COCs (chemicals driving risks):
  - As ecological risk-based numbers
  - For use as Interim Remediation Goals (IRGs) for interim offshore monitoring program
  - For use across the Shipyard offshore areas
  - Using Round 2 data (late spring) when metals are more bioavailable
    - Chemical data for sediment and sediment porewater (metals)
    - Whole sediment and porewater toxicity tests

PRGs used as IRGs are as follows:

- Copper = 486 mg/kg
- Nickel = 124 mg/kg
- Acenaphthylene = 210 µg/kg
- Anthracene = 1,236 µg/kg
- Fluorene = 500 µg/kg
- HMW PAHs = 13,057 µg/kg

Baseline Report

- Primary objective:
  - Recommend appropriate season for future monitoring events
- Other evaluations conducted
  - Presented general observations regarding data
  - Calculated BAFs/BSAFs
  - Calculated coefficients of variation
  - Compared sediment results to IRGs
  - Recommended discontinuation of lobster sampling
  - Conducted a risk evaluation for dioxins

9

Comparison to IRGs

- Six monitoring stations had average sediment concentrations that exceed an IRG in at least one round:
  - M01 (offshore Site 34)
    - PAHs
  - M03 and M04 (offshore Site 32)
    - Copper at M03
    - Copper, nickel, and anthracene at M04
  - M11 (offshore Sites 6 and 29)
    - Copper and nickel
  - M12 and M13 (offshore the dry-dock area)
    - PAHs

Baseline Report (cont.)

- Three changes made to the program based on Baseline Report recommendations
  - Late summer season (August) selected as preferred sample collection time
    - Few, if any seasonal differences were observed in AVS and SEM data
  - SEM and AVS analyses discontinued after Round 5
    - Data not necessary for decisions for monitoring program
  - Lobster sampling and analyses discontinued after Round 5
    - Chemical concentrations in lobster samples generally similar across stations and within reference station concentrations

Baseline Report (cont.)

- Coefficient of Variation
  - CVs were calculated for the following groups
    - Chemicals across all stations for each round (IOMP)
    - Chemicals across AOCs for all rounds (IOMP/1996-97)
      - IOMP included subtidal and intertidal locations
      - 1996-97 sampling only included intertidal locations
  - CVs greatest for all parameters in Back Channel across all rounds and pesticide/PCBs at all AOCs
  - Most of the average CVs were 0.8 or lower for PAHs and inorganics – similar to historic samples
    - PAHs and inorganics are primary COCs

Baseline Report (cont.)

- Dioxin Risk Screening Evaluation
  - Compared TEQ values to PNS-specific human health screening levels and literature ecological screening levels
    - A few mussel TEQs > RME HH screening level
    - No mussel TEQs > CTE HH screening level
    - No sediment TEQs > RME or CTE screening level
    - 2 reference station lobster TEQs > RME HH screening level; none above CTE screening level
    - 1 sediment TEQ > Eco screening level
  - Dioxin concentrations in sediment, mussel, and lobster collected from monitoring stations do not appear to be at levels that result in unacceptable risks to human or ecological receptors

Baseline Report (cont.)

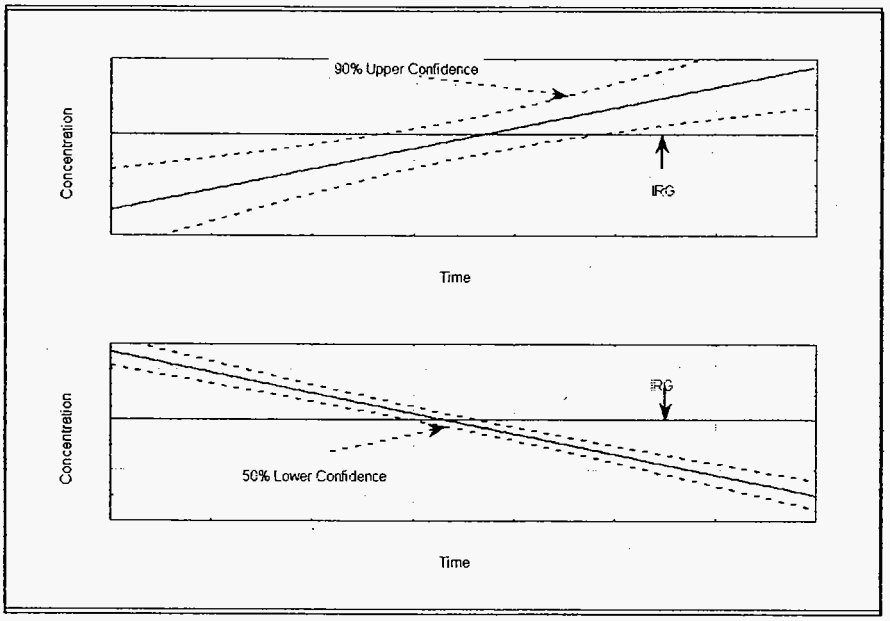
- Sediment Observations Based on Habitats
  - Intertidal area concentrations > than subtidal area
  - Non-saltmarsh intertidal area concentrations > saltmarsh concentrations
  - Non-eelgrass subtidal area concentrations similar to eelgrass concentrations
- Minor TOC and grain size differences between RS and MS
  - Slightly lower TOC in reference samples

Evaluation of Rd 1-7 Data

- Determine need to sample certain monitoring stations more frequently than every five years
- Determine the need for additional scrutiny at certain monitoring stations
- Determine if adequate numbers of sediment samples are being collected at each monitoring station
- Determine if the monitoring program should be modified for subsequent sampling rounds

Data Use/Interpretation

- Trending will be used to evaluate the data.
  - Confidence limits take into account the variability of the data and number of samples
  - 90 percent upper confidence limit on concentration trends less than IRGs
  - 50 percent lower confidence limit on concentration trends greater than IRGs.
  - Average concentration of three samples used for MS trending.
- Average concentration of all reference samples for RS trending.

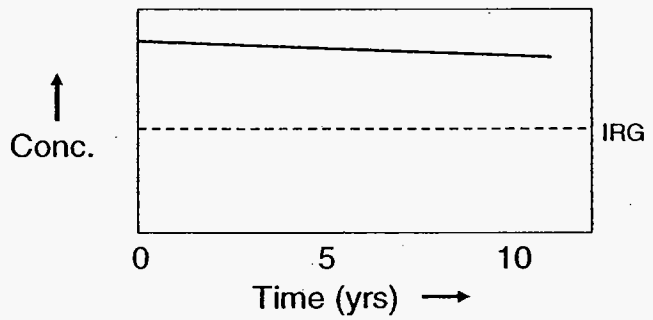


Evaluation of Sampling Frequency

- Rapidly changing trends imply greater monitoring frequency to detect changes
- No additional monitoring between 5-year sampling event when trend line
  - Starts above and remains above IRG
  - Starts below and remains below IRG
- Additional monitoring between 5-year sampling event (years 6 and 8) when trend line
  - Starts above and crosses below IRG within 5 years
  - Starts below and crosses above IRG within 5 years
- Examples of different possibilities on next four slides

Evaluation of Sampling Frequency (cont.)

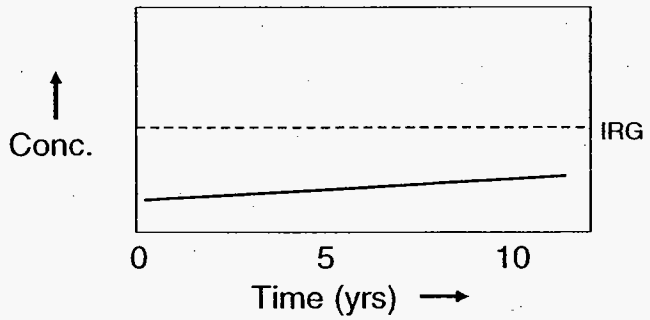
Trend line (50% LCL) starts at [COC] > IRG and stays "high"



No additional monitoring between 5-year sampling events

Evaluation of Sampling Frequency (cont.)

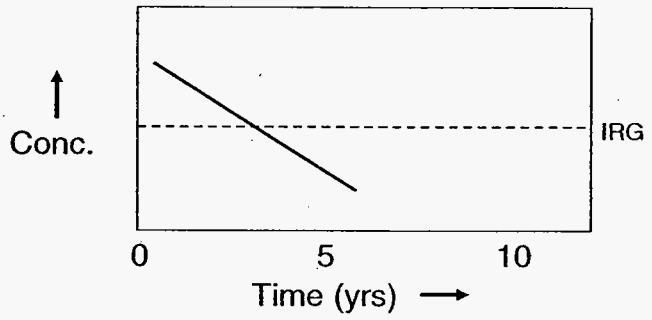
Trend line (90% UCL) starts at [COC] < IRG and stays "low"



No additional monitoring between 5-year sampling events

Evaluation of Sampling Frequency (cont.)

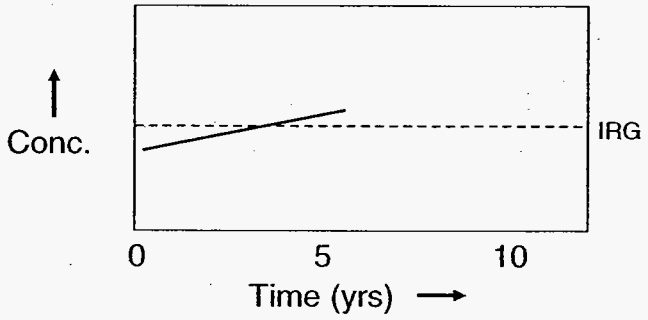
Trend line (50% LCL) starts at [COC] > IRG and crosses IRG within 5-yr decision window



Additional sampling in years 6 and 8

Evaluation of Sampling Frequency (cont.)

Trend line (90% UCL) starts at [COC] < IRG and crosses IRG within 5-yr decision window



Additional sampling in years 6 and 8

Need for Additional Scrutiny

- Decision for additional scrutiny will be determined by evaluating the following items:
  - MS trend line above or below the IRG
  - MS and RS trends increasing, decreasing, or staying the same
  - MS trend line slope compared to RS trend line slope
  - Professional judgment

Discussion Points

- Discontinuing dioxin analysis
- Weighting of Monitoring Station Data
- Coefficient of Variance
- Other issues

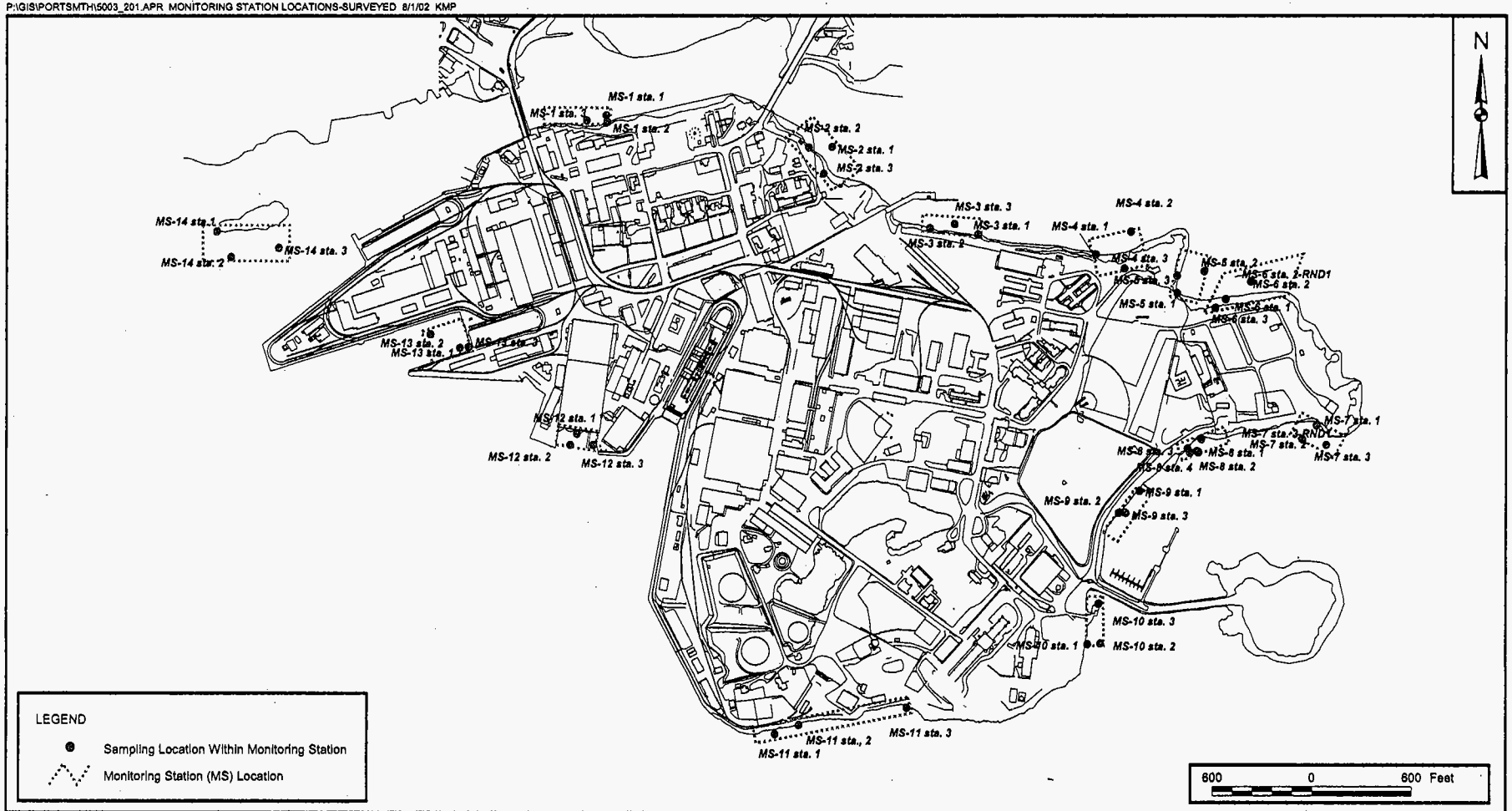
Attachments

- Figure 1 – Monitoring Station Locations
- Figure 2 – Reference Station Locations
- Acronym List
- Table of PAH and PCB congener lists
- Example Trending Plots
- References

ACRONYMS AND ABBREVIATIONS

AOC	Area of Concern
AVS	Acid Volatile Sulfide
BAF	Bioaccumulation Factor
BSAF	Biota Sediment Accumulation Factor
COC	Chemical of Concern
CTE	Central Tendency Exposure
CV	Coefficients of Variation
DEFT	Decision Error Feasibility Trials
EERA	Estuarine Ecological Risk Assessment
HH	Human Health
HMW	High Molecular Weight
IOMP	Interim Offshore Monitoring Program
IRG	Interim Remediation Goal
LCL	Lower Confidence Limit
MS	Monitoring Station
NOAA	National Oceanic and Atmospheric Administration
OU4	Operable Unit 4
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
PNS	Portsmouth Naval Shipyard
PRG	Preliminary Remediation Goal
Rd	Round
RME	Reasonable Maximum Exposure
ROD	Record of Decision
RS	Reference Station
SEM	Simultaneously Extracted Metals
SW	Surface Water
TEQ	Toxic Equivalent Concentrations
TOC	Total Organic Carbon
UCL	Upper Confidence Limit
USEPA	U.S. Environmental Protection Agency
WHO	World Health Organization

P:\GIS\PORTSMTH\5003_201.APR MONITORING STATION LOCATIONS-SURVEYED 8/1/02 KMP



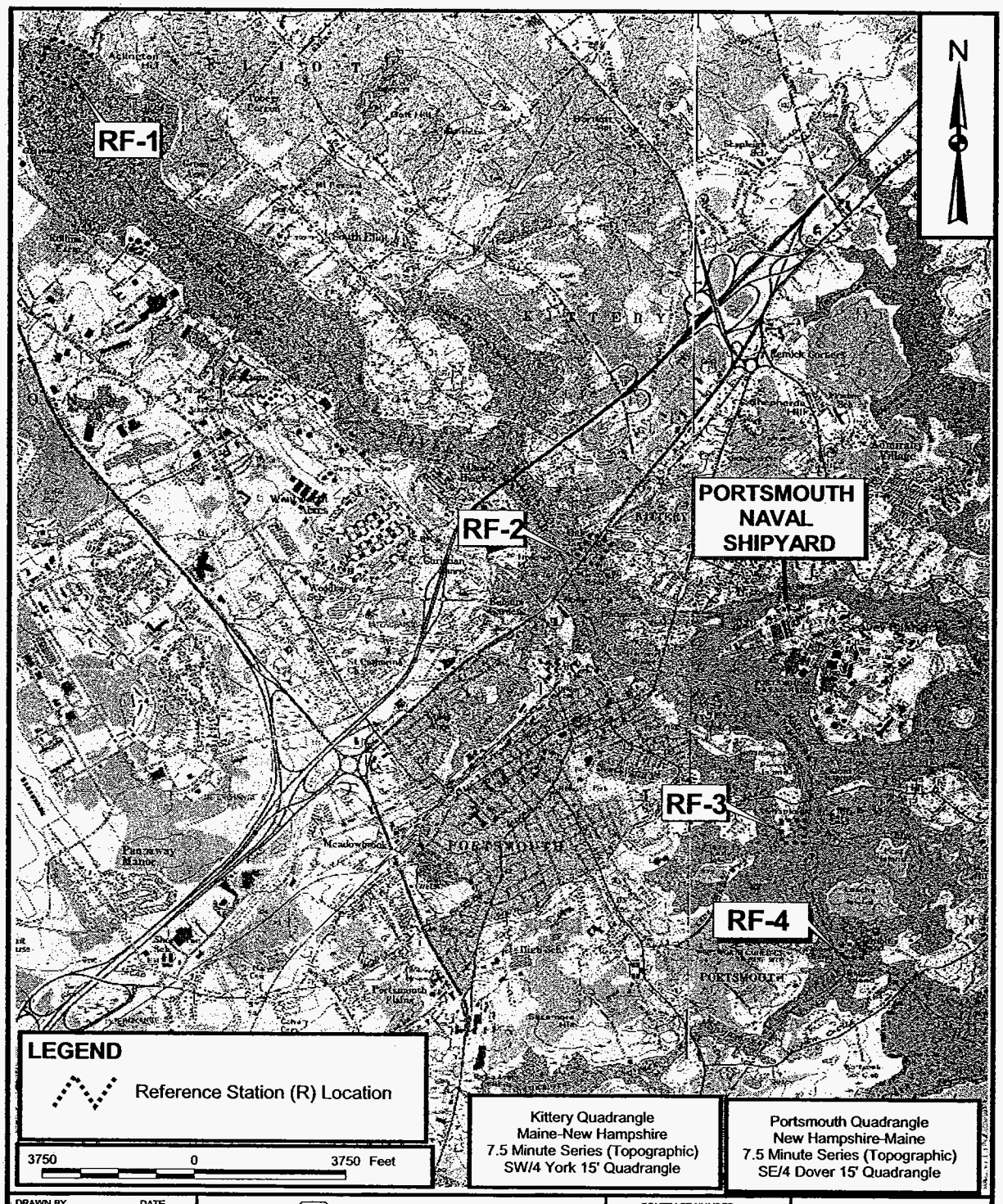
NO.	DATE	REVISIONS	BY	CHKD	APPD	REFERENCES

DRAWN BY J. LAMEY	DATE 12/15/99
CHECKED BY A. BERNHARDT	DATE 11/06/01
COST/SCHEDULE-AREA	
SCALE AS NOTED	

Tetra Tech NUS, Inc.

OVERVIEW OF MONITORING STATION LOCATIONS
INTERIM OFFSHORE MONITORING PROGRAM
PORTSMOUTH NAVAL SHIPYARD, KITTERY, MAINE

CONTRACT NUMBER 2833	OWNER NUMBER CTO 0815
APPROVED BY —	DATE —
APPROVED BY —	DATE —
DRAWING NO. FIGURE 1	REV 0



DRAWN BY	DATE	Tetra Tech NUS, Inc.	CONTRACT NUMBER	OWNER NO.
J. BELLONE	15-JULY-99		5003	0201
CHECKED BY	DATE	OVERVIEW OF REFERENCE STATION LOCATIONS INTERIM OFFSHORE MONITORING PROGRAM PORTSMOUTH NAVAL SHIPYARD, KITTERY, MAINE	APPROVED BY	DATE
D. COHEN	7/26/99		—	—
COST/SCHEDULE-AREA			APPROVED BY	DATE
			—	—
SCALE			DRAWING NO.	REV
AS NOTED			FIGURE 2	0

P:\GIS\PORTSMOUTH\5003_201.APR 1/30/02 ACS USGS MONITORING REFERENCE LOCATIONS LAYOUT

ATTACHMENT TABLE

LIST OF PAHS INCLUDED IN TOTAL PAH CALCULATIONS AND LIST OF PCB CONGENERS INCLUDED IN ANALYTICAL SUITE UPDATE ON INTERIM OFFSHIRE MONITORING PROGRAM FOR OU4 PORTSMOUTH NAVAL SHIPYARD, KITTERY, MAINE

List of PAHs included in the calculation of LMW PAHs, HMW PAHs and total PAHs	
Low Molecular Weight PAHs	**High Molecular Weight PAHs**
2-Methylnaphthalene	Benzo(a)anthracene
Acenaphthene	Benzo(a)pyrene
Acenaphthylene	Chrysene
Anthracene	Dibenzo(a,h)anthracene
Fluorene	Fluoranthene
Naphthalene	Pyrene
Phenanthrene	

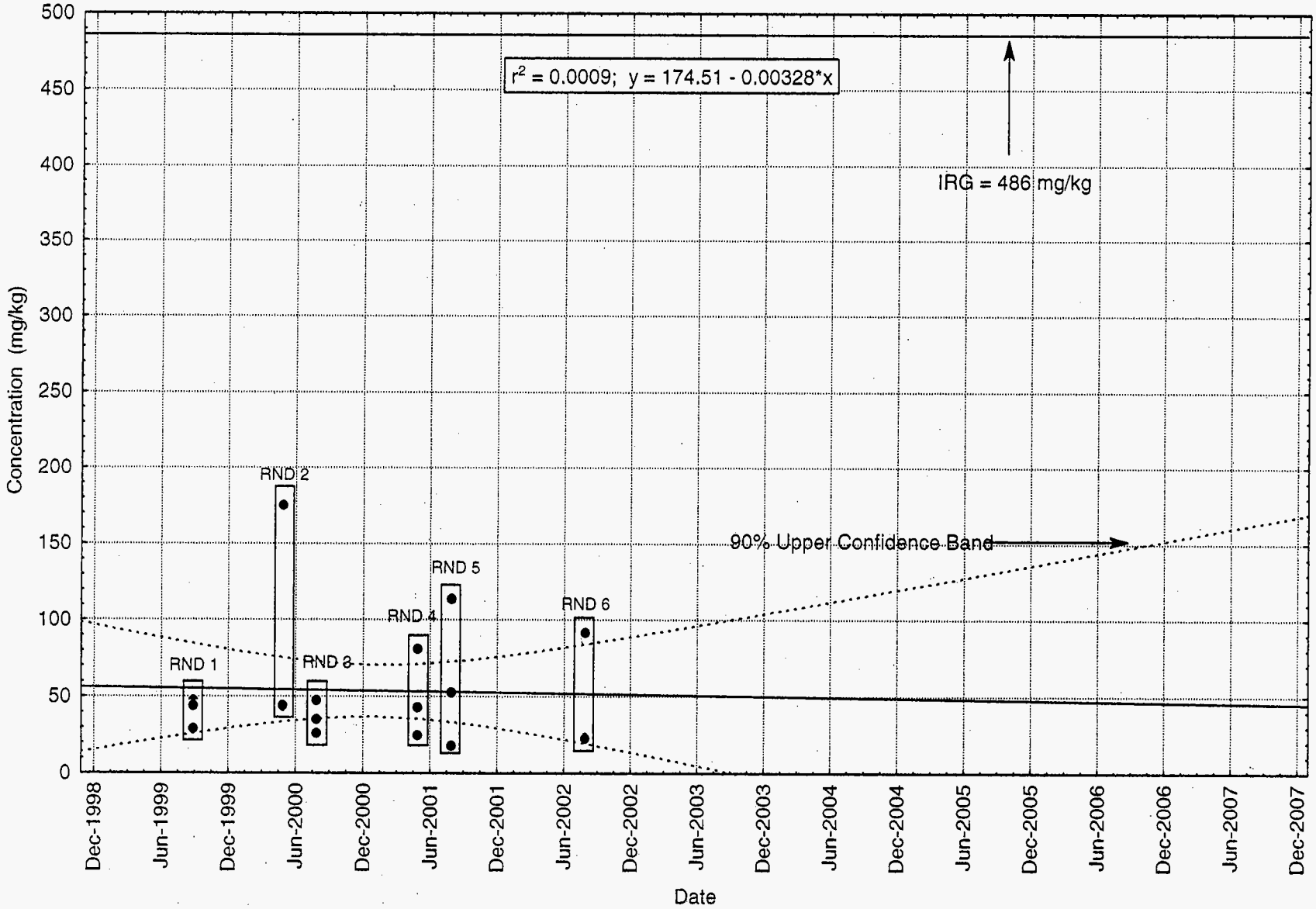
18 NOAA PCBs included in the calculation of total PCBs	
PCB-101/90	PCB-187
PCB-105	PCB-195/208
PCB-118	PCB-206
PCB-128	PCB-209
PCB-138 /160	PCB-28
PCB-153/132	PCB-44
PCB-170/190	PCB-52
PCB-18/17	PCB-66
PCB-180	PCB-8/5

10 WHO PCBs that are not included in the calculation of total PCBs	
PCB-114	PCB-169
PCB-126	PCB-189
PCB-149/123	PCB-201/157/173
PCB-156	PCB-77
PCB-167	PCB-81

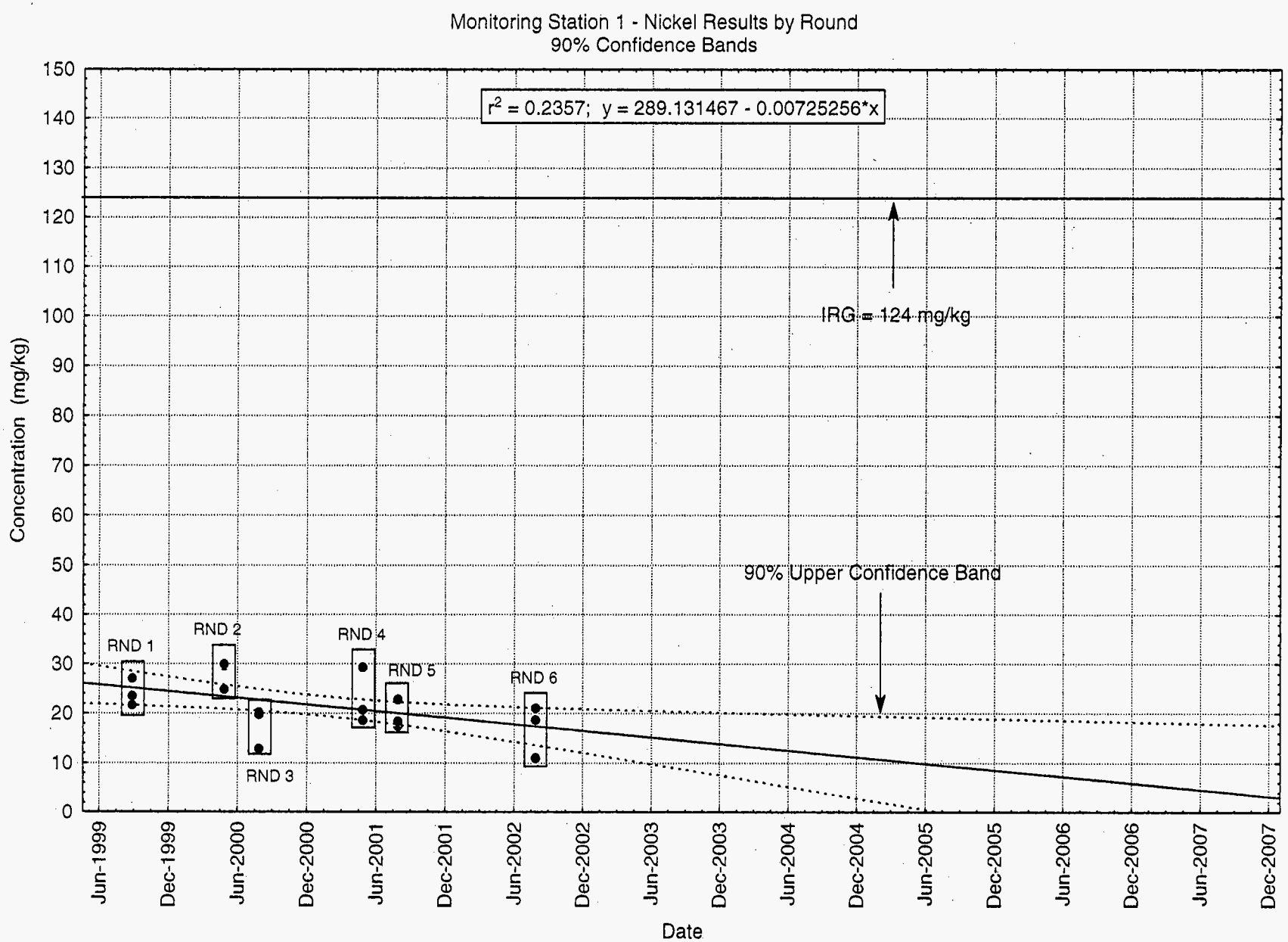
Acronyms:
HMW = High Molecular Weight
LMW = Low Molecular Weight
NOAA = National Oceanic and Atmospheric Administration
PAH = Polycyclic Aromatic Hydrocarbon
PCB = Polychlorinated Biphenyl
WHO = World Health Organization

PRELIMINARY

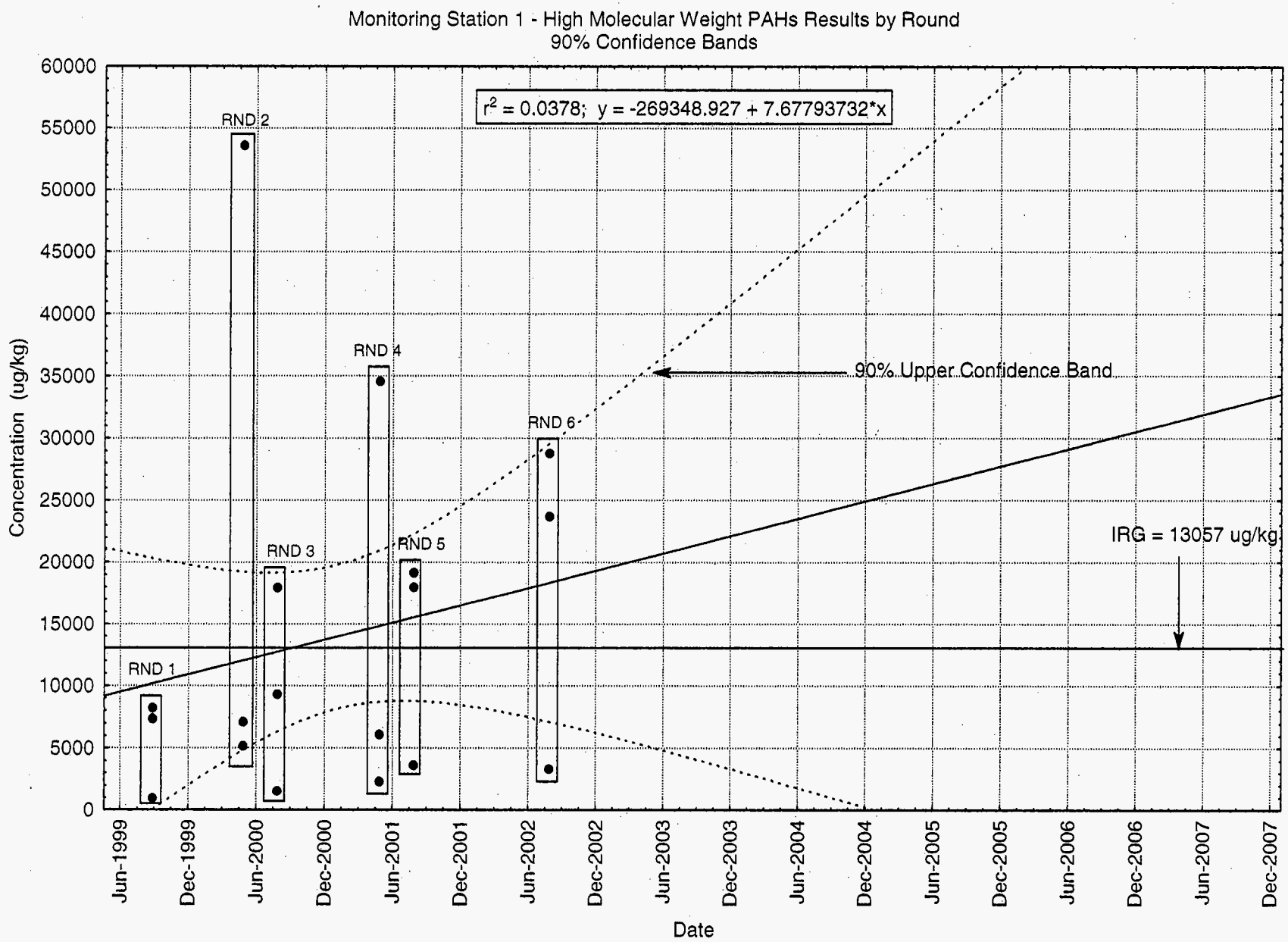
Monitoring Station 1 - Copper Results by Round
90% Confidence Bands



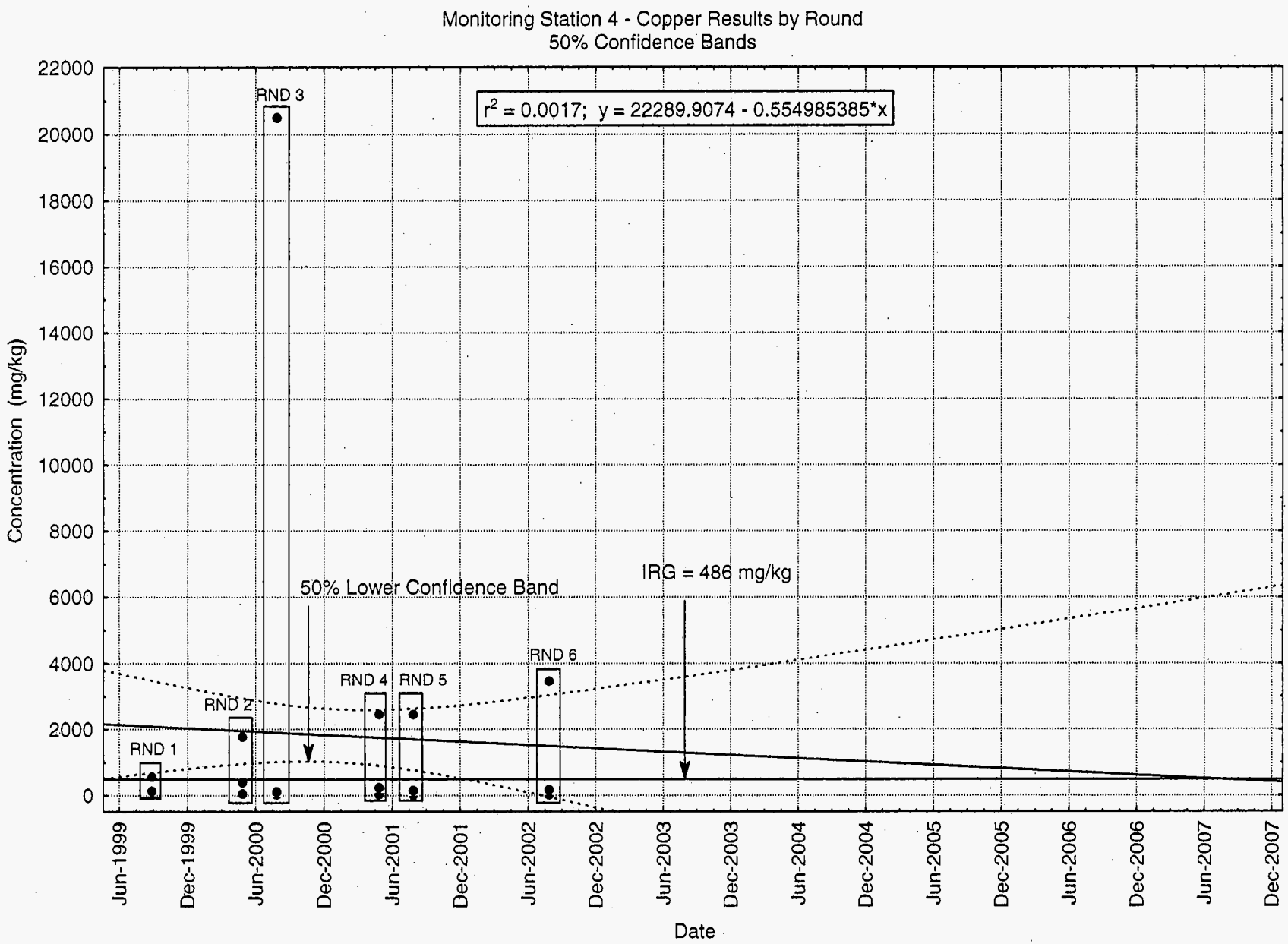
PRELIMINARY



PRELIMINARY

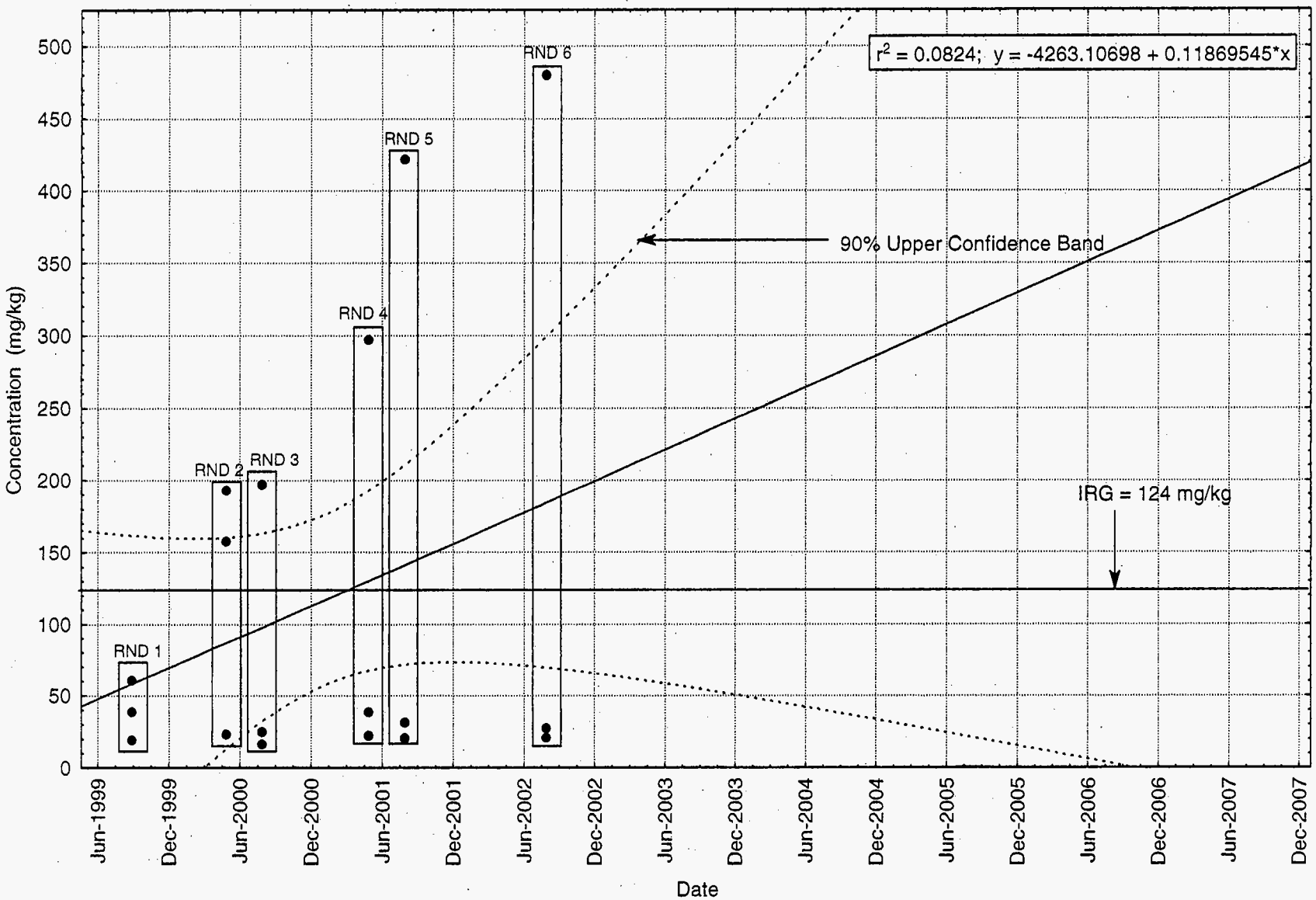


PRELIMINARY

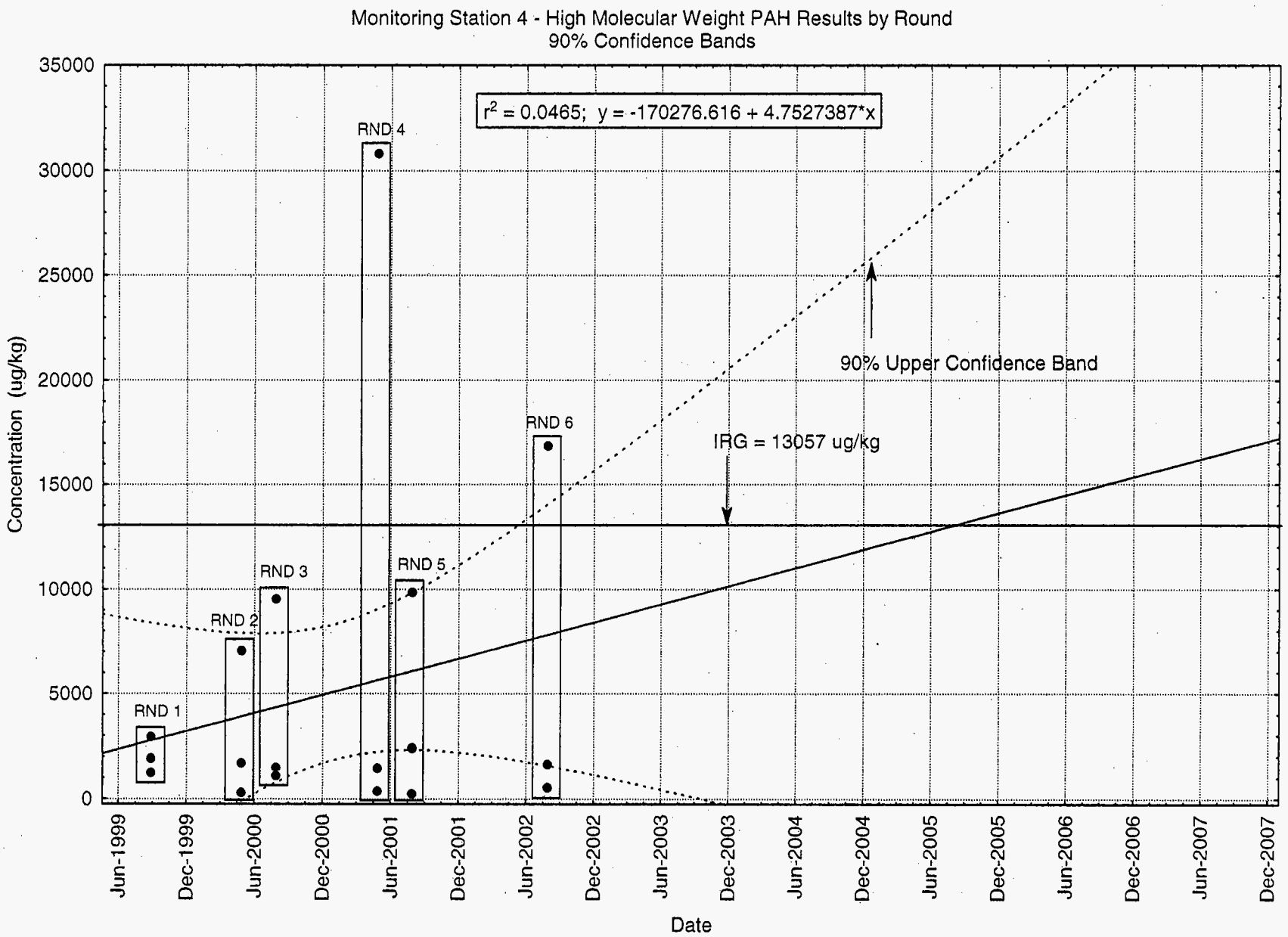


PRELIMINARY

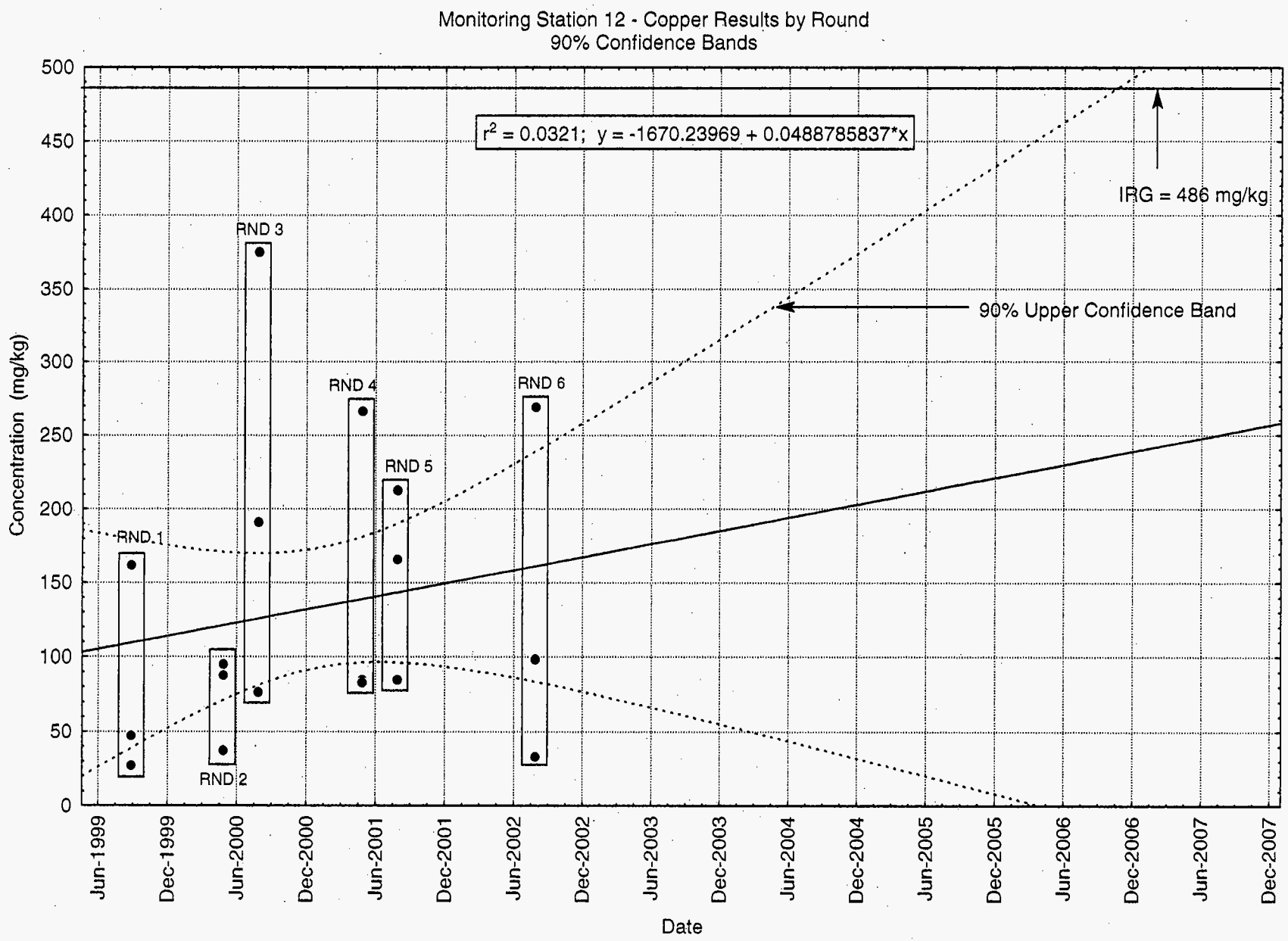
Monitoring Station 4 - Nickel Results by Round
90% Confidence Bands



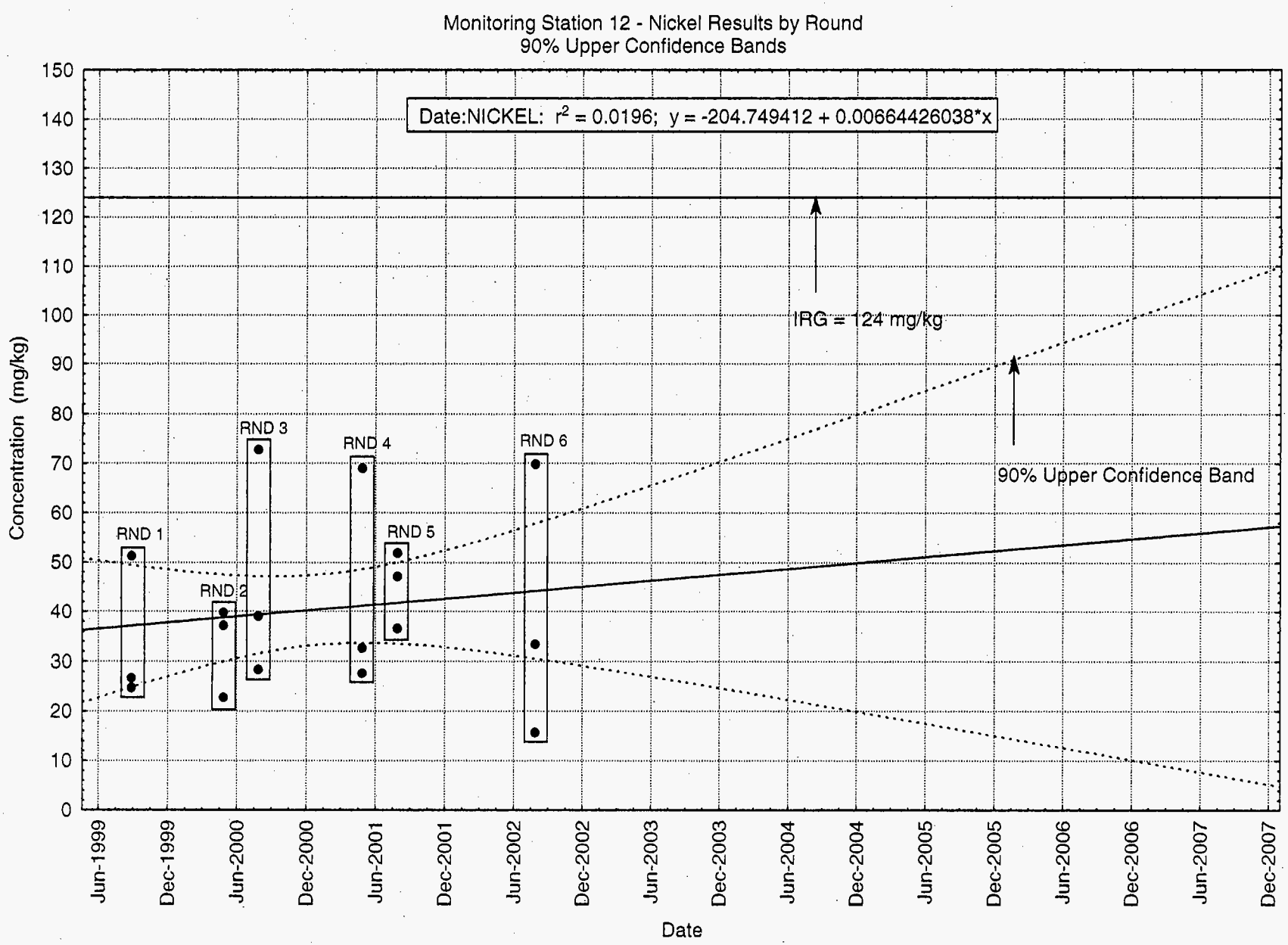
PRELIMINARY



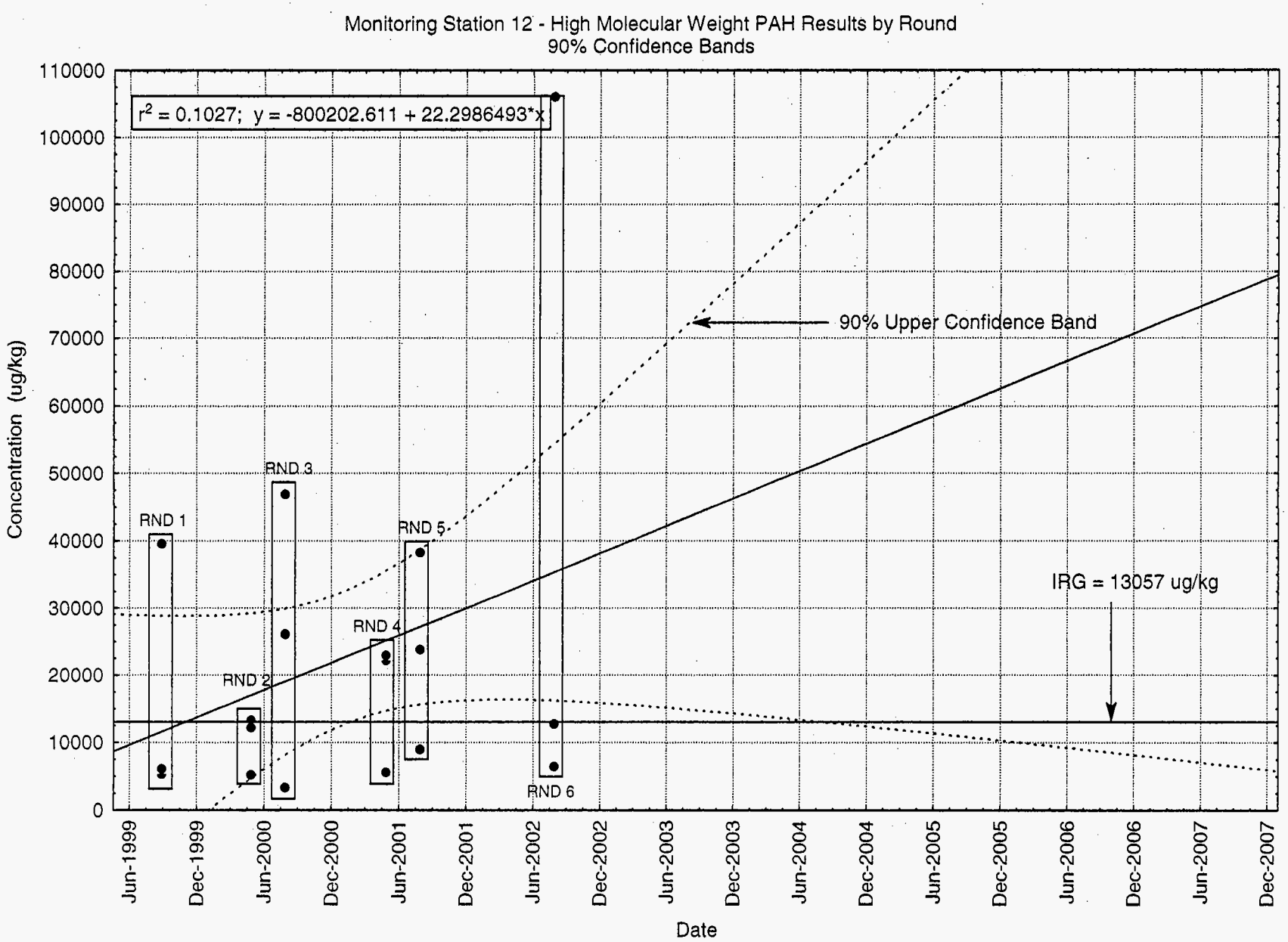
PRELIMINARY



PRELIMINARY



PRELIMINARY



REFERENCES

McLaren/Hart, 1994. Final Human Health Risk Assessment Report for Offshore Media, Portsmouth Naval Shipyard, Kittery Maine. McLaren/Hart Environmental Engineering Corporation, Albany, New York, May.

Navy, 1999. Interim Record of Decision for Operable Unit 4 for Portsmouth Naval Shipyard, Kittery, Maine, May.

NCCOSC (Naval Command, Control, and Ocean Surveillance Center), 2000. Estuarine Ecological Risk Assessment for Portsmouth Naval Shipyard, Kittery, Maine, Narragansett, Rhode Island, May.

TtNUS (Tetra Tech NUS, Inc), 1999. Interim Offshore Monitoring Plan Operable Unit 4 for Portsmouth Naval Shipyard, Kittery, Maine. King of Prussia, Pennsylvania, October.

TtNUS, 2000. Interim Offshore Monitoring Program Round 1 Data Package for Portsmouth Naval Shipyard, Kittery, Maine. King of Prussia, Pennsylvania, February.

TtNUS, 2000. Interim Offshore Monitoring Program Round 2 Data Package for Portsmouth Naval Shipyard, Kittery, Maine. King of Prussia, Pennsylvania, October.

TtNUS, 2000. Seep/Sediment Summary Report for Data Collected between December 1996 and November 1997 for Portsmouth Naval Shipyard, Kittery, Maine. King of Prussia, Pennsylvania, August.

TtNUS, 2001. Interim Offshore Monitoring Program Round 3 Data Package for Portsmouth Naval Shipyard, Kittery, Maine. King of Prussia, Pennsylvania, January.

TtNUS, 2001. Interim Offshore Monitoring Program Round 4 Data Package for Portsmouth Naval Shipyard, Kittery, Maine. King of Prussia, Pennsylvania, September.

TtNUS, 2001. Preliminary Remediation Goals for Operable Unit 4 for Portsmouth Naval Shipyard, Kittery, Maine. King of Prussia, Pennsylvania, November.

TtNUS, 2002. Interim Offshore Monitoring Program Round 5 Data Package for Portsmouth Naval Shipyard, Kittery, Maine. King of Prussia, Pennsylvania, February.

TtNUS, 2002. Baseline Interim Offshore Monitoring Report for Operable Unit 4 for Portsmouth Naval Shipyard, Kittery, Maine. King of Prussia, Pennsylvania, July.

TtNUS, 2003. Interim Offshore Monitoring Program Round 6 Data Package for Portsmouth Naval Shipyard, Kittery, Maine. King of Prussia, Pennsylvania, March.

SAPL Input for January 8, 2004 OU4 Technical Meeting

In November 2003, the Seacoast Anti-Pollution League (SAPL) provided the Navy with comments on the September 26, 2003, *Recommendations for Resolution of Selected Items Prior to the Round 1 through 7 Report, Interim Offshore Monitoring Program.* SAPL received the Navy's responses to comments, along with the Proposed Discussion Topics for the January 8th OU4 Technical Meeting, on December 24, 2003. In the two December documents, the Navy requested that additional information be provided prior to or at the meeting regarding: 1) SAPL's position that dioxin should be retained as an analytical parameter, and 2) SAPL's concerns with data limitations and the Navy's and statistical analysis.

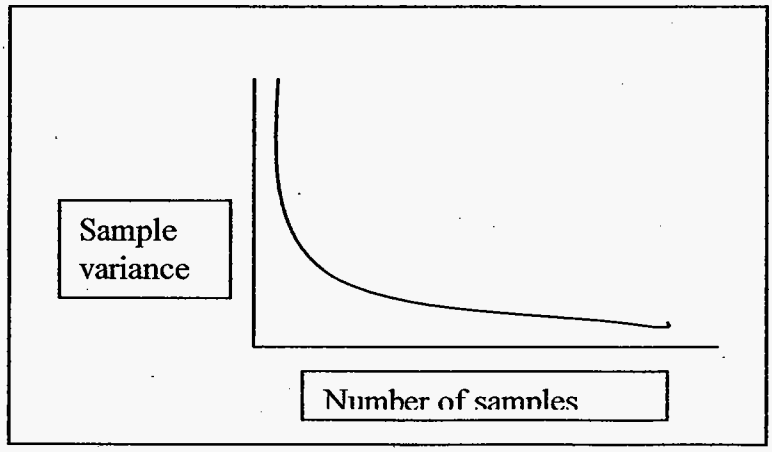
Because the September 26, 2003 OU4 Recommendations did not provide sufficient detail about the Navy's statistical approach, SAPL will illustrate its concerns with comments on the September 19, 2003 *Technical Memorandum, Evaluation of Dioxin Results for Ash and Recommendation Regarding Additional Dioxin Analysis for Site 34.* SAPL believes that its concerns regarding the basis for the Navy's proposal to drop dioxin analysis at Site 34 also apply to the Navy's approach to OU4.

The last four columns on Table 4 [of the September 19, 2003 Technical Memorandum] tabulate the findings from parametric and non-parametric statistical tests based on data tabulated in columns 2, 3, and 4 of the same table. This table appears to be offered as support for the argument "<u>that the site samples and the background samples are statistically similar</u>" as proposed in the text of the memorandum. It appears from the tables that the Navy is using estimates based on very small sample sizes. In Table 4, statistics are calculated based on THREE site samples for which measurements exceed detection levels in 0, 2 or 3 times ,depending on the isomer or isomer average. It is not surprising that the tabulation in the last column of Table 4 shows that 18 tests do not <u>statistically</u> exceed background using this test rationale with small sample sizes.

This information is meaningful only if a difference is found; otherwise it is meaningless. Whether failure to reject that null hypothesis, " <u>that the test sites are different from the background,"</u> is due to small sample numbers that inflate the variances in the test group, or due to actual differences, cannot be determined from these data. That is because the T value is . dependent on the estimates of the variance of the data and the sample size when there are small numbers. For these extremely small data sets, <u>sample size dominates that calculation</u>.

Variance = Sum (Xi –Mean)/ sample number – 1

When sample numbers are 2 and 3 the divisor is 1 or 2 respectively.



As shown above, very small sample sizes give inordinately large variances. The true population variance is only shown as the sample size increases in the figure. The number does not need to be large but the when the number is very small the variance is over-estimated

In the next step of the hypotheses testing, the T statistic is calculated. The variance is divided into the differences between the mean of the sample and the mean of the reference group. The magnitude of the T values determines whether the difference is statistically significant. When the variance is "over-estimated", the T value declines and the conclusion reached is, "THE HYPOTHESIS THAT THE SAMPLES ARE DIFFERENT IS REJECTED". That does not mean that the sample and reference groups are the same but only that the given data sets do not detect a difference. To decide to accept that the data are from the same population is a Type II error, rejection of the null hypothesis when it is true. Studies with inordinately small sample sizes lack the power necessary to reject the null hypothesis that the samples are the same.

Notice the following:

$$T = \frac{\text{sqrt N (sample mean} - \text{reference mean)}}{\text{Variance}}$$

Thus if the variance is very large, the T value declines and the test loses power. When sample numbers are very small the size of the variance reflects uncertainty due to the small number and the uncertainty in the actual variability of the unknown and reference populations.

That is exactly the case with the data shown in Table 4. It also will be a problem in the testing described in the Recommendations for OU 4.

The tests for the distribution shown in columns 13-15 found in Table 4 do not correct for this weakness.

These data can only be used to statistically to test whether the site is above background statistically. Given the size of the sampling data, no statistical statement can be advanced with respect to whether the site is below background or equal to background.

There are three other serious problems with using these data to support statistical inferences in the manner shown.

1. Although there are 13 isomers tested at the laboratory, they are not independent of each other because they all came from the same three grab samples. They do not characterize the site differences.
2. The TEQ and isomer totals shown are taken from the samples above, and statistical testing of these samples a second time also cannot be considered an independent test.
3. The dates when the background samples are taken are different with respect to season, tide level and perhaps other factors. This increases the likelihood that the variance of the reference sample is increased.

Table 4 tabulations also show some use of non-parametric tests. These tests also lack power to reject the null hypothesis. They are very useful when they show differences statistically, as they do in some cases. They lack the power to support the conclusion that there are no differences.

This testing strategy is useful for screening when positive differences are found, but if there are not positive differences more information is needed before one can accept that the samples are essentially from the same population. This referred to as the "Type II error" or power of the test statistic. There is methodology to determine Type II errors. The Navy should explore those methods and logic.

SAPL does not believe that the findings in the September 19, 2003 Technical Memorandum can be used to support the dropping of dioxin from the testing at Site 34. With regard to the OU4 Recommendations, SAPL does not have confidence that the Navy's basis for decision-making is markedly better than for Site 34.

105OU4Jan8TechMtgSAPLinput.JA4