Click here to return to Administrative Record Index

USACE CONTRACT NO. DACW33-94-D-0002
TASK ORDER NO. 017
TOTAL ENVIRONMENTAL RESTORATION CONTRACT

all 8 1/2 x 11
double sided
except where noted

Superfund Records Center
SITE: New Bedford
BREAK: 6.4
OTHER: Doc # 25709

TECHNICAL MEMORANDUM
COMPARISON OF PCB NOAA CONGENER WITH TOTAL HOMOLOGUE GROUP CONCENTRATIONS
NEW BEDFORD HARBOR SUPERFUND SITE
Bristol County, Massachusetts

May 2001

Prepared for
U.S. Army Corps of Engineers
New England District
Concord, Massachusetts

2001-017-0119
5/10/01

USACE CONTRACT NO. DACW33-94-D-0002
TASK ORDER NO. 017
TOTAL ENVIRONMENTAL RESTORATION CONTRACT

TECHNICAL MEMORANDUM
COMPARISON OF PCB NOAA CONGENER WITH TOTAL HOMOLOGUE GROUP CONCENTRATIONS
NEW BEDFORD HARBOR SUPERFUND SITE
Bristol County, Massachusetts

May 2001

Prepared for

U.S. Army Corps of Engineers
New England District
Concord, Massachusetts

Prepared by

Foster Wheeler Environmental Corporation
133 Federal Street
Boston, MA 02110

Revision	Date	Prepared By	Approved By	Pages Affected
0	5/10/01	H. Douglas		All

2001-017-0119
5/10/01

TABLE OF CONTENTS

EXECUTIVE SUMMARY ES-1

1.0 INTRODUCTION 1-1
- 1.1 Report Organization 1-2
- 1.2 Procedure for Collection and Splitting Samples 1-2
- 1.3 Analytical Methodology 1-3
- 1.4 Data Review and Validation 1-4

2.0 DATA COMPARISON RESULTS 2-1
- 2.1 Correlation Approach 2-1
- 2.2 Best Fit Regression Equations for the Entire Concentration Range (Phase I Data) 2-2
  - 2.2.1 Regression Analysis for Low Concentrations (Phase I) 2-5
  - 2.2.2 Comparison of Equations for the Lower Concentration Range (Phase I) 2-7
  - 2.2.3 Evaluation of Regression Analysis for Higher Concentrations (Phase I) 2-9
  - 2.2.4 Phase I Correlation Summary 2-11
- 2.3 Regression Results for Phase I and Phase II Data Set 2-12
  - 2.3.1 Phase I and Phase II Low Concentration Data Correlation 2-13
  - 2.3.2 Summary of Regression Results for Phase I and II Data Sets 2-18
- 2.4 Comparison of NOAA Congener to Aroclor Data 2-20
- 2.5 Comparison of Aroclor to Homologue Group Results 2-23

3.0 RECOMMENDATIONS FOR FUTURE DATA EVALUATION 3-1

4.0 REFERENCES 4-1

LIST OF FIGURES

Figure 2-1 Total Homologue Groups as a Function of NOAA Congeners for Scenario #1 2-3
Figure 2-2 PCB Homologues as a function of NOAA Congeners for Scenario #2 2-4
Figure 2-3 PCB Homologues as a Function of NOAA Congeners for Scenario #1 2-14
Figure 2-4 PCB Homologues as a Function of NOAA Congeners for Scenario #2 2-15
Figure 2-5 PCB Homologues as a Function of NOAA Congeners for Scenario #1 (Low Concentrations) 2-16
Figure 2-6 Total PCB Homologues as a Function of NOAA Congeners for Scenario #2 (Low Concentrations) 2-17
Figure 2-7 Measures of Agreement for Calculated Total PCBs (2.5*NOAA Congeners) And Total Homologue Groups 2-19
Figure 2-8 PCB Aroclors as a Function of NOAA Congeners for Scenario #1 2-21
Figure 2-9 PCB Aroclors as a Function of NOAA Congeners for Scenario #2 2-22
Figure 2-10 PCB Aroclors as a Function of PCB Homologues 2-24
Figure 2-11 Measures of Agreement for Total PCB Homologues and Total PCB Aroclors 2-26

LIST OF TABLES

Table 2-1 Best Fit Regression Lines for Different Scenarios 2-2
Table 2-2 Regression Equations for the Low Concentration Data Range (Phase I) 2-6
Table 2-3 Percent Difference Categories for Different Slopes (m) (Phase I) 2-8

TABLE OF CONTENTS - *Continued*

Table 2-4	Summary of Bias (Measured as %D) Using a Different Slope (M) for Low Concentrations	2-8
Table 2-5	Summary of Calculated PCB Results that Disagree Relative to 1 ppm (Phase I)	2-9
Table 2-6	Summary of Calculated Values that Differ Relative to Decision Criteria (Phase I)	2-10
Table 2-7	Summary of Percent Difference Distribution for Higher Concentrations (Phase I)	2-10
Table 2-8	Summary of %D for Higher Concentrations (Phase I)	2-11
Table 2-9	Regression Equations for the Combined Phase I and Phase II Data Sets	2-12
Table 2-10	Regression Equations for Low Concentration Range (Phase I and Phase II Data)	2-13
Table 2-11	Regression Equations for the Comparison of NOAA Congeners to Aroclors	2-20
Table 2-12	NOAA Congeners to Aroclor Regression Equations for the Low Concentration Range	2-23
Table 2-13	Regression Equations for the Comparison of Aroclors-to-Homologue Groups	2-25

LIST OF ATTACHMENTS

Attachment A	Phase I Data – Linear Regression Outputs and Sequential Standard Residuals for Scenario #1 (Homologue-to-Congener Ration <1 Removed)
Attachment B	Phase I Data – Linear Regression Outputs and Sequential Standard Residuals for Scenario #2 (Ratios <1 Included)
Attachment C	Phase I Data – Data Table Showing Linear Regression Input and Outliers
Attachment D	Phase I Data – Percent Difference Tables and Figures
Attachment E	Phase I and II Data – Linear Regression Outputs and Sequential Standard Residuals for Scenario #1 (Ratios <1 not Included)
Attachment F	Phase I and II Data – Linear Regression Outputs and Sequential Standard Residuals for Scenario #2 (Ratios <1 Included)
Attachment G	Phase I and Phase II Data – Linear Regression Outputs and Sequential Standard Residuals for Scenario #1 (Ratios <1 Not Included)
Attachment H	Phase I and Phase II Data – Linear Regression Outputs and Sequential Standard Residuals for Scenario #2 (Ratios <1 Included)
Attachment I	Phase I and Phase II Data – Linear Regression Outputs and Sequential Standard Residuals (Ratios <1 Included)

EXECUTIVE SUMMARY

Foster Wheeler Environmental (Foster Wheeler) prepared this Technical Memorandum to document an analysis of whether a robust, site-specific relationship exists between the PCB concentrations measured using two different analytical methods for quantifying sediment PCB concentrations at the New Bedford Harbor (NBH) Superfund site. Analysis led to identification of a linear relationship between the sum of the concentration of the ten PCB homologue groups and the sum of the concentrations of the 18 PCB congeners on the NOAA list. This identified relationship has only been evaluated with regard to its applicability to sediment samples collected from New Bedford Harbor.

This evaluation was performed in two phases as split sample comparison data quantified using both analytical methods became available. The initial draft correlation report dated January 2001 was based on the analysis of 95 split sample result pairs from the Phase I sediment sampling event. Since the draft correlation report was submitted, an additional 66 result pairs from the Phase II sampling event became available. Consequently, a total of 161 result pairs were evaluated for this draft final report. A total of 115 split samples also were analyzed for Aroclors. These results are also briefly referenced and discussed.

Foster Wheeler evaluated the relationship between the sum of the total homologue groups and the sum of NOAA congeners using linear regression. Depending on the data set used for the regression analysis and whether the regression relationship is constrained to pass through the original point (i.e., zero NOAA congeners and zero total homologue groups), different equations were fit to the data to explore the strength of the relationship between the results obtained from the two analytical methods. The two scenarios that were selected for independent evaluation were:

- **Scenario #1:** Data points were removed where the sum of the homologue groups was less than the sum of the congeners (i.e., the ratio of measured homologues to measured congeners was less than 1), because this is a physically impossible condition. Statistical outliers also were removed.

- **Scenario #2:** Only the data points determined to be statistical outliers were removed from the data set, and those with homologue-to-congener ratios less than 1 were considered valid data points (see Section 2.1 for the rationale for this inclusion).

Scenarios #1 and #2 were evaluated using both the Phase I data set (95 samples) and the combined Phase I and II data set (161 samples).

The regression equations generated for both scenarios without the constraint of passing through the origin (i.e., the "non-zero y-intercept" case) had positive y-intercepts greater than the lowest cleanup goal of 1 ppm (i.e., greater than the 1 ppm action level for residential areas). As such, no total PCB concentration less than these y-intercept values could ever be calculated regardless of the measured NOAA congener results. Therefore, the non-zero y-intercept equations could not be used to calculate total PCBs near the lowest cleanup goal of 1 ppm.

To address this concern, the low PCB concentrations must be calculated using a zero-intercept regression equation or two separate equations (each fit to a specific range of sediment PCB concentrations). In order to calculate total PCBs for the entire concentration range of interest, two possible regression approaches were tested:

1) Use a linear regression model forced through the origin for the entire concentration range. Using one equation with the regression line forced through zero would greatly simplify the

future approach to determining total PCB concentrations and monitoring the linear relationship as more data become available.

2) Use a positive y-intercept equation to calculate total PCBs greater than 1 ppm, and use a second equation applicable for calculating lower concentrations.

To explore the possibility of identifying separate equations for two different concentration ranges, the lower concentration data (near or below 1 ppm total PCBs) for each scenario were used to generate linear regression equations applicable to this concentration range. The maximum concentration for the lower range was defined by filtering out the higher concentration data from the data set until three criteria were met: 1) the resulting regression equation produced results up to and including the 1 ppm homologue concentration (equivalent to total PCBs); 2) the y-intercept for the equation was generally consistent with the low concentration values; (i.e., markedly less than 1 ppm); and 3) the resulting correlation coefficient indicated a reasonably robust relationship between the congeners and homologue results.

Based on the evaluation of the Phase I data, it was recommended that the zero y-intercept equation with a slope of 2.5 be used to calculate total PCBs from the sum of NOAA congeners over the entire concentration range of interest. This relationship was recommended as technically and statistically valid, and would be significantly more practical and straightforward to implement and monitor during future project work. The following considerations supported this recommendation:

- Based on the Phase I data set, the zero y-intercept regression is as statistically valid as the non-zero y-intercept model. Both the actual data set and the theoretical basis for this correlation support the premise that the "true" relationship is defined by a y-intercept that is either zero or very slightly positive. Analytical detection limits for both methods are less than 0.002 ppm and the data set shows general agreement for the lower concentrations (i.e., the total homologues are typically greater than the total congeners and a median ratio of 2.7). There is no apparent reason (analytical or otherwise) that the form of the relationship would be different for concentrations near 1 ppm (total PCBs).

- Use of a single regression equation for the entire PCB concentration range of concern would simplify implementation and minimize the potential for confusion and error to support NBH project decisions.

- The use of two different regression equations based on concentration would necessitate determining which concentration range applies to each equation. PCB sediment data in the lower range of concern (e.g., total PCBs at or below a 1 ppm action level) are rather limited and this decision would be somewhat subjective and possibly to change based on a larger data set. This possibility creates significant documentation issues regarding which relationships were and potentially are used for specific project decisions at particular points in time.

Following submittal of the Phase I correlation results, the Phase II data were added to provide a larger data set (resulting in a total of 161 analytical result pairs). This larger data set was used to re-calculate and re-evaluate the four possible linear relationships described above for the Phase I data set. The regression equations from this larger data set were found to be similar to those calculated for the Phase I data. The zero y-intercept equations for both scenarios had slopes of 2.5, confirming this form as the best fit zero y-intercept relationship.

To further assess whether the zero y-intercept approach was appropriate for defining the relationship, the positive y-intercept equation and separate relationships based on concentration range were also generated for the combined Phase I and II data sets. The positive y-intercept and low concentration range

relationships for the larger data set were quantitatively different from those generated using only the Phase I data set, verifying the concern that adding new data is more likely to have a greater impact on the more complex (segmented) regression relationships. Overall, the regression equations identified for the two separate concentration ranges became more similar to each other, and the slopes and y-intercepts grew closer to those of the zero y-intercept model for the larger data set than for the limited Phase I data set. These results add confidence that a zero y-intercept model is an appropriate relationship and that the addition of more data points is likely to lead to further convergence on a relationship closely approximated by the identified zero y-intercept model. At a minimum, these results support the premise that a single zero y-intercept relationship to define the entire concentration range is less likely to change as more data become available than using two different equations depending on concentration.

To further test and evaluate this relationship, the zero y-intercept equation fit to the entire concentration range was used to calculate total PCBs based on the NOAA congener measurements and these concentrations were compared to the total homologue data using the percent difference metric. Overall the comparisons agreed well with over 90% of the data being "in agreement" (defined to be the case where both results are less than 0.5 ppm or greater than 50 ppm, or a percent difference less than 100%). Only 15 of 161 comparisons (9%) had poor agreement (defined as a percent difference greater than 100 %). Overall, this represents a good correlation and a strong basis for future comparisons. A similar evaluation approach is recommended as a relatively quick way to assess agreement for subsequent data and to reduce the frequency of updating the correlation relationship developed in this memorandum.

The statistical approach and assumptions used for this data evaluation are based on the premise that the relationship between the NOAA congeners and the sum of the homologue groups does not vary at different concentration levels. For the reasons stated above, it is recommended that the regression equation with a slope of 2.5 and y-intercept of zero be used for this initial data set, with continued monitoring of the effect on cleanup decisions, especially for the lower concentration range. As a larger data set becomes available, the relationship should be re-examined relative to the new data set. If a larger data set indicates a need to make adjustments in the calculation of total PCBs, the regression approach will be updated accordingly.

In conclusion, it is recognized that this regression equation to convert the sum of the 18 NOAA congeners to total PCBs equivalent (based on the total homologue groups) is site-and media-specific. Therefore, this relationship is only applicable for the New Bedford Harbor Superfund site and only for sediment analyses. As discussed above, the appropriate regression equation may be subject to modification over time as more data become available even though the identified relationship appears to be very robust. Also, if a similar relationship is required, for other matrices (e.g., air or groundwater), a separate regression relationship should be developed and tested. Any other use of this regression without independent validation could lead to erroneous conclusions/decisions that would be unacceptable.

1.0 INTRODUCTION

Foster Wheeler Environmental Corporation (Foster Wheeler) prepared this Technical Memorandum to statistically characterize the correlation relationship between the 18 NOAA PCB congeners and total PCB concentrations (as defined by the sum of the homologue group concentrations) in sediment samples from New Bedford Harbor. Based on this characterization, a regression equation has been developed for future correlation of site PCB congener and homologue data. This memo incorporates comments received from EPA on the earlier Technical Memorandum entitled *Sediment Sample Analytical Method Comparison,* dated May 10, 2000 which evaluated several PCB analytical methods and various geographical comparisons, and supercedes that memorandum.

The earlier comparison memorandum was submitted in draft form and included an evaluation of three analytical methods for reporting PCBs in sediment. These methods were 1) NOAA and WHO congeners using gas chromatography with an electron capture detector (GC/ECD) and EPA Method 8082; 2) Aroclors also using EPA Method 8082; and 3) total homologue groups using mass spectrometry in selected ion monitoring (SIM) mode. The relationship between congeners in a group and the results for homologues (e.g., NOAA di-congeners to total di-homologue) was also evaluated. The different relationships for different geographical areas of the Harbor also were briefly discussed. Based on subsequent discussions and comments on the earlier memorandum, the principal focus of this evaluation is on the statistical relationship between the sum of the 18 NOAA congeners and the total homologue analytical results using relevant portions of the same data set. The Aroclor data from the earlier memorandum are also briefly summarized.

EPA's input to determining an analytical approach to PCB analysis for this project was provided at a meeting held on July 20, 2000 with the EPA, USACE, ENSR and Foster Wheeler and in subsequent phone conversations with Mr. William Nelson (EPA Narragansett Laboratory), Ms. Barbara Bergen (EPA Narragansett Laboratory), Mr. David Dickerson (EPA Region I, RPM), and Ms. Helen Douglas (Foster Wheeler). The conclusions reached in these discussions are summarized as follow:

- PCB concentrations in sediment will be determined by analysis for 18 NOAA congeners using gas chromatography (GC) with an electron capture detector (ECD). This method was chosen to provide data most comparable to other analytical methods being used for other sample media and to minimize analytical difficulties associated with weathering and/or differential migration of individual congeners. This method also can be used to provide relatively cost-effective, rapid turnaround analysis at an on-site or an off-site laboratory during future construction activities, if needed.

- In order to compare sediment sample results with ROD defined total PCB cleanup goals, a scientifically defensible mathematical relationship between the sum of the NOAA congeners and total PCBs is needed.

- For the purposes of developing the relationship between the sum of the NOAA congeners and total PCBs, the best method for determining total PCBs is analysis of total homologue groups by GC/mass spectrometry selective ion monitoring (MS SIM).

Accordingly, this memo focuses on the statistical relationship between existing sediment data for total NOAA congeners analyzed by GC/ECD and total homologue groups by GC/MS SIM. Recommendations are provided for a mathematical regression equation to be used by the project for calculating total PCBs relative to ROD defined action levels. At the request of other reviewers, this memo also documents the relationship between Aroclor and homologue data and Aroclor and NOAA congeners. The results discussed in this memo and the equations derived to define the relationships between different analytical

methods for PCBs are specific to sediment samples from New Bedford Harbor. These equations are not applicable and should not be used for other sites or matrices.

1.1 Report Organization

Initial work for this correlation study was conducted following the Phase I sampling program conducted in the fall of 1999 and initially reported in May 2000. Based on reviewer comments, the focus of the correlation was adjusted as mentioned above and a draft version of this memorandum issued in January 2001. Concurrently with discussions regarding the correlation results, additional samples (Phase II) were collected to further define the extent of PCB contamination at New Bedford Harbor. These data are now available and are included in Section 2.3 of this memorandum and may be of the most interest to many readers. In order to document the sequence of developing the correlation equation, the regression results are discussed in chronological order in Section 2 of this memorandum. Note that evaluation of the Phase I data generated several valid regression equations to define the relationship between analytical results. Of these possible equations, a single y-intercept equation was recommended for use until the additional Phase II data became available to substantiate this recommendation.

In addition, several reviewers requested that the correlation results for Aroclors and congeners and Aroclors and homologue groups also be documented. These are also included in Section 2. An outline of this memorandum is as follows:

1.0 Introduction
- 1.1 Report Organization
- 1.2 Procedure for Collection and Splitting Samples
- 1.3 Analytical Methodology
- 1.4 Data Review And Validation

2.0 Data Comparison Results
- 2.1 Correlation Approach
- 2.2 Best Fit Regression Equations For the Entire Concentration Range (Phase I Data)
  - 2.2.1 Regression Analysis for Low Concentrations (Phase I)
  - 2.2.2 Comparison of Equations for the Lower Concentration Range (Phase I)
  - 2.2.3 Evaluation of Regression Analysis for Higher Concentrations (Phase I)
  - 2.2.4 Phase I Correlation Summary
- 2.3 Regression Results for Phase I and Phase II Data Set
  - 2.3.1 Phase I and Phase II Low Concentration Data Correlation
  - 2.3.2 Summary of Regression Results for Phase I and II Data Sets
- 2.4 Comparison of NOAA Congener to Aroclor Data
- 2.5 Comparison of Aroclor to Homologue Group Results

3.0 Recommendations for Future Data Evaluation

Data tables and figures documenting the regression results are included in the Attachments and are referenced in the text, where appropriate.

1.2 Procedure for Collection and Splitting Samples

Foster Wheeler collected samples from various locations throughout the Harbor to further delineate areas of sediment above the ROD defined cleanup goals. Sampling locations were selected by EPA and USACE, and were chosen to fill data gaps identified following earlier sampling events. Samples were collected in one-foot intervals from the mud line surface to depths of up to four feet, depending on

sampling location and the data need. Samples were identified sequentially by location with a suffix designating the bottom depth of the one-foot interval (i.e., -2 designates the 1 to 2 foot interval).

Samples were collected using various methodologies depending on the nature of the sediments. Equipment included a hand auger, manually driven split spoon, and vibracore sampler (used mostly for off-shore locations accessed by boat). Equipment was selected to maximize sample recovery and ensure the representativeness of the depth interval being sampled. Samples were composited over the one foot interval by mixing in a disposable aluminum pie plate with a stainless steel spoon. Following mixing, the sample was apportioned into jars, with the total number of jars dependent on the analyses.

Approximately 700 samples were collected during the Phase I sampling event (autumn 1999) used for the initial comparison. Of these, 95 samples were selected (every seven to eight samples) for this correlation study and were analyzed by the three different methods described in Section 1.3 below. Note that originally 100 samples were chosen for this correlation; however, some samples were not collected due to the proximity to the high power Com Electric cable crossing, refusals, and slight modifications in the sampling plan. The number of samples available for this correlation study was reduced accordingly. In addition, samples were split at a frequency of one in twenty for QA/QC analyses (field duplicates, MS/MSDs, and USACE split samples analyses).

Another 715 samples were collected during the Phase II sampling event. Sixty-six (66) of these were split for congener and homologue analysis for a total of 161 congener and homologue comparisons. In addition, 115 samples were analyzed for Aroclors in addition to congeners and homologue groups. These data are also included in this correlation report.

1.3 Analytical Methodology

Samples were analyzed using the three analytical methods described below. The methods were selected following discussions with EPA, USACE, and Foster Wheeler, and were intended to provide the most accurate methodology at a reasonable cost to the project.

Total Homologue Groups – This method uses gas chromatography (GC) in combination with low resolution mass spectrometry (LRMS) to selectively identify and quantify PCB groups based on their specific mass. Results are reported for each homologue group (i.e., total mono-through deca-PCBs). The total homologue group method was expected to provide the most accurate measure of total PCBs as it reports PCBs by mass with minimal potential for falsely high data or missed compounds. The drawbacks to this method are that it requires highly specialized equipment, software, and highly trained analysts. Accordingly, it is relatively expensive and impractical to implement for rapid turnaround analysis or in an on-site laboratory if the project requires it at a later date.

Congener Analysis – Selected PCB congeners were analyzed using EPA Method 8082 (GC/ECD). The congener method identifies selected individual congeners (18 NOAA congeners) using retention time with second column confirmation for both identification and quantification. Note that the initial study also included the WHO list of congeners in addition to the NOAA congeners. These were determined to be of such low concentration relative to the total homologue concentrations that statistical comparisons were difficult. The previous May 2000 comparison memo recommended that the WHO congeners not be used to develop total PCB concentrations relative to cleanup goals and they were not included in subsequent analyses.

Because the congener method reports individual congeners based on retention time, false negatives due to weathering are minimized. This method is subject to potential false positives from target and non-target analytes. In some ways, this is more of a concern than for the Aroclor method because pattern

recognition is not used to selectively choose peaks for quantification. On the whole, the effects of potential false positives on the total PCB concentration are minimized by the use of second column confirmation (with the lower of the two values is reported) and the premise that the proportion of interferences from non-target PCBs is proportionally the same for all sediment samples collected from the Harbor. Therefore, any interferences would be included in the correlation equation defining the relationship with total homologue groups.

Congener analysis was selected as the primary analytical method for New Bedford Harbor sediment samples. By performing this correlation study, it was intended that a relationship between congeners and homologue groups would become a part of a defensible, site specific means of performing total PCB analysis for sediment samples for the lifetime of this project (possibly ten years or more). The method was selected as a cost effective alternative to homologue group analysis and less likely to mis-identify or mis-quantify PCBs due to weathering than individual Aroclor identification and quantification. This method could be relatively easily implemented in an on-site laboratory at a later date and is likely to remain consistent with analytical advances over the upcoming years.

Aroclors – Analysis for individual Aroclors was also included in the initial correlation study. Based on comments from EPA, this analysis will no longer be used to evaluate sediment data relative to project cleanup goals. For this study, Aroclors were analyzed by the same laboratory that performed the congener analysis (homologue analysis was performed by a second laboratory), reducing some of the variability due to the field and laboratory handling process. For this reason, the Aroclor results are interesting and included for comparison with other total PCB values. In this context, Aroclor results are discussed in Sections 2.3 and 2.4 of this memorandum.

Aroclor analysis is a gas chromatography/electron capture detector (GC/ECD) analysis that identifies the Aroclor mixture of congeners by retention time and pattern recognition. Target compounds are identified primarily by pattern recognition with concentration confirmed using second column confirmation. Quantification is performed by external standard technique using three to five chromatographic peaks that are representative of the Aroclor selected. This is the conventional method for PCB analysis and the method used for much of the historical data collected from the Harbor. The disadvantages to using this method are that more than one Aroclor is present in the Harbor and they are, in some cases, heavily weathered. This method is also subject to interferences from non-target compounds. These method limitations pose difficulties both with identification and quantification creating the potential for falsely biased high results due to overlapping and/or interfering peaks and falsely biased low results due to mis-identification are the result of weathering or interferences.

1.4 Data Review and Validation

Data were reviewed for compliance with relative QC criteria to determine its acceptability for use in achieving project objectives. The primary objective for this sediment data was to identify sediments requiring dredging and/or excavation due to exceedences of action levels specified in the ROD. A secondary objective was to use this data in this comparison study to determine the relationship between the results from the different PCB analytical methods.

Data were given a "checklist" review for compliance with QC criteria. This review was based on the Region I, EPA NE Data Validation Functional Guidelines for Evaluating Environmental Analyses, December 1996 criteria. This brief review was intended to identify significant QC exceedences which may significantly affect the reported sample results and to provide information on the quality of the data in more detail than an EPA Region I Tier I validation, but was not intended to provide as much detail as a Tier II validation. Approximately 10 percent of the data were validated in accordance with EPA Tier II procedures to identify potential systematic QC issues that may not be readily apparent from the less

detailed checklist review. In addition, approximately 5 percent of the samples were split for analysis by the government designated QA laboratory.

The results from the data reviews, validations, and USACE QA split sample comparisons will be discussed in more detail in the final Sediment Sample Report(s), to be submitted under separate cover and will include data validation memoranda in a separate Appendix. The data review/validation process identified two significant findings with the data relative to the use of these sample results. These are summarized in the sections that follow.

In the majority of cases, field duplicate results met the 50% RPD criteria defined by the EPA Region I data validation guidelines as acceptable for soil/sediment precision. In some cases, field duplicates did not agree well and no cause for the discrepancy due to laboratory methodology was identified from the review. The variability appears to be matrix related and is much more apparent with the higher concentration samples.

The second significant finding affected the congener analysis and was identified by the independent calibration check standard. This standard is prepared from a source other than that used for the calibration standards during a typical analytical run sequence and is used to verify that the calibration standards accurately reflect concentration. For some of the analyses, the laboratory purchased a prepared calibration check standard that included non-target congeners in addition to target congeners. Some of these non-target analytes coelute with target analytes on the chromatography column causing an apparently high (twice as high = 200%) recovery for some of the target congeners.

This does not have a direct affect on reported concentrations since the check standard was not used to calculate sample results. It does indicate that non-target congeners could be present in the samples, which would cause a falsely biased high or false positive result for a specific congener. For the purpose of determining total PCBs, it can be assumed that the ratio of non-target to target congeners is the same for each sample. Therefore, for the purposes of this split sample comparison, the potential bias is the same for each sample and would be included in any regression calculations, negating the bias in the total PCB value.

Other QC exceedences were identified during the data review which were relatively minor in nature, potentially affecting individual sample results and generally not applicable to the data set as a whole. Some of the QC exceedences were associated with poor matrix and surrogate spike recoveries, especially for high concentration samples. These appear to be related to the interferences from elevated concentrations of target and non-target analytes and, in some cases, the significant dilutions required to bring extract concentrations within the instrument calibration range. This type of QC exceedence is consistent with the earlier conclusion that the data are likely to more variable at the higher concentrations.

2.0 DATA COMPARISON RESULTS

Sediment samples were collected for PCB analysis during two sampling events referred to as Phase I and Phase II. During each event, samples were analyzed for NOAA congeners as the primary method of analysis with approximately 10 percent of the samples split and also analyzed for total homologue groups. Split samples collected during Phase I also were analyzed for individual Aroclors. These split sample results were used to generate the regression equations discussed below. Phase I sampling generated a total of 95 split sample comparisons used to develop a regression equation recommended for subsequent work and submitted in a draft version of this memorandum in January 2001. Based on results from the Phase I comparison data, a single regression equation was recommended for use in calculating total PCBs from the eighteen NOAA congeners. In addition, it was mutually decided by the Project Team that subsequent split sample data should also be evaluated to determine whether the same or a similar regression relationship was applicable to the larger data set. In order to maintain the sequence of the development of the regression equations, the following subsections discuss the results from Phase I and then from Phase II.

2.1 Correlation Approach

Sediment PCB data were compared using a linear regression analysis of possible relationships between total PCB homologue groups (y-axis or dependent variable) as a function of the sum of the NOAA congeners (x-axis or independent variable). The entire data set was narrowed to include concentrations within the range of interest by removing points where one or both results were non-detect and/or where both results were greater than 50 ppm, with the understanding that these values exceeded the highest cleanup goal. Data were then compared using two scenarios:

- Scenario #1: Those points where the total homologue group concentrations were less than the total measured NOAA congener concentrations (i.e., the ratio of homologues to congeners is less than 1) were removed from the data set as physically implausible results.

- Scenario #2: The entire data set within the range of interest was used for the linear regression analysis, including those points where the homologue concentrations were less than the congener concentrations. Although it is understood that the NOAA congeners are a subset of the total homologues, these samples were split in the field and analyzed by two different laboratories, therefore, field variability is also inherent in this comparison. Under this scenario, those points where the congeners are greater than the homologues were assumed to be statistically valid and were included as a part of the complete data set. Furthermore, to exclude these points would remove the error on only one side of the comparison relationship, artificially increasing the proportion of homologues relative to congeners.

Least square linear relationships were identified both with a y-intercept and by constraining the line to pass through the origin (i.e., the point of zero NOAA congeners and zero homologues). Confining the line through zero is consistent with the premise that the NOAA congeners are a smaller, yet significant subset of the total PCBs and that they are directly proportional (linear) to one another. That is, if some of the NOAA congeners are present, total homologues will be detected and reported. It is possible to have total homologues detected without NOAA congeners (and this is the case for seven comparisons in this study). This is possible because the NOAA congeners represent only a percentage (approximately 40%) of the total PCBs and therefore could be present at such low concentrations that they were not detected and yet total homologues would have a high enough concentration to be detected. Also, the total homologue method is slightly more sensitive (has a lower reporting limit) than the congener analysis.

Data review and validation and this subsequent evaluation did not identify reasons why the y-intercept for the line defining the relationship between the NOAA congeners and homologues should not be a small positive number (significantly less than 1 ppm) or zero. The data set documents that both analytical methods have the capacity to report PCBs to concentrations less than 0.002 ppm, significantly less than the lowest project action level of 1 ppm. In addition, the data does appear to reflect a relationship at these lower concentrations (the majority of homologue results are greater than congeners with a median ratio of 2.5; the median ratio for the low concentration Phase I data set was 2.7)). Non-detect values for both methods are low relative to the 1 ppm action level and for comparison purposes can be considered equivalent to zero.

This evaluation determined two different relationships between the congeners and homologue groups depending on the concentration (discussed in more detail below). An analytical reason for this difference was not determined. However, it is noted that in this evaluation, total congeners and total homologue groups are defined as the sum of detected concentrations, with non-detect values for individual analytes treated as zero. Non-detect values are usually insignificantly small relative to the total when detected concentrations are elevated, but can be close to detected concentrations for the lower analytical range.

The data sets for the two scenarios described above were statistically evaluated using linear regression (both with and without the y-intercept term). Standardized residuals were plotted for each regression result and those with a standardized residual absolute value greater than 4 were removed as outliers (Neter and Wasserman, Applied Linear Statistical Models, Richard D. Irwin, Inc., 1974 Section 3.6 and 4.2). (Note: data points exhibiting a standardized residual of greater than 4 reflect values estimated to have less than 0.005% chance of occurrence based on the distribution of the remaining data.) These outliers were removed and the resulting data regressed once again and the standardized residuals re-plotted and tested for outliers. This process was repeated until the absolute value of the residuals was less than 4.

2.2 Best Fit Regression Equations for the Entire Concentration Range (Phase I Data)

Using the process discussed above, the best fit linear regression equations estimated for the relationship between the sum of the NOAA congeners and total homologue groups, for the Phase I data, are illustrated on Figures 2-1 and 2-2 and summarized in Table 2-1.

Table 2-1
Best Fit Regression Lines for Different Scenarios

Scenario (Data Base Adjustments)	Number of Data Points in Final Phase I Comparison (n)	Line Constrained to Pass Through Origin	Line with a y-intercept	
		Slope (95% CI)	Slope (95% CI)	y-intercept (95% CI)
# 1 (7 ratio <1 pts + 4 outliers removed)	55	2.5 (2.3 to 2.8) $\Delta = 0.50$ $R^2 = 0.82$	2.3 (2.1 to 2.6) $\Delta = 0.58$ $R^2 = 0.83$	4.0 (0.7 to 7.3) $\Delta = 6.56$
# 2 (1 ratio < 1 pt + 4 outliers removed)	61	2.4 (2.1 to 2.7) $\Delta = 0.52$ $R^2 = 0.77$	2.2 (1.9 to 2.5) $\Delta = 0.62$ $R^2 = 0.78$	3.3 (-0.17 to 6.8) $\Delta = 6.93$

CI = Confidence Interval
Δ = Range of the 95% Confidence Interval
R^2 = Correlation Coefficient

Figure 2-1
Total Homologue Groups as a Function of NOAA Congeners
Scenario #1

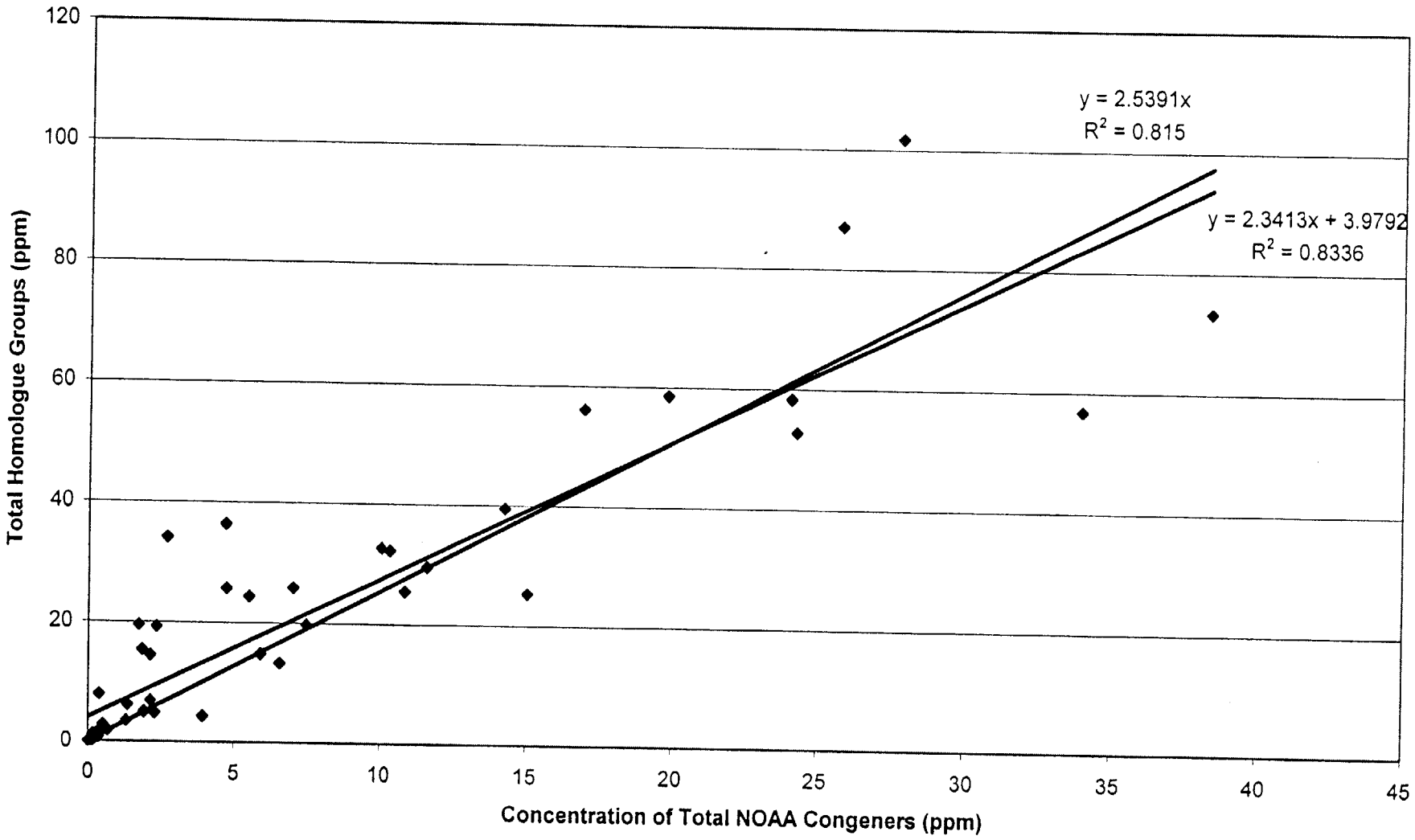
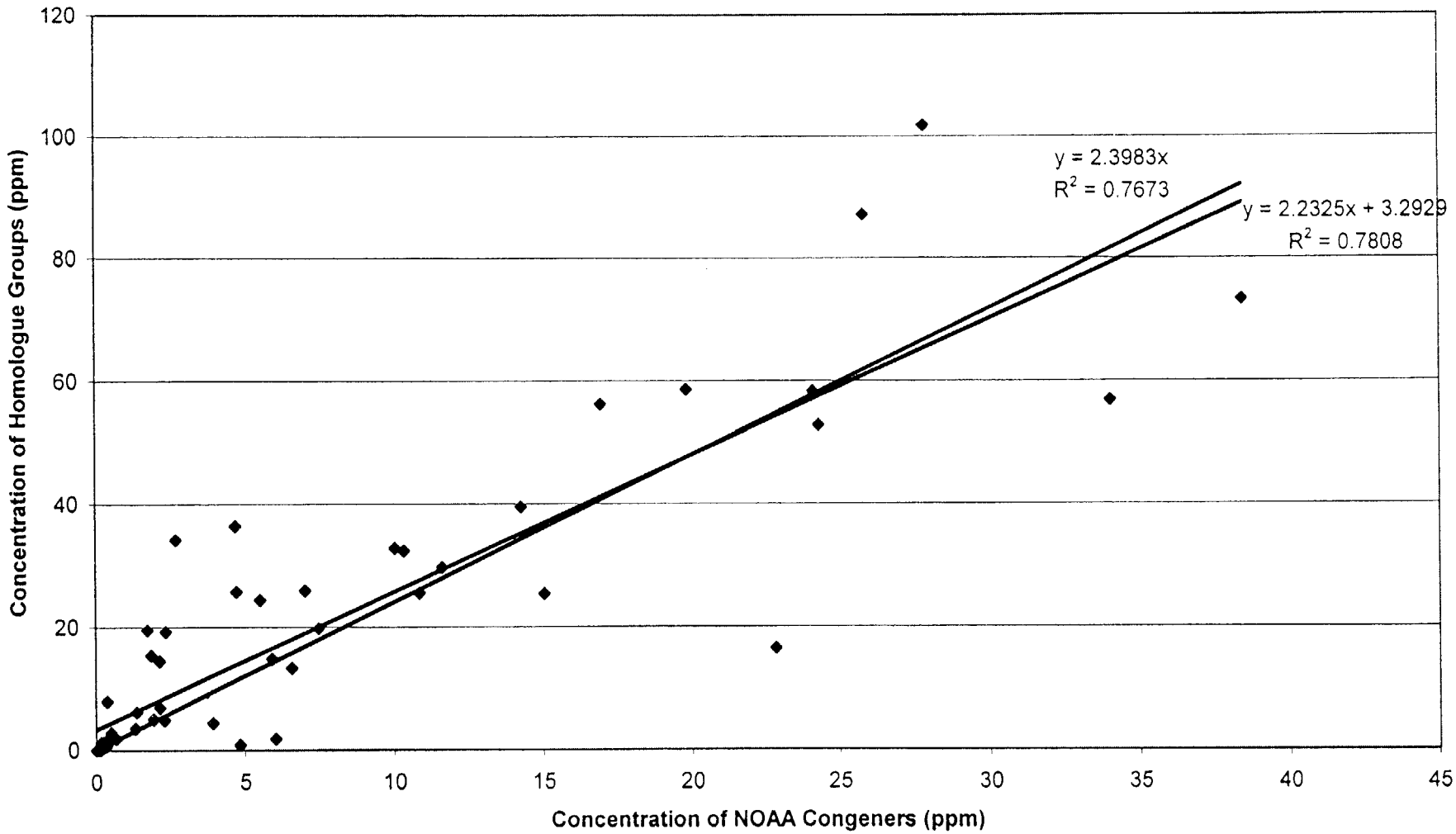


Figure 2-2
PCB Homologues as a Function of NOAA Congeners
Scenario #2



Regression and outlier tests of both data sets are included in Attachments A (Scenario #1) and B (Scenario #2). The cover page of each attachment summarizes the outliers for each data set and the correlation coefficients. The data table summarizing the input into each regression model is included in Attachment C.

For reference, the regression and outlier tests generated from Scenario #2 with three outliers removed is included as Attachment B-3. The R^2 correlation coefficient generated from this data set is low (R^2=0.27) indicating very poor correlation and the line appears skewed to one side of the data points. Removal of the highest congener point provides a much better correlation (R^2=0.73) and appears to define a linear relationship more consistent with the majority of the data points (see Attachment B-2). Note that removal of this high point is consistent with the premise that the error associated with field and laboratory variability at higher concentrations is more important to defining the slope than at lower concentrations. Removing this point causes one more point to fall as an outlier, resulting in the same four outliers as identified in Scenario #1, effectively causing the data sets for both Scenarios to be the same, except for those points where the homologue to congener ratio is less than 1. Removing this single data point (congeners = 134 ppm) has the added benefit of focusing the line on homologue concentrations (and therefore total PCB concentrations) less than 50 ppm, ensuring that the data set is most appropriate in the range of project action levels (1, 10, 25, and 50 ppm). This point was removed for subsequent outlier tests using both the y-intercept and zero-intercept models.

As shown in Table 2-1, the correlation coefficients for each of the regression lines indicate a reasonable linear relationship with similar slopes. This indicates that each line is a generally good approximation of the actual relationship between the congeners and homologues, and supports the premise that the actual relationship and the error associated with it will narrow somewhat as the number of data points included in the analysis is increased. The regression equations with zero as a y-intercept have a significant practical advantage over those with the positive y-intercept in that they can be used to calculate total PCB concentrations over the entire applicable range.

Both regression equations with a positive y-intercept share a common feature, in that the y-intercept is greater than 3 ppm. Accordingly, these relationships cannot be used to generate total PCB concentrations less than 1 ppm, the lowest cleanup goal applicable to residential direct contact with the sediment. Because neither line with a y-intercept is practically useful for the 1 ppm cleanup goal This necessitates a choice of using a regression equation with a zero-intercept or a separate equation applicable to the lower concentration range.

The equation generated using regression through the origin appears to be as technically and statistically valid as the second approach, and provides the most straightforward approach relative to future implementation and monitoring. To verify the applicability of each approach and to evaluate the significance to the project of selecting the zero-intercept approach, the lower concentration data near and below 1 ppm total PCBs were evaluated in more detail. This evaluation and a comparison of the different regression equations are discussed in the following sections.

2.2.1 Regression Analysis for Low Concentrations (Phase I)

The non-zero y-intercept regression models discussed above generate y-intercepts greater than 3 ppm and are, therefore, unsuitable for determining concentrations near the low 1 ppm action level. Because neither line with a positive y-intercept is practically useful for this application, a regression relationship with a zero intercept or a separate equation fit to the lower concentration range was indicated to be needed. To evaluate the significance of using the different equations and to determine whether constraining the line to pass through the origin is appropriate, the regression relationship for the lower concentration data was separately evaluated.

The lower concentration ranges for each of the two Scenarios were plotted and regression analysis performed to evaluate the best fit line for this lower concentration range. The data sets chosen for these evaluations were selected somewhat subjectively by removing the higher concentration data points until three criteria were met:

1) the resulting regression equation produced results up to and including the 1 ppm homologue concentration (equivalent to total PCBs);

2) the y-intercept was generally consistent with the low concentration values (i.e., markedly less than 1 ppm); and

3) the correlation coefficient indicated a reasonably robust relationship. Outliers were tested as described above and a single outlier removed for both Scenarios. The resulting data sets included 24 and 27 comparisons for Scenario #1 and Scenario #2, respectively.

The results from the regression analyses are included as Attachments A-5 and B-6. The final data sets are included in Attachment C and include the range from the lowest concentrations reported for both analyses up to approximately 1.5 ppm for congeners and 5 ppm for homologues. The initial regression line is shown in Attachment A-6 and identifies one outlier. Attachments A-5 and B-6 show the regression plotted without the outlier. The resulting equations for these lower concentration data are presented in Table 2-2.

Table 2-2
Regression Equations for the Low Concentration Data Range (Phase I)

Scenario (Database Adjustments)	Number of Data Points in Final Phase I Comparison (n)	Line Constrained to Pass Through Origin Slope (95% CI)	Line with a y-intercept	
			Slope (95% CI)	y-intercept (95% CI)
Scenario #1 low concentrations (1 outlier removed)	24	3.6 (3.2 – 4.1) $\Delta = 0.9$ $R^2 = 0.87$	3.7 (3.1 – 4.3) $\Delta = 1.2$ $R^2 = 0.87$	-0.04 (-0.3 - 0.2) $\Delta = 0.5$
Scenario #2 low concentrations (1 outlier removed)	27	3.6 (3.2 – 4.1) $\Delta = 0.9$ $R^2 = 0.88$	3.7 (3.2 – 4.3) $\Delta = 1.1$ $R^2 = 0.88$	-0.07 (-0.3 – 0.2) $\Delta = 0.5$

CI = Confidence Interval
Δ = Range of the 95% Confidence Interval
R^2 = Correlation Coefficient

As shown in Table 2-2, the slopes (m) generated using these two forms of the relationship are similar and the y-intercepts are close to zero. Based on the similarity of these equations, it is appropriate to simplify further evaluation and select the zero y-intercept model (m = 3.6) as the most appropriate equation for further comparison with the zero y-intercept equations determined for the entire data range (see Table 2-1).

2.2.2 Comparison of Equations for the Lower Concentration Range (Phase I)

****Note: The work discussed in this section was conducted to verify that the practical application of the single zero y-intercept equation (see Section 2.2.1) provided project decisions similar to those generated from using the other applicable equations. The conclusion to use the single zero y-intercept equation and the related expectation that addition of the Phase II data set would support this recommendation is supported by the Phase II results reported in Section 2.3. The work in this section is included to document the evaluation process conducted for Phase I with this note that, to some extent, it has been updated by the conclusions based on the Phase II results.**

To verify the technical validity of each regression approach and to evaluate the significance to the project of selecting the simpler relationship to implement (zero-intercept equation), the data were further evaluated for agreement and bias and to assess the impact to reported data relative to project action levels. Agreement and bias were assessed by taking the same data set (including outliers and those points excluded in Scenario #1) and using the percent difference metric to compare calculated total PCB values with the homologue concentrations. Percent difference is a useful and common analytical tool in evaluating the agreement between two results. Smaller percent differences indicate better agreement, percent differences greater than 100 % were considered poor agreement. Percent difference also provides a measure of bias, the positive differences indicating that the calculated value is the greater of the two, negative percent difference indicating that the homologue concentration is the greater of the two.

The effects of using the full concentration range regression lines forced through the origin (slope = 2.5 or 2.4) versus using a second regression equation (slope = 3.6) applicable only for the lower end of the concentration range were considered by applying these equations to the NOAA congener data for this Phase I data set and comparing the results to the homologue data. The comparisons and the results are discussed below.

The three regression equations applicable to the 1 ppm action level (slopes of 2.4, 2.5, and 3.6 with zero y-intercepts) were used for the Phase I data set to calculate total PCBs from the NOAA congeners and the resulting product(s) compared with the total homologue results using the percent difference (%D) equation. The number of comparisons for each equation and percent difference category are illustrated in Figure D-1 and summarized in Table 2-3. For the purposes of this evaluation, percent differences (%D) within ±100 % were considered reasonable agreement. Accordingly, the number of comparisons presented in Table 2-3 is cumulative for each category less than 100% (i.e., the number of comparisons <50%D also are included in the <75%D category, and these also are included in the <100%D category). The percent difference equation was used to evaluate the same data set (28 comparisons, including outliers) that was used to generate the slope of 3.6 as discussed in Section 2.2 above.

As shown in Table 2-3, the slopes of 2.5 and 2.4 provide total PCBs that agree slightly better with total homologues than the slope of 3.6 with 14 to 18 percent of the 28 total comparisons outside of 100 %D for the 2.4 and 2.5 slope, respectively, versus 25 percent (7 out of 28) outside of 100% D for the 3.6 slope. Of course, the slope of 3.6 is the largest and therefore, the most conservative measure of concentration.

Table 2-3
Percent Difference Categories for Different Slopes (m) (Phase I)

% Difference (absolute value) Category	Number of Comparisons (total = 28)		
	m = 3.6	m = 2.5	m = 2.4
<50%	15	13	13
<75%	18	19	18
<100%	21	23	24
>100%	7	5	4

Number of comparisons is cumulative for %Ds less than 100

To provide a measure of the potential bias associated with the different regression relationships, the proportion of positive and negative percent differences was evaluated and is presented in Table 2-4. A positive %D indicates the total PCB concentration calculated using the regression relationship was greater than the measured total homologue concentration and a negative %D indicates that the measured homologue concentration was greater.

Table 2-4
Summary of Bias (Measured as %D) Using a Different Slope (m) for Low Concentrations

%D Category	Number of Comparisons		
	m = 3.6	m = 2.5	m = 2.4
Positive %D (calculated total PCB concentration greater than total homologues)	17	13	12
Negative %D (total homologue concentration greater than the calculated total PCB concentration)	11	15	16

As shown in Table 2-4, the slope of 3.6 is a conservative measure of total PCB concentration with the highest percentage of calculated total PCB values greater than the measured homologue concentrations. The slope of 2.5 provides a more equally balanced data set with a slight bias toward the negative percent difference. The 2.4 slope appears to have a negative bias (highest potential for false negatives). Accordingly, the 2.5 slope was chosen as preferable to the 2.4 slope for further evaluation of the low concentration data.

To further evaluate the significance of using a slope of 3.6 rather than 2.5 to define the correlation relationship for the low concentrations, the potential impacts to project decisions were reviewed by comparing the different calculated total PCB values relative to the project action level of 1 ppm. In the data set reviewed (28 comparisons), this affects the decision relative to the 1 ppm action level (i.e., the decision for cleanup differs depending on the calculation) for three samples as shown in Table 2-5. Note that the results included in Table 2-5 and discussed below were identified based only on concentration, not based on the sample-specific cleanup goal (i.e., the cleanup goals for the samples in Table 2-5 may not be 1 ppm).

Note that the calculated total PCBs disagree with the total homologue group concentration relative to the 1 ppm action level for two additional points (homologues > 1 ppm, calculated value < 1 ppm) regardless of which equation is used (see Table D-1).

Table 2-5
Summary of Calculated PCB Results that Disagree Relative to 1 ppm (Phase I)

Sample ID	Homologue conc. (ppm)	Calculated ppm (%D) Using: slope = 2.5		slope = 3.6	
0010-2	1.0	0.71	(-36 %D)	1.0	(0 %D)
0002-2	0.76	0.88	(14 %D)	1.3	(49 %D)
0161-1	7.9	0.94	(-157 %D)	1.4	(-142 %D)

Based on this evaluation of the low concentration data, both the 2.5 and the 3.6 slopes appear to generate reasonable agreement with the reported homologue concentrations. The slope of 2.5 generates slightly better agreement, having fewer percent differences greater than 100% and slightly less overall bias measured as a similar distribution of positive and negative percent differences, with a slight negative bias.

Because it is a larger number, the equation with a slope of 3.6 is, by definition, a more conservative equation for calculating total PCBs. This is manifested in a larger percentage of positive percent differences and has the potential to alter project decisions relative to the 1 ppm cleanup goal as exemplified in Table 2-5 above. Note that the data in Table 2-5 illustrate that this level of conservatism may generate total PCB concentrations higher than those reported for homologues (i.e., has the greatest potential for false positives).

2.2.3 Evaluation of Regression Analysis for Higher Concentrations (Phase I)

****Note: The work discussed in this section was conducted to verify that the practical application of the zero y-intercept equation provided project decisions similar to those generated from using the other applicable equations. The conclusion to use the single zero y-intercept equation and the related expectation that addition of the Phase II data set would support this recommendation is supported by the Phase II results reported in Section 2.3. The work in this section is included to document the evaluation process conducted for Phase I with this note that, to some extent, it has been updated by the conclusions based on the Phase II results.**

The impact of using the different regression equations for total PCB concentrations greater than 1 ppm was evaluated in the same manner as used above for the lower concentrations. This included comparing the Phase I data relative to project action levels and by calculating percent difference for the calculated values and total homologues. Total PCBs were calculated using three equations, two equations (one from each Scenario) with a positive y-intercept and the zero-intercept equation with the slope of 2.5 (selected over the 2.4 slope because of less negative bias for the lower concentrations, as discussed above). The data range is shown in Table D-2 and was selected based on the initial Phase I data set minus those comparisons included for the near 1 ppm range in Section 2.2.1 above. The resulting range extends from above 1.5 ppm congeners (approximately 5 ppm total PCBs) to greater than 50 ppm homologues (total PCBs).

The total PCBs values calculated using the different equations vary slightly. The equation 2.3*NOAA+4 calculates the most conservative (highest) values for total PCBs over the entire concentration range of interest (total PCBs = 10 to 50 ppm). The zero y-intercept equation of 2.5*NOAA calculates the lowest concentration for total PCB concentrations less than 30 ppm. At concentrations above 30 ppm, the zero y-intercept equation with a slope of 2.5 generates a higher value than 2.2*NOAA + 3.3 and lower than 2.3*NOAA + 4 (see Table D-2). Each equation provides the same decision (i.e., greater than or less than) relative to the higher project action levels (10, 25, and 50 ppm) for this data set with the exception of the two comparisons shown in Table 2-6.

Table 2-6
Summary of Calculated Values That Differ Relative to Decision Criteria (Phase I)

Sample ID	Homologue (ppm)	Calculated Total PCB Concentration (ppm) Using:		
		2.5* NOAA	2.3* NOAA+4.0	2.2* NOAA+3.3
0093-1	4.4	9.9	13	12
0159-2	59	50	50	47

Note that an additional fourteen comparisons in this data range disagree with the homologue concentration relative to the project action levels of 10, 25, and 50 ppm regardless of the equation used to calculate total PCBs. This indicates that there is significantly more variability in the analytical data (including field preparation) than in the regression equations used to calculate total PCBs.

Data were also evaluated for agreement using percent difference in the same manner as for the lower concentrations discussed above. For this comparison, the regression equation with a zero y-intercept (2.5* NOAA) was compared with the two positive y-intercept equations. The results of the percent difference evaluation are shown in Attachment D-2 and summarized in Table 2-7.

Table 2-7
Summary of Percent Difference Distribution for Higher Concentrations (Phase I)

%D (absolute value) Category	Number of Comparisons		
	2.5* NOAA	2.3* NOAA +4.0	2.2* NOAA +3.3
<50	21	21	21
<75	23	26	25
<100	25	29	29
>100	11	7	7

Number of Comparisons <100 %D is Cumulative

Both equations with positive y-intercepts give slightly better agreement using percent difference as a measure of agreement. However, as discussed above and summarized in Table 2-6 the magnitude of this difference has relatively little practical impact on project decisions that would be made based on the calculated total PCB results.

Percent difference results also were evaluated to provide an indication of bias. A positive percent difference indicated the calculated value was greater than the measured homologue concentration and a negative percent difference indicated the homologue group result was the greater of the two. The results are summarized in Table 2-8.

Each of the three equations shows some negative bias. The equation 2.3*NOAA+4 provides the least amount of overall bias despite the fact that this equation calculates the highest total PCB values of the three. The slope of 2.5 has the most bias, the calculated results being less than the total homologue concentrations. The equation 2.2*NOAA+3.3 also has a negative bias.

Table 2-8
Summary of %D for Higher Concentrations (Phase I)

%D Category	Number of Comparisons Calculated Using:		
	2.5*NOAA	2.3*NOAA +4	2.2*NOAA +3.3
Positive %D (calculated total PCB concentration greater than total homologues)	13	17	15
Negative %D (total homologue concentration greater than the calculated total PCB concentration)	23	19	21

Of the three equations used to evaluate this data set, the equation 2.3*NOAA+4 provides the best overall agreement with the least amount of bias for this concentration range. As shown in Table 2-6 above, the slightly better agreement for these higher level concentrations obtained from using an equation with a non-zero y-intercept is not likely to have a significant impact on actual project decisions. Furthermore, overall project remediation technical needs do not warrant the added level of conservatism and poorer agreement for the lower concentrations (in the 1 ppm total PCB range) that a second equation would necessitate.

2.2.4 Phase I Correlation Summary

Based on the results from the Phase I data set and with concurrence from the USACE and EPA, it was decided to use the zero y-intercept relationship with a slope of 2.5 as the equation for calculating total PCBs until the data from Phase II were available for evaluation. This equation was chosen primarily for the following reasons:

- The zero y-intercept equation is as appropriate as the non-zero y-intercept regression model given the premise that the NOAA congeners are a representative subset of the total PCBs in New Bedford Harbor sediments and that the relationship between the two is linear.

- The single zero y-intercept equation would be significantly simpler to implement than the "if/then" statement required for the non-zero y-intercept approaches. A single equation reduces the possibility of errors associated with documenting, using, and updating multiple equations. This choice also eliminates the somewhat subjective decision of exactly which concentration range was applicable to the lower concentration equation.

- Of the two zero-intercept equations, the equation with a slope of 2.5 (Scenario #1) provided better agreement and less bias than the Scenario #2 equation (slope = 2.4) for total PCB concentrations near the 1 ppm action level.

- The use of two separate equations depending on concentrations did not appear to provide significantly better agreement and predictability for the overall data set than the single zero y-intercept equation.

The decision-making implications of using a single equation with a y-intercept of zero versus using two different equations depending on concentration were evaluated. The NOAA congeners from the Phase I data set were used to calculate total PCBs using the different equations and the results were compared to the measured total homologue group concentrations. The two calculated and one measured total PCB concentrations were then compared to the project action levels (1, 10, 25, and 50 ppm). In 73 out of 95

cases, both of the calculated total PCB concentration and the measured sum of the homologue group concentrations were the same with respect to their relationship to the nearest project action levels (e.g., all three values were between 10 and 25 ppm) implying that the same project decision would be indicated using either regression approach. In 17 of the 95 cases, both regression-based calculated concentrations were similar relative to the project action levels, but both differed from the measured total homologue concentration (e.g., both regression estimates were greater than 1 ppm while the measured concentration was less than 1 ppm). In only 5 of the 95 cases, the results from one of the regression relationships was similar to the measured concentration relative to project action levels while the other regression relationship led to a different project decision. Based on these results, it was concluded that using the two different equations depending on concentration did not produce significantly different conclusions from the single zero-intercept equation.

In evaluating the different regression equations, the resulting effect on total PCB concentrations, and the associated project decisions, the zero y-intercept regression equation and the two different equations depending on concentration appear to provide the project with a similar level of predictive power. Based on this conclusion, it was recommended that the zero y-intercept equation with a slope of 2.5 be used as the simplest to implement, monitor, document, and update as more data are collected. The results from the combined Phase I and II data set further support this conclusion and are discussed in the following section.

2.3 Regression Results for Phase I and Phase II Data Set

Based on the regression results from the Phase I data set, a single correlation equation with a slope of 2.5 and a zero y-intercept was recommended as technically defensible for the concentration range of interest and most practical for use for on-going work. It was recommended that the regression analysis be performed again on a larger data set as more data became available. Since submittal of the draft memorandum in January 2001, Phase II data are now available for inclusion in the regression analysis. These data were incorporated into the data set and the results discussed in this section.

For this evaluation, the Phase I and Phase II data sets were combined and the data sets applicable to each Scenario chosen in the same manner as for the Phase I data and discussed in Section 2.1 above. That is, non-detect results and those where both total concentrations were greater than 50 ppm were removed from the data set as not applicable to the concentration range of interest. The data sets were further divided using the two Scenarios discussed in Section 2.1, where Scenario #1 excludes those points where the total homologues are less than the total congeners (ratio of homologues-to-congeners <1) and Scenario #2 includes these as valid points. Least squares linear regression equations were plotted with positive y-intercepts and with y-intercepts forced through zero. Points with absolute value standardized residuals greater than 4 were removed as outliers. Results for Scenario # 1 are included in Attachment E and those for Scenario # 2 are included in Attachment F. The resulting equations are illustrated on Figures 2-3 and 2-4 and summarized in Table 2-9.

Table 2-9
Regression Equations for the Combined Phase I and Phase II Data Sets

Scenario (Database Adjustments)	Zero y-intercept equation (R^2)	Positive y-intercept equation (R^2)
Scenario #1 (Homologue to congener ratios <1 not included)	2.5*NOAA ($R^2 = 0.84$)	2.4*NOAA+2.4 ($R^2 = 0.85$)
Scenario #2 (Homologue to congener ratios <1 included)	2.5*NOAA ($R^2 = 0.84$)	2.4*NOAA+2.0 ($R^2 = 0.84$)

The regression equations generated using this larger data set are in closer agreement with each other than those previously generated using the smaller Phase I data set. The zero y-intercept equation for both Scenarios has a slope of 2.5, consistent with that chosen based on the Phase I results. The two equations with positive y-intercepts are also similar to each other and have y-intercepts smaller than for the more limited Phase I data set. However, as for the Phase I data set, the positive y-intercepts are greater than 1 ppm and are therefore not practically useful for those samples where concentrations relative to 1 ppm are of importance.

2.3.1 Phase I and Phase II Low Concentration Data Correlation

The lower concentration range data for the Phase I and Phase II data set were evaluated by removing higher concentration data points at the same point chosen for the Phase I data set (sum of NOAA congeners = 1.4 ppm) and sequentially testing and removing outliers. As noted in Section 2.2.1, the selection of NOAA=1.4 ppm as the upper end of this range was determined somewhat subjectively by removing the higher concentration data points until three criteria were met:

1) the resulting regression equation produced results up to and including the 1 ppm homologue concentration (equivalent to total PCBs);

2) the y-intercept was generally consistent with the low concentration values (i.e., markedly less than 1 ppm); and

3) the correlation coefficient indicated a reasonably robust relationship. To maintain consistency and evaluate changes to the regression as a result of increasing the number of data points, the same point was selected as the upper concentration range for these subsequent regression calculations. Linear regression equations for each Scenario are included on Figures 2-5 and 2-6, respectively, and summarized in Table 2-10. Supporting data tables and the regression analyses are included in Attachments E and F, respectively.

Table 2-10
Regression Equations for Low Concentration Range (Phase I and Phase II Data)

Scenario (Database Adjustments) [n = number of comparisons]	Zero y-intercept equation (R^2)	y= intercept equation (R^2)
Scenario #1 (Homologue to congener ratios <1 not included) [n = 51]	2.9*NOAA (R^2 = 0.82)	3.0*NOAA+(-0.02) (R^2 = 0.82)
Scenario #2 (Homologue to congener ratios <1 included) [n = 58]	2.9*NOAA (R^2 = 0.82)	3.0*NOAA+(-0.06) (R^2 = 0.82)

The larger data set generates equations with a lower slope than was identified using the smaller Phase I data set (reference Table 2-2) and the slope is closer to the slope of 2.5 calculated for the entire data set with a zero y-intercept equation. Using the larger data set of both Phase I and Phase II data, the calculated relationship describing the lower concentrations appears to be approaching an equation similar to that going through zero for the entire data set. For these reasons, the conclusion to use the single zero y-intercept equation for the entire concentration range of interest appears to be valid and remains the recommendation for the project.

Figure 2-3

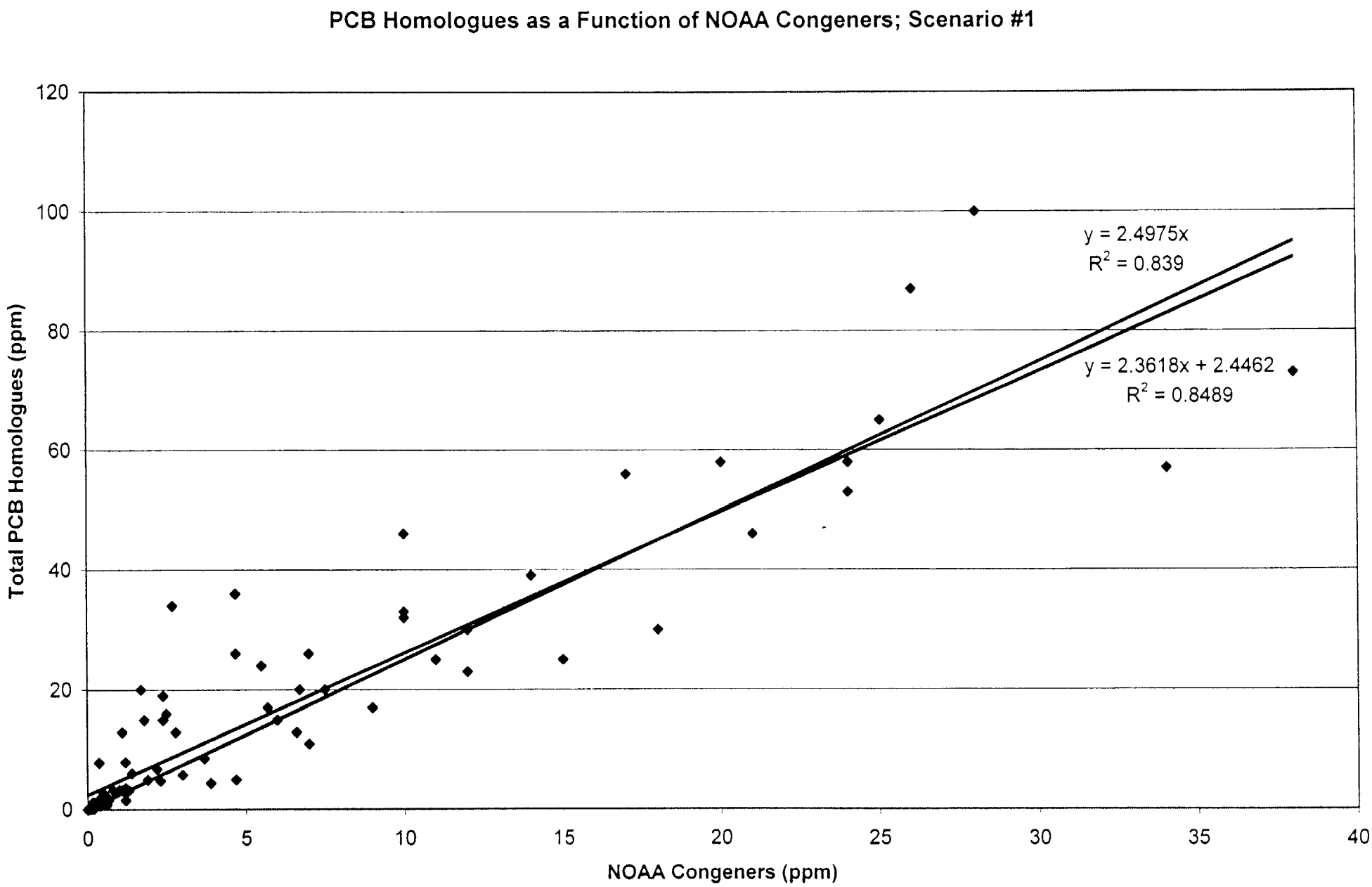
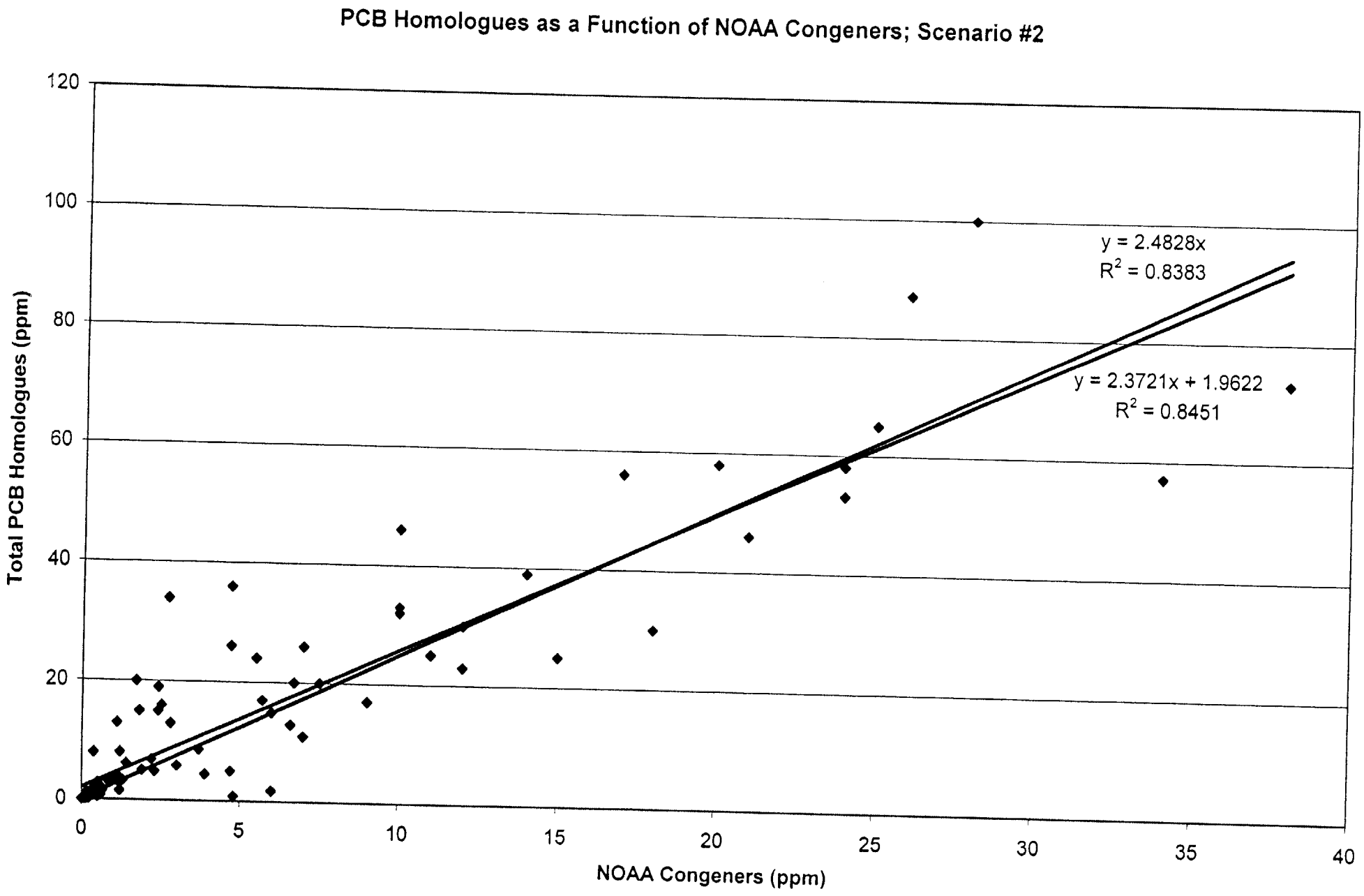

PCB Homologues as a Function of NOAA Congeners; Scenario #1
Total PCB Homologues (ppm)
NOAA Congeners (ppm)
0
20
40
60
80
100
120
0
5
10
15
20
25
30
35
40
y = 2.4975x
R2 = 0.839
y = 2.3618x + 2.4462
R2 = 0.8489

Figure 2-4



4/12/01

dra r F- al)

Figure 2-5

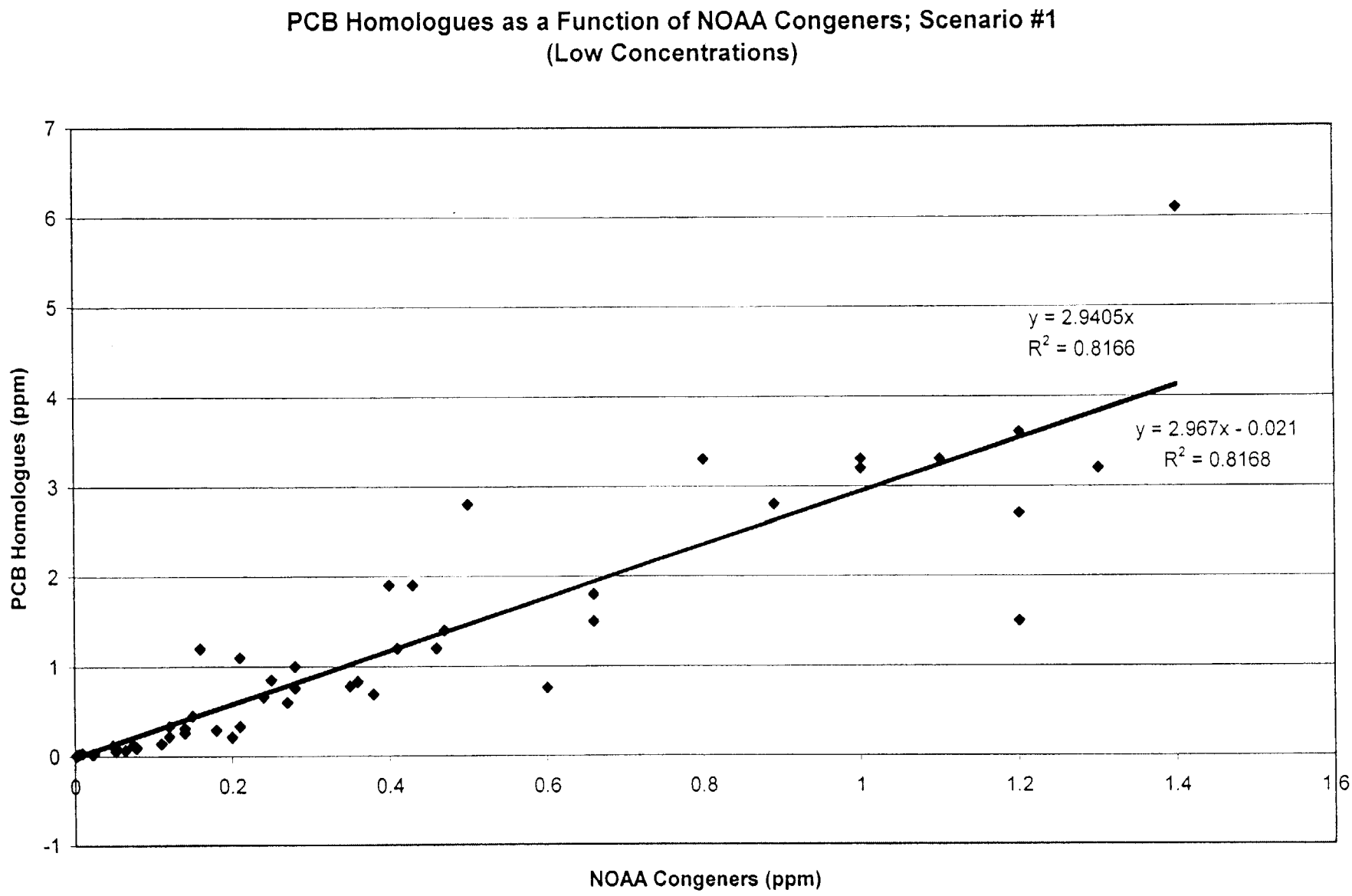
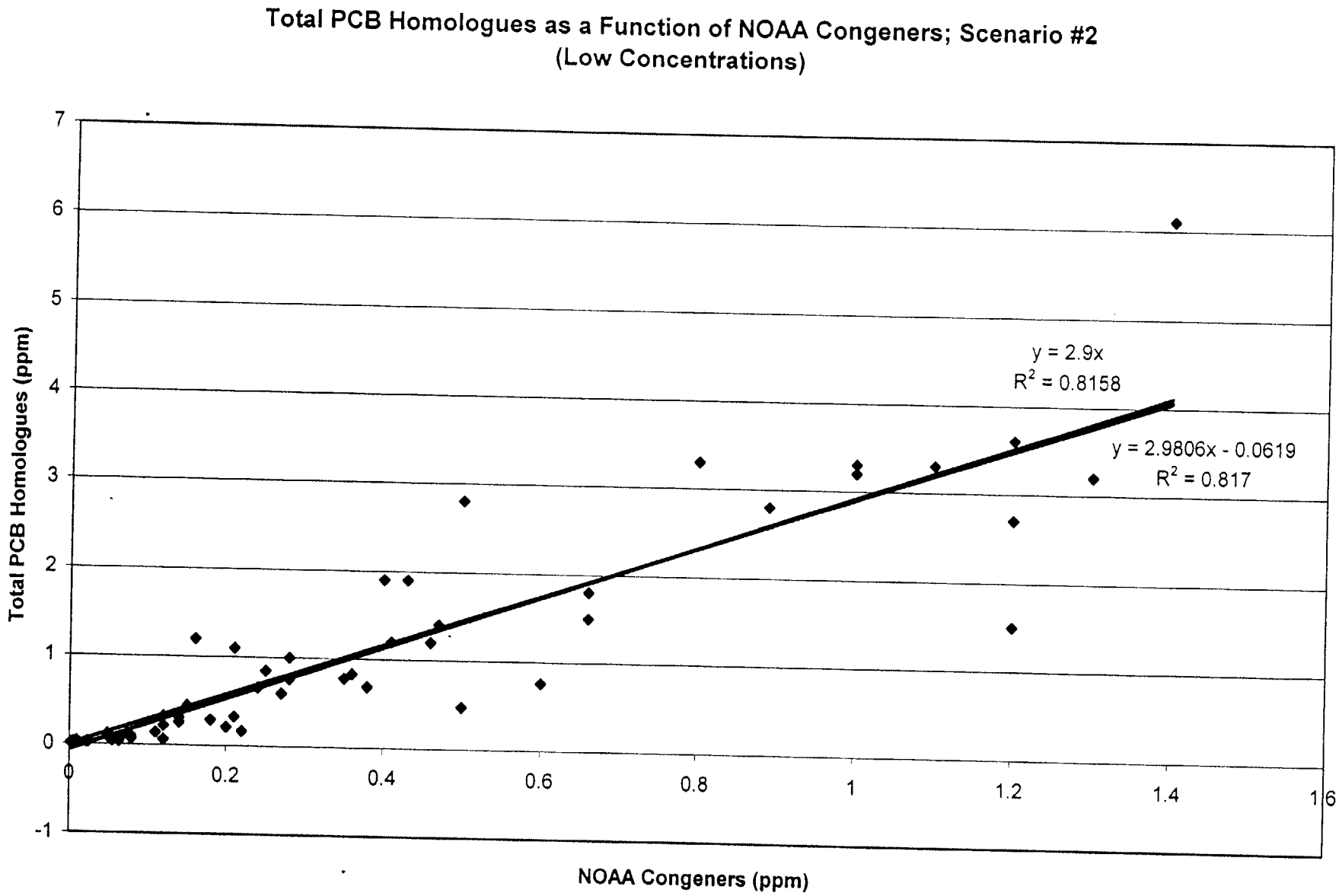

PCB Homologues as a Function of NOAA Congeners; Scenario #1
(Low Concentrations)
y = 2.9405x
R2 = 0.8166
y = 2.967x - 0.021
R2 = 0.8168
PCB Homologues (ppm)
NOAA Congeners (ppm)
7
6
5
4
3
2
1
0
-1
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6

Figure 2-6



4/12/01

A[illegible] 'raft F L[illegible] nal)

2.3.2 Summary of Regression Results for Phase I and II Data Sets

Following submittal of the Phase I correlation results with the recommendation to use a zero y-intercept equation with a slope of 2.5, Phase II data were added to provide a larger data set (161 comparisons). This larger data set was then used to calculate the four possible linear relationships described above for the Phase I data analysis. The regression results from this larger data set were similar to those calculated for the Phase I data. The zero y-intercept equations for both scenarios had slopes of 2.5, confirming this as the best fit zero y-intercept relationship.

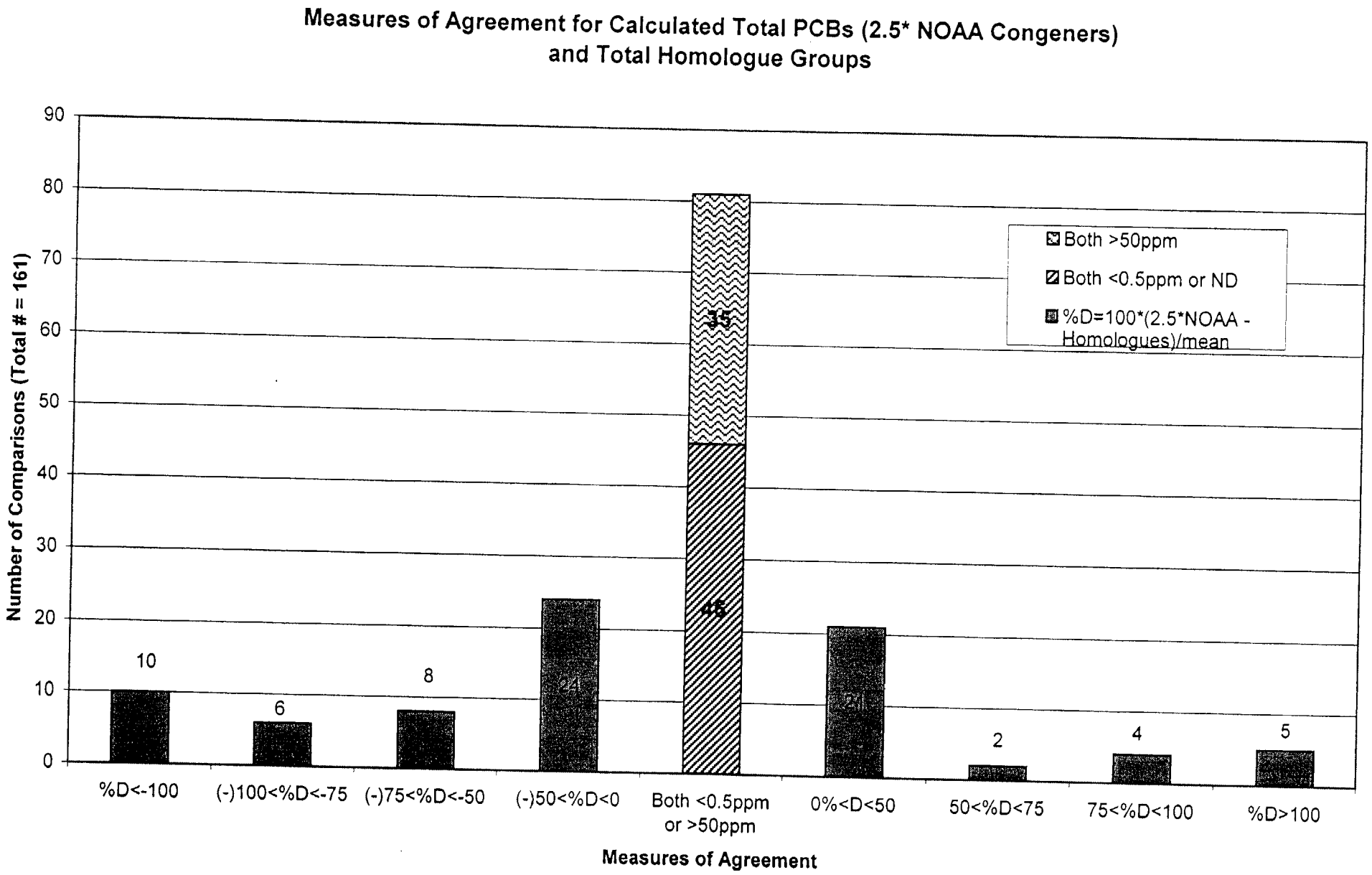
To further assess that the zero y-intercept approach was appropriate for defining the relationship, the positive y-intercept equation and separate equation for the lower concentrations also were identified and evaluated for the larger Phase I and II data set. The non-zero intercept equations shared similar limitations to those from the smaller Phase I data set, in that two equations would be necessary to calculate concentrations for the entire concentration range of interest and that the selection of applicable concentration ranges was somewhat subjective. The positive y-intercept and low concentration equations for the larger data set were different from those generated from the Phase I data set, amplifying the concern that adding new data is more likely to have a greater impact on the more complex regression relationships. Overall, the regression equations for the two concentration ranges are more similar to each other and the slopes and y-intercepts closer to the zero y-intercept model for the larger data set than for the limited Phase I data.

These results support the premise that a zero y-intercept model is an appropriate definition of this relationship and that the addition of more data points is likely to lead to greater convergence on a relationship closely approximated by the zero y-intercept model.

At a minimum, these results confirm the premise that a single zero y-intercept relationship to define the entire concentration range is less likely to change as more data become available than using two different equations depending on concentration. To verify this approach and to evaluate its applicability to the available data set, the equation with a slope of 2.5 and a zero y-intercept was used to calculate total PCBs for comparison with the measured total homologue data using the percent difference metric.

The results of this comparison are summarized in Figure 2-7. A total of 81 of 161 comparisons (50%) were above 50 ppm or below 0.5 ppm. A total of 55 comparisons (34%) had relatively good agreement with percent differences less than 75%. An additional 10 comparisons had reasonable agreement (%D <100%) and only 15 comparisons (9%) had poor agreement greater than 100% D. Overall, this appears to represent a good correlation and a strong basis for future comparisons. A similar evaluation approach is recommended in Section 3 below as a relatively quick way to assess agreement for subsequent data and to reduce the frequency of updating the correlation relationship developed in this memorandum.

Figure 2-7



4/12/01

entD nal

2.4 Comparison of NOAA Congener to Aroclor Data

The initial planning for this correlation study included an evaluation of the relationship between the NOAA congeners and total Aroclors. Accordingly, 115 split samples from the Phase I and Phase II sampling also were analyzed for PCB Aroclors in addition to congeners and homologue groups. As discussed in Section 1.3, the Aroclor and congener analyses were performed by the same laboratory from the same sample jar, possibly eliminating some of the variability introduced by the field splitting procedure.

To document the results of this comparison, the NOAA congeners and Aroclor data were compared in the same manner as the congener and total homologue group results discussed in Sections 2.2 and 2.3, above. The data were evaluated using the two data set Scenarios and excluding those points where both results totaled greater than 50 ppm and where one or both analytical results were non-detect. Outliers were similarly defined as those points with standard residuals greater than four and were sequentially removed. Tables and graphs illustrating the linear regression analyses and the outlier tests are included in Attachments G (Scenario #1) and H (Scenario #2). The resulting final regression lines are included as Figures 2-8 and 2-9, for Scenarios #1, and #2, respectively, and summarized in Table 2-12.

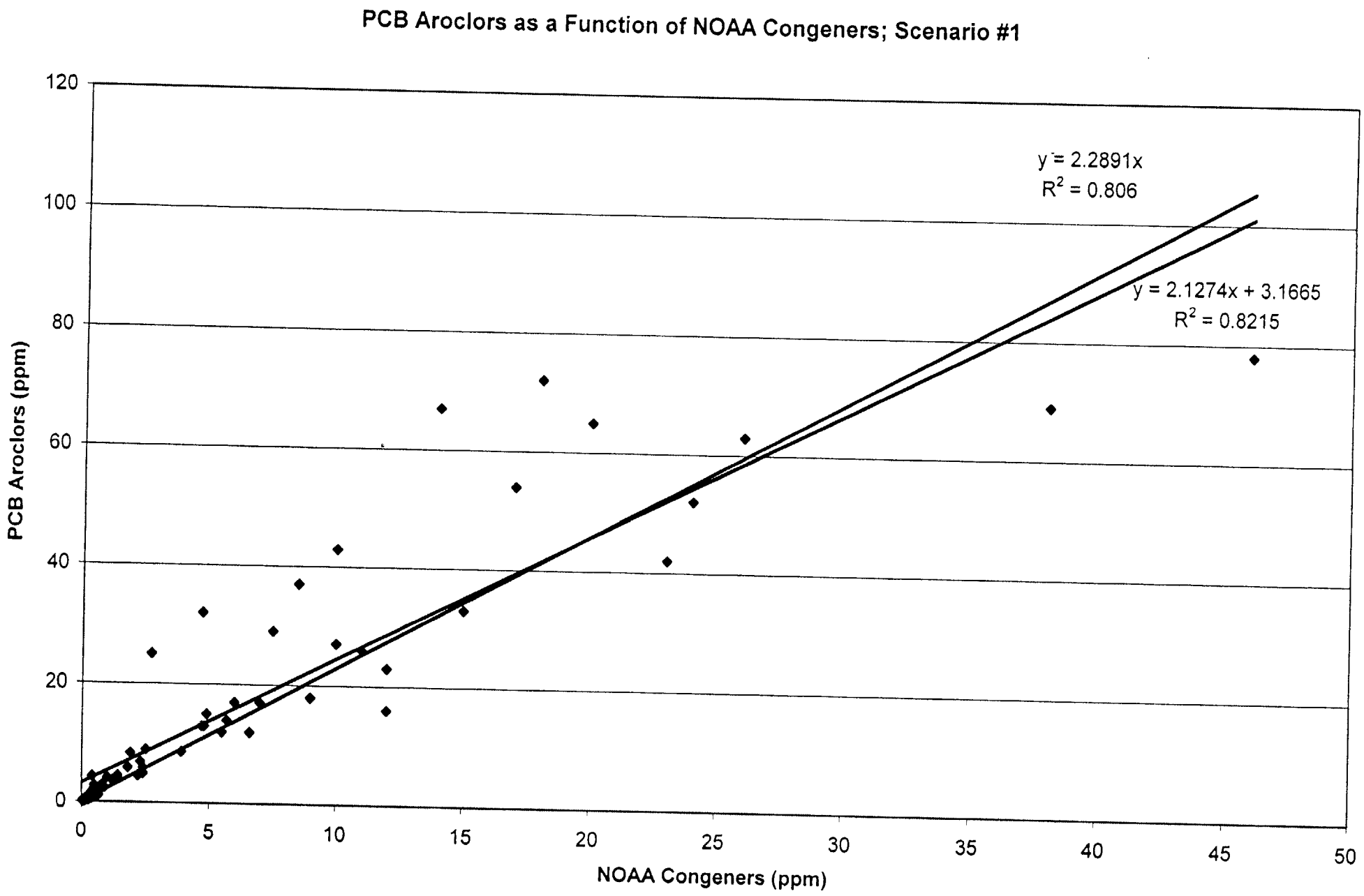
Table 2-11
Regression Equations for the Comparison of NOAA Congeners to Aroclors

Scenario (Database Adjustments)	Number of Data Points in Final Comparison (n)	Line Constrained to Pass Through Origin Slope (95% CI)	Line with a y-intercept Slope (95% CI)	y-intercept (95% CI)
Scenario #1 (Aroclor to congener ratios <1 not included)	68	2.3 (2.1 to 2.5) $\Delta = 0.40$ $R^2 = 0.81$	2.1 (1.9 to 2.4) $\Delta = 0.50$ $R^2 = 0.82$	3.2 (0.5 to 5.8) $\Delta = 5.3$
Scenario #2 (Aroclor to congener ratios <1 included)	75	2.1 (1.9 to 2.3) $\Delta = 0.40$ $R^2 = 0.75$	2.0 (1.7 to 2.2) $\Delta = 0.50$ $R^2 = 0.77$	3.1 (0.34 to 5.9) $\Delta = 5.6$

CI = Confidence Interval
Δ = Range of the 95% Confidence Interval
R^2 = Correlation Coefficient

As was determined for the congener-to-homologue group regression equations, the congener-to-Aroclor equations with the positive y-intercepts have y-intercepts greater than the lowest action level of 1 ppm. As was done for congener-to-homologue comparisons, the lower concentration data points were plotted separately to evaluate the difference in the relationship for the lower concentrations. For ease of reference, the lower concentration data set was chosen to cover the same range of independent variable (ND to sum of NOAA congeners = 1.4 ppm) range as selected in Section 2.2.1 above and illustrated in Attachments G-L-1 and H-L-1. As noted above, this selection was somewhat subjective and is presented here only to provide an indication of the difference in regression equations for the lower concentration range. The equations for the lower concentration data are presented in Table 2-13.

Figure 2-8



4/12/01

Figure 2-9

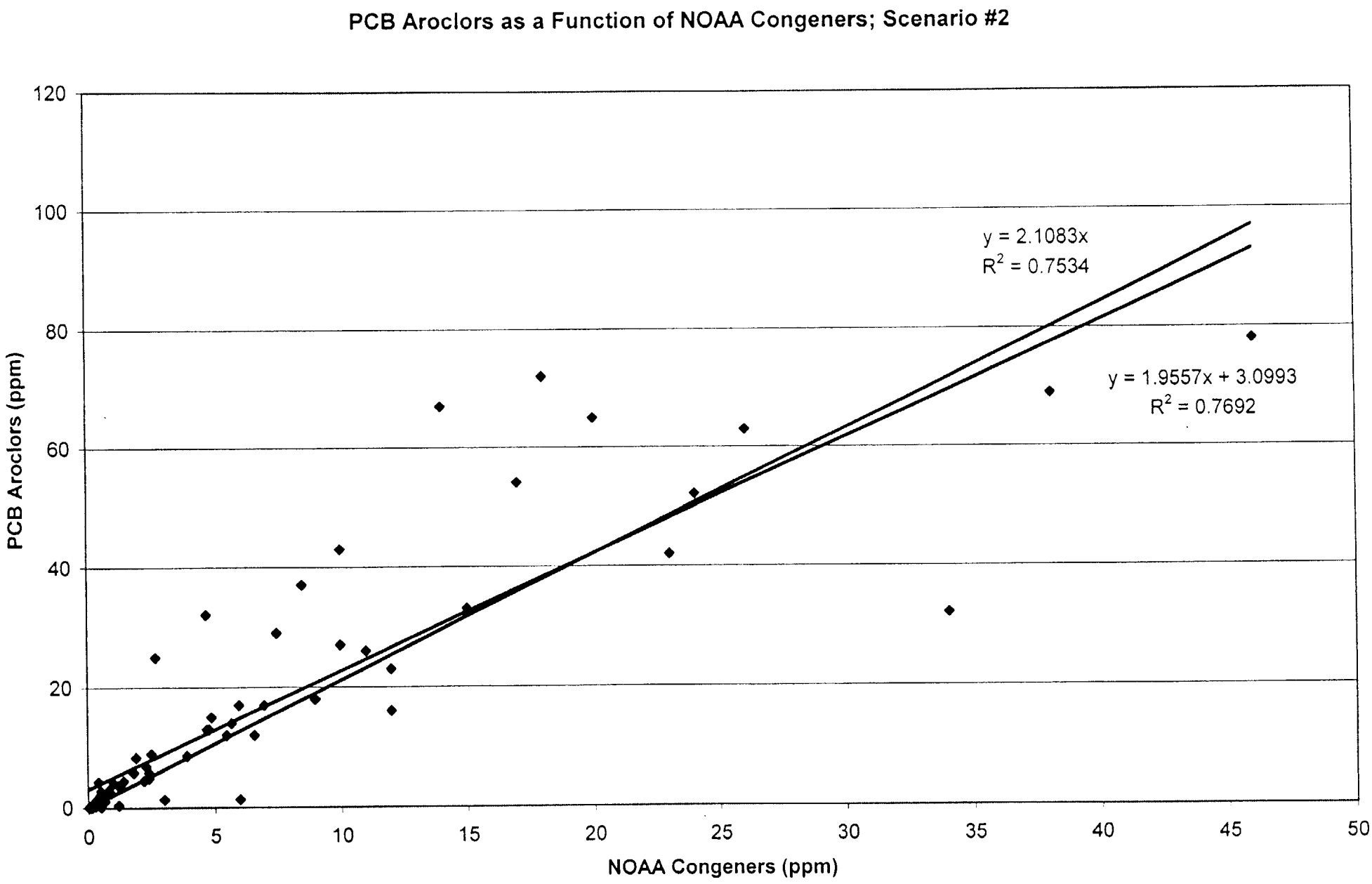

PCB Aroclors as a Function of NOAA Congeners; Scenario #2
y = 2.1083x
R2 = 0.7534
y = 1.9557x + 3.0993
R2 = 0.7692
PCB Aroclors (ppm)
NOAA Congeners (ppm)
0
20
40
60
80
100
120
0
5
10
15
20
25
30
35
40
45
50

Table 2-12
NOAA Congeners to Aroclor Regression Equations for the Low Concentration Range

Scenario (Database Adjustments)	Number of Data Points in Final Comparison (n)	Line Constrained to Pass Through Origin Slope (95% CI)	Line with a y-intercept	
			Slope (95% CI)	y-intercept (95% CI)
Scenario #1 (Aroclor to congener ratios <1 not included)	31	3.2 (2.8 – 3.5) $\Delta = 0.70$ $R^2 = 0.88$	3.4 (2.9 – 3.8) $\Delta = 0.90$ $R^2 = 0.88$	-0.14 (-0.39 – 0.10) $\Delta = 0.49$
Scenario #2 (Aroclor to congener ratios <1 included)	36	2.8 (2.3 – 3.4) $\Delta = 1.1$ $R^2 = 0.56$	2.8 (1.9 – 3.6) $\Delta = 1.7$ $R^2 = 0.56$	0.04 (-0.41 – 0.49) $\Delta = 0.9$

CI = Confidence Interval
Δ = Range of the 95% Confidence Interval
R^2 = Correlation Coefficient

The equations defining the relationships between the NOAA congeners and the Aroclors are similar to those relating the NOAA congeners to the sum of the homologue groups. The equations for the lower concentrations also are similar, with significantly lower correlation coefficients for Scenario #2. The lower correlation coefficients for Scenario #2 appear to be primarily the result of two points [(0.41, 4.3) and (1.2, 0.43)] that do not test as outliers although they appear to be inconsistent with the overall regression pattern characteristic of the data set (see Attachment H-L-1). The (1.2, 0.43) pair was removed from the Scenario #1 data set because the Aroclor concentration was less than the total congener concentration. With this data point excluded, the (0.41, 4.3) pair tests as an outlier and the resulting regression equations have lower slopes and significantly better correlation coefficients than Scenario #2. Because the Project Team has determined that the NOAA congener/homologue group relationship will be used to define total PCBs, no further evaluation of the NOAA congener-to-Aroclor relationship was performed.

2.5 Comparison of Aroclor to Homologue Group Results

In addition to the NOAA congener to Aroclor comparison discussed in Section 2.4, the Aroclor data were used for comparison to the sum of the total homologue groups. A total of 115 split samples were available for this comparison. Of these, the data set chosen for this comparison was selected by excluding those points where one or both results were non-detect and/or where both results were above the concentration range of interest (greater than 50 ppm). The data sets and the graphs and linear regression results are included in Attachment I. Outliers were defined as those points where the standard residuals were greater than 4 and were sequentially removed as shown on the Attachment I cover sheet. The resulting regression equations are shown on Figure 2-10 and summarized in Table 2-14.

Figure 2-10

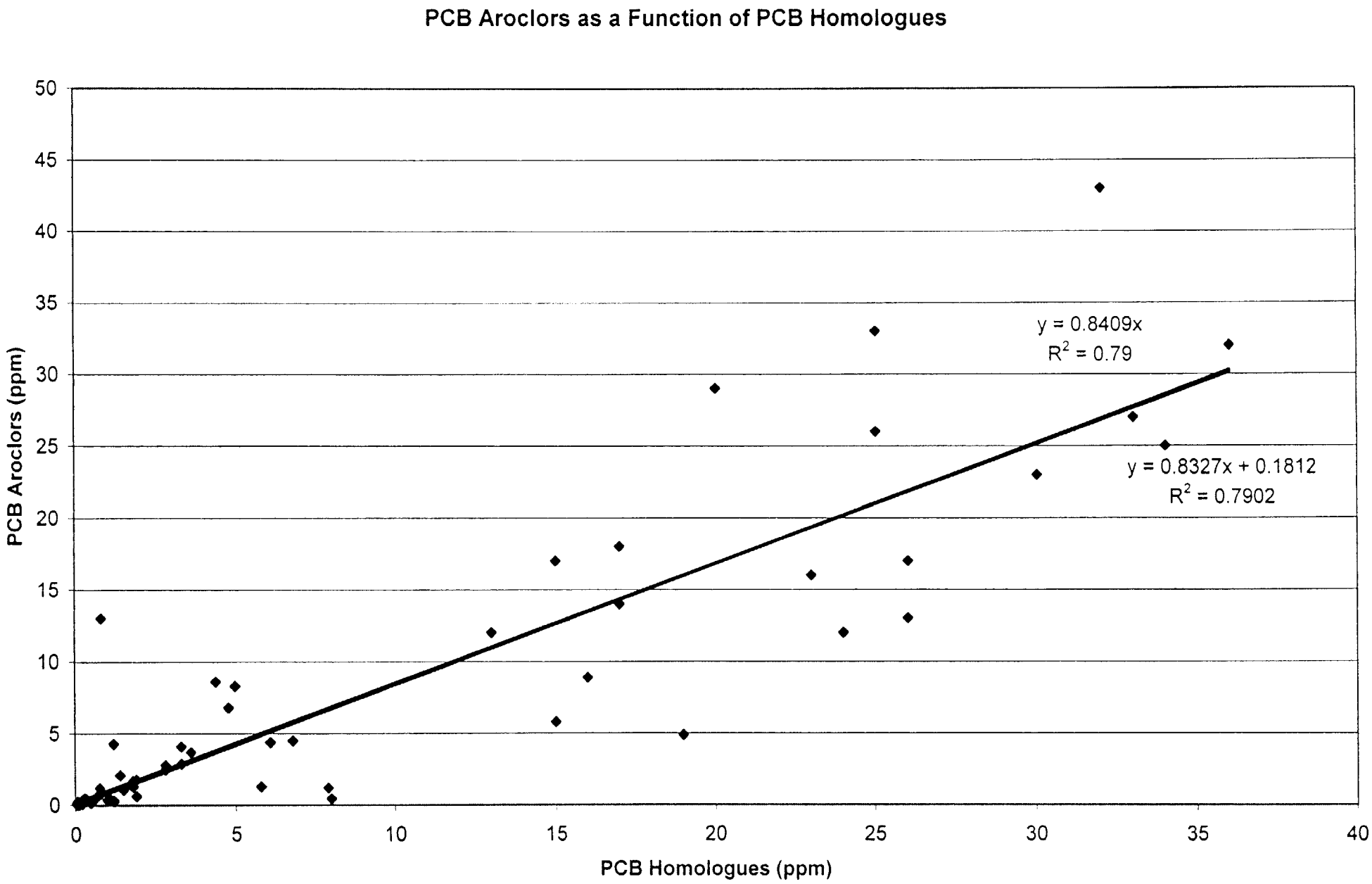

PCB Aroclors as a Function of PCB Homologues
PCB Aroclors (ppm)
PCB Homologues (ppm)
0
5
10
15
20
25
30
35
40
45
50
y = 0.8409x
R2 = 0.79
y = 0.8327x + 0.1812
R2 = 0.7902

Table 2-13
Regression Equations for the Comparison of Aroclors-to-Homologue Groups

Number of Data Points in Final Comparison (n)	Line Constrained to Pass Through Origin Slope (95% CI)	Line with a y-intercept	
		Slope (95% CI)	y-intercept (95% CI)
64	0.84 (0.76 – 0.93) $\Delta = 0.17$ $R^2 = 0.79$	0.83 (0.72 – 0.94) $\Delta = 0.22$ $R^2 = 0.79$	0.18 (-1.3 – 1.7) $\Delta = 3.0$

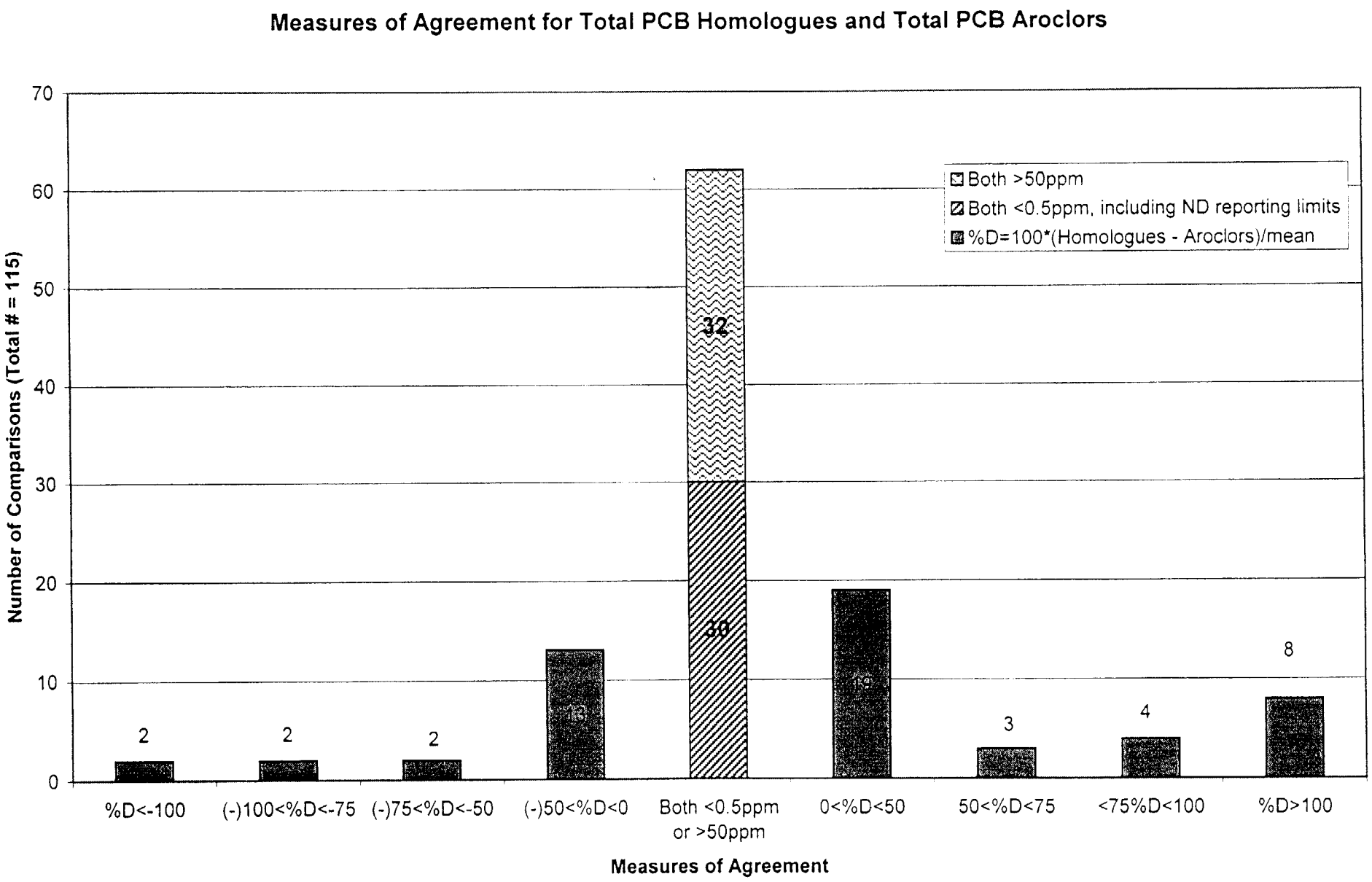
CI = Confidence Interval
Δ = Range of the 95% Confidence Interval
R^2 = Correlation Coefficient

As shown by the correlation coefficients, this relationship shows generally good agreement. The slopes of the equations and the 95% confidence intervals illustrate a slight bias toward the homologue groups tending to be slightly greater than the Aroclors. The two equations are similar and the equation with a positive y-intercept has an intercept reasonably close to zero relative to the lowest action level of 1 ppm. Both the y-intercept and zero-intercept equations appear to be appropriate for determining total PCBs over the entire concentration range, including those points near 1 ppm.

For reference to the relationship between the NOAA congeners and the sum of the homologue groups, the Aroclors and homologue groups were compared using the same agreement criteria (i.e., agreement is defined as both results are less than 0.5 ppm including ND values, or both are greater than 50 ppm or the percent difference is less than 100%). The results are illustrated on Figure 2-11. Data and calculations are included in Attachment D. Of the 115 comparisons, 54 percent (62 out of 115) were in agreement relative to low and high concentrations (<0.5 ppm or >50 ppm). An additional 30% (37 out of 115) had percent differences (absolute value) less than 75%, indicating good agreement, five percent (6 comparisons) had reasonable agreement (percent differences greater than 75% and less than 100%. Ten out of 115 (9%) comparisons had poor agreement, measured as percent differences greater than 100%. The majority (64%) of the percent differences are positive, further confirming that the homologue concentrations tend to be slightly more than the Aroclor concentrations.

As discussed above, the Project Team has decided to use NOAA congeners and the equation developed for the congeners to homologue group relationship as a measure for total PCBs. Accordingly, no further evaluation of the Aroclor data was performed.

Figure 2-11


Measures of Agreement for Total PCB Homologues and Total PCB Aroclors
Number of Comparisons (Total # = 115)
70
60
50
40
30
20
10
0
Both >50ppm
Both <0.5ppm, including ND reporting limits
%D=100*(Homologues - Aroclors)/mean
32
30
2
2
2
13
19
3
4
8
%D<-100
(-)100<%D<-75
(-)75<%D<-50
(-)50<%D<0
Both <0.5ppm or >50ppm
0<%D<50
50<%D<75
<75%D<100
%D>100
Measures of Agreement

3.0 RECOMMENDATIONS FOR FUTURE DATA EVALUATION

Several recommendations are proposed regarding future sediment PCB analytical data evaluation efforts for New Bedford Harbor. These recommendations are based on the results for the statistical evaluations discussed above, considered in conjunction with the practical needs associated with implementing the PCB sediment remediation effort. Recommendations are made with the intent to develop procedures that allow the project to use NOAA congener data for establishing PCB concentrations as rapidly as possible during future remediation activities. To achieve this purpose, systematic split sample comparisons using percent difference equations are recommended as measures of data quality that can be implemented in an expeditious and cost-effective manner.

The following approach is recommended for the next phases of investigation sediment sampling. Note that this approach was developed only for sediment samples collected from New Bedford Harbor and is therefore, only applicable to this matrix and this site. Other interpretations of this relationship or applications to other matrices or sites is inappropriate and could result in erroneous data and associated conclusions.

1. Use the linear regression equation with a zero y-intercept form and a slope of 2.5 (2.5*NOAA) to calculate total PCBs for comparison with project action levels.

2. In order to confirm that this equation is an accurate representation of total PCBs, continue to analyze approximately 7 to 8% of the samples for total homologue groups in addition to the 18 NOAA congeners. Continue to evaluate the agreement between the calculated total PCB value (using a slope of 2.5) and the measured homologue group analytical results using percent difference (%D). Recommended acceptance criteria for this data comparison include the following:

   - Both values are less than 0.5 ppm, including ND reporting limits

   - Both values are greater than 50 ppm

   - Where one or both of the concentrations are between 0.5 and 50 ppm, absolute value of %D <= 100%

   Less than 10% (approximately 9% of the available Phase I and II data set) of the data set is expected to exceed these criteria (see Figure 2-7). Results that do not meet these criteria are attributed primarily to the field variability. Provided the proportion of exceedences remains less than 10%, the data will be considered acceptable and the correlation equation with a slope of 2.5 considered the most applicable description of the relationship between the NOAA congeners and total PCBs. If a larger proportion of comparisons exceed this criteria or a one-way bias appears apparent, the data will be reviewed in more detail to determine possible causes and make recommendations for an alternative comparison approach, if appropriate.

   - As an additional measure of data quality, approximately 20 percent of the sediment samples will also be split for USACE analysis for NOAA congeners.

   - To verify the appropriateness of the correlation equation (2.5*NOAA), additional comparisons should be periodically added to the data set and the regression relationship recalculated using the same process as described in this memo. Comparisons will be added as practical for the project. For example, the next set of comparison results will be from the

Phase III investigation collected, analyzed and reported following the 2001 sampling season. Results will be discussed with USACE and EPA and summarized in a brief memo with detailed attachments (similar to Attachments E and F included with this memo). Assuming that the larger data set will produce similar correlation relationships, the frequency of updating the correlation results may be decreased based on consensus with USACE and EPA. If the results of subsequent regression analyses conclude that a different equation is more appropriate, recommendations will be made for future calculation of total PCBs. It is recommended any future equations be used only for subsequent sample data and not to revise data reported using the current equation. Similarly, any change to the current equation or approach will be discussed with and approved by USACE and EPA prior to implementation.

4.0 REFERENCES

Axys Analytical Services, Ltd. *SOPs and Sample Results for the Extraction and Analysis of PCB Homologue Groups Using Low Resolution Mass Spectrometry.*

Dixon, Wilfrid J. and Frank J. Massey, Jr. Introduction to Statistical Analysis, Third Edition. McGraw-Hill Book Company. 1951

ENSR Corporation. *Chemical Quality Assurance Reports, New Bedford Harbor Superfund Site, New Bedford, Massachusetts. Prepared for the USACE.*

EPA, 1994. *Guidance for the Planning for Data Collection in Support of Environmental Decision Making Using the Data Quality Objective Process,* September 1994, EPA/600-R-96/055 (EPA QA/G-4).

EPA, 1995. *Guidance for the Preparation of Standard Operating Procedures (SOPs) for Quality-Related Documents,* November 1995, EPA/600-R-96/027 (EPA QA/G-6).

EPA, 1996b. *Region I, EPA-NE Data Validation Functional Guidelines for Evaluating Environmental Analyses,* December 1996.

EPA, 1997. *EPA Requirements for Quality Assurance Project Plans for Environmental Data Operations,* November 1997 (EPA QA/R-5).

EPA, 1998a. *EPA Quality Manual for Environmental Programs,* 5360, July 1998.

EPA, 1998b. *Region I, EPA-New England Compendium of Quality Assurance Project Plan Requirements and Guidance,* Draft Final September 1998.

EPA, 1998c. *EPA Requirements for Quality Assurance Project Plans for Environmental Data Operations,* EPA QA/R-5, October 1998.

Foster Wheeler Environmental, 2000. *Sampling and Analysis Plan (Revision 11). New Bedford Harbor Superfund Site, New Bedford, Massachusetts.* December 2000.

Foster Wheeler Environmental, 2001. *Draft Sediment Sampling Report, New Bedford Harbor Superfund Site, New Bedford, Massachusetts.* April 2001.