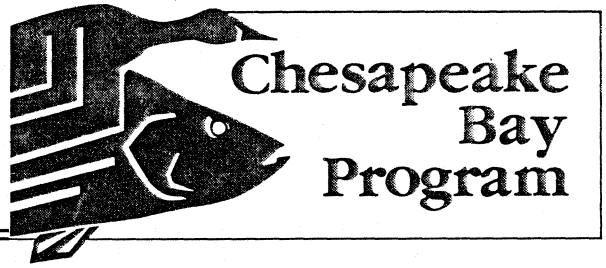
#10

Chesapeake Bay Program
Coordinated Split Sample Program

Report Series: No. 1
July 25, 1989

Split Sample Water Quality Results from Laboratories Participating in the Chesapeake Bay Program: 1985–1989



Split Sample Water Quality Results from Laboratories Participating in the Chesapeake Bay Program: 1985–1989

Chesapeake Bay Program
Coordinated Split Sample Program

Report Series: No. 1

July 25, 1989

Produced by
Peter W. Bergstrom

Computer Sciences Corporation
Chesapeake Bay Program
410 Severn Avenue
Suite 113
Annapolis, Maryland 21403
(301) 266-6873

FOREWORD

The reader of this review and evaluation must recognize some of the factors which affect the apparent performance of the participating laboratories in the split sampling exercises under consideration. First, it must be understood that the split sample results evaluated herein were efforts on the part of the participating laboratories to examine inter-laboratory variability, often without the full spectrum of controls necessary to establish true field splits and without the system of checks to focus on the specific source of the observed variability. That is to say that in many cases the variables associated with the study were not sufficiently controlled to assign the observed variability to the performance of the laboratory in question. Natural system variability, sediment laden samples, different holding times, and disparate analytical methods are among the variables known to be active in many of the attempts at split sample analyses.

Further, the reader must be very aware that the laboratories under consideration in this review are inherently different in their missions. While all engage in environmental analyses, some are research laboratories which specialize in low level estuarine and marine analyses while others are process laboratories which service a wide variety of analytical programs and matrices. To these process laboratories, low level environmental analyses and certainly estuarine matrices represent a small fraction of their work. Analytical systems are set up to address effluent samples and are then modified to handle the environmental samples which often require detection limits which are orders of magnitude lower than the routine work. In these modifications, sometimes compromises have been required, compared to a system which is optimized for the lowest level analytical work.

It is our hope that this review and the results of the current Chesapeake Bay Program Coordinated Split Sampling Program will be used together by laboratory and program staff to identify the sources of error associated with the analytical systems from the time of sample collection through handling, analysis, and data reporting. This will allow informed choices to be made in system design by each participating program to bring the quality of the data in line with the needs of the data users.

Bettina Fletcher
Chesapeake Bay Program Quality Assurance Officer

ACKNOWLEDGMENTS

The analysis of these data was improved with suggestions from Dr. Ray Alden and Steve Sokolowski, both at Old Dominion University. Helpful comments on earlier drafts of the manuscript came from Rich Batiuk and Tina Fletcher of the EPA-Chesapeake Bay Program, and CSC staff members Lacy Nasteff, Lowell Bahner, Ricky Price, Susan Brunenmeister, Chao-Hsi Chang, and Bob Reynolds.

INTRODUCTION

Split sample water quality testing is an important part of the Quality Assurance (QA) procedures of the Chesapeake Bay Program. Two types of split samples have been used (terms used are from EPA 1984):

1. Field split samples: One large water sample is split into smaller aliquots (bottles) for distribution to each analysis laboratory, as soon after sample collection as possible. This gives an estimate of inter-laboratory precision (agreement) after the sample collection stage. This is the most common type used (Table 1).

2. Co-located samples: Each aliquot is collected separately, but at the same time and same sampling location. This gives an estimate of inter-laboratory precision for the entire measurement system, including sample collection. This has been used in two programs (Table 1).

Split sample results complement the other QA data collected, which estimate intra-laboratory precision and accuracy. Although split sample results do not measure inter-laboratory accuracy directly (the closeness of the observed results to the true results), they can be used to estimate this indirectly when compared to other QA data. Split sample testing is especially important in a multi-state, multi-agency program such as the Chesapeake Bay Program, to ensure comparability of results.

In the past, several split sample programs have been conducted among laboratories in the Chesapeake Bay Program (Table 1), but these programs have not been mandated or coordinated by the EPA's Chesapeake Bay Liaison Office (CBLO). Methods have varied, and most of the data have never been analyzed statistically. Recently, the Monitoring Subcommittee of the Chesapeake Bay Program decided to implement a Chesapeake Bay Coordinated Split Sample Program, using field split samples, to begin in May or June 1989. As part of the planning for that program, an analysis of the data from the past Chesapeake Bay split sample programs was undertaken. This report gives the results of the analysis, for the following purposes:

1. To refine data submission and data analysis methods to be used in the Coordinated Split Sample Program, which is best done by looking at past split sample data;

2. To provide general estimates of inter-laboratory precision in the past. The methods used were too varied, and the data are too incomplete, to provide definitive estimates of this. It is hoped that these estimates will stimulate new investigations and communication among the laboratories involved, and that the result will be higher inter-laboratory agreement in water quality testing results in the future.

A preliminary version of this report was discussed at a meeting of the Virginia participants in the Coordinated Split Sample Program in Richmond on March 14, 1989. Many of the results differ in this version, because the data have been supplemented and corrected since then by the analysis laboratories.

DATA DESCRIPTION

The five split sample programs analyzed are listed in Table 1, and the laboratory abbreviations used below are spelled out there. Parameter abbreviations are spelled out the first time they are used in the text, and they are also listed in the CBP Water Quality Data Management Plan (CBP 1989).

Two-way Split Samples

1. VIMS and ODU (part of the Mainstem/Tidal Tributaries Component)

The data were analyzed in eight periods, starting in April 1985 (Cruise 17) and ending in January 1989 (Cruise 92). Most of the samples were collected by ODU, but some were collected by VIMS (Table 2). The stations used varied, but all were in the lower (Virginia) portion of the Bay; Station LE5.5, at the mouth of the James River, was the only station sampled during all time periods. The data were from the quarterly status reports from VIMS and ODU, except 1986 data were from Steve Sokolowski, and Kevin Curling sent Total Dissolved Nitrogen (TDN) data from VIMS. All the parameter names were changed to the CBP names (Table 2). The data from VIMS were not censored at the Method Detection Limit (MDL) when submitted, but this was done as part of the analysis using MDLs from Kevin Curling (see below).

There was only one aliquot sent to each laboratory per sample, so the results had to be pooled over sampling stations and over time to get large enough sample sizes for analysis (Table 2). For samples collected by ODU, inter-laboratory differences were not tied to particular sampling stations (see below), so data were pooled across the different stations used on the same cruise. During two periods when VIMS and ODU both collected samples (Cruises 17-31 and 32-40), the results from the samples collected by the two laboratories were analyzed separately. Data were also pooled over groups of cruises to achieve a sample size of about 10-20 pairs of data per comparison, covering periods of 3 to 8 months (Table 2). These groups of cruises represented periods when the stations and parameters included in the program did not change.

2. CBL and MDHMH (part of the Mainstem/Tidal Tributaries Component)

The data analyzed were from samples collected at Stations CB1.1 and CB2.2 in 1988. There were two aliquots, and thus two pairs of data, per

laboratory per station and cruise. Sample sizes were large enough to allow for analysis of the data from the two stations separately. Data from different cruises were combined before and after a break in the data in June.

Three-way Split Samples

1. CRL/DCRA, DCLS, and MDHMH (Tidal Potomac Component)

The data analyzed were from co-located samples collected at Station PMS-10 (Key Bridge) by DCECD, beginning in March 1986. Two gallon cubitainers were collected for DCLS and MDHMH, then the routine samples sent to CRL/DCRA were collected at the same depth (surface) and site a few minutes later. There was one aliquot, and one pair of data, per sample date (usually monthly). The data from DCLS were so incomplete that this was essentially a two-way split sample program between CRL/DCRA and MDHMH.

Four-way Split Samples

1. DCLS, HRSD, VIMS, and ODU (Virginia Mainstem/Tributaries Component)

The data analyzed were from four split samples from Station TF5.5 (Hopewell) collected between August 15, 1988 and January 11, 1989. The data from DCLS, VIMS, and ODU were corrected by the laboratories after they were submitted.

Fourteen parameters were compared; the goal of four-way split samples was achieved for five parameters (Table 6). Three-way split data could be analyzed for three parameters, and two-way split data were analyzed for six parameters (Table 6).

2. CBL, MDHMH, ODU, and VIMS (Mainstem/Tidal Tributaries Component)

The data analyzed were from four split samples from Station CB5.3 (near the Maryland-Virginia state line) in 1987-88 (Cruises 67, 81, 88, and 91). The samples were collected by CBL. Three field split aliquots were sent to each laboratory from each sample, the same number planned for the Coordinated Split Sample Program.

DATA ANALYSIS METHODS

Paired sample tests were used to analyze the two-way split sample results, because of the matching or positive correlation of samples that is inherent in the split sample design (Zar 1984). The paired t-test could not be used because the difference variables were usually not normally distributed, even after transformation. A non-parametric equivalent of the

paired t-test, the Wilcoxon paired-sample test, was substituted (Zar 1984), using StatView 512+ (BrainPower Inc. 1986) on a Macintosh for the analysis. This test requires a minimum of six pairs of data to find any significant differences.

A non-parametric repeated measures Friedman ANOVA was used on the four-way split data from Maryland and Virginia, which had three aliquots (replicates) per cruise. The test was done with StatView 512+ (BrainPower Inc. 1986), using the methods in Zar (1984), and it requires a minimum of three replicates per laboratory. The same method will be used to test the Coordinated Split Sample data when more than two laboratories are compared.

The four-way split sample data from TF5.5 in Virginia could not be analyzed with the Friedman test because there were unequal numbers of replicates, so an independent sample ANOVA had to be substituted, even though its assumptions are not strictly met by these data. For this reason, and because of the unequal cell sizes and some missing data, the results from this data set are tentative. The data were analyzed using a two-way ANOVA, using the General Linear Models (GLM) procedure (SAS Institute Inc. 1985). Multiple comparison posterior tests among three or more means were done with Duncan's Multiple Range Test, when the main effect for laboratories was significant (P for Type III SS < 0.05).

There were some values below Method Detection Limits (MDL), producing data that are "censored on the left" (Gilbert 1987, p. 177). Rather than exclude these values, non-parametric tests based on ranks were used, and any values below the MDL for that laboratory were set equal to the lowest MDL of the laboratories compared. For example, if the MDL at VIMS was 0.01 and at ODU it was 0.005, all values below 0.01 from VIMS were set to 0.005. Thus, all values that were below the MDL were tied for the lowest rank. This made it possible to keep those values without biasing the results of the analysis, and is the method recommended by Gilbert (1987, p. 252) for comparing data with "moderate" numbers of values below detection limits. When a majority of the values for one laboratory were below detection limits in a two-way comparison, or if two or more laboratories had values below detection limits in a three-way or higher comparison, no statistical test was done, since very little was known about the concentrations.

Several participants in the split sample programs suggested that inter-laboratory differences should exceed the MDL in absolute magnitude before any remedial action is needed, even if the difference is statistically significant. Large inter-laboratory differences are certainly more cause for concern than small ones, especially when differences are expressed as a percentage of the mean concentration. However, this use of the MDL applies only to single pairs of measurements, not to the long series of paired measurements analyzed here. If a single pair of measurements of the same water sample differ, it could be due to random error, especially if the difference is smaller than the MDL. However, when a series of split sample measurements consistently differ in

the same direction and with similar magnitudes, random error becomes a much less likely explanation, and systematic bias becomes more likely. For example, if the lab duplicates for a parameter always showed lower results from the first than the second aliquot, one would suspect bias, such as a problem with the sample splitting method, rather than random error. By definition, random errors should have a mean of zero. The P values from statistical tests express how likely it is that the observed differences are due to random errors, and when the P value is below 0.001 (as it was in five comparisons in Table 2, for example), random errors are a very unlikely cause for the observed differences. Any statistically significant difference found in a split sample program shows a consistent disagreement of laboratory results, which may indicate the presence of bias. Smaller differences probably indicate smaller amounts of bias, but they still merit attention if they are statistically significant in the same direction during more than one or two time periods.

RESULTS AND DISCUSSION

Two-way Split Samples

1. VIMS and ODU (LE5.5 and other stations)

Three of the eleven parameters compared (Total Kjeldahl Nitrogen--Whole [TKNW], Total Phosphorus [TP], and Total Dissolved Phosphorus [TDP]) differed significantly between VIMS and ODU during at least three periods (Table 2). Three other parameters (Nitrite/Nitrate [NO23], Silica [SI], and Ammonium [NH4]) had one or two significant differences each. (In this and all subsequent tables, significant inter-laboratory differences are shown by one or more asterisks.) Measurements from ODU were significantly higher than those from VIMS in 17 of the 20 significant differences found in these parameters (Table 2); the samples were mostly collected by ODU, and may have been analyzed sooner by them. This difference was not an artifact of differing detection limits, since values below the Method Detection Limit (MDL) from both laboratories were set to the same value (see above). One period, Cruises 68-74, had no significant differences for any parameter; this was the first period after analytical method changes that started with Cruise 68 (October 5, 1987). Several new parameters were added at Cruise 68, and none of them have had any significant inter-laboratory differences since then (Table 2).

The consistently higher results from ODU for TP and TDP were pointed out to Steve Sokolowski by Robert Siegfried in a letter dated June 10, 1986, apparently based on results through Cruise 32 (January 1986). The number of stations sampled was expanded to four in Cruise 52 (and later reduced to two in Cruise 68) to see if the inter-laboratory differences in these parameters depended on the station. The results separated by station for this period (Table 3) show that there was some station effect, but

every station had some significant differences. Stations LE5.5, CB7.4N, and CB7.3 had similar rates of significant differences in this small sample (67% of the comparisons made), while significant differences were less frequent at Station CB6.4 (22% of the comparisons made). The occurrence of significant laboratory differences in TP and TDP was apparently not related to concentration: plots of percent difference vs. concentration (Fig. 1 a) and the Probability (P) value of the Wilcoxon test vs. concentration (Fig. 1 b) for TP and TDP show no obvious concentration effects. Further investigation by the laboratories involved is needed to understand the possible causes of these differences, which persisted after the method changes at Cruise 68 (Table 2).

2. CBL and MDHMH (Stations CB1.1 and CB2.2)

Nine of the eleven parameters compared differed significantly between CBL and MDHMH during at least two periods (Table 4). Measurements from CBL were significantly higher for Total Organic Carbon (TOC), Dissolved Organic Carbon (DOC), Total Suspended Solids (TSS), and Orthophosphate (PO4F) (positive differences in the table), while those from MDHMH were significantly higher for SI, NO23, Nitrite (NO2), Nitrate (NO3), TP, and TDP (negative differences). For SI and TSS, significant differences occurred in one period and not the other. There were no parameters that differed at one station and not the other; Station CB1.1 is in the tidal fresh zone, while CB2.2 had a surface salinity of 0 to 5.5 o/oo (mean 2.4 o/oo) during this period. The series was too short to determine if any of the differences were related to concentration. These samples were collected by CBL.

Three-way Split Samples

1. CRL/DCRA, DCLS, and MDHMH (Station PMS-10)

Three of the nine parameters compared differed significantly between CRL/DCRA and MDHMH during at least two periods, and three more parameters had one significant difference each (Table 5). Measurements at MDHMH were higher (negative differences) in all of the 10 significant differences in these six parameters. Since this program used co-located rather than field split samples, there was an added source of variability (sample collection) not included in field split samples. Its magnitude is unknown.

The results from three samples in 1986 could be compared among CRL/DCRA, DCLS, and MDHMH for the parameters TOC, TKNW, NO23, and TP. There were no significant inter-laboratory differences in this small data set, using the Friedman ANOVA.

Four-way Split Samples

1. DCLS, HRSD, VIMS, and ODU (Station TF5.5)

Eight parameters were compared among three or four laboratories (Table 6), and two (NO23 and TP) showed significant inter-laboratory differences (Table 7). The values for NO23 were significantly higher at DCLS, and significantly lower at ODU, than at the other laboratories (Table 7). ODU was usually the last laboratory to receive its aliquots, which might cause the lower NO23 values from ODU. Of the six two-way split parameters, two (Total Nitrogen [TN] and Total Dissolved Nitrogen [TDN]) showed significant differences (Table 7). However, as noted above, missing values and unequal cell sizes may have affected these results.

2. CBL, MDHMH, ODU, and VIMS (Station CB5.3)

There were significant inter-laboratory differences ($P < 0.03$) for two or more cruises for five of the nine parameters compared (TP, Particulate Phosphorus [PHOSP], TDP, SI, and NH4; Table 8). The same five parameters also had significant inter-laboratory differences when the data were combined over all the cruises with complete data for that parameter (see the last row in each section, Table 8). Due to the small sample sizes, pairwise (or multiple) comparisons among laboratory means could only be done for these pooled data, and their results are shown by the rows of letters in Table 8. In this notation, any two laboratory means that do **not** have the same letter below them differ significantly ($P < 0.05$).

The results from MDHMH were higher than those from the other laboratories in 18 of the 25 significant differences (including the results pooled over two or more cruises), by up to 219% more than the overall mean concentration. In the other seven cases, one of the other three laboratories had the most different values (Footnote 3, Table 8). The higher values from MDHMH were not the only cause of significant differences, however, because 13 of the 21 significant differences (in four-way comparisons) were still significant when data from MDHMH were excluded (Footnote 4, Table 8). This table shows that combining results over several cruises can mask differences found in separate cruises, since the results for PHOSP showed significant differences only when compared over single cruises. This demonstrates the advantage of having three aliquots per laboratory per cruise, so that pooling results over cruises is unnecessary.

The results in this table are similar to those in Tables 2 and 4, which compared pairs of the same laboratories. As in the previous tables, measurements from ODU were usually higher than those from VIMS, and those from MDHMH were generally higher than those from CBL (for the parameters included in both tables).

CONCLUSIONS AND RECOMMENDATIONS

Although the methods, parameters, and sample sizes of the five split sample programs were too varied to make any quantitative comparisons among their results, some general patterns are discernible. Certain parameters tended to have more frequent and larger inter-laboratory differences than others, particularly TP and TDP. Since TP is usually calculated from (TDP + PHOSP), the problem is mainly with TDP. Three parameters, NO23, NH4, and TDP, were sometimes below detection limits, which limits the usefulness of split sample comparisons. One laboratory, MDHMH, appeared to have more frequent and larger inter-laboratory differences than other laboratories, although it was also involved in more of the split sample programs than some laboratories. The advantage of conducting split sample programs consistently over several years is shown by the VIMS/ODU program in Table 2, since differences sometimes changed over time.

Detailed investigations aimed at determining the causes of inter-laboratory differences are beyond the scope of this report. However, clues concerning possible causes can be obtained by data analysis (Table 3 and Fig. 1), and plots of the raw data, both as sequential line plots and box plots (Fig. 2a and 2b), can also suggest causes or corrective actions. For example, the line plot (Fig. 2a) shows that there was low intra-laboratory precision in the ODU results for Cruise 88 (observations 7-9 on the plot). Raw data from any of the split sample programs are available to any involved program or laboratory personnel for further analysis.

Based on these results, the following measures are desirable as part of future split sample programs:

1) Ensure consistent measurement and reporting of parameters. The laboratories should try to analyze the same parameters, with the same number of replicates, and report them with CBP names. Only two parameters, TP and NO23, were reported by all five split sample programs. To achieve consistent and complete reporting of parameters, laboratories should follow the revised guidelines for data submission (May 1989), using the standard data submission form.

2) Report the actual results for values below detection limits, while also flagging them with the "<" code. This would make it possible to analyze the actual results, which is recommended by Gilbert (1987) under some circumstances, or set them to a fixed value lower than any measured values in the data set, as was done with the data analyzed with the Wilcoxon paired-sample test and the Friedman ANOVA. It would be very helpful if each laboratory sent a list of current MDLs with each split sample data submission.

3) Modify the components of the Coordinated Split Sample Program so that they all contain at least three laboratories. This could be done by

merging some of the fall line components, or adding a laboratory to existing components. The Wilcoxon paired-sample test cannot be used with fewer than six replicates, so the results from any component with two laboratories can only be analyzed over pairs of cruises. Also, determining the sources of differences is easier when three or more laboratories are compared. In components with three or more laboratories, two-way comparisons will be avoided by comparing directly measured to computed parameters, if the data for the computations are available. Ideally, the results from each component of the program will resemble the format in Table 8, with at least three laboratories in each comparison, and three replicates (aliquots) per cruise from each laboratory.

REFERENCES

BrainPower, Inc. 1986. StatView 512+. BrainPower, Inc., Calabasas, CA.

CBP (Chesapeake Bay Program), 1989. Water Quality Data Management Plan, Revision 1. Chesapeake Bay Program, Annapolis, MD.

EPA, Quality Assurance Management and Special Studies Staff. 1984. Chapter 5: Calculation of precision, bias, and method detection limit for chemical and physical measurements. US EPA, Washington, D.C.

Gilbert, R. O. 1987. Statistical methods for environmental pollution monitoring. Van Nostrand Reinhold, New York.

SAS Institute, Inc. 1985. SAS User's Guide: Statistics. Version 5 Edition. SAS Institute, Inc., Cary, NC.

Zar, J. 1984. Biostatistical analysis. Prentice-Hall, Englewood Cliffs, N.J.

Table 1

Chesapeake Bay Split Sample Programs, 1985-1989

Component[1]	Laboratories[2]	Dates	Aliquots/sample[3]
Mainstem/Tidal Tributaries (VA)	ODU & VIMS	4/85 - 1/89	1
Mainstem/Tidal Tributaries (MD)	CBL & MDHMH	2/88 - 12/88	2
Tidal Potomac[4]	CRL/DCRA, DCLS, & MDHMH	3/86 - 1/89	1
VA Mainstem/Tributaries	DCLS, HRSD, VIMS, & ODU	8/88 - 1/89	1-3
Mainstem/Tidal Tributaries	CBL, MDHMH, ODU, & VIMS	9/87 - 12/88	3
Potomac Fall Line[5]	USGS & OWML	?	1

[1] Component (or part of a component) of the Coordinated Split Sample Program. The Mainstem/Tidal Tributaries Component will also include DCLS. All use field split samples, except as noted.

[2] ODU = Old Dominion University, VIMS = Virginia Institute of Marine Science, CBL = Chesapeake Biological Laboratory (Solomons), MDHMH = Maryland Department of Health and Mental Hygiene (Baltimore), CRL/DCRA = Central Regional Laboratory (EPA, Annapolis), with analyses run there by DC Dept. of Consumer & Regulatory Affairs personnel, DCLS = Division of Consolidated Laboratory Services (Richmond), HRSD = Hampton Roads Sanitation District (Newport News), USGS = U.S. Geological Survey, OWML = Occoquan Watershed Monitoring Laboratory.

[3] An aliquot is a bottle split from a larger sample in the field; each laboratory received 1 to 3 aliquots per sample.

[4] Uses co-located samples, collected by the DC Environmental Control Division (DCECD) at PMS-10. Due to missing data from DCLS, it was mainly a two-way split sample program between CRL/DCRA and MDHMH.

[5] This program is done at Chain Bridge with co-located sampling; data are being analyzed by S. Kenney at Maryland Dept. Environment (MDE).

Table 2

Results of two-way split samples, VIMS and ODU, 1985-88

CBP name	Cruises	N[1]	Mean diff.[2]	Mean conc.[3]	% diff.[4]	Significance[5]
TKNW	17-31	15 (1)	-0.180	0.4962	36%	**
TKNW	17-31[6]	14 (8)	-0.055	0.325	-	
TKNW	32-40	9 (1)	-0.240	0.521	46%	**
TKNW	32-39[6]	8 (2)	-0.159	0.458	35%	*
TKNW	41-51	9 (1)	-0.122	0.428	29%	**
TKNW	52-61	26 (4)	-0.070	0.347	20%	**
TKNW	62-67	24 (4)	-0.176	0.479	37%	***
TP[7]	17-31	15 (1)	-0.042	0.073	58%	***
TP	17-31[6]	26 (8)	-0.023	0.041	57%	***
TP	32-40	9 (1)	-0.032	0.068	47%	**
TP	32-39[6]	8 (2)	-0.0054	0.025	-	
TP	41-51	9 (1)	-0.0019	0.065	-	
TP	52-61	35 (4)	-0.017	0.052	33%	**
TP	62-67	20 (4)	-0.0033	0.054	-	
TP	68-74	12 (2)	0.0083	0.031	-	
TP	76-85	20 (2)	-0.0074	0.040	19%	*
TP	86-92	14 (2)	-0.012	0.038	31%	**
TDP[8]	32-40	5 (1)	-[9]	-	-	-
TDP	32-39[6]	8 (2)	-[9]	-	-	-
TDP	41-51	10 (1)	0.00010	0.035	-	
TDP	52-61	32 (4)	-[9]	-	-	-
TDP	62-67	20 (4)	0.0075	0.027	28%	***
TDP	68-74	14 (2)	0.0006	0.022	-	
TDP	76-85	20 (2)	-0.0078	0.020	39%	**
TDP	86-92	14 (2)	-0.0091	0.022	41%	**
PHOSP	68-74	12 (2)	0.0052	0.017	-	
PHOSP	76-85	20 (2)	-0.00070	0.020	-	
PHOSP	86-92	14 (2)	-0.0016	0.016	-	
NO23	32-40	9 (1)	0.0080	0.034	24%	*
NO23	32-39[6]	8 (2)	-[9]	-	-	-
NO23	41-51	9 (1)	-[9]	-	-	-
NO23	52-61	10 (4)	-0.0055	0.023	-	
NO23	62-67	20 (4)	-[9]	-	-	-
NO23	68-74	14 (2)	0.0059	0.027	-	
NO23	76-85	16 (2)	-[9]	-	-	-
NO23	86-92	14 (2)	0.0058	0.028	21%	*

Table 2 (continued)

CBP name	Cruises	N[1]	Mean diff.[2]	Mean conc.[3]	% diff.[4]	Significance[5]
SI	17-31	15 (1)	-0.0010	0.371	-	
SI	17-31[6]	14 (8)	0.0274	0.230	-	
SI	32-40	9 (1)	-0.0500	0.143	35%	**
SI	32-39[6]	8 (2)	0.0020	0.047	-	
SI	41-51	9 (1)	-0.019	0.374	-	
SI	52-61	19 (4)	0.0006	0.0617	-	
SI	62-67	24 (4)	-0.016	0.285	-	
SI	68-74	14 (2)	-0.0334	0.319	-	
SI	76-85	20 (2)	0.016	0.233	-	
SI	86-92	14 (2)	-0.0252	0.284	-	
NH4	17-31	15 (1)	-0.013	0.051	25%	**
NH4	17-31[6]	12 (8)	0.0076	0.031	-	
NH4	32-40	9 (1)	-0.01	0.0235	43%	*
NH4	32-39[6]	6 (2)	-[9]	-	-	-
NH4	41-51	9 (1)	0.0081	0.072	-	
NH4	52-61	19 (4)	-[9]	-	-	-
NH4	62-67	20 (4)	-0.0024	0.037	-	
NH4	68-74	14 (2)	0.021	0.055	-	
NH4	76-85	17 (2)	0.0115	0.050	-	
NH4	86-92	10 (2)	-0.0028	0.051	-	
PC[10]	68-74	12 (2)	-0.147	1.00	-	
PC	76-85	16 (2)	-0.093	0.57	-	
PC	86-92	14 (2)	-0.12	0.53	-	
PN	68-74	12 (2)	-0.016	0.078	-	
PN	76-85	18 (2)	-0.0061	0.067	-	
PN	86-92	13 (2)	0.027	0.076	-	
TDN[11]	68-74	10 (2)	0.071	0.32	-	
TDN	76-85	13 (2)	0.0066	0.25	-	
TDN	86-92	8 (2)	-0.018	0.29	-	
TN	68-74	10 (2)	0.050	0.48	-	
TN	76-85	12 (2)	-0.00085	0.33	-	
TN	86-92	7 (2)	0.019	0.37	-	

Table 2 (continued)

Dates for each period were:
17-31 = 4/23/85-12/85,
32-40 = 1/86-6/9/86,
41-51 = 6/25/86-12/86,
52-61 = 1/87-6/15/87,
62-67 = 7/6/87-9/14/87,
68-74 = 10/87-2/88 (there was no Cruise 75),
76-85 = 3/22/88-8/22/88,
86-92 = 9/9/88-1/89.

1 Number of pairs compared in that period (number of stations per period).
2 Mean difference between laboratory values, (VIMS - ODU), in mg/l.
3 Mean concentration, from VIMS and ODU values (both censored at MDL).
4 Mean difference as % of mean concentration, significant differences only.
5 Significance of Wilcoxon paired-sample test (Zar 1984): * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$.
6 Samples were collected by VIMS; all other samples were collected by ODU.
7 Reported as TPO4 by ODU.
8 Reported as DPO4, DP, or TP(f) by ODU.
9 Too many values were below detection limits to make a comparison.
10 PC and PN are approximately equivalent to Particulate Organic Carbon (POC) and Particulate Organic Nitrogen (PON) as used by the CBP.
11 Reported as Dissolved Persulfate Nitrogen (DPN) by ODU, and as TN by VIMS. Omitted from summary tables made by R. Hoffman.

Table 3

Effect of sampling station on the significance of differences

VIMS-ODU data from Table 2, Cruises 52-92

Parameter[1]	Cruises	Stations[2] LE5.5	CB7.4N	CB7.3	CB6.4	All
TKNW	52-61	*	*	ns	ns	**
TKNW	62-67	ns	ns	*	*	***
TP	52-61	**	*	*	ns	**
TP	68-74	ns	-	-	ns	ns
TP	76-85	**	-	-	ns	*
TP	86-92	*	-	-	*	**
TDP	68-74	ns	-	-	ns	ns
TDP	76-85	*	-	-	ns	***
TDP	86-92	*	-	-	ns	**

[1] Only parameters with at least one significant difference after Cruise 52 are listed. Only groups of cruises with N ≥ 6 for each station are included.

[2] * = P < 0.05, ** = P < 0.01, *** = P < 0.001, ns = not significant (P > 0.05), all for Wilcoxon paired-sample test (Zar 1984). - = no data.

Table 4

Results of two-way split samples, CBL and MDHMH, 1988

CBP name	Cruises	Station	N[1]	Mean diff.[2]	Mean conc.[3]	% diff.[4]	Significance[5]
SI	73-79	CB1.1	12	-0.018	1.48	-	
SI	82-89	CB1.1	16	-0.071	1.02	7%	***
SI	73-79	CB2.2	12	0.022	1.60	-	
SI	82-89	CB2.2	15	-0.141	1.11	13%	**
TOC	73-79	CB1.1	9	1.89	3.06	62%	**
TOC	82-89	CB1.1	14	1.46	3.39	44%	***
TOC	73-79	CB2.2	9	1.21	2.59	47%	**
TOC	82-89	CB2.2	16	1.15	3.28	35%	*
DOC	73-79	CB1.1	9	0.78	2.29	34%	**
DOC	82-89	CB1.1	14	0.714	2.72	26%	***
DOC	73-79	CB2.2	9	0.52	1.98	26%	**
DOC	82-89	CB2.2	16	0.591	2.84	-	
TSS	73-79	CB1.1	12	4.79	16.9	28%	**
TSS	82-89	CB1.1	16	-3.51	9.31	-	
TSS	73-79	CB2.2	12	5.49	17.9	31%	*
TSS	82-89	CB2.2	15	-2.51	10.6	-	
NH4	73-79	CB1.1	10	0.0017	0.057	-	
NH4	82-89	CB1.1	12	0.0023	0.037	-	
NH4	73-79	CB2.2	12	0.0073	0.12	-	
NH4	82-89	CB2.2	14	-0.0019	0.11	-	
NO23	73-79	CB1.1	11	-0.122	1.15	11%	*
NO23	82-89	CB1.1	13	-0.120	1.09	11%	**
NO23	73-79	CB2.2	12	-0.174	1.21	14%	**
NO23	82-89	CB2.2	16	-0.022	0.563	-	
NO2	73-79	CB1.1	12	-0.0027	0.015	21%	**
NO2	82-89	CB1.1	16	-0.0041	0.039	10%	***
NO2	73-79	CB2.2	12	-0.0032	0.016	19%	**
NO2	82-89	CB2.2	16	-0.0026	0.046	-	
NO3	73-79	CB1.1	11	-0.12	1.1	10%	*
NO3	82-89	CB1.1	13	-0.12	1.0	11%	**
NO3	73-79	CB2.2	12	-0.17	1.2	14%	**
NO3	82-89	CB2.2	15	-0.020	0.52	-	

Table 4 (continued)

CBP name	Cruises	Station	N[1]	Mean diff.[2]	Mean conc.[3]	% diff.[4]	Significance[5]
TP	73-79	CB1.1	12	-0.021	0.066	32%	**
TP	82-89	CB1.1	16	-0.018	0.046	39%	**
TP	73-79	CB2.2	12	-0.021	0.066	32%	**
TP	82-89	CB2.2	15	-0.014	0.045	31%	**
TDP	73-79	CB1.1	11	-0.022	0.026	85%	**
TDP	82-89	CB1.1	16	-0.012	0.019	63%	**
TDP	73-79	CB2.2	12	-0.024	0.031	79%	**
TDP	82-89	CB2.2	16	-0.032	0.037	86%	**
PO4F	73-79	CB1.1	12	0.000017	0.010	-	
PO4F	82-89	CB1.1	16	-[6]	-	-	-
PO4F	73-79	CB2.2	12	0.00027	0.014	-	
PO4F	82-89	CB2.2	15	0.0039	0.011	36%	**

Dates of periods were: 73-79 = 2/88-5/25/88, 82-89 = 7/7/88-11/88.

[1] Number of pairs of data, two pairs (aliquots) per cruise and station.

[2] Calculated as (CBL value - MDHMH value), in mg/l.

[3] Mean concentration, from both CBL and MDHMH values.

[4] Laboratory difference as a % of mean concentration for significant differences only (- listed for non-significant differences).

[5] Significance of the Wilcoxon paired-sample test (Zar 1984): * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$.

[6] Too many PO4F values from MDHMH were below detection limits (14 of 16) to make a comparison meaningful.

Table 5

Results of two-way split samples, CRL/DCRA and MDHMH

1986-88, Station PMS-10 (Key Bridge)

CBP name	Year	N[1]	Mean diff.[2]	Mean conc.[3]	% diff.[4]	Significance[5]
TOC	86	7	0.26	3.8	-	
TOC	87-88	7	0.56	4.3	-	
TKNW	86	7	-0.233	0.617	38%	*
TKNW	87-88	6	-0.0810	0.660	-	
NO23	86	7	-0.046	0.71	6%	*
NO23	87	10	-0.091	1.5	6%	*
NO23	88	9	-0.11	1.1	10%	*
NO2	86	6	0.0017	0.016	-	
NO2	87	10	-0.0070	0.016	45%	*
NO2	88	10	-0.0020	0.013	-	
NO3	86	6	-0.052	0.73	7%	*
NO3	87	10	-0.062	1.5	-	
NO3	88	9	-0.11	1.1	10%	*
TP	86	7	-0.026	0.057	46%	*
TP	87-88	6	-0.005	0.13	-	
PO4F	86	6	-0.011	0.018	-	
PO4F	87	9	-0.016	0.038	42%	*
PO4F	88	6	-0.010	0.029	34%	*
TSS	86	5	-3.2	9.4	-	
TSS	87-88	7	42	56	-	
NH4	87	10	0.010	0.043	-	
NH4	88	9	-0.0020	0.041	-	

1 Number of pairs of data, one pair (aliquot) per sampling date.
2 Calculated as (CRL/DCRA value - MDHMH value), in mg/l.
3 Mean concentration, from both CRL/DCRA and MDHMH values.
4 Laboratory difference as a % of mean concentration for significant differences only.
5 Significance, Wilcoxon test: *=P < 0.05, **=P < 0.01, ***=P < 0.001.

Table 6

Parameter names and sample sizes for four-way split samples

Virginia Mainstem/Tributaries Component

Station TF5.5, August 1988 to January 1989 (four samples)

CBP name	Other names used	Sample sizes			
		D[1]	H	O	V
PO4F	OP, O-PO4	8	11	0	10
NH4	NH3, Ammonia	8	11	8	7
NO23	NO2+3, NO3+2, NO3/NO2	8	11	9	10
NO3		8	9	0	10
NO2		7	9	0	9
TKNW	TKNU	8	12	0	0
TN		0	0	9	7
PN	Particulate-N	0	0	9	9
TDN		0	0	12	8
TP	TPU	8	9	10	9
TDP	TPF, Dissolved PO4	8	9	10	9
PHOSP	PP, Particulate PO4	8	9	12	12
SI	Silicates	8	0	0	10
PC	Particulate-C	0	0	9	9

[1] D = DCLS, H = HRSD, O = ODU, V = VIMS. Sample size is the number of data points for three or four sample dates (cruises). Three aliquots per cruise.

Table 7

Results of four-way split samples, Station TF5.5, 1988-89

Virginia Mainstem/Tributaries Component

CBP name	Mean concentrations (mg/l)				% diff.[1]	Signif.[2]
	DCLS	HRSD	ODU	VIMS		
NH4	0.28	0.25	0.17	0.12	-	
NO23	1.1	0.96	0.83	0.95	11%, 12%	***
NO3	0.87	0.91	-	0.88	-	
NO2	0.040	0.043	-	0.052	-	
TKNW	1.2	1.6	-	-	-	
TN	-	-	2.0	1.6	25%	**
PN	-	-	0.46	0.45	-	
TDN	-	-	1.6	1.1	33%	**
TP	0.24	0.23	0.25	0.26	-	*[3]
PHOSP	0.11	0.096	0.13	0.12	-	
TDP	0.13	0.14	0.12	0.13	-	
PO4F	0.12	0.12	-	0.13	-	
SI	1.76	-	-	1.60	-	
PC	-	-	3.15	3.42	-	

[1] Laboratory differences as a % of the mean concentration, for significant differences only.

[2] Significance, based on P value from ANOVA (Type III SS): * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$. Laboratory means that are not connected by an underline differ significantly (means compared only when overall P was significant). Results are tentative due to unequal cell sizes.

[3] Overall P was significant, yet there were no pairwise differences found by the Duncan test.

Table 8

Results of four-way split samples, Station CB5.3, 1987-88

Mainstem/Tidal Tributaries Component

CBP name	Cruise	Mean Concentration (mg/l) CBL	MDHMH	ODU	VIMS	Max. % diff.[1]	Significance[2]
TP	67	0.033	0.053	0.016	0.020	73%	**[4]
TP	81	0.015	0.067	-	0.020	97%	*
TP	88	0.023	0.050	0.044	0.020	46%	*[4]
TP	91	0.016	0.070	0.017	0.017	-	
TP	67,88,91	0.024	0.058	0.026	0.019	83%	**
		AAAAA		AAAAAAAAAAAAA			
PHOSP	81	0.013	0.004	-	0.012	59%	*
PHOSP	88	0.016	0.010	0.029	0.010	78%[3]	*
PHOSP	91	0.008	0.027	0.010	0.008	104%	**[4]
PHOSP	81-91	0.012	0.018	0.019	0.0090	-	
		AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA					
TDP	67	0.012	0.050	0.015	0.005[6]	144%	**[4]
TDP	81	0.0020	0.063	0.0090	0.005[6]	219%	**[4]
TDP	88	0.0070	0.040	0.014	0.005[6]	142%	**
TDP	91	0.0080	0.043	0.0080	0.005[6]	-	
TDP	67-91	0.0072	0.049	0.012	0.005[6]	168%	***[4]
		AAAAAA	BBBBBBBBBBBBBB		AAAAAA		
		cccccc		cccccc			
PO4F	67	0.0023[5]	0.0047	-	0.002[6]	57%	*
SI	67	0.113	0.30	0.217	0.187	47%	**[4]
SI	81	0.037	0.10	0.183[5]	0.095[5]	76%	*[4]
SI	88	0.22	0.433	-	0.26	42%	*
SI	91	0.153	0.40	0.214	0.223	62%	*
SI	67,81,91	0.101	0.267	0.205	0.168	45%[3]	***[4]
		AAAAA	BBBBBBBBBBBBB		AAAAA		
NO23	67	0.0044[5]	-[6]	-	-[6]	-	-
NO23	81	0.004[5]	-[6]	-[6]	-[6]	-	-
NO23	88	0.0060	-[6]	-[6]	-[6]	-	-
NO23	91	0.067	0.070	0.049	0.060	20%[3]	**[4]

Table 8 (continued)

CBP name	Cruise	Mean Concentration (mg/l) CBL	MDHMH	ODU	VIMS	Max. % diff.[1]	Significance[2]
PC	81	1.02	0.54[5]	0.630	0.546	-	
PC	88	0.691	0.32[5]	0.707	0.578	44%	*
PC	91	0.480	0.27[5]	0.450	0.372	-	
PC	81-91	0.731	0.378	0.599	0.499	32%[3]	**[4]
		AAAAA		AAAAA			
			BBBBBBBBBBBBBBBBBBBB				
PN	81	0.182	0.150	0.078	0.113	43%[3]	**
PN	88	-	0.133	0.129	0.107	-	
PN	91	-	0.100	0.062	0.068	-	
PN	81-91	-	0.128	0.090	0.096	-	
			AAAAAAAAAAAAAAAAAAAA				
NH4	67	0.016[5]	0.016	0.006	0.028	70%[3]	*[4]
NH4	81	-[6]	0.013	0.012	-[6]	-	-
NH4	88	0.0070	0.024	0.0080	0.020	63%	*[4]
NH4	91	0.039	0.061	0.051	0.040	28%	*
NH4	67-91	0.025	0.029	0.020	0.026	20%[3]	*
		AAAAA		AAAAAAAAAAAA			
			BBBBBBBBBBBBBBBBBBBB				
TKNW	67	-	1.21[5]	0.36	0.399	-	

Pairs of means that do **not** have the same letter below them **differ** significantly ($P < 0.05$, using ranks from Friedman test, for means calculated over three or more cruises only).

Dates of cruises were: 67 = 9/14/87, 81 = 6/20/88, 88 = 10/3/88, 91 = 12/5/88.

[1] Maximum % difference, for significant ($P < 0.03$) comparisons only. Let X = mean concentration over all laboratories, Y = laboratory mean that differs most from X; maximum % difference = (100)(Y - X) / X.

[2] Significance of Friedman non-parametric ANOVA, using the exact probability table (Table B.13) in Zar (1984): * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$.

[3] The most different mean was from ODU, VIMS, or CBL, rather than from MDHMH as in all other cases of significant differences.

[4] Three-way comparison without MDHMH was also significant in these cases.

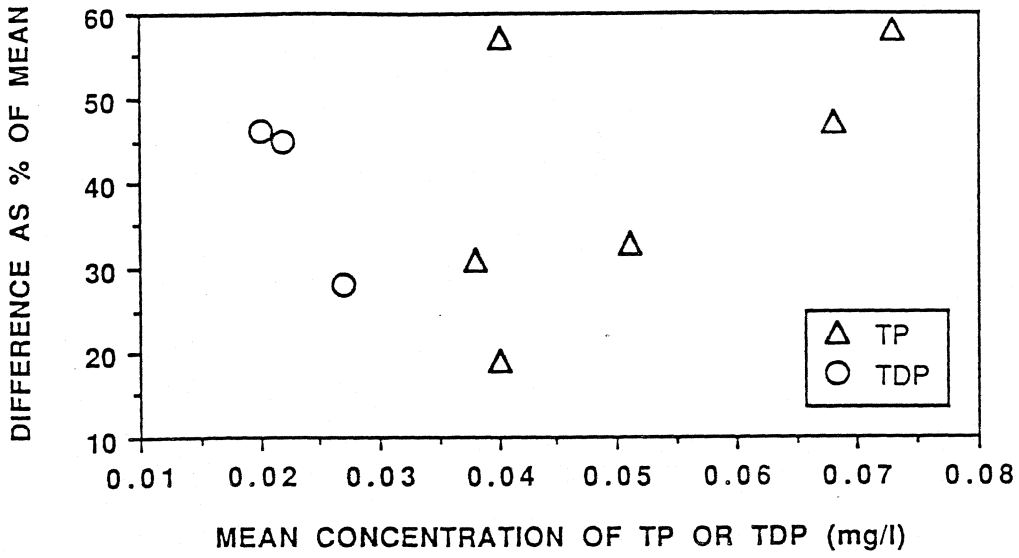
[5] Low precision among the 3 aliquots (SD/mean = CV = 39% to 145%).

[6] Values below detection limits were reduced to a value ≤ the lowest measured value; Friedman test was not calculated when this was done for two or more laboratories, because of the uncertain ranks.

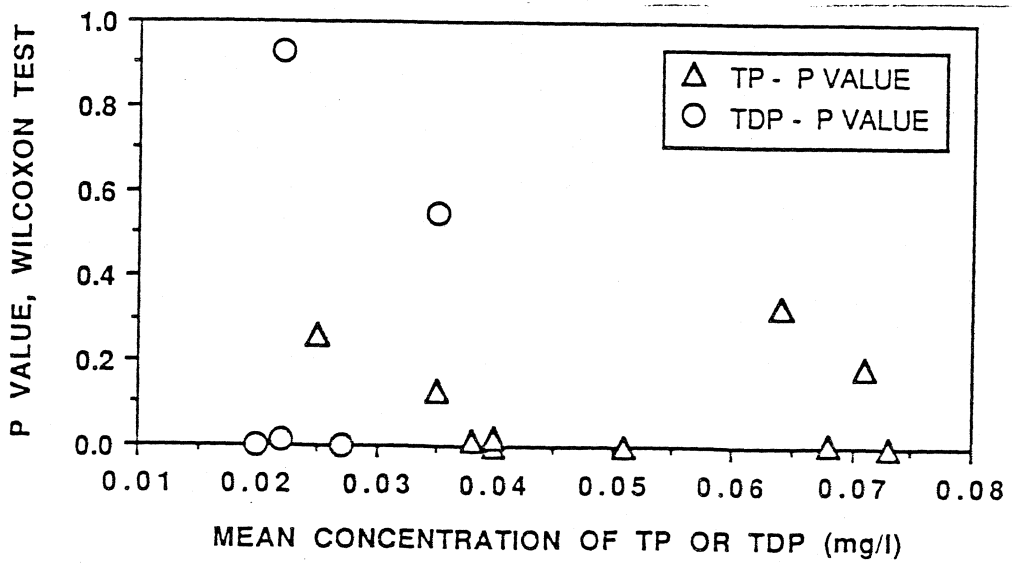
Figure 1

Percent difference and Significance vs. Concentration

TP and TDP Data from Table 2 (VIMS and ODU)



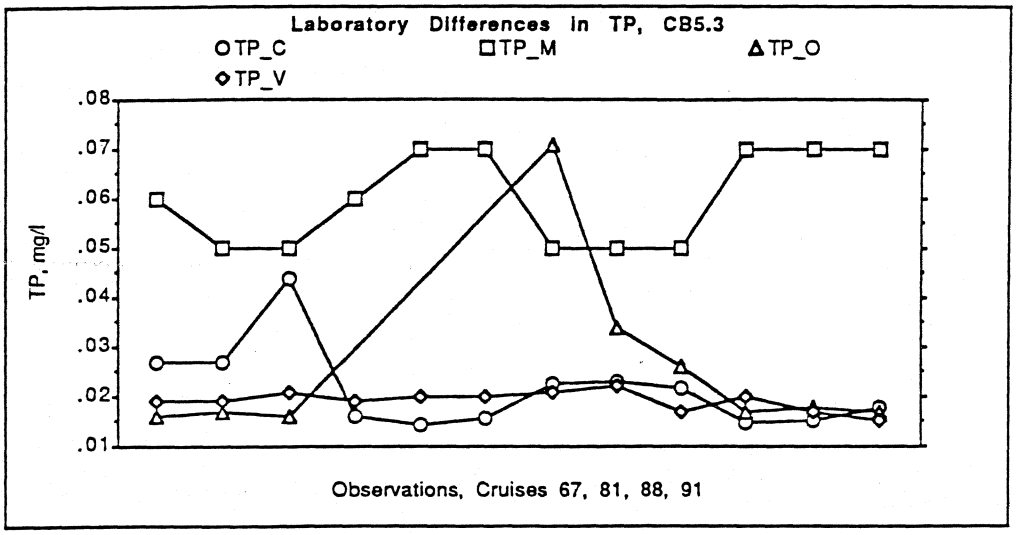
a. Percent difference vs. concentration for significant ($P < 0.05$) inter-laboratory differences between VIMS and ODU split sample results for TP and TDP, 1985-1989.



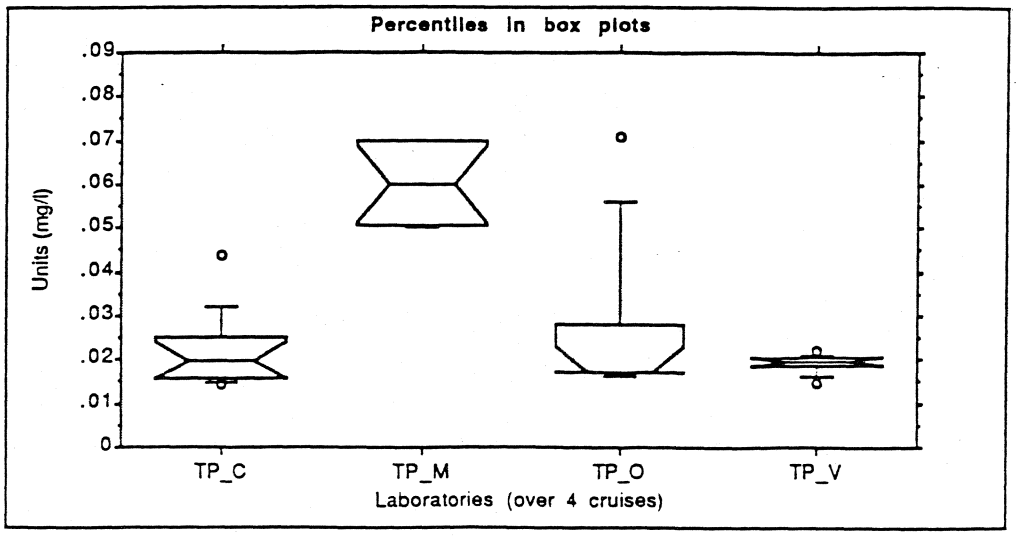
b. P value (from Wilcoxon paired-sample test) vs. concentration for TP and TDP from VIMS and ODU, 1985-1989.

Figure 2

Plots of raw TP data from CB5.3; CBL, MDHMH, ODU, and VIMS



a. Sequential line plot of the raw data for TP from CB5.3, showing the results from the three aliquots from the four cruises in order.



b. Notched box plots of the same data showing the 10th, 25th, 50th (or median, with notch), 75th, and 90th percentiles, with observations outside those limits shown as circles (both plots were made with StatView 512+).

COMPUTER SCIENCES CORPORATION

APPLIED TECHNOLOGY DIVISION (301) 266-6873
410 SEVERN AVENUE, SUITE 110 • ANNAPOLIS, MARYLAND 21403

Date : July 25, 1989

Subject : Split Sample Water Quality Results

From : Lacy Nasteff
Site Manager
Computer Sciences Corporation

To : Distribution

Attached is the report titled "Split Sample Water Quality Results from Laboratories Participating in the Chesapeake Bay Program: 1985-1989" produced by Peter Bergstrom of Computer Sciences Corporation. If you have any questions regarding this report, please contact Peter at 301-266-6873.

cc:
L. Nasteff, CSC
L. Bahner, CSC
R. Batiuk, EPA/CBLO
K. Mountford, EPA/CBLO
P. Bergstrom, CSC
S. Brunenmeister, CSC
C. Chang, CSC
T. Fletcher, EPA/CRL
R. Price, CSC
R. Reynolds, CSC
E. Stigall, EPA/CBLO
R. Magnien, MDE
B. Michael, MDE
C. Zimmerman, CBL
A. Bober, MDHMH
A. Lepley, MDHMH
R. Hoffman, VWCB
C. Cook, VWCB
N. Roadcap, DCLS
B. Neilson, VIMS
B. Salley, VIMS
R. Alden, ODU
S. Sokolowski, ODU
H. Karimi, DCRA
J. Collier, DCRA
D. Donnelly, EPA/CRL
O. Villa, Jr., EPA/CRL
P. Zawodny, EPA/CRL
P. Krantz, EPA/CRL
S. Myers, MWCOG
D. Francis, HRSD