



**Statistical analysis and summary of Analytical  
Round Robin #1 - a data comparability study**

**Report prepared by Linda Sedlacek  
May 2008**

Samples collected February 25 2008  
in Pensacola Bay, FL and preserved, split, and prepared for shipment  
at the U.S. EPA Gulf Ecology Laboratory, Gulf Breeze, FL

# Table of Contents

1. Introduction	3
2. Statistical methodology	5
3. Results and Discussion	7
A. Chlorophyll a	7
B. Total phosphorous	8
C. Orthophosphate	11
D. Ammonia	11
E. Total Nitrate + Nitrite	13
F. Total Kjeldahl Nitrogen	14
G. Biological Oxygen Demand	14
H. Carbonate Biological Oxygen Demand	15
I. Total Organic Carbon	16
J. Dissolved Organic Carbon	17
K. Dissolved $\text{NO}_2 + \text{NO}_3$	18
L. Dissolved $\text{NO}_2$	19
M. Dissolved $\text{NO}_3$	19
4. Conclusions	19

## 1. Introduction

The Gulf of Mexico Alliance initiated its first analytical round robin on February 25, 2008 in an effort to address the needs presented by the first Action Plan to the Water Quality Priority Issue Team. In a July, 2007 amendment to the Governors' Action Plan, the Water Quality Team was charged to "Conduct a round-robin analysis to establish baseline data comparability of the water quality parameter(s) selected for Gulf-wide standardization in WQ-3:4."

At a Gulf of Mexico Alliance workshop in September 2007 the states established standard analytical methods for a core set of analytes (see Table 1) for adoption by Gulf monitoring programs.

This first analytical round robin looks at the variability in these analytes for Gulf of Mexico state and federal labs before the new standards are employed. (Future round robins will be open to any labs that sample the Gulf of Mexico waters that wish to participate and all analyses will be performed using the adopted methods.)

Table 1. Preservation methods for round robin analytes.

<b>Analyte</b>	<b>Acid preserved</b>	<b>0.45-<math>\mu</math>m filtered</b>
Chl a	No	No
BOD	No	No
CBOD	No	No
Total NO <sub>2</sub> +NO <sub>3</sub>	Yes	No
NH <sub>3</sub>	Yes	No
TKN	Yes	No
TP	Yes	No
TOC	Yes	No
Dissolved NO <sub>2</sub> +NO <sub>3</sub>	Yes	Yes
DOC	Yes	Yes
OrthoP	No	Yes

Eight state laboratories from four of the Gulf of Mexico states and one federal laboratory participated in this round robin. Two sites near U.S. EPA Gulf Breeze Laboratory in Florida were sampled on February 25, 2008. Sites were selected in an effort to provide one site that was high in nutrients and another site that was low in nutrients. However, because of heavy rains for the three days prior to sampling, both sites were diluted and had very low salinities. Field measurements from the sites are listed below. Please note that the field-measurement equipment for site B failed, so little information is available.

Table 2. Field measurements taken at each site.

	Site A	Site B
<b>Temperature</b>	17.99 °C	
<b>Salinity</b>	12.32	0
<b>Depth</b>	0.5 m	Surface water
<b>pH</b>	7.91	
<b>Turbidity</b>	~3.5 NTU (range 0.0-16.2 NTU)	

Jan Kurtz of the U.S. EPA Gulf Ecology Laboratory in Gulf Breeze, FL hosted the round-robin event. Participants included Steve Wolfe, Jessica Patronis, and Linda Sedlacek of the Florida Department of Environmental Protection and Danny Wiegand from the U.S. EPA as well as two assistants from the Gulf Breeze Laboratory.

For each participating laboratory, samples were split in the following manner:

- 1 L of unfiltered, unpreserved sample for chl a
- 1 L of unfiltered, unpreserved sample for biological oxygen demand
- 1 L of unfiltered, unpreserved sample for carbonate biological oxygen demand
- 125 mL of unfiltered, acid-preserved sample for total nitrate+nitrite, ammonia, Total Kjeldahl Nitrogen (TKN), total phosphorous, and total organic carbon
- 125 mL of 0.45- $\mu$ m filtered, acid-preserved sample for dissolved nitrate+nitrite and dissolved organic carbon
- 125 mL of 0.45- $\mu$ m filtered, unpreserved sample for orthophosphate and dissolved nitrite

For each lab, four replicates of each type were created from Site A and three replicates from site B. Samples were kept in a walk-in cooler at 4 °C until shipment. The samples were put on ice and shipped in coolers. All samples were received in good condition (i.e., still chilled).

Laboratories were given 45 days to send in their results. This paper presents the results and analyses of variability among the labs for each analyte. The names of the laboratories will not be included for the initial round robins. This round robin is not to judge how accurate a particular laboratory's results are; its purpose is to examine the degree of data comparability present to determine the potential need to standardize methods across the Gulf of Mexico. These results will later be compared to a future round robin after standardized methodology is implemented to see how much of the variability was a result of the differences in methods used.

## 2. Statistical methodology

The analyses conducted were patterned after those used in previous Florida Department of Environmental Protection round robins (e.g., Niu and Tintle, 2006; Statistical analysis and summary of the total phosphorous round robin XVI inter-laboratory comparison program) that used methodology set forth by Lin and Niu (<ftp://ftp.dep.state.fl.us/pub/labs/assessment/roundrobin/err-method.pdf>). These analyses are only possible when more than six laboratories are included in each step of the analysis. That was not always the case in this round robin. Table 3 lists the analytes and the number of laboratories that carried out each. As noted in table 3, a few analytes not listed above were analyzed (e.g., Dissolved NO<sub>2</sub> and Dissolved NO<sub>3</sub>). This addition is because some of the laboratories routinely ran these analyses in conjunction with the analytes chosen. Laboratories also ran other analyses, but the number of labs that carried out these analyses was less than four. In all cases, scatter plots show all data values for all laboratories.

Table 3. A list of the analytes measured during this round robin and the number of laboratories that ran each (n). One of the labs carried out the chl a analysis using two methods, so for chl a, n=10.

<u>Analyte</u>	<u>n</u>
Chl a	10
TP	8
OrthoP	7
NH <sub>3</sub>	7
Total NO <sub>x</sub>	6
TKN	5
BOD	4
CBOD	4
TOC	4
DOC	4
Dissolved NO <sub>x</sub>	4
Dissolved NO <sub>2</sub>	4
Dissolved NO <sub>3</sub>	4

As can also be seen in Table 3, only five of the eleven analytes that were chosen as a core group were analyzed by six or more laboratories, so only five analytes will be incorporated into the statistical analyses. Useful information can be found in scatterplots for each of the analytes listed above.

In short, the statistical analyses used are based on a one-way layout linear model. The linear model has the form:

$$Y_{ij} = \mu + a_i + \varepsilon_{ij}, i = 1, \dots, p_*; j = 1, \dots, r, \quad (1)$$

Where  $Y_{ij}$  is the result of the  $i$ th laboratory on the  $j$ th replicate at a given sample site for a given analyte,  $p_*$  is the number of participating laboratories without any outliers, and  $r$  is the number of replicates from each laboratory. Random errors ( $\varepsilon_{ij}$ ) are assumed to be independently and normally distributed with a mean of zero and variance  $\sigma^2$ .

The first step in this analysis is to identify and remove outliers. The procedure for identifying outliers is based on calculated residuals ( $\varepsilon_{ij}$ ), the upper and lower quantiles ( $Q_U$  and  $Q_L$ ), the interquartile range ( $IQR$ ), and the upper and lower outer limits ( $O_U$  and  $O_L$ ) for the sample measurements at a single site for a given analyte. The residuals of each laboratory for each site and analyte were compared to the  $O_U$  and  $O_L$ . If a residual for a given laboratory is greater than  $O_U$  or less than  $O_L$  then the sample measurement corresponding to this residual is considered to be an outlier. Any laboratory with an outlier is excluded from further analyses for that site and that analyte.

The second step in this analysis is to assess the influence of each laboratory on the linear regression model. The Cook-Weisberg distance ( $D_i$ ) was used to assess the influence of the  $i$ th laboratory.

$$D_i = \frac{(\hat{\beta}_1 - \hat{\beta})(X'X)(\hat{\beta}_1 - \hat{\beta})}{p_*s_*^2} = \frac{r(\hat{Y}_{i\bullet} - \hat{Y}_{\bullet\bullet})^2}{p_*s_*^2}, \quad (2)$$

Where  $\hat{\beta}$  is the vector-parameter estimate in the linear model based on measurements from the  $p_*$  laboratories,  $\hat{\beta}_1$  is the vector-parameter estimate without using the measurement from the  $i$ th laboratory and  $s_*^2$  is the sample variance of the experimental error terms calculated based on the residuals from laboratories without outliers. Laboratories with a Cook-Weisberg distance greater than three, which corresponds to a p-value of  $\sim 0.001$ , are considered highly influential. Laboratories with a Cook-Weisberg distance greater than 10, which corresponds to a p-value of  $2.24 \times 10^{-10}$ , are considered extremely influential. Laboratories that are considered highly or extremely influential are removed from further analyses for that site and that analysis.

The data were then transformed as needed to meet assumptions based on normality and constant variance. The remaining laboratories were then rated from 0 to 5 with 5 having average values for that site and that analyte based on their t-value. For a detailed description of how the scores were determined, see Table 4.

Table 4. An explanation of how laboratories were rated in this report.

Rating	Defining characteristics
5	Absolute t-value between 0.00 and 2.00
4	Absolute t-value between 2.01 and 4.00
3	Absolute t-value greater than 4.00
2	Cook-Weisberg distance between 3.00 and 10.00
1	Cook-Weisberg distance greater than 10.00
0	One or more outliers

### 3. Results and Discussion

#### A. Chlorophyll a

Nine labs participated in the chlorophyll a (chl a) analysis. One of the laboratories used two different methods to run their chl a analysis, so technically there are ten replicates for this analysis. Figure 1 shows a scatter plot of the results for each replicate by each lab and each site.

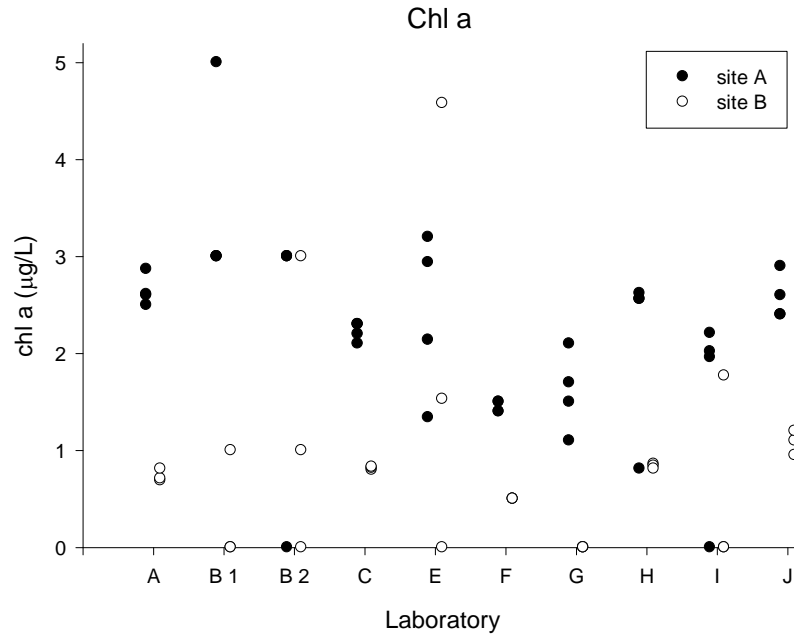


Figure 1. A scatter plot showing the variability in the measurements made for chlorophyll a by the laboratories. Values that lie below the x-axis were below the detection limit.

From Figure 1, one can see that, in general, site A had greater chl a values than site B. In some cases, the values for the two sites overlapped within a laboratory (e.g., lab E and lab I) whereas in most cases the sites were very different (e.g., lab A and B). Labs B, E, G, and I had values below their detection limit. These values were replaced with a zero for the scatter plot, but these laboratories will not be included in the analyses. Thus, only eight replicates are included in the analysis.

For site A, no values were determined to be outliers, so all laboratories remained in the analysis. No laboratories were found to be highly influential either. Most laboratories scored a 5, which indicates that the values for a given laboratory tended to be near the consensus mean. Values are given in Table 5.

Table 5. The summary statistics for chl a analyses from site A. C-W distance is the Cook-Weisberg distance for each laboratory. *t*-values are based on the average measurement with respect to the consensus mean value for that analyte and site. Scores range from 5 to 0 and were based on definitions created by Lin and Niu where a 5 indicates a laboratory that is close to the consensus mean. For more information see Table 4.

Laboratory	Mean ( $\mu\text{g/L}$ )	C-W Distance	<i>t</i> -value	Score
A	2.65	0.023	1.193	5
B (method1)	3.50	0.548	4.306	3
C	2.23	0.347	-0.336	5
E	2.41	0.161	0.319	5
F	1.45	1.953	-3.158	4
G	1.60	1.540	-2.612	4
H	2.14	0.461	0.650	5
J	2.58	0.050	0.938	5

For site B, only five replicates remained after excluding the outliers, so no analyses could be carried out.

## B. Total Phosphorous

Eight laboratories participated in the total phosphorous (TP) analysis portion of this round robin. From the scatter plot in Figure 2, one can see

that the values for site A and site B overlapped within most laboratories. The spread of values within a given laboratory was sometimes rather large ( $\sim 0.035$  mg P/L in some cases) while small in other laboratories ( $\sim 0.010$  mg P/L). Laboratory J had four values under its detection limit ( $=0.20$  mg P/L), and so it was not included in the analysis. Thus, only seven laboratories were included in the analysis.

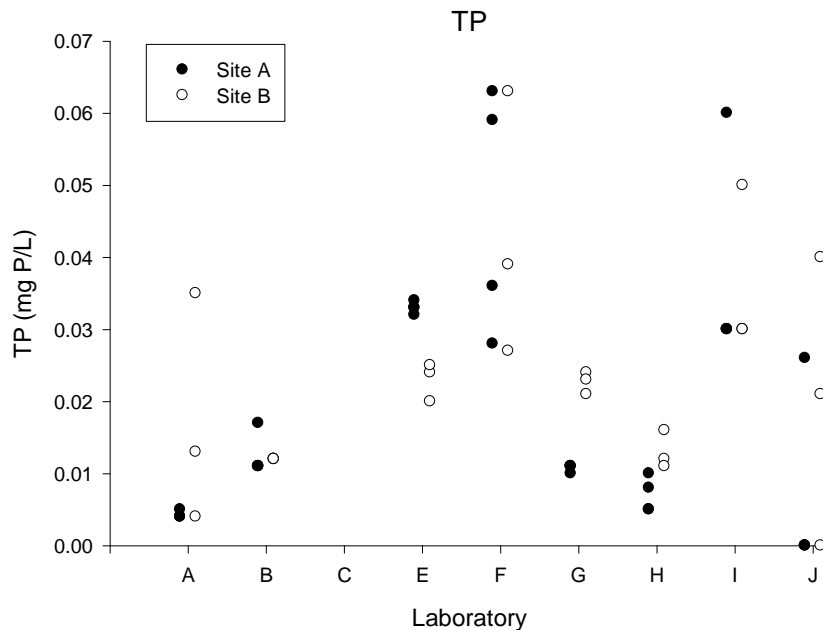


Figure 2. A scatter plot showing the variability in the measurements made for total phosphate by the participating laboratories. Values that lie below the x-axis were below the detection limit.

Tables 6 and 7 give the values for each laboratory as well as summary statistics. For site A and B, no laboratories were identified as outliers, so all seven laboratories were included in the remaining portions of the analyses.

For site A, laboratory F was determined to be highly influential in regards to the consensus mean, so it was removed from further analyses. The remaining six laboratories were scored. Scores were relatively low, ranging from 3 to 4 (laboratory F was scored as 2 because of the highly influential values). The scores are probably all low because they were equally above or below the consensus mean.

Table 6. The summary statistics for total phosphate analyses from site A. C-W distance is the Cook-Weisberg distance for each laboratory. *t*-values are based on the average measurement with respect to the consensus mean value for that analyte and site. Scores range from 5 to 0 and were

based on definitions created by Lin and Niu where a 5 indicates a laboratory that is close to the consensus mean. For more information see Table 4.

Laboratory	Mean ( $\mu\text{g/L}$ )	C-W Distance	<i>t</i> -value	Score
A	0.004	1.702	-5.344	3
B	0.013	0.472	-2.392	4
E	0.033	0.720	4.940	3
F	0.047	3.459	Highly influential	2
G	0.011	0.669	-3.019	4
H	0.007	1.177	-4.295	3
I	0.038	1.406	6.550	3

For Site B, no laboratories were found to be influential to the consensus mean. Scores were all high, with a majority of them getting a score of 5. These values were probably all high because the values within a laboratory tended to be more consistent.

Table 7. The summary statistics for total phosphate analyses from site B. C-W distance is the Cook-Weisberg distance for each laboratory. *t*-values are based on the average measurement with respect to the consensus mean value for that analyte and site. Scores range from 5 to 0 and are based on definitions created by Lin and Niu where a 5 indicates a laboratory that is close to the consensus mean. For more information see Table 4.

Laboratory	Mean ( $\mu\text{g/L}$ )	C-W Distance	<i>t</i> -value	Score
A	0.017	0.117	-0.338	5
B	0.012	0.578	-1.308	5
E	0.023	0.004	0.693	5
F	0.043	1.473	4.332	4
G	0.023	0.007	0.633	5
H	0.013	0.495	-1.147	5
I	0.037	0.657	3.180	4

### C. Orthophosphate

Seven laboratories participated in the orthophosphate (orthoP) analysis. From the scatter plot in Figure 3, one can see that four of the laboratories had values below the detection limit. As such, the number of laboratories that could be included in this analysis for Site A is only three and for Site B is five, which is too few for further analyses.

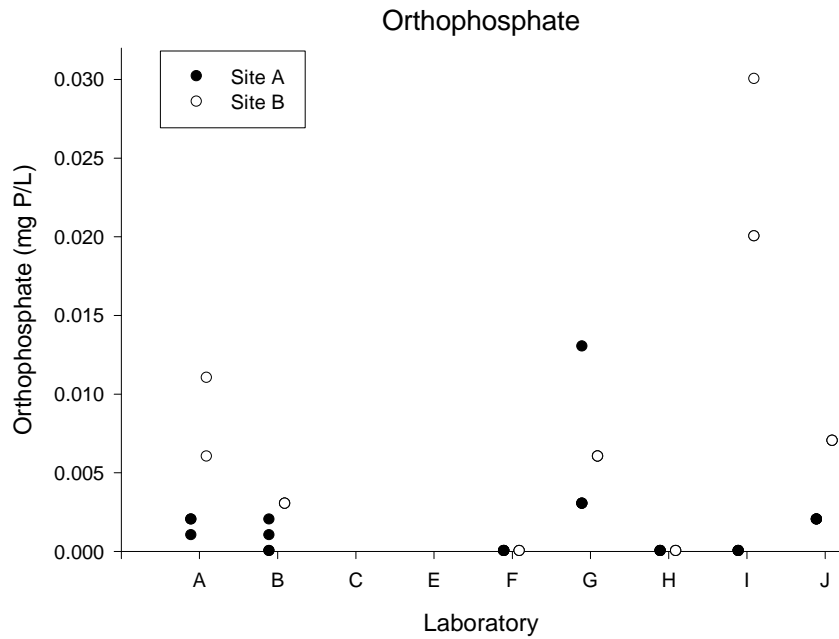


Figure 3. A scatter plot showing the variability in the measurements made for orthophosphate for each laboratory. Values that lie on the x-axis were below the detection limit.

### D. Ammonia

Seven laboratories participated in the ammonia ( $\text{NH}_3$ ) analysis. From the scatter plot in Figure 4, one can see that about half of the laboratories reported values that did not overlap between sites and the other half reported values that completely overlapped between sites. One value for lab G was below the detection limit. This value was replaced with the detection limit for the scatter plot and for the analyses because it is a single value that was close to the other data values for that laboratory.

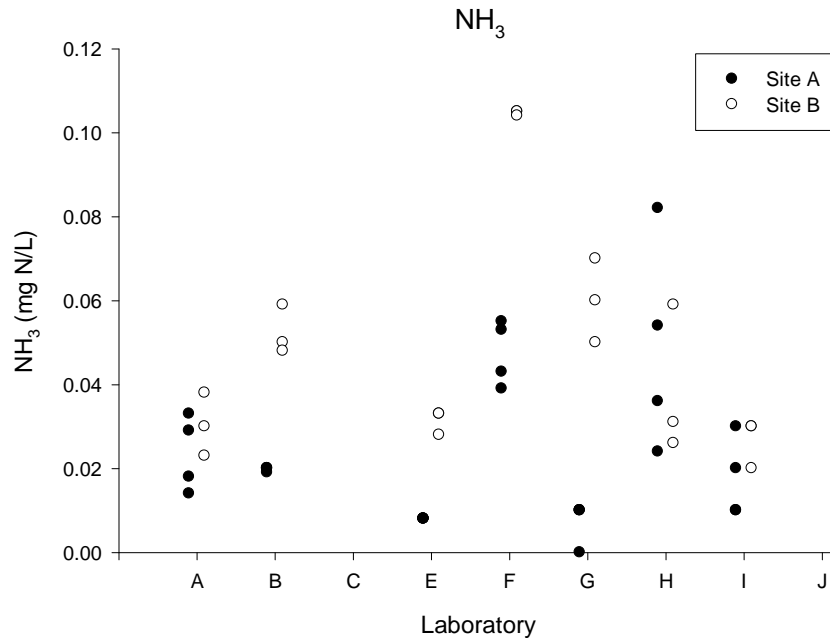


Figure 4. A scatter plot showing the variability in the measurements made for ammonia by each laboratory.

Table 8 gives the values for each laboratory as well as summary statistics for the one site that had enough replicates for each step of the analyses. For site A, laboratory H had an outlier value, so it will be excluded from further analyses. Six laboratories will remain for further analyses. For site B, no laboratories were identified as outliers, so all seven laboratories will be included in the remaining portions of the analyses.

For site A, laboratory F was found to be highly influential (Cook-Weisberg distance = 9.100). Only five laboratories remained for the remaining analyses, which is too few to continue.

For Site B, laboratory F was found to be highly influential, so it was not included in the calculation of the  $t$ -value. All remaining laboratories scored relatively well, suggesting that they the values for all of these laboratories were relatively consistent.

Table 8. The summary statistics for ammonia analyses from site B. C-W distance is the Cook-Weisberg distance for each laboratory.  $t$ -values are based on the average measurement with respect to the consensus mean value for that analyte and site. Scores range from 5 to 0 and are based on definitions created by Lin and Niu where a 5 indicates a laboratory that is close to the consensus mean. For more information see Table 4.

Laboratory	Mean ( $\mu\text{g/L}$ )	C-W Distance	<i>t</i> -value	Score
A	0.03	1.889	-1.854	5
B	0.05	0.054	2.457	4
E	0.03	1.744	-1.710	5
F	0.10	16.814	Extremely influential	1
G	0.06	0.638	3.978	4
H	0.04	0.594	-0.235	5
I	0.03	2.774	-2.636	4

### E. Total Nitrate + Nitrite

Six laboratories participated in the total nitrate + nitrite (Total  $\text{NO}_x$ ) analysis. From the scatter plot in Figure 5, one can see that the measurements between site A and B did not overlap within most laboratories (laboratory A being the exception). All laboratories were included in the analysis.

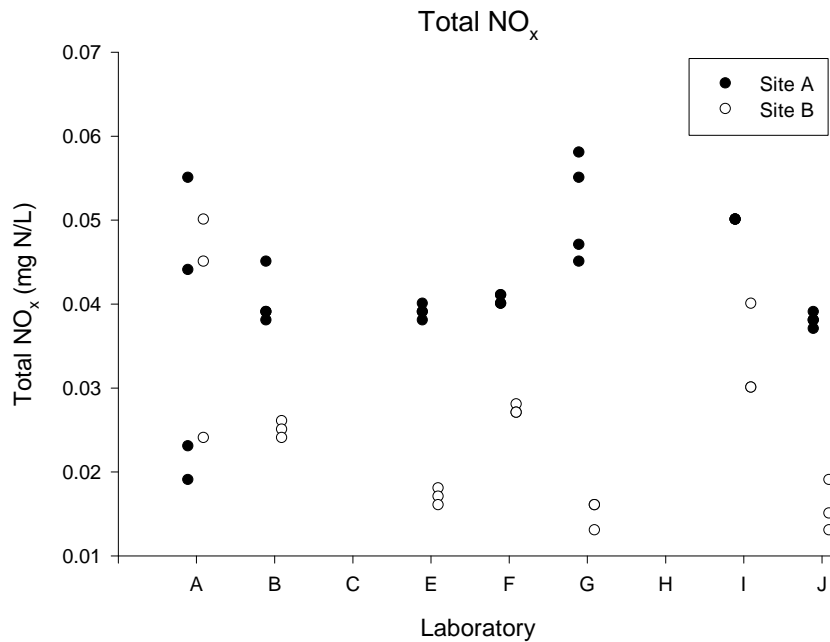


Figure 5. A scatter plot showing the variability in the measurements made by each laboratory for total  $\text{NO}_2 + \text{NO}_3$ .

For site A, laboratory H had an outlier value, so it was excluded from further analyses. Five laboratories remain, which was not enough for further analyses. For site B, laboratory A had outlier values, so it was

excluded from further analyses. Five laboratories remain, which was not enough for further analyses.

### F. Total Kjeldahl Nitrogen

Five laboratories participated in the Total Kjeldahl Nitrogen (TKN) analysis. From the scatter plot in Figure 6, one can see that within every laboratory no overlap existed between the sites. The distance between site A and site B was fairly constant between laboratories. Too few laboratories participated for any analyses to be carried out.

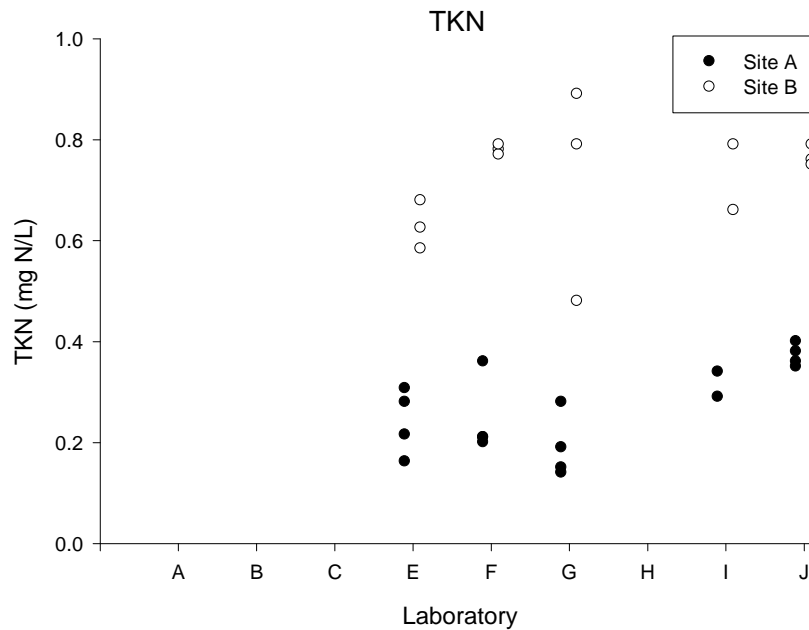


Figure 6. A scatter plot showing the variability in the measurements made by each laboratory for TKN.

### G. Biological Oxygen Demand

Four laboratories participated in the Biological Oxygen Demand (BOD) analysis portion. From the scatter plot in Figure 7, it is evident that half of the laboratories had values that were below their detection limit. This result could be a consequence of the extreme rain events. Those laboratories that did have values above their detection limit differed greatly in their values, but the sites did overlap for both laboratories. Too few laboratories participated for further analyses.

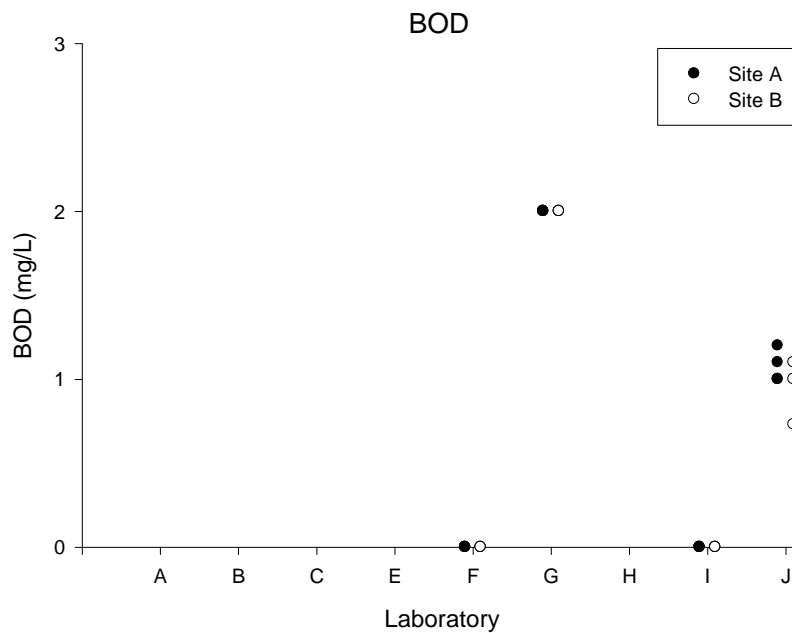


Figure 7. A scatter plot showing the variability in the measurements made by each laboratory for biological oxygen demand. Values given as zero were below that laboratory's detection limit.

## H. Carbonaceous Biological Oxygen Demand

Four laboratories participated in the Carbonaceous Biological Oxygen Demand (CBOD) analysis portion. From the scatter plot in Figure 8, it is evident that the values were below the detection limit for about half of the laboratories. The samples were likely too diluted by the extreme rain events. For the remaining two laboratories, values for the two sites overlapped and were very similar. The values between laboratories were also very similar. Too few laboratories participated for further analyses.

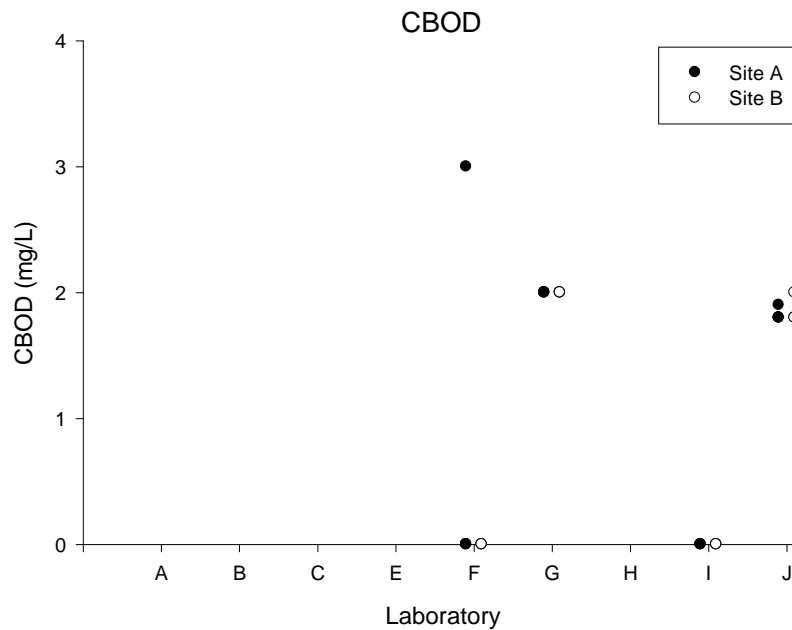


Figure 8. A scatter plot showing the variability in the measurements made by each laboratory for carbonate oxygen demand. Values given as zero were below that laboratory's detection limit.

### I. Total Organic Carbon

Four laboratories ran analyses for total organic carbon (TOC). From the scatter plot in Figure 9, one can see that site B always had higher values than site A. None of the laboratories had overlapping values for TOC. In all cases, site B had a higher TOC value than site A, and the difference between site A and site B was fairly consistent across laboratories. Too few laboratories participated in this analysis for further consideration.

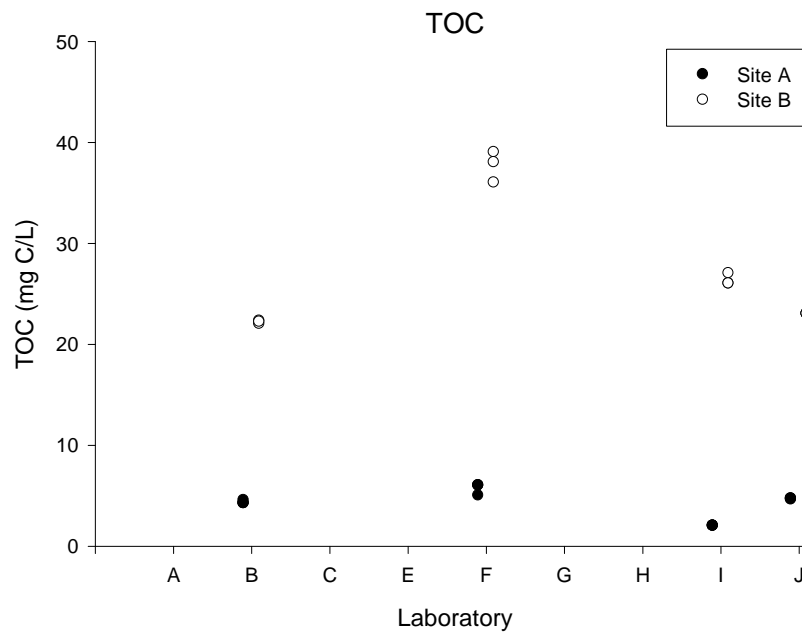


Figure 9. A scatter plot showing the minor variability in the values among the laboratories for total organic carbon.

#### J. Dissolved Organic Carbon

Four laboratories participated in the dissolved organic carbon (DOC) portion of the round robin. In figure 10, one can see that the values for DOC for the two sites did not overlap for any laboratory. As in TOC, site B had higher values for DOC than site A, and the difference between the two sites was fairly consistent across laboratories. Too few laboratories participated in this portion for statistical analyses to be carried out.

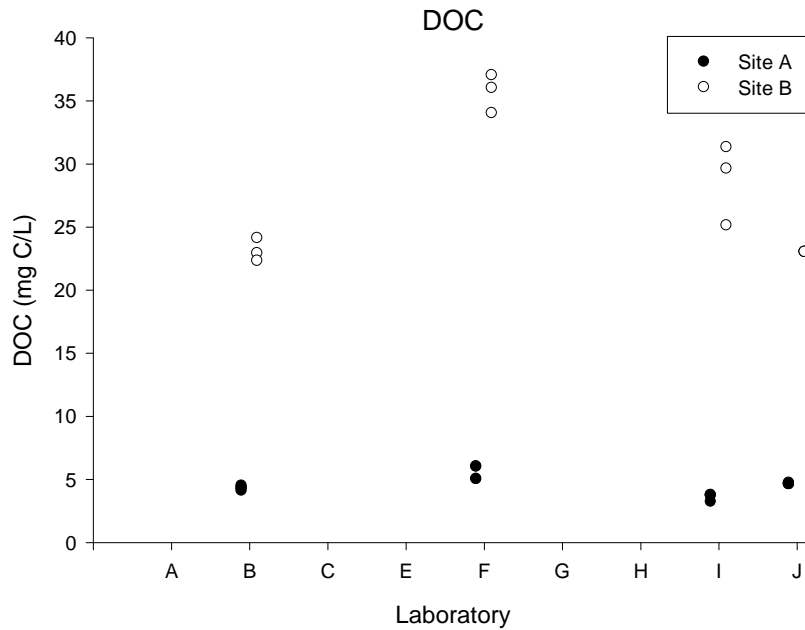


Figure 10. A scatter plot showing the minor variability in the values among the laboratories for dissolved organic carbon.

### K. Dissolved $\text{NO}_2 + \text{NO}_3$

Four laboratories analyzed for dissolved  $\text{NO}_2 + \text{NO}_3$  (Dissolved  $\text{NO}_x$ ). As Figure 11 shows, the sites did not overlap for any of the laboratories. The difference between the two sites did differ between the laboratories (0.01 to 0.025 mg N/L). However, in all cases site A dissolved  $\text{NO}_x$  was higher than at site B. Too few laboratories participated for any statistical analyses.

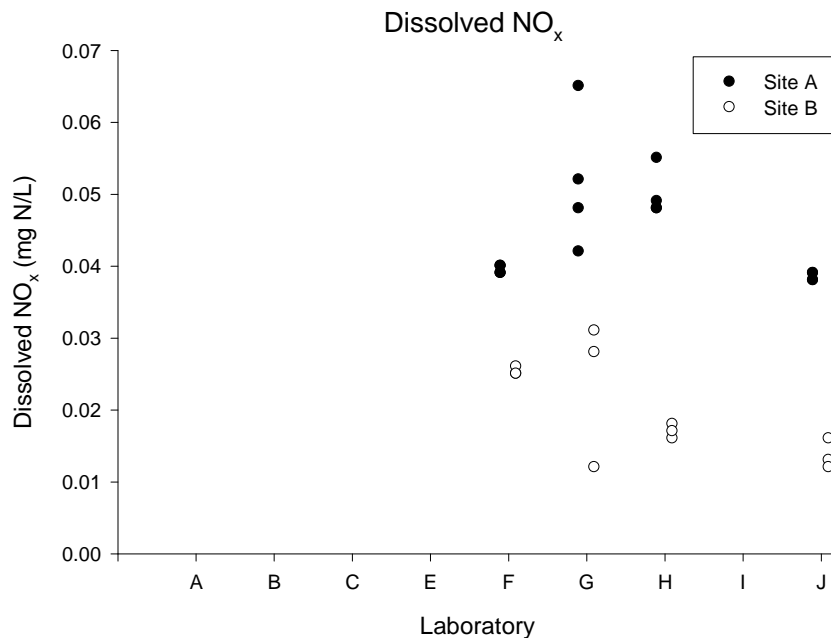


Figure 11. A scatterplot showing some variability in the values for dissolved  $\text{NO}_2 + \text{NO}_3$  among the laboratories.

#### L. Dissolved $\text{NO}_2$

Four laboratories analyzed for dissolved  $\text{NO}_2$ . As shown in Figure 12, the data is quite variable. For laboratory J, the values were below their detection limit of 0.004 mg N/L. For laboratory F, one value from site B was below their detection limit. For laboratory F and H the values for sites A and B overlapped considerably, whereas for laboratory G the values were well separated between sites. Too few laboratories participated for statistical analyses to be carried out.

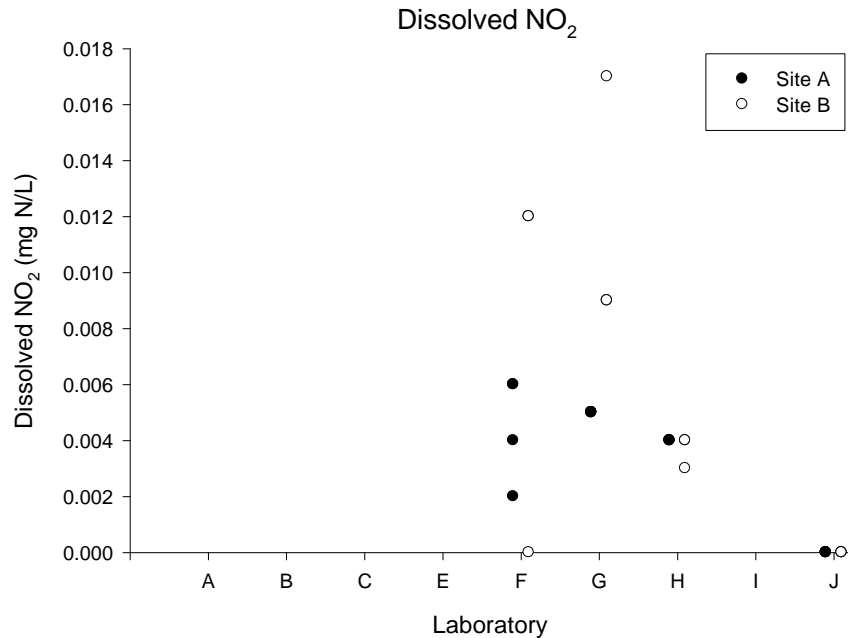


Figure 12. A scatter plot showing the variability in the values for dissolved  $\text{NO}_2$  among the laboratories. Values given as zero were below a laboratory's detection limit.

#### M. Dissolved $\text{NO}_3$

Four laboratories provided values for dissolved  $\text{NO}_x$  and dissolved  $\text{NO}_2$ . Dissolved  $\text{NO}_2$  was subtracted from dissolved  $\text{NO}_x$  to provide a value for dissolved  $\text{NO}_3$ . Because so few laboratories participated, and this values are calculated, it is not further considered here.

### 4. Conclusions

I was able to complete three analyses. Table 9 lists all of the scores and averages the scores for each laboratory. Two of the laboratories were only scored once, so no average was computed. Most average scores were

relatively high (~4), suggesting that the values measured by most laboratories were relatively close to the consensus mean. Some of the analyses clearly yielded very different results. The variability in values within laboratories was sometimes high, and the variability between laboratories was also large at times. In some analytes, values measured for sites A and B were clearly different for some laboratories while in other laboratories the values greatly overlapped. Such differences are very worrisome because the conclusions drawn from the data supplied by one laboratory would be quite different from those drawn from the data provided by another laboratory. In future round robins, we will standardize our methodology in the hopes to reduce such variability in the data values obtained across the Gulf of Mexico.

Table 10. Scores for each laboratory for the suite of analytes that could be analyzed completely. TP: total phosphate; NH<sub>3</sub>: ammonia. Letters in parentheses are the sites locales. Avg: Average; n/a: not applicable, either the laboratory did not carry the analysis out or the values were below the detection limit; --: too few scores to average.

Laboratory	Chl a (A)	TP (A)	TP (B)	NH <sub>3</sub> (B)	Avg
A	5	3	5	5	4.5
B	3	4	5	4	4.0
C	5	n/a	n/a	n/a	--
E	5	3	5	5	4.5
F	4	2	4	1	2.8
G	4	4	5	4	4.3
H	5	3	5	5	4.5
I	n/a	3	4	1	2.7
J	5	n/a	n/a	n/a	--