

Field Round Robin #1  
St. Petersburg, FL – June 4th

## **Field Round Robin #1 Report on findings**

### **Introduction**

Many groups routinely take measurements in ambient waters of the Gulf of Mexico. However, each group uses slightly different methodologies and equipment, leading to an unknown amount of variability in the data collected. This lack of data comparability has been the subject of many discussions, not just for the Gulf of Mexico. The Gulf of Mexico Alliance (GOMA) identified the need to assess this variability and to explore ways to decrease variability in the data values based solely on methodology. The GOMA chose to carry out round robins to assess the variability and then use subsequent discussions to help improve data comparability. One area that was identified as potentially having a lack of data comparability is in the data values from samples collected from a single site by different samplers and from measurements taken in the field (=field measurements). In the case of these field round robins, samplers are the focus.

For water samples, samplers split the water sample and prepared them for three different analytes. For one analyte, the water is not acid-preserved or filtered. For the second analyte, the water is acid-preserved but not filtered. For the third analyte, the water is not acid preserved, but it is filtered.

For field measurements, the samplers took the following field measurements: dissolved oxygen, temperature, salinity, conductivity, pH, turbidity, and photosynthetically active radiation.

The first field round robin was held in conjunction with the First Annual Monitoring Forum at the Florida Water Resources Institute in St. Petersburg, FL on 4 June 2008. Samplers from all five Gulf of Mexico States were represented as well as some individuals from the Regional Ambient Monitoring Program (RAMP; local county agencies that carry out their own round robins). A total of 10 sampling groups participated in the exercise. The data collected were graphed and then statistically analyzed with the same methodology used in the round robins carried out by the Florida Department of Environmental Protection and the first of the analytical round robins carried out by the Gulf of Mexico Alliance.

### **Methods**

RAMP provided a large tub (four feet tall and four feet in diameter) for sampling. The tub was filled with ambient seawater and a pump was used to keep the water well mixed. It was decided to collect water samples from the tub because then one did not have to worry about laboratories sampling water from different heights or from different water

Field Round Robin #1  
St. Petersburg, FL – June 4th

masses passing by the dock over the time that water was sampled. A sample was taken before and after the samplers took their samples to assess the contamination that might have occurred during sampling. A water bucket was provided for rinsing of equipment prior to sampling. The order of sampling was recorded so that changes over time in the concentrations of nutrients could be mapped. Each laboratory split the samples for three analytes. The water for the first analyte was neither acid-preserved nor filtered. The water for the second analyte was acid-preserved but not filtered. The water for the third analyte was not acid-preserved but it was filtered. For each analyte, three replicate bottles were filled. The water samples were then placed in a cooler and iced. The coolers were sent overnight to the Florida Department of Environmental Protection's Central Laboratory for analysis. The un-preserved, un-filtered sample was to be analyzed for nitrate, but the holding time was too short. Given the options, it was decided that alkalinity would be measured instead. The acid-preserved, un-filtered sample was analyzed for ammonia. The un-preserved, filtered sample was analyzed for orthophosphate. Participating samplers are listed in no particular order in Table 1.

Table 1. Samplers participating in round robin. Each check mark represents a participant from that laboratory participated in a given portion of the field round robin.

Agency	Sampling portion	Field measurement portion
Alabama Department of Environmental Management	✓	✓
Florida Department of Environmental Protection – SW district office		✓
Florida Department of Environmental Protection – CAMA – Big Bend Seagrasses AP	✓	
Florida Department of Environmental Protection – CAMA – Charlotte Harbor AP	✓	✓
Louisiana Department of Environmental Quality	✓	✓
Mississippi Department of Environmental Quality	✓	✓
Texas Commission on Environmental Quality	✓ ✓	✓
Environmental Protection Commission of Hillsborough County	✓	✓
Lee County Environmental Lab	✓	✓

Field Round Robin #1  
St. Petersburg, FL – June 4th

Sarasota County	✓	✓
-----------------	---	---

The samplers then participated in the field measurement portion (see list above). All samplers were asked to carry out their routine field measurements. The measurements were to be taken off of the dock (surface water and bottom water) as well as in the tub. None of the laboratories brought their equipment to take measurements for photosynthetically active radiation or turbidity. One laboratory measured LDO%.

The data values from the sampling and field-measurement portion were then graphed and the variability statistically analyzed with the same methodology used by the round robins for the Florida Department of Environmental Protection and the first analytical round robin for the Gulf of Mexico Alliance. In short, the statistical analyses used are based on a one-way layout linear model. The linear model has the form:

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

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. 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.

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

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

Field Round Robin #1  
St. Petersburg, FL – June 4th

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 2. 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

For these analyses, at least six laboratories must be included in all steps in order to complete the analyses. Also, multiple measurements must be available for each analyte. Unfortunately, field measurements are typically taken only once, and so no replicate measurements were made. Therefore, for this field round robin, only scatter plots of the data are available, but the data clearly show areas for improvement. In the future, multiple measurements will need to be made by each sampler.

Samplers were assigned letter designations. The letter designations were not changed within the sampling portion or within the field-measurement portion, but the designations did change between the two portions. The change was made because one of the groups that carried out the sampling portion split into two groups for the field-measurement portion. Also, one group only did the sampling portion and another group only did the field-measurement portion. The identities of the samplers are not revealed to others, so that samplers do not feel judged by their results. This round robin is to be utilized to help achieve data comparability and is a tool for groups to speak freely about what they are and not

Field Round Robin #1  
St. Petersburg, FL – June 4th

comfortable with in their methodology, not as a way to grade laboratories on their results.

### Results and discussion

**Sampling portion.** Each sampling team collected a sample from the tub. That sample was then treated in three different ways: un-preserved, un-filtered; acid-preserved, un-filtered; and un-preserved, filtered. Three replicates of each “treatment” were sent to the Florida Department of Environmental Protection’s Central Laboratory. The un-preserved, un-filtered sample was to be analyzed for nitrate, but because they could not make the holding time they were analyzed for alkalinity. The acid-preserved, un-filtered sample was analyzed for ammonia. The un-preserved, filtered sample was analyzed for orthophosphate. Statistical analyses were run on the samples to test for variability.

*Alkalinity.* From Figure 1, it is evident that alkalinity did not vary much within or between laboratories. The value before and after sampling was not that different (pre-sampling = 126 and post-sampling = 128). Nine sampling groups participated in this portion of the round robin.

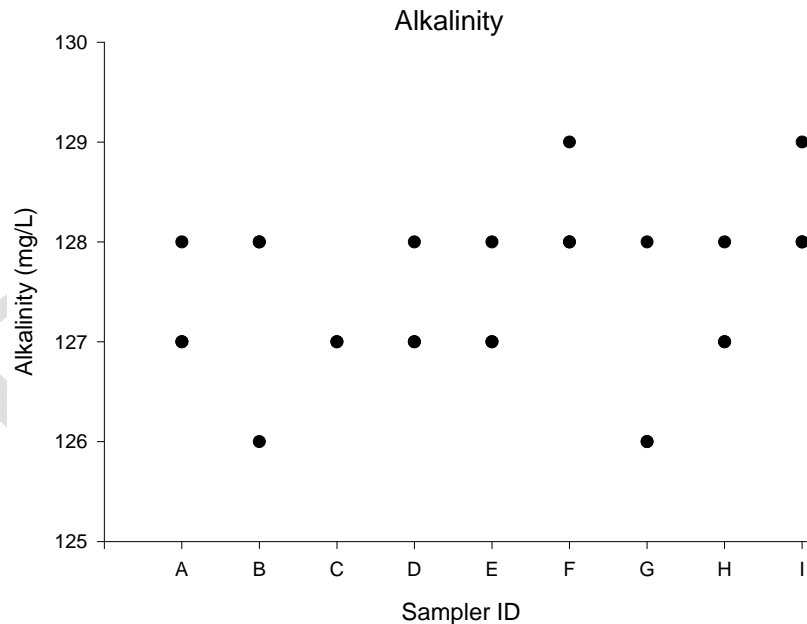


Figure 1. A scatter plot of alkalinity measured in samples taken by different sampling groups.

The results from the statistical analyses are given in Table 3. No samplers were identified as producing outlier values, and no samplers were found to be highly influential. All but one sampling group scored a 5, indicating that the values were very close to the consensus mean.

Field Round Robin #1  
St. Petersburg, FL – June 4th

Table 3. The summary statistics for alkalinity analyzed from the tub. 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 2.

Sampler	Mean (mg/L)	C-W Distance	<i>t</i> -value	Score
A	127.3	0.007	0.000	5
B	127.3	0.007	0.000	5
C	127.0	0.120	-0.721	5
D	127.3	0.007	0.000	5
E	127.3	0.007	0.000	5
F	128.3	0.480	2.863	4
G	126.7	0.367	-1.909	5
H	127.3	0.007	0.000	5
J	128.3	0.480	1.909	5

*Ammonia.* A scatter plot of the ammonia values obtained from samples collected is given in Figure 2. As you can see, multiple samples had ammonia values below the detection limit (indicated by a value of 0 on the scatter plot) of the Florida Department of Environmental Protection's Central Laboratory (=0.010 mg/L). Most samplers had relatively low values (<0.025 mg/L), but sampler A had one value of 0.038 and sampler E had values ranging from 0.046 to 0.068. The value before and after sampling was somewhat different (pre-sampling = 0.025 and post-sampling = below detection limit). Nine sampling groups participated in this portion.

Field Round Robin #1  
St. Petersburg, FL – June 4th

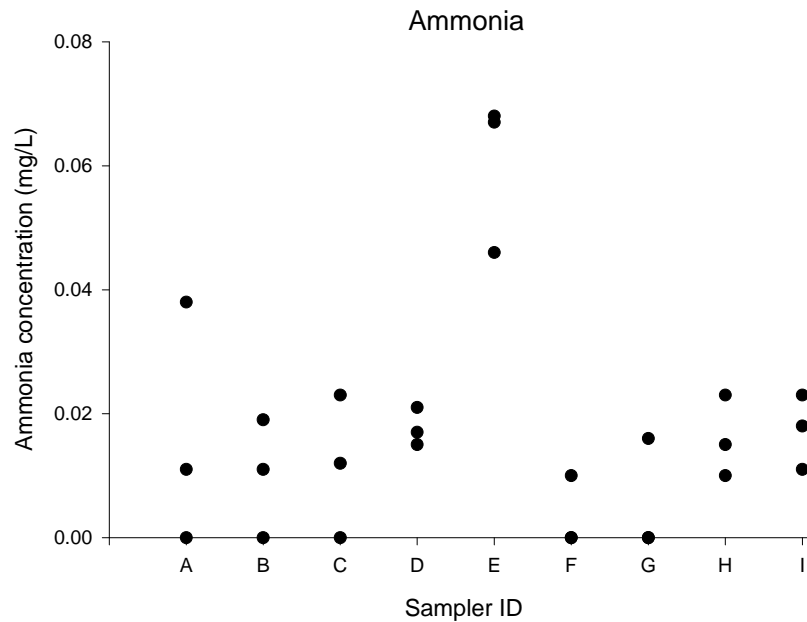


Figure 2. A scatter plot of the ammonia values measured in samples taken by different sampling groups. Values of 0 indicate that the values were below the detection limit (0.010 mg/L).

The results from the statistical analyses are summarized in Table 4. Samples that were below the detection limit were given a value of the detection limit. This method is somewhat questionable, but since the values that were above the detection limit tended to be close to 0.010 mg/L, it was decided that it was the best course of action. No samplers had ammonia values that were found to be outliers, but sampler E was found to be highly influential. As such, sampler E was omitted from further analyses for this analyte. All of the remaining samplers scored a five, suggesting that their values were very close to the consensus mean. Sampler E scored low because of the highly influential values.

Table 4. The summary statistics for ammonia analyzed from the tub. 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 2.

Field Round Robin #1  
St. Petersburg, FL – June 4th

Sampler	Mean (mg/L)	C-W Distance	<i>t</i> -value	Score
A	0.020	0.001	1.100	5
B	0.013	0.240	-0.800	5
C	0.015	0.137	-0.300	5
D	0.018	0.032	0.500	5
E	0.060	8.343	Highly influential	2
F	0.010	0.532	-1.801	5
G	0.012	0.343	-1.200	5
H	0.016	0.089	0.000	5
J	0.017	0.041	0.400	5

The Florida Department of Environmental Protection's Central Laboratory compensates for the pH of the samples when measuring ammonia concentrations. They provided a table (see Table 5) of the amount of compensation required to run the samples, and the unusually high amount of compensation that was required for sampler E's values may partially explain why sampler E was considered highly influential.

Table 5. The amount of adjustment required to compensate for too much or too little acid added to the ammonia samples. "None" indicates that it was correctly acidified. "2 H<sup>+</sup>" indicates that it was under-acidified. "3 OH<sup>-</sup>" indicates that it was over-acidified. "39 OH<sup>-</sup>" indicates that it was extremely over-acidified.

Sampler	Adjustment		
	Replicate 1	Replicate 2	Replicate 3
A	None	2 H <sup>+</sup>	None
B	2 H <sup>+</sup>	2 H <sup>+</sup>	None
C	None	2 H <sup>+</sup>	2 H <sup>+</sup>
D	2 H <sup>+</sup>	2 H <sup>+</sup>	2 H <sup>+</sup>
E	39 OH <sup>-</sup>	39 OH <sup>-</sup>	39 OH <sup>-</sup>

Field Round Robin #1  
St. Petersburg, FL – June 4th

F	None	None	None
G	None	2 H <sup>+</sup>	None
H	3 OH <sup>-</sup>	2 H <sup>+</sup>	None
J	3 OH <sup>-</sup>	None	2 H <sup>+</sup>

*Orthophosphate.* A scatter plot of the orthophosphate values for each sampler is give in Figure 3. Values did not appear to vary much either within or between samplers. The value before and after sampling were not that different (pre-sampling = 0.069 and post-sampling = 0.073). Samplers B and C did not participate in this portion, so only seven samplers participated.

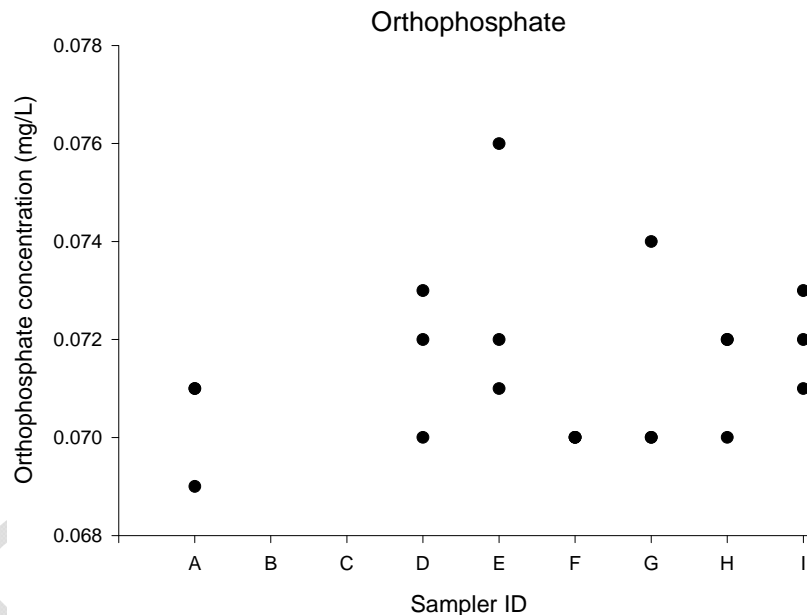


Figure 3. A scatter plot of the orthophosphate values measured from samples taken by different sampling groups.

A summary of the results of the statistical analyses are in Table 5. None of the samplers were found to have ammonia values that were outliers, and the samplers were not found to be highly influential to the consensus mean. All samplers except for Sampler F had a score of 5, suggesting that the values were all very close to the consensus mean.

Table 5. The summary statistics for orthophosphate analyzed from the tub. C-W distance is the Cook-Weisberg distance for each laboratory. *t*-values are based on the average measurement with respect to the

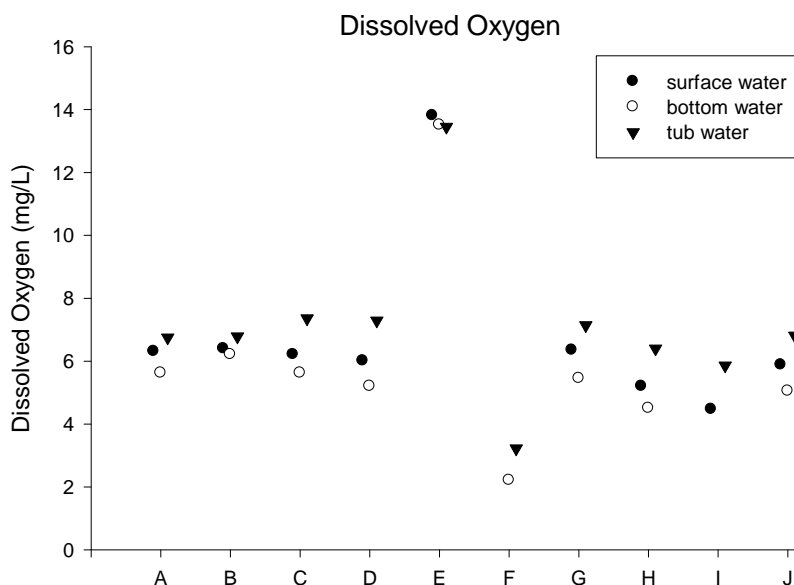
Field Round Robin #1  
St. Petersburg, FL – June 4th

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 2.

Sampler	Mean (mg/L)	C-W Distance	<i>t</i> -value	Score
A	0.070	0.180	-1.888	5
D	0.072	0.013	0.000	5
E	0.073	0.429	1.888	5
F	0.070	0.312	-2.359	4
G	0.071	0.000	-0.472	5
H	0.071	0.000	-0.472	5
J	0.072	0.063	0.472	5

**Field-measurement portion.** The following measurements were taken: dissolved oxygen, temperature, conductivity, salinity, and pH. Only one measurement was taken in a given location, so no statistical analyses could be carried out. However, scatter plots of the data are provided and will be discussed.

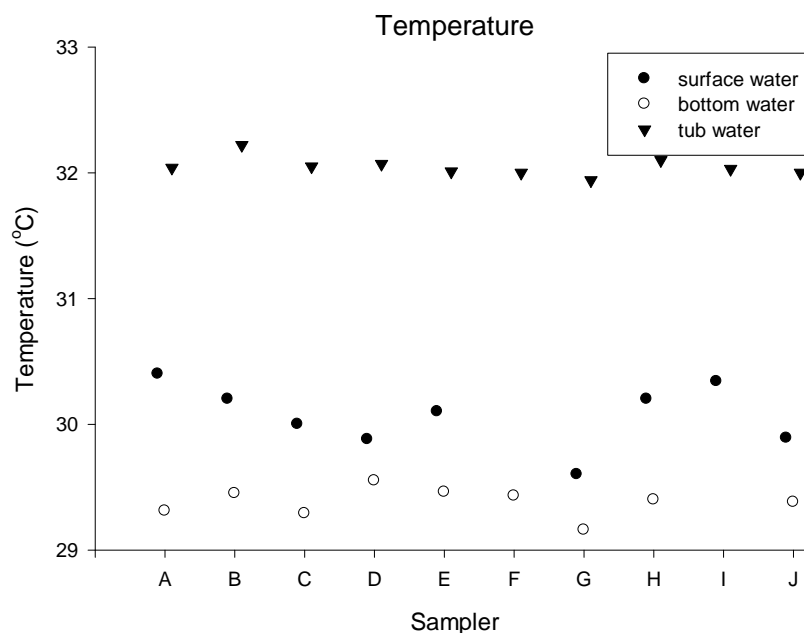
*Dissolved Oxygen.* A scatter plot of the values obtained for dissolved oxygen is given in Figure 4. In general, the value in the tank is the highest and the value for the bottom water is lowest for most samplers. Most samplers appear to have obtained values around 6 mg/L. However, sampler E has an unusually high value of ~13.6 mg/L, and sampler F has a slightly lower average value of ~2.7 mg/L. These values suggest that work towards data comparability is required to reduce the variability.



Field Round Robin #1  
St. Petersburg, FL – June 4th

Figure 4. A scatter plot of dissolved oxygen values obtained by ten samplers. The diamonds are values measured in surface waters; the squares are values measured in bottom waters, and the triangles are values measured in the tank. Sampler F did not take a measurement from surface waters, and Sampler I did not take a measurement from bottom waters.

*Temperature.* Figure 5 is a scatter plot of temperature values. Temperatures within the tank (31.94°C to 32.22°C – a difference of 0.28°C) and from the bottom water (29.16 to 29.55°C – a difference of 0.39°C) was relatively consistent between samplers. However, variability did exist in temperature values obtained from surface water (29.60 to 30.40°C – a difference of 0.80°C). These differences could have been caused by micro-scale differences in the actual temperatures, but they could have also been caused by differences in the depth of the measurement. Temperature varies greatly with depth, the depth of the measurement may explain some of the variability. In the future, it would be beneficial to standardize how one determines what depth to measure when doing surface and bottom water measurements. However, the basic pattern that tank temperature was much higher than the other values and that bottom water temperature was lower than the other values held true for all samplers.



Field Round Robin #1  
St. Petersburg, FL – June 4th

Figure 5. A scatter plot of temperature values obtained by ten samplers. The diamonds are values measured in surface waters; the squares are values measured in bottom waters, and the triangles are values measured in the tank. Sampler F did not take a measurement from surface waters, and Sampler I did not take a measurement from bottom waters.

*Conductivity.* Conductivity values are given in Figure 6. For all samplers, the conductivity measured in bottom water and the tank was similar whereas conductivity from surface water was lower. Sampler A had unusually low conductivity values. For the other samplers, surface water values ranged from 45,971  $\mu\text{s}/\text{cm}$  to 48,698  $\mu\text{s}/\text{cm}$ , a difference of 2,727  $\mu\text{s}/\text{cm}$ ; bottom water values ranged from 47,470  $\mu\text{s}/\text{cm}$  to 49,408  $\mu\text{s}/\text{cm}$ , a difference of 1,938  $\mu\text{s}/\text{cm}$ ; and tank values ranged from 47,558  $\mu\text{s}/\text{cm}$  to 49,678  $\mu\text{s}/\text{cm}$ , a difference of 2,120  $\mu\text{s}/\text{cm}$ . Once again, variability was greatest in the surface water measurements, suggesting that either small-scale variability was a factor or that differences in the depth measured could be important factors to consider in the future.

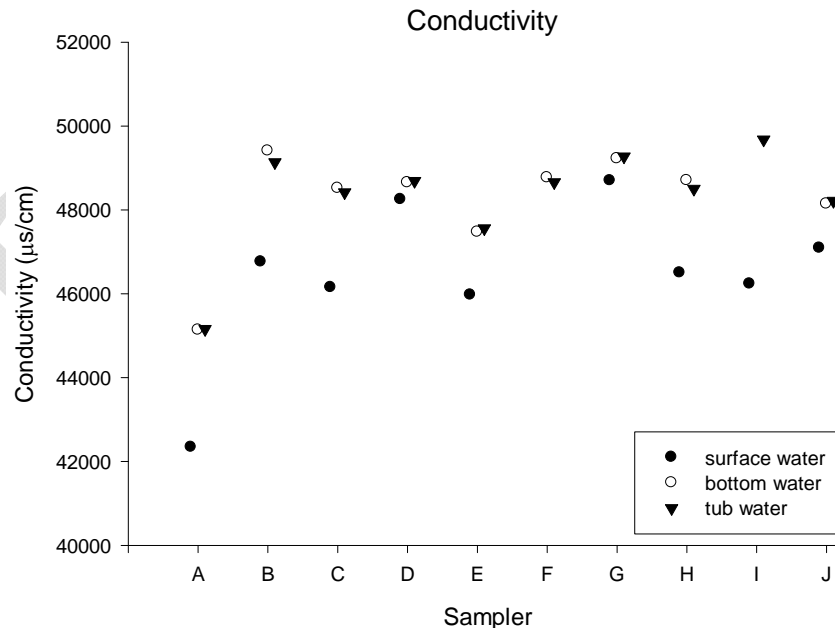


Figure 6. A scatter plot of conductivity values obtained by ten samplers. The diamonds are values measured in surface waters; the squares are values measured in bottom waters, and the triangles are values measured in the tank. Sampler

Field Round Robin #1  
St. Petersburg, FL – June 4th

F did not take a measurement from surface waters, and  
Sampler I did not take a measurement from bottom waters.

*Salinity.* Figure 7 is a scatter plot of salinity values for the different samplers. Similar to conductivity, salinity values were generally the same for bottom water and the tank and much lower for surface water. Sampler A had unusually low salinity values again. For the remaining samplers, surface water sample values ranged from 29.61 to 31.62, a difference of 20.1; bottom water sample values ranged from 30.77 to 32.15, a difference of 1.38; and tank values ranged from 30.74 to 32.29, a difference of 1.55. As in salinity and temperature, the variability in the surface water samples is the greatest, supporting the suggestions above that small-scale differences and depth of measurement are important components of the variability.

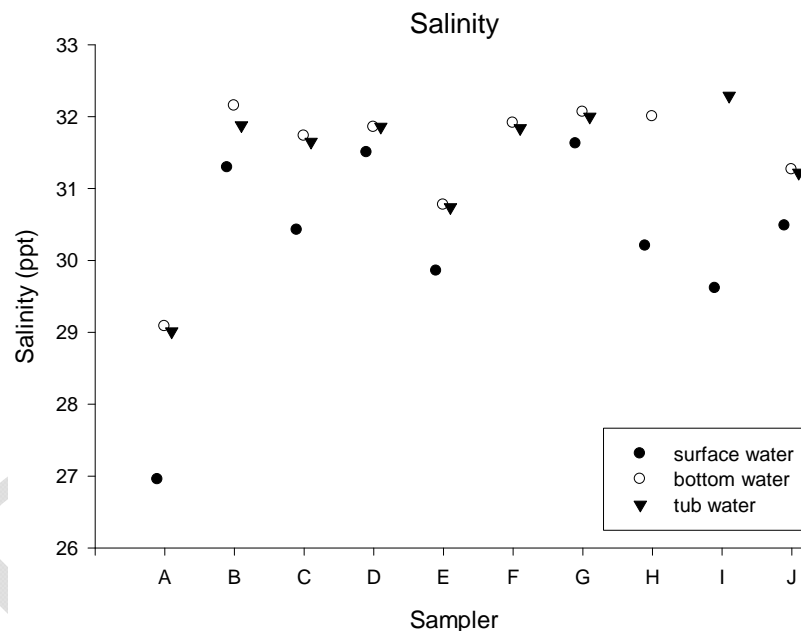


Figure 7. A scatter plot of salinity values obtained by ten samplers. The diamonds are values measured in surface waters; the squares are values measured in bottom waters, and the triangles are values measured in the tank. Sampler F did not take a measurement from surface waters; Sampler H did not record this measurement from the tank, and Sampler I did not take a measurement from bottom waters.

*pH.* pH values are shown in Figure 8. Samplers are listed roughly in the order in which they took their measurements. Oddly enough, the values that they recorded suggest a decrease of 0.34 pH units over an ~30 minute time period, which does not make sense. I suspect that this pattern only showed up by chance. However, Samplers E and F have

Field Round Robin #1  
St. Petersburg, FL – June 4th

unusually low pH values. In general, tank water pH was higher than the other values, but no pattern exists for whether surface water or bottom water was higher in pH. It appears that the Gulf of Mexico Alliance needs to continue to work towards improving data comparability for pH measurements.

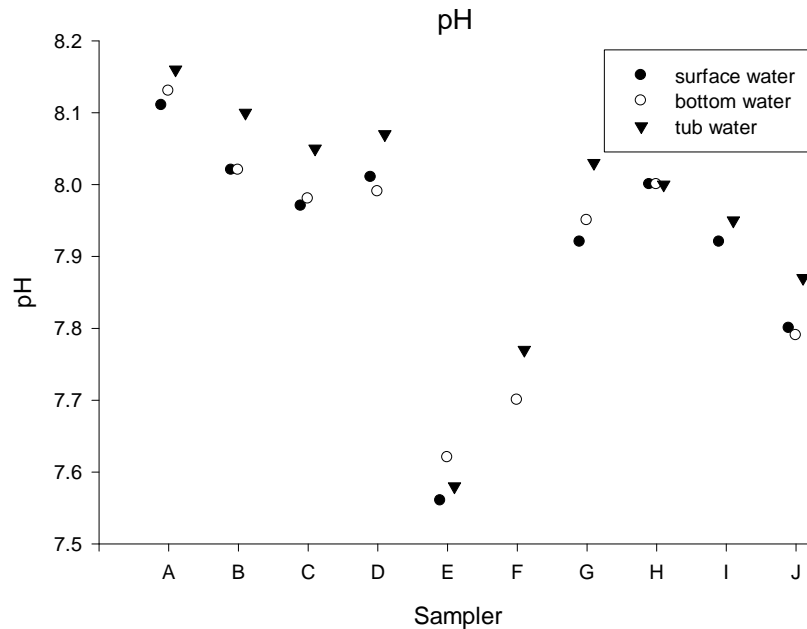


Figure 8. A scatter plot of pH values obtained by ten samplers. The diamonds are values measured in surface waters; the squares are values measured in bottom waters, and the triangles are values measured in the tank. Sampler F did not take a measurement from surface waters, and Sampler I did not take a measurement from bottom waters.

## Conclusions

**Sampling portion.** Nine different sampling groups participated in the first field round robin. The samples that they collected and prepared for analytical analysis yielded somewhat different results. The un-preserved, un-filtered samples and the un-preserved, filtered samples were very close to the consensus mean, suggesting that little bias was introduced by sampling technique for these samples. However, the un-preserved, un-filtered samples were analyzed for alkalinity, and that analyte is not as easily influenced by contaminants that might exist in the sampling containers. In the future, a nutrient, such as nitrate, will be analyzed instead.

The acid-preserved, un-filtered samples yielded very different results. One sampling group was found to be highly influential, and this result

Field Round Robin #1  
St. Petersburg, FL – June 4th

may be because of the sample being extremely over-acidified. How to determine the amount of acid to add to a sample should be discussed.

**Field-measurement portion.** Ten different sampling groups participated in this portion. (One of the sampling groups split into two for this portion.) The sampling groups measured dissolved oxygen, temperature, conductivity, salinity, and pH. Only one measurement of each analyte was taken at each location, so no statistical analyses could be carried out. In the future, three to four measurements will be taken at each location. All discussions are based on observations from the scatter plots. For dissolved oxygen, most of the samplers had very similar values. The exceptions were samplers E (with an extremely high value) and F (with a somewhat low value). Temperature was fairly constant across sampling groups. Conductivity and Salinity was fairly constant across samplers, but sampler A had unusually low values. pH was fairly constant across samplers with the exception of samplers E (with an extremely low value) and F (with a somewhat low value).

It was later announced that some of the samplers had not been able to calibrate before they took their measurements. It would be prudent to allow time for the samplers to calibrate before they take their measurements, and to provide a location for qualifiers on the sampling sheet.

Also, one pattern that did present itself in the field-measurement portion of this round robin was that the surface waters had an unusually high variability in their temperature, conductivity, and salinity values relative to those taken from the bottom water or from the tank. This variability could be from the small-scale variability or it could be from differences in how samplers decide what depth to measure. A discussion of how samplers decide what depth to measure at would be wise.