Appendix 5C-4: Evaluation of Sampling Methods for Total Phosphorus

Pete Rawlik

INTRODUCTION

Environmental monitoring data contains variability generated by inherent natural variation from an aquatic environment, small-scale chronic interferences with sampling, and rare, larger-scale acute spikes caused by sample entrainment of materials not normally found in ambient waters. Smaller scale interferences are associated with local runoff from water management features, wildlife activities, or vegetation impacts and they can be difficult to separate from natural background noise. These frequent events become more influential as concentrations approach low background levels; 5 micrograms per liter (µg/L) added at 100 µg/L marsh level is not materially influential, but the same extraneous amount added at 15 µg/L represents a 33 percent interference. In addition to these three types of data variation associated with water samples themselves, these are also systematic data errors associated with the methods used to take and process samples. The project summarized here attempts to gather specific information on potential sources of variation linked to water quality samples and sampling systems. This information can eventually guide sampling improvements so that acute spikes and chronic interferences will not ordinarily cause unexplained variation in water quality data.

The South Florida Water Management District (SFWMD or District) operates multiple stormwater treatment areas (STAs), which are constructed wetlands designed to remove total phosphorus (TP). The collection of representative samples is critical to evaluating the true performance of the STAs. Surface water sampling must provide representative data as free from interferences as possible, particularly at low TP concentrations expected from the STAs. Unexplained variability from various forms of extraneous materials collected in the samples themselves or generated by flaws in the sampling process must be understood before they can be eliminated. Most of the *Science Plan for the Everglades Stormwater Treatment Areas* (SFWMD 2013b) projects seek to improve STA performance; this study seeks to understand factors influencing sample representativeness used to assess performance. Using multiple sampling methods, cameras, and probes, sampling interferences and malfunctions were documented in this **Remote Environmental Sampling Test Project (Project REST)** (Rawlik 2016). Information gathered from the study can then be used to refine sampling protocols and methods to improve the quality of data used for tracking STA performance.

Four sampling methods were compared in Project REST: (1) grabs, samples that represent a discrete point in time and are not automated; (2) autosampler composite flow (ACF), in which autosamplers are set to collect for specific volumes of flow through a structure and composited into one bottle for single week; (3) autosampler discrete time (ADT), in which autosamplers are set to collect at a set frequency and composited into one bottle representing a single day; and (4) a remote phosphorus analyzer (RPA). The RPA combines automated sampling equipment with an automated micro-laboratory and is able to collect large quantities of TP data at specific locations for extended periods of time. The ACF method creates flow-weighted data and can be used to calculate compliance for the water quality based effluent limitation

(WQBEL) pertaining to the STAs. These methods together provided comparative information on potential interferences both with water samples and with the sampling process being used.

In many instances, all methods can be expected to produce very similar long-term data. However, significant short-term deviations in data from grabs and autosamplers have been documented and these have the potential to impact estimates of concentrations, associated flows, and water quality performance. Previous technical analyses have provided little insight into the causes of such differences. Unexplained variation in TP data tends to produce bias on the high side because in the STA discharges and many other Everglades sites, the concentrations of TP are skewed to the right meaning that most values are low and a small proportion are much higher producing the skewed distribution. Furthermore, many values are relatively close to background levels (circa 10 µg/L) and interferences produce higher values. These higher concentrations cannot usually be counterbalanced; samples with very low or no TP simply do not occur in STA marsh water. Therefore, the influence of high values from acute or chronic interferences cannot be counterbalanced, forcing upward bias in the TP data set used to assess STA performance. Project REST seeks to gather information to minimize these sampling interferences.

METHODS

The REST study leveraged TP concentration data streams from four sampling methods with varying assumptions and levels of complexity: existing weekly grabs; existing ACF samplers; new ADT samplers collecting every two hours for discrete daily bottles; and RPA, which analyzed samples every two hours. An additional method, of taking grab samples through the autosampler pumps weekly was used to validate the equipment installation and assure that the equipment was not contaminated. All sampling followed standard District field and lab protocols and was conducted for a period of 34 weeks during 2014. Monitoring was supplemented with three cameras: one directly above the equipment intake screens looking at the surface of the water (**Figures 1** and **2**) one underwater and focused on the sampling equipment intake screens (**Figures 3** and **4**), and, one far afield, looking at the sampling platform as a whole. Probes were also installed to monitor temperature, pH, specific conductance, and turbidity.

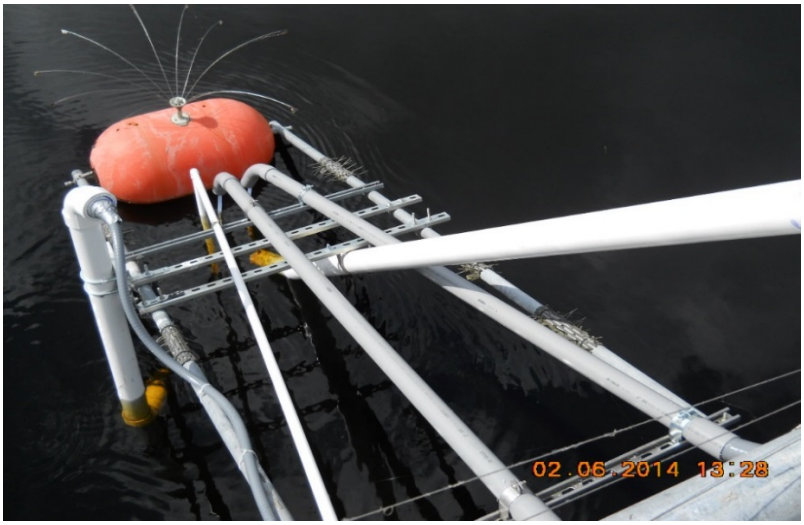

Figure 1. Sampling equipment at G310 mounted on a floating armature.

Figure 2. Sampling equipment at G390B mounted on the structure wing wall and demonstrating anhinga feces collecting on floating barriers.

Figure 3. The sampling equipment at G310 including, from left to right, intakes for the RPA, ACF, and ADT. The diagonal, slotted pipe houses in situ probes, including temperature, pH, specific conductance, and turbidity.

Figure 4. The sampling equipment at G390B including, from left to right, intakes for the ACF, RPA, and ADT. The pipe housing the in situ probes, including temperature, pH, specific conductance, and turbidity, is off screen to the right.

This array of equipment was operated at two structure locations. The first location was G310, located at structure G-310, a large pump station and the primary outflow from STA-1 West (STA-1W) that is comparable to discharge structures at STA-1 East (STA-1E) and STA-2. The second location was G390B, located at G-390B, a small culvert interior to STA-3/4 associated with the Periphyton-based Stormwater Treatment Area (PSTA) Project. This location is comparable to discharge structures at STA-3/4 and STA-5/6. Quality assurance and routine maintenance was performed in accordance with the SFWMD *Field Sampling Quality Manual* currently in place at that time (SFWMD 2014). All samples were analyzed by the District's laboratory (SFWMD 2013a).

RESULTS: POTENTIAL SAMPLING INTERFERENCES

Underwater cameras documented suspended particles at both locations. At G310, the suspension of particles was linked to the operation of the pump station, with more particles being mobilized at higher flow velocities. At G390B, particles were seen on the surface caught in eddies of the culvert (**Figure 5**). Floating aquatic vegetation and uprooted plants were observed by cameras at both locations, but dominated at G310 where they often encircled the sampling intakes, sometimes for days at a time (**Figure 6**). These events did not appear to increase TP concentrations. Fresh submerged aquatic vegetation may be too cohesive to contaminate the sampler and could even serve as a filters, preventing particles from entering the sampler.

Figure 5. Ashy material floating on the water column at G390B, material is encircled in red.

Figure 6. Vegetation wrapped around the sampling equipment at G310, compare to **Figure 3**.

Wildlife was found to be a potential source of sampling interference. Visits by turtles were common and they often interacted with the equipment, creating clouds of detritus. Vultures roosted overnight at G390B, leaving fecal deposits that would work their way through the grating and into the water column below. Small flocks of cattle egrets (*Bulbulcus ibis*) and various herons were observed, but did not seem to cause significant impacts. On the other hand, anhingas (*Anhinga anhinga*) were often seen swimming, hunting, feeding, and defecating around structures and sampling equipment. Their feces are a potential acute problem when directly deposited in the water column in front of sampling equipment (**Figure 7**), or a longer-term chronic problem when accumulated feces on infrastructure (**Figure 2**) are washed off the structure, sampling equipment, or floating vegetation barriers by rain. In any case, recent studies in the Everglades found high concentrations of TP in wading bird feces ranging from 21to 57 grams per kilogram and this range may be conservative based on other data (2 to 8 percent) for white ibis (*Eudocimus albus*) and wood storks (*Mycteria americana*). Because the entire levee and associated infrastructure serves as a wildlife attractor, corridor, and habitat, observed wildlife processes likely turn the levees into functional non-point sources of TP. Note that, while wildlife have the potential to interfere with collection of a representative sample as documented here, no extreme interferences or spikes were apparent in this study. This suggests the chances of collecting an extreme aberrant value are relatively low, yet smaller magnitude interferences may be more commonplace and difficult to separate from normal water quality ambient variations.

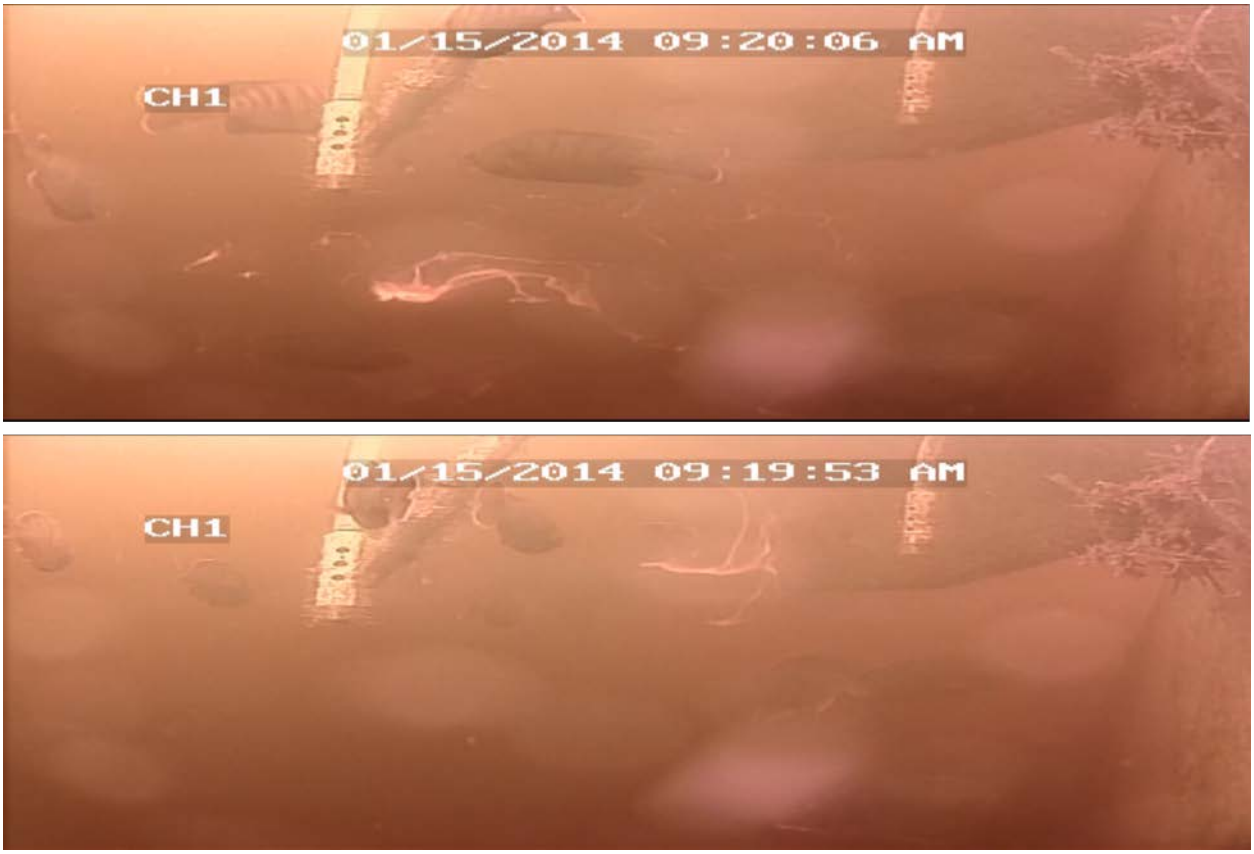

Figure 7. Anhinga feces dissolving into the water column near the intakes of the sampling equipment at G390B.

RESULTS: INTERFERENCES WITH SAMPLING METHODS

An analysis of sampling completeness found large differences between methods. At G310, grab samples and time-proportional samples properly collected greater than 99 percent of the expected samples. In contrast, the RPA collected 76 percent of expected samples and failures were usually associated with the pump. Analysis of the entire year of flow-based ACF results found that the method functioned properly for 90 percent of the time, but only represented 65 percent of the flow. This highlights the fact that flow is not evenly distributed over the year and that short-term failures at critical times strongly bias the resulting data sets. When actually used to calculate annual flow-weighted means, such missing ACF data would be replaced with the results of grab sampling. Results from grab samples are considered less than optimal as they are a single data point representing a week of flow.

A completeness analysis at G390B found that the grab samples and the ADT (timed autosamplers) collected 100 percent of the expected samples, while the micro-lab RPA collected 71 percent with losses once again being attributed to pump failures. Initial analysis found that the flow-based ACF collected 100 percent of the samples expected, but there were other problems. The ACF was generating results for weeks with no flow or even negative flow and did not appear to be triggering at the correct flow volume. Investigation found that a change in structure size was recorded in the District's standard flow database, but not in the database used to trigger the sampler. As a result, the sampler program was overestimating flow. Another issue with the ACF had to do with back flow, which is possible at open culverts such as G390B. In the flow database, once a volume of water has passed through the structure it is recorded as a positive value. Unfortunately, if that volume of water backflows, the volume in the database is not reduced.

Regardless of how much negative flow occurs, only positive flow is counted toward the flow trigger. Thus, the sampler responds to this overestimation of flow. Further investigation of the triggering systems at other structures is warranted.

A final issue at G390B concerned noise (fluctuations in equipment that give the appearance of flow) that is generated in preliminary flow data that the trigger database interprets as flow. Through the quality assurance process, this noise is often identified and reset to zero. Unfortunately, this has no impact on the trigger database and samples cannot retroactively be eliminated. When entire weeks of data are set to zero, the impact of this 'phantom flow' may be negligible, but if only portions of the week are impacted, too many samples are collected at inappropriate times. While this may have a limited impact on overall annual results, it does impact weekly values and could influence operational decisions. More importantly, it highlights the fact that the flow-proportional system is prone to sampling errors at locations with small headwater to tailwater differences.

Deployed probe data were analyzed for completeness at both sites. More than 85 percent of the expected data were collected and missing data was attributed to either programming errors or power failures. This is less than the SFWMD completeness target of 95 percent, but given the probes are deployed and maintained on a weekly basis, even small temporal losses of data might be considerable.

Analysis of water quality data showed that the autosampler pump and tubing were not a factor driving data differences. Analysis of data at both sites confirmed a previously reported diel pattern in STA TP concentrations. Over the course of a day, concentrations increased by up to 25 percent, tending to be higher at midday. The RPA data also demonstrated that the TP data distribution can be different based on structure type. The RPA data at G310 is skewed to the right (**Figure 8**) with the peak (17–21 µg/L) representing 58 percent of the data, and higher values accounting for 35 percent, and leaving only 7 percent of results lower than the peak. In contrast, the distribution of RPA data at G390B (**Figure 9**) showed no clear peak, but rather a plateau from 14 to 34 µg/L accounting for 79 percent of the data. As a result of these data differences, an interference derived value of 40 µg/L would influence the average value more at G310.

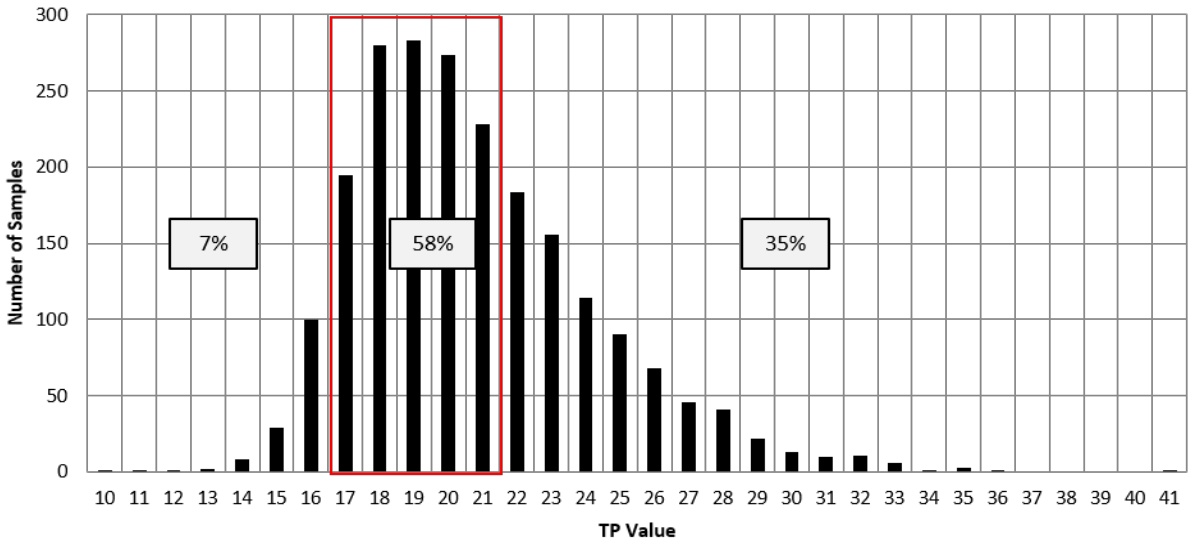

Figure 8. Frequency distribution of RPA TP values at G310 during 34 weeks of sampling at a two-hour frequency.

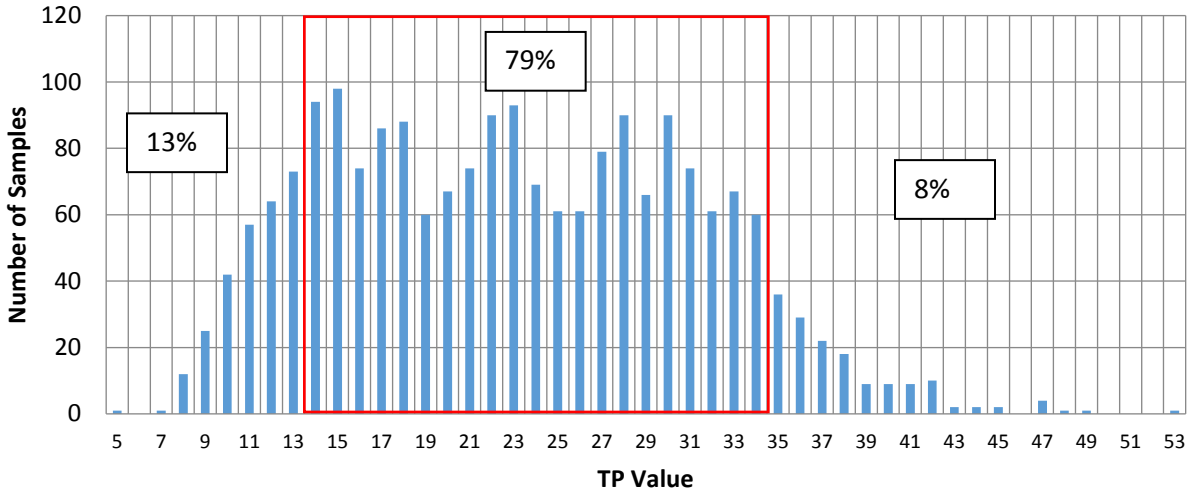

Figure 9. Frequency distribution of RPA TP values at G390B during 34 weeks of sampling at a two-hour frequency.

A comparison of loads calculated using RPA data and multiple substeps in time found that time-based sampling was comparable to flow-based sampling for deriving loads It had been expected that this calculation process would allow extraneous high TP events in short-term flows to be isolated and their impact minimized, but no significant difference in the results was found. Similarly, the weekly results from the flow-based ACF were compared to the weekly means of the ADTs (**Figures 10** and **11**), as well as the flow-weighted means derived from the daily ADTs and daily flow values. In both cases, results were again highly comparable. Statistical analysis found that the data was normally distributed but heteroscedastic ($p = 0.537$). An analysis of variance (ANOVA) on ranks found no significant difference ($p = 0.432$). When equal variance is assumed the data is still not significantly different using a paired t-test ($p = 0.12$ for both the ADT and flow-weighted ADT data). These findings suggest that the more complex ACF process could be replaced with ADT data with a concurrent increase in completeness and cost-effectiveness.

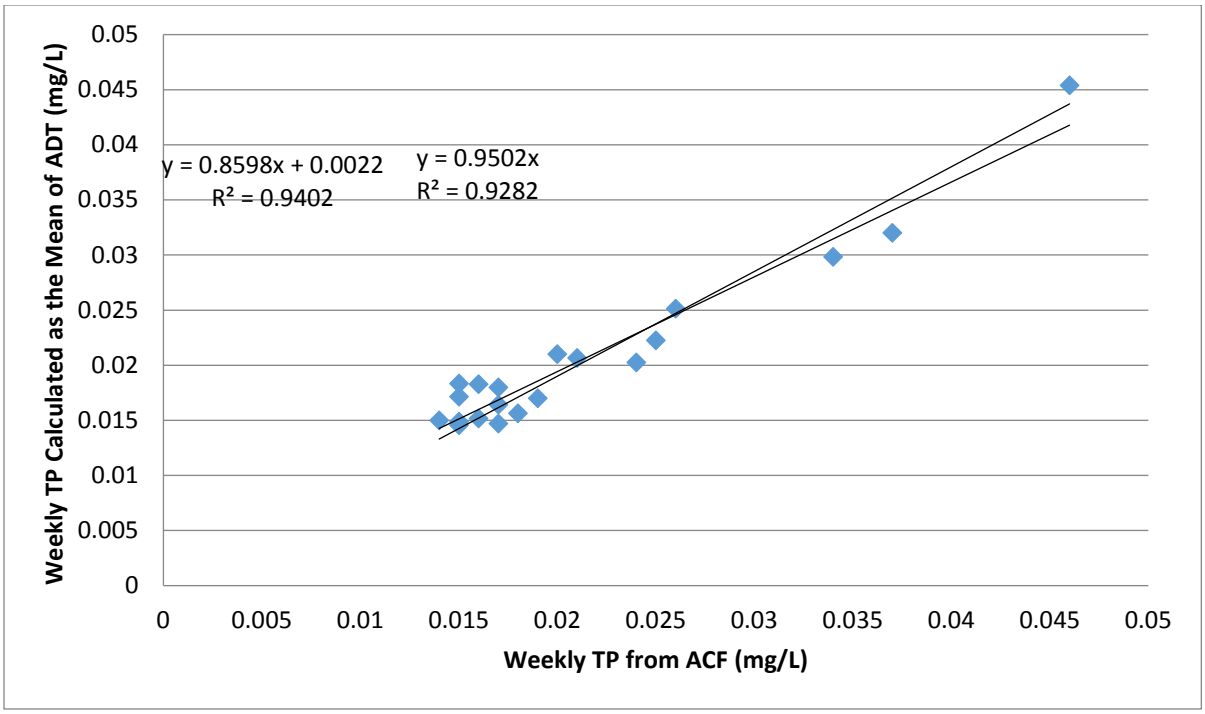

Figure 10. Weekly TP results from ACF and mean ADT results with one outlier removed. Two equations are presented: one forced through zero and one not. (Note: mg/L – milligrams per liter.)

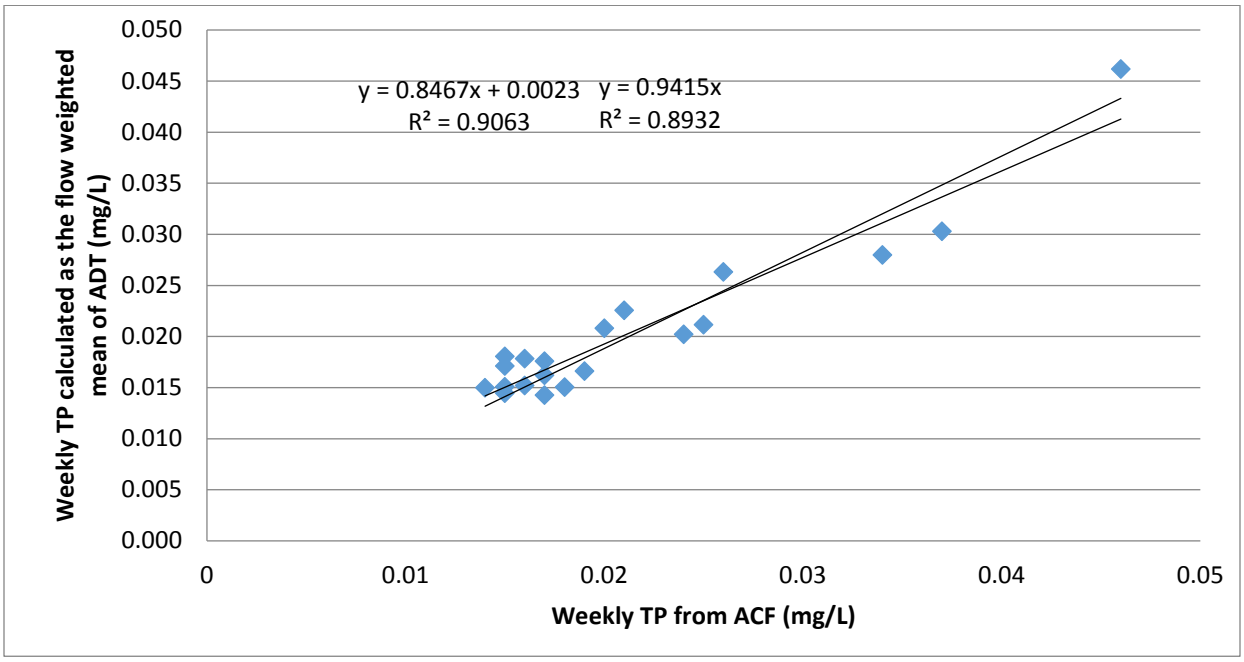

Figure 11. Weekly TP results from ACF and flow-weighted mean ADT results with one outlier removed. Two equations are presented, one forced through zero and one not.

CONCLUSIONS AND RECOMMENDATIONS

The primary conclusions of Project REST are as follows:

1. Grab and ADT methods appeared to be more reliable in collecting samples than flow-based ACF. Under some conditions, the ACF method may be biased and did not meet completeness targets. Data from time-based sampling can be used for flow-weighted calculations without sacrificing information.
2. Infrastructure and levees function as attractive habitat for wildlife, which makes them potential sources for TP. Some animals like anhingas and turtles were observed to interfere directly and routinely with sampling systems. Extreme aberrant samples were not documented during the study period.
3. Masses of submerged aquatic vegetation often collected at the sample intake screens, but the influence of these events on TP results was not clear.
4. The relationship between grab and autosampler results and the relative impact of extraneous values is influenced by the distribution of the data at a particular structure and tends to produce bias on the high side.
5. Data from the RPA method indicates a mid-day peak in TP concentrations.
6. Systematic problems were also found. Reverse, low, or phantom flows interfere with ACF triggering and can lead to unrepresentative sampling, particularly at sites with small headwater to tailwater differences.
7. Completeness estimates differ depending on whether they are based on time coverage and flow representation. Completeness levels for data from probes were lower than expected and the current completeness targets for probes deserve attention.

The primary recommendations of Project REST project are as follows:

1. A process to check the flow-proportionality of autosamplers should be implemented for important water management structures.
2. The placement of ACFs at structures at which reverse, low, and phantom flows are prevalent should be discouraged. A method for preventing reverse, low, and phantom flows through structures with mandatory autosamplers should be considered.
3. For autosamplers, completeness calculations should be based on the amount of flow represented, rather than span of time sampled.
4. Wildlife and biogeochemical processes on infrastructure should be considered as sources of TP to the adjacent marsh and water column. Methods for limiting the wildlife impacts on surface water quality sample collection warrant further study.
5. A focused study comparing ACF data with ADT data (or RPA data as a surrogate) should be carried out at multiple locations with the goal of validating the use of alternative methods.
6. Findings from Project REST should be further evaluated at other water management structures to validate or quality initial study results.

LITERATURE CITED

Rawlik, P. 2016. *Results from Project REST: Remote Environmental Sampling Test*. SFWMD WR-2016-003, South Florida Water Management District, West Palm Beach, FL. In review.

SFWMD. 2013a. Chemistry Laboratory Quality Manual, SFWMD-LAB-QM-2013-01. South Florida Water Management District, West Palm Beach, FL.

SFWMD. 2013b. *Science Plan for the Everglades Stormwater Treatment Areas*. South Florida Water Management District, West Palm Beach, FL. June 2013. Available online at www.sfwmd.gov/rs_scienceplan.

SFWMD. 2014. *Field Quality Sampling Manual*. QM-001-08.1, South Florida Water Management District, West Palm Beach, FL.