Restoration Strategies Regional Water Quality Plan – Science Plan for the Everglades Stormwater Treatment Areas

Evaluation of Sampling Methodologies

Results from Project REST
Remote Environmental Sampling Test

WR-2017-002

Peter Rawlik
Water Quality Monitoring Section
Water Quality Bureau
Water Resources Division
South Florida Water Management District

April 2017

EXECUTIVE SUMMARY

INTRODUCTION

Environmental monitoring data contains variability generated by four sources: inherent natural variation from an aquatic environment, variation added by the sampling process itself, small-scale chronic interferences with sampling, and rare, large magnitude acute spikes caused by sample entrainment of materials not normally found in ambient waters. Data variations added during sampling are dealt with by established quality assurance and quality control processes in the field and laboratory, and will not be investigated further as part of this project. 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. 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 2013a) 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)**. 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 with the water quality based effluent limits (WQBEL). 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, 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, one underwater and focused on the sampling equipment intake screens, and one far afield, looking at the sampling platform as a whole. Probes were also installed to monitor temperature, pH, specific conductance, and turbidity.

This array of equipment was operated at two monitoring stations located near structures. The first location was G310, located at structure G-310, a large pump station and the primary outflow from STA-1 West that is comparable to discharge structures at STA-1 East 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 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 2014a). All collected samples were analyzed by the District's laboratory (SFWMD 2013b).

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. 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. 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.

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 and various herons were observed, but did not seem to cause significant impacts. On the other hand, anhingas 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, or a longer-term chronic problem when accumulated feces on infrastructure 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 and wood storks. 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.

RESULTS: INTERFERENCES WITH SAMPLING METHODS

An analysis of sampling completeness (number expected versus number collected) 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 RPA sampling and analysis system 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 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 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.

A comparison of loads calculated using RPA data and multiple sub-time steps 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, 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.

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. However, extreme aberrant samples with acute interferences 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 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.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS, ABBREVIATIONS, AND UNITS OF MEASUREMENT

µg/L	micrograms per liter
AC	alternating current
ACF	autosampler composite flow
ACT	autosampler composite time
ADT	autosampler discrete time
ANOVA	analysis of variance
District	South Florida Water Management District
DVR	digital video recorder
ft^3	cubic feet
ft^3/s	cubic feet per second
GP	grab pumped
NOB	no bottle sample
NTU	nephelometric turbidity unit
Project REST	Remote Environmental Sampling Test Project
QA	quality assurance
RPA	remote phosphorus analyzer
SFWMD	South Florida Water Management District
STAs	stormwater treatment areas
TP	total phosphorus
WQBEL	water quality based effluent limit

INTRODUCTION

BASIS FOR PROJECT

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. Environmental monitoring data should be expected to contain variability generated by inherent natural variation, some variation associated with water sampling itself, but also small-scale chronic interferences with sampling. Interferences of low concentrations can be associated with local runoff from water management features, wildlife activities, or vegetation impacts and can be difficult to separate from natural background noise. These chronic 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 contrast, interferences of high concentrations result in acute spikes in water quality data and can be associated with aberrant events that are difficult to identify.

In addition to these three types of data variation associated natural sources, there are also small errors associated with taking and processing samples. This source of data variation is dealt with routinely through quality assurance and quality control processes in the field and laboratory and will not be investigated during this study. Finally, there are potential errors generated by failures in automated sampling systems. These institutional errors are often systematic in nature but may be just as influential on analytical results as all other variations in the sampled water itself. A look inside automated sampling processes is an important part of this study.

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 research projects associated with the STAs seek to improve performance, but no project has been focused on understanding factors influencing the sampling used to assess that performance. Using multiple sampling methods, cameras, and probes, sampling interferences, and malfunctions were documented in this study, the **Remote Environmental Sampling Test Project (Project REST)**. The project attempts to gather specific information on potential sources of variation linked to water quality samples and sampling systems. It is hoped that information gathered from the study can then be used to refine sampling protocols and methods to improve sampling methods, data quality, data interpretation, and reporting such that unexplained variation associated with chronic and acute interferences will be limited and not unduly influence data used for tracking and improving STA performance.

The use of multiple sampling methods and procedures to generate flow-weighted means for comparative purposes using autosamplers and even flow-proportional samples collected over shorter time frames and analyzed discretely, might provide information to understand discrepancies in the data. It is possible that one of these methods with or without modifications, or another yet to be implemented or explored, would provide more accurate representation of surface water TP. Additionally, while the theory of using machines to automatically sample in a flow-proportional manner, and therefore create a representative sample, might appear straightforward, and be relatively easy on a small scale over short time periods, this might not be the case at larger scales and longer periods. One of the primary questions Project REST examined is the efficacy of such sampling and if the sampling process is generating the number of expected samples. Only by moving beyond the actual numbers and examining the mechanics and logistics that generated those numbers can we understand whether the results mean what we think they mean. Amongst other things, Project REST seeked to look past the results and inside the process to test the assumption that the values generated by complex, automated systems are superior to those collected by simpler methods.

BACKGROUND ON EXISTING SAMPLING METHODS

In general, the District uses three primary methods to collect surface water samples: grabs (single point measurements); autosampler composite flow (ACF), which are flow-proportional samples collected using a robotic pump and composited over approximately a week; and autosampler composite time (ACTs), which are time proportional samples collected using a robotic pump and composited over approximately one week. Additionally, variations on these methods exist in which the composite sample can be done at different time steps, including hourly samples collected into a daily composite known as autosampler discrete time (ADT). Variations in grab samples exist as well, and can include direct collection into the analytical bottle, use of a van Dorn, use of a peristaltic pump, and in some cases use of the robotic peristaltic pump used for the collection of ACF samples.

Over the past two decades, increasing numbers of ACFs have been used to create single TP value for the entire week of flow at a structure, which is then used to generate an annual flow-weighted mean concentration as an estimate of discharge concentrations. For the STAs, water quality compliance is based on flow-weighted means calculated from ACFs with grab samples collected as backups in case of autosampler failure. Over the long term (e.g., months to years), the methods have been shown to produce very similar data and/or trends, although autosampler results tend to be slightly higher than grab values. However, significant deviations in data from grabs and autosamplers have been documented for short periods at some stations. These rare, chronic events, in which TP results from grabs and autosamplers vary widely, are of particular concern as they have the potential to impact long-term estimates of loads and concentrations and increase uncertainty in the process especially at low ambient concentrations.

It would not be possible or practical here to review the large literature on sampling methods for water quality or even to list out all the prior investigations done by the District on the differences between autosamplers and grabs. However, one significant study, Millian, Swain & Associates, Inc. (2007) deserves discussion. Under this study, ten time-proportional discrete autosamplers (ADTs) were deployed and the samples analyzed for TP. Of the ten stations, two did not show a statistical difference in concentration between the grab and the last bottle of the week (the composite closest in time to the grab), or the weekly average. In contrast, the other eight stations showed either a difference between the grab and the last bottle or the weekly average. A load analysis for eight of the ten stations found that only four were significantly different statistically, and in all cases, grab calculated loads were lower than those calculated from autosamplers. This pattern was also reflected in the concentration data. It must be remembered, however, that statistical significance can be found but have no real impact on the use of data for environmental management, and there are notable differences in the costs of sampling using different methods.

The fundamental conclusion of the report acknowledged that grab and autosampler results were different, but provided no information about the factors causing the deviation. However, the generally held explanation is provided in the literature on urban stormwater monitoring:

> Data obtained from too few grab samples are highly variable, particularly for industrial monitoring programs, and subject to greater uncertainly because of experimenter error and poor data-collection practices. In order to use stormwater data for decision making in a scientifically defensible fashion, grab sampling should be abandoned as a credible stormwater sampling approach for virtually all applications. It should be replaced by more accurate and frequent continuous sampling methods that are flow weighted. Flow-weighted composite monitoring should continue for the duration of the rain event (National Research Council 2009).

Ackerman et al. (2011) had similar conclusions:

> Volume-paced microsampling and targeted volume-paced sampling with analysis of discrete samples provide alternatives that improve accuracy without costing as much as pollutograph sampling. The common features of both these approaches are: 1) use of volume pacing, not time pacings, 2) their ability to capture a range of different storm types (i.e., sizes and timing), and 3) their inclusion of multiple discrete samples. Numerous authors have previously documented that volume

> based sampling is more accurate than time-based because it provides better representation of the overall storm.

Thus it is often concluded that flow-proportional monitoring provides the most accurate (meaning close to the actual value) estimation of loading during stormwater events based on the fundamental notion that grab sampling programs have an insufficient number of samples at an inappropriate frequency to properly “capture the action”. Following this line of reasoning, when District staff sees discrepancies in TP values between grabs and ACFs, the tendency has been to conclude that the ACF has appropriately characterized the TP for the week, but that the in situ conditions had changed by the time that the grab was taken.

However, Ackerman et al. (2011) provide specific suggestions about how such sampling should be carried out to capture “event mean concentrations,” which would be comparable to the composite data collected by the ACFs in the STAs:

> By targeting the volumetric pacings based on anticipated storm size, sampling is better able to capture a representative portion of the storm. Given the error inherent in weather predictions, it is preferable to overestimate (i.e., storm is smaller than expected) than to underestimate when setting the sample pacings. Although more costly, analyzing discrete samples as opposed to compositing into a single sample allows for better representation over the course of a storm, results in more accurate EMCs [Event Mean Concentrations] and provides greater flexibility if the storm does not materialize as predicted.
>
> Costs can be partially reduced by shortening the duration of sampling or reducing the number of discrete samples analyzed from ten to four. Because concentrations are typically higher during the early portion of storms . . . ending sampling when flow is 50% of peak flow can reduce costs with little overall effect on accuracy of the EMCs because at that point the majority of the pollutant mass and storm volume have flowed past the station. However, this requires accurate assessment of timing of peak flow in the field; inaccurate determination of peak flow may introduce additional bias.

This design of dynamic storm-based sampling is similar to that proposed by Abtew et al. (1997), which proposed autosamplers with variable trigger volumes based on the flow rates of structures during the current and preceding week. Ackerman et al. (2011) warned that the “volume-paced composite sampling resulted in overestimates of EMC, particularly for large storms.” Indeed, Ackerman et al. (2011) suggests that the sampling strategy might have to be different for different kinds of runoff, particularly those with lower and longer amplitudes.

This is particularly true as the grab samples approach the detection limit of the analytical method (for SFWMD, this is 2 µg/L). At such low levels, the distribution of data is no longer normal and is skewed to the right (higher values). In simple terms, water free, or nearly free, from phosphorus is rarely found, and when it is, it is rarely sustained, and therefore such results are quickly obfuscated by higher concentrations. Essentially, rare low concentrations are lost in the compositing process, their signal being overwhelmed by that of the more prevalent higher concentrations. For example, imagine a composite of four samples that has a TP concentration of 14 µg/L. It is reasonable to assume that the concentrations of each of those samples is 14 µg/L. It is also possible, that other combinations could produce the same result, including a set of samples with values of 1, 11, 17, and 27 µg/L. Such scenarios are likely given the limited amount of data concerning water quality between sampling events, and suggests that to understand water quality results, an understanding of sampling design and processes is needed.

Additionally, unexplained variation in TP data likely 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 only higher values. These higher concentrations cannot usually be counterbalanced; samples with 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

That the sampling strategy may need to be tailored to different kinds of runoff raises another question concerning confounding factors not through inclusion, but rather by absence. Most of the studies on stormwater sampling focus on urban, riverine, or industrial systems, but the STAs are shallow wetlands with significant amounts of emergent, submergent, and floating macrophytes, periphyton, and algae, a condition not common in the studied systems. Also likely absent from the referenced studies, particularly in urban and industrial settings, are wildlife including fish, birds, amphibians, reptiles (turtles, snakes, and alligators), and large mammals all of which are common in STAs and could impact sampling equipment and water quality. The impact of water body type, vegetative detritus, and wildlife on sampling equipment deployed over the long term (months to years) is essentially unknown.

The literature suggests some important issues that must be addressed to assure that the flow-proportional monitoring is accurate. The assumption of studies advocating for flow-proportional monitoring is that the sampling is being done in response to short-lived storm events (hours to days) that have clear beginnings and ends, and that the flows and loads mimic that same cycle. South Florida’s cycle of wet and dry seasons seem antithetical to this design, and while storm events do occur in this subtropical region, the resulting flows tend to last for extended periods. This condition may be exacerbated by how those flows are managed, essentially artificially metered, into and out of the STAs, which tend to flatten peak flow and extend the duration of the event.

Indeed, the volume-based program developed by SFWMD for use in its ACF monitoring program is essentially insensitive to storm size, and makes no attempt to define individual storm events. In theory, the sampling program should be redesigned with new trigger volumes based on expected storm magnitude, taking into account intensity, geographic distribution, and duration. The District’s program does not do this, instead triggers are set at volumes that assure that during high flow events the samplers themselves will not overflow, but at the same time collect a sufficient number of samples to process and analyze. However, this regime means that for extended periods of low flow, very few samples are collected. Additionally, for extended periods (greater than a week) of high flow, the storm event might be split between multiple samples. Simply put, the District’s flow-based monitoring is neither dynamic (adjusted for storm size) or event-based (focused on definable storms). Unlike the systems described in the literature, which are dominated by punctuated storm events with definable beginnings and ends, the South Florida environment and the STAs can be viewed as a season of almost constant rain punctuated with higher flow events from localized and regional storms followed by a dry season. This difference also results in a system that attracts and sustains biological communities and conglomerations that are not present in more controlled settings. Consequently, while it is tempting to assume that the differences in ACF and grab samples is solely attributable to differences in sampling frequency, there are other factors that must be examined.

AMBIENT CONDITIONS AT EACH SITE

Project REST collected detailed data and information on TP sampling at two distinctly different water management structures. The primary focus of the Project REST was the G-310 structure (**Figure 1**), a large pump station that functions as the primary discharge of STA-1 West. Sitting at the end of an approximately nine-kilometer long discharge canal, the structure consists of two 950-cubic foot per second (ft^3/s) diesel pumps, two 470-ft^3/s diesel pumps, and two 100-ft^3/s electric pumps for a total capacity of 3,040 ft^3/s. The canal collects treated water from three wetland treatment flow-ways with a total area of 6,670 acres covered by a mix of emergent wetland plants and submerged aquatic vegetation. The annual operational targets for this acreage are to treat 131,400 acre-feet of water and sequester 25,800 kilograms of phosphorus. The sampling site for the G-310 structure, site G310, is a platform approximately 1,200-meters upstream of the G-310 structure. The sampling platform with equipment jut out approximately ten meters from the eastern levee of the discharge canal (SFWMD 2014b). Flow is calculated by difference between two stage gauges on either side of the structure.

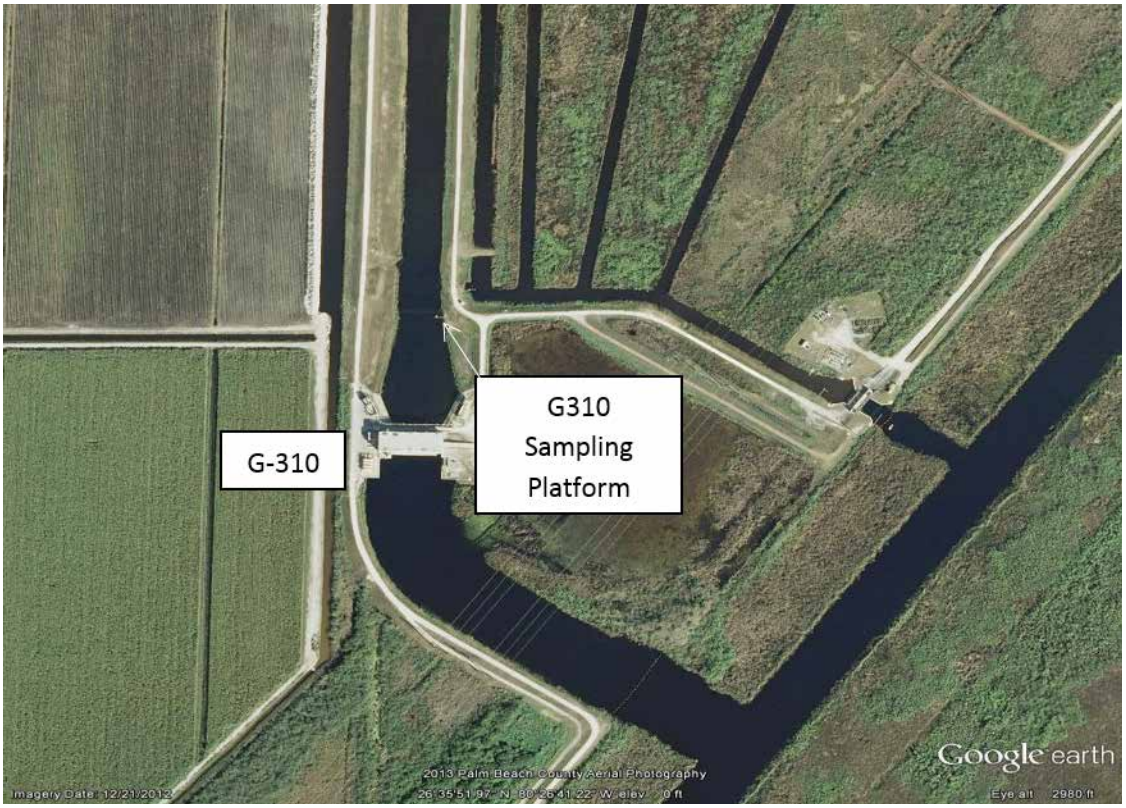

Figure 1. Structure G-310, the primary discharge of STA-1 West, and the G310 sampling platform.

A second location, structure G-390B for the Periphyton-based Stormwater Treatment Area Project, was leveraged (**Figure 2**). G-390B, which according to documentation (SFWMD 2014c) is a gated 6-foot by 6-foot box culvert with a peak capacity of 210 ft^3/s. The primary source of water to G-390B is a 222-acre constructed wetland cell dominated by submerged aquatic vegetation, but ringed with emergent vegetation and a collection canal. The sampling site for the G-390B structure, site G390B, is the end of the wing wall, approximately seven meters from the gates, but generally considered within the footprint of the structure itself. A vegetation barrier is in place directly in front of the wing walls. The structure discharges to an 11-acre constructed wetland cell dominated by submerged aquatic vegetation and periphyton, but subdivided with emergent vegetation. Distribution and collection canals at the upper and lower ends of the cell also exist (SFWMD 2014c). Flow is calculated as the difference between two stage gauges on either side of the structure.

Figure 2. Structure G-390B, the inflow to the Periphyton-based Stormwater Treatment Area Project, and the G390B sampling site.

WATER QUALITY MONITORING EQUIPMENT AND OPERATIONS

The REST Project used observations and data collected for four sampling techniques and several other sampling-related operations. Each of these methods was conducted with varying numbers of inherent assumptions and operational complexity. The basic project objective was to document how well these approaches deliver the desired data and to identify factors that influence sampling success. The study period was originally intended to span one year from October 2013 through October 2014. However, as a result of various delays, the study at G390B began in February 2014 and ended in October 2014, while the work at G310 lasted from April 2014 to December 2014. In both cases representative periods of both the wet and dry season were still captured. All sampling and monitoring followed quality guidance set out in the District's *Field Sampling Quality Manual* (SFWMD 2014a). The discussion below provides a basic understanding of sampling methods and the associated language.

GRAB SAMPLING

Grab sampling is generally considered the simplest of sample collection and can be as easy as filling a sample bottle directly, or involve the use of Niskin bottles or sampling poles in order to collect away from the shoreline or other undesirable locations. Grab sampling is the least complex method and only assumes that an individual sample is representative to the ambient water at the time of sampling, this is one of the fundamental concerns with weekly grab samples. The utility of grab samples over a week during which ambient conditions likely change must be acknowledged. Weekly grab samples of surface water for nutrients and related parameters, including TP were leveraged for this project. The existing quality assurance program (SFWMD 2014a) for grab samples was leveraged for this study, and included standardized equipment decontamination procedures, quarterly equipment blanks, and one replicate sample per quarter.

TIME-PROPORTIONAL AUTOSAMPLERS

Autosamplers are robots that consist of tubing, a peristaltic pump, and an array of sampling bottles that can be filled in a variety of ways depending on programming. These methods include time-proportional discrete samples, or autosampler discrete time (ADT), in which aliquots are collected from each sampling site at a set interval and then composited to create samples in a single bottle for a set period of time.

In this study, aliquots were collected every two hours to create bottles that were representative of a single day. The daily bottles were then analyzed separately. ADT programs can be desirable as they rely solely on a power supply, the autosampler, and the technician who programs it. Samples from autosamplers are collected by staff weekly. ADT sampling assumes that the time-based aliquots are taken as programmed and that 12 aliquots per day reflect daily water mass well without flow weighting. The existing quality assurance program (SFWMD 2014a) for autosamplers was used for this study, and included standardized equipment decontamination procedures, quarterly replacement of tubing, and an associated equipment blank.

FLOW-PROPORTIONAL AUTOSAMPLERS

In contrast, flow-proportional composite samples, or autosampler composite flow (ACF), are created by programming the autosampler to respond to a specified trigger volume. Aliquots are collected in multiple bottles and then composited into a single sample that is submitted for analysis. While this method is preferred for collecting flow-proportional samples, it relies on a complex interplay of independent parts including the power supply, the technician, the autosampler, a communications network, stage measurements, flow calculation (stream gauging), determination of a trigger volume, and a flow totalizer. Flow-proportional sampling assumes that all parts of the autosampling and flow measurement process are functioning as designed and that multiple weekly aliquots provide consistent flow proportionality and are representative of the flow-adjusted concentration moving through the structure during the week. ACF sampling is the most complex approach tested.

Trigger volumes are determined primarily based on the maximum capacity of the structure and are calculated such that the maximum volume is distributed over the capacity of the entire bottle set without exceeding it. Adjustments to the trigger volume can be made in response to seasonality or operational limits. The measurement of volume is usually accomplished by measuring the difference between upstream and downstream stage and then applying these to a flow algorithm. Alternatively, a similar result can be accomplished using pump revolutions.

The flow data are transmitted and accumulated in what can be thought of as a sample flow totalizer, a database that accumulates flow and then at the set flow volume, sends a signal to the autosampler. Importantly, this database is not the same as the official database that contains quality assured flow data for each structure. Rather, the totalizer is a distinct and separately maintained array, with two important aspects that must be understood. First, the totalizer only accumulates positive flow; it does not accumulate negative flow, nor does it subtract out negative flows. Second, the sampling technician visiting the site weekly does not have access to the totalizer and cannot reset the values contained within it. Consequently, flow measurements that have not triggered the ACF during one week are carried over to the next week. During periods of no or low flow, it may take weeks or months for the totalizer to reach a trigger volume.

At each site, the existing ACF was leveraged with the G390B autosampler programmed to collect every 310,000 cubic feet, and the sampler at G310 programmed to collect at every 5,200,000 cubic feet. Measurements of positive flow volume were cumulative over time. However, negative flows, prevalent at G390B but not at G310, are, according to documentation (Abtew et al. 1997), subtracted from the positive flow. Composite samples were collected and processed for analysis weekly. Flow-proportional sampling assumes that all parts of the autosampling and flow measurement process are functioning as designed and that multiple weekly aliquots provide consistent flow proportionality and were representative of the flow-adjusted concentration moving through the structure during the week. The existing quality assurance program (SFWMD 2014a) for autosamplers was used for this study, and included standardized equipment decontamination procedures, quarterly replacement of tubing, and an associated equipment blank.

REMOTE PHOSPHORUS ANALYZER

At each site, a remote phosphorus analyzer (RPA) was installed. The RPA is an experimental piece of equipment that combines the functionality of a sampler with rapid chemical analysis for TP. While the equipment and methodology are still in the experimental stages, the preliminary results show promise and appear to be adequate for research (Struve et al. 2012). The RPA at each site was programmed to sample every two hours in conjunction with the ADT. On occasion, the RPA at G390B was adjusted to sample every three hours to support other work being carried out in the area. Maintenance and quality assurance for this piece of equipment is detailed in Struve et al. (2012). One critical difference in this piece of equipment, as compared to autosamplers, is that the intake line is not a piece of replaceable tubing, but rather an integrated piece of polyvinyl chloride (PVC) that is not regularly changed.

PUMPED SAMPLES

Both the ACF and ADT draw samples through tubing that is immersed at depth in the water column. This tubing is changed quarterly per quality assurance (QA) guidelines (SFWMD 2014a). There was concern that the tubing itself could impact TP results. In order to document that this was not the case, weekly samples were taken through the tubing and autosampler pumps for comparison with the grab samples. The waters taken through the autosampler pumps will be termed grab pumped (GP) samples.

DATA SONDES

Deployed probes measuring temperature, specific conductivity, and pH were originally programmed to collect at an hourly time step. As a result of changes in power supply, the probes at G390B were capable of collecting every fifteen minutes. This change in frequency was not possible at G310. In situ probe sets were changed weekly with units calibrated to District standards (SFWMD 2014a).

NO BOTTLE SAMPLE

Several circumstances can produce unusable samples or failures in the sampling process. A term associated with such gaps is ‘no bottle sample’ (NOB), which are misfired alternatives to valid samples. In general, there are four ways to generate a NOB: 1) if there was no flow at the structure and the sampling is dependent on flow such as an ACF or a grab sample if flowing; 2) if there was flow, but it was insufficient to reach a trigger volume; 3) in autosamplers, the bottles are pre-preserved with acid, and in events in which only one trigger is received, the volume of water may not be enough to raise the pH to a level that will not damage the analytical machinery (such samples are rejected by the laboratory); and 4) sampling equipment failure generated by problems with communications, power, pumps, and programming. Of these four types of NOBs, only sampling failure is considered unacceptable, while in the three other types, the equipment is deemed to be in working order and the sample will count toward efficacy of the monitoring system. On occasions when either the totalizer or communications equipment is taken off line, the ACF can be temporarily reprogrammed to function as a time-proportional unit either discretely (ADT) or more commonly, as a composite (ACT). In these cases, the actual equipment is not able to function as designed by intentional actions and the samples will not be counted toward efficacy of the sampling technique.

CAMERAS

Two aboveground cameras were also installed at each site, one focused on the water's surface above the sampling point, and one focused on the platform and equipment surrounding it. Initially, motion sensor cameras programmed to collect at a regular frequency as well as in response to motion were used, but these were found to have insufficient internal power supplies. Instead, waterproof security cameras were installed in the same places. These, in combination with the underwater cameras, were connected to a four-channel digital video recorder (DVR) with the intent of collecting images continuously. DVRs were removed from the field and replaced on a weekly basis. Imaging was downloaded and reviewed on a weekly basis, which regularly took three days. In practice, this created a massive amount of data and practical decisions regarding collection and archiving had to be made, with imaging at night showing primarily black scenes (particularly underwater) being discarded.

EQUIPMENT PLACEMENT

At G310, the three surface water intakes were installed in close proximity to each other using a floating arm attached to the platform as an anchor point (**Figure 3**). All three intakes consisted of PVC pipe that followed the support pipes of the floating arm and then turned 90 degrees down. Two of the PVC pipes served as housing for Tygon tubing connected to distinct autosamplers, while the third served as a direct supply line to the RPA. The camera was positioned perpendicular to flow such that positive flow was from right to left (**Figure 4**).

Figure 3. Sampling equipment at G310 mounted on a floating armature.

Figure 4. 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, conductance, and turbidity.

At G390B, the three surface water intakes were installed in close proximity (less than one meter) to each other using the wing wall as an anchor point. All three intakes consisted of PVC pipe that was connected to the wing wall at a slight angle (**Figure 5**). Two of the PVC pipes served as housing for Tygon tubing, which led to distinct autosamplers, while the third served as a direct supply line to an RPA (**Figure 6**). These were adjacent to the probes. The underwater camera was focused on the three intakes. The camera was positioned such that the intakes were upstream and positive flow was toward the camera.

Figure 5. Sampling equipment at G390B mounted on the structure wing wall.

Figure 6. 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, conductance, and turbidity, is off screen to the right.

FIELD OBSERVATIONS ON THE SAMPLING PROCESS

Events that could interfere with representative sampling could occur above or below the water surface and at the sampling intake, or close by on structures or embankments. To get information from all these locations, three cameras were installed at each site (G310 and G390B) to document events above and below the water surface when field crews were not present. These cameras included one underwater camera aimed at the sampling intakes (**Figure 4** and **Figure 6**), one camera above water focused on the surface of the water above the sampling intakes (**Figure 3** and **Figure 5**), and one camera set further back giving a synoptic view of the sampling area.

Initially, motion sensitive field cameras took photographs every minute and in response to motion were used above water, but these cameras were found to consume too much battery power to be cost effective. In their place, waterproof security cameras were installed. These have the option to manipulate not only the frame rate, but the resolution as well and tied into an alternating current (AC) power supply. All three cameras at each site were then connected to programmable DVRs.

In order to streamline transfers of imaging, two DVRs were assigned to each site and swapped out weekly. This duplication was necessary because transferring an entire week’s worth of images from each site required about three days, although this process could be run in parallel for each site. The rate limiting factor here was that the DVRs would only allow transfers of data from a single day at a time, and often each transfer could only be set to move twelve hours at a time. Adjustment of frame rate and resolution helped reduce the size of files and increase transfer rates. In this process, more than twelve terabytes of data were collected and to conserve storage space, some collected imaging had to be discarded. Discarded images included those collected by the underwater cameras at night when the recordings were essentially black, also discarded were a large amount of night imaging of the surface water and the wider field views, but only after the imaging was screened.

PARTICLES AND FLOW DYNAMICS

Photographic evidence confirms large changes in particle number and type at different times of day and with flow dynamics. Particles at G390B appeared to consist of small pieces of detritus, likely periphyton and unidentifiable pieces of macrophytes. Also visible were finer particles, some of which were likely phytoplankton, and others which, given a limited amount of locomotion, were probably zooplankton. Objects that were likely zooplankton seemed to be more prevalent around dawn or dusk, but whether this was real or simply a result of the angle or intensity of light cannot be determined. Such particles were seen regardless of flow through the structure.

In contrast, particles at G310 seemed directly tied to operation of the pump. When there was no flow, there were almost no particles to be seen (**Figure 4**). When pumps were turned on particles became apparent almost immediately and grew in size and number as pumping rate increased (**Figure 7**). At lower flow rates, the particles at G310 appear to be detritus, primarily decaying vegetation, as velocity increased, however, the ability to resolve the individual particles became difficult and it is possible that some of the material seen was not detritus, but rather mineral sediments.

Figure 7. G310 during a period of high flow showing particles in motion.

Observations of particle motion documented unanticipated changes in the direction of flow. While the dominant flow was north to south toward the pump station, on occasion the flow was south to north, often in response to either the development of an eddy, or the cessation of flow or change in flow operations, which created temporary waves away from the structure. In the case of the cessation of flow, such waves could often be seen in the imaging for some time, presumably moving back and forth along the canal until they finally dissipated. Particles moving backwards during these reverse flow events did not settle out, quickly indicating that they were likely primarily detritus. Field crews have observed such eddies at other structures during flow events (e.g, S-333).

Particles were also seen from the cameras trained on the surface waters at G390B (**Figure 8**). These accumulations of dust appeared to originate from within the bay of the structure itself, and appear to be fine debris knocked off of the structure itself during operations. Investigation of the underside of the structure found dried mats of algae, mold and fungi, dead insects, spider webs, nests of mud daubers, dried animal feces, and other detrital material. It appears that some of this material would, given sufficient impetus, fall off the structure and onto the surface of the water. This material tended to be captured in the eddy and circulate several times in the structure bay on the surface, often passing over or near to the intakes several times before no longer being able to be observed. Clearly, more consideration of hydrodynamics within the influence of the structure is warranted as we seek to improve the representativeness of samples.

Figure 8. Ashy material floating on the water column at G390B.

VEGETATION

At G390B, the primary vegetation seen to impact the sampling site was detached cattail (*Typha* sp.); other emergent vegetation tended to be captured by the weed barrier, rather than by the sampling equipment. However, although the weed barrier intercepted floating debris, the result was only temporary. The constant pressure of flow, combined with wave and wind action, worked at the captured macrophytes and over the course of days and weeks, broke them down into smaller pieces. This detritus was either able to fall over the weed barrier, or pass through it.

At G310, the primary vegetation seen to impact the sampling site appeared to be southern naiad (*Najas* sp.). These plants tended to accumulate on the sampling intake pipes in huge conglomerations; in fact, masses covering hundreds of square feet around the sampler were common. These mats of vegetation were often so thick that the sampling equipment itself was obscured. While the primary source of such masses was from the north, masses often arrived from the south as well, either caught in the eddy or during back flow events caused by changes in flow operations. Additionally, some masses seemed to arrive driven by the wind. In several instances large masses of either cattail or southern naiad were seen to impact the site shortly after being cleaned by staff and persisted for days if not the entire week after. In one notable event, a mass of vegetation impacted a site a day after the sampling crew left, and stayed in place until just minutes before they returned a week later, leaving no evidence that there was ever anything amiss (**Figure 9** through **Figure 13**). Once vegetation began to accumulate on the equipment, it was rare for the site to clear itself, though this did occur, primarily when low flows were suddenly and dramatically increased. Change in flow direction also could clear the site, but this was rare.

When impacted with southern naiad and under low to no flow conditions, the act of collecting a sample, particularly the purging needed to prime the pump, could be seen to disturb debris captured amongst the leaves, causing it to rain down toward the sampler intakes. This was not observed during high flow events as the purging bubbles and raining debris tended to be carried away from the intakes by the flow. It was also interesting to note that the breakdown of southern naiad and subsequent uptake by the sampling equipment was not observed.

Figure 9. G310 on the morning of May 23, 2014, showing no impact from vegetation.

Figure 10. G310 on the morning of May 23, 2014, showing the arrival of uprooted cattail from downstream of the platform.

Figure 11. G310 on May 26, 2014, showing the cattail mass persisting.

Figure 12. G310 on the morning of May 27, 2014, showing the cattail mass beginning to break up.

Figure 13. G310 the morning of May 27, 2014 showing staff arriving with no evidence of impacts from uprooted vegetation.

However, cattails at G310 were seen to wrap themselves around the equipment (**Figure 14**) and essentially reduce or eliminate the available surface area of the intake to collect a sample. In one case, the complete blockage of the RPA intake was blamed for a resulting failure of the associated pump. While cattails appeared to be less likely to entangle on the equipment, they also appeared to be less frangible than southern naiad, and therefore persisted on the equipment in a different manner.

Figure 14. Vegetation wrapped around the sampling equipment at G310.

INVERTEBRATES

Few invertebrates were captured by the underwater cameras, and those that were seen consisted primarily of small snails. Surface cameras captured a wide variety of insect life including several hatches of what appeared to be mayflies (Ephemeroptera). Camera lenses were also prime habitat for spiders, which made webs on a nightly basis, possibly drawn by the camera lights. Similarly, wasps, particularly mud daubers, made cases on the cameras. In one set of images at the G-390B structure, a pair of large worm-like invertebrates moved up the structure wing wall and then into the landscape surrounding the area (**Figure 15**). It is suspected that these were an invasive species of giant land planarians (*Bipalium* sp.).

Figure 15. Possible *Bipalium* planarians crawling up the wing wall at the G-390B structure.

FISH

At G390B, large-bodied fish including largemouth bass (*Micropterus salmoides*), sunfish (Centrarchidae), cichlids (Cichlidae), and gar (Lepisosteidae) were frequently observed and almost always in close proximity to the sampling intakes (**Figure 16**). On regular occasions, large schools of such animals were seen to pass through the field of vision. On multiple occasions, various species of sunfish were seen to strike the intakes and the intake tubing in what appeared to be feeding behaviors. Gar were also seen, but were never seen to feed. Also common were sailfin catfish (*Pterygoplichthys multiradiatus*), which would often attach themselves to the intakes, and were also often seen among the rocks that formed the base of the levee. Small fish were rare at G390B.

Figure 16. Large-bodied fish at G390B.

At G310, large-bodied fish were almost non-existent, and only gar and sailfin catfish were seen with any regularity. Gar were never seen to feed, but the sailfin catfish were often attached to the intakes and the supporting infrastructure. Small fish including mosquitofish (*Gambusia* sp.), least killifish (*Heterandria* *formosa*), and sailfin mollies (*Poecilia latipinna*) were plentiful, and often congregated underneath the autosampler float in large schools (**Figure 17**).

Figure 17. Small-bodied fish at G310.

REPTILES

No crocodilians or snakes were ever seen to approach the sampling equipment at the two sampling sites, although some alligators were seen in the distance; however common visitors to both sites were aquatic turtles (*Pseudemys* sp.). At G310, turtles were seen to graze the accumulated material on the intake pipes and rest beneath sampling arm float (**Figure 18**). Turtles were often seen to dig through accumulated vegetation at G310. At G390B, turtles grazing on the equipment were also observed. Also documented at G390B was the use by turtles of the angled sampling pipes as a ladder to climb out of the water, and then as a slide to return to the water. In all cases, the activity of turtles was seen to significantly disturb the detritus accumulated on the sampling infrastructure and put a significant amount of material into the water column.

Figure 18. A turtle grazing on sampling equipment at G310.

MAMMALS

No mammals were seen in the water column, on the sampling platform, or in the camera field. The only humans seen were those authorized to collect samples or service the equipment.

BIRDS

While songbirds and wading birds, including cattle egrets (*Bubulcus ibis*) and a variety of herons (Ardeidae; **Figure 19**), were regularly seen at both sites, the most common visitor species were anhingas (*Anhinga anhinga*). Indeed, anhingas were observed almost on a daily basis, usually as solitary individuals that often occupied the area for hours on end. During these periods, they carried out a variety of activities that might influence sampling. Typical activities included swimming after prey, using the bird deterrents on the structures as tools to scrape fish off their bills, and not unexpectedly, defecating in the water column and on the infrastructure, particularly the weed barrier at G390B. The video images captured show that waste products remain cohesive for several minutes while in the water column (**Figure 20**). This cohesiveness may also explain why some infrastructure tends to accumulate large quantities of bird feces.

Figure 19. Egrets resting on infrastructure at the G-390B structure and G390B monitoring site.

Figure 20. Anhinga feces dissolving into the water column near the intakes of the sampling equipment at G390B.

There is an interesting diurnal process here that should be noted. Because of the subtropical nature of South Florida's climate, rainfall tends to be concentrated in the afternoons. Consequently, anhingas tended to be observed in the morning and early afternoon, and this was the time period during which feces would accumulate on equipment. A significant amount of this accumulation would then be washed from the infrastructure during the afternoon rains (**Figure 21**).

Vultures (Cathartidae) were also regularly seen at the G-390B structure, but only in the early mornings, sometimes as individuals, but also in groups. These birds did not occupy the sampling equipment, but rather roosted on the guard rails and equipment on the culvert itself. These animals would often defecate before departing, leaving behind material, some of which would eventually work its way through the floor grates of the structure and fall into the water column.

Figure 21. Progression of the accumulation and removal of anhinga droppings on the vegetation barrier at the G-390B structure and G390B monitoring site during a single day.

DISCUSSION OF FIELD OBSERVATIONS ON THE SAMPLING PROCESS

Much of the activity captured by the camera can be classified into four broad categories of potential concern for interfering with sampling representativeness. These are biofouling, bioturbation, excretion, and local runoff. All these activities can produce chronic interferences with water quality sampling.

Biofouling of the intakes occurred in two ways. First there was material, primarily periphyton and detritus, that accumulated directly on the intakes and surrounding infrastructure. Second, there were floating mats of vegetation that periodically surrounded the intakes. In both cases, the material could potentially function to interfere with water quality monitoring by acting as a source, or acting as a filter to keep particles out and absorb dissolved nutrients. It is important to note here that southern naiad may reproduce through fragmentation (Tarver et al. 1978); therefore, it is likely that the masses of plant material surrounding the intakes were still active and potentially taking nutrients out of the water column. Similarly, the periphyton growing on the equipment might also be decreasing TP concentrations. Bioturbation of the material located on or entangled with the intake equipment was observed relatively often. While fish, particularly sailfin catfish, were seen to feed on and disturb detritus, the biggest offenders by far were turtles, which created clouds of detritus when feeding or climbing on the sampling equipment. Biofouling and bioturbation both have the potential to generate increases in concentrations in samples only if the material is small enough, or can be rendered small enough, to be collected by the intake tubes during sampling.

Excretion by animals in the area surrounding the sampler intake is of critical importance. Fish tended to cluster around the intakes, and fish feces are known to be a source of phosphorus (Geesey et al. 1984). While it might be unlikely for the feces of any one fish to contaminate a sample, the density of fish might increase that potential, particularly at low flow rates. Similar things could be said about turtles. More dramatic, were the deposits of semi-liquid feces from anhingas, which appeared to slowly disperse in the water column, rather than sinking. However, as with fish feces, the chance that a bird dropping would be deposited just as the sampler was collecting is relatively low. Even so, the number of samples collected is so high that the possibility of interference must be considered. If we assume that the chances of sampling feces is only 0.05%, and an average autosampler collects about 2,000 aliquots a year, then 1 sample per year should be expected to be contaminated in such a manner. Analyses performed in the Everglades have determined that phosphorus concentrations in wading bird feces range up to 57 grams per kilogram (Irick et al. 2015) and thus could significantly impact sampling results.

While the excretion products of fish, turtles, and birds potentially generate some rare but acute interferences when immediately sampled, the waste products from birds can also generate chronic interferences long after the excretion event. Anhinga and vulture feces were observed to accumulate on infrastructure such as vegetation barriers and the water control structure itself. This creates a reservoir of waste that over the course of time could be moved into the water column by weathering (wind and rain). This runoff of high nutrient pools into the water column might be a critical factor in understanding localized but chronic, low-level increases in nutrient concentrations. Bird feces will produce large acute spikes in nutrient concentrations if entrained; such an event was not documented during Project REST.

Indeed, bird droppings might not be the only source to consider when discussing runoff and low-level, chronic interferences. The levees are made out of primarily shell rock that generally contains trace levels of phosphorus and are often covered with vegetation that serves as habitat for wildlife. Thus the levees and associated flora and fauna can be considered potential sources of phosphorus to the nearby water column. More dramatically, the water control structures often serve as habitat for biota as well, including growths of algae and bacteria, spiders and their prey, wasps and mud daubers, and even bats (Chiroptera) and birds. Consequently, the structures might be viewed as localized microhabitat that can become coated with a variety of detrital material. Various factors including rainfall, wind, water level, structure operations, and associated maintenance could all contribute to this detritus entering the water column. Once in the water column, even material that originated downstream of the sampling point may impact the sampler multiple times by being caught up in eddies, or through reverse flow.

As with all wetlands, the interface between the water and the land is an ecotone, a transition between two communities. However, in these constructed wetlands the transition is not gradual, but rather sharp with levees and structures rising at steep angles from collection canals. Consequently, the actual interface between the two systems is small, and likely generates an uneven exchange. This is by design, the STAs were not meant to serve as natural wetlands with well blended or rough interfaces, but rather with the intent of holding the water and biota in place. As a result, the exchange between the wetland and levees/structures can be considered as almost entirely in one direction, with material and energy flowing from the land into the wetland, but with very little moving in the opposite direction. If that material is phosphorus rich, including decaying terrestrial vegetation, animal waste, sediment, and calcium carbonate rock dust (suspended in the water column), then the levee itself becomes a source of TP to the water column. Studies in Louisiana found that runoff from bare levees could reach as high 2 milligrams per liter of phosphorus (Burwell 2009). The levee could be seen as the source of a chemical edge effect that potentially impacts water quality results. Extending this concept to structures, which tend to have even greater vertical surfaces, as well as highly stable (and therefore attractive to avifauna) horizontal structures, the potential for increased flow of energy and material off the land and into the water column around structures becomes apparent.

Another issue that must be discussed is representativeness, and this functions at two scales. First there is the issue of biofouling interfering with localized water quality. While the influence of periphyton and macrophytes on the sampling intakes on water quality may not be discernable, it is clearly not a representative condition. Consequently, a method or design that limits the accumulation of such material should be investigated to minimize chronic interference with sampling. On a larger scale, the impact of the infrastructure and the levee needs to be considered when installing sampling equipment, particularly at compliance stations. However, this presents a conundrum. If the levee and structure are themselves sources of phosphorus to the water column, and these are representative samples, is it appropriate to try and exclude that impact, or should steps be taken to limit such impacts? Comparable work on roads through wetlands use swales to reduce material entering the water column from the road bed, something similar might be appropriate in the STAs.

Ultimately, the images captured by the cameras suggest that the structures and levees are not simple control and transfer points, but rather complex systems in their own right that act as sources of nutrients that likely impact water quality in elevated levels of complexity that have not generally been examined. Unfortunately, this study was unable to detect a signal in the water quality data from these various events. This suggests that while there is a potential for wildlife to impact samples, it is relatively rare. This likely is because acute contamination events are relatively rare and ephemeral and that the timing of sampling must be just right to capture them. A review of historical water quality data would likely reveal rare elevated values that are incongruous with previous or subsequent values. Rarity is welcome, but acute inferences can have major impacts on an annual average value if no quality assurance issue is found to allow the value to be eliminated from the annual mean.

EFFICACY OF MONITORING

Environmental monitoring is fundamentally a data stream ending with the capture of useful data and information for meeting objectives. In this section, we will analyze monitoring efficacy, broadly defined as the capacity to produce information to meet programmatic needs. For some purposes, efficacy is known as data completeness. However as will be shown, “completeness” lacks the depth and flexibility needed to encompass the factors that should be considered when successful monitoring events do not produce usable analytical results. Our goal is to dissect monitoring processes and find causes for loss of output performance. The analysis will show that understanding programmatic needs and defining efficacy in compatible terms is essential to designing procedures to assure that acceptable standards are met. Finally, the limitations of various methods and the inherent failure types will be shown to greatly influence efficacy. Therefore, these factors will be considered in establishing efficacy or completeness targets and new standard operating procedures to maximize monitoring efficacy.

INITIAL EFFICACY TARGETS

In general, the stated completeness target for monitoring is 95% (SFWMD 2014a), and this is assumed to be based on time, meaning for example, that each year at least 50 of the 52 weekly samples will be collected. The original goal of Project REST was to have an entire year of monitoring by several methods to analyze comparatively at two stations. However, delays in budgeting, equipment ordering, delivery, installation, and activation greatly reduced the period available for analysis. In both locations, several factors influenced project initiation and sample collection. Consequently, the period available for review of efficacy was reduced at both locations from 52 weeks to 34 weeks. At G310, that period ran from April 21, 2014, to December 15, 2014, while the period at G390B ran from February 3, 2014, to September 29, 2014. This period of investigation can be expressed in a variety of ways that are relevant to the sampling frequencies at each station: 34 weeks, 35 Mondays, 238 days, or 5,736 hours.

Using these temporal intervals and completely successful monitoring, one would expect 35 weekly grab samples; a maximum of 34 weekly flow-proportional composited samples; 238 daily time-proportional samples; and 5,736 hourly measures of in situ parameters (temperature, conductivity, and turbidity). However, conditions at G390B allowed for the increase of frequency of in situ measurements to every fifteen minutes, so that the potential number of in situ measures was increased to 16,296 events.

EFFICACY AT G310

A summary of sampling efficacy at G310 is presented in **Table 1**.

Table 1. Efficacy of monitoring at G310.

Method	Expected Count	Actual Count	Percentage Success
Grab samples	35	35	100%
ADT Samples	238	235	99%
ADT Acceptable NOBs	0	3	<1%
Data Sonde Events	5,736	5,162	90%
RPA	2,856	2,169	76%
ACF Samples	34	20	59%
ACF Acceptable NOBs		8	23%
ACF converted to ACT		6	18%

Sample Completeness

The sample completeness results for G310 revealed some surprises. Not unexpectedly, grab samples and ADT samples showed near perfect sampling efficacy. This is reasonably attributable to the rather simple technology with few underlying assumptions involved to collect these types of samples. No appreciable samples were missing from either method, although three samples from the ADT were too acidic to analyze and classified as NOBs.

Similarly, the deployed data sonde functioned effectively, missing only 10% of its events. The loss of 143 data events (2.5%) were attributed to a single programming error that resulted in missing almost an entire week's worth of data. The remaining 7.5% of the data events (430) were lost as a result of power failures, and occurred in three distinct weeks. As a point of reference, a single week's worth of hourly data would be approximately 168 events.

Evaluating the ACF data (flow-proportional weekly composites) highlights some of the issues with flow-proportional sampling and the working definition of completeness. Sampling numbers suggest that since the autosampler collected samples only 20 times, and then had 8 acceptable NOBs, the percent completeness is only 82%. During the other 18% of the time (six weeks in July and August), flow-proportionality was lost and sampling was changed to time-based triggering (**Figure 22**). This was done to compensate for a failure in the radio communication network that delivers trigger signals to the sampler. Although staff were informed about this and the sampler was set to collect on a time-based algorithm, such a situation cannot be considered as meeting completeness targets as the sampling equipment did not perform as required or designed.

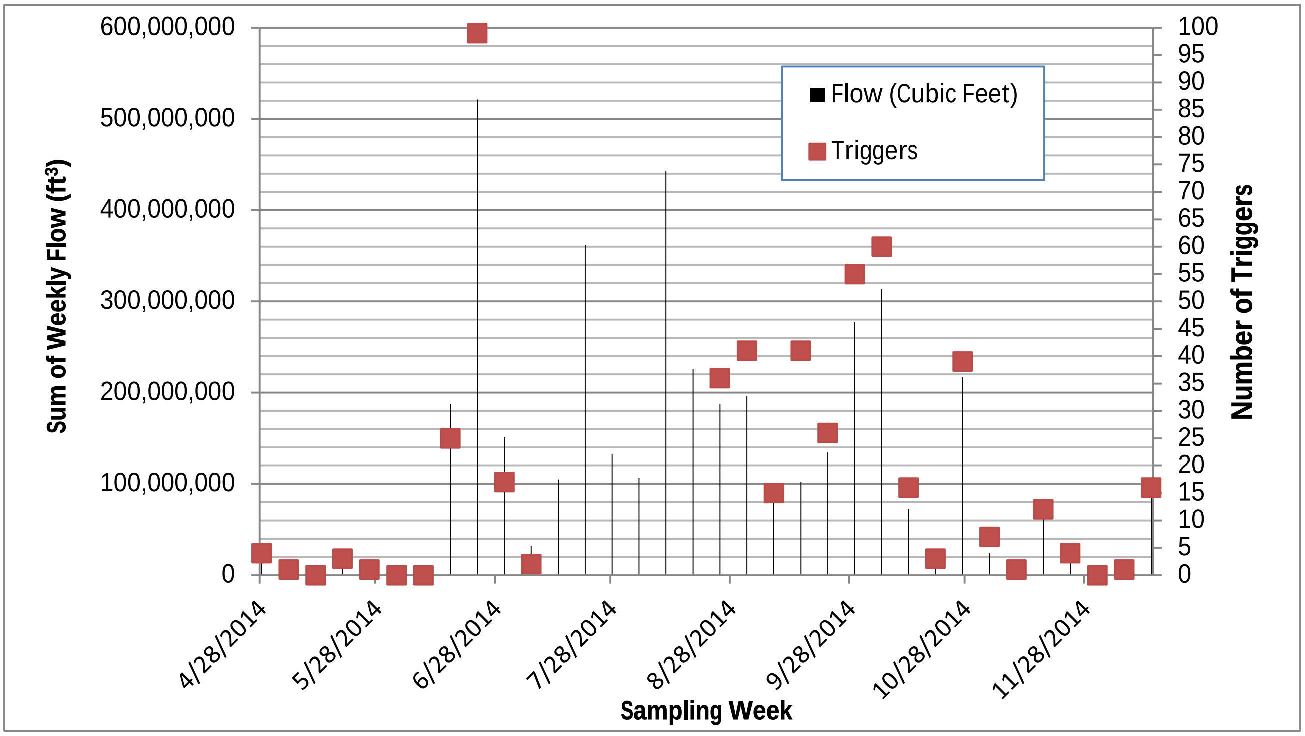

Figure 22. Weekly flow and sampling triggers at G310 for 34 weeks, and the failure of triggers during six weeks in July and August. (Note: ft^3 – cubic feet.)

Magnitude of Flow and Its Representation

However, magnitude of flow and its representation must also be considered when analyzing success in this form of sampling. Expanding the analysis of the performance to look at flow shows that for the 34-week period, the 20 weeks the ACF collected samples represented 59% of the time and 65.5% of the flow. The eight NOB samples represented 23% of the time, but only 0.5% of the flow. The six weeks the ACF was switched to an ACT was only 18% of the time, but 34% of the flow. All totaled, the ACF performed as required 82% of the time, but only successfully represented 66% of the flow. In other words, the flow-proportional monitoring program was offline and an alternative method of sampling was used for only one-fifth of the time, but for one-third of the flow. When actually used to calculate annual flow-weighted means, such missing ACF data would be replaced with the results of grab sampling. However, results of grab samples are considered less than optimal as they are only a single data point attempting to represent a significant period of time and flow.

Spurred by these results, a similar analysis was undertaken for the entire year. This revealed that the ACF collected samples for 29 weeks (56% of the time) and documented 64% of the flow. It produced NOBs for 16 weeks (31%) of the time representing less than 1% of the flow. The conversion to ACT was limited to six weeks, which represented 12% of the year and 25% of the flow. One additional week was a complete failure of the equipment caused by a loss of power, which represented 2% of the time, but 10% of the flow. All in all, the equipment functioned properly for 45 weeks or 87% of the time, but only captured data on 65% of the flow. Thus an argument could be made that the data at this site was only 65% complete as opposed to 87% complete.

Given that flow is not evenly distributed in time, this situation is likely a regular occurrence. A review of the flow data finds the monitoring program at G310 extremely inefficient, with 99% of the flow for 2014 being captured in just 36 weeks, 95% of the flow being captured in just 25 weeks, and 90% of the flow being captured in 20 weeks. It would be highly efficient to simply target the critical weeks and reduce monitoring for the other half of the year, but this analysis can only be done in retrospect, and not a priori. It is reasonable to expect that the majority of work will be done during the summer rainy season, but in 2014, a single week in February represented 10% of the flow for the entire year. A structure by structure analysis that identified weekly flow values that were likely too small to be more than 0.5% of the annual flow might be helpful in directing and managing resources.

As a QA check, the number of triggers created each week was regressed to the flow volume and found to have a relationship in which flow was 5.13 million times the number of triggers effectively validating the designated trigger volume of 5.2 million cubic feet (ft^3).

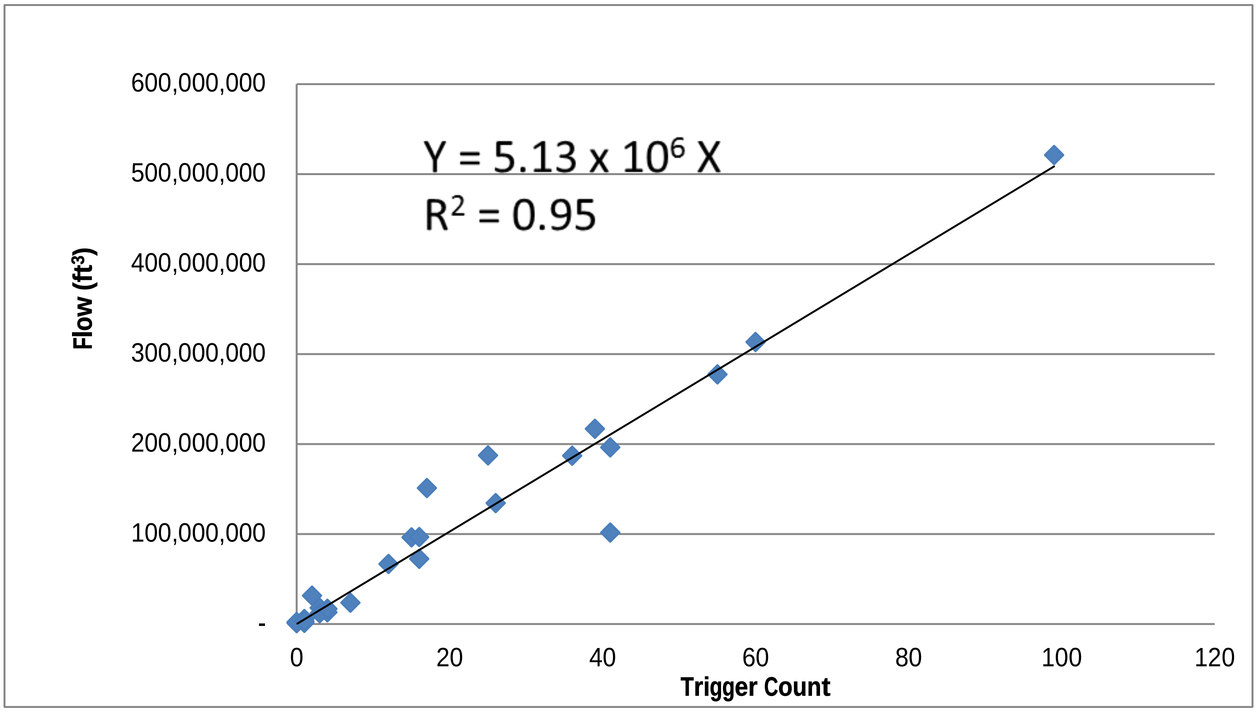

Figure 23. Comparison of trigger counts to flow at G310 confirming the trigger volume of 5.2 million ft^3.

EFFICACY AT G390B

A summary of sampling efficacy at G390B is presented in **Table 2**.

Table 2. Efficacy of monitoring at G390B.

Method	Expected Count	Actual Count	Percentage Success
Grab Samples	35	35	100%
ADT Samples	238	238	100%
Data Sonde Events	16,296	14,243	87%
RPA	2,856	2,031	71%
ACF Samples	34	31	91%
ACF Acceptable NOBs		3	9%

Sample Completeness

The sample completeness results for G390B revealed some surprises. Not unexpectedly, grab and ADT samples showed perfect sampling efficacy. This performance is reasonably attributable to the rather simple technology with few underlying assumptions involved to collect these types of samples.

Similarly, the deployed data sonde functioned effectively missing only 13% of its events. Review found that one entire week of hourly data was lost as a result of a programming error, representing 3% of the monitoring period, but because frequency increased partway through the study, this only represented 1% of all the data. The remaining data losses were the result of power failures that occurred part way through the week and ranged from one to three days of losses each.

As mentioned earlier, reviewing the ACF data (flow-proportional weekly composites) highlights some of the issues with ACF samplers and challenges in defining completeness with a more complex sampling system. A simple look suggests that since the autosampler collected samples 31 out of 34 weeks and had 3 weeks of acceptable NOBs, the percent completeness is 100%.

Magnitude of Flow and Its Representation

However, a review of the trigger-flow relationship (**Figure 24**) revealed serious concerns. Although the sampling data suggests only three events with NOBs, implying three weeks with no or low flow, review of the flow data finds nine weeks with no flow (for two of these weeks, the autosampler collected samples, but the sampling crew failed to record the number of triggers generated) and one week with negative flow producing a single trigger. Perhaps the most confusing of all results are three dates (March 31, April 7, and May 26) in which eight triggers were generated but the weekly flow was reported to be zero. This information generates a serious question: How do weeks with no or negative flows generate triggers?

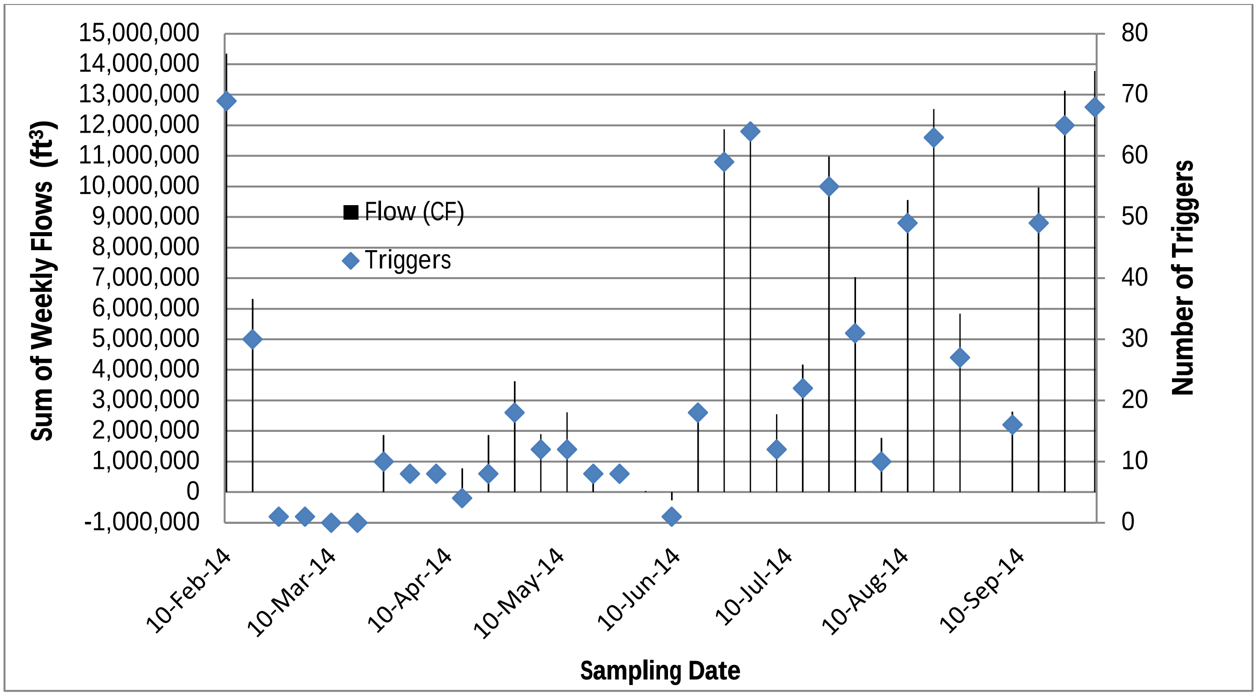

Figure 24. Weekly flow and sampling triggers at G390B for 34 weeks.

Discussions with the staff experts on maintenance of the hydrology data revealed a likely explanation. Just as water quality data have a QA program, so too does the data stream for hydrology. At G390B, the hydrology QA program reviews the instantaneous flow calculations using stage data, and if the differences between upstream and downstream stage gauges are below a certain threshold, the preliminary flow data generated by that difference is set to zero in the database. Unfortunately, the flow totalizer uses the preliminary stage data to calculate trigger volumes, and the analytical results from the sampling cannot be retroactively adjusted once the flow data in the primary database is quality assured. It is important to stress that the process by which preliminary flow values are quality assured to zero is not strictly a flow volume issue, but rather is based on site-specific uncertainty and difference between stage gauges. Consequently, in some cases, flows that result in triggers will be quality assured as real flow, while other flows of similar magnitude will be set to zero. These events that generate water quality samples but are then quality assured out of existence will be referred to by the term “phantom flow”.

That negative flow can trigger the autosampler is the result of an entirely different process. It should be recalled that the flow data in **Figure 24** is a weekly sum. In some weeks, the structure experienced periods of time during which positive flow occurred, while during other periods, the structure experienced negative flow. Contrary to documentation (Abtew et al. 1997), there is no evidence that negative flows are subtracted from the positive flow. Therefore, the negative flows have no impact on the flow total. Consequently, water movement back and forth through a structure will generate positive flow and, therefore, a flow trigger, even if flows for that week sum to a negative value. In simplified terms, once sufficient positive flow results in a sampling trigger, no amount of negative flow can impact the flow totalizer or retroactively eliminate a sample from the autosampler. Fortunately, for the dates with zero or negative flow, the number of aliquots in each sample were too low to dilute the preservation acid and the samples had to be disposed. What is not clear is the level of negative flow that impacts higher level positive flows, which would be masked by the weekly totals.

An investigation of the triggering process documented a cascade of technical issues. As a QA check for the triggering process, the number of triggers created each week was regressed with flow volume

(**Figure 25**). The resulting relationship with flow was 199,000 times the number of triggers. Notably, the trigger volume specified for the original design of the structure was much larger at 312,000 ft^3, so the regression effectively invalidated the designated trigger volume.

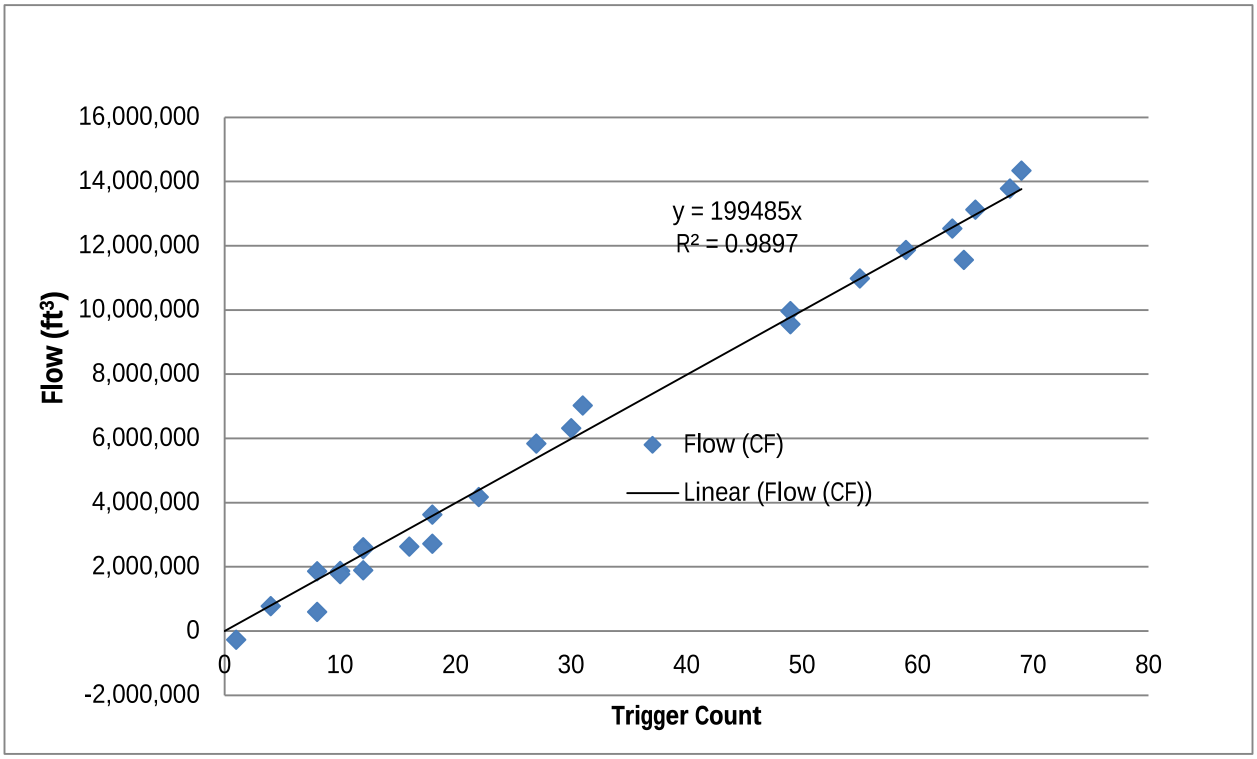

Figure 25. Comparison of trigger counts to flow at G390B, invalidating the trigger volume of 312,000 ft^3.

A review of history at G390B revealed the reason for this large discrepancy and several other issues. In October 2011, the outflow culvert for this structure was modified from a 6-foot by 6-foot square to a 36-inch round, reducing the structure's flow capacity dramatically. While this change in capacity was captured in the hydrologic database, it was not changed in the program that calculates trigger volumes for the autosampler. As a result, the autosampler triggering program thought 312,000 ft^3 had passed, in actuality only around 200,000 ft^3 had gone through. Additionally, with the reduction in size, the trigger volume should also have been adjusted. This new trigger volume should have been 73,000 ft^3, but it was not implemented in the sampling system. The revised trigger volume would have resulted in a greater number of samples being collected at much lower flow volumes. In round numbers, for every 2,000,000 ft^3, the autosampler collected ten aliquots when it should have taken 27 aliquots. Consequently, the water quality at G390B could be viewed as being substantially underrepresented by the sampling program. These problems can be corrected by reprogramming the trigger-flow relationship, but it illustrates the sort of issues that can arise in a complex sampling system. Ways to improve QA for autosampling will need to be considered.

DISCUSSION OF SAMPLING EFFICACY

Grab sample collections depend purely on human action and proved to be the most reliable sampling method, meeting expectations 100% of the time. As the complexity of automated systems increases, so does the rate of sampling failure. Autosamplers programmed to collect at set time intervals were also highly effective. Time-based sampling in this setting is relatively simple. It relies only on a single person properly programming it each week, and is usually supported by a secure power supply and backup batteries.

In contrast, the deployed data sondes had a substantial number of failures, primarily as a result of power outages, and reducing their completeness to 90% or less. According to language presented in monitoring plans, the target for completeness is 95%. However, this number was intended for probe sets that are carried on day trips only, after which they are subject to maintenance. So for long-term deployments, the 95% completeness target might not be applicable. Indeed, for long-term deployments, this target might be problematic. Consider that for a 52-week deployment, each week represents 2% of the time, and therefore, a failure of just 19 days would not meet the completeness target at 95%. Similarly, in a 7-day deployment, a failure of just one day would result in not meeting completeness targets. In the case of deployments of probes, it might be advised to reduce the target to a lower value unless improvements can be derived by careful review of problems with sampling by these systems.

When operating properly, the analytical results of the RPA have been shown to be statistically indistinguishable from other methods and the device provides unique temporal coverage. Still, the results from this study suggest that significant advances in equipment stability need to be achieved before this method is fully deployable. In theory, the RPA has some unique advantages. This equipment can be deployed and left in place for long periods of time without maintenance and the analyzer can take samples with real-time results and more frequently than staff visits would allow. However, with the less than 80% completeness estimated in this investigation, it appears that relying on laboratory staff to service and maintain the equipment on a quarterly basis may not be sufficient. If the technology is to be relied on for compliance with permit requirements, then it might be in the interest of efficacy to create more secure power supplies, invest in more robust pumps, and train field staff in routine maintenance of the equipment.

The analysis of the ACF data collection system reveals several significant issues concerning completeness and loss of information. First, calculating completeness of an ACF program in terms of the number of samples or even data points (including no-flow events) may be inappropriate. This approach to sampling can easily collect 95% of the samples expected during low flow events, but the last 5% could represent substantial flows that dominate the resulting flow-based data set. Instead, it might be more appropriate to calculate completeness in terms of flow captured by the sampler. This becomes particularly clear at G310 where 18% of the samples represented 34% of the data during peak flows. During low flows, 22% of the data represented only 0.5% of the flow. Bear in mind also, that any use of flow assumes that the system is working properly and programmed correctly for its trigger volume.

This leads to the second issue regarding interpretation and use from data from ACF composites. If the ACF sampling system is properly programmed and the trigger properly calculated and executed, the volumes of flow missing from the ACF results should be relatively small, and are unlikely to have a significant effect on the flow-weighted mean. However, when flow-weighted means are calculated, data analysts routinely supplement the low flow, NOB autosampler results with data from grab samples. This practice seems to represent a fundamental disconnect between the designs of the ACF program with rare or no samples at very low flows, and the data end users who seem to desire data for all flows, no matter how insignificant.

A third issue revealed by the REST study also relates to the NOB samples. At G310, a small number of NOB events actually had samples in them, but the sample volume was too small to offset the acid used as preservative and, therefore, the samples were not able to be analyzed by the laboratory. In practice, this means that at stations where this occurs, the minimum flow volume needed to create a viable sample is at least twice the trigger volume, and not the trigger volume itself. However, as these still represent small quantities of flow it may not be a critical issue, but review of the flow volume trigger in combination with changing the sample volume size might be able to resolve this issue.

Another issue here is derived from the hydrology QA program. In the process of data review, when the preliminary data used to trigger the autosampler are based on very small upstream to downstream, increments of stage can be set to zero, creating in the flow totalizer phantom flow. While reasonable, this practice negates work by both the sampling team and the analytical laboratory. Additionally, if the phantom flows are significant they can create an inordinate amount of aliquots in the autosampler that should not have been collected, thereby skewing the analytical results.

Similar issues arise with the treatment of negative flows. It was believed that negative flows were subtracted out of the autosampler triggering program, but the REST analysis shows no evidence of this. In actuality, negative flows are set to zero and therefore do not impact the autosampler triggering program directly. This situation could lead to a condition in which the autosampler triggers on a given positive flow, does not trigger on the same volume when it returns, but then triggers once again when the same water becomes positive flow once more. In this manner, the same volume of water could be sampled multiple times, at least in theory, and the subsequent analytical result would be skewed.

Low flows can also skew results. If a flow volume is insufficient to trigger a sample, the volume carries over from one week to the next. In the new week the carried over volume may be sufficient to result in more triggers than should occur, thereby skewing analytical results, particularly in events with few triggers.

Reviewing the flow-based autosampling program in terms of accuracy, here defined as the difference between the desired flow volume trigger and the actual flow volume trigger used to collect samples, G310 worked quite well in terms of accuracy, but the sampler deployed at G390B did not. In fact, G390B had multiple layers of failure using a triggering volume and flow equation designed for a specific-sized structure. The structure was modified, but the flow equation and trigger volume were not updated.

Evaluating the flow-based autosampling program in terms of precision, here defining precision as the variability in flow volume between events, indicates that both negative flows and phantom flows likely decrease precision. This decrease in precision also likely skews the timing and number of aliquots, thereby skewing the analytical results. Ultimately, this suggests that at structures with questionable preliminary flow data, and/or the possibility of negative flow, the ability to properly implement a flow-proportional autosampler is questionable.

These issues can be related to the complexity of the system relative to organizational responsibilities and process design. The official flow program is maintained in one database, while the autosampler flow totalizer is based in another. Maintaining identical flow programs in both pathways can be problematic, increasing the chances of failure. The other issue is the lack of an auditable process for using the structure size and flow equation to calculate a trigger volume and then implementing that in the sampler. Likewise, there is no auditable process for changing the trigger volume or flow equation.

This complexity reveals a fundamental organizational factor. Consider that the flow equation is calculated in one work unit, the trigger volume is calculated in a second unit, they are implemented by a third group, and the equipment is regularly used and maintained in part by a fourth unit having no authority over any of the others, while a fifth group carries out chemical analysis, which is interpreted and reported by a sixth group. Thus, responsibility for autosampler operations and reporting is spread over a wide span of the District creating opportunity for breakdowns in maintenance, communication, operational understanding, and control, which leads to a potential failure to understand and interpret results in context.

VALIDATION OF INSTALLATION

Fundamental to the study is the concept that the grab sample, RPA, ACF, and ADT are sampling from the same water mass. To assure that the sampling equipment intakes were installed as close as possible, and to validate the sampling comparative accuracy, multiple samples were taken during a single day. The rate limiting step here was the RPA sampling that takes approximately an hour to cycle through its sampling and analysis process. However, in order to create a more robust database, all other samples were collected every 45 minutes. All other samples, including routine grabs and grab samples collected using the autosampler tubing and peristaltic pumps, were preserved on site and then analyzed by the SFWMD laboratory. The results for G310 are shown in **Table 3**, while the results for G390B are shown in **Table 4**.

Table 3. Validation samples for G310.

Time of Sample	TP Grab	TP ACT	TP ACF	TP RPA	Mean	Maximum	Minimum	Standard Deviation	20% of Mean	Pass?
7:00	36	35	36	39	37	39	35	1.73	7	Yes
7:45	36	36	35		36	36	35	0.58	7	Yes
8:30	36	37	36	42	38	42	36	2.87	8	Yes
9:15	37	39	35		37	39	35	2.00	7	Yes
10:00	35	40	36	39	38	40	35	2.38	8	Yes
10:45	37	37	38		37	38	37	0.58	7	Yes
11:30	36	41	37	42	39	42	36	2.94	8	Yes
12:15	36	36	37		36	37	36	0.58	7	Yes
13:00	36	36	36	42	38	42	36	3.00	8	Yes
13:45	35	39	39		38	39	35	2.31	8	Yes
14:30	34	34	34	43	36	43	34	4.50	7	Yes
15:15	35	37	41		38	41	35	3.06	8	Yes
16:00	36	38	36	43	38	43	36	3.30	8	Yes
16:45	36	37	38		37	38	36	1.00	7	Yes
17:30	37	36	38		37	38	36	1.00	7	Yes
Mean	36	37	37	41						
Maximum	37	41	41	43						
Minimum	34	34	34	39						
Mode	36	36	36	42						
Geometric Mean	36	37	37	41						
Standard Deviation	1	2	2	2						

Table 4. Validation Samples for G390b.

Time of Sample	TPO_4 Grab	TPO_4 ACT	TPO_4 ACF	TPO_4 RPA	Mean	Maximum	Minimum	Standard Deviation	20% of Mean	Pass?
8:00	17	18	17	19	17	19	17	0.6	3	Yes
8:45	17	16	17		17	17	16	0.6	3	Yes
9:30	16	18	17	21	17	21	16	1.0	3	Yes
10:15	18	17	17		17	18	17	0.6	3	Yes
11:00	18	17	18	18	18	18	17	0.6	4	Yes
11:45	18	17	19		18	19	17	1.0	4	Yes
12:30	19	18	18	19	18	19	18	0.6	4	Yes
13:15	19	18	18		18	19	18	0.6	4	Yes
14:00	18	17	17	21	17	21	17	0.6	4	Yes
14:45	20	16	17		18	20	16	2.1	4	Yes
15:30	17	16	17	18	17	18	16	0.6	3	Yes
16:15	18	16	17		17	18	16	1.0	3	Yes
17:00	17	16	17	19	17	19	16	0.6	3	Yes
17:45	17	17	17		17	17	17	0.0	3	Yes
18:30	17	16	17		17	17	16	0.6	3	Yes
Mean	18	17	17	19						
Maximum	20	18	19	21						
Minimum	16	16	17	18						
Mode	17	16	17	19						
Geometric Mean	18	17	17	19						
Standard Deviation	1.0	0.8	0.6	1.3						

At both sites, differences between the grab and pumped sampling methods were relatively small, averaging 1 to 2 µg/L, but ranging up to 4 µg/L at G390B and 6 µg/L at G310. At G310 the range between paired pumped samples was as high as 4 µg/L. However, all of these results were within 10% of the mean, well below a 20% metric. This 20% metric was chosen as this value is used to validate field duplicates. If field duplicates (two distinct samples taken from the same site in rapid succession) are within 20% of each other, they are considered to be statistically indistinguishable for QA purposes. Thus, all the results from the grab samples are functionally within the acceptable field sampling error.

The RPA data produced higher minimums, means, and maximums at both stations. However, at both stations, 5 of the 7 samples were within 10% of the mean of the grab and pump data, and all results were within 20%. Despite this, it must be noted that RPA results at G390B were on average 2 µg/L higher than the grab results and ranged up to 5 µg/L higher. Similarly, at G310 RPA, results were on average 5 µg/L higher than grab results and ranged up to 9 µg/L higher. This may indicate some systematic bias in the sampling or in the analytical methods. Given that the grabs and pumped grabs showed a potential difference of up to 4 or 6 µg/L depending on location, the systematic differences created by using distinct sampling and analytical processes should not be unexpected.

The integrity of the autosamplers was maintained through weekly checks on the autosamplers by collecting pumped (grab pumped or GP) samples from each unit and comparing the TP values to those of the grab. **Figure 26** presents the data from G310. As shown, the three results showed similar trends through time, except for only one event (July 28, 2014) in which the grab sample resulted in a value of 5 µg/L, while the GP samples produced values of 17 µg/L (removed from the figure). Correlation analysis of the GP versus the grab resulted in linear r^2 values greater than 0.75. The same analysis repeated with the July 28, 2014, data removed linear r^2 values greater than 0.90.

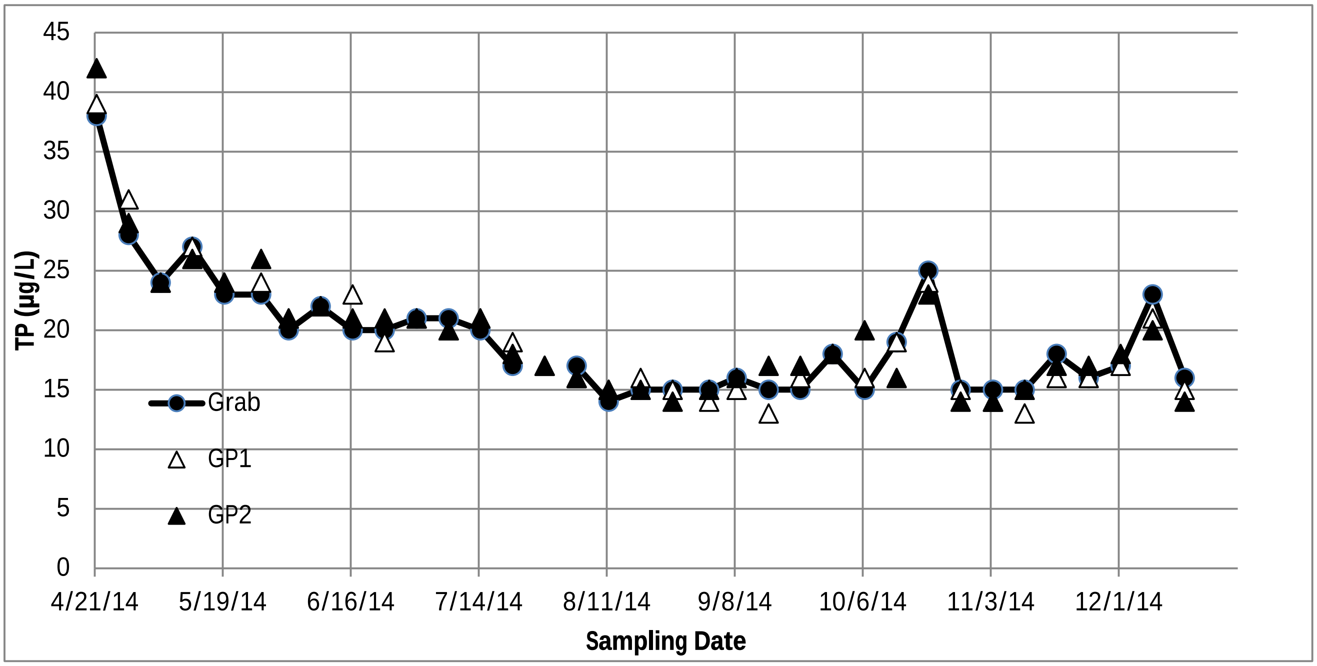

Figure 26. Comparison of TP concentrations from weekly grab and GP samples at G310, with one outlier removed.

Figure 27 shows the same results for G390B and again shows a strong relationship between the three sample types. For this data set, the linear r^2 values were greater than 0.91. In combination, these data strongly suggest that the sampling equipment associated with the autosamplers did not influence or bias the sampling results, and that the installation of the equipment was sufficiently close as to supply comparable estimates for the water mass.

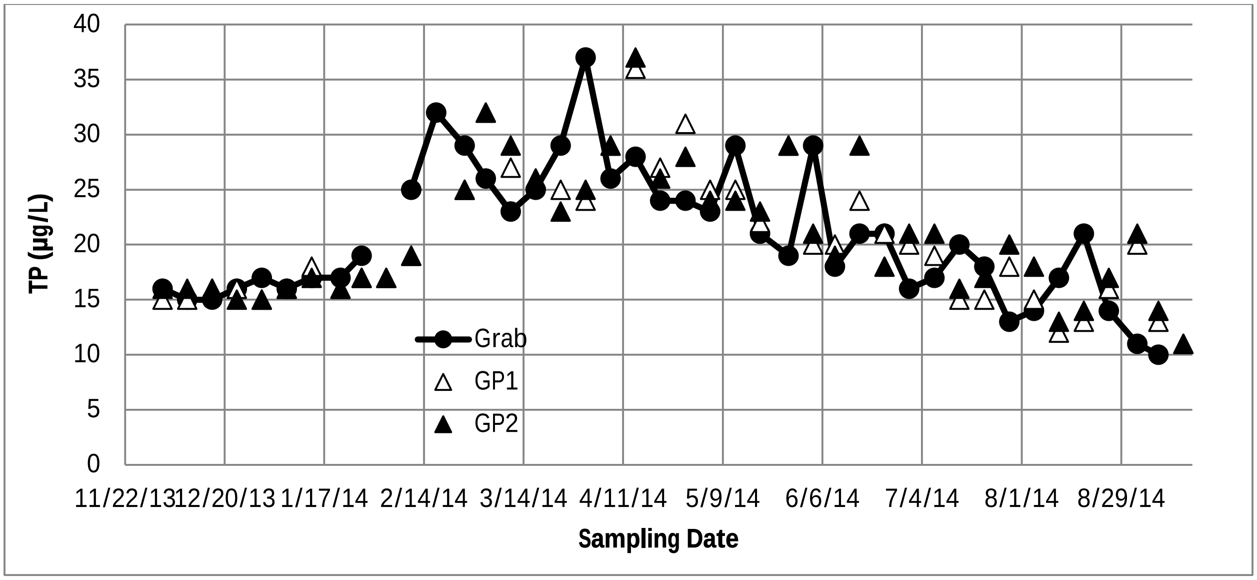

Figure 27. Comparison of TP concentrations from weekly grab and GP samples at G390B.

However, as a further check on comparability of data, the daily values produced using the G310 ADT were compared to the mean values generated from the G310 RPA data for that same day. Unfortunately, the relationship was not as strong as would have been preferred with approximately 60% of the RPA daily means being more than 20% different than the ADT value for the same day. An investigation into this difference revealed a possible procedural factor. The clock on the RPA was not adjusted for daylight savings time, while the clock on the autosampler was. Since almost the entire period of the study was within the period of daylight savings time it must be concluded that the paired collections for the RPA and the ADT were actually one hour apart. Thus, while in the initial setup sampling variability was somewhat controlled by sampling at the same time, leaving only analytical variability as a factor, in the study variability was impacted not only by method differences, but also by differences in sampling time generated by the occurrence of daylight savings time.

DATA TRENDS

Data trends for G310 for the deployed data sonde, the RPA, the grab, ACF, and ADT are shown in the following graphics (**Figure 28** through **Figure 33**), limited to the core 34 weeks identified previously.

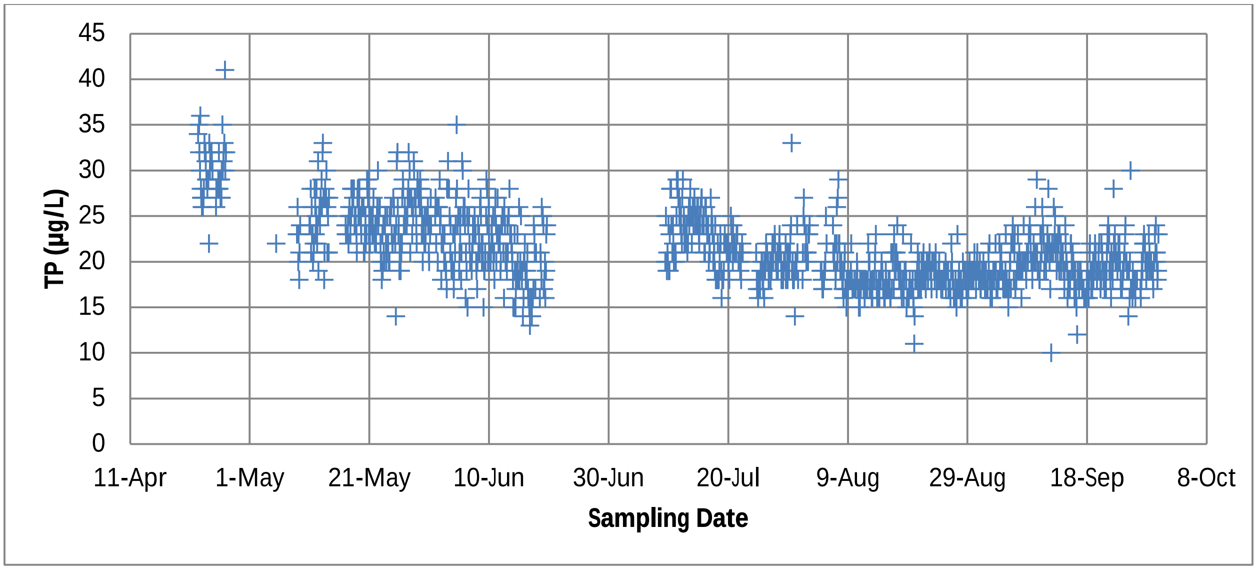

Figure 28. TP data collected from the RPA every two hours at G310, and showing equipment failure in June and July.

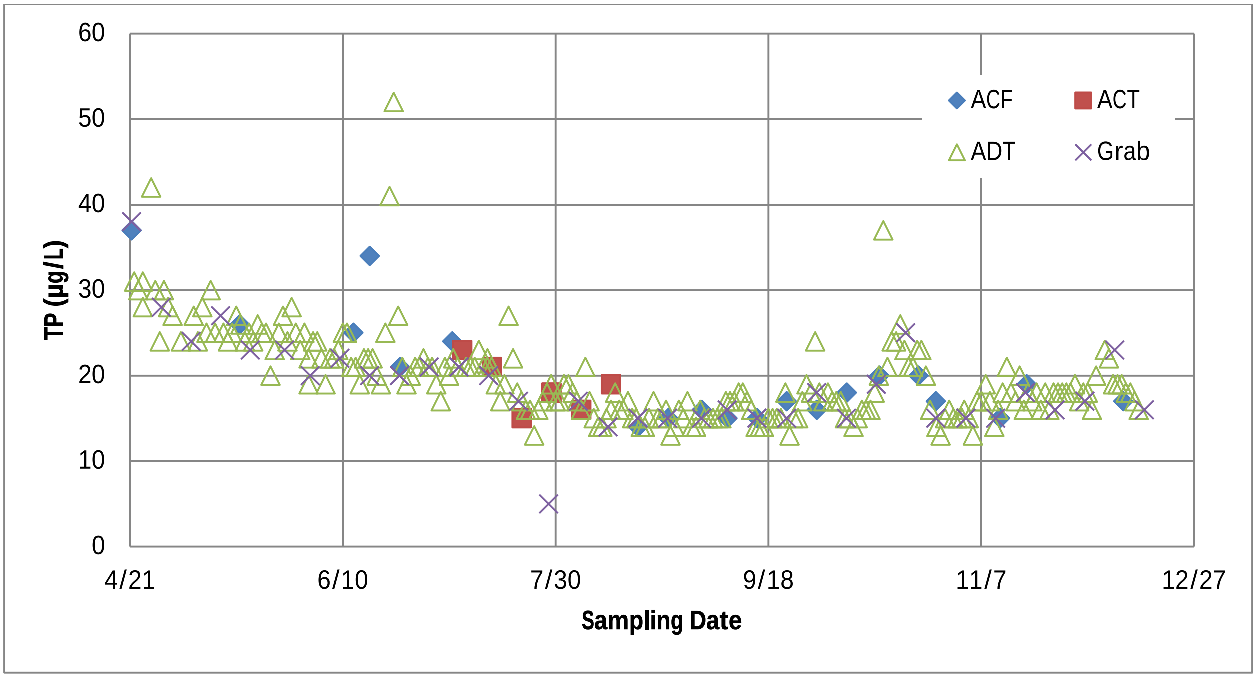

Figure 29. TP from autosamplers and grabs, with one ADT value of 111 µg/L on May 2 removed. Sampling dates for autosamplers are final collection time.

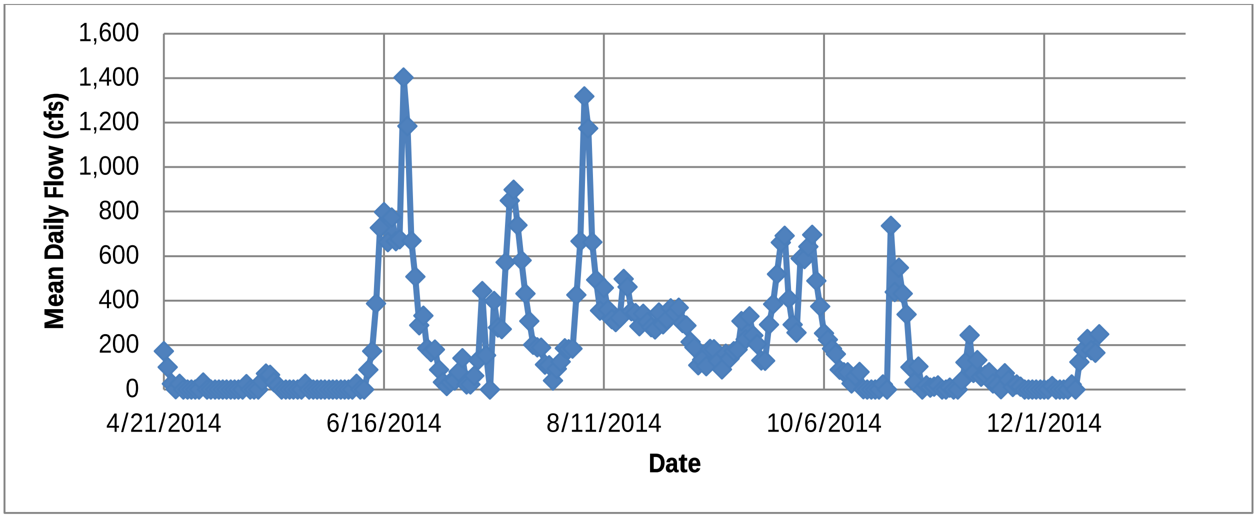

Figure 30. Mean daily flow in cubic feet per second (cfs) at G310.

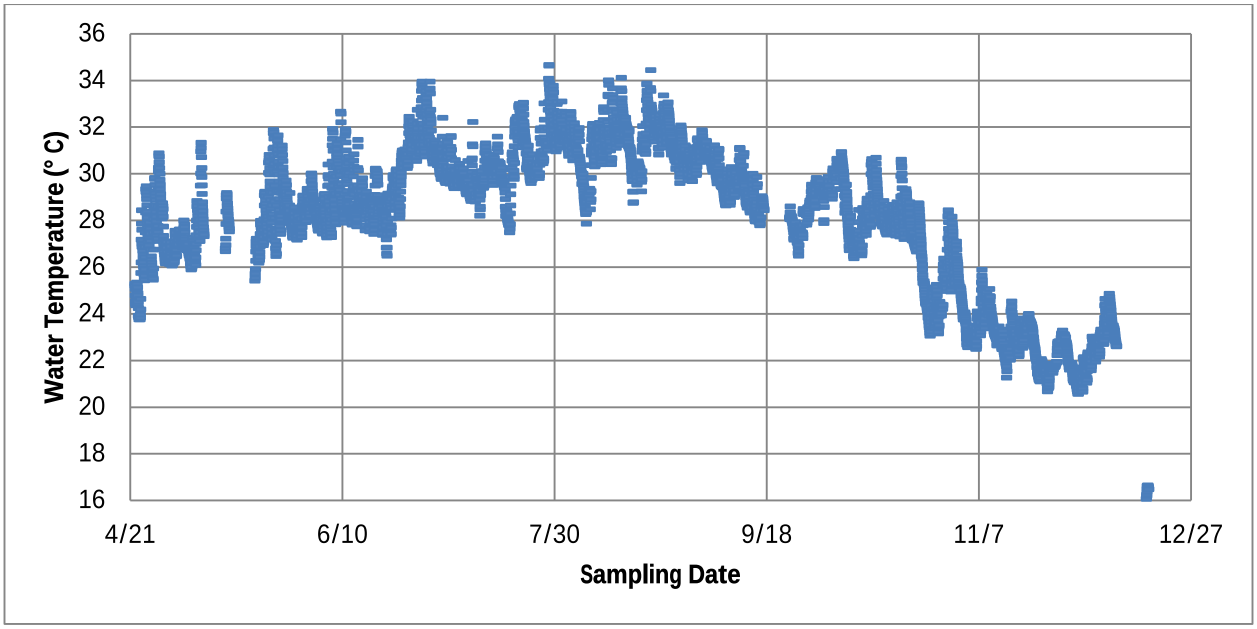

Figure 31. Water temperature in degrees Celsius (°C) at G310 measured every two hours.

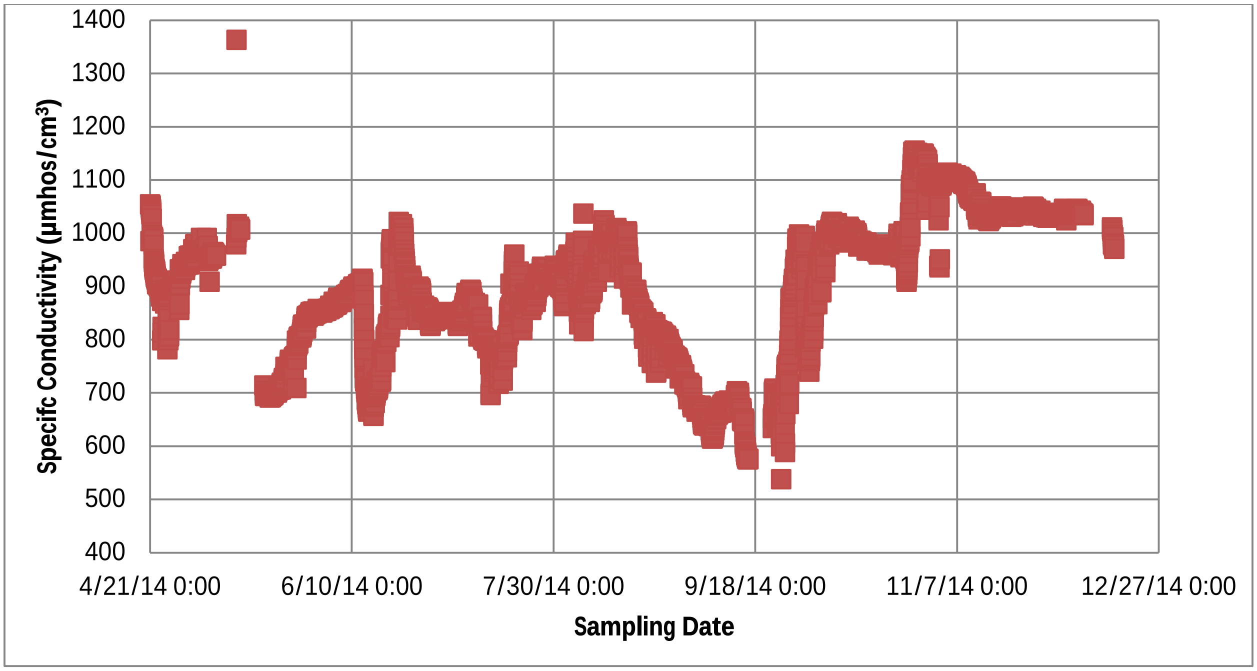

Figure 32. Specific conductivity in micromhos per cubic centimeter (µmhos/cm³) at G310 collected every two hours.

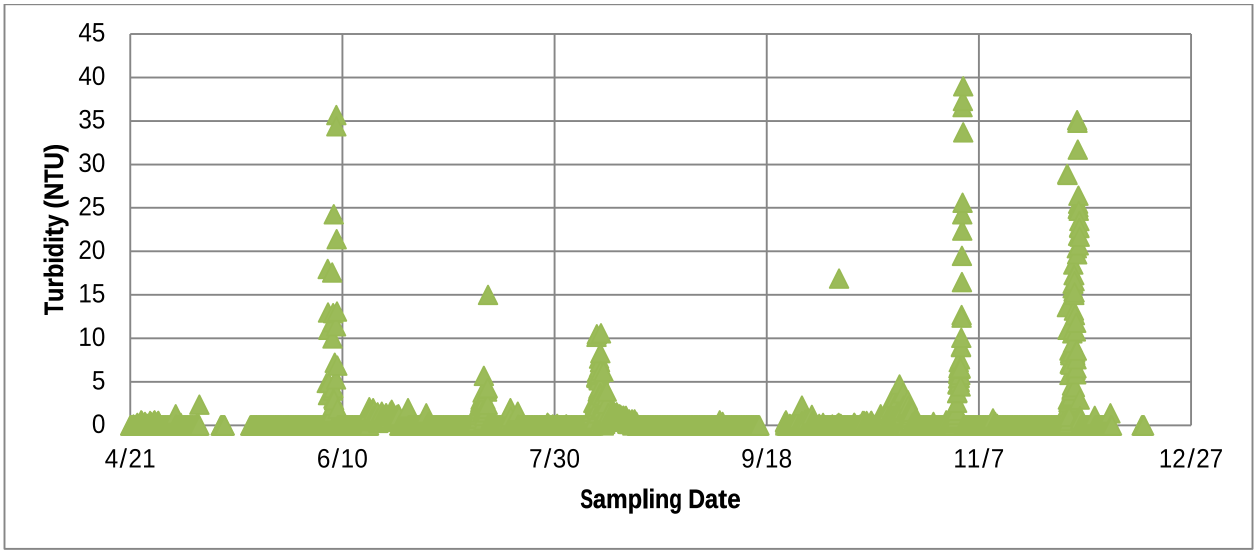

Figure 33. Turbidity in nephelometric turbidity units (NTU) at G310 collected every two hours.

As shown in **Figure 28** and **Figure 29**, TP values at G310 ranged from 10 to 60 µg/L (with one outlier grab sample dropping down to 5 µg/L). In general, results from all four methods showed the same trends, though the results from the RPA did not range as high as the other methods. This seems to contradict previous observations on paired data, which suggested that RPA data might be biased higher than laboratory analysis. Additionally, it should be noted that the results from the ADTs ranged higher than results from grabs and ACFs. These higher TP concentrations do not appear to be associated with days of peak flows but rather seem to occur on days when flow is staging down (**Figure 30**). It may be possible that, in these cases, the ADT is capturing samples when the structure is transitioning from one state of operation to another creating a seiching effect that would include brief periods of still water or even a back flowing wave. Under these conditions, particles that would normally be moving too fast to be captured might be moving slow enough to be taken up by the autosampler.

Water temperatures showed the influence of both diel processes and seasonal processes, and this is particularly noticeable in the period from August to December, which shows daily increases in temperature during the day, and decreases at night, while at the same time slowly dropping in response to a seasonal change (**Figure 31**).

Turbidity measurements showed that the vast majority of the data were below 5 nephelometric turbidity units (NTUs), with the dominant value being less than 1 NTU (**Figure 33**). Periodic spikes in turbidity were short lived and ranged up to 40 NTUs. While these spikes in turbidity seem to occur during increases in flow (**Figure 30**), an increase in flow was not always accompanied by an increase in turbidity, and suggests that there might be other factors in play.

Not apparent in the full data set is a diel cycle in RPA TP concentrations that can be seen in daily data or in the average of several consecutive days. As shown in **Figure 34**, TP concentrations from1800 to 0600 averaged below 23 µg/L, while values collected between 0800 and 1600 averaged above 23 µg/L. While the entire range only spans from 21 µg/L to 25 µg/L the periodicity, although slight, appears real.

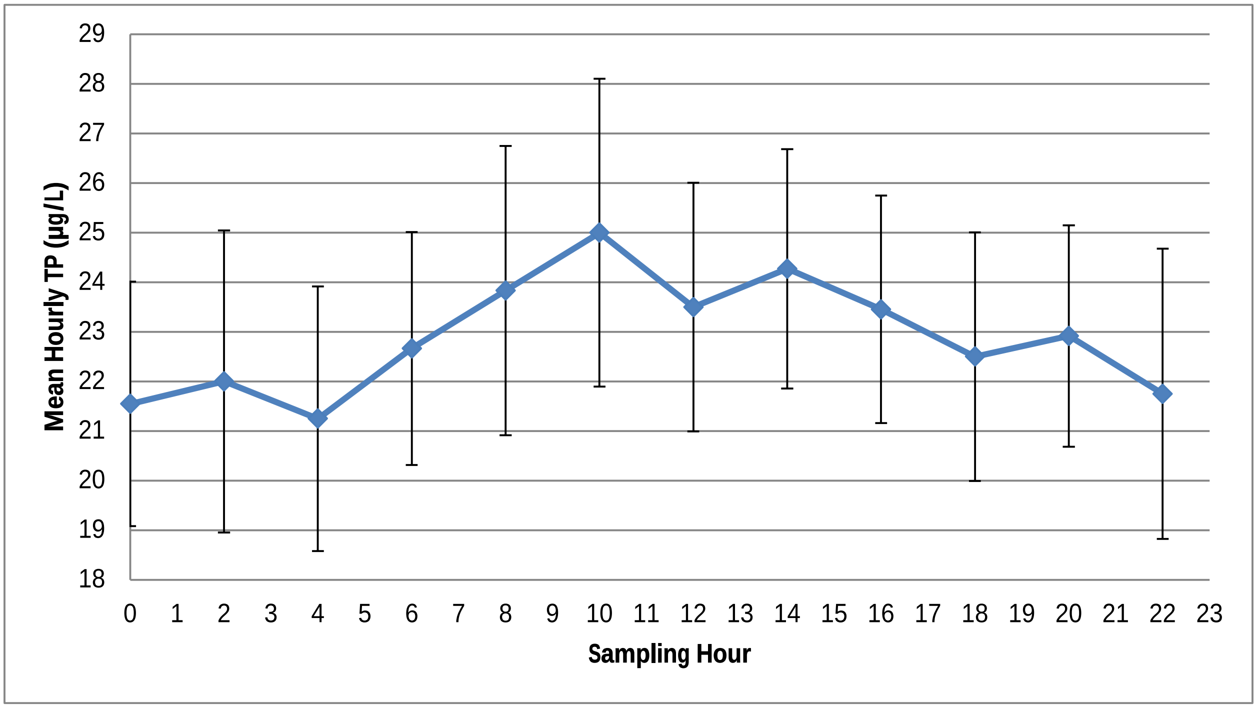

Figure 34. Average hourly TP data at G310 from 12 days in July 2014 with standard deviation bars

Data trends for G390B for the deployed data sonde, RPA, grab, ACF, and ADT are shown in the following graphics (**Figure 35** through **Figure 40**), limited to the core 34 weeks identified previously.

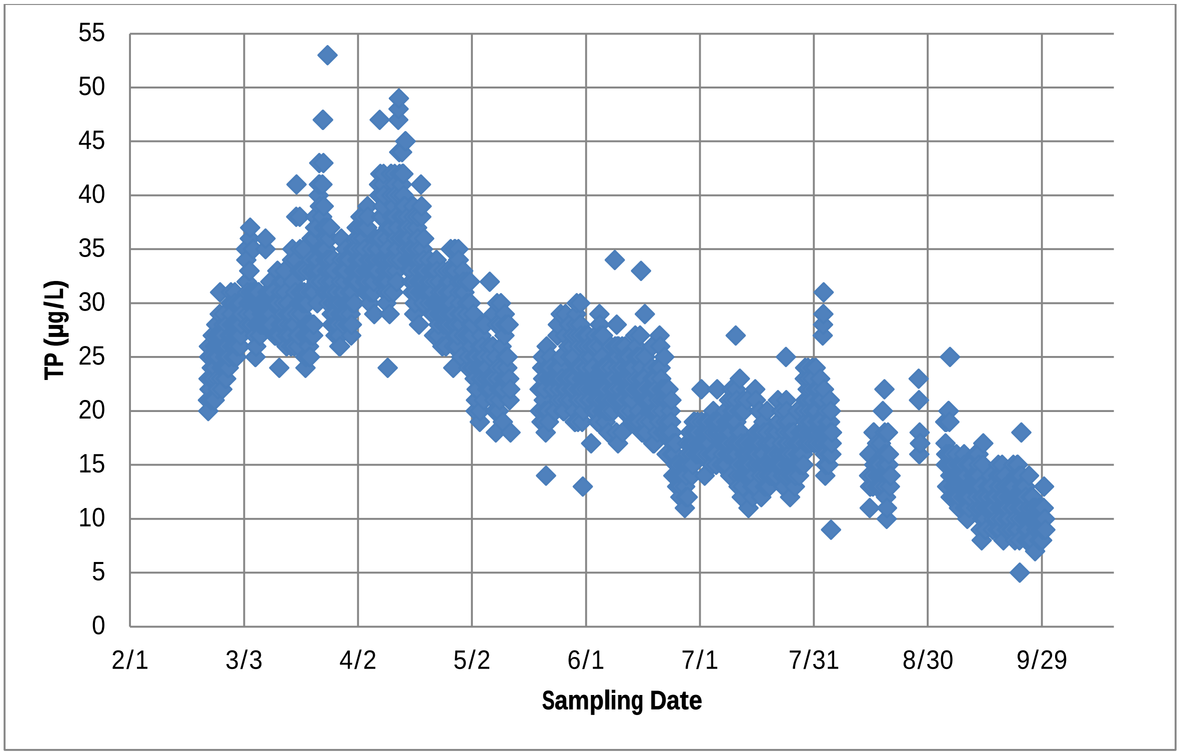

Figure 35. TP data collected and analyzed via RPA every two hours at G390B.

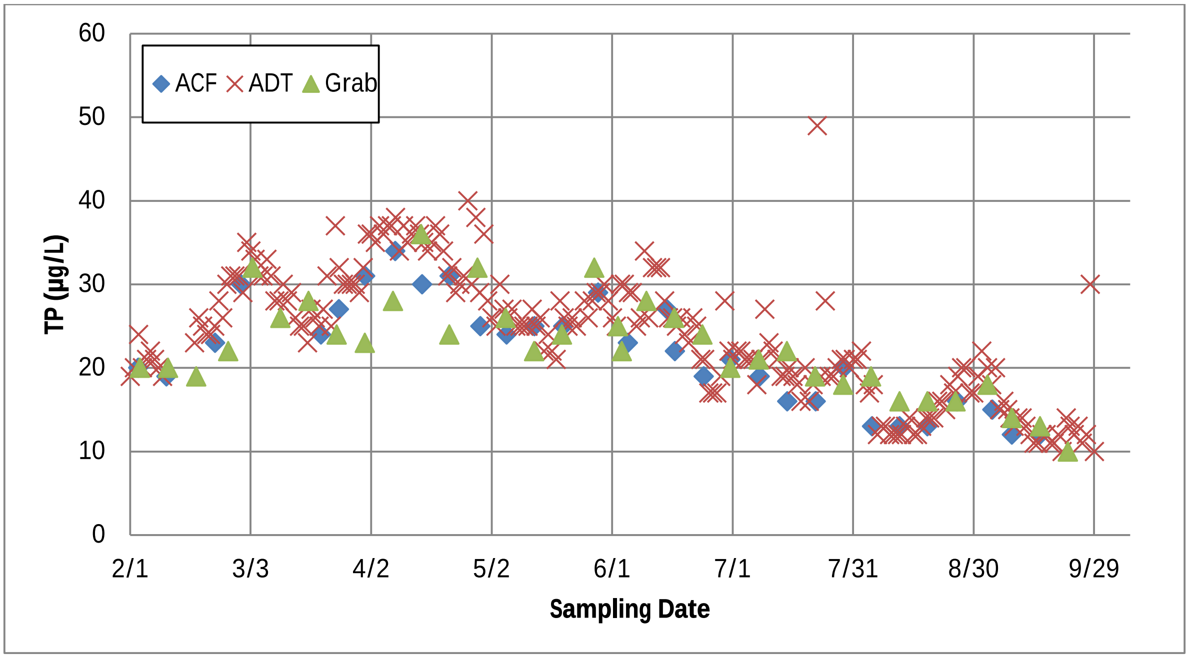

Figure 36. TP data collected by grab, ACF, and ADT at G390B.

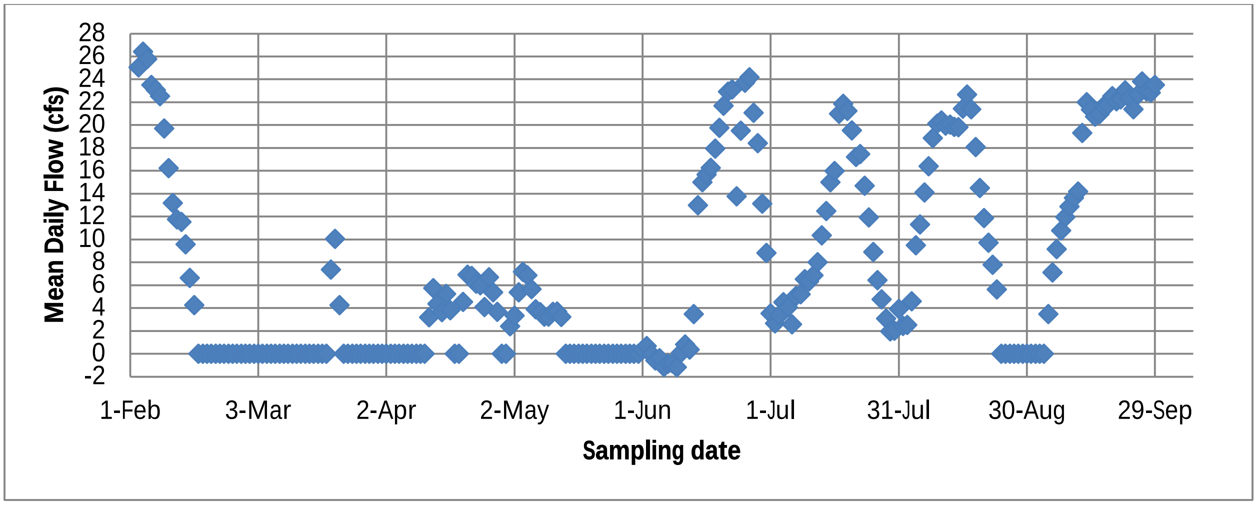

Figure 37. Mean daily flow in cubic feet per second (cfs) at G390B.

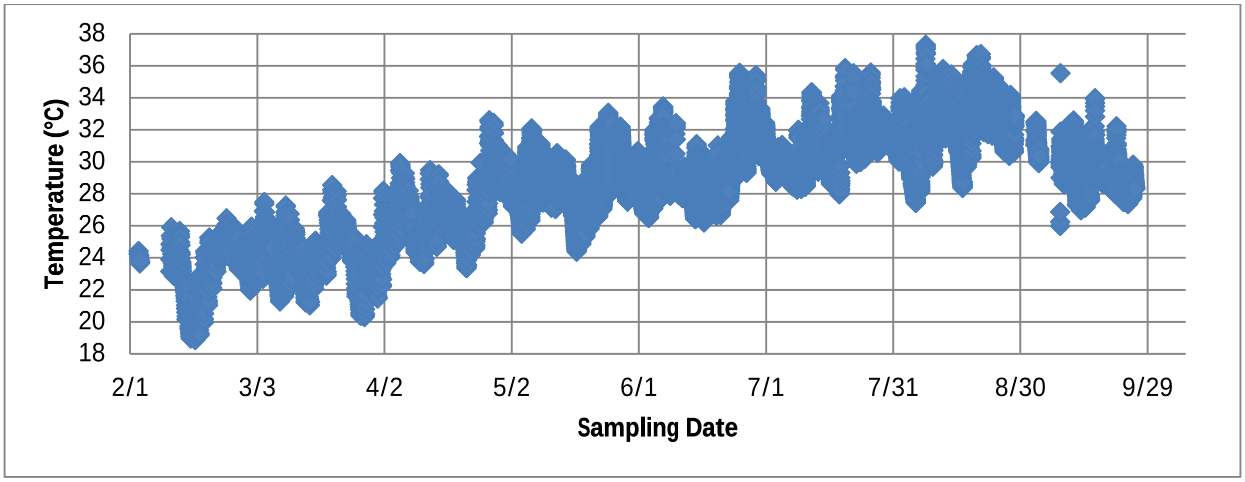

Figure 38. Temperature in degrees Celsius (°C) at G390B.

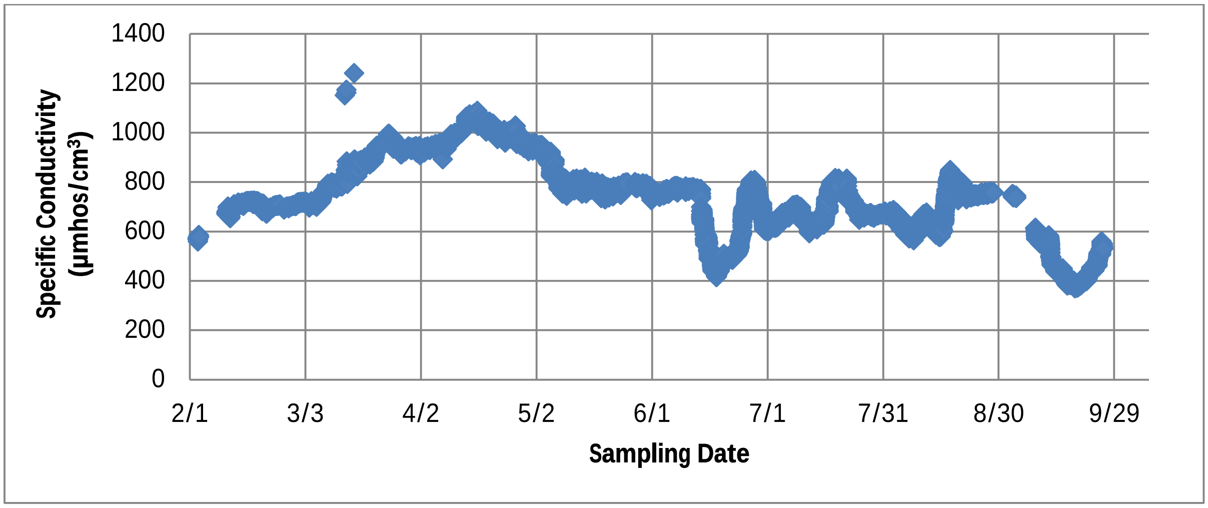

Figure 39. Specific conductivity in micromhos per cubic centimeter (µmhos/cm^3) at G390B.

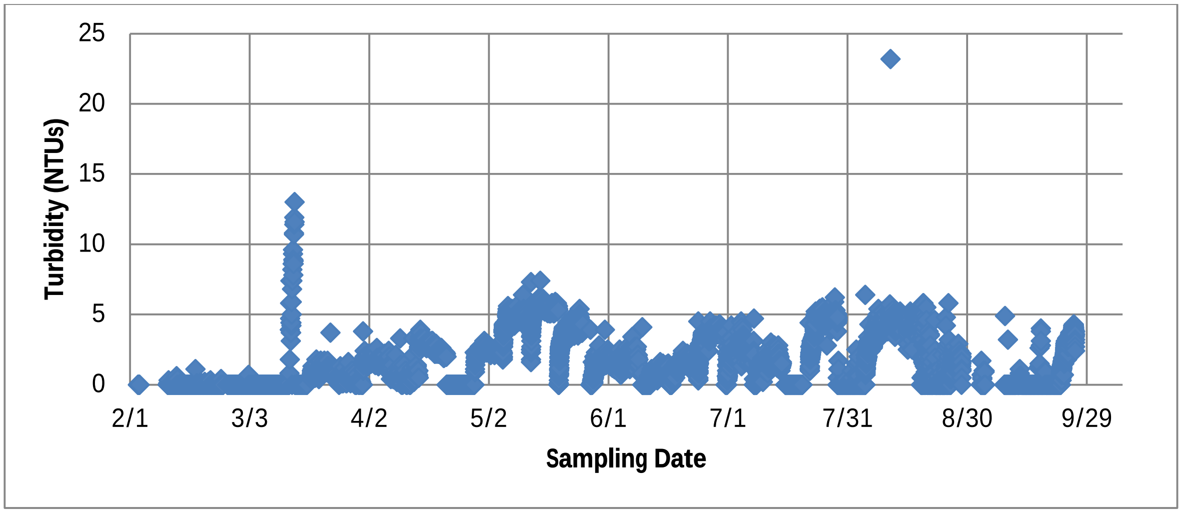

Figure 40. Turbidity at G390B.

As shown in **Figure 35** and **Figure 36** TP values ranged from 5 to 55 µg/L (with one RPA sample with a value of 5 µg/L). Data clouds in these figures suggest that variation from natural sources, sampling and chronic interferences produced a range of about +- 5 µg/L and a few acute events are visible in both data sets. In general, results from all four methods showed the same trends, with the results from the RPA trending slightly higher and lower than the other methods. This seems to contradict previous observations, which suggested that RPA data might be biased higher than laboratory analysis. Additionally, it should be noted that the results from the ADTs ranged higher than results from grabs and ACFs. These higher TP concentrations do not appear to be associated with days of peak flows.

Water temperature showed both diel and seasonal trends (**Figure 38**), but neither specific conductivity (**Figure 39**) nor turbidity (**Figure 40**) showed any relationship to flow or season. Investigation of periodic spikes in the turbidity data once again suggest that increased flow was a factor in increased turbidity, but not always (**Figure 37** and **Figure 40**).

Not apparent in the mean daily flow data (**Figure 37**) is the actual extent of reverse flow at the structure. In order to highlight this issue, the mean hourly data is shown in **Figure 41**.

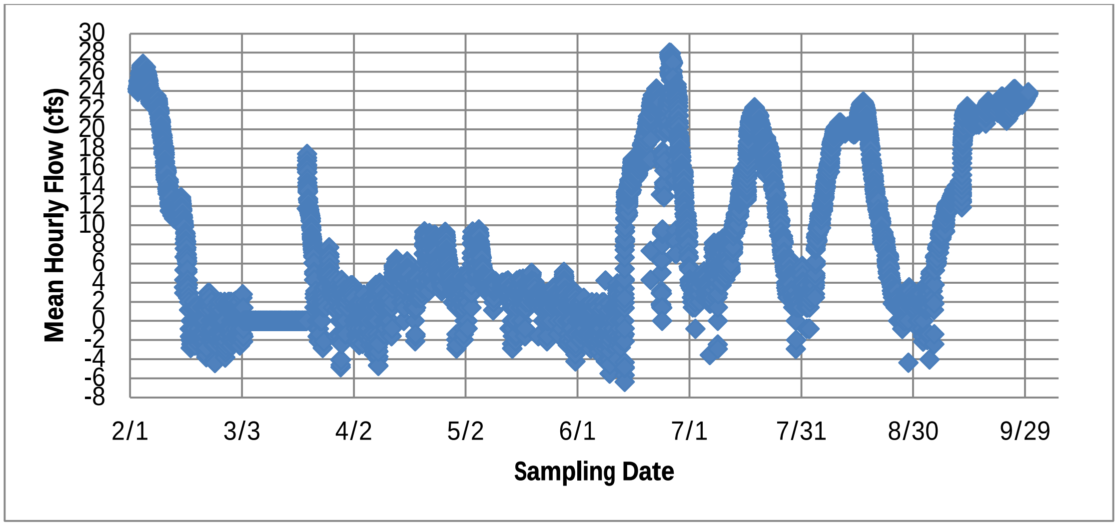

Figure 41. Mean hourly flow in cubic feet per second (CFS) at G390B showing extensive negative flow.

EVIDENCE OF DIEL PATTERNS

As mentioned in the previous section, there is some suggestion in the TP data that a diel pattern exists in water column concentrations. In order to highlight this pattern, data from G390B for each hour of each day was averaged together for each month (**Figure 42**) creating a mean for each hour of the day by each month.

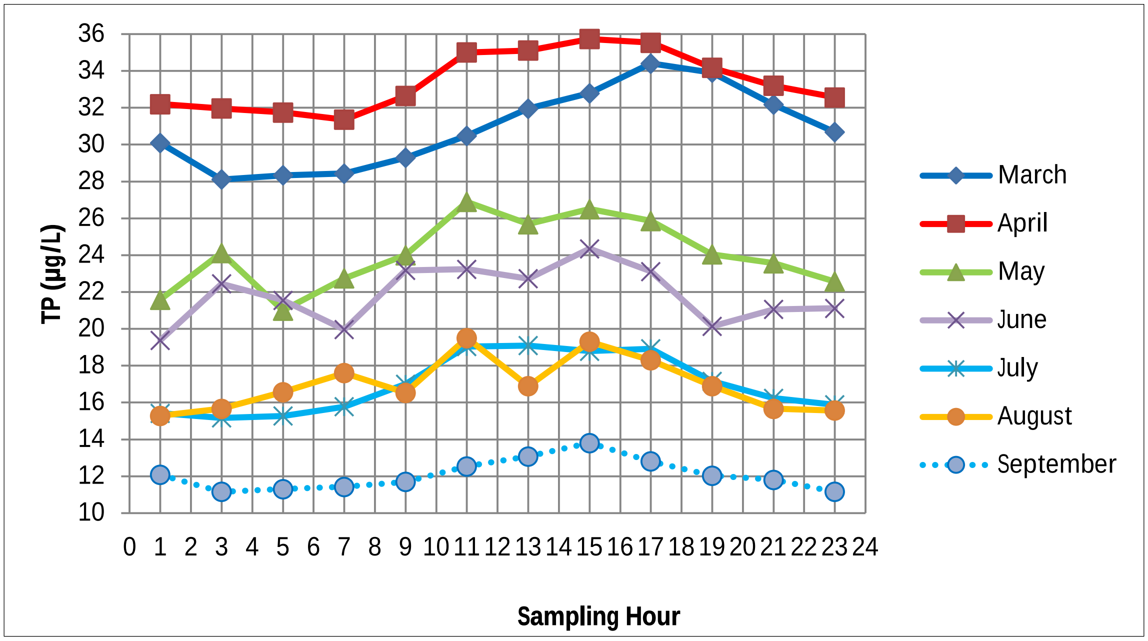

Figure 42. Mean hourly RPA TP at G390B by hour separated out by month.

At first it was thought that this pattern was the result of processes in the marsh influencing the behavior of TP at G390B. However, as shown by **Figure 43**, a similar pattern exists at G310 that is significantly disconnected from the marsh.

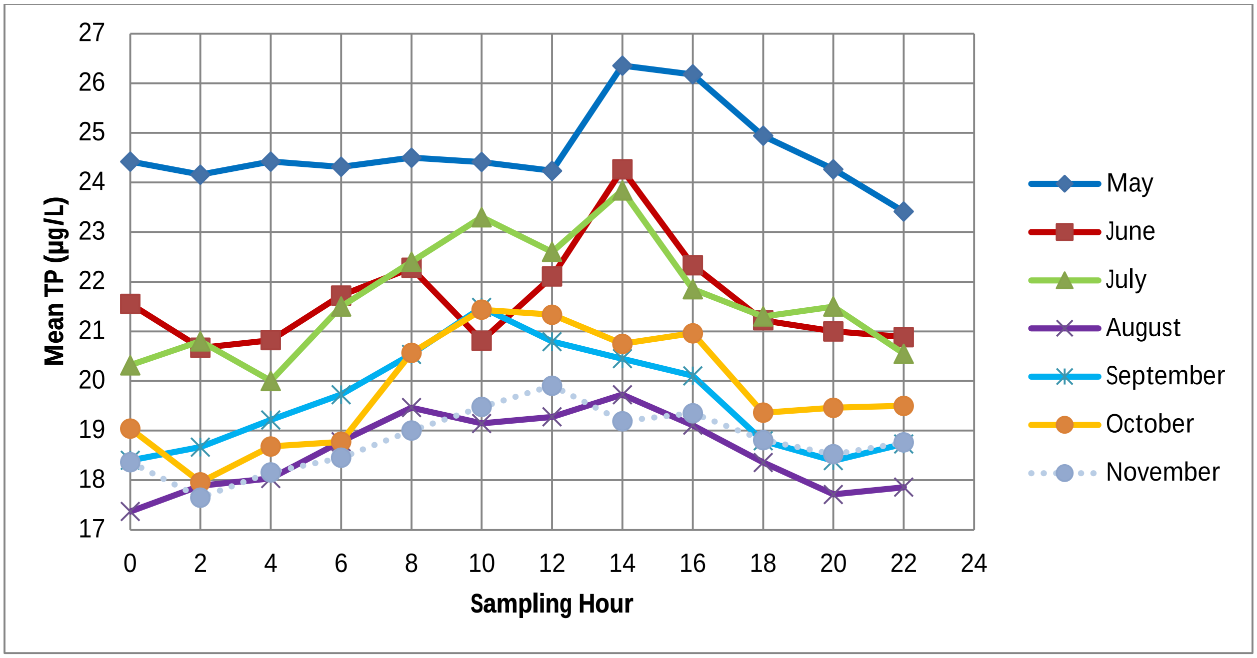

Figure 43. Mean hourly RPA TP at G310 by hour separated out by month

While the diel cycle is clear at both cases, and over several months, the actual extent is difficult to see. **Figure 44** and **Figure 45** present the same data with each data set normalized to the lowest average hourly value for the month.

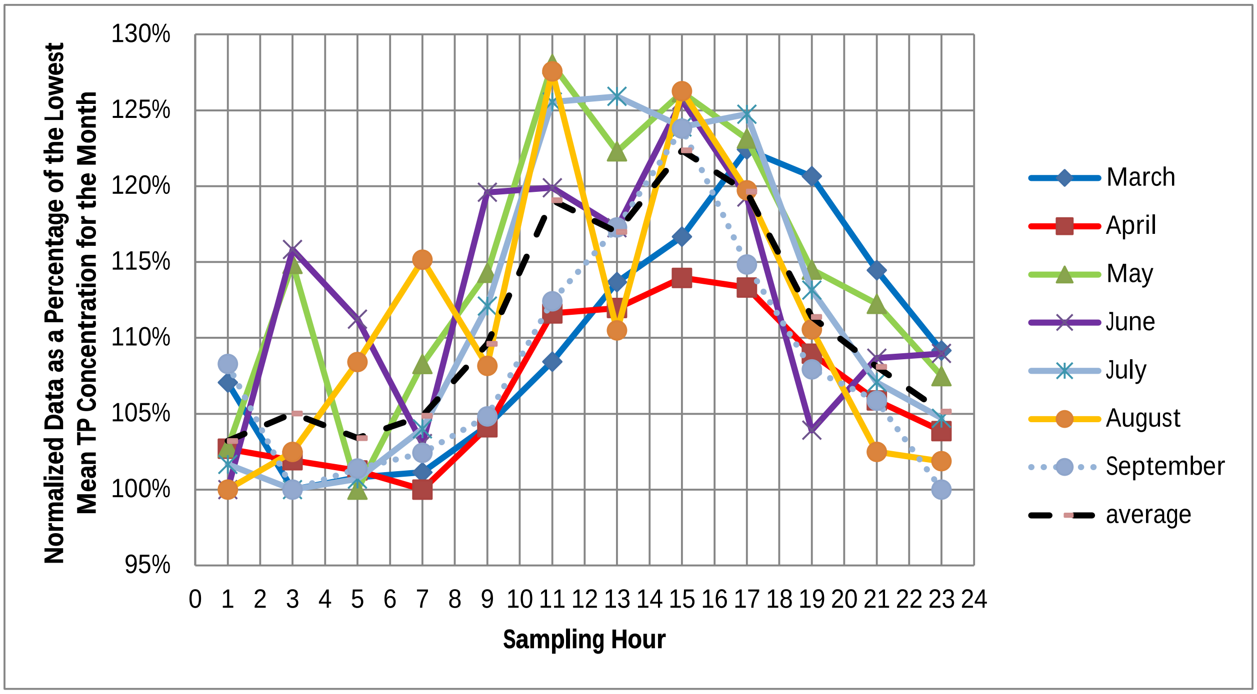

Figure 44. Normalized mean hourly diel changes in TP at G390B.

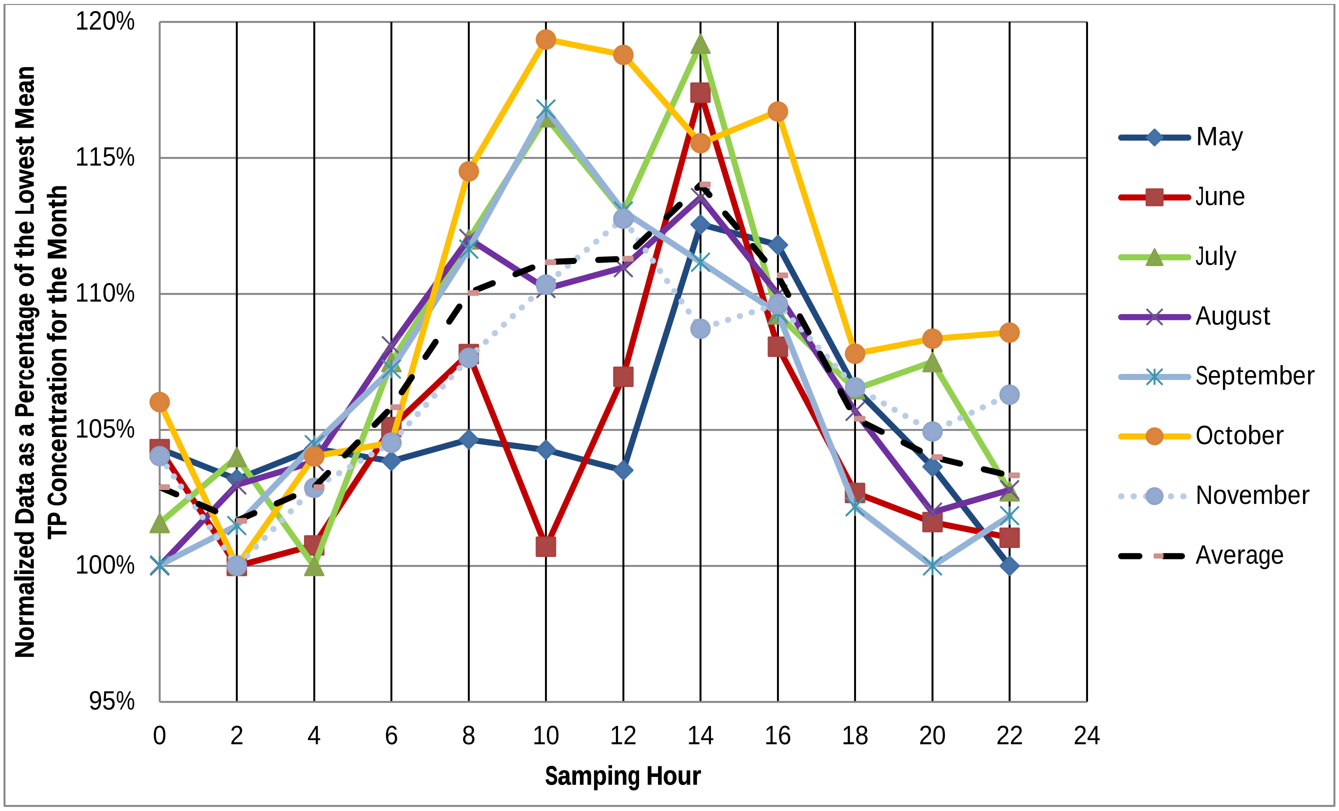

Figure 45. Normalized mean hourly diel changes in TP at G310.

As shown, the mean TP value can change by up to about 25% during the course of the day, with the lowest values being in the night and the highest values peaking between 1200 and 1400. The exact cause of this is unknown, but it presents an interesting issue for representative sampling and downstream TP loading. If the diel cycle is real and independent of flow events, it might be possible to manipulate discharge activity to periods of the day when TP is lower and reduce downstream loading slightly.

It might also suggest an issue with the timing of grab sampling. It is the nature of the water quality sampling process that most stations are visited between 1000 and 1400 with very few samples taken before 0800 or after 1600, reflecting the core operating hours of 0700 to 1700 of most staff, and the need to account for travel time to and from the sampling sites and sample processing time. Consequently, most grab samples could be considered to be collected at the peak TP concentrations of this possible diel cycle. This seems to contradict the generally accepted concept that autosamplers tend to produce higher concentrations than grab samples. However, if grab samples were biased toward the earlier part of the day than pump activity and, therefore, autosampler events based on flow, this might explain such a pattern. The Project REST RPA and other data deserves additional analysis on circadian patterns in concentration.

IMPLICATIONS OF THE DISTRIBUTION OF RPA DATA

TP data in south Florida marshes is widely recognized to be highly skewed. An examination of the RPA data (**Figure 28**) will show that while most of the data seem to cluster near a central tendency, there are points that are substantially higher than most, and others, much fewer, that are substantially lower than the majority. This can be highlighted by showing the distribution of the data at G310 (**Figure 46**). The skewed distribution of the RPA data at G310 might partially explain why autosampler data are occasionally higher than grab samples from the same time period. Just five values centered around the most common value of

19 µg/L account for 58% of the data, ranging from 17 to 21 µg/L. Results higher than these central values account for 35% of the data, while values lower account for only 7% of the data. Consequently, if one were to randomly collect grab samples less than one in ten would be less than 17 µg/L. In contrast, an autosampler, which composites multiple aliquots during the course of the week has a much lower chance of having all samples lower than 17 µg/L, and therefore averaging out in the physical composite as less than 17 µg/L. For example, if the chances of getting a sample less than 17 µg/L were 10%, the chances of getting two samples in the autosampler that were both less than 17 µg/L is only 1%, and only becomes more difficult as the number of aliquots increases. Additionally, the fact that autosampler events with single aliquot samples are rarely analyzed biases the autosampler data set away from lower values.

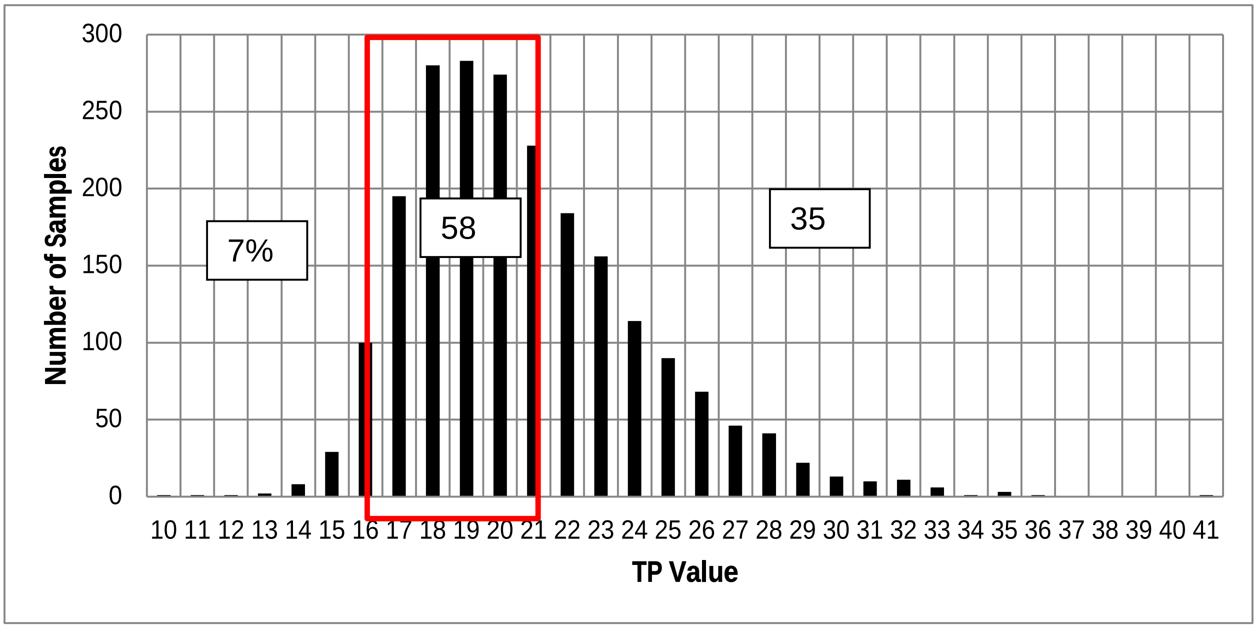

Figure 46. Frequency distribution of RPA TP values at G310 during 34 weeks of sampling every 2 hours.

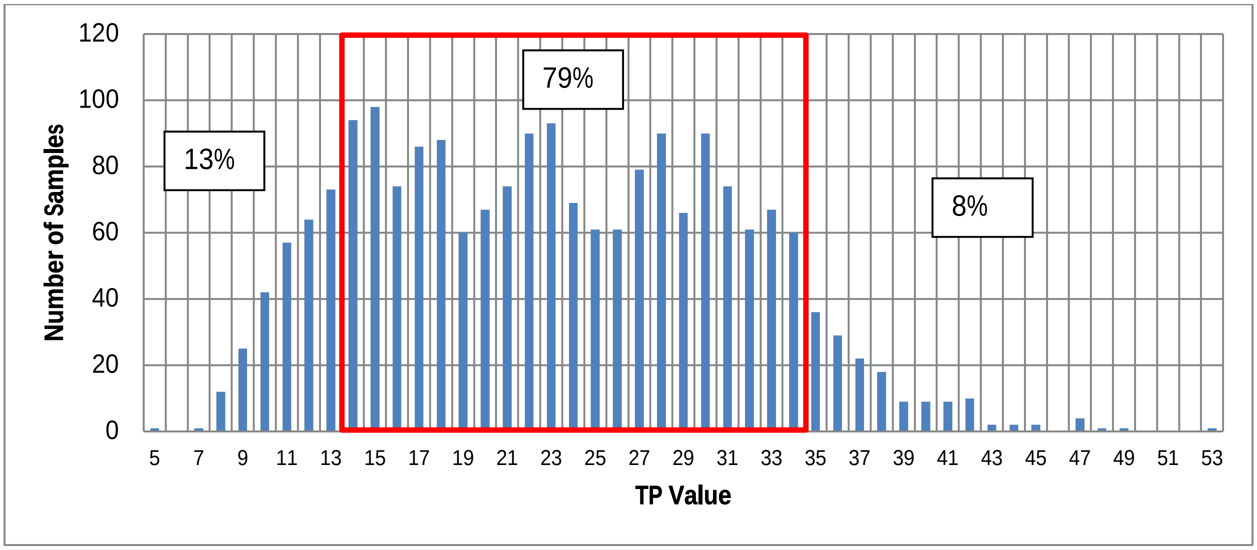

Figure 47. Frequency distribution of RPA TP values at G390B during 34 weeks of sampling every 2 hours.

A similar analysis at G390B shows a remarkable difference (**Figure 47**). The distribution of values at G390B does not have a single peak, but rather multiple peaks, or a rough plateau that ranges between 14 to 34 µg/L. This plateau accounts for 79% of the data with 8% of the data ranging above 34 µg/L, and 13% ranging below 14 µg/L. While the shape of this curve is unusual, it is not unexpected. G390B is an open culvert with a significant amount of reverse flow, which would allow for volumes of water that had passed through the structure to return and then pass through again. Such activity would obviously skew the distribution of the data, generating more events for each value than there actually should be as compared to a pump or one-way gate.

Another factor that may be involved with differences between grabs and autosamplers may be sampling time. For the 34 weeks of sampling at G310, the grab sample was collected in a three-hour window between 0700 and 1000, with an average time of 0834. In contrast, the sampling in the flow-proportional autosampler was widely scattered, but averaged a time of 1206. This difference in mean sampling time may factor into the variation between autosamplers and grabs because of the observed diel trends as shown in **Figure 44**. This note should serve as a point of discussion when analyzing water quality data.

Another issue to consider is the highly stratified, non-random manner in which samples are collected. Water quality sampling is often thought of as a somewhat random process, but this is simply not true for a number of reasons. First, because water quality sampling is a routine program, no work is done on Saturday or Sunday. Additionally, because sampling staff work 10-hour days, little work is carried out on Friday. Therefore, the vast majority of samples are collected on Monday through Thursday. Similarly, staff rarely begin traveling before 0700 and are usually back from the field by 1700. If one factors in a reasonable amount of travel time of one hour, the window for sampling most stations is between 0800 and 1600. Furthermore, most sampling trips are repeated in a similar order, the sampling window for a particular station is usually only a few hours wide. Thus, the sampling program is highly stratified, rather than random. Similar factors likely influence the operation of structures. Consequently, the stratified sampling trips considered with the diel patterns observed in the data may contribute to skewing the grab data one way or another.

COMPARISON OF DAILY TO SUB-DAILY LOAD ESTIMATES FROM RPA DATA AND IMPLICATIONS FOR ADT AND ACF COMPARISONS

One factor that may contribute to error in the calculation of loads and annual flow-weighted means is the reliance on mean daily flow instead of sub-daily flow measurements. Historically, evaluating this factor was not possible because the frequency of sampling either a weekly grab or a weekly flow composited sample did not allow for accurate sub-daily TP concentrations to be paired with the flow data. However, with the data from the RPA, that impediment no longer exists and comparisons of loads calculated from mean daily flows and mean daily TP concentrations, as well as at higher frequencies became possible.

Using G310 15-minute flow data, the flows were developed and included 24-hour, 12-hour, 8-hour, 6-hour, and 4-hour time steps. In order to calculate loads for these intervals, the RPA data set was screened and limited to days where there was flow and at least 11 data points. This strategy produced 138 dates for which loads could be calculated. **Figure 48** shows the result of this calculation and reveals that contrary to expectations, the use of sub-daily flow data paired with sub-daily TP data did not substantially impact results.

This negative result might reveal a practical application. In the process of dividing the days into smaller time steps, periods of time without flow were eliminated along with the RPA results during these periods. Thus the day divided into six 4-hour periods may have some of those periods set to zero in terms of load because the flow is zero. However, the loads calculated at the 4-hour time step do not appear to be substantially different than the loads calculated using all the data. This suggests that the 12 samples taken during days of flow still approximate the load even if the flow only occurred during a limited portion of the day. In other words, a simple time-proportional sample collected during the day of flow may adequately approximate more flow-targeted sampling. Elimination of flow-dependent monitoring might increase efficacy by reducing sampling program complexity.

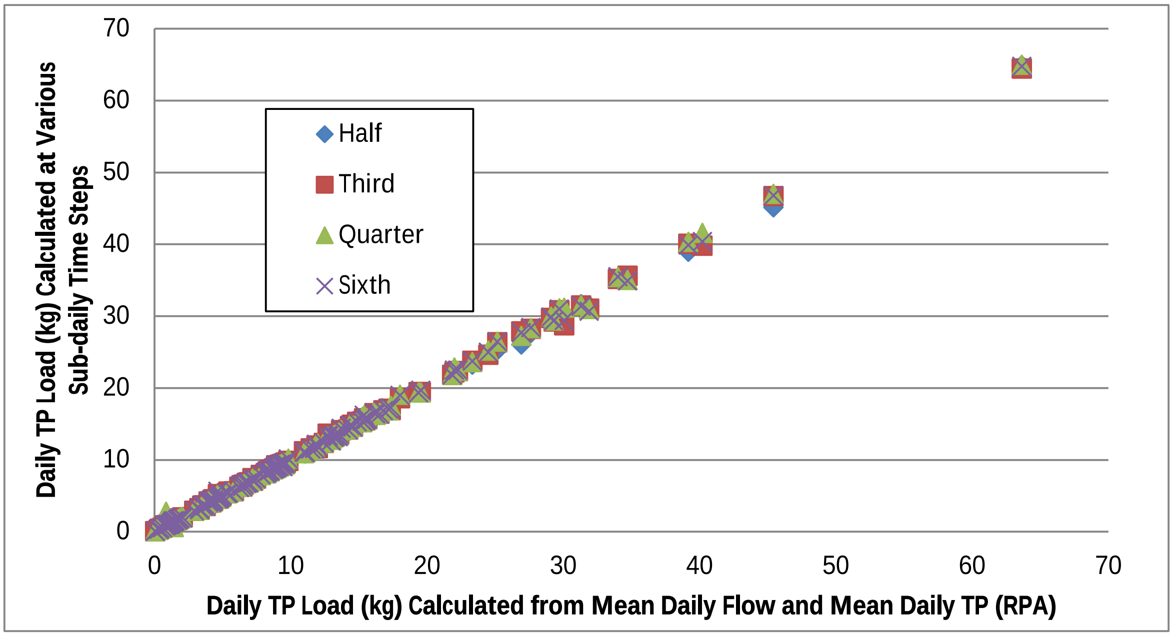

Figure 48. Comparison of mean daily flow load calculations in kilograms (kg) and sub-daily calculations at G310.

Taking a suggestion from the presentation of RPA data in **Figure 48**, the weekly ACF data was compared to the mean ADT data for the same week. **Figure 49** shows these results and seems to indicate that the mean ADT data gave a close approximation of the results from the ACF.

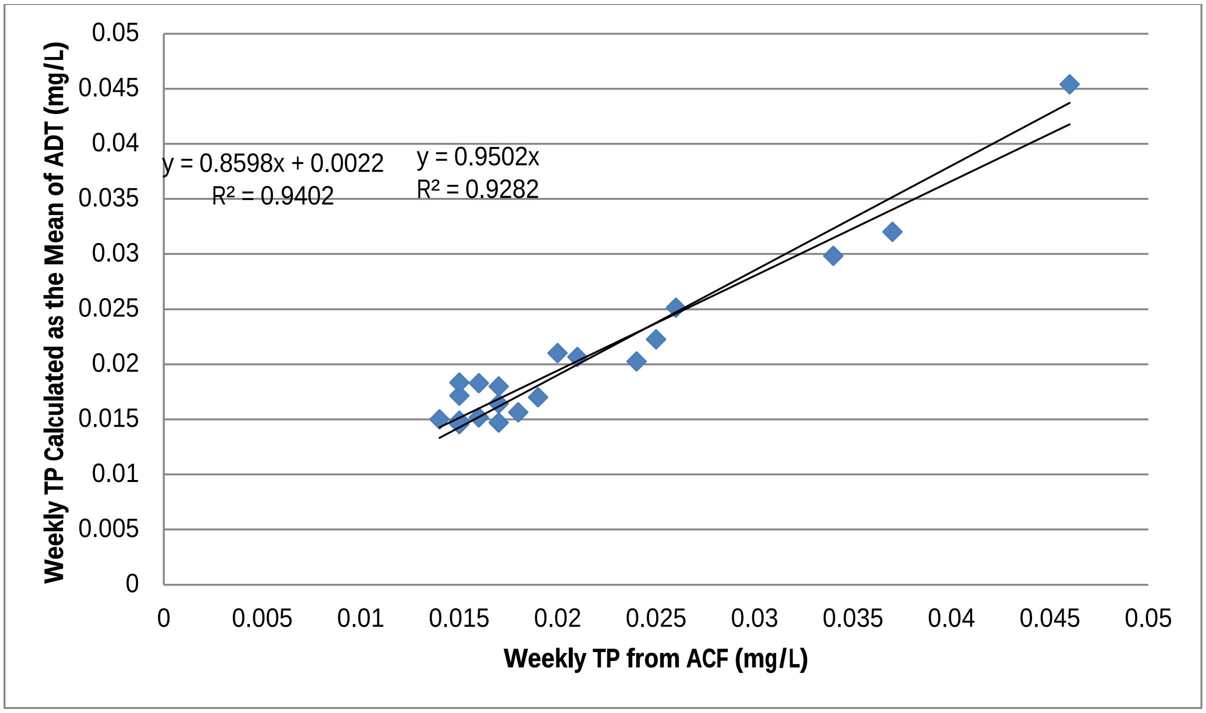

Figure 49. Weekly TP results from ACF and mean ADT results with on outlier removed.

However, it is theoretically possible to refine the ADT data by weighting the daily results with the flow of the day those samples were collected, this analysis is presented in **Figure 50**.

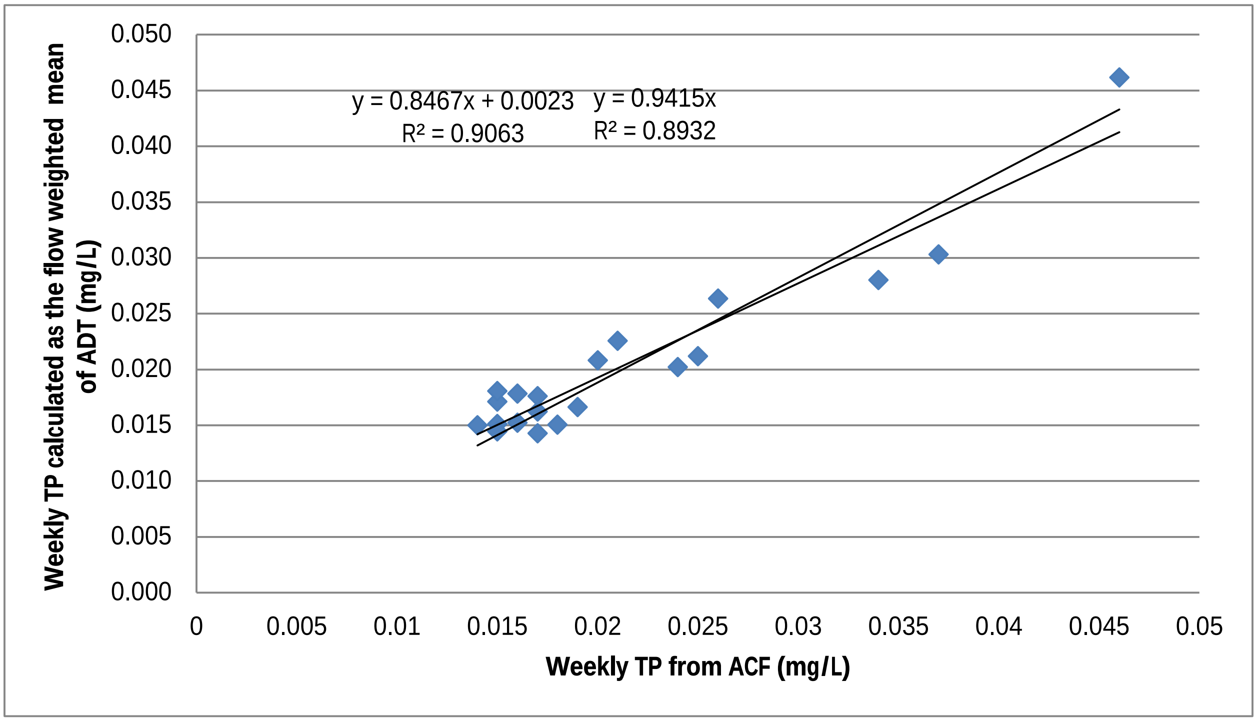

Figure 50. Weekly TP results from ACF and flow-weighted mean ADT results with one outlier removed.

As shown, flow weighting the ADT data showed no appreciable difference in the estimation of the TP concentration when using the ACF as the standard of comparison. Indeed, the argument could be made that the strength of the relationship decreased, albeit slightly. Since there was no improvement by weighting the ADT data, it suggests that daily discrete samples may not be needed to flow weight the samples, and that weekly composites (ACT) might be sufficient and comparable to ACF data.

Statistical analysis found that this data was normally distributed, but heteroscedastic ($p = 0.537$). An analysis of variance (ANOVA) on Ranks found no significant difference ($p = 0.432$) between the ACF, ADT, and flow-weighted ADT. 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.

PROJECT REST CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

The conclusions are as follows:

1. The programming supporting the flow-proportional autosampler can be flawed and should be validated.
2. Flow-proportional autosamplers at structures with low flows and the potential for reverse flow may not be producing accurate results, and may be biased by sampling the same waters repeatedly or by missing some events and oversampling others.
3. Flow-proportional autosamplers at structures with significant amounts of phantom flow may be biased by false positive sampling triggers.
4. This study showed that flow-proportional autosamplers are not meeting completeness targets and that, depending on how completeness is defined, up to 35% of the expected samples are not collected.
5. Grab and ADT methods appeared to be the most reliable in collecting expected samples.
6. Infrastructure was found to potentially function as habitat for biota and for biogeochemical processes. This includes weed barriers, platforms, sampling pipes, structures, and levees. As habitat, infrastructure likely serves as both an acute and chronic source of TP to the local water column influencing discharge concentrations in the following ways:
    a. Fauna, particularly turtles and fish, may elevate TP levels through bioturbation.
    b. Birds, particularly anhingas and to a lesser extent wading birds and vultures, may elevate TP levels through defecation and bioturbation.
    c. The structures themselves provide habitat for insects and spiders and other flora and fauna. These species may influence TP concentrations through their waste and other products.
    d. Vegetation control on levees either by herbicides or mowing may also be a factor.
    e. Weathering of the levee material should itself be considered a source of phosphorus.
7. Vegetation, particularly masses of submerged aquatic vegetation, often impacted the sampler but the influence of these events on TP results was not evident.

8. As expected, particles increased during flow events but the impact of this on turbidity and TP results was not evident.
9. The relationship between grab and autosampler results appears to be influenced by the distribution of the data, as well as by the sampling method of the autosampler; compositing samples made the odds of reporting a low concentration extremely unlikely.
10. The RPA results are comparable to results from other methods, but are likely biased high because of different analytical methods.
11. The RPA is still in the experimental stage and some functionality failures must be overcome to make it compatible with the needs of compliance monitoring.
12. TP results from daily time-proportional samples did not appear to be substantially different from flow-proportional samples.
13. Daily time-proportional results and mean weekly time-proportional results calculated from ADT data are highly comparable to ACF results. Weighting the ADT data with daily flow showed no appreciable change in this relationship.
14. While the ACF may be most representative manner in which to estimate flow-weighted concentrations, the ADT results are comparable and have a greater efficacy. This fact combined with the lower field labor and troubleshooting costs of the ADT may make time-proportional sampling more effective.
15. Data from the RPA sampling systems demonstrates a diel cycle with values peaking near midday.

RECOMMENDATIONS

Recommendations are as follows:

1. A process to document and periodically check the flow-proportionality of autosamplers should be developed.
2. The placement of ACFs at structures at which reverse, low, and phantom flows are prevalent should be discouraged.
3. A method for preventing reverse, low, and phantom flows through structures with mandatory autosamplers should be investigated.
4. Completeness calculations should be based on the amount of flow represented, rather than time.
5. Biogeochemical processes on levees and infrastructure should be considered as sources of TP to the adjacent marsh and water column.
6. Concepts for limiting the influence of levee and infrastructure processes on the water column should be developed and investigated for practicality and functionality.
7. Methods for limiting the impact of vegetation, turtles, and birds on the sampling equipment, including redesign, should be investigated.
8. Deterrents for fish and turtles, such as harvesting, should be considered.
9. A study of particle sources and dynamics in response to flow regimes should be considered.
10. Refinements to RPA analyses and maintenance should continue.
11. The use of deployed cameras both above and under the water to evaluate conditions when staff is not present should become a tool to troubleshoot problem stations.

12. A focused study comparing ACF data with ADT data should be carried out at multiple locations with the goal of validating the use of ADT and ACT methods as equivalent and cost-effective alternatives.

REFERENCES

Abtew, W., L.J. Lindstrom and T. Bechtel. 1997. *Flow-Proportional Sampling From Variable Flow Canals*. WRE Technical Paper 351, South Florida Water Management District, West Palm Beach, FL.

Ackerman, D., E.D. Stein and K.J. Ritter. 2011. Evaluating stormwater sampling approaches using a dynamic watershed model. *Environmental Monitoring and Assessment* 180(1):283-302.

Burwell Jr., R.W. 2009. *Nutrient and Sediment Losses from Surface Runoff during Bermuda Grass (*Cynodon dactylon *L.) Establishment on a Levee Embankment*. School of Plant, Environmental and Soil Sciences, Louisiana State University, Baton Rouge, LA.

Geesey, G.G., G.V. Alexander, R.N. Bray and A.C. Miller. 1984. Fish fecal pellets are a source of minerals for inshore reef communities. *Marine Ecology Progress Series* 15:19-25.

Irick, D.L., B. Gu, Y.C. Li, P.W. Inglett, P.C. Frederick, M.S. Ross, A.L. Wright and S.M.L. Ewe. 2015. Wading bird guano enrichment of soil nutrients in tree islands of the Florida Everglades. *Science of the Total Environment* 532:40-47.

Millian, Swain & Associates, Inc. 2007. *Statistical Analysis of Autosamplers vs. Grab Water Quality Data Collected at 10 Inflow Points to Everglades National Park*. Submitted under work order STO61294-WO03 to the South Florida Water Management District, West Palm Beach, FL.

National Research Council. 2009. *Urban Stormwater Management in the United States*. The National Academies Press, Washington, DC. DOI:10.17226/12465.

SFWMD. 2013a. *Science Plan for the Everglades Stormwater Treatment Areas*. South Florida Water Management District, West Palm Beach, FL.

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

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

SFWMD. 2014b. *Operational Monitoring Plan for Eastern Flowway Stormwater Treatment Area 1W*. Field-MP-047-03, South Florida Water Management District, West Palm Beach, FL.

SFWMD. 2014c. *Operational Monitoring Plan for Eastern Flowway Stormwater Treatment Area 3/4*. Field-MP-045-05, South Florida Water Management District, West Palm Beach, FL.

Struve, D., R. Walker and M. Zhou. 2012. 2*012 Update to the 2008 Remote Phosphorus Analyzer Project Report*. South Florida Water Management District, West Palm Beach, FL.

Tarver, D.P., J.A. Rodgers, M.J. Mahler and R.L. Lazor. 1978. *Aquatic and Wetland Plants of Florida. Bureau of Aquatic Plant Research and Control*. Florida Department of Natural Resources, Tallahassee, FL.