University of Maryland
CENTER FOR ENVIRONMENTAL SCIENCE

Post Office Box 38
146 Williams Street
Solomons, MD 20688-0038
(410) 326-4281
Fax: (410) 326-7302
http://www.umces.edu/cbl

A Power Analysis Web Tool to Enhance Monitoring Studies

Prepared For:
Chesapeake Bay Trust
Restoration Research Award Program

Prepared By:
Dong Liang
Lora A. Harris
Solange Filoso
Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science

Joshua Thompson
Watershed Protection & Restoration Program, Anne Arundel County Department of Public Works

September 15th 2024

APPALACHIAN LABORATORY   CHESAPEAKE BIOLOGICAL LABORATORY   HORN POINT LABORATORY

AN INSTITUTION OF THE UNIVERSITY SYSTEM OF MARYLAND

Report #8, Final Report
September 15, 2024

Contents

Report #8, Final Report
September 15, 2024

Executive Summary

The effectiveness of Best Management Practices (BMP) to achieve their intended goals varies, which makes it difficult to link the implementation of watershed restoration activities to reduced loads of nutrients and sediment to the Chesapeake Bay. The objectives of this project are to synthesize the best available monitoring data from seven urban sites with high temporal frequency of data collection in the Chesapeake Watershed and to develop an open-source tool to guide BMP monitoring studies and enhance restoration research. We also aim to adopt a co-development approach with practitioners, scientists, and modelers.

High-frequency and stream restoration monitoring data were collected from seven urban systems. Over 2 million in-situ 15-minute nutrient estimates were modeled in a Bayesian hierarchical approach to estimate the statistical powers (the probability of detecting restoration signals) of monitoring and sampling designs recommended by practitioners. Results show that the statistical powers of the monitoring design depend on spatial scale, watershed characteristics, and pollutants of interest. Despite this uncertainty, we found that (1) Pre-restoration monitoring at a nearby similar site (control) is efficient (power > 0.8) at detecting restoration improvement signals from noisy flow and concentration data. (2) Automated storm sampling should be used, especially for pollutants more susceptible to hydrological variations (e.g. TSS). (3) Flow normalization is a powerful way to detect certain BMP effects, especially those that affect concentration. (4) End-user engagement and co-development of monitoring tools (e.g. in choices of sampling methods, delivery of shiny-app) could enhance monitoring efforts by learning from the stakeholder and delivering R-Shiny tools.

I. Introduction

Sound decision-making for the management and rehabilitation of degraded environments such as urban watersheds requires an understanding of the cause-effect relationships between environmental management interventions and ecological responses. However, demonstrating causal relationships between interventions and responses in environmental systems is challenging not only because of the natural variability in environmental responses but also because of the difficulties associated with implementing rigorous experiments and monitoring programs at the watershed scale (Nichols et al. 2017). Consequently, our ability to robustly estimate the probability that an ecological recovery resulted from the management intervention designed to mitigate the impairment becomes limited.

The precautionary principle implies that environmental impact assessments favor the null hypothesis of no impact over the alternative of positive impacts when the monitoring data are lacking (Underwood and Chapman 2003). This principle has resulted in a body of literature that has evaluated the power of monitoring designs for quantifying environmental impacts (Osenberg et al. 2006, Osenberg et al. 1994, Underwood 1994), with studies mostly focused on the before-after (BA) Control Impact (CI) design. Early examples include power analysis based on BACI design using paired t-tests (Osenberg et al. 1994) or parametric tests for non-normal data (Fryer and Nicholson 1993). Power analysis requires prior estimates of the impacts and the natural variability of the system under management actions (Johnson et al. 2015). These estimates are difficult to obtain in general, and in particular for water quality monitoring in urban watersheds in Maryland.

The increasing availability of in situ water quality sensors provides insights into the challenges in urban water quality monitoring (Liu et al. 2021). The relation between stream discharge and constituent concentration (c-Q) can reveal the complex processes generating the hydrograph and chemograph (Duncan et al. 2017) during storms. For large storm events, nitrate load decreased initially due to dilution in response to the discharge. The extent of dilution was more evident in growing seasons due to high base flow concentrations. Over the longer temporal scale, the slope of weekly c-Q trended toward zero, indicating chemostasis (Duncan et al. 2017). Various hydrological, biogeochemical, and antecedent environmental conditions have been used to predict the event scale hysteresis patterns (Liu et al. 2021). Watershed characteristics include land use, slope, and soil type (Aguilera and Melack 2018). Event scale biogeochemical properties such as precipitation can also affect the c-Q relationship (Miller et al. 2021). These large data sets can be synthesized in various machine learning and statistical tools to classify the hysteresis effects (Ayers et al. 2021, Hamshaw et al. 2018). These models can predict the errors in estimating load and BMP estimation effectiveness based on hysteresis patterns at urban watershed scales.

The Bayesian hierarchical approach offers a flexible framework for optimizing the statistical powers of monitoring studies, although they have been less accessible and/or applied in stream monitoring contexts. The Bayesian framework accounts for spatiotemporal masking of restoration effects while avoiding erroneous impact assessment based on correlated data (Benedetti-Cecchi 2001). One recent example was proposed to account for spatial variation and temporal autocorrelation within a BACI design (Liang et al. 2019). Power analysis in the Bayesian framework is flexible in considering realistic sampling designs and monitoring methods, but it must rely on simulation and is computationally slow (Underwood and Chapman 2003). Moreover, the amount of effort to set up the simulation and interpretations of results remains a burden for the general application of power analysis (Green and MacLeod 2016, Johnson et al. 2015). To make the Bayesian power analysis useful broadly, practitioners need user-friendly software and clear roadmaps for implementation (Fisher et al. 2019).

We applied a Bayesian framework to study the power of a variety of sampling designs commonly used to monitor urban watersheds. The model synthesizes big data collected by in-situ sensors to quantify BMP effects and adjust for the natural variability of pollutant loads in urban watersheds. Uncertainty in the monitoring studies was quantified through a Monte Carlo approach. This power analysis tool focuses on chemical monitoring in the urban context and considers the effects of storm event characteristics based on in-situ high-frequency data. We created a rigorous, and accessible, statistical method and web platform for evaluating monitoring data in service of restoration outcomes. Combining the Bayesian framework with the monitoring designs and packaging these in a software tool creates, first, confidence that the approach is useful and, second, a way for practitioners to apply the method to their data. The tool will enhance restoration research science by maximizing the value of future water quality monitoring studies and enable them to answer critical questions about how to improve stream health.

II. Methods – Data Collection

1. USGS Data Collection

USGS discharge and water chemistry data were retrieved using the R package dataRetrieval (Hirsch and De Cicco 2015, Hirsch et al. 2010). We collected unit-level discharge (Q, NWIS[1]#00060). We also collected grab sampling of water chemistry data: total nitrogen (TN, NWIS# 00600), total phosphorus (TP, NWIS# 00665), and total suspended sediment (TSS, NWSI# 80154). Lastly, we collected in-situ high-frequency data of nitrate (NO3, NWIS# 99133) and turbidity (NWIS# 63680) from a subset of sites. We collected unit level and water quality data between 2000 and 2022. Data collection was conducted in parallel using a high performance computer. At USGS-gauged sites (Table 1), discharge records from USGSs were collected using standard technology and methods (Groffman et al. 2004). Grab samples were collected weekly using a standard protocol. No attempt was made to account for storm sampling.

2. Baltimore Ecosystem Studies Data Collection

We expanded the previously funded project and completed a collection of six urban stream sites from the Baltimore Ecosystem Studies. The data set spans 1998 and 2019. Water chemistry data consists of weekly grab samples of total nitrogen and total phosphorus. In-situ high-frequency nitrate data were also available from the Dead Run BES sites (NWIS# 01589330). We extracted a subset of the high-frequency nitrate data between 2012 and 2015 from Duncan et al. (2017). SUNA sensors were used to conduct continuous measurements of nitrate. An EXO2 (YSI/Xylem) Sonde was outfitted with temperature, specific conductance, dissolved oxygen, and turbidity sensors. Data collection frequency was at least 15 minutes for discharge and turbidity and 30 minutes for nitrate.

3. Restoration research monitoring data

We collected additional data sources available from one restoration site in Linnean Drive, District of Columbia, USA (Linnean), and a nearby control site (Spring Valley) with existing pre and post-construction monitoring data (Williams and Filoso 2023). The monitoring design included both baseflow and storm event-based sampling. At the Linnean restoration site, baseflow samples were collected monthly with 12-hour antecedent dry periods. Stormflow was sampled twice quarterly. Stormflow samples were collected at 15-minute intervals (i.e. 4 samples per 1-L bottle hourly) using an automated sampler. A subset of the 1-L bottles was volume-integrated for chemical analysis. The number of bottles ranged from 5 to 10 depending on the duration of the event. These bottles were evenly distributed across the duration of the event to ensure coverage of the entire rising and falling limbs of hydrographs. Chemistry data associated with each bottle were applied to the discharge after the initial automated sampling.

---

1. National Water Information System (NWIS) uses codes to describe sites and water quality parameters to aid its interpretation. Please refer to the NWIS help system for details.

4. Catchment characteristics

Catchment characteristics such as factors of size, impervious cover treated, and land use were collected for selected basins collected from a database previously developed by co-PI Thompson with the Exponent team. Additional sites not included in the Exponent project (#16925) were characterized using the StreamStats website (USGS 2016). Additional catchment characteristics such as stream order, slope, and position in the watershed were extracted from the stream restoration monitoring database developed by co-PI Filoso.

III. Methods – Data Processing

1. Hydrological data

In-situ collection frequency of hydrological data varied slightly between sites and across the years from every 5 minutes to every 15 minutes. Data were tabulated to identify missing values (i.e. sensor maintenance). To facilitate analyses, all high-frequency data were aligned to the nearest 15-minute intervals. The sample size of each site was tabulated based on the number of 15-minute intervals out of the possible total number of intervals.

2. Chemistry data

No quality assurance was applied to the USGS data due to their existing quality assurance process. The downloaded chemistry data were assumed as grab samples occurring at a specific time point. Each sample was linked with the closest 15-minute interval of the discharge records. The in-situ nutrient and sediment measurements from USGS also varied in their collection frequency. Thus, in-situ nutrient and sediment measurements from USGS were aligned to the nearest 15-minute interval and matched with the available in-situ discharge record. The in-situ nitrate data from Dead Run (NWIS# 01589330) were available in a 30-minute interval. To facilitate data analyses, all 15-minute intervals (centroid) contained with each 30-minute interval

Table 1: Total number of hydrological and chemistry measurements from the fifteen sites monitored by the United States Geological Survey. NWIS= National Water Information System. The numbers are reported in 15-minute intervals between 1998 and 2022. Q (NWIS#00060): discharge, TN (NWIS#00600): total nitrogen, NO3 (NWIS#99133):in-situ nitrate, TP (NWIS#00665):total phosphorus. TSS (NWIS#80154): total suspended sediment, turbidity (NWIS#63680).

NWIS#	Station	Q	TN	NO3	TP	TSS	Turbidity
01589330	Dead Run, Franklintown, MD	796,503	696	96,745	693	0	0
01589290	Scotts Level Branch, Rockdale, MD	567,151	0	0	0	514	234,151
01648010	Rock Creek, Joyce Rd, DC	344,670	473	81,522	476	479	321,732
01649500	Northeast Branch, Riverdale, MD	805,460	460	0	462	585	447,925
01650800	Sligo Creek, Takoma Park, MD	471,018	320	0	321	316	288,123
0167889257	Storm Drain, Garrett Drive, Hampton, VA	244,113	153	0	153	0	215,525
0204279245	Storm Drain, Newport News, VA	253,919	184	0	184	0	230,188

were assigned the corresponding in-situ nitrate value.

We analyzed sites previously considered by the Exponent Project and the site Difficult Run near Great Falls, VA (NWIS# 01646000). In addition, we considered the Baltimore Ecosystem Study sites previously considered in the CBT Project (Project# 13973). Thus, hydrological and chemistry (both grab and in-situ) were analyzed for 55 sites. Sites with less than 20% impervious cover were identified from the geo-database. The site at Difficult Run (NWIS# 01646000) has zero percent impervious surface cover in the watershed. Given the urban focus of the project, these sites were removed.

Table 1 lists the sample sizes for all urban sites collected from USGS NWIS. No long-term water quality information was found for the following three sites: Long Quarter Branch at Lutherville, MD (NWIS #01583800), Moore's Run Trib. Near Todd Ave (NWIS #01585225), Moore's Run at Radecke Ave (NWIS# 0158230). Load calculation could not be performed and hence these three sites were removed from the study. The water quality samples were limited (n<100) from NWIS for the following four sites: Herring Run (NWIS# 01585219), Jones Falls (NWIS# 01589485), Little Falls Pump Station (NWIS# 01646500), Beaverdam Creek (NWIS# 01651730). The small sample size will limit the accuracy of load calculation and these four sites were removed from the study.

Table 1 also lists the sample sizes for sites collected from BES. All BES sites contain long-term records of water quality samples and were included in the study. Thus, we will analyze previously-collected monitoring data for constituent concentration and discharge from fifteen watersheds in the Chesapeake Bay region (Table 1). All sites contain long-term records of chemistry data and hydrological samples. Most of the sites also contain a high-frequency sampling of nutrients or sediment. This includes six sites from the Baltimore Ecosystem Studies and nine sites from USGS NWIS.

3. Event delineation

Event delineation refers to the identification of the start and the stop of storm events. Event delineation could involve the separation of base flow (Jeong et al. 2012). For example, Thompson et al. (2018) applied the method of Lyne and Hollick (Ladson et al. 2013) to separate the base flow and quick flow in streamflow. A percent threshold of quick flow can be used to delineate the stormflow events. However, the performance of the base flow separation method, when compared with the labor-intensive chemical tracer-based method, can generate uncertainty, which could impact the quality of the downstream data analyses (Liu et al. 2021).

Many studies instead have used abrupt changes in the discharge as an indicator of the boundaries of storm events. Vaughan et al. (2017) defined a storm event using base flow separation, and a minimal rise criterion determined the start of each storm, while the end of the storm was chosen manually as a point on the falling limb when discharge approached the antecedent or interevent level. Lloyd et al. (2016) defined a storm event as a response to rainfall where discharge increased by at least 20% of the baseflow. But the quality of the baseflow estimate remains undefined. Bowes et al. (2015) defined a storm as one that resulted in >60% increase in flow during high flow periods (December to April) and >40% increase in flow during low flow periods (May to November). Site-specific rainfall data could also be utilized to delineate the events (Tunaley et al. 2017). The performance of these methods varies by season and site conditions, as well as catchment size, and to the best of our knowledge, there is not an existing R package that implements these delineation methods.

Building upon these existing approaches, we implemented a semi-automated storm delineation method in R in the appendix. Storms were delineated by tracking sequences of positive and negative changes in the flow. Several minimum change criteria were applied to determine the start of a storm event. The criteria were based on both the magnitude of the increase as well as the relative increase of discharge. To facilitate parameterization of the magnitude threshold across watersheds of different sizes, flow data were normalized according to the watershed area into water yields.

To identify the start of the rising limb, discharge increases between 0.02 and 0.10 $m^3s^{-1}km^{-2}$ coupled percent discharge increases between 1% and 5% were considered a rise (Miller et al. 2021, Smith and Smith 2015). The small threshold was designed for identifying rising limbs during base-flow periods, while the larger threshold was designed for those during storm-flow periods. Discharge sequences with either small changes in magnitude or relative changes were deemed null or uncertain. A positive change below these thresholds was also deemed an uncertain or null period which represented minor fluctuations in the discharge time series. We tracked runs of either rising, null or uncertain changes until a definite falling limb was reached.

The falling limb was identified with similar methods. Thresholds for an apparent decline to be considered part of a falling limb were less stringent, because the slopes of the falling limb tended to be smaller in magnitude than those in the rising limb. We tracked discharge sequences of either falling or uncertain changes until the next definite rising limb was reached. The end of the falling limb was chosen automatically based on the duration of the falling limb, and when the falling limb approached the antecedent or interevent levels. The falling limb was cut to at most four hours after the peak, following a similar study of the hydrological responses of urban streams (Miller et al. 2021). A falling limb with a duration of less than one hour was considered an uncertain or null discharge sequence, and not considered the end of the event. The intermediate falling limbs between one and four hours in duration were classified as an end of the event or not based on the relative magnitude of yield level with respect to the adjacent peaks.

In longer events with multiple peaks and more complex variations of the hydrograph, the algorithm results were sensitive to the parameters and could involve uncertainty. Thus, these automatic delineation results were manually reviewed to correct the automated decisions of the falling limbs. Small segments of missing values in the discharge time series were linearly interpolated before event delineation. The algorithm ignored segments of non-missing discharge records shorter than 3 hours. When missing values occurred in the falling limb, the corresponding rising limbs were ignored unless the majority of the falling limb was observed. A vignette is attached that demonstrates the semi-automated workflow of delineating storm events in the appendix.

Table 2 lists the storm delineation results from fifteen urban watersheds in the Chesapeake Bay region. Eight sites contained over 20 years of discharge records, four sites contained around 7 years of discharge records, and three sites contained between 10 and 15 years of discharge records. The hydrographs at the three storm drain locations were less flashy than the other stream monitoring stations. The majority of the hydrograph at the three storm drain locations was less than 0.25 $m^3s^{-1}km^{-2}$ in yield. The distribution of storm sizes differed across the other 12 stations that are not storm drains. Large storms (> 0.50 $m^3s^{-1}km^{-2}$ yield) represented between 1% and 15% of the delineated storms. The site NWIS #01589330 at Dead Run at Franklin Town, MD represented the largest proportion and number of large storms (n=274, p=16%), followed by the site NWIS #01589197 at Gwynn’s Falls near Delight, MD (n=98, p=11%). The site NWIS #01583580 recorded the least number of large storms (n=4, p=1%).

Table 2: Number of delineated quick flow events according to the average water yield from the included study sites.

			quick flow water yield ($m^3s^{-1}km^{-2}$)					
site_no	From	to	0.00 to 0.10	0.10 to 0.25	0.25 to 0.50	0.50 to 1.00	1.00 to 3.51	Total
01583580	1999-11	2022-09	189 (53%)	150 (42%)	14 (4%)	4 (1%)	0 (0%)	357
01589180	1998-10	2019-06	466 (45%)	348 (33%)	143 (14%)	62 (6%)	25 (2%)	1044
01589197	1998-10	2022-09	256 (28%)	398 (44%)	151 (17%)	79 (9%)	19 (2%)	903
01589290	2005-10	2022-09	153 (24%)	276 (43%)	139 (22%)	50 (8%)	17 (3%)	635
01589300	1998-10	2022-09	66 (19%)	165 (48%)	83 (24%)	23 (7%)	4 (1%)	341
01589330	1998-10	2022-09	231 (14%)	752 (45%)	407 (24%)	186 (11%)	88 (5%)	1664
01589352	1998-10	2022-09	363 (43%)	316 (37%)	124 (15%)	39 (5%)	11 (1%)	853
01636845	2016-10	2020-07	11 (12%)	33 (35%)	38 (40%)	10 (11%)	3 (3%)	95
01648010	2012-10	2022-09	26 (28%)	45 (48%)	21 (22%)	2 (2%)	0 (0%)	94
01649500	1998-10	2022-09	168 (38%)	166 (37%)	87 (20%)	23 (5%)	0 (0%)	444
01650800	2008-10	2022-09	184 (27%)	316 (46%)	126 (18%)	45 (7%)	21 (3%)	692
01651000	1998-10	2022-09	151 (27%)	280 (50%)	97 (17%)	21 (4%)	9 (2%)	558
0167889257	2015-07	2022-09	183 (71%)	70 (27%)	3 (1%)	0 (0%)	0 (0%)	256
0167891721	2016-11	2022-09	171 (43%)	182 (46%)	40 (10%)	2 (1%)	0 (0%)	395
0204279245	2015-04	2022-09	346 (55%)	230 (37%)	47 (8%)	1 (0%)	0 (0%)	624

4. Load estimation

The high-frequency in-situ measurements of nitrate and turbidity were temporally aggregated into 15-minute intervals. The aggregated data were matched with the nearest discharge and associated water chemistry data from grab sampling and assessed as surrogates for the concentration of nitrate, and TP or TSS respectively (Table 1). For each watershed and the chemical variable, a robust regression model was selected to account for outliers based on the adjusted R^2 (Mächler and Ruckstuhl 2006). Model coefficients were used to predict a 15-minute time series of concentration, discharge, and loads. Data were processed by water years. To avoid extrapolation, only water years when there were at least 80% of the discharge and high-frequency concentration estimates were retained in the following analyses.

A predictive modeling approach of events loads/yields was developed based on the hydrograph characteristics. Modeling was conducted on the storm events delineated. Flow-weighted mean concentrations (FWMC) were modeled in a linear regression with selected hydrograph characteristics (Ayers et al. 2021). We considered the storm characteristics such as time of rise, initial discharge rate, peak discharge, mean discharge, cumulative discharge, relative duration of rising limb, water yield, and slope of the initial phase of the recession limb. Variance inflation factors were computed to reduce collinearity among the predictors in the regression. The model reduction was conducted based on the backward selection using the Akaike Information Criteria (AIC). An optimal predictive model in terms of AIC was used to predict the FWMC of each storm event and pollutant within the watersheds.

Grab samples collected during baseflow conditions at Linnean were analyzed in regression modeling. The pollutant concentration was modeled using potential predictors including a quadratic polynomial function of the log discharge, a quadratic trend in days, and sinusoidal functions capturing the potential seasonality of the loads. These predictors were commonly applied in the LOADEST method (Runkel et al. 2004). Backward selection was used to simplify the model. The selected model was used to predict baseflow loads for a specific pollutant and watershed.

5. Power study

Sampling frame

At each USGS site, the in-situ high-frequency data served as the sampling frame to evaluate various sampling methods. Uncertainty in the discharge records was assumed at 5% in terms of coefficient of variation and was propagated via a Monte Carlo framework. The uncertainty of the high-frequency load estimates was quantified via the predictive standard deviation from the respective robust regression model relating the grab sampled chemistry data and in-situ values.

At the Linnean site, the optimal regression models selected for storm FWMC and baseflow grab samples were implemented in a hierarchical Bayesian framework. The Bayesian model was completed by the Penalized Complexity prior onto the scale parameters of the random effects and parameterized to be non-informative (Simpson et al. 2017). The posterior inference was implemented using the Integrated Nested Laplace Approximation (INLA) using the R Package R-INLA (Rue et al. 2009). Markov chain Monte Carlo sampling was conducted to capture the model-based uncertainty of FWMC and baseflow concentrations during periods when sampling was not conducted. Within each storm event, given the predicted FWMC, high-frequency concentration was imputed based on a fixed coefficient of variation of high-frequency concentration data from a USGS site with similar catchment characteristics during the same month.

Monitoring Design and sampling methods

The monitoring schemes involved a before-after design with long-term monitoring at the convergence zone of the catchment, where part of the catchment undergoes restoration. The second scheme followed the Control Impact design, which involves monitoring multiple paired sites within the catchment. The third scheme followed the Before After Control Impact (BACI) design, which involves monitoring the paired sites both before and after the restoration.

The grab sampling methods (Thompson et al. 2021) evaluated include: (1) Monthly: sample taken every month; (2) Weekly: sample taken every week; (3) Weekly+Storm: sample taken every week with one daily sample taken from a storm. (4) MDE MS4 Permittee Requirements: Samples taken from 12 storms with monthly samples taken during episodes of extended low flow. Automated samples include: (5) Seven-Hour: Automated sample taken every 7 hours. (6) Flow-paced: Automated sample taken when the flow exceeds a cumulative threshold (threshold set to yield an average of 80 sampling pumps per week and aggregated into weekly composite); Sampling was implemented by water years. We also evaluated the hybrid sampling method which involved monthly grab sampling during baseflow and flow-paced automated storm sampling conducted at the Linnean site (Williams and Filoso 2023). The event delineation detailed in the report was used for the implementation of the hybrid storm+base sampling, and other methods that involve storm sampling.

Grab sampling simulation was restricted to regular work hours. Grab sampling times were selected using a stratified random design with strata defined by the revisit schedule (e.g. monthly). FMWC was estimated using a stratified estimator from the simulated grab sampling. Automated sampling (e.g. Seven-Hour, Flow-paced) was implemented throughout the water year regardless of the working hours. Due to the discharge variabilities, the flow-paced sampling was programmed to target approximately 80 samples per week at a weekly discharge that was around the 90th percentile of the weekly discharge distribution in the current water year. Multiple aliquots were simulated when a high flow was observed during a 15-minute interval that exceeded the required pace. The weekly average concentration was used to approximate the FMWC for the flow-paced sampling.

Simulation Framework

Monte Carlo simulation was performed to evaluate the precision of a sampling method and monitoring design. For each Monte Carlo replicate, a finite population of high-frequency loads was simulated from the fitted model for each pollutant and site. Environmental impacts of the restoration activities were simulated by altering the concentrations after the restoration. For USGS sites, a pulse percent of reduction was applied to the concentrations after the restoration. We evaluated powers for up to five years of pre-restoration monitoring and ten years of post-restoration monitoring. The simulated effect varied from 10% to 50% to capture various levels of restoration activities. For the Linnean site, the percent reduction in the concentration was applied such that the reduction at the restored site was a certain percent of the reduction observed at the control site. This reduction was simulated to between 50% and 150% of the observed restoration effects at the Linnean site for each pollutant. When the observed effects were contrary to the hypothesized restoration (i.e. when restored sites were associated with higher loads than the control site), the effect size was truncated to 1%.

The simulated data were sampled using a given sampling method, and the reduction of yields and restoration effects were then estimated according to the monitoring design. The sampling variability of each method was estimated by Monte Carlo integration of the residuals between the estimated effects and the simulated effects. Sampling variability for non-BACI design also included bias in the estimated effects due to a lack of appropriate control data. We considered the sampling variability for comparing the difference in raw loads, versus the differences in FWMC due to restoration.

IV. Work Product

1. Co-development Process

Initial contacts were made with regulators (MDE) and local jurisdictions to identify common modeling and data visualization needs. We met with the regulators (Jeff White, Bel Martinez da Matta, and Shannon McKenrick from MDE) to discuss existing and future monitoring guidance documents from MDE. The documents reviewed include the NPDES MS4 Permits Website which includes the most recently issued permits; Section G. Assessment of Controls of the permit, which includes the details on monitoring requirements; the 2021 MS4 Monitoring Guidelines, which were developed to supplement the MS4 permits with additional details on the requirements and to provide recommendations on best practices; the ICPRB MS4 data study as research deliverables, which are in the first drop down "Is watershed restoration through stormwater permits (MS4) working?", and the MD TMDL Data Center.

The regulators recommended we conduct site visits to local jurisdictions. Given the busy sampling schedules of the local jurisdictions, a centrally located workshop/listening session, as we originally proposed in our project narrative, seemed less feasible than site-visiting individual permittees based on their selected monitoring sites. Consequently, we reached out to Anne Arundel, Baltimore, and Montgomery counties.

To prepare for the visit, we developed a standardized questionnaire to query the needs of the water quality monitoring programs, the logistics of monitoring within each county, the potential design features of this application that are useful for respective counties, and some specific questions regarding statistical errors. A sample of the questionnaire is given in the appendix A in page 3. We visited the Cowhide Branch Flow-weighted monitoring station with Chris Victoria and Janis Markusic from Anne Arundel county. We reviewed their monitoring program and asked questions about Anne Arundel County’s efforts to monitor restoration sites. We also discussed the cost of the monitoring program and the kind of errors or confidence levels of acceptable monitoring programs.

We also met with Kevin D Brittingham to review the monitoring program and asked questions about Baltimore County's effort. In addition, we met with B. Madigan to review the monitoring programs at Carroll county. Alignment of completion of the analysis and delivery of a workshop for end users did not occur within the timeline of the grant. However, we are working to deliver this as a webinar this fall and following up with counties who have expressed interest in sharing the shiny app resource alongside this final report.

2. Shiny interface and results

The simulation-based power analysis was implemented as an open-source R package with a Shiny interface to facilitate broader applications. The R Shiny App and storm delineation tool can be found on GitHub (https://github.com/dliang-cbl/stormstats). A similar application was developed by the Exponent project team on detecting the impact of management actions on water quality. In this project, a prototype of the application was developed based on the existing R Shiny app developed by the Exponent team. The prototype app was developed through simulations from the Baltimore Ecosystem Studies data (Liang et al. 2019) and high-frequency in-situ data from USGS.

Ten USGS stations and one restoration site in Washington D.C. were included in the power study (Table 1), which included one station for nitrate, five stations for TP, and four stations for TSS. The duration of the monitoring ranged from 2 to 13 monitoring years. According to our inclusion criteria, at least 80% of the days were covered with high-frequency monitoring data.

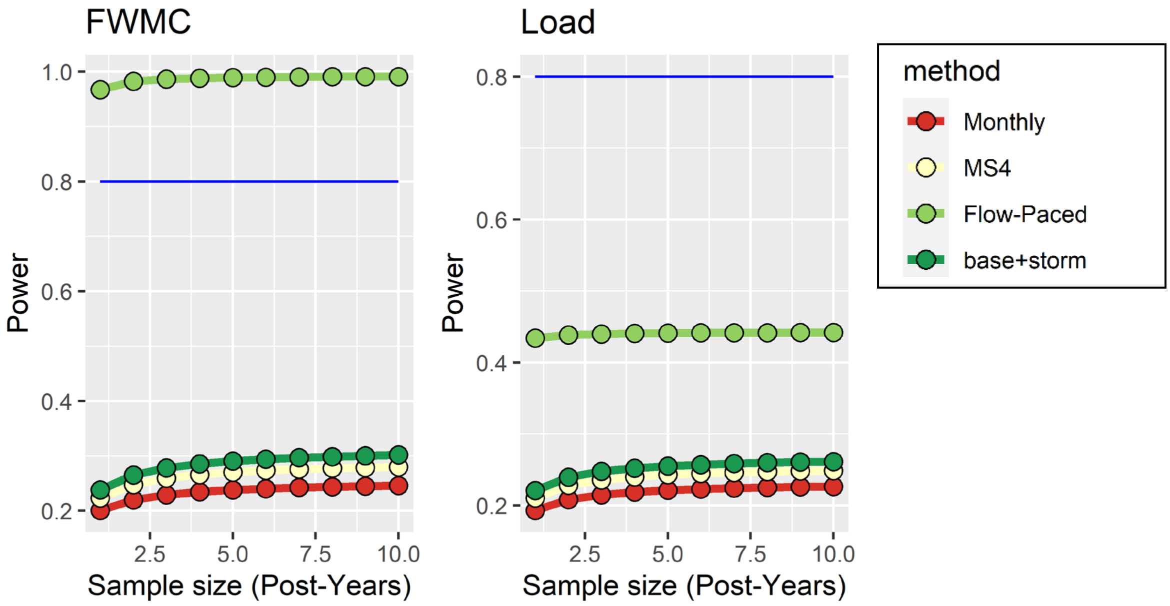

Figure 1 Statistical power in detecting a 20% reduction of total phosphorus concentration in Rock Creek, District of Columbia, by various sampling frequencies. Left panel: comparing Flow Weighted Mean Concentration, right panel: comparing loads. The results were based on Before-After monitoring design and 1-year pre-restoration monitoring, a type I error of 0.10. Blue line denotes 80% power.

Overall thousands of storms were available for the power study. The Shiny Applets are available from the following URLs (https://stormstats.shinyapps.io/Stormstats_Linnean/ and https://stormstats.shinyapps.io/Stormstats_USGS/)

From the Rock Creek, District of Columbia results, a 20% reduction of total phosphorus concentration was not detectable by many methods at 80% power (Figure 1). However, using flow paced sampling, a reduction was detectable above 80% power when assessing flow-weighted mean concentration. Power was largest for the automated sampling method, followed by the hybrid sampling of baseflow and automated storm flow, the two grab sampling methods were associated with the least power. Comparing flow-weighted mean concentration was more powerful in detecting the reduction, than comparing the raw loads.

In Dead Run, Franklin Town Maryland, the simulated 20% reduction of nitrate concentration was detectable at 80% power (Figure 2). Powers were higher than 80% when directly comparing FWMC. Powers were lower when comparing loads. None of the methods was able to detect the reduction when comparing raw loads. Automated sampling was the most powerful sampling method, followed by the MS4 sampling. Hybrid sampling was able to detect the reduction but was not as powerful as simpler MS-4 grab sampling.

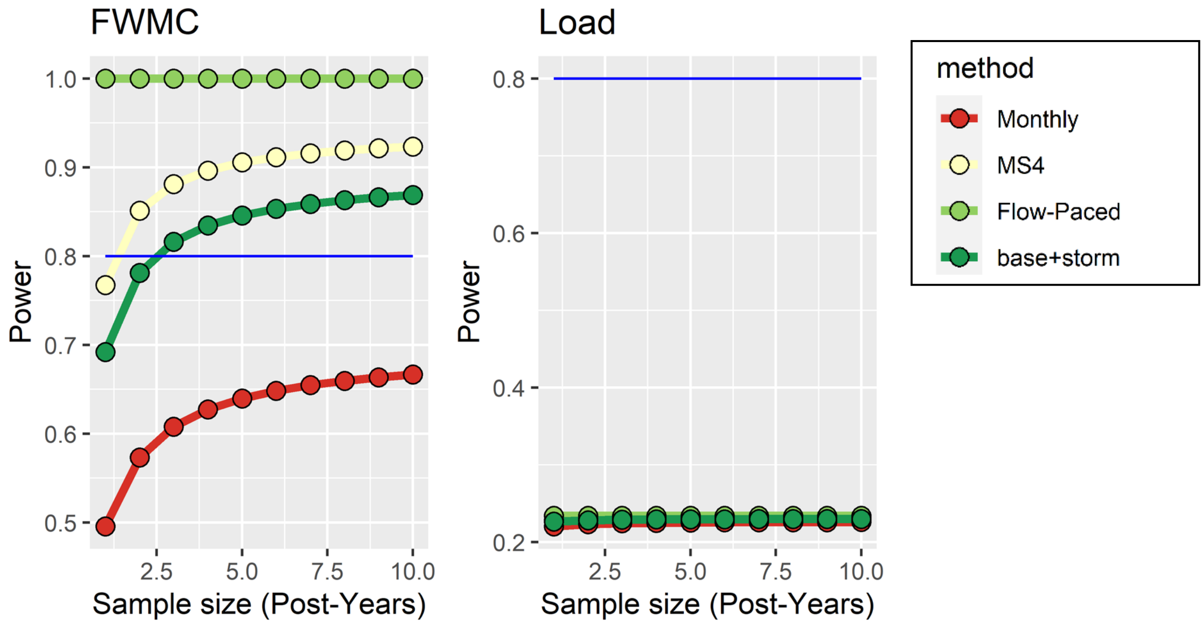

Figure 2: Statistical power in detecting a 20% reduction of nitrate concentration in Dead Run, Franklin Town Maryland, by various sampling frequencies. Left panel: comparing Flow Weighted Mean Concentration, right panel: comparing loads. The results were based on Before-After monitoring design and 1-year pre-restoration monitoring, a type I error of 0.10.

In Northeast Branch, Riverdale MD, a 20% reduction of total suspended solids concentration was not detectable at 80% power with the sampling design considered (Figure 3). In this watershed, again flow normalization through calculating the FWMC is more powerful at detecting the reduced concentration than raw pollutant loads. Automated sampling and hybrid sampling were more powerful than grab sampling.

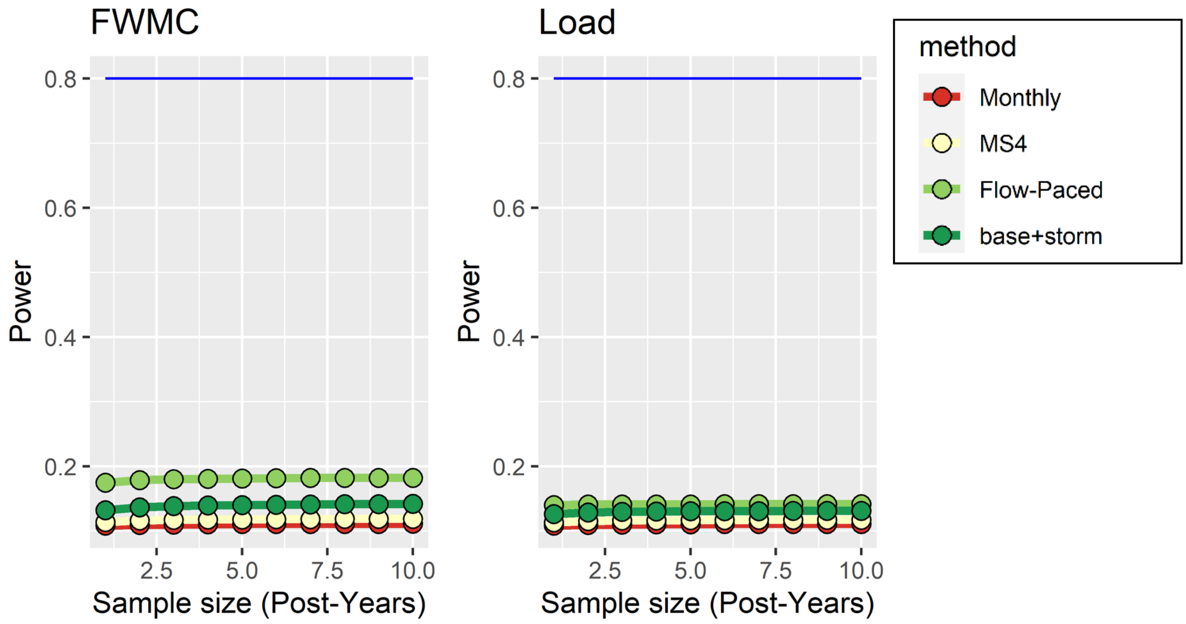

Figure 3: Statistical power in detecting a 20% reduction of total suspended solids concentration in Northeast Branch, Riverdale MD, by various sampling frequencies. Left panel: comparing Flow Weighted Mean Concentration, right panel: comparing loads. The results were based on Before-After monitoring design and 1 year pre-restoration monitoring, a type I error of 0.10.

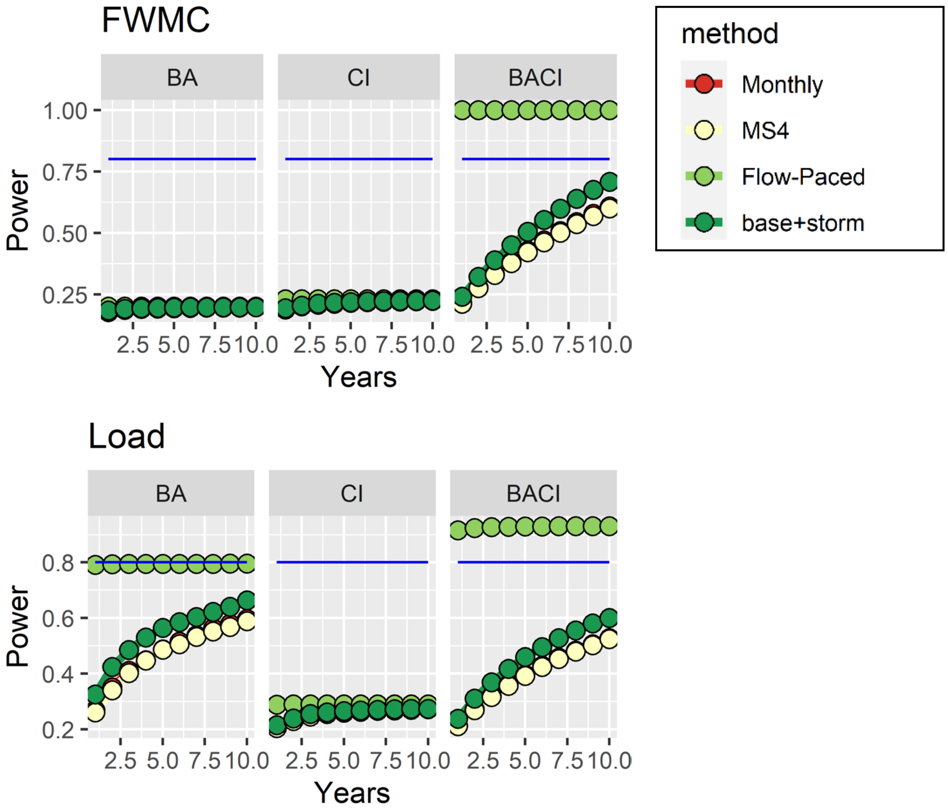

Figure 4: Statistical power in detecting a reduction of nitrate concentration in Linnean, District of Columbia, by various sampling frequencies. Top panel: comparing Flow Weighted Mean Concentration, bottom panel: comparing loads. Restoration effects were simulated at 55% of the observed reduction in flow weighted mean concentration of nitrate. BA: Before-After monitoring design, CI: Control-Impact monitoring design. Results were based on equal years of pre-restoration and post-restoration monitoring for BA and BACI design, a type I error of 0.10.

Simulated effects of nitrate concentration reduction were detectable only with the BACI design and automated flow-paced sampling method (Figure 4). The BACI design was more powerful than BA or CI designs when other factors such as monitoring duration, and sampling methods were fixed. Within the BACI design, comparing FWMC was marginally more powerful than comparing raw loads. The converse was true for BA and CI designs, when comparing raw loads appeared to be more powerful than comparing FWMC.

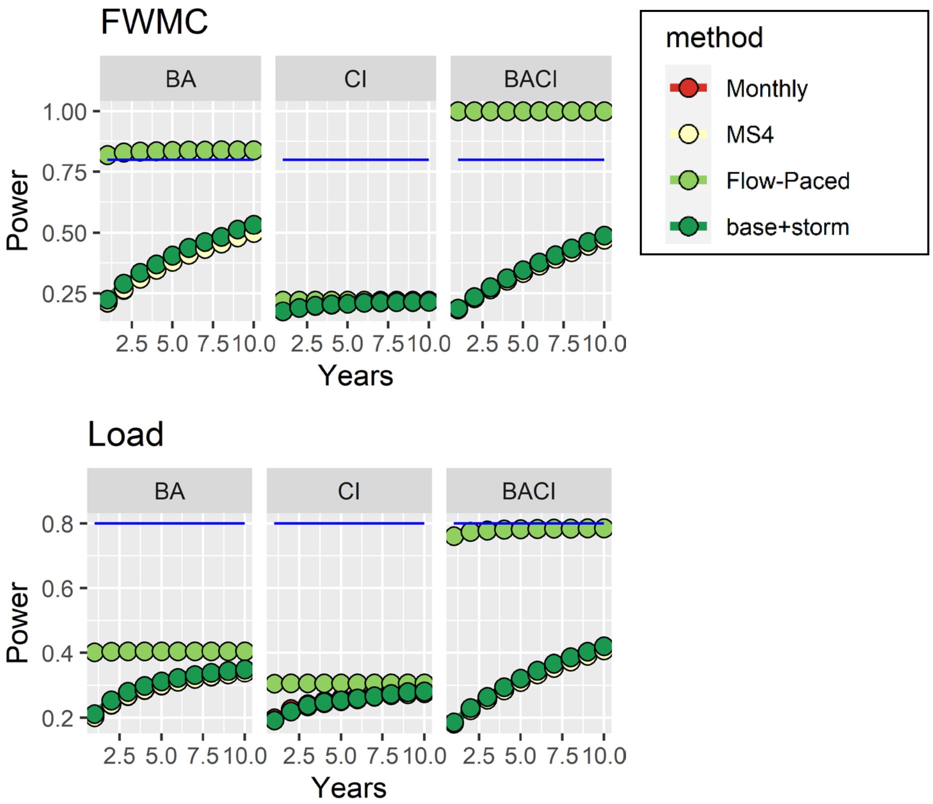

Figure 5: Statistical power in detecting a reduction of total nitrogen concentration in Linnean, District of Columbia, by various sampling frequencies. Top panel: comparing Flow Weighted Mean Concentration, bottom panel: comparing loads. Restoration effects were simulated at 55% of the observed reduction in flow weighted mean concentration of nitrate. BA: Before-After monitoring design, CI: Control-Impact monitoring design. Results were based on equal years of pre-restoration and post-restoration monitoring for BA and BACI design, a type I error of 0.10.

For total nitrogen (TN) at the same Linnean site in the District of Columbia, the BACI design was more powerful than the CI or BA design (Figure 5). For TN in this site, FWMC was a more statistically powerful metric to measure the simulated load reduction than raw loads. Specifically, reduction in FWMC was detected by BACI design when measured directly at 80% power using flow-paced sampling but was not detectable when comparing the raw loads using the same sampling method. Hybrid sampling was not as powerful as the flow-paced sampling.

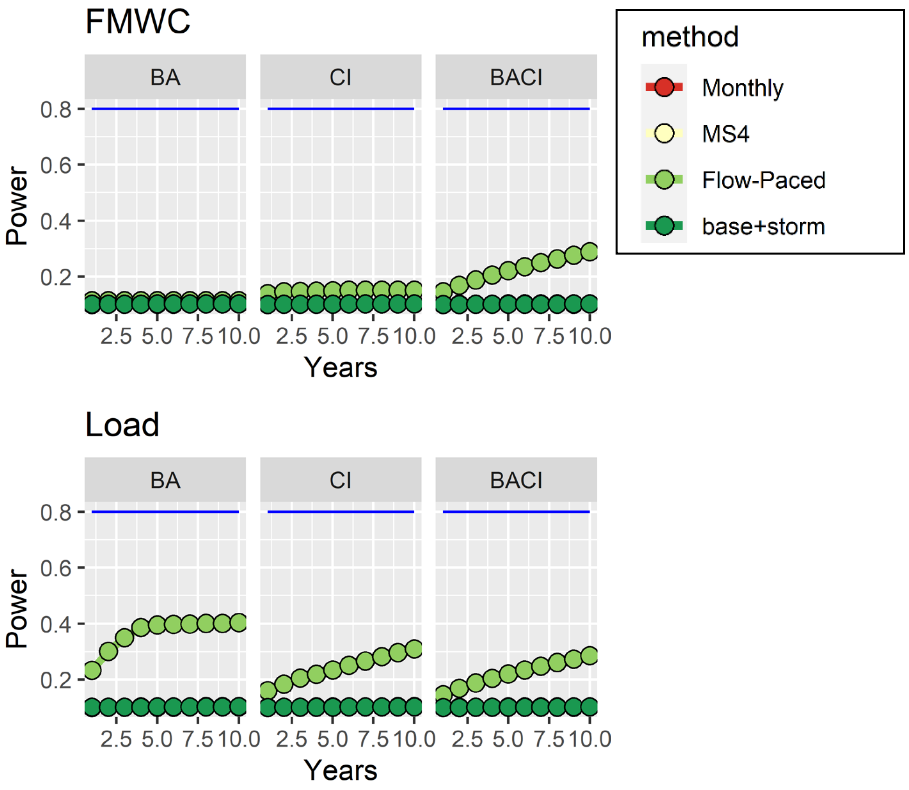

Figure 6: Statistical power in detecting a reduction of total suspended solids concentration in Linnean, District of Columbia, by various sampling frequencies. Top panel: comparing Flow Weighted Mean Concentration, bottom panel: comparing loads. Restoration effects were simulated at 55% of the observed reduction in flow weighted mean concentration of nitrate. BA: Before-After monitoring design, CI: Control-Impact monitoring design. Results were based on equal years of pre-restoration and post-restoration monitoring for BA and BACI design, a type I error of 0.10.

Total suspended solids at the same Linnean site in the District of Columbia were not detectable with any of the design criteria considered (Figure 6). Among the power calculations, BA design coupled with raw loads generated the most powerful tests, but only at around 40% power.

V. Discussion

The effectiveness of Best Management Practices (BMP) to achieve their intended outcomes varies, which makes it difficult to link watershed restoration activities to reduced loads of nutrients and sediment to the Bay. We developed an R tool to guide BMP monitoring studies and enhance restoration research. High frequency and stream restoration monitoring data collected by previous efforts were used to develop the tool using commonly used monitoring and sampling designs. The software tool was co-developed by practitioners, scientists, and modelers. Statistical modeling results showed that powerful monitoring plan can be designed to quantify load reductions from restoration in urban watersheds.

In-situ data enable reliable quantification of the sampling variability of commonly used sampling methods. Grab sampling tends to perform poorly relative to automated approaches. The relative performance of grab sampling is especially poor for total phosphorus (TP) and suspended sediments, which is more susceptible to the hydrological variation during storm events (Janke et al. 2014). Simulated restoration effects of TP were not detectable at sufficient power (Thompson et al. 2018). Automated sampling of nutrient concentration reduced the chance of missing large TP runoff and was associated with higher power for detecting the simulated restoration effects. The hybrid approach that involves grab sampling during baseflow conditions, and automated stormflow sampling also presented improved power. However, the improvements were inconsistent across watersheds and criteria pollutants. This could be the result of a large baseflow contribution to total loads. When baseflow is a major contributor to total loads, the simulation suggests that the monthly baseflow sampling of the hybrid sampling approach could generate less precise estimates than alternative sampling methods with higher sampling frequency during base-flow.

This study took advantage of monitoring studies at the Linnean stream site, Washington DC, USA before planned restoration both at the site and a nearby control. The Linnean data enabled us to evaluate the impact of various monitoring designs. Results indicate that BACI monitoring at one additional watershed as a control yielded higher power than alternative designs that do not follow the BACI principles (Fisher et al. 2019). Simpler designs were associated with biased estimates of the simulated restoration effects. The extent of bias differed across the parameters considered. In general, the bias in a before-after design resulted from the natural temporal variability of pollutant loads, which can be typically measured by comparing a nearby control watershed before and after the restoration. The bias in a control impact design increases as the natural spatial variability of the pollutant loads, which can be measured by the spatial contrast of pollutants before the restoration (Osenberg et al. 2006).

While BACI studies with flow-paced monitoring are powerful, it is associated with more sampling efforts, costs and other challenges. First, an appropriate control site must be identified. In the Linnean site, despite careful consideration of the control site selection and the spatial adjacency, the natural spatial variability of pollutants confounded the restoration signal and hence resulted in reduced power of the control impact design. This highlights the challenges in identifying an appropriate control watershed in monitoring studies. Second, the samples must be promptly filtered after collection or acid-preserved to enable analysis of both nutrients and suspended solids. Thirdly, the flow-paced sampling efforts limit the potential to evaluate the effects of discharge rates on concentration at a higher frequency (Thompson et al. 2021). Lastly, the data from flow-paced sampling would limit future meta-analyses to evaluate future generations of sampling methods. Despite these challenges, the sampling design typically requires only weekly visits to the sites. This level of personnel efforts is feasible for most local jurisdictions and research communities to implement.

Uncertainty in the pollutant loads is largely driven by the unpredictable storm events. Results from most sites show that comparing flow-weighted mean concentration was more powerful at detecting the same restoration signal than those comparisons based on raw loads (Zhang and Hirsch 2019). The results demonstrate that flow normalization is effective in reducing uncertainty and improving the power of detecting a signal. Approaches such as WRTDS are useful in this regard to enhance the monitoring studies (Hirsch et al. 2010). While flow-paced automated sampling provides an effective way to estimate loads (Thompson et al. 2021), sampling methods for flow-weighted mean concentration also require an annual estimate of discharge. Like the flow-paced sampling, annual estimation of discharge can be implemented using automated sampling devices. Both automated methods would probably require the same time resources as a weekly manual sampling but provide much higher power. If the local jurisdiction can invest in the initial setup of these automated sampling approaches and receive sufficient training, sampling for flow-weighted mean concentration is a potential solution for regulatory monitoring.

The studies are subject to a few limitations. Our study just considered pulse percent reduction of restoration effects. For relatively short studies, the pulse effects could serve as an approximation to the different trajectories of water quality after restoration such as gradual improvement, exponential improvement, or seasonal declines and improvement. For longer studies, however, this pulse approximation would be less accurate, and more complex and non-linear trends of water quality should be considered in the simulation.

The best management practices and effects considered are specific to the watershed and restoration techniques considered. Thus, results obtained cannot be generalized to other BMPs. At the Linnean site, the TSS effects were barely detectable because the observed effects of restoration were negligible. Simulated TSS in other sites, however, were detectable (Liang et al. 2019). The included urban watersheds varied in size and other watershed characteristics, but the sample sizes were still limited. Thus, the spatial variability of power cannot be generalizable to other sites. Users of the tool must manually identify the most similar watersheds to estimate the powers of their potential monitoring streams. The tool is not able to generalize power results to select stream sections for restoration. We could generalize the tool such that existing in-situ data can be incorporated from a specific site to recommend monitoring schemes. In addition, the tool could be generalized to spatially predict the restoration effects on a particular stream segment based on the restoration design, watershed characteristics, and other explanatory factors. Predictions can be used to derive location-specific recommendations based on spatial power analyses. With technological advancement and the increasing availability of continuous nutrient concentration, we expect more sites and better estimates of the spatial variability of powers (Liu et al. 2021).

The monitoring studies can also be enhanced by engaging end users during the development process. Early engagement with the regulators would provide insights into common modeling and data visualization needs. These needs are usually driven by the regulatory monitoring requirements and recommendations for best practices. We conducted site visits to several restoration research locations. These visits provide further insights into the financial constraints of monitoring and incentives for optimizing monitoring design from a user's perspective. Understanding the practical and logistical constraints of modeling would enhance the design and usability of empirical modeling studies and associated power analysis tools.

Monitoring efforts varied across the MS-4 jurisdictions we interviewed. Permit is still a major incentive for monitoring. Since pollutant loads are written in the permit language, most jurisdictions conduct load monitoring. Loads however are subject to the high natural variability of discharge. Thus, the local jurisdictions need tools for better quantification of the uncertainty in the loads. Some jurisdictions went beyond the permit requirement to monitor for BMP effectiveness. However, due to limited staff time and resources, the corresponding data analysis was mostly done in-house. We are only aware of one CBT-funded effort to enhance the sustainability of these data to address the TMDL concerns at a broader spatial scale. Consequently, there is still a need for a unified cyberinfrastructure that can easily synthesize data across the jurisdictions generated by disparate analytical software, meta-data formats, and presentation results.

In conclusion, technical advancement in continuous nutrient measurements coupled with discharge records offers great potential to understand the monitoring effectiveness of restoration studies. Appropriate choice of the control watershed, coupled with an automated sampling of pollutants before and after the restoration provides the highest probability of detecting reasonable load reduction. This result highlights the importance of considering monitoring designs and sampling methods in monitoring studies, as well as the values of co-development of cyberinfrastructures with multiple stakeholders involved in regulatory and restoration monitoring. Our study indicates that sampling for loads is not optimal for addressing the BMP effectiveness at the sites considered. Flow normalization might be a better solution for BMP monitoring. Power analysis tools developed in this project are a first step to applying open science and open data approach to modernize the data analyses conducted at the local jurisdictions and make the existing data more useful to address TMDL and the uncertainty in load reduction.

- Sampling for flow-weighted mean concentration has greater statistical power than sampling for loads.
- Before-After Control-Impact studies with flow-paced sampling are powerful and feasible for local jurisdictions
- A similar open-data approach to co-develop tools will enhance local jurisdiction monitoring.

VI References

Aguilera, R. and Melack, J.M. (2018) Relationships among nutrient and sediment fluxes, hydrological variability, fire, and land cover in coastal California catchments. Journal of Geophysical Research: Biogeosciences 123(8), 2568-2589.

Ayers, J.R., Villarini, G., Schilling, K. and Jones, C. (2021) Development of statistical models for estimating daily nitrate load in Iowa. Science of the Total Environment 782, 146643.

Benedetti-Cecchi, L. (2001) Beyond BACI: optimization of environmental sampling designs through monitoring and simulation. Ecological applications 11(3), 783-799.

Bowes, M., Jarvie, H., Halliday, S.J., Skeffington, R., Wade, A., Loewenthal, M., Gozzard, E., Newman, J. and Palmer-Felgate, E. (2015) Characterising phosphorus and nitrate inputs to a rural river using high-frequency concentration–flow relationships. Science of the Total Environment 511, 608-620.

Duncan, J.M., Welty, C., Kemper, J.T., Groffman, P.M. and Band, L.E. (2017) Dynamics of nitrate concentration-discharge patterns in an urban watershed. Water Resources Research 53(8), 7349-7365.

Fisher, R., Shiell, G.R., Sadler, R.J., Inostroza, K., Shedrawi, G., Holmes, T.H. and McGree, J.M. (2019) epower: An r package for power analysis of Before-After-Control-Impact (BACI) designs. Methods in Ecology and Evolution 10(11), 1843-1853.

Fryer, R. and Nicholson, M. (1993) The power of a contaminant monitoring programme to detect linear trends and incidents. ICES Journal of Marine Science 50(2), 161-168.

Green, P. and MacLeod, C.J. (2016) SIMR: an R package for power analysis of generalized linear mixed models by simulation. Methods in Ecology and Evolution 7(4), 493-498.

Groffman, P.M., Law, N.L., Belt, K.T., Band, L.E. and Fisher, G.T. (2004) Nitrogen fluxes and retention in urban watershed ecosystems. Ecosystems 7(4), 393-403.

Hamshaw, S.D., Dewoolkar, M.M., Schroth, A.W., Wemple, B.C. and Rizzo, D.M. (2018) A new machine-learning approach for classifying hysteresis in suspended-sediment discharge relationships using high-frequency monitoring data. Water Resources Research 54(6), 4040-4058.

Hirsch, R.M. and De Cicco, L.A. (2015) User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data, US Geological Survey.

Hirsch, R.M., Moyer, D.L. and Archfield, S.A. (2010) Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs 1. JAWRA Journal of the American Water Resources Association 46(5), 857-880.

Janke, B.D., Finlay, J.C., Hobbie, S.E., Baker, L.A., Sterner, R.W., Nidzgorski, D. and Wilson, B.N. (2014) Contrasting influences of stormflow and baseflow pathways on nitrogen and phosphorus export from an urban watershed. Biogeochemistry 121(1), 209-228.

Jeong, J.J., Bartsch, S., Fleckenstein, J.H., Matzner, E., Tenhunen, J.D., Lee, S.D., Park, S.K. and Park, J.H. (2012) Differential storm responses of dissolved and particulate organic carbon in a mountainous headwater stream, investigated by high-frequency, in situ optical measurements. Journal of Geophysical Research: Biogeosciences 117(G3).

Johnson, P.C., Barry, S.J., Ferguson, H.M. and Müller, P. (2015) Power analysis for generalized linear mixed models in ecology and evolution. Methods in Ecology and Evolution 6(2), 133-142.

Ladson, A.R., Brown, R., Neal, B. and Nathan, R. (2013) A standard approach to baseflow separation using the Lyne and Hollick filter. Australasian Journal of Water Resources

17(1), 25-34.

Liang, D., Harris, L.A., Testa, J.M., Lyubchich, V. and Filoso, S. (2019) Detection of the effects of stormwater control measure in streams using a Bayesian BACI power analysis. Science of the Total Environment 661, 386-392.

Liu, W., Birgand, F., Tian, S. and Chen, C. (2021) Event-scale hysteresis metrics to reveal processes and mechanisms controlling constituent export from watersheds: a review. Water research, 117254.

Lloyd, C.E., Freer, J.E., Johnes, P.J. and Collins, A. (2016) Using hysteresis analysis of high-resolution water quality monitoring data, including uncertainty, to infer controls on nutrient and sediment transfer in catchments. Science of the Total Environment 543, 388-404.

Mächler, M. and Ruckstuhl, A. (2006) Robust Statistics Collaborative Package Development:'robustbase', p. 119.

Miller, A.J., Welty, C., Duncan, J.M., Baeck, M.L. and Smith, J.A. (2021) Assessing urban rainfall-runoff response to stormwater management extent. Hydrological Processes 35(7), e14287.

Nichols, S.J., Peat, M. and Webb, J.A. (2017) Challenges for evidence-based environmental management: what is acceptable and sufficient evidence of causation? Freshwater Science 36(1), 240-249.

Osenberg, C.W., Bolker, B.M., White, J.-S.S., St Mary, C. and Shima, J.S. (2006) Statistical issues and study design in ecological restorations: lessons learned from marine reserves. Foundations of restoration ecology 280.

Osenberg, C.W., Schmitt, R.J., Holbrook, S.J., Abu-Saba, K.E. and Flegal, A.R. (1994) Detection of environmental impacts: natural variability, effect size, and power analysis. Ecological applications 4(1), 16-30.

Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the royal statistical society: Series b (statistical methodology) 71(2), 319-392.

Runkel, R.L., Crawford, C.G. and Cohn, T.A. (2004) Load Estimator (LOADEST): A FORTRAN Program for Estimating Constituent Loads in Streams and Rivers, Reston, Virginia.

Simpson, D., Rue, H., Riebler, A., Martins, T.G. and Sørbye, S.H. (2017) Penalising model component complexity: A principled, practical approach to constructing priors.

Thompson, J., Pelc, C., Brogan III, W. and Jordan, T. (2018) The multiscale effects of stream restoration on water quality. Ecological Engineering 124, 7-18.

Tunaley, C., Tetzlaff, D. and Soulsby, C. (2017) Scaling effects of riparian peatlands on stable isotopes in runoff and DOC mobilisation. Journal of Hydrology 549, 220-235.

Underwood, A. (1994) On beyond BACI: sampling designs that might reliably detect environmental disturbances. Ecological applications 4(1), 3-15.

Underwood, A. and Chapman, M. (2003) Power, precaution, Type II error and sampling design in assessment of environmental impacts. Journal of Experimental Marine Biology and Ecology 296(1), 49-70.

USGS (2016) U.S. Geological Survey, The StreamStats program.

Vaughan, M.C., Bowden, W.B., Shanley, J.B., Vermilyea, A., Sleeper, R., Gold, A.J.,Pradhanang, S.M., Inamdar, S.P., Levia, D.F. and Andres, A.S. (2017) High-frequency dissolved organic carbon and nitrate measurements reveal differences in storm hysteresis and loading in relation to land cover and seasonality. Water Resources Research 53(7), 5345-5363.

Williams, M.R. and Filoso, S. (2023) Changes in hydrology and pollutant loads from stream restoration in an urban headwater catchment. Journal of Hydrology 618, 129164.

Zhang, Q. and Hirsch, R.M. (2019) River water-quality concentration and flux estimation can be improved by accounting for serial correlation through an autoregressive model. Water Resources Research 55(11), 9705-9723.

Report #8, Final Report
September 15, 2024

Appendix A. Questionnaire to assess monitoring needs of the water quality programs

Lora A. Harris compiled the following important questions to ask the local water monitoring program.

1. What kind of stream restoration monitoring are you carrying out?
2. Does it include:
    a. Automated Flow-weighted composite sampling with stilling wells and rating curves?
    b. Hierarchical sampling of baseflow and storms with stilling well?
    c. ?
3. Can you estimate how much one of your stations costs to support for 1 year of sampling? If possible, please break into analyte chemistry costs versus labor for data collection versus labor for interpretation and administration
4. What is the intent/goal of your stream monitoring?
5. Within your department what incentivizes you to do this monitoring?
6. How is it supported financially?
7. What resources are available "in-house" in county government? Do you do all of the monitoring yourselves or do you contract with outside groups? Does stormwater tax revenue provide support for this work?
8. How much error are you comfortable having in your estimates of watershed loads from this monitoring approach? You can report below in terms of confidence?
9. How have you designed your monitoring efforts in the past?
10. How likely are you to use a web-based power analysis tool to evaluate the statistical power associated with your monitoring program?

Appendix B: Semi-automated event delineation R package manual and vignette.