Final Report for Project Sponsored by the Pooled Monitoring Initiative's Restoration Research program from 2016 to 2019 managed by the Chesapeake Bay Trust and this project was supported by the Maryland Department of Natural Resources

<u>Project Title:</u>

Evaluating the Effectiveness and Sustainability of Novel Stream Restoration Designs for Coastal Plain Streams in Maryland: Integrating Existing and New Data from Restoration Monitoring.

Solange Filoso

FINAL REPORT
January 15, 2020

Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, Solomons, MD 20688, USA

Table of Contents (with hyperlinks)

List of Figures and Tables

Figures

Tables

Table 1. Study streams and information about their major watershed, restoration method, and situation of their previously existing monitoring datasets. The text marked in red indicate streams with new field data collected during the present study.

Table 2. Study streams and key characteristics of their surrounding watershed. The text marked in red indicate streams with new field data collected during the present study. The experimental design used during monitoring is indicated in the last column on the right.

Table 3. Parameters measured and method detection limits.

Table 4. Categories of rainfall depths associated with different 24-hour storm events in Anne Arundel County, MD.

Table 5. Annual average discharges of streams monitored during this project. Values are expressed in cubic meters per year (m^3).

Table 6. Number of samples collected biweekly to monthly in each stream in a control and restored stream before and after restoration. Base flow samples were collected biweekly to monthly while stormflow samples were collected in a number of different rainfall events during the monitoring period. Each storm event sampled generated dozens of samples which were analyzed.

Table 7. Results from multiple regression analyses to predict restoration effectiveness based on multiple independent variables (imperviousness in the catchment, drainage area of catchment and stream slope).

Executive Summary

The objective of this project was to compile and synthesize pre- and post-restoration data from different streams monitored in the Coastal Plain and Piedmont physiographic provinces of Maryland to estimate load reduction rates, compare the range of efficacy among projects, and determine potential factors affecting efficacy. Almost half of the streams selected for the comparative analyses had incomplete monitoring datasets, hence, another important objective of the project was to collect additional monitoring data to complement such datasets. In some cases, datasets were incomplete because they included only pre-restoration data while in other cases they were incomplete because post-restoration monitoring ended before the project was potentially mature.

The streams selected were restored with different methods and have different characteristics, but they were monitored with comparable methodologies and experimental approaches. In cases where differences in monitoring strategies emerged, datasets were examined for bias as well as for potential outliers, in a validation process.

Results show that, except for one, all the restored streams analyzed were effective at reducing TN loads while the effectiveness to reduce TP and TSS loads was more variable. Furthermore, the comparative analyses showed that restored headwater streams were more effective than restored lowland channels at reducing loads of TN, TP and TSS and also performed consistently well for all pollutants.

Besides position of the restored stream in the watershed, which is associated with the channel slope, restoration performance for reduction of TP and TSS was correlated with level of development in the watershed, characterized by percent imperviousness, and catchment size. Together, these three variables have a significant predictive power to determine restoration effectiveness not only at reducing TP and TSS loads but also TN.

Introduction and Background

The transport of excess sediment and nutrients in rivers is a major driver of degradation of the Chesapeake Bay ecosystem. Consequently, efforts to improve the Bay health emphasize the reduction of riverine loads to tidal waters, where stream restoration has become an important tool.

Small streams naturally have the functional capacity to retain nutrients and moderate the transport of water and sediments to downstream waters, but such capacity is largely lost when streams become degraded by land use change in their drainage area and other anthropogenic activities. Small streams constitute about two-thirds of fluvial drainage networks, so restoring their functional capacity to retain pollutants can potentially contribute to substantial load reductions to the Bay and help meet TMDL goals.

In the past decade in Maryland, a growing number of stream restoration projects have been implemented throughout the state, especially in urbanized watersheds. To assess the effectiveness of these projects at improving water quality and reducing nutrient and sediment loads to downstream waters, a number of monitoring programs have been carried out in recent years, but results have been mixed.

A number of factors are likely to influence the performance outcome of restoration projects, including how they are monitored and evaluated. In this study, I proposed to compare the performance of different restoration projects regarding water quality improvement and load reduction for total nitrogen (TN), total phosphorus (TN) and total suspended solids (TSS) to determine how it varies among projects and the potential factors influencing it.

A number of factors can influence project performance, including restoration design, stream order, position of the restoration project in the watershed, and characteristics of the surrounding watershed such as land use, catchment size, and topography or channel slope. The objective of this comparative study is to determine the most influential factors while reducing the probable influence of the approaches used to monitor the different projects. Accordingly, only restoration projects monitored with similar methodologies and comparable experimental approaches were included in the analyses. Improving our understanding about factors affecting stream restoration performance can not only help prioritize restoration projects but also help set more accurate expectations for the outcomes.

Objectives

The **first objective** of this project was to assess the variability in stream restoration effectiveness by comparing results from different monitoring studies that used comparable methods. Once the variability was determined, the **second objective** was to examine the influence of key factors such as position of the restoration project in the watershed, pollutant concentrations in streamflow before restoration/upstream, channel slope, watershed size, and level of development in the watershed characterized by percent imperviousness.

Given the number factors that can potentially influence stream restoration effectiveness, this project sought to include data from as many monitored streams as possible. However, the number of streams monitored with comparable methods was limited, hence, a large part of this project was devoted to conducting complementary monitoring in streams with limited datasets but which had been monitored before either before restoration or before and after with comparable methods.

In total, 9 study streams selected for this project were monitored by my group and one stream (Red Hill) was monitored by a different group. In the future, as more comparable datasets become available, the comparative analysis can be expanded to improve robustness.

Methods

Study Streams Selection

Streams for the comparative analyses were selected according to key attributes of the monitoring data. Essentially, the data had to be:

a. collected with similar methodologies and for at least 1 year for baseline conditions (control) as well as in the treatment site (restored stream).
b. collected during base flow and stormflow conditions.

c. Stormflow sampling had to include several individual storms of different sizes throughout the year, representing the typical annual rain distribution for the region.
d. Stream flow had to be recorded continuously for the duration of the project.
e. Rainfall volume had to be recorded continuously in or near the study catchment.

In total, 10 streams were selected (Table 1). Six streams (Linnean, Park Dr., Carriage Hills, Howard's Br., Wilelinor and Red Hill Cr.) had complete existing datasets, hence, the work related to these sites involved only data compilation, verification, and analyses. The remaining four streams (Saltworks, Cypress, Dividing and Church creeks) had only pre-restoration or limited post-restoration data, meaning that new field data had to be collected for the comparative analyses.

Table 1. *Study streams and information about their major watershed, restoration method, and situation of their previously existing monitoring datasets. The text marked in red indicates streams with new field data collected during the present study.*

Study Site	Major Watershed	Restoration Method	Additional new data needed
Dividing Cr.	Magothy	RSC	YES
Saltworks	Severn	RSC	YES
Church Cr.	South	SVR	YES
Cypress Cr.	Magothy	RSC	YES
Howard's Br.	Severn	SWC	NO
Wilelinor	South	RSC	NO
Linnean	Rock Cr.	RSC	NO
Park Drive	Anacostia	RSC	NO
Carriage Hills	Severn	SPC	NO
Red Hill Br.	Patuxent	NCD	NO

Field Experimental Designs

Each stream selected was monitored using the design that best matched the availability of resources at the time of monitoring or, in some cases, the fact that restoration had been implemented long before monitoring started. Therefore, different experimental designs were used to monitor the selected streams. However, despite differences, all streams were monitored for baseline conditions (control) and after treatment (restored).

Restoration designs used included: 1) Before and after restoration above and below the targeted reach in an **Above-Below-Before-After** sampling design, 2) before and after restoration below the targeted reach using a **Before-After** sampling design, 3) before and after restoration in the targeted stream as well as in a control stream using the **BACI** design, 4) after restoration above and below the targeted reach in an **Above-Below** sampling design, or 5) after restoration below the targeted stream and from a control stream in a similar adjacent catchment using the **paired catchment** design. A detailed description of each design and information about the advantages and limitations of each one is provided below.

1. Above-Below-Before-After Design

This design allows monitoring stream chemistry and discharge above and below the reach targeted for restoration before and after restoration is implemented (Figure 1). This is a standard "impact analysis" design where parameter values are compared before versus after restoration has been implemented. The "before" data provide baseline or temporal control conditions, but evidence for causal links is limited by the lack of spatial controls. In other words, while changes are measured, it is often difficult to determine whether the changes would have occurred independently of restoration. This design is also difficult to assess restoration effectiveness if the changes observed are not large.

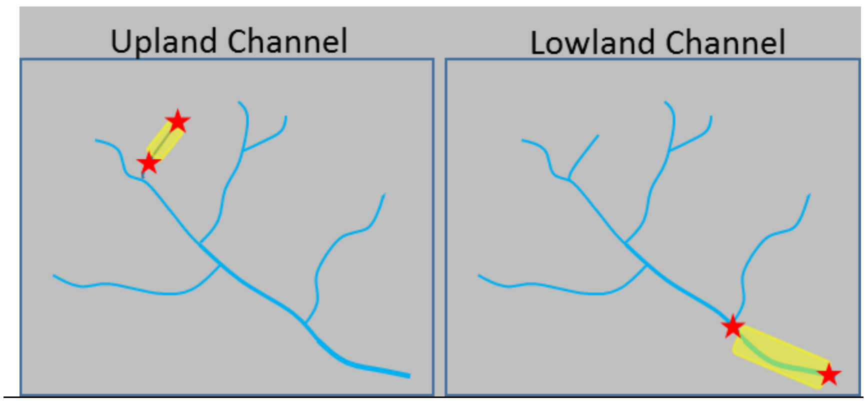


***Figure 1**. Schematic of monitoring design Above-Below Before-After, where the control site is represented by the site above the restored reach, either on an upland channel (left panel) or lowland channel (right panel).*

2. Before-After Design

This design consists of monitoring stream chemistry and discharge at a downstream location before and after restoration is implemented (Figure 2). The pre-restoration data provide information about the "baseline" conditions of the stream before restoration, against which restoration effectiveness is assessed after implementation is completed. This approach is commonly used when financial resources are limited, but this makes it difficult to attribute observed changes to the restoration given that the monitoring just tracks change over time, without controlling for external factors that may have influenced the changes observed. However, if all conditions in the watershed (including land use, runoff, and flow) remain the relatively stable over the monitoring period, changes from baseline conditions can be attributable to the restoration, especially if the monitoring is done over an extended period of time. Potentially, the use of multivariate regression analyses can help isolate possible covariates and confounding factors that influenced changes from baseline conditions after restoration and improve confidence in the results.

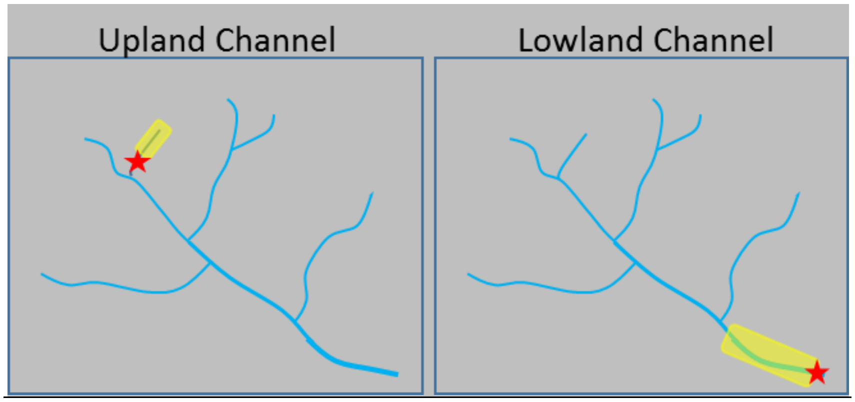


Figure 2. *. Schematic of monitoring design Before-After, where the control site is represented by the site monitored before restoration, either in an upland channel (left panel) or lowland channel (right panel). The monitoring sites before and after restoration are the same.*

3. Before/After/Control/Impact (BACI) Design

This is the preferred design to assess the impact of restoration, where parameter values are monitored before versus after activities have been implemented, and a "control" site serves as a reference site (Figure 3). The "before" data provide baseline or temporal control conditions and the control site provides some measure of whether natural changes coincide with changes seen in the impact site. This design also allows assessment of whether the trend of a response is towards the reference condition.

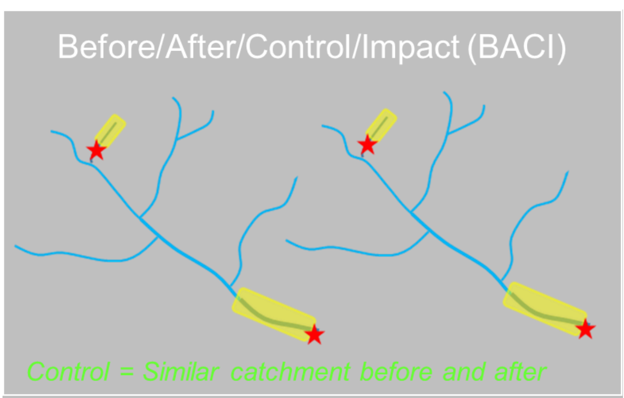


Figure 3. *Schematic of monitoring the BACI design, where the control site is represented by both, a site below restoration monitored before implementation and also a reference site in a similar catchment.*

4. Above-Below Design

This control/impact monitoring design (Figure 4) is commonly used to evaluate in-stream changes from restoration when the project has been implemented already. Estimates of load reduction in the restored reach are based on comparisons of upstream and downstream values normalized by the drainage areas, but the results should be interpreted with caution if land use of the drainage area associated with the downstream site is drastically different from land use upstream.

Common statistical approaches used with the above-and-below design is the t-test of the differences between paired observations at the above and below stations. Parametric and nonparametric (distribution free) t-test approaches are appropriate.

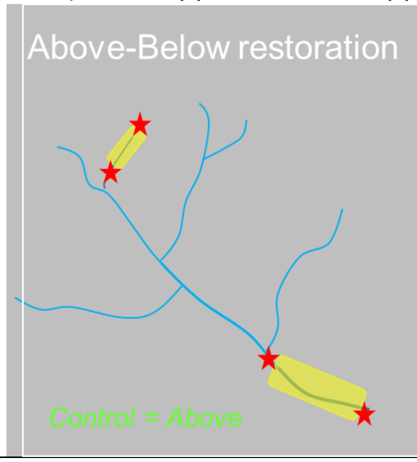


Figure 4. *Schematic of monitoring design Above-Below, where the control site is represented by the upstream site above the restored reach. Both the control and downstream sites are monitored together only after restoration.*

5. Paired Catchments

This design (Figure 5) requires a minimum of two catchments paired for control/impact analysis, where the catchments have the same characteristics, so differences in the restored stream can be attributed to the restoration. The stream in the control catchment serves as a check over year-to-year or seasonal climatic variations and receives no changes in management practices during the study. Ideally, monitoring in both catchments starts before restoration is implemented to confirm that the paired watersheds behave in similar ways prior to implementation of restoration. When before data are used, this is also considered a before/after/control/impact (BACI) design and provides strong inference. This approach is often used when activities encompass a small headwater watershed, and there is no opportunity for sampling above the site, as it is the case of upland headwater streams.

Figure 5. *Schematic of pair catchments monitoring design, where the control site is represented by a nearby catchment with characteristics similar to the restored site before restoration.*

Description of Study Streams Selected

Streams with Complete Pre- and Post-restoration Datasets

Howard's Branch

Howard's Branch is a lowland alluvial channel within a wide valley, draining an area of 98 ha composed of low-density residential, transportation, and forest land uses (Figure 6). Only 17% of the watershed is impervious, but some of the impervious area is directly connected to the stream channel through stormwater drainage pipes, generating large amounts of surface runoff during storm events. The stream restoration project was implemented in 2002 and it consisted of capping fine sediment that had been deposited in the channel valley from a failed dam with a layer of white sand, and of grading the stream banks and placing structural controls along the channel to control and guide water flows. The objectives of restoration were channel stabilization and erosion control, wetland creation, stream-floodplain reconnection and creation of topographic conditions conducive to cedar tree propagation. The restored stream has been transformed into a stream-wetland complex.

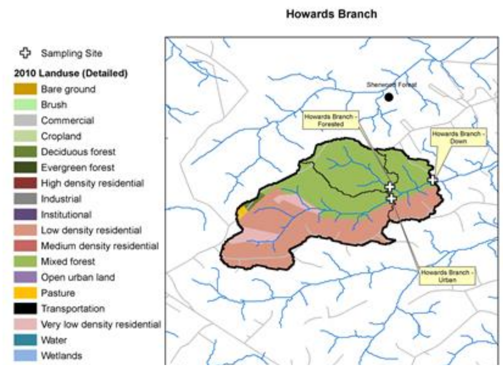


Figure 6. *Drainage area boundary and land use at the Howards Branch study site.*

Wilelinor

Wilelinor is also a lowland alluvial channel with a wide valley (Figure 7). The stream drains a developed watershed of about 106 ha with a mix of commercial, industrial and high-density residential land use. Prior to restoration, the stream valley was impounded by two failing stream ponds that captured surface runoff from an adjacent highway and upstream development.

The Wilelinor restoration was implemented in 2004, and involved mass grading and placement of structural controls of water flows. The objectives of the restoration were channel stabilization, wetland creation, stream-floodplain reconnection and creation of topographic conditions conducive to cedar tree propagation. Like Howards Br., the restored stream at Wilelinor has been transformed into a stream-wetland complex.

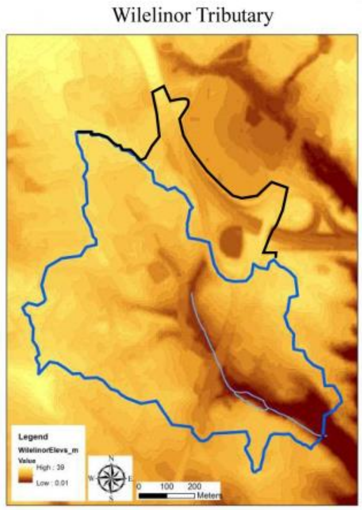


Figure 7. *Topographic boundaries (dark blue lines) of the Wilelinor Tributary watershed derived from high resolution elevation data. Additional area conveying runoff to WIL from a highway storm drain system is delineated in black. The reconfigured reaches coincide with the bifurcated light blue stream line segments shown in the lowland valleys.*

Carriage Hills

The Carriage Hills restoration was implemented in a highly incised headwater channel on the west side of the Severn River watershed (Figure 8). The channel drained an area of approximately 5.6 ha of low-density residential land. Adjacent to this catchment, a similar one in size and land use served as a control site. The streams draining the control and restored catchments were monitored simultaneously in a pair-catchment experimental design.

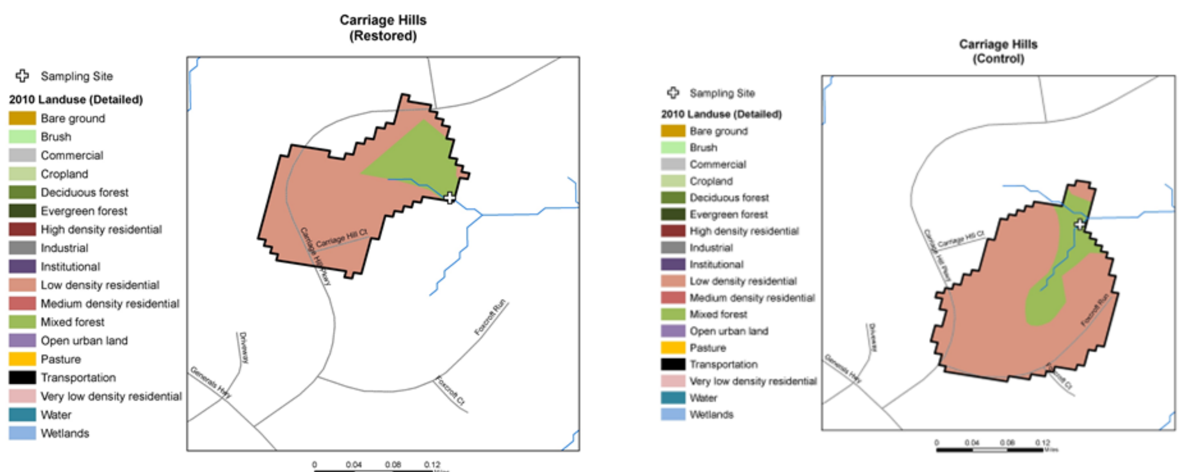


Figure 8. *Topographic boundaries of the catchments in Carriage Hills drained by the restored stream (left panel) and the Control stream (right panel). The restored and control streams are first-order channels of the same stream downstream.*

Linnean

The regenerative streamwater conveyance at Linnean was constructed in an upland headwater channel draining 14 hectares of mixed land use and cover in an urban region of Maryland, north of Washington DC. The monitoring included sampling the restoration channel and an adjacent control stream before and after restoration. Land use in the control catchment was similar to that of the restored channel, but the drainage area was smaller because the delineation of the control catchment was based on elevation derived from a topographic map, not the sewer drainage network as with the restoration catchment (Figure 9).

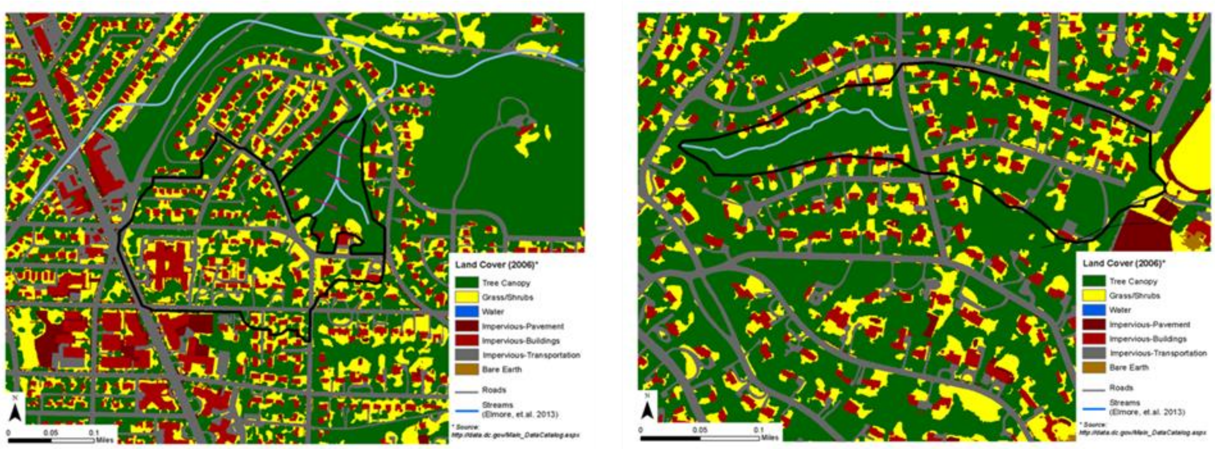


***Figure 9**. Map of the land use surrounding the restored stream (left) and the control stream (right). The catchments boundaries are delineated with a black line and the stream reaches marked with blue lines.*

Park Dr.

The restored stream at Park Dr. is a headwater channel that had been deeply incised due to urban development in its drainage area. Monitoring of the project consisted of sampling during the post-construction period at the outflows of the restored reach and of a control stream (Figure 10).

***Figure 10**. Land cover / land use in the Park Drive control (left) and restored (right) catchments; boundaries are delineated by black lines. Sampling sites where discharge was measured using Parshall flumes are indicated by white circles located at the outflow of each stream reach.*

Red Hill Branch

The project reach is a first order stream located in an urbanized watershed with predominantly residential land use. Prior to restoration, the stream channel was unstable and incised, probably because of increased surface flows from development in the watershed. The stream restoration project involved the stabilization of an approximate 2,100-linear foot segment of an unnamed tributary to Red Hill Branch located immediately downstream of a stormwater outfall, as well as an additional 300 feet of outfall stabilization. The restoration was based on the “natural channel” design approach, and involved re-grading of the stream bed and banks, the use of stone toe protection, bioenhancement techniques, imbricated riprap for bank protection, the installation of grade control structures, and a comprehensive planting plan for the site.

Streams with Incomplete Post-restoration Datasets (Needed Complementary Monitoring)

Church Creek

Church Creek has been rated in the top tier for restoration prioritization based on Anne Arundel County’s watershed assessment and the South River Federation’s strategic plan. Its watershed was 42% impervious and impacted by population growth and development as it sits downstream of five major shopping centers and two major highways (Figure 11). The monitoring was based on the Before-After restoration design with a site downstream of the restoration reach.

Before this project, the monitoring dataset for this site included only pre-restoration information. Complementary monitoring was necessary to provide post-restoration data.

***Figure 11**: Land use map and aerial photograph of the Church Creek subwatershed to the stream sampling station. The watershed boundary delineation provided by AA Co. was used to calculate land use areas (yellow line, left panel). The site where samples were collected and discharge measured downstream of the restoration reach is marked in yellow on the aerial photograph (right panel).*

Dividing Creek

Dividing Cr. is located within the Magothy River watershed, which is the most developed in Anne Arundel County. Land use in the surrounding watershed of Church Cr. is predominately commercial, institutional and residential. About 31% of the watershed is composed of impervious cover (Figure 12).

Similar to Church Cr., the restoration project at Dividing Cr. had been monitored only during the pre-restoration period before this complementary monitoring. Monitoring was based on a Before-After-Upstream-Downstream design, with two sites monitored before and after restoration.

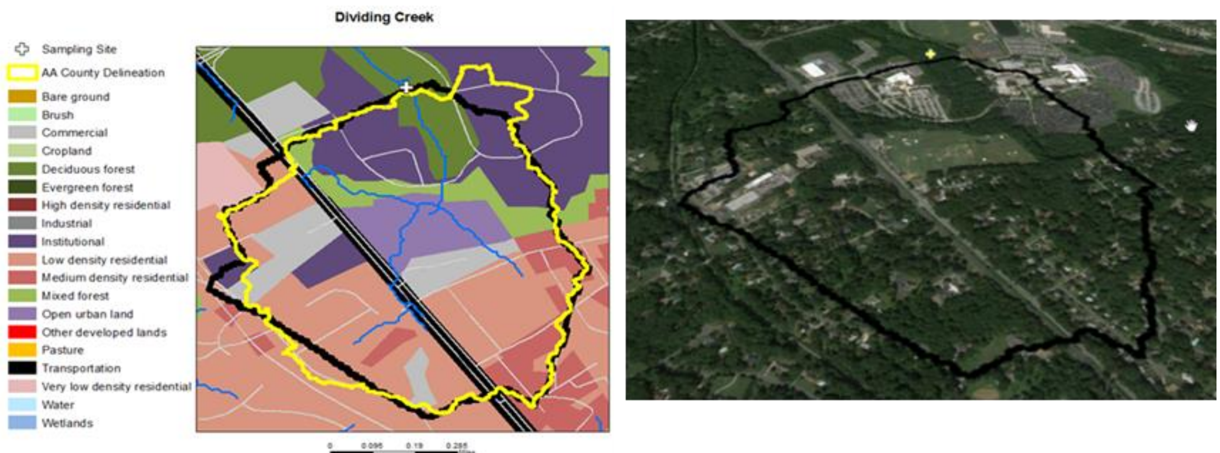


***Figure 12**: Land use and aerial photograph of the Dividing Creek sub-watershed to the stream sampling station. The watershed boundary delineation provided by AA Co. was used to calculate land use statistics (left panel). Up- and down-stream sampling stations are marked in yellow in the aerial photograph (right panel).*

Cypress Creek

The stream channel monitored in Cypress Cr. is the main stem of North Cypress Creek, also within the Magothy River watershed. The Cypress Cr. watershed is one of the most developed in AACo., with more than 50% of its area cover by impervious surface. Consequently, prior to restoration, the targeted channel was highly degraded and unstable, with areas of severe bank erosion.

Before the complementary monitoring, this site had been monitored before and after restoration. Before restoration, sampling included a site upstream and one downstream of the restoration reach. After restoration, monitoring continued only in the downstream site due to limitation of resources. The monitoring site downstream was located at the end of the catchment near the tidal boundary (Figure 13). The upstream site was located about 2500 ft. upstream of the downstream site (close to Richie Hwy); monitoring at the upstream site was discontinued in 2011.

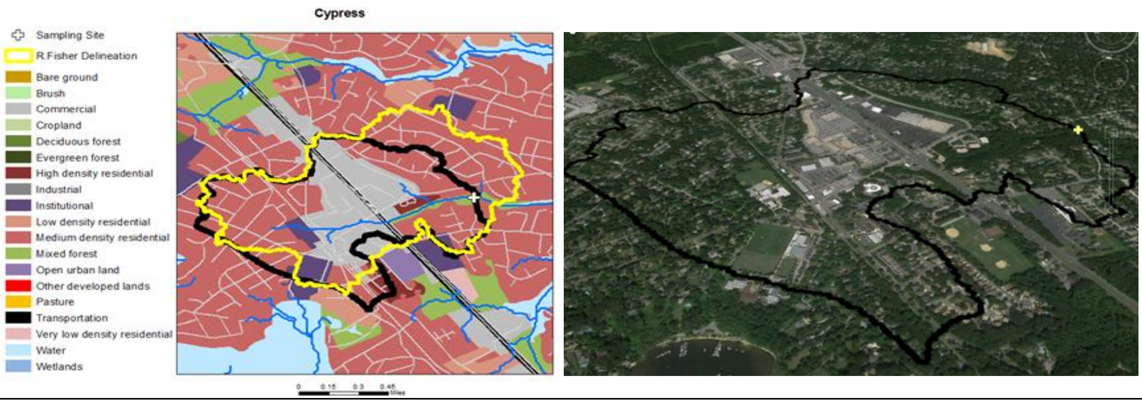


***Figure 13**: Land use in the Cypress Creek sub-watershed to the stream sampling station and aerial photography. The watershed boundary delineation provided by AA Co. was used to calculate land use statistics (left panel). Up- and down-stream sampling stations are marked in yellow in the aerial photograph (right panel)*.

Saltworks

Saltworks Creek drains an area of extensive development, with more than 70% of impervious surface in its surrounding watershed (Figure 14). Consequently, the stream had become a distressed urban waterway associated and increase in the sediment loads and the deposition of silt in the tidal zone.

To reduce pollutant loads and improve the overall health of the stream, in 2013 a critical reach of the stream adjacent to a shopping mall and two major roads was restored using the regenerative stormwater conveyance system design (RSC). The reach was monitored a few years ago before and after restoration, downstream of the end of the reach targeted for restoration. However, the post-restoration dataset covered a period of less than a year after restoration implementation, when the stream was still recovering from construction. The complementary monitoring post-restoration was done to provide additional post-restoration data after the project had time to mature.

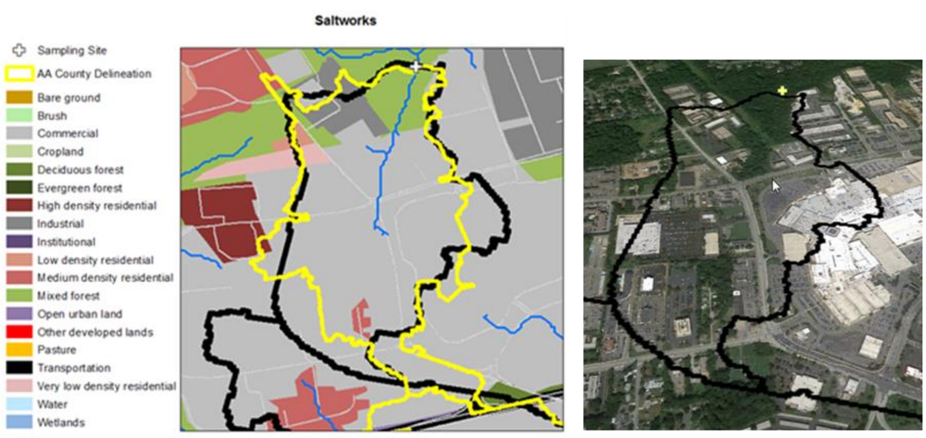


***Figure 14**: Land use map and aerial photograph of Saltworks Creek sub-watershed to the stream sampling station marked in yellow in the aerial photograph (right panel). The watershed boundary delineated in yellow provided by AA Co. was used to calculate land use statistics (left panel).*

Monitoring Methodology

The experimental design used to monitor each selected streams is listed in Table 2 below, along with some of the key characteristics of their catchment area.

Table 2. *Study streams and key characteristics of their surrounding watershed. The text marked in red indicates streams with new field data collected during the present study. The experimental design used during monitoring is indicated in the last column on the right.*

Study Site	Catchment area (ha)	% Impervious-ness	Slope	Position in Watershed	Experimental Design
Dividing Cr.	89	32	Low-gradient	Lowland	Above-Below-Before-After
Saltworks	49	55	Low-gradient	Lowland	Before-After
Church Cr.	227	46	Low-gradient	Lowland	Before-After
Cypress Cr.	143	46	Low-gradient	Lowland	Before-After
Howard's Br.	96	17	Low-gradient	Lowland	Above-Below
Wilelinor	106	40	Low-gradient	Lowland	Above-Below
Linnean	13	22	Mid-gradient	Headwaters	Above-Below-Before-After
Park Drive	1.3	28	High-gradient	Headwaters	BACI
Carriage Hills	6	19	High-gradient	Headwaters	Paired Catchment
Red Hill Br.	50	25 (est.)	Mid-gradient	Headwaters	Above-Below-Before-After

Field Methods

Water sampling

Baseflow samples were collected manually (grab sample) in the control and restored sites every two to four weeks. Discharge was measured immediately before water samples were collected at each stream site. Water samples were collected in 1-L pre-leached polyethylene amber bottles and an aliquot was immediately filtered through pre-rinsed glass fiber filters (Whatman 0.45 μm) to separate dissolved from particulate nutrient phases. Filtered water samples were stored in high-density polyethylene bottles and kept on ice and in the dark to be transported to the laboratory (within 6 h) where they were frozen until analyzed for dissolved N and P forms. Particulate material in the remaining unfiltered samples was collected on pre-combusted (500°C for 1.5 h) glass-fiber filters following the recommended sampling protocols used by EPA for the determination of PN and PF. According to the protocol, a measured volume of sample water (no less than 200 mL) is poured into a filtration apparatus in increments of 50 mL and filtered with a vacuum pressure no greater than 10 in. Hg. When the filter is saturated, it is removed from the base of the apparatus using forceps, folded in half and stored in labeled glassine envelopes to be frozen and stored until analyses can be performed. Total suspended solids (TSS) samples were collected by passing a known volume of sample through a pre-weighed glass-fiber filter. Concentrations were calculated as the weight of the filter after collection of sample, divided by the volume of water filtered. The remaining unfiltered water was kept on ice and in the dark for the collection of total suspended solids (TSS) in pre-weighed glass-fiber filters in the laboratory. Dissolved and particulate samples were stored in a freezer at -20°C prior to analysis. Duplicate field samples were routinely collected about every 30 samples, while blank field samples were collected in 10% of the sampling campaigns.

Stormflow samples were collected in the same sites that baseflow samples were collected but sampling was done on an event basis. At least two storm events were sampled every two to three months in each site throughout the monitoring period in order to encompass storm sizes of different sizes and seasons. Sampling was done using automated samplers (ISCO 6712) configured to collect 24 or more samples per event, either as discrete or time-integrated composite samples (NRC 2008).

Samplers were turned on manually before the beginning of each storm, but water collection started only when stream stage reached approximately 2.0 cm above the precedent baseflow level. Collection was done in 10 to 20 min intervals for the duration of the storm event, while stream stage remained elevated above baseflow levels. When a storm event lasted for more than 6 to 12 hours, ISCO bottles were replaced so sampling could continue throughout the falling limb of the hydrograph. During warm months, ISCOs were filled with dry ice to keep samples cold until they could be retrieved after the end of the storm event. After collection, all samples were stored in a dark cooler to be transported to the laboratory where they were filtered within 24 hours and subsequently frozen until lab analyses.

Hydrological measurements

Instantaneous discharge was measured using the cross-sectional area method (Gordon et al. 2004) at each sampling station immediately before baseflow samples were collected, and also during a wide range of stage levels. Data loggers (Pressure Transducers HOBO U20-001-04) installed in each sampling station recorded stage levels in 5- to 10-minute intervals during the monitoring period. Wherever possible, a calibrated H-flume or a weir was installed with pressure transducers for more accurate stream flow data.

The data collected by the PTs installed in the stream channels were corrected for changes in atmospheric pressure with data recorded by an unsubmerged PT installed nearby each stream. All PTs were inspected and data generally downloaded every 60 to 72 days.

Laboratory Analyzes

All water samples collected were analyzed for a suite of chemical constituents, including total nitrogen (as the sum of total dissolved nitrogen (TDN), dissolved organic nitrogen (DON), nitrate (NO_3), ammonium (NH_4), and N in particulate form (PN)), total dissolved phosphorus and phosphate (TDP, PO_4), and total suspended solids (TSS).

Nitrate was determined using an ion chromatograph (Dionex IC 1000), following the EPA 300.0 Method for Inorganic Anions by Ion Chromatography. In this method, a small volume of water sample is injected into the ion chromatograph, and the anions of interest are separated and measured, using a system comprised of a guard column, separator column, suppressor device, and conductivity detector. When the anions are injected into the ion chromatograph, they are separated on an analytical column and a sodium carbonate/bicarbonate solution serves as a mobile phase. Once separated, the anions pass through an anion suppressor where they are converted to their highly conductive acid forms and the sodium carbonate/ bicarbonate eluent is converted to very weakly conductive water. The separated anions in their acid forms are measured by conductivity and are identified on the basis of retention times compared to standards. Ammonium (NH_4) was determined using the Berthelot Reaction method (Kerouel and Aminot 1987) in which a blue-colored compound similar to indophenol forms when a solution of ammonium salt is added to sodium phenoxide, followed by the addition of sodium hypochlorite (Kerouel & Aminot, 1987). Total dissolved nitrogen (TDN) concentrations were determined on filtered water samples using the persulfate digestion method, where all N is converted into nitrate (D'Elia et al. 1977), and then passed through a granulated copper-cadmium column to be reduced to nitrite. The nitrite is determined by diazotizing with sulfanilamide and coupling with N-1-naphthylethylenediamine dihydrochloride to form a colored azo dye. Color is proportional to nitrogen concentration.

The potassium persulfate method was also used for determining total dissolved P (TDP) concentrations. The method consists of oxidizing organic and inorganic P to orthophosphate under heated acidic conditions. Ammonium molybdate and potassium antimony tartrate then react in an acid medium with dilute solutions of orthophosphate to form an antimony-phosphomolybdate complex, which is reduced to an intensely blue-colored complex by ascorbic acid. Color is proportional to orthophosphate concentration.

Dissolved inorganic orthophosphate (PO_4) was determined following the EPA Method 365.1 (1979). In the method, ammonium molybdate and potassium antimony tartrate react in an acid medium with dilute solutions of orthophosphate to form an antimonyphosphomolybdate complex which is reduced to an intensely blue colored complex by ascorbic acid. Color is proportional to orthophosphate concentration.

Particulate N, or the nitrogen content of the particulate matter retained on the GFF, was measured with a Costhech ECS 2010 CHNSO Elemental Analyzer. In this method, samples are combusted in pure oxygen (O_2) under static conditions. Products of combustion are passed over suitable reagents in the combustion tube where complex oxidation occurs. In the reduction tube, oxides of nitrogen (N) are converted to molecular N. The carbon dioxide (CO_2), water vapor and N are mixed and released into the thermal conductivity detector where the concentrations of the sample gases are measured.

Total suspended solids (TSS) concentrations were determined by filtering a known volume of well mixed sample through a 47 mm pre-weighed GFF. Prior to filtration, each GFF was dried at 105°C overnight, cooled at room temperature in desiccators, and weighed. Following filtration, filters were frozen at -20 °C until analysis. Upon analysis, filters were dried at 105°C overnight, cooled at room temperature in desiccators, and weighed. Concentrations were calculated as the weight of the filter (minus the filter's weight) divided by the volume of water filtered. Results are expressed in mg/L. The detection limits associated with all the analytical methods described above are listed in Table 3.

Table 3. *Parameters measured and method detection limits.*

Parameter	Method Detection Limit
Nitrate + Nitrite, mg/L as dissolved N	0.01 mg/L
Ammonium, mg/L as dissolved N	0.004 mg/L
Orthophosphate, mg/L as PO4-P	0.0007 mg/L
Total Dissolved Nitrogen (mg/L as N)	0.006 mg/L
Total dissolved P, mg/L as PO_4-P	0.0015 mg/L
Particulate Nitrogen (mg/L) as N	0.006 mg/L
Total Suspended Solids (mg/L)	2.4 mg/L

Precipitation Measurements

Rain depth was measured continuously during the monitoring period using a tipping bucket rain gage (Onset HOBO RG3-M) and a plastic Tenite bulk volume collector installed in the vicinity of each study stream. The rain data were used to characterize rain duration and intensity during the course of the monitoring study and compare with stream discharge (e.g. runoff producing rainfall). The tipping bucket rain gauges were inspected and loggers downloaded monthly, whereas the plastic gauge was routinely monitored on a more regular basis.

Wunderground weather stations near the study streams were used to obtain surrogate precipitation data for the sites where a tipping bucket rain gauge could not be installed.

Data Analyses

Precipitation Data

All the hourly precipitation data collected during the pre- and post-restoration monitoring periods in the study streams included in the comparative analyses were compiled and used to calculate daily rainfall depths for the different streams included in the analyses. Daily rainfall values were subsequently divided into six categories of rainfall depths (Table 4) to determine the frequency as well as the relative contribution of the different categories to total precipitation in the monitoring period.

Table 4. *Categories of rainfall depths associated with different 24-hour storm events in Anne Arundel County, MD.*

Rainfall characteristic	Rainfall Depth (in)
< 1-year storms	<= 0.25
	> 0.25, <= 0.5
	> 0.5, <= 1
	> 1, <= 2
1- to 5-year storms	> 2, <= 5
10- to 100-year storms	> 5

Discharge Calculations

Instantaneous discharge (m^3) measured in each sampling station was computed as the product of flow velocity (m/s) and the cross-sectional area (m^2) of the channel where stream velocity was measured. The relationship between instantaneous discharge and different stage levels was subsequently used to transform stage data into continuous record of stream discharge.

Except for Red Hill (which was monitored by Tetra-Tech and KCI), annual discharge in the study streams was computed as the sum of the continuous discharge data within a year. Base flow discharge was estimated by adding all the continuous discharge at or below the maximum base flow levels measured in the field during the year. Stormflow discharge was estimated as the difference between annual and base flow discharge.

Calculating Mean Concentrations and Loads

For each study stream, the mean concentrations of TN, TSS and TP in base flow were calculated from concentrations measured in water samples collected biweekly or monthly in field campaigns during base flow conditions. Average stormflow concentrations were estimated by first calculating the flow-weighted mean concentration for each storm event sampled before and after restoration (or control stream) (Equation 1), and then averaging them weighted by rainfall depths.

Flow-weighted mean concentrations were calculated as:

$$FWMC = (\sum C_i\, Q_i) / \sum Q_i \qquad (1)$$

where Ci represent concentrations of individual samples collected during the storm event and Q_i represent the discharge at the time interval interval *i* when the sample was collected. The denominator is the sum ($\sum Q_i$) of interval stormflow discharge during the event sampled.

Base flow loads were estimated as a product of the estimated annual base flow discharge and mean base flow concentrations. Stormflow loads were estimated as a product of the average flow-weighted mean concentrations of storm events and the estimated annual stormflow discharge.

Preparing Datasets for Comparative Analyses

The datasets compiled and synthesized for the comparative analyses included hourly rainfall, stream discharge and concentrations of TN, TP and TSS, as well as information about key characteristics of watersheds surrounding the selected streams, which could influence restoration performance and pollutant load reduction capacity. Except for Red Hill Cr., the data on stream discharge and concentrations were used to calculate loads in the selected streams according to the methods described above. Before that, however, the monitoring data compiled and which had been collected previously were manually validated based on methods proposed by EPA (https://www.epa.gov/sites/production/files/2015-06/documents/g8-final.pdf).

In the validation process, datasets were examined for bias as well as for potential outliers in order to ascertain the influence of anomalies on the comparative analyses. For instance, extreme concentration values of total suspended solids (TSS) were checked against other constituents analyzed and against discharge and rainfall data to determine if they were potential outliers. Furthermore, rainfall depths of the events sampled for stormflow in the streams before and after restoration were examined to see if they were comparable in size and frequency, and also representative of the typical annual rainfall distribution in the region.

To minimize potential bias generated from differences in differences in rain sizes and frequency of storms sampled for stormflow before restoration or in control sites and after restoration, rain events were categorized into six different groups of rain size and given weights according to their relative contribution (%) to average annual rainfall in the region. Subsequently, the median flow-weighted mean concentrations of stormflow events associated with the same rain size category were calculated. The median values from each rain category were used to calculate the rain-weighted mean concentrations of stormflows before and after restoration.

Assessing Restoration Effectiveness

Restoration effectiveness was assessed by comparing average pollutant concentrations in restored and their respective control sites, and by comparing average annual loads. Comparisons of loads included annual base flow and stormflow loads and total loads.

Restoration was considered effective if total annual pollutant loads decreased significantly with stream restoration in comparison control or/and pre-restoration sites.

The effectiveness of restoration among different streams was assessed by comparing annual yields, which are annual loads normalized by the stream drainage area. Effectiveness among streams was also assessed by calculating the percent change in annual loads associated with restoration.

Statistical Analyses

T-test and non-parametric Wilcoxon-Mann-Whitney test were used to test differences between base flow concentrations before restoration or control sites and after restoration, after the data were examined for normality. Likewise, flow-weighted mean concentrations from the different storm events sampled during the monitoring periods in the different streams were examined for normality and subsequently tested for differences between control and post-restoration concentrations using the same tests.

The correlations between watershed characteristics listed in Table 3 above and restoration effectiveness, determined as the difference between loads measured during pre-restoration or in control sites and loads measured post-restoration, were evaluated using Spearman correlation analyses. Subsequently, linear regressions were used to examine the degree of relationship between percent changes and the significantly influential variables. Multiple regression was also used to determine if the relationship between restoration performance (dependent variable) and watershed characteristics improved if two or more independent variables are used together as predictive variables.

Results

Results from Complementary Monitoring Effort

As explained before, four streams were monitored during the present project to complement limited existing datasets collected previously to assess the effectiveness of restoration on pollutant loads. Two of these streams, Saltworks and Cypress creeks, were monitored before and after restoration implementation, but monitoring results showed no significant improvements in water quality and load reductions probably because the post-restoration monitoring period ended too soon after restoration was implemented. The present project collected additional post-restoration data after the projects had time to become fully established in order to make the datasets more suitable for the comparative analyses.

The other two streams, Church and Dividing creeks, had only pre-restoration data because delays in restoration implementation made post-restoration monitoring unfeasible within the time frame of previous monitoring efforts. In this project, post-restoration data was collected for over a year in order for changes associated with restoration to be assessed and the datasets included in the comparative analyses.

Rainfall Depths of Stormflows Sampled

In Saltworks and Cypress creeks, the previous post-restoration monitoring period was limited to only 6 and 8 months, respectively, and did not include stormflow sampling during storm categories that contribute most of the total precipitation volume. The complementary monitoring added another two years of post-restoration data and improved the range of storm sizes sampled (Figure 15). In total, 9 storm events were sampled in each stream during the complementary post-restoration monitoring period in addition to 21 and 19, baseflow samples, respectively.

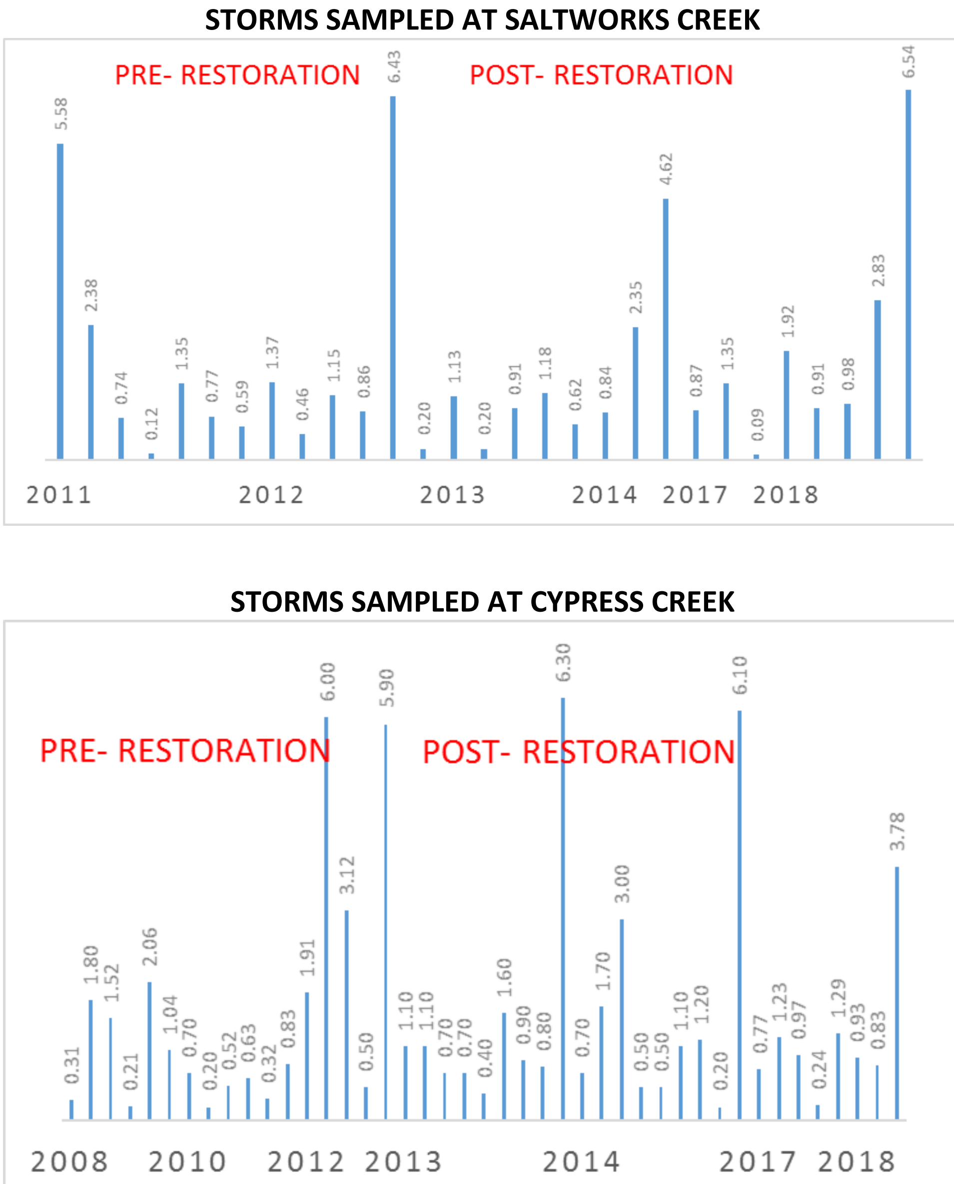


Figure 15*. Rain depths for storm events sampled for stormflow before and after restoration at Saltworks Cr. (****top panel****) and Cypress Cr. (****bottom panel****).*

For Church and Dividing creeks, no post-restoration data existed before the complementary monitoring period, when samples were collected for consecutive 18 months. During this period, eight different storm events ranging from less than 0.2 inches to over 7 inches were sampled for stormflow in each stream (Figure 16).

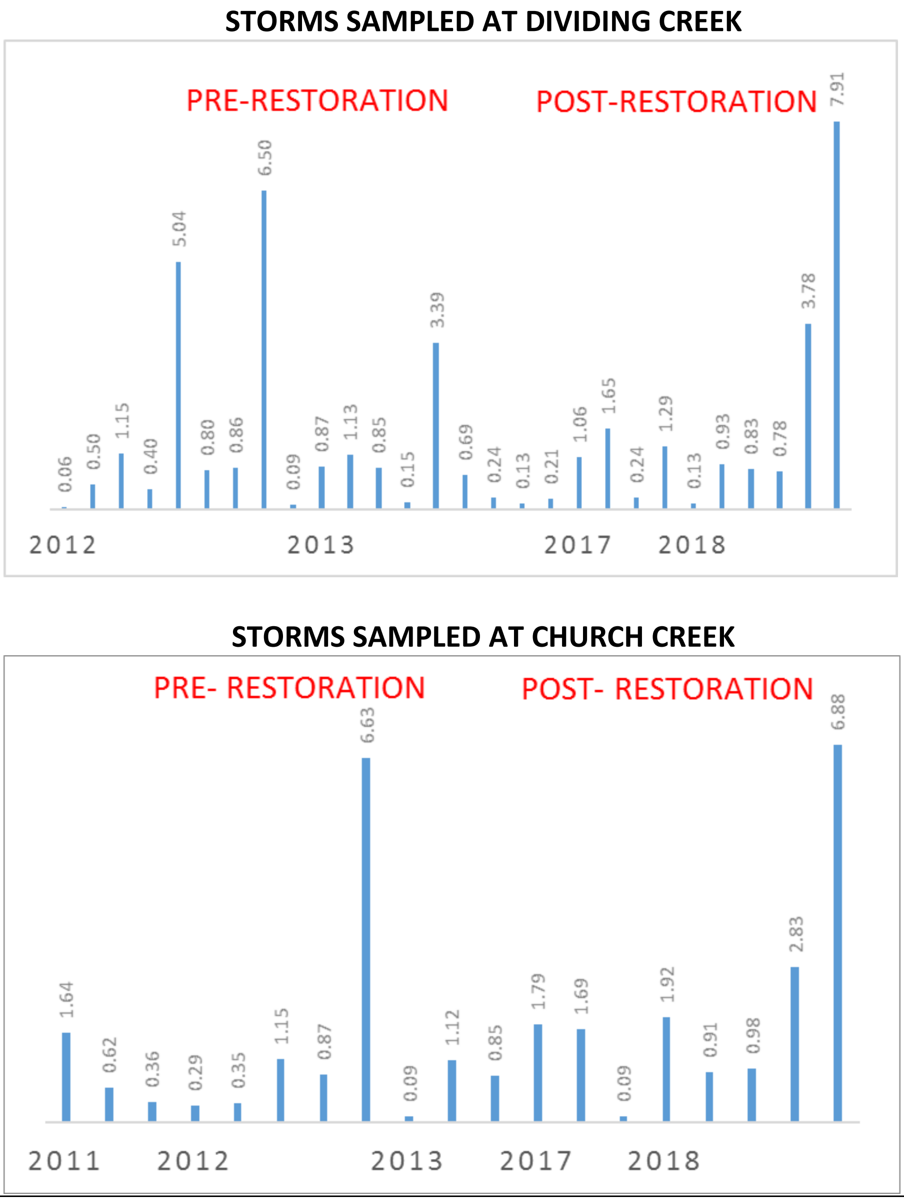


Figure 16*. Rain depths for storm events sampled for stormflow before and after restoration at Dividing Cr. (****top panel****) and Church Cr. (****bottom panel****).*

Concentrations

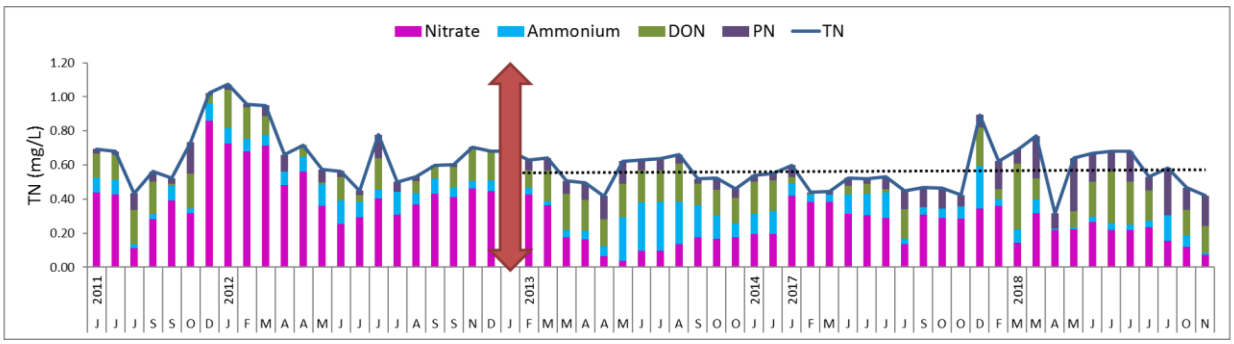
Base Flow

Total Nitrogen

During base flow conditions, pre-restoration TN concentrations in Saltworks averaged about 0.65 mg/L, decreasing to an average of approximately 0.5 mg/L after restoration. This decrease was not statistically significant (t-test, $p>0.05$), while the additional data from the complementary monitoring period suggest that TN concentrations are actually trending higher in the post-restoration period (Figure 17, upper panel).

In contrast, restoration of Cypress Creek resulted in significantly lower TN concentrations (t-test $p<0.05$) and a decreasing trend extending into the complementary monitoring period. (Figure 17, lower panel). In both streams, however, the predominant form of N changed from nitrate, which is typical of urban streams, to dissolved organic N (DON), which suggests that restoration either promoted denitrification or limited the nitrification process in base flow.

BASEFLOW NITROGEN IN SALTWORKS CREEK



BASEFLOW NITROGEN IN CYPRESS CREEK

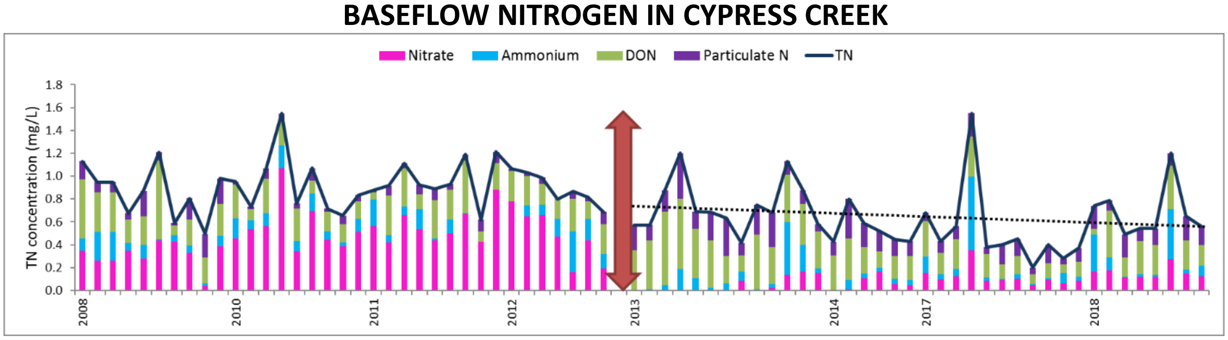


Figure 17. *Concentrations of TN partitioned into different N forms for water samples collected during base flow conditions at Saltworks Creek (**top panel**) and Cypress Creek (**bottom panel**) in the pre-restoration and post-restoration phases of the project. The period marking the restoration implementation and separating the pre- and post-restoration monitoring periods is marked by a red arrow. The black dotted line indicates the trend in concentrations during the post-restoration period.*

For Dividing and Church creeks, restoration also reduced the average base flow TN concentrations but the changes were not statistically significant (t-test, $p>0.05$). In fact, the post-restoration data from both streams revealed an increasing trend in TN concentrations since restoration implementation, especially due to increases in DON concentrations in 2018

(Figure 18). Decaying riparian forest vegetation may have acted as an additional source of N to these streams, at least during the monitoring period. Restoration in Dividing Creek, in particular, was relatively recent at the time of post-restoration monitoring.

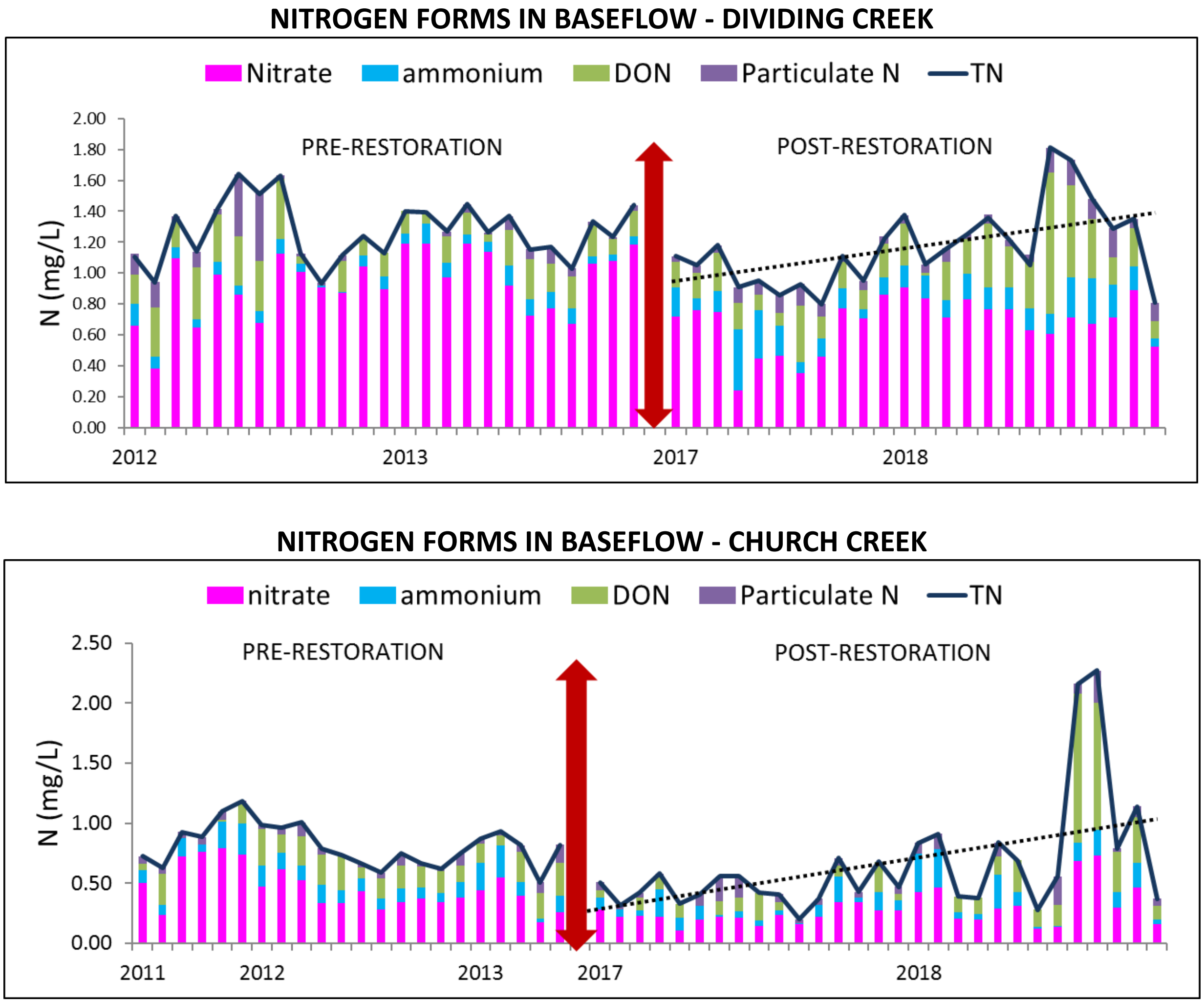


Figure 18. *Concentrations of TN partitioned into different N forms in water samples collected during base flow conditions at Dividing Creek (**top panel**) and Church Creek (**bottom panel**) in the pre-restoration and post-restoration phases of the project. The period marking the restoration implementation and separating the pre- and post-restoration monitoring periods is marked by a red arrow. The black dotted line indicates the trend in concentrations during the post-restoration period.*

Total Suspended Solids

The effects of restoration on TSS concentrations in baseflow were different from the effects on TN. In Saltworks, for instance, TSS concentrations increased dramatically immediately after restoration implementation, decreasing a few months later to levels similar to those in pre-restoration (Figure 19, upper panel). The decreasing trend continued throughout the complementary monitoring period. Yet, the mean TSS concentration remained significantly higher after restoration (t-test, $p<0.05$).

In Cypress Creek, post-restoration concentrations were significantly higher than pre-restoration concentrations (t-test, $p<0.05$). The higher concentrations were led by periodic high peaks observed throughout the post-restoration monitoring period (Figure 19, lower panel). However, the magnitude of peaks decreased with time and resulted in a decreasing trend in TSS concentrations from the time of restoration implementation to the complementary monitoring period.

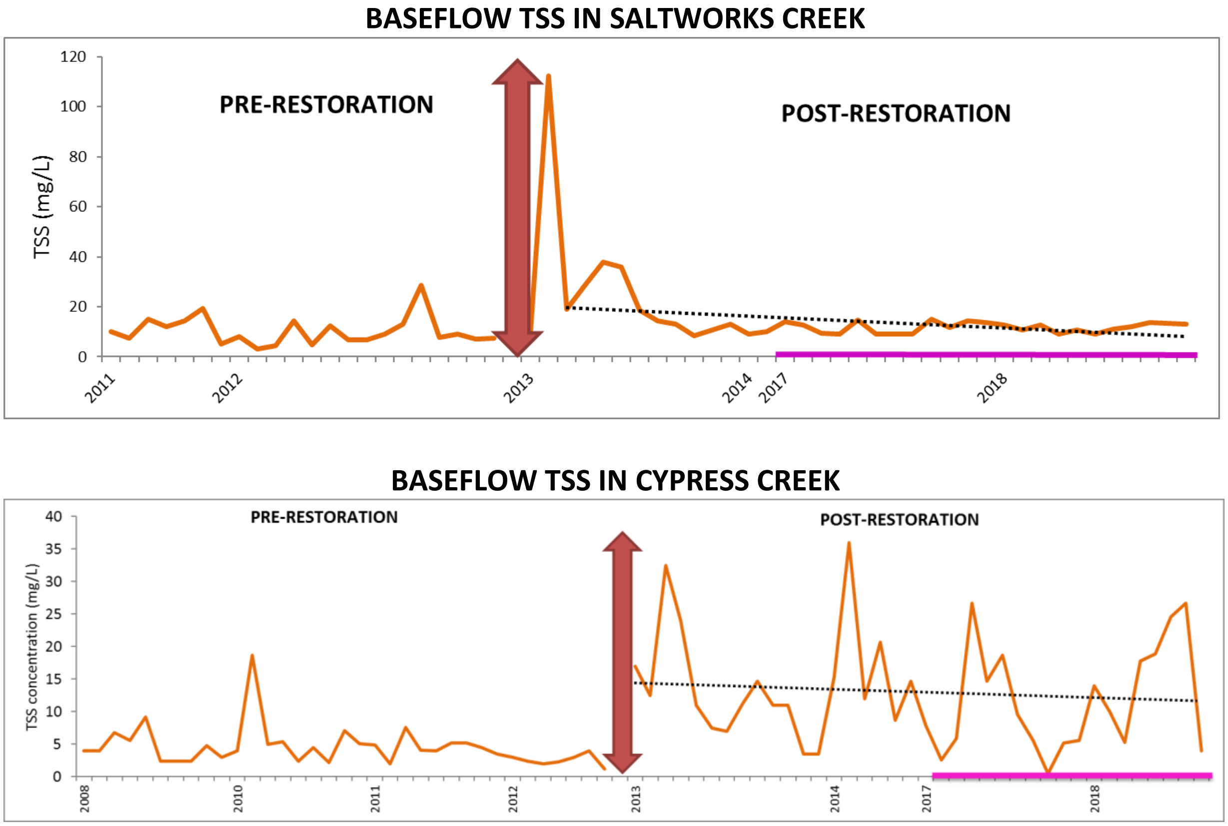
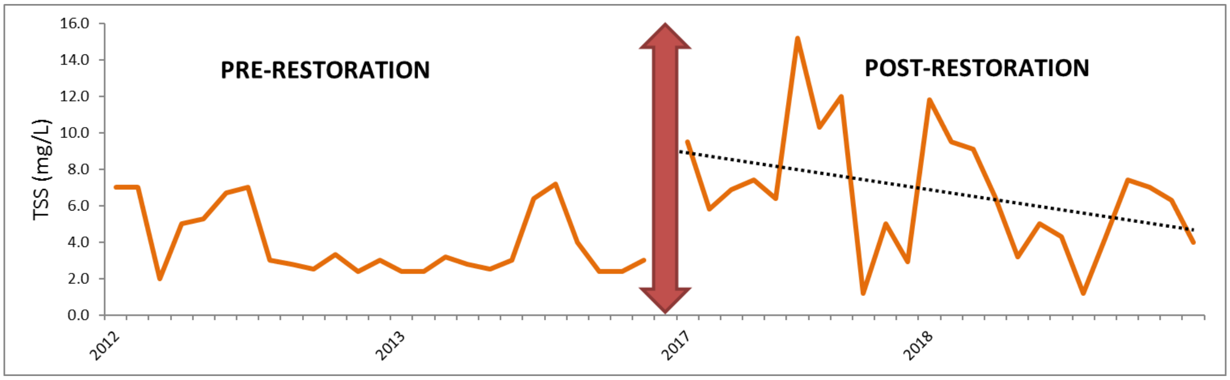


Figure 19. *Concentrations of TSS in base flow at Saltworks Cr. (**upper panel**) and Cypress Cr. (**lower panel**). The period marking the restoration implementation and separating the pre- and post-restoration monitoring periods is marked by a red arrow. The black dotted line indicates the trend in concentrations during the post-restoration period. The pink line indicates the complementary monitoring period of the present project.*

Restoration of Dividing and Church creeks was followed by high peaks of TSS concentrations in base flow (Figure 20). However, while at Dividing Creek such variability resulted in significant higher concentrations, at Church Creek there was no significant change ($P>0.05$). On the other hand, the post-restoration data revealed a decreasing trend in TSS concentrations in Diving Creek and the opposite at Church Creek.

BASEFLOW TSS IN DIVIDING CREEK



BASEFLOW TSS IN CHURCH CREEK

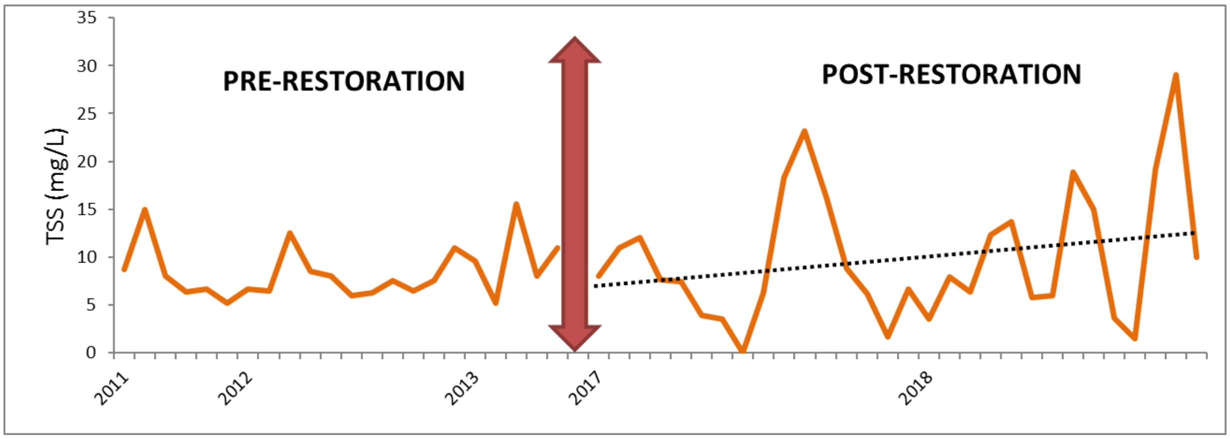
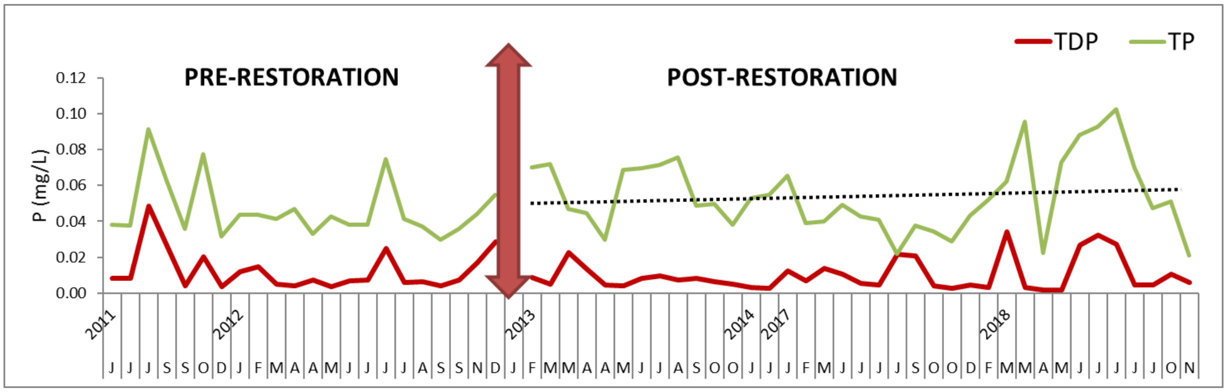


Figure 20. *Concentrations of TSS in base flow at Dividing Cr. (**upper panel**) and Church Cr. (**lower panel**). The period marking the restoration implementation and separating the pre- and post-restoration monitoring periods is marked by a red arrow. The black dotted line indicates the trend in concentrations during the post-restoration period.*

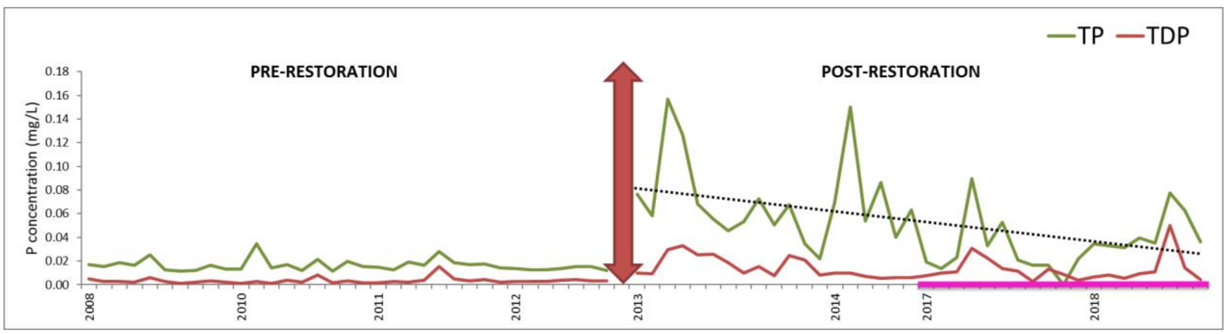
Total Phosphorus

The impact of restoration on base flow concentrations of TP was mostly negative (Figure 21). Except for a significant decrease in Cypress Creek ($p<0.05$), base flow concentrations in all streams increased after restoration. The increase was significant only for Church Creek ($p<0.005$), but the trend was positive in all streams with increasing concentrations.

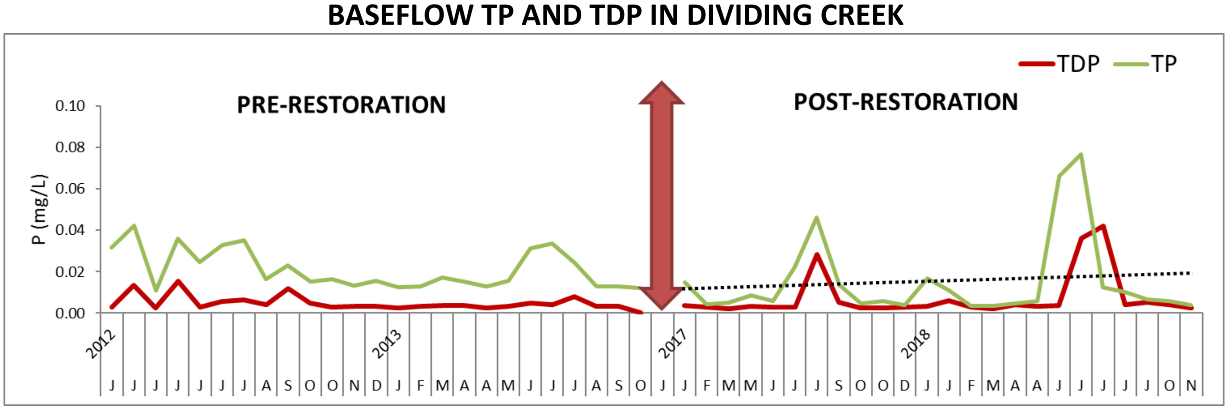
BASEFLOW TP AND TDP IN SALTWORKS CREEK



BASEFLOW TP AND TDP IN CYPRESS CREEK



BASEFLOW TP AND TDP IN DIVIDING CREEK



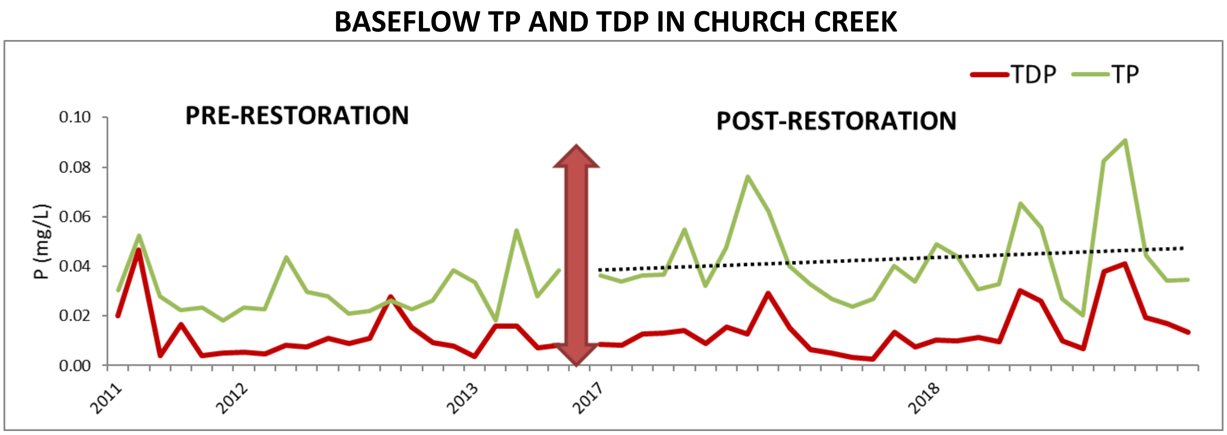


Figure 21. *Concentrations of TP (green) and TDP (red) in base flow in the streams monitored during this project (Saltworks, Cypress, Dividing and Cypress creeks). The period marking the restoration implementation and separating the pre- and post-restoration monitoring periods is marked by a red arrow. The black dotted line indicates the trend in concentrations during the post-restoration period.*

Stormflow

Changes in stormflow concentrations associated with restoration were examined by comparing flow-weighted means of storms of different sizes sampled during the pre- and post-restoration monitoring periods. The objective was to determine how restoration impacted stormflow concentrations on a range of storm sizes representative of typical storm sizes observed on an annual basis.

Total Nitrogen

The median flow-weighted mean concentrations of TN in stormflow before restoration ranged around 1.5 mg/L in the monitored streams, except in Saltworks where concentrations were slightly lower (Figure 22). After restoration, the median concentrations decreased to levels around 1 mg/L or less. Except for Saltworks, concentration reductions were statistically significantly ($p<0.005$). After restoration, there was no correlation between concentrations and storm sizes, meaning that larger storms were not necessarily associated with the highest concentrations. In general, larger storms had lower concentrations after than before restoration.

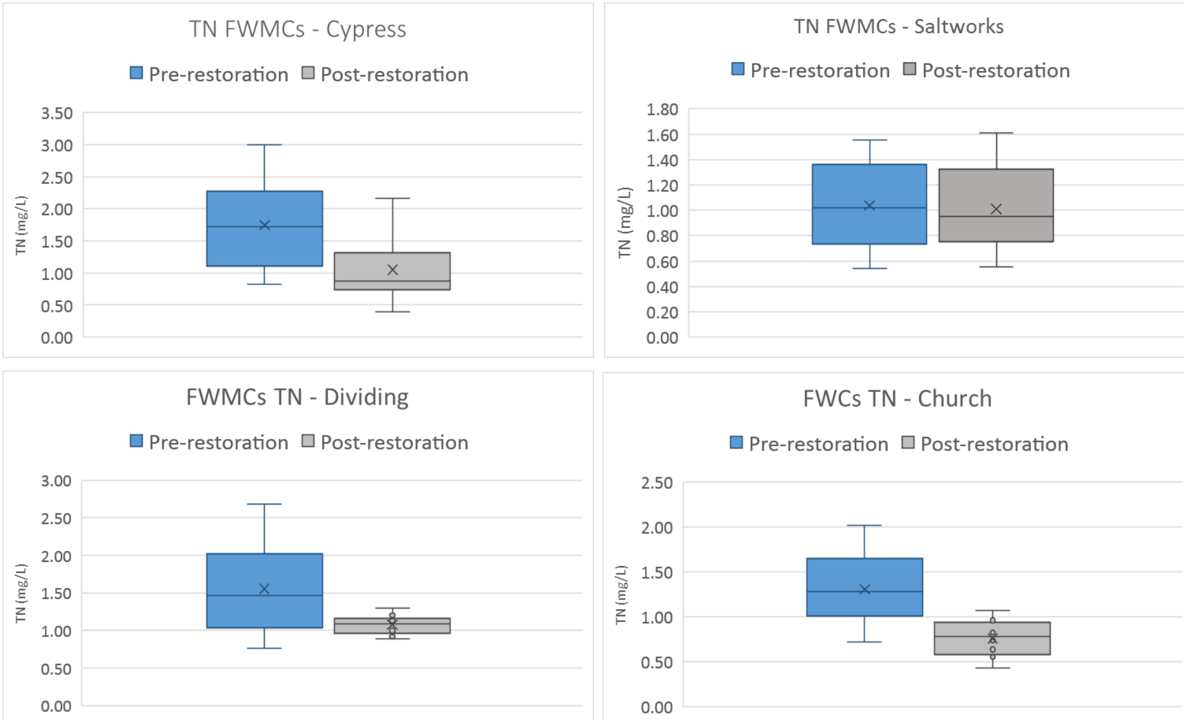


***Figure 22**. Box plots of FWM TN concentrations in stormflow sampled during several storms before and after restoration in the streams monitored during this project. For each stream, the median flow-weighted mean concentration is represented by a line across the box. This line indicates that 50% of the concentrations were greater than the median value and 50% lower. The top and bottom "whiskers" represents the maximum and minimum concentrations above and below the median, respectively.*

Total Suspended Solids

For total suspended solids (TSS) (Figure 23), restoration did not result in a significant decrease in stormflows (p>0.01). The median FWM concentration at Cypress, Dividing and Church creeks decreased, but not enough to be statistically significant. At Saltworks, concentrations increased with restoration.

After restoration, there was no correlation between concentrations and storm sizes, meaning that larger storms were not necessarily associated with the highest concentrations. In general, larger storms had lower concentrations after than before restoration.

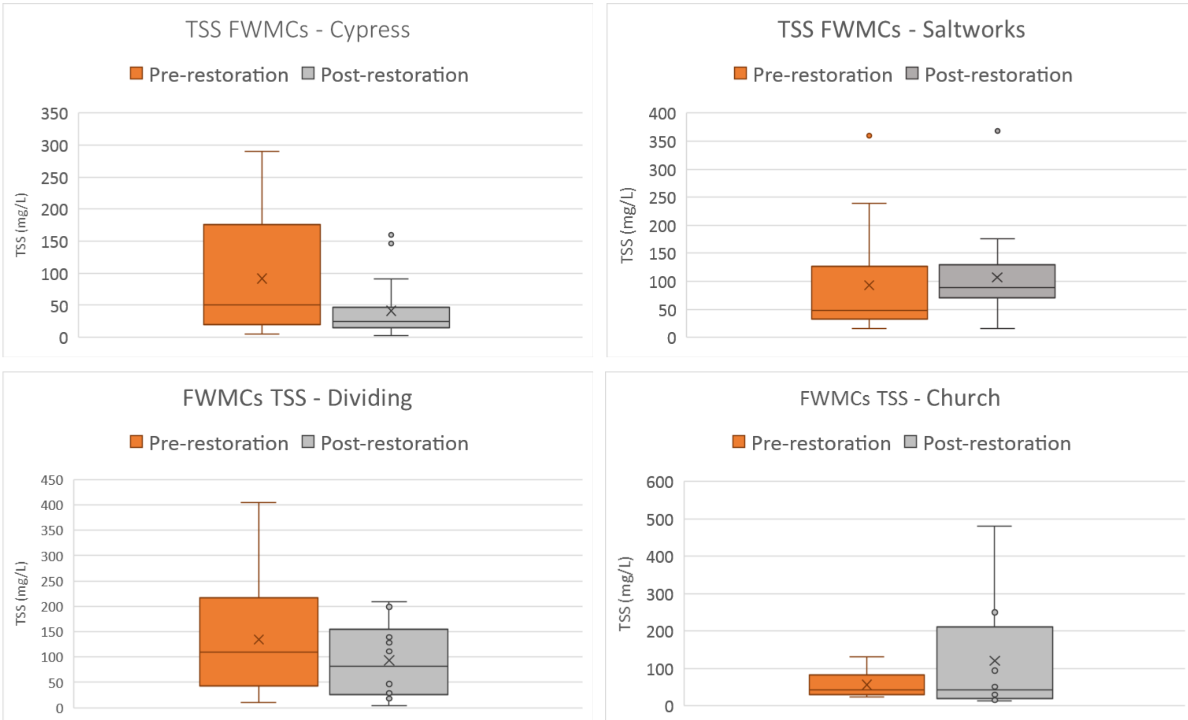


***Figure 23**. Box plots of FWM TSS concentrations in stormflow sampled during several storms before and after restoration in the streams monitored during this project. For each stream, the median flow-weighted mean concentration is represented by a line across the box. This line indicates that 50% of the concentrations were greater than the median value and 50% lower. The top and bottom "whiskers" represents the maximum and minimum concentrations above and below the median, respectively. Outliers are represented by circles outside whiskers.*

Despite non-significant decreases in flow-weighted mean concentrations during, at Cypress and Saltworks creeks restoration changed the correlation between concentrations and storm sizes. Before restoration, stormflow TSS concentrations in these streams were positively correlated with storm size but restoration eliminated such correlation, meaning that larger storms no longer associated with the highest TSS concentrations (Figure 24).

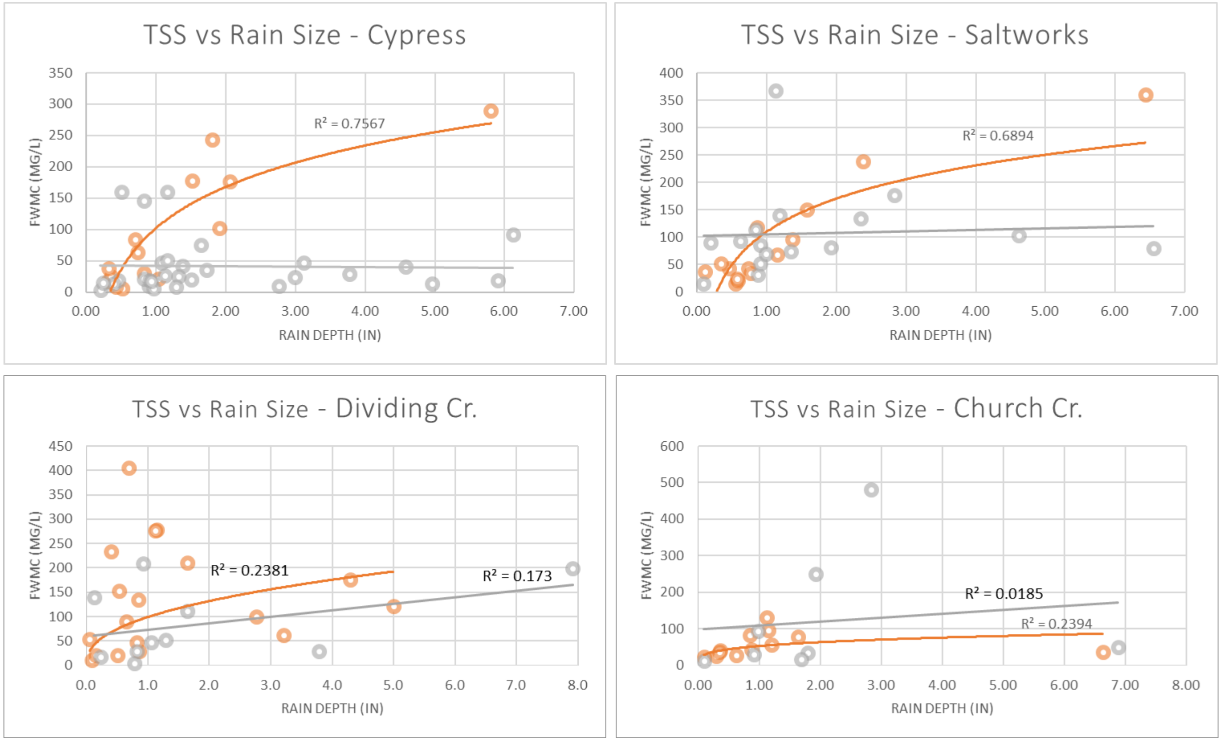


***Figure 24**. Correlations between stormflow TSS FWMCs and rain depths for storms collected at the streams with extended monitoring data.*

Total Phosphorus

The median flow-weighted mean concentrations of TP in stormflow of the streams monitored either stayed the same or increased significantly with restoration (Figure 25), as it was the case in Saltworks ($p<0.005$). After restoration, there was no correlation between concentrations and storm sizes, meaning that larger storms were not necessarily associated with the highest concentrations. In general, larger storms had lower concentrations after than before restoration.

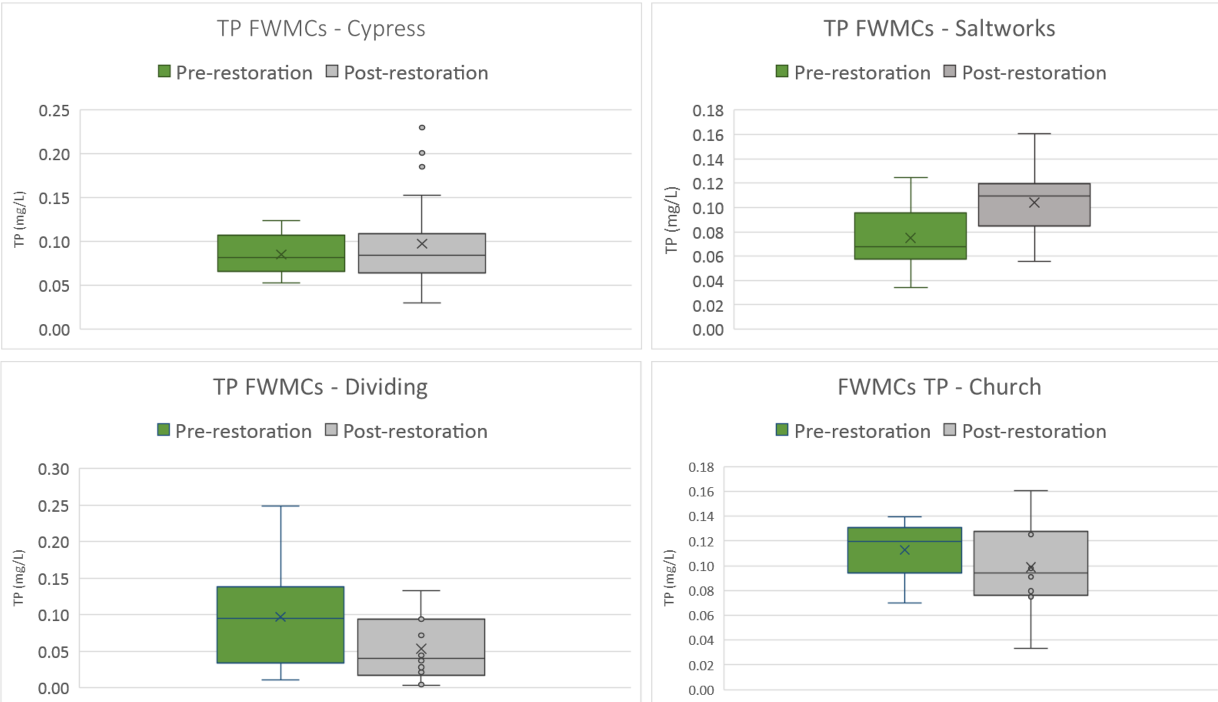


***Figure 25**. Box plots of FWM TP concentrations in stormflow sampled during several storms before and after restoration in the streams monitored during this project. For each stream, the median flow-weighted mean concentration is represented by a line across the box. This line indicates that 50% of the concentrations were greater than the median value and 50% lower. The top and bottom "whiskers" represents the maximum and minimum concentrations above and below the median, respectively.*

Stream Hydrology

The annual discharges estimated for each stream based on average volume of annual rain during the monitoring periods are shown in Table 5. The largest average annual discharge was observed at Church Creek, where the total volume was about twice as large as in the other streams. Restoration generally decreased the total discharge volume by 3 to 5% in all streams, probably because of higher evapotranspiration rates in warmer months with growing wetland and floodplain vegetation.

***Table 5**. Annual average discharges of streams monitored during this project. Values are expressed in cubic meters per year (m^3 y^{-1}).*

Stream	Pre-restoration	Post-restoration
	m^3 yr^{-1}	
Cypress Cr.	513	487
Saltworks Cr.	379	360
Dividing Cr.	438	416
Church Cr.	1128	1094

The four streams monitored during this project are Coastal Plain lowland channels with wide valleys and gentle slope, where the relative contribution of stormflow to annual discharge should not outweigh that of base flow. Consequently, restoration of these streams sought to reestablish a more balanced hydrologic regime by increasing water residence time and, consequently, the prevalence of base flow conditions to provide more opportunities for nutrient and sediment retention. Figure 26 below shows that the relative contribution of base flow to total discharge increased between 5% at Cypress Cr. to 13% at Saltworks.

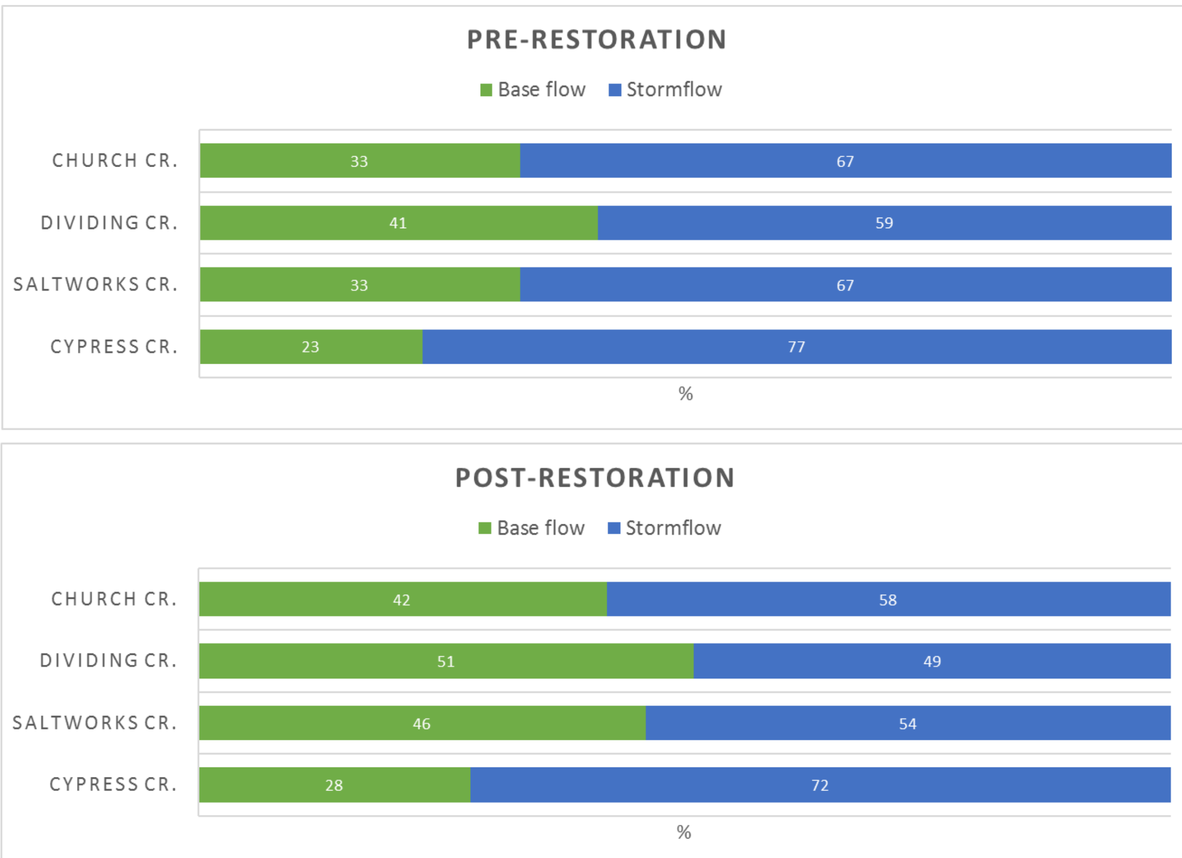


***Figure 26**. Relative contribution of base flow (green) and stormflow (blue) to annual discharge before restoration (upper panel) and after restoration (lower panel) in the streams monitored in this project. Values are expressed as percentage of total discharge (%).*

Loads

As shown above, restoration increased the relative contribution of base flow to total runoff and also reduced the amount of total annual runoff. Furthermore, concentrations of TN in both base flow and stormflow decreased and resulted in a significant decrease in TN loads in all streams, especially because of substantial decreases in stormflow loads (Figure 27).

In absolute terms, the largest reductions occurred in Church Creek followed by Cypress, Dividing and Saltworks creeks. In relative terms, Cypress Cr. had the largest reduction (50%) followed by Church, Saltworks and Dividing creeks (43%, 36%, 28%, respectively).

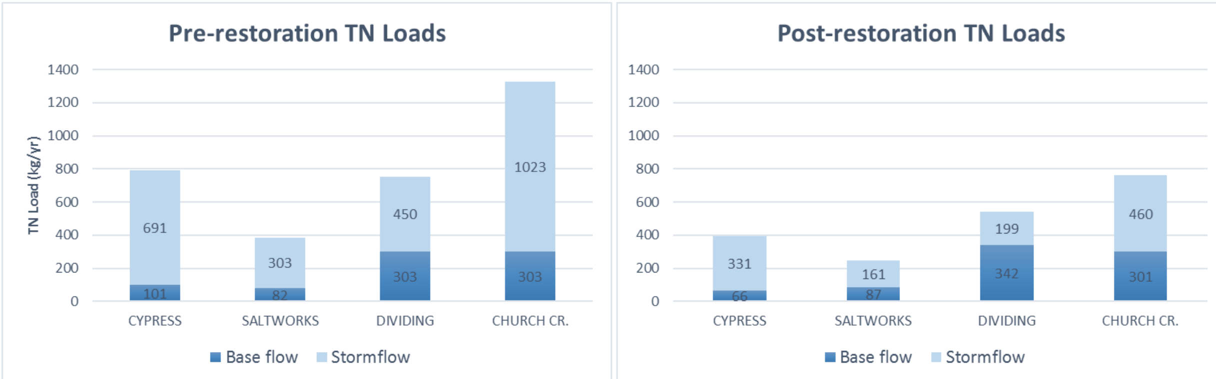


Figure 27. *Annual loads of TN in the monitored streams partitioned as base flow and stormflow loads before and after restoration (left and right panels, respectively).*

For TP, restoration resulted in load reduction in three out of four streams monitored in this project. Total P loads decreased in Cypress, Dividing and Church creeks, where concentrations in stormflow decreased significantly with restoration, but no significant changes occurred in Saltworks Creek (Figure 28). In general, TP loads in baseflow increased but did not affect total loads noticeably.

Figure 28. *Annual loads of TP in the monitored streams partitioned as base flow and stormflow loads before and after restoration (left and right panels, respectively).*

The impact of restoration on TSS loads was more pronounced than on TN and TP, despite the fact that stormflow loads increased slightly in the stream that had the largest total loads prior to restoration (Figure 29). In the remaining streams, TSS loads decreased by about half as stormflow loads decreased substantially with restoration. In all streams, however, TSS loads in base flow increased usually due to higher concentrations of particulate matter other than suspended sediment.

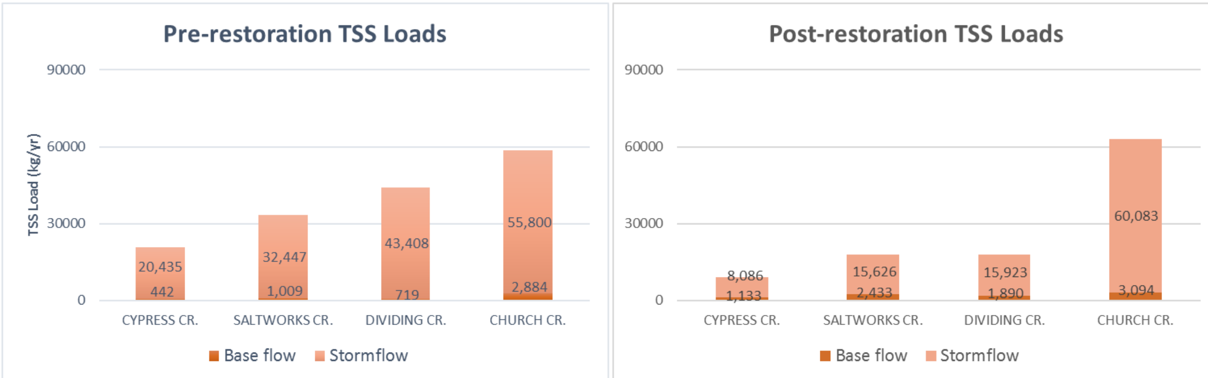


Figure 29*. Annual loads of TSS in the monitored streams partitioned as base flow and stormflow loads before and after restoration (left and right panels, respectively).*

Comparative Analyses

Data Compiled from All the Streams Selected

A summary of the number of samples collected during base flow and stormflow conditions in each stream before and after restoration or in a control site is shown in Table 6. The number base flow samples reflect the number of days of biweekly to monthly water sample collection and discharge measurements in the monitoring periods. The number of stormflow samples indicates the number of storm events sampled, not the number of individual water samples collected during the storms.

The streams selected for the comparative analyses were monitored in different periods of time and for different time lengths. However, on average, they were all monitored at least one year before restoration (or in a control stream) and one year after restoration. The average monitoring length pre- and post-restoration was 1.5 years.

Table 6*. Number of samples collected biweekly to monthly in each stream in a control and restored stream before and after restoration. Base flow samples were collected biweekly to monthly while stormflow samples were collected in a number of different rainfall events during the monitoring period. Each storm event sampled generated dozens of samples which were analyzed. Streams monitored during the present project are indicated by red and bold letters.*

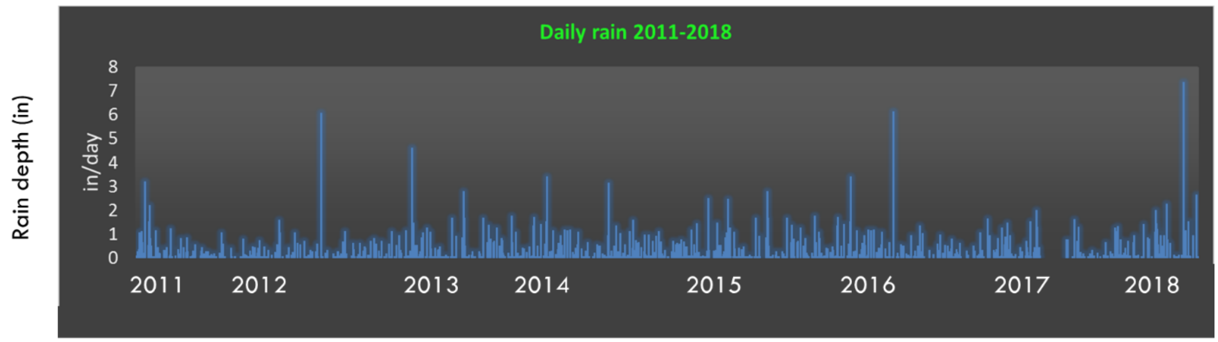
Stream	Site	Base Flow Samples (#)	Stormflow Events Sampled (#)	Total #	Monitoring Length (yr)
CARRIAGE HILLS	Control	30	15	45	1.5
	Restored	26	15	41	1.5
LINNEAN	Pre- rest	21	16	37	1.0
	Post-rest	16	16	32	2.0

PARK DR.	Control Restored	18 20	24 23	45 43	1.0 1.5
RED HILL BR.	Pre- rest Post-rest	16 24	14 36	30 60	1.0 4.5
CYPRESS CR.	Pre- rest Post-rest	34 34	12 28	46 62	1.5 3.0
SALTWORKS CR.	Pre- rest Post-rest	23 36	13 16	36 52	1.5 1.5
DIVIDING CR.	Pre- rest Post-rest	27 23	17 10	44 33	1.5 1.0
CHURCH CR.	Pre- rest Post-rest	21 22	11 8	32 30	1.5 1.0
HOWARDS BR.	Control Restored	49 49	19 19	68 68	2.5 2.5
WILELINOR	Control Restored	15 39	15 15	30 54	2.0 2.0

Precipitation Data

The daily rainfall depths recorded during the pre- and post-restoration monitoring periods in the streams selected for the comparative analyses ranged from detection level (~0.05 in) to more than 7 inches (Figure 30, upper panel).

The frequency of the different rainfall depth categories as well as their relative contribution to total precipitation during this monitoring period are shown in the pie charts (Figure 30, lower panel). The majority of storm events (63%) were equal or smaller than 0.25 inches. However, these frequent storm events accounted for only about 12% of the total rainfall volume. More than 50% of the total rainfall was from rain events > 0.5 inches. A third of the total volume was from events larger than 1 inch, which accounted for only about 10% of the events. Therefore, the weight of these events is disproportionally large and stormflow sampling was designed to reflect this fact.



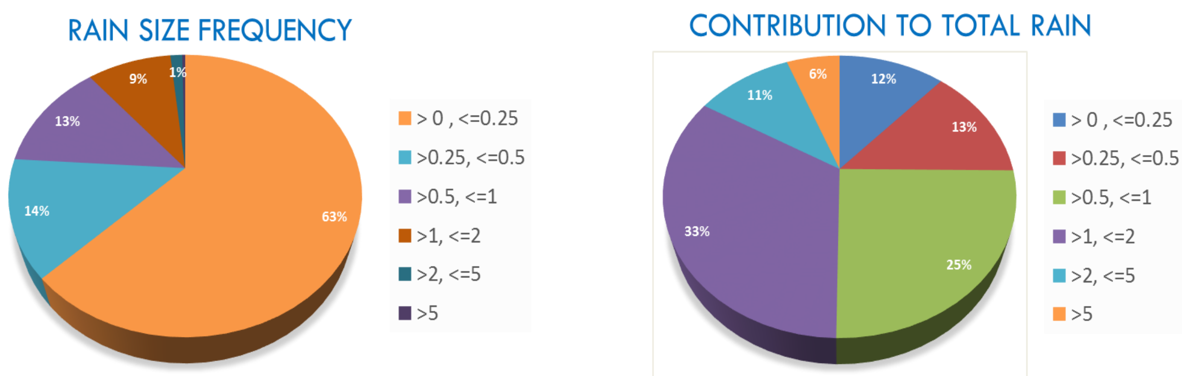


Figure 30. Average daily rainfall recorded in watersheds surrounding the streams included in the comparative analyses (**upper panel**). In the lower panel, the pie chart on the **left** shows the frequency of different rain sizes separated into six categories. The pie chart on the **right** shows the relative contribution of rain events of different sizes to the average volume of annual rainfall in the region.

Effects of Restoration on Stream Discharge

The average annual discharge in the selected streams varied from about 11 million liters to more than 1 billion liters (Figure 31). As expected, the lowest discharge values are associated with upland headwaters streams with relatively small drainage areas.

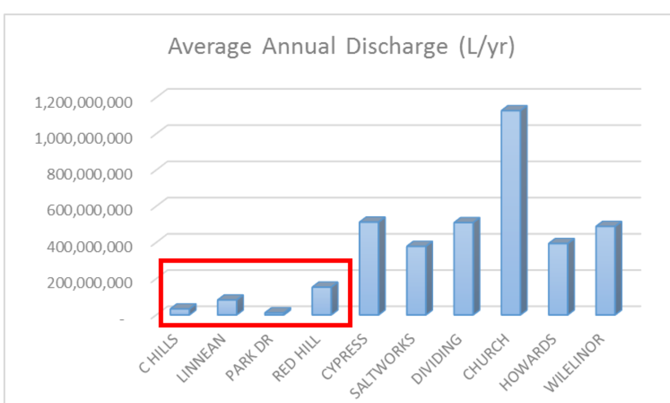


***Figure 31**. Average annual discharge (L/year) in the streams selected for the comparative analyses. Headwater streams positioned at the top of the watershed are marked with a red rectangle.*

Before restoration or in the control sites, stormflow discharge accounted for more than 60% of the annual discharge in the majority of streams. After restoration, the relative contribution of base flow to annual discharge increased in most streams and resulted in a more balanced contribution of base flow and stormflow to annual discharge (Figure 32). In 30% of the streams, base flow accounted for more than 50% of the annual flow after restoration.

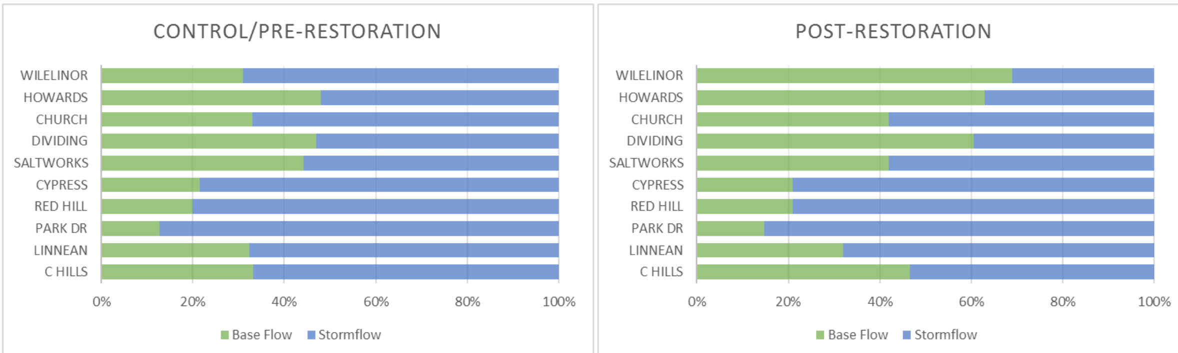


Figure 32*. Percent contribution of base flow and stormflow to annual discharge in the streams included in the comparative analyses before restoration (left) and after restoration (right).*

Comparison of Concentrations Before and After Restoration

Base Flow

The range and distribution of base flow concentrations before restoration or in control streams and after restoration are summarized in the box plots below. For TN, the range of concentrations in pre- and post-restoration ranged from detection limit levels to almost 5 mg/L. The highest concentrations were observed in Carriage Hills, Linnean and Red Hill, which are all upland headwater channels. The post-restoration data revealed substantially lower concentrations in some streams but not in others. In fact, post-restoration TN concentrations actually increased in a couple of streams (Park Dr. and Wilelinor) (Figure 33). Also, the difference in concentrations observed in Red Hill, Howard's Br., and Dividing and Church creeks were not significant. Red Hill was the only upland headwater stream restored with the NCD.

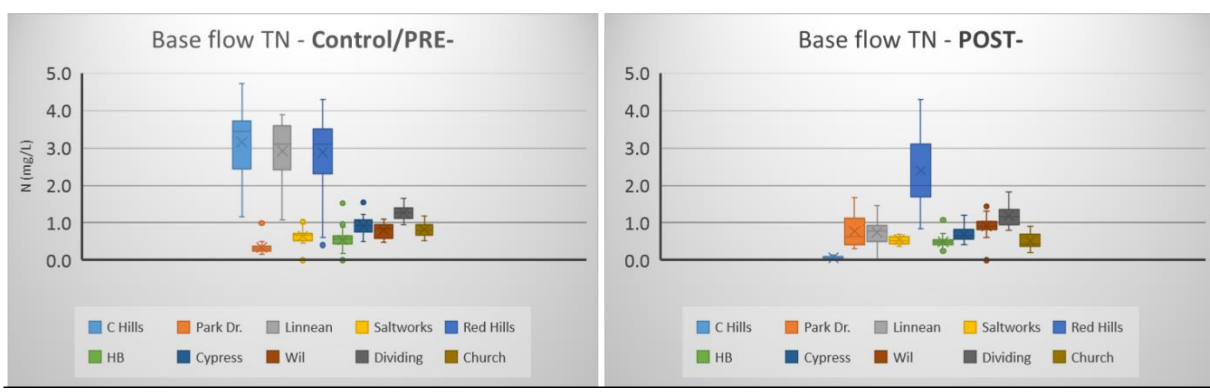


Figure 33*. Box plots of base flow TN concentrations in the streams included in the comparative analyses in Control or Pre-restoration sites (left panel) and Post-restoration sites (right panel). Carriage Hills, Park Dr., Linnean and Red Hill are headwater channels and the others are lowland channels.*

Total P concentrations were generally lower in the pre-restoration monitoring period and control streams than in the post-restoration (Figure 34). The exception was Red Hill, where the median concentration went from 0.12 mg/L before restoration to 0.07 mg/L after it. Concentrations at Dividing Cr. also decreased in the post-restoration period, but not significantly (t-test, p>0.05).

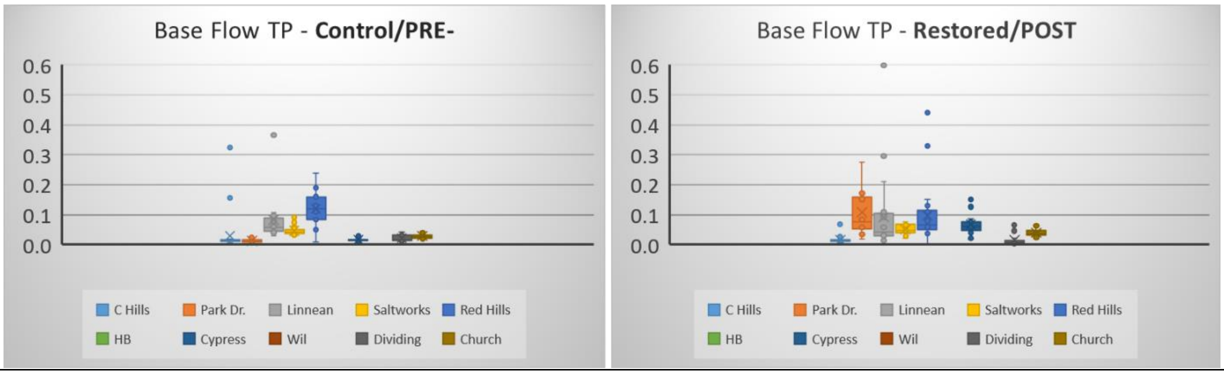


Figure 34*. Box plots of base flow TP concentrations in the streams included in the comparative analyses in Control or Pre-restoration sites (left panel) and Post-restoration sites (right panel). Carriage Hills, Park Dr., Linnean and Red Hill are headwater channels and the others are lowland channels.*

Concentrations of TSS in base flow were generally higher in the restored streams (Figure 35), with a significant difference in Park Dr., Saltworks, Cypress and Dividing creeks (**t-test**, p<0.05). The only streams where concentrations decreased significantly with restoration was Red Hill and Howards Br. (**t-test**, p<0.05).

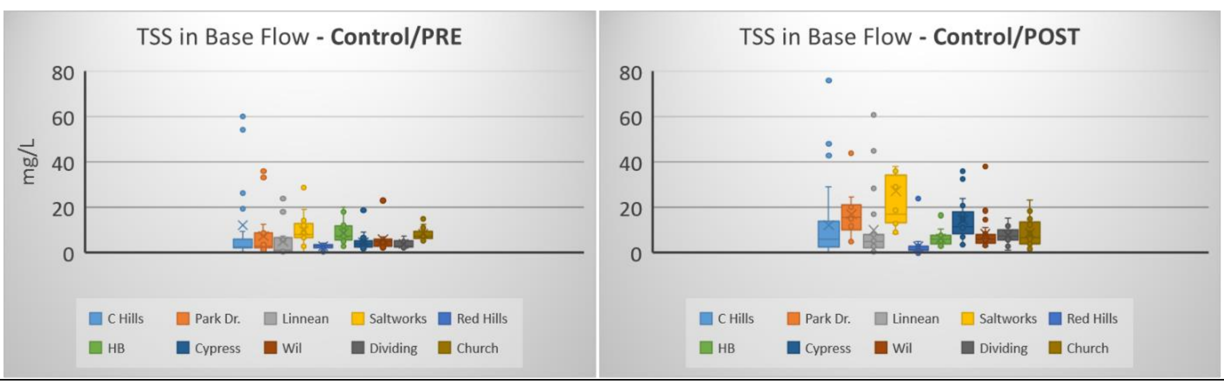


Figure 35*. Box plots of base flow TSS concentrations in the streams included in the comparative analyses in Control or Pre-restoration sites (left panel) and Post-restoration sites (right panel). Carriage Hills, Park Dr., Linnean and Red Hill are headwater channels and the others are lowland channels.*

Stormflow

Flow-weighted mean concentrations in stormflow also changed with restoration in most streams for most pollutants. For TN, the median flow-weighted mean concentrations were significantly lower in six out of the 10 restored streams analyzed (Figure 36). The exceptions included Red Hill, Park Dr., Saltworks Cr. and Wilelinor. Two of these streams are upland channels and two lowland.

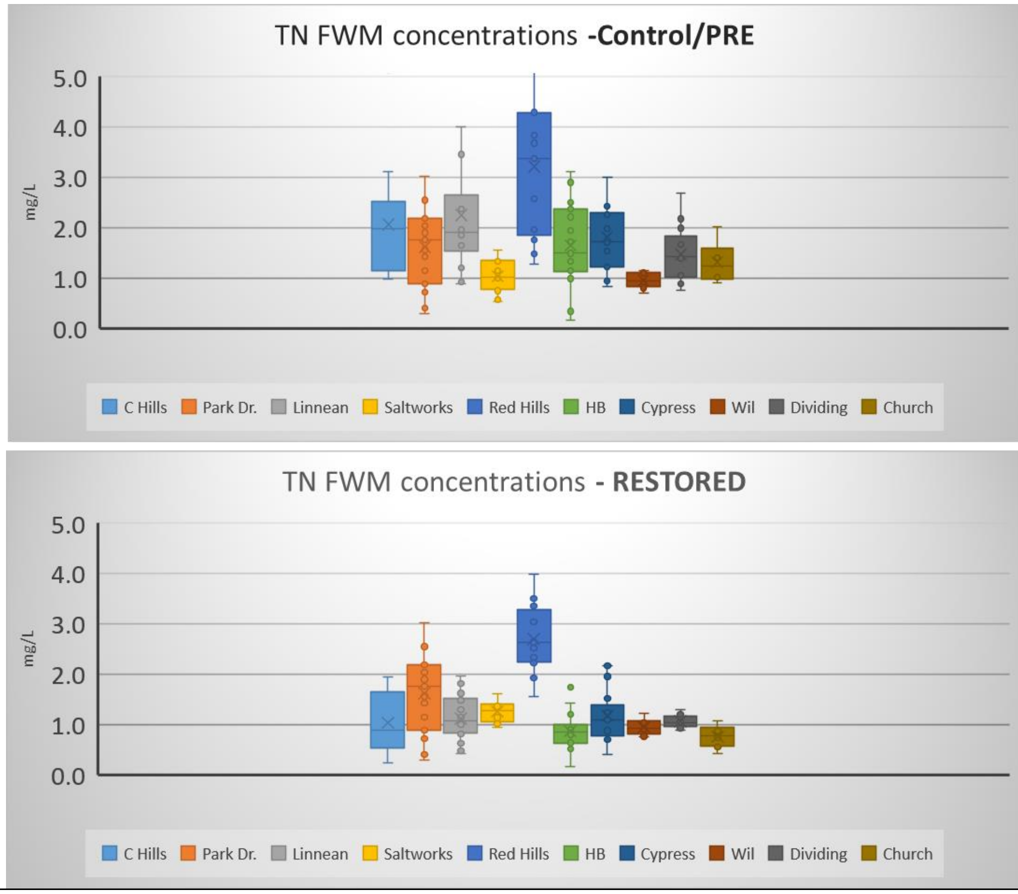


Figure 36*. Box plots of stormflow TN flow-weighted mean concentrations in the streams included in the comparative analyses in Control or Pre-restoration sites (left panel) and Post-restoration sites (right panel). C. Hills, Park Dr., Linnean and Red Hill are headwater channels and the others are lowland channels.*

For TP, flow-weighted mean concentrations in stormflows were lower after restoration in every stream except for Saltworks Cr. (Figure 37). However, concentrations were significantly lower only in the upland channels and in Dividing Creek (t-test, $P<0.05$). In Saltworks, concentrations were significantly higher after restoration.

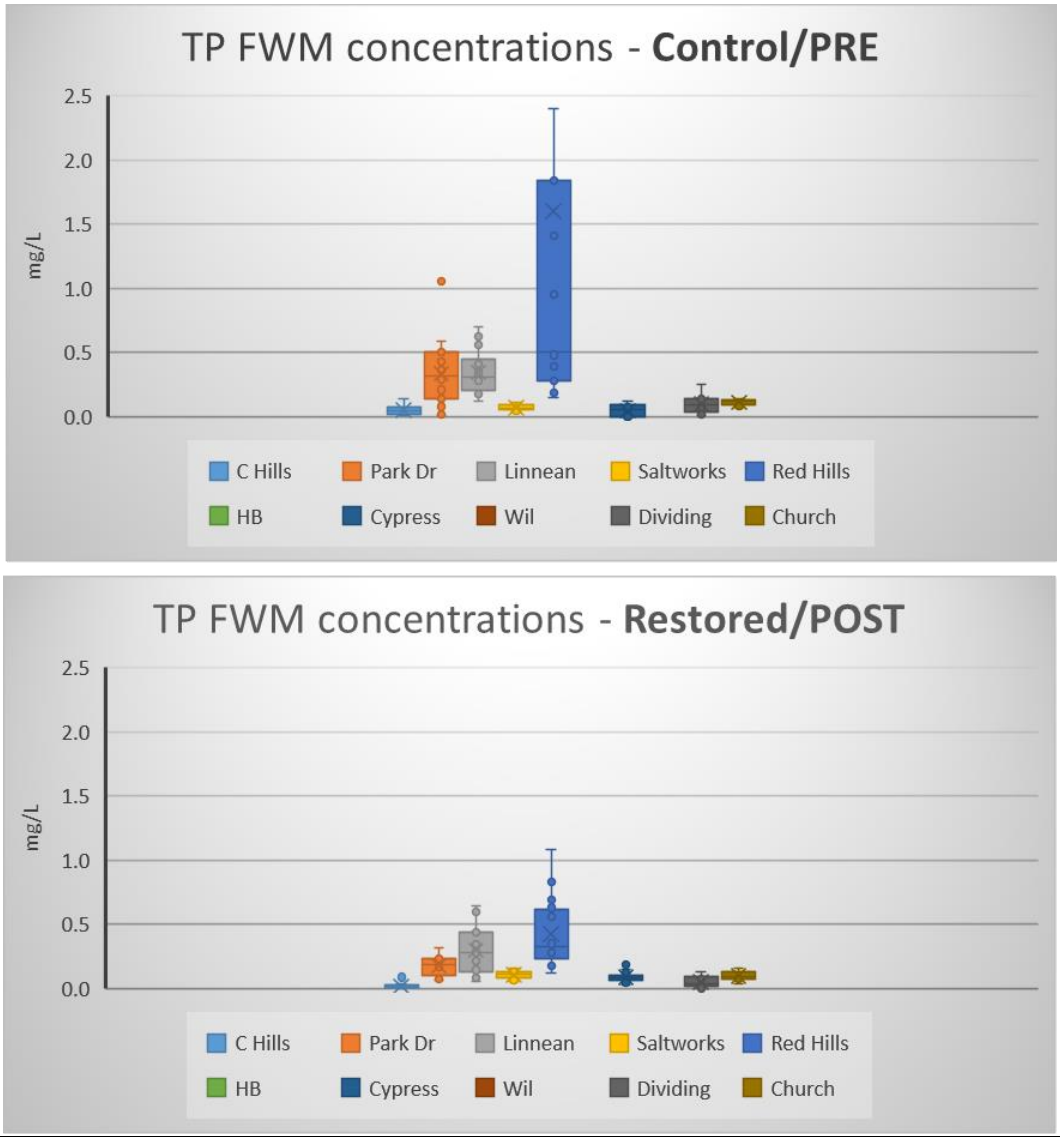


***Figure 37**. Box plots of stormflow TP flow-weighted mean concentrations in the streams included in the comparative analyses from control or Pre-restoration sites (left panel) and Post-restoration sites (right panel). Carriage Hills, Park Dr., Linnean and Red Hill are headwater channels and the others are lowland channels.*

In the case of TSS, flow-weighted mean concentrations in stormflows decreased in all channels except Saltworks and Church creeks, where they increased (Figure 38). The decreases were statistically significant only in headwater channels (i.e., Carriage Hills, Park Dr., Linnean and Red Hill) and in Howard's Branch. In the remaining lowland channels differences between pre-restoration or control stream and post-restoration TSS concentrations were not significant (t-test, $p>0.05$).

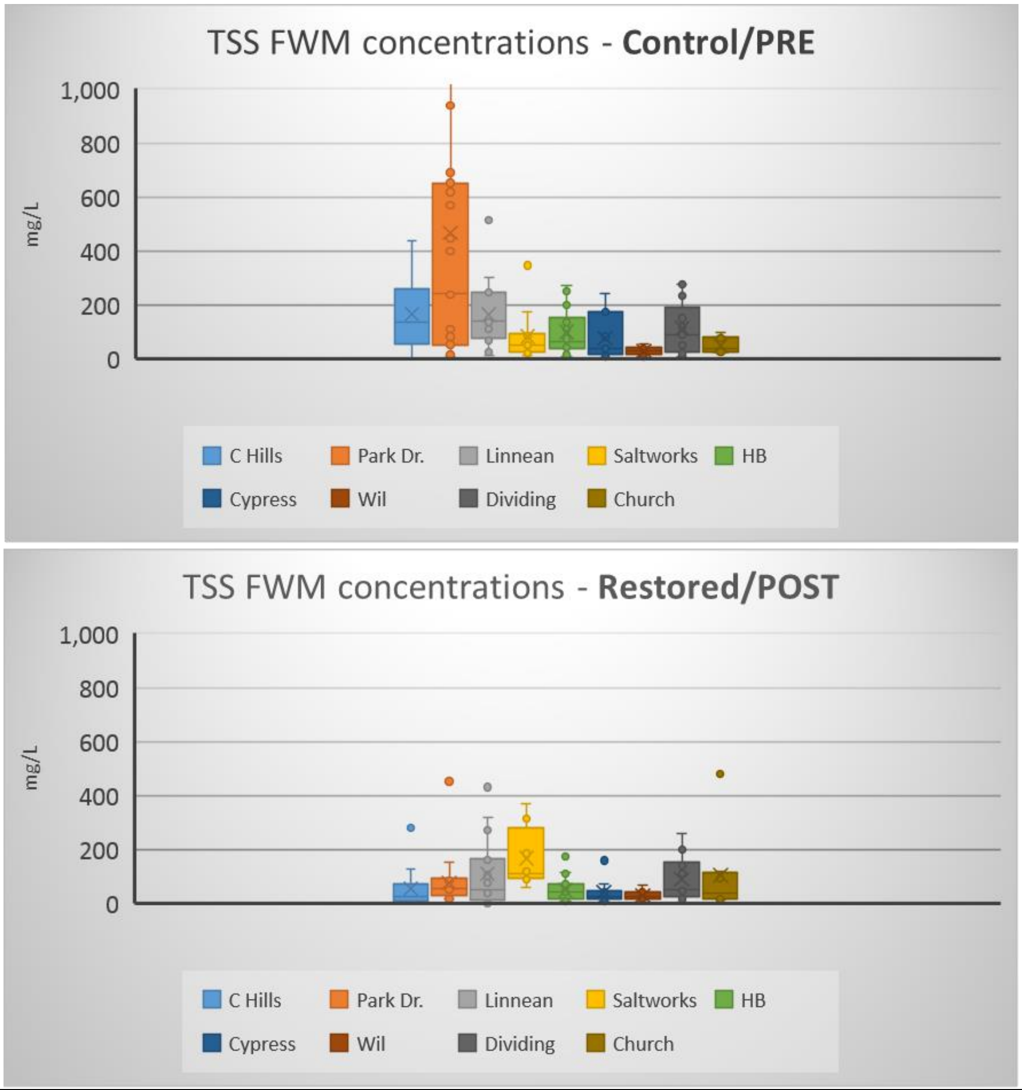


Figure 38*. Box plots of stormflow TSS flow-weighted mean concentrations in the streams included in the comparative analyses in Control or Pre-restoration sites (left panel) and Post-restoration sites (right panel). Carriage Hills, Park Dr., Linnean and Red Hill are headwater channels and the others are lowland channels.*

Comparison of Loads and Yields Before and After Restoration

The concentration and discharge data from the selected streams revealed some significant differences between the pre- and post-restoration phases. Accordingly, there were also differences in pollutant loads in restored versus unrestored or control streams. In general, loads were relatively lower after restoration than prior to it or in the control streams. In large part, the reduce loads can be attributed to lower stormflow loads in the post-restoration datasets (Figure 39).

The largest load reduction for TN was observed in Church Cr., which also had the highest total load prior to restoration. Loads of TP and TSS in Church Cr. were also relatively high and could be attributed to the fact that the channel has the largest annual discharge as it drains a relatively large watershed (Table 2). However, Red Hill drains a relatively small catchment but had loads comparable or higher than those of other lowland channels. Loads of TP and TSS were particularly high in Red Hill.

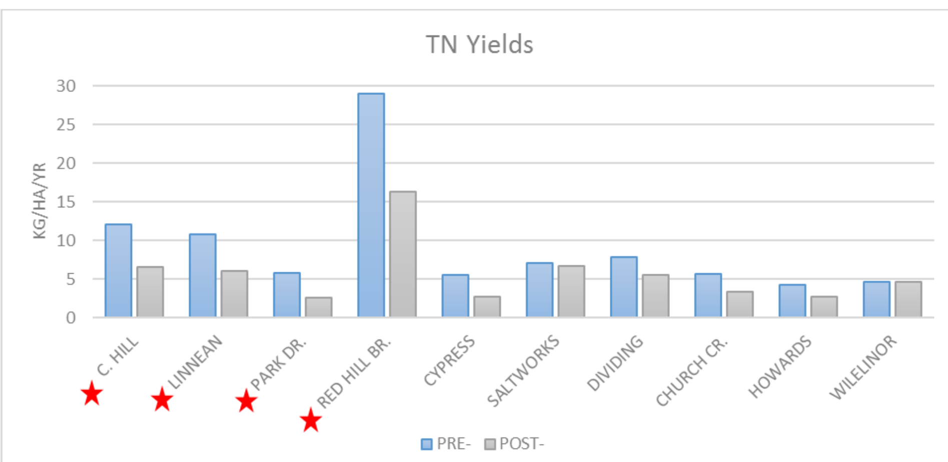
Figure 39. *Comparisons of average annual TN, TP and TSS loads during pre-restoration or in control sites (left panels) and post-restoration (right panels) in the streams included in the comparative analyses. Stream names marked with a red star indicate that the stream is an upland channel, positioned at the top of the watershed.*

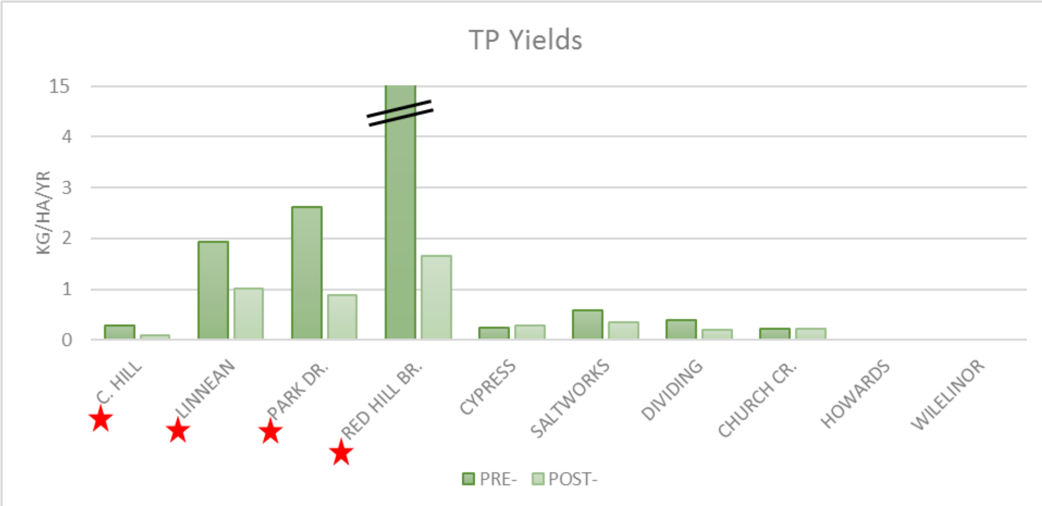
Loads normalized by the stream drainage area or stream yields provide a better comparison of nutrients and TSS loads among streams with different watershed areas.

Yields of TN in upland channels varied from about 5 kg/ha/yr in Park Dr. to nearly 30 kg/ha/yr in Red Hill Br. prior to restoration (Figure 40, upper panel). Estimates of yields for Maryland range between 1.5 kg/ha/yr to more than 12.5 kg/ha/yr, with the highest values associated with agricultural areas (CBP, click on link below). This means that the yields in Red Hill B, which were estimated using TN loading data collected by a third party, are in the high range for the region and for the entire Chesapeake Bay watershed, for this matter. In contrast, yields estimated for the other streams are within a normal range predicted for the region in Maryland (https://www.chesapeakebay.net/what/maps/delivered_yield_of_total_nitrogen_all_sources).

Except for Wilelinor, total N yields clearly decreased in all the channels with restoration. Monitoring at Wilelinor was initiated after restoration and used the upstream-downstream catchment approach, meaning that the control site was upstream of the restored reach. Therefore, the increase in TN concentrations observed might have been caused by additional sources of N in the restored reach.

In contrast to TN, TP and TSS yields decreased in the upland channels but not in all lowland channels. In some lowland channels, TP and TSS yields either increased or did not change substantially after restoration (Figure 40, lower panels), suggesting that the expected outcomes for restoration are likely to be more certain for upland than lowland channels. Furthermore, the magnitude of yield reductions in upland channels was greater than in lowland channels for both TP and TSS.





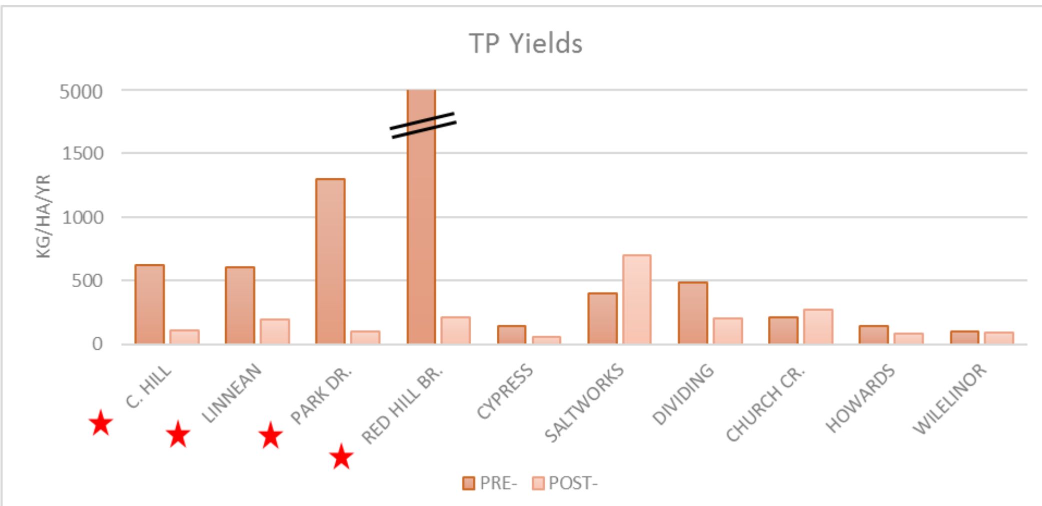


***Figure 40**. Comparisons of average annual TN, TP and TSS yields in pre-restoration or control streams and post-restoration streams. Stream names marked with a red star indicate that the channel was positioned at the top of the watershed (upland channel).*

Restoration Effectiveness and Performance Evaluation

When changes in loads are compared in relative terms such as percentage of pre-restoration or control stream loads, it becomes clear that the outcome of restoration was mostly positive, i.e., most pollutant loads decreased considerably in all streams (Figure 41). However, there was variability.

For TN, restoration seemed to have been quite effective at reducing loads, with an average of 28% and 46% lower loads in restored than unrestored lowland channels and upland channels, respectively (Figure 41, upper panel). Wilelinor had the worst performance for TN but, again, this was probably an artifact of the monitoring design, as explained before. Accordingly, TN data from Wilelinor were removed from the analyses where potential factors influencing load reductions in restored streams were examined.

For TP, some streams reduced more than 65% of the loads observed in the pre-restoration period or in control streams while others became a source of TP with restoration (Figure 41, mid panel). A similar pattern was observed for TSS, where some restored streams had nearly 100% lower loads post-restoration than in pre-restoration or control sites but other streams exported more TSS after restoration (Figure 41, bottom panel).

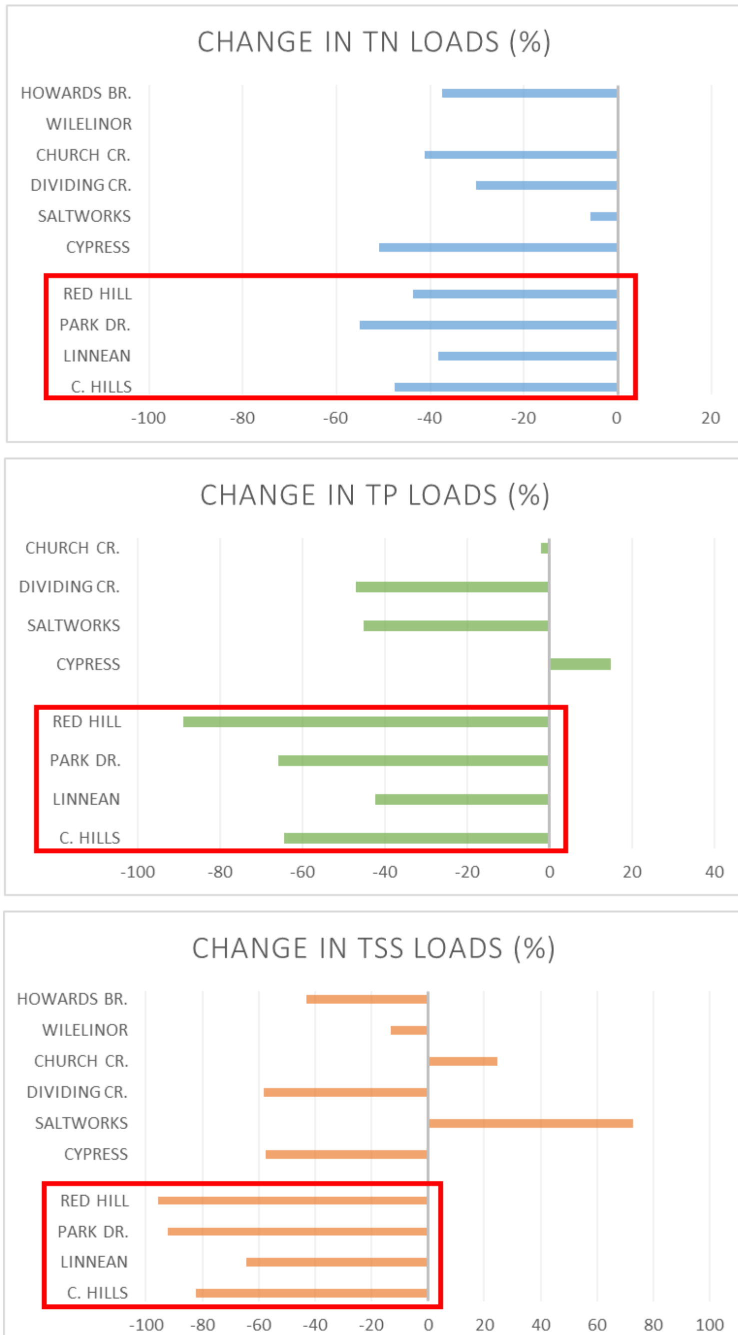


Figure 41. *Percent changes in TN, TP and TSS loads associated with stream restoration calculated as the percent difference between loads in the pre-restoration or control sites and in post-restoration. Negative values indicate a net loss while positive values indicate net gain.*

Potential Explanatory Factors for Performance Variability

Figure 41 above showed that there was variability in the effectiveness of restoration among projects but this variability was different for each pollutant. For TN, restoration effectiveness measured in terms of load reductions went from no change in loads to a 55% reduction, for TP it went from a 15% gain to 89% reduction, and for TSS from a 73% gain to a 96% reduction.

Such variability can be attributed, in part, by differences in monitoring designs, as it was probably the case for Wilelinor as explained above. The monitoring design and water sampling method used for Red Hill Br. could also have affected some of the results, especially if excessive amounts of stream bed material were collected with the water samples in the upstream site of the restored reach. The estimated loads for Red Hill Br. were relatively high, suggesting that differences in monitoring methods could have created an artifact.

Other potential influences include differences in the total rainfall volume in the monitoring period before and after restoration, differences in storm sizes sampled for stormflow before and after restoration, and external factors such as watershed characteristics. The potential influence of differences in total rainfall was minimized by standardizing the annual discharge before and after restoration in each stream to the typical amount of annual rainfall in the study region. For example, if the average annual rainfall in the post-restoration monitoring period was 20% lower than in the post-restoration period, discharge was adjusted by 20%. Likewise, differences in storm sizes sampled for stormflow before and after restoration were minimized by attributing weights to the different storms sampled according to size categories and relative contribution to total annual rainfall, as described before. Therefore, the only major factor that could have influenced the performance of the selected restored streams was watershed characteristics.

The percent load reductions estimated for the different streams were plotted against key watershed characteristics such as catchment size and percent imperviousness cover, which are easy to measure, to examine correlations. Percent load reductions were also plotted against potential indicators of performance such as magnitude of loadings and stormflow concentrations before restoration or above the restored reach.

Percent load reductions of TN, TP and TSS were negatively correlated with percent imperviousness in the watershed (Figure 42), with the correlation between TSS and imperviousness being the most significant and indicating that restoration projects draining more impervious watersheds have a relatively higher chance of being effective at reducing TSS loads. There was no significant correlation between imperviousness and TP load reductions (r=0.28).

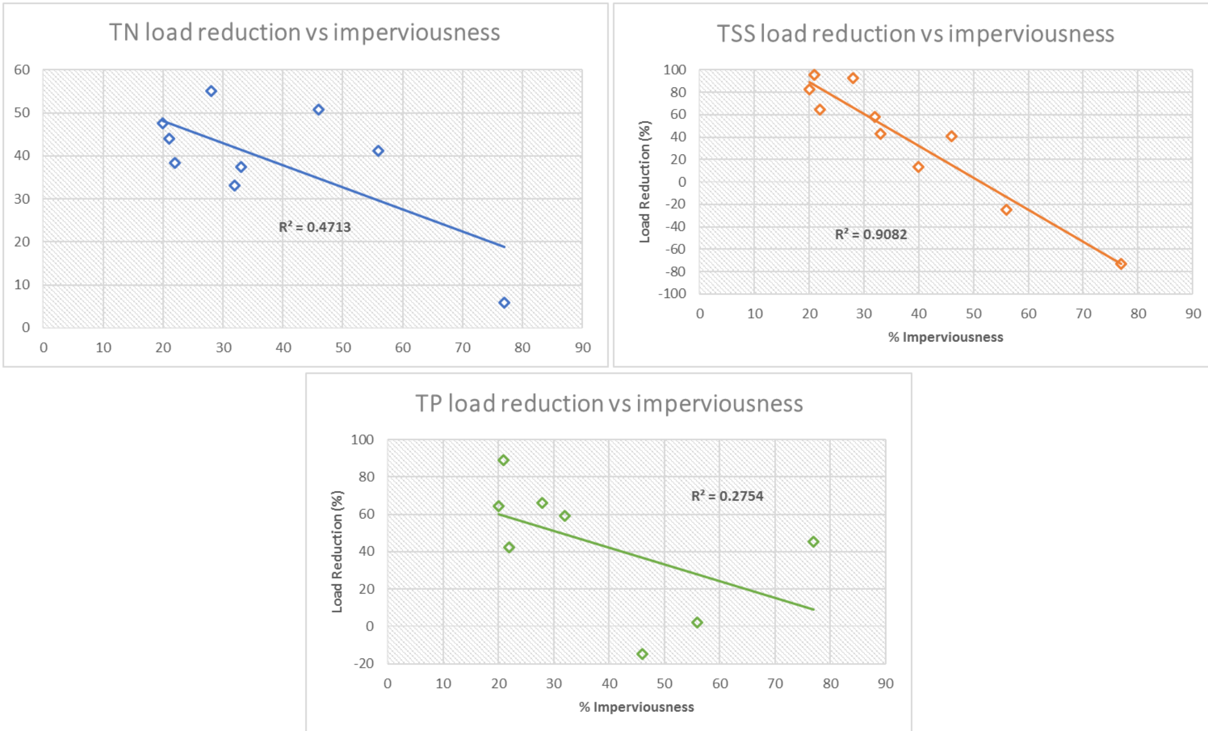


Figure 42. *Relationship between level of imperviousness in the catchment and restoration effectiveness, measured by magnitude of load reduction in the restored stream in comparison to a control site or the stream before restoration. Load reductions are expressed as percentage of initial loads. Negative values indicate an increase in loads.*

The size of the watershed was also negatively correlated with percent load reductions of TP and TSS (Figure 43), suggesting that restoration projects draining smaller watersheds are more likely to be effective at reducing loads of these pollutants. Such correlations also indicate that the position of the restored stream in the watershed and the stream gradient or slope are important factors influencing restoration effectiveness for TP and TSS loads given that the headwater channels monitored are associated with relatively small watersheds and steeper gradients than lowland channels.

It is important to mention, however, that the correlations between load reductions of TP and TSS and drainage area became more significant (i.e. r>0.50) when data from Saltworks were removed from the analyses. Saltworks performance for TSS and TP was substantially lower than expected for the size of its drainage area, suggesting there were other factors that were more influential.

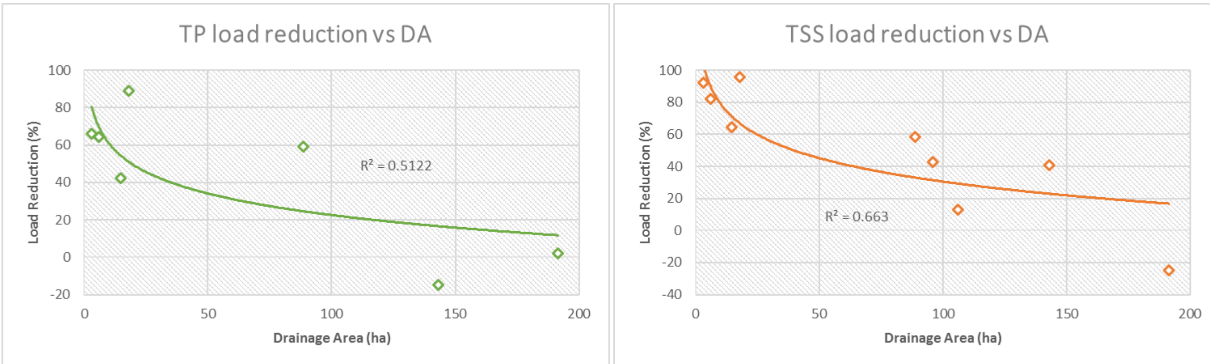


***Figure 43**. Relationship between size of the drainage area of the restored stream and restoration effectiveness, measured by magnitude of load reduction in the restored stream in comparison to a control site or the stream before restoration. Load reductions are expressed as percentage of initial loads. Negative values indicate an increase in loads.*

For TN, none of the watershed characteristics examined correlated significantly to load reduction in the selected streams. However, when the watershed characteristics examined are combined in a multiple regression, the relationship with percent reduction of TN loads is significant, suggesting that they can be used as predictors of restoration performance for TN (Table 7).

***Table 7**. Results from multiple regression analyses to predict restoration effectiveness based on multiple independent variables (imperviousness in the catchment, drainage area of catchment and stream slope).*

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       10.22385    1.16486   8.777 0.000318 ***
Imperviousness    -0.28157    0.04657  -6.047 0.001783 **
Catchment          0.04658    0.01374   3.391 0.019430 *
SlopeLow-gradient -2.69047    1.47890  -1.819 0.128535
SlopeMed-gradient  4.81216    1.07618   4.472 0.006570 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.96 on 5 degrees of freedom
Multiple R-squared:  0.9153,    Adjusted R-squared:  0.8476
F-statistic: 13.51 on 4 and 5 DF,  p-value: 0.006863
```

Other potential predictors of restoration performance, such as average annual concentration in base flow and stormflow in the restored stream prior to restoration or in the control streams were examined, but correlations were not significant either for TN. In contrast, load reductions of TP and TSS were positively correlated to stormflow concentrations (Figure 44), highlighting the importance of reducing stormflow loads in stream restoration projects to reduce the export of these pollutants.

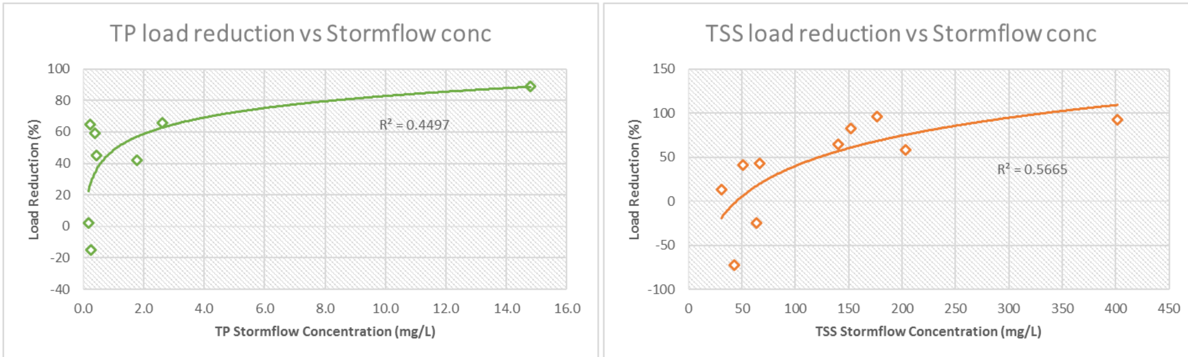


***Figure 44**. Relationship between average stormflow concentrations the pre-restoration period or control streams and restoration effectiveness, measured by magnitude of load reduction in the restored stream in comparison to a control site or the stream before restoration. Load reductions are expressed as percentage of initial loads. Negative values indicate an increase in loads.*

Conclusions

- All restoration projects but one included in the comparative analyses were effective at reducing TN concentrations, loads and yields. The exception was Wilelinor, but such different outcome was probably caused by limitations with the monitoring design used in that site.
- Restoration performance was more variable for TSS and TP, except in headwater streams.
- The magnitude of pollutant loads in headwater streams was smaller than in lowland channels, consequently, load reductions were smaller as well. When loads are normalized by drainage area, changes in loading rates in restored headwater streams were more pronounced than in lowland streams.
- Restoration performance was also influenced by the level of imperviousness in catchment, drainage area, and channel slope; climate change and higher frequency of larger storms may have a negative impact on the performance of projects based on these influential factors.
- Restoration design is probably another factor affecting performance of the streams included in the analyses, but datasets from more natural channel design are needed to properly examine this.
- The performance of lowland channels is more variable than the performance of upland channels, the data collected in the complementary monitoring period provide some evidence that lowland channels may take longer to mature after restoration and may perform better in the future than presently
- More studies on the long-term trajectories of restoration projects are needed.