Urban Water Quality: Socio-Economic Distribution of Stream Degradation, and the Influence of Climate on BMP Performance

Isabelle Horvath
Marquette University

URBAN WATER QUALITY: SOCIO-ECONOMIC DISTRIBUTION OF STREAM DEGRADATION, AND THE INFLUENCE OF CLIMATE ON BMP PERFORMANCE

by

Isabelle Horvath, M.S.

A Dissertation submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Milwaukee, Wisconsin
May 2023

ABSTRACT

URBAN WATER QUALITY: SOCIO-ECONOMIC DISTRIBUTION OF STREAM DEGRADATION, AND THE INFLUENCE OF CLIMATE ON BMP PERFORMANCE

Isabelle Horvath, M.S.

Marquette University, 2023

Urban water quality impairments have long burdened urban aquatic ecosystems in a phenomenon termed “urban stream syndrome”. The symptoms of this “syndrome” include physical changes in stream morphology and water level, biologically stressed environments, and perturbations in ecosystem processes. Despite years of research and costly investments in restoration and rehabilitation, urban waterways are still plagued by degradation. Improvement in urban water quality will require multi-faceted efforts, including progress on 2 key fronts 1) increased understanding of current impairments, and 2) increased knowledge about urban stormwater infrastructure like best management practices (BMPs). Towards both topics, engineers have quantified the influence of controls that can be manipulated in restoration and design to improve water quality, such as imperviousness, real time controls, and soil amendments in BMPs. However, urban waterways remain degraded, and there are still many unknowns regarding the distribution of stream quality impairments. Improvement of urban water quality requires better quantification of water quality distribution and quantification of factors that impact stormwater infrastructure performance**.** This dissertation provided a diagnosis of the extent of spatial variability in urban stream syndrome in an urban watershed in metropolitan Detroit, and informed stormwater infrastructure performance variabilities through the execution of three objectives. First, the distribution of stream quality variability across socio-economic groups was evaluated through a partnership with volunteer science. Second, nutrient management BMP variability was assessed on a regional climate scale, and variable performance between climates was quantified. Finally, the influence of storm characteristics on BMP nutrient management was assessed and performance under different types of storms was quantified. This research showed that high poverty areas are disproportionately burdened by poor stream water quality and identified phosphorus leaching vulnerability for BMPs in arid climates and during intense storm events.

ACKNOWLEDGMENTS

Isabelle Horvath, M.S.

My advisor Dr. Anthony Parolari imparted the assuring wisdom multiple times through my graduate school experience that graduate school is not just about a dissertation, but about my growth as a person and a scientist. Dr. Parolari ensured that at this conclusion of my graduate school experience, I am graduating not only with this dissertation body of work to show, but also immeasurable personal growth through improved research and critical thinking skills, increased awareness of myself and my goals, and a supportive network of caring and intelligent colleagues. Dr. Parolari both challenged me and gave me the space and conditions to grow, I will be eternally grateful for his mentorship.

Marquette University has provided an environment of security, growth, reflection, joy, and friendship. I am grateful for the opportunities to engage in faith and Jesuit teachings on campus through Gesu Parish. Financial security and encouragement to explore fields outside my department were also integral to my experience, both of which were supported by the Graduate Assistantship in Areas of National Need fellowship from the Department of Education. Finally, my Marquette experience would not be the same without the many people who made this journey a happy and supportive one. Thank you to the faculty who fostered engaging and challenging classrooms, and especially to Dr. Mayer and McDonald who also lent their time and wisdom towards making my dissertation a work I am extremely proud of. I am especially grateful for the support of Dr. Jeff Starke, whose thoughtful words of encouragement and belief in me fueled me though the hardest days.

Thank you also to the amazing students who welcomed me to Marquette and Milwaukee with a warm embrace and filled my years here with laughter, engaging academic discussions, and the roots of lifelong friendships. The students I have mentored, Elaina Simms, Andrew Hiestand, E Swartz, Kamila Turczewski, and Grace Kidd, can not possibly have learned as much from me as I learned from them. The evolving crew of hydrology lab students, Laine Havens, Liz Regier, Tamim Sharior, Joe Naughton, Spencer Sebo, Nate Hay, CJ Gunawardana, Caitlin Lulay, Colin Wilson, Wasif Bin Mamoon, Matt Dupasquier, Joe Branca, Greg Dieter, and Jefferson Ibhagui have not only ensured that lunchtime is never boring, but they were there for me on the birthdays, the bad days, and every day in between. I am especially grateful for Donald Ryan, my bus buddy and playlist curator for his friendship, support, and impeccable music taste. And finally, there are no words for my gratitude for Duyen Lam, my roommate, lab mate, cat-coparent, pandemic bubble, and doubtless lifelong friend. There's no one else I would rather cry with on Teams.

I am also grateful for my experience as a summer fellow with the Cooperative Institute of Great Lakes Research. This experience was made possible by funding to CIGLR through the NOAA Cooperative Agreement with the University of Michigan (NA17OAR4320152). Additional thanks to Friends of the Rouge and their volunteers for sharing their passion and data collection, without which Objective one would not have

been possible. Thank you specifically to Dr. Casey Godwin, Dr. Craig Stowe, and Sally Petrella for the window into Great Lakes research, and the power of community science. Thank you especially to Dr. Tim Maguire for exceeding every bar for mentorship and extending so much knowledge and kindness with me from spatial modeling tools to pick-me-up ice cream.

Finally, none of my growth or dissertation work would have been possible without the boundless support of my friends and family. Thank you to Erin Cuddihy and Scott Stewart for keeping me smiling, near or far, I felt your love and support always. Finally, thank you to my parents, who believed in me and supported my every academic pursuit even when it took me hundreds of miles away. Thank you for fostering my scientific curiosity, creativity, and love of learning and water, without which I would not have achieved all that I have so far in life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Human activity and environmental systems are interconnected. Over one third of Earth's surface is impacted by anthropogenic landcover alterations (Vitousek, Mooney, Lubchenco, & Melillo, 1997) and these landcover changes are connected to water quality and river ecosystem health (Allan, 2004). Specifically, urbanization is connected to water quality through the generation of non-point source pollution, which exports pollutants and elevated stormwater volumes from the landscape into surface water bodies through stormwater runoff. After flushing the urban landscape, this pollutant-laden stormwater impinges surface water bodies, causing negative effects like altered streambank morphology and perturbed biological and ecosystem processes. This harm is not one-directional. This means that not only does human activity negatively impact urban surface waters, but the damage to these waters diminish their value in human systems (called "ecosystem services") (Garcia et al., 2016). The term "urban stream syndrome" broadly defines this relationship between dense anthropogenic activity and the negative effect on stream quality and diminished ecosystem services provided by urban waterways (Booth, Roy, Smith, & Capps, 2016; Walsh et al., 2005; Withers & Jarvie, 2008).

A foundational review of urban streams imparts the closing message that "the challenge for stream ecologists in furthering our understanding of streams in urban areas is to not only better understand interactions between catchments and stream processes, but to integrate this work with social, economic, and political drivers of the urban environment" (Walsh et al., 2005). Interactions between catchments and stream processes have been broadly assessed, but less is known about social, economic, and political interactions with urban stream degradation. Particularly, there is a lack of information

regarding the distribution of stream quality degradation on people of different socio-economic demographics. In the United States, the Environmental Protection Agency (EPA) monitors spatial connections between environmental conditions and demographic indicators through the "EJscreen" platform(United States Environmental Protecion Agency, 2021). However, establishing these connections between urban water quality and socio-economic demographics is challenging because stream systems are highly spatially variable, complex connected networks. Therefore, furthering our understanding of urban stream quality degradation and the distribution of this degradation across socio-economic demographics requires tools that both supply robust spatial data, and account for the complex connectedness of stream networks. Objective 1 (Chapter 3) of this dissertation addresses both challenges to develop increased understanding of the socio-economic distribution of urban stream quality degradation in the Rouge River watershed in metropolitan Detroit.

Objective 1, titled, ***Evaluation of the socio-economic distribution of stream quality***, was a collaboration with the non-profit organization, Friends of the Rouge (FOTR), formed to explore a common research interest in socioeconomic and environmental patterns in relation to water quality. This objective addressed the aforementioned challenges, a need for robust local data and need for a tool appropriate for connected networks, by partnering with community science data collection and employing a spatial stream network (SSN) model, respectively. Together these methods enabled testing of the hypothesis that water quality degradation in metropolitan Detroit is not distributed uniformly across communities of varying poverty levels. Additionally, a model was developed and used to predict water quality under varying manipulated watershed

conditions to evaluate the relationship between poverty and predicted water quality. This work provided evidence that urban stream quality degradation in the Rouge River watershed inequitably burdens areas with lower household incomes.

The knowledge gained from Objective 1 leads to the logical next question: "What can be done to reduce the burden of stream degradation on areas inequitably affected by it?" In the case of the Rouge River watershed, and many other watersheds, nonpoint sources are managed with green stormwater infrastructure (Madden, 2010; "River Restoration," n.d.; Sonne, 2014; Tackett, 2010). Green storm water infrastructure is an umbrella term encompassing systems like bioretention, rain gardens, and bioswales, that are also called stormwater best management practices (BMPs). These infrastructure practices are unified by their design goal: to capture stormwater runoff and provide volume storage and or water quality treatment. Yet, the reliability of BMP water quality management, especially for nutrients, is variable (Clary, Jones, Leisenring, Hobson, & Strecker, 2020; Gold, Thompson, & Piehler, 2019; Lintern, McPhillips, Winfrey, Duncan, & Grady, 2020; Park, Kang, Jung, & Roesner, 2015). Nonpoint source nutrient management in urban areas is essential for preventing urban stream degradation, as elevated nutrients cause eutrophic conditions, infamous for odorous algae and unsafe swimming and fishing conditions. Therefore, lack of nutrient management reliability from BMPs leads to a disconnect between research and financial investments into stormwater BMPs, and tangible improvements in water quality from nonpoint source pollution (Lintern et al., 2020). Increased understanding of stormwater BMP nutrient management function is necessary to make these infrastructure practices more effective, and ultimately reduce urban water quality degradation.

Variable BMP nutrient management is caused by the variety of actors that influence nutrient removal and retention processes in BMPs. Compared to other engineered water quality treatment systems like a wastewater or drinking water treatment plant, green stormwater infrastructure often has little or no energy-supported processes and thus has limited control over temperature regulation, retention times (except in the case of real-time control), or employment of screening mechanisms that require backwash. Further, the physical, chemical and biological treatment processes present in BMPs (like sedimentation, ion exchange with soil, and metabolic microbial processes, respectively) are subject to the influence of the conditions of the local environment, like seasonality, long-term water balances, manifold inflow rates from variable storms and associated changes in soil moisture and oxygen properties (Davis, 2007; Horvath, Pulvermacher, & Parolari, 2022; Hunt, Davis, & Traver, 2012; Li & Davis, 2014). Many studies have examined these climate- and weather-associated conditions on BMP nutrient management at the site-level, but few studies have sought to broadly define the influences of these conditions across BMPs. Objectives 2 and 3 (Chapters 4 and 5, respectively) fill this knowledge gap by using data from the International BMP Database to define relationships between climate, storm events, and BMP nutrient management.

Objective 2, titled, ***Effects of Regional Climate and BMP Type on Stormwater Nutrient Concentrations in BMPs: A Meta-Analysis,*** evaluated the impact of climate on changes in stormwater nutrient concentrations through three types of vegetated BMPs. Meta-analysis was applied to aggregate the direction and magnitude of nutrient concentration changes induced by BMPs in different climates. It was hypothesized that BMP nutrient concentration change was different in wet and dry climates. This

hypothesis was tested by analyzing BMP influent and effluent nutrient concentrations with data from the International Stormwater BMP Database. To evaluate this hypothesis, this objective conducted a meta-analysis of data from the BMP database and evaluated changes in N and P concentrations through BMPs in wet and dry climate regimes. This objective showed that both climate and BMP type influence nutrient management. Specifically, dry climate BMPs leach dissolved inorganic and total phosphorus more consistently and at a higher magnitude than wet climate BMPs, and bioretention leaches more DIP than grass strips and swales.

Objective 3, titled, ***Storm impact on BMP nutrient treatment,*** quantified the water quality performance of BMPs under five distinct types of storms. Clustering algorithms were applied to group storms into like groups based on storm event attributes, including total precipitation depth, return period, peak hourly intensity, and antecedent dry period. It was hypothesized that BMP water quality management is different for different types of storm events. To assess this hypothesis, BMP water quality data from five BMP types were collected from the International Stormwater BMP database and matched with traits of the storm events that generated the stormwater quality data. The relationship between the individual storm traits and influent and effluent water quality was assessed through metrics of concentration, volume, and load. Then the storm traits were used to create clusters of like storms, and the changes in nutrient concentration and load between influent and effluent were assessed for different storm groups. This objective provided insight into the dominance of storm depth and return period on altering influent and effluent volumes and loads, and identified storm conditions that make BMPs more vulnerable to poor nutrient removal and retention performance. This

research provides evidence that BMPs elevate TP and DIP concentrations across storm events and identifies that TN performance is most vulnerable for large storms, DIN and TP loss are most vulnerable for arid climate storms, and DIP loss is most vulnerable for intense storms.

2. STATE OF THE SCIENCE

2.1 Urban Water Quality Degradation

Water quality degradation in surface waters is a high public interest, costly environmental challenge that remains a persistent issue despite research and restoration efforts. Water quality impairments are numerous, encompassing pollution via heavy metals, nutrients, temperature, sediment, and salinity – to name a few (Akhtar, Syakir Ishak, Bhawani, & Umar, 2021). Among this non-exhaustive list of impairments, the impact of nutrients alone was estimated to cost $2.2B in the U.S. in 2008, where costs were incurred through eutrophication impact on property value, recreation revenue loss, endangered species conservation costs, and drinking water treatment costs (Dodds et al., 2009). In addition to high fiscal costs, there are high social costs associated with degraded water quality because of public interest in recreational value and safety concerns related to fishing, swimming, and drinking from water bodies (Dodds et al., 2009; Dodds & Smith, 2016; Heisler et al., 2008). High costs of impairment and public interest motivate funding via research, restoration, and policy change towards improving water quality (Olmstead, 2009). Yet, water bodies remain impaired (Keiser, Kling, & Shapiro, 2018; Oliver et al., 2017; Stets et al., 2020).

Urban areas are particularly susceptible to water quality degradation, a phenomenon driven by dense anthropogenic activity and labeled with the term “urban stream syndrome” (Walsh et al., 2005). Symptoms of this “syndrome” in urban streams include high nutrient loading (Grimm et al., 2005; Meyer, Paul, & Taulbee, 2005; Wahl, McKellar, & Williams, 1997; Withers & Jarvie, 2008), biochemical oxygen demand loading (Mallin, Johnson, Ensign, & MacPherson, 2006), highly variable flows

(Blaszczak, Delesantro, Urban, Doyle, & Bernhardt, 2019), and highly variable temperature profiles (Walsh et al., 2005), contributing to hypoxia and other damaging impacts that alter urban stream function (Meyer et al., 2005). These symptoms demand human attention because of the diminished ecosystem services resulting from urban stream syndrome (Booth et al., 2016; Withers & Jarvie, 2008).

Impairment is so pervasive in urban watersheds that biogeochemical processes in urban watersheds are distinct from non-urban watersheds (Hobbie et al., 2017; Kaye, Groffman, Grimm, Baker, & Pouyat, 2006; Van Meter, Chowdhury, Byrnes, & Basu, 2019). Specifically, urban streams are impacted by dissolved pollutants, especially orthophosphate, nitrate, sodium, and chloride (Manning, Rosemond, Benstead, Bumpers, & Kominoski, 2020; Stets et al., 2020). These urban-sourced pollutants can have severe environmental impacts within and beyond urban centers, like impacts on drinking water security and treatment (Steffen et al., 2017), and pollutant spread from stormwater reaching 100 km^2 into coastal waters (Ahn et al., 2005). The culmination of these environmental impacts from anthropogenically altered urban biogeochemistry requires management practices to mitigate the impacts of human perturbations in the urban environment. These disturbances to urban biogeochemical processes can have important implications not only on the environment, but on people as well, as historically, environmental disturbances are inequitably distributed on socioeconomically and racially repressed people (Bullard, 1996). Management of urban pollutants is particularly challenging because they do not have clear, direct sources; in other words, they are from non-point pollution (Meyer et al., 2005; Walsh et al., 2005).

Nonpoint sources of pollutants are a challenge to manage both from a physical and regulatory perspective. Nonpoint sources include pollution from sources that have no clear direct origin, and do not enter water bodies through an identifiable discharge point, like a pipe, making regulation complex under the Clean Water Act (33 U.S.C. §1251). In general, non-point sources are not directly regulated by the Clean Water Act, but urban stormwater is often regulated through a municipal separate sewer system (MS4) permit if stormwater is conveyed separately from sewage. Nonpoint sources of pollution are as difficult to treat as they are to regulate, as the dispersed nature of their transport makes conventional water treatment difficult. In some regions, stormwater is conveyed to treatment plants through combined sewers, leading to overflows of sewage and stormwater called combined sewer overflows (CSOs). Where MS4 permits are required, stormwater can be treated with decentralized treatment methods like green stormwater infrastructure (also called stormwater best management practices, or BMPs). However, green stormwater infrastructure water quality treatment performance is complex and subject to interactions from many design factors and outside stressors (Jalali & Rabotyagov, 2020; Vogel et al., 2015). The ecological, and decentralized nature of green storm water infrastructure means that the infrastructure practices are subject to the influence of many complex and interacting drivers like vegetation health, maintenance (or lack thereof), biological suitability, variable influent composition, and climate (Lintern et al., 2020; Valenca et al., 2021).

If surface water pollution is a disease (as the aptly coined “urban stream syndrome” implies), monitoring urban stream conditions and green stormwater infrastructure performance are two useful “diagnostics” to identify sources and severity

of the syndrome. The subsequent literature review sections present the state of the science in urban stream syndrome and green stormwater infrastructure water quality performance and identify needs for 1) improved methods of quantifying the distribution of stream quality degradation burdens on different socio-economic groups, 2) quantification of regional climate impact on BMP water quality treatment, and 3) quantification of storm characteristic impact on BMP water quality treatment.

2.2 Urban Stream Syndrome: Distribution of Burden

Urban streams provide valuable ecosystem and social services. As of 2005, the top five motivations for urban stream restoration projects were 1) water quality improvements, 2) riparian zone management, 3) improved in-stream habitat, 4) fish passage, and 5) bank stabilization (Bernhardt et al., 2005). Value of urban streams is also driven by social value of urban streams, which can provide opportunities for personal connections to nature in an otherwise developed landscape, and serve as locations for environmental education and recreation (Bernhardt et al., 2005; M. A. Kenney, Wilcock, Hobbs, Flores, & Martínez, 2012). It is also notable that local community interests regarding urban streams and rivers may vary based on resident location in the watershed and income. In Portland Oregon, a positive correlation was found between community involvement, higher income, and higher perceived flood risk in lower streams, whereas in the headwaters, streams were perceived as being valuable for property protection (Hong & Chang, 2020). Thus, urban streams offer both environmental and social values, and while there is community interest in urban streams, this interest likely varies among members of the community.

Just as human interest in streams in diverse, urban streams themselves are also diverse, a reality sometimes overlooked by the ubiquity implied by the term “urban stream syndrome.” Fundamental groundwork popularizing urban stream syndrome outlined symptoms of urban stream syndrome and postulated causes of these symptoms, but may overgeneralize the similarities between urban streams (Booth et al., 2016; Walsh et al., 2005). The urban stream syndrome framework helped to diagnose a broad universal relationship between increasingly degraded stream conditions and increasing impact from urbanization. This broad relationship, while helpful in communicating the importance of urbanization on stream quality, overgeneralizes the diversity of conditions amongst urban streams (Booth et al., 2016; Walsh et al., 2016). Urban streams vary widely from one another because of differences in climate (Hale, Scoggins, Smucker, & Suchy, 2016), infrastructure (Parr, Smucker, Bentsen, & Neale, 2016), and varying funding priorities and social values (Booth et al., 2016; Capps, Bentsen, & Ramírez, 2016). One underexplored local context for urban stream quality is the distribution of water quality impairments on watershed residents. There is a lack of information about the relationship between socio-economic traits and water quality impairments in different communities. Understanding disproportionate environmental burdens on the community scale is an essential component to understanding the extent and impact of urban stream syndrome.

Socio-economic factors and environmental conditions are often linked, and urban stream conditions are likely to follow a similar relationship. The study of relationships between socio-economic and environmental characteristics is referred to as “environmental justice”, and this field aims to ensure that environmental burdens are distributed on community members justly (United States Environmental Protecion

Agency, 2021). Environmental burdens such as degraded air quality (Anderson, Kissel, Field, & Mach, 2018; Miranda, Edwards, Keating, & Paul, 2011), harmful chemical exposures (Bevc, Marshall, & Picou, 2007), inequitable land use zoning, and environmental law and regulation enforcement (Bullard, 1996) have all been shown to disproportionately burden racially and economically repressed people. In water-related environmental concerns, injustices against racially and economically repressed communities are present in flood risk maps (Maantay & Maroko, 2009; Meenar, Fromuth, & Soro, 2018), distribution of green infrastructure services (Taguchi et al., 2020; Wolch, Byrne, & Newell, 2014), drinking water infrastructure risks (VanDerslice, 2011), and urban stream degradation (Sanchez et al., 2014).

In urban streams, a series of model developments identified spatial relationships between socio-economic parameters and stream quality. Spatial clustering and spatial correlations between modeled stream health indicators and socio-economic measures in the Saginaw River watershed in Michigan initially revealed mixed correlations, with the strongest correlation existing between a stream biological index and household size, at 0.18 (Sanchez et al., 2014). Other known relationships between socio-economic parameters and stream health were correlations of 0.15 between poor water quality and diverse communities, and between good water quality and higher education (Sanchez et al., 2014). However, these correlations have low magnitudes, which may be a product of the way spatially aggregated zones were formed in analysis, or from geographic scale (Sanchez et al., 2014). Successive models expanded on the relationships identified by Sanchez et al (2014) by improving the predictability between socio-economic factors and environmental drivers with spatial clustering via regression and confirmatory factor

analysis (Sanchez et al., 2015). Further model improvement was achieved by limiting the level of spatial dependency to smaller scales of socio-economic data like census tracts and block groups rather than county, resulting in observing a highest correlation of -0.48 between a biological water stream quality index and single to multi-unit housing ratio (Daneshvar et al., 2016). Finally, more robust and consistent relationships between socio-economic variables and stream quality were found by applying a two-phase approach to variable selection in environmental justice modeling, resulting in the highest correlation of 0.59 between stream quality and the female-led household metric (Daneshvar, Nejadhashemi, Zhang, & Herman, 2018).

These methodological improvements in environmental justice modeling in urban stream networks have contributed toward the identification of strong relationships and addressed issues of spatial complexity with various modeling tools, but there are barriers to repeating these methods in other watersheds. The models in the Saginaw River watershed were built using stream quality data from fish and macroinvertebrate surveys collected by the state of Michigan (Daneshvar et al., 2016; Sanchez et al., 2015, 2014). In watersheds without state-supported observational datasets, collecting stream quality data with enough spatial coverage for effective modeling is a major barrier in adopting these environmental justice models. A solution to this data paucity challenge is the partnership of advancing methods in stream modeling with volunteer science data collection (also called “citizen science”).

2.2.1 Spatial Stream Networks

Streams are complex networks, connected both via overland spatial proximity and along flow paths. These spatial relationships mean that straightforward regression is

insufficient in modeling streams because stream spatial connectedness violates the assumptions of independence between data points in a regression. Spatial stream network (SSN) modeling offers a solution to this modeling challenge by directly encompassing different possible spatial relationships into models. SSNs offer a framework where flow connected and flow unconnected streams are fit to a single model, thus allowing for hypothesis testing with watershed characteristics as predictor variables (Isaak et al., 2014; Peterson & Ver Hoef, 2014; Peterson et al., 2013; Ver Hoef, Peterson, Clifford, & Shah, 2014). SSN and associated modeling tools have been applied to a range of stream modelling applications like surface water isotope variations (McGill, Steel, Brooks, Edwards, & Fullerton, 2020), fish genetic diversity in southern France (Paz-Vinas et al., 2018), and fecal contamination in streams in Northeast Scotland (Neill et al., 2018) and central North Carolina (Holcomb, Messier, Serre, Rowny, & Stewart, 2018). Spatial stream network methods have been previously applied with citizen science data (Kielstra, Chau, & Richardson, 2019) and macroinvertebrates in streams (Frieden, Peterson, Angus Webb, & Negus, 2014; Pond, Krock, Cruz, & Ettema, 2017). Thus, SSN models have widespread use forming network-cognizant modeling structures and have a precedent of being applied with both volunteer science and bioindicator data. However, this type of modeling requires a high volume of data with widespread coverage of the stream network. Volunteer science is one potential means for collecting the robust datasets SSNs require.

2.2.2 Volunteer Science

Volunteer science involves the participation of volunteer community members in the collection of data in their local community. Other names for this community

engagement and data collection process have slightly different implications of intentions. "Citizen science" refers to the same process of community members contributing to a dataset, but implies that data collectors must be citizens, an unrelated qualification for participation (Cooper et al., 2021). The terms "community science," and "community monitoring" may sound similar to this data collection process but are used to describe a slightly different scientific process, where a hypothesis or research motive originates directly from the community and the process of developing and answering research questions is driven primarily by the community. Alternately, "volunteer science" involves the participation of community members in data collection towards answering a question that they may not have formed themselves, but rather originated from the priority of an "outside" group like a not-for-profit group, conservation organization or research group.

Benefits of collaborative work with volunteer scientists include: development of spatially and temporally robust datasets, contribution of local knowledge from residents, enhanced community relations by collaboratively addressing research questions that align with the goals of volunteer scientists, and additional co-benefits of community engagement and education (Buytaert et al., 2014; Jollymore, Haines, Satterfield, & Johnson, 2017; Krabbenhoft & Kashian, 2020; Njue et al., 2019; Taylor et al., 2021). In addition to these benefits, incorporating community collaboration and community data when studying environmental justice topics makes the science more accessible to community members with local knowledge. Their contribution and local expertise can shape the analysis – a key gap missing from environmental justice work performed without direct connection to the community being studied (Lee, 2020; Mah, 2017).

2.2.3 Volunteer Science & Bioindicators

Volunteer science has a long-standing role in aquatic science research through bioindicator monitoring. Bioindicator monitoring burgeoned in the 1980s, and was consequently adopted in state and federal aquatic quality monitoring programs like EPA's Rapid Bioassessment Protocol (Barbour, Gerritsen, Snyder, & Stribling, 1999; Barbour, Stribling, & Verdonschot, 2006). These monitoring programs typically involve the observation and collection of benthic macroinvertebrate species as indicators of aquatic quality. The frequency and diversity of observed species tell the story of the quality of streams through the known tolerances of various benthic macroinvertebrates to pollution in their habitat. Bioindicators have been employed in environmental impact studies in the Laurentian Great Lakes region (Burlakova et al., 2018) and globally (Bae, Kil, & Bae, 2005; Del Arco, Ferreira, & Graca, 2012; Graham & Taylor, 2018; Patang, Soegianto, & Hariyanto, 2018).

The pervasiveness and reliability of bioindicator proxies makes them a fitting metric for volunteer science monitoring. Volunteer scientists can be effectively trained to collect these organisms and benthic macroinvertebrates act as biological indicators of stream health (Bae et al., 2005; M. Kenney, Sutton-Grier, Smith, & Gresens, 2010). Volunteer science-collected bioindicator observations have been shown to uphold reliability as a means of aquatic quality assessment (Del Arco et al., 2012; M. Kenney et al., 2010; Vitousek et al., 1997; Walsh et al., 2005; Walsh, Sharpe, Breen, & Sonneman, 2001). The precedent of bioindicator monitoring among volunteer scientist groups has resulted in the development of indices of aquatic quality composed of weighted scores of

species frequency and diversity (Firehock, K. and West, 1995). These indices are accepted as reliable indicators of aquatic conditions (Engel & Voshell, 2002).

Specific benefits of benthic macroinvertebrate indicator data are that they provide long-term insight into stream quality, are reliably procured by citizen scientists, and have a precedent for use in socio-economic stream quality modeling. Presence or absence of sensitive bioindicator taxa reflect long-term stream conditions, rather than the "snapshot" conditions shown by grab samples and chemical analysis (Infante, David Allan, Linke, & Norris, 2009; Lenat, 1988). Further, citizen scientists can be effectively trained to collect these organisms and benthic macroinvertebrates act as biological indicators of stream health (Bae et al., 2005; M. Kenney et al., 2010). Finally, the use of bioindicators as a metric for stream quality degradation has a precedent in environmental justice water quality modeling (Daneshvar et al., 2018, 2016; Sanchez et al., 2015, 2014).

Towards the goal of increasing understanding of the nuanced relationships between socio-economic and environmental parameters, a combination of spatial modeling methods and volunteer science bioindicator monitoring offers a feasible approach for the expansion of environmental justice model development in new cities.

2.3 Green Stormwater Infrastructure Water Quality Performance

BMPs have become ubiquitous stormwater infrastructure tools, yet performance reviews continuously show mixed or failing water quality treatment performance (Carey et al., 2013; Clary et al., 2020; Gold et al., 2019; Lintern et al., 2020; Park et al., 2015). BMPs are designed to alleviate flood risk and provide water quality treatment through biogeochemical retention mechanisms to reduce the impact of non-point source pollution. While BMP water quantity management is generally acceptable, water quality treatment

has not seen the same broad success. For example, improvement of nutrient concentrations in urban BMPs has variable or uncertain impact on the watershed scale (Barrett, 2008; Dietz, 2007; Gold et al., 2019; Jefferson et al., 2017; Walsh, Imberger, Burns, Fletcher, & Bos, 2022).

The lack of performance reliability and subsequent lack of impact on urban water quality, are due to the many unknowns that persist in BMP function (Lintern et al., 2020). These unknowns include, but are not limited to: maintenance (Erickson, Taguchi, & Gulliver, 2018), distribution within the watershed (Hung, Harman, Hobbs, & Sivapalan, 2020), design controls (Costello, Hartung, Stoll, & Jefferson, 2020; Davis, Shokouhian, Sharma, & Minami, 2006), local landscape (Krimsky, Lusk, Abeels, & Seals, 2021; Simpson, Winston, & Brooker, 2022), lag times between implementation and impact (Meals, Dressing, & Davenport, 2010), and climate (Horvath et al., 2022; Mullins et al., 2020). Each of these unknowns has been shown to have some impact on BMP water quality performance at the site level. Generalizing the broad impact of these factors remains difficult because understanding of these individual and interrelated climate and design factors on BMP performance must be studied (Lintern et al., 2020). Further, each controlling factor includes multiple biotic and abiotic processes (Hsieh, Davis, & Needelman, 2007; Li & Davis, 2016). For example, both BMP design and climate influence BMP N retention because climate impacts soil moisture content, which drives biological removal, but this biological removal is also aided by design of a shallow saturated zone and N can be removed by other design factors like high vegetation presence (Valenca et al., 2021).

Estimating water quality treatment performance is further complicated by the unique biogeochemical processes that occur in BMPs to retain the breadth of pollutants entering BMPs in stormwater. Generally, particulate contaminants are well removed through sedimentation and filtration, and dissolved contaminants are either leached or retained via adsorption, precipitation, ion exchange, or biological processes (LeFevre et al., 2015; Y. Liu et al., 2017). The type of stormwater contaminant also influences removal efficiencies, where metals are well retained, and nutrients are unreliably removed depending on the nutrient and its form (LeFevre et al., 2015). Varying removal pathways for particulate and dissolved N and P lead to a divergence in removal efficiency. For example, particulate P may be removed by sedimentation and filtration, but dissolved P requires sorption or plant uptake for removal (Hunt et al., 2012; LeFevre et al., 2015). Additionally, effluent P speciation may also depend on shifts between dissolved organic and dissolved inorganic forms (Liu & Davis, 2014; Yan et al., 2016). Similarly, particulate N may be removed through sedimentation and filtration, but dissolved N removal is driven by microbially-mediated nitrification and denitrification (Hsieh et al., 2007; LeFevre et al., 2015; Li & Davis, 2014).

Understanding the varying removal pathways for target pollutant species is key to designing infrastructure that uses these pathways to optimize removal. For example, dissolved phosphorus (P) is well removed by sorption, so amendment materials like sorptive metals added to soil media can be used in BMP engineered soil media to target dissolved P removal (Erickson, Gulliver, & Weiss, 2012; Erickson, Weiss, & Gulliver, 2007; Marvin, Passeport, & Drake, 2020; Yan et al., 2016). On the other hand, dissolved N as nitrate is removed well by denitrification, so nitrate removal can be targeted with

BMPs designed with deep internal water storage and selection of engineering soil media with low infiltration rates (Hunt et al., 2012). Therefore, studying stormwater nutrient speciation can assist in understanding BMP function by revealing potential nutrient removal pathways.

While BMPs can be designed to target pollutants with certain removal pathways, these removal pathways are susceptible to perturbations made by the local environment of the BMP, namely, the climate and weather of the BMP's local environment. Climate impact on BMP water quality performance has been assessed by studying metrics of vastly different scales like seasonality of temperature and soil moisture, seasonal weather conditions like freezing and snowmelt, and storm-specific metrics like antecedent dryness, and storm size (Buffam, Mitchell, & Durtsche, 2016; Horvath et al., 2022; Hoss, Fischbach, & Molina-Perez, 2016; Roseen, Robert et al., 2009; Smith et al., 2020; Sohn, Kim, Li, & Brown, 2019; Valenca et al., 2021). Each of these metrics exist at different temporal scales, ranging from daily changes in local weather, to seasonal changes, to broad assessments of regional climate. Observations from these studies provide evidence that changes in storm characteristics or climate at each of these scales contributes to alterations in BMP performance. Beyond the inherent importance of understanding the nuances of these climate metrics in contributing to water quality treatment of BMPs, the context of climate change makes understanding these drivers even more important as the water quality performance of BMPs in the face of an altered climate (and associated storm conditions) is critical (Pyke et al., 2011; Xie, Chen, & Shen, 2015).

2.3.1 Regional Climate

In natural soil systems, regional climate impacts how nutrients are processed and retained. In the context of this dissertation work, regional climate is defined as the ratio of decades-long averages of potential evapotranspiration to precipitation, or the Budyko aridity index (Yin, Calabrese, Daly, & Porporato, 2019). The influence of regional climate on biogeochemical processing has been assessed in natural soils by comparing nutrient availability in soils along precipitation gradients. These climate assessments showed that nutrient retention and processing of similar soils varies along precipitation gradients (Aranibar et al., 2004; Austin & Vitousek, 1998; Austin et al., 2004; Feyissa et al., 2021; Hou et al., 2018). Both N (Welter, Fisher, & Grimm, 2005; Yahdjian & Sala, 2010) and P (Buckingham, Neff, Titiz-Maybach, & Reynolds, 2010; Hou et al., 2018; Ippolito et al., 2010) availability vary with regional climate-related variables like annual precipitation patterns, temperature, soil characteristics, and vegetation.

In arid climates, both N and P tend to be more available in more arid soils (Austin & Vitousek, 1998; Austin et al., 2004; Emadi, Baghernejad, Bahmanyar, & Morovvat, 2012; Ippolito et al., 2010; McCulley, Burke, & Lauenroth, 2009). The high availability and subsequent leaching of N in arid climates is attributed to the rapid conversion of organic N to inorganic N in arid conditions, a form which can be converted to N-oxide via denitrification, and lost from soil to the atmosphere (Aranibar et al., 2004). The leaching of N was also attributed to the accumulation of N in soils between storm events, and precipitation-induced leaching (Austin et al., 2004). The causes of P leaching from soils are more complicated, where P availability is controlled differently by temperature and precipitation (Hou et al., 2018). Generally, high temperatures are associated with less

P availability due to weathering loss and plant demand (Dixon, Chadwick, & Vitousek, 2016; Hou et al., 2018). Precipitation affects P differently depending on species, where higher precipitation is associated with decreasing biologically available P and mineral P, but higher organic P (Hou et al., 2018). P availability is also associated with the presence of different P-sorbing compounds in soil and biological processing, both of which can vary with changing climates (Buckingham et al., 2010; Emadi et al., 2012; Hou et al., 2018; Ippolito et al., 2010). This climate-driven variability in nutrient cycling and retention in native soils provides evidence that a similar pattern may exist in BMPs that rely on soil ecosystems for nutrient treatment.

The same processes that drive N and P availability in natural soils exist in BMPs. N has been shown to leach following dry periods in a stormwater wetland (Horvath et al., 2022) and N transformations via nitrification and denitrification have been observed in infiltration basins (L. McPhillips & Walter, 2015; Morse, McPhillips, Shapleigh, & Walter, 2017). P availability in BMPs is influenced by soil media composition, where the presence of compost elevates available P concentrations (Hurley, Shrestha, & Cording, 2017). Further, adsorption of P by P-binding materials is a promising control being explored for enhanced P retention, a process that parallels the P absorption occurring in natural soils (Erickson et al., 2012; Marvin et al., 2020). Thus, there is evidence of processes that vary with climate existing in BMPs, providing a logical foundation to hypothesize that climate influences the availability of nutrients in BMPs.

Site-level observations provide evidence of the influence of climate conditions on the individual water quality performance of BMPs. Specifically, BMPs effectiveness has been attributed to short-term weather variability, storm intensity, antecedent dryness, and

seasonality (Buffam et al., 2016; Chaubey, Chiang, Gitau, & Mohamed, 2010; Horvath et al., 2022; Sohn et al., 2019). A Soil and Water Assessment Tool (SWAT) model forecasting 25 years of BMP implementation in a pasture-dominated watershed indicated that weather can cause BMP nutrient retention performance to vary significantly (Chaubey et al., 2010). Another study showed elevated mobile nitrate in soils at an urban constructed stormwater wetland following a long inter-storm period (Horvath et al., 2022). Longer term climatic observations have also been made linking seasonal changes to BMP nutrient treatment performance, although reports of the direction of seasonality differ. For example, dry summer conditions stimulated depressed summertime nitrate levels in a roadside infiltration trench due to water table withdrawal and subsequent denitrification (Mullins et al., 2020), but in a green roof, elevated N, P and C levels were observed during summer and were attributed to intense summer precipitation events and temperatures (Buffam et al., 2016). Despite these site-level observations indicating climate impacts BMP water quality treatment, the broad impacts of climate on BMP performance through treatment mechanisms are not yet known.

Evidence of the need for better understanding of the impact of climate on BMP nutrient management is the growing body of work indicating that climate is one key control on BMP hydrologic effectiveness. Multiple models have shown that regional climate impacts BMP hydrologic performance due to the regionality of rainfall and temperature regimes (Cook, VanBriesen, & Samaras, 2019; Jennings, 2016; Lammers et al., 2022; Voter & Loheide, 2021). These works have shown greater volume and peak flow reduction in dry climates (Lammers et al., 2022), rainfall index correlations with BMP performance vary depending on the rainfall patterns of a region (Cook et al., 2019),

and effectiveness ranging from 51% to 99% among comparable rain gardens due to climatic precipitation traits (Jennings, 2016). One model concluded that climate frameworks are necessary for understanding BMP functionality as the current basis of design, precipitation intensity as return periods, is too simple and BMP function is driven by nuanced timing of water and energy availability (Voter & Loheide, 2021). Each of these modeling studies focused on the hydrologic function of BMPs rather than the water quality function. The same conclusions regarding climate influence on BMP hydrologic functionality are not likely to translate directly to water quality because design decisions for water quantity and quality management may not align (Hunt et al., 2012; Lammers et al., 2022).

There is evidence that climate impacts the water quality performance of BMPs, yet broad relationships between climate and BMP water quality performance are not yet defined. Evidence of climate impact on soil-based BMPs includes the known relationships between climate and soil nutrient availability, field level assessments of climate-related traits impact on BMPs, and modeling assessments of the variable hydrologic performance of BMPs in different climates. A review of non-point source nutrient management with stormwater BMPs identified climate as one “unknown” in BMP performance, and identified the widely ranging means by which climate has been assessed in BMPs, calling for a meta-analysis of climate’s influence on BMP water quality performance (Lintern et al., 2020). However, there remains a need for a broad analysis of the influence of climate, especially on the impact of climate on biogeochemical processes of dissolved N and dissolved P removal.

2.3.2 Storm Characteristics

Like regional climate, BMP soil processing mechanisms are also impacted by the characteristics of a storm. Storm characteristics may include variables like precipitation intensity, duration, storm depth, antecedent dryness, and temperature (Sohn et al., 2019). The same soil conditions that are altered by climate, changing soil water content and subsurface flow, which in turn affect infiltration time and soil oxygen content, are also impacted by different storm characteristics. The key difference between climate and storm influence on soil conditions are the time scales that the soil alterations occur (Brady & Weil, 1996). Storm-derived changes in soil conditions stimulate changes in soil biogeochemistry that either induce or reduce water constituent removal and retention in natural and agricultural soils (Wu, Peng, Qiao, & Ma, 2018; Q. Zhu, Castellano, & Yang, 2018). For example, rainfall intensity is positively related to erosion and N and P loss in agricultural soils in the Mediterranean (Martínez-Mena et al., 2020) and in agricultural, grass, and woodland soils in China (Yao et al., 2020). Therefore, the presence and prevalence of mechanisms governing nutrient transport in soils, and the evidence of these mechanisms being driven by storm characteristics, together provide support that nutrient management in soil based BMPs is affected by storm characteristics.

The pollutant removal and retention capacity of BMP soils in the face of dynamic storm characteristics has been explored in observational site-level studies, with inconclusive results. For example, a study of 34 storms in vegetated swales in Texas found that storm size did not affect the suspended solids removal rate (Barrett, Walsh, Joseph F Jr., & Charbeneau, 1998). However, in eight bioretention cells in Burlington, VT, high storm size and peak flow rates were strongly correlated with nutrient leaching

(Shrestha, Hurley, & Wemple, 2018). Yet another contrary result was observed in bioretention test columns, where the impact of rainfall intensity was mixed - high intensities sometimes induced denitrification and enhanced nitrate removal (H. Wang, Gan, Zhang, Yu, & Zhu, 2021). The impact of storm characteristics on dissolved N removal is especially complex, where some studies report that high rainfall intensities result in poor removal due to rapid transport though soil, and others indicate that dissolved N removal via denitrification is achieved under high intensity, low antecedent dryness conditions (He, Qin, Wang, Ding, & Yin, 2020; Lopez-Ponnada, Lynn, Ergas, & Mihelcic, 2020; H. Wang et al., 2021).

Models of the relationship between storm characteristics and BMP water quality performance tend to show more consistent results than observational studies. In Beijing, China, a model showed that the time to peak within storms of the same intensity and duration had a negative relationship with suspended solids removal. This model also found that within a 10-year period, increasing storm duration decreased the percent of suspended solids reduction in BMPs (Gong et al., 2021). A model calibrated from BMPs in California showed that water quality standards for total suspended solids and total copper were more likely to be exceeded at high storm sizes (Ackerman & Stein, 2008). Finally, the EPA's SUSTAIN model was used to show that pollutant removal effectiveness of seven BMP types decreased, each at different rates, with increasing storm size in a Chesapeake Bay tributary watershed in Maryland (Hoss et al., 2016). Generally, models agree that increasing intensity, storm size, and time to peak each negatively affected water quality treatment performance of BMPs.

In addition to model evidence of storm controls on natural and BMP soils, the strong relationship between stormwater concentrations and storm characteristics provides further support of an underlying causal structure between storms and BMP performance. Stormwater contaminants like suspended solid particle size distribution, gross solids weight, and leaf litter mass all increase with increasing storm intensity (Winston & Hunt, 2017). Runoff particle size distribution is also related to rainfall intensity, storm depth, and duration (Charters, Cochrane, & O'Sullivan, 2015). Additionally, rainfall amount and intensity were found to be two of the three variables most strongly related to event loads in a database of 343 rain events in Minneapolis and St. Paul, MN (Brezonik & Stadelmann, 2002). This influence of storm characteristics on stormwater concentrations (BMP influent) provides support that BMP water quality treatment performance is likely to vary with changing storm characteristics, since influent concentrations are related to effluent concentrations (Barrett et al., 1998; Barrett, 2008).

Support that BMP performance would be influenced by storm-driven controls is extensive, there is still not a clear understanding of whether the influence of different storm characteristics on BMP performance can be summarized broadly across BMPs (Lintern et al., 2020). A meta-analysis summarized the impacts of storm characteristics on BMP flood mitigation performance, but noted that while 46 articles pertaining to climate relevant traits were identified by the review, only 17% of them contained information about pollutant reduction relating to storm characteristics (Sohn et al., 2019). Further, extensive storm and BMP water quality performance data in the United States are both accessible in public datasets, so data acquisition is not a barrier to studying the broad impact of storm characteristics on BMP water quality performance. Clearly

defining the relationship between storm characteristics and BMP water quality performance is increasingly critical as climate change impacts are realized. Namely, storms are expected to grow in intensity and the traditional intensity duration frequency (IDF) applied in stormwater engineering is expected to change drastically (J. Zhu, 2013). Further, nutrient loading from watersheds are expected to increase due to extreme precipitation conditions (Bai, Shen, Wang, Chen, & He, 2020; Ouyang, Parajuli, Feng, Leininger, & Wan, 2018). Thus, a comprehensive understanding of the relationship between storm characteristics and BMP water quality performance is an essential element supporting improved management of stormwater quality in a changing climate.

2.4 Summary of Research Needs

This dissertation explored three factors associated with urban surface water quality management: 1) urban stream monitoring and socio-economic distribution of stream quality, 2) the impact of regional climate on BMP performance and, 3) the impact of storm characteristics on BMP performance.

2.4.1 Objective 1: Evaluate socio-economic distribution of stream quality

Environmental justice models are a key component of understanding the spatial distribution of environmental burdens like urban stream degradation. Previous urban stream environmental modeling work has identified that spatial clustering, variable selection, and spatial scales are integral to creating valuable models. However, these models are difficult to transfer to watersheds beyond the one used to develop them, because spatially and temporally robust spatial data are required to develop adequate models. Using volunteer science data in combination with spatial modeling methods can address this lack of transferability without compromising model applicability.

2.4.2 Objectives 2 & 3: Investigate the influence of regional climate on BMP pollutant removal performance

BMPs are important controls of water quality as nonpoint source treatment mechanisms, but their pollutant removal performance is notoriously unreliable. Removing nutrients is a key challenge in urban water quality, and there is evidence that nutrient removal mechanisms in BMPs are likely to be influenced by factors driven by climate. Climate influence on BMPs can be categorized into 1) the impact of regional climate variability, and 2) storm related traits. An increased understanding of the influence of these parameters on BMPs would help to inform design of BMPs to perform more robustly, with suitable controls for their climate.

3. EVALUATION OF THE SOCIO-ECONOMIC DISTRIBUTION OF STREAM QUALITY

3.1 Rationale & Summary

Urban stream quality is compromised by anthropogenic influence, and the burden of this dampened water quality on different community members in a watershed is unknown. Understanding the full impact of urban stream degradation and planning for restoration action must consider where, and upon whom, the impacts of urban stream syndrome are most severe. Previous analyses have found weak trends, with contradictory positive and negative relationships between stream quality characteristics and socio-economic traits (Sanchez et al., 2014). Results from this objective improve on past analyses by using a spatial modeling tool designed for stream network analysis and partnered with a volunteer science group to together apply advanced modeling techniques with a robust stream quality dataset to address a common interest in the socio-economic distribution of stream quality degradation.

This chapter investigated the distribution of urban steam degradation in the Rouge River watershed. The Rouge River watershed is in Southeast Michigan, covering parts of the city of Detroit and its Western suburbs. Since 2001, stream quality on the Rouge Rivers and its tributaries has been monitored by the non-profit group Friends of the Rouge (FOTR). The goal of this work was to answer the question asked by FOTR volunteers, “who is affected by degraded water quality in the Rouge River”? To address this question, water quality was estimated with bioindicators of stream quality (counts of benthic macroinvertebrates) from the FOTR volunteer science monitoring program. This stream quality data was modeled with a combination of environmental variables and

socio-economic variables to assess the explanatory power of socio-economic traits towards predicting stream quality and explore the relationships between stream quality and environmental and socio-economic traits. It was hypothesized that the distribution of stream degradation in the Rouge River watershed is not uniform across communities of different socio-economic status.

The research addressing this objective is published in *Journal of Hydrology* (Horvath et al., 2022).

3.2 Methods

3.2.1 Study Area

The study took place in the Rouge River watershed in part of metropolitan Detroit, MI, and its northern and western suburbs. The watershed area is approximately 1200 km^2 and includes 204 km of stream segments composed of the Rouge River and its tributaries (Figure 1). The watershed drains into the Detroit River, which connects Lake St. Clair and Lake Erie in the Laurentian Great Lakes. The Rouge River watershed is highly urbanized, with 85% developed, 4% agricultural, and 6% forested landcover (NLCD, 2019). These landcover types are spatially heterogeneous across the watershed, with a general trend of increasing urbanization towards the outlet in the southeast. The northernmost and westernmost fringes of the watershed contain the most forested and agricultural landcover. The Rouge River twenty-year mean annual discharge is 147 million m^3 $year^{-1}$ (4.67 m^3 $second^{-1}$) (US Geological Survey, 2016). Landcover and hydrologic conditions within the various tributaries in the Rouge River watershed are diverse. The relatively undeveloped and rural headwaters contain the least impacted streams. The Rouge River stream segments span all levels of anthropomorphic alteration.

For example, some Northern streams are from groundwater-fed pristine segments with trout habitats ("Our Watershed," n.d.). However, closer to the outlet, multiple stream segments are encased in concrete channels. Johnson Creek on the western edge of the watershed (Middle Branch) is a protected trout stream, but creeks near the watershed outlet are submerged in pipes underground (Tonquish Creek) or channelized in concrete (lower River Rouge). The EPA identified the Rouge River as an Area of Concern under the Great Lakes Water Quality Agreement of 1987 and cited nine Beneficial Use Impairments in the watershed (Selzer, 2008).

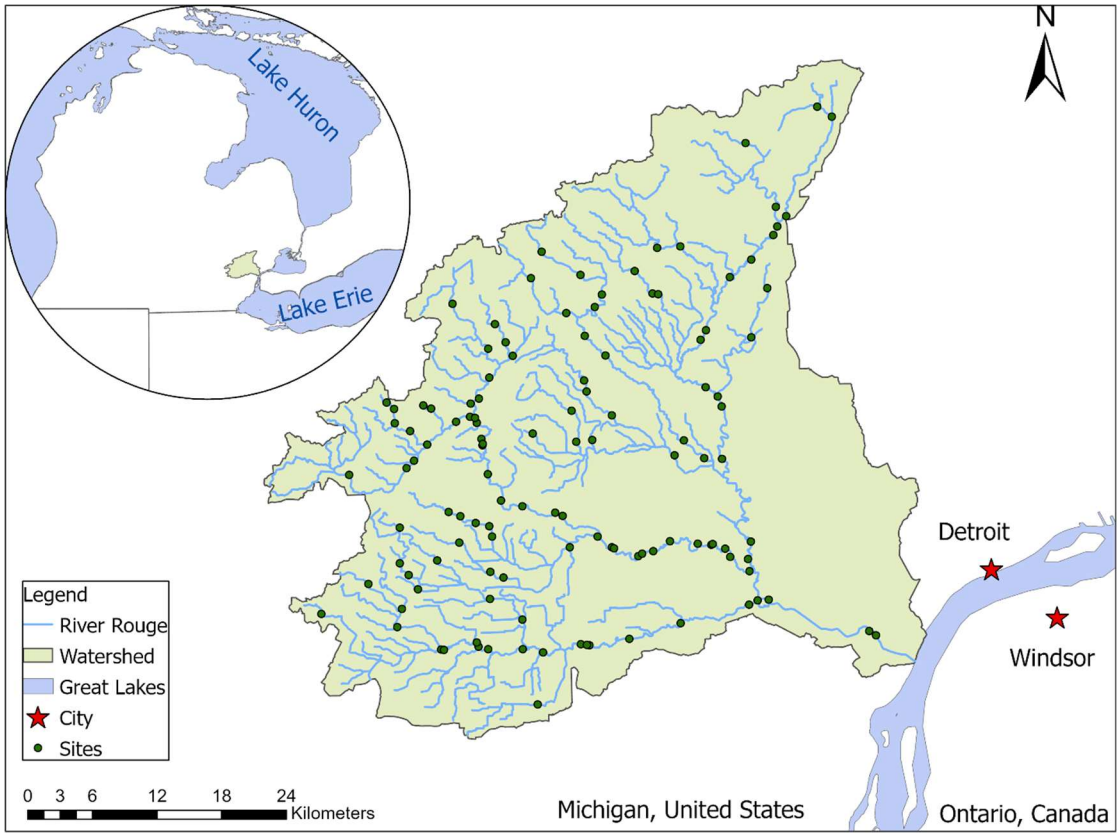


Figure 1: Map of the Rouge River watershed. The Rouge River watershed includes metropolitan Detroit and its Western suburbs. Volunteer science benthic macroinvertebrate data were collected sporadically at 122 observation sites along the Rouge River

3.2.2 Volunteer Science Stream Quality Index Data

Benthic macroinvertebrates are bioindicators of stream health and quality, and they are relevant in environmental impact studies near the Rouge River (Burlakova et al., 2018) and globally. Here, volunteer scientists collected benthic macroinvertebrate data as a bioindicator of water quality. This bioindicator data was applied as a composite score, "bio integrity", and was used in place of a chemical descriptor of water quality because bioindicator monitoring was a longstanding water quality monitoring method practiced by the FOTR community partner. Justification of the application of bioindicators as a water quality proxy, and their prevalence in volunteer science hydrologic monitoring is provided in the State of the Science section.

Macroinvertebrate species and frequencies were collected by FOTR volunteers. FOTR collected benthic macroinvertebrate data with volunteer scientists participating in biannual (Spring and Fall) "bug hunts". FOTR started collecting benthic macroinvertebrate data in 2001, and data collection is ongoing. Prior to collection and identification events, volunteers were trained as "bug hunt" team leaders in workshops led by both FOTR and a local biologist. Samples were collected from a rotating subset of 122 sampling locations (Figure 1). Trained volunteer scientist leaders surveyed instream habitats for benthic macroinvertebrates (riffle, cobble, pool, overhanging vegetation, undercut banks) with "D"-frame nets (Brua et al., 2011). Macroinvertebrates were preliminarily identified in the field, to order level. Four to five specimens of all but clams, mussels, snails, and crayfish were preserved in ethanol and later identified in the lab by FOTR staff and the local biologist to check field identifications and identify to family level. This method varies slightly from other benthic macroinvertebrate

observation studies wherein protocol dictate that all specimens are saved for post-hoc lab identification (Barbour et al., 1999).

The sensitivity of benthic macroinvertebrates and their frequencies of observation were converted to a Stream Quality Index (SQI) using the MiCorps' Macroinvertebrate Datasheet (Appendix Figure A1). SQI categorizes macroinvertebrates (mainly by order) into three levels: "sensitive", "somewhat sensitive" and "tolerant," based on pollution sensitivity and rates them as rare (1-10 individuals) or common (11 or more). Common "sensitive" organisms like mayflies are scored higher than common "tolerant" organisms. A higher SQI score reflects higher numbers of sensitive species like stonefly nymphs (*Plecoptera*) and hellgrammites (*Megaloptera*), indicating higher water quality. This study considers biannual SQI observations from spring 2001 through spring 2021 (n=1,655 site visits).

3.2.2.1 SQI Quality Assurance

All FOTR volunteer science SQI collection was completed using a quality assurance project plan reviewed by the Michigan Department of Environment, Great Lakes, and Energy (Michigan EGLE), the Michigan Department of Natural Resources, the Michigan Clean Water Corps (MiCorps), the Wayne County Department of Public Services, and FOTR (Petrella, 2020). FOTR checked SQI scores year to year and flagged data points that differed from past observations. Flagged points were reviewed for human error. Yearly observations of SQI were also checked against local knowledge and reported biannually. A validation study found that SQI calculated in the Rouge River and nearby Clinton River by volunteer scientists produced comparable, but more conservative estimates of stream quality than quantitative data collected by professional scientists

(Krabbenhoft & Kashian, 2020). The SQI is a water quality index used by monitoring groups in Michigan developed by the Michigan Department of Environmental Quality (now, Michigan EGLE) through their grant-funded program to engage volunteer science groups in benthic macroinvertebrate monitoring around the state. MiCorps is a statewide network that assumed oversight of the state-backed volunteer science monitoring program in 2003 (*Michigan Clean Water Corps: About*, n.d.). The establishment of the SQI metric in Michigan follows the popularization of bioindicators for water quality monitoring at the state and federal level in the late 1980s due in part to guiding programs like EPA's Rapid Bioassessment Protocol (Barbour et al., 1999, 2006).

3.2.3 Spatial Stream Network Modeling

Data paucity was overcome by building a Spatial Stream Network (SSN) model for SQI. This modeling step was performed to expand the spatial coverage of SQI data. The model was developed with two frameworks each shaped to test the overall hypothesis. The modeling frameworks were: 1) socio-economically extrinsic, and 2) socio-economically intrinsic, each of which was proceeded by a different statistical analysis for appropriate hypothesis testing under the framework. Methods regarding each modeling framework are provided in section 3.2.5 Hypothesis Testing. Both modeling frameworks use the same environmental variables and socio-economic variable and apply the same process of spatial model development and selection. Environmental variables, including landcover and stream characteristics, are applied as explanatory variables (for both intrinsic and extrinsic modeling frameworks). Poverty data was used for statistical comparison in the extrinsic modeling framework and as an additional explanatory

variable in the intrinsic modeling framework. Spatial modeling used geospatial modeling tools to model potential spatial relationships between points in a stream network system.

3.2.3.1 Environmental Variables

Landcover is a strong driver of in-stream conditions, where anthropogenic land uses, whether urban or agricultural, degrade stream quality (Brabec et al., 2002; Carlisle et al., 2009a; Chen et al., 2016; Epps & Hathaway, 2021; Tong & Chen, 2002). Degraded stream quality affects population size and diversity of benthic macroinvertebrate communities, which are sensitive to degraded stream conditions (Carlisle et al., 2009b; Walsh et al., 2001; Wang et al., 2018). Thus, sediment regulation (lack of degradation from sedimentation) and percent imperviousness watershed area were used as landcover characteristics to predict invertebrate population derived SQI. These parameters were obtained from the U.S. EPA StreamCat database and were available for each individual stream segment (Hill et al., 2016). Three different poverty metrics were weakly positively correlated with another water quality index in the neighboring watershed of the Saginaw Bay basin (Sanchez et al., 2014). Poverty was obtained from the U.S. Census Bureau's 2016 American Community Survey data.

Imperviousness is a measured value indicating the mean percent of landcover that is classified as an anthropogenic surface such as pavement, roads, and buildings (Figure 2b). The imperviousness variable is an average of the mean percent of impervious landcover within a stream segment's immediate and upstream drainage area as reported for all available yearly landcover datasets in the National Land Cover Database (NLCD) during the timeframe of water quality observations (2001, 2004, 2006, 2008, 2011, 2013, 2016, and 2019) (Dewitz & U.S. Geological Survey, 2021).

Sediment regulation is a modeled parameter on a scale of 0 to 1 that was developed to summarize sedimentation using instream and out-of-stream parameters in the StreamCat database (Hill et al., 2016; Thornbrugh et al., 2018) (Figure 2a). Sedimentation describes inorganic particle retention and size alteration due to transport to and within streams (Flotemersch et al., 2016; Thornbrugh et al., 2018). The sediment regulation parameter was calculated considering observed values of stressors relative to maximum stress level for five major stressors: 1) presence and volume of reservoirs; 2) stream channelization and levee construction; 3) alteration and changes to riparian vegetation; 4) frequency of mines, frequency of forest cover loss, and density of roads; and 5) agriculture presence weighted by soil erodibility (Flotemersch et al., 2016; Hill et al., 2016; Thornbrugh et al., 2018). This variable was developed for all streams and rivers in the conterminous United States, so it is expected that variability of this parameter within a single region would not express a high range of variability on the 0-1 scale developed for the full dataset, however, this does not mean that variability does not exist within a smaller region like a watershed (Thornbrugh et al., 2018).

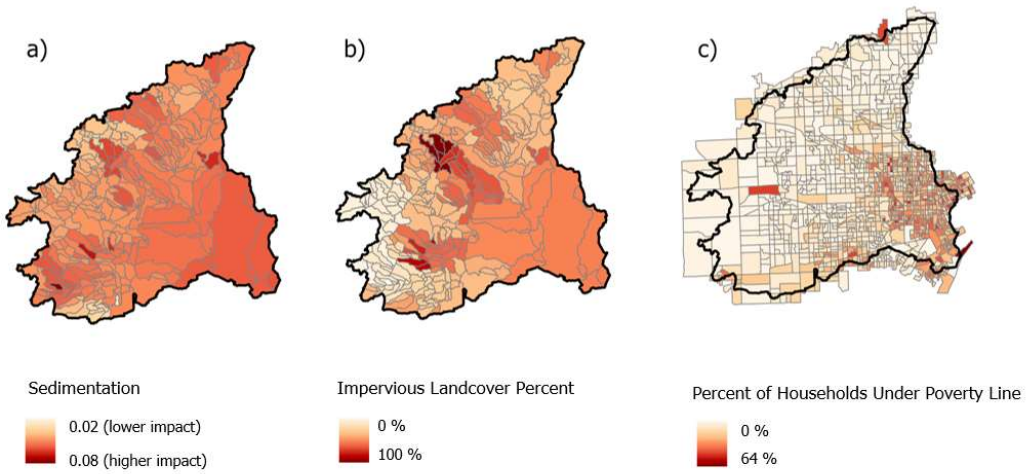


Figure 2: Relevant characteristics in the Rouge River watershed. Sedimentation (a) is a modeled parameter from 0-1 where 0 indicates low impact of sediment within a sub watershed, imperviousness (b) as the average percent of landcover identified as impervious, and percent of the population living under the poverty line (c) plotted in original data format as percentages within census tracts.

3.2.3.2 Socio-Economic Variable: Poverty

Poverty associated with each stream segment reflects census-tract level percentages of households living below the poverty line, an annual household income of $31,661 (U.S. Census Bureau (US Census), 2020). Poverty information was obtained as census tract-based and converted to the average poverty in the topographical boundary (catchment) of each stream segment. These catchment-level values were then averaged with upstream catchments to express the percentage of households below the poverty line in the entire upstream drainage area of each stream segment. Poverty as census tract-based measurements ranged from 0% to 91%, and when converted to upstream watershed-based, ranged from 0.2% to 24.5% of households in the catchment and upstream watershed residing below the poverty line.

3.2.3.3 Spatial Modeling Considerations

In addition to multiple explanatory variables, the SSN also considers spatial relationships between sites in models. Spatial relationships are categorized into either flow-connected or flow-unconnected relationships, based on whether there is a direct flow path connecting two sites. These relationships consider three autocovariance functions: tail-up, tail-down, and Euclidean distance. Tail-up autocovariance exists only between flow-connected sites, and they represent a weighted moving average function in the upstream direction. Tail-down autocovariance may exist under either flow-connected or flow-unconnected conditions, and they represent a weighted moving average function in the downstream direction. Euclidean distance may be considered in flow-unconnected relationships when autocovariance is not restricted to in-channel distances between sites (Garreta et al., 2010; Isaak et al., 2014; Ver Hoef & Erin, 2010). The weighting model for these tail-up and tail-down autocovariances can be calculated with linear, exponential, spherical, Mariah, and Epanech weights (Garreta et al., 2010; Ver Hoef & Erin, 2010). Euclidean autocovariance weighting included standard spatial covariance models: spherical, exponential, Gaussian, and Cauchy. The suitability of these various spatial autocovariances differs depending on the nature of the stream metric. For example, chemical data would be most likely to follow flow-connected tail-down autocovariance because chemical transport in a stream network is driven by transport in the channel, and in the downstream direction. However, macroinvertebrate-derived data may be represented with both flow-connected and flow-unconnected relationships since benthic macroinvertebrates have preferential travel along stream channels, but they can travel in

both in upstream and downstream directions and can also move outside of the confinement of stream channels (Isaak et al., 2014).

The SSN was implemented by using the Spatial Tools for the Analysis of River Systems (STARS) and SSN tools in ArcMap 10.8.1, R version 3.6.1, and RStudio version 1.2.5019, respectively (Peterson & Ver Hoef, 2014; Ver Hoef et al., 2014). SSN models were made with sediment regulation, imperviousness, and poverty as independent variables. The dependent variable was log mean SQI. Mean SQI was calculated as the mean SQI observation at a site through time. Means were taken to simplify temporally diverse data, because only 9% of sites observed a linear change ($p < 0.05$) in SQI over time, and this change was mixed, with 7 sites increasing and 4 sites decreasing SQI. Mean SQIs were logged to ensure normal distribution. All explanatory variables were normalized using min-max normalization to redistribute values from 0-1 based on the ranges of these variables measured at observation sites. This was done to standardize model covariates to the same scale. SSN models were constructed with multiple combinations of tail up, tail down, and Euclidean distance autocovariances to encompass the three possible spatial relationships between observation sites (Isaak et al., 2014; Ver Hoef et al., 2014). A final SSN model was then selected by comparing models with the evaluators: Akaike information criterion (AIC), coefficient of determination (R^2), and root mean square error (RMSE) calculated from leave one out cross validation (LOOCV). The best performing SSN of SQI as a function of the environmental variables and socio-economic variables was further evaluated by comparing it to two simpler models. The first simple model omitted the spatial component of the SSN and the second simple model omitted the socio-economic variable. Additionally, SQI could decrease

downstream along flowlines because of physical stream attributes associated with high flows and greater depth. To account for this, the best performing model was reparametrized with a random effect for stream order. Again, models with and without the stream order random effect were compared via AIC, R^2, and RMSE.

3.2.4 Hypothesis Testing

The hypothesis, that stream quality is not distributed uniformly across areas of variable poverty rates, was tested under both socio-economically extrinsic and intrinsic modeling frameworks. The principal difference between the modeling frameworks (and their associated statistical analyses) is the inclusion or exclusion of a socio-economic variable, percent of households below the poverty line (poverty), as an explanatory variable in the SSN model. The socio-economically extrinsic model does not use poverty as a model variable, but instead predicts stream quality then compares stream quality predictions in areas of high poverty and low poverty to evaluate the relationship between stream quality and poverty. This method uses a principal component analysis to ensure that only streams with similar landcover characteristics are compared. The socio-economically intrinsic method uses poverty as an explanatory variable in the model, then examines the statical value of the poverty variable in predicting stream quality. The relationship between poverty and stream quality is then explored by predicting water quality under different watershed conditions, where the same poverty conditions true to the Rouge River watershed are used to exemplify the degree of variability in stream quality that is associated solely with variable poverty rates.

3.2.4.1 Socio-Economically Extrinsic Hypothesis Testing

To test the hypothesis in a socio-economically extrinsic framework, modeled stream quality was compared between poverty segments in relatively rich and relatively poor areas, for streams classified as having similar land use attributes (Figure 3).

First, subsets of stream segments were made based on the classification of a stream segment into either a rich or poor poverty group. Poverty data (percentage of households below the poverty line in the land draining to a steam segment) was defined as rich or poor based on the distribution of poverty rates observed in the Rouge River watershed. Those stream segments whose drainage areas were classified as falling within the 90^{th} percentile of the range of poverty were considered high poverty, or poor stream segments. Similarly, stream segments with drainage areas classified within the 10^{th} percentile of the range of poverty were considered low poverty, or rich stream segments.

After quantifying the spatial distribution and range of percentages of household below the poverty line, stream segments with similar landcover attributes were identified. This step was included to isolate the influence of poverty. In other words, this step was added to reduce the influence of landcover on the comparison of stream segments so that, for example, forested rural segments were not inadvertently compared to developed urban stream segments. Forest cover, urban open space, low intensity urbanization, medium intensity urbanization, and high intensity urbanization as defined by NLCD were used to characterize each river segment. NLCD landcover categories are highly correlated, thus landcover for each stream segment catchment was reprojected as orthogonal principal components via PCA. The use of a PCA allowed for multivariate similarity and does not require assumptions of landcover data distribution. Thus, the PCA was able to identify

data points useful towards assessing the hypothesis, by identifying stream segments with like land use between the high and low poverty groups. Similar stream segments were identified with a plot of the first two PCs of stream segment catchment landcover and used for hypothesis testing. The extent of high poverty points and low poverty points were used to form polygons bounding poverty groups. The low poverty and high poverty sites located within the two intersecting polygons were defined as having sufficiently similar landcover for comparison. Finally, ANOVA was used to estimate the difference in SQI between those intersecting high poverty segments and low poverty sites.

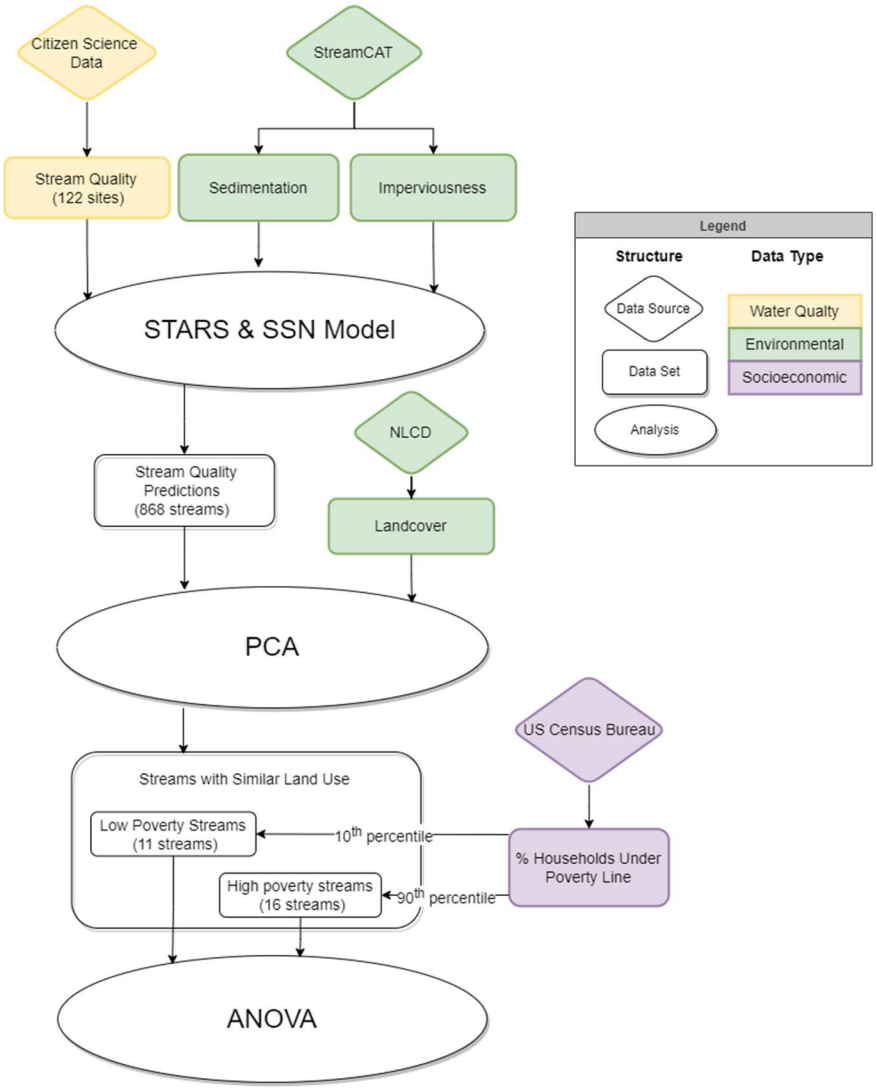


Figure 3: Flow diagram of socio-economically extrinsic methods, highlighting data inputs and analysis methods.

3.2.4.2 Socio-economically Intrinsic Method

To test the hypothesis in a socio-economically intrinsic framework, stream quality was modeled with three explanatory variables, two environmental variables and one socio-economic variable. The hypothesis was tested by evaluating the statistical power of the socio-economic variable within the model. The extent of the relationship between the socio-economic variable and water quality was then demonstrated by predicting water

quality under different plausible environmental conditions in the watershed, with the socio-economic variable held constant.

First, a model was selected and evaluated. Models were built with three explanatory variables, applying different flow path and spatial relationship combinations. The pre-determined model selection (detailed in section 3.2.4.3) was used to identify the best fitting spatial socio-economically intrinsic model. The p-value of the socio-economic variable was used to determine the value of the relationship between the socio-economic variable and stream quality, where a p-value <0.1 indicates a moderate relationship and a p-value <0.05 indicates a strong relationship.

After testing the hypothesis, the extent of the relationship was evaluated by applying the model to make predictions under varying environmental conditions in the watershed (subsequently, hypothetical watershed conditions). SQI was predicted every 800 m of all stream segments in the Rouge River watershed. SQI predictions were made under four watershed conditions: true (observed) conditions, and three levels of hypothetical watershed conditions – good, standard, and poor conditions (Figure 4). Each hypothetical watershed condition used manipulated values of imperviousness and sediment regulation and observed values of poverty. The values of imperviousness and sediment regulation conditions assigned to the "good", "standard" and "poor" labels were selected to represent a range of values that are realistic for the watershed. Good conditions were defined as imperviousness at 25% of the range of imperviousness observations (18% imperviousness) and 75% of the range of sediment regulation (0.96). Standard conditions were defined as imperviousness at 50% of the range of imperviousness observations (35% imperviousness) and 50% of the range of sediment

regulation (0.94). Poor conditions were defined as imperviousness at 75% of the range of imperviousness (53% imperviousness) and 25% of the range of sediment regulation (0.92). Imperviousness and sediment regulation intervals were opposite one another because increasing imperviousness is associated with poor environmental conditions, while increasing sediment regulation indicates higher integrity, or lack of impact from sedimentation, and is thus associated with better environmental conditions. These intervals were made to demonstrate the impact of poverty on SQI under different environmental conditions that were reasonable in the context of the ranges of imperviousness and sediment regulation observed in the watershed. Linear models of predicted SQI and poverty were generated based on the four conditions above. The slopes of these linear models were then compared.

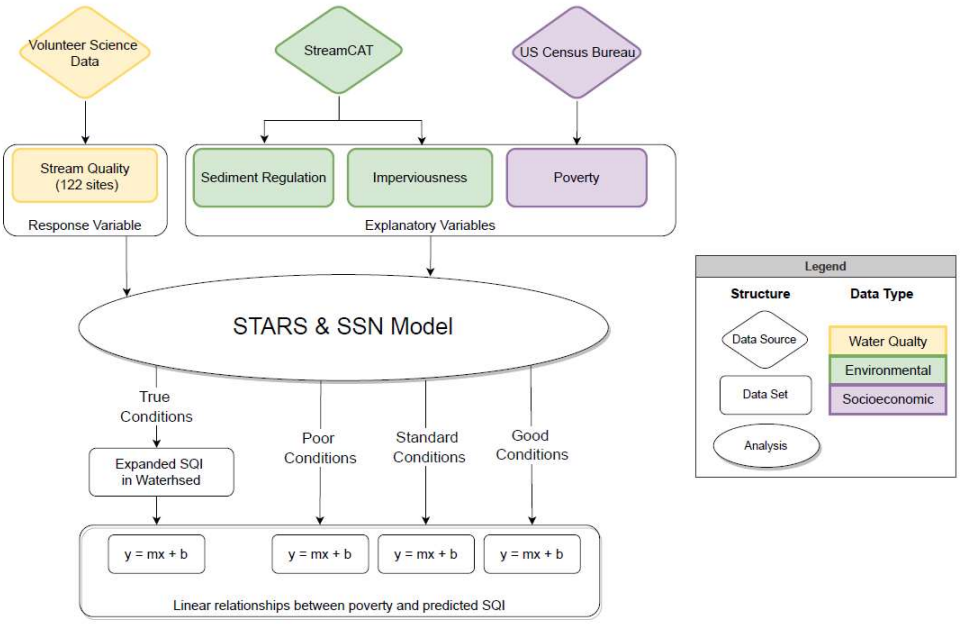


Figure 4: Flow diagram of socio-economically intrinsic methods, highlighting data inputs and analysis methods.

3.3 Results

There was a total of 1,655 SQI observations from 122 sites over 20 years of biannual observation. Average SQI measurements per site ranged from 14 to 48 (Figure 5a). Stream quality was generally worse on the main branch and near the watershed outlet. However, poor quality was also observed in some headwater streams. The highest quality was observed in headwater streams on the western edge of the watershed.

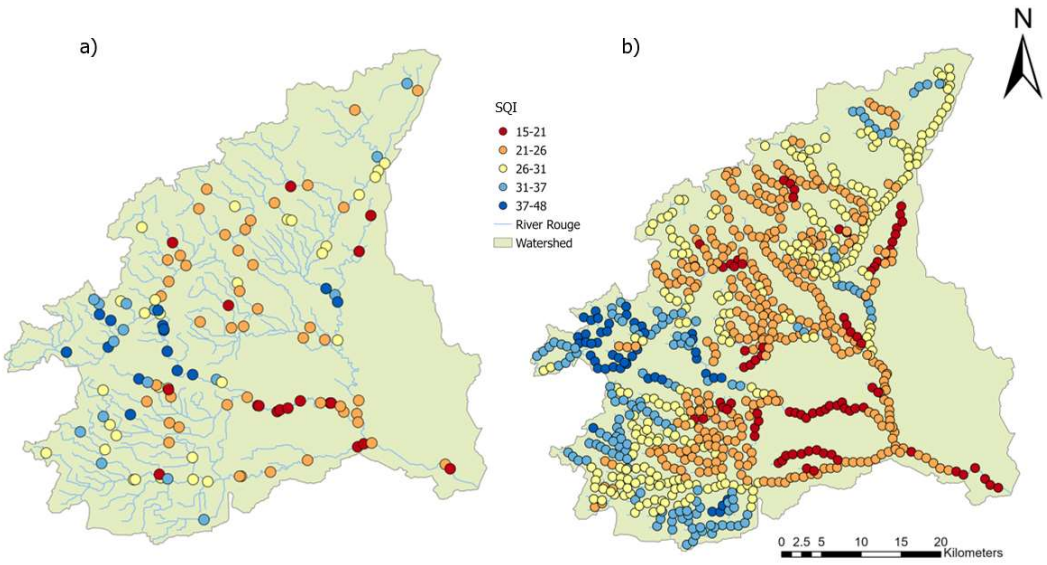


Figure 5: Observed and modeled SQI data. SQI measures were collected for sites in the Rouge River watershed by the volunteer science organization Friends of the Rouge. Observations of SQI (a) compared to modeled SQI along every 800 m of stream (b).

3.3.1 Socio-Economically Extrinsic Model Performance

The selected extrinsic SSN model used sedimentation regulation and imperviousness in a multivariate spatial regression model with tail up and Euclidean correlations (AIC = -86.68, R^2 = 0.64, RMSE = 3.14). This model outperformed a model with stream order as a random effect and imperviousness and sedimentation as fixed effects (AIC = -84.68, R^2 = 0.64, RMSE = 3.14), and a model with stream order, sedimentation, and imperviousness as fixed effects (AIC = -80.30, R^2 = 0.63, RMSE = 3.17). The SSN model was used to predict SQI at 868 prediction points in the Rouge River watershed (roughly one prediction point for every 800 m of stream segment). Predicted SQI ranged from 15.76 to 44.83 (Figure 5b), a narrower range than observed SQI. The average prediction standard error was 1.18. The model yielded a negative coefficient with imperviousness (-0.66 ± 0.23), indicating lower SQI with higher

imperviousness. The model had a positive coefficient with sedimentation regulation (8.53 ± 3.66), indicating higher SQI where sedimentation is less prevalent. Sedimentation regulation and imperviousness were weakly correlated (Pearson's correlation coefficient = 0.12).

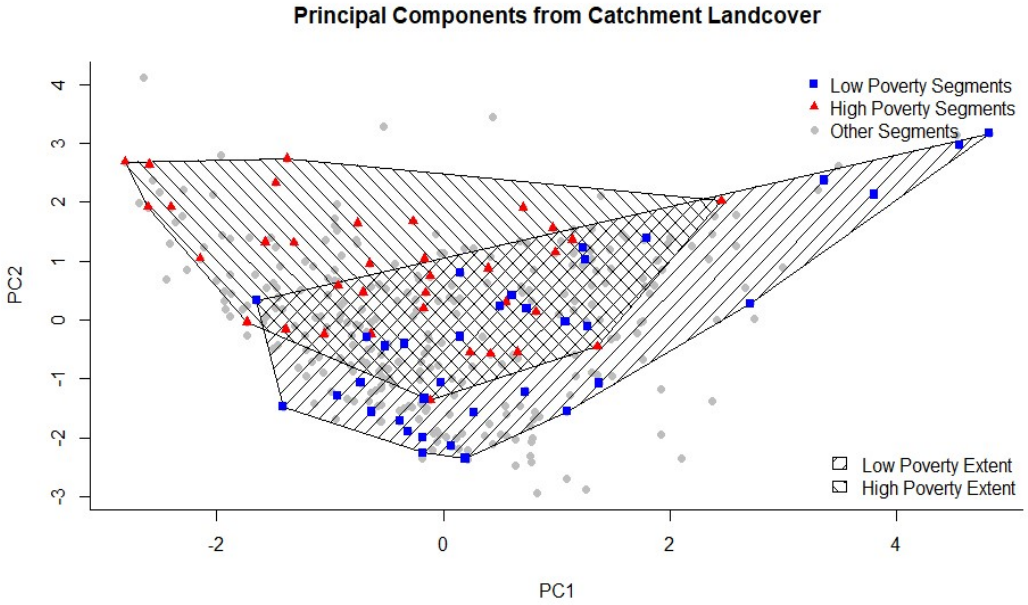


Figure 6: Principal components one and two from a PCA of land cover types in the Rouge River watershed. The five land cover classifications used for the PCA were reprojected as composite variables via PCA for each stream segment in the Rouge River watershed (grey circles). The segments were then classified as rich or poor based on their distribution, so that stream segments in the 10th percentile of houses below the poverty rate were labeled "low poverty" (or rich, labeled with blue squares). Segments with poverty percentages in the 90th percentile of the distribution of percentage of households below the poverty rate among all stream segments in the watershed were labeled as "high poverty" (or poor, labeled with red triangles). The extent of the high and low poverty segments was then outlined for both groups with a polygon. The intersection of the polygons defines stream segments with like landcover, thus outlines the high and low poverty segments suitable for comparison.

The PCA of the Rouge River watershed landcover reprojected the five most dominant NLCD categories: forest, urban open space, low intensity urbanization, medium intensity urbanization and high intensity urbanization. These categories were

condensed because the less common landcover types results in a PCA with low cumulative variance coverage from the first two principal components (PCs), thus over amplifying the influence of landcover with little spatial coverage. The PCA with the five primary landcover types had a cumulative 70% of the variance in the data represented by the first two PCs (36% by the first and 34% by the second). Each PC represented a different compression (i.e., dimension reduction) of landcover descriptor composite variables. There were 27 stream segments belonging in the poverty groups of interest (high or low poverty) with similar landcover types. There were 16 high poverty stream segments and 11 low poverty streams stream segments (Figure 6). The high poverty streams had a median SQI value of 26.0 (Q25 = 24.3, Q75 = 31.4), classified as fair stream conditions. The low poverty streams had a median SQI value of 32.5 (Q25 = 29.8, Q75 = 34.4), on the threshold of fair and good stream condition (Appendix Figure 1). The ANOVA between these high and low poverty segments of like landcover estimated that low poverty segments have an SQI 4.1 ± 3.6 index points greater than high poverty segments (Figure 7). Additionally, the p-value indicates that this difference in SQI has a low likelihood of being produced by chance (3%).

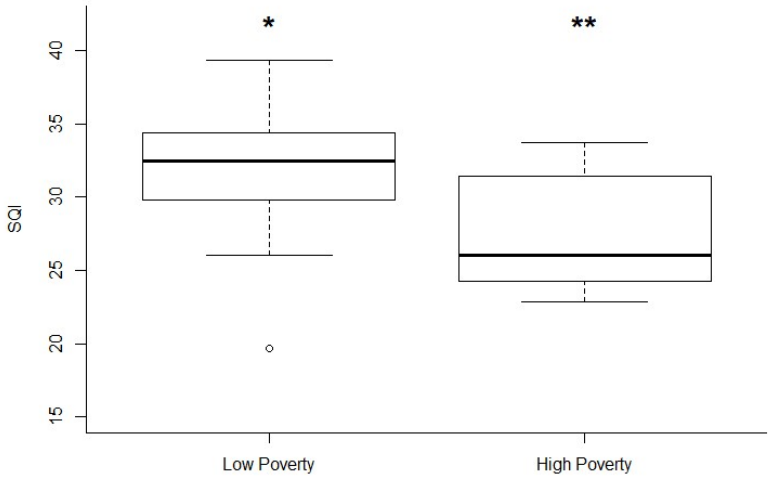


Figure 7: Comparison of SQI between high and low poverty stream segments of like landcover. Different numbers of stars indicate statistical difference between box plots.

3.3.2 Socio-economically Intrinsic Model Performance

The best performing SSN model used sediment regulation, imperviousness, and poverty in a multivariate spatial regression model with a linear-sill tail-down autocovariance and no random effect on stream order (Supplementary Table 1). This model captures 36% of the variability in SQI ($R^2 = 0.36$). The model AIC was the lowest of all attempted variants of the model (Table 1). The RMSE indicates that prediction error is 3.14 SQI points, or about 10% of the range of observed SQI values. The explanatory variables are correlated with one another, however, variance inflation factors (VIF, (Helsel & Hirsch, 1992) for sedimentation, imperviousness, and poverty were low (1.23, 1.23, and 1.16, respectively). These are close to the ideal value (VIF ~1, Helsel & Hirsch, 1992) and below the cutoff value applicable for SSN models (VIF <5, Isaak et al., 2017), thus suitable for hypothesis testing. Imperviousness and poverty had negative

relationships with SQI with model coefficients -0.28 (p = 0.01) and -0.23 (p = 0.05), respectively. Sediment regulation had a positive relationship, model coefficient 0.30 (p = 0.07), which can be interpreted as a mild positive relationship between higher stream quality and high stream integrity from sedimentation. The linear sill tail-down autocovariance indicates that both flow-connected and flow unconnected relationships exist in the SQI data, and that these relationships are linear and point downstream. This means that between two SQI observations the downstream point is influenced by the upstream point and that relationship decreases linearly with increasing distance between the points.

Table 1: Model selection parameters for the best performing model and parallel models excluding spatial modeling methods, socio-economic data, and including stream order.

Model Name	Spatial Relationship	Variables	Random Effect	AIC	R^2	RMSE
Simple	-	Imperviousness Sediment Regulation Poverty	-	-48.01	0.40	4.11
Spatial Environmental Only	Linear Sill Tail-down	Imperviousness Sediment Regulation	-	-83.33	0.31	3.16
Spatial Random Effects	Linear Sill Tail-down	Imperviousness Sediment Regulation Poverty	Stream Order	-81.77	0.36	3.16
Spatial	Linear Sill Tail-down	Imperviousness Sediment Regulation Poverty	-	-83.77	0.36	3.14

This spatial socio-economic environmental model outperformed the simple model, spatial model fit with only environmental predictors, and spatial model with a random effect for stream order. The simple model had a higher R^2 value (Table 1), but

lower AIC and RMSE. The spatial environmental-only model had a slightly higher AIC, lower R^2, and higher RMSE compared to the best model (Table 1). The random effects spatial model had the same R^2 as the spatial model without random effects, but it had worse AIC and RMSE metrics. Comparing the RMSE values across models highlights the value of modeling SQI with SSN models, as the RMSE for the simple model was about one SQI index point higher than the RMSE for either of the spatial models, indicating worse ability of the simple model to capture the true variability in SQI data (Figure 8). This difference in model performance indicated by the RMSE value justifies that the spatial model outperforms the simple model, despite the higher R^2 of the simple model, as the simple model makes predictions with less accuracy than the spatial model. Poverty adds predictive power to the model, as demonstrated by the improvement in all model evaluators when poverty is included in the spatial model compared to the spatial environmental only model.

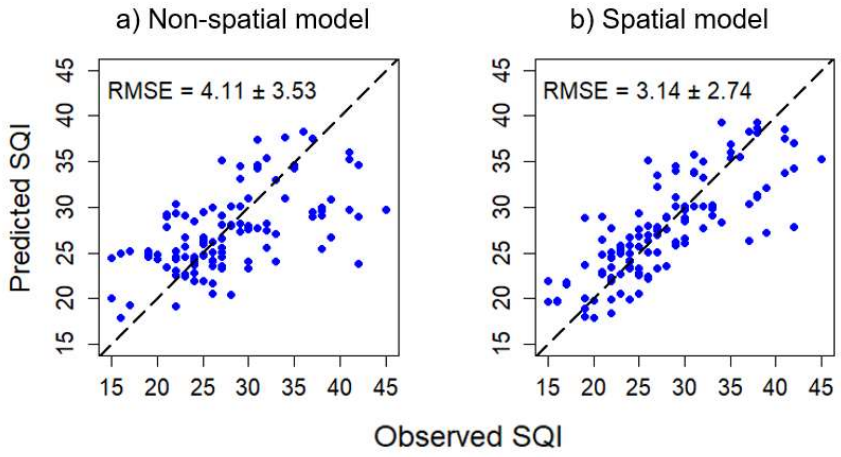


Figure 8: Leave one out cross validation (LOOCV) results compared for a non-spatial model (a) containing the same predictor variables as a spatial model with socio-economic and environmental variables (b). Root mean square error (RMSE) and the standard deviation of this calculation is printed on each plot, showing higher RMSE and standard deviation for the simple non-spatial model than for the spatial model.

Adding stream order as a random effect did not improve model performance. The stream order random effect model had a higher AIC and RMSE, and a comparable R^2 as the best performing model. This showed that the relationship between SQI and explanatory variables did not vary based on the stream order. In other words, small streams should not be modeled differently than larger branches. This provides support that stream order and associated downstream trends do not explain water quality in the watershed better than sediment regulation, imperviousness, and poverty without stream positioning information.

3.3.3 Predictions Under Potential Scenarios

The spatial environmentally intrinsic SSN model was used to predict SQI every 800 m of stream segment in the Rouge River watershed under both true and hypothetical watershed conditions. Under true (observed) sediment regulation and imperviousness

conditions, SQI predictions ranged from 15.76 (poor) to 44.83 (good). This range is slightly more conservative than the observed SQIs. The average prediction standard error was 1.17. The slope between poverty and predicted SQI was negative and indicated that a stream segment with 10% higher poverty in its upstream watershed drainage area would have a 3.62 lower SQI. This 3.62 change in SQI is equivalent to a 10% change in the observed range of water quality, which translates to about a 1% decrease in water quality for every 1% increase in poverty.

Under observed and hypothetical (poor, standard, and good) watershed conditions, poverty and predicted SQI also had negative relationships (Figure 9). The magnitude of this negative relationship increased with increasingly positive watershed conditions. Under poor watershed conditions (53% imperviousness, 0.92 sediment regulation) a 10% increase in poverty would result in a decrease in SQI by 2.87. Under standard watershed conditions (35% imperviousness, 0.94 sediment regulation) a 10% increase in poverty would decrease SQI by 3.61. Finally, under good watershed conditions (18% imperviousness, sediment regulation = 0.96) a 10% increase in poverty would decrease SQI by 4.53.

The slope between poverty and SQI is the same for the manipulated standard conditions and true conditions because the standard conditions were selected as the median values of true conditions. The poor, standard, and good condition scenarios have high precision amongst predictions compared to the true conditions. This difference in precision is due to the homogenous assignment of watershed conditions in the manipulated scenarios, compared to the naturally varying environmental observations for the true condition predictions.

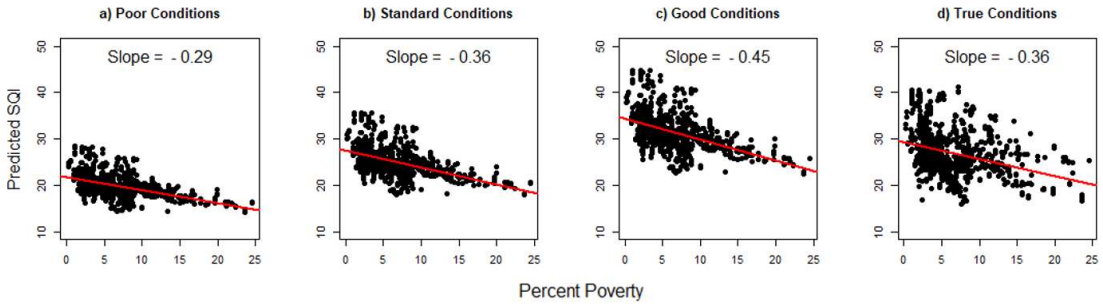


Figure 9: Relationships between predicted SQI and poverty under hypothetical poor (a), standard (b) and good (c) watershed conditions, compared to the relationship under true watershed conditions (d). The slope of the linear relationship between predicted SQI and poverty is plotted under each scenario.

3.4 Discussion

3.4.1 Socio-Economically Intrinsic vs. Extrinsic Modeling Approach

The socio-economically extrinsic and intrinsic models both arrived at similar conclusions, demonstrating that there is a general relationship of higher poverty catchments containing less healthy stream quality. However, the socio-economically intrinsic model arrives at this conclusion with simple hypothesis testing and less propagation of error, whereas the extrinsic model requires complex hypothesis testing, potentially ambiguous categorization of poverty, additional variable introduction, and generally more complex methodology.

The primary difference between the model approaches is the inclusion or exclusion of poverty from the model. The benefit of the inclusion of poverty in the model is that the p value for the coefficient of poverty can be used to directly gauge the relationship between poverty and stream quality (within the context of other environmental variables known to be important drivers of urban stream quality).

However, without the inclusion of poverty in the model, the relationship between poverty and stream quality must be assessed using the model predictions with other statistical analyses (PCA and t-test). Thus, despite the similar outcomes, the socio-economically extrinsic model introduces more complexity and potential for error propagation from one statistical analysis to the next, potentially convoluting the interpretation of results.

The statistical methods introduced to test the hypothesis in the socio-economically extrinsic model introduces ambiguity and assumptions to the method. First, the process of comparing model predictions in rich and poor areas requires a method to define poverty groups. Here, the 90^{th} and 10^{th} percentile interval of the poverty range were used to define high poverty and low poverty groups, respectively. This assumption helps define distinct groups but reduces direct interpretability of the results. For example, the context to understand whether this statistical distinction is meaningful from an economic standpoint is not explored in this work. A second fault introduced by the extrinsic modeling method is the introduction of landcover types for use in the PCA. This method was added to identify similar stream segments, so by using landcover types in the PCA it is assumed that landcover is what defines similarity and differences amongst stream segments in the Rouge River watershed. This assumption may not be true, especially in a watershed with diverse stream orders (1-6), and diverse channel conditions (from natural streams to channelized river). Thus, while the PCA and t-test methods of hypothesis testing are valid analyses, their associated ambiguity and assumptions are best avoided.

The socio-economically intrinsic model avoids complication and assumption from complex hypothesis testing, does not propagate error between statistical methods, and is

overall simpler to conduct and interpret. The duration of the discussion focuses on the interpretation and impact of the socio-economically intrinsic model.

3.4.2 Degraded Water Quality in Higher Poverty Areas

The identified negative relationship between water quality and poverty provides information about the spatial distribution of water quality degradation. The negative (socio-economically intrinsic) model coefficient between stream quality and poverty provides statistical evidence that stream quality is associated with socioeconomic factors, in addition to known relationships between stream quality and environmental factors like sediment regulation and imperviousness.

The observed decrease of stream quality in higher poverty catchments provides support that urban stream degradation is inequitably distributed. It is important to emphasize that the negative relationship does not prove a causal relationship; it provides statistical support that environmental degradation of water quality disproportionately affects impoverished communities. Explicitly, it is incorrect to interpret that high poverty causes poor water quality. While a cause-effect relationship may exist, this analysis does not articulate an underlying causal structure. Previous research provides support for potential casual structures. For example, inequity in access and proximity to parks has been shown for poor communities (Rigolon, Browning, & Jennings, 2018), and park land is one tool used to impede stormwater runoff from polluting streams (Cettner, Ashley, Viklander, & Nilsson, 2013). Further, communities of color and of low income are disproportionally located near highest-intensity chemical polluters (Collins, Munoz, & Jaja, 2016), and proximity to pollution generation may also explain stream conditions in high poverty areas of the Rouge River watershed as well.

Local knowledge and spatial setting contextualize the relationship between poverty and water quality, highlighting the co-occurrence of high poverty and high density of combined sewer overflows (CSOs) and low stream density in the same region. The highest poverty area in the watershed is in the Southeast region of the watershed (Figure 10). Observations of SQI in this area included 23 sites, with an average SQI of 24, a "fair" rating. While this SQI score is relatively low, it fails to express other stream quality issues in this area. The segment of the Rouge River bordering the highest density poverty area contains 21 uncontrolled CSO outfalls, making this area subject to flashy water levels and at risk to acute degradation events post rainfall as is true in cities with similar drainage systems, like Philadelphia, PA, and Chicago, IL (Miskewitz & Uchrin, 2013; Quijano, Zhu, Morales, Landry, & Garcia, 2017). Further, tributary streams in this area are sparse, having been removed from their historical locations and remaining only in historical records, now called "ghost streams" (Figure 10). The lack of tributary streams in this area is an example of water inequality, as this high poverty area is deprived of natural stream presence entirely.

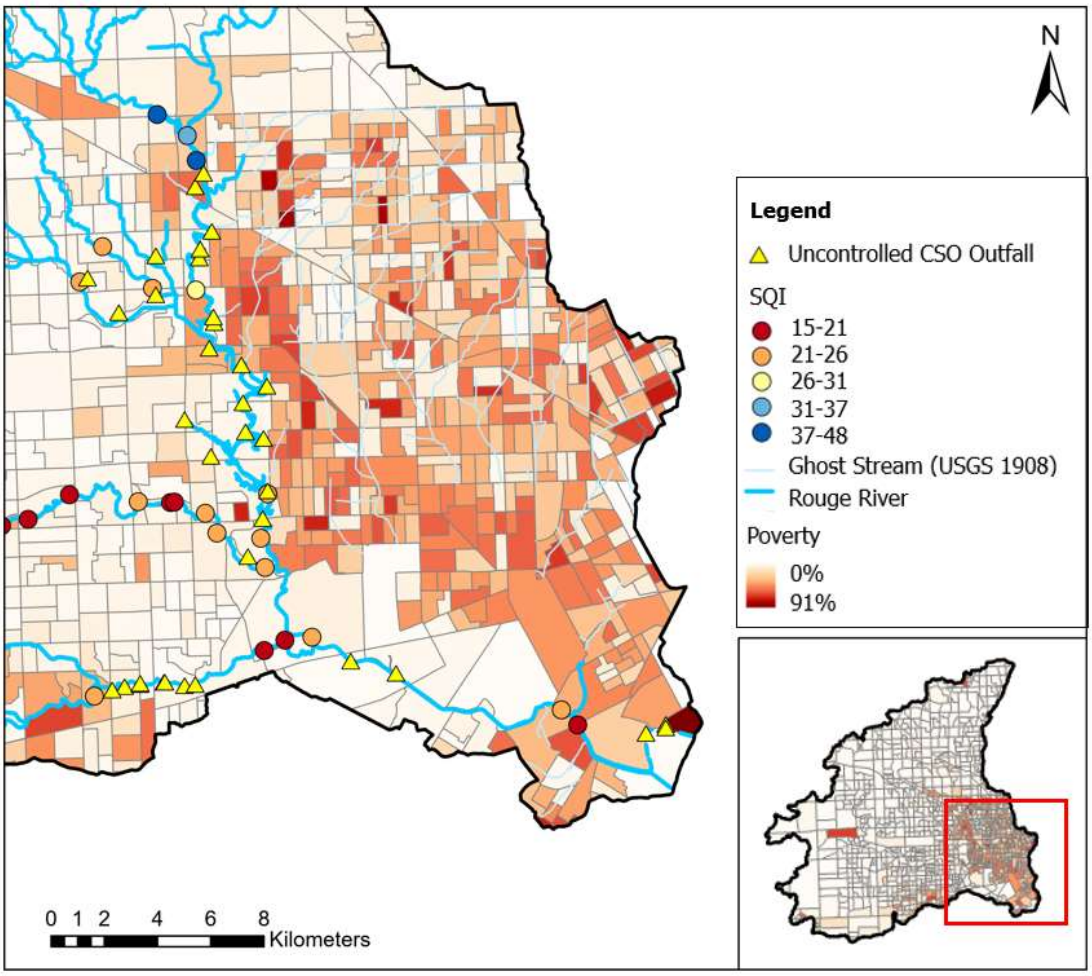


Figure 10: Map of poverty rates in census tracts in the Southeast portion of the Rouge River watershed. Locations of uncontrolled Combined Sewer Overflow (CSO) outfalls and historic locations of ghost streams are shown.

This lack of naturally formed stream channels is also a limit of the analysis – lack of natural drainage boundaries in high poverty areas, as well as highly urbanized areas, compromise the catchment-level units of analysis. In these areas, measurements of sediment regulation and imperviousness may not properly represent the land being drained to stream segments since stormwater infrastructure in a combined sewer system would carry stormwater to a wastewater treatment plant, or in an overflow event, may convey water to stream segments that would not have naturally received that water. To

estimate water quality more accurately in the high poverty area of the Rouge River, future work would need to consider conversion of naturally delineated drainage areas (delineation by elevation) to drainage areas defined by stormwater infrastructure (Achleitner, Möderl, & Rauch, 2007; House et al., 1993; Tscheikner-Gratl et al., 2019).

Other limits of the poverty analysis are the quality of U.S. Census data, and the assumptions made in converting poverty data from census tract to catchment-based units. A limitation of environmental justice datasets is low survey responses and lack of internal community involvement in surveying (Lee, 2020; Mah, 2017). Increased involvement of local community members in environmental justice data collection is necessary for increased understanding of the disproportionate water quality burdens across socioeconomic groups. A second layer of potential error in U.S. Census data was introduced when this data was converted from census tracts to drainage area. This conversion was made by assuming that poverty was distributed homogenously within census tracts. This assumption is an over-generalization that could lead to inaccuracy in calculating poverty rates in units of catchments. Scales of socioeconomic data resolution are influential in improving stream health modeling performance (Daneshvar et al., 2016), so future modeling efforts would benefit from a more realistic conversion of socioeconomic data from census-area to area units more conducive to water quality modeling.

3.4.3 Volunteer Science Data Applicability

Volunteer science-collected water quality data was key to executing this work. The term volunteer science was selected intentionally over similar titles (i.e., citizen science, community science, community-based monitoring) because volunteers collected

data and volunteerism was entirely unrelated to citizen status (contrary to the implication of the term citizen science), and the community was not involved in all stages of the research (as is common in community science), but rather, was primarily involved in data collection (Cooper et al., 2021).

Our work serves as an example of a mutually beneficial partnership between formal research and volunteer science. Labor, cost, time, and local knowledge would have prevented this research without volunteer science collaboration, which provided a temporally and spatially robust dataset. For the volunteer science data collecting group FOTR, technical and resource hurdles stand in the way of the spatial model building and analysis needed to interpret complex spatial correlation in river data. This mutually beneficial partnership between scientists and the local community offers the exchange of knowledge and perspective from interested parties who come from diverse backgrounds and motivations (Taylor et al., 2021), and is one reason why volunteer science has recently become more prevalent in aquatic science and hydrology research (Kielstra et al., 2019; Krabbenhoft & Kashian, 2020; Maguire & Mundle, 2020). An additional co-benefit of FOTR volunteer science is that data collection events are used to engage volunteer scientists in the watershed, raise awareness about river conditions, and advocate for the need to clean up the Rouge River.

Despite the benefits offered to both scientists and volunteer science groups, there are obstacles that prevent the widespread use of volunteer science data. These obstacles include scientific community acceptance, data validity and governance, research problem definition, and in the case of water quality – observation tool expense and access (Buytaert, Dewulf, De Bièvre, Clark, & Hannah, 2016; Buytaert et al., 2014). The most

common critique of volunteer science is data validity (Jollymore et al., 2017). Means to overcome this obstacle include volunteer scientist training and understanding of volunteer science volunteerism motivation, which increases the reliability (Alender, 2016; Buytaert et al., 2014; Jollymore et al., 2017).

In volunteer science organized by FOTR, volunteer training and internal quality assurance checks are the primary means of data quality assurance. Volunteer training consists of two tiers of volunteers, those being trained for general assistance and those trained as team leaders. Team leaders are trained in the classroom on field sampling methods and identification to genus or species. To become a team leader, a volunteer must first attend a sampling day as a regular volunteer, proceeded by the classroom training, then the volunteer attends multiple sampling events with an experienced team leader to ensure thorough sampling and proper adherence to procedure. Team leaders repeat classroom trainings after a few years to stay refreshed on methodology. On sampling events, team leaders conduct all sampling, and general volunteers assist in picking through samples. Team leaders also collect some specimens as vouchers, to verify identification in the lab.

Quality assurance of collected data occurs by comparison to historical observations. Newly collected SQI data are compared to historical observations of SQI at the same site, and large deviations are investigated by returning to voucher species and field notes, to determine potential error. A reliability study on FOTR volunteer science data concluded the SQI data used here is a conservative estimate of water quality as traditionally measured numerically by scientists (Krabbenhoft & Kashian, 2020). The macroinvertebrate preservation method used by FOTR may be one potential source of

this discrepancy, as only 4-5 representative specimens are preserved for post-hoc identification rather than preserving all samples, as recommended by other benthic macroinvertebrate sampling (Barbour et al., 1999).

3.4.3.1 Lessons from Friends of the Rouge

The long-term operation of volunteer science at FOTR has resulted in many learned experiences that can benefit other communities, including the scientific community. FOTR volunteer science has three take-aways from the success of their on-going sampling initiative: an effective sampling model for their community, accessible data sharing, and openness to exploring and supporting co-benefits from the sampling process.

Over the 20+ years of volunteer science events, FOTR has explored different sampling efforts, arriving on large, concentrated sampling days as the most successful model. Initially, FOTR provided training and equipment and expected trained volunteers to monitor sites on their own time and report findings to FOTR. This model failed to engage volunteers, and consequently FOTR altered their sampling events to group sampling days with the trainees leading untrained volunteers. This structure allows for wide community participation, with over 100 volunteers attending monitoring days. Success of this method is measured through volunteer retention, and influence of volunteering experience on community members. Many volunteers return year after year, some for as long as 20 years. Volunteers have reported that at these group sampling events they learn about stream ecology and urban rivers. A valued metric reported by volunteers is the impact of these volunteering events on children, where multiple children

who participated in sampling events with their parents indicated a desire to pursue a degree in sciences because of the experience.

FOTR also attributes their success to their commitment to ensure that the data is useful and made available to stakeholders. Following each monitoring event, a report is made available to all volunteers, and state and local agencies, including the communities who are now providing some of the funding to support monitoring. FOTR makes the data freely available to academic institutions for research use, which has resulted in multiple journal publications (Krabbenhoft & Kashian, 2020; Maguire & Mundle, 2020) and several Master's students theses.

Volunteer science events conducted by FOTR have also resulted in unsuspected co-benefits that arose because of the observations made by community members spending time in streams. Inspired by questions from volunteers about pipes while sampling, team leaders are now trained in illicit discharge elimination and have been responsible for reporting spills, sewage leaks, erosion issues, and more that might have never been noticed otherwise. In addition to these stream condition observations, volunteers have also observed the presence of species beyond the benthic macroinvertebrates they intended to study. These observations have included the observation of a native species that was not known to be present in the area, and observations of multiple species invasive to the state of Michigan. These observations have been used to help track the spread of species, especially invasive species.

3.4.4 Spatial Modeling

The SSN and STARS tools were useful in modeling stream water quality in the Rouge River from volunteer science water quality data, and spatial relationships in

stream systems. STARS and SSN tools have been applied to a range of stream modeling applications like surface water isotope variations (McGill et al., 2020), fish genetic diversity in southern France (Paz-Vinas et al., 2018), and fecal contamination in streams in Northeast Scotland (Neill et al., 2018) and central North Carolina (Holcomb et al., 2018). SSN methods have been previously applied with volunteer science data (Kielstra et al., 2019) and macroinvertebrates in streams (Frieden et al., 2014; Pond et al., 2017). This project uniquely combines volunteer science-collected macroinvertebrate data into a spatial model, which together were able to overcome challenges in data paucity and stream connectivity.

Water quality in the Rouge River was modeled with two environmental explanatory variables: imperviousness and sediment regulation. Both variables reflect some degree of anthropogenic activity; and together they show that human behavior affects stream quality through different avenues. Imperviousness is directly related to human populations and densities, where high imperviousness is associated with high human density and is known to cause streamlined pollution conveyance via stormwater and increased flashiness, elevated temperatures, and higher BOD (Blaszczak et al., 2019; Grabowski, Watson, & Chang, 2016; Mallin, Johnson, & Ensign, 2009). The negative imperviousness coefficient modeled here aligns with the emphasis placed on impervious sources as a key driver of water resources impacts in previous research (Arnold & Gibbons, 1996; McGrane, 2016; Salerno, Viviano, & Tartari, 2018). Sediment regulation is estimated through factors directly or indirectly driven by humans, like reservoir presence and volume, stream channelization, riparian vegetation, and agriculture weighted by soil erodibility (Thornbrugh et al., 2018). The positive coefficient associated

with sediment regulation indicates an increase in sensitive benthic macroinvertebrate species associated with high sediment regulation. This relationship was expected as benthic macroinvertebrates thrive in well oxygenated water, with low proportions of fine substrate (Kaller & Hartman, 2004; Von Bertrab, Krein, Stendera, Thielen, & Hering, 2013). The use of imperviousness and sediment regulation helped to build the stream quality SSN model.

Our methodology using an SSN model to explore the relationship between stream quality and poverty builds upon existing analyses of the socioeconomic influence of stream quality. Previous analyses explored regression relationships and spatial clustering between stream environment indicators and variables describing historically disadvantaged populations. These studies found mixed correlation results, revealing negative trends between a stream health index and both household size and poverty (Daneshvar et al., 2016; Sanchez et al., 2014). The strength of correlations between socioeconomic and stream health indices was improved by applying spatial clustering (Sanchez et al., 2015) and tailoring the resolution of spatial analysis (Daneshvar et al., 2016). In general, higher resolution data produced higher correlations (Daneshvar et al., 2016; Sanchez et al., 2015). The method of parameter estimation for environmental justice modeling has also been performed with many explanatory variables categorized as ecological, socioeconomic, and physiological (Daneshvar et al., 2018). This work's methodology avoided the ambiguity associated with correlation calculations and complexity of clustering methods by using both socioeconomic and environmental variables, and a spatial model designed for stream networks. The spatial modeling framework applied in past models was conditional autoregressive modeling, which

considers the spatial influence of neighboring points (Daneshvar et al., 2016; Sanchez et al., 2015, 2014). The modeling approach with SSN expands on this consideration of neighboring points, by including relationships that exist on stream flow paths. While the model identifies weaker statistical relationships than those observed in past models (Sanchez et al., 2015, 2014), the simplicity and interpretability of the SSN model provides a straightforward means of expressing the complex relationship between socioeconomic parameters and urban stream quality. Ultimately, this work aligns with previous environmental justice models, all finding negative relationships between historically underserved groups and water quality via stream health indices.

3.5 Conclusions

Urban stream syndrome remains a prevalent environmental concern, and this work shows how degraded stream water quality disproportionately burdens higher poverty areas in metropolitan Detroit. These results show that under similar environmental conditions, streams in higher poverty areas have lower stream quality. Volunteer science-collected data provided a robust understanding of stream quality in the Rouge River, and spatial modeling methods enabled the incorporation of stream interdependencies in stream quality modeling. In further analyses of the socioeconomic distribution of water quality degradation, partnership with volunteer science groups should be explored, as these groups may have shared interests in understanding water quality in their community and this work can provide fruitful new insights into the dynamic relationship between streams and the environmental and anthropogenic communities through which they run.

These findings highlight that disparate stream conditions exist in the Rouge River watershed in a manner that disproportionally burdens high poverty areas. Remedying this inequitable stream syndrome will require targeted use of water quality management tools to induce stream quality improvements in areas of impairment. Stormwater best management practices (BMPs) are infrastructure able to provide this targeted management. FOTR have already embraced the implementation of BMPs in their mission, supporting community members through stormwater trainings, rain barrel and rain garden design classes and rain barrel construction workshops ("River Restoration," n.d.). However, as presented in the state of the science, urban stormwater BMPs have variable management performance driven in part by the yet unknown influence of environmental factors. Chapters 2 and 3 contribute to the larger picture of improving conditions in urban waterways by investigating the influence of regional climate and storm characteristics on BMP water quality management.

4. EFFECTS OF REGIONAL CLIMATE AND BMP TYPE ON STORMWATER NUTRIENT CONCENTRATIONS IN BMPS: A META-ANALYSIS

4.1 Rationale & Summary

Stormwater BMPs are common infrastructure tools for nonpoint source pollutant treatment, however, their performance is unreliable due to the many factors influencing treatment performance. Climate is one of these factors that may influence nutrient removal processes in stormwater BMPs, but the influence of climate is not yet quantified broadly (Lintern et al 2020). Climate impacts the flood mitigation function of BMPs (Lammers et al., 2022; Sohn et al., 2019; Voter & Loheide, 2021), but its impact on nutrient treatment function is not well understood (Gold et al., 2019; Lintern et al., 2020; Sohn et al., 2019). Previous research has evaluated site-level BMP response to climate-driven factors such as storm characteristics (e.g., storm volume), soil responses (e.g., antecedent moisture content) and seasonality (e.g., freezing patterns and temperature changes) (Buffam et al., 2016; E Daly, Deletic, Hatt, & Fletcher, 2012; Hoss et al., 2016; Lintern et al., 2020; Roseen, Robert et al., 2009; Valenca et al., 2021). This chapter presents predominate patterns in BMP nutrient removal performance under different climate regimes, thus furthering the field of stormwater infrastructure by 1) assessing performance of BMPs on a large scale, and 2) providing information about the impact of regional climate on BMP nutrient treatment performance.

The goal of this chapter was to evaluate the impact of climate on changes in stormwater nutrient concentrations and speciation through three types of soil-based BMPs. Meta-analysis was applied to address high heterogeneity in nutrient concentration changes across storms and sites and to generalize the direction and magnitude of nutrient

concentration changes induced by BMPs in different climates. Large-scale climate was defined by the Budyko dryness index, the ratio of annual potential evapotranspiration to precipitation that quantifies the relative aridity (or humidity) of a site. It was hypothesized that BMP nutrient removal performance was different in wet and dry climates. This hypothesis was tested by analyzing BMP influent and effluent nutrient concentrations with data from the International Stormwater BMP Database. The objectives were 1) to conduct a meta-analysis of data from the BMP database and evaluate changes in N and P species and 2) to compare BMP nutrient transformations between sites in different wet and dry climate regimes. To support these objectives, multiple sources of the climatic patterns observed were contextualized.

4.2 Methods

4.2.1 Data Acquisition

Stormwater event influent and effluent concentrations were obtained from the International Stormwater BMP Database (http://www.bmpdatabase.org). The BMP database was initialized in 1996, and this work applies version 12-29-2019 obtained in 2019. In addition to water quality data, BMP information such as latitude and longitude, type, purpose, and soil media data were retrieved. Data in the International BMP Database is reported through a combination of literature review and self-reporting, such that data quality assurance is beyond the control of research applying the Database.

Data was filtered by geography, BMP type, and pollutant type. Only sites in the contiguous United States were considered. BMP types with N and P concentrations

reported for the influent to effluent were identified in the BMP Database, then narrowed to bioinfiltration, grass strip, and grass swale BMP types because they share common vegetated soil ecosystem features for stormwater treatment (Clary et al., 2020). These three types of BMPs are subsequently referred to as vegetated BMPs. BMP selection was also restricted by the number of available paired influent and effluent concentrations ($n > 2$). Data was restricted to the nutrients N and P (total and dissolved inorganic measures). N species selected from the Database were total N (TN), nitrate (NO_3^-) as N, nitrite (NO_2^-) as N, $NO_2^- + NO_3^-$ as N, NO_x as N (referring to the sum of NO_3^- and NO_2^-), ammonia (NH_3) as NH_3, ammonia as N, and ammonium (NH_4^+) as N. P species selected were total P (TP), orthophosphate as P, orthophosphate as PO_4^{3-}, and dissolved orthophosphate as P. Concentrations of these analytes may not reflect loads, as flow data were sparse in the International BMP Database. This limits interpretation of overall water quality impact, but concentration is an informative measure because it can 1) inform interpretation of BMP treatment processes and 2) have important ecological implications.

N and P species were translated into parameters of interest by converting units to mg/L as N or as P, removing overlaps, and categorizing as: TN, dissolved inorganic N (DIN), TP, dissolved inorganic phosphorus (DIP), DIN:TN, DIP:TP, DIN:DIP and TN:TP. TN and TP were simply all entries of paired influent and effluent TN and TP. DIN was calculated by summing nitrate, nitrite, and ammonium or ammonia that co-occurred on the same date and time at the same BMP. Orthophosphate was considered to be equivalent to DIP. Dissolved inorganic fractions were calculated as DIN divided by TN, and DIP divided by TP (DIN:TN and DIP:TP, respectively). N:P ratios were calculated as DIN divided by DIP (DIN:DIP) to represent a dissolved N:P ratio, and TN

divided by TP as a total N:P ratio (TN:TP). Dissolved inorganic fractions and N:P ratios were calculated by pairing observations at the same site, during the same storm event, for both influent and effluent.

The difference between the means of all influent and effluent observations at a BMP (called the within-group mean difference, MD) was calculated for each analyte at each BMP as

$$MD = \overline{x_{eff}} - \overline{x_{inf}}$$

where $\overline{x_{eff}}$ is the mean of the effluent values and $\overline{x_{inf}}$ is the mean of the effluent values (Harrer, Cuijpers, Furukawa, & Ebert, 2021). The MD for dissolved inorganic ratios and N:P ratios were standardized, and thus labeled as SMD, calculated by

$$SMD = \frac{\overline{x_{eff}} - \overline{x_{inf}}}{s_{inf}}$$

where s_{inf} is the standard deviation of the influent (Harrer et al., 2021). This standardization accounted for differences in scale in the measurement of dissolved inorganic and total N and P, and between N and P. Hedge's g correction was applied to the SMD to correct for systematic overestimation caused by standardization of sites with 20 or fewer paired observations (Harrer et al., 2021). The standard error of each MD and SMD was calculated for within-group MDs as

$$SE_{MD} = \sqrt{\frac{s_{inf}^2 + s_{eff}^2 - (2\, r_{inf,eff}\, s_{inf}\, s_{eff})}{n}}$$

where s_{inf} and s_{eff} are the standard deviations of influent and effluent, respectively; $r_{in\ ,eff}$ is the correlation between the influent and effluent groups at a BMP; and *n* is the

number of paired observations of influent and effluent (Borenstein, Hedges, Higgins, & Rothstein, 2009; Harrer et al., 2021). For SMDs the standard error was calculated as

$$SE_{SMD} = \sqrt{\frac{2(1 - r_{inf,eff})}{n} + \frac{SMD^2}{2n}}$$

(Becker, 1988; Harrer et al., 2021).

4.2.2 Climate Definition and Assignment

Regional climate for each site was classified as either water-limited or energy-limited, as determined with the Budyko dryness index. The Budyko dryness index is an indicator of aridity, calculated as the ratio of annual average potential evapotranspiration to precipitation, i.e., PET/P (Edoardo Daly, Calabrese, Yin, & Porporato, 2019a). A Budyko index greater than 1 indicates a water-limited, or dry climate, and a value less than 1 indicates an energy-limited, or wet climate. Budyko dryness index values were calculated using PET and P values from the Climate Research Unit gridded Time Series (CRU TS) monthly gridded climate dataset with a spatial resolution of 0.5° latitude by 0.5° longitude (Harris, Osborn, Jones, & Lister, 2020). The BMP sites were assigned the Budyko dryness index values of the spatial grids in which they were located.

Correlation tests were applied to parse the contributions of climate and influent concentration on BMP nutrient removal. The influence of climate on stormwater influent concentrations was assessed by comparing influent and effluent concentrations between wet and dry climates. Additionally, the relationship between climate index and influent concentration was measured with Kendall's rank correlation (τ).

4.2.3 Meta Analysis and Subgroup Analysis

A meta-analysis was conducted to determine the effect of BMPs on nutrient concentrations aggregated across sites. In this work, meta-analysis refers to a statistical analysis used to pool quantitative information regarding a treatment across many studies into a single numerical estimate. Agglomeration across studies is done by calculating a treatment effect, called effect size, at each site then using a weighted average to calculate the single estimate of the treatment, called a pooled effect size (Harrer et al., 2021). Here, the effect size is the change in concentration between influent and effluent at a BMP, and the treatments being investigated are vegetated BMPs. Meta-analysis has been previously employed in a stormwater context to study hydrologic responses of low impact development (LID) to storm frequency (Sohn et al., 2019). Effect sizes were calculated as within-group MD or SMD and the standard error of within-group MD (SE_{MD}, SE_{SMD}) because influent and effluent data is not independent, and these metrics account for their dependency (Equations 1-4). For DIN, TN, DIP and TP, a positive effect size indicates higher effluent versus influent concentrations, and a negative effect size indicates lower effluent versus influent concentrations, in units of mg/L. The term leaching described positive effect sizes, and retention was used to describe negative effect sizes. These terms refer only to changes in concentration between influent and effluent and not net mass change, as influent and effluent flow data were not available for all datapoints and thus restrict the size of the dataset. For DIN:TN, DIP:TP, DIN: DIP, and TN:TP, the effect size is unitless and reflects the effect of BMPs on relative speciation of the stormwater from influent to effluent. In general, SMD = 0.2, 0.5, and 0.8 indicate small, moderate, and high effect, respectively (Cohen, 1988; Harrer et al., 2021).

The pooled effect size describes the overall difference between influent and effluent across all *k* number of BMPs in the study. Effect sizes were pooled with a random effects model because high heterogeneity was expected due to the multiple known sources of variance between BMPs. For example, engineered soil media types, drainage area ratios, drainage area land use, vegetation type and health, influent concentrations, maintenance, and inlet/forebay pretreatment are a few attributes that may vary among BMPs classified within the same type of BMP. The random effects model assumes that these real (but difficult to simultaneously summarize) differences between BMPs also contribute to the overall effect (Harrer et al., 2021). The variance associated with the random effects model (τ^2) was estimated with the Paule-Mendel method. It was selected as an alternative to the common Restricted Maximum Likelihood estimator, which can be biased when the number of studies is small and bias is high (Harrer et al., 2021; Jackson, Angeliki, Martin, Andrea, & Baker, 2017; Veroniki et al., 2016). The Knapp-Hartung adjustment of τ^2 was applied to correct for false positives within highly heterogeneous data (Harrer et al., 2021; Inthout, Ioannidis, & Borm, 2014). Pooled effect was calculated using a weighted average, where the weight was an inverse variance weight for random-effect models (Harrer et al., 2021). The weight of an individual site (w_k) was calculated as

$$w_k = \frac{1}{s_k^2 + \tau^2}$$

where s_k is the variance at a site (SE_{MD} or SE_{SMD}) and τ^2 is the variance associated with the random effects model. The weighting factor reduces the influence of sites with a low number of paired influent and effluent observations (n), and low precision (high variance).

In addition to reporting the pooled effect size for each analyte, additional statistics were calculated to contextualize the result. A 95% confidence interval (95% CI) for the pooled effect size, and p value of the effect size are calculated. A measure of the variance in pooled effect sizes was reported as Higgins and Thompson's I^2 statistic of between-study heterogeneity, where I^2 = 25, 50, and 75% indicate low, medium, and high heterogeneity, respectively (Higgins & Thompson, 2002). These I^2 values measure the percentage of variability caused by true difference in effect between studies (not sampling error).

The impact of climate on nutrient retention in BMPs was measured using subgroup analysis. Subgroup analysis investigates heterogeneity in meta-analysis results by testing whether observational differences between subgroups align with a scientific hypothesis. Subgroup analysis provides statistical evidence (in the form of an omnibus Q-test between subgroups (Cochran, 1954; Harrer et al., 2021)) to accept or reject a hypothesis that questions if two groups are distinct. The hypothesis was tested with a subgroup analysis between climate subgroups (wet and dry) and among BMP types (bioretention, grass strip, grass swale) to investigate if unequal weighting of BMP types coincided with climate groups to create a confounding effect. The subgroup analysis used the same statistical parameter as the initial meta-analysis but was expanded to include a fixed effects model. Pooled effect sizes within subgroups were calculated identically to meta-analysis pooled effect size, then the pooled effect sizes between subgroups were compared (Borenstein & Higgins, 2013) and results between climate and BMP type subgroups were analyzed. Subgroup analysis was applied to all eight analytes described above. A visual overview of the approach is provided in Figure 11.

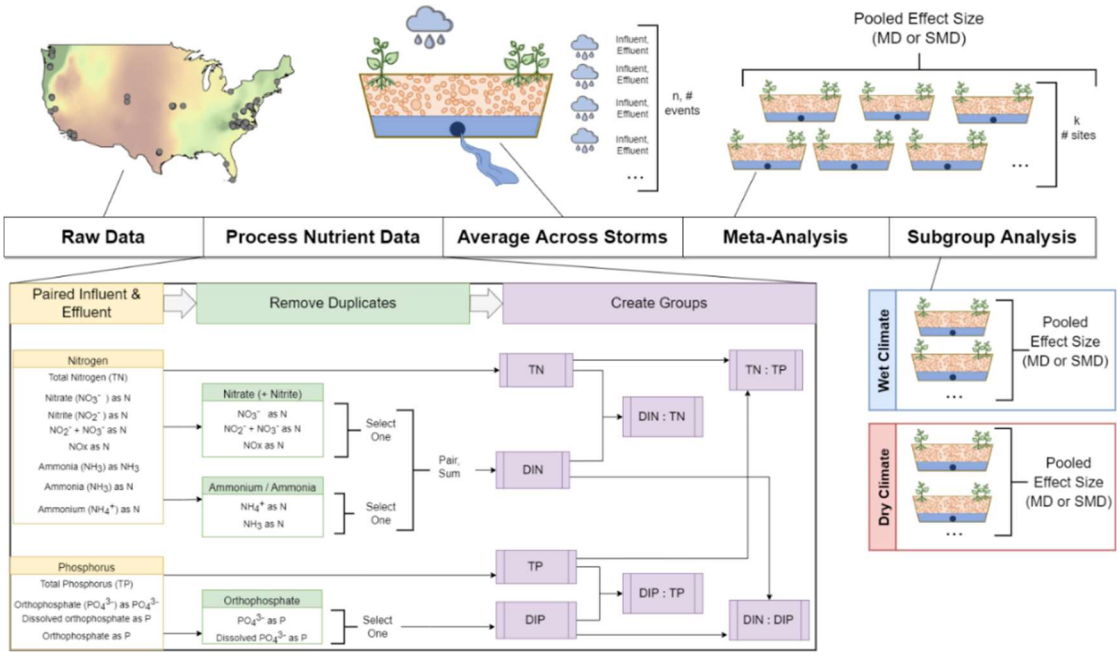


Figure 11: Process used to compile and analyze water quality data from the International Stormwater BMP Database. Paired influent and effluent data from the International Stormwater BMP Database were obtained and analytes were grouped and summed as necessary to calculate DIN and DIP. Then dissolved to total ratios and N:P ratios were calculated. Influent and effluent concentrations of each analyte were averaged per site, and the standard error and number of storms were recorded. Meta-analysis was then conducted to pool influent and effluent averages across BMPs, and finally meta-analysis was repeated for wet and dry climate subgroups of BMPs.

4.3 Results

4.3.1 Data Composition

The BMP data was temporally and spatially diverse, covering 91 sites in 15 states, from 1982 to 2018. Observations were most dense in the mid-Atlantic, where 27 sites were in North Carolina alone (Figure 12). However, Budyko dryness indices were diverse even within state boundaries (i.e., range 0.77 – 1.0 in North Carolina). There was a close climatic distribution of sites, where 42 sites were in dry climates, and 49 sites were in wet climates (Appendix figure A1). This provides a suitable distribution of data between

climate groups for subgroup analysis. There were 2692 paired influent and effluent observations (n) in total. Influent and effluent concentrations for TN and TP were of similar magnitudes but with slightly more conservative quartile ranges and median values compared to the International BMP Database 2020 summary (Table 2).

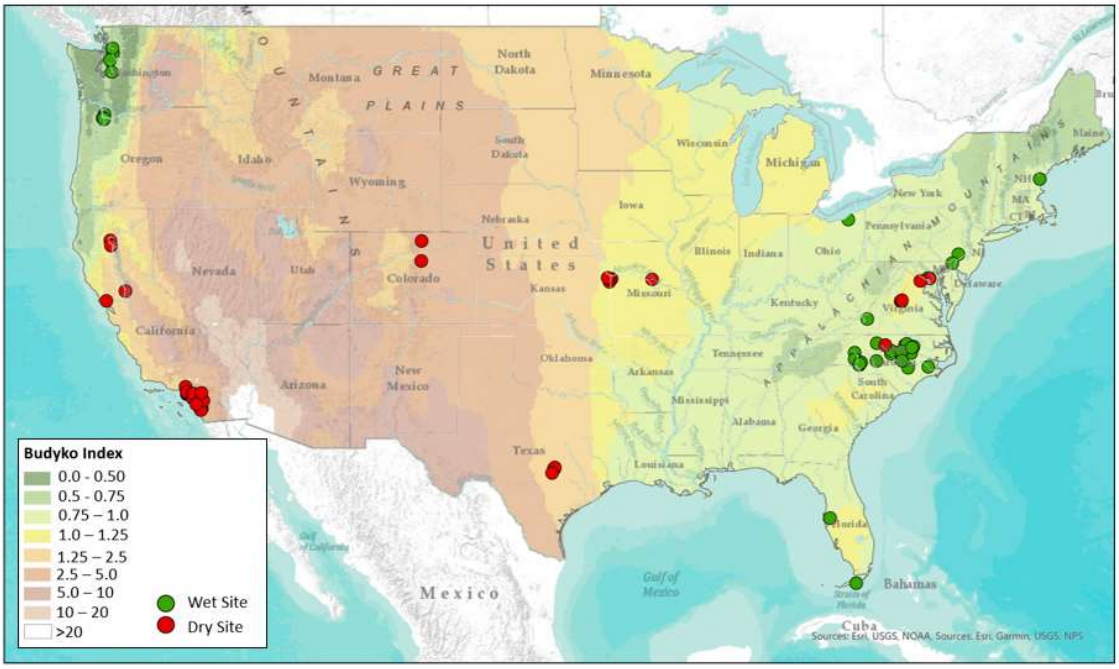


Figure 12: Map of Budyko Dryness index in the contiguous United States with all BMP sites used in the meta-analysis plotted by assigning them into the wet or dry climate grouping.

Table 2: Summary of data composition as influent and effluent. Counts of paired observations (n) and number of BMPs (k) are provided for DIN, TN, DIP, and TP. The ranges of nutrient observations in influent and effluent are provided for the interquartile range (25^{th} to 7^{th} percentile), and the median is provided with a 95% confidence interval. Asterisks indicate data from the 2020 International BMP Database Summary Report, reflecting data for grass swales, grass strips, and bioretention (Clary et al. 2020).

Analyte	# Paired Observations (n)	# BMPs (k)	Interquartile Range		Median (95% CI)	
			Influent (mg/L)	Effluent (mg/L)	Influent (mg/L)	Effluent (mg/L)
DIN	537	40	0.30 – 1.11	0.20 – 0.93	0.54 (0.49 – 0.60)	0.41 (0.36 – 0.47)
TN	522	37	0.60 – 1.83	0.56 – 1.67	1.0 (0.92 – 1.11)	0.96 (0.56 – 1.67)
DIP	543	45	0.01 – 0.10	0.03 – 0.33	0.04 (0.03 – 0.04)	0.10 (0.08 – 0.12)
TP	1090	86	0.09 – 0.30	0.11 – 0.48	0.17 (0.15 – 0.18)	0.22 (0.21 – 0.23)
TN*	913 (influent) 937 (effluent)	51 (influent) 54 (effluent)	0.70 – 1.95	0.60 – 1.64	1.15	0.95
TP*	2317 (influent) 2004 (effluent)	131 (influent) 133 (effluent)	0.08-0.34	0.10 – 0.44	0.17	0.22

4.3.2 Meta-Analysis Results

Meta-analysis revealed varying degrees of change in nutrient concentrations between influent and effluent for each nutrient species, as well as for nutrient ratios. All eight nutrient types examined had moderate to high heterogeneity among the BMP practices analyzed, indicating that the majority of the variation in effect sizes was due to between-site heterogeneity, such that meta-analysis was an appropriate tool (Higgins & Thompson, 2002).

Meta-analysis showed that BMPs are likely to leach DIP and TP and are likely to either leach or retain DIN and TN (Figure 13). The DIP and TP effect sizes were positive (indicating leaching) and significantly different from zero ($p < 0.01$). The DIP leaching

effect was larger than the TP leaching effect (Table 3). In contrast, both DIN and TN effect sizes were negative and not significantly different from zero (p = 0.65, 0.75, respectively), indicating a weak tendency toward retention. The pooled retention effect for DIN and TN were similar in magnitude, and the confidence interval for both analytes indicates both net leaching and retention are likely for different sites (Table 3).

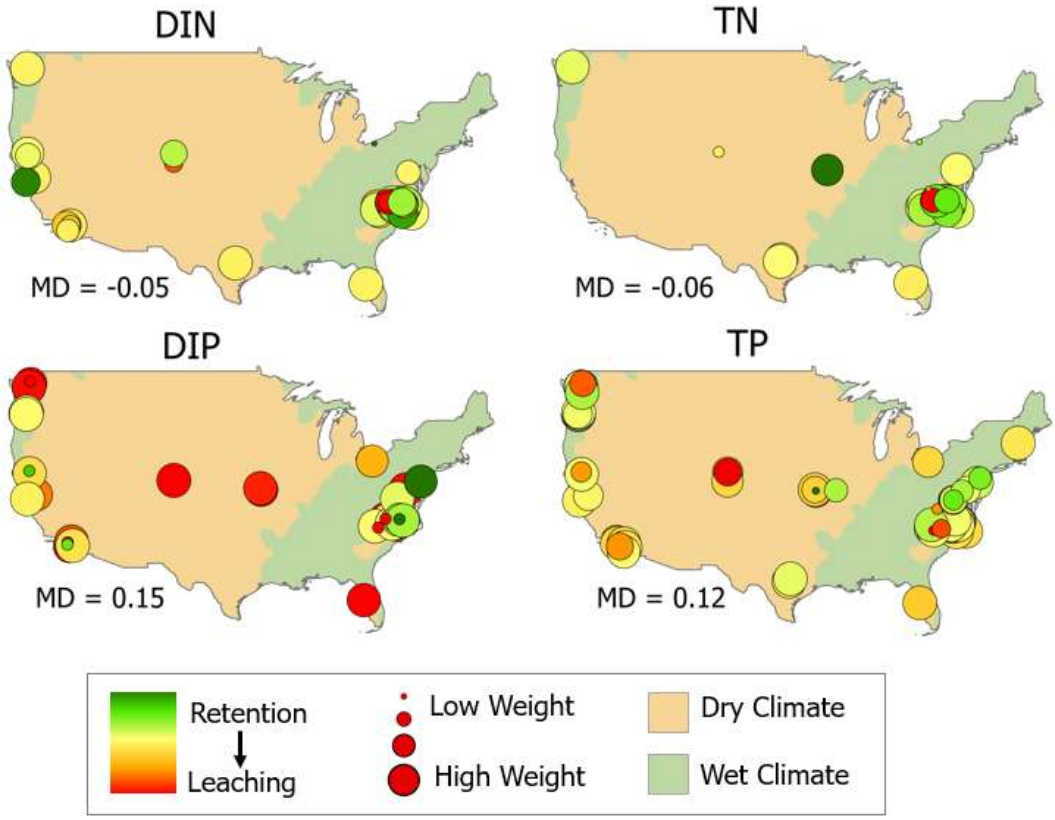


Figure 13: Pooled nutrient effect sizes (g) and site-level effect sizes for (a) DIN, (b) TN, (c) DIP and (d) TP, ranging from net retention to leaching (green to red). Sites with larger weights in the pooled effect size calculation are shown with larger markers, while lower weights are shown with smaller marker sizes.

Table 3: Meta-analysis and subgroup meta-analysis results for change in DIN, TN, DIP and TP through BMPs. The nutrient retention performance was estimated as the inverse variance weighted mean difference within BMP influent and effluent concentrations with a random effects model. The number of paired observations (n), sites (k), effect size (MD), 95% confidence interval of MD (95% CI), likelihood of a MD result being due to chance (p, MD), likelihood that the difference between two subgroups is due to chance (p, between subgroups), and heterogeneity (I^2) are listed for each constituent combined and separated into climate subgroups. NA indicates data no available because one subgroup composes 80% or more of the total data.

		n	k	MD	95% CI	p, MD	p, between	I^2
DIN	Combined	537	40	-0.05	-0.27 to 0.17	0.65		89.6%
	Wet	391	27	-0.04	-0.34 to 0.26	0.79	0.89	91.4%
	Dry	146	13	-0.07	-0.43 to 0.29	0.68		81.9%
	Grass Strip	164	14	-0.10	-0.44 to 0.25	0.56	0.20	76.7%
	Bioretention	173	17	0.16	-0.28 to 0.60	0.45		88.6%
	Grass Swale	200	9	-0.29	-0.61 to 0.04	0.08		95.5%
TN	Combined	522	37	-0.06	-0.41 to 0.30	0.75		84.6%
	Wet	429	31	NA	NA	NA	NA	81.2%
	Dry	93	6	NA	NA	NA		93.2%
	Grass Strip	84	7	0.32	-1.52 to 2.17	0.68	0.80	62.4 %
	Bioretention	205	19	0.03	-0.68 to 0.73	0.94		86.9 %
	Grass Swale	233	11	-0.11	-0.26 to 0.05	0.17		79.6 %
DIP	Combined	543	45	0.15	0.06 to 0.23	<0.01		96.2%
	Wet	318	25	0.06	-0.01 to 0.13	0.07	0.03	91.3%
	Dry	225	20	0.24	0.09 to 0.40	<0.01		97.8%
	Grass Strip	184	17	0.12	0.04 to 0.19	<0.01	0.01	95.1%
	Bioretention	213	19	0.22	0.01 to 0.42	0.04		97.5%
	Grass Swale	146	9	0.03	-0.01 to 0.06	0.12		88.3%
TP	Combined	1090	86	0.12	0.05 to 0.19	<0.01		91.7 %
	Wet	575	46	0.03	-0.02 to 0.08	0.21	0.01	86.9 %
	Dry	515	40	0.21	0.08 to 0.35	<0.01		91.7 %
	Grass Strip	307	26	0.12	0.03 to 0.22	0.01	0.79	90.9 %
	Bioretention	436	33	0.13	-0.02 to 0.28	0.10		92.0 %
	Grass Swale	347	27	0.08	-0.02 to 0.18	0.11		89.8 %

BMPs increased the ratio of DIP:TP and had no statistically meaningful change on DIN:TN in aggregate (Table 3). BMPs had a strong positive pooled effect on DIP:TP ($p < 0.01$), indicating that DIP comprised a higher fraction of TP in effluent than influent

(Table 3). Heterogeneity was high among BMPs measuring DIP:TP change, indicating a relatively wide range of effect sizes that BMPs had on DIP:TP ratios.

BMPs decreased DIN:DIP, but did not significantly affect TN:TP ratios, indicating higher dissolved inorganic P relative to N in effluent than influent ($p < 0.01$ Table 3). The magnitude of the DIN:DIP decrease (SMD = -0.44) was small to moderate (Table 3).

4.3.3 Climate Effect on Stormwater Influent

Stormwater influent in arid climates was higher in DIN ($p \leq 0.01$), DIP ($p = 0.03$), and TP ($p = 0.02$) than influent in wet climates (Figure 14). The majority (77 to 91%) of the lowest quantile of influent concentrations were observed in wet climates. A majority (59-70%) of the highest quantile of influent concentrations were observed in dry climates for DIN, TP, and DIP. However, TN had the opposite trend where only 30% of the highest quantile influent concentrations were observed in dry climates. Correlations between standardized Budyko dryness indices and influent concentrations suggested that these parameters are not strongly related (Kendall's $\tau = 0.18, 0.37, 0.11, 0.27$ for TN, DIN, TP, and DIP, respectively).

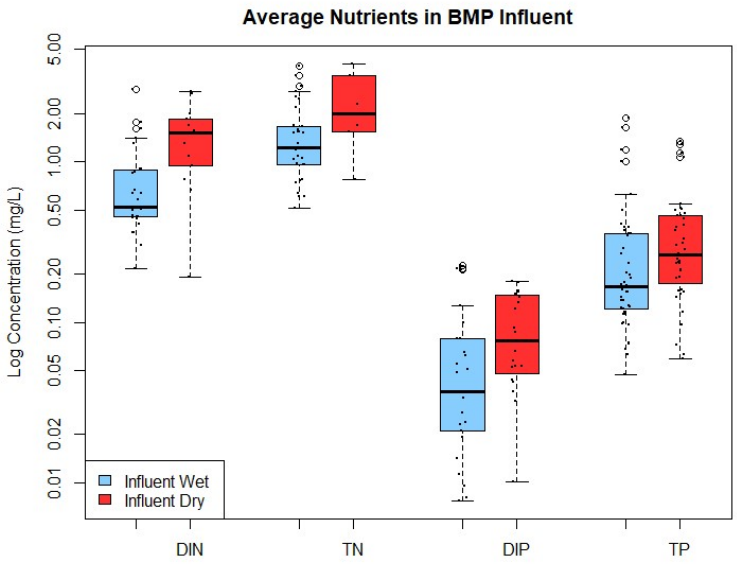


Figure 14: Average influent concentrations per BMP in wet and dry climates for DIN (27 wet, 13 dry), TN (31 wet, 6 dry), DIP (25 wet, 20 dry), and TP (46 wet, 40 dry).

4.3.4 Climate and BMP Type Effect on BMP nutrients

Climate had a strong effect on TP and DIP retention in BMPs, inconclusive results for TN, and no observed effect on DIN. Climate classifications were composed by unbalanced BMP types for dry climate DIN, TN, and DIP (Figure 15).

BMPs in dry climates had higher DIP and TP leaching than BMPs in wet climates ($p = 0.03$ and $p = 0.01$, respectively, Table 3). The positive MD for DIP and TP indicated leaching in both wet and dry climates (Figure 16a). The climate effect on TN retention was inconclusive due to the low number of sites with TN data in dry climates. DIN effect sizes spanned the positive and negative ranges in both wet and dry climates, with a higher tendency of leaching in dry climates. However, this difference was not distinguishable from a difference caused by chance ($p = 0.89$). In summary, N retention differences were not observed, but DIP and TP both leached to a higher degree in dry climates than wet

climates. Visual summaries of subgroup analysis are provided with forest plots in the Appendix (Appendix Figures B1.1-B1.4).

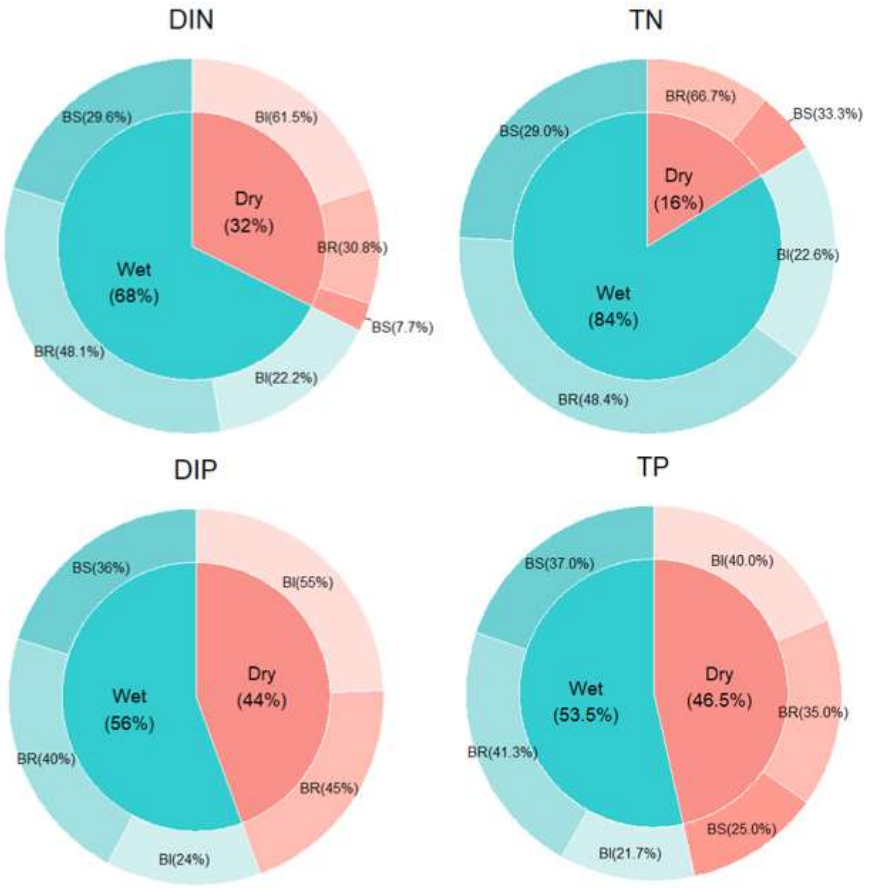


Figure 15: Distribution of the three BMP types reporting DIN, TN, DIP, or TP: grass strips (BI), bioretention (BR), and grass swales (BS), between wet and dry climates.

BMP types had distinct effect sizes for DIP removal, but not for DIN, TN, or TP. All three BMP types had a general leaching effect on DIP, but this effect was the most severe for bioretention with a MD of 0.22 mg/L leached DIP. Grass strips leached the next most (MD = 0.12 m/L), followed by grass swales (0.03 mg/L). Notably, grass swales were the least observed BMP type for DIP (k = 9) and all observations were in wet

climates. The dry climate group did not contain balanced composition of BMP types for DIN (underrepresented grass swales), TN (underrepresented grass strips) and DIN (underrepresented grass swales) (Figure 15).

Table 4: Meta-analysis and subgroup meta-analysis results for change in DIN:TN, DIP:TP, DIN:DIP, and TN:TP through BMPs. The number of paired observations (n), sites (k), effect size (SMD), 95% confidence interval of SMD (95% CI), likelihood of a SMD result being due to chance (p, SMD), likelihood that the difference between two subgroups is due to chance (p, between subgroups), and heterogeneity (I^2) are listed for each constituent combined and separated into climate subgroups. NA indicates data no available because one subgroup composes 80% or more of the total data.

	Climate	n	k	SMD	95% CI	p, SMD	p, between subgroup	I^2
DIN:TN	Combined	426	29	-0.18	-0.69 to 0.32	0.47		84.7 %
	Wet	386	27	NA	NA	NA	NA	85.0 %
	Dry	40	2	NA	NA	NA		0 %
	Grass Strip	76	6	-0.70	-2.59 to 1.18	0.38	0.03	86.8 %
	Bioretention	154	14	0.39	-0.21 to 0.99	0.18		75.6 %
	Grass Swale	196	9	-0.78	-1.70 to -0.78	0.09		88.5 %
DIP:TP	Combined	541	45	0.67	0.38 to 0.97	<0.01		81.5%
	Wet	331	26	0.59	0.21 to 0.97	<0.01	0.49	71.5 %
	Dry	210	19	0.80	0.29 to 1.32	<0.01		86.7 %
	Grass Strip	172	17	0.38	0.10 to 0.67	0.01	0.10	68.4 %
	Bioretention	208	18	1.26	0.45 to 2.07	<0.01		83.8 %
	Grass Swale	161	10	0.42	0.14 to 0.71	0.01		55.9 %
DIN:DIP	Combined	285	25	-0.44	-0.74 to -0.15	<0.01		64.3 %
	Wet	227	16	-0.38	-0.85 to 0.08	0.10	0.59	75.8 %
	Dry	58	9	-0.52	-0.77 to -0.26	<0.01		0
	Grass Strip	78	10	-0.13	-0.72 to 0.47	0.64	0.01	72.4 %
	Bioretention	108	11	-0.44	-0.79 to -0.08	0.02		32.0%
	Grass Swale	99	4	-0.98	-1.59 to -0.38	0.01		50.9%
TN:TP	Combined	484	36	-0.15	-0.59 to 0.29	0.50		84.3 %
	Wet	411	30	NA	NA	NA	NA	84.2 %
	Dry	73	6	NA	NA	NA		84.1 %
	Grass Strip	84	7	-0.13	-1.01 to 0.74	0.73	0.15	80.4 %
	Bioretention	204	19	0.15	-0.71 to 1.01	0.71		85.7 %
	Grass Swale	196	10	-0.56	-0.84 to -0.29	<0.01		55.9 %

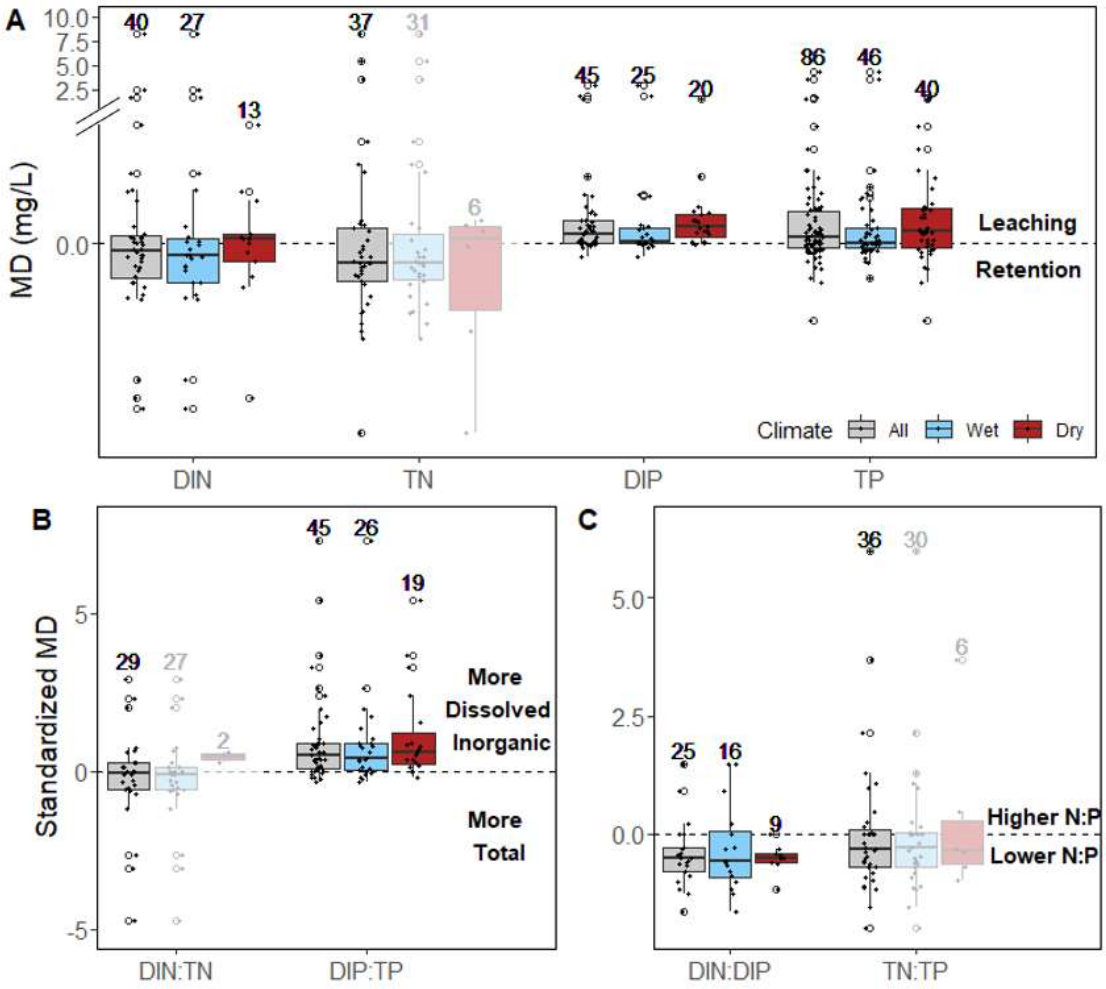


Figure 16: Effect of BMPs on a) retention of nutrients (DIN, TN, DIP, and TP), b) dissolved fraction (DIN:TN and DIP:TP), and c) N:P ratio (DIN:DIP and TN:TP) between influent and effluent. Effect of climate on nutrient retention effect size is shown with clustered boxplots of all sites (grey), wet climate sites (blue), and dry climate sites (red), grouped as defined by Budyko climate definitions. The number of sites per group (k) is labeled above each boxplot. MD and SMD are thresholds for direction of change between effect > 0 and effect < 0. Positive and negative values correspond with (a) leaching and retention, (b) more and less dissolved N or P in effluent than influent, and (c) higher or lower N:P ratio in effluent than influent, respectively. Translucent boxes indicate limitations of analysis because one subgroup composes 80% or more of the total data.

DIP:TP increased through BMPs in both wet and dry climates (Figure 16b). The wet and dry subgroup SMDs were comparable to the SMD for the two climate regimes

combined and there was no difference in the effect of P speciation change between wet and dry climates (

Table ***4***). Therefore, climate was not detected as a driver of the shift of DIP:TP through BMPs. A similar result was observed for BMP type, where all three BMP types favored more DIP in effluent than influent. This shift was highest for bioretention, and moderate for grass swales and grass strips. It is more likely that DIP:TP differences align with BMP type ($p = 0.10$) than climate groups ($p = 0.49$).

The effect size of DIN:DIP ratio change in wet and dry climates were both negative, indicating lower DIN:DIP ratios in effluent (Figure 16c). The effect size for wet climates was slightly lower in magnitude than in dry climates, but this difference may be due to chance ($p = 0.59$,

Table ***4***). However, there is a clear difference in DIN:DIP change among BMP types. All three BMP types shift towards lower DIN:DIP ratios, but this change is strongest in grass swales (followed by bioretention, then grass strips). Grass swales and bioretention indicate DIN:DIP decreases for the whole 95% CI, but grass strips encompass DIN:DIP ratio increase as well as decrease, with higher heterogeneity and ambiguity in the MD, indicated by the higher I^2 and high p value, respectively.

There was no conclusive shift in DIN:TN (Figure 16b) and TN:TP (Figure 16c) between influent and effluent. There were insufficient observations of both DIN and TN, or both TN and TP in dry sites for conclusive results ($k < 5$) (

Table *4*). Both DIN:TN and TN:TP analyses yielded noticeable differences in MDs among BMP types, particularly for DIN:TN. Despite the measured differences among BMP types, the MDs of the BMP types themselves have low confidence, obscuring the interpretation of these effect sizes as truly different from one another. For example, DIN:TN is different among BMP types ($p = 0.03$), but the effect sizes for the BMP groups themselves have low confidence ($p = 0.38$, 0.18, 0.09 for grass strips, bioretention and grass swale, respectively) so the influence of BMP type is uncertain.

N speciation transformations from influent to effluent did not vary by climate regime. Broadly, DIN composition shifted towards nitrate (+ nitrite) as a larger component of DIN in effluent (Figure 17). In both wet and dry climates, influent ammonia/ammonium constituted just under half of the DIN concentration (46% and 48% in wet and dry climates, respectively). In effluent, ammonia/ammonium composed about 36% of DIN in both wet and dry climates. The pooled effect size across both climate groups indicated that BMPs shift DIN towards higher nitrate (+nitrite) composition in effluent ($p = 0.01$). This effect was present in both wet ($p = 0.06$) and dry ($p = 0.06$) climates at similar magnitudes (SMD = -0.36, -0.46, wet and dry, respectively). Thus, there was no meaningful distinction in DIN composition shift between wet and dry climates ($p = 0.75$).

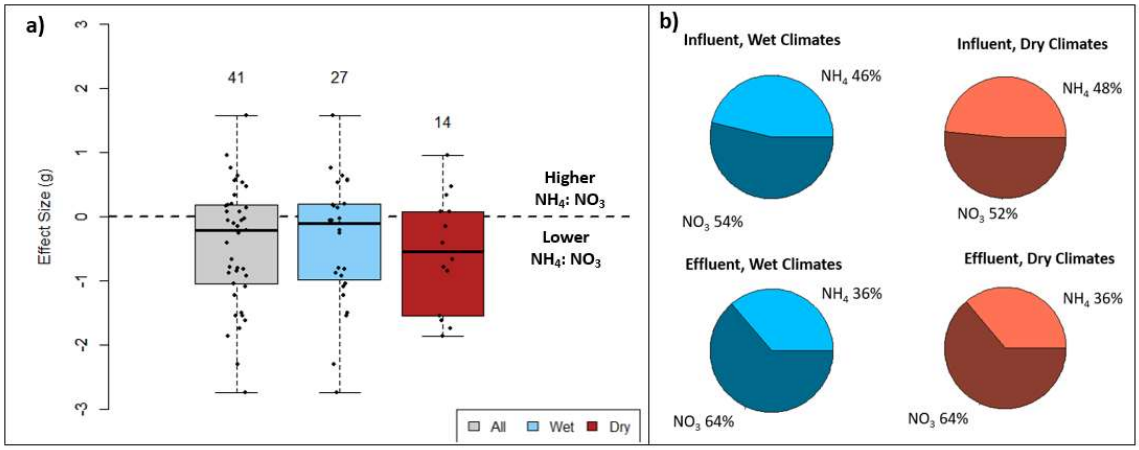


Figure 17: Shift in DIN composition from influent to effluent in wet and dry climates illustrated by (a) effect sizes, g, as box plots and (b) pie charts. NH_4 indicates either NH_4^+ or NH_3 and NO_3^- is the sum of NO_3^- and NO_2^-, condensed for simplicity as NH_4^+ and NO_3^- are the dominate species of their respective DIN subgroups.

4.4 Discussion

4.4.1 Drivers of Climate-Varying Performance

BMPs leach DIP and TP more frequently in arid climates than in wet climates. Meta-analysis revealed this observational trend but cannot attribute a definitive cause to this difference. The pattern of arid BMP DIP and TP leaching aligns with distinct patterns in arid climate P cycling, a direct driver related to climate–driven biogeochemistry. The results also align with indirect explanations, including elevated stormwater nutrient influent concentrations in dry climates, BMP type differences, and regional stormwater design goals. Each of these respective explanatory structures are explained in detail in sections 4.4.1.1 – 4.4.1.4.

4.4.1.1 Arid Climate Biogeochemistry

In arid climates, biogeochemical cycling is driven by low soil moisture and long inter-storm periods (Austin et al., 2004). For natural soils of similar composition and

geography, this arid soil hydrology and climatology induce different cycling of N and P, ultimately leading to higher availability and leaching of both N and P from arid soils. Results indicate higher leaching of DIP and TP from arid soils; therefore, patterns of leaching P species may be due to this arid climate biogeochemistry.

Natural soils have higher availability of inorganic P in lower precipitation regimes (Austin & Vitousek, 1998; Emadi et al., 2012; Ippolito et al., 2010). Soil P availability depends on whether aridity is temperature- or precipitation-driven (Hou et al., 2018). Temperature is negatively related to soil P availability, where higher temperatures are associated with more weathered soils and plant uptake (Dixon et al., 2016; Hou et al., 2018). The relationship between precipitation and available soil P is more complex, where higher precipitation is associated with lower available P and mineral P, but higher organic P (Hou et al., 2018). Attributing a cause to this pattern has proven difficult, where weathering and plant uptake (Austin & Vitousek, 1998; Hou et al., 2018; Ippolito et al., 2010), presence of different P-binding soil compounds (i.e., calcium, iron, aluminum, manganese, or the oxides of these cations) (Buckingham et al., 2010; Emadi et al., 2012), and biological processes influencing organic transformations (Hou et al., 2018; Siebers, Sumann, Kaiser, & Amelung, 2017) have all been hypothesized as causes independently and in combination.

Soil P mineralization and cation compound binding in BMP soils were not explicitly explored here; however, parallels may be drawn between these processes and BMP performance. Increased P leaching has been observed for BMPs using media containing high P (Hatt, Fletcher, & Deletic, 2009; Hunt, Jarrett, Smith, & Sharkey, 2006; L. McPhillips, Goodale, & Walter, 2018), especially compost, which can act as a P

source in BMPs (Hurley et al., 2017; Kranz, Heitman, Rivers, & Mclaughlin, 2022). Thus, if BMPs have heightened P content in soil media, and temperature-driven arid climates are known to leach P in native soils, then it is foreseeable that BMPs in temperature-stressed arid climates that also use high P-containing engineered media may export P at higher rates. On the other hand, the sorption of P on cations, such as iron, has been found to enhance P retention in BMPs (Erickson et al., 2012, 2007). Therefore, leaching of DIP in arid climates can be explained by either lack of P-binding materials or the saturation of these materials in engineered soil media. None of the BMPs reporting engineered soil media in this study included details about the presence of enhanced P–binding soil media, so this explanation could not be pursued further. Overall, the potential drivers of higher arid climate P leaching, mineralization and varying cation compound availability, are both explanations that align with existing knowledge of BMP processes and support the general trend of P leaching observed in arid climates.

4.4.1.2 Influent Concentrations

In addition to the biogeochemical precedent for elevated DIP and TP leaching in dry climates, influent concentration may have a separate or compounding role in BMP performance. Both nutrient forms that leached more in arid climates (DIP and TP) also have higher influent concentrations than wet climate BMPs (Figure 3). Elevated stormwater influent concentrations in dry climates may be a co-occurring driver with climate or a derivative of climate. Influent concentration has been shown to drive BMP removal efficiency (Barrett, 2008). This suggests that influent concentration is an underlying cause of different BMP removal in different climates. However, the low correlations between climate and influent concentration in this study indicate that influent

alone cannot fully explain the observation of varying BMP removal in different climates. In fact, the tendency for dry climates to have higher influent concentrations may be a product of the arid-climate biogeochemical drivers influencing stormwater influent concentrations. Thus, the observed effect of DIP and TP leaching in dry climates may be due to the compounding effects of poor dry climate nutrient retention in BMP soils and elevated influent concentrations in arid climates.

4.4.1.3 BMP Type

The attributes of individual BMPs may vary greatly, and the cumulative influence of these attributes is likely to cause variance in overall nutrient treatment performance. By grouping BMPs by type, some of these attributes may be considered en masse. For example, rather than examining underdrain size, vegetation species, or longitudinal slope as independent variables, examining BMP type groups together the dozens of specific design attributes into similar groups based on the infrastructure goal and typical design. The BMP Database defines grass swales as shallow and vegetated, grass strips as vegetated areas designed for lateral flow conveyance, and bioretention as shallow vegetated basins with an underdrain (Clary et al., 2020).

Difference in performance by BMP type is evident in DIP data from both the International BMP Database and in this dataset. In the 2020 database summary, difference in median of all recorded (unpaired) influent and effluent DIP concentrations was greatest for bioretention (leached 0.24 mg/L), followed by grass swales (0.07 mg/L), then grass strips (0.06 mg/L) (Clary et al., 2020). In this analysis, bioretention also leached the most DIP on average (0.22 mg/L), although relative leaching from grass strips and swales were flipped compared to the Database report. This observation of DIP

leaching from bioretention is also evident in the effect size changes for nutrient ratios involving DIP, where DIP:TN and DIN:DIP indicate shifts toward more DIP in effluent being strongest for bioretention. Decreased bioretention performance aligns with design attributes of bioretention such as soil media high in P and high potential for biomass (vegetation) decomposition, both of which are associated with variable or poor nutrient removal (Hunt et al., 2006; Muerdter, Wong, & LeFevre, 2018; Shrestha et al., 2018; Skorobogatov, He, Chu, Valeo, & Duin, 2020). Design details in this study provide support that the engineered soil media selection may fall along either divisions by BMP type or climate, as ten studies report a detailed account of soil media selection, eight of which contained high organic media (compost, loamy fill, organics, or peat). Of these eight, six were bioretention and two were grass strips, where all bioretention and one of the strips were in wet climates (seven wet, one dry).

The influence of BMP type on removal performance may confound the climate observations because of the coinciding occurrence of a BMP type in a single climate regime. This is the case for the observation of higher DIP leaching in dry climates. Only grass strips and bioretention BMPs observed DIP leaching in dry climates (Figure 15). Grass swales had the lowest DIP removal among the three BMP types. Therefore, the overlapping BMP type and climate group observations preclude distinguishing whether the absence of high performing grass swale BMPs from the dry climate groups overinflates their leaching or if dry climates drive reduced DIP removal performance and the performance of grass swales is overinflated by the absence of dry climate swales.

4.4.1.4 Regional BMP Design Goals

The observed trends in performance by climate may also reflect human priorities and designs. Different regions may have different design standards, stormwater regulations, and performance goals adjusted to the needs of their region. For example, in New Mexico's Middle Rio Grande Municipal Separate Storm Sewer System (MS4) permit, standards require that storms up to the 90th percentile must be retained for newly developed sites and up to 80th percentile for reconstructed sites (Environmental Protecion Agency, 2016). In contrast, Vermont requires both volume management for 90% of storms, and water quality treatment of 80% total suspended solids (TSS) and 60% TP (Environmental Protecion Agency, 2016). Thus, stormwater infrastructure practices in New Mexico, a state entirely classified as dry climate in this work, may focus on volume retention to meet this regulation whereas Vermont, a state classified entirely as wet, may design BMPs with considerations for both volume retention and water quality treatment. However, it is notable that there are many exceptions to this trend with states requiring both volumetric water retention and water quality treatment in majority wet (New York, Massachusetts, Delaware, Connecticut, New Jersey) and dry (Wisconsin, Minnesota, Colorado) climates (Environmental Protecion Agency, 2016). In this dataset, ~45% (41) of BMPs specified a design goal or purpose, 11 of which only explicitly mention water quality goals, and 30 of which directly mention water quantity as well as quality treatment goals. Of these 41 BMPs, 15 were in dry climates, indicating that some dry climate BMPs in this study specified water quality treatment goals.

4.4.2 Meta-Analysis Application

The application of meta-analysis for investigating BMP performance allows for the compilation of data from many BMPs with a method commonly used in many scientific fields. There are limitations to meta-analysis which pertain to the interpretation of this analysis. These limitations are provided in section 4.4.2.1. Despite limitations, this methodology should be adopted by other researchers studying BMP treatment performance. Recommendations for future applications of meta-analysis for the study of BMPs are provided in section 4.4.2.2.

4.4.2.1 Interpretations of Meta-Analysis Limitations

First, meta-analysis quality is limited by the quality of the data being pooled, and data quality may be impacted by publishing bias and geographic limitations. International BMP Database entries can be self-reported, so there is potential for both positive and negative bias from researchers submitting data by dampening the true severity of leaching or overstating retention capacity. While there is no evidence to show unpublished data in the International BMP Database is less reliable than published data, research bias is possible in both published and unpublished work. In this study, 69 of the 91 total individual sites included data that was published either in a report or journal article. Geographic distribution of data limited the ability to execute analysis of TN (or TN containing ratios) in dry climates.

Second, heterogeneity is calculated, but not attributed to a cause. A pillar of meta-analysis is that a true effect size exists across studies, and that the measured effect deviates from the true effect size due to heterogeneity. In the context of BMP water quality treatment this means that a BMP would be expected to have a consistent treatment

effect on influent, and that any deviation from this true effect must be due to factors that impact the treatment mechanisms in BMPs. Meta-analysis quantifies the amount of deviation that exists (via I^2) but it does not give insight into the sources of that variability (Higgins & Thompson, 2002). To ascribe causation of variability, hypotheses can be formed based on the attributes of sites which have a logical foundation to be a driver of variable performance. Subgroup analysis can then test hypotheses, and rejection or failed rejection of the hypotheses can provide support that the attribute investigated caused variability in the treatment effect. It is therefore important to clarify that BMP treatment is different in wet and dry climates for DIP and TP and that treatment is different among grass swale, grass strip, and bioretention BMP types for DIP, DIN:TN, and DIN: DIP, but meta-analysis cannot conclude that these differences are caused by climate or BMP type. Any causational explanations may only be supported by meta-analysis results, not proven by them.

4.4.2.2 Recommendations for Future Meta-Analysis Applications for BMPs

Meta-analysis is an effective tool to analyze overall performance of BMPs. Meta-analysis allows for the compilation of multiple observations relating to an overarching research question (Harrer et al., 2021). Three major benefits of meta-analysis in a stormwater BMP application are 1) use for finding a pooled effect over many studies 2) encompassing between-study heterogeneity in the calculation of an effect size, and 3) simplicity of use.

First, there is extensive documentation of BMP-level performance in the International BMP Database (Clary et al., 2020). Data submission instructions and data storage formatting ensure clear record of relevant data, including the influent and effluent

concentrations necessary for meta-analysis of removal performance. It is useful to estimate overall treatment effects of BMPs, and additional information logged in the database like design, age, and maintenance can help analyze factors that may drive differences in a true effect between BMPs. This can be done with tools that help parse out contributors of variability in true effect sizes, like subgroup analysis or meta-regression.

Second, variable removal performance, ranging from net removal to net leaching, is common in BMP nutrient removal studies (LeFevre et al., 2015). There is evidence that site-level intrinsic BMP characteristics and extrinsic factors are both important for understanding BMP nutrient removal, and meta-analysis acknowledges and quantifies this between-site heterogeneity (Lintern et al., 2020; Valenca et al., 2021).

Third, the meta-analysis process is accessible and widely used. The availability of free meta-analysis packages in R make the method approachable and accessible. This R package is paired with easily accessible documentation and interpretation assistance. Further, the wide use of meta-analysis in a range of scientific fields make this tool recognizable for a broad audience.

Meta-analysis, however, is not without limitations, which impact potential future use of meta-analysis for BMPs. Researchers looking to apply meta-analysis for BMPs are limited in what removal performance drivers are suitable for subgroup analysis. Meta-analysis is limited in its application to BMPs, where only variables that are homogenous at a BMP are appropriate as sub-group analysis groups. Since random effects models in meta-analysis consider variability within a study (here, BMP) level, it is assumed that variables inherent to individual sites contribute to the variability in effect and is addressed in meta-analysis. However, only some of this variability can be assessed

further with subgroup analysis. Subgroups can be defined for factors that are the same for all observations at one BMP, like the regional climate of a site, fill material used, or drainage area ratio (DAR), all of which are constant for all paired observations at a BMP. However, factors that vary within a BMP, like seasonal changes in temperature or precipitation characteristics, cannot be directly investigated with subgroup analysis.

4.4.3 Implications

These results have implications on our understanding of the impact of BMPs in the urban landscape. First, decreased N and increased P concentration in BMP effluent is a result with important ecological implications because of the significance of N:P ratios on biological growth limitations. Second, results regarding BMP performance in different climates may be built upon to develop a stronger understanding of how BMP nutrient management will be affected by a changing climate. Finally, by examining paired concentration change of analytes with different removal pathways, these results provide insight about the varying nutrient transformation processes driving the observed changes in concentration.

4.4.3.1 N:P Ratio Implications

BMPs tend to decrease both dissolved and total N:P ratios from influent to effluent. The ratio of N and P is an important ecological parameter as N:P ratios can indicate biological growth limitation (Koerselman & Meuleman, 1996). Globally, N:P ratios have increased due to anthropogenic perturbations (anthropogenic inputs at 19:1 in 1980s increased to 30:1 in 2020) and varying mobilities of N and P in the environment (Penuelas, Janssens, Ciais, Obersteiner, & Sardans, 2020; Penuelas & Sardans, 2022).

Results here show that BMPs decrease N:P ratios in stormwater, indicating that BMPs may partially offset increased N:P ratios in surface waters.

However, N:P ratio of BMP effluent may have ecological consequences for downstream bodies of water because BMP effluent observed here contained N:P ratios lower than those typically observed in water bodies. N:P ratios of US streams and rivers are diverse, with the interquartile range (IQR) of DIN:DIP ranging from 22:1 to 187:1, and IQR of TN:TP ranging from 18:1 to 59:1 (Manning et al., 2020). These ranges may also vary based on climate (Grimm et al., 2005) and disturbance level (Green & Finlay, 2010). Reported observations of N:P ratios in streams and rivers from four different sources report median and mean DIN:DIP of 45:1 to 51:1, and median N:P ratios of 24:1 to 55:1 (Figure 18). Each of these estimations are above the N:P ratios found in BMP effluent (DIN: DIP = 20:1, TN:TP =15:1) (Figure 18). Thus, discharge of low N:P BMP effluent may supply P to P-limited streams and rivers, potentially stimulating eutrophication. To estimate the extent of impact of low N:P BMP effluent on streams with higher N:P ratio, the context of load must be considered. N:P ratios alone reflect ratios of concentrations, therefore, the mass of N and P in streams may exist on a magnitude much larger or smaller than the mass of N and P discharged from BMP effluent, giving the effluent variable likelihood of altering the N:P ratio of the stream. For example, a stream with high masses of N and P would not likely be ecologically altered by BMP effluent entering with low masses of N:P, but a stream with low existing N and P mass may be highly perturbed by high masses of N and P entering via BMP effluent.

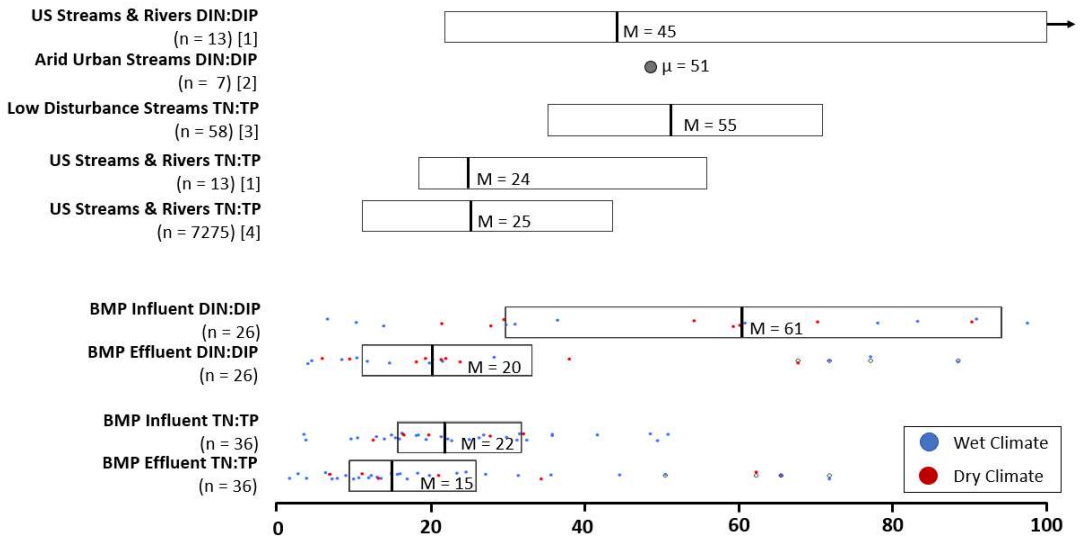


Figure 18: Comparison of N:P ratios in natural waters [1] (Manning et al., 2020) [2] (Grimm et al., 2005), [3] (Green & Finlay, 2010), [4] (Maranger, Jones, & Cotner, 2018) and BMP influent and effluent in this study. M and μ indicate measures of many N:P ratios as either a median or mean, respectively. Colored points indicate the Budyko dryness index classification of sites in this study as either wet (blue) or dry (red).

4.4.3.2 Climate Change

This work investigates the differences in BMP nutrient management in different climates, but it does not directly address how climate change may have an impact on the nutrient management of BMPs. Climate change directly impacts two key factors related to BMP water quality treatment performance: altering stormwater runoff characteristics (e.g., volume, intensity, frequency), and nutrient volumes. It is predicted that climate change will drive more intense, frequent extreme precipitation events, and that urban watershed level impacts will include elevated nutrient loading (Alamdari, Sample, Steinberg, Ross, & Easton, 2017; Marshall & Randhir, 2008). Translating these watershed-level changes to impacts on BMP nutrient management remains a challenge because BMPs are not typically designed for extreme storms, and little monitoring data is

available regarding BMP water quantity or quality performance (L. E. McPhillips, Matsler, Rosenzweig, & Kim, 2021).

Models offer insight into how BMPs may perform under projected climate scenarios through modeling BMP hydrologic performance under various extreme storm conditions, and through watershed-level investigations of agricultural nutrient management. First, the hydrologic performance of BMPs is expected to yield increased discharge volume and discharge peaks, the degree of which is expected to vary by BMP type (M. Wang, Zhang, Cheng, & Keat, 2019). Second, models of agricultural watersheds show mixed predictions about BMP treatment effectiveness under anticipated elevated loads depending on the climate scenario, agricultural land use type, and BMP type (Jayakody, Parajuli, & Cathcart, 2014; Jeon, Ki, Cha, Park, & Kim, 2018; Teshager, Gassman, Secchi, & Schoof, 2017). For example, a Soil and Water Assessment Tool (SWAT) model of a watershed in Iowa estimated that climate change would undermine nitrate reductions by 65% (Teshager et al., 2017), but a similar SWAT model in Mississippi indicated that TN removal was unaffected by climate change and TP removal efficiency increased (Jeon et al., 2018), and a third SWAT analysis of watersheds draining to Lake Erie showed that BMPs became less effective at TP retention 3 of 4 agricultural watersheds (Bosch, Evans, Scavia, & Allan, 2014). Thus, models have yet to establish a clear understanding of the impact of climate change on BMP effectiveness.

The results of this chapter contribute foundational knowledge about nutrient management of BMPs in different climates, necessary information to assess the impact of climate change on BMP water quality treatment. The context that BMP P management varies by regional climate provides a baseline of expectations for BMP performance. This

information is necessary to compare BMP performance to future climate scenarios and establishes that varying P management in different climate regions is expected under future climate conditions. Further. since definition of climate region was established through the Budyko aridity index, this result may also indicate that as a region's climate shifts towards a high Budyko index (more PET than P over long timescales) then decreased P management efficiency may be expected.

4.5 Conclusions

Towards the first objective, meta-analysis found variability of N removal through BMPs as well as increases in DIP composition of TP and decreased DIN:DIP ratios across climate boundaries. Addressing the second objective confirmed the hypothesis that nutrient treatment is different for BMPs in wet and dry climates for the analytes DIP and TP. Contextualizing these observations with analyses of BMP type, influent concentrations, and regional design goals provided evidence of potential causational explanations for this lower relative P performance in dry climates.

This objective highlighted that vegetated BMPs are vulnerable to DIP and TP concentration increases in dry climates, a vulnerability that can be addressed to make vegetated BMP P treatment performance more reliable and watershed-level nutrient reductions more attainable (Lintern et al., 2020). First. BMPs have a key role in altering stormwater P composition toward DIP and decreasing DIN:DIP ratios across both wet and dry climates. Second, P treatment in BMPs is distinct between wet and dry climates, indicating a functional difference in how BMPs process DIP and TP between wet and dry climates. Identifying the drivers behind this arid-intensified P leaching remains a challenge for assessing and designing for BMPs across diverse climates. Further

investigation into driving mechanisms, especially parsing the contributions of biogeochemical processes, elevated influent, and regulations in arid climates, will inform more robust and reliable BMP design.

The influence of regional climate on BMP nutrient management illustrates one impact of the external environment on BMP function. In this section, regional climate was examined through the lens of the Budyko index, and thus reflects an energy and water balance over extended periods of time. However, influences of the external environment on BMP nutrient management may also exist on shorter timescales. In objective 3, the influence of weather is examined by exploring the relationship between storm characteristics and storm-level water quality performance.

5. STORM IMPACT ON BMP NUTRIENT TREATMENT

5.1 Rationale & Summary

Storm characteristics impact factors key to BMP performance, like soil moisture, soil oxygen, filter bypass, and vegetative activity. Current literature has synthesized the impact of storm characteristics on flood mitigation but notes the lack of literature about the influence of storm characteristics on BMP water quality treatment (Sohn et al., 2019), noting that hydrologic and water quality performance likely differ (Lammers et al., 2022).

Studying the impact of storm characteristics across many BMPs requires methods that summarize highly variable storm data and pair storm data with BMP performance metrics. Statistical clustering methods have begun to be applied to urban precipitation data to characterize like storm events (Kajewska-Szkudlarek, 2020; Mikołajewski et al., 2022). The combined prevalence of statistical cluster methodology for storm data and wealth of stormwater quality data before and after treatment in BMPs together provide the means and medium for robust analysis of the impact of storm characteristics on BMP water quality treatment performance. Quantifying the impact of different storm types on BMP water quality performance is an important gap in knowledge of BMP function and must be addressed to design more robust BMPs for the current and future diversity of storm events.

This chapter quantifies the impact of storm characteristics on BMP water quality performance through the lenses of concentration change and load treatment. This is done through two tasks: (1) identifying relationships between storm characteristics and concentration, volume, and load; and (2) comparing concentration and load treatment

performance of BMPs among storm groups identified by statistical clustering. Both tasks were performed on influent and effluent concentration, volume, and load data, and on the removal fraction, measuring the paired change in these water quality metrics. It was hypothesized that BMPs perform water quality treatment differently for distinct storm groups. The first task showed that influent and effluent volumes and loads increase with increasing storm depth and return period. The second task identified five distinct storm clusters, where vulnerabilities to certain storm groups were identified for each nutrient analyte. To contextualize the results, the value of viewing storm characteristics as clustered storm groups was discussed, the context of concentration and load analyses were presented with a comparison of volume reduction in this work compared to other published work, and finally, potential explanatory drivers of storm-group performance susceptibility for specific analytes were presented.

5.2 Methods

5.2.1 Data Curation

Two different groups of data were curated for this objective, water quality treatment performance data from BMPs and storm characteristics data.

5.2.1.1 BMP Performance

BMP performance data was obtained from the International Stormwater BMP Database (https://bmpdatabase.org/), the same large-scale database of stormwater BMPs used in Objective 2 (henceforth, Database). For this objective, version 02-08-2021 was used, obtained in 2022. Water quality data and details pertaining to the location and type of BMP were obtained, including influent and effluent concentration and volumes, site latitude and longitude and multiple BMP site descriptors. Data characterizing storm

events associated with each BMP performance measurement entry were also obtained from the BMP database.

Data selection was filtered by geography, stormwater pollutant type, sample method, and BMP type. Data was restricted to BMPs located in the contiguous United Sates. Water quality data selected included total and dissolved inorganic species of nitrogen (N) and phosphorus (P). Like methods applied in Objective 2, Total N (TN) and Total P (TP) were taken directly from the Database. Dissolved inorganic N (DIN) was calculating by summing nitrate and nitrite measures (recorded in the Database as "nitrate (NO_3^-) as N", "nitrite (NO_2^-) as N", "$NO_2^- + NO_3^-$ as N", "NO_x as N (referring to the sum of NO_3^- and NO_2^-)") with an ammonia or ammonium measure (in the Database as "ammonia (NH_3) as NH_3", "ammonia as N", and "ammonium (NH_4^+) as N"). Dissolved inorganic phosphorus (DIP) was taken from the Database as analytes labeled either "orthophosphate as P", "orthophosphate as PO_4^{3-}", or "dissolved orthophosphate as P". Influent and effluent volume type was restricted to surface runoff, and water quality sample measures were restricted to flow-weighted event mean concentrations. Entries were only considered if they contained both concentration and volume data for influent and effluent. BMP types considered were grass swales, grass strips, bioretention, detention basins, and wetland channels. Grass swales, grass strips and bioretention were selected as BMP types common in their application of engineered soil media as a primary design component, and because of known susceptibility to nutrient leaching, as identified in previous database reports (Barrett, 2008; Clary et al., 2020), and in Objective 2. Detention basins and wetland channels were added to the analysis to provide comparisons of effects on BMP types with distinct designs from the three soil-based types.

BMP performance was analyzed by examining influent and effluent directly, as well as the difference between the two. Influent and effluent concentration, volume, and loads were examined. Concentration and volume data were taken directly from the Database, and load was calculated as the product of concentration and volume. The metric reflecting the difference between influent, and effluent was the effluent fraction, computed independently as concentration, volume, and load effluent over influent fractions. Effluent fraction was used as a BMP performance metric because this metric does not assume correlation between influent an effluent. Assumption of correlation between influent and effluent water quality is presumed if percent change is used to evaluate change in water quality, thus making percent change metrics poor determinants of performance (Davis, 2007; McNett, Hunt, & Davis, 2011). In general, a smaller effluent fraction indicates higher removal whereas a large effluent fraction indicates low removal or leaching for effluent fractions > 1. In figures, effluent fraction is presented in log scale such that negative log effluent fractions indicate net removal and positive log effluent fractions indicate leaching.

Influent and effluent and effluent fractions are presented for each concentration, volume, and load. Both concentration and load were studied because water quality performance can be considered through either concentration or load, depending on the goal of the BMP. Generally, concentrations are easier to compare to other water quality regulations like point source discharge permits, and can be important ecological thresholds (Balderas Guzman, Wang, Muellerklein, Smith, & Eger, 2022). Loads, on the other hand, are often more useful for quantifying the contribution of BMPs towards

watershed-level nutrient management (Hobbie et al., 2017). Volume was studied to assist in contextualizing divergences between concentration and load.

Storm characteristics were also downloaded from the Database when available. These metrics included total storm depth and peak hourly intensity. Storm events that generated influent and effluent concentration and volume data but did not have storm depth or peak hourly intensity data recorded were filled in with data external to the Database, presented in section 5.2.1.2 below.

5.2.1.2 Storm Characteristics

Storm characteristics considered in this analysis included total storm depth, peak hourly intensity, return period, and antecedent dry period. Storm depth and peak hourly intensity data were taken from the Database when available and supplemented with the National Oceanic and Atmospheric Administration (NOAA) U.S. hourly precipitation data, and Automated Surface Observing System (ASOS) hourly precipitation data. The antecedent dry period was calculated entirely from NOAA hourly precipitation data. NOAA hourly precipitation data was retrieved from the three precipitation gauges closest to a BMP (by Euclidean distance) and within 100 km. If only 1 or 2 gauges were within 100 km of the BMP, only 1 or 2 gauges were used. At each gauge, a 3-day window (the day of the reported storm event, and 1 day before and after) was applied, and the total depth was recorded as the highest sum of 24-hour data (midnight to midnight) of the three days in the window. The final total depth value was calculated by applying a weighted average to the reported data, with weights assigned by the Euclidean distance between the BMP and reporting precipitation gauge so that closer gauges had heavier weights. Peak hourly intensity was calculated as the highest hourly precipitation interval

reported within the 24-hour window selected for the total depth calculation. Final peak hourly precipitation intensity was calculated as a distance-weighted average of the gauges used.

Return period was calculated using an event's total depth to interpolate a return period from reported intensity duration frequency (IDF) curves. IDF curves were retrieved from the NOAA Precipitation Frequency Data Sever. The nearest 24 hr IDF curve was identified, and the return period of the storm was calculated by linearly interpolating between the reported storm depths for the 1-, 2-, 5-, 10-, 25-, 50-, 100-, 500-, and 1000-year storms. The return period was calculated for the depth reported directly in the Database when available, with the NOAA retrieved depth if reported depth was not available. IDF curves for the states of Washington and Oregon were not available in the NOAA Precipitation Frequency Data Server, so they were taken from tabulated return period values in a regional report in Washington (Demissie & Mortuza, 2016) and the NOAA Atlas 2 report for Oregon (Miller, Frederick, & Tracey, 1973). Finally, the calculated return period values were summarized in a final value with a distance-weighted average.

Both return period and retrieved depth may be overestimated. The moving 3-day window used to retrieve storm depths selected the largest recorded 24-hour depth in the window. It is possible that there was more than one true storm to occur within the 3-day window, in which case, the larger event would be favored. As return period was calculated using depth, both depth and return period values may be overestimated.

The antecedent dry period was calculated as the difference between the date of the selected storm event in the total depth calculation and the last previously recorded rainfall

event, with adjustments for the 3-day window. Rainfall was considered an event if there was 0.05 cm or greater total precipitation within a day (24 hours, midnight to midnight). If two storm events occurred within the 3-day window applied to find total depth, then the antecedent dry period was recorded as 0.1. For antecedent storms outside the window, the number of antecedent dry days was recorded as the number of days between the recorded storm date and the previous event, minus 0.5 days, since the data may have been taken from the 0-1 range of the dryness window and 0.5 is the mean of this range. The final antecedent dry period was calculated as a distance-weighted average of the three nearest precipitation gauges within 100 km. Gaps in recorded hourly data were identified and overcome using standard deviation of antecedent dry period for long dry periods. Any storm events with an antecedent dry period identified as a statical outlier (> 8.7 days) and whose standard deviation was greater than the mean antecedent dry period value, was recalculated by dropping the gauge responsible for the long dry period, as this value may be representing absence of recorded data rather than a true dry period. This process proceeded until either the standard deviation fell below the outlier value and the standard deviation was lower than the mean dry period, or until only one dry day value remained. There were 177 storms for which antecedent dry period estimates were dropped due to suspected period of inactivity at the precipitation gauge. This process may result in underestimating the length of dry days, especially if the true antecedent dry period was long and variable between nearby gauges.

5.2.1.3 Climate Consideration

Regional climate and storm characteristics are intertwined variables, where climate is represented by long-term averages of precipitation and evapotranspiration, and

storm characteristics describe single, short-term weather events, but whose statistics are related to the regional climate. Therefore, climate in this analysis is not considered when evaluating storm characteristics themselves but is used in the clustering analysis to assist in the clustering of like storm events. Climate was quantified through the Budyko aridity index, a ratio of annual potential evapotranspiration to precipitation, i.e., PET/P (Edoardo Daly, Calabrese, Yin, & Porporato, 2019b). The dryness index for each BMP was calculated using the Climate Research Unit gridded Time Series (CRU TS) monthly gridded climate dataset with a spatial resolution of 0.5° latitude by 0.5° longitude (Harris et al., 2020), the same method applied in Objective 2.

5.2.2 Water Quality Analysis

The impact of storm characteristics on BMP nutrient management was quantified through a two-step process. First water quality parameters (concentration, volume, and load) observed in influent and effluent were examined, and second the paired change between these water quality parameters were examined. This two-step framework displayed ranges of expected concentrations, volumes, and loads of nutrients in influent and effluent of BMPs and demonstrated how influent and effluent changed independently. In contrast, the second method, examining paired change through the removal fraction, analyzed the actual degree of change observed between paired influent and effluent from the same storm at the same BMP.

This two-step framework was repeated under two different means of examining storm characteristics, 1) isolated storm traits, and 2) storm groups. Isolated storm traits include the four factors: storm depth, return period, peak hourly intensity, and antecedent dry period. Storm groups classify storms by their similarity among the four isolated traits,

and their regional climate index. Thus, the two-step framework of water quality analysis is repeated for two means of examining storm characteristics, resulting in 4 major components of water quality performance analysis. Methods to perform water quality analysis by isolated storm traits are presented in methods section 5.2.2.1, and methods for analysis by storm group are presented in methods section 5.2.2.2.

5.2.2.1 Water Quality Analysis for Isolated Storm Traits

Water quality analysis by isolated storm trait was evaluated with boxplots of increasing magnitude of each storm trait. Groups of storm characteristics were defined using the quantiles of each storm characteristic. Each storm characteristic was broken into five bins, representing the first, second, third, and fourth quantiles (Q1, Q2, Q3, and Q4, respectively), as well as high-range outliers High range outliers were defined using the equation:

$$Outlier \geq Q3 + (1.5 \times IQR)$$

Where *Q3* is the third quartile and *IQR* is the interquartile range (i.e., the difference between the first and third quantile). The bounds of each quantile bin for each storm characteristics are presented in Table 5.

Table 5: Boxplot group boundaries for the storm characteristics depth, return period, peak hourly intensity, and antecedent dry period based on quantile distributions.

Trait	Q1	Q2	Q3	Q4	Outliers
Depth (cm)	< 1.03	1.03 - 1.8	1.8 - 3.12	3.12 – 36.26	≥ 36.27
Return Period (yr)	< 0.17	0.17 - 0.30	0.30 - 0.46	0.46 – 0.90	≥ 0.90
Peak Hourly Intensity (cm/hr)	< 0.36	0.36 - 0.68	0.68 - 1.23	1.23 – 2.53	≥ 2.53
Antecedent Dry Period (days)	≤ 0.50	0.50 - 1.86	1.86 - 4.94	4.94 - 11.61	≥ 11.61

Assessment of influent and effluent water quality parameters by isolated storm traits was conducted by examining presence and direction of change across boxplots of increasing trait magnitude. These trends were examined for influent and effluent measures of concentration, volume, and load for TN, DIN, TP, and DIP. For example, general positive or negative relationships between water quality parameters (i.e., influent concentration of DIP) and increasing quantile bins (first quantile to second quantile of total storm depth) were noted.

Assessment of the paired change in concentration, volume, and load of each nutrient was performed with boxplots of the log removal faction. Boxplots were used to determine the general change in concentration, volume, and load as leaching or retention for the subsequent interquartile range bins of isolated storm traits. Further, a Tukey test at $\alpha = 0.05$ was used to compare the log removal faction of each analyte across the binned intervals of storm traits.

5.2.2.2 Water Quality Analysis for Clustered Storm Groups

Water quality analysis by storm groups were also evaluated with boxplots of increasing magnitude of each storm trait. However, where the boxplots of isolated storm groups needed to be repeated for each storm group separately, and bins were created by quantile, here, all storm traits are assessed together through storm groups identified with clustering methods. Two clustering algorithms were applied: k-means and hierarchical clustering. The selected clustering assignment was used to separate the BMP water quality treatment performance data by the type of storm that generated the storm event. Finally, the water quality treatment performance of BMPs for each of the four nutrient analytes was assessed for each storm group.

Clustering algorithms are common unsupervised machine learning tools used to partition multivariate data into groups (Steinley, 2006). The goal of clustering algorithms is to separate data into groups, called clusters, that share more similar attributes between samples within clusters than samples in different clusters (James, Witten, Hastie, & Tibshirani, 2021). It is assumed that the clusters follow a natural but hidden structure (Xu & Wunsch, 2005). In this analysis, k-means and hierarchical clustering algorithms were used to cluster 5-dimensional storm attribute data into distinct clusters subsequently called storm groups. K-means creates clusters by assigning k number of centroid points, and assigning the samples to the nearest centroid, then moving the centroid points and re-assigning clusters iteratively to reduce the within-cluster variation for all clusters (James et al., 2021). Hierarchical clustering applies a different approach, forming similar groups in subsequent steps by starting at groups of size 1 and building larger groups (i.e. agglomerative nesting), or by starting with a single group of all observations and splitting

into smaller groups (i.e. divisive analysis) (Murtagh, 2017; Murtagh & Contreras, 2012). These clustering algorithms were selected because both are common, and comparisons between the methods can be made easily by comparing cluster assignments for the same k number of clusters (James et al., 2021).

Storm characteristic and climate data were prepared for clustering by evaluating collinearity and standardizing all values. Collinearity between storm characteristics was compared to ensure distinct information contributions from each metric. Collinearity was evaluated with a correlation matrix of the storm characteristics: depth, peak hourly intensity, return period, antecedent dry period, and Budyko aridity index. Storm metrics were then standardized using z-score standardization:

$$Z_n = \frac{x_n - \mu}{\sigma}$$

Where $Z_n$is the standardized value of sample n, x_n is the value of sample n, μ is the mean of samples, and σ is the standard deviation of the samples. This standardization was performed to reduce the influence of difference of scales on the development of clusters.

Multiple options are available for optimizing the formation of k-means and hierarchical clustering, including a distance calculation, hierarchical linkage method, and the desired number of clusters, k. In this analysis, Euclidean distance was used as the distance measure to determine similarity for k-means and both agglomerative and divisive hierarchical clustering. The hierarchical linkage was selected from four common methods including: average, single, complete and Ward's (Murtagh, 2014). Ward's hierarchical linkage method was selected because it produced the best clustering structure (as measured by the agglomerative coefficient). Each algorithm was run with 2 to 7 desired final clusters. Outliers were handled by manually designating storms to an outlier

cluster if they were repeatedly identified as a distinct cluster away from all other observations with both hierarchical and k-mean algorithms (Gan & Kwok-Po Ng, 2017). Selection of a final clustering algorithm and optimal *k* value were determined by assessing the agreement of k-selection values between two optimal cluster selection methods: elbow method and silhouette method (Shahapure & Nicholas, 2020; Syakur, Khotimah, Rochman, & Satoto, 2018).

The result of the clustering analysis was an assignment of every storm entry in the dataset into a group of like storms. The storm groups were then compared to assess influent and effluent water quality for each storm group, and the changes in water quality parameters between storm groups. Assessment of influent and effluent water quality parameters by storm groups was conducted by examining the ranges median of concentrations, volumes, and loads across storms, and noting any observations that varied from other storm groups. Assessment of the paired change in concentration, volume, and load of each nutrient was performed with boxplots of the removal faction. For this analysis, each boxplot was tested with a one-sample, two-tailed Wilcoxon sign rank test for statistical difference from 1, the removal fraction indicating no change between influent and effluent. Difference from 1 was evaluated where $\alpha < 0.05$ was interpreted as significant difference and $\alpha < 0.10$ was interpreted as marginal difference, so that a distribution statistically different from one (either smaller or larger) indicates a meaningful change from influent to effluent (retention, or leching, respectively).

5.3 Results

5.3.1 Water Quality Data Characterization

In total, 1871 water quality entries met the BMP data curation conditions. These entries came from 90 BMPs across 82 sites (Figure 19). Sites had a wide geographic spread, spanning 58 cities in 18 states. The stormwater quality entries were composed of 887 unique storm events, meaning many storms contained measurements for more than one of the four considered analytes, as each analyte was considered a different water quality entry. TP was the most observed analyte, then DIP, then TN, and DIN was the least observed analyte. Bioretention was the most observed BMP type, followed by detention basins, wetland channels and grass swale, then grass strips (Table 6).

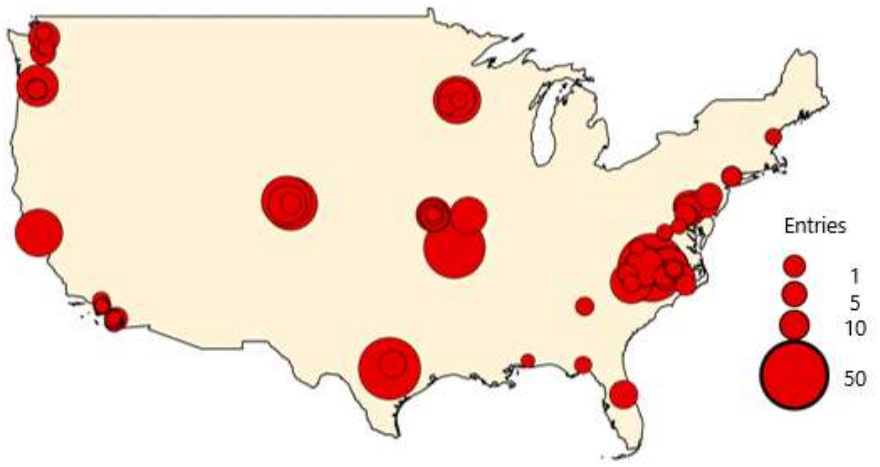


Figure 19: Map of BMP sites used, showing larger point sizes for sites with more water quality data entries.

Table 6: Counts of BMP water quality entries in total (ALL), and by analyte (TN, DIN, TP, and DIP) for each of the 5 BMP types considered: grass strips (BI), bioretention (BR), grass swales (BS), detention basins (DB) and wetland channels (WC).

Analyte	Total Entries	BI	BR	BS	DB	WC
TN	330	6	125	38	108	53
DIN	301	20	147	54	55	25
TP	785	56	244	126	263	96
DIP	455	68	164	43	93	87
ALL	1871	150	680	261	519	261

The magnitudes of concentrations, volumes, and loads fell into similar groups for bioretention, grass strips, and grass swales, while detention basins and wetlands each had much higher magnitudes of volumes and loads. This prompted the compilation of the bioretention, grass strip and grass swale BMP types into a single BMP type labeled “bio-composite” BMPs. This compilation is justified not only by the common ranges of volumes and loads received by these BMP types, but also by the common design and nutrient treatment. Bioretention, grass swales, and grass strips all share common design features, relying on soil biogeochemistry and vegetation for water quality treatment, and commonly are referred to with other, overlapping terms (e.g., “bioswales”, “rain garden”) (Choat & Bhaskar, 2020). Further, these BMP types are alike in their variable nutrient processing, resulting in increasing nutrient concentrations from influent to effluent especially for P species (Objective 2, and Clary et al., 2020). These ample similarities justify the presentation of these three BMP types in subsequent section of this objective as a single agglomerated BMP type.

5.3.2 Storm Characteristic Characterization

Storm characteristics were generally right skewed with a long right tail, showing high frequency of low-magnitude data by the zero-bound (Figure 20). Such distributions are expected for storm data, for which small rain events and low intensities are more common than large, intense events. Depth data was reported for 86% of entries, so only 14% of entries were filled in with retrieved precipitation data. For the 14% (262) of entries with retrieved depths, 226 entries used three precipitation gauges to develop the average, nine entries used two gauges and 27 entries used one gauge. The distance-weighted average distance between a BMP and the precipitation gauge(s) used ranged from 5.6 to 91.0 km (median = 23.9 km, mean = 35.5 km). Peak hourly intensity, however, was only reported for 18% of entries. The other 82% (1539 entries) used three precipitation gauges for 1332 storms, two gauges for 70 storms, and one gauge for 147 storms. The distance-weighted average distance between gauge(s) and a BMP ranged from 3.0 to 94.9 km (median = 41.0 km, mean = 41.1 km). Return periods were calculated with the reported or retrieved depth values and tabulated returns from IDF curves. The distance between BMPs and the location of the tabulated IDF values ranged from 0.40 to 5.6 km. Finally, all antecedent dry period entries were calculated from precipitation gauges, where 24% of entries required reduced gauge counts because of unavailable data in the time preceding the observed event. With this adjustment, 66.8% of entries used three precipitation gauges, 16.6% used two precipitation gauges, and the remaining 16.6% used one precipitation gauge. The average weighted distance between BMPs and precipitation gauges used for antecedent dry period estimations was 3.06 - 92.2 km (median = 38.8 km, mean = 39.6 km).

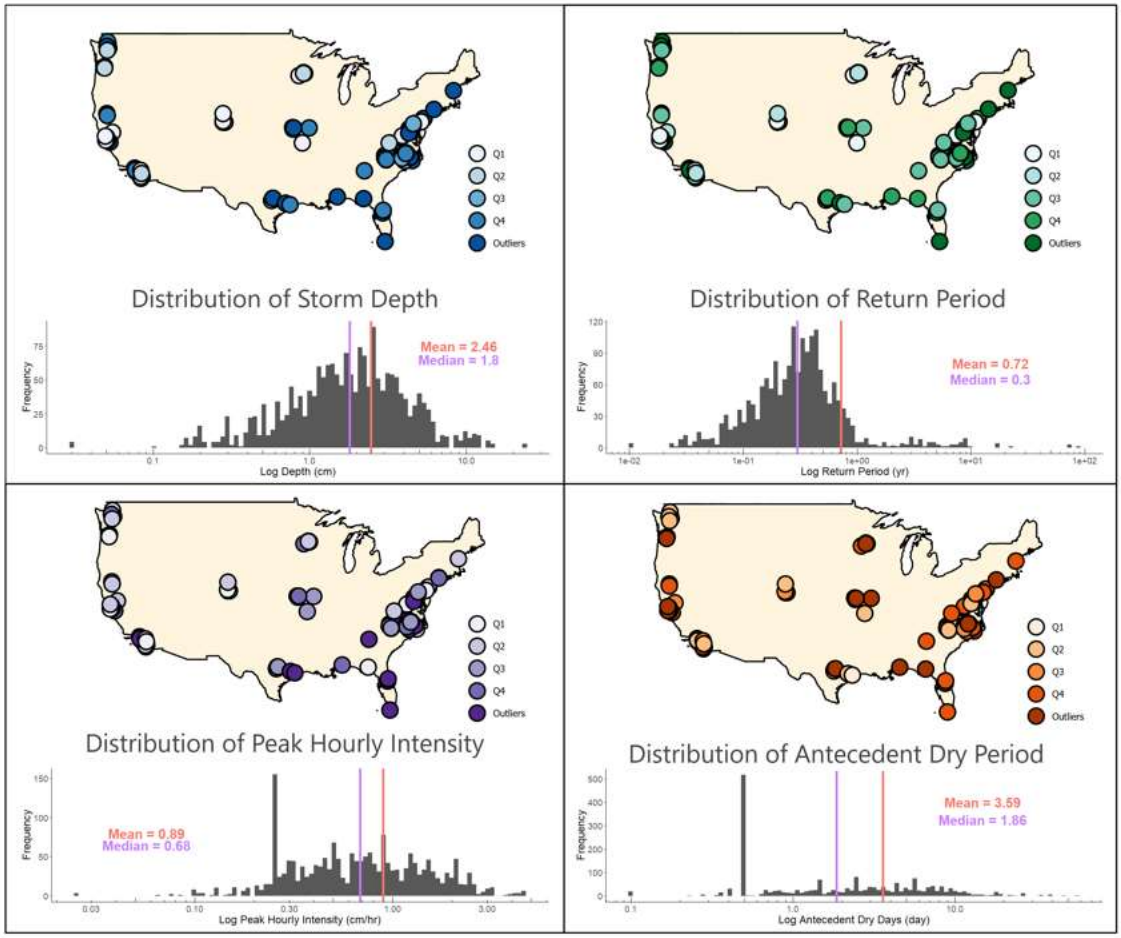


Figure 20: Maps of the distribution of 1) depth, 2) return period, 3) peak hourly intensity, and 4) antecedent dry period. Points indicate distribution of storm traits as falling within the first, second, third, or fourth quantile or outliers, with increasingly dark shading for higher values. Histograms below maps present the range of the storm trait observations in log scale, with vertical lines indicating the mean and median values (in orange, and purple, respectively).

There were no strong correlations among storm characteristics or the regional dryness index. The highest correlation was 0.50 between depth and return period. A high correlation between these variables was expected because return period was calculated using depth. However, the relatively low correlation coefficient between these two variables indicates that return period has distinct information from depth, brought about from the IDF curves which added the context of regionality to precipitation depth values.

Since none of the storm traits or regional dryness index were highly correlated, all traits were used in the clustering algorithms.

5.3.3 Influent and Effluent Evaluation by Isolated Storm Traits

There was a general positive relationship between both influent and effluent loads and increasing storm depth and return period for all BMP types and analytes. No broad trends were clear across all BMP types and analytes for peak hourly intensity and antecedent dry period. No clear unidirectional change in the load reduction was observed across grouped stormwater traits. However, there were some distinctions between certain quantile and outlier groups indicative of varying removal performance for different storms depth and return period ranges.

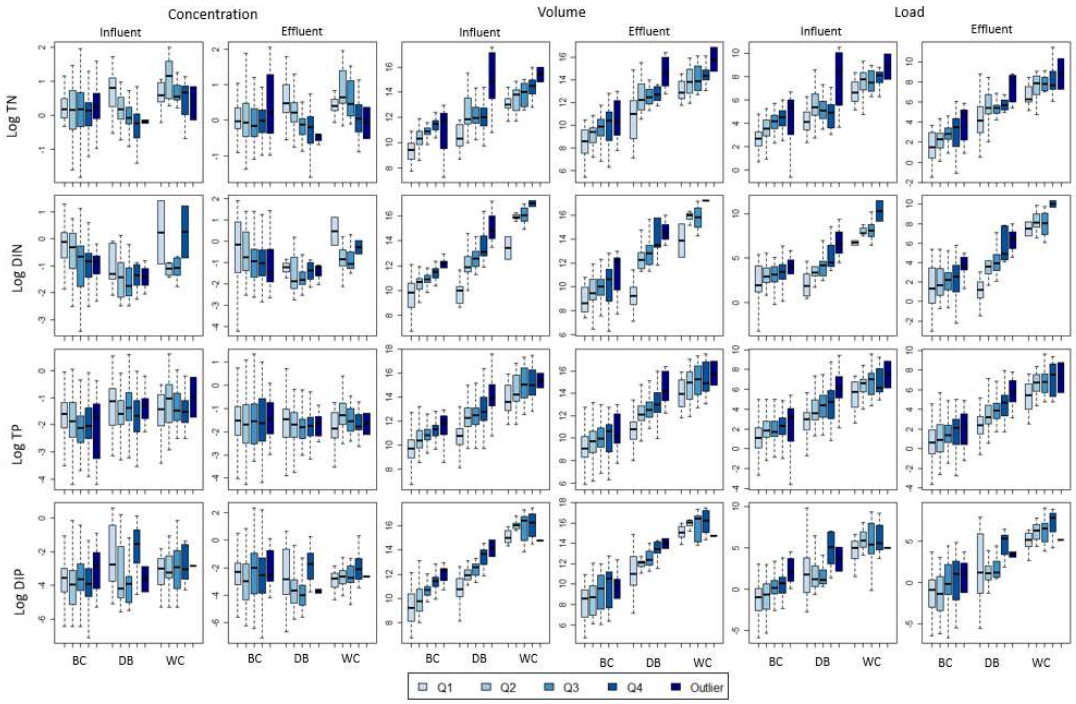


Figure 21: Boxplots of log-scaled influent and effluent concentrations, volumes, and loads at increasing quantiles of storm depth for each of the analytes: TN, DIN, TP, and DIP, at each of the three condensed BMP types: bio-composite (BC), detention basin (DB) and wetland channel (WC). Storm depth is displayed with boxplots of increasing depth values through five groups: four quantiles and a high-range outlier group, where darker shaded boxplots indicate higher depth. Boxplots show the interquartile range of each water quality parameter shaded in a box with whiskers extending to the minimum and maximum water quality values without outliers.

Depth had a clear relationship with each analyte, where increasing depth was associated with increasing volumes and loads for both influent and effluent. Concentrations of TP and DIP did not change with increasing depth, but influent and effluent TN concentrations in detention basins and DIN concentrations of bio-composite BMPs both decreased with increasing storm depth (Figure 21). These negative relationships between depth and concentrations may be due to dilution as stormwater volume increases. Volumes of influent and effluent had positive relationships with increasing depth for all BMP types. These relationships are expected as storm depth is a

direct source of influent volume. Finally influent and effluent loads follow the same positive relationship as volume with higher loads for higher storm depths.

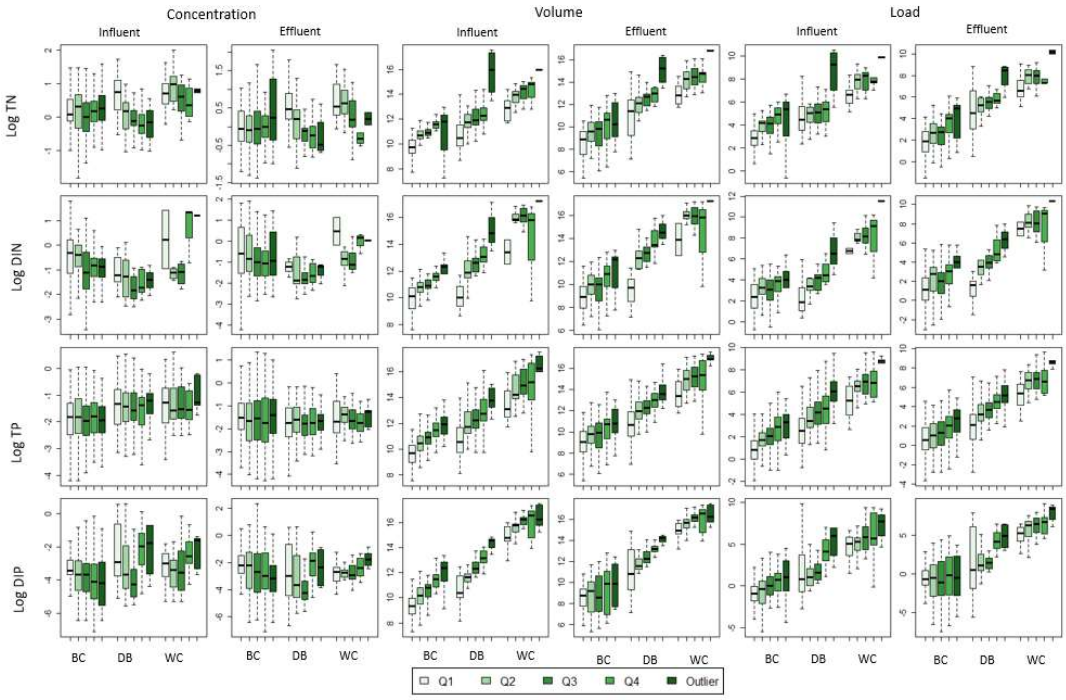


Figure 22: Boxplots of log-scaled influent and effluent concentrations, volumes, and loads at increasing return periods for each of the analytes: TN, DIN, TP, and DIP, at each of the three condensed BMP types: bio-composite (BC), detention basin (DB) and wetland channel (WC). Return period is displayed with boxplots of increasing return period through five groups: four quantiles and a high-range outlier group, where darker shaded boxplots indicate higher return period. Boxplots show the interquartile range of each water quality parameter shaded in a box with whiskers extending to the minimum and maximum water quality values without outliers.

The relationships between water quality metrics and return period follow many of the same relationships illustrated in the plots of water quality metrics against depth, where generally influent and effluent concentrations are largely unaffected by increasing return period, but both volume and loads of influent and effluent increase with increasing return period. In detention basins, TN influent concentrations decreased with increasing return period, and decreased for TN effluent concentrations, and the difference in

concentrations was smaller for effluent than influent (Figure 22). DIN influent concentrations at bio-composite BMPs decreased with increasing return period, and at the same BMPs, effluent also decreased from Q1 to Q3, but Q4 and return period outliers have similar concentrations as Q3. Both DIN and DIP influent and effluent in detention basins appeared to form a “u” shape between boxplots, showing a decrease in concentration until the third quantile (a return period of 0.30 – 0.46 years), after which, concentrations increased with increasing return period. This “u” shape may be the result of dilution bringing down influent concentrations as storms grow larger, then crossing an inflection point where nutrient accumulation surpasses any accumulation from increased storm volume associated with increase return period. Finally, for TP and DIP at bio-composite and wetland channel BMPs there was no apparent relationship between return period and influent or effluent concentration. Volume and load for influent and effluent had positive relationships with return period for all three BMP types.

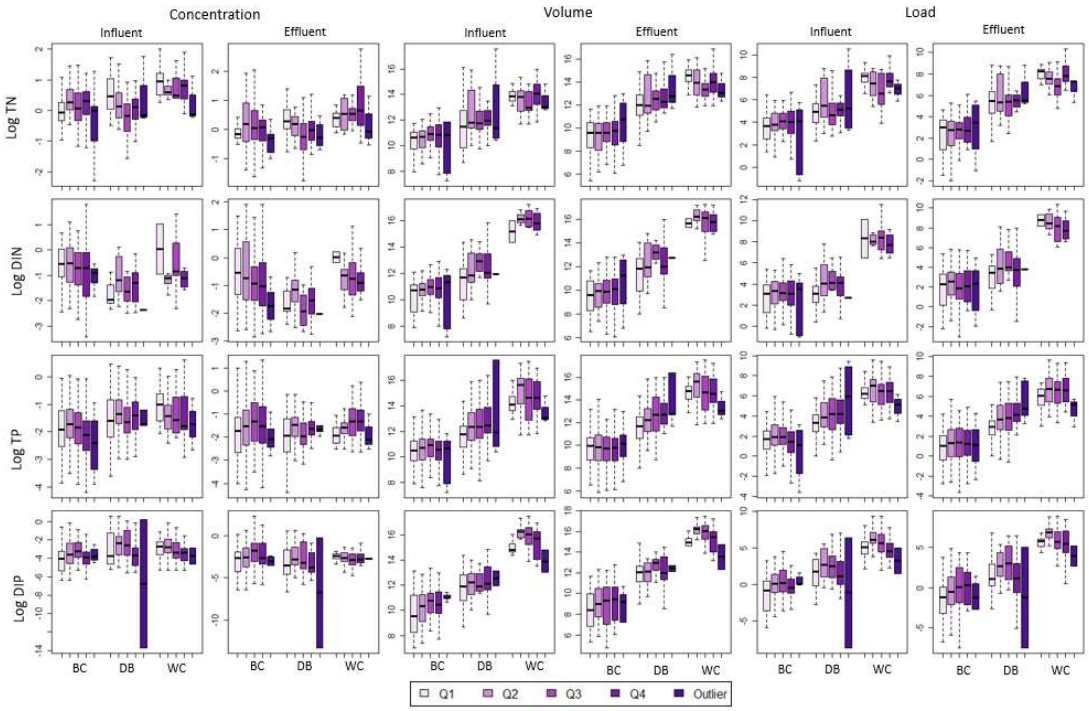


Figure 23: Boxplots of log-scaled influent and effluent concentrations, volumes, and loads at increasing peak hourly intensity for each of the analytes: TN, DIN, TP, and DIP, at each of the three condensed BMP types: bio-composite (BC), detention basin (DB) and wetland channel (WC). Peak hourly intensity is displayed with boxplots of increasing intensity through five groups: four quantiles and a high-range outlier group, where darker shaded boxplots indicate higher intensity. Boxplots show the interquartile range of each water quality parameter shaded in a box with whiskers extending to the minimum and maximum water quality values without outliers.

Relationships between water quality metrics and peak hourly intensity were not unidirectional or shared between multiple analytes or BMP types. Some individual trends are discernible. For example, influent TN concentration for detention basins forms a “u” shape of decreasing concentration from Q1 to Q3, then increasing from Q3 to outliers (Figure 23). Effluent TN concentrations for detention basins form a different relationship, with deceasing concentrations with increasing intensity across all 5 groups. Detention basins also have a positive relationship with influent and effluent volume for TP observations. However, even such specific patterns are not discernable for many boxplot

groups. For example, the boxplot whiskers of the effluent concentrations and loads for TP in bio-composite BMPs are largely overlapping amongst boxes, showing data with similar ranges, even if the median values may appear to follow an “n” shape, with increasing concentration from Q1 to Q3 and decreasing again from Q3 to outliers. Overall, the relationships between peak hourly precipitation intensity and analyte concentrations, volumes and loads are complex without clear universal trends.

Relationships observed between antecedent dry periods and water quality parameters had no clear or re-occurring relationships across BMP types and analytes. However, there were some distinct relationships for specific analytes and BMPs. DIN effluent concentration for detention basins, and TP concentrations of influent and effluent for bio-composite and detentions basins all generally increased with increasing antecedent dry periods (Figure 24). However, these observations are not true for every sequential boxplot. For example, TP effluent concentrations decreased from Q3 to Q4 despite otherwise increasing with increasing dry period. Effluent volume for TP observations also tended to decease with increasing dry period, although this did not occur for every subsequent boxplot, and the whiskers for many plots cover a similar range indicating no distinct changes in volume range by increasing antecedent dry period. There were no clear relationships between load and antecedent dry period length.

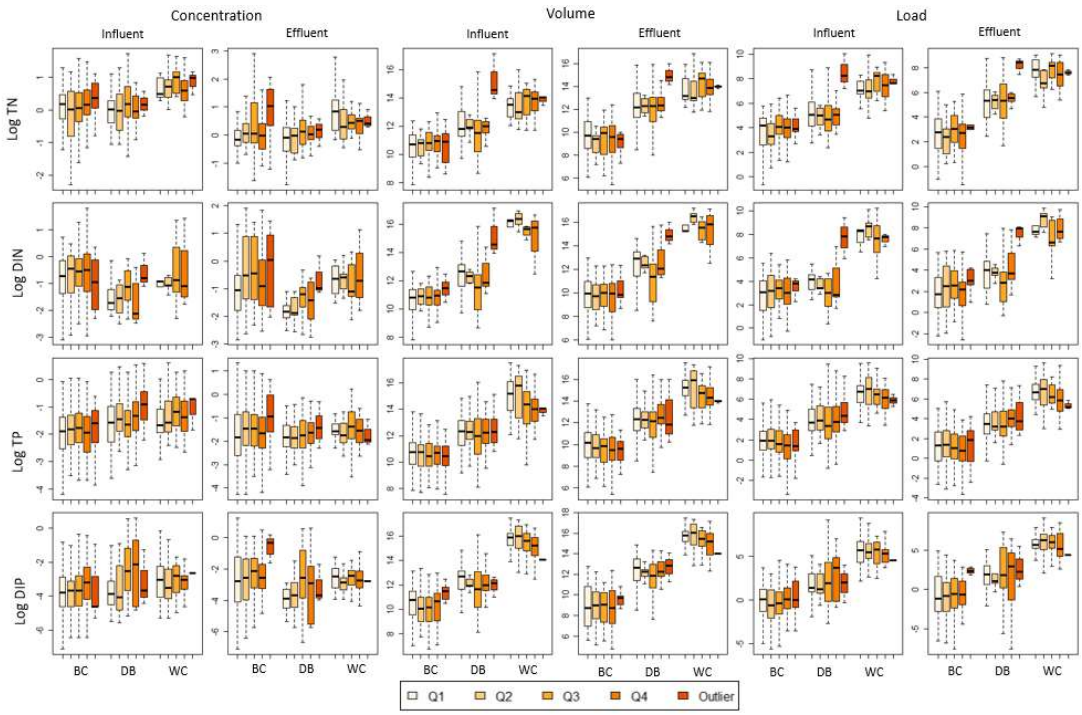


Figure 24: Boxplots of log-scaled influent and effluent concentrations, volumes, and loads at increasing antecedent dry periods for each of the analytes: TN, DIN, TP, and DIP, at each of the three condensed BMP types: bio-composite (BC), detention basin (DB) and wetland channel (WC). Antecedent dry period is displayed with boxplots of increasing dry days through five groups: four quantiles and a high-range outlier group, where darker shaded boxplots indicate longer dry periods. Boxplots show the interquartile range of each water quality parameter shaded in a box with whiskers extending to the minimum and maximum water quality values without outliers.

The investigation of relationships between isolated storm traits and BMP water quality parameters resulted in three major take-aways. First, relationships between depth and load and return period and load more closely resemble the relationships between these storm traits and volume than they resemble the concentration relationships. This observation provides justification that volume change, and the factors that drive variable volumes are controls of influent and effluent loads. Second, many of the trends that appear in influent are also present in effluent. This consistency makes it difficult to distinguish the influence of storm characteristics on the changes between influent and

effluent. Finally, only depth and return period had clear universal impacts on water quality parameters, where peak hourly intensity and antecedent dry period did not.

5.3.4 Water Quality Treatment by Isolated Storm Traits

The five quantile-based precipitation depth groups did not have shared trends in differences from influent to effluent across increasing precipitation depths groups. TN had different volumes and TN loads between precipitation outliers and all other precipitation groups for bio-composite BMPs, where the outlier precipitation group had higher log removal fractions than the other groups (indicating more occurrences of effluent volume and load greater than influent volume and load). For TN there was no difference between any precipitation groups for detention basins and wetland channels. The same observation was made with TP, where TP volume and load in the outlier groups were different from all other groups (again, lower reduction of volume and load). DIN also had different volume and load changes for outlier precipitation than all other precipitation groups for bio-composite BMPs. However, DIN also had a notable difference between Q1 and Q4 for detention basins, where precipitation in Q4 (3.12-36.27 cm) had a higher log removal fraction than in Q1 (<1.03 cm). This difference is visible in the Figure 25, where the boxplot for Q4 for DIN in detention basins is above the threshold of no change between influent and effluent, indicating net increases in volume and load from influent to effluent. Comparatively, the entire interquartile range for the corresponding Q1 is below this threshold, indicating confidence that volume and load are reduced for storm depths in Q1. Finally, DIP was not significantly different between any depth groups at any BMP types.

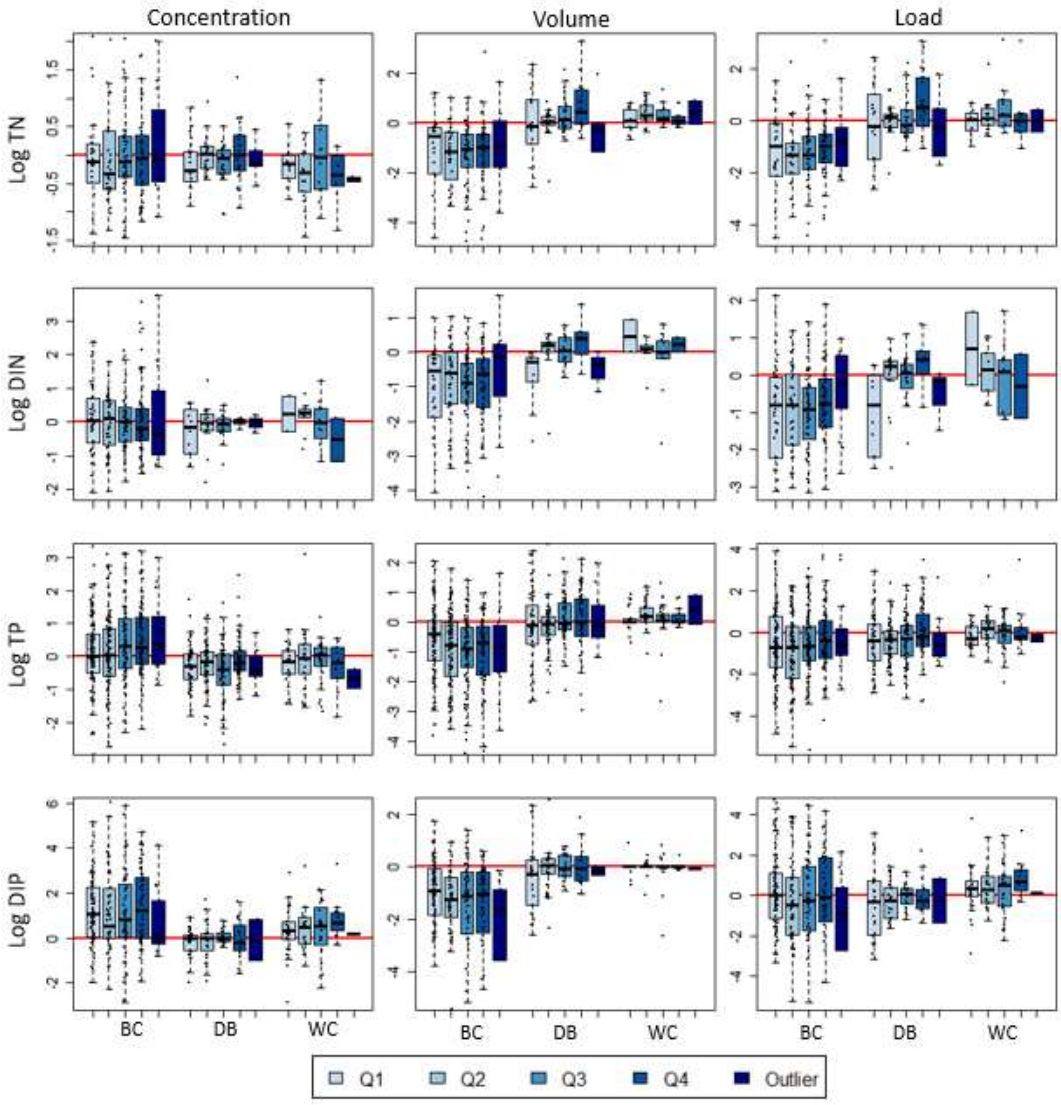


Figure 25: Boxplots comparing the log concentration, volume, and load removal fractions across groups of increasing storm depth for each analyte: TN, DIN, TP, and DIP, at each of the three condensed BMP types: bio-composite (BC), detention basin (DB) and wetland channel (WC). Precipitation depth is displayed with boxplots of increasing dry days through five groups: four quantiles and a high-range outlier group, where darker shaded boxplots indicate higher depths. Boxplots show the interquartile range of each water quality parameter shaded in a box with whiskers extending to the minimum and maximum water quality values without outliers. Black circles represent all observed points, which are plotted on top of boxplots to convey the counts and distribution of data summarized within each boxplot. The red horizontal line is the logged removal threshold, above which effluent values are higher than influent values, indicating leaching; and below which effluent values are lower than influent values, indicating retention.

The other three storm characteristics also did not reveal broad trends in differences across boxplots, but some isolated differences for certain analyte and BMP type combinations were observed. These specific cases of differences in performance within boxplot groups are presented below. As the format for these plots is largely the same, figures parallel to Figure 25 for depth are included in Appendix C1 for return period (Appendix C1.1), peak hourly intensity (Appendix C1.2), and antecedent dry period (Appendix C1.3).

Removal fractions varied between the low and high return period groups for specific analyte and BMP type combinations. The removal fractions for TN concentration were not distinct amongst return periods for any BMP types and was only distinct between outlier return periods (> 0.9 years) and all other groups of return periods for volume and load removal fractions in bio-composite BMPs (Appendix C1.1). In these cases, the outlier return period group had higher log removal fractions than all other return period groups. Despite displaying the highest removal fractions of any return period group, even this outlier return period group had a majority of observed log removal fractions less than 0, indicating net retention. DIN concentrations in bio-composite BMPs were similar across all return period groups, but for DIN volume and load change, outliers had higher log removal (greater tendency towards leaching) for bio-composite BMPs. While only outliers were different from other return period groups in a statistically meaningful way, there is an observed trend of increasing log removal fraction (progressively worse DIN load removal) across groups of increasing return period. For DIN in detention basins, volume change with DIN measurements had a higher log removal fraction than the other return period groups. No wetland channel DIN-related

changes were different among return period groups. For both TP and DIP, the only difference between return period groups was between Q1 and Q4, and Q1 and outliers for volume reductions in bio-composite BMPs. In this case, both Q4 and outliers had higher log removal fractions than Q1, but Q4 and outliers were not different from each other or from any other groups.

Removal fractions were not distinct between peak hourly intensity groups for concentration removal, but sometimes varied for volume and load removal. TN load and volume removal was different for outlier peak intensities than across the other peak intensity groups for bio-composite BMPs. Detentions basin TN volume and load reduction did not vary by peak intensity group, but volume associated with TN measurements in wetland channels was different between Q1 (<0.36 cm/hr) and Q4 (1.23 – 2.53 cm/hr), where there was net volume removal for more events in Q4. The wetland channel volume changes associated with TN measurements also displayed a general trend of improved volume reductions with higher peak hourly intensity rates. DIN volume and load removal fractions in bio-composite BMPs in the outlier group had higher DIN load leaching than for any other quantile group. In detention basins, DIN load removal was different between Q2 and outliers, Q3 and Q4, and Q4 and outliers. There was no difference in DIN removal fractions in wetland channels by peak intensity group. TP for bio-composite BMPs, like TN and DIN, displayed different removal fractions between outliers and all other peak intensity groups for volume and load. TP did not vary between any group of intensity for detention basins, but in wetland channels volumes associated with TP removal were different between Q1 and Q3, and Q1 and Q4. In both these cases, Q1 peak intensities were associated with higher volume leaching tendencies than Q3 or

Q4, a pattern of higher volume leaching for lower peak intensity values that can be seen in the slight negative trend across boxplots for DIN volume removal in wetland channels. Finally, DIP was not different between any peak intensity groups for any BMP type.

Differences in removal fraction by antecedent dry period groups were scarce, and most differences were for DIN. No TN or TP removal fractions of concentration, volume, or load varied among any antecedent dry period groups for any BMP type. DIN concentration in bio-composite BMPs varied among outlier antecedent dryness (>11.61 days) and all other dry period groups. In detention basins, DIN varied among different dryness groups for concentration, volume, and load, where concentration removal varied between Q1 and Q4, volume removal varied between Q3 and Q4, and load removal varied between Q1 and Q4, Q2 and Q4 and Q3 and Q4. In each of these cases, the group associated with shorter antecedent dryness period had a strong tendency toward higher removal than Q4, indicating worse removal at dryness periods falling in the Q4 group (4.94 - 11.61 days). DIN did not vary between any dry period groups in wetland channels. DIP only had a different removal fraction for DIP concentration removal in bio-composite BMPs, where Q3 and outlier boxplots were different with outliers having a higher tendency towards increasing BIP concentrations through bio-composite BMPs.

5.3.4 Storm Cluster Assignment

A clustering assignment created by the k-mean algorithm with k = 5 was selected. Two outliers were manually labeled as outliers because they were identified as their own cluster with k-means at k = 2, and 4-7, and by hierarchical clustering at k = 2 – 7. These storms also displayed clear extreme values of storm characteristics used in the algorithm, notably their average depth was 17.5 cm and average return period was 82.21 years

(Table 7). These storms were labeled as outliers and were removed from the clustering analysis. The remaining storms were assigned into storm groups using the k-means algorithm at k = 5 because both the elbow and silhouette method identified 5 as the optimal number of clusters (Appendix Figure C2.1). Comparatively, the hierarchical algorithm recommended k = 4 with the elbow method and k = 5 with the silhouette method. The silhouette coefficient for the selected clustering method (k-means at k = 5) was 0.40.

The storm cluster assignments were visualized with a principal component analysis to reduce the 5 dimensions used to create the clusters into 2 dimensions for comprehensible observations of clusters. PC1 was composed primarily of depth, return period and peak intensity, and is thus labeled as the precipitation dimension. PC2 was composed of the climate index and antecedent dry period and was named the dryness dimension. Thus, plotting with these two dimensions as axes reveals that most storms are low in both dimensions, and clusters are formed as characteristics ascribed to either dimension increase (Figure 26).

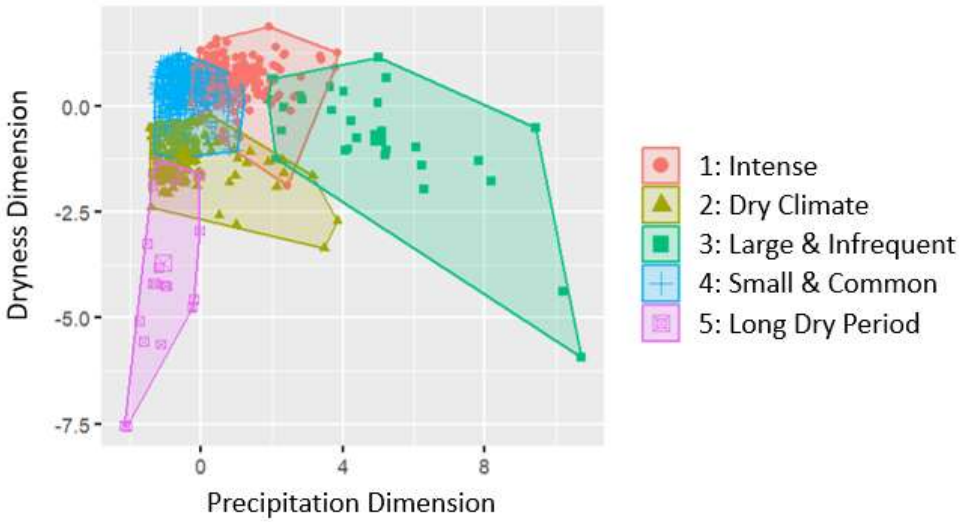


Figure 26: Labeling of storm clusters using PCA visualization to simplify dimensionality in two dimensions: precipitation and dryness. Each storm cluster assignment from the k-means clustering algorithm with $k = 5$ is labeled. The intense storm group is in red circles, dry climate storm group in gold triangles, large & frequent storms in green squares, small & frequent storms in blue plusses and long dry period storms in unfilled pink squares.

The storm cluster groups followed logical groupings summarized by the titles: intense, dry climate, large & infrequent, small & common, and long dry period (Table 7). The intense storm group contained the second highest storm depth and the lowest number of antecedent dry days, as well as the highest average peak hourly intensity. The dry climate group contained the highest average Budyko aridity index, a higher return period than other groups with similar depths, and the second lowest peak hourly intensity. The large & infrequent group contained both the largest depth and return period averages and the second highest peak hourly intensity. Conversely, the small & common storm group contained a low storm depth and return period and had most wet-climate leaning dryness index. Finally, the last storm group, long dry period, contained the longest antecedent dry period, and second driest average climate close behind the dry climate group. The most

frequently observed storm type was the small & frequent storm group. Intense and dry climate groups were more like the small & frequent storm groups than the large & infrequent group and the long dry period group. Both the large & infrequent group and the long dry period group were more variable within their traits than the other clusters (Figure 26).

Table 7: Average characteristics of clustered storm groups for each of the five clustered groups and the manually labeled outliers. Counts of the number of unique storm events in each cluster (n) and the average depth, return period, peak hourly intensity, antecedent dry days, and climate index are provided for each storm group. A summary name is provided for each group based on its defining attributes.

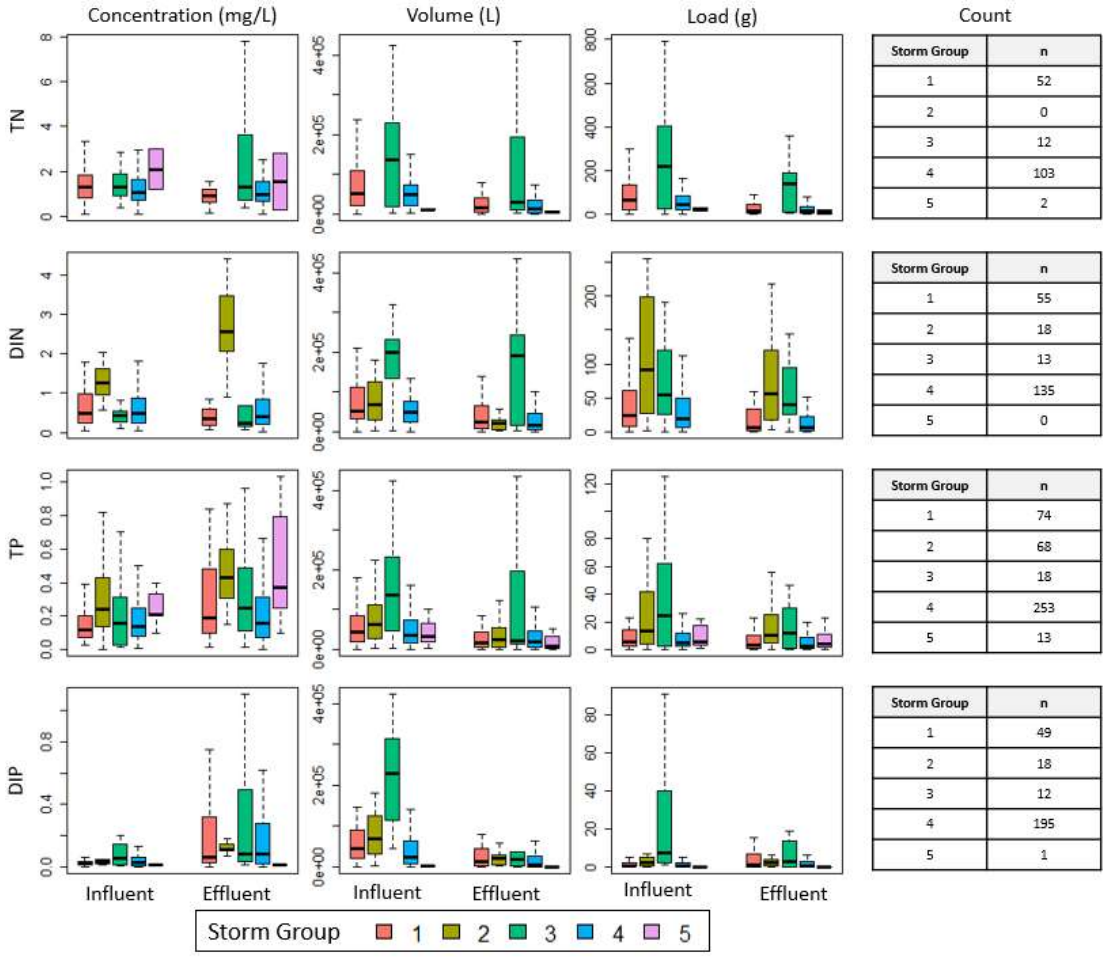
Group	n	Depth (cm)	Return Period (yr)	Peak Hourly Intensity (cm/hr)	Antecedent Dry Days (days)	Climate Index	Summary Name
1	170	3.22	0.46	1.81	2.52	1.17	Intense
2	163	1.74	0.55	0.51	2.88	2.87	Dry Climate
3	31	10.05	5.25	1.51	3.26	1.11	Large & Infrequent
4	493	1.71	0.29	0.55	3.22	0.95	Small & Common
5	28	1.68	0.35	0.47	30.38	2.45	Long Dry Period
Outlier	2	17.50	82.21	2.40	5.53	1.69	Outlier

5.3.5 Influent and Effluent Evaluation by Storm Group

The concentration, volume, and load variability were compared among the five identified storm groups for each nutrient analyte, revealing a general trend of high influent and effluent volumes and loads for large & infrequent storms. This analysis was conducted on all three of the BMP type groups (bio-composite, detention basin, and wetland channel). However, only results pertaining to bio-composite BMPs are presented

in main text (Figure 27) as these BMP types were also explore in Objective 2. Figures containing for the other BMP types are provided in Appendix C3, Figure C3.1 for detention basins and Figure C3.2 for wetland channels.

Large & infrequent storms had the most different volume and TN loads in influent and effluent from any other storm group (Figure 27). Concentrations of TN ranged from 0.8 mg/L to 3 mg/L across influent and effluent in for all storm groups. Long dry period storms had the highest median TN influent concentration (2.1 mg/L) and effluent concentration (1.6 mg/L) among storms groups. However, this result was based on only two observations of long dry period storms, and thus may be a product of low sample size. The volume of storms monitoring TN was more variable among storm groups for influent than effluent. For example, median influent volumes ranged from ~11,000 L (long dry period storms) to ~136,000 L (large & infrequent storms), while effluent volumes tended to be smaller in magnitude and range, with a lowest median effluent volume of ~4,000 L (long dry period storms) and highest median volume of ~29,000 L (large & infrequent storms). Like volumes, TN loads were also generally largest for large & infrequent storms and smallest for long dry period storms, for both influent and effluent. Median loads for the intense storm group and small & common storms were similar in effluent (median 15 g for intense storms and 16 g for small & common storms), despite having varied influent loads (63 g in intense storms, 45 g in small & common storms). Overall, TN volumes and loads from large & infrequent storms were notably higher than other storms' influent and effluent observations.



Storm Group	n
1	52
2	0
3	12
4	103
5	2

Storm Group	n
1	55
2	18
3	13
4	135
5	0

Storm Group	n
1	74
2	68
3	18
4	253
5	13

Storm Group	n
1	49
2	18
3	12
4	195
5	1

Figure 27: Comparison of TN, DIN, TP and DIP concentrations, volumes, and loads in influent and effluent by storm groups for bio-composite BMPs. Counts of the number of paired influent and effluent data points per storm group are also provided in accompanying tables. Storm group labels 1-5 represent the 5 storm groups assigned by the clustering algorithm: 1) intense, 2) dry climate, 3) large & infrequent, 4) small & common, and 5) long dry period. Boxplots display the interquartile range of observations with whiskers indicating minimum and maximum values without outliers.

Storm groups with the highest water quality measures of DIN varied for concentration, volume, and load. Median concentrations of DIN ranged from 0.2 - 1.2 mg/L across all influent storm groups, and three of the four observed effluent storm groups. Effluent concentrations in dry climate storms were the sole storm group with higher concentrations than the other storm groups, with a median concentration of 2.5

mg/L, roughly double the median influent concentration for dry storms. Volume observations also indicated one storm group with higher values than other groups, the large & infrequent storm group. Large & infrequent storms had a median influent volume around four times as large as median influent volumes for the other storm groups (200,000 L for large and infrequent, compared to 53,000, 68,000, and 49,000 for intense, dry climate, and small & common storms, respectively). This trend was even more extreme for effluent volumes than influent, as the median effluent volume for large & infrequent storms was roughly 8 times larger than the effluent volumes of the other three monitored storms (192,000 L for large and infrequent, compared to 25,000 L, 22,000 L, and 15,000 L for intense, dry climate, and small & common storms, respectively). Unlike TN where the highest volume storm groups yielded the highest loads, for DIN the storm group with the highest loads was the storm group with the highest concentrations, the dry climate group. While dry climate storms had the highest median DIN load for influent (92g) and effluent (56g), this storm group was closely followed by the large & infrequent storm group (influent DIN load 54g, effluent 40g). In contrast, both intense and small & frequent storms had influent loads of ~20g, and effluent loads ~5g. In general, the dry climate storms had highest influent and effluent DIN concentrations, large & infrequent storm had highest influent and effluent storm volumes, and both these storm groups had elevated loads compared to the other 2 observed storm groups.

TP observations were like TN observations in that concentrations were variable by storm group, but large & infrequent storms had the highest influent and effluent volumes and loads. Median concentrations of TP influent (range 0.12 – 0.24 mg/L) were generally lower than the range of median effluent concentrations (0.16-0.43 mg/L). Dry

climate storm TP concentrations had the highest influent and effluent median concentration, but the interquartile range of the long dry period storm group effluent concentration exceeded the interquartile range of the dry climate storm group effluent TP concentrations. The median influent volume for TP measurement was highest for large & infrequent storms at 136,000 L, more than double the volume of the next largest median volume from a storm group (63,000 L for dry climate storms). However, the median effluent volume for large & infrequent storms was similar to the median volumes of the other storm groups (21,000 L for large & infrequent storm, other groups ranged from 8,000 to 24,000 L). Loads of TP were higher in influent and effluent of dry climate and large & infrequent storms than the other 3 storm types. Influent TP loads were 24 g and 13 g and effluent loads were 11 g and 10 g for the large & infrequent and dry climate storms respectively. Median influent loads in intense, small & common storm groups, and long dry period storms were all ~5g, and median effluent loads were all ~3g. Overall, TP had higher concentrations in effluent than influent, high volumes in large & frequent storms, and high loads in dry climate and large & infrequent storms.

DIP concentrations and loads were the lowest of the 4 analytes, and large & infrequent storms generally had the highest concentrations, volumes, and loads. Median DIP concentrations for all storms but long dry period (which had only one storm observation) varied from 0.02 to 0.05 mg/L in influent and 0.05 to 0.11 mg/L in effluent. In both influent and effluent ranges, large & infrequent storms had the highest median concentrations. Large and infrequent storms also had the highest influent volume associated with DIP (median 228,000 L), which was much higher than the other median influent volumes (22,000 – 68,000 L). Effluent volumes were much smaller and more to

each other similar, ranging from 6,000 L (small & common storms) to 22,000 L (dry climate storms). Finally, median DIP loads were highest in large and infrequent storms for both influent (7.5 g) and effluent (2.7 g). Influent loads for large & infrequent storms were more variable than the other storm groups, with an interquartile range of 2.0 to 39.9 g. In summary, the large & infrequent storm group most frequently had the highest and most variable concentration, volumes, and loads of DIP.

5.3.6 Water Quality Treatment by Storm Group

The five storm groups performed similar concentration, volume, and load removal except for the large & infrequent storm group which had lower volume removal, and subsequent lower load reductions than the other storms for TN, DIN, and TP.

Concentration removal fractions were highly variable within storm clusters, and although median values of removal fraction varied, the overall distribution of removal fractions was not significantly different for any storm group for any analyte (Figure 28). While there was consistency in removal fractions across storms measuring the same analyte, there were differences in what that removal fraction was by analyte. TN and DIN concentrations were closely split by the threshold between retention and leaching while TP and DIP concentrations had more consistently high effluent concentration removal fractions indicating more frequent elevated effluent concentrations compared to influent concentrations.

The median effluent volume fraction was consistently less than 1 for all storms and analytes, indicating consistent median volume reduction across storms. Notably, some storm events did indicate higher effluent volume than influent volume. The two most severe combinations of volume leaching were for DIP volume measurements in

intense storms, where 75.5% of entries had higher flow volumes measured for effluent than influent, and DIN observations of large & infrequent storms, where 46.2% of entries had higher flow volumes measured for effluent than influent.

The load reduction of TN, DIN, and TP was different for large, infrequent storms (group 3) vs. all other storm groups (except for long dry period group for TN, which was similar). Like median effluent volume fractions, all median effluent load fractions were less than 1, indicating a tendency to remove nutrient loads across all storms and analytes. However, leaching of load was more common than leaching of volume by a small margin for DIN and TN, and by a larger margin for DIP and TP. For example, out of 426 TP events, 24% leached volume, and 36% leached TP load (Appendix Table C4.2). Across most storms, concentration removal was the least reduced metric with variable leaching and retention; volume and load were consistently reduced, but to different degrees for different storms.

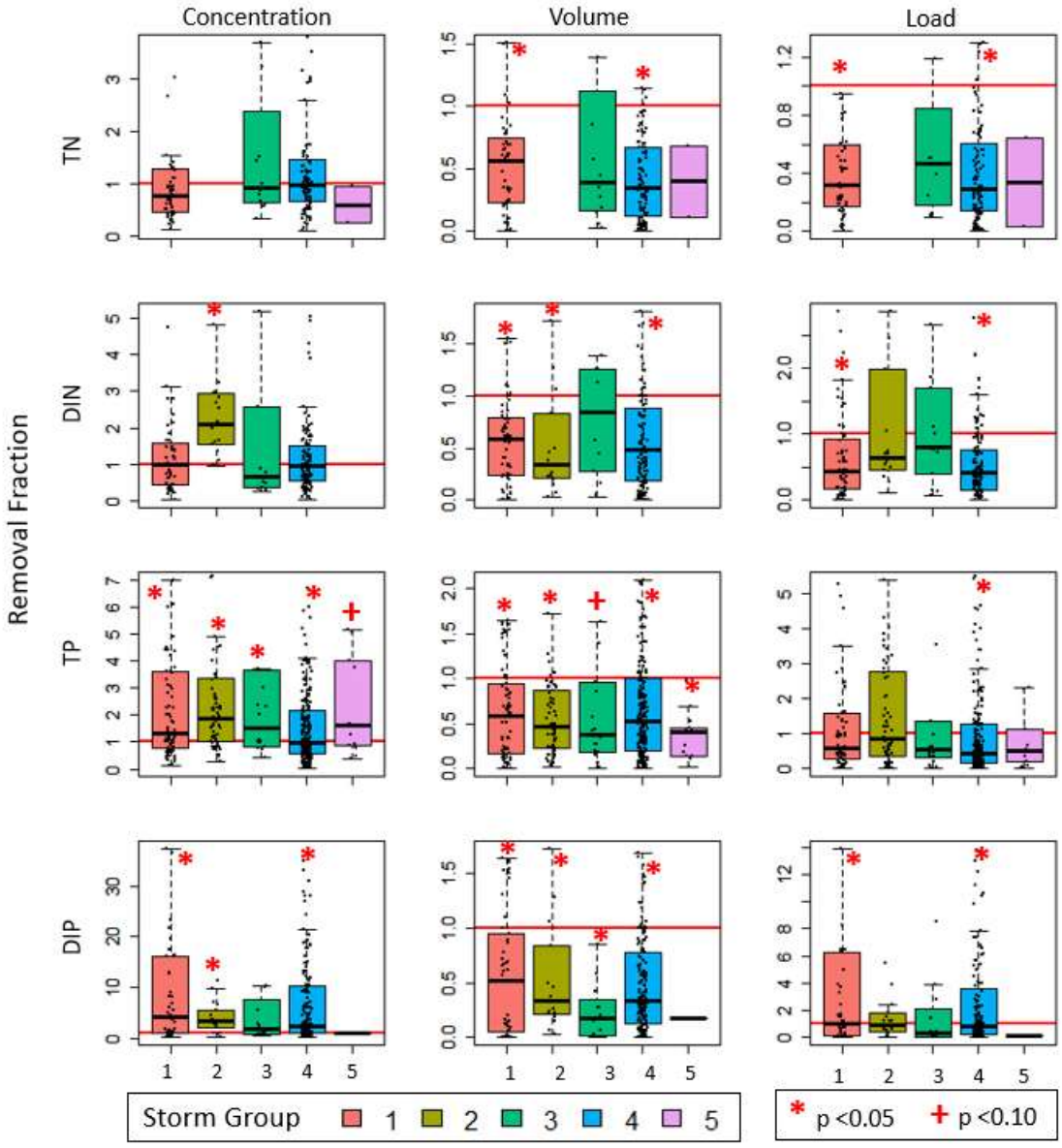


Figure 28: Comparison of removal fractions by storm groups for TN, DIN, TP, and DIP by concentration, volume, and load measures for bio-composite BMPs. Boxplots of the distribution of removal fraction for each storm group display the interquartile range with whiskers extending to the minimum and maximum values without outliers. All removal fractions are plotted over boxplots in black filled circles to display the frequency and distribution of observations. Markers over boxplots indicate p-values results of a statistical test that a distribution is not equal to 1, the threshold of leaching and retention.

Intense storms reduced TN and DIN concentrations but increased DIP and TP concentrations. Volume under intense storms was reduced by slightly less than half for all

analyte entries (effluent reduced to 52 – 58% of influent volume, Appendix Table C4.3). Loads under intense storms were reduced for TN, DIN, and TN. However, intense storms increased DIP load, the only storm group to have a median increase in load (Table 8). For DIP, intense storms were responsible for both the worst median concentration change, and median load change, where there was leaching of DIP both through concentration and load.

Dry climate storms induced the worst concentration reduction for DIN and TP and the worst TP load reduction. Dry climate storms approximately doubled the DIN and TP and more than tripled the DIP concentration in effluent compared to influent (Table 8). However, volume for these storms was consistently reduced by more than half. This resulted in load reductions for all measured analytes, although compared to load reductions for other storms groups, the dry climate storms load reduction was worst for TP and second worst for DIP and DIN. This highlights a tendency of concentration increase and low P load reductions for dry climate storms. There were no observations of a dry climate group storm with TN data.

Large & infrequent storms induced the worst observed N load removals but was one of the better P load removal storm groups (Table 8). TN and DIN concentrations decreased during large and infrequent storm events, while TP and DIP concentrations both increased. Volume change was highly variable for large and intense storms between the different analytes. Load reductions were worst across any storms for both TN and DIN, but for TP this storm group had the second-best load reduction and DIP had the best reduction of any storm group.

Table 8: Median removal fractions by storm group for concentration, volume, and load of each analyte. Removal fractions greater than one indicate median leaching tendency (effluent values higher than influent values) while removal fractions less than one indicate median retention tendency (effluent values lower than influent values). A dash indicates no storms observed, and NA indicates less than 3 storms were observed.

Analyte	Metric	1: Intense	2: Dry Climate	3: Large & Infrequent	4: Small & Common	5: Long Dry Period
TN	Concentration	0.77	-	0.93	0.97	NA
	Volume	0.57	-	0.39	0.35	NA
	Load	0.32	-	0.46	0.30	NA
DIN	Concentration	0.98	2.10	0.67	0.95	-
	Volume	0.59	0.34	0.85	0.49	-
	Load	0.45	0.64	0.81	0.41	-
TP	Concentration	1.33	1.87	1.53	0.95	1.62
	Volume	0.58	0.47	0.39	0.53	0.40
	Load	0.58	0.87	0.54	0.44	0.53
DIP	Concentration	4.20	3.22	1.68	2.17	NA
	Volume	0.52	0.34	0.18	0.34	NA
	Load	1.06	0.98	0.35	0.79	NA

Storms with long antecedent dry periods could only be analyzed for TP. Concentration was increased by long antecedent dry period storms, and this increase was the second most severe TP concentration leaching, behind only dry climate storms (Table 8). Volume was reduced to less than half of the influent volume, and TP load was subsequently reduced to about half of the influent load.

5.4 Discussion

5.4.1 Quantile vs. Clustering

While the isolated storm trait analysis and clustering analysis are similar methods to study the impact of storms on BMP nutrient management, the clustering analysis

accomplished both a simpler and more realistic means of studying the impact of storms on BMP nutrient management. The clustering analysis reduced the number of components to track statistical relationships from the four used in this analysis to one composite grouping that encompassed all four storm traits plus consideration of climate. The number of statistical tests applied for each method demonstrates the impact of this simplification. For the isolated storm trait (quantile boxplot) analysis, Tukey's test comparing each of the 5 bins (4 quantiles and an outlier bin) against each other, accounted for 10 statistical tests of difference for each of 4 storm traits, summing to 40 statistical tests to understand one analyte at one BMP type. The clustering method only required a maximum of 10 statistical tests to compare each of the five storm groups.

Clustering not only simplified the number of relationships and statistical differences to consider, but clustered storm groups are also more realistic for true storm conditions. True storms in practice cannot be simplified through one sole driver, as their other attributes would never be held constant to view that attribute in isolation. For example, examining precipitation depth alone across multiple storms cannot provide reliable information about the influence of precipitation depth as these storms inherently have different intensities, times-to-peak, antecedent dry periods, or other weather or climatic attributes that are relevant to BMP performance (Voter & Loheide, 2021). While these attributes have been the primary focus of past research, this framework was developed to align with the BMP design process of fitting infrastructure to design storms which are typically return period- or depth-based. This method may not reflect true storm conditions (Voter & Loheide, 2021). Clustering encompasses the true complex multi-

dimensionality of storms by creating groups that encompass multiple storm characteristics.

Analysis of isolated storm characteristics helped to provide baseline expectations of how these characteristics contribute to overall storm performance. Most notably, analysis of isolated traits revealed strong positive relationships between both precipitation depth and return period, and increasing volumes and loads of both influent and effluent. This relationship, and other relationships specific to individual BMP types and analytes, helped to provide context to the observations in the storm clusters based on the contribution of the different storm traits to their storm cluster group. For example, storm group 1, defined by its high intensity, has notably poor volume reduction for DIP observations (volume leached for 75.5% of storms with DIP observations, Appendix C4.3). This observation is supported by the boxplot of volume removal fraction for DIP volumes, which shows a general trend of progressively worse volume removal for increasing peak hourly intensity across quantiles.

5.4.2 Unexpectedly Consistent Volume Treatment

Volume is an important component of water quality treatment, responsible for load reductions through infiltration. Previous research has agreed that increasing storm size, viewed either though increasing total precipitation depth (Giese, Rockler, Shirmohammadi, & Pavao-zuckerman, 2019; Hoss et al., 2016; Lammers et al., 2022; Shrestha et al., 2018) or increasing return period (Fry & Maxwell, 2017; Hu, Zhang, Li, Yang, & Tanaka, 2019) is negatively related to BMP volume reduction performance on the site, catchment, or watershed scale. However, in the quantile distributions of both depth and return period examined here, there was little change in percent volume

reduction with increasing depth or return period (Table 9). For example, all ranges of storm depth (quantiles 1-4 and outliers) volume reduction percent ranged from mid-twenty percent volume reduction (or lower) to about 80% volume reduction.

Table 9: The interquartile range (25-75%) of volume reduction percentages for quantile groups of depth, return period, peak hourly intensity, and antecedent dry period. Quantile groups are assigned based on the range of values observed for each storm characteristics, as presented in section 5.2.2.1.

	Q1	**Q2**	**Q3**	**Q4**	**OUTLIERS**
Depth	3 – 81%	14 – 84%	27 – 84%	20 – 84%	15 – 84%
Return Period	23 – 80%	14 – 84%	30 – 86%	20 – 85%	33 – 82%
Peak Hourly Intensity	22 – 82%	8 – 79%	18 – 86%	16 – 86%	-132 – 65%
Antecedent Dry Period	9 – 78%	11– 84%	14– 85%	21 – 88%	57 – 88%

This contrasting observation may be indicative that the negative relationship observed in other studies does not follow a linear trend when applied across BMPs or that volume reduction is so different at different BMPs that this trend is lost to the influence of other BMP attributes.

While many studies report a negative trend with increasing storm size and decreasing volume reduction performance, these conclusions tend to be clear linear or exponentially decreasing relationships for model studies of the impact of BMPs at the catchment or watershed scale (Giese et al., 2019; Lammers et al., 2022; Palla & Gnecco, 2015). These models may rely on a process-based framework that oversimplifies the storage depth or pore water saturation of BMPs, key means of BMP volume reduction (Palla & Gnecco, 2015; Qin, Li, & Fu, 2013). For studies forming this conclusion from

field observations, a threshold is generally used to separate large and small storms and summarize performance between these groups. Such thresholds have included storms above and below 2 cm (Woznicki, Hondula, & Jarnagin, 2018), 1 cm, (Shrestha et al., 2018), or comparison of a 2- and 100-year storm (Hu et al., 2019). These studies may have found differences across their stated thresholds indicating that larger storms have lower volume reduction than small storms. However, this does not mean that there is a stepwise decrease in performance with increasing storm depth or return period. Thus, model inaccuracy and simplified field observations may be responsible for repeated reporting of overgeneralized relationships between volume reduction and storm depth or return period.

The absence of evidence of difference in volume reduction in this study may also be due to the comparison of many different BMPs. While bio-composite BMPs share common traits like soil media fill, and vegetative features, they likely vary in specific design parameters, soil median characteristics, drainage area characteristics, vegetation species, maintenance, and ages, all of which have been shown to play a role in BMP hydrologic performance (Ahiablame, Engel, & Chaubey, 2012; Skorobogatov et al., 2020; Vijayaraghavan et al., 2021). Thus, while the influence of storm size may be seen at the site level, or across controlled models, it is possible that this relationship is not strong enough to be observed in volume removal rates alone when comparing multiple sites with varying other design and local attributes. This explanation can be explored in future work, where individual sites in the BMP database could be isolated and analyzed with the same framework applied in this work to evaluate if the anticipated relationships

between storm depth and return period are apparent within observations of the same BMP.

5.4.3 Water Quality Performance Drivers Associated with Storm Groups

Examining the removal fractions and percent of leaching storm events through the lenses of concentration and load revealed certain storm types for which each analyte was most vulnerable to leaching conditions (Table 10). Potential drivers of these poor water quality treatment conditions are discussed by each analyte.

Table 10: Comparison of the storm groups that caused the poorest treatment condition for each analyte through different metrics: concentration removal, load removal, frequency of concentration increase event, and frequency of load leaching event.

	TN	DIN	TP	DIP
Concentration Removal	4	2	2	1
Load Removal	3	3	2	1
Frequency of Concentration Leaching Event	3	2	2	2
Frequency of Load Leaching Event	3	3	2	1

5.4.3.1 TN Vulnerability to Large & Infrequent Storms

TN was most vulnerable to large & infrequent storms. TN had the lowest concentration reduction for small & common storms, but under large & infrequent storm conditions, TN had its lowest load reduction and the highest likelihood of a concentration or volume leaching event. This is likely due to the hinderance of volume reduction for large & infrequent storms, which were also the storms with volumes reduced the least for TN. As TN is primarily reduced through sedimentation, hydraulic retention time is key to TN removal (Mai et al., 2022; Valenca et al., 2021). Thus, these large & infrequent storms may overflow bioretention systems, which are designed for small & frequent

storms (L. E. McPhillips et al., 2021), and therefore not allow for sufficient retention time to induce sedimentation as flow instead bypasses to piped drainage.

5.4.3.2 DIN Concentration and Load Divergence

DIN concentration and load removal metrics were both most susceptible to leaching for different storm types. DIN concentration had the lowest removal margin and was most frequently found to have net concentration increases in dry climate storms. Like TN, DIN loads leached most, and leaching events were most likely to occur in large & infrequent storms.

DIN concentration increases in dry climate storms may be due to the influence of arid climate biogeochemistry, which promotes a higher of partitioning of N as the highly mobile DIN form than humid climate soils (Austin et al., 1999; Austin & Vitousek, 1998; McCulley et al., 2009). This DIN accessibility is attributed to the rapid conversion of organic N to inorganic N in arid climates (Aranibar et al., 2004). Additionally, arid climate soils have reduced N-oxide gas emissions due to extended dry periods preventing prolonged anoxic conditions for denitrification (Aranibar et al., 2004). This hampered arid-climate denitrification would hamper a major pathway for nitrate removal in BMPs (Morse et al., 2017; Norton, Harrison, Keller, & Moffett, 2017), and result in an accumulation of nitrate in soils, a highly mobile form of N susceptible to leaching. Thus, the combined climatic effects of increased organic to inorganic N conversion and dampened denitrification likely lead to more DIN concentration leaching for arid climate storms. This concentration risk likely does not translate to load because of high-performing volume reduction for arid climate storms, which had the highest volume reduction of any DIN storm group.

DIN load leaching is likely a product of high inflow volumes from large storm events. Volume reduction for the large storm group was lowest of any storm group, and there was a positive relationship between increasing storm depth and DIN load in influent, highlighting that more DIN enters bio-composite BMPs during large storm events, and that volume reduction is dampened compared to other storms. Further, DIN is composed in part by nitrate, an anion well known to be labile in soils. Thus, the DIN leaching from large storms may be caused by the mobilization of DIN from the drainage area and flushing of DIN from the BMP engineered soil media, the reduced volume reduction from large storms, or the compounding effect of the two.

5.4.3.3 TP Dry Climate Vulnerability

All four metrics of storm group performance indicated worst TP removal from arid climate storms. The observation of poor TP removal in arid climate storms agrees with the results from Objective 2, finding higher DIP and TP concentration increases through arid-climate BMPs. Therefore, the same explanations of those concentration increases are likely also true here, where P is higher in arid climate effluent than influent because of arid climate P weathering or reduced plant uptake, higher P in arid climate influent than humid climate influent, or the influence of designs with high P-containing engineered fill media confounding the results in arid-climate BMPs. Unlike the DIN observations where high volume reductions negate elevated DIN concentrations, TP volume reductions were similar in arid climate storms to the other storm groups. This is unexpected as volume reduction for similar storms should not be different depending on which analyte was measured in a storm event. Further work is needed to determine why volume reduction for TP-monitoring storms in the arid climate storms group did not see

increased volume reduction as expected both from other analytes in this work, and in other work (Lammers et al., 2022).

5.4.3.4 DIP High Intensity Vulnerability

DIP was most likely to have events with leaching DIP concentrations in arid climate storms, but all other performance metrics indicated the worst leaching and most likely DIP leaching events in intense storms. The high frequency of DIP leaching in arid storms is likely due to many of the same drivers discussed for TP in section 5.4.3.3. Interestingly, only DIP had lowest performance metrics for the intense storm group. DIP retention in BMPs is closely related to the engineered soil media's soil P index (or soil P content) and cation exchange potential of soil, as adsorption is widely regarded as the best means of DIP retention (Hunt et al., 2006; LeFevre et al., 2015; Li & Davis, 2016). It is therefore likely that high intensity storm groups either contribute to elevated P loss of soil P from the engineered media, or that reversible surface-level adsorbed DIP is released in high-intensity storm events. The elevated DIP load leaching and higher frequency of load leaching for intense storms may also have to do with the compounding effects of poor hydrologic treatment in intense storm events, as volume reduction for DIP-observing storms in this study was lowest for intense storms, and many other studies have noted a negative relationship between intensity and BMP hydrologic performance (Jennings, 2016; Lammers et al., 2022; Sun, Pomeroy, Li, & Xu, 2019).

5.4.4 Water Quality as Concentration vs. Load

When contextualizing the water quality vulnerabilities here, it is important to highlight the different patterns and vulnerabilities when water quality is viewed through the context of concentration versus load. Federal and state permits restricting pollution

discharge to water bodies from point sources typically provide limits as concentrations, and concentrations are also typically used when stating thresholds of aquatic ecological impact. In the larger goal of protecting water bodies, total maximum daily loads (TMDLs) are used to assist in developing point source discharge limits, and as motivation of BMP implementation. Point source regulations as concentrations are feasible because they can easily be converted to loads because of consistent or monitored effluent volumes. However, stormwater volumes in influent and effluent are variable, so concentrations are not as easily indicative of loads. For stormwater BMPs, quantifying water quality performance through load reduction is directly beneficial because quantification of loads more easily translates to contributions towards TMDL goals. Further, stormwater post construction standards (often expected to be achieved through BMPs) use vague language when stating water quality standards. For example, in Wisconsin, 80% total suspended solids reduction is required without specifying concentration or load, and in Michigan the requirement is to treat the first inch of runoff or 90% of all runoff, with no definition of what "treat" means. Thus, both concentration and load measurements of BMP performance are important for assessing BMPs from a regulatory and systematic perspective.

This work demonstrates that BMP performance efficiencies and vulnerabilities differ depending on the context of concentration or load. The preliminary plot of concentration, volume and load against individual storm traits showed clear relationships with volume and load to depth and return period, but no relationship with concentration. Further, the storms most likely to cause leaching varied depending on if that leaching was considered as concentration elevation or load elevation for TN and DIN. Therefore,

future work should monitor water quality through both frameworks, or be very explicit about which framework is being used as conclusions regarding concentration and load cannot be assumed to be the same.

5.5 Conclusions

Results from this objective show that storm conditions impact stormwater influent, effluent, and treatment efficiency for nutrient concentrations, volumes, and loads. This conclusion was reached by addressing two tasks, analyzing stormwater influent and effluent concentrations, volumes, loads, and the changes in these metrics across variable storm traits, and among 5 clustered storm groups. Performing both tasks demonstrated that clustering methods can reduce the quantity of statistical tests and offer more realistic means of comparing BMP performance under varying storm conditions. Towards the objective's goal of quantifying the impact of storm characteristics on BMP nutrient management, this work revealed consistent volume treatment across diverse storm conditions, and a divergence of performance trends depending on the water quality framework of concentration or load. Finally, this objective identifies and contextualizes the storm condition where nutrient treatment is least effective, , identifying TN vulnerability to large & infrequent storms, DIN and TP vulnerability to arid climate storms, and DIP vulnerability to intense storms.

6. CONCLUSIONS

Water is essential to life. Yet in urban areas where most of the world's population resides, waterways are plagued by urban stream syndrome. As with any syndrome, combatting it must entail both knowledge of a syndrome's spread, and understanding of treatments. This dissertation increases scientific understanding of the spread of urban stream syndrome in a socio-economic context and contributes to increased understanding of the treatment of urban stream syndrome with stormwater BMPs in diverse environmental conditions.

6.1 Key Findings

Objective 1 explored anthropogenic effects on water quality and the heterogeneous distribution of degraded streams in the urban watershed of the Rouge River in metropolitan Detroit, Michigan. Benthic macroinvertebrate data collected by volunteer scientists was aggregated into a Stream Quality Index (SQI) to define long-term water quality patterns. Spatial dependence of the data was assessed with spatial stream network models incorporating socio-economic and environmental predictors. The best model included poverty as an explanatory variable with a negative relationship with stream quality. SQI predictions under true watershed conditions revealed a 1% decrease in SQI associated with 1% increase in poverty. This work demonstrated the benefits of volunteer science and spatial modeling methods for urban stream modeling. The finding of inequitably distributed water quality impairment in urban streams underscores the importance of focused restoration in economically oppressed urban areas.

Objective 2 conducted a meta-analysis of vegetated BMPs in the International Stormwater BMP Database and compared influent and effluent nitrogen and phosphorus

concentrations to quantify BMP effect on nutrient management across climates. BMP effect on nutrient concentration change was compared between vegetated BMPs in wet and dry climates. Meta-analysis with subgroup analysis determined differences between wet and dry climates and among vegetated BMP types. Across both wet and dry climates BMPs leached DIP and TP, increased the fraction of dissolved inorganic P (DIP:TP), and decreased dissolved N:P ratios. Dry climate BMPs leached DIP and TP more consistently and at a higher magnitude than wet climate BMPs, and bioretention leached more DIP than grass strips and swales.

Objective 3 quantified the impact of diverse storm conditions on BMP water quality performance through both a concentration and load framework. Like Objective 2, water quality treatment performance data from the International Stormwater Database was used in conjunction with external storm characteristic data. Quantile-based comparisons of water quality metrics and storm characteristics showed that depth and return periods increased volumes and loads of influent and effluent for all N and P species. Clustering storm characteristics revealed five distinct storm types, which featured different BMP treatment between large & infrequent storm event and all other storms. However, closer investigation of the removal performance and likelihood of leaching events revealed that each analyte had a distinct vulnerability to a storm group. This objective revealed that arid climate storms are most likely to produce elevated DIN and TP concentrations relative to load, but that high hydrologic performance for these storms negated the concentration leaching effect. Further, TN, DIN, and TP load removal is achieved for all storm types, but is least effective for large & infrequent storms for N

species and for arid climate storms for TP. The worst storm condition for DIP was intense storm, which tended to leach DIP load.

6.2 Future Work

Urban water quality degradation is a complex problem, and this work advances our understanding of the socio-economic distribution of urban stream syndrome, and the reliability of BMPs in diverse climates and storm conditions. However, this work offered three major preliminary findings and further work is necessary to contextualize their impact, and ultimately work to apply this knowledge toward improving urban water quality.

First, objective one provided new information about the disparate distribution of degraded stream quality in high poverty areas. A model showed that this relationship existed across multiple future watershed scenarios, where watershed characteristics (sedimentation clarity and impervious land use) were universally improved across the watershed. However, it would be a worthwhile task to model the resulting water quality if watershed conditions were improved in targeted high-poverty areas. This could be done by modeling higher quality sedimentation and impervious land use conditions specifically in high poverty areas and examining the modeled change in stream quality under these improved conditions. This research could also integrate socio-economic water quality disparities with watershed-level BMP implementation. Existing modeling efforts have studies the impact of BMP placement for watershed- level water quality improvements (Hung et al., 2020; Pennino, McDonald, & Jaffe, 2016), but modeling hot-spot BMP implementation with socio-economic considerations would be a beneficial next step towards creating more equitable watersheds.

The second objective illustrated that phosphorus concentrations in stormwater increase through BMPs and shows that this increase is more severe in dry climates. Defining this sensitivity of P concentration management in dry climates is an important contribution towards understanding urban nutrient management in stormwater under a changing climate. However, understanding the climate variability of BMP water quality performance is complex and much work is still needed to fully understand this topic. Namely, water quality management should also consider the load reduction of BMPs, and more information is needed about the causational drivers behind the observations presented in Objective two. First, objective two focuses entirely on concentrations as the water quality metric and does not make any conclusions regarding load change. Concentrations are impacted by volume reduction and nutrient retention mechanisms and can't be assumed to indicate trends in load treatment and reduction. Therefore, it would be beneficial to repeat the work presented in objective two with load data to determine the impact of climate on BMP load performance, a water quality metric important to monitor for total maximum daily load regulations. Second, objective two presented multiple causational structures that may explain the concentration changes observed including elevated influent concentrations, arid climate biogeochemistry, and varying BMP goals, designs, and regional dominance. However, distinguishing between these causational structures remains difficult, thus further research regarding BMP nutrient treatment performance is needed, especially for nitrogen species in arid climates, where data was limited in this study. In particular, ample data already exists to characterize stormwater quality by region, and could be extended to a study of variable stormwater

chemical composition by regional aridity (Balderas Guzman et al., 2022; Voter & Loheide, 2021).

Finally, objective three addressed a key gap from objective two – by presenting results about varying concentrations, loads, and changes in both concentration, load, and volume in different storm conditions. This objective contains two major contributions, both of which require further research. First, this objective showed that while concentration increase is common in multiple different storm types, this leaching of concentration largely did not translate to loads. This conversation of stormwater treatment through both the lenses of concentration and load is present in some literature regarding site-level performance but not yet presented explicitly for a large regional study (Hunt et al., 2012; Selbig, 2016; Smith et al., 2020). Second, objective three also presents a storm condition most vulnerable to poor nutrient removal performance for each analyte examined, but this data summarizes the performance of many BMPs of different ages, designs, types, drainage area characteristics, to name a few factors which may influence BMP nutrient management (Lintern et al., 2020). Thus, the conclusions of this objective should be compared to case studies of BMPs with many records of water quality treatment performance in diverse storm conditions, to determine whether the same conclusions are visible at the site level, where many of these conditions would be ubiquitous across storms.

Overall, this dissertation presented multiple contributions that further our understanding of urban water quality distribution and BMP nutrient treatment performance. The future work suggested in this section offers multiple next steps to further contextualize the implications and implementation of this work.

BIBLIOGRAPHY

Achleitner, S., Möderl, M., & Rauch, W. (2007). CITY DRAIN (C) - An open source approach for simulation of integrated urban drainage systems. *Environmental Modelling & Software*, *22*(8), 1184–1195. https://doi.org/https://doi.org/10.1016/j.envsoft.2006.06.013

Ackerman, D., & Stein, E. D. (2008). Evaluating the Effectiveness of Best Management Practices Using Dynamic Modeling. *Journal of Environmental Engineering*, *134*(8), 628–639. https://doi.org/10.1061/(asce)0733-9372(2008)134:8(628)

Ahiablame, L. M., Engel, B. A., & Chaubey, I. (2012). Effectiveness of low impact development practices: Literature review and suggestions for future research. *Water, Air, and Soil Pollution*, *223*(7), 4253–4273. https://doi.org/10.1007/s11270-012-1189-2

Ahn, J. H., Grant, S. B., Surbeck, C. Q., Digiacomo, P. M., Nezlin, Nikolay, P., & Jiang, S. (2005). Coastal Water Quality Impact of Stormwater Runoff from an Urban Watershed in Southern California. *Environmental Science and Technology*, *39*, 5940–5953.

Akhtar, N., Syakir Ishak, M. I., Bhawani, S. A., & Umar, K. (2021). Various natural and anthropogenic factors responsible for water quality degradation: A review. *Water (Switzerland)*, *13*(19). https://doi.org/10.3390/w13192660

Alamdari, N., Sample, D. J., Steinberg, P., Ross, A. C., & Easton, Z. M. (2017). Assessing the Effects of Climate Change on Water Quantity and Quality in an Urban Watershed Using a Calibrated Stormwater Model. *Water*, *9*(464), 1–24. https://doi.org/10.3390/w9070464

Alender, B. (2016). Understanding volunteer motivations to participate in citizen science projects: A Deeper look at water quality monitoring. *Journal of Science Communication*, *15*(3), 1–19. https://doi.org/10.22323/2.15030204

Allan, J. D. (2004). Landscapes and riverscapes: The influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, *35*(2002), 257–284. https://doi.org/10.1146/annurev.ecolsys.35.120202.110122

Anderson, C. M., Kissel, K. A., Field, C. B., & Mach, K. J. (2018). Climate Change Mitigation, Air Pollution, and Environmental Justice in California. *Environmental Science and Technology*, *52*(18), 10829–10838. https://doi.org/10.1021/acs.est.8b00908

Aranibar, J. N., Otter, L., Macko, S. A., Feral, C. J. W., Epstein, H. E., Dowty, P. R., … Swap, R. J. (2004). Nitrogen cycling in the soil-plant system along a precipitation gradient in the Kalahari sands. *Global Change Biology*, *10*(3), 359–373. https://doi.org/10.1111/j.1365-2486.2003.00698.x

Arnold, C. L. J., & Gibbons, C. J. (1996). Impervious Surface Coverage: The Emergence of a Key Environmental Indicator. *Journal of the American Planning Association*,

62(2). https://doi.org/https://doi.org/10.1080/01944369608975688

Austin, A. T., Sala, O. E., Schulze, E. D., Farquhar, G. D., Miller, J. M., Schulze, W., … Williams, R. J. (1999). Foliar δ15N is negatively correlated with rainfall along the IGBP transect in Australia. *Australian Journal of Plant Physiology*, *26*(3), 293–298.

Austin, & Vitousek, P. M. (1998). Nutrient dynamics on a precipitation gradient in Hawai'i. *Oecologia*, *113*(4), 519–529. https://doi.org/10.1007/s004420050405

Austin, Yahdjian, L., Stark, J. M., Belnap, J., Porporato, A., Norton, U., … Schaeffer, S. M. (2004). Water pulses and biogeochemical cycles in arid and semiarid ecosystems. *Oecologia*, *141*(2), 221–235. https://doi.org/10.1007/s00442-004-1519-1

Bae, Y. J., Kil, H. K., & Bae, K. S. (2005). Benthic macroinvertebrates for uses in stream biomonitoring and restoration. *KSCE Journal of Civil Engineering*, *9*(1), 55–63. https://doi.org/10.1007/bf02829098

Bai, X., Shen, W., Wang, P., Chen, X., & He, Y. (2020). Response of Non-point Source Pollution Loads to Land Use Change under Different Precipitation Scenarios from a Future Perspective. *Water Resources Management*, *34*, 3987–4002. https://doi.org/https://doi.org/10.1007/s11269-020-02626-0

Balderas Guzman, C., Wang, R., Muellerklein, O., Smith, M., & Eger, C. G. (2022). Comparing stormwater quality and watershed typologies across the United States: A machine learning approach. *Water Research*, *216*(February), 118283. https://doi.org/10.1016/j.watres.2022.118283

Barbour, M. T., Gerritsen, J., Snyder, B. D., & Stribling, J. B. (1999). *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macriinvertebrates, and Fish, Second Edition. EPA 841-B-99-002. Rapid Bioassessment Protocols for Use in Streams and Rivers*. Washington, D. C. Retrieved from http://www.epa.gov/OWOW/monitoring/techmon.html

Barbour, M. T., Stribling, J. B., & Verdonschot, P. F. M. (2006). The multihabitat approach of USEPA's rapid bioassessment protocols: Benthic macroinvertebrates. *Limnetica*, *25*(3), 839–850. https://doi.org/10.23818/limn.25.58

Barrett, M. E. (2008). Comparison of BMP Performance Using the International BMP Database. *Journal of Irrigation and Drainage Engineering*, *134*(5), 556–561. https://doi.org/10.1061/(asce)0733-9437(2008)134:5(556)

Barrett, M. E. ., Walsh, P. M. ., Joseph F Jr., M., & Charbeneau, R. J. (1998). PERFORMANCE OF VEGETATIVE CONTROLS FOR TREATING HIGHWAY RUNOFF. *Journal of Environmental Engineering*, *124*(November), 1121–1128.

Becker, B. J. (1988). Synthesizing Standardized Mean-Change Measures. *British Journal of Mathematical and Staistical Psychology*, *41*(2), 257–278.

Bernhardt, E. S., Palmer, M. A., Allan, J. D., Alexander, G., Barnas, K., Brooks, S., … Sudduth, E. (2005). Synthesizing U.S. River Restoration Efforts. *Science*, *308*(April), 636–638.

Bevc, C. A., Marshall, B. K., & Picou, J. S. (2007). Environmental justice and toxic exposure: Toward a spatial model of physical health and psychological well-being. *Social Science Research*, *36*(1), 48–67. https://doi.org/10.1016/j.ssresearch.2005.11.001

Blaszczak, J. R., Delesantro, J. M., Urban, D. L., Doyle, M. W., & Bernhardt, E. S. (2019). Scoured or suffocated: Urban stream ecosystems oscillate between hydrologic and dissolved oxygen extremes. *Limnology and Oceanography*, *64*(3), 877–894. https://doi.org/10.1002/lno.11081

Booth, D. B., Roy, A. H., Smith, B., & Capps, K. A. (2016). Global perspectives on the urban stream syndrome. *Freshwater Science*, *35*(1), 412–420. https://doi.org/10.1086/684940

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis* (1st ed.). John Wiley & Sons Ltd.

Borenstein, M., & Higgins, J. P. T. (2013). Meta-Analysis and Subgroups. *Prevention Science*, *2*, 134–143. https://doi.org/10.1007/s11121-013-0377-7

Bosch, N. S., Evans, M. A., Scavia, D., & Allan, J. D. (2014). Interacting effects of climate change and agricultural BMPs on nutrient runoff entering Lake Erie. *Journal of Great Lakes Research*, *40*, 581–589. https://doi.org/http://dx.doi.org/10.1016/j.jglr.2014.04.011 0380-1330/©

Brady, N. C., & Weil, R. R. (1996). *Soil Phosphorous and Potassium*. (A. Kupchik, M. Carnis, & C. Stewart, Eds.), *The Naure and Properties of Soils* (Eleventh). Upper Saddle River, New Jersey: Prentice-Hall Inc.

Brezonik, P. L., & Stadelmann, T. H. (2002). Analysis and predictive models of stormwater runoff volumes, loads, and pollutant concentrations from watersheds in the Twin Cities metropolitan area, Minnesota, USA. *Water Research*, *36*(7), 1743–1757. https://doi.org/10.1016/S0043-1354(01)00375-X

Buckingham, S. E., Neff, J., Titiz-Maybach, B., & Reynolds, R. L. (2010). Chemical and textural controls on phosphorus mobility in drylands of southeastern Utah. *Biogeochemistry*, *100*, 105–120. https://doi.org/10.1007/s10533-010-9408-7

Buffam, I., Mitchell, M. E., & Durtsche, R. D. (2016). Environmental drivers of seasonal variation in green roof runoff water quality. *Ecological Engineering*, *91*, 506–514. https://doi.org/10.1016/j.ecoleng.2016.02.044

Bullard, R. D. (1996). Environmental Justice: It's More Than Waste Facility Siting. *Social Science Quarterly*, *77*(3), 493–499. Retrieved from https://www.jstor.org/stable/42863495

Burlakova, L. E., Kovalenko, K. E., Schmude, K. L., Barbiero, R. P., Karatayev, A. Y., & Lesht, B. M. (2018). Development of new indices of Great Lakes water quality based on profundal benthic communities. *Journal of Great Lakes Research*, *44*(4), 618–628. https://doi.org/10.1016/J.JGLR.2017.11.004

Buytaert, W., Dewulf, A., De Bièvre, B., Clark, J., & Hannah, D. M. (2016). Citizen

Science for Water Resources Management: Toward Polycentric Monitoring and Governance? *Journal of Water Resources Planning and Management*, *142*(4), 01816002. https://doi.org/10.1061/(asce)wr.1943-5452.0000641

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., … Zhumanova, M. (2014). Citizen science in hydrology and water resources: Opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science*, *2*(October), 1–21. https://doi.org/10.3389/feart.2014.00026

Capps, K. A., Bentsen, C. N., & Ramírez, A. (2016). Poverty, urbanization, and environmental degradation: Urban streams in the developing world. *Freshwater Science*, *35*(1), 429–435. https://doi.org/10.1086/684945

Carey, R. O., Hochmuth, G. J., Martinez, C. J., Boyer, T. H., Dukes, M. D., Toor, G. S., & Cisar, J. L. (2013). Evaluating nutrient impacts in urban watersheds : Challenges and research opportunities. *Environmental Pollution*, *173*, 138–149. https://doi.org/10.1016/j.envpol.2012.10.004

Cettner, A., Ashley, R., Viklander, M., & Nilsson, K. (2013). Stormwater management and urban planning: Lessons from 40 years of innovation. *Journal of Environmental Planning and Management*, *56*(6), 786–801. https://doi.org/10.1080/09640568.2012.706216

Charters, F. J., Cochrane, T. A., & O'Sullivan, A. D. (2015). Particle size distribution variance in untreated urban runoff and its implication on treatment selection. *Water Research*, *85*, 337–345. https://doi.org/10.1016/j.watres.2015.08.029

Chaubey, I., Chiang, L., Gitau, M. W., & Mohamed, S. (2010). Effectiveness of best management practices in improving water quality in a pasture-dominated watershed. *Journal of Soil and Water Conservation*, *65*(6), 424–437. https://doi.org/10.2489/jswc.65.6.424

Choat, B. E., & Bhaskar, A. S. (2020). Spatial Arrangement of Stormwater Infiltration Affects Subsurface Storage and Baseflow. *Journal of Hydrologic Engineering*, *25*(11), 1–13. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002005

Clary, J., Jones, J., Leisenring, M., Hobson, P., & Strecker, E. (2020). *International Stormwater BMP Database: 2020 Summary Statistics*. Retrieved from www.waterrf.org

Cochran, W. G. (1954). Some Methods for Strengthening the Common χ 2 Tests. *Bioethics*, *10*(4), 417–451. https://doi.org/10.2307/3001616

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.

Collins, M. B., Munoz, I., & Jaja, J. (2016). Linking "toxic outliers" to environmental justice communities. *Environmental Research Letters*, *11*(1). https://doi.org/10.1088/1748-9326/11/1/015004

Cook, L. M., VanBriesen, J. M., & Samaras, C. (2019). Using rainfall measures to

evaluate hydrologic performance of green infrastructure systems under climate change. *Sustainable and Resilient Infrastructure*, *6*(3–4), 156–180. https://doi.org/10.1080/23789689.2019.1681819

Cooper, C. B., Hawn, C. L., Larson, L. R., Parrish, J. K., Bowser, G., Cavalier, D., … Wilson, S. (2021). Inclusion in citizen science: The conundrum of rebranding. *Science*, *372*(6549), 1386–1388. https://doi.org/10.1126/science.abi6487

Costello, D. M., Hartung, E. W., Stoll, J. T., & Jefferson, A. J. (2020). Bioretention cell age and construction style infl uence stormwater pollutant dynamics. *Science of the Total Environment*, *712*, 1–10. https://doi.org/10.1016/j.scitotenv.2019.135597

Daly, E, Deletic, A., Hatt, B. E., & Fletcher, T. D. (2012). Modelling of stormwater biofilters under random hydrologic variability : a case study of a car park at Monash University , Victoria (Australia). *Hydrological Processes*, *26*, 3416–3424. https://doi.org/10.1002/hyp.8397

Daly, Edoardo, Calabrese, S., Yin, J., & Porporato, A. (2019a). Linking parametric and water-balance models of the Budyko and Turc spaces. *Advances in Water Resources*, *134*(June). https://doi.org/10.1016/j.advwatres.2019.103435

Daly, Edoardo, Calabrese, S., Yin, J., & Porporato, A. (2019b). Linking parametric and water-balance models of the Budyko and Turc spaces. *Advances in Water Resources*, *134*(June), 103435. https://doi.org/10.1016/j.advwatres.2019.103435

Daneshvar, F., Nejadhashemi, A. P., Zhang, Z., & Herman, M. R. (2018). Assessing the relative importance of parameter estimation in stream health based environmental justice modeling. *Journal of Hydrology*, *563*(February), 211–222. https://doi.org/10.1016/j.jhydrol.2018.06.004

Daneshvar, F., Nejadhashemi, A. P., Zhang, Z., Herman, M. R., Shortridge, A., & Marquart-Pyatt, S. (2016). Evaluating stream health based environmental justice model performance at different spatial scales. *Journal of Hydrology*, *538*, 500–514. https://doi.org/10.1016/j.jhydrol.2016.04.052

Davis, A. P. (2007). Field Performance of Bioretention : Water Quality. *Environmental Engineering Science*, *24*(8). https://doi.org/10.1089/ees.2006.0190

Davis, A. P., Shokouhian, M., Sharma, H., & Minami, C. (2006). Water Quality Improvement through Bioretention Media : Nitrogen and Phosphorus Removal. *Water Envirornment Research*, *78*(3), 284–293. https://doi.org/10.2175/106143005X94376

Del Arco, A. I., Ferreira, V., & Graca, M. A. S. (2012). The performance of biological indicators in assessing the ecological state of streams with varying catchment urbanisation levels in Coimbra, Portugal. *Limnetica*, *31*(1), 141–154.

Demissie, Y., & Mortuza, R. M. (2016). *Development and Update of Rainfall and Runoff Intensity-Duration-Frequency Curves for Washington State Counties in Response to Observed and Anticipated Extreme Rainfall and Snow Events*.

Dietz, M. E. (2007). Low impact development practices: A review of current research

and recommendations for future directions. *Water, Air, and Soil Pollution*, *186*(1–4), 351–363. https://doi.org/10.1007/s11270-007-9484-z

Dixon, J. L., Chadwick, O. A., & Vitousek, P. M. (2016). Climate-driven thresholds for chemical weathering in postglacial soils of New Zealand. *Journal of Geophysical Research: Earth Surface*, *121*, 1619–1634. https://doi.org/10.1002/2016JF003864

Dodds, W. K., Bouska, W. W., Eitzmann, J. L., Pilger, T. J., Pitts, K. L., Riley, A. J., … Thornbrugh, D. J. (2009). Eutrophication of U . S . Freshwaters : Damages. *Environmnetal Science & Technology: Policy Analysis*, *43*(1), 12–19. https://doi.org/10.1021/es801217q

Dodds, W. K., & Smith, V. H. (2016). Nitrogen , phosphorus , and eutrophication in streams. *Inland Waters*, *6*(2), 155–164. https://doi.org/10.5268/IW-6.2.909

Emadi, M., Baghernejad, M., Bahmanyar, M. A., & Morovvat, A. (2012). Changes in soil inorganic phosphorous pools along a precipitation gradient in northern Iran. *Int. J. Forest, Soil and Erosion, 2012 2 (3): 143-147*, *2*(3), 134–147.

Engel, S. R., & Voshell, J. R. (2002). Volunteer biological monitoring: Can it accurately assess the ecological condition of streams? *American Entomologist*, *48*(3), 164–177. https://doi.org/10.1093/ae/48.3.164

Environmental Protecion Agency. (2016). *Summary of State Post Construction Stormwater Standards*. Retrieved from https://www.epa.gov/sites/default/files/2016-08/documents/swstdsummary_7-13-16_508.pdf

Erickson, A. J., Gulliver, J. S., & Weiss, P. T. (2012). Capturing phosphates with iron enhanced sand filtration. *Water Research*, *46*(9), 3032–3042. https://doi.org/10.1016/j.watres.2012.03.009

Erickson, A. J., Taguchi, V. J., & Gulliver, J. S. (2018). The Challenge of Maintaining Stormwater Control Measures : A Synthesis of Recent Research and Practitioner Experience. *Sustainability*, *10*(3666), 1–15. https://doi.org/10.3390/su10103666

Erickson, A. J., Weiss, P. T., & Gulliver, J. S. (2007). Enhanced sand filtration for storm water phosphorus removal. *Journal of Environmental Engineering*, *133*(5), 485–497. https://doi.org/10.1061/40856(200)384

Feyissa, A., Yang, F., Wu, J., Chen, Q., Zhang, D., & Cheng, X. (2021). Soil nitrogen dynamics at a regional scale along a precipitation gradient in secondary grassland of China. *Science of the Total Environment*, *781*, 146736. https://doi.org/10.1016/j.scitotenv.2021.146736

Firehock, K. and West, J. (1995). A Brief History of Volunteer Biological Water Monitoring Using Macroinvertebrates Author (s): Karen Firehock and Jay West Source : Journal of the North American Benthological Society , Vol . 14 , No . 1 (Mar ., 1995), Published by : The University of. *Journal of the North American Benthological Society*, *14*(1), 197–202.

Frieden, J. C., Peterson, E. E., Angus Webb, J., & Negus, P. M. (2014). Improving the predictive power of spatial statistical models of stream macroinvertebrates using

weighted autocovariance functions. *Environmental Modelling and Software*, *60*, 320–330. https://doi.org/10.1016/j.envsoft.2014.06.019

Fry, T. J., & Maxwell, R. M. (2017). Evaluation of distributed BMPs in an urban watershed — High resolution modeling for stormwater management. *Hydrological Processes*, (November 2016), 2700–2712. https://doi.org/10.1002/hyp.11177

Gan, G., & Kwok-Po Ng, M. (2017). k-means clustering with outlier removal. *Pattern Recognition Letters*, *90*, 8–14. https://doi.org/https://doi.org/10.1016/j.patrec.2017.03.008

Garcia, X., Barceló, D., Comas, J., Corominas, L., Hadjimichael, A., Page, T. J., & Acuña, V. (2016). Placing ecosystem services at the heart of urban water systems management. *Science of the Total Environment*, *563–564*, 1078–1085. https://doi.org/10.1016/j.scitotenv.2016.05.010

Giese, E., Rockler, A., Shirmohammadi, A., & Pavao-zuckerman, M. A. (2019). Assessing Watershed-Scale Stormwater Green Infrastructure Response to Climate Change in Clarksburg , Maryland. *Journal of Water Resources Planning and Management*, *145*(2017), 1–13. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001099

Gold, A. C., Thompson, S. P., & Piehler, M. F. (2019). Nitrogen cycling processes within stormwater control measures : A review and call for research. *Water Research*, *149*, 578–587. https://doi.org/10.1016/j.watres.2018.10.036

Gong, Y., Fu, H., Li, H., Chen, Y., Zhang, W., Wu, L., & Li, Y. (2021). Influences of time scale on green stormwater infrastructure's effect on suspended solids in urban rainfall runoff. *Journal of Hydrology*, *598*(May), 126439. https://doi.org/10.1016/j.jhydrol.2021.126439

Grabowski, Z. J., Watson, E., & Chang, H. (2016). Using spatially explicit indicators to investigate watershed characteristics and stream temperature relationships. *Science of the Total Environment*, *551–552*, 376–386. https://doi.org/10.1016/j.scitotenv.2016.02.042

Graham, M., & Taylor, J. (2018). *Development of Citizen Science Water Resource Monitoring Tools and Communities of Practice for South Africa, Africa and the World*.

Green, M. B., & Finlay, J. C. (2010). Patterns of hydrologic control over stream water total nitrogen to total phosphorus ratios. *Biogeochemistry*, *99*(1), 15–30. https://doi.org/10.1007/s10533-009-9394-9

Grimm, N. B., Sheibley, R. W., Crenshaw, C. L., Dahm, C. N., Roach, W. J., & Zeglin, L. H. (2005). N retention and transformation in urban streams. *Journal of the North American Benthological Society*, *24*(3), 626–642. https://doi.org/10.1899/04-027.1

Hale, R. L., Scoggins, M., Smucker, N. J., & Suchy, A. (2016). Effects of climate on the expression of the urban stream syndrome. *Freshwater Science*, *35*(1), 421–428. https://doi.org/10.1086/684594

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. . (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Boca Raton, FL and London: Chapman & Hall/CRC Press.

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, *7*(1), 1–18. https://doi.org/10.1038/s41597-020-0453-3

Hatt, B. E., Fletcher, T. D., & Deletic, A. (2009). Hydrologic and pollutant removal performance of stormwater biofiltration systems at the field scale. *Journal of Hydrology*, *365*(3–4), 310–321. https://doi.org/10.1016/j.jhydrol.2008.12.001

He, K., Qin, H., Wang, F., Ding, W., & Yin, Y. (2020). Importance of the submerged zone during dry periods to nitrogen removal in a bioretention system. *Water (Switzerland)*, *12*(3). https://doi.org/10.3390/w12030876

Heisler, J., Glibert, P. M., Burkholder, J. M., Anderson, D. M., Cochlan, W., Dennison, W. C., … Suddleson, M. (2008). Eutrophication and harmful algal blooms : A scientific consensus. *Harmful Algae*, *8*, 3–13. https://doi.org/10.1016/j.hal.2008.08.006

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/SIM.1186

Hobbie, S. E., Finlay, J. C., Janke, B. D., Nidzgorski, D. A., Millet, D. B., & Baker, L. A. (2017). Contrasting nitrogen and phosphorus budgets in urban watersheds and implications for managing urban water pollution. *PNAS*, *114*(16), 4177–4182. https://doi.org/10.1073/pnas.1618536114

Holcomb, D. A., Messier, K. P., Serre, M. L., Rowny, J. G., & Stewart, J. R. (2018). Geostatistical Prediction of Microbial Water Quality Throughout a Stream Network Using Meteorology, Land Cover, and Spatiotemporal Autocorrelation. *Environmental Science and Technology*, *52*(14), 7775–7784. https://doi.org/10.1021/acs.est.8b01178

Hong, C. Y., & Chang, H. (2020). Residents' perception of flood risk and urban stream restoration using multi-criteria decision analysis. *River Research and Applications*, *36*(10), 2078–2088. https://doi.org/10.1002/rra.3728

Horvath, I. R., Pulvermacher, L., & Parolari, A. J. (2022). Seasonal Hydroclimatic and Soil Biogeochemical Drivers of N and P Availability in a Constructed Stormwater Wetland. *Journal of Sustainable Water in the Built Environment*, *8*(1), 04021018. https://doi.org/10.1061/jswbay.0000962

Hoss, F., Fischbach, J., & Molina-Perez, E. (2016). Effectiveness of Best Management Practices for Stormwater Treatment as a Function of Runoff Volume. *Journal of Water Resources Planning and Management*, *142*(11), 05016009. https://doi.org/10.1061/(asce)wr.1943-5452.0000684

Hou, E., Chen, C., Luo, Y., Zhou, G., Kuang, Y., Zhang, Y., … Wen, D. (2018). Effects

of climate on soil phosphorus cycle and availability in natural terrestrial ecosystems. *Global Change Biology*, *24*(8), 3344–3356. https://doi.org/10.1111/gcb.14093

House, M. A., Ellis, J. B., Herricks, E. E., Hvitved-Jacobsen, T., Seager, J., Lijklema, L., … Clifforde, I. T. (1993). Urban drainage - Impacts on receiving water quality. *Water Science and Technology*, *27*(12), 117–158. https://doi.org/10.2166/wst.1993.0293

Hsieh, C., Davis, A. P., & Needelman, B. A. (2007). Nitrogen Removal from Urban Stormwater Runoff Through Layered Bioretention Columns. *Water Environment Research*, *79*(12), 2404–2411. https://doi.org/10.2175/106143007X183844

Hu, M., Zhang, X., Li, Y., Yang, H., & Tanaka, K. (2019). Flood mitigation performance of low impact development technologies under different storms for retro fitting an urbanized area. *Journal of Cleaner Production*, *222*, 373–380. https://doi.org/10.1016/j.jclepro.2019.03.044

Hung, F., Harman, C. J., Hobbs, B. F., & Sivapalan, M. (2020). Assessment of Climate, Sizing, and Location Controls on Green Infrastructure Efficacy: A Timescale Framework. *Water Resources Research*, *56*(5), 1–25. https://doi.org/10.1029/2019WR026141

Hunt, Davis, A. P., & Traver, R. G. (2012). Meeting Hydrologic and Water Quality Goals through Targeted Bioretention Design. *Journal of Environmental Engineering*, *138*(6), 698–707. https://doi.org/10.1061/(asce)ee.1943-7870.0000504

Hunt, W. F., Jarrett, A. R., Smith, J. T., & Sharkey, L. J. (2006). Evaluating Bioretention Hydrology and Nutrient Removal at Three Field Sites in North Carolina. *Journal of Irrigation and Drainage Engineering*, *132*(6), 600–608. https://doi.org/10.1061/(ASCE)0733-9437(2006)132

Hurley, S., Shrestha, P., & Cording, A. (2017). Nutrient Leaching from Compost : Implications for Bioretention and Other Green Stormwater Infrastructure. *Journal of Sustainable Water Built Environmnet*, *3*, 1–8. https://doi.org/10.1061/JSWBAY.0000821

Infante, D. M., David Allan, J., Linke, S., & Norris, R. H. (2009). Relationship of fish and macroinvertebrate assemblages to environmental factors: Implications for community concordance. *Hydrobiologia*, *623*(1), 87–103. https://doi.org/10.1007/s10750-008-9650-3

Inthout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology*, *14*(1), 1–12. https://doi.org/10.1186/1471-2288-14-25/TABLES/6

Ippolito, J. A., Blecker, S. W., Freeman, C. L., McCulley, R. L., Blair, J. M., & Kelly, E. F. (2010). Phosphorus biogeochemistry across a precipitation gradient in grasslands of central North America. *Journal of Arid Environments*, *74*(8), 954–961. https://doi.org/10.1016/j.jaridenv.2010.01.003

Isaak, D. J., Peterson, E. E., Ver Hoef, J. M., Wenger, S. J., Falke, J. A., Torgersen, C. E., … Monestiez, P. (2014). Applications of spatial statistical network models to stream data. *Wiley Interdisciplinary Reviews: Water*, *1*(3), 277–294. https://doi.org/10.1002/wat2.1023

Jackson, D., Angeliki, A., Martin, V., Andrea, L., & Baker, R. (2017). Paule - Mandel estimators for network meta - analysis with random inconsistency effects. *Research Synthesis Methods*, *8*, 416–434. https://doi.org/10.1002/jrsm.1244

Jalali, P., & Rabotyagov, S. (2020). Quantifying cumulative effectiveness of green stormwater infrastructure in improving water quality. *Science of the Total Environment*, *731*, 138953. https://doi.org/10.1016/j.scitotenv.2020.138953

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). Unsupervised Learning. In *An Introduction to Statistical Learning* (2nd ed., pp. 497–552). New York, NY: Spinger. https://doi.org/https://doi.org/10.1007/978-1-0716-1418-1_12

Jayakody, P., Parajuli, P. B., & Cathcart, T. P. (2014). Impacts of climate variability on water quality with best management practices in sub-tropical climate of USA. *Hydrological Processes*, *28*(23), 5776–5790. https://doi.org/10.1002/hyp.10088

Jefferson, A. J., Bhaskar, A. S., Hopkins, K. G., Fanelli, R., Avellaneda, P. M., & McMillan, S. K. (2017). Stormwater management network effectiveness and implications for urban watershed function: A critical review. *Hydrological Processes*, *31*(23), 4056–4080. https://doi.org/10.1002/hyp.11347

Jennings, A. A. (2016). Residential Rain Garden Performance in the Climate Zones of the Contiguous United States. *Journal of Environmental Engineering*, *142*(12). https://doi.org/10.1061/(asce)ee.1943-7870.0001143

Jeon, D. J., Ki, S. J., Cha, Y. K., Park, Y., & Kim, J. H. (2018). New methodology of evaluation of best management practices performances for an agricultural watershed according to the climate change scenarios: A hybrid use of deterministic and decision support models. *Ecological Engineering*, *119*(May), 73–83. https://doi.org/10.1016/j.ecoleng.2018.05.006

Jollymore, A., Haines, M. J., Satterfield, T., & Johnson, M. S. (2017). Citizen science for water quality monitoring: Data implications of citizen perspectives. *Journal of Environmental Management*, *200*(2017), 456–467. https://doi.org/10.1016/j.jenvman.2017.05.083

Kajewska-Szkudlarek, J. (2020). Clustering approach to urban rainfall time series prediction with support vector regression model. *Urban Water Journal*, *17*(3), 235–246. https://doi.org/https://doi.org/10.1080/1573062X.2020.1760319

Kaller, M. D., & Hartman, K. J. (2004). Evidence of a threshold level of fine sediment accumulation for altering benthic macroinvertebrate communities. *Hydrobiologia*, *518*(1–3), 95–104. https://doi.org/10.1023/B:HYDR.0000025059.82197.35

Kaye, J. P., Groffman, P. M., Grimm, N. B., Baker, L. A., & Pouyat, R. V. (2006). A distinct urban biogeochemistry ? *TRENDS in Ecology and Evolution*, *21*(4), 192–

199. https://doi.org/10.1016/j.tree.2005.12.006

Keiser, D. A., Kling, C. L., & Shapiro, J. S. (2018). The low but uncertain measured benefits of US water quality policy. *PNAS*, *116*(12), 5262–5269. https://doi.org/10.1073/pnas.1802870115

Kenney, M. A., Wilcock, P. R., Hobbs, B. F., Flores, N. E., & Martínez, D. C. (2012). Is Urban Stream Restoration Worth It? *Journal of the American Water Resources Association*, *48*(3), 603–615. https://doi.org/10.1111/j.1752-1688.2011.00635.x

Kenney, M., Sutton-Grier, A., Smith, R., & Gresens, S. (2010). Benthic macroinvertebrates as indicators of water quality: The intersection of science and policy. *Terrestrial Arthropod Reviews*, *2*(2), 99–128. https://doi.org/10.1163/187498209x12525675906077

Kielstra, B. W., Chau, J., & Richardson, J. S. (2019). Measuring function and structure of urban headwater streams with citizen scientists. *Ecosphere*, *10*(4). https://doi.org/10.1002/ecs2.2720

Koerselman, W., & Meuleman, A. F. M. (1996). The Vegetation N : P Ratio : a New Tool to Detect the Nature of Nutrient Limitation. *Journal of Applied Ecology*, *33*(6), 1441–1450. https://doi.org/10.2307/2404783

Krabbenhoft, C. A., & Kashian, D. R. (2020). Citizen science data are a reliable complement to quantitative ecological assessments in urban rivers. *Ecological Indicators*, *116*(April), 106476. https://doi.org/10.1016/j.ecolind.2020.106476

Kranz, C. N., Heitman, J. L., Rivers, E. N., & Mclaughlin, R. A. (2022). Influence of compost application rate on nutrient and heavy metal mobility : Implications for stormwater management. *Journal of Envionmental Quality*, (December 2021), 1–13. https://doi.org/10.1002/jeq2.20403

Krimsky, L. S., Lusk, M. G., Abeels, H., & Seals, L. (2021). Sources and concentrations of nutrients in surface runoff from waterfront homes with different landscape practices. *Science of the Total Environment*, *750*, 142320. https://doi.org/10.1016/j.scitotenv.2020.142320

Lammers, R. W., Ph, D., Miller, L., Bledsoe, B. P., Ph, D., & Asce, M. (2022). Effects of Design and Climate on Bioretention Effectiveness for Watershed-Scale Hydrologic Benefits. *Journal of Sustainable Water in the Built Environment*, *8*(4), 1–16. https://doi.org/10.1061/JSWBAY.0000993

Lee, C. (2020). *A Game Changer in the Making? Lessons from States Advancing Environmental Justice Through Mapping and Cumulative Impact Strategies*. *Environmental Law Reporter* (Vol. 5). Washington D.C. Retrieved from https://www.eli.org/sites/default/files/docs/50.10203.pdf

LeFevre, G. H., Paus, K. H., Natarajan, P., Gulliver, J. S., Novak, P. J., & Hozalski, R. M. (2015). Review of Dissolved Pollutants in Urban Storm Water and Their Removal and Fate in Bioretention Cells. *Journal of Environmental Engineering*, *141*(1), 04011050. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000876.

Lenat, D. R. (1988). Water Quality Assessment of Streams Using a Qualitative Collection Method for Benthic Macroinvertebrates. *Freshwater Science*, *7*(3). https://doi.org/https://doi.org/10.2307/1467422

Li, & Davis, A. P. (2014). Urban Stormwater Runoff Nitrogen Composition and Fate in Bioretention Systems. *Environmental Science & Technology*, *48*, 3403–3410. https://doi.org/10.1021/es4055302

Li, J., & Davis, A. P. (2016). A unified look at phosphorus treatment using bioretention. *Water Research*, *90*, 141–155. https://doi.org/10.1016/j.watres.2015.12.015

Lintern, A., McPhillips, L., Winfrey, B., Duncan, J., & Grady, C. (2020). Best Management Practices for Diffuse Nutrient Pollution : Wicked Problems Across Urban and Agricultural Watersheds. *Environmental Science & Technology*, *54*(15), 9159–9174. https://doi.org/10.1021/acs.est.9b07511

Liu, J., & Davis, A. P. (2014). Phosphorus Speciation and Treatment Using Enhanced Phosphorus Removal Bioretention.

Liu, Y., Engel, B. A., Flanagan, D. C., Gitau, M. W., McMillan, S. K., & Chaubey, I. (2017). A review on effectiveness of best management practices in improving hydrology and water quality: Needs and opportunities. *Science of the Total Environment*, *601–602*, 580–593. https://doi.org/10.1016/j.scitotenv.2017.05.212

Lopez-Ponnada, E. V., Lynn, T. J., Ergas, S. J., & Mihelcic, J. R. (2020). Long-term field performance of a conventional and modified bioretention system for removing dissolved nitrogen species in stormwater runoff. *Water Research*, *170*, 115336. https://doi.org/10.1016/j.watres.2019.115336

Maantay, J., & Maroko, A. (2009). Mapping urban risk: Flood hazards, race, & environmental justice in New York. *Applied Geography*, *29*(1), 111–124. https://doi.org/10.1016/j.apgeog.2008.08.002

Madden, S. (2010). *Choosing Green Over Gray: Philadelphia's Innovative Stormwater Infrastructure Plan*. Massachusetts Institute of Technology.

Maguire, T. J., & Mundle, S. O. C. (2020). Citizen Science Data Show Temperature-Driven Declines in Riverine Sentinel Invertebrates. *Environmental Science and Technology Letters*, *7*(5), 303–307. https://doi.org/10.1021/acs.estlett.0c00206

Mah, A. (2017). Environmental justice in the age of big data: challenging toxic blind spots of voice, speed, and expertise. *Environmental Sociology*, *3*(2), 122–133. https://doi.org/10.1080/23251042.2016.1220849

Mai, J., Shen, H., Tolson, B. A., Gaborit, É., Arsenault, R., Craig, J. R., … Waddell, J. W. (2022). The Great Lakes Runoff Intercomparison Project Phase 4: The Great Lakes (GRIP-GL). *Hydrology and Earth System Sciences*, *26*(13), 3537–3572. https://doi.org/10.5194/hess-26-3537-2022

Mallin, M. A., Johnson, V. L., & Ensign, S. H. (2009). Comparative impacts of stormwater runoff on water quality of an urban, a suburban, and a rural stream. *Environmental Monitoring and Assessment*, *159*(1–4), 475–491.

https://doi.org/10.1007/s10661-008-0644-4

Mallin, M. A., Johnson, V. L., Ensign, S. H., & MacPherson, T. A. (2006). Factors contributing to hypoxia in rivers, lakes, and streams. *Limnology and Oceanography*, *51*(1 II), 690–701. https://doi.org/10.4319/lo.2006.51.1_part_2.0690

Manning, D. W. P., Rosemond, A. D., Benstead, J. P., Bumpers, P. M., & Kominoski, J. S. (2020). Transport of N and P in U.S. streams and rivers differs with land use and between dissolved and particulate forms. *Ecological Applications*, *30*(6). https://doi.org/10.1002/eap.2130

Maranger, R., Jones, S. E., & Cotner, J. B. (2018). Stoichiometry of carbon, nitrogen, and phosphorus through the freshwater pipe. *Limnology And Oceanography Letters*, *3*(3), 89–101. https://doi.org/10.1002/lol2.10080

Marshall, E., & Randhir, T. (2008). Effect of climate change on watershed system: A regional analysis. *Climatic Change*, *89*(3–4), 263–280. https://doi.org/10.1007/s10584-007-9389-2

Martínez-Mena, M., Carrillo-López, E., Boix-Fayos, C., Almagro, M., García Franco, N., Díaz-Pereira, E., … de Vente, J. (2020). Long-term effectiveness of sustainable land management practices to control runoff, soil erosion, and nutrient loss and the role of rainfall intensity in Mediterranean rainfed agroecosystems. *Catena*, *187*(November 2019), 104352. https://doi.org/10.1016/j.catena.2019.104352

Marvin, J. T., Passeport, E., & Drake, J. (2020). State-of-the-Art Review of Phosphorus Sorption Amendments in Bioretention Media: A Systematic Literature Review. *Journal of Sustainable Water in the Built Environment*, *6*(1), 03119001. https://doi.org/10.1061/jswbay.0000893

McCulley, R. L., Burke, I. C., & Lauenroth, W. K. (2009). Conservation of nitrogen increases with precipitation across a major grassland gradient in the Central Great Plains of North America. *Oecologia*, *159*(3), 571–581. https://doi.org/10.1007/s00442-008-1229-1

McGill, L. M., Steel, E. A., Brooks, J. R., Edwards, R. T., & Fullerton, A. H. (2020). Elevation and spatial structure explain most surface-water isotopic variation across five Pacific Coast basins. *Journal of Hydrology*, *583*(July 2019), 124610. https://doi.org/10.1016/j.jhydrol.2020.124610

McGrane, S. J. (2016). Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: a review. *Hydrological Sciences Journal*, *61*(13), 2295–2311. https://doi.org/10.1080/02626667.2015.1128084

McNett, J. K., Hunt, W. F., & Davis, A. P. (2011). Influent Pollutant Concentrations as Predictors of Effluent Pollutant Concentrations for Mid-Atlantic Bioretention. *Journal of Environmental Engineering*, *137*(September), 790–799. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000373

McPhillips, L. E., Matsler, M., Rosenzweig, B. R., & Kim, Y. (2021). What is the role of green stormwater infrastructure in managing extreme precipitation events?

Sustainable and Resilient Infrastructure, *6*(3–4), 133–142. https://doi.org/https://doi.org/10.1080/23789689.2020.1754625

McPhillips, L., Goodale, C., & Walter, M. T. (2018). Nutrient Leaching and Greenhouse Gas Emissions in Grassed Detention and Bioretention Stormwater Basins. *Journal of Sustainable Water Built Environment*, *4*(1), 1–10. https://doi.org/10.1061/JSWBAY.0000837.

McPhillips, L., & Walter, M. T. (2015). Hydrologic conditions drive denitrification and greenhouse gas emissions in stormwater detention basins. *Ecological Engineering*, *85*, 67–75. https://doi.org/10.1016/j.ecoleng.2015.10.018

Meals, D. W., Dressing, S. A., & Davenport, T. E. (2010). Lag Time in Water Quality Response to Best Management Practices: A Review. *Journal of Environmental Quality*, *39*(1), 85–96. https://doi.org/https://doi.org/10.2134/jeq2009.0108

Meenar, M., Fromuth, R., & Soro, M. (2018). Planning for watershed-wide flood-mitigation and stormwater management using an environmental justice framework. *Environmental Practice*, *20*(2–3), 55–67. https://doi.org/10.1080/14660466.2018.1507366

Meyer, J. L., Paul, M. J., & Taulbee, W. K. (2005). Stream ecosystem function in urbanizing landscapes. *Journal of the North American Benthological Society*, *24*(3), 602–612. https://doi.org/10.1899/04-021.1

Mikołajewski, K., Ruman, M., Kosek, K., Glixelli, M., Dzimińska, P., Ziętara, P., & Licznar, P. (2022). Development of cluster analysis methodology for identification of model rainfall hyetographs and its application at an urban precipitation field scale. *Science of the Total Environment*, *829*. https://doi.org/10.1016/j.scitotenv.2022.154588

Miller, J. F., Frederick, R. H., & Tracey, R. J. (1973). Precipitation Frequency Atlas for Western United States, Volume X - Oregon. Washington D.C.: NOAA.

Miranda, M. L., Edwards, S. E., Keating, M. H., & Paul, C. J. (2011). Making the environmental justice grade: The relative burden of air pollution exposure in the United States. *International Journal of Environmental Research and Public Health*, *8*(6), 1755–1771. https://doi.org/10.3390/ijerph8061755

Miskewitz, R., & Uchrin, C. (2013). In-Stream Dissolved Oxygen Impacts and Sediment Oxygen Demand Resulting from Combined Sewer Overflow Discharges. *Journal of Environmental Engineering*, *139*(10), 1307–1313. https://doi.org/10.1061/(asce)ee.1943-7870.0000739

Morse, N. R., McPhillips, L. E., Shapleigh, J. P., & Walter, M. T. (2017). The Role of Denitrification in Stormwater Detention Basin Treatment of Nitrogen. *Environmental Science & Technology*, *51*, 7928–7935. https://doi.org/10.1021/acs.est.7b01813

Muerdter, C. P., Wong, C. K., & LeFevre, G. H. (2018). Emerging investigator series: the role of vegetation in bioretention for stormwater treatment in the built environment:

pollutant removal, hydrologic function, and ancillary benefits. *Environmental Science: Water Research & Technology*, *5*, 592–612. https://doi.org/DOI https://doi.org/10.1039/C7EW00511C

Mullins, A. R., Bain, D. J., Pfeil-McCullough, E., Hopkins, K. G., Lavin, S., & Copeland, E. (2020). Seasonal drivers of chemical and hydrological patterns in roadside infiltration-based green infrastructure. *Science of the Total Environment*, *714*, 136503. https://doi.org/10.1016/j.scitotenv.2020.136503

Murtagh, F. (2014). Ward 's Hierarchical Agglomerative Clustering Method : Which Algorithms Implement Ward ' s Criterion ? *Journal of Classification*, *295*(October), 274–295. https://doi.org/10.1007/s00357-

Murtagh, F. (2017). Algorithms for hierarchical clustering : an overview , II. *WIREs Data Mining Knowl Discov*, *7*(December), 1–16. https://doi.org/10.1002/widm.1219

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering : an overview. *WIREs Data Mining Knowl Discov*, *2*(February), 86–97. https://doi.org/10.1002/widm.53

Neill, A. J., Tetzlaff, D., Strachan, N. J. C., Hough, R. L., Avery, L. M., Watson, H., & Soulsby, C. (2018). Using spatial-stream-network models and long-term data to understand and predict dynamics of faecal contamination in a mixed land-use catchment. *Science of the Total Environment*, *612*, 840–852. https://doi.org/10.1016/j.scitotenv.2017.08.151

Njue, N., Stenfert Kroese, J., Gräf, J., Jacobs, S. R., Weeser, B., Breuer, L., & Rufino, M. C. (2019). Citizen science in hydrological monitoring and ecosystem services management: State of the art and future prospects. *Science of the Total Environment*, *693*. https://doi.org/10.1016/j.scitotenv.2019.07.337

Norton, R. A., Harrison, J. A., Keller, C. K., & Moffett, K. B. (2017). Effects of storm size and frequency on nitrogen retention , denitrification , and N2O production in bioretention swale mesocosms. *Biogeochemistry*, (134), 353–370. https://doi.org/10.1007/s10533-017-0365-2

Oliver, S. K., Collins, S. M., Soranno, P. A., Wagner, T., Stanley, E. H., Jones, J. R., … Lottig, N. R. (2017). Unexpected stasis in a changing world: Lake nutrient and chlorophyll trends since 1990. *Global Change Biology*, *23*(12), 5455–5467. https://doi.org/10.1111/gcb.13810

Olmstead, S. M. (2009). The Economics of water quality. *Review of Environmental Economics and Policy*, *4*(1), 44–62. https://doi.org/10.1093/reep/rep016

Our Watershed. (n.d.). Retrieved from https://therouge.org/about-us/our-watershed/

Ouyang, Y., Parajuli, P. B., Feng, G., Leininger, T. D., & Wan, Y. (2018). Application of Climate Assessment Tool (CAT) to estimate climate variability impacts on nutrient loading from local watersheds. *Journal of Hydrology*, *563*(April), 363–371. https://doi.org/10.1016/j.jhydrol.2018.06.017

Palla, A., & Gnecco, I. (2015). Hydrologic modeling of Low Impact Development

systems at the urban catchment scale. *Journal of Hydrology*, *528*, 361–368. https://doi.org/10.1016/j.jhydrol.2015.06.050

Park, D., Kang, H., Jung, S. H., & Roesner, L. A. (2015). Reliability analysis for evaluation of factors affecting pollutant load reduction in urban stormwater BMP systems. *Environmental Modelling and Software*, *74*, 130–139. https://doi.org/10.1016/j.envsoft.2015.08.010

Parr, T. B., Smucker, N. J., Bentsen, C. N., & Neale, M. W. (2016). Potential roles of past, present, and future urbanization characteristics in producing varied stream responses. *Freshwater Science*, *35*(1), 436–443. https://doi.org/10.1086/685030

Patang, F., Soegianto, A., & Hariyanto, S. (2018). Benthic Macroinvertebrates Diversity as Bioindicator of Water Quality of Some Rivers in East Kalimantan, Indonesia. *International Journal of Ecology*, *2018*. https://doi.org/10.1155/2018/5129421

Paz-Vinas, I., Loot, G., Hermoso, V., Veyssière, C., Poulet, N., Grenouillet, G., & Blanchet, S. (2018). Systematic conservation planning for intraspecific genetic diversity. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1877). https://doi.org/10.1098/rspb.2017.2746

Pennino, M. J., McDonald, R. I., & Jaffe, P. R. (2016). Watershed-scale impacts of stormwater green infrastructure on hydrology , nutrient fluxes , and combined sewer over flows in the mid-Atlantic region. *Science of the Total Environment*, *565*, 1044–1053. https://doi.org/10.1016/j.scitotenv.2016.05.101

Penuelas, J., Janssens, I. A., Ciais, P., Obersteiner, M., & Sardans, J. (2020). Anthropogenic global shifts in biospheric N and P concentrations and ratios and their impacts on biodiversity, ecosystem productivity, food security, and human health. *Global Change Biology*, *26*(4), 1962–1985. https://doi.org/10.1111/gcb.14981

Penuelas, J., & Sardans, J. (2022). The global nitrogen-phosphorus imbalance. *Science*, *375*(6578). https://doi.org/10.1126/science.abl4827

Peterson, E. E., & Ver Hoef, J. M. (2014). Stars: An ArcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. *Journal of Statistical Software*, *56*(2), 1–17. https://doi.org/10.18637/jss.v056.i02

Peterson, E. E., Ver Hoef, J. M., Isaak, D. J., Falke, J. A., Fortin, M. J., Jordan, C. E., … Wenger, S. J. (2013). Modelling dendritic ecological networks in space: An integrated network perspective. *Ecology Letters*, *16*(5), 707–719. https://doi.org/10.1111/ele.12084

Pond, G. J., Krock, K. J. G., Cruz, J. V., & Ettema, L. F. (2017). Effort-based predictors of headwater stream conditions: comparing the proximity of land use pressures and instream stressors on macroinvertebrate assemblages. *Aquatic Sciences*, *79*(3), 765–781. https://doi.org/10.1007/s00027-017-0534-3

Pyke, C., Warren, M. P., Johnson, T., LaGro, J., Scharfenberg, J., Groth, P., … Main, E. (2011). Assessment of low impact development for managing stormwater with

changing precipitation due to climate change. *Landscape and Urban Planning*, *103*(2), 166–173. https://doi.org/10.1016/j.landurbplan.2011.07.006

Qin, H., Li, Z., & Fu, G. (2013). The effects of low impact development on urban fl ooding under different rainfall characteristics. *Journal of Environmental Management*, *129*, 577–585. https://doi.org/10.1016/j.jenvman.2013.08.026

Quijano, J. C., Zhu, Z., Morales, V., Landry, B. J., & Garcia, M. H. (2017). Three-dimensional model to capture the fate and transport of combined sewer overflow discharges: A case study in the Chicago Area Waterway System. *Science of the Total Environment*, *576*, 362–373. https://doi.org/10.1016/j.scitotenv.2016.08.191

Rigolon, A., Browning, M., & Jennings, V. (2018). Inequities in the quality of urban park systems: An environmental justice investigation of cities in the United States. *Landscape and Urban Planning*, *178*(January), 156–169. https://doi.org/10.1016/j.landurbplan.2018.05.026

River Restoration. (n.d.).

Roseen, Robert, M., Ballestero, Thomas, P., Houle, James, J., Avellaneda, P., Briggs, J., Fowler, G., & Wildey, R. (2009). Seasonal Performance Variations for Storm-Water Management Systems in Cold Climate Conditions. *Journal of Environmental Engineering*, *135*(March), 128–137. https://doi.org/10.1061/(ASCE)0733-9372(2009)135

Salerno, F., Viviano, G., & Tartari, G. (2018). Urbanization and climate change impacts on surface water quality: Enhancing the resilience by reducing impervious surfaces. *Water Research*, *144*, 491–502. https://doi.org/10.1016/j.watres.2018.07.058

Sanchez, G. M., Nejadhashemi, A. P., Zhang, Z., Marquart-Pyatt, S., Habron, G., & Shortridge, A. (2015). Linking watershed-scale stream health and socioeconomic indicators with spatial clustering and structural equation modeling. *Environmental Modelling and Software*, *70*, 113–127. https://doi.org/10.1016/j.envsoft.2015.04.012

Sanchez, G. M., Nejadhashemi, A. P., Zhang, Z., Woznicki, S. A., Habron, G., Marquart-Pyatt, S., & Shortridge, A. (2014). Development of a socio-ecological environmental justice model for watershed-based management. *Journal of Hydrology*, *518*(PA), 162–177. https://doi.org/10.1016/j.jhydrol.2013.08.014

Selbig, W. R. (2016). Evaluation of leaf removal as a means to reduce nutrient concentrations and loads in urban stormwater. *Science of the Total Environment*, *571*, 124–133. https://doi.org/10.1016/j.scitotenv.2016.07.003

Shahapure, K. R., & Nicholas, C. (2020). Cluster Quality Analysis Using Silhouette Score. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 747–748). Sydney, Australia. https://doi.org/10.1109/DSAA49011.2020.00096

Shrestha, P., Hurley, S. E., & Wemple, B. C. (2018). Effects of different soil media, vegetation, and hydrologic treatments on nutrient and sediment removal in roadside bioretention systems. *Ecological Engineering*, *112*(December 2017), 116–131.

https://doi.org/10.1016/j.ecoleng.2017.12.004

Siebers, N., Sumann, M., Kaiser, K., & Amelung, W. (2017). Climatic Effects on Phosphorus Fractions of Native and Cultivated North American Grassland Soils. *Soil Science Society of America Journal*, *81*(2), 299–309. https://doi.org/https://doi.org/10.2136/sssaj2016.06.0181

Simpson, I. M., Winston, R. J., & Brooker, M. R. (2022). Effects of land use, climate, and imperviousness on urban stormwater quality: A meta-analysis. *Science of The Total Environment*, *809*, 152206. https://doi.org/10.1016/j.scitotenv.2021.152206

Skorobogatov, A., He, J., Chu, A., Valeo, C., & Duin, B. Van. (2020). Science of the Total Environment The impact of media , plants and their interactions on bioretention performance : A review. *Science of the Total Environment*, *715*, 136918. https://doi.org/10.1016/j.scitotenv.2020.136918

Smith, J. S., Winston, R. J., Tirpak, R. A., Wituszynski, D. M., Boening, K. M., & Martin, J. F. (2020). The seasonality of nutrients and sediment in residential stormwater runoff: Implications for nutrient-sensitive waters. *Journal of Environmental Management*, *276*(August), 111248. https://doi.org/10.1016/j.jenvman.2020.111248

Sohn, W., Kim, J., Li, M., & Brown, R. (2019). The influence of climate on the effectiveness of low impact development : A systematic review. *Journal of Environmental Management*, *236*(February), 365–379. https://doi.org/10.1016/j.jenvman.2018.11.041

Sonne, B. (2014). Managing Stormwater by Sustainable Measures: Preventing Neighborhood Flooding and Green Infrastructure Implementation in New Orleans. *Tulane Environmental Law Journal*, *27*(2), 323–350.

Steffen, M. M., Davis, T. W., McKay, R. M. L., Bullerjahn, G. S., Krausfeldt, L. E., Stough, J. M. A., … Wilhelm, S. W. (2017). Ecophysiological Examination of the Lake Erie Microcystis Bloom in 2014: Linkages between Biology and the Water Supply Shutdown of Toledo, OH. *Environmental Science and Technology*, *51*(12), 6745–6755. https://doi.org/10.1021/acs.est.7b00856

Steinley, D. (2006). K -means clustering : A half-century synthesis. *The British Psychological Society*, *59*, 1–34. https://doi.org/10.1348/000711005X48266

Stets, E. G., Sprague, L. A., Oelsner, G. P., Johnson, H. M., Murphy, J. C., Ryberg, K., … Riskin, M. L. (2020). Landscape Drivers of Dynamic Change in Water Quality of U.S. Rivers. *Environmental Sceince and Technology*, *54*, 4436–4343. https://doi.org/10.1021/acs.est.9b05344

Sun, Y., Pomeroy, C., Li, Q., & Xu, C. (2019). Impacts of rainfall and catchment characteristics on bioretention cell performance. *Water Science and Engineering*, *12*(2), 98–107. https://doi.org/10.1016/j.wse.2019.06.002

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best

Customer Profile Cluster. In *IOP Conf. Ser.: Mater. Sci. Eng.* (pp. 1–6). https://doi.org/10.1088/1757-899X/336/1/012017

Tackett, T. (2010). Seattle's Implementation of Green Stormwater Infrastructure to the Maximum Extent Feasible. In *Low Impact Development 2010: Redefining Water in the City*. San Francisco. https://doi.org/https://doi.org/10.1061/41099(367)101

Taguchi, V. J., Weiss, P. T., Gulliver, J. S., Klein, M. R., Hozalski, R. M., Baker, L. A., … Nieber, J. L. (2020). It is not easy being green: Recognizing unintended consequences of green stormwater infrastructure. *Water*, *12*(2). https://doi.org/10.3390/w12020522

Taylor, J., Graham, M., Louw, A., Lepheana, A., Madikizela, B., Dickens, C., … Warner, S. (2021). Social change innovations, citizen science, miniSASS and the SDGs. *Water Policy*, *00*(0), 1–10. https://doi.org/10.2166/wp.2021.264

Teshager, A. D., Gassman, P. W., Secchi, S., & Schoof, J. T. (2017). Simulation of targeted pollutant-mitigation-strategies to reduce nitrate and sediment hotspots in agricultural watershed. *Science of the Total Environment*, *607–608*, 1188–1200. https://doi.org/10.1016/j.scitotenv.2017.07.048

Thornbrugh, D. J., Leibowitz, S. G., Hill, R. A., Weber, M. H., Johnson, Z. C., Olsen, A. R., … Peck, D. V. (2018). Mapping watershed integrity for the conterminous United States. *Ecological Indicators*, *85*(December 2017), 1133–1148. https://doi.org/10.1016/j.ecolind.2017.10.070

Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., … Tait, S. (2019). Recent insights on uncertainties present in integrated catchment water quality modelling. *Water Research*, *150*, 368–379. https://doi.org/10.1016/j.watres.2018.11.079

United States Environmental Protecion Agency. (2021). EJSCREEN. Retrieved February 9, 2022, from https://www.epa.gov/ejscreen

Valenca, R., Le, H., Zu, Y., Dittrich, T. M., Tsang, D. C. W., Datta, R., … Mohanty, S. K. (2021). Nitrate removal uncertainty in stormwater control measures: Is the design or climate a culprit? *Water Research*, *190*, 116781. https://doi.org/10.1016/j.watres.2020.116781

Van Meter, K. J., Chowdhury, S., Byrnes, D. K., & Basu, N. B. (2019). Biogeochemical asynchrony : Ecosystem drivers of seasonal concentration regimes across the Great Lakes Basin. *Limnology and Oceaonography*, *9999*, 1–15. https://doi.org/10.1002/lno.11353

VanDerslice, J. (2011). Drinking water infrastructure and environmental disparities: Evidence and methodological considerations. *American Journal of Public Health*, *101*(SUPPL. 1), 109–114. https://doi.org/10.2105/AJPH.2011.300189

Ver Hoef, J. M., Peterson, E. E., Clifford, D., & Shah, R. (2014). SSN: An R Package for Spatial Statistical Modeon Stream Networks. *Journal of Statistical Software*, *56*(3). https://doi.org/10.18637/jss.v056.i03

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., … Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in. *Research Synthesis Methods*, *7*(September 2015), 55–79. https://doi.org/10.1002/jrsm.1164

Vijayaraghavan, K., Kumar, B., Gerrit, M., Hong, S., Tsen-tieng, D. L., Davis, A. P., … Balasubramanian, R. (2021). Bioretention systems for stormwater management : Recent advances and future prospects. *Journal of Environmental Management*, *292*(April), 112766. https://doi.org/10.1016/j.jenvman.2021.112766

Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science*, *277*(5325), 494–499. https://doi.org/10.1126/science.277.5325.494

Vogel, J. R., Moore, T. L., Coffman, R. R., Rodie, S. N., Hutchinson, S. L., McDonough, K. R., … McMaine, J. T. (2015). Critical Review of Technical Questions Facing Low Impact Development and Green Infrastructure: A Perspective from the Great Plains. *Water Environment Research*, *87*(9), 849–862. https://doi.org/10.2175/106143015x14362865226392

Von Bertrab, M. G., Krein, A., Stendera, S., Thielen, F., & Hering, D. (2013). Is fine sediment deposition a main driver for the composition of benthic macroinvertebrate assemblages? *Ecological Indicators*, *24*, 589–598. https://doi.org/10.1016/j.ecolind.2012.08.001

Voter, C. B., & Loheide, S. P. (2021). Climatic controls on the hydrologic effects of urban low impact development practices. *Environmental Research Letters*, *16*(6). https://doi.org/10.1088/1748-9326/abfc06

Wahl, M. H., McKellar, H. N., & Williams, T. M. (1997). Patterns of nutrient loading in forested and urbanized coastal streams. *Journal of Experimental Marine Biology and Ecology*, *213*(1), 111–131. https://doi.org/10.1016/S0022-0981(97)00012-9

Walsh, C. J., Booth, D. B., Burns, M. J., Fletcher, T. D., Hale, R. L., Hoang, L. N., … Wallace, A. (2016). Principles for urban stormwater management to protect stream ecosystems. *Freshwater Science*, *35*(1), 398–411. https://doi.org/10.1086/685284

Walsh, C. J., Imberger, M. J., Burns, M. J., Fletcher, T. D., & Bos, D. G. (2022). Dispersed Urban-Stormwater Control Improved Stream Water Quality in a Catchment-Scale Experiment. *Water Resources Research*, *58*. https://doi.org/10.1029/2022WR032041

Walsh, C. J., Roy, A. H., Feminella, J. W., Cottingham, P. D., Groffman, P. M., & Morgan, R. P. (2005). The urban stream syndrome: Current knowledge and the search for a cure. *Journal of the North American Benthological Society*, *24*(3), 706–723. https://doi.org/10.1899/04-028.1

Walsh, C. J., Sharpe, A. K., Breen, P. F., & Sonneman, J. A. (2001). Effects of urbanization on streams of the Melbourne region, Victoria, Australia. I. Benthic macroinvertebrate communities. *Freshwater Biology*, *46*(4), 535–551. https://doi.org/https://doi.org/10.1046/j.1365-2427.2001.00690.x

Wang, H., Gan, H., Zhang, Z., Yu, Z., & Zhu, D. Z. (2021). Purification Efficiency of Bioretention with Modified Media under Varied Rain Intensity and Drying Conditions. *Journal of Environmental Engineering*, *147*(4), 04021009. https://doi.org/10.1061/(asce)ee.1943-7870.0001868

Wang, M., Zhang, D., Cheng, Y., & Keat, S. (2019). Assessing performance of porous pavements and bioretention cells for stormwater management in response to probable climatic changes. *Journal of Environmental Management*, *243*(April), 157–167. https://doi.org/10.1016/j.jenvman.2019.05.012

Welter, J. R., Fisher, S. G., & Grimm, N. B. (2005). Nitrogen transport and retention in an arid land watershed: Influence of storm characteristics on terrestrial-aquatic linkages. *Biogeochemistry*, *76*(3), 421–440. https://doi.org/10.1007/s10533-005-6997-7

Winston, R. J., & Hunt, W. F. (2017). Characterizing Runoff from Roads: Particle Size Distributions, Nutrients, and Gross Solids. *Journal of Environmental Engineering*, *143*(1), 04016074. https://doi.org/10.1061/(asce)ee.1943-7870.0001148

Withers, P. J. A., & Jarvie, H. P. (2008). Delivery and cycling of phosphorus in rivers: A review. *Science of the Total Environment*, *400*(1–3), 379–395. https://doi.org/10.1016/j.scitotenv.2008.08.002

Wolch, J. R., Byrne, J., & Newell, J. P. (2014). Landscape and Urban Planning Urban green space , public health , and environmental justice : The challenge of making cities ' just green enough .' *Landscape and Urban Planning*, *125*, 234–244. https://doi.org/10.1016/j.landurbplan.2014.01.017

Woznicki, S. A., Hondula, K. L., & Jarnagin, S. T. (2018). Effectiveness of landscape - based green infrastructure for stormwater management in suburban catchments. *Hydrological Processes*, (January), 2346–2361. https://doi.org/10.1002/hyp.13144

Wu, L., Peng, M., Qiao, S., & Ma, X. (2018). Assessing impacts of rainfall intensity and slope on dissolved and adsorbed nitrogen loss under bare loessial soil by simulated rainfalls. *CATENA*, *170*, 51–63. https://doi.org/https://doi.org/10.1016/j.catena.2018.06.007

Xie, H., Chen, L., & Shen, Z. (2015). Assessment of Agricultural Best Management Practices Using Models: Current Issues and Future Perspectives. *Water (Switzerland)*, *7*(3), 1088–1108. https://doi.org/10.3390/w7031088

Xu, R., & Wunsch, D. I. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Nueral Networks*, *16*(3), 645–678. https://doi.org/10.1109/TNN.2005.845141

Yahdjian, L., & Sala, O. E. (2010). Size of precipitation pulses controls nitrogen transformation and losses in an arid Patagonian ecosystem. *Ecosystems*, *13*(4), 575–585. https://doi.org/10.1007/s10021-010-9341-6

Yan, Q., Davis, A. P., & James, B. R. (2016). Enhanced Organic Phosphorus Sorption from Urban Stormwater Using Modified Bioretention Media : Batch Studies. *Journal of Environmental Engineering*, *142*(4), 1–11.

https://doi.org/10.1061/(ASCE)EE.1943-7870.0001073

Yao, Y., Liu, J., Wang, Z., Wei, X., Zhu, H., Fu, W., & Shao, M. (2020). Responses of soil aggregate stability, erodibility and nutrient enrichment to simulated extreme heavy rainfall. *Science of the Total Environment*, *709*, 136150. https://doi.org/10.1016/j.scitotenv.2019.136150

Yin, J., Calabrese, S., Daly, E., & Porporato, A. (2019). The Energy Side of Budyko: Surface-Energy Partitioning From Hydrological Observations. *Geophysical Research Letters*, *46*(13), 7456–7463. https://doi.org/10.1029/2019GL083373

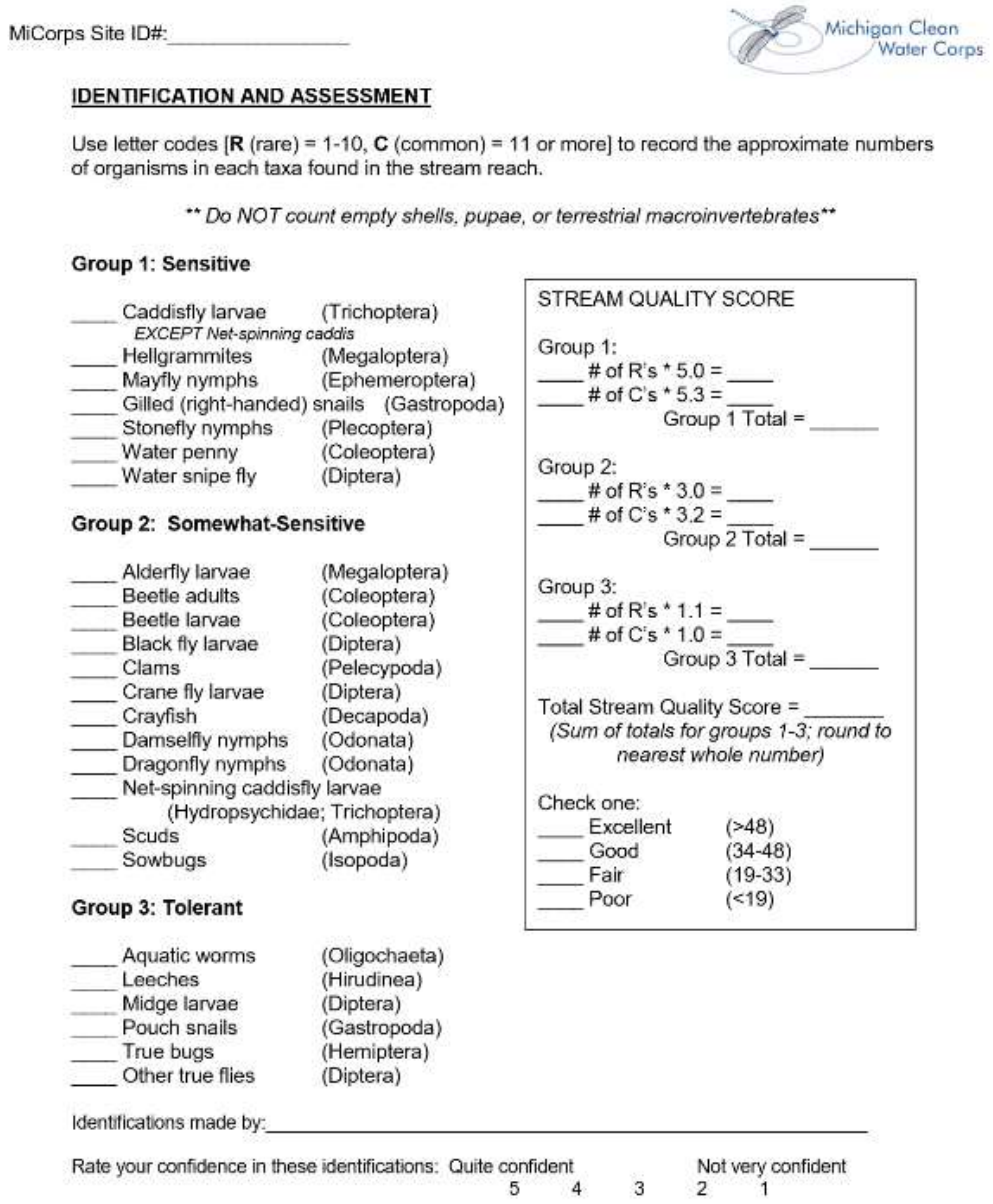
Zhu, J. (2013). Impact of Climate Change on Extreme Rainfall across the United States. *Journal of Hydrologic Engineering*, *18*(October), 1301–1309. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000725

Zhu, Q., Castellano, M. J., & Yang, G. (2018). Coupling soil water processes and the nitrogen cycle across spatial scales: Potentials, bottlenecks and solutions. *Earth-Science Reviews*, *187*, 248–258. https://doi.org/https://doi.org/10.1016/j.earscirev.2018.10.005

APPENDIX

Appendix A: Objective 1 Supplemental Information

Appendix A1: Stream Quality Index Scorecard



MiCorps Site ID#:_____

Michigan Clean Water Corps

IDENTIFICATION AND ASSESSMENT

Use letter codes [**R** (rare) = 1-10, **C** (common) = 11 or more] to record the approximate numbers of organisms in each taxa found in the stream reach.

*** Do NOT count empty shells, pupae, or terrestrial macroinvertebrates***

Group 1: Sensitive

____ Caddisfly larvae (Trichoptera)
EXCEPT Net-spinning caddis
____ Hellgrammites (Megaloptera)
____ Mayfly nymphs (Ephemeroptera)
____ Gilled (right-handed) snails (Gastropoda)
____ Stonefly nymphs (Plecoptera)
____ Water penny (Coleoptera)
____ Water snipe fly (Diptera)

Group 2: Somewhat-Sensitive

____ Alderfly larvae (Megaloptera)
____ Beetle adults (Coleoptera)
____ Beetle larvae (Coleoptera)
____ Black fly larvae (Diptera)
____ Clams (Pelecypoda)
____ Crane fly larvae (Diptera)
____ Crayfish (Decapoda)
____ Damselfly nymphs (Odonata)
____ Dragonfly nymphs (Odonata)
____ Net-spinning caddisfly larvae (Hydropsychidae; Trichoptera)
____ Scuds (Amphipoda)
____ Sowbugs (Isopoda)

Group 3: Tolerant

____ Aquatic worms (Oligochaeta)
____ Leeches (Hirudinea)
____ Midge larvae (Diptera)
____ Pouch snails (Gastropoda)
____ True bugs (Hemiptera)
____ Other true flies (Diptera)

STREAM QUALITY SCORE

Group 1:
____ # of R's * 5.0 = ____
____ # of C's * 5.3 = ____
Group 1 Total = _____

Group 2:
____ # of R's * 3.0 = ____
____ # of C's * 3.2 = ____
Group 2 Total = _____

Group 3:
____ # of R's * 1.1 = ____
____ # of C's * 1.0 = ____
Group 3 Total = _____

Total Stream Quality Score = _____
(Sum of totals for groups 1-3; round to nearest whole number)

Check one:
____ Excellent (>48)
____ Good (34-48)
____ Fair (19-33)
____ Poor (<19)

Identifications made by:_____

Rate your confidence in these identifications: Quite confident 5 4 3 2 1 Not very confident

Figure A29: SQI calculation sheet developed by the Michigan Clean Water Corps from their Macroinvertebrate Datasheet (pre 2020) ("Stream Macroinvertebrate Datasheet," n.d.).

Appendix B: Objective 2 Supplemental Information

Appendix B1: Forest plots of meta-analysis results.

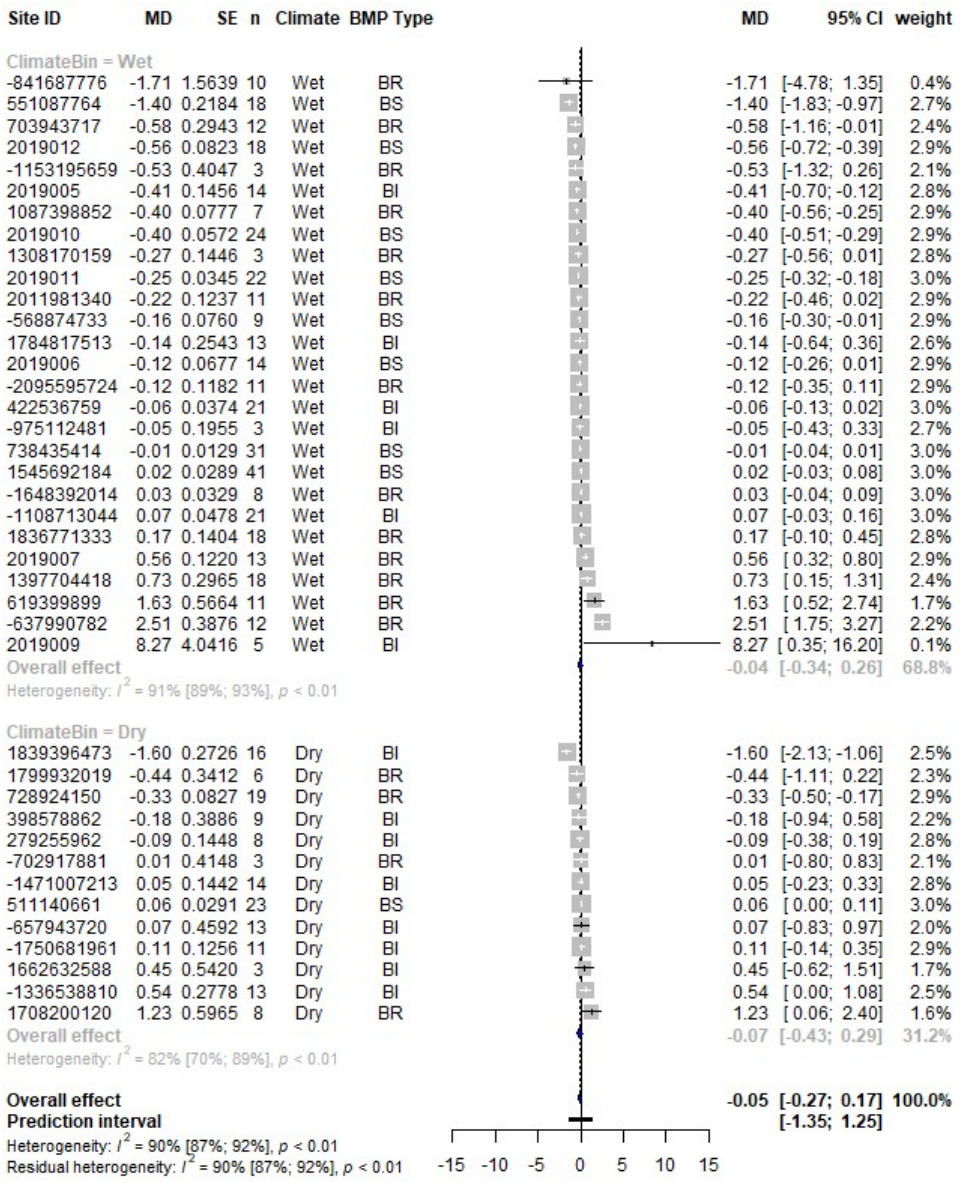


Figure B1.1: Forest plot of DIN subgroup analysis.

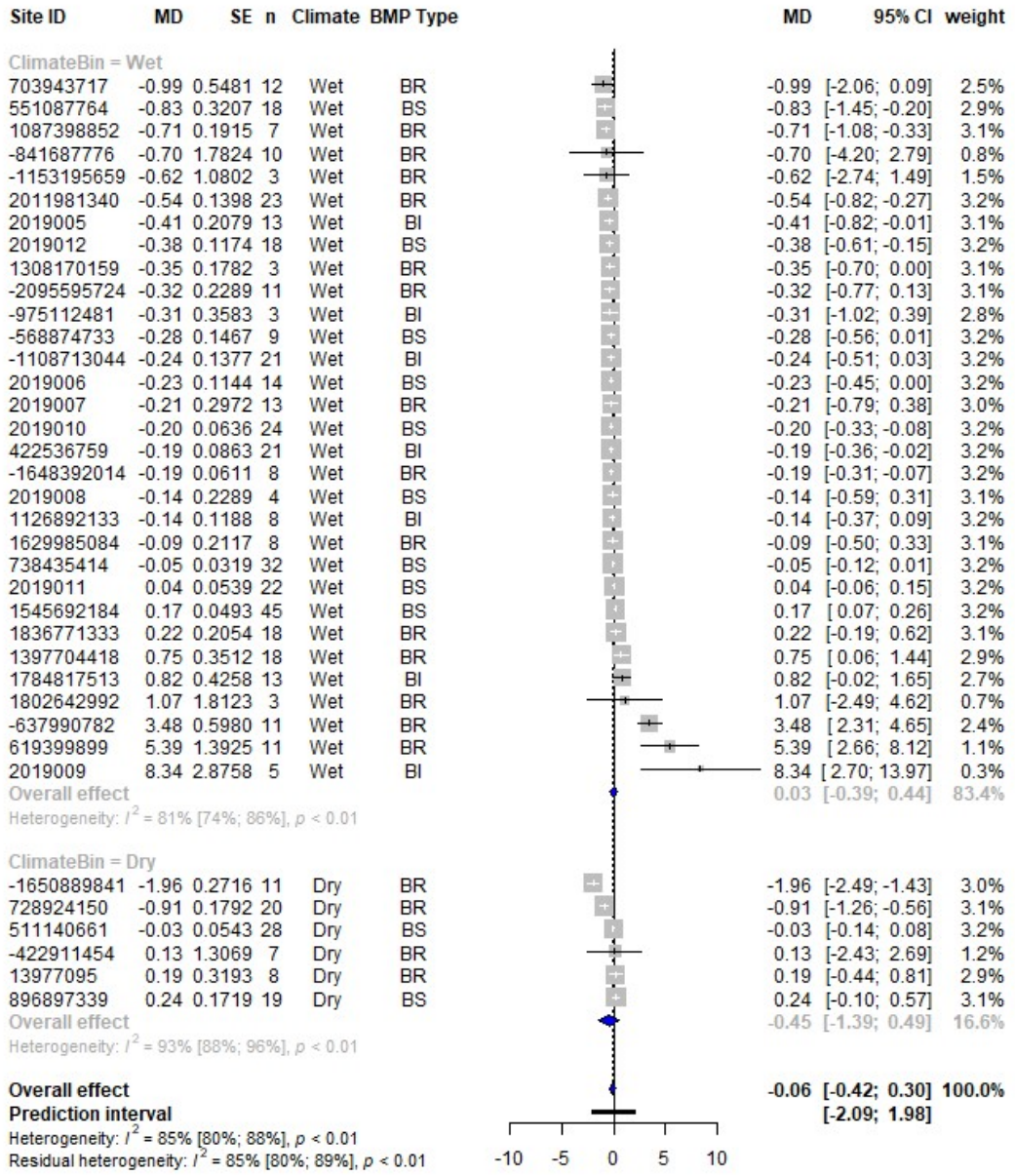


Figure B1.2: Forest plot of TN subgroup analysis

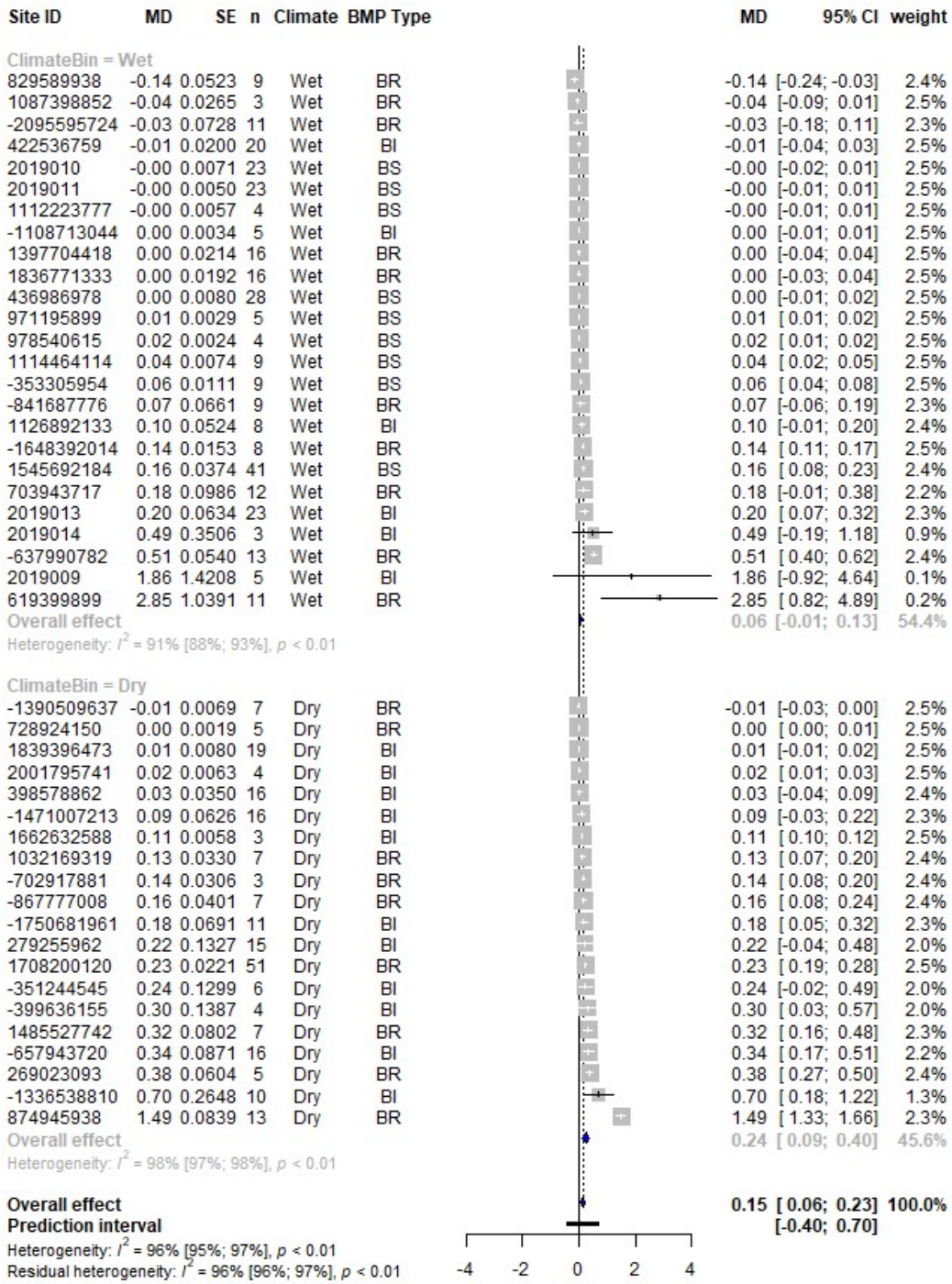


Figure B1.3: Forest plot of DIP subgroup analysis

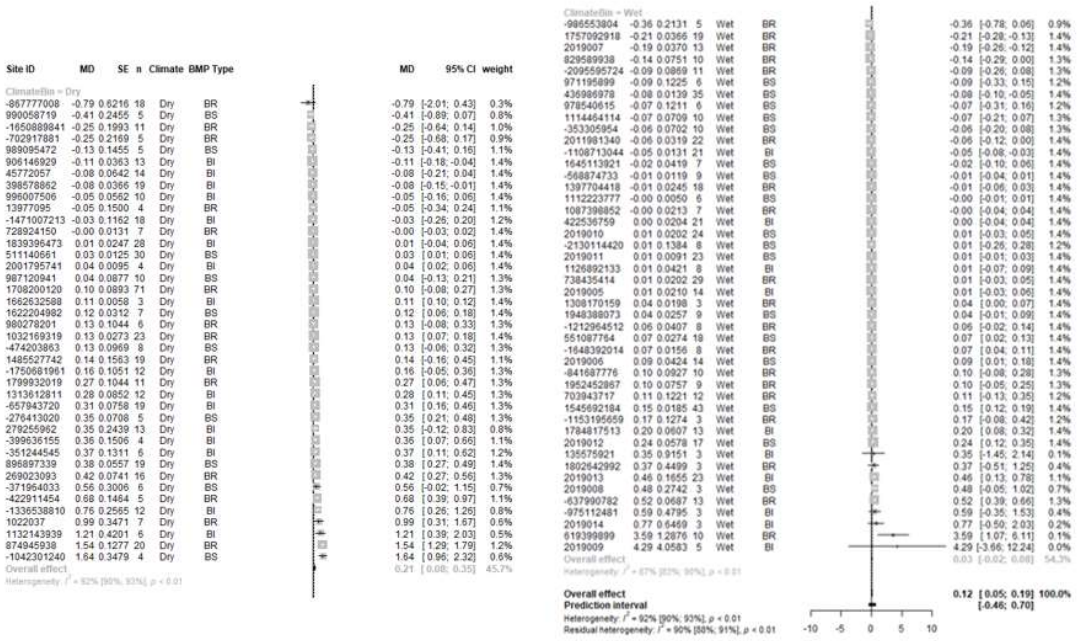


Figure B1.4: Forest plot of TP subgroup analysis.

Appendix C: Objective 3 Supplemental Information

Appendix C1: Performance of BMP concentration, volume, and load removal against isolate storm traits.

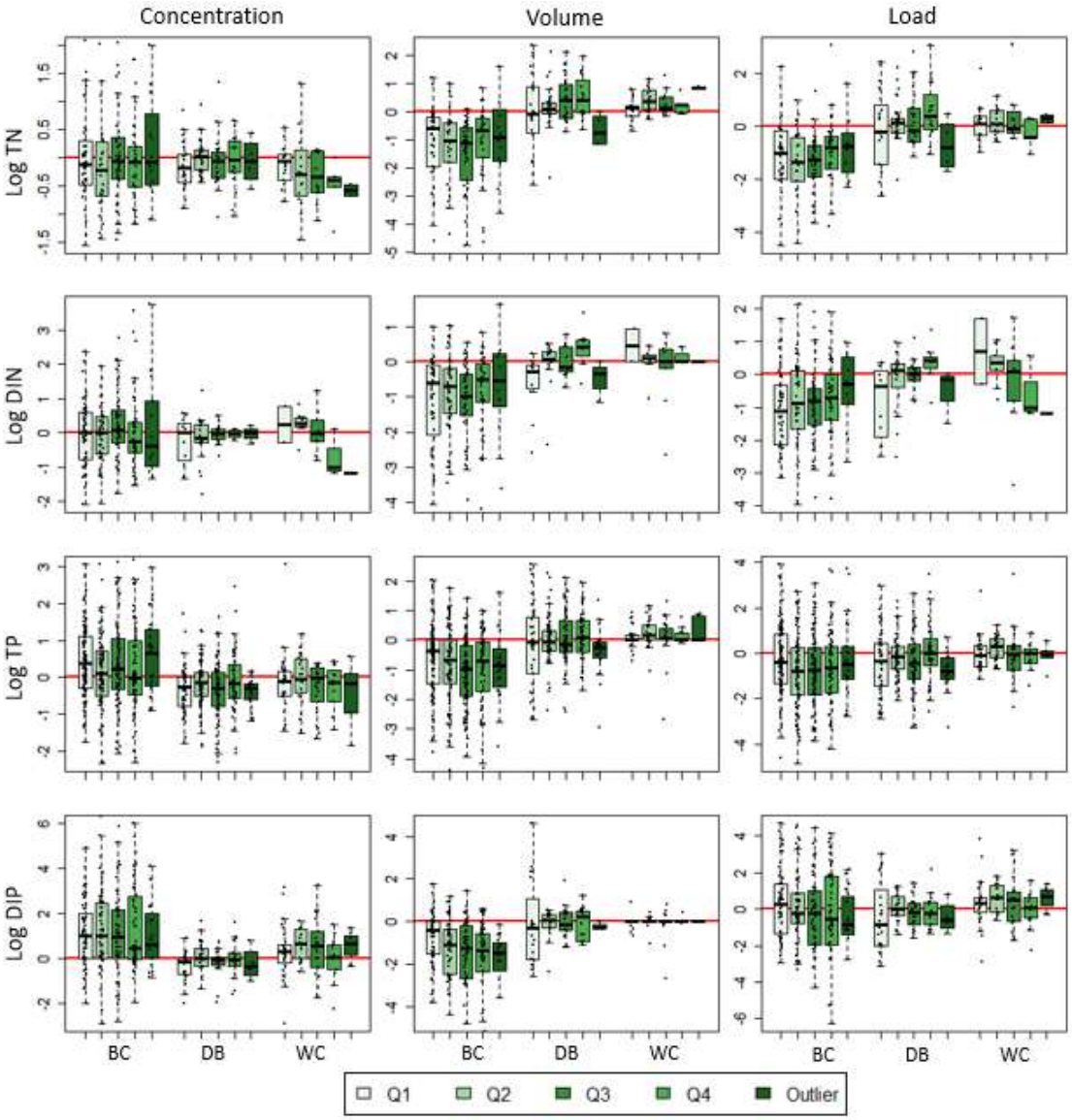


Figure C1.1. Boxplots comparing the log concentration, volume, and load removal fractions across groups of increasing return period for each analyte: TN, DIN, TP, and DIP, at each of the three condensed BMP types: bio-composite (BC), detention basin (DB) and wetland channel (WC). Return period is displayed with boxplots of increasing return interval through 5 groups: 4 quantiles and a high-range outlier group, where darker shaded boxplots indicate higher return intervals. Boxplots show the interquartile range of each water quality parameter shaded in a box with whiskers extending to the minimum and maximum water quality values without outliers. Black circles represent all observed points, which are plotted on top of boxplots to convey the counts and distribution of data summarized within each boxplot. The red horizontal line is the logged removal threshold, above which effluent values are higher than influent values, indicating leaching; and below which effluent values are lower than influent values, indicating retention.

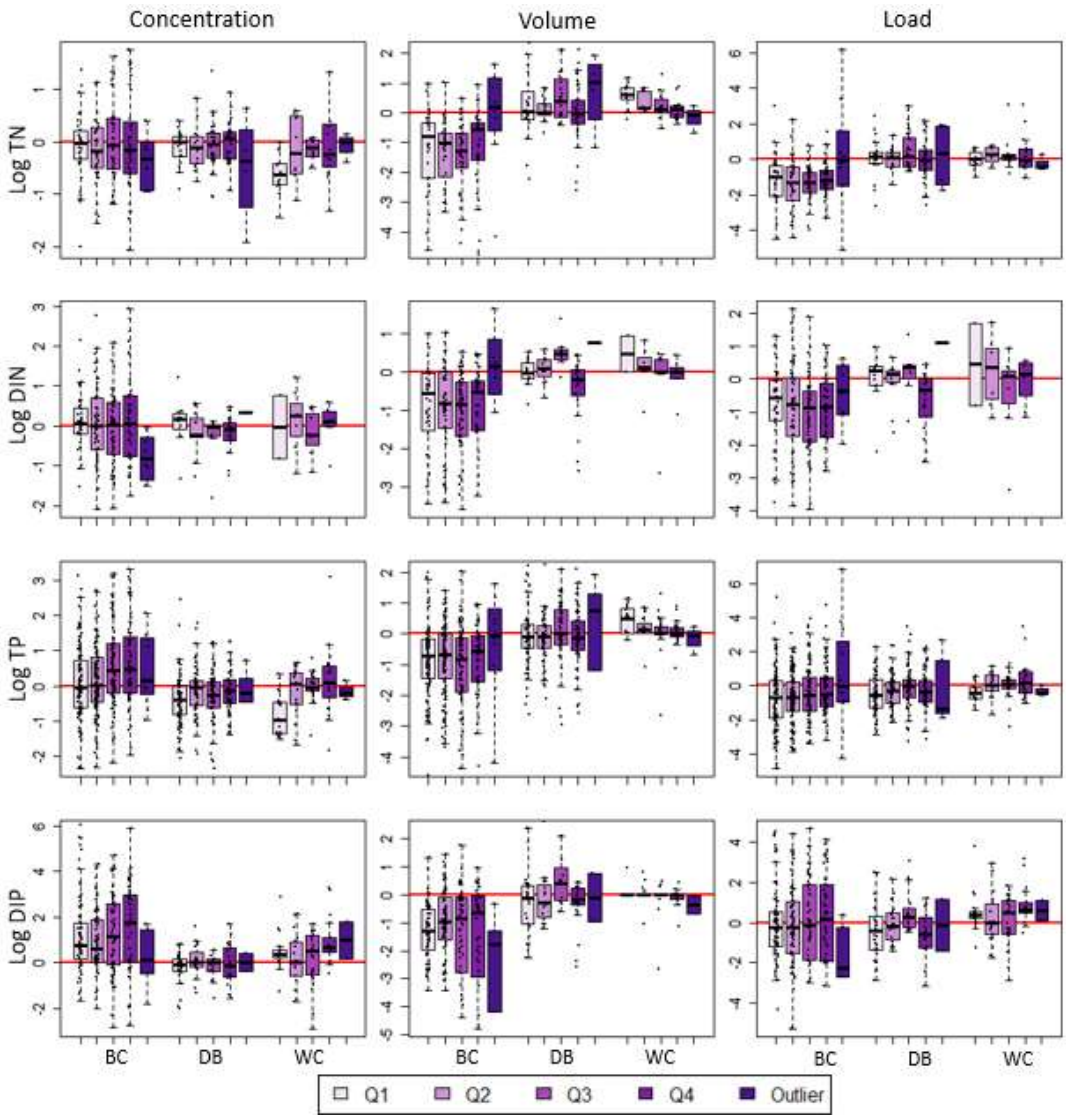


Figure C1.2: Boxplots comparing the log concentration, volume, and load removal fractions across groups of increasing peak hourly intensity for each analyte: TN, DIN, TP, and DIP, at each of the three condensed BMP types: bio-composite (BC), detention basin (DB) and wetland channel (WC). Peak hourly intensity is displayed with boxplots of increasing return interval through 5 groups: 4 quantiles and a high-range outlier group, where darker shaded BOXPLOTS indicate higher intensities. Boxplots show the interquartile range of each water quality parameter shaded in a box with whiskers extending to the minimum and maximum water quality values without outliers. Black circles represent all observed points, which are plotted on top of boxplots to convey the counts and distribution of data summarized within each boxplot. The red horizontal line is the logged removal threshold, above which effluent values are higher than influent values, indicating leaching; and below which effluent values are lower than influent values, indicating retention.

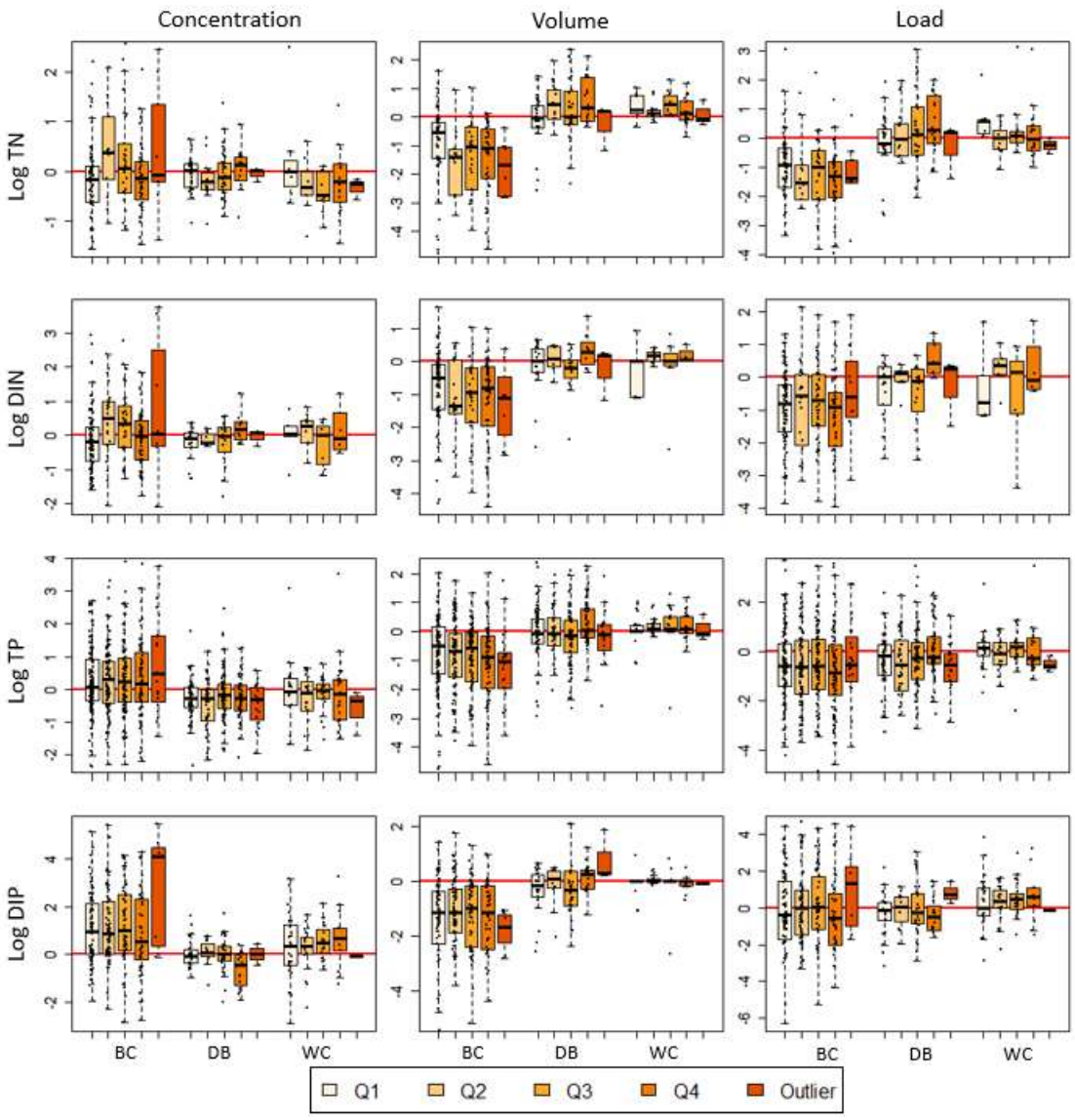


Figure C1.3: Boxplots comparing the log concentration, volume, and load removal fractions across groups of increasing antecedent dry period for each analyte: TN, DIN, TP, and DIP, at each of the three condensed BMP types: bio-composite (BC), detention basin (DB) and wetland channel (WC). Antecedent dry period is displayed with boxplots of increasing return interval through 5 groups: 4 quantiles and a high-range outlier group, where darker shaded boxplots indicate longer dry periods. Boxplots show the interquartile range of each water quality parameter shaded in a box with whiskers extending to the minimum and maximum water quality values without outliers. Black circles represent all observed points, which are plotted on top of boxplots to convey the counts and distribution of data summarized within each boxplot. The red horizontal line is the logged removal threshold, above which effluent values are higher than influent values, indicating leaching; and below which effluent values are lower than influent values, indicating retention.

Appendix C2: Clustering Selection.

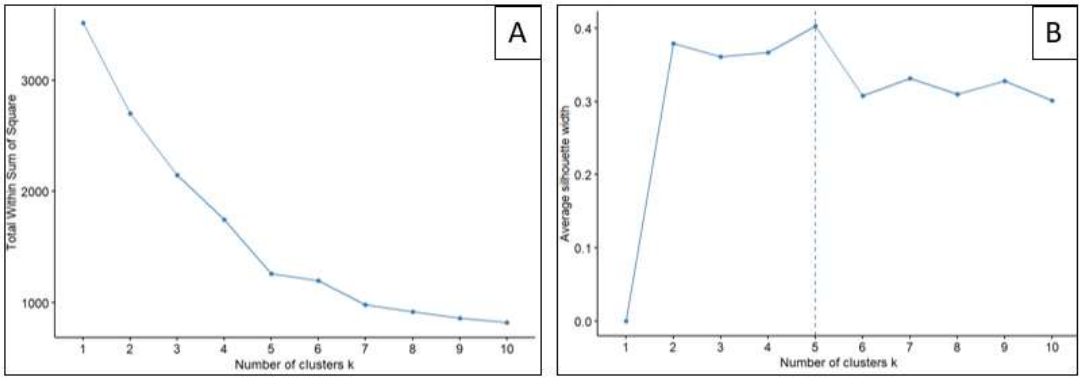
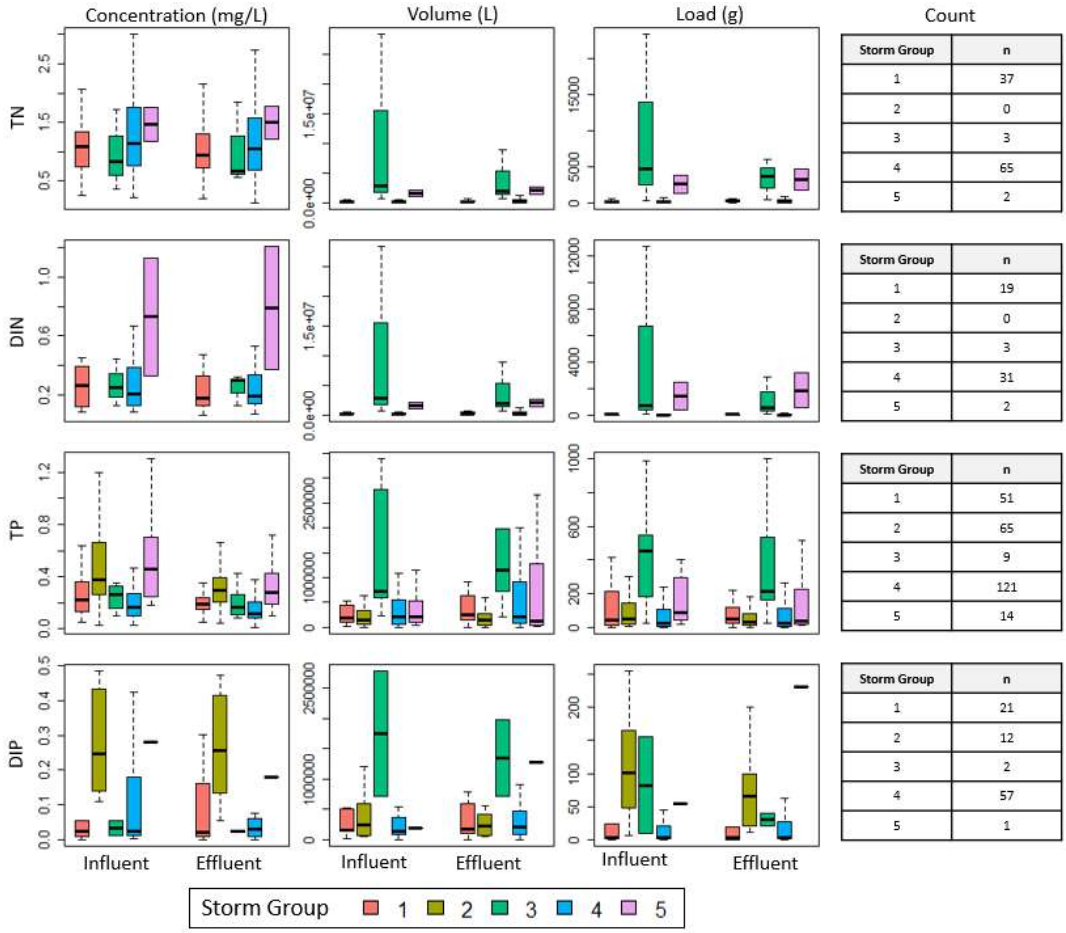


Figure C2.1: Elbow plot (A) and silhouette plot (B) indicating the ideal selection of a k-value for use in statistical clustering.

Appendix C3: Influent and effluent comparison by storm group for detention basin and wetland channels.



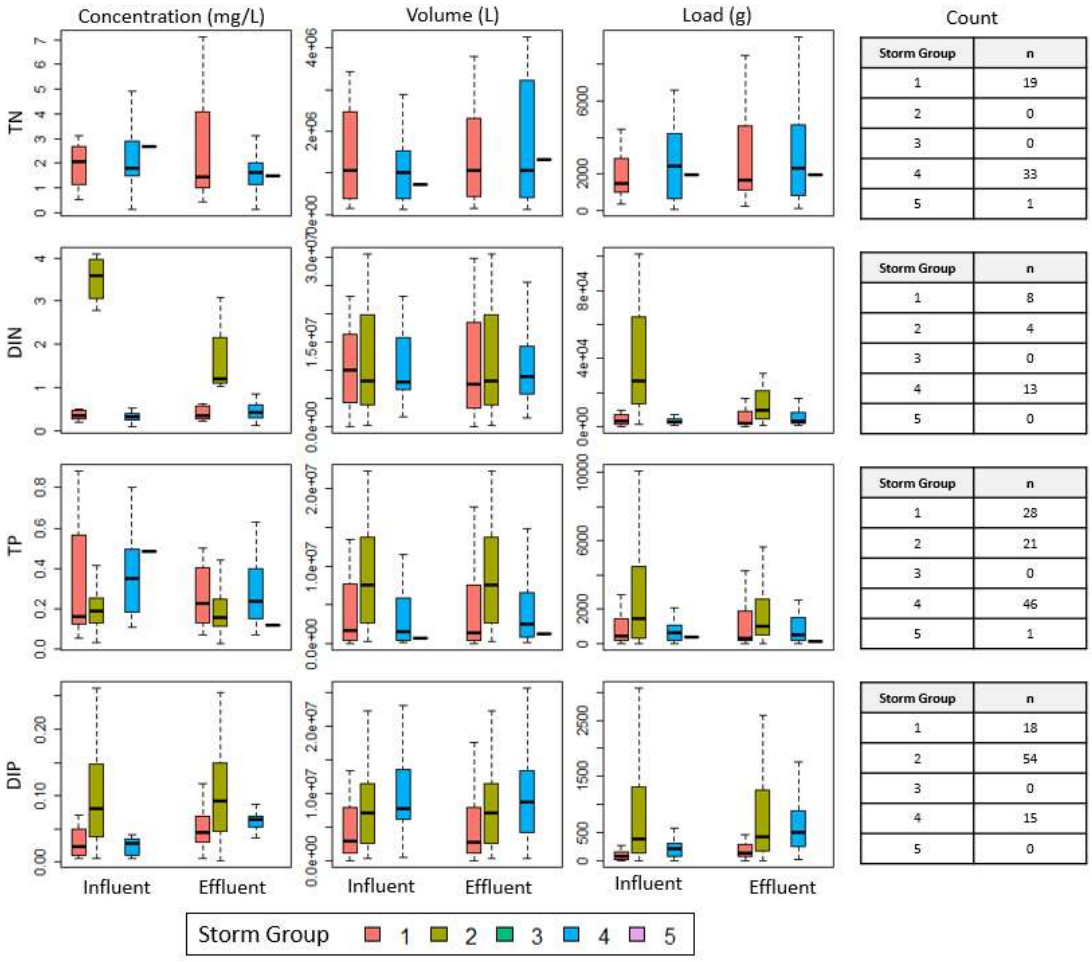
Storm Group	n
1	37
2	0
3	3
4	65
5	2

Storm Group	n
1	19
2	0
3	3
4	31
5	2

Storm Group	n
1	51
2	65
3	9
4	121
5	14

Storm Group	n
1	21
2	12
3	2
4	57
5	1

Figure C3.1 Comparison of TN, DIN, TP and DIP concentrations, volumes, and loads in influent and effluent by storm groups for detention basins. Counts of the number of paired influent and effluent data points per storm group are also provided in accompanying tables. Storm group labels 1-5 represent the 5 storm groups assigned by the clustering algorithm: intense, dry climate, large & infrequent, small & common, and long dry period. Boxplots display the interquartile range of observations with whiskers indicating minimum and maximum values without outliers.



Storm Group	n
1	19
2	0
3	0
4	33
5	1

Storm Group	n
1	8
2	4
3	0
4	13
5	0

Storm Group	n
1	28
2	21
3	0
4	46
5	1

Storm Group	n
1	18
2	54
3	0
4	15
5	0

Figure C3.2 Comparison of TN, DIN, TP and DIP concentrations, volumes, and loads in influent and effluent by storm groups for wetland channels. Counts of the number of paired influent and effluent data points per storm group are also provided in accompanying tables. Storm group labels 1-5 represent the 5 storm groups assigned by the clustering algorithm: intense, dry climate, large & infrequent, small & common, and long dry period. Boxplots display the interquartile range of observations with whiskers indicating minimum and maximum values without outliers.

Appendix C4: Tables providing summary statistics about of removal fraction for bio-composite BMPs.

Table C4.1: P-values for one-sample Wilcoxon sign rank test indicating the likelihood that a distribution of concentration, volume, or load removal fraction's true mean value is not equal to one. P-values are provided for each analyte TN, DIN, TP, and DIP, for each storm group. Green shaded boxes indicate strong or marginal likelihood of retention (removal fraction statistically less than 1) and orange shaded boxes indicate strong or marginal likelihood of leaching (removal fraction statistically higher than 1). Strong likelihood is considered p-values <0.05, and marginal likelihood is considered p-values <0.10.

Analyte	Metric	1: Intense	2: Dry Climate	3: Large & Infrequent	4: Small & Common	5: Long Dry Period
TN	Concentration	0.19	-	0.45	0.49	NA
	Volume	<0.01	-	0.30	<0.01	NA
	Load	<0.01	-	0.27	<0.01	NA
DIN	Concentration	0.73	<0.01	1	0.60	-
	Volume	<0.01	0.02	0.64	<0.01	-
	Load	<0.01	0.87	0.95	<0.01	-
TP	Concentration	<0.01	<0.01	0.03	<0.01	0.06
	Volume	<0.01	<0.01	0.08	<0.01	<0.01
	Load	0.48	0.15	0.47	<0.01	0.59
DIP	Concentration	<0.01	<0.01	0.12	<0.01	NA
	Volume	<0.01	0.02	<0.01	<0.01	NA
	Load	0.08	0.80	0.72	0.05	NA

Table C4.2: Percentages of events that resulted in net leaching or removal of concentration, volume, and load for each of the 4 analytes: TN, DIN, TP, and DIP.

Analyte	n	WQ Metric	% Leaching	% Retain
TN	169	Concentration	43	57
		Volume	11	89
		Load	12	88
DIN	221	Concentration	50	50
		Volume	21	79
		Load	22	78
TP	426	Concentration	58	42
		Volume	24	76
		Load	33	67
DIP	275	Concentration	80	20
		Volume	19	81
		Load	45	55

Table C4.3: Percentages of entries that caused leaching and retention by storm group for each of analyte: TN, DIN, TP, and DIP.

Analyte	**WQ Metric**	**1: Intense**	**2: Dry Climate**	**3: Large & Infrequent**	**4: Small & Common**	**5: Long Dry Period**
		Leach/ Retain	Leach/ Retain	Leach/ Retain	Leach/ Retain	Leach/ Retain
TN	Concentration	32.7 / 67.3	-	50.0 / 50.0	48.5 / 51.5	0 / 100
	Volume	13.5 / 86.5	-	25.0 / 75.0	8.7 / 91.3	0 / 100
	Load	11.5 / 88.5	-	25.0 / 75.0	11.7 / 88.3	0 / 100
DIN	Concentration	45.5 / 54.5	94.4 / 5.6	30.8 / 69.2	48.1 / 51.9	-
	Volume	16.4 / 83.6	22.2 / 77.8	46.2 / 53.8	20.7 / 79.3	-
	Load	21.8 / 78.2	33.3 / 66.7	38.5 / 61.5	18.5 / 81.5	-
TP	Concentration	67.6 / 32.4	77.9 / 22.1	66.7 / 33.3	49.0 / 51.0	61.5 / 38.5
	Volume	24.3 / 75.7	13.2 / 86.8	22.2 / 77.8	27.7 / 72.3	0 / 100
	Load	36.5 / 63.5	48.5 / 51.5	27.8 / 72.2	28.5 / 71.5	30.8 / 69.2
DIP	Concentration	81.6 / 18.4	88.9 / 11.1	41.7 / 58.3	79.5 / 20.5	100 / 0
	Volume	75.5 / 24.5	22.2 / 77.8	0 / 100	17.9 / 82.1	0 /100
	Load	53.1 / 46.9	50.0 / 50.0	33.3 / 66.7	43.6 / 56.4	0 / 100