6-7-2020

A Multi-Scale Assessment of the Relationship Between the Riparian Landscape and the Health of Streams in Portland, Oregon

Dylan Esmonde
Portland State University

A Multi-Scale Assessment of the Relationship Between the Riparian Landscape and the Health of Streams in Portland, Oregon

By Dylan Esmonde
Portland State University
June 7th, 2020

A report submitted in partial fulfillment of the requirements for the degree of Master of Environmental Management

Project Committee:

Joseph Maser Ph.D. – Portland State University - Department of Environmental Science and Management

Chris Prescott – City of Portland Bureau of Environmental Service

Patrick Edwards Ph.D. – Portland State University Department of Environmental Science and Management

Abstract

The riparian areas adjacent to streams are sensitive and important environments that have a great deal of influence on the overall health of the streams they border. In cities, riparian areas are routinely impacted and degraded by property developments, urban infrastructure, and other alterations to the natural environment. In this study, I worked with the City of Portland to determine how a wide range of landscape features occurring in riparian areas affect the overall health of streams at multiple scales. Data were collected from buffer areas of widths ranging from 15 ft to 500 ft that were digitally generated around 1-km reaches upstream of 59 long term monitoring stations. In these buffer areas, I aggregated information on 20 variables relating to land use, impervious surfaces, transportation, vegetation, and hydrologic control structures. Using a random forest variable selection algorithm and regression analysis tools, I created models that described the health of the urban streams (represented by macroinvertebrate IBI) as a function of the urban landscape variables in the multiple riparian buffer areas surrounding the 59 study reaches. In the process of model refinement and analysis, I determined that the 50 ft and 500 ft buffer models were sufficient at describing the spatial variability in the datasets. I found that the length of piped stream in both of the buffer areas was negatively correlated with stream health but that it had an especially strong influence in the 50 ft buffer areas. In these same buffer areas, the mean canopy height was positively correlated with stream health, but its effect was much more pronounced in 50 ft buffer areas. Additionally, my analysis showed that impervious surfaces in riparian areas within 50 ft of streams had a strong, negative impact on the health of the 59 stream reaches studied. These findings suggest that health of streams in the study area could be improved by planting more trees in riparian zones, daylighting piped waterways, and by removing impervious surfaces from areas within 50 ft of streams.

Table of Contents

Introduction

Urban streams are highly affected by their surrounding landscapes and the direct and indirect actions of individuals, businesses, and society in general. Researchers coined the term "urban stream syndrome" to describe the suite of problems faced by urban streams (Walsh et al. 2005). The symptoms include a flashier hydrograph, elevated concentrations of nutrients and contaminants, altered channel morphology, and reduced biotic richness due to increased dominance of tolerant species (Walsh et al. 2005). In many urban environments, stream courses have been modified so that their main functions are to transport waste and storm water out of population centers (Kaye et al. 2006). These types of urban streams often are structurally reinforced, realigned to suit development, and disconnected from the overall hydrology and ecology of the area (Paul and Meyer 2001).

The urban area of Portland, Oregon is host to many intermittent and perennial streams, including higher order streams—such as Johnson Creek, in Southeast Portland—that are rearing and migration corridors for salmonids (Ward et al. 1994). Some Portland streams have rich biological communities thriving in cool, clean water, while other streams in the area are plagued by legacy pollutants, modifications, and other persisting human impacts. While the health and quality of Portland's streams vary, individual citizens, conservation groups, and the local government have shown substantial interest in preserving and improving these sensitive aquatic habitats (Shandas and Messer 2008). Additionally, the federal Clean Water Act and Endangered Species Act regulations drive many of the restoration and monitoring activities that occur in Portland streams.

The Bureau of Environmental Services (BES), at the City of Portland (COP) has been collecting data on Portland's stream conditions through their Portland Area Watershed Monitoring and Assessment Program (PAWMAP) since 2010 (City of Portland 2011c). PAWMAP is based on the Environmental Monitoring and Assessment Program (EMAP) created by the U.S. Environmental Protection Agency (EPA) to use statistically rigorous methods to systematically assess the conditions of Portland streams. The BES collects measurements of the biological, physical, and chemical parameters of streams through standardized, quantitative methods. The BES uses the data gathered to monitor streams, to identify watersheds in need of corrective measures, and to create annual reports on the health of Portland's urban waterways (City of Portland 2018).

Study Objectives and Project Partner Interest

In addition to being used to monitor changes in Portland's watersheds, the PAWMAP dataset also offered new opportunities for research into other questions about Portland's streams. In speaking with BES project lead Chris Prescott and reviewing PAWMAP reports together, we agreed that the data could be used to better understand how the range of activities and landscape modifications occurring in riparian areas affected their conditions and overall health. We then outlined the following goals and objectives for this study, which were to:

- Understand how much of a stream's health can be explained by the features of the surrounding urban landscape

- Learn how the alterations and development in the riparian zone surrounding urban streams affect the overall health and ecological function of these streams
- Determine if the significance of impacts of landscape features changes when the landscape area under consideration is analyzed at different scales; and to
- Create statistically sound and replicable models that describe the relationship between landscape features and stream health at multiple scales.

To accomplish these goals and to provide the BES with data and findings that are useful to their mission, I planned to model the PAWMAP stream health data in the context of the urban riparian landscape. This study will also complement other work being done at the BES that involves examining these same issues throughout the entire area of a watershed and at radial distances from PAWMAP sampling points. An understanding of the significance of landscape features at various scales will help inform future BES work attempting to balance the needs of the modern urban society in the Portland area while protecting, maintaining, and improving the ecological communities and functions.

Scope of Work

The stream health data I used in this project was collected as part of PAWMAP over a four-year period, beginning in 2010. This data was collected from the watersheds of tributaries to the Willamette River; Johnson Creek, Tryon Creek, Tualatin Creek, Fanno Creek and the Columbia Slough at 128 randomly selected points. At each of these 128 sampling points, the parameters collected were of biological, physical, and chemical categories. This work was done to get a comprehensive, wide-ranging dataset for analysis of several facets of the urban streams and to evaluate the relative severity of the many stressors affecting these systems.

For this project, I utilized the macroinvertebrate data collected for PAWMAP in the field, along with existing GIS datasets, to explore connections between the conditions of the altered landscape and the conditions of Portland streams. Using measurements of macroinvertebrate assemblages as a proxy indicator for overall stream health, the relationship between the urban landscape and the health of the stream was explored and analyzed. I utilized the disparate datasets to create predictive models that described stream health as a function of the urban landscape across a range of spatial scales.

PAWMAP History and Background

The 2005 Portland Watershed Management Plan (PWMP) is a policy directive that steers the work of the BES in regard to streams and overall watershed health. While the city has always operated in certain ways to comply with federal laws and regulations, PWMP explicitly expanded the BES mission to protect fish and wildlife while enhancing the less tangible aspects of livability, vitality, and aesthetics that nature provides in urban areas. To fulfill the broad vision laid out in the PWMP, the comprehensive PAWMAP was designed and launched by the BES in 2010. The data generated by PAWMAP would be used by the BES to establish baseline conditions; evaluate

the success of the plan over time; and provide data to inform policy and planning going forward (City of Portland 2011).

PAWMAP is a quantitative, science-based program that standardizes data collection methods and analysis within Portland's watershed. While the focus of PAWMAP is on aquatic systems, it is designed to view streams in the context of their watersheds and the overall urban landscape. The city-wide, standardized protocol is designed to establish baseline conditions, identify problematic stream systems, minimize overlap in governmental monitoring efforts and ensure the city complies with local, state, and federal rules and regulations (City of Portland 2018).

The BES aspires to use PAWMAP to integrate monitoring of physical, biological, and chemical facets of a watershed to provide comprehensive data that will inform future management and restoration efforts. For a stream, these facets are observed in the hydrology, water quality, geomorphology, living organisms, and habitat forms. The quantity and diversity of certain aquatic insects observed is a key indicator of a stream's health that is used in the analysis of this collected data.

PAWMAP was modeled on the EPA's Environmental Monitoring and Assessment Program (EMAP), which began in 1990 in the Office of Research and Development (ORD) (City of Portland 2011). The goals of the EMAP were essentially the same as those of PAWMAP, but at a much larger scale. Study areas defined in EMAP include ten roughly equally sized groupings of states: the Mississippi River Basin, the U.S. coastline, and the entire area west of the Rocky Mountains. The overlap and variable sizes in the study areas allowed for an analysis and exploration of the data that would link the measured in-stream health and function to the characteristics of hydrologically contributing areas of varying dimensions (Messer et al. 1991).

EMAP attempted to standardize data collection methods across state agencies, the federal government, and tribal organizations (Messer et al. 1991). The "Field Operations Manual for Wadeable Streams" (Peck et al. 2005b) created for the program detailed collection methods and equipment to be used and provided sets of forms to keep data organized, but it also allowed for modifications based on local variability. The data generated in EMAP was used for ensuring compliance with the Clean Water Act and the Endangered Species Act. It was also used to characterize the condition of aquatic ecological systems at a variety of scales and provide data that could be used for further analysis by governmental and institutional researchers as well as for private business (Messer et al. 1991).

In 2006, the EMAP program concluded operations at the ORD, but was continued in a modified way as the National Aquatic Resource Surveys (NARS) within the EPA's Office of Water. While the monitoring work continues as NARS, PAWMAP is chiefly built on the work done by the EPA with EMAP. Connecting PAWMAP to work done with EMAP and to the years of research done on stream impacts and function allowed for the program to be grounded in established methods and the findings of other studies.

Literature Review

Streams and rivers flow through many towns and cities of the Pacific Northwest (PNW) region (Naiman et al. 1992). The concentration of humans along these waterways have impacted stream systems in a variety of ways. Direct alterations to the streams, in addition to the land uses

and activities in the surrounding watershed, have led to changes to the physical and chemical properties of waters flowing through these areas (Walsh et al. 2005). For this reason, city governments, land management agencies and academic researchers have a substantial interest in understanding the mechanisms and degrees of impacts between streams and the features of the surrounding urban landscape (Beschta et al. 1997; Naiman et al. 1992; Olson et al. 2007).

Impervious Area

As towns and cities in the PNW became established and expanded, the natural landscape was gradually covered by buildings, roads, and other types of impervious surfaces (Chang 2007a). These impervious surfaces prevent waters from infiltrating soils and the groundwater table, which then may lead to an increase in overland flow volume in a watershed. (Arnold and Gibbons 1996; May et al. 1998). Additionally, the reduction in infiltration rates can create higher velocity overland flows that reach streams more quickly. The effects of impervious surfaces on streams are seen in the increases in flow volume, gauge height, and in the shortened time between a rainfall event and peak stream levels (Alberti and Booth 2007).

Because water flowing across impervious surfaces can reach streams faster, a stream's volume and velocity can suddenly and dramatically increase as a result of a rainfall event. Streams that have these characteristics are commonly known as "flashy" streams. Flashy streams can lead to greater erosion of the bed and stream banks, removal of habitat structures, and even removal ("flushing out") of established biota. The high energy of these flashy streams straightens channels and removes gravel and sediment in urban stream systems that are separated from replenishing sources of these geologic materials (Booth and Henshaw 2001). This can create streams that are less morphologically complex and poorly suited for forming stable and complex ecosystems. A reduction in infiltration also can lead to reductions in baseflow volumes (Morley and Karr 2002) which has been observed in in the PNW. This is thought to be of less significance to aquatic ecology than the increase in flow volume and timing due to frequency and severity of these flushing out events (Wang et al. 2001).

May et al. (1998) found that impervious areas in a watershed have a strong, positive correlation with the 2-year storm to baseflow ratio (which is a way of describing the flashiness of stream). In their study that analyzed multiple watersheds in the Puget Sound Lowlands, the authors found that the watersheds with only 5% impervious area had streams with significantly reduced index of biological integrity values. The index of biological integrity (IBI) is a metric that weights the relative abundance of indicator species of concern and integrates these quantities into a single value that conveys ecological conditions so that sampling sites can be compared to one another (Karr and Chu 2006). May et al. (1998) concluded that the reduction in diversity was likely caused by the flashy flows, and the negative effects could be seen in both the macroinvertebrate and salmonid communities. Another study conducted in the Puget Sound Lowlands examined the relationship between impervious area in the urban watersheds and channel stability. The researchers found that the watersheds with greater than 10% impervious area had generally unstable channels. They found that this lack of stability led to greater rates of erosion, removal of habitat, and unpredictable flow regimes (Morley and Karr 2000).

Roads and Stream Crossings

Roads are impervious surfaces that are an essential and omnipresent feature of urban environments. A road can act like a stream channel in that it concentrates and transports stormwater down-gradient on a confining, impervious surface (Coffin 2007). This concentration of flow has shown to be effective at gathering automobile-generated pollutants from the road surface and transporting them to streams through outfalls or stormwater facilities (Jones et al. 2000). By concentrating flow on a road surface and into collection systems, the natural flow pattern of stormwater is altered, bypassing riparian areas that would allow groundwater percolation, plant uptake, and filtering (Hatt et al. 2004). In certain instances, these changes in flow patterns and the reductions in pollutant filtering ability have been linked to a net decrease in the volume and quality of water available for basic biological processes, but that is not the trend generally observed in PNW streams (Wheeler et al. 2005; Coffin 2007).

Booth et al. (2004) found a direct, negative correlation between impervious areas (which included a large quantity of roads) and the IBI values measured in urban streams in the Puget Sound region. In this region, it was found that the impervious area was correlated with many other aspects of urban development and that it should not be the only metric used to predict stream health. As the slope of a road is logically tied to its ability to transport surface flows, the interaction between the two variables was evaluated in the Puget Sound area study (Booth et al. 2004). While the issue of traffic intensity was considered, a suitable, scale-appropriate dataset was not found (Booth et al. 2004). Other urban area studies found the total suspended solid and heavy metal loads increased in streams in watersheds with greater road densities (Lane and Sheridan 2002; Wojciechowska et al. 2019).

Stream crossings (structures that allow the passage of automobiles over stream channels) have also been linked to overall stream degradation in cities (Forman and Alexander 1998). This is thought to occur because the road bypasses riparian areas, which discharges the pollutants generated from automobiles directly to streams (Trombulak and Frissell 2000). The physical obstructions created by stream crossings in the PNW have also been linked to a reduction in aquatic organism diversity, dispersal, and overall migratory opportunities, due to the practice of piping streams under road crossings. Additionally, crossing structures built over streams often constrict and concentrate flows so thoroughly that aquatic organisms and habitat are flushed out of a reach by increased flows occurring after relatively small rainfall events (Castro and Beavers 2016).

Municipal Zoning

In the urban environment, the landscape is modified and utilized by humans in many ways (Allan 2004). This range of modifications and utilizations are broadly defined as "land use," which is a core component of planning and land management (Allan 2004). Land use management policies can be used to protect natural resources by defining the allowed activities and developments that can be undertaken in a given area. This tool can be used to physically separate activities and developments from vulnerable resources such as streams and other water bodies (Paul and Meyer 2001). While any activity or development on the landscape will have some effect on the environment, the degree and severity vary based upon the formal land use designation.

To address this variance in impact, urban ecologists use the categories of industrial, commercial, or residential to study the environmental impacts of different types of land use (Chew and McMahon 1999; Brezonik and Stadelman 2002; Hatt et al. 2004). While there are more refined and nuanced subsets within each, these broad categories are used by land management agencies in the Pacific Northwest region (Nelson and Moore 1993; Franczyk and Chang 2009). Of these three categories, industrial land has been shown to generate the greatest amount of pollutants in urban areas (Stepenuck 2002; Nedeau et al. 2003). These pollutants include materials directly generated by industry, as well as the associated support services of trucks and trains that produce petrochemicals, heavy metals, and other organic and inorganic pollutants as they transport products from industrial land use areas to consumers (Groffman et al. 2003; Allan 2004; Wiłkomirski 2011). Additionally, many industrial operations generate waste byproducts that are hazardous and must be dealt with in specific ways. Even when best practices are followed, these waste byproducts can make their way into the local ecosystem (Lowe and Evans 1995).

Conversely, population density and full-time building occupancy is often less of a factor in industrial areas compared to residential areas. Because businesses mainly utilize and occupy industrial areas during a business's operating hours, the pollutant-generating activities originating from some industrial operations only occur during a limited number of hours in a day. This lessens the potential impact of the industrial land use class (Bernhardt and Palmer 2007). Residential areas often have vegetated landscapes which can increase infiltration, but the maintenance of yards and gardens with fertilizers increases the nutrient load running off residential properties (Groffman et al. 2004). These scenarios show that land use does not have a clear, linear relationship with the generation of pollutants and altered flow, but instead that there are trends and patterns that vary within localities and study areas (Sorrano et al. 1996; Snyder et al. 2003).

There are many studies that examine varieties of land use and their respective impacts on streams, but few that focus on the different uses within an urban area. One such study, conducted in the Raleigh, North Carolina metropolitan area used the classifications of urban, forested, and agricultural land use and found that streams in the urban areas had significantly higher suspended sediment and metal concentrations after storm events. As in this project, macroinvertebrates were used as a metric to quantify overall stream health (Lenat and Crawford 1994). A study by Sonoda and Yeakley (2007) examined the phosphorus (P) levels in streams along the rural to urban gradient in the SE Portland area. They concluded that the mineral composition of soils in rural areas were strongly correlated with P concentrations in the streams while land use patterns were the primary drivers of P levels in stream waters in the urban areas.

Other studies have observed similar relationships between land use patterns and stream pollution, but also concluded that the impact of land use varies greatly by watershed and region and that there are issues of covariation with other anthropogenic and natural landscape features which made it difficult to establish a clear causal relationship between land use, pollutants, and stream health (Walsh et al. 2005). Additionally, the impacts of historical land utilizations were deemed too difficult to reconcile with current uses (Allan 2004).

Riparian Vegetation

In riparian areas, trees, shrubs, and grasses provide shade, sediment retention, leaf litter for consumption, and habitat structures for instream fauna (Osborne 1993). Broadly defined, a riparian area is a zone surrounding the stream on which the soils, biota, and overall ecology is influenced by and dependent on the flow of a stream (Elmore and Beschta 1987). Riparian areas have a direct and complex impact on the biological, chemical, and physical parameters of the streams that flow through them (Ekness and Randhir 2007). Streams with heavily vegetated riparian areas exhibit fewer of the detrimental impacts associate with urban development (e.g., increased pollutant loads, flashier flows, and reductions in biological diversity) than urban streams with sparse riparian vegetation (Lenat and Crawford 1994).

Riparian areas in urban environment can act as buffers that intercept the runoff and associated pollutants flowing towards a stream. Ions and metals such as zinc, chromium, and copper, can be captured by soils and vegetation, which reduce the quantities of these substances that might reach the stream (Hupp et al. 1993; Dow et al. 2006). These chemicals, generated from a range of anthropogenic sources such as transportation, landscaping, and industrial processes, can be toxic to aquatic organisms (Pascoe 1986; Rolline 1998). Riparian areas also can intercept nutrients such as nitrogen and phosphorus. Unlike metals, these nutrients (from sources such as the fertilizers used in landscaping) are utilized by vegetation in the system which increases plant growth. An excess of these nutrients entering a stream can lead to eutrophication, spurring on rapid and abundant algae growth—also known as an "algal bloom" (Carpenter 1998). While algae are a necessity in aquatic ecosystems, algal blooms can overwhelm lentic and lotic systems, leading to oxygen depletion, the creation of harmful toxins, and an overall decrease in biodiversity (Smith 1999).

In the PNW, riparian environments often provide energy inputs to a stream through litterfall. Litterfall, consisting of leaves, needles, and small pieces of wood, are shredded, processed, and consumed by macroinvertebrates. Litterfall is a major food source in macroinvertebrate communities; it provides them energy to thrive and grow. Macroinvertebrate communities then function as a source of food for other organisms, including fish, reptiles, and amphibians (Benfield 1997; France 1998). Macroinvertebrates and other insects living on riparian plants above streams can also be a substantial source of food energy for organisms in the stream. This can come about through organisms falling into the stream or, in a more dramatic example, fish jumping out of the water to capture their prey from the vegetation above (Baxter et al. 1997; Kawaguchi et al. 2005).

Wetlands are ecosystems that are dependent on saturated soils. These saturated soils allow for a range of hydrophilic plants to grow. The hydrologic connectivity between streams and their surrounding landscapes can allow for soils to develop in the riparian corridors that allow for these wetlands ecosystems to form (Chopin et al. 2002; Titus et al. 2006). A study conducted in Oregon and Washington found that wetlands in watersheds are directly responsible for reductions in total suspended solids (TSS), fecal coliform, and nitrates in streams. In that study, the proximity of a wetland to a stream was shown to be a major factor, where the wetlands closer to the stream were more effective at providing these types of filtering and uptake ecosystem services (Thorn et al. 2001). This same study found that altered flow regimes and land development has led to a drastic decrease in the size of wetland area in the two states. Wetlands

in Oregon have also proven to be effective agents of denitrification and sewage processing, but their filtering and processing capacity can be easily exceeded by heavy rainfall and surging stormwater runoff in urban areas (Song et al. 2012). Seasonality and altered flow regimes have been tied to the ability of PNW wetlands to perform these services, which suggests that wetlands cannot be expected to provide a filtering function in all conditions on all days of the year (Reinelt et al. 1998; Shaffer et al. 1999).

Stormwater Infrastructure

In PNW cities, streams are sometimes diverted into underground piping systems that control the direction, volume, and gradient of flow. Streams are transported underground to allow for development of property, to control flooding, and to make stream response to rainfall events more predictable (Hatt et al. 2003). Piping streams can be an effective way to control flows in a given section of a watershed, but this practice can lead to increased flow rates and volumes in the natural channel downstream (Walsh et al. 2004). By piping streams through sections of a watershed, these segments of a stream become disconnected from the surface and subsurface flows that surround them (Walsh et al. 2005). Additionally, riparian soils that become disconnected from streams become drier, which can reduce the denitrification abilities of plants. Riparian soils that previously acted as nutrient sinks can even become net nutrient sources (Groffman et al. 2003).

While the redirecting of flows underground can be a useful practice from an urban development perspective, the isolation of a stream from its environment leads to segmentation of aquatic habitats, alteration of ecological processes, and to an overall loss of biotic diversity in a stream system (Hughes et al. 1995; Ladson et al. 2006).

Municipal Separate Storm Sewer Systems (MS4) are types of infrastructure facilities that collect surface waters from the urban landscape and transport them to streams, lakes, and other waterbodies (City of Portland 2011a). These systems are regulated by the Clean Water Act because the effluents from MS4 outfalls generally contain pollutants of concern. For this reason, the volumes and quality of water are monitored at key points in the system to assess the discharge of pollutants into waterbodies (Minan 2005).

Though MS4 represent an attempt to regulate and monitor pollutants in watersheds, they also can be viewed as a direct point source from where the concentrated pollutant load of a drainage flows into a stream. The receiving waterbodies below MS4 outfalls routinely experience increased levels of organic and inorganic pollutants, fine sediments, pathogenic bacteria and increased temperatures (Kuivila et al. 2000, Van Metre et al. 2000). The concentration of diffuse flows into MS4 facilities and then into waterbodies have even been shown to cause detrimental geomorphic alterations to stream channel morphology (City of Portland 2011b). While the pollutants flowing out of MS4 facilities into streams are closely regulated and monitored, the pollutants that they carry can still have both acute and chronic detrimental impacts to stream biota at all trophic levels. (Carpenter 1998, Dalu et al. 2017) In the COP, bioswales are commonly used to intercept MS4 flows so that nutrients and metals can be absorbed by soils and plants before the waters discharge to ground or surface waters (City of Portland 2011). A study that examined the effectiveness of stormwater best management practices (BMP) on Portland's Columbia Slough watershed recognized that there were problematic levels of pollutants being

discharged through MS4 facilities, but also found that relatively simple measures of community engagement and education could be taken that would lead to substantial reduction in the total pollutant load in MS4 effluents. The measures taken by the COP included working with landowners to design and implement improved erosion control practices, conducting workshops on the benefits of naturescaping on private property, and informing businesses about the importance of properly recycling of oils and greases (Hottenroth et al. 1999).

Macroinvertebrates

Since the 1980s, there has been a trend toward using macroinvertebrates as a proxy for overall stream health (Rosenberg and Resh 1993; Wright et al. 2000). Some macroinvertebrates are tolerant to pollutants and extreme conditions, while others have only a narrow range of conditions they can live and thrive in. Because macroinvertebrates have this range in tolerances of environmental stressors, it is possible to gain some understanding of a stream's health based on which macroinvertebrate organisms are present in an aquatic environment (Hubler 2008). This is done by comparing the macroinvertebrate community composition of a study reach (the "observed" score) to that of a previously sampled reference site (the "expected" score). The reference sites are in streams reaches that have not been degraded by human activities. The ratio of the observed value to the expected value provides an estimate how impacted a given stream reach is (Hubler 2008).

Study Area and Reach Length

A stream "reach" is a unit of analysis that refers to a segment of a stream examined in a study. Studies on stream ecology and geomorphology have used reaches with lengths from 100 m to 28,000 m (Kasprak et al. 2012; Kasprak et al. 2018; Wang et al. 2006). The length of a stream reach is not standard and can even vary within a single study. These reaches of varying lengths can be based on one or more features that are common along a given length of stream. These features include (but are not limited to) slope, bed composition, pools, flow rates, habitat units, and overall channel geometry (Somerville 2010; Trainor and Church 2003). There is not a standardized length in this field, but instead lengths are chosen that are applicable to the landscape under consideration, the desired resolution in data, and the study question at hand (Brown et al. 2009).

The study area delineated in a watershed research project is the spatial extent of the landscape from which data can be collected for analysis. Some study area delineations are influenced solely by the questions at hand, while others are based on the same questions, but are also influenced by the available data, funding, and practical constraints. Other studies have a more experimental approach where a very large area is delineated so that a wider, more diverse range of features from the landscape can be considered (Wang et al. 1997). The lack of a standardized approach poses a challenge, but a rationale or method is always described and is based on the goals of the study or the findings of others.

Some researchers have focused on an entire upstream watershed as their study area, while others have used buffers of varying widths to define their area of interest (Frissell at al.

1986; Osborne and Kovacic 1993). The watershed approach aims to capture the effects of the entire hydrologically contributing area upstream of a sampling point. A buffer delineation approach captures some of the same watershed contributions, but primarily aims to quantify impacts of riparian conditions and developments and activities in areas near streams. (Jackson et al. 2015; Solins et al. 2018). There is not a common method for selecting either reach length or spatially delineating the unit of analysis for studying a stream. Though this process could be standardized across the discipline, an experimental approach based on the specific concerns of a study and the local landscape may still be the most appropriate.

While my review of the current literature found patterns and some consistency in findings (i.e., human development as the major driver of stream degradation), the methods, experimental design, and analyses varied greatly. Urban streams and their surrounding landscapes are complex systems full of confounding and non-linear relationships between variables (Booth et al. 2004). This review also demonstrated that studies are hard to replicate and that their findings are not universal. This was not seen as a weakness or fault, but something to consider when adapting methods documented in the literature for this project.

The selection of predictive variables in this project was informed by the study objectives, previous research, and practical limitations. This approach grounded the project in established research approaches and findings, but also allowed for customization based upon the specific goals of PAWMAP and the interests of the BES. Based on these considerations, I selected 20 landscape variables that included both man-made and natural facets of the riparian environment. These variables are listed in the subsequent Methods section of this report (Table 2).

Methods

Study Area

The City of Portland is situated at the confluence of the Willamette and Columbia Rivers in Northwest Oregon (Figure 1). The Willamette River divides the city between west and east. The west side features higher elevations and steeper slopes which bound the city's urban core. The more residential eastside has a gentler relief that is dotted by extinct volcano cinder cones. The path of the Willamette and Columbia Rivers from Portland to the Pacific Ocean is free flowing, undammed and is designated as a core rearing and migratory corridor for a variety of salmonid species (Friessen et al. 2005).
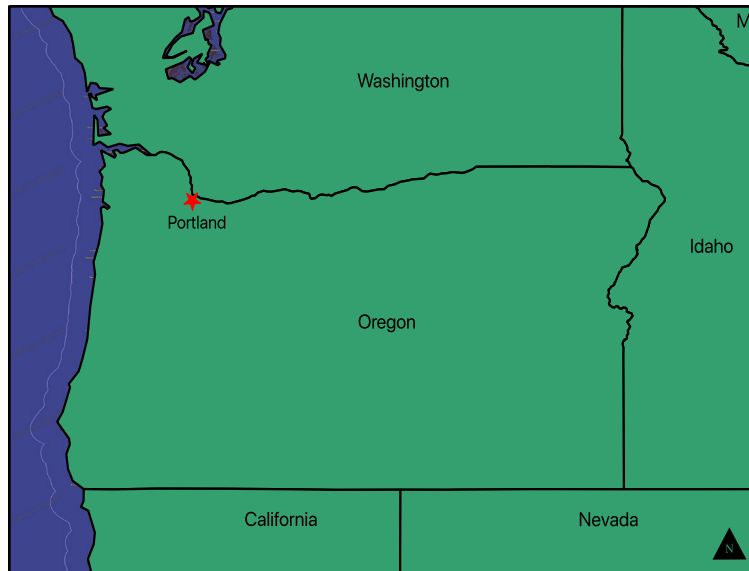
Figure 1: The red star on the map represents the location of the City of Portland within the State of Oregon.

Climate

Portland's temperate climate is greatly influenced by its proximity to the Pacific Ocean. The cool ocean, 113 km to the west of the city, helps regulate temperature variance which creates a climate of wet, mild winters and dry, warm summers. The average rainfall in the summer is 114 mm while the winter months collectively experience 355 mm of rainfall, on average. The average winter high temperature is 7.2° C while the low is 2.2° C. The average high temperatures in the summer is 25.6° C with low of 13.3° C (NOAA 2018). Due to its moderate temperatures and relatively low elevations, there is rarely snow accumulation. Because of this, local streams are driven by rainfall and runoff (Chang 2007a). While a majority of the streams in the study area are perennial, they are quite responsive to rainfall events which can lead to flooding and extreme channel and habitat alteration during the rainy winter months. The lower Willamette River area has historically had a great deal of flooding, but channelization and upstream dam controls have reduced flood frequency (Chang 2007b).

Soils and Geology

The soils in the study area are generally silt loams while the underlying geology is generally basalt. The study area geology is, broadly, sandstone and alluvium from historic flood deposits and volcanic activity (DOGAMI 1991).

Vegetation

Portland's temperate climate, long summer daylight hours, and high levels of rainfall allows for an assortment of perennial vegetation to grow. Areas along streams have been protected allowing for the establishment of riparian plant communities (Rios-Touma et al. 2015). In less developed areas, such as the parks and large swathes of the West Hills neighborhoods,

thriving forests that support complex ecological communities can be observed (Hennings and Edge 2003). Even in the more developed and urbanized areas, there are large mature trees lining streets. Palustrine and lacustrine wetlands can be observed in the low gradient areas on the eastside of the city (Kentula et al. 2004).

Population and History

Portland is a mid-sized American city with a population of about 580,000 people and an area of 376 km^2. Settled by ethno-European immigrants in the 1830s, Portland was incorporated as a city in 1845. With the influx of non-indigenous peoples in the early 19th century, the area has been greatly altered by the extraction of resources and landscape modifications required to build a city. While incrementally expanding, the Portland Urban Growth Boundary has slowed the rate of urban sprawl into surrounding natural areas and encouraged denser development which has led to greater population density (Jun 2004).

PAWMAP Sampling Stations

To provide statistical rigor for analysis, BES staff sited PAWMAP stations using a spatially balanced random approach. To do this BES staff used a generalized random tessellation survey design based on methods from Stevens and Olsen (2004). In ArcGIS (Version 10.4.1, Esri Inc. 2016), an algorithm selected 32 points along a digitized, vector representation of the stream network in the study area. The 32 random points on the stream network generated within ArcGIS became the initial 32 stations for PAWMAP. In each of the following three years, the same process was used to generate 32 additional PAWMAP station locations (Table 1).

Year	Stations Sampled			
FY 10-11	panel 1			
FY 11-12		panel 2		
FY 12-13			panel 3	
FY 13-14				panel 4
FY 14-15	panel 1			
FY 15-16		panel 2		
FY 16-17			panel 3	
FY 17-18				panel 4

Table 1: From 2010 to 2018, PAWMAP stations were sampled on a yearly rotating basis. Over this time period, each station was sampled twice (BES 2017).

This four-year process established the final 128 sampling stations used in PAWMAP (Figure 2). Out of these 128 stations, I chose to omit 49 from this study because they are intermittent streams or were on the Columbia Slough. These were omitted from this study because the stations on intermittent streams and on the Columbia Slough do not have flow regimes that are compatible with the standardized EPA macroinvertebrate sampling protocol (Peck et al. 2005b) (Figure 3).
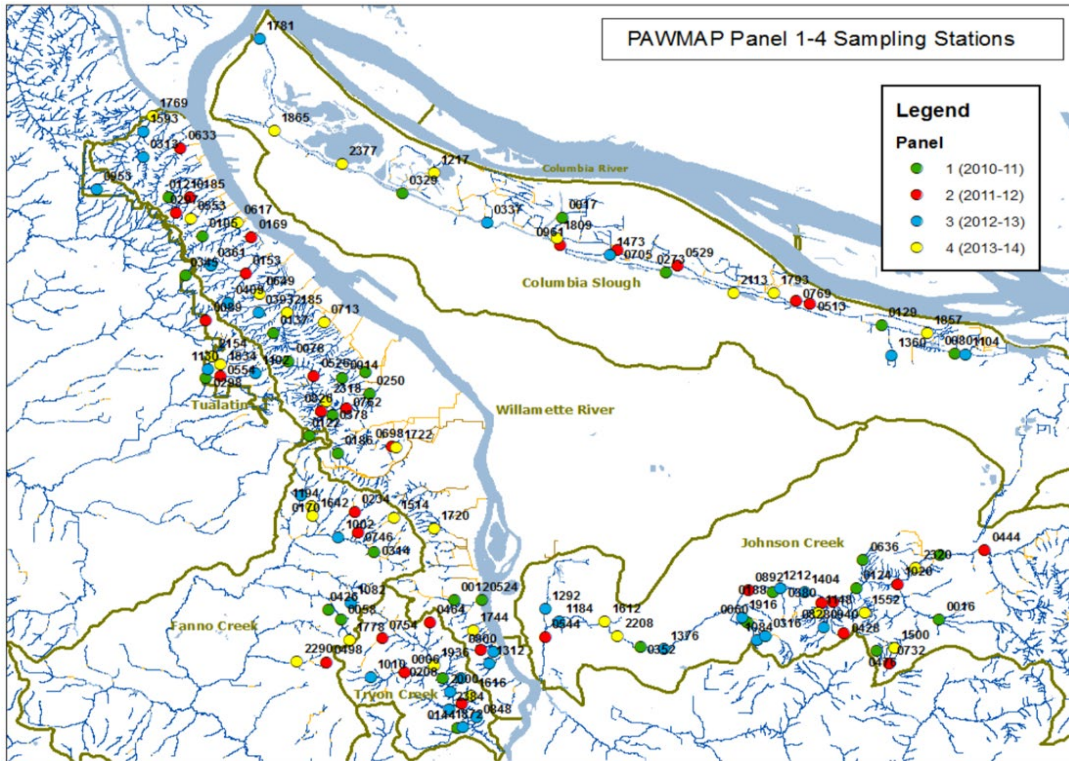

Figure 2: PAWMAP sampling stations are located on streams within Portland's four major watersheds. All PAWMAP sites were initially considered for inclusion during the design of this study, but sites on intermittent streams and the on Columbia Slough were not ultimately used because they could not be sampled with the EPA protocol.
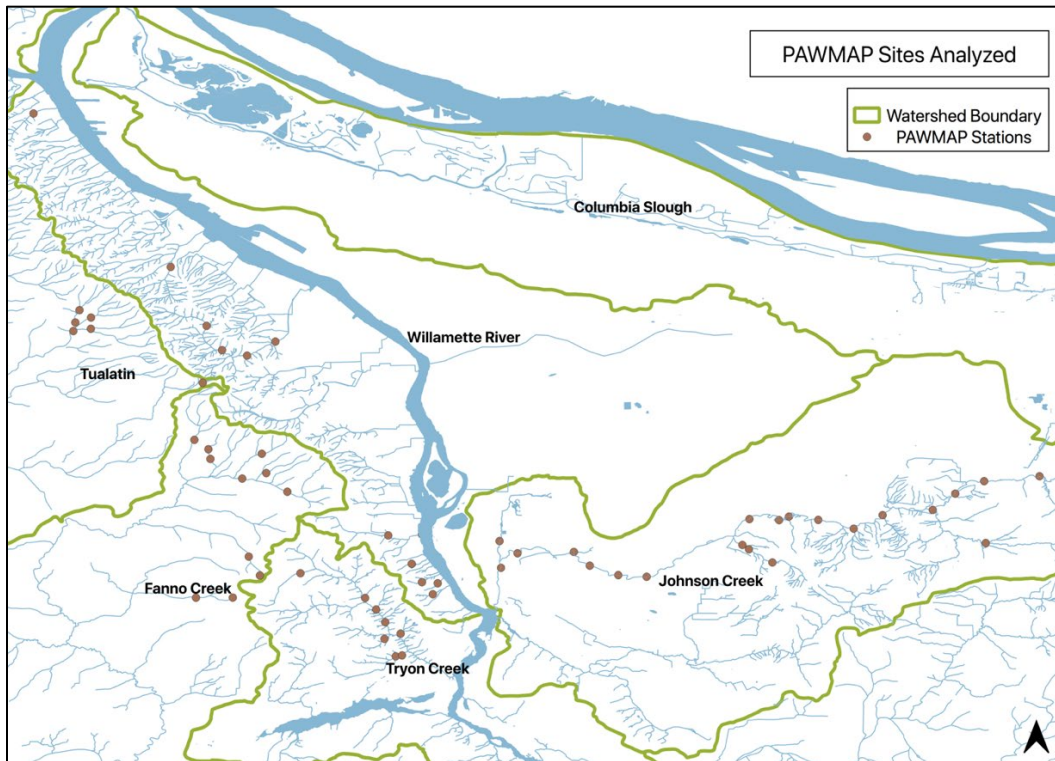
Figure 3: After removing the Columbia Slough and intermittent stream stations from the database, there were 59 PAWMAP stations remaining for analyze.

Unit of Analysis

In delineating the area of interest for analyzing the impacts of landscape variables on the PAWMAP stations, I considered my review of the literature, but also the interests of the BES. EPA researchers conducted a similar process of using their well-informed, professional judgment in selecting EMAP study buffer area delineations (Messer et al. 1991). For the analysis of the data gathered, PAWMAP project lead Chris Prescott decided on three types of delineated areas on the landscape that would provide information about the urban study area (Figure 4). The three types of areas were:

- The hydrologically contributing watershed area above a sampling station
- A buffered area around a fixed length reach upstream of the sampling station
- A radial area around the sampling station created irrespective of topography or hydrologic connectivity

Analyzing the health of a stream within the context of its contributing watershed area is the most traditional research approach and it is one seen in many classic studies (Triska et al. 1989; Carlson et al. 1990; Likens et al. 2002;). It is based on the idea that the health of the watershed can be assessed by examining the water flowing down, across, and out via streams and rivers. The watershed area is generated through processing digital elevation model (DEM) topography data to determine the hydrologically contributing area to the sampling station. The radial approach

17

puts less emphasis on hydrology and more on the avian and terrestrial fauna, and on the non-aquatic ecology in the urban environment surrounding the sampling station. This is an approach seen in more recent studies (Deacon et al. 2005; Wang 2001).

The second, lateral buffer-based approach, incorporates some aspects of the other two methods, but is designed to look at different facets of the landscape. Stream connectivity and form are factors in the buffer's geometry, but the overall size is essentially limited by the buffer width chosen and not by the size of the watershed. The goal was not to find the best study area delineation method, but instead to use the three delineations to answer three different questions. In the case of this study, Chris Prescott and I decided on a lateral buffer area delineation method because it encompassed only the riparian areas that abut the sample reaches associated with PAWMAP stations.
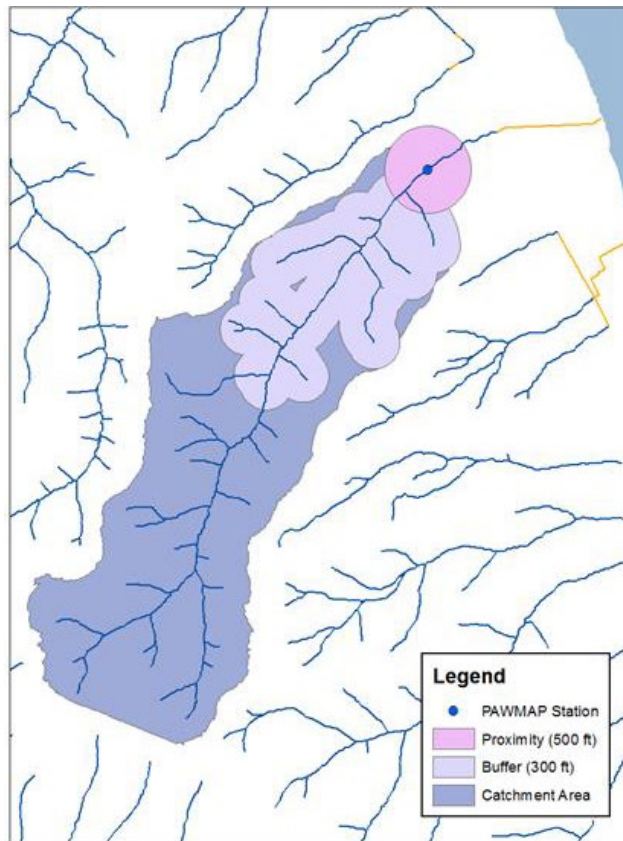

Figure 4: This map displays and the spatial extent and overlap for the watershed (catchment), radial (proximity) and buffers sample area delineations used in the greater PAWMAP project. All three polygons were generated for the same PAWMAP station. In this project, only the buffer delineation method was used.

Buffer Area Dimensions and Reach Length

For this study, Chris Prescott and I decided on widths of 15, 30, 50, 100, 200, 300 and 500 ft to delineate the riparian buffer areas. For new development in Portland, there is no single buffer width established that restricts development. Instead, COP land-use specialists choose widths based on regulations and site-specific conditions. The BES has established buffer width

recommendations that address specific ecological functions including preservation of allochthonous energy inputs, channel dynamics, water quality, microclimate stability and wildlife habitat. These suggested width values range from 10 ft to 984 ft and are based upon stream functions of concern. A 300 ft buffer is the general width designated for environmental protection zones in the City of Portland, but that width varies case by case (City of Portland 2001).

In this study, the buffer widths were selected to encompass the range of values cited in the literature as widths to consider for protecting specific ecosystem functions. Our selection of these seven widths was somewhat arbitrary and emphasized a finer level of gradation in the 15 ft – 100 ft range, but also to include larger values to capture potential impacts from areas that are outside of the immediate riparian area. The reach length in this study was set at 1 km (3280.84 ft). While this metric again is somewhat arbitrary, it was a length chosen by Chris Prescott based on best professional judgment and project goals. Additionally, it is a reach length utilized in many other urban stream studies (Roth et al. 1996; Harding et al. 1997; Peterson et al. 2001) (Figure 5).
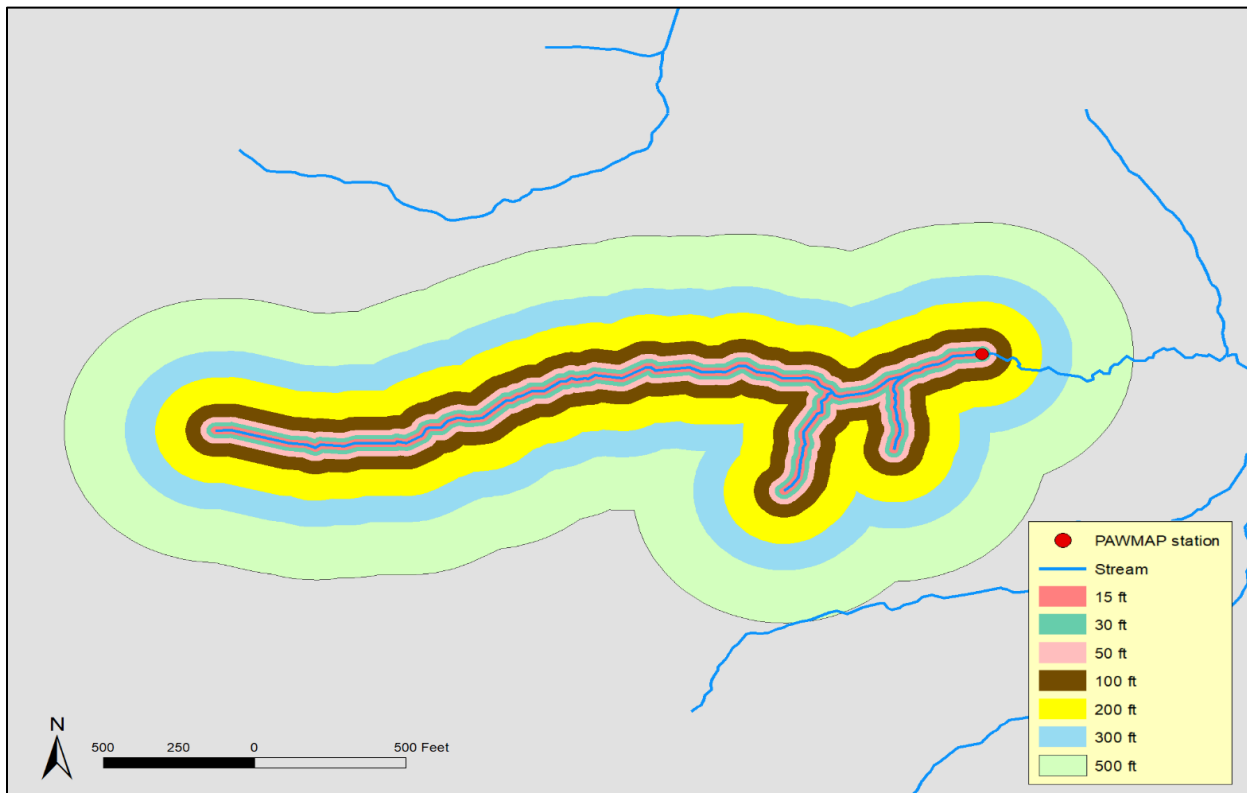

Figure 5: The stream reaches upstream of each PAWMAP station were encompassed by buffers of varying widths. The widths values used to generate the buffer areas ranged from 15 to 500 ft in width while the upstream reach length was fixed at 1 km (3280.84 ft).

Stream Health Metrics

The Oregon Department of Environmental Quality (DEQ) created the Predictive Assessment Tool for Oregon (PREDATOR) model to form a standardized, quantitative value or "score" for streams in Oregon that represents an overall measure of stream health based on

macroinvertebrate assemblages (Hubler 2008). The DEQ developed the PREDATOR model by sampling macroinvertebrates from stream reaches that were deemed, by best professional judgment, to be relatively undisturbed. The DEQ identified many of these reaches across the state and within bioregions. These relatively undisturbed sample locations were the reference sites to which other sites were compared (Hubler 2008).

Before sampling, DEQ personnel established a consistent level of practical taxa identification. Taxa with well-established species identification were grouped by the more general genus, while other taxa that could not be confidently grouped were removed from the study (Hubler 2008). This process created a uniform level of taxa resolution of 75 groupings or "operational taxonomic units" (OTU) that would allow for consistency in sampling and analysis.

While all macroinvertebrates were enumerated, the established OTU were used in the creation of the PREDATOR model. A simple "presence/absence" scheme was used in the process; if there was any occurrence of a taxa from an OTU found, that sampling site would be considered to be one that hosted an organism from that particular OTU. If an OTU occurred at less than 50% of the sites, it was removed from analysis due to a lack of representation across the study area. To create the final expected (E) values, the measured occurrence probabilities of the 75 OTU at reference sites were summed by the project's researchers (Hubler 2008).

To create E values for streams that had not been surveyed, the sampled sites were grouped and analyzed through cluster analysis. (Hubler 2008). This process weighted the environmental conditions and landscape characteristics of the areas surrounding the sampled stream reaches. With these conditions quantified and identified, anthropogenic variables were then removed to create the model that best predicted stream health based upon the natural surroundings. After a process of calibration, exploration, validation and refinement, a unique model for Oregon's three ecoregions was created (Hubler 2008). This was done to take into account the geographic and climatological dissimilarities across the state. (Hubler 2008) determined that an optimal PREDATOR model for the Willamette Valley would only utilize the day of the year and the longitude as predictive variables.

Using these two variables, E values for all streams in the Willamette Valley ecoregion were generated. Using the same methods that were used to characterize reference sites, a PREDATOR score could then be generated in any stream in the area by dividing the observed (O) assemblage of the defined OTU at a chosen site by the model-generated expected (E). With a range of zero to one (0-1), this O/E value quantified a given stream reach's overall "health" in relation to a theoretically undisturbed stream from the same ecoregion. With 0 being the most disturbed and 1 being undisturbed, this model provided a simple metric for assessing streams in the area.

Macroinvertebrate Sampling and Processing

Using methods described in the "EPA Field Operations Manual Wadeable Streams," benthic macroinvertebrate assemblages were collected from PAWMAP sites by BES staff and myself (Peck et al. 2005b). A surber net with a 1 ft x 1 ft opening featuring a 500 µm mesh screen was placed in the center of the stream 1 m downstream of a riffle. The personnel collecting the samples then used a scrub brush to agitate the materials on the rocks directly upstream of the surber net for 30 seconds. The brush dislodged macroinvertebrates, algae and periphyton from the rocks which were then caught in the surber net. By using 30 seconds as the period for

scrubbing and agitating the rocks, the sampling efforts could be standardized across PAWMAP sites. After the 30 seconds of active sampling were completed, the surber net was inverted and the contents were poured into a sieve bucket with 500 μm openings. A spray bottle was used to dislodge all benthic samples from a sieve bucket and into HDPE sample jars. In the jars, samples were preserved with a 95% ethanol solution.

This process was conducted at all 59 PAWMAP sites based on the aforementioned sampling schedule rotation. The sampled macroinvertebrates were enumerated and identified to the species level by laboratory technicians. Using this data, PREDATOR scores at the PAWMAP sites were then generated.

Aggregation of Spatial Data

Informed by a review of the relevant literature and with the guidance of Chris Prescott, we decided on 1 km to be the standard stream reach length used in this study. COP GIS specialists created a modified ArcGIS "Network Analysis Tool" to delineate stream reaches for analysis. I used the reach length of 1 km as an input for the tool to create a new GIS layer consisting of 59 reaches above the PAWMAP sites. Flow direction data in in the stream network dataset was also used as an input for the GIS tool to ensure that the reaches were oriented upstream.

Along these 1 km reaches, I created buffer polygons that delineated the analysis area for each station via the "Buffer" tool in ArcGIS. I configured the tool to generate buffers with equal areas on each side and to have rounded ends (Figure 4). A geometry function within the ArcGIS tables calculated the dimensions of these buffers and I then added their area values to the master attribute table where they could be used in further calculations. I repeated this process for each of the seven buffer widths resulting in seven total buffer polygon datasets. ArcGIS automatically linked the station ID of each PAWMAP site to their respective buffers. This ID provided a unique value to link other variables generated to their respective PAWMAP station.

With the buffer polygons created for all PAWMAP stations, I next used the "Intersect" tool in ArcGIS on the variables of the polygon classes to create a subset that included only the area of a feature in question that was inside the buffer area. I used a similar process to select point and line features from within the buffer areas. Utilizing the common station ID, I joined the attributes of the subset feature created by the "Intersect" tool, as well as the PREDATOR scores, to their respective PAWMAP buffers.

Using a feature height model (FHM) raster layer and vegetation data, I created the canopy height dataset. The FHM is a LiDAR sourced layer that displays the elevation of the highest object at a given pixel above the earth's surface topography. If there was not an object at that point, the FHM would equal the DEM, but if there was an object (i.e., a tree or a building), the FHM would be equal to that object's elevation above the DEM of the earth's surface. Using a BES-sourced polygon layer that categorized Portland canopy composition, coniferous and deciduous cover areas were isolated and saved as new polygon layers. Utilizing the "Raster Clip" tool in ArcGIS, I trimmed the extent of the city-wide FHM layer to match the extent of the deciduous and coniferous polygons areal layers.

This manipulation produced an FHM representing the heights of the trees in the two canopy polygon classification layers. A BES forestry staff member and I had a discussion about

maximum tree height of the two genera considered in this study. The BES forestry staff member stated that the maximum deciduous tree height in Portland was 135 ft while the maximum coniferous tree height was 242 ft. I then used these two values as inputs for the "Raster Calculator" tool in ArcGIS. I used this tool to produce a new raster layer that set the pixel values in the FHM that exceeded the maximum plausible tree height to "nodata". In LiDAR datasets, there are often small errors from the data collection process that lead to errors in the elevation data generated (Lefsky et al. 2002). Additionally, the boundaries in the canopy cover layers were likely not entirely accurate and may include areas with taller objects (i.e., buildings or telephone poles) instead of trees. This process that I devised removed pixels from the two tree canopy raster layers to create a new FHM model with height values based on the locally observed maximum heights of the two genera under consideration.

My next step was to use the "Zonal Statistics as Table" tool in Arc to extract data from the newly created canopy height raster layers. The tables generated with this tool determined the maximum, minimum, and mean canopy heights of the coniferous and deciduous category trees within the seven buffer areas. The six canopy metrics were then joined to the spatially overlapping buffer areas utilizing the same method that I used to extract data from the point, line, and polygon features.

Finally, I calculated the "density" of each non-canopy height feature by dividing the area, line, and point count variables by the area of their associated buffers. This division created a density value for these landscape features that was either a ratio of area/area (for areal features), length/area (for line features), or count/area (for point features). I then checked the seven tables of the seven buffer widths and 20 variables for calculation errors, missing data, and organizational issues and then exported the data as a .csv file for analysis (Table 2).

Variable	GIS File Type	Dimensions	Source	Model Alias
PREDATOR score	Point	0 – 1 value	PAWMAP stream surveys	macro_pred
Commercial land	Polygon	Area/Area	Metro RLIS database	commerical_area
Industrial land	Polygon	Area/Area	Metro RLIS database	industrial_area
Residential land	Polygon	Area/Area	City of Portland - BES	residential_area
Street length	Line	Length/Area	City of Portland DOT	street_length
Stream crossings	Point	Count/Area	City of Portland - BES	streamX
Impervious surface	Polygon	Area/Area	City of Portland - BES	impervious_area
Piped stream	Line	Length/area	City of Portland - BES	piped_stream
MS4 outfalls	Point	Count/Area	City of Portland - BES	MS4
Unsewered area	Polygon	Area/Area	City of Portland - BES	unsewered
Wetland area	Polygon	Area/Area	City of Portland - BES	wetland_area
Vegetation area	Polygon	Area/Area	City of Portland - BES	veg_area
Canopy area	Polygon	Area/Area	City of Portland - BES	canopy_area
Maximum canopy height	Raster	Average in buffer area	City of Portland - BES	canopy_height_max
Mean canopy height	Raster	Average in buffer area	City of Portland - BES	canopy_height_mean
Coniferous canopy area	Polygon	Area/Area	City of Portland - BES	conifer_area
Maximum coniferous canopy height	Raster	Max value	City of Portland - BES	canopy_height_conifer_max
Mean coniferous tree canopy height	Raster	Average in buffer area	City of Portland - BES	canopy_height_conifer_mean
Deciduous area	Polygon	Area/Area	City of Portland - BES	deciduous_area
Maximum deciduous canopy height	Raster	Max value	City of Portland - BES	canopy_height_decid_max
Mean deciduous canopy height	Raster	Average in buffer area	City of Portland - BES	canopy_height_decid_mean

Table 2: This table displays the metadata for the 20 landscape predictor variables and the PREDATOR response variable analyzed in this study. The GIS format, dimensionality, data source and R alias is listed for each variable.

Data Analysis and Modelling

I used the RStudio integrated development environment software package (Version 1.1.463, RStudio Team 2015) to process and analyze the PREDATOR data and the landscape variables from the seven buffer width datasets. Using RStudio (henceforth referred to as "R"), I first examined the data graphically and statistically to identify potential errors in data organization, outliers, and to get a general sense of distributions. Next, I created correlation matrices to identify variables with potential collinearity issues within the seven datasets that

could be problematic in the modeling process. These correlation matrices conveyed the strength and direction of correlation between the 20 landscape variables (Figure 6).
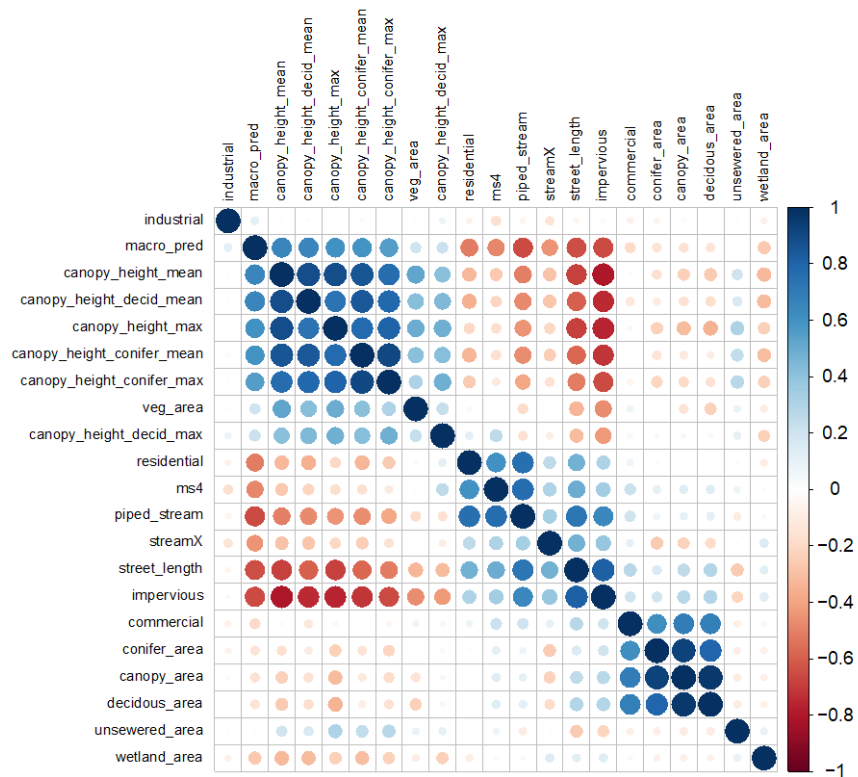

Figure 6: The correlation matrix for the complete 500 ft buffer data set displays positive correlation with shades of blue and negative correlations with shades of red. The size of the circle represents the strength of each correlation. This matrix shows that there are positive and negative correlations of varying strength amongst the 20 landscape variables.

I examined the correlation matrices and took note of trends and patterns within each array and across the seven buffer width datasets. I saw that there were routinely strong correlations (both positive and negative) between variables that were measuring similar aspects of the landscape. For example, variables pertaining to the land use, infrastructure, and other man-made or "unnatural" features of the landscape were strongly correlated. Variables that were metrics of the "natural" landscape features — vegetation, forested area, wetlands, and canopy heights — were also generally correlated with each other in all datasets. The correlations within both of these landscape feature groups were not just strong overall, but also varied in direction. This was an unsurprising finding because, for example, an increase in one zoning in an area leads to a decrease in another zoning area and wetlands generally don't occur in areas with large amounts of coniferous trees. Additionally, it was unsurprising to see that the canopy height metrics were strongly, positively correlated with each other because the mean canopy height values would be greater when the maximum canopy height values were higher.

Informed by these initial observations of the matrices, I then created subsets of the data to focus on the variables that had the highest absolute correlation values so that issues of collinearity could be addressed. I observed that that the canopy height metrics and the other

vegetation related variables were highly correlated at all buffer widths, but the correlation was especially pronounced in the 300 ft buffers (Figure 7).
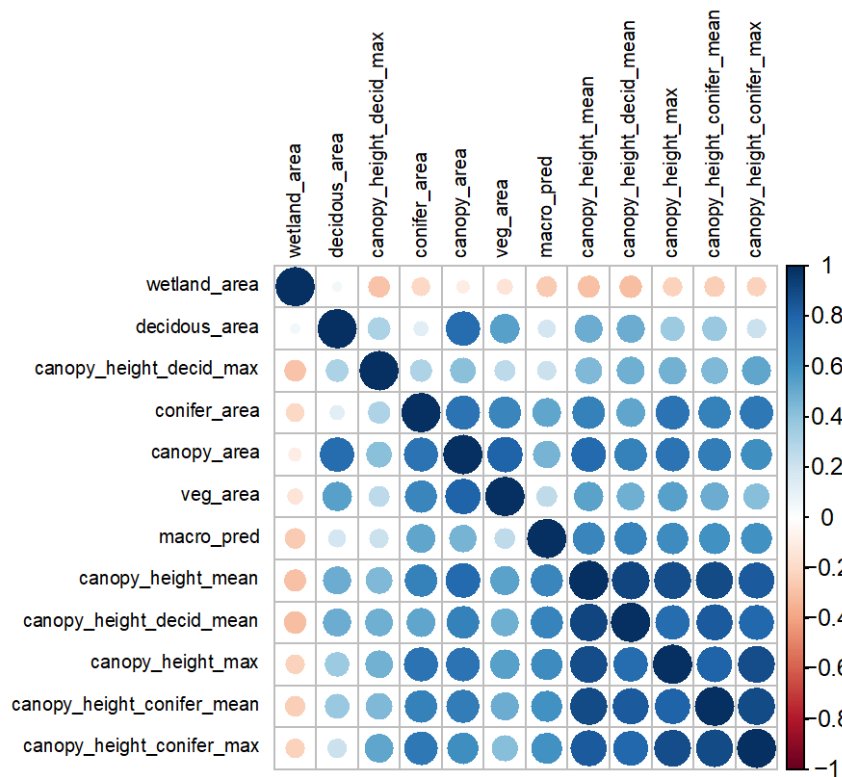

Figure 7: The canopy height and vegetation variables displayed positive correlations at all scales, but the strongest correlations were observed in the 300 ft buffer width dataset.

I also observed that there was a moderate level collinearity in the infrastructure and land use variables amongst all buffer width data sets, but again it was most pronounced in the 300 ft buffer dataset (Figure 8).
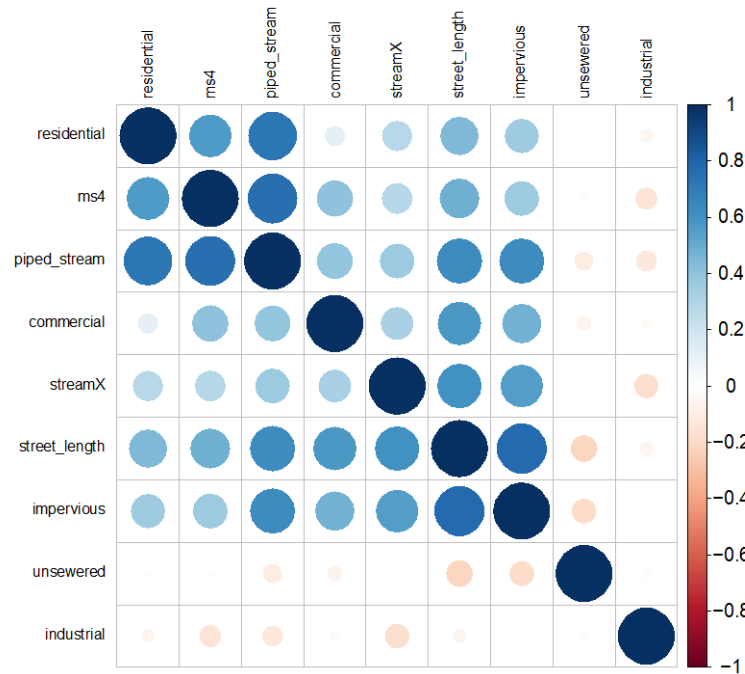
Figure 8: Most of the land use and built-environment variables were positively correlated in the 300 ft buffer area datasets. The exception was with the unsewered and industrial area features, which had weak, positive correlations with the other seven variables in the 300 ft buffer areas.

I discussed these persistent and predictable instances of collinearity in the datasets with project advisor Patrick Edwards, PhD and we decided that it would be advantageous to reduce the overall number of variables to be used in the modelling and analysis stage of this project. We decided that this reduction should be done to increase statistical power and to preserve degrees of freedom in the regression models. To do this, I considered the mechanisms of impact of landscape features, the correlation values observed, and the project's interest in analyzing a diverse set of predictive variables. I examined the datasets through the lens of these considerations and I finally choose to remove eight variables from further consideration. The remaining 12 variables that were to be analyzed in the seven buffer width data sets were:

- Canopy area
- Commercial area
- Impervious area
- Industrial area
- Mean canopy height
- MS4 outfalls
- Piped stream length
- Residential area
- Stream crossings
- Street length
- Vegetated area
- Wetland area

While there were still varying levels of collinearity observed amongst these 12 landscape features, these new subsets of variables in the seven datasets had fewer redundant landscape features and less feature pairs with covariance values greater than .05 (Figure 9).
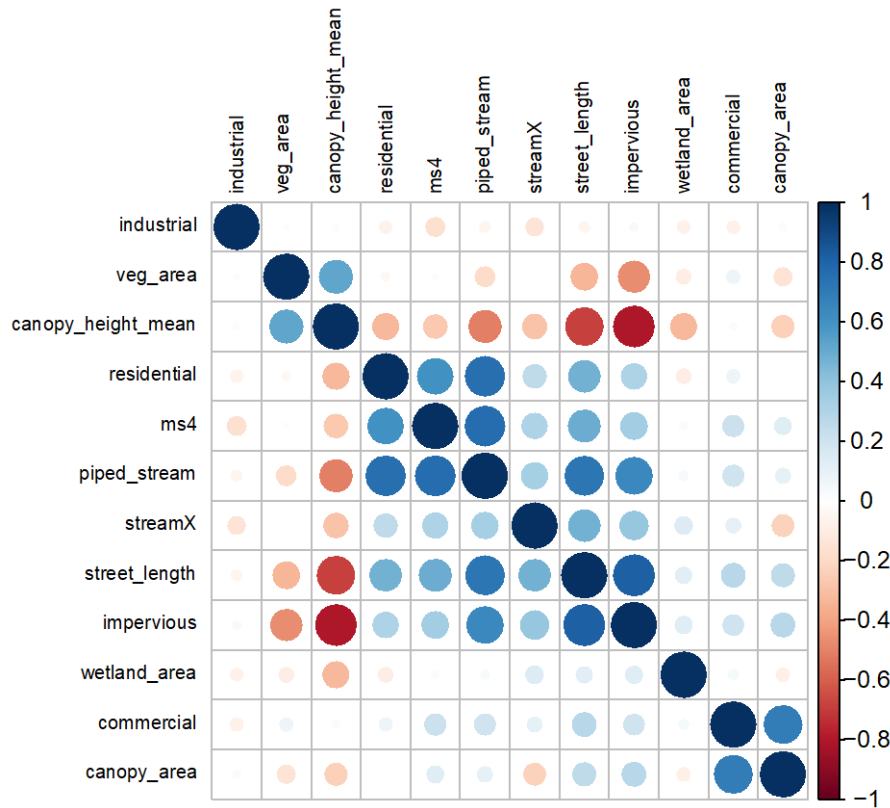

Figure 9: The correlation matrix for the reduced 500 ft dataset displays positive and negative correlations within the subset of 12 variables.

My final step in preparing and refining the data for analysis was to determine if the buffer dataset groups were substantially different from each other. If two buffer width datasets were not substantially different from each other, the analysis process would include redundant data and the extraneous inclusion would limit the predictive power of the models due to statistical inference issues arising from multiple comparison problems.

For example, if I used all seven buffer width datasets, a Bonferroni correction would have been necessary. I would have divided my target alpha value (.05) by the number of datasets (7). While this would have been a sound method to use, I found that the new, corrected alpha value of .007 would have made it all but impossible to create a statistically sound model with the data used in this study. For this reason, I decided that I would need to focus my analysis on fewer than seven buffer widths datasets.

Though I could have used a one-way ANOVA test to determine if there was a statistical difference between variables of the same type in the different datasets, I could not identify a method to determine if an entire dataset was statistically different from another. Lacking an established method to make this determination, my project advisors suggested that I should

examine the means, standard errors, and distributions of the variables across the datasets to identify potential redundancy.

Some variables had means that did not vary greatly across the buffer width range (Figure 10) while others were quite skewed (Figure 11). Other variables had means that did not display a clear trend across the buffer-width range (Figure 12). I examined these graphs for all 12 variables and took note of the trends in means and in the overlap in standard errors ranges.
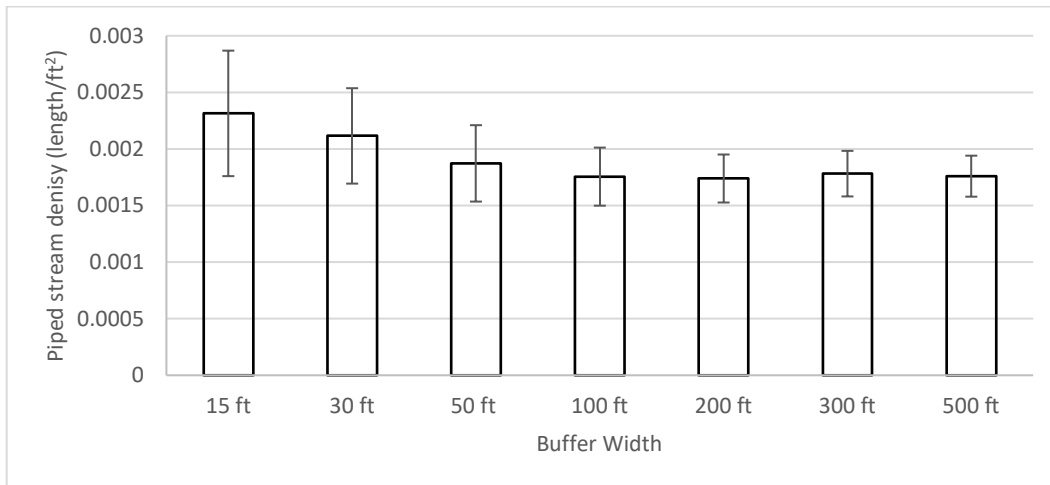

Figure 10: The concentration of piped stream length was relatively consistent across the entire range of buffer widths. The standard error ranges for all seven piped stream length density values overlapped.
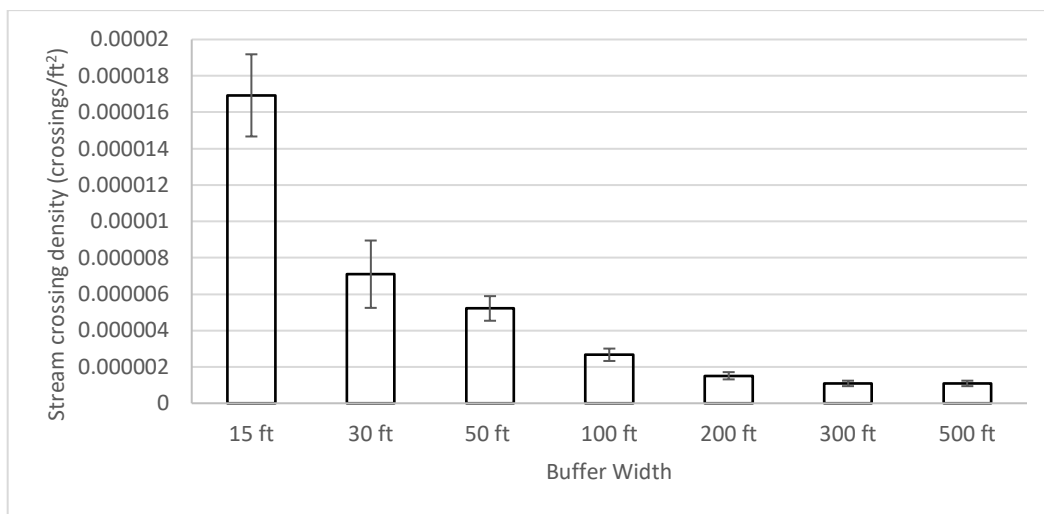

Figure 11: This bar graph shows that stream crossings were highly concentrated in the 15 ft buffer areas. Additionally, it shows that the crossing density values were strongly skewed right and there was very little overlap in the standard error values around the means.
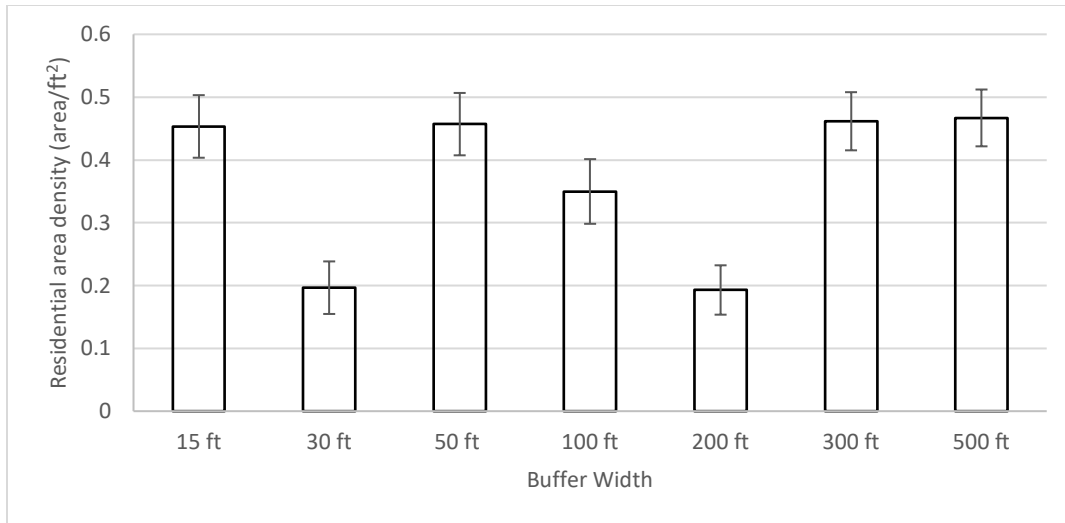
Figure 12: The density of residential area varied in a seemingly random way across the range of buffer widths.

From my examination of the graphed means and distributions I found that the 15, 30, 100, and 300 ft groups were generally not greatly dissimilar from the means of the variables in the 50, 200, and 500 ft datasets. Conversely, I found that the standard error ranges of the 12 variables in the 50, 200, and 500 ft datasets rarely overlapped.

To gain further insight into the differences between variables in the 50, 200, and 500 ft data, I examined the histograms for all the variables across the three buffer width datasets. I took note of the centers, dispersions and shapes displayed in distributions of these histograms. While some of the distributions in the three datasets were similar in shape and range (Figure 13), others displayed great differences (Figure 14). Other variables in the three datasets had distributions that were similar in some buffer widths, but very different in others (Figure 15).
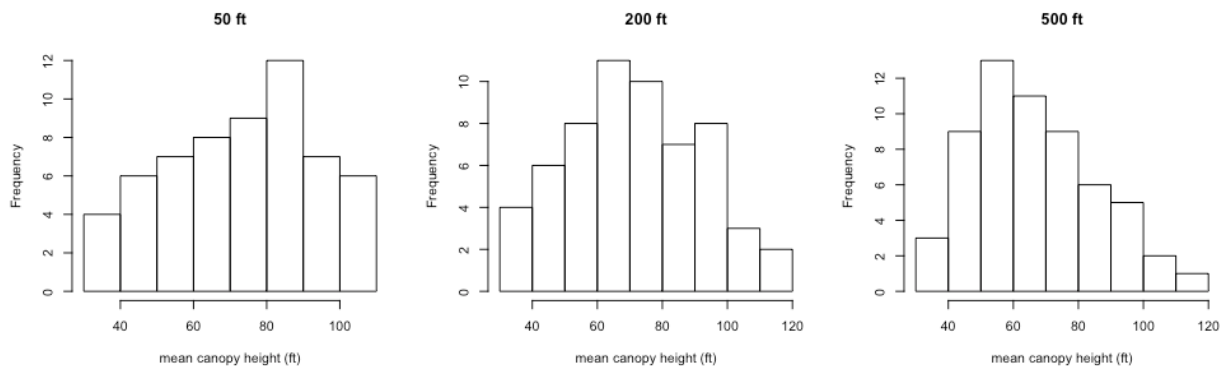

Figure 13: The mean canopy height distributions were generally similar in the 50 ft, 200 ft and 500 ft buffer width datasets.
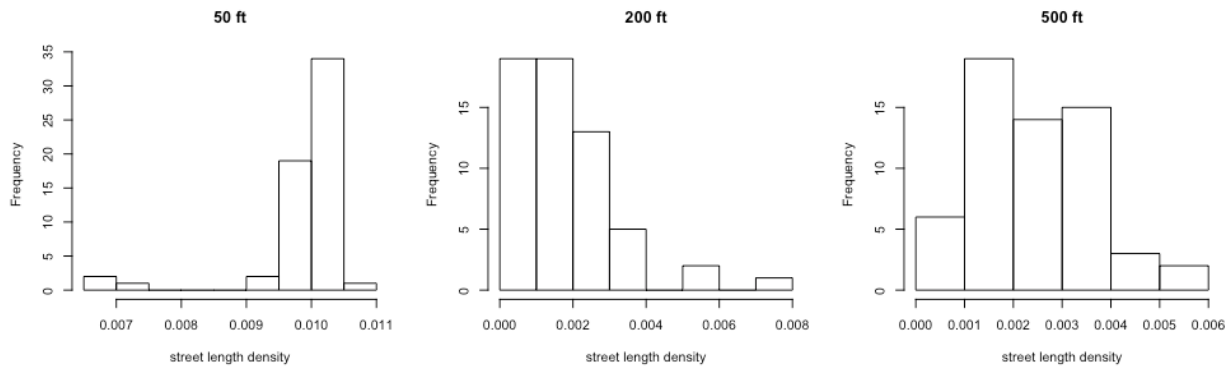
29

Figure 14: The distribution of street length density values varied greatly across the 50 ft, 200 ft and 500 ft datasets.
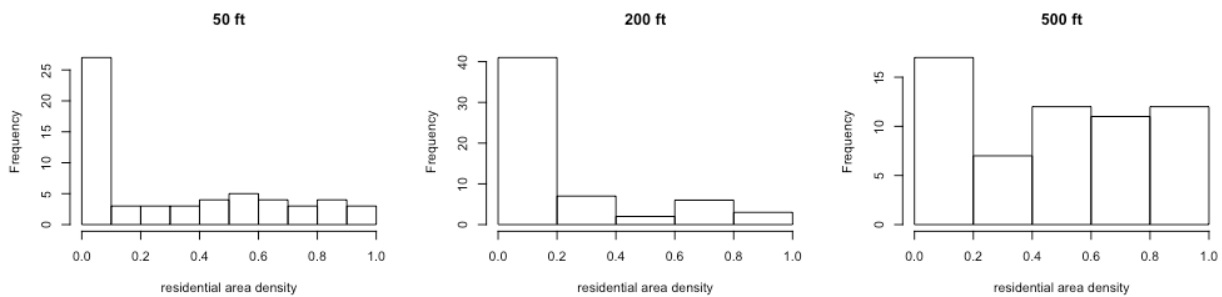

Figure 15: The distribution of the residential area density values was similar in the 50 ft and 200 ft buffer widths.

I then performed one-way ANOVA tests on the 12 variables in these three datasets. From these statistical tests, I determined that a preponderance of variables in the 200 ft buffer width data set were not significantly different ($p>.05$) from the same variables in the 50 ft and 500 ft groups. For this reason, I made a decision to remove the 15, 30, 100, 200, and 300 ft buffer width datasets from further analysis. While this process that led me to choose to use only the 50 ft and 500 ft datasets going forward was somewhat subjective, it was the approach that my advisors and I agreed upon — an approach that would allow for me to make statistically powerful predictive models that addressed the goals of this study.

Variable Selection Using Random Forest Analysis

To create multiple linear regression models for the 50 ft and 500 ft datasets, I needed to determine which variables would be the best predictors to use. To do this, I used a random forest regression analysis algorithm to identify the strongest predictive variables in the 50 ft and 500 ft datasets. Random forest analysis is a relatively modern machine learning algorithm that has been shown to be effective at analyzing large ecological datasets (Breiman 2001). Random forest analysis algorithms are useful and appealing because they are robust to many of the assumptions that must be satisfied in linear regression (e.g., normality of residuals) and because they can deal with missing data and nonlinearity issues without using transformations. Random forest models are somewhat of a "black-box" because it is hard to concisely report the structure of the trees or

30

which variables were used at which nodes. Because of this, random forest model results are generally presented as the proportion of variance explained and the out-of-bag error rate values (Breiman 2001).

A full explanation of random forest analysis techniques can be found in Breiman (2001) and De'Ath (2007), but a brief explanation of the process is provided here. Using a dataset with multiple predictors and a single response variable as the input in the algorithm, a subset of predictive variables is randomly selected to generate a large number (500+) of decision trees (Prasad et al. 2006). In these trees, the number of explanatory variables considered for further tree splitting at each node is set by the user, but it is generally equal to the square-root of the total number of variables under consideration (Breiman 2001). With the maximum number of explanatory variables considered for splits at each node set, a large number of decision trees are created wherein a random group of explanatory variables is used at each split. For this reason, each decision tree generated will be different, but will consist of a subset of variables drawn from the initial complete set. By running this process many times and aggregating the results, an optimized decision tree is formed. Subsets from the complete data set can then be used to calculate the effectiveness of the classification ability of the model (Breiman 2001).

As I stated before, random forest analysis is somewhat of a "black box" process. The analysis has many outputs, but the two of concern for this project are the out-of-bag error rate and the variable importance value. The out-of-bag error rate lets the user know if the model is overfitting and if it is actually effective at classifying the data that the model was based upon. The out-of-bag error rate is a percentage that conveys how often a subset of the initial data is misclassified when run through the model. This dimensionless variable importance value (sometimes referred to as the "mean decrease in accuracy") quantifies each variables contribution to a model's overall predictive power. This allows a user to rank the importance of variables and then to select the most important predictors for further analysis and modelling (Breiman 2001).

For this project, I utilized the random forest algorithm script in R to identify the relative importance of the 12 variables under consideration. The number of trees generated ranged from 570 to 10,000 and the number of variables used at node splits ranged from 3 to 5. The final configuration, parameters, explanatory power, and diagnostic plots for all models were recorded during this process. To interpret the outputs of the random forest analysis, I created variable importance plots in R with the "varImpPlot" function. The importance is expressed as the mean decrease in Gini value which is the relative importance that individual variables contribute to the random forest models (Figure 16, Figure 17). A greater Gini value indicates that a variable contributes a greater amount of predictive power to the Random forest model (Breiman 2001).
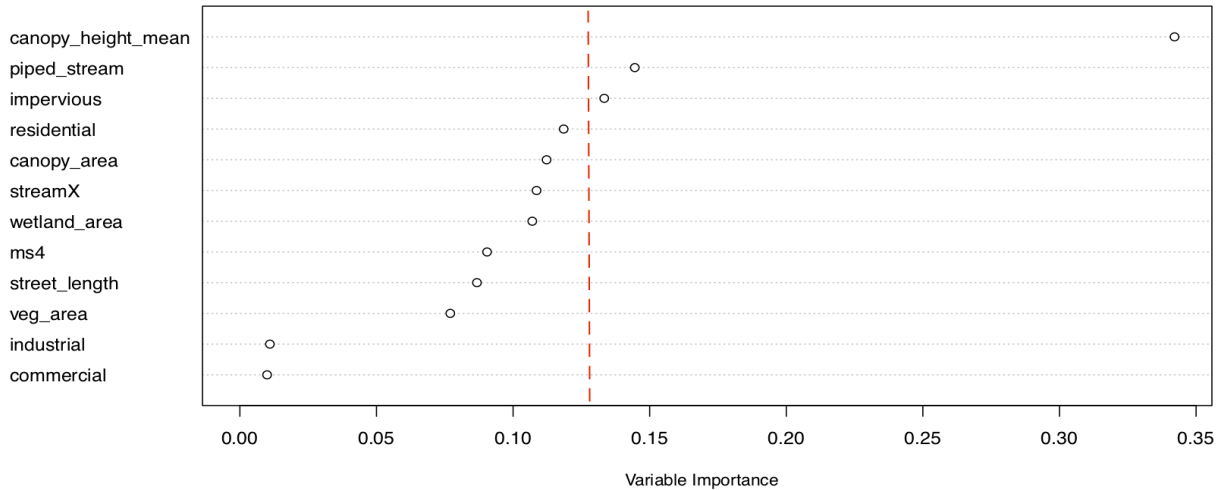
Figure 16: The variable importance plot for the 50 ft buffer random forest model indicates that the mean canopy height contributes the most to the predictive power of the model. Conversely, the commercial land density is shown to contribute the least to the predictive power of the same random forest model. The red line represents the cutoff threshold that I decided on for choosing variables from the 50 ft buffer areas going forward. Only the variables to the right of the line were used in further analysis.

Figure 17: The variable importance plot for the 500 ft buffer Random forest model indicates that the mean canopy height contributes again contributes the most in terms of predictive power, while the commercial area density contributes the least to the predictive power to the model. The red line represents the cutoff threshold that I decided on for choosing variables from the 500 ft buffer areas going forward. Only the variables to the right of the line were used in further analysis.
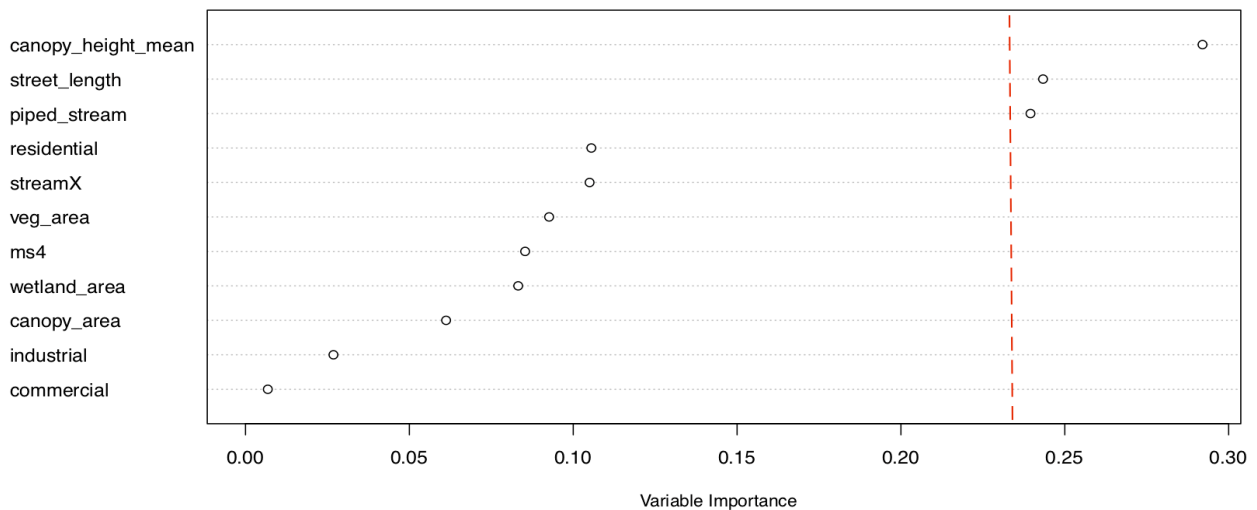
The variable importance plots helped me visually identify predictors that could be excluded from further analysis. In my inspection of the plots for the 50 and 500 ft random forest models, I saw that there was a wide range of importance values amongst the 12 variables in the two random forest models. (Figure 16, Figure 17). While these values varied greatly, there was a pattern of grouping or clustering amongst the variables in both plots. Examining the two variable importance plots, I took note in the breaks between these clusters and then selected the

variables that stood out as having especially high contributions to the predictive power of the models. Informed by these variable importance plots, I chose to use the mean canopy height, piped stream length, and impervious area variables for the 50 ft buffer data analysis, and the mean canopy height, piped stream length, and street length for the 500 ft buffer data analysis

Regression Analysis and Modelling

I used a script in R to create multiple linear regression models for the two buffer width datasets using the variables identified in the random forest analysis as being of greatest influence. These two models contained 3 variables that had generally Gaussian distributions. Next, I used the Akaike Information Criterion (AIC) script in R to test the goodness of fit of the two models (Akaike 1987). The AIC created permutations of linear regression models that used the variables identified in the random forest analysis to optimize the balance between the goodness of fit and the simplicity of the model. This process helped me create refined models that were less complex, more robust, and had higher correlation coefficients (aka R^2).

Finally, I used a normalization function in R to scale the two models so that the coefficient values could be compared and interpreted along a common scale. This process allowed me to produce two statistically sound models that expressed the PREDATOR scores as a function of landscape characteristics that I could use to understand how the urban landscape in riparian areas effect stream health at different spatial scales.

Results

The two final linear regression models that I created contained only the variables of piped stream length, mean canopy height, and impervious area. Each model contained a single natural landscape variable and 1-2 anthropogenic landscape variables. The models had R^2 values that were within +/- 10% of the overall R^2 mean of .529(Table 3). The maximum R^2 value was observed in the 500 ft buffer model (R^2 = .556) and the minimum was observed in the 50 ft model (R^2 = .502). While these R^2 values may seem somewhat low, it is similar to values seen in other studies involving the study of ecological variables across a broad landscape (Jackson et al.; 2015; Solins et al. 2018).

Buffer Width (ft)	Model formula	p-value	Correlation coefficient (adjusted R^2)	Degrees of freedom	Residual standard error
50	macro_pred = -0.296* piped_stream - 0.2578*impervious + 0.452* canopy_height_mean	4.977e-09	0.502	F(20.45,3)=55 F-statistic: 20.45 on 3 and 55 DF	0.706 on 55 degrees of freedom
500	macro_pred = -0.443*piped_stream + 0.428*canopy_height_mean	4.847e-11	0.556	F(37.39.,2)=56 F-statistic: 37.39 on 2 and 56 DF	0.665 on 56 degrees of freedom

Table 3: This table displays the formulas for the two scaled regression models and their significance levels and goodness of fit diagnostic values.

In both models, I found that the piped stream length variable was negatively correlated with the PREDATOR scores while the mean canopy height was positively correlated with the PREDATOR scores. In the 50 ft model, impervious area was negatively correlated with the PREDATOR scores. I also noticed that the variables common to both buffer width models had correlations going in the same directions. Both regression models had p-values less than .05 which indicated a significant relationship between the variables and PREDATOR value (Table 3). Additionally, all individual variables within the models had significance values well below the .05 level threshold. This indicated that all variables used made significant contributions to the predictive power of the models.

Discussion

The final models showed that the piped stream length, the mean canopy height, and impervious area have varying levels of impact on stream health in the 50 ft and 500 ft buffer areas around the 59 sampled stream reaches. Within the 50 ft buffers areas, the impact of piped streams was about 87% of that of impervious area (Table 3). The mean canopy height in the 50 ft models was positively correlated with stream health and its beta coefficient of .452 indicated that the variable was of nearly double the influence of each of the unnatural landscape features (Table 3) This finding shows that the man-made features in the 50 ft buffer areas were drivers of stream health degradation, but that a substantial amount of their influence could offset by a greater mean canopy height. In the 500 ft model, I found that the coefficients for the piped stream length and mean canopy height variables were nearly exact opposites (-.443 vs .428) This meant that the negative influence of piped stream length on stream health was essentially equal to the positive influence of the mean canopy height. This was an interesting finding because it showed that stream health in the 500 ft buffer areas was being driven almost equally by a natural and an unnatural landscape feature. It was also reassuring to see that directions of correlation were consistent in the two models. Interpretation of the models would have been more difficult

if, for example, piped stream length was positively correlated with PREDATOR scores in the 50 ft model and negatively correlated in the 500 ft model.

Because the size of the buffers increased linearly with widths, the entire 50 ft buffer width data set was included as 1/10th of the 500 ft buffer width data set. I think that this study benefitted from having this proportion of overlap being relatively small. For example, if I created models for just the 300 ft and 500 ft datasets, the concentration of landscape variables would likely be much more similar due to the similarities in buffer area sizes. Even though five of the seven initially selected buffer width datasets were not ultimately analyzed, the two widths that I did analyze covered very different parts of the riparian area. For this reason, I think that my reduction in datasets from seven to two did not ultimately prevent me from addressing the project's goal of understanding how a stream's health is influenced by the urban landscape at a range of spatial scales.

Mean Canopy Height

In both models, I found that mean canopy height was positively correlated with PREDATOR scores. In the 500 ft model, I observed that the beta coefficient was nearly equal to the mean canopy height beta coefficient in the 50 ft model (Table 3). This was an interesting finding because it suggests that the importance of riparian tree height does not vary with distance from the stream channel. Instead, it indicates that the shade providing function that trees provide near stream channels may not be as important as the overall presence of a canopy with a greater mean height. While I cannot identify the specific mechanism of impact that the mean canopy height had on stream health in this study area, my findings were consistent with the positive correlation between canopy height and stream health in other studies (Hawkins et al. 1983, Sridhar et al. 2004).

Seeing that there was a positive correlation between tree height and stream health in both the 50 ft and 500 ft buffer areas, it would be logical to think the taller trees would be the best predictors of stream health due to their ability to provide the greater amounts of shade to a stream. While the tall trees were surely providing more shade to the stream than the shorter trees, I think that the *mean* and not the *maximum* canopy height was found to be the best predictor because the maximum values were generally outliers that did not represent the overall canopy composition well. In a given buffer area, the maximum canopy height value could have theoretically been just one single tall tree amongst many shorter trees. For this reason, I think the mean canopy height was an effective predictive variable in the final models because it best conveyed the general canopy conditions in buffer areas.

The mean canopy height might have also been the best predictor from the 20 considered because it was the best proxy for other vegetation or natural (i.e. non man-made) features. For example, an area with a high average canopy height could indicate that the vegetation in a buffer area had not been routinely disturbed or altered. In one of these less disturbed areas, it would be unsurprising to find more overall vegetation, wetlands, and less invasive species. Conversely, the mean canopy height might have been identified as a strong predictor in the modelling process because it conveyed information about a lack of urban development and overall human impact in the riparian buffer areas.

To see if the data might support my hypothesis, I revisited both the full 50 ft and 500 ft buffer data sets and found that the mean canopy height values in both buffer width datasets were positively correlated with a majority of the other nine natural landscape features. (Figure 18, Figure 19).
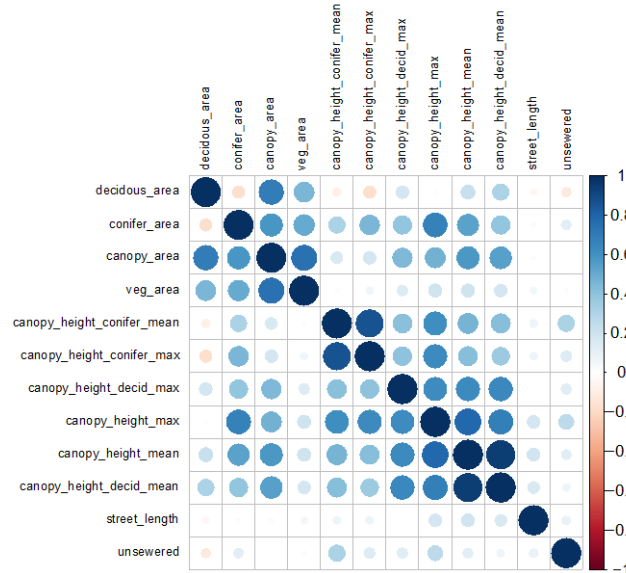

Figure 18: The mean canopy height was positively correlated with 11 other features in the 50 ft buffer width dataset. Out of these 11 features, 9 of them were of the generally natural or vegetation related.
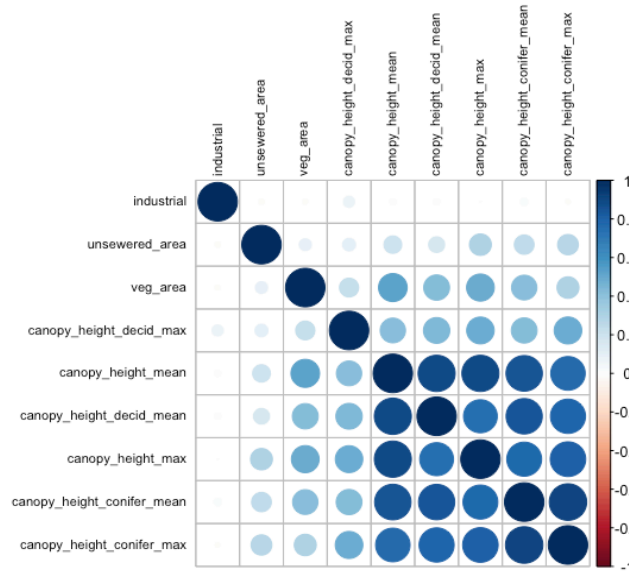
Figure 19: In the complete 500 ft buffer width dataset, the mean canopy height was positively correlated with eight other landscape variables. Out of these eight, six were of the generally natural or vegetation related.

Even though many of the natural (and some unnatural) variables had positive correlations with the PREDATOR scores, the correlations I observed in the above figures suggest that the mean canopy height was a strong predictive proxy because it was strongly correlated with the other

natural features. I suspect that collinear relationships between mean canopy height and other natural features were identified by the random forest and AIC algorithms and for that reason, only the mean canopy height variable was used in the two final regression models.

Piped Stream Length

In the two models I created, concentration of piped streams was negatively correlated with PREDATOR scores in both the 50 ft and 500 ft buffer areas. In the 500 ft model, the negative influence of piped stream length on stream health was about 50% greater than the influence observed in the 50 ft model (Table 3). This was an interesting finding because subsurface piping infrastructure is only found on streams and an increase in buffer width area did not generally lead to a proportional increase in piped stream density. The practice of piping streams in urban areas is known to be detrimental to stream health, so my findings here were not too unexpected. The negative correlations I observed in both buffer width models suggests that stream health could be improved by removing the piping infrastructure that isolate streams from their immediate surroundings.

While my finding of a negative correlation between piped stream length and stream health reflected the findings of others (Hatt et al. 2004; Neal et al. 2016), I wondered why this variable was identified by the random forest and AIC algorithms as being a strong predictor. In the initial buffer width datasets, there were many other man-made or unnatural landscape features and land-use area variables that had negative correlation with PREDATOR scores. Some of these variables, such as residential area and street length, had similarly strong negative correlations with PREDATOR scores.

I thought that the piped stream length might have been included in the final models because it was a proxy for other unnatural features that were negatively correlated with PREDATOR scores, as was the case with canopy height. I created a correlation matrix for the features positively correlated with piped stream length in the 50 ft buffer areas and I found that correlations were not especially strong amongst these variables. From this matrix, I observed that the strongest correlations were between residential areas and outfalls, and that the rest of the positive correlations were barely greater than zero (Figure 20). Piped streams in the 500 ft buffer areas were strongly correlated with many other landscape features, but nearly half of these were a canopy height or a vegetated area type variable (Figure 21).

Figure 20: In the 50 ft buffer areas, the concentration of piped streams was correlated with only two features. The remaining six positive correlations between piped stream length and other landscape features were barely greater than zero.
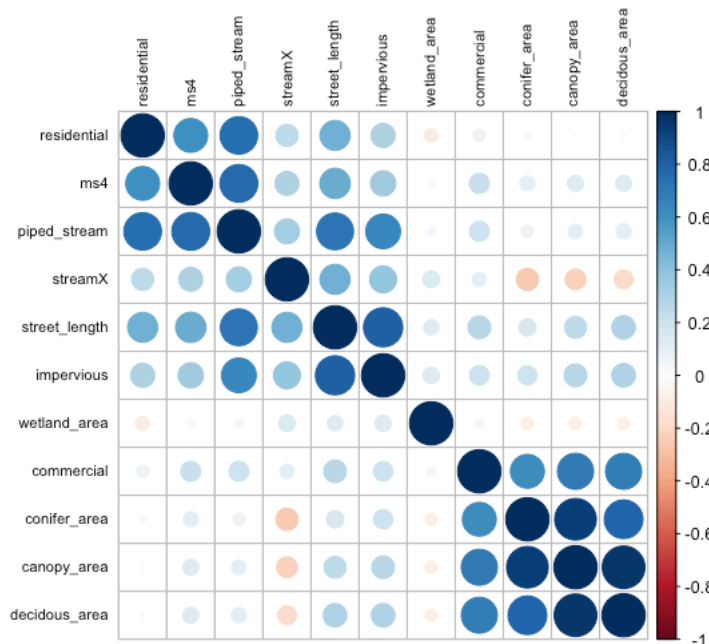
Figure 21: The piped stream length variable was positively correlated with 10 other features in the 500 ft buffer areas. Out of these 10 positive correlations, only six of were with other unnatural-type landscape variables.

The lack of correlation between piped stream length and the other 19 variables initially considered indicated to me that piped stream length was not serving as a proxy for other features in the regression models. Instead, it indicated that piped stream length is a strong predictor of stream health and that it is not consistently associated with either natural or unnatural features in the riparian buffer areas. For this reason, I think it would be an interesting landscape feature to analyze in detail in future studies.

Impervious Area

In the 50 ft foot buffer areas, I found that the concentration of impervious area was negatively correlated with PREDATOR scores of the PAWMAP stream reaches. Again, I was not surprised by this finding, as it is a connection that has been well documented in the literature (Roy et al. 2005; Ladson et al. 2006). In the 50 ft buffer area model, impervious area had similar level of negative influence on stream health as piped stream length. 55 out of the 59 total PAWMAP stations had some degree of impervious area within 50 ft of their streams. For these reasons, the prevalence and strong influence of impervious surfaces in narrow buffer areas near streams should be of interest for future researchers and environmental managers alike.

I expected impervious area density to be a strong predictive variable in this study because it would be an overall metric for urban development based on my assumption that the majority of lands zoned residential, commercial, or industrial would also be impermeable. Additionally, it seemed reasonable to assume that the density of streets in a buffer area would also be conveyed through the overall impervious area variable. I examined the positive correlation between impervious area in the 50 ft buffer areas and found that the variable was correlated with two out of the three zoning features and not at all with street length (Figure 22).
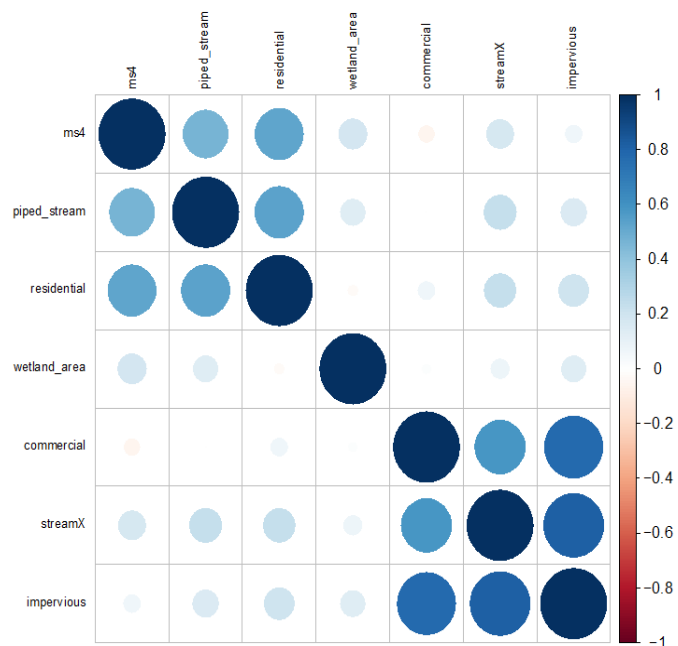
Figure 22: The concentration of impervious area was positively correlated with six other landscape variables in the 50 ft buffer areas. All but one of these six variables were unnatural, man-made features.

As was the case with piped stream length, this finding indicates that impervious area is a strong, somewhat independent predictor of stream health in areas immediate to Portland's urban streams. The lack of inclusion of the impervious area variable in the 500 ft buffer models was surprising, however. Even though the concentration of impervious area was greater in the 500 ft buffer areas, it was ultimately not found to be useful to include in the model. I also thought

it was interesting that no other land use or land cover variable was used in the 500 ft model. This is because the mean canopy height was a strong proxy and pipe stream length was not a proxy, but a strong independent predictive variable.

Reflections of Study Design and Analysis Methods

Even though I thought it was important to understand why these three variables were included in the final models, I also recognized that it would be impossible to fully understand their relationship to other landscape features and their individual functions in the urban environment. Though I found piped stream length, mean canopy height, and impervious area to be strong predictors of stream health, it should also be considered that they are proxies for landscape features that were not considered in this study. For example, the three variables could be conveying information related to historic timber harvest, invasive vegetation, hydrologic alterations to the watershed, pollutants, and socioeconomic factors.

A future study could focus on the variables used in the two regression models by examining a new suite of possibly related landscape features. A closer look, informed by research and speculation/exploration could look at the landscape variables mentioned above, and many more. Additionally, a more nuanced study could also be done at an individual PAWMAP site to identify the specific mechanism of impact the variables of piped stream length, mean canopy height, and impervious area have on stream health.

Statistical Issues

This study was designed to collect a large dataset from a large study area and with systematic and replicable methods. While these methods I used were effective at aggregating data from multiple sources to address the objectives of the study, I thought of multiple ways that future researchers could conduct similar studies that would better account for the hydrologic connectivity, spatial relationships, and statistical problems associated with analyzing landscape scale data.

While the area of the buffers increased proportionally with the increase of buffer width, the occurrence/density of landscape features did not have such a linear relationship. The nonlinear relationship between features calculated as a density (all features but the canopy height values) was expected and it is, broadly, what I analyzed in this study. This was a logical and expected finding – it would be very odd if the variables considered in this study were uniformly distributed across the landscape. I did find that there were variables that were not well distributed across the entire range of buffer widths. In the 15 ft buffer areas, only 7 out of the total 59 buffers had occurrences of industrially zoned land. Conversely, the densities of piped stream and MS4 outfalls was the highest in the narrower buffer widths. There were so few occurrences of piped stream and MS4 outfalls in the 300 ft and 500 ft range width areas that mean density values were barely above zero. Even though the data on these highly skewed, zero-inflated variables was properly collected, I think that a future study could address these types of distributions in the early stage of the analysis process by using ordinary least squares, negative binomial, or zero-inflated Poisson regression models. All three of these methods are known to

be effective at analyzing datasets that contain a number of "zero" values. Because some of the predictors had a nonlinear relationship with PREDATOR scores, polynomial regression techniques could also be used to model these variables. If polynomial regression techniques were used, the variables with a nonlinear relationship to stream health (e.g. piped stream length, industrial area, MS4 count) could be better represented and utilized more effectively to create stronger buffer width models.

Another (though more subjective) approach would be to establish a user-defined percentage of observations with a "zero" value that would be tolerated in a given variable. If that threshold was exceeded, the variable would be removed from further consideration/analysis. In this study, I did not use any of these methods and I don't think the zero-inflated variables were ultimately a problem in my statistical analyses. I think this was the case because the variables that were selected for my models in the 50 ft and 500 ft buffers areas did not have a great number of "zero" values. I suspect that they were not included because these variables were identified in the random forest analysis as having low predictive power.

Buffer Area Over-Reach

In examining the initial datasets, I noticed that there were variables (such as piped stream length, MS4, and stream crossings) that had concentrations that increased in the widest buffer areas. While this would be a logical finding for land-cover type variables, it did not make sense for those three variables that only occurred on or very near streams. I examined the spatial distribution of many variables and came to the conclusion that the buffer area polygons were over-collecting data for some of the PAWMAP sites.

Because the buffers that I used in this study were generated to cover areas perpendicular to streams without consideration of watershed topography, features from the landscape that were not hydrologically connected to PAWMAP sites were often included in the datasets. For example, some of the 500 ft buffer polygons captured not just the piped stream in the riparian stream network associated with the PAWMAP station, but also piped stream segments from a neighboring watershed. The unexpected increase in piped stream density in the greater width datasets can be seen on a simple plot of the concentration of a feature across the seven buffer widths (Figure 23).
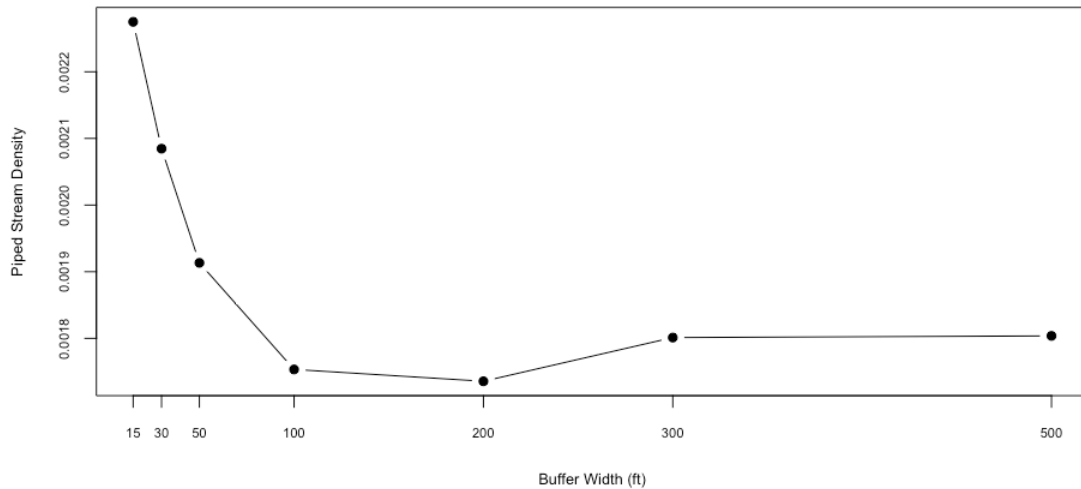
Figure 23: In the areas closer to streams, the piped stream density decreased as the buffer widths increased, but then increased in the wider buffer ranges. This indicated that these wider buffer areas were likely collecting features from neighboring watersheds.

Instances of buffers collecting data from adjacent watershed occurred mostly in the 300 ft and 500 ft datasets and only with the three variables that occur only on, or adjacent to, a stream. These cases of over collection were not something that this study was designed to avoid. The related watershed-scale study being conducted at the BES will likely mitigate this issue by focusing on hydrologic connectivity instead of the riparian buffer areas as a unit of analysis.

I did think of a way that the riparian buffer area approach that I used in this study could be altered to deal with this issue. I examined the spatial distribution of all of the landscape features. I noticed that some features were in buffer areas below PAWMAP stations, while others were in a buffer area, but not in the contributing watershed (Figure 24).
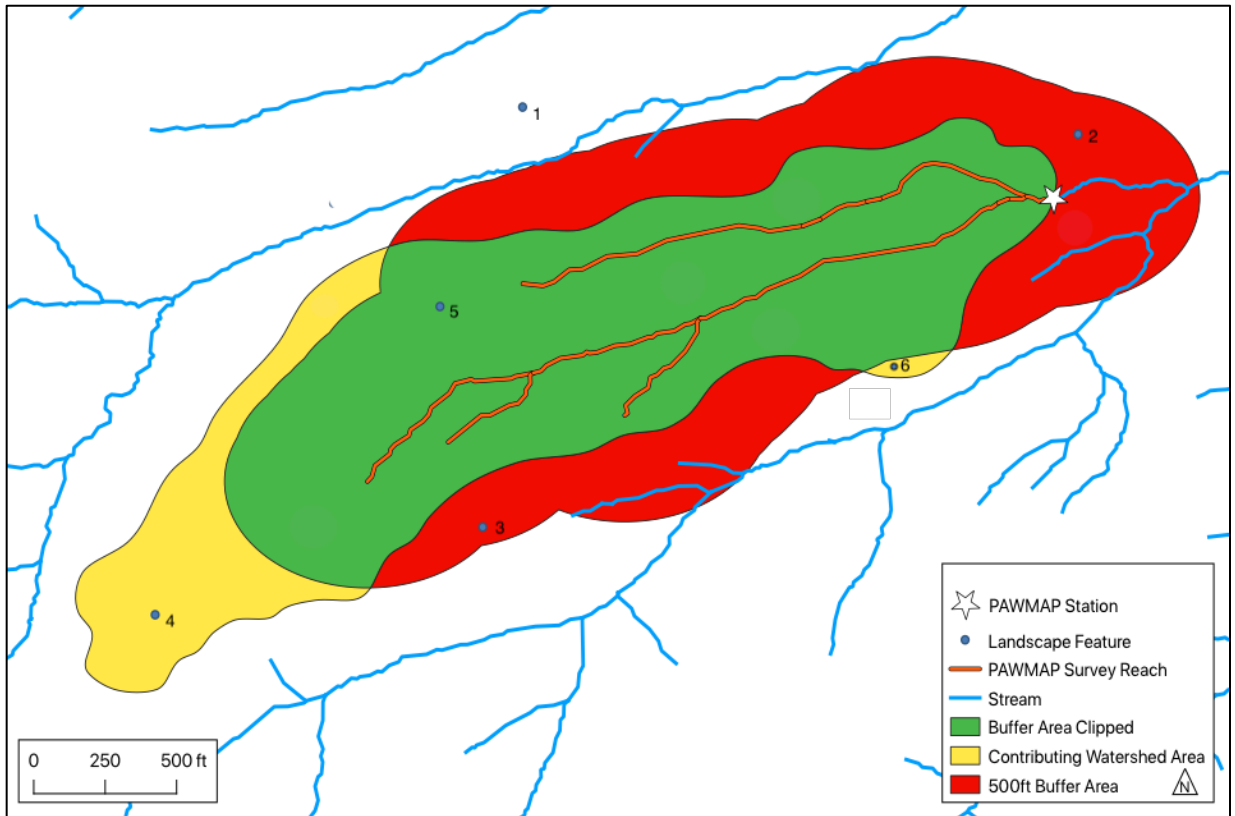
Figure 24: The buffers around stream reach can encompass features from areas outside of the watershed of a PAWMAP station. Additionally, features located in a station's watershed can be excluded from a study because they are not within the reach's buffer area. In this figure, the 500 ft buffer area around the reach upstream of station #0617 covers parts of the landscape to the north, south and west that are not in the station's watershed. Because feature #1 was not in the buffer area or the contributing watershed area, it would not be counted in this study. Feature #3 is in the buffer area, but not in the watershed. Feature #2 is inside the rounded end of the buffer that is at a lower elevation than that of the PAWMAP station. These two points would have been counted in this study. Features #4 and #6 were hydrologically connected to the PAWMAP station but were outside the 500 ft buffer area around the stream reach and therefore would not have been counted in this study. Feature #5 was located in the watershed and the 500 ft buffer area and would have been counted in this study. *Note: I manually created the theoretical landscape feature points and watershed area in this figure to convey the problems and gaps in this project's spatial data collection methods in a clear manner.*

One way to address this issue of data over-collection would be to trim or clip the buffer polygons by the extent of the watershed that contributed flow to each PAWMAP station. If the watershed area was larger, the buffer area would not change, but it would change if the watershed boundaries overlapped with the riparian buffer areas under consideration. The watershed areas would need to be created in ArcGIS by using a DEM to create hydrologic network models and watershed areas that used the PAWMAP point stations as an outlet (or "pour") point.

In clipping the buffer area polygons with a PAWMAP station's contributing area, only features that were hydrologically connected to the sampled stream reach would be enumerated in a dataset. Though this would cause the area of the buffer to not be linearly related to its width, the landscape variables could be used in the same way as I did in this study because their values would still be calculated as the quantity of the feature per unit area.

43

With the same DEM used to create PAWMAP watersheds described above, another analysis method could be used to analyze the influence of landscape features on stream health. The length of the overland flow path across a DEM between a point of the landscape and a sampled stream reach could be derived for all points in a study area. This new hydrologic proximity raster layer (generated with inverse distance weighting or Kriging interpolation methods) would consider the topography of watersheds and the proximity of landscape features in relation to sampled stream reaches. The derived layer could then be used to scale or weigh the landscape feature values based upon their proximity to the sampled reach. Similar methods of weighing the influence of landscape features based upon proximity to streams have been described in works by Ver Hoef et al. (2006) and Peterson et al. (2007). Though processes like these would involve more advanced data processing of datasets in GIS, the resulting product of the analysis could address the goals of this study in a way that would not depend on discrete, user defined riparian area buffer widths.

While more advanced statistical analysis and spatial modelling tools could be used to examine stream health (in the context of the urban environment), I feel that the methods I used were very effective at analyzing the landscape data. The methods I used in this study helped me determine that piped streams had negative impact on stream health in the buffer areas within 50 ft and 500 ft around the sampled streams, while the mean canopy height in these same areas had a positive impact on the stream health. Additionally, the amount of impervious area in the 50 ft buffer areas was found to have a negative influence on stream health.

Conclusion

In this study, I used classical regression models that were informed by modern machine learning algorithms to create models that described the health of Portland's streams in the context of the urban riparian areas that they flow through. The models I created from large, disparate data sources showed that man-made landscape alterations and facets of the natural environment act as opposing forces that both strongly drive the in-stream health. Even though my analysis explained roughly 50% of the variance in stream health, the models still provided insight into the connections between streams and the urban landscape. The moderate predictive power (represented by R^2 values) of the models also implies there are other drivers of stream health in the riparian environment that were not examined here that could be identified in future studies.

Though I suggested other methods for studying the influence of riparian areas on urban streams, the moderate predictive power of the two models could also indicate that stream health in the study area is not strongly driven by the riparian environment. Instead, it may be that stream health is being primarily driven by environmental conditions and urbanization features that would be better examined at the watershed or even landscape scale. If that is the case, the methods and analysis techniques I used in this study could still be used to address some of the same basic questions about the relationship between the stream health and the urban environment.

It is my hope that these findings on the relative import of natural and unnatural features in riparian areas may provide practical data that can be of use to COP environmental managers. With regards to the piping of streams, the practice should be stopped, and the many plans for

"daylighting" Portland streams should continue, and further, be given priority. While the mean canopy height in the riparian area cannot be changed with ease in an immediate time frame, efforts to preserve and promote healthy growth in the current tree stands in riparian areas should continue in conjunction with revegetation and invasive species management efforts. Even though impervious areas are ubiquitous in the urban environment, the extent of impervious surfaces created in new developments should be minimized, and stormwater runoff management features and other green spaces should be included when possible. Additionally, regulations could be established to prohibit the creation of new impervious surfaces within 50 ft of streams. Incentives and technical assistance could also be provided to property owners to encourage the removal of existing impervious surfaces.

This study shows how a disparate range of alterations in the Portland, Oregon landscape impact stream health at a range of scales, and that urban streams are responsive to conditions and activities occurring in their surrounding riparian environments. It also shows how the continually growing PAWMAP dataset can be used to investigate a specific question about Portland's watersheds and aquatics ecosystems. Though I made no absolute or universal conclusions about the nature of streams in the urban environment, I hope that the methods and findings described in this report will be of use to BES scientists, policymakers, and in the ongoing research being conducted by my fellow watershed hydrologists.

Works Cited

Akaike, H. (1987). Factor analysis and AIC. In Selected Papers of Hirotugu Akaike (pp. 371-386). Springer, New York, NY.

Alberti, M., Booth, D., Hill, K., Coburn, B., Avolio, C., Coe, S., and Spirandelli, D. (2007). The impact of urban patterns on aquatic ecosystems: An empirical analysis in Puget lowland sub-basins. Landscape and urban planning, 80(4), 345-361.

Allan, J. D. (2004). Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecological Systems., 35, 257-284.

Arnold Jr, C. L., and Gibbons, C. J. (1996). Impervious surface coverage: the emergence of a key environmental indicator. Journal of the American planning Association, 62(2), 243-258.

Baxter, C. V., Fausch, K. D., and Carl Saunders, W. (2005). Tangled webs: reciprocal flows of invertebrate prey link streams and riparian zones. Freshwater biology, 50(2), 201-220.

Benfield, E. F. (1997). Comparison of litterfall input to streams. Journal of the North American Benthological Society, 16(1), 104-108.

Bernhardt, E. S., and Palmer, M. A. (2007). Restoring streams in an urbanizing world. Freshwater Biology, 52(4), 738-751.

Beschta, Robert L. (1997). Riparian shade and stream temperature: an alternative perspective. Rangelands 19.2, 25-28.

Booth, D. B., and Henshaw, P. C. (2001). Rates of channel erosion in small urban streams. Water Science and application, 2, 17-38.

Booth, D. B., Karr, J. R., Schauman, S., Konrad, C. P., Morley, S. A., Larson, M. G., & Burges, S. J. (2004). Reviving urban streams: Land use, hydrology, biology, and human behavior 1. JAWRA Journal of the American Water Resources Association, 40(5), 1351-1364.

Brezonik, P. L., and 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.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Brown, L. R., Cuffney, T. F., Coles, J. F., Fitzpatrick, F., McMahon, G., Steuer, J., ... & May, J. T. (2009). Urban streams across the USA: lessons learned from studies in 9 metropolitan areas. Journal of the North American Benthological Society, 28(4), 1051-1069.

Carlson, J. Y., Andrus, C. W., & Froehlich, H. A. (1990). Woody debris, channel features, and macroinvertebrates of streams with logged and undisturbed riparian timber in northeastern Oregon, USA. Canadian Journal of Fisheries and Aquatic Sciences, 47(6), 1103-1111.

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., and Smith, V. H. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological applications, 8(3), 559-568.

Castro, J. M., and Beavers, A. (2016). Providing aquatic organism passage in vertically unstable streams. Water, 8(4), 133.

Chang, H. (2007a). Streamflow characteristics in urbanizing basins in the Portland metropolitan area, Oregon, USA. Hydrological Processes 21 (2): 211–22.

Chang, H. (2007b). Comparative streamflow characteristics in urbanizing basins in the Portland Metropolitan Area, Oregon, USA. Hydrological Processes: An International Journal, 21(2), 211-222.

Chapin, D. M., Beschta, R. L., & Shen, H. W. (2002). Relationships between flood frequencies and riparian plant communities in the upper Klamath Basin, Oregon: Riparian ecology and management in multi-land use watersheds: Part II. Journal of the American Water Resources Association, 38(3), 603-617.

City of Portland, Oregon Bureau of Planning, (2001), "Discussion Draft: Inventory of Significant Riparian and Wetland Resources, Skyline West Resource Sites" Healthy Portland Streams: A River Renaissance Project.

City of Portland, Bureau of Environmental Services, (2018). Portland Area Watershed Monitoring and Assessment Program (PAWMAP): Report on the First Four Years of Data (FY 2010-11 to FY 2013-14).

City of Portland (2011a), Bureau of Environmental Services: Portland Area Watershed Monitoring and Assessment Program (PAWMAP), https://www.portlandoregon.gov/bes/article/489038 (last accessed: July 10, 2018).

City of Portland (2011b), Stormwater Management Plan: National pollutant discharge elimination system (NPDES) municipal separate storm sewer system (MS4) discharge permit (Permit number: 101314).

City of Portland (2011c), Environmental Services:: Portland watershed management plan, www.portlandoregon.gov/bes/article/338860 (last accessed: July 10, 2018).

Coffin, A. W. (2007). From roadkill to road ecology: a review of the ecological effects of roads. Journal of transport Geography, 15(5), 396-406.

Dalu, T., Wasserman, R. J., Tonkin, J. D., Mwedzi, T., Magoro, M. L., & Weyl, O. L. (2017). Water or sediment? Partitioning the role of water column and sediment chemistry as drivers of macroinvertebrate communities in an austral South African stream. Science of The Total Environment, 607, 317-325.

De'Ath, G. (2007). Boosted trees for ecological modeling and prediction. Ecology, 88(1), 243-251.

Deacon, J. R., Soule, S. A., & Smith, T. E. (2005). Effects of urbanization on stream quality at selected sites in the Seacoast region in New Hampshire, 2001-03. US Department of the Interior, US Geological Survey.

DOGAMI. (1991). Geologic Map of the Portland Quadrangle. Map. [ca. 1:24,000].

Dow, C. L., Arscott, D. B., & Newbold, J. D. (2006). Relating major ions and nutrients to watershed conditions across a mixed-use, water-supply watershed. Journal of the North American Benthological Society, 25(4), 887-911.

Ekness , P., and Randhir, T. (2007). Effects of Riparian Areas, Stream Order, and Land Use Disturbance on Watershed-Scale Habitat Potential: An Ecohydrologic Approach to Policy 1. JAWRA Journal of the American Water Resources Association, 43(6), 1468-1482.

Elmore, W., and Beschta, R. L. (1987). Riparian areas: perceptions in management. Rangelands Archives, 9(6), 260-265.

Forman, R. T., and Alexander, L. E. (1998). Roads and their major ecological effects. Annual review of ecology and systematics, 29(1), 207-231.

France, R. L. (1998). Colonization of leaf litter by littoral macroinvertebrates with reference to successional changes in boreal tree composition expected after riparian clear-cutting. The American midland naturalist, 140(2), 314-324.

Franczyk, J., and Chang, H. (2009). The effects of climate change and urbanization on the runoff of the Rock Creek basin in the Portland metropolitan area, Oregon, USA. Hydrological Processes: An International Journal, 23(6), 805-815.

Frissell, C. A., Liss, W. J., Warren, C. E., and Hurley, M. D. (1986). A hierarchical framework for stream habitat classification: viewing streams in a watershed context. Environmental management, 10(2), 199-214.

Groffman, P. M., Bain, D. J., Band, L. E., Belt, K. T., Brush, G. S., Grove, J. M., ... and Zipperer, W. C. (2003). Down by the riverside: urban riparian ecology. Frontiers in Ecology and the Environment, 1(6), 315-321.

Groffman, Peter M., Neely L. Law, Kenneth T. Belt, Lawrence E. Band, and Gary T. Fisher. Nitrogen fluxes and retention in urban watershed ecosystems. Ecosystems 7, no. 4 (2004): 393-403.

Harding, J. S., et al. "Stream biodiversity: the ghost of land use past." Proceedings of the national academy of sciences 95.25 (1998): 14843-14847.

Hatt, B. E., Fletcher, T. D., Walsh, C. J., and Taylor, S. L. (2004). The influence of urban density and drainage infrastructure on the concentrations and loads of pollutants in small streams. Environmental management, 34(1), 112-124.

Hawkins, C. P., Murphy, M. L., Anderson, N. H., & Wilzbach, M. A. (1983). Density of fish and salamanders in relation to riparian canopy and physical habitat in streams of the northwestern United States. Canadian Journal of Fisheries and Aquatic Sciences, 40(8), 1173-1185.

Hennings, L. A., and Edge, W. D. (2003). Riparian bird community structure in Portland, Oregon: habitat, urbanization, and spatial scale patterns. Condor, 288-302.
.
Hubler, S. (2012). PREDATOR: development and use of RIVPACS-type macroinvertebrate models to assess the biotic condition of wadeable Oregon streams (November 2005 models). DEQ08-LAB-0048-TR Version 1.1. Hillsboro, OR: Oregon Department of Environmental Quality. http://159.121. 9.10/lab/techrpts/docs/10-lab-004. pdf. (last accessed: August 13th, 2019).

Hughes, J. M., S. E. Bunn, D. M. Kingston, and D. A. Hurwood. (1995). Genetic differentiation and dispersal among populations of Paratya australis (Atyidae) in rainforest streams in southeast Queensland, Australia. Journal of the North American Benthological Society 14:158–173.

Hupp, Cliff R., Michael D. Woodside, and Thomas M. Yanosky. "Sediment and trace element trapping in a forested wetland, Chickahominy River, Virginia." *Wetlands* 13, no. 2 (1993): 95-104.

Jackson, C. R., Leigh, D. S., Scarbrough, S. L., and Chamblee, J. F. (2015). Herbaceous versus forested riparian vegetation: narrow and simple versus wide, woody and diverse stream habitat. River research and applications, 31(7), 847-857.

Jones, J. A., Swanson, F. J., Wemple, B. C., & Snyder, K. U. (2000). Effects of roads on hydrology, geomorphology, and disturbance patches in stream networks. Conservation Biology, 14(1), 76-85.

Jun, M. J. (2004). The effects of Portland's urban growth boundary on urban development patterns and commuting. Urban Studies, 41(7), 1333-1348.

Karr, James R., and E. W. Chu. "Seven foundations of biological monitoring and assessment." Biologia Ambientale20.2 (2006): 7-18.

Kasprak, A., Magilligan, F. J., Nislow, K. H., and Snyder, N. P. (2012). A Lidar-derived evaluation of watershed-scale large woody debris sources and recruitment mechanisms: Coastal Maine, USA. River Research and Applications, 28(9), 1462-1476.

Kasprak, A., Sankey, J. B., Buscombe, D., Caster, J., East, A. E., & Grams, P. E. (2018). Quantifying and forecasting changes in the areal extent of river valley sediment in response to altered hydrology and land cover. Progress in Physical Geography, 1, 26.

Kawaguchi, Y., Taniguchi, Y., and Nakano, S. (2003). Terrestrial invertebrate inputs determine the local abundance of stream fishes in a forested stream. Ecology, 84(3), 701-708.

Kaye, J. P., Groffman, P. M., Grimm, N. B., Baker, L. A., and Pouyat, R. V. (2006). A distinct urban biogeochemistry. Trends in Ecology and Evolution, 21(4), 192-199.

Kentula, M. E., Gwin, S. E., and Pierson, S. M. (2004). Tracking changes in wetlands with urbanization: sixteen years of experience in Portland, Oregon, USA. Wetlands, 24(4), 734-743.

Kuivila, K. M., Hladik, M. L., Ingersoll, C. G., Kemble, N. E., Moran, P. W., Calhoun, D. L., ... & Gilliom, R. J. (2012). Occurrence and potential sources of pyrethroid insecticides in stream sediments from seven US metropolitan areas. Environmental science & technology, 46(8), 4297-4303.

Ladson, A. R., Walsh, C. J., & Fletcher, T. D. (2006). Improving stream health in urban areas by reducing runoff frequency from impervious surfaces. Australasian Journal of Water Resources, 10(1), 23-33.

Lane, P. N., and Sheridan, G. J. (2002). Impact of an unsealed forest road stream crossing: water quality and sediment sources. Hydrological processes, 16(13), 2599-2612.

Lefsky, M. A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002). Lidar remote sensing for ecosystem studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. BioScience, 52(1), 19-30.

Lenat, D. R., and Crawford, J. K. (1994). Effects of land use on water quality and aquatic biota of three North Carolina Piedmont streams. Hydrobiologia, 294(3), 185-199.

Likens, G. E., Driscoll, C. T., Buso, D. C., Siccama, T. G., Johnson, C. E., Lovett, G. M., ... & Bailey, S. W. (1998). The biogeochemistry of calcium at Hubbard Brook. Biogeochemistry, 41(2), 89-173.

Lowe, E. A., & Evans, L. K. (1995). Industrial ecology and industrial ecosystems. Journal of cleaner production, 3(1-2), 47-53.

May, C. W., Horner, R. R., Karr, J. R., Mar, B. W., & Welch, E. B. (1998). The cumulative effects of urbanization on small streams in the Puget Sound Lowland Ecoregion. In Proceedings of the Puget Sound Research.

Minan, J. H. (2005). Municipal Separate Storm Sewer System (MS4) Regulation Under the Federal Clean Water Act: The Role of Water Quality Standards. San Diego L. Rev., 42, 1215

Messer, J. J., Linthurst, R. A., and Overton, W. S. (1991). An EPA program for monitoring ecological status and trends. Environmental Monitoring and Assessment, 17(1), 67-78

Morley, S. A., and Karr, J. R. (2002). Assessing and restoring the health of urban streams in the Puget Sound basin. Conservation Biology, 16(6), 1498-1509.

Naiman, R. J., Beechie, T. J., Benda, L. E., Berg, D. R., Bisson, P. A., MacDonald, L. H., … & Steel, E. A. (1992). Fundamental elements of ecologically healthy watersheds in the Pacific Northwest coastal ecoregion. In Watershed management (pp. 127-188). Springer, New York, NY.

Nedeau, E. J., Merritt, R. W., and Kaufman, M. G. (2003). The effect of an industrial effluent on an urban stream benthic community: water quality vs. habitat quality. Environmental pollution, 123(1), 1-13.

NOAA (2018). Climate of Portland, OR. https://www.wrh.noaa.gov/pqr/pdxclimate/introduction.pdf. (last accessed July 19, 2018)

Olson, D. H., & Weaver, G. (2007). Vertebrate assemblages associated with headwater hydrology in western Oregon managed forests. Forest Science, 53(2), 343-355.

Osborne, L. L., and Kovacic, D. A. (1993). Riparian vegetated buffer strips in water-quality restoration and stream management. Freshwater biology, 29(2), 243-258.

Pascoe, D., Evans, S. A., and Woodworth, J. (1986). Heavy metal toxicity to fish and the influence of water hardness. Archives of Environmental Contamination and Toxicology, 15(5), 481-487.

Paul, M. J., and Meyer, J. L. (2001). Streams in the urban landscape. Annual review of Ecology and Systematics, 32(1), 333-365.

Peterson, B. J., Wollheim, W. M., Mulholland, P. J., Webster, J. R., Meyer, J. L., Tank, J. L., … & McDowell, W. H. (2001). Control of nitrogen export from watersheds by headwater streams. Science, 292(5514), 86-90.

Peterson, E. E., Theobald, D. M., & Ver Hoef, J. M. (2007). Geostatistical modelling on stream networks: developing valid covariance matrices based on hydrologic distance and stream flow. Freshwater biology, 52(2), 267-279.

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems, 9(2), 181-199. a watershed from septic sources. Journal of Environmental Quality, 4(2), 249-252.

Reinelt, L., Horner, R., & Azous, A. (1998). Impacts of urbanization on palustrine (depressional freshwater) wetlands—research and management in the Puget Sound region. Urban Ecosystems, 4(2), 219-236.

Rios-Touma, B., Prescott, C., Axtell, S., and Kondolf, G. M. (2015). Habitat restoration in the context of watershed prioritization: the ecological performance of urban stream restoration projects in Portland, Oregon. River Research and Applications, 31(6), 755-766.

Roth, N. E., Allan, J. D., & Erickson, D. L. (1996). Landscape influences on stream biotic integrity assessed at multiple spatial scales. Landscape ecology, 11(3), 141-156.

Roy, A. H., Faust, C. L., Freeman, M. C., & Meyer, J. L. (2005). Reach-scale effects of riparian forest cover on urban stream ecosystems. Canadian Journal of Fisheries and Aquatic Sciences, 62(10), 2312-2329.

Shaffer, P. W., & Ernst, T. L. (1999). Distribution of soil organic matter in freshwater emergent/open water wetlands in the Portland, Oregon metropolitan area. Wetlands, 19(3), 505-516.

Shandas, V., and Messer, W. B. (2008). Fostering green communities through civic engagement: community-based environmental stewardship in the Portland area. Journal of the American Planning Association, 74(4), 408-418.

Smith, V. H., Tilman, G. D., and Nekola, J. C. (1999). Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. Environmental pollution, 100(1-3), 179-196.

Snyder, C. D., Young, J. A., Villella, R., and Lemarie, D. P. (2003). Influences of upland and riparian land use patterns on stream biotic integrity. Landscape Ecology, 18(7), 647-664.

Solins, J. P., Thorne, J. H., and Cadenasso, M. L. (2018). Riparian canopy expansion in an urban landscape: Multiple drivers of vegetation change along headwater streams near Sacramento, California. Landscape and Urban Planning, 172, 37-46.

Somerville, D. E. (2010). Stream assessment and mitigation protocols: a review of commonalities and differences. Contract No. GS-00F-0032M, US Environmental Protection Agency, Washington, DC.

Song, K., Kang, H., Zhang, L., & Mitsch, W. J. (2012). Seasonal and spatial variations of denitrification and denitrifying bacterial community structure in created riverine wetlands. Ecological Engineering, 38(1), 130-134.

Sonoda, K., & Yeakley, J. A. (2007). Relative effects of land use and near-stream chemistry on phosphorus in an urban stream. Journal of environmental quality, 36(1), 144-154.

Sridhar, V., Sansone, A. L., LaMarche, J., Dubin, T., & Lettenmaier, D. P. (2004). Prediction of stream temperature in forested watersheds. Journal of the American Water Resources Association, 40(1), 197-213.

Stepenuck, K. F., Crunkilton, R. L., and Wang, L. (2002). Impacts of urban land use on macroinvertebrate communities in southeastern Wisconsin streams 1. JAWRA Journal of the American Water Resources Association, 38(4), 1041-1051.

Stevens Jr, D. L., & Olsen, A. R. (2004). Spatially balanced sampling of natural resources. Journal of the American Statistical association, 99(465), 262-278.

Titus, J. H., Christy, J. A., VanderSchaaf, D., Kagan, J. S., & Alverson, E. R. (2006). Native wetland, riparian, and upland plant communities and their biota in the Willamette Valley, Oregon.

Tong, S. T., & Chen, W. (2002). Modeling the relationship between land use and surface water quality. Journal of environmental management, 66(4), 377-393.

Trainor, K., & Church, M. (2003). Quantifying variability in stream channel morphology. Water Resources Research, 39(9).

Triska, F. J., Kennedy, V. C., Avanzino, R. J., Zellweger, G. W., & Bencala, K. E. (1989). Retention and transport of nutrients in a third-order stream in northwestern California: Hyporheic processes. Ecology, 70(6), 1893-1905.

Trombulak, S. C., and Frissell, C. A. (2000). Review of ecological effects of roads on terrestrial and aquatic communities. Conservation biology, 14(1), 18-30.

Van Metre, P. C., Mahler, B. J., & Furlong, E. T. (2000). Urban sprawl leaves its PAH signature. Environmental Science & Technology, 34(19), 4064-4070.

Ver Hoef, J. M., Peterson, E., & Theobald, D. (2006). Spatial statistical models that use flow and stream distance. Environmental and Ecological statistics, 13(4), 449-464.

Walsh, C. J., Roy, A. H., Feminella, J. W., Cottingham, P. D., Groffman, P. M., and 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.

Wang, X. (2001). Integrating water-quality management and land-use planning in a watershed context. Journal of environmental management, 61(1), 25-36.

Wang, L., Lyons, J., Kanehl, P., and Gatti, R. (1997). Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. Fisheries, 22(6), 6-12.

Wang, L., Lyons, J., Kanehl, P., and Bannerman, R. (2001). Impacts of urbanization on stream habitat and fish across multiple spatial scales. Environmental Management, 28(2), 255-266.

Wang, L., Seelbach, P. W., and Lyons, J. (2006). Effects of levels of human disturbance on the influence of catchment, riparian, and reach-scale factors on fish assemblages. In American Fisheries Society Symposium (Vol. 48, No. 48, pp. 641-664).

Ward, D. L., Nigro, A. A., Farr, R. A., and Knutsen, C. J. (1994). Influence of waterway development on migrational characteristics of juvenile salmonids in the lower Willamette River, Oregon. North American Journal of Fisheries Management, 14(2), 362-371.

Wheeler, A. P., Angermeier, P. L., and Rosenberger, A. E. (2005). Impacts of new highways and subsequent landscape urbanization on stream habitat and biota. Reviews in fisheries science, 13(3), 141-164.

Wooster, J., and Hilton, S. (2004). Large woody debris volumes and accumulation rates in cleaned streams in redwood forest in southern Humboldt County, California. Research Note PSW-426. Albany, CA: US Dept. of Agriculture, Forest Service, Pacific Southwest Research Station. 16 p, 426.

Wojciechowska, E., Nawrot, N., Walkusz-Miotk, J., Matej-Łukowicz, K., & Pazdro, K. (2019). Heavy metals in sediments of urban streams: contamination and health risk assessment of influencing factors. Sustainability, 11(3), 563.

Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (2000). Assessing the biological quality of fresh waters. RIVPACS and other techniques. Freshwater Biological Association, Ambleside, England.

Young, K. D., & Thackston, E. L. (1999). Housing density and bacterial loading in urban streams. Journal of Environmental Engineering, 125(12), 1177-1180.