Optimising the stream habitat assessment for the Bay of Plenty Regional Council

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Conservation Biology

at Massey University, Palmerston North,

New Zealand.

Adele Bittner

2021

Abstract

Multiple stream habitat attributes are evaluated using qualitative and quantitative assessment methods during the current Bay of Plenty Natural Environmental Monitoring Network (BOP NERMN) stream habitat survey. These assessments are carried out by a team of different students each summer. Because qualitative assessments typically require less time (shorter in-field assessment duration) than quantitative measurements, using only qualitative assessment methods are likely to be more economical. However, the results of qualitative assessments are thought to be more subjective. Therefore, the year-to-year change of the surveying team (inter-annual observer variability) could influence the BOP NERMN stream habitat monitoring results. In this thesis, I aimed to determine whether omitting quantitative metrics and using only qualitative metrics could be an appropriate option for creating a more economical BOP NERMN stream habitat survey, considering the metrics' in-field assessment duration and inter-annual observer variability.

First, I investigated whether qualitative and quantitative metrics captured stream habitat attributes similarly by assessing the relationship between the two approaches through Spearman rank correlation tests. The Spearman rank correlation analyses revealed that all qualitative metrics, apart from the 'RHA riparian shade', were significantly correlated to at least one quantitative metric. Furthermore, I timed the in-field assessment duration of qualitative and quantitative metrics, with the result that, on average per site, all qualitative metrics were evaluated within 9 minutes, and all quantitative metrics were measured in 17 minutes. I investigated inter-annual observer variability by comparing each metric's percent coefficient of variation (CV) through Bayesian hierarchical linear models and found that data from most metrics (69%) had high levels of inter-annual observer variability (CV estimate > 30%), regardless of whether metrics were quantitative or qualitative. Lastly, I applied value models to evaluate the performance trade-offs between infield-assessment duration and inter-annual observer variability of qualitative and quantitative metrics. More than half (56%) of all metrics performed relatively equally in relation to their in-field assessment duration and inter-annual observer variability, regardless of whether metrics were qualitative or quantitative. These results suggest there is currently no clear reason to favour qualitative metrics in the BOP NERMN stream habitat survey. Overall, this research suggests the existing BOP NERMN stream habitat assessment protocol requires further refinement to reduce inter-annual observer variability.

Acknowledgements

I would like to acknowledge several people who have helped me accomplish my academic achievements. Firstly, I would like to express my gratitude to my parents, Wolfgang and Brigitte Bittner, who gave me the freedom to choose my own path through life, even if it involved many detours, and who never doubted I chose the correct way. I would also like to thank my Kiwi family, the Lammers family, for always being my support network in this country I now call home. Through living with them, I turned from the naïve and pampered only child I once was, into a responsible young adult.

The biggest THANK YOU to my supervisor, Prof Doug Armstrong, who replied to my emails, shared his academic experience and gave constructive feedback in an incredibly short timeframe; and to my co-supervisor, Dr Alastair Suren, who was the inspiration behind this research, shared his data for this study, and provided valuable comments to improve this thesis. I truly appreciate the time you two spent reading my drafts and answering my questions. I know you two only want the best for me, and I am so grateful for your support throughout the years! Furthermore, I would like to acknowledge the Bay of Plenty Regional Council for employing a temporary field technician and providing equipment in order for me to collect supplementary data.

I would also like to express my gratitude to Grace Kirk and Adam Canning for proofreading countless assignments throughout my undergraduate studies before I discovered the 'read aloud' function in Microsoft Word. Additionally, I wish to thank Adam for his tutoring, advice and mental support. Many more friends have been part of a strong support network, and together we pushed each other through our years at university. Thank you all so much for your encouraging words, heart-warming drawings and good banter. Last but definitely not least, I would like to acknowledge Brian Levine for supporting me throughout this study by keeping me in good spirits.

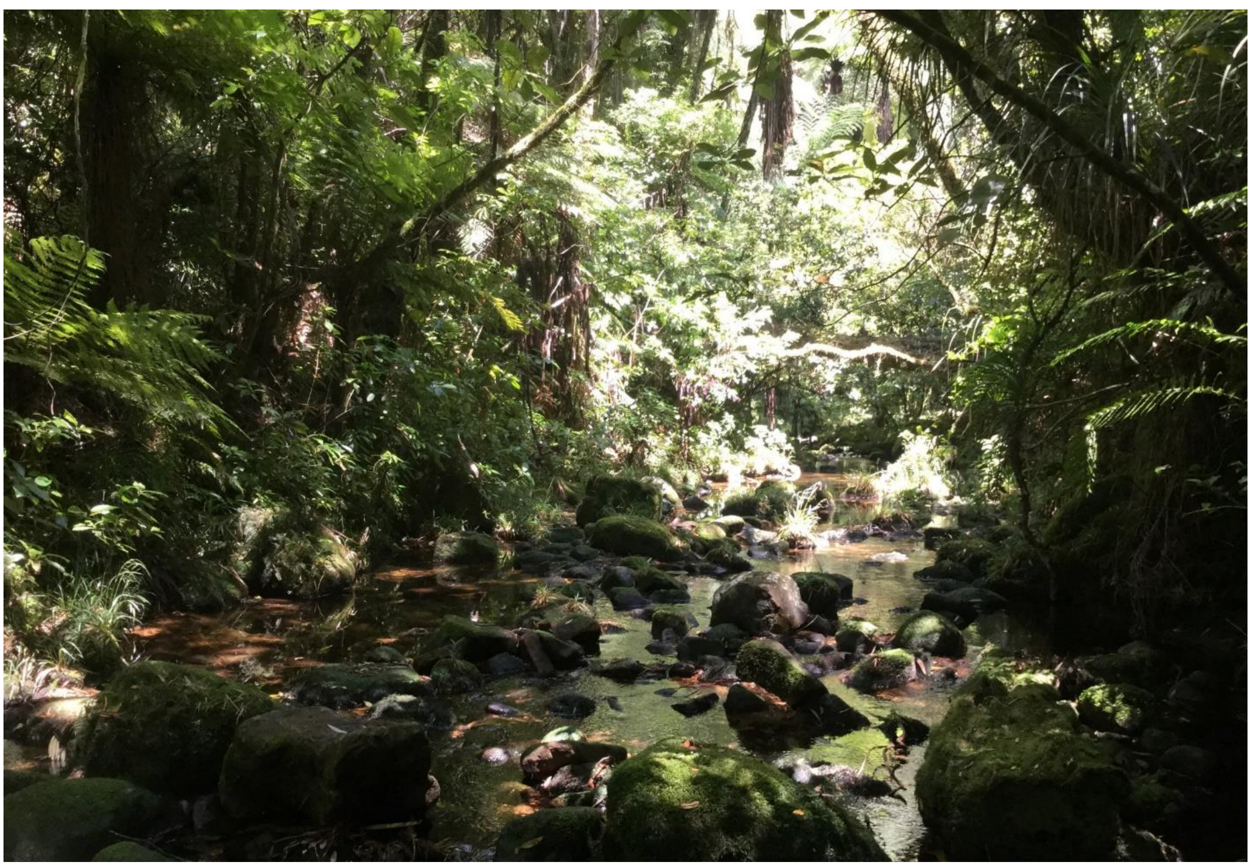
Table of Content

List of Abbreviations

Abbreviation	Meaning
aka	also known as/ also referred to
Av_Bank_U	‘average bank undercut’
Av_Sed_Dep	‘average sediment depth’
BOP RC	Bay of Plenty Regional Council
BOP_NERM_#	Monitoring site code
BU	Bank Undercut
CV	Percent coefficient of variation
CV_Vel	‘coefficient of variation flow velocity’
Flo_Het	‘flow heterogeneity’
FMP	Field Measurement Precision
IAOV	Inter-annual Observer Variability
IFAD	In-field Assessment Duration
GPS	Global Positioning System
HB	Hard Bottomed
MCI	Macroinvertebrate Community Index
NA	Not Applicable/ Not Available
NERMN	Natural Environment Regional Monitoring Network
NPS – FM 2020	National Policy Statement for Freshwater Management
OV	Overhanging Vegetation
QL	Qualitative
QT	Quantitative
RC_SID	Regional Council sampling sites
RHA	Rapid Habitat Assessment
RHA_Ch_Alt	‘RHA channel alternation’
RHA_Fish	‘RHA fish cover’
RHA_Hyd	‘RHA hydraulic heterogeneity’
RHA_Inv	‘RHA invertebrate habitat’
RHA_Sed	‘RHA fine sediment deposition’
RMA	Resource Management Act
SAM 2	Sediment Assessment Method 2
SB	Soft Bottomed
SDM	Structured Decision Making
Sed	data set for deposited fine sediment attribute
Sed_brm	Model name in R
Shuffle	‘shuffle index’
Std_Dep	‘standard deviation stream depth’
Std_Vel	‘standard deviation flow velocity’
Std_WetWid	‘standard deviation wetted width’
Sub_Div	‘substrate diversity’
Sub_Ind	‘substrate index’
SoE	State of the Environment

Chapter 1:

General Introduction



Native forest site in the Bay of Plenty. Summer 2018/19

1.1 Stream habitat and influential factors

"A river doesn't just carry water, it carries life."

- Amit Kalantri

This thesis was focused on streams and the habitat they provide to aquatic biota. Stream habitat is made up of physical, chemical and biological features, as well as the water which flows through the habitat (Harding et al., 2009; Kaufmann et al., 1999), and is formed by the interactions between topography, geology, climate and land use (NIWA, 2019). Harding et al. (2009) has found that habitat quantity and quality affect the diversity, abundance and distribution of aquatic communities. An intact stream habitat provides shelter, predator protection and residence for organisms (Elosegi et al., 2011), as well as spawning sites (Harding et al., 2009). A diverse range of habitat types can sustain an array of diverse biota, as many aquatic species have unique habitat preferences (NIWA, 2019). A healthy stream system is one in which habitat extends from the waterbody to its surrounding floodplains, while in an unhealthy stream habitat, the physical form of the stream can no longer sustain a large variety of aquatic organisms (Clapcott, 2015). Thus, the type and quality of physical habitat strongly influence aquatic species (NIWA, 2019). This study focuses on the physical aspects of stream[1] habitats and biological attributes that influence or provide resources to aquatic organisms.

Because streams are open systems, they are closely connected to each other and their surrounding environments (Cloern, 2007). For example, water circulates through streams via the hydraulic cycle, from the atmosphere to the oceans, and organisms may transfer between aquatic and terrestrial ecosystems (Harding et al., 2009). Therefore, streams, and the habitat they provide, are influenced by environmental factors (Beechie et al., 2010), which act at different spatiotemporal scales (Frissell et al., 1986). For instance, rare geological events of large magnitude can bring about fundamental changes to entire catchments, whereas geomorphic events of high reoccurrence and low magnitude may alter streams at smaller spatial scales, including reach-, transect- or microhabitat segments

[1] Thereafter, the term 'stream(s)' refers to streams and rivers, unless stated otherwise.

(Frissell et al., 1986). At a stream's reach and transect level, environmental factors can affect channel morphology and riparian zone conditions (Beechie et al., 2010; Frissell et al., 1986). Channel morphology and riparian zone, in turn, play an integral part in regulating a riverine ecosystem and have strong influences on stream habitat conditions (Beechie et al., 2010; Harding et al., 2009).

1.1.1 Effects of channel morphology

Channel morphology affects aquatic habitat as it influences flow velocity, substrate size and flow types (Beechie et al., 2010; Miserendino et al., 2011). Flow velocity is the speed at which water travels through a waterbody and is a crucial attribute of stream habitat (NIWA, 2019). Fast flow velocity is beneficial to aquatic biota, as it boosts nutrient uptake in plants as well as dissolved oxygen uptake in animals and delivers more food for aquatic organisms than slow currents (NIWA, 2019). An average flow velocity of less than 0.3m/s is expected to have adverse effects on aquatic biota, as slow flow is associated with the deposition of sediment and periphyton growth (Harding et al., 2009). Therefore, fast-flowing streams often provide habitat to a larger diversity of aquatic biota than sluggish streams (NIWA, 2019). However, slow-flowing pools provide important habitat for certain aquatic biota. Good stream habitat comprises an array of flow velocities, as various aquatic species have adapted to distinct velocity ranges (Elosegi et al., 2011). For example, Jowett et al. (1996) have found that fish have different velocity preferences and that fish abundance and variety was consistent with their flow type preferences. Nevertheless, during flood events, fish and macroinvertebrates may be displaced, macrophytes may be uprooted, and stream substrate may be moved (NIWA, 2019). Consequently, substrate composition is tightly linked to flow velocity (Parsons et al., 2002).

The substrate composition has major influences on habitat quality and quantity (Harding et al., 2009). The streambed provides aquatic fauna with refuge places, food forage opportunities, and egg deposition and incubation spaces (Gebrekiros, 2016; Parsons et al., 2002). Several aquatic species are adapted to specific substrate types (Elosegi et al., 2011), as requirements for substrate composition, or accessibility of interstitial spaces vary between species (Parsons et al., 2002). Generally, a streambed with a diverse substrate composition provides suitable habitat for a variety of fish and invertebrates, whereas a uniform streambed does not (Parsons et al., 2002). Streambeds comprising of primarily boulders and cobbles offer surfaces for macroinvertebrates to crawl on or under and are therefore said to support a greater diversity of macroinvertebrates than streambeds made up of fine substrate (NIWA, 2019). Additionally, a combination of water depth, flow velocity and substrate composition defines different flow types (Yang, 1971), to which many aquatic biotas have specific preferences (NIWA, 2019). Filter feeders, for instance, are more commonly found in riffle habitat than in pools (Harding et al., 2009). Commonly assessed flow types during stream habitat assessments are riffles, runs and pools (Yang, 1971). Therefore, researchers may draw conclusions from the inspection of substrate composition and flow types as to which aquatic communities may exist in a stream (NIWA, 2019).

1.1.2 Effects of the riparian zone

The characteristics of the riparian vegetation play an essential part in stream protection (Davies and Nelson, 1994). A riparian zone that is covered in mature vegetation provides many ecological benefits to stream habitats, such as reducing sediment input, nutrient and contaminants uptake, organic matter and wood inputs as well as the provision of shade and overhanging vegetation (Allan, 2004; Bauer and Ralph, 2001; Miserendino et al., 2011).

A riparian zone will have the greatest benefits to stream health when the vegetation canopy is dense and closed (Suren et al., 2017). A mature canopy cover and a thick leaf litter layer of riparian vegetation can intercept heavy precipitation (Simon and Collison, 2002), thereby reducing the amount of surface runoff entering a stream (Suren et al., 2017). Reduced surface runoff decreases flow velocity, reduces the risk of bank erosion or undercutting, as well as decreases in-stream sediment-, nutrient- and contaminant loads. Sediments as well as other nutrients and contaminants that get carried towards the stream in surface runoff, can be trapped and taken up by the roots of riparian plants (Dosskey et al., 2010; Hill, 1996). Additionally, the root system of intact riparian vegetation limits bank erosion, as it stabilises the ground (Beechie et al., 2010; Dosskey et al., 2010). Sediment and nutrients can have negative impacts on stream biota. Sediment deposited on and in between the stream's substrate can smother macroinvertebrate habitat (Gayraud and Philippe, 2003) and may decrease the survival of fish eggs and larvae (Gebrekiros, 2016). Sediment can also clog up fishes' gills (NIWA, 2019). Increased nutrient loads can also accelerate algal growth, which may reduce macroinvertebrate habitat (Matthaei et al., 2010).

Riparian vegetation supplies a stream with organic matter and dead wood (Dosskey et al., 2010; Elosegi et al., 2011). Organic matter is an integral energy input for a riverine food web, as it provides food for macroinvertebrates and microorganisms (Bundschuh and McKie, 2016). These small organisms, in turn, provide an important food source for fish (Baxter et al., 2005). In-stream leaf packs and dead wood also provide habitat and protection for aquatic biota (Allan, 2004).

Furthermore, dead wood modifies the flow regime in a stream by altering flow velocities and creating different flow types such as pools and cascades (Beechie et al., 2010; Elosegi et al., 2011).

A dense riparian and overhanging stream vegetation provides shading for a stream (Elosegi et al., 2011). Shading influences water temperature and the amount of sunlight reaching a stream (Allan, 2004). As warm and sunny conditions accelerate plant growth, shading has a major impact on algae growth in streams (Harding et al., 2009). Algae growing on the stream substrate alters habitat availability for macroinvertebrates. Additionally, overhanging stream vegetation offers food sources, as terrestrial insects residing on the riparian vegetation may fall into the stream (Allan et al., 2003) and provides spawning sites for fish (Harding et al., 2009).

However, small scale ecological functions of the riparian zone are frequently outweighed by the large-scale impacts occurring in a catchment (Suren et al., 2017). Studies have shown that the riparian zone width (Davies and Nelson, 1994) and the ratio of riparian length to catchment size (Suren et al., 2017) are the dominant factors influencing riparian ecological services and functions. For example, narrow riparian zones (≤ 10 m width) do not significantly protect a stream from impacts, whereas broader riparian strips, 30 to 100 m wide, have been found to do so effectively (Davies and Nelson, 1994). Furthermore, riparian zones may not fulfil their services and functions if they make up a small strip of land within a much larger landscape dominated by human activities (Suren et al., 2017).

1.1.3 Effects of land use

Human activities disrupt ecological services and functions that maintain a healthy riverine ecosystem, and the relationship between human activities and their effects on stream habitat can be complex (Miserendino et al., 2011). Stream habitat conditions are influenced by large and small spatiotemporal environmental factors (Fernandez et al., 2011), which complicate pinpointing the effects of human activities on stream habitat (Barquín Ortiz and Martinez-Capel, 2011). As catchments gather and funnel water downstream, individual human activities and changes can accumulate into major environmental pressures in a catchment, even if these activities seem to have minor effects on their own (Ministry for the Environment and Stats NZ, 2020).

The global shift from natural environments to human-dominated land use is impacting ecosystems worldwide (Allan, 2004). Human land use influences channel morphology, sediment loads and riparian vegetation (Jowett et al., 1996). Jowett et al. (1996) found that the abundance of some fish species varies amongst different land uses, and this variation is often attributed to differences in physical stream habitat. Waterway and land management activities have severely altered streams worldwide within the last 100 years (Beechie et al., 2010). For example, approximately 20% of the world's electricity supply was being generated by large dams in the early 2000s, and 40% of the world's food was produced on irrigated land (Barquín Ortiz and Martinez-Capel, 2011). Not only do dams affect downstream flow regimes and alter the downstream transportation of sediments and nutrients; in New Zealand, damns also interrupt the ability of native migratory fish species to complete their life cycle (Suren et al., 2017). Worldwide, more than 40% of aquatic biodiversity has been lost, and ecological services and functions have been compromised due to human activities (Barquín Ortiz and Martinez-Capel, 2011). It is estimated that water shortage caused by increased human demands and climate change (Beechie et al., 2010), as well as the loss of ecosystem services and functions, may affect 40% of people worldwide by 2050 (Barquín Ortiz and Martinez-Capel, 2011).

In New Zealand, the scope of human land use transformation is significant, with only 30% of native bush remaining. Twenty-four percent of land is used for cropping, horticulture and pasture, 7% of land is covered in exotic forest plantations, and urban settings stretch over 1% of land area (Ministry for the Environment, 2010). Agricultural and urban land uses are the most environmentally degrading activities, although forestry operations of exotic forest plantations can also damage stream habitat conditions (Ministry for the Environment, 2010). Even though urban land generally only occupies a small part of a total catchment area, it disproportionally influences its immediate and distant environment compared to other human land uses (Miserendino et al., 2011).

As agricultural land use increases, water and stream habitat quality decreases, impacting the in-stream species composition (Allan, 2004). Agricultural land use affects stream habitat in many ways, as nonpoint inputs of sediments, nutrients and pesticides are increased, riparian zone and stream channels are degraded, and water flows may be modified (Matthaei et al., 2010). Additionally, agricultural streams tend to have compromised bank stability and increased sediment and contaminant loads (Dosskey et al., 2010), as the native riparian vegetation of most streams running through agricultural land has been removed (Ministry for the Environment, 2010). In agricultural settings, often stream channels have been or are continually modified to reduce the effects of floods (Schoof, 1980). In New Zealand, an ongoing farm practice to reduce the effects of floods is to dredge plant and bed materials from streams and drains, which destroys stream habitat and directly removes and kills aquatic biota (Christensen et al., 2017). Additionally, channelisation and realignment of streams can modify and/or unify substrate compositions and flow types, which otherwise would provide a diversity of habitats for aquatic species (Miserendino et al., 2011). Water extractions for irrigation purposes reduce flow velocity and water depth and thus directly reduce habitat as the wetted stream width narrows (Dewson et al., 2007).

While agricultural and urban land uses share some pressures affecting stream habitat, including the lack of riparian services and functions as well as channel modification, urban settings also exert unique stresses on streams (Allan, 2004). For example, impervious surfaces and stormwater conveying systems increase erratic hydrology due to enhanced runoff in urban areas (Paul and Meyer, 2001). Additionally, urban runoff is associated with increased amounts and varieties of pollutants, such as heavy metals, oils and chemicals, compared to other land uses (Paul and Meyer, 2001). In New Zealand, many streams in urban areas are polluted with pathogens (Ministry for the Environment and Stats NZ, 2020). Lined channels commonly found in urban settings may reduce habitat structures, decrease channel morphologies, and restrict interactions between the stream and the riparian zone (Miserendino et al., 2011; Paul and Meyer, 2001). A study conducted in Patagonia found that urban land use produced the most significant changes to habitat conditions, riparian quality, nutrient loads and invertebrate metrics, compared to pine plantations, pastoral and native land cover (Miserendino et al., 2011).

Forestry operations, including logging and roads crossing streams, can negatively affect stream habitat as a result of increased stream sediment loads, altered stream morphology and hydrology, as well as impacts to aquatic biota (Davies and Nelson, 1994). Additionally, the removal or input of logging debris can also cause changes to stream morphology and biota. Davies and Nelson (1994) found that logging increased the amount of deposited sediment, algal cover and logging debris in streams with narrow riparian zones (< 30m width). The effects of logging activities on streams lined with narrow riparian zones were also associated with decreased shading above the stream and higher water temperature (Davies and Nelson, 1994).

In the Bay of Plenty, irrigation, contaminated discharges and hydraulic dams are the predominant human activities causing pressures on stream conditions (Suren et al., 2017). With over 1000 granted water take consents in the Bay of Plenty region, urban and agricultural activities are largely sustained by irrigation (Suren et al., 2017). In 2017 nearly 430 consents permitted discharges of contaminants to be released into water, originating from agricultural and industrial activities, as well as from stormwater or wastewater treatment plants (Suren et al., 2017). Eight hydroelectrical dams are operating within the Bay of Plenty, and together these dams have a capacity of approximate 190MW (Suren et al., 2017).

1.2 Regulating human activities in New Zealand

1.2.1 Environmental legislation framework

As land uses can have major impacts on the environment (Allan, 2004), human activities need to be regulated to protect natural resources (Ministry for the Environment, 2020b). To safeguard renewable resources in a coordinated and comprehensive approach, governments implement environmental legislative frameworks, including natural resource management policies (Ministry for the Environment, 2020b). A natural resource management policy can be defined as “any action deliberately taken to manage human activities with a view to prevent, reduce, or mitigate harmful effects on nature and natural resources, and ensuring that man-made changes to the environment do not have harmful effects on humans or the environment” (McCormick, 2001).

In New Zealand, the Resource Management Act 1991 (RMA) is the main piece of environmental legislation (Ministry for the Environment, 2021). The RMA outlines how New Zealand’s natural and physical resources should be managed and includes sections on the sustainable management of air, soil, freshwater and coastal marine areas, as well as on regulations for land use and infrastructure. The RMA directs sustainable management by providing National Environmental Standards and Policy Statements. National Environmental Standards are regulations that define standards, methods or other requirements for human activities. Meanwhile, National Policy Statements define objectives and policies to achieve sustainable management under the RMA. Regional councils must compose Regional Policy Statements and Regional Plans containing objectives that align with the national policy statements and environmental standards that are effective at that time. The Regional Policy Statements must specify environmental issues significant for the particular region and must include action plans to address those issues.

Decisions on resource consents and permits are made in line with local plans, national directions, and RMA objectives (Ministry for the Environment, 2021). Examples for resource consents include consents occupying the coast, executing activities in rivers, obtaining natural water, or discharging contaminants into the environment. This framework allows for most decisions being made by local governments.

1.2.2 Freshwater management in New Zealand

In September 2020, new amendments to the national directions for freshwater management came into force, as it was found that previous regulations could not halt declining freshwater quality in many of New Zealand's catchments (Ministry for the Environment and Ministry for Primary Industries, 2020a). The Essential Freshwater Package is part of the updated national direction for managing New Zealand's freshwaters, which aims to stall further degradation of freshwater quality, make rapid improvements within the next five years, and restore waterways and ecosystems to a healthy state within a generation.

The National Environmental Standards for Freshwater 2020 (NES-FW 2020) include measures to halt the degradation of freshwater quality, while the National Policy Statement for Freshwater Management 2020 (NPS-FM 2020) provides national guidelines for regional councils to prepare their regional policy statements and plans (Ministry for the Environment, 2020a). The NPS-FM 2020 requires regional councils to give effect to the Te Mana o Te Wai, which is the fundamental concept underpinning the NPS-FM 2020 (Ministry for the Environment and Ministry for Primary Industries, 2020b). Under the Te Mana o Te Wai, regional councils must consult with tangata whenua (people of the land) and communities to set long-term visions for freshwater resources and actively involve tangata whenua in managing freshwater. Additionally, new national documents included in the Essential Freshwater Package are the Stock Exclusion Regulations and Measurement and Reporting of Water Takes Regulations (Ministry for the Environment and Ministry for Primary Industries, 2020a).

1.2.3 Bay of Plenty Natural Environmental Resource Monitoring Network programme

As regional councils must develop Regional Policy Statements and Regional Plans containing objectives in line with the National Policy Statements, the Bay of Plenty Regional Council developed the Natural Environmental Monitoring Network (BOP NERMN) programme (Suren et al., 2017). Put simply, the BOP NERMN programme was designed to fulfil the state of the environment monitoring requirements under the RMA (Donald, 2014). The program has been running since 1989 and has now progressed to include over 1000 monitoring sites across a broad range of natural resources. The programme's original goal was "to provide scientifically defensible information on the important physical, chemical and biological characteristics of the natural resources of the Bay of Plenty region as a basis for the preparation of Bay of Plenty Regional Council policies and plans, and the monitoring of their suitability and effectiveness" (Donald, 2014). Under the BOP NERM programme, natural resources are grouped into four modules and subsequent categories. These groups comprise of land, water, air and ecology. Subsequent categories of the ecology module are freshwater, marine, wetlands and terrestrial subcategories. The freshwater ecology category was introduced to the BOP NERM programme in 1992 to assess ecological conditions and detect trends in streams' ecosystem health throughout the region (Donald, 2014). The assessment of ecological health in streams includes collecting aquatic macroinvertebrates and assessing habitat conditions. As of 2020, 134 freshwater ecology monitoring sites across the entire region were included in the BOP NERMN programme (Ministry for the Environment, 2017). Fifty-five sites were located in agricultural settings, eight sites were found in urban areas, 24 sites were situated in exotic forest plantations, and 47 were located in native bush. Different summer students sample these sites between December and February each year (Ministry for the Environment, 2017).

1.3 State and trends of stream habitat and health in New Zealand

In New Zealand, 70 major river systems make up more than 425,000 kilometres in length (Ministry for the Environment and Stats NZ, 2020). Around 440 billion m^3 of water flows through all rivers and streams across New Zealand. A statistical summary, prepared by Stats NZ (2020), has shown that habitat conditions were excellent in 22.2%, good in 57.1% and fair in 20.7% of the assessed sites between 2013/14 and 2018/19. Assessed sites were located in native, pastoral, exotic and urban land areas. Of the monitored sites, streams in native locations had the highest proportion of streams scoring excellent or good habitat conditions (97.4%), followed by pastoral sites (75.4%) and exotic forest sites (65.05%). Urban sites had the least proportion of streams scoring excellent or good habitat conditions (52.2%).

Furthermore, a report by Clapcott et al. (2019) found that on average New Zealand’s stream ecosystem health was impaired and identified issues concerning water quality, physical habitat, ecological processes, and biodiversity. Impaired streams exhibited compromised and contaminated water quality and altered physical habitats that could not support aquatic communities any longer. In addition, reduced water quantity in streams was found to decrease the connectivity and dispersal of biota, and impaired ecological processes could no longer process carbon and nutrients. Lastly, the report identified that the diversity of aquatic life had reduced (Clapcott et al., 2019). While a report on the state and trends in river health of the Bay of Plenty between 1992 and 2014 found that overall stream health had not changed considerably at each site over that period (Suren et al., 2017). Stream health conditions at sites located in catchments dominated in native or exotic plantation forests were the most pristine, compared to streams running through agricultural landcover, which displayed intermediate stream health conditions, and streams draining urban catchments, which exhibited the lowest stream health conditions (Suren et al., 2017).

1.4 Stream habitat assessments

Environmental attributes must first be assessed and monitored in order to report on the state and trends of the environment and evaluate the effectiveness of policy plans (Harding et al., 2009; Young et al., 2018). Environmental assessments involve the evaluation of indicators that are associated with ecosystem health. However, indicators are often affected by a broad range of ecosystem factors, making it a complex concept (Young et al., 2018). Therefore, defining the cause of ecosystem degradation and, subsequently, identifying appropriate mitigation strategies is difficult because of the complexities involved with the broad range of ecosystem factors affecting indicators (Young et al., 2018). For example, the Macroinvertebrate Community Index (MCI) is used to pinpoint the cause of degradation in an ecosystem (Young et al., 2018). However, as the MCI is affected by various stream habitat attributes, such as fine deposited sediment, algae bloom, or water temperature, it is challenging to ascribe any change in the MCI at a site to a specific cause (Young et al., 2018). To combat this issue, it is good practice to assess multiple stream habitat attributes, in addition to the annual collection of macroinvertebrates, as periodic stream habitat assessments support identifying changes in habitat conditions.

Stream habitat assessments can also help identify human activities that may have caused stream habitat conditions to decline (NIWA, 2019), increase the knowledge of stream ecosystem functions, and improve the success of management efforts (Barquín Ortiz and Martinez-Capel, 2011). Consequently, stream habitat assessments have become more popular over the years (Barquín Ortiz and Martinez-Capel, 2011) and have evolved from merely capturing stream habitat attributes to appreciating that stream habitat attributes are interrelated, driven predominantly by physical processes and altered by human activities (Gurnell et al., 2020).

Several reviews of stream habitat assessments have been conducted that captured the predominant characteristics of stream habitat assessments (Belletti et al., 2015; Fernandez et al., 2011). For example, Belletti et al. (2015) examined 73 stream habitat assessment methods in use worldwide and identified that all assessments evaluated stream channel attributes, most assessments included attributes of the riverbanks and riparian areas and three-quarters of reviewed stream assessments extended to the surrounding floodplains. The channel attributes most recorded included channel dimensions, substrate composition, flow types, and artificial features, whereas most commonly recorded attributes of the riparian zone and stream bank included characteristics of bank structures and the presence of artificial elements (Belletti et al., 2015). Regarding the assessment of floodplain characteristics, land use and the presence of river landforms were most commonly recorded, whereas larger-scale features, such as catchment or valley characteristics, were seldomly documented (Belletti et al., 2015). In New Zealand, current habitat assessments for wadeable streams focus on describing the riparian zone, channel morphology, habitat conditions and deposited fine sediments (Clapcott, 2015).

Regarding stream habitat assessment methods, in a review of river habitat characterisation methods used in Europe, North America and Australia, Fernandez et al. (2011) found that 60% of the revised field survey methods involved rapid assessment techniques (aka qualitative assessment methods), in which habitat attributes are assessed qualitatively. Rapid assessments often involve allocating scores related to the quality or quantity of certain stream habitat attributes (Fernandez et al., 2011). In contrast to qualitative assessment methods stand the quantitative survey techniques, within which quantitative measurements of habitat attributes are taken, using surveying equipment (Queirós et al., 2017).

1.5 Research objectives

Multiple stream habitat attributes are evaluated using qualitative and quantitative assessment methods during the current Bay of Plenty Natural Environmental Monitoring Network (BOP NERMN) stream habitat survey. These assessments are carried out by a team of different students each summer. Typically, qualitative assessments require less time to be carried out (shorter in-field assessment duration) than quantitative measurements. Consequently, using only qualitative assessment methods are likely to be more economical for the BOP NERMN stream habitat survey. However, qualitative assessments are thought to be more subject to inter-observer variability and thus may reduce the ability to detect changes or trends at monitored sites. Therefore, a consideration that should be addressed when designing a monitoring programme is whether the data quality derived from field surveys is credible. Data must meet quality standards to address the monitoring goals of a given project. For the BOP NERMN stream habitat monitoring, it is essential that the analysis of collected data can detect environmental trends. However, the year-to-year change of the surveying team personnel (inter-annual observer variability) could influence the monitoring results of the BOP NERMN stream habitat survey.

Therefore, the objectives of this study are:

- to investigate whether qualitative and quantitative metrics capture stream habitat attributes similarly.
- to examine in-field assessment duration of metrics to determine whether qualitative metrics are indeed assessed in a shorter time frame than quantitative metrics.
- to estimate inter-annual observer variability amongst qualitative and quantitative metric data over assessed years.
- to evaluate performance trade-offs between in-field assessment duration and inter-annual observer variability amongst qualitative and quantitative metrics in order to determine whether omitting quantitative metrics and using only qualitative metrics could be an appropriate option for creating a more economical BOP NERMN stream habitat survey.

1.6 Thesis structure

This thesis consists of four chapters as briefly outlined below:

Chapter 1: General introduction

In this chapter, stream habitat and factors influencing stream habitat conditions are discussed. Factors influencing stream habitat include the effects of channel morphology, riparian zone, and diverse human land uses. In particular, the effects of agricultural, urban and forestry land uses on stream habitat conditions are delineated. Because of the significant impacts of human activities, land uses need to be regulated in order to safeguard the environment and natural resources. As such, the environmental legislation framework of New Zealand is defined, and state and trends of New Zealand's stream habitat conditions are outlined. Additionally, a brief introduction to stream habitat assessments is provided. Lastly, the stimulus for this thesis is described, and the structure of this thesis is outlined.

Chapter 2: Correlations amongst qualitative and quantitative metrics

This chapter investigates whether qualitative and quantitative metrics capture stream habitat attributes similarly by assessing the relationship between the two approaches through Spearman rank correlation tests. If the correlation between these metrics is significant, assessing only qualitative metrics could potentially be an appropriate option for creating a more economical BOP NERMN stream habitat survey. However, to assess the appropriateness of metrics resulting in significant correlations, their performance trade-offs in relation to in-field assessment duration and inter-annual observer variability is investigated in Chapter 3.

Chapter 3: Choosing metrics based on in-field assessment duration and inter-annual observer variability

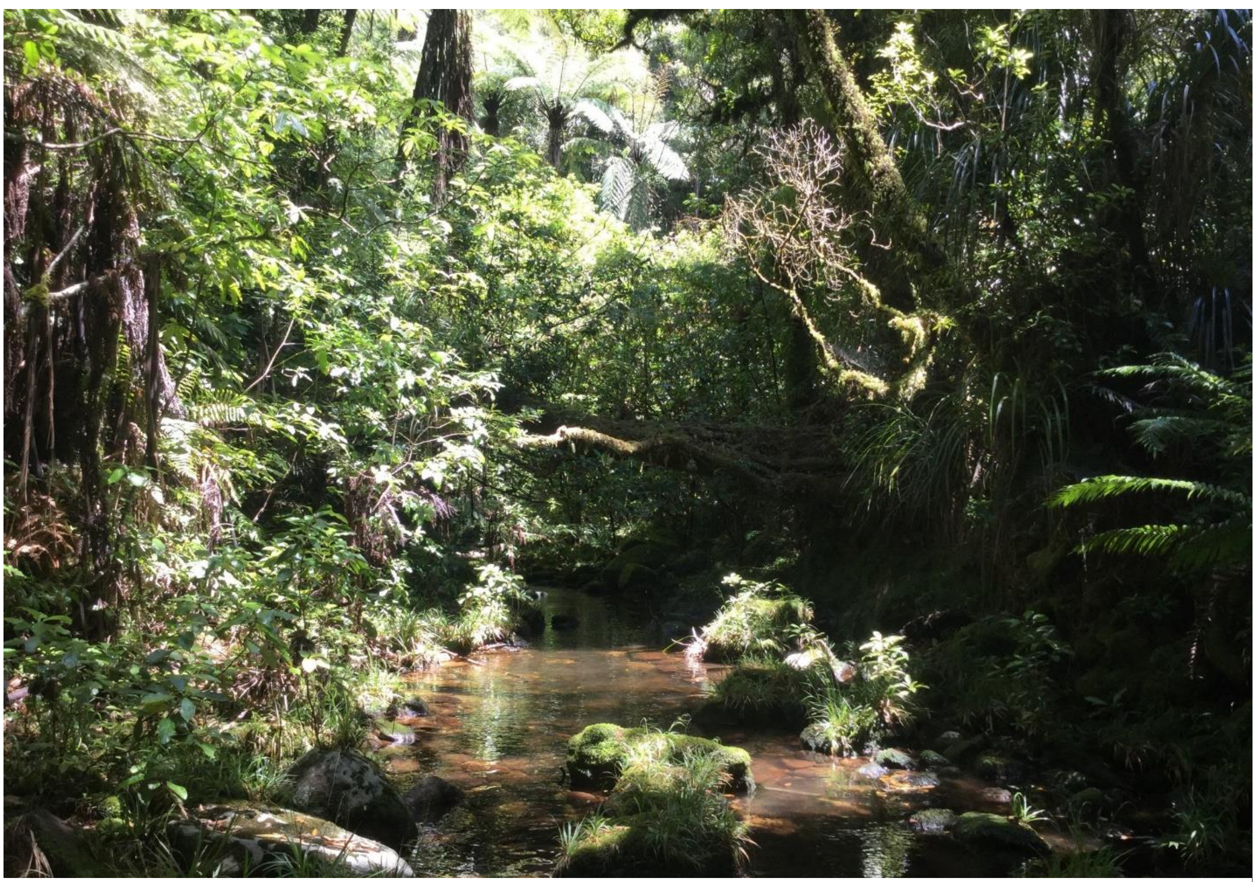
The focus of this chapter is threefold. First, the in-field assessment duration of qualitative and quantitative metrics is examined. Second, the degree of inter-annual observer variability amongst qualitative and quantitative metric data is investigated by comparing the metrics' percent coefficient of variation (CV) through Bayesian hierarchical linear models. Lastly, to assess the performance trade-offs between in-field assessment duration and inter-annual observer variability of qualitative and quantitative metrics, the previous findings of this study are combined into value models as per Gregory et al. (2012). The results of the value models illuminate the performance of metrics to determine whether omitting quantitative metrics and using only qualitative metrics is an appropriate option for creating a more economical BOP NERMN stream habitat survey.

Chapter 4: General discussion

This chapter provides an overview of the findings of this thesis. Management recommendations on ways to optimize the stream habitat survey for the BOP NERMN stream habitat monitoring are also outlined, and future research directions based on the findings from Chapters 2 and 3 are suggested.

Chapter 2 and 3 present research results and refer to data collected by different summer students each season between 2012/13 and 2019/20, except for in-field assessment duration data. To avoid duplication, a comprehensive description of field methods is given in Chapter 2. Therefore, when reading Chapter 3, refer back to Chapter 2 for the field method description.

Chapter 2:
Correlations amongst qualitative and quantitative metrics



Native forest site in the Bay of Plenty. Summer 2018/19

2.1 Introduction

The Resource Management Act (RMA) 1991 requires regional councils to uphold the sustainable management of their region's natural and physical resources (Ministry for the Environment, 1991). Under section 35 (2)(a), regional councils are required to monitor and report on the state and trends of the environment across their region and to document the effectiveness of their regional management plans (Ministry for the Environment, 1991). To meet the obligations under s35 (2)(a) of the RMA, the Bay of Plenty Regional Council established the Bay of Plenty Natural Environment Regional Monitoring Network (BOP NERMN) programme (Suren et al., 2017). Part of the BOP NERMN programme includes annual stream invertebrate and habitat monitoring of approximately 130 sites, which provides a representative overview of stream health across the Bay of Plenty (Suren et al., 2017).

Multiple habitat attributes are evaluated through qualitative and quantitative assessment methods during the annual BOP NERMN stream habitat monitoring by a team of students that changes year to year. The Rapid Habitat Assessment (RHA) developed by Clapcott (2013 and 2015) and the shuffle index (Clapcott et al., 2011) are the qualitative assessment methods used in the BOP NERMN stream habitat survey. The RHA is a qualitative assessment at the reach scale and is frequently applied by councils to support their duties under the RMA (Clapcott et al., 2020). The RHA provides a quick and easy site-based assessment of physical stream habitat conditions (Clapcott, 2013). Each metric included in the assessment was chosen because of its importance to stream biota and is scored on a numerical scale, with high scores indicating pristine or near-pristine stream habitat conditions. In contrast, low scores indicate degraded stream habitat conditions. All metrics are weighted equally in the summation of the total score (Clapcott, 2013). The total score, aka Habitat Quality Score, can be compared to habitat condition categories, where scores of 0-25, 26-50, 51-75, and 76-100 represent poor, fair, good and excellent stream habitat conditions, respectively (Clapcott et al., 2020). The shuffle index, the second qualitative assessment included in the BOP NERMN stream habitat survey, is an assessment of suspendable sediment, during which deposited sediment on and in between the first layer of the streambed is quantified (Clapcott et al., 2011). The quantitative metrics included in the BOP NERMN stream habitat survey are measurements related to the stream channel, streambank and riparian zone as well as substrate composition and deposited fine sediment.

Generally, qualitative and quantitative assessment methods are seen as two different paradigms (Sale et al., 2002; Steckler et al., 1992) and can be distinguished between the type of data collected (ordinary vs. interval data, respectively; Bazeley, 2004). Additionally, both methods comprise advantages and disadvantages (Table 2.1).

Qualitative research aims to provide contextual understanding (Steckler et al., 1992), by describing processes and identifying meaning (Sale et al., 2002). Qualitative stream assessments usually involve visual observations which are ascribed to categories (Harding et al., 2009) or scores related to the quality or quantity of certain habitat attributes (Fernandez et al., 2011). The advantages of qualitative assessments include instant contextual understanding of survey results (Steckler et al., 1992) and the rapid and more economic in-field assessment duration, compared to more intensive quantitative field surveys (Harding et al., 2009; Weiß et al., 2008). Because of the rapid assessment method, more study sites can be evaluated per unit of effort using a qualitative assessment approach (Hannaford and Resh, 1995). Additionally, information obtained by qualitative habitat surveys is easily understood and interpreted (Harding et al., 2009), as there is no complex statistical analysis required to obtain results, enabling rapid management decisions and report preparation (Hannaford and Resh, 1995). However, the data quality of qualitative evaluations is often lower, compared to quantitative measurements, because of the rapid assessment approach (Fernandez et al., 2011). There are also concerns about the subjective interpretation or inter-observer variability and the lack of precision of qualitative assessment methods, which limits their application in model analysis (McGinnity et al., 2005).

Within the quantitative paradigm, researchers use methods adapted from physical sciences, including formal instruments for data collection (Queirós et al., 2017) and sophisticated statistical data processing (Steckler et al., 1992). Quantitative assessment methods are commonly used for in-stream habitat survey protocols (Fernandez et al., 2011), in which actual measurements of habitat attributes are taken (Harding et al., 2009), often using surveying equipment (Queirós et al., 2017). The strengths of quantitative assessment methods lie within their more accurate (Queirós et al., 2017), precise (McGinnity et al., 2005; Weiß et al., 2008) and reliable (Steckler et al., 1992) data output while avoiding subjectivity, compared to qualitative assessment methods (Fernandez et al., 2011). Because quantitative data is interpreted during statistical analysis instead of in the field, as is the case with qualitative assessments, quantitative methods offer greater flexibility to interpret or re-interpret data (Kaufmann et al., 1999). Quantitative data can not only be used for purposes such as the development of predictive models (Queirós et al., 2017; Steckler et al., 1992) and indices (Fernandez et al., 2011), but quantitative data can also aid ground-truthing of other assessment methods, including qualitative assessments or data from aerial images (McGinnity et al., 2005). However, statistical analysis for interpretation can also be seen as a disadvantage, as stakeholders must be aware of the context of quantitative data and their analysis to be able to understand the results of quantitative assessments (Queirós et al., 2017; Steckler et al., 1992). Other downfalls of quantitative assessments are the elevated time commitment to collect data (Weiß et al., 2008; Kaufmann et al., 1999) and the requirements for surveying equipment (Queirós et al., 2017; Steckler et al., 1992).

Table 2.1: Summary of advantages and disadvantages of qualitative and quantitative stream habitat assessments found in the literature review.

Advantages	**Disadvantages**	**References**
	Qualitative stream habitat assessments	
Rapid assessment		Fernandez et al., 2011; Harding et al., 2009; Weiß et al., 2008
More economic		Hannaford and Resh, 1995; Harding et al., 2009; Weiß et al., 2008
More study sites can be visited per unit effort		Hannaford and Resh, 1995
No statistical analysis, instant results		Hannaford and Resh, 1995
Provides contextual understanding		Queirós et al., 2017; Steckler et al., 1992
Easily understood results		Harding et al., 2009
	Lower data quality	Fernandez et al., 2011
	Higher subjectivity	Fernandez et al., 2011
	Lower precision	Harding et al., 2009; McGinnity et al., 2005
	Difficult to generalise	Queirós et al., 2017
	Quantitative stream habitat assessments	
Higher accuracy		Harding et al., 2009; Queirós et al., 2017; Steckler et al., 1992
Higher precision		Harding et al., 2009; Kaufmann et al., 1999; McGinnity et al., 2005; Weiß et al., 2008
More reliable		Queirós et al., 2017; Steckler et al., 1992
Lower subjectivity		Fernandez et al., 2011; Harding et al., 2009
Data can be used for other purposes, i.e. predictive models or ground-truthing of other assessment methods		Fernandez et al., 2011; Harding et al., 2009; Kaufmann et al., 1999; McGinnity et al., 2005

Generalisable		Steckler et al., 1992
	Require more time	Kaufmann et al., 1999; Weiß et al., 2008
	Require instruments	Queirós et al., 2017; Steckler et al., 1992
	Require analysis for interpretation	Queirós et al., 2017; Steckler et al., 1992

During this chapter, I investigated whether qualitative and quantitative metrics captured stream habitat attributes similarly by assessing the relationship between the two approaches. Spearman rank correlation tests were performed between qualitative and corresponding[2] quantitative metrics of stream habitat attributes. If the correlation between these metrics was significant, omitting quantitative metrics and assessing only qualitative metrics could potentially be an appropriate option for creating a more economical BOP NERMN stream habitat survey. However, to decide whether quantitative metrics should be removed from the BOP NERMN stream habitat survey, in-field assessment duration and degree of inter-annual observer variability between qualitative and quantitative metric data were also considered and were examined in Chapter 3.

2.2 Methods

2.2.1 Study site

The Bay of Plenty is situated in the mid-region of the east coast of New Zealand's North Island, stretching approximately 1,231 km^2 from Cape Runaway in East Cape to Waihi Beach in the west (Suren et al., 2017). The Wairoa, Kaituna, Tarawera, Rangitāiki, Whakatāne, Waioeka, Motu and Raukokore are the eight largest rivers in the Bay of Plenty and hundreds of smaller rivers and streams run through the region (Ministry for the Environment, 2017). The Bay of Plenty's main water bodies combined carry more than 211.5 m^3/s of water. Rivers and streams in the region typically flow from their headwaters north towards the sea.

2.2.2 Field methods

Approximately[3] 130 wadeable stream[4] and river sites across the Bay of Plenty region were assessed annually as part of the BOP NERMN programme (Figure 2.1). Assessments were carried out by two

[2] Corresponding quantitative metrics refer to quantitative metrics corresponding to a qualitative metric, both of which describe the same habitat attribute. For example, the measurement of average sediment depth is a corresponding quantitative metric to the qualitative metric 'shuffle index'. Both metrics describe the stream habitat attribute of deposited fine sediment.
[3] The exact number of sites assessed each year varies, as some sites become inaccessible, dry up, etc.
[4] Hereafter, the term 'stream(s)' refers to streams and rivers, unless stated otherwise.

different students between November and February each year. I was part of the sampling team in the summers of 2018/19 and 2019/20.

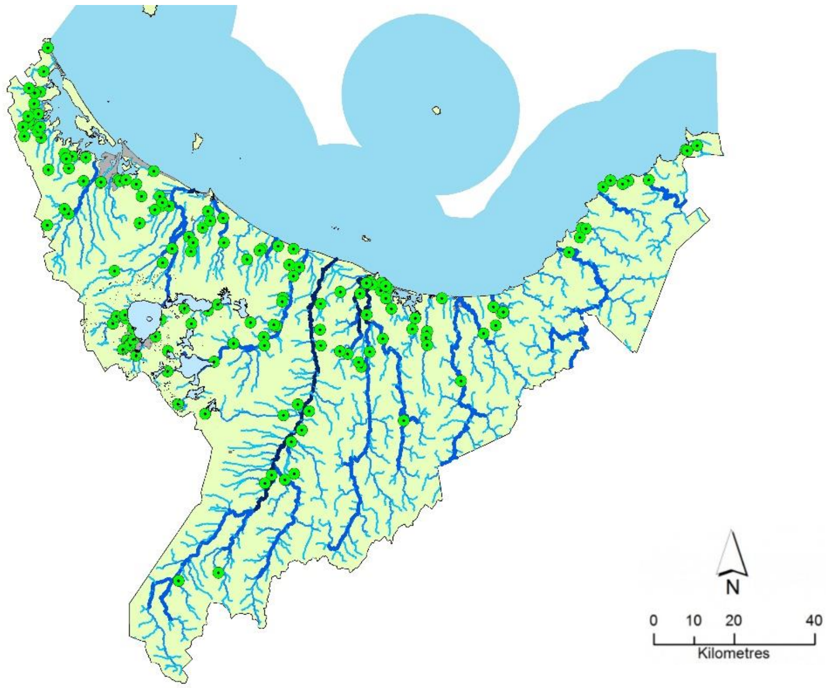


Figure 2.1: Overview of the Bay of Plenty Natural Environment Regional Monitoring Network (BOP NERMN) stream invertebrate and habitat monitoring sites (green dots). Image retrieved from Alastair Suren.

The BOP NERMN stream invertebrate and habitat monitoring sites were located within various land cover categories (agriculture, urban, native bush and exotic forest) and were marked with global positioning system (GPS) points to re-locate the exact sampling spots for assessments in subsequent years. The sampling sites were of two different streambed substrate types, namely hard and soft-bottomed substrates. A hard-bottomed (HB) stream was defined as one where > 50% of the streambed was classified as gravel or greater, whereas a soft-bottomed (SB) stream was defined as one where > 50% of the streambed was classified as sand or smaller.

During the BOP NERMN monitoring, an invertebrate sample was collected in accordance with the National Environmental Monitoring Standards for macroinvertebrate collection (NEMS, 2020), and a habitat assessment was conducted at each site annually. This thesis was focused on the stream habitat assessment only. Stream habitat attributes were evaluated using quantitative and qualitative metrics, some of which were assessed at five equally spaced transects placed up the study site, while others were assessed across the entire study reach (Table 2.2). All metrics were possible to obtain in shallow streams, i.e. less than 0.5 m deep, whereas in deeper (> 0.5 m), fast-flowing rivers, assessments were obtained from the wadeable portion of the rivers only.

Table 2.2: Overview of quantitative and qualitative metrics collected during the BOP NERMN stream habitat survey, some of which were assessed at each transect of a study site, while others were assessed across the entire study reach.

Metric type	**Metric**	**Measured where**
Quantitative	Wetted width	Each transect
	Bank to bank width	Each transect
	Stream depth	Three locations across each transect
	Sediment depth	Three locations across each transect
	Flow Velocity	Three locations across each transect
	Overhanging vegetation	Left and right bank at each transect
	Bank undercut	Left and right bank at each transect
	Flow heterogeneity	Entire reach
	Wolman pebble count (resulting in substrate index and substrate diversity)	Entire reach
Qualitative	Shuffle index	Three locations across each transect
	Rapid Habitat Assessment (RHA)	Entire reach

Measurements at transects

Upon arrival at a study site, the study reach was first split into five transects of equally spaced distances (yellow boxes and lines in Figure 2.2). In accordance to the suggestion that the length of a sampling site should be a function of the stream size (McGinnity et al., 2005), the length of the study reach was determined by multiplying the stream's width (in m) by 20. However, a minimum reach length of 40 m or a maximum reach length of 100 m applied to streams less than 2 m in width and rivers wider than 5 m, respectively. Where sampling sites were located close to a bridge or other artificial construction, assessments were conducted upstream from the structure.

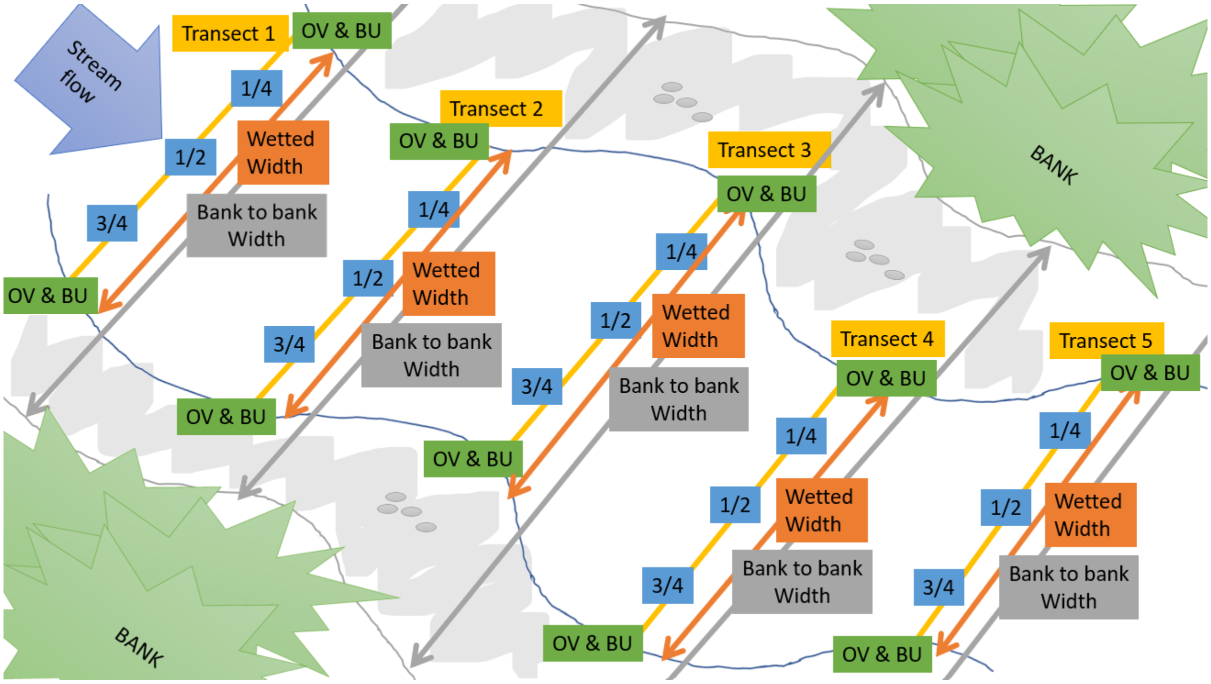


Figure 2.2: Overview of quantitative metrics, and the qualitative shuffle index, obtained during the BOP NERMN stream habitat survey at five equally spaced transects (marked as yellow boxes and lines) placed up a study reach. Quantitative metrics were taken a) across the entire transect (i.e. measurements of the wetted with and bank-to-bank width, marked as orange and grey boxes and lines, respectively), b) at both bank sides of the transect (i.e. overhanging vegetation (OV), bank undercut (BU), marked as green boxes), or c) at three points (¾, ½, ¼) across the transect (i.e. water depth, sediment depth, flow velocity and shuffle index, marked as blue boxes).

Each transect was further divided into ¼, ½, and ¾ points across the stream width, at which water depth, sediment depth, flow velocity and shuffle index were recorded (blue boxes in Figure 2.2). Water depth was recorded in that the vertical distance between the streambed and water surface was measured with a measurement staff. Then, the measurement staff was pushed into the streambed, and again the vertical distance between the streambed and water surface was recorded (second measurement). The water depth measurement was subtracted from the second measurement to obtain the sediment depth measurement.

In the summer seasons of 2012/13 and 2013/14, the velocity was measured using the ruler technique (Harding et al., 2009), while from the summer season 2014/15 onwards, a velocity meter (Global Water Flow Probe, 3.7-6’, FP111) was utilised.

To conduct the shuffle index, a white tile was placed 1 m downstream from where the shuffle index assessment was carried out. Then, the streambed substrate was disturbed by shuffling both feet in a left and right direction for 5 s, digging them into the substrate. Once the substrate had been disturbed, the resultant sediment plume was scored on a scale from 1 to 5, depending on the visibility of the tile and plume duration. The shuffle index has been part of the assessment since the summer season of 2016/17.

Additionally, overhanging vegetation and bank undercut were measured on both sides of the channel at each transect (green boxes in Figure 2.2). Overhanging vegetation was defined by its functional importance to stream biota, i.e. providing cover for fish. Therefore, vegetation higher than hip height when standing in the stream was not considered as overhanging vegetation for this assessment. Emergent or submerged aquatic vegetation growing in the stream was also not considered overhanging vegetation. Overhanging vegetation measurements were obtained as described in Figure 2.3.

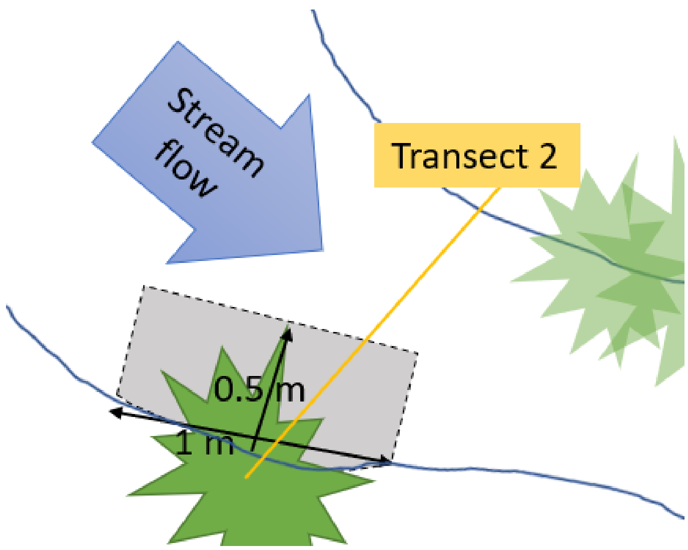


Figure 2.3: Explanation of overhanging vegetation measurement. Overhanging vegetation measurements were obtained by first measuring the longest part of the vegetation within a 1 m wide band perpendicular to the channel at a transect. An estimation was made as to how much of this area (grey area) was covered by overhanging vegetation. This percentage cover estimate was multiplied by the longest vegetation measurement. In this illustration, the longest part of the overhanging vegetation within one meter is 0.5 m, but the rest of the vegetation within the observed area is only covering around 35%. This means that the final overhanging vegetation measurement is 0.17 m^2 (0.5 m x 0.35 m).

For the bank undercut measurement, the longest part of the undercut within a 1 m wide band perpendicular to the channel at a transect was recorded. The wetted and bank-to-bank widths were obtained (orange and grey boxes and lines, respectively, in Figure 2.2). The bank-to-bank width was the width between the stream banks, i.e. the area that is wetted during an annual flood event. Signs of erosion and trapped debris in the riparian vegetation aided with identifying the bank-to-bank width. However, bank-to-bank width data were not available for statistical analysis.

Measurements at entire study reach

The Wolman pebble count (Clapcott et al., 2011) was used to quantify the substrate sizes of a streambed. This method is a quantitative assessment during which 100 substrate measurements were taken across the study reach, using the Wentworth scale of rock particle sizes. These recordings were converted to the quantitative metric ‘substrate index’ (Sub_Ind; Jowett, 1993), whereby:

$$Sub\_Ind = [(0.8 * bedrock) + (0.7 * boulder) + (0.6 * large\ cobbles) + (0.5 * small\ cobbles) + (0.4 * coarse\ gravel) + (0.3 * fine\ gravel) + (0.2 * sand) + (0.1 * silt/mud)]$$

The larger the value of the ‘substrate index’, the larger the substrate sizes present in the streambed, whereas the quantitative metric ‘substrate diversity’ was simply a measure of the number of different substrate size classes in each stream.

The flow heterogeneity of the reach was obtained by recording the dominant flow types (e.g., riffle, run, pool) up the stream every meter (e.g. 0 to 7 m: run, 7 to 10 m: riffle, etc.). Overall, the quantitative metric ‘flow heterogeneity’ was the sum of how many different flow types were present in the stream.

For the qualitative Rapid Habitat Assessment (RHA), either the RHA 2013 (Clapcott, 2013) or the RHA 2015 (Clapcott, 2015) protocols were applied. In the summer of 2013/14 and from 2017/18 onward, the RHA 2013 was used, whereas the RHA 2015 was used in 2015/16 and 2016/17. The switch back to the original RHA 2013 protocol was based on an analysis by Snelder et al. (2019) that showed that the RHA 2013 scores explained much more variability in the BOP NERMN macroinvertebrate data than the RHA 2015, for data collected between the 2012/13 summer and 2016/17 summer.

The RHA 2013 protocol included nine habitat metrics, whereas ten habitat metrics were assessed in the RHA 2015 protocol (Table 2.3; Clapcott, 2015). The 'RHA invertebrate habitat - diversity and abundance' metric in the RHA 2013 protocol was split into two separate metrics, assessing invertebrate habitat diversity and invertebrate habitat abundance separately. Likewise, the metric 'RHA fish cover - diversity and abundance' in the RHA 2013 was also split in the RHA 2015. The metric 'RHA channel alternation' was excluded from the RHA 2015 protocol. Additionally, the scoring scale was reduced from 20 to 1 as in the RHA 2013, to 10 to 1 in the RHA 2015.

Table 2.3: Differences between RHA metrics included in the 2013 and 2015 RHA protocols, resulting from protocol updates in 2015 (Clapcott, 2015). The 'RHA invertebrate habitat diversity and abundance' as well as the 'RHA fish cover diversity and abundance' were split into separate categories, and the 'RHA channel alternation' was removed.

Qualitative RHA metrics	**RHA 2013 protocol**	**RHA 2015 protocol**
RHA fine sediment deposition	✓	✓
RHA invertebrate habitat - diversity and abundance	✓	✗
RHA invertebrate habitat diversity	✗	✓
RHA invertebrate habitat abundance	✗	✓
RHA fish cover - diversity and abundance	✓	✗
RHA fish cover diversity	✗	✓
RHA fish cover abundance	✗	✓
RHA hydraulic heterogeneity	✓	✓
RHA bank stability	✓	✓
RHA bank vegetation	✓	✓
RHA riparian buffer (width)	✓	✓
RHA riparian shade	✓	✓
RHA channel alternation	✓	✗

2.2.3 Statistical analysis

The average and standard deviations of all quantitative habitat metrics were calculated, except for the metrics 'substrate index', 'substrate diversity' and 'flow heterogeneity', for which the measured values were used. Additionally, the coefficient of variation of flow velocity was calculated. These calculations resulted in the quantitative metrics of 'average bank undercut', 'average overhanging vegetation', 'average sediment depth', 'standard deviation stream depth', 'standard deviation flow velocity', 'standard deviation wetted width' and 'coefficient of variation flow velocity'.

Data of eight seasons (2012/13 through to 2019/20) were analysed. Across this timeframe, 935 stream observations were made, comprising streams of different substrate types and landcovers (Table 2.4). The exact number of sites assessed each year varies, as some sites become inaccessible, dry up, or due to other reasons.

Table 2.4: Summary of stream observations across eight seasons (2012/13 through to 2019/20), divided into substrate type (hard and soft-bottom streams) and landcovers (native forest, agriculture, exotic forest and urban). There are no hard-bottomed urban stream sites included in the BOP NERMN monitoring programme (NA).

Substrate type	Native Forest	Agriculture	Exotic Forest	Urban	Grand Total
Hard	286	80	76	NA	442
Soft	60	326	75	32	493
Grand Total	346	406	151	32	935

Qualitative vs. quantitative metrics

The Rapid Habitat Assessment (RHA) 2015 protocol scores were adjusted to match the RHA 2013 protocol scores. The adjustment was made by doubling the RHA 2015 scores, except for the 'RHA invertebrate habitat diversity' and 'RHA invertebrate habitat abundance' as well as 'RHA fish cover diversity' and 'RHA fish cover abundance'. These metrics were summed up to represent the combined 'RHA invertebrate habitat - diversity and abundance'[5] and 'RHA fish cover - diversity and abundance'[5] metrics of the RHA 2013. As the 'RHA channel alternation' was excluded from the RHA 2015, missing values were replaced with average values of the RHA 2013 protocol from across years at a site. Missing data of the metrics following metrics were substituted using the same technique; all RHA metrics for the season 2012/13 (as the RHA was introduced to the BOP NERMN stream habitat survey in 2013) and 'RHA riparian shade' (seasons 2015/16 to 2017/18). In addition, as the 'shuffle index' was introduced to the BOP NERMN stream habitat survey in 2016, missing data for the years prior to 2016 were substituted with average values of the years 2016 to 2020 at a site.

[5] For simplicity reasons, thereafter the metric 'RHA invertebrate habitat - diversity and abundance' is termed 'RHA invertebrate habitat'. Likewise, the metric 'RHA fish cover - diversity and abundance' is termed 'RHA fish cover' from here onwards.

Spearman rank correlation tests were performed between qualitative and corresponding[6] quantitative metrics to investigate whether metrics captured a given stream habitat attribute similarly (Table 2.5) using R (R Core Team, 2021).

The Spearman rank correlation does not carry any assumption about the distribution of the data and was the appropriate test here as none of the stream habitat metrics conformed to normality. The critical value for significant correlations between metrics was selected as $p < 0.05$. If the correlations between qualitative and corresponding quantitative metrics were significant, omitting quantitative metrics and assessing only qualitative metrics could potentially be an appropriate option for creating a more economical BOP NERMN stream habitat survey.

Table 2.5: Summary of investigated correlations between qualitative and corresponding[6] quantitative metrics to examine whether metrics captured the same stream habitat attribute. Stated abbreviations are used throughout this thesis.

Stream habitat attributes	**Qualitative metrics**	**Corresponding quantitative metrics**
Deposited fine sediment[7]	RHA fine sediment deposition (hard and soft-bottomed streams) (RHA_Sed(HBandSB)) and RHA fine sediment deposition (hard-bottomed streams only)[8] (RHA_Sed(HBonly)	Average sediment depth (Av_Sed_Dep)
		Substrate index (Sub_Ind)
		Substrate diversity (Sub_Div)
	Shuffle index (Shuffle)	Av_Sed_Dep
		Sub_Ind
		Sub_Div
Invertebrate habitat	RHA invertebrate habitat (RHA_Inv)	Av_Sed_Dep
		Sub_Ind

[6] The term 'corresponding quantitative metrics' refers to quantitative metrics corresponding to a qualitative metric, both of which describe the same habitat attribute. For example, 'average sediment depth' is a quantitative metric corresponding to the qualitative metric 'shuffle index'. Both metrics describe the stream habitat attribute of deposited fine sediment.

[7] Even though the RHA Fine sediment deposition and 'shuffle index' are both associated with the stream habitat attribute of deposited fine sediment, they were investigated as two separate qualitative metrics.

[8] Because Clapcott (2013) suggested that the RHA fine sediment deposition was only suitable for hard-bottomed (HB) streams, I created an additional dataset that excluded soft-bottomed (SB) streams

		Sub_Div
		Flow heterogeneity (Flo_Het)
Fish cover	RHA fish cover (RHA_Fish)	Sub_Ind
		Sub_Div
		Average overhanging vegetation (Av_Veg)
		Average bank undercut (Av_Bank_U)
		Flo_Het
Hydraulic/flow heterogeneity	RHA hydraulic heterogeneity (RHA_Hyd)	Standard deviation wetted width (Std_WetWid)
		Standard deviation stream depth (Std_Dep)
		Standard deviation flow velocity (Std_Vel)
		Coefficient of variation flow velocity (CV_Vel)
		Flo_Het
Bank stability/erosion	RHA bank stability (RHA_Bank_Stab)	Av_Bank_U
Riparian vegetation	RHA bank vegetation (RHA_Bank_Veg)	Av_Veg
Riparian buffer	RHA riparian buffer	There was no data available to validate the RHA riparian buffer attribute
Riparian shading	RHA riparian shade	Av_Veg
Channel alternation	RHA channel alternation (RHA_Ch_Alt)	Std_WetWid
		Std_Dep
		Std_Vel
		CV_Vel
		Flo_Het

The following is a brief overview of the rationale for assigning qualitative and corresponding quantitative metrics to the stream attributes in Table 2.5.

- Deposited fine sediment: As the ‘substrate index’ and ‘substrate diversity’ provide estimates of the substrate sizes (Jowett, 1993), they can also be used to indicate fine sediment deposition (Clapcott et al., 2011).

- Invertebrate habitat: Many macroinvertebrates live between and on stream substrate, and thus, fine deposited sediment decreases habitat availability (NIWA, 2019). Therefore, quantitative metrics of 'average sediment depth', 'substrate index' and 'substrate diversity' can be used to assess invertebrate habitat. Additionally, flow types, which are constituted of combinations of water depth, velocity and substrate composition, define different types of habitat diversity (Parsons et al., 2002). Riffles, runs, pools, rapids, backwaters and cascades count to the common flow types found in streams (Parsons et al., 2002), to which many aquatic biotas have specific preferences (Jowett and Richardson, 1990). Therefore, a measure of the different flow types, herein referred to as 'flow heterogeneity', can also be used to assess invertebrate habitat.
- Fish cover: Not only do bank undercuts, overhanging vegetation and stream substrate offer shelter for fish (Kaufmann et al., 1999), but fish also rely on a variety of flow types for refuge, as well as for ambushing prey such as drifting invertebrates (Gebrekiros, 2016). Therefore, fish cover can be quantitatively assessed via the 'substrate index' and 'substrate diversity' as well as through the 'average bank undercut', 'average overhanging vegetation' and 'flow heterogeneity' metrics.
- Hydraulic heterogeneity and channel alternation[9]: Both of these attributes are related to a stream's morphology. Stream morphology, in turn, is related to stream features such as width, flow velocity and flow types (Harding et al., 2009). Variations of these features can indicate how complex a stream channel is, whereas uniformity in these features is related to a modified channel with low hydraulic heterogeneity. Therefore, statistical measures of variability, including the standard deviation or coefficient of variation, of width and flow velocity measurements and records of 'flow heterogeneity' offer good tools to examine the hydraulic heterogeneity and channel alternation of a stream.
- Bank stability: As unstable banks may display bank undercuttings (Harding et al., 2009), measurements of bank undercuts may be used to assess bank stability.
- Riparian shading: A dense riparian and overhanging stream vegetation provides shading above the stream (Elosegi et al., 2011), and thus measurements of overhanging vegetation can be used to assess stream shading.

[9] Even though these stream habitat attributes are separate attributes in the RHA, the rationale underlying the allocation of corresponding quantitative metrics are the same and therefore are discussed together, for simplicity.

- Riparian vegetation: Measurements of overhanging vegetation can also be used to assess bank vegetation.

2.3 Results

The Spearman rank correlation analyses revealed that all qualitative metrics, apart from the 'RHA riparian shade', were significantly correlated to at least one of their corresponding quantitative metrics. However, the models explained little of the variation within the data, ranging from the highest correlation being $r_s(935) = -0.39$ for 'shuffle index' vs. 'substrate index' (Figure 2.7) and lowest being $r_s(935) = -0.089$ 'RHA bank stability' vs. 'average bank undercut' (Figure 2.21). Although because of the large sample size, the weak significant correlations were statistically representative of the population and did not originate from chance factors (Fowler et al., 2013).

In regards to the stream habitat attribute deposited fine sediment, the 'RHA fine sediment deposition' showed a significant positive correlation to 'substrate index' and a significant negative correlation to 'average sediment depth' ($r_s(935) = 0.28$, $p < 0.001$, Figure 2.4 and $r_s(935) = -0.29$, $p < 0.001$, Figure 2.5; respectively) for both, hard and soft-bottomed streams (HB and SB data set). For hard-bottomed streams only (HB only dataset), there was also a significant correlation between the 'RHA fine sediment deposition' and 'substrate index' ($r_s(442) = 0.18$, $p < 0.001$, Figure 2.6). Of interest was the observation that there was a stronger correlation between the 'RHA fine sediment deposition' and 'substrate index' for the combined HB and SB dataset than the dataset of HB streams only, even though Clapcott et al. (2013) stated that the application of the RHA should be limited to hard-bottomed streams. Additionally, the 'shuffle index' showed a significant negative correlation to the 'substrate index' and a significant positive correlation to 'average sediment depth' ($r_s(935) = -0.39$, $p < 0.001$, Figure 2.7 and $r_s(935) = 0.27$, $p < 0.001$, Figure 2.8; respectively). Neither of the qualitative metrics related to the stream habitat attribute of deposited fine sediment (i.e. 'RHA fine sediment deposition' and 'shuffle index') were significantly correlated to the quantitative metric 'substrate diversity'.

Concerning habitat attributes associated with in-stream habitat, the 'RHA invertebrate habitat' was significantly correlated to all of its corresponding quantitative metrics (i.e. 'substrate index': $r_s(935) = 0.3$, $p < 0.001$, Figure 2.9; 'substrate diversity': $r_s(935) = 0.16$, $p < 0.001$, Figure 2.10; 'average sediment depth': $r_s(935) = -0.22$, $p < 0.001$, Figure 2.11 and 'flow heterogeneity': $r_s(935) = 0.26$, $p < 0.001$, Figure 2.12), while the 'RHA fish cover' was significantly correlated to two of its five corresponding quantitative metrics (i.e. 'substrate index': $r_s(935) = 0.13$, $p < 0.001$, Figure 2.13 and 'average bank undercut': $r_s(935) = 0.13$, $p < 0.001$, Figure 2.14). The corresponding quantitative metrics that did not result in significant correlations to the 'RHA fish cover' were the 'substrate index', 'average overhanging vegetation' and 'flow heterogeneity'.

With reference to attributes related to channel morphology, the qualitative metrics of 'RHA hydraulic heterogeneity' and the 'RHA channel alternation' shared the same corresponding quantitative metrics. However, the correlation analyses revealed that the 'RHA hydraulic heterogeneity' was significantly correlated to all of its correlating quantitative metrics (i.e. 'standard deviation wetted width': $r_s(935) = 0.21$, $p < 0.001$, Figure 2.15; 'standard deviation flow velocity': $r_s(935) = 0.14$, $p < 0.001$, Figure 2.16; 'coefficient of variation flow velocity': $r_s(935) = 0.15$, $p < 0.001$, Figure 2.17; 'standard deviation stream depth': $r_s(935) = 0.15$, $p < 0.001$, Figure 2.18 and 'flow heterogeneity': $r_s(935) = 0.38$, $p < 0.001$, Figure 2.19), whereas the 'RHA channel alternation' was significantly correlated to only one correlating quantitative metric (i.e. 'flow heterogeneity': $r_s(935) = 0.17$, $p < 0.001$, Figure 2.20).

Regarding stream habitat attributes associated with the riparian zone vegetation, there was a significant negative correlation between the 'RHA bank stability' and 'average bank undercut' ($r_s(935) = -0.073$, $p = 0.027$, Figure 2.21) as well as a significant negative correlation between the 'RHA riparian bank vegetation' and 'average overhanging vegetation' ($r_s(935) = -0.089$, $p = 0.0072$, Figure 2.22). No significant correlation was found between the 'RHA riparian shade' and 'average overhanging vegetation.

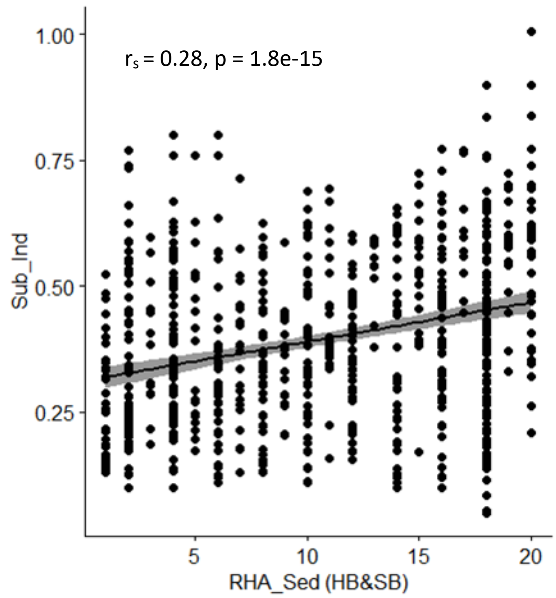


Figure 2.4: Spearman rank correlation between the 'RHA fine sediment deposition' for hard and soft-bottomed streams (RHA_Sed (HBandSB)) and 'substrate index' (Sub_Ind)

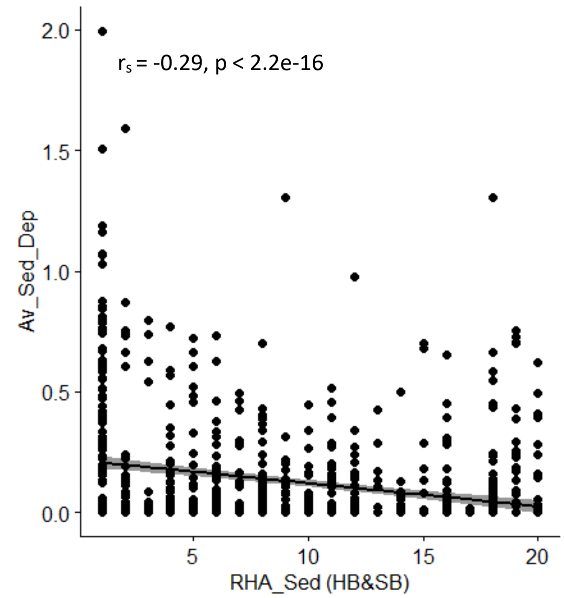


Figure 2.5: Spearman rank correlation between the 'RHA fine sediment deposition' for hard and soft-bottomed streams (RHA_Sed (HBandSB)) and 'average sediment depth' (Av_Sed_Dep, y-axis scale: 0.5 = 5 cm)

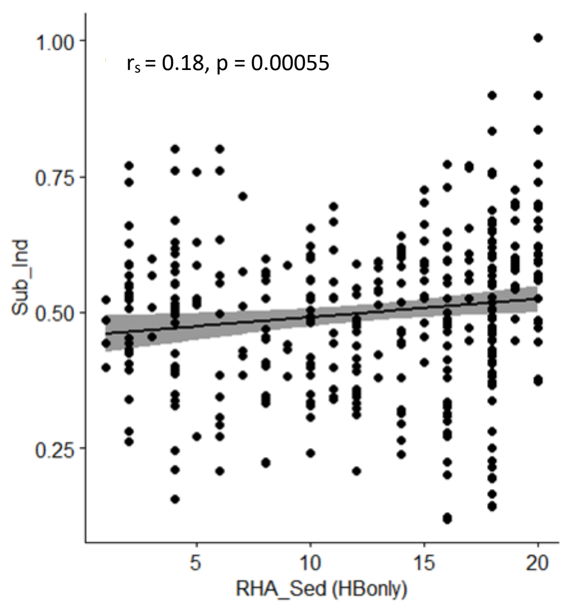


Figure 2.6: Spearman rank correlation between the 'RHA fine sediment deposition' only including hard-bottomed streams (RHA_Sed (HBonly) and 'substrate index' (Sub_Ind)

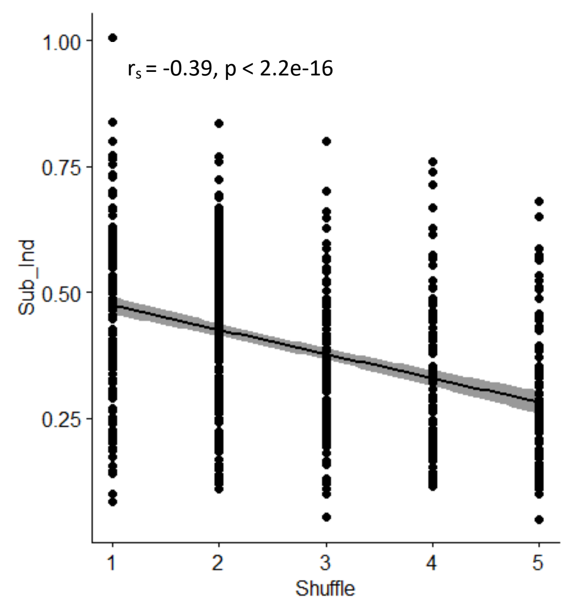


Figure 2.7: Spearman rank correlation between the 'shuffle index' (Shuffle) and 'substrate index' (Sub_Ind)

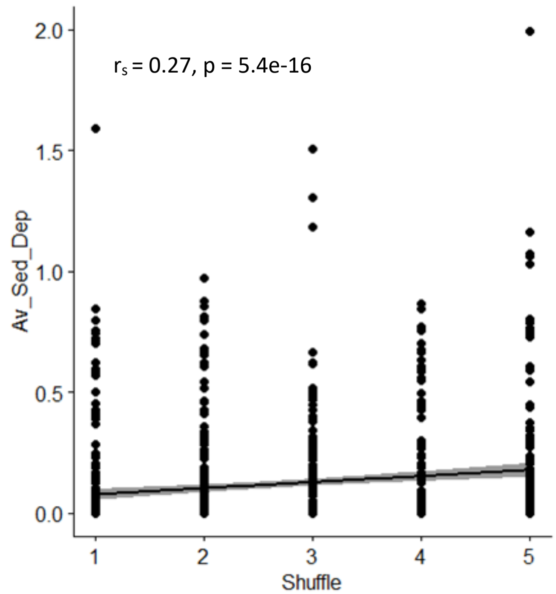


Figure 2.8: Spearman rank correlation between the 'shuffle index' (Shuffle) and 'average sediment depth' (Av_Sed_Dep, y-axis scale: 0.5 = 5cm)

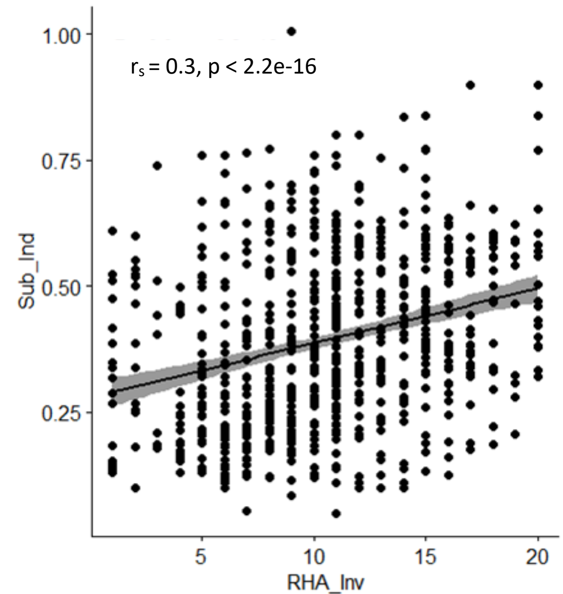


Figure 2.9: Spearman rank correlation between the 'RHA invertebrate habitat' (RHA_Inv) and 'substrate index' (Sub_Ind)

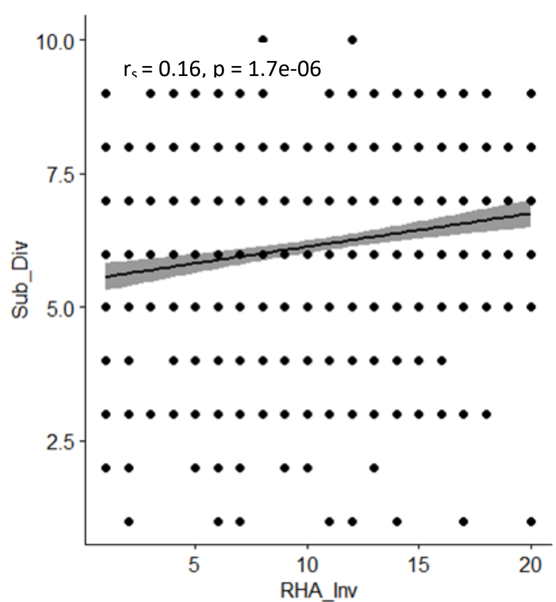


Figure 2.10: Spearman rank correlation between the 'RHA invertebrate habitat' (RHA_Inv) and 'substrate diversity' (Sub_Div)

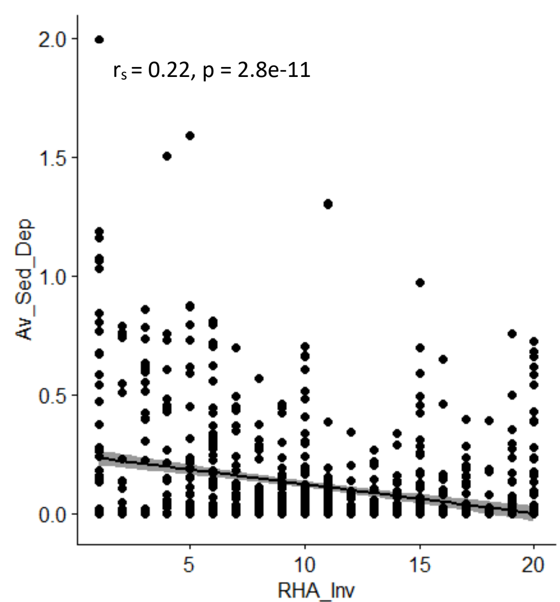


Figure 2.11: Spearman rank correlation between the 'RHA invertebrate habitat' (RHA_Inv) and 'average sediment depth' (Av_Sed_Dep, y-axis scale: 0.5 = 5cm)

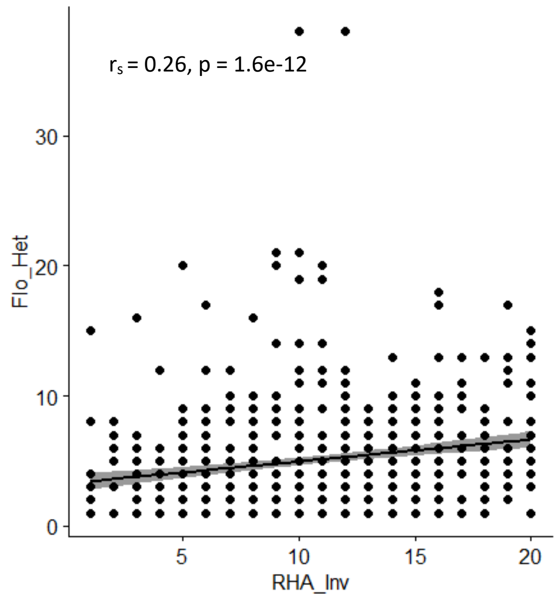


Figure 2.12: Spearman rank correlation between the ‘RHA invertebrate habitat’ (RHA_Inv) and ‘flow heterogeneity’ (Flo_Het)

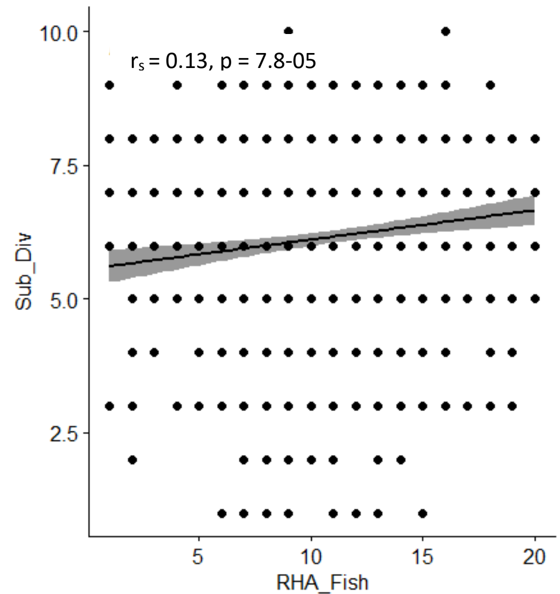


Figure 2.13: Spearman rank correlation between the ‘RHA fish cover’ (RHA_Fish) and ‘substrate diversity’ (Sub_Div)

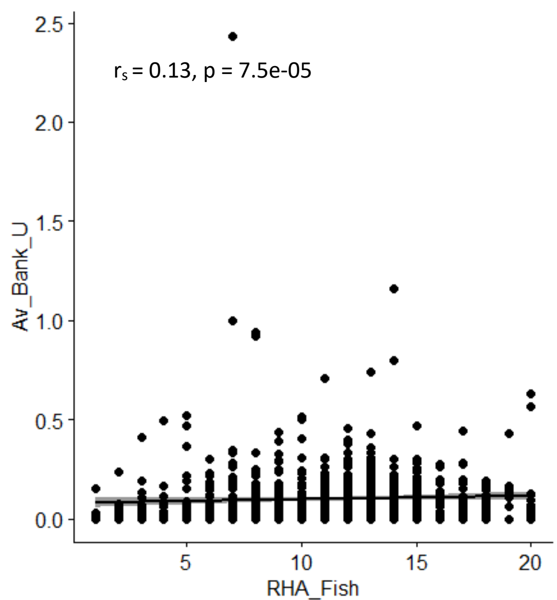


Figure 2.14: Spearman rank correlation between the ‘RHA fish cover’ (RHA_Fish) and ‘average bank undercut’ (Av_Bank_U, y-axis scale: 0.5 = 5cm)

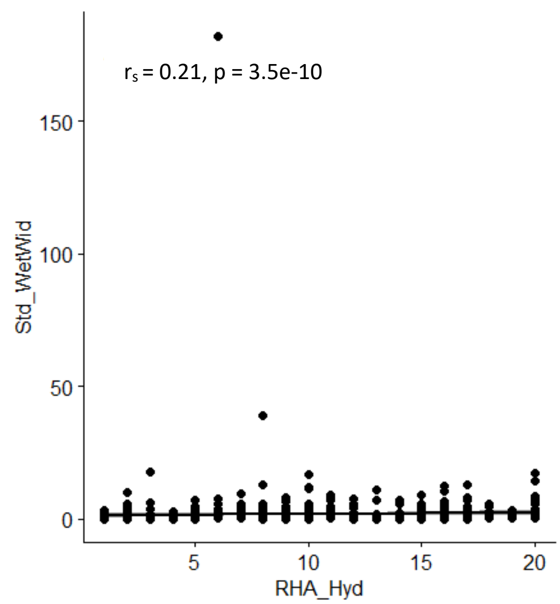


Figure 2.15: Spearman rank correlation between the ‘RHA hydraulic heterogeneity’ (RHA_Hyd) and ‘standard deviation wetted width’ (Std_WetWid)

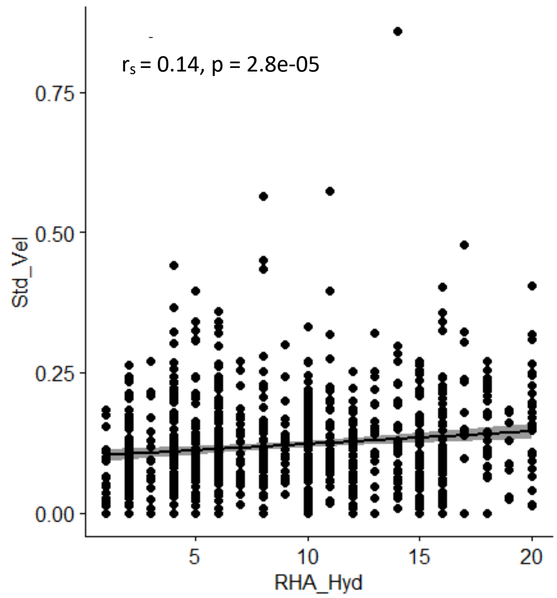


Figure 2.16: Spearman rank correlation between the 'RHA hydraulic heterogeneity' (RHA_Hyd) and 'standard deviation flow velocity' (Std_Vel)

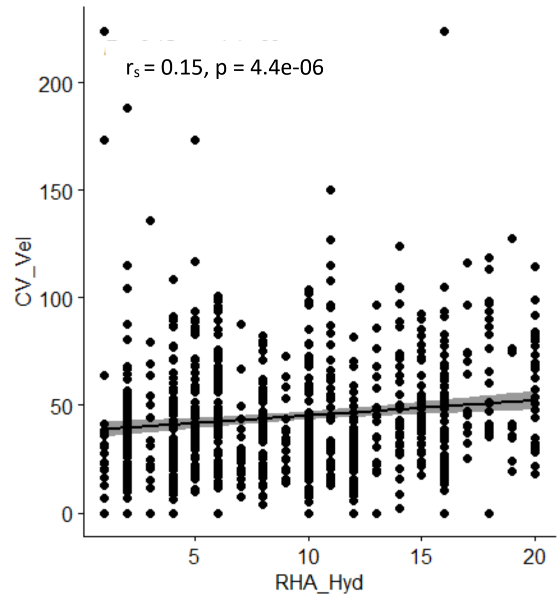


Figure 2.17: Spearman rank correlation between the 'RHA hydraulic heterogeneity' (RHA_Hyd) and 'coefficient of variation flow velocity' (CV_Vel)

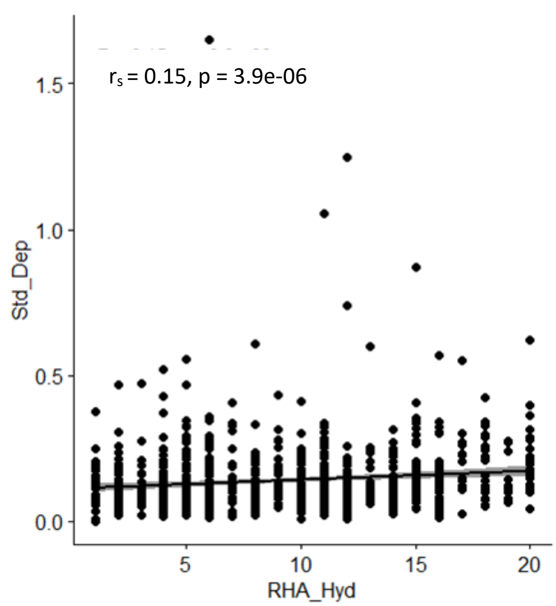


Figure 2.18: Spearman rank correlation between the 'RHA hydraulic heterogeneity' (RHA_Hyd) and 'standard deviation stream depth' (Std_Dep)

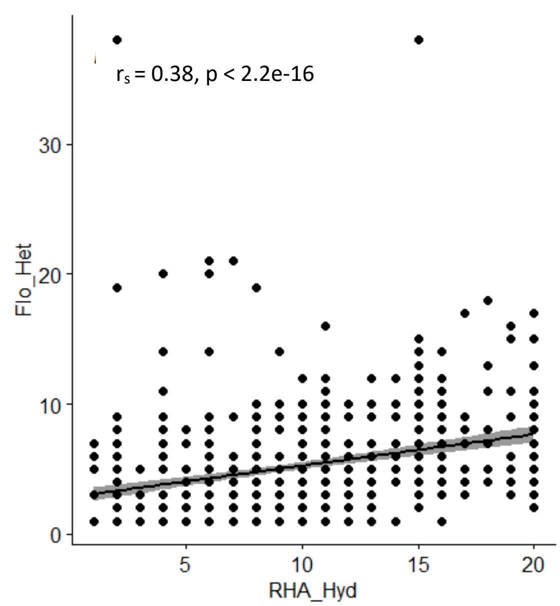


Figure 2.19: Spearman rank correlation between the 'RHA hydraulic heterogeneity' (RHA_Hyd) and 'flow heterogeneity' (Flo_Het)

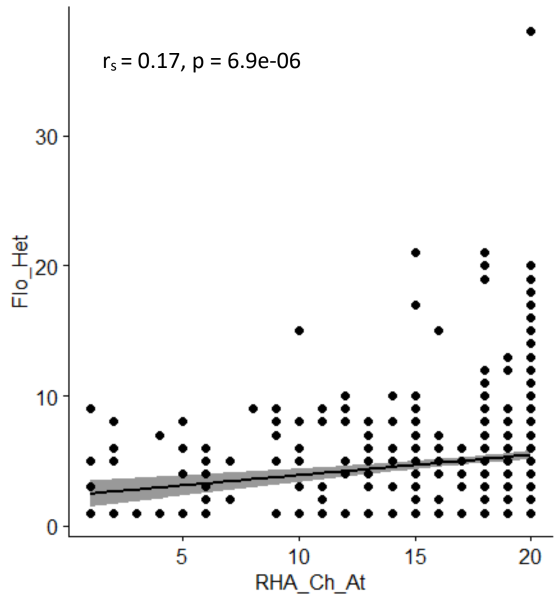


Figure 2.20: Spearman rank correlation between the 'RHA channel alternation' (RHA_Ch_Alt) and 'flow heterogeneity' (Flo_Het)

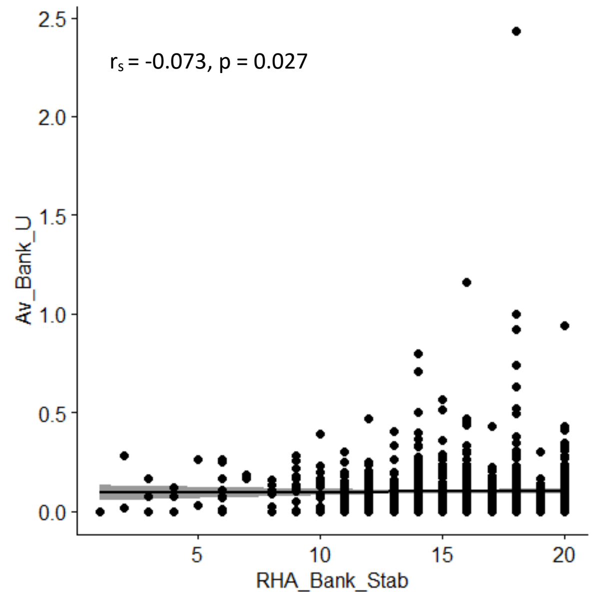


Figure 2.21: Spearman rank correlation between the 'RHA bank stability' (RHA_Bank_Stab) and 'average bank undercut' (Av_Bank_U, y-axis scale: 0.5 = 5cm)

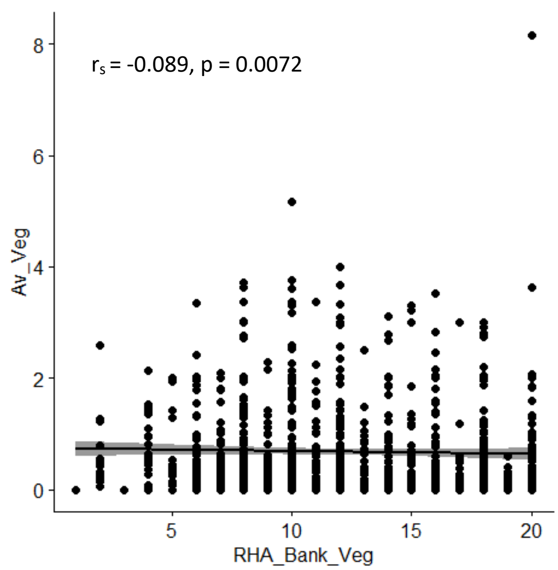


Figure 2.22: Spearman rank correlation between the RHA Riparian vegetation (RHA_Hyd) and 'average overhanging vegetation' (Av_Veg, y-axis scale: 2 = 20cm^2)

2.4 Discussion

The correlation analyses revealed that all qualitative metrics, apart from the 'RHA riparian shade', were significantly correlated to at least one of their corresponding quantitative metrics (Table 2.6). One aspect to note here is that even though the qualitative metric 'RHA bank vegetation' and the quantitative metric 'average overhanging vegetation' were significantly correlated, the negative direction of the relationship was unanticipated. It was expected that longer overhanging vegetation would score higher 'RHA bank vegetation' scores, and therefore would show a positive relationship towards each other. Consequently, the significant correlation between 'RHA bank vegetation' and 'average overhanging vegetation' was excluded from further analysis. All other directions of significant correlations were plausible.

The correlation analyses results suggest that qualitative metrics of the stream habitat attributes of deposited fine sediment, invertebrate habitat, fish cover, hydraulic/flow heterogeneity, bank stability and channel alternation have the potential to be appropriate options to assess stream habitat attributes more economically and thus corresponding quantitative metrics could theoretically be removed from the BOP NERMN stream habitat survey. Further, it was inconclusive whether the measurement of bank-to-bank width should be excluded from the BOP NERMN survey, as no data were available for statistical analysis.

Table 2.6: Summary of investigated correlations between qualitative and corresponding quantitative metrics of stream habitat attributes evaluated during the BOP NERMN stream habitat survey. ~~Strikethrough~~ indicates no significant correlation was found between metrics (*for the stream attribute riparian vegetation, there was a significant negative correlation between metrics, but this relationship was regarded as irrational, so those variables were excluded from further analysis). See Table 2.5 for abbreviations.

Stream habitat attribute	Qualitative metrics	Corresponding quantitative metrics
Deposited fine sediment	RHA _Sed (HBandSB) and RHA _Sed (HB only)	Av_Sed_Dep
		Sub_Ind
		~~Sub_Div~~
	Shuffle	Av_Sed_Dep
		Sub_Ind
		~~Sub_Div~~
Invertebrate habitat	RHA_Inv	Av_Sed_Dep
		Sub_Ind

		Sub_Div
		Flo_Het
Fish cover	RHA_Fish	~~Sub_Ind~~
		Sub_Div
		~~Av_Veg~~
		Av_Bank_U
		~~Flo_Het~~
Hydraulic/flow heterogeneity	RHA_Hyd	Std_WetWid
		Std_Dep
		Std_Vel
		CV_Vel
		Flo_Het
Bank stability/erosion	RHA_Bank_Stab	Av_Bank_U
~~Riparian vegetation~~ *	~~RHA_Bank_Veg~~	~~Av_Veg~~
~~Riparian Shading~~	~~RHA riparian shade~~	~~Av_Veg~~
Channel alternation	RHA_Ch_Alt	~~Std_WetWid~~
		~~Std_Dep~~
		~~Std_Vel~~
		~~CV_Vel~~
		Flo_Het

Assessment approaches, i.e. what to measure (which attributes) and how (qualitative vs. quantitative methods), should be aligned with monitoring goals (Bazeley, 2004; Harding et al., 2009; NIWA, 2019). Additionally, data derived from field surveys must be credible and of the right quality to identify changes in monitored systems adequately (Bazeley, 2004). The BOP NERMN stream habitat monitoring goal is to collect data for State of the Environment (SoE) reports and inform the effectiveness of regional plans in the most economical way. Therefore, it is essential that analysis of collected data can detect environmental trends. However, as the surveying team's annual change could influence the monitoring results of the BOP NERMN stream habitat survey, the subsequent chapter investigated the degree of inter-annual observer variability amongst quantitative and qualitative metric data. Further, as the BOP NERMN stream habitat data collection should be as economical as possible, the in-field assessment duration amongst quantitative and qualitative metrics was also examined.

The main complication of the investigation in this chapter was that qualitative and quantitative assessments did not always relate to the exact same habitat metric. For example, within the RHA protocol, the attribute of bank vegetation is targeted towards assessing the riparian vegetation composition and ground cover. In contrast, the corresponding quantitative metric of overhanging vegetation only encompassed measurements of vegetation overhanging the stream and did not consider any other aspects of the riparian vegetation. Previous research on comparing the accuracy between different assessment types (i.e. qualitative vs. quantitative) has been carried out before, with the awareness that different metrics are, to some degree, assessing different aspects of attributes (Morrison, 2016). Consequently, different conclusions might be drawn from qualitative and quantitative data (Morrison, 2016). Therefore, instead of comparing the accuracy of different assessment types, in this study, the Spearman rank correlation analyses were used to assess whether qualitative and corresponding quantitative metrics captured the same stream habitat attribute similarly. Nevertheless, this investigation should provide some foundation for deciding whether it is appropriate to use only qualitative metrics when evaluating stream habitat attributes during the BOP NERMN stream habitat survey, rather than choosing metrics on an arbitrary basis.

Chapter 3:
Choosing metrics based on in-field assessment duration and inter-annual observer variability

Native forest site in the Bay of Plenty. Summer 2018/19

3.1 Introduction

Monitoring programmes are usually initiated to inform decision-making, as was the Bay of Plenty Natural Environment Regional Monitoring Network (BOP NERMN) monitoring programme. However, raw data resulting from surveys must first be analysed and interpreted to improve and inform understandings of the question under investigation (Houston and Hiederer, 2009). Effective decision-making depends on reliable data that detects changes within ecosystems and on being able to identify the causes of those changes (Ferretti, 2011). Therefore, monitoring methods must be standardised, objective and repeatable to achieve reliable data (Kolada et al., 2014). The ease with which decisions can be made greatly depends on how data are collected, managed, and interpreted (Houston and Hiederer, 2009).

When data is of poor quality, even the most advanced statistical analyses are ineffective (Ferretti, 2011), and there will be limited ability to detect changes or trends (Morrison, 2016). Thus, poor-quality data can reduce the robustness and reliability of assessments (Goodenough et al., 2020). Consequently, the decision-making process can be severely compromised, leading to inappropriate decisions (Ferretti, 2011). For example, if a waterbody is incorrectly classified into a higher ecological status, its degradation may remain unnoticed (Kolada et al., 2014). The quality of data, information and decisions are all interconnected in the environmental decision-making process (Ferretti, 2011). Decisions must be defensible, and therefore the better the data quality, the better the defensibility of a decision (Ferretti, 2011). Sound science should form the foundation of environmental policy, and therefore good quality data are essential (Ferretti, 2011; Kaufmann et al., 1999).

Data quality may be influenced by uncertainties around spatial, temporal and methodological variations (Carstensen and Lindegarth, 2016; Ferretti, 2011; Morrison, 2016; Roper et al., 2002). For example, spatial variation could include random variation among samples taken at different locations within a study reach, while temporal variation could originate from random inter-annual or seasonal differences (Carstensen and Lindegarth, 2016). Uncertainty of methodological variations can stem from sampling and statistical methods as well as from different monitoring equipment and field staff involved in the monitoring programme (Carstensen and Lindegarth, 2016). Monitoring programs running over an extended time period usually involve different sampling equipment, changing processing practices for samples and data, as well as different observers (Carstensen and Lindegarth, 2016). Field staff changes are unavoidable in long-term monitoring programmes, and the effects on data quality vary (Carstensen and Lindegarth, 2016). For example, inter-observer variability amongst different technicians analysing hydrochemistry is relatively low compared to taxonomists analysing macroinvertebrates or phytoplankton samples (Carstensen and Lindegarth, 2016). Harding et al. (2009) suggested that inter-observer variability is the most probable source of error in stream habitat monitoring. Further, Roper et al. (2002) evaluated physical stream habitat attributes evaluated in stream monitoring and found that the most critiqued aspect of stream monitoring was the fact that different field staff, applying the same protocol, frequently arrived at different results.

Sources of inter-observer variability commonly include field staff traits such as mental and physical fatigue, level of enthusiasm (Morrison, 2016) and personal biases, i.e. subjectivity (Cherrill, 2016; Hogle et al., 1993; Morrison, 2016). The mental state of an observer can be influenced by team members or supervisors, who may distract or stimulate the observer (Morrison, 2016). The state of mind of field staff should not be underestimated, as all monitoring methods depend on their integrity and attitude (Morrison, 2016). Observations of field staff are also influenced by their background and level of expertise (Goodenough et al., 2020; Harding et al., 2009). Experienced observers with an understanding of the environment of their region can effectively identify stream habitat conditions (Hannaford et al., 1997). However, temporary staff and volunteering groups are increasingly employed to conduct monitoring (Hannaford et al., 1997), which may lead to increased inter-observer variability over time. Other concerns regarding inter-observer variability include inconsistent use of monitoring protocols (Goodenough et al., 2020; Hogle et al., 1993) and the difficulty to detect habitat change using stream habitat attributes (Roper et al., 2002). At the core of the issues discussed above lies the variability related to the surveying approaches of stream habitat attributes and the effects of this variation on the conclusion relative to these attributes (Roper et al., 2002).

To identify which surveying approaches are the most effective for a monitoring programme, the value model, a tool used in structured decision-making (SDM), can form the basis for enlightening the decision (Gregory et al., 2012). In SDM, considered alternatives (e.g. different surveying approaches) and their consequences as well as decision-making objectives (e.g. low inter-observer variability) are carefully defined. Through the value model, decision-making objectives, as well as trade-offs in performances of alternatives, are summarised and evaluated (Gregory et al., 2012). Ultimately, through the value model, an overall ‘performance score’ for each alternative included in the model is calculated, considering the relative importance-weight placed on each objective

(Gregory et al., 2012). The higher the 'performance score' of an alternative, the better the alternative's performance is in achieving the objectives, with the 'performance score' scale ranging from 0 to 1.

Multiple stream habitat attributes are evaluated using qualitative and quantitative assessment methods during the current BOP NERMN stream habitat survey. These assessments are carried out by a team of different students each summer. Typically, qualitative assessments require less time to be carried out (shorter in-field assessment duration) than quantitative measurements. Thus, using only qualitative assessment methods is likely to be more economical for the BOP NERMN stream habitat survey. However, qualitative assessments are thought to be more subject to inter-observer variability, and thus, the year-to-year change of the surveying team personnel (inter-annual observer variability) could influence the BOP NERMN stream habitat monitoring results. In this thesis, I aimed to determine whether omitting quantitative metrics and using qualitative metrics only could be an appropriate option for creating a more economical BOP NERMN stream habitat survey, taking into account the metrics' in-field assessment duration and inter-annual observer variability.

In this chapter, I a) examined the in-field assessment duration of assessing qualitative and quantitative metrics, b) investigated the degree of inter-annual observer variability amongst qualitative and quantitative metric data and c) combined my findings into value models to assess the performance trade-offs of qualitative and quantitative metrics, in relation to in-field assessment duration and inter-annual observer variability. The resulting 'performance scores' illuminated the performance of metrics and therefore informed the decision of whether it would be appropriate to remove quantitative metrics from the BOP NERMN stream habitat survey.

a) In-field assessment duration

To examine whether qualitative assessments were indeed more economical in terms of in-field assessment duration, I followed the BOP NERMN stream habitat assessment protocol but, additionally I recorded the time it took to collect qualitative and quantitative metric data.

b) Inter-annual observer variability

I investigated the degree of inter-annual observer variability amongst quantitative and qualitative metric data. However, data variability over time can originate from a mixture of inter-annual observer variability and actual changes in stream habitat attributes. Therefore, to minimise the effects of actual changes of habitat attributes over time and to maximise the ability to detect only inter-annual observer variability, this analysis was restricted to habitat data collected from native forest sites only. Limiting the analysis to native forest site data was based on the assumption that stream habitat attributes were not likely to change considerably in these sites over time – at least not over the seven years over which the RHA was applied (Alastair Suren, personal communication, 28 May 2021). Thus, if I detected a high degree of variability in data at native forest sites over time, then I assumed that this was due to inter-annual observer variability. However, if the degree of data variability was low in native forest sites over time, it was likely that inter-annual observer variability was low. Although it is somewhat problematic to test this assumption, the examination of photographs from two native forest sites (Figure 3.1 to Figure 3.6) clearly shows what appeared to be very similar stream habitat conditions over time.

For example, the fallen log at site BOP_NERMN_006, circled in red, did not move between 2015/16 and 2018/19 and all stream habitat conditions assessed during the BOP NERMN stream habitat survey appeared stable (Figure 3.1 to Figure 3.3).

Figure 3.1: Native forest site BOP_NERMN_006 in 2015/16. The log, circled in red, did not move between 2015/16 and 2018/19.

Figure 3.2: Native forest site BOP_NERMN_006 in 2016/17. The log, circled in red, did not move between 2015/16 and 2018/19.

Figure 3.3: Native forest site BOP_NERMN_006 in 2018/19. The log, circled in red, did not move between 2015/16 and 2018/19.

At site BOP_NERMN_11003, there was no visible change in the road cutting, circled in red, or any other stream habitat conditions assessed between the BOP NERMN stream habitat survey seasons 2014/15 and 2019/20 (Figure 3.4 to Figure 3.6).

Figure 3.4: Native forest site BOP_NERMN_110003 in 2014/15. No apparent change happened to the road cutting, circled in red, between 2014/15 and 2019/20.

Figure 3.5: Native forest site BOP_NERMN_110003 in 2018/19. No apparent change happened to the road cutting, circled in red, between 2014/15 and 2019/20.

Figure 3.6: Native forest site BOP_NERMN_110003 in 2019/20. No apparent change happened to the road cutting, circled in red, between 2014/15 and 2019/20.

I calculated the percent coefficient of variation (CV) for each metric at each native sampling site to compare the degree of inter-annual observer variability of qualitative and quantitative metric data. Then, I fitted a Bayesian hierarchical linear model for each habitat attribute to compare the CV values among the metrics for that attribute. I chose the Bayesian hierarchical linear model to include the different sampling sites as random factors.

c) Value models

Lastly, I created a value model for each stream habitat attribute as per Gregory et al. (2012) to assess the performance trade-offs between in-field assessment duration and inter-annual observer variability of qualitative and quantitative metrics of the given stream habitat attribute. Herein I included the in-field assessment duration and the degree of inter-annual observer variability as objectives, and as alternatives, I included the metrics found to have significant correlations in Chapter 2.

3.2 Methods

For details of the study site and field methods see 'Methods' in Chapter 2.

3.2.1 Statistical analysis

In-field assessment duration

A field technician and I measured the time taken to collect qualitative and quantitative metric data (in-field assessment duration) in order to determine whether qualitative assessments are indeed carried out in a shorter time frame. In-field assessment duration was documented at selected streams that were part of the BOP NERMN monitoring sites. We sampled 17 streams, as further outlined in Table 3.1. However, due to time constraints, we were not able to obtain data from each substrate type and landcover.

Table 3.1: Number of sites at which time data was obtained to investigate in-field assessment duration amongst metrics, including streams of different substrate types (hard and soft bottomed streams), streams running through different landcovers (native and exotic forests, agriculture and urban) and small streams (< 3rd order) and larger rivers (5th order or more). Due to time constraints, it was not possible to obtain data from each substrate type and landcover (NA).

Substrate type	**Native Forest**		**Agriculture**	**Exotic Forest**	**Urban**	**Grand Total**
	River	**Stream**	**Stream**	**Stream**	**Stream**	
Hard bottomed	2	3	NA	5	NA	10
Soft bottomed	NA	4	1	1	1	7
Grand Total	2	7	1	6	1	17

We timed (in minutes) the collection of all qualitative metric data such as the RHA metrics and the 'shuffle index', as well as quantitative metric data of 'flow heterogeneity', the Wolman pebble count (from which the 'substrate index' and 'substrate diversity' are derived) and measurements at each transect (wetted width, flood bank width, stream depth, sediment depth, flow velocity, overhanging vegetation and bank undercut). Note that it was not feasible to time the collection of individual quantitative metric data at each transect, so these transect-based measurements represented the total time for these seven metrics. However, it is expected that if these transect-based measurements are retained in the BOP NERMN stream habitat survey, they will remain a collective unit rather than collecting only one or two of these metrics. Finally, we were not able to measure the in-field assessment duration of the 'shuffle index' at river sites as it was not possible to conduct the 'shuffle index' at deep and/or very fast flowing sites.

Inter-annual observer variability

For each stream habitat attribute, I fitted a Bayesian hierarchical linear model to data of metrics found to have significant correlations in Chapter 2 (Table 3.2).

Table 3.2: Metrics of significant correlations found in Chapter 2 and list of missing data. A Bayesian hierarchical linear model was fitted to data of the qualitative metric and corresponding quantitative metrics of each stream habitat attribute to compare their degree of inter-annual observer variability. If no data were available for a metric in a particular season (Missing data), data for that season were also excluded from the corresponding metrics. Stated abbreviations are used throughout this thesis.

<table>
<tr><th>Stream habitat attributes</th><th>Qualitative metric</th><th>Corresponding quantitative metrics</th><th>Missing data</th></tr>
<tr><td rowspan="4">Deposited fine sediment[10]</td><td rowspan="2">RHA fine sediment deposition (HB and SB data set; RHA_Sed)</td><td>Substrate index (Sub_Ind)</td><td rowspan="2">Substrate index: 2014/15</td></tr>
<tr><td>Average sediment depth (Av_Sed_Dep)</td></tr>
<tr><td rowspan="2">Shuffle index (Shuffle)</td><td>Sub_Ind</td><td rowspan="2">Shuffle index: 2013/14 - 2016/17
Substrate index: 2014/15</td></tr>
<tr><td>Av_Sed_Dep</td></tr>
<tr><td rowspan="4">Invertebrate habitat</td><td rowspan="4">RHA invertebrate habitat (RHA_Inv)</td><td>Av_Sed_Dep</td><td rowspan="4">Substrate index: 2014/15</td></tr>
<tr><td>Sub_Ind</td></tr>
<tr><td>Substrate diversity (Sub_Div)</td></tr>
<tr><td>Flow heterogeneity (Flo_Het)</td></tr>
</table>

[10] Even though the RHA fine sediment deposition and 'shuffle index' are both associated with the stream habitat attribute of deposited fine sediment, they were investigated as two separate qualitative metrics.

Fish cover	RHA fish cover (RHA_Fish)	Sub_Div	
		Average bank undercut (Av_Bank_U)	
Hydraulic/flow heterogeneity	RHA hydraulic heterogeneity (RHA_Hyd)	Standard deviation wetted width (Std_WetWid)	Flow heterogeneity: 2016/17
		Standard deviation flow velocity (Std_Vel)	
		Coefficient of variation flow velocity (CV_Vel)	
		Standard deviation stream depth (Std_Dep)	
		Flo_Het	
Bank stability/ erosion	RHA bank stability (RHA_Bank_Stab)	Av_Bank_U	
Channel alternation	RHA channel alternation (RHA_Ch_Alt)	Flo_Het	RHA channel alternation: 2015/16 – 2017/18 Flow heterogeneity: 2016/17

First, to compare the degree of inter-annual observer variability in qualitative and quantitative metric data, I calculated the percent coefficient of variation (CV) for each metric at all BOP NERMN native forest sites over time. If no data were available for a metric in a particular season, I also excluded data for that season from the corresponding metrics (Table 3.2). For example, as there were no data available for the quantitative metric 'substrate index' for the 2014/15 season, I also omitted this season's data from the qualitative metric 'RHA fine sediment deposition' and the quantitative metric 'average sediment depth'. Analysing data of the same seasons avoided the CV comparisons being confounded by differences in variability amongst seasons.

Then, for each stream habitat attribute investigated, I created a long format datasheet, including columns of the native forest sites (RC_SID), qualitative and quantitative metrics of the given stream habitat attribute (Metric) and their CV values (CV; Table 3.3).

Table 3.3: Excerpt of a long format datasheet for the deposited fine sediment attribute, including the metrics 'RHA find sediment deposition' (RHA_Sed), 'substrate index' (Sub_Ind) and 'average sediment depth' (Av_Sed_Dep). Columns include the native forest sites (RC_SID), qualitative and quantitative metrics of the said stream habitat attribute (Metric) and their CV values (CV). Note the vast majority of data is excluded from this excerpt (…).

RC_SID	**Metric**	**CV**
BOP_NERMN_006	RHA_Sed	59.93
…	…	…
BOP_NERMN_006	Sub_Ind	26.27
…	…	…
BOP_NERMN_006	Av_Sed_Dep	0
…	…	…

For each habitat attribute, I then modelled the data to compare the CV values among the metrics for that attribute. To do this, I fitted Bayesian hierarchical linear models where the CV value was a function of the dummy-coded factor METRIC using the package *brms* (Bürkner, 2017) in R (R Core Team, 2021), following the tutorial by Franke and Roettger (2019). I included the factor METRIC as a fixed effect and the factor SAMPLING SITES (RC_SID) as a random effect in the models. I chose to apply Bayesian hierarchical linear models, as they enabled the inclusion of random effects for SAMPLING SITES. The following is an example of the R code to compare the CV values of the qualitative metric 'RHA fine sediment deposition' and quantitative metrics of 'substrate index' and 'average sediment depth':

Sed_brm <- brm(CV~Metric+(1|RC_SID),data = Sed).

The resulting CV estimates of metrics were assigned to different degrees of inter-annual observer variabilities, ranging from low (0-20), moderate (20.1-30), to high (30.1-100). This scale was adapted from Roper et al. (2002), who stated that when the sampling variance of a metric is less than 20 percent of the total variability, the metric can be classified as reliable.

Value models

Lastly, I created a value model for each stream habitat attribute as per Gregory et al. (2012) to assess the performance trade-offs of qualitative and quantitative metrics in relation to in-field assessment duration and inter-annual observer variability. (Table 3.4). The key objectives for selecting metrics for the BOP NERMN stream habitat survey were their low inter-annual observer variability and short in-field assessment duration. Performance measures for these two objectives were CV estimates and minutes, respectively. The metrics of a stream habitat attribute found to have significant correlations in Chapter 2 (Table 3.2) were listed as the alternatives in the value model. To select appropriate metrics for the BOP NERMN stream habitat survey, Alastair Suren addressed trade-offs among alternatives by assigning importance-weights to the objectives (personal communication, 28 May 2021). In other words, more importance was placed on metrics having lower inter-annual observer variability (weight: 0.6) than being assessed in a short timeframe (weight: 0.4). As the performance measures were reported in different units, their values were normalised using the following formula:

$$Normalised\ score = \frac{score - worst\ score}{best\ score - worst\ score}$$

Lastly, the normalised scores were combined in the value model by calculating the 'performance score' of the alternatives, whereby:

$$performance\ score = (0.6 * normlised\ score\ of\ CV) + (0.4 * normalised\ score\ of\ min)$$

Table 3.4: Example of a value model comparing alternative metrics for assessing the stream habitat attribute deposited fine sediment (i.e. the qualitative metric 'RHA fine sediment deposition' (RHA_Sed) and the quantitative metrics of 'substrate index' (Sub_Ind) and 'average sediment depth' (Av_Sed_Dep)). Alastair Suren assigned importance-weights. Objectives of the alternative metrics were low inter-annual observer variability and short in-field assessment duration, with performance measures of CV estimates and minutes, respectively. In this example, the best metric to assess the stream habitat attribute fine deposited sediment, with respect to the objectives, would be the 'substrate index', as it scored the highest 'performance score' of 0.93.

Objectives	**Performance measures**	**Alternatives**			**Importance - weights**	**Normalised scores**		
		RHA_Sed	Sub_Ind	Av_Sed_Dep		RHA_Sed	Sub_Ind	Av_Sed_Dep
low inter-annual observer variability	CV estimate	47.93	20.8	137.89	0.6	0.77	1.00	0.00
short in-field assessment duration	minutes	4	5	10	0.4	1.00	0.83	0.00
					Performance score	0.86	0.93	0.00

3.3 Results

3.3.1 In-field assessment duration

On average, it took 4 and 5 minutes to collect qualitative metric data of the RHA protocol and the ‘shuffle index’, respectively, across the entire study reach (Table 3.5). The time to collect quantitative metric data of ‘flow heterogeneity’, ‘substrate index’ and ‘substrate diversity’ (derived from the Wolman pebble count) and the quantitative transect measurements (i.e. measurements of wetted stream width, bank to bank width, overhanging vegetation, bank undercut, stream and sediment depth as well as flow velocity) across the entire study reach was on average 2, 5 and 10 minutes, respectively. Thus, it took on average 9 minutes to collect qualitative metric data, 17 minutes to collect quantitative metric data and 28 minutes to collect all metric data included in the BOP NERMN stream habitat survey.

3.3.2 Inter-annual observer variability

Data of the quantitative metric ‘substrate diversity’, as well as the qualitative metrics ‘shuffle index’, ‘RHA bank stability’ and ‘RHA channel alternation’, showed low inter-annual observer variability (CV estimate ≤ 20%) and data of the quantitative metric ‘substrate index’ showed moderate inter-annual observer variability (CV estimate ≤ 30%; Table 3.5). Data of the remaining 11 metrics showed high inter-annual observer variability (CV estimate > 30%). Data of most metrics (69%) had high inter-annual observer variability levels, and inter-annual observer variability was independent of the metric type, i.e. quantitative vs. qualitative.

3.3.3 Value models

More than half (56%) of all metrics scored relatively high ‘performance scores’ (≥ 0.8), with a few exceptions including the qualitative metric ‘RHA channel alternation’ (‘performance score’ of 0.6) and the quantitative metrics of ‘average sediment depth’, ‘average bank undercut’, ‘standard deviation wetted width’ and ‘standard deviation stream depth’ (‘performance score’ ≤ 0.56; Table 3.5). The metrics with ‘performance scores’ ≥ 0.8 are the quantitative metrics of ‘substrate index’, ‘substrate diversity’ and ‘flow heterogeneity’ as well as the qualitative metrics of ‘RHA fine sediment deposition’, ‘RHA invertebrate habitat’, ‘RHA fish cover’, ‘RHA bank stability’, ‘RHA hydraulic heterogeneity’ and the ‘shuffle index’.The quantitative metric of ‘flow heterogeneity’ scored high ‘performance scores’ in describing the stream habitat attributes of invertebrate habitat and hydraulic/flow heterogeneity (0.85 and 1, respectively), but scored low in describing the stream habitat attribute of channel alternation (0.4 'performance score').

Table 3.5: Average in-field assessment duration (IFAD) in minutes (min) to assess quantitative (QT) and qualitative (QL) stream habitat metrics at one site. The in-field assessment duration of the seven quantitative transect measurements (i.e. wetted stream width, bank to bank width, overhanging vegetation, bank undercut, stream and sediment depth as well as flow velocity) could not be taken separately from each other. Further, the table shows the distribution of the metrics' coefficient of variation in percent estimate (CV estimate %) and its lower and upper limits of the 95 percent confidence interval (L 95% CI, U 95% CI, respectively) over time at native forest sites. Data of metrics are considered to have low degree of inter-annual observer variability (IAOV) if the CV estimate is ≤ 20%, moderate IAOV if the CV estimate is ≤ 30% and high IAOV if the CV estimate is > 30 %. Lastly, the results of the value tables are presented in the form of the 'performance scores'. Metrics are considered to have high 'performance scores' if their score is ≥ 0.8. For metric abbreviations, see Table 3.1

Stream habitat attribute	Metric	Metric type	IFAD (min)	Coefficient of variation			Degree of IAOV	Per-formance scores
				CV estimate %	L 95% CI	U 95% CI		
Deposited fine sediment	RHA_Sed	QL	4	47.93	33.49	62.02	High	0.86
	Sub_Ind	QT	5	20.8	1.53	40.68	Moderate	0.93
	Av_Sed_Dep	QT	10	137.89	118.04	158.04	High	0.00
	Shuffle	QL	5	18.43	1.21	35.73	Low	1.00
	Sub_Ind	QT	5	20.56	3.48	38.38	Moderate	0.98
	Av_Sed_Dep	QT	10	86.43	73.89	98.5	High	0.00
Invertebrate habitat	RHA_Inv	QL	4	34.92	20.49	49.55	High	0.81
	Sub_Ind	QT	5	20.5	5.96	34.35	Moderate	0.84
	Sub_Div	QT	5	18.41	3.58	33.22	Low	0.85
	Av_Sed_Dep	QT	10	127.02	116.56	137.15	High	0.00
	Flo_Het	QT	2	44.93	30.66	59.58	High	0.85
Fish cover	RHA_Fish	QL	4	38.34	28.11	48.6	High	0.87
	Sub_Div	QT	5	18.21	8.23	28.2	Low	0.93
	Av_Bank_U	QT	10	109.62	101.97	117.04	High	0.00
Hydraulic/ Flow hetero-geneity	RHA_Hyd	QL	4	47.25	36.5	57.71	High	0.90
	Std_WetWid	QT	10	79.54	68.9	90.28	High	0.00
	Std_Vel	QT	10	50.15	39.76	60.87	High	0.54
	CV_Vel	QT	10	49.37	41.23	57.44	High	0.56
	Std_Dep	QT	10	59.77	49.12	70	High	0.37

	Flo_Het	QT	2	47.14	36.51	57.99	High	1.00
Bank stability/ erosion	RHA_Bank_Stab	QL	4	18.57	7.3	29.95	Low	1.00
	Av_Bank_U	QT	10	109.56	100.88	118.24	High	0.00
Channel Alternation	RHA_Ch_Alt	QL	4	12.25	4.45	19.94	Low	0.60
	Flo_Het	QT	2	44.82	38.92	50.84	High	0.40

3.4 Discussion

The investigation of in-field assessment duration between qualitative and quantitative metrics did not show consistent results. Across the entire study reach, the time required to record data of the quantitative metric ‘flow heterogeneity’ was the lowest, with an average of 2 minutes, whereas data of the quantitative transect measurements took the greatest amount of time to obtain, with an average of 10 minutes per stream. Consequently, this finding is partially consistent with Kaufmann et al.’s (1999) and Weiss et al.’s (2008) statements that quantitative assessments require more time than qualitative assessments. Further, the average in-field assessment duration for the quantitative Wolman pebble count (resulting in the ‘substrate index’ and ‘substrate diversity’) and the qualitative ‘shuffle index’ was 5 minutes, whereas for the qualitative RHA protocol the average in-field assessment duration was 4 minutes.

It must be noted here that the field technician and I were familiar with the assessment methods and sampling sites. Therefore, it is likely that in-field assessment duration would be longer for personnel unfamiliar with the study sites or assessment methods. Further, we conducted the RHA after we collected quantitative data. Consequently, we had already created a “mental image” of the stream habitat before commencing the RHA, which could have reduced the RHA in-field assessment duration. Ideally, we should have carried out the RHA before obtaining other measurements, but the methodology behind the data collection process meant that that was not done.

Further, the investigation into the degree of inter-annual observer variability amongst qualitative and quantitative metric data indicated that most metrics’ data had high inter-annual observer variability, independent of their metric type. The four metrics with low inter-annual observer variability included the qualitative ‘RHA channel alternation’, ‘shuffle index’, ‘RHA bank stability’ and the quantitative metric ‘substrate diversity’, followed by the quantitative metric ‘substrate index’, which was the only metric scoring moderate inter-annual observer variability (Table 3.5).

In the following paragraphs, I discuss my findings concerning inter-annual observer variability and compare my results to other authors' findings regarding observer variability of qualitative and quantitative metrics (Table 3.6). However, these comparisons should be viewed with caution because reviewed studies were not restricted to native forest sites, and observer variability was investigated amongst multiple observers carrying out assessments in the same year (intra-annual observer variability), contrasting this study, where inter-annual observer variability was examined. Additionally, study details varied amongst the reviewed literature (see Appendix).

The low inter-annual observer variability of the 'RHA channel alternation' metric data was probably because native forest sites are typically unmodified. As such, all observers were consistently scoring this metric high each year. Hannaford and Resh (1995) found high intra-annual observer variability, and Clapcott (2015) found moderate intra-annual observer variability in qualitative assessments of channel alternation, but these studies were not restricted to native forest sites.

The low inter-annual observer variability of the 'shuffle index' data could have been influenced by the fact that the 'shuffle index' is scored on a scale from 1 to 5, compared to the RHA scoring of 1 to 20 or the continuous measurements of quantitative metrics. For example, Roper and Scarnecchia (1995) suggested that when observers were required to distinguish between fewer habitat types, the classification of stream habitats would show less variability. This statement could also explain the low inter-annual observer variability of the 'shuffle index' data, as the 'shuffle index' has the narrowest assessment scale amongst metrics included in the BOP NERMN stream habitat survey. Therefore, observers were required to distinguish between fewer scoring scales.

Interestingly, three out of the five metrics scoring low to moderate inter-annual observer variability (i.e. the 'shuffle index', 'substrate index' and 'substrate diversity') were associated with substrate size. However, data of the remaining qualitative metric 'RHA fine sediment deposition', also associated with substrate size, had high inter-annual observer variability. Two of the reviewed studies found high intra-annual observer variability in the qualitative assessment of deposited fine sediment, i.e. substrate embeddedness, in streams (Hannaford and Resh, 1995; Wang et al., 1996), which is reflected in the finding of the 'RHA fine sediment deposition', but not in the 'shuffle index'. Regarding the quantitative assessment of substrate composition, Archer et al. (2004) found low intra-annual observer variability, which does align with my results of the 'substrate diversity' but not of the 'substrate index', which showed moderate inter-annual observer variability in this study.

All other qualitative metrics scored high inter-annual observer variability in this study, whereas previous studies have found mixed results (Table 3.6). Additionally, Hannaford and Resh (1995) found that qualitative habitat assessments did not produce consistent results, as variability was ascribed to a combination of interpretation of the written descriptions (viewer error) and natural intra-site variability.

Unexpectedly, the quantitative metrics of 'average sediment depth' and 'average bank undercut' had by far the highest CV estimates (Table 3.5). Because these measurements were made with a ruler, it was anticipated that they would have a relatively low degree of inter-annual observer variability. Additionally, the high inter-annual observer variability in data of stream depth and bank undercut measurements found in this study does not align with previous findings. For example, Wang et al. (1996) found moderate intra-annual observer variability in the measurements of sediment depth, and regarding measurements of bank undercuts, Wang et al. (1996), Newcombe et al. (2007) and Archer et al. (2004) found low intra-annual observer variability.

However, it appeared that data of quantitative metrics measured with a ruler displayed the highest inter-annual observer variability in this study. Metrics measured on a continuous scale may leave more room for observer variability than metrics scored on a discrete scale. Further, the sediment depth measurement technique in soft-bottomed streams could have influenced inter-annual observer variability. Sediment depth measurements were taken by pushing the measurement staff into the streambed substrate, and as such, some observers may have firmly pushed the measurement staff into the soft substrate, whereas others may have used less force. The high inter-annual observer variability in data of bank undercut measurements could have been caused by spatial variations within a study reach rather than a result of measurement variations between observers, as bank undercuts are rarely uniform within a study site. The location at which an undercut was measured was therefore likely to influence data variability.

A study by Hannaford and Resh (1995) found that natural intra-site variations had an influence on intra-annual observer variability. They ascribed within-site variation to the fact that different observer pairs did not conduct their assessments from the exact same point (Hannaford and Resh, 1995). In this study, intra-site variability is possible to have happened for all measurements at cross-sections in streams, as it is very unlikely that these would have been in exactly the same place year-to-year. Hannaford and Resh (1995) suggested observers should obtain a series of observations along the study reach to reduce within-site variability. Although this method is already part of the qualitative BOP NERMN stream habitat assessment, inter-annual observer variability in quantitative measurement data may be improved by taking more measurements along the study reach. By increasing the number of transects across the stream, a better representation of measured factors such as width, depth, sediment depth, and bank undercut at a site may be obtained. However, this potential for lower inter-annual observer variability is offset by the increased in-field assessment duration.

Table 3.6: Literature review summary of intra-annual observer variability in stream habitat attributes, divided into qualitative and quantitative metrics. Cited studies were not restricted to native forest sites, investigated intra-annual observer variability, and involved diverse methodological approaches (see Appendix). While not all listed stream habitat attributes are relevant to this chapter, they are all relevant to the BOP NERMN stream habitat survey overall.

Stream habitat attribute	Intra-annual observer variability	Reference
Qualitative Metrics		
Fish cover	Moderate	Clapcott, 2015
Bank stability	High	Hannaford and Resh, 1995
	High	Clapcott, 2015
	Low	Wang et al., 1996
Substrate composition	Low	Hannaford and Resh, 1995
	Moderate	Wang et al., 1996
Fine sediment/ embeddedness	High	Hannaford and Resh, 1995
	Moderate	Clapcott, 2015
	High	Wang et al., 1996
Canopy cover	Moderate	Hannaford and Resh, 1995
	Moderate	Clapcott, 2015
Channel alternation	High	Hannaford and Resh, 1995
	Moderate	Clapcott, 2015
Riparian vegetation cover	High	Hannaford and Resh, 1995
	Low	Clapcott, 2015
Riparian vegetation type	High	Wang et al., 1996
	Low	Clapcott, 2015
Riparian width	Moderate	Hannaford and Resh, 1995
	Moderate	Clapcott, 2015
Hydraulic heterogeneity	High	Clapcott, 2015
Invertebrate habitat	Low	Clapcott, 2015
Quantitative metrics		
Bank stability	High	Archer et al., 2004
Fish cover/ bank undercut	Low	Newcomb et al., 2007
	Low	Wang et al., 1996

	Low	Archer et al., 2004
Overhanging vegetation	Moderate	Wang et al., 1996
Wetted width	Low	Newcomb et al., 2007
	Low	Wang et al., 1996
Bank-to-bank width	High	Newcomb et al., 2007
Stream depth	Low	Newcomb et al., 2007
	Low	Wang et al., 1996
Sediment depth	Moderate	Wang et al., 1996
Substrate size/ composition	Low	Archer et al., 2004

Generally, high CV estimates reduce the power in detecting trends in stream habitat conditions and weaken the interpretation reliability of results from a single survey (Archer et al., 2004). In contrast, lower CV estimates indicate that each individual value is similar to the mean (Archer et al., 2004). Nevertheless, several researchers have suggested that adequate training can decrease inter-observer variability (Hannaford et al., 1997; Morrison, 2016; Parsons et al., 2002; Roper et al., 2002). However, each year, summer students spent at least two weeks in the field learning how to collect habitat data, which presumably would have minimised inter-annual observer variability. Still, the results of this study suggested that a prolonged training period may be required, as this study found high inter-annual observer variability in data of most BOP NERMN stream habitat survey metrics.

Besides training, awareness of monitored attributes and their functions in stream systems and a thorough understanding of assessment terminology and definitions should ensure measurements and observations are taken in the appropriate format and are consistent across sampling teams (Parsons et al., 2002). Parson et al. (2002) identified the following points to be demonstrated and synchronised during training, including the sequence of work at a sampling site, identification of flow levels, cross-section measurements, assessment and interpretation of each individual stream habitat attribute and estimation of distance. However, all these points are addressed in the BOP NERMN habitat surveying training. For assessments including the estimation of covered areas, such as the areas covered in fine sediment, observers can be trained with samples of known percentage cover (Killourhy et al., 2016; Morrison, 2016). For example, computer calibrations can offer field staff the experience required to accurately estimate substrate proportions (Killourhy et al., 2016). As such, computer simulations may be introduced to the BOP NERMN habitat surveying training. Additionally, Morrison (2016) suggested feedback systems should be included in the training of field staff as there is evidence that feedback lets observers express uncertainties in their estimates more accurately. Though, summer students conducting the BOP NERMN stream habitat survey were always able to communicate any uncertainties they encountered with their supervisor.

Lastly, if observers did not produce acceptably consistent results after training, Morrison (2016) suggested rejecting observers from a field team to achieve acceptable data quality. This suggestion was supported by the fact that estimates of a single observer could result in outliers in the dataset, which have the potential to decrease overall precision (Morrison, 2016). However, from the perspective of a Regional Council that employs summer students to undertake their routine State of the Environment (SoE) monitoring work every summer, this would not be a particularly workable solution. It is an ongoing challenge for councils that employ casual contractors each year to ensure high-quality data is collected by sufficiently trained staff to minimise inter-annual observer variability. More frequent quality assurance and quality control checks may be implemented in future years to help minimise inter-annual observer variability.

Turning attention to the results of the value models, which revealed that more than half (56%) of all metrics scored relatively high 'performance scores' to assess stream habitat attributes, independent of their metric type, considering the two objectives of low inter-annual observer variability and short in-field assessment duration (Table 3.5). However, on the one hand, the reduced inter-annual observer variability of some quantitative metrics outweighed the extra time taken to assess stream habitat attributes. For example, the reduced inter-annual observer variability of the quantitative metrics 'substrate index' and 'substrate diversity' outweighed the extra time taken to assess the stream habitat attributes of deposited fine sediment, invertebrate habitat and fish cover, compared to the qualitative RHA metrics. It became apparent that the 'substrate index' and 'substrate diversity' consistently scored high 'performance scores' and therefore could be great metrics to assess multiple attributes effectively. The in-field assessment duration for the Wolman pebble count is not particularly high relative to its potential of evaluating multiple attributes.

On the other hand, 'performance score' results revealed that several quantitative metrics were dominated by their qualitative alternatives, scoring worse in both objectives. For example, the quantitative metrics of 'average sediment depth', 'average bank undercut', and 'standard deviation wetted width' were exposed as the least efficient metrics, scoring zero in their 'performance scores'. In addition, the quantitative metrics 'standard deviation flow velocity', 'coefficient of variation flow velocity' and 'standard deviation stream depth' also scored relatively low 'performance scores'.

In one incidence, a metric had different 'performance scores' depending on the stream habitat attribute it was assigned to. This metric was the quantitative metric 'flow heterogeneity', which had high 'performance scores' in describing the stream habitat attributes of invertebrate habitat and hydraulic/flow heterogeneity, but had a low 'performance score' in describing the stream habitat attribute of channel alternation. The low 'performance score' was caused by the very low inter-annual observer variability of the alternative metric 'RHA channel alternation' outweighing the short in-field assessment duration of the quantitative metric 'flow heterogeneity'.

Ultimately, I sought to determine whether it was appropriate to exclusively rely on qualitative metrics, taking into account in-field assessment duration and inter-annual observer variability to create a more economical BOP NERMN stream habitat survey. However, the results of the 'performance scores' suggested that there is no clear reason to favour qualitative metrics, as more than half (56%) of all metrics performed relatively equally considering their in-field assessment duration and inter-annual observer variability.

The main limitation of the analysis in this chapter was that temporal variability was considered minimised by basing the analysis on the assumption that native forest sites did not change over the considered time period and that all variability was due to differences among observers. Although, given the relatively short time period these data were collected over (< 7 years), and the visual observations that most habitat attributes in native forest streams appeared constant, this assumption did not seem unreasonable. However, to conduct this study in a more scientifically sound manner, it would have been appropriate to have multiple observers carrying out the BOP NERMN stream habitat survey simultaneously to measure intra-annual observer variability that is not confounded with potential temporal variability. Nevertheless, even though the data analysed in this study were not intended to investigate any type of inter-observer variability, it was hoped that the results of this analysis would provide empirically based insights to choose the most appropriate metrics for future sampling.

Chapter 4:

General Discussion

Native forest site in the Bay of Plenty. Summer 2018/19

Multiple stream habitat attributes are evaluated using qualitative and quantitative assessment methods during the current Bay of Plenty Natural Environmental Monitoring Network (BOP NERMN) stream habitat survey. These assessments are carried out by a team of different students each summer. Typically, qualitative assessments require less time to be carried out (shorter in-field assessment duration) than quantitative measurements. Consequently, using only qualitative assessment methods are likely to be more economical for the BOP NERMN stream habitat survey. However, qualitative assessments are thought to be more subject to inter-observer variability and thus may reduce the ability to detect changes or trends at monitored sites. Therefore, a consideration that should be addressed when designing a monitoring programme is whether the data quality derived from field surveys is credible. Data must meet quality standards to address the monitoring goals of a given project. For the BOP NERMN stream habitat monitoring, it is essential that the analysis of collected data can detect environmental trends. However, the year-to-year change of the surveying team personnel (inter-annual observer variability) could influence the BOP NERMN stream habitat monitoring results. In this thesis, I aimed to determine whether omitting quantitative metrics and using qualitative metrics only could be an appropriate option for creating a more economical BOP NERMN stream habitat survey, taking into account the metrics' in-field assessment duration and inter-annual observer variability. This chapter provides an overview of the findings of this thesis. Furthermore, management recommendations for optimising the stream habitat assessment for the BOP NERMN stream habitat monitoring are outlined, and suggestions for further research directions are specified.

4.1 Summary of findings

4.1.1 Correlations amongst qualitative and quantitative metrics

Firstly, I used correlation analyses to determine whether qualitative and corresponding[11] quantitative metrics captured the same stream habitat attributes similarly. If the correlation between these metrics was significant, assessing only qualitative metrics could potentially be an appropriate option for creating a more economical BOP NERMN stream habitat survey, and thus quantitative metrics could theoretically be omitted.

[11] Corresponding quantitative metrics refer to quantitative metrics corresponding to a qualitative metric, both of which describe the same habitat attribute. For example, the measurement of average sediment depth is a corresponding quantitative metric to the qualitative metric 'shuffle index'. Both metrics describe the stream habitat attribute of deposited fine sediment.

- The correlation analyses revealed that all qualitative metrics, apart from the Rapid Habitat Assessment (RHA) riparian shade, were significantly correlated to at least one of their corresponding quantitative metrics.
- However, it was not anticipated that there would be a negative correlation between the metrics of ‘RHA bank vegetation’ and ‘average overhanging vegetation’ because longer overhanging vegetation was expected to result in higher ‘RHA bank vegetation’ scores. Therefore, the qualitative metric ‘RHA bank vegetation’ and the corresponding quantitative metric were excluded from further analysis.

The correlation analyses suggested that it was potentially appropriate to assess the stream habitat attributes of deposited fine sediment deposition, invertebrate habitat, fish cover, hydraulic/flow heterogeneity, bank stability, and channel alternation using only qualitative metrics and that most quantitative metrics had the potential to be removed from the survey. Quantitative metrics that could potentially be omitted included measurements of sediment depth, wetted width, flow velocity, stream depth, bank undercut, Wolman pebble count (i.e. ‘substrate index’ and ‘substrate diversity’) and flow heterogeneity.

However, to decide whether it was appropriate to remove quantitative metrics from the BOP NERMN stream habitat survey, I also considered the performance trade-offs of qualitative and quantitative metrics in relation to their in-field assessment duration and degree of inter-annual observer variability.

4.1.2 In-field assessment duration

A field technician and I measured the time taken to collect qualitative and quantitative metric data to investigate whether qualitative assessments were indeed more economical in terms of in-field assessment duration than quantitative metrics. Across the entire study reach, on average, it took:

- 4 minutes to carry out the qualitative RHA protocol
- 5 minutes to conduct the qualitative ‘shuffle index’
- 2 minutes to record quantitative ‘flow heterogeneity’ data
- 5 minutes to carry out the quantitative Wolman pebble count (of which the ‘substrate index’ and ‘substrate diversity’ are derived)

- 10 minutes to collect data of the seven quantitative transect measurements (i.e. measurements of wetted stream width, bank to bank width, overhanging vegetation, bank undercut, stream and sediment depth as well as flow velocity)

Therefore, the statement that qualitative metrics require less time to be obtained was only partially supported. Across the entire study reach, the quantitative metric 'flow heterogeneity' had the shortest in-field assessment duration, whereas data of the quantitative transect measurements took the longest to obtain.

4.1.3 Inter-annual observer variability

I investigated the degree of inter-annual observer variability amongst quantitative and qualitative metric data, as the year-to-year change between surveying team personnel could influence the monitoring results of the BOP NERMN stream habitat survey. I calculated the percent coefficient of variation (CV) of all metrics and then fitted a Bayesian hierarchical linear model to compare the CV values among the qualitative metric and corresponding quantitative metrics of a given stream habitat attribute.

- Data of the qualitative metrics 'RHA channel alternation', 'RHA bank stability', 'shuffle index' and the quantitative metric 'substrate diversity' showed low inter-annual observer variability.
- Data of the quantitative metric 'substrate index' showed moderate inter-annual observer variability.
- Data of the remaining qualitative and quantitative metrics showed high inter-annual observer variability (i.e. 'RHA deposited fine sediment', 'RHA invertebrate habitat', 'RHA fish cover', 'RHA hydraulic heterogeneity', 'average sediment depth', 'flow heterogeneity', 'average bank undercut', 'standard deviation wetted width', 'standard deviation flow velocity' and 'coefficient of variant flow velocity')
- The low inter-annual observer variability of the 'RHA channel alternation' metric data was likely caused by the fact that native forest sites are usually unmodified.
- The low inter-annual observer variability of the 'shuffle index' metric data could have been influenced by the fact that the 'shuffle index' is scored on a narrower scale (1 to 5), compared to the RHA scoring scale (1 to 20) or the continuous measurement scale of quantitative metrics.

The results of the Bayesian hierarchical linear models suggested that data of most metrics (69%) had high levels of inter-annual observer variability, independent of the metric type.

4.1.4 Value models

Lastly, to conclude as to whether omitting quantitative metrics and using only qualitative metrics could be an appropriate option for creating a more economical BOP NERMN stream habitat survey, I crated value models as per Gregory et al. (2012) to assess the performance trade-offs of qualitative and quantitative metrics in relation to in-field assessment duration and inter-annual observer variability.

- More than half (56%) of all metrics performed relatively equivalent in relation to their in-field assessment duration and inter-annual observer variability, with 'performance scores' ≥ 0.8, regardless of whether they were qualitative (i.e. 'shuffle index', 'RHA fine sediment deposition', 'RHA invertebrate habitat', 'RHA fish cover', 'RHA hydraulic heterogeneity' and 'RHA bank stability') or quantitative (i.e. 'substrate index', 'substrate diversity' and 'flow heterogeneity'[12]).
- The qualitative metric of 'RHA channel alternation' reached a 'performance score' of 0.6. However, this score was still higher than the corresponding quantitative metric 'flow heterogeneity'[12] ('performance score' 0.4), and therefore 'RHA channel alternation' is the preferred metric describing the stream habitat attribute of channel alternation.
- The quantitative metrics of 'average sediment depth', 'average bank undercut', 'standard deviation wetted width', 'standard deviation stream depth', 'standard deviation flow velocity', 'coefficient of variation flow velocity' and 'flow heterogeneity'[12] scored low 'performance scores' (≤ 0.56).

Ultimately, I sought to determine whether it was appropriate to exclusively rely on qualitative metrics, taking into account in-field assessment duration and inter-annual observer variability to create a more economical BOP NERMN stream habitat survey. However, the results of the 'performance scores' suggest that there is currently no clear reason to favour qualitative metrics in the BOP NERMN stream habitat survey, as more than half of all metrics performed relatively

[12] The quantitative metric of 'flow heterogeneity' scored high 'performance scores' in describing the stream habitat attributes of invertebrate habitat and hydraulic/flow heterogeneity, but scored low in describing the stream habitat attribute of channel alternation. This was because the alternative metric 'RHA channel alternation' had very low inter-annual observer variability.

equivalent in relation to their in-field assessment duration and inter-annual observer variability. Overall, this research suggests the need for further refinement of the BOP NERMN stream habitat assessment protocol to achieve consistent data collection across surveying teams, as inter-annual observer variability was generally high in most metrics, regardless of their metric type.

4.2 Management recommendations

The BOP NERMN programme was initiated with the objectives to monitor and report the state and trends of the environment across the Bay of Plenty region and to document the effectiveness of the council's regional plans. To achieve these objectives collected data must be of sufficient quality. Generally, the best way to achieve good data quality is to keep errors at the data collection phase at a minimum (Houston and Hiederer, 2009). The ideal solution to increase data quality and avoid observer variability would be employing the same observer(s) for all data collection (Morrison, 2016). However, the reality is that employing the same observers is impossible for many long-term or spatially large-scale monitoring programmes (Morrison, 2016). In addition to good training, as discussed in Chapter 3, another realistic approach to increase data quality and decrease inter-annual observer variability could include the use of quality assurance and quality control systems.

Quality checks of assessments should be carried out throughout a monitoring programme, especially if the field team includes new observers (Harding et al., 2009). For example, collecting replicates improves data quality, as they can verify the precision of the data collected (NIWA, 2019). A divergence of about 10% in replicate measurements could indicate differences in how individual observers carry out measurements (NIWA, 2019). Additionally, continuous evaluation and calibration of observer performance are recommended to reduce observer variability (Morrison, 2016). For instance, an accreditation scheme that involves completing a training course and an assessment is currently in place for field staff conducting stream habitat surveys on behalf of the UK's Environmental Agency (Cherrill, 2016). Additionally, UK organisations analysing macroinvertebrate samples are now subject to external auditing, which has reduced error rates in taxonomic identification (Cherrill, 2016). A pan-European comparison showed that taxonomic errors for macroinvertebrate identification were lower in the UK compared to other countries that had no auditing in place (Cherrill, 2016). An introduction of an accreditation scheme would very likely provide a reliable and transparent method to assess observers' performance and enhance professional practices of the BOP NERMN stream habitat monitoring.

4.3 Future research

Next to refining the current the BOP NERMN stream habitat assessment protocol, further research should investigate how to incorporate quality assurance and quality control systems into the BOP NERMN stream habitat monitoring to ensure monitoring objectives are met. Quality assurance and quality control strategies could include data proofing/reviewing (Herron et al., 2004) and cross-user validation (Harding et al., 2009). Additionally, as variability is an element of all sampling protocols (Roper et al., 2002), instead of eliminating a metric because of its degree of inter-annual observer variability, a different approach could be to incorporate uncertainty estimates of inter-annual observer variability during the analysis of the BOP NERMN stream habitat monitoring data.

Lastly, further research on a national scale could involve developing a mandatory countrywide accreditations scheme, involving a uniform training programme and assessment, for observers carrying out the annual SoE monitoring for regional councils. This accreditation scheme could support standardised and high-quality SoE sampling across New Zealand.

References

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

Allan, J. D., Wipfli, M. S., Caouette, J. P., Prussian, A., and Rodgers, J. (2003). Influence of streamside vegetation on inputs of terrestrial invertebrates to salmonid food webs. *Canadian Journal of Fisheries and Aquatic Sciences, 60*(3), 309-320. doi:10.1139/f03-019

Archer, E. K., Roper, B. B., Henderson, R. C., Bouwes, N., Mellison, S. C., and Kershner, J. L. (2004). Testing common stream sampling methods for broad-scale, long-term monitoring. *Gen. Tech. Rep. RMRS-GTR-122. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain research Station. 15 p., 122*.

Barquín Ortiz, J., and Martinez-Capel, F. (2011). Preface: Assessment of physical habitat characteristics in rivers, implications for river ecology and management. *Limnetica, 30*(2), 159-167.

Bauer, S. B., and Ralph, S. C. (2001). Strengthening the use of aquatic habitat indicators in Clean Water Act programs. *Fisheries, 26*(6), 14-25.

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.

Bazeley, P. (2004). Issues in mixing qualitative and quantitative approaches to research. *Applying qualitative methods to marketing management research, 141*, 156.

Beechie, T. J., Sear, D. A., Olden, J. D., Pess, G. R., Buffington, J. M., Moir, H., Pollock, M. M. (2010). Process-based Principles for Restoring River Ecosystems. *Bioscience, 60*(3), 209-222. doi:10.1525/bio.2010.60.3.7

Belletti, B., Rinaldi, M., Buijse, A. D., Gurnell, A. M., and Mosselman, E. (2015). A review of assessment methods for river hydromorphology. *Environmental Earth Sciences, 73*(5), 2079-2100. doi:10.1007/s12665-014-3558-1

Bundschuh, M., and McKie, B. G. (2016). An ecological and ecotoxicological perspective on fine particulate organic matter in streams. *Freshwater biology, 61*(12), 2063-2074.

Bürkner, P. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*(80(1)), 1-28. doi:10.18637/jss.v080.i01

Carstensen, J., and Lindegarth, M. (2016). Confidence in ecological indicators: a framework for quantifying uncertainty components from monitoring data. *Ecological indicators, 67*, 306-317.

Cherrill, A. (2016). Inter-observer variation in habitat survey data: investigating the consequences for professional practice. *Journal of Environmental Planning and Management, 59*(10), 1813-1832.

Christensen, K., Pathirage, D., Halliday, P., O'Brien, R., and Forrest, G. (2017). Minimising the Environmental Effects of Dredging - Wharemauku Stream. *Paper presented at the Stormwater Conference.* https://www.waternz.org.nz/Article?Action=ViewandArticle_id=1310

Clapcott, J. (2013). *Rapid Habitat Assessment Workshop. Prepared for Hawkes Bay Regional Council*. Cawthron Report No. 2445. 7 p.

Clapcott, J. (2015). *National rapid habitat assessment protocol development for streams and rivers. Prepared for Northland Regional Council.* Cawthron Report No. 2649. 29 p. plus appendices.

Clapcott, J., Casanovas, P., and Doehring, K. (2020). *Indicators of freshwater quality based on deposited sediment and rapid habitat assessment*. Prepared for the Ministry for the Environment. Cawthron Report No. 3402. 21 p. plus appendices.

Clapcott, J., Goodwin, E., Williams, E., Harding, J., McArthur, K., Schallenberg, M., Young, R., Death, R. (2019). *Technical report on the prototype New Zealand river ecosystem health score. Prepared for Ministry for the Environment*. Cawthron Report No. 3332. 48 p. plus appendices.

Clapcott, J., Young, R., Harding, J., Matthaei, C., Quinn, J., and Death, R. (2011). *Sediment assessment methods: protocols and guidelines for assessing the effects of deposited fine sediment on in-stream values*. Cawthron Institute, Nelson, New Zealand.

Cloern, J. E. (2007). Habitat connectivity and ecosystem productivity: Implications from a simple model. *The American Naturalist, 169*(1), E21-E33.

Davies, P., and Nelson, M. (1994). Relationships between riparian buffer widths and the effects of logging on stream habitat, invertebrate community composition and fish abundance. *Marine and Freshwater Research, 45*(7), 1289-1305.

Dewson, Z. S., James, A. B., and Death, R. G. (2007). A review of the consequences of decreased flow for instream habitat and macroinvertebrates. *Journal of the North American Benthological Society, 26*(3), 401-415.

Donald, R. (2014). Review of the NERMN Programme 2014. Retrieved from Whakatane: https://cdn.boprc.govt.nz/media/368036/review-of-the-nermn-programme-2014.pdf

Dosskey, M. G., Vidon, P., Gurwick, N. P., Allan, C. J., Duval, T. P., and Lowrance, R. (2010). The role of riparian vegetation in protecting and improving chemical water quality in streams 1. *JAWRA Journal of the American Water Resources Association, 46*(2), 261-277.

Elosegi, A., Flores, L., and Diez, J. (2011). The importance of local processes on river habitat characteristics: A Basque stream case study. *Limnetica, 30*(2), 0183-0196.

Fernandez, D., Barquin, J., and Raven, P. J. (2011). A review of river habitat characterisation methods: indices vs. characterisation protocols. *Limnetica, 30*(2), 217-234. Retrieved from <Go to ISI>://WOS:000298769700005

Ferretti, M. (2011). Quality assurance: a vital need in ecological monitoring. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources, 6*(011), 1-14.

Fowler, J., Cohen, L., and Jarvis, P. (2013). Practical statistics for field biology. John Wiley & Sons.

Franke, M., and Roettger, T. B. (2019). Bayesian regression modeling (for factorial designs): A tutorial.

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. doi:10.1007/bf01867358

Gayraud, S., and Philippe, M. (2003). Influence of bed-sediment features on the interstitial habitat available for macroinvertebrates in 15 French streams. *International Review of Hydrobiology: A Journal Covering all Aspects of Limnology and Marine Biology, 88*(1), 77-93.

Gebrekiros, S. (2016). Factors affecting stream fish community composition and habitat suitability. *Journal of Aquaculture and Marine Biology, 4*(2), 00076.

Goodenough, A. E., Carpenter, W. S., McTavish, L., Blades, B., Clarke, E., Griffiths, S., Wilson, L. (2020). The impact of inter-observer variability on the accuracy, precision and utility of a commonly-used grassland condition index. *Ecological indicators, 117*, 106664.

Gregory, R., Failing, L., Harstone, M., Long, G., McDaniels, T., and Ohlson, D. (2012). *Structured decision making: a practical guide to environmental management choices*: John Wiley and Sons.

Gurnell, A. M., Scott, S. J., England, J., Gurnell, D., Jeffries, R., Shuker, L., and Wharton, G. (2020). Assessing river condition: A multiscale approach designed for operational application in the context of biodiversity net gain. *River Research and Applications, 36*(8), 1559-1578.

Hannaford, M. J., Barbour, M. T., and Resh, V. H. (1997). Training reduces observer variability in visual-based assessments of stream habitat. *Journal of the North American Benthological Society, 16*(4), 853-860.

Hannaford, M. J., and Resh, V. H. (1995). VARIABILITY IN MACROINVERTEBRATE RAPID-BIOASSESSMENT SURVEYS AND HABITAT ASSESSMENTS IN A NORTHERN CALIFORNIA STREAM. *Journal of the North American Benthological Society, 14*(3), 430-439. doi:10.2307/1467209

Harding, J. S., Clapcott, J., Quinn, J. M., Hayes, J. W., Joy, M. K., Storey, R. G., Boothroyd, I. K. D. (2009). *Stream Habitat Assessment Protocols for wadeable rivers and streams in New Zealand*: University of Canterbury, School of Biological Sciences.

Herron, E., Green, L., Stepenuck, K., and Addy, K. (2004). Building credibility: Quality assurance and quality control for volunteer monitoring programs. *University of Rhode Island and University of Wisconsin http://www. usawaterquality. org/Volunteer/Outreach/BuildingCredibilityVI. pdf*.

Hill, A. R. (1996). Nitrate removal in stream riparian zones. *Journal of Environmental Quality, 25*(4), 743-755. doi:10.2134/jeq1996.00472425002500040014x

Hogle, J. S., Wesche, T. A., and Hubert, W. A. (1993). A test of the precision of the habitat quality index model II. *North American Journal of Fisheries Management, 13*(3), 640-643.

Houston, T. D., and Hiederer, R. (2009). Applying quality assurance procedures to environmental monitoring data: a case study. *Journal of environmental monitoring, 11*(4), 774-781.

Jowett, I. G. (1993). A method for objectively identifying pool, run, and riffle habitats from physical measurements. *New Zealand journal of marine and freshwater research, 27*(2), 241-248.

Jowett, I. G., and Richardson, J. (1990). Microhabitat preferences of benthic invertebrates in a New Zealand river and the development of in-stream flow-habitat models for Deleatidium spp. *New Zealand journal of marine and freshwater research, 24*(1), 19-30.

Jowett, I. G., Richardson, J., and McDowall, R. (1996). Relative effects of in-stream habitat and land use on fish distribution and abundance in tributaries of the Grey River, New Zealand. *New Zealand journal of marine and freshwater research, 30*(4), 463-475.

Kaufmann, P. R., Levine, P., Robison, E. G., Seeliger, C., and Peck, D. V. (1999). Quantifying Physical Habitat in Wadeable Streams. In *EPA 620/R-99/003. Environmental Monitoring and Assessment Program*. Corvallis, Oregon: U.S. Environmental Protection Agency.

Killourhy, C. C., Crane, D., and Stehman, S. V. (2016). Precision and accuracy of visual estimates of aquatic habitat. *Freshwater Science, 35*(3), 1062-1072.

Kolada, A., Ciecierska, H., Ruszczyńska, J., and Dynowski, P. (2014). Sampling techniques and inter-surveyor variability as sources of uncertainty in Polish macrophyte metric for lake ecological status assessment. *Hydrobiologia, 737*(1), 265-279.

Matthaei, C. D., Piggott, J. J., and Townsend, C. R. (2010). Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. *Journal of Applied Ecology, 47*(3), 639-649. doi:10.1111/j.1365-2664.2010.01809.x

McCormick, J. (2001). *Environmental Policy in the European Union*. New York: Palgrave.

McGinnity, P., Mills, P., Roche, W., and Müller, M. (2005). A Desk Study to Determine a Methodology for the Monitoring of the 'Morphological Condition'of Irish Rivers for the Water Framework Directive.

Resource Management Act 1991, (1991).

Ministry for the Environment. (2010). Land: Land use. *Environmental Snapshot January 2010.* Retrieved from https://www.mfe.govt.nz/sites/default/files/media/Land/land-use_0_0.pdf

Ministry for the Environment. (2017). *Bay of Plenty – Te Moana a Toi-te-Huatahi*. Retrieved from http://www.mfe.govt.nz/

Ministry for the Environment. (2020a*). A new Freshwater Planning Process: Technical guidance for councils.* Retrieved from Wellington: https://www.mfe.govt.nz/sites/default/files/media/Fresh%20water/a-new-freshwater-planning-process-technical-guidance-for-councils.pdf

Ministry for the Environment. (2020b). *Tauākī Whakamaunga Atu | Statement of Intent 2020-2025.* Retrieved from Wellington: https://environment.govt.nz/assets/Publications/Files/Ministry-for-the-environment-statement-of-intent-2020-2025-final.pdf

Ministry for the Environment. (2021). *Understanding the RMA and how to get involved. An everyday guide to the Resource Management Act: 1.1*. Retrieved from Wellington: https://environment.govt.nz/assets/Publications/Files/1.1-understanding-the-rma.pdf

Ministry for the Environment, and Ministry for Primary Industries (Producer). (2020a). *Essential Freshwater: Overview factsheet*. Retrieved from https://www.mfe.govt.nz/sites/default/files/media/Fresh%20water/essential-freshwater-overview-factsheet.pdf

Ministry for the Environment, and Ministry for Primary Industries (Producer). (2020b). *Te Mana o te Wai factsheet*. Retrieved from https://www.mfe.govt.nz/sites/default/files/media/Fresh%20water/essential-freshwater-te-mana-o-te-wai-factsheet.pdf

Ministry for the Environment, and Stats NZ. (2020). *New Zealand's Environmental Reporting Series: Our freshwater 2020.* Retrieved from https://environment.govt.nz/publications/our-freshwater-2020/

Miserendino, M. L., Casaux, R., Archangelsky, M., Di Prinzio, C. Y., Brand, C., and Kutschker, A. M. (2011). Assessing land-use effects on water quality, in-stream habitat, riparian ecosystems and biodiversity in Patagonian northwest streams. *Science of the total environment, 409*(3), 612-624.

Morrison, L. W. (2016). Observer error in vegetation surveys: a review. *Journal of Plant Ecology, 9*(4), 367-379.

NEMS. (2020). *National Environmental Monitoring Standards: Macroinvertebrates. Collection and Processing of Macroinvertebrate Samples from Rivers and Streams*. Retrieved from https://bucketeer-54c224c2-e505-4a32-a387-75720cbeb257.s3.amazonaws.com/public/Documents/NEMS-Macroinvertebrates-v1.0.0.pdf

New Zealand Government. (2020). *National Policy Statement for Freshwater Management 2020*.

Newcomb, T. J., Orth, D. J., and Stauffer, D. F. (2007). Habitat evaluation. *Analysis and interpretation of freshwater fisheries data. American Fisheries Society, Bethesda, Maryland*, 843-886.

NIWA. (2019). *SHMAK Stream Health Monitoring and Assessment Kit User Manual*. Retrieved from Christchurch: http://www.niwa.co.nz/freshwater/tools/shmak

Parsons, M., Thoms, M., and Norris, R. (2002). Australian river assessment system: AusRivAS physical assessment protocol. *Monitoring river health initiative technical report, 22*.

Paul, M. J., and Meyer, J. L. (2001). Streams in the urban landscape. *Annual Review of Ecology and Systematics, 32*, 333-365. doi:10.1146/annurev.ecolsys.32.081501.114040

Plafkin, J. L. (1989). Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish. United states Environmental protection Agency, Office of Water.

Queirós, A., Faria, D., and Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies*.

R Core Team. (2021). R: A language and environment for statistical computing. In Vienna, Austria: R Foundation for Statistical Computing.

Roper, B. B., Kershner, J. L., Archer, E., Henderson, R., and Bouwes, N. (2002). AN EVALUATION OF PHYSICAL STREAM HABITAT ATTRIBUTES USED TO MONITOR STREAMS 1. *JAWRA Journal of the American Water Resources Association, 38*(6), 1637-1646.

Sale, J. E., Lohfeld, L. H., and Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and quantity, 36*(1), 43-53.

Schoof, R. (1980). ENVIRONMENTAL IMPACT OF CHANNEL MODIFICATION 1. *JAWRA Journal of the American Water Resources Association, 16*(4), 697-701.

Simon, A., and Collison, A. J. (2002). Quantifying the mechanical and hydrologic effects of riparian vegetation on streambank stability. *Earth Surface processes and landforms, 27*(5), 527-546.

Stats NZ. (2020, 16.04.2020). *Freshwater physical habitat*. Retrieved from https://www.stats.govt.nz/indicators/freshwater-physical-habitat

Steckler, A., McLeroy, K. R., Goodman, R. M., Bird, S. T., and McCormick, L. (1992). Toward integrating qualitative and quantitative methods: an introduction. In: Sage Publications Sage CA: Thousand Oaks, CA.

Suren, A., Nistelrooy, D. V., and Fergusson, V. (2017). *State and trends in river health (1992-2014) in the Bay of Plenty: Results from 22 years of the NERMN stream bio-monitoring programme* Retrieved from Whakatane: https://cdn.boprc.govt.nz/media/610374/state-and-trends-in-river-health-1992-2014-in-the-bay-of-plenty_-results-from-22-years-of-the-nermn-stream-bio-monitoring-programme.pdf

Wang, L., Simonson, T. D., and Lyons, J. (1996). Accuracy and precision of selected stream habitat estimates. *North American Journal of Fisheries Management, 16*(2), 340-347. doi:10.1577/1548-8675(1996)016<0340:Aaposs>2.3.Co;2

Weiß, A., Matouskova, M., and Matschullat, J. (2008). Hydromorphological assessment within the EU-Water Framework Directive—trans-boundary cooperation and application to different water basins. *Hydrobiologia, 603*(1), 53-72.

Yang, C. T. (1971). Formation of riffles and pools. *Water Resources Research, 7*(6), 1567-1574.

Young, R., Wagenhoff, A., Holmes, R., Newton, M., and Clapcott, J. (2018). *What is a healtvy river?* Retrieved from https://www.cawthron.org.nz/media/publications/pdf/2018_10/Cawthron_-_Healthy_River_Report_-_PRINT_003.pdf

Appendix: Study details of literature cited in Table 3.7., Chapter 3.

This appendix offers an outline of the study details of the literature cited in Table 3.7., Chapter 3. First, general study methods are summarised, including details of the observers and observer variability analysis. Second, Table A1 contains details of metric definitions and study results, as well as my interpretations of the findings of cited studies. Because it was challenging to paraphrase study methods and metric definitions without misrepresenting original information, the following text contains multiple quotations.

In the study by Clapcott (2015), without prior training, two to three observers rated nine stream habitat attributes on a scale from 1 (very degraded conditions) to 20 (pristine conditions), following the Rapid Habitat Assessment protocol (RHA; by Clapcott, 2013) at 17 sites between February and July 2014. The resulting scores given by different observers were compared, with the result that intra-annual observer variability was the lowest for the stream habitat attributes of invertebrate habitat and bank vegetation, and the highest for the stream habitat attributes of hydraulic heterogeneity and bank stability. To fit these findings into my literature review, I considered intra-annual observer variability of a) invertebrate habitat and bank vegetation[13] as low, b) fine sediment, fish cover, riparian buffer (riparian width), riparian shade and channel alternation as moderate, and c) hydraulic heterogeneity and bank stability as high (Table A1).

In the study by Archer et al. (2004), observers were trained for two weeks before the sampling season. Eighteen observers were split into three groups, and all groups started the stream measurements at the same point in a stream. Intra-annual observer variability was estimated by calculating the coefficient of variation amongst groups and the 95% confidence interval. In addition, the percent variation ascribed to groups relative to the overall variability amongst sampled streams was calculated. For metric descriptions and results, see 'Study details' in Table A1.

In the study by Hannaford and Resh (1995), 28 students were trained to conduct the Rapid-Bioassessment Protocol III survey (RBP; by Plafkin, 1989) at three sites. Sampling efforts were completed within one week. Each stream habitat attribute included in the RBP III "was rated as optimal, sub-optimal, marginal, or poor" (rating categories) and written descriptions in the protocol supported the assignation of stream habitat attribute ratings (Hannaford and Resh, 1995). During

[13] The attribute bank vegetation in the RHA (Clapcott, 2013) includes descriptions of both the type and cover of the riparian vegetation.

the statistical analysis, “a coefficient of variation (CV = SD/mean) was calculated for each metric at each site as a measure of variability. The overall precision of each metric among the three sites was evaluated using the multiple correlation coefficient (R2) from the ANOVA models” (Hannaford and Resh, 1995). For metric descriptions and results, see ‘Study details’ in Table A1.

In the study by Wang et al. (1996), six observers estimated stream habitat attributes in three streams at approx. 100-m long sampling stations during summer 1991. Intra-annual observer variability was examined by “comparing the confidence interval of means from six observers with the field measurement precision for each habitat variable” (Wang et al., 1996). The field measurement precision (FMP) for each stream habitat attribute (aka habitat variable) “was defined as the nearest decimal unit to which a habitat value was measured or visually estimated in the field and represented our ability to measure this variable under field conditions” (Wang et al., 1996). Field observations by experienced personnel familiar with the stream habitat assessment procedure provided the basis for the FMPs. For metric descriptions and results, see ‘Study details’ in Table A1.

In their article ‘Habitat Evaluation’, Newcomb et al. (2007) remark that measurements of bank undercut, wetted width and stream depth can be precise and repeatable, while measurements of bank-to-bank width can be imprecise and of low repeatability. Therefore, I considered intra-annual observer variability of bank undercut, wetted width and stream depth measurements as low, and of bank-to-bank width measurements as high (Table A1)

Table A1: Literature review summary of intra-annual observer variability in stream habitat attributes, divided into qualitative and quantitative metrics. Details of metric definitions and study results, as well as my interpretations of the findings of cited studies are outlined (Study details), with some study details being described in the text above (NA). Wang et al. (1996) compared measurements and visual estimates of stream habitat attributes to the field measurement precision (FMP) of the given stream habitat attribute.

Stream habitat attribute	Intra-annual observer variability	Reference	Study details
Qualitative metrics			
Fish cover	Moderate	Clapcott, 2015	NA
Bank stability	High	Hannaford and Resh, 1995	**Metric**: Upper bank stability was rated from optimal (10 points, low erosion) to poor (2 points, high erosion). **Results**: Upper bank stability assessments were highly variable. **My interpretation**: Bank stability ratings were distributed across all four rating categories (i.e. optimal, sub-optimal, marginal, poor); therefore, I assigned high intra-annual observer variability to the bank stability ratings in my summary.
	High	Clapcott, 2015	NA
	Low	Wang et al., 1996	**Metric**: Bank erosion left and right bank: "Percentage of bank with bare soil along the transect: if the bank was not discernible, percentage of bare soil within 5 m of stream edge: visually estimated along transect line" (Wang et al., 1996). **Results**: Bank erosion (left and right) visual estimates did not exceed FMP. **My interpretation**: Therefore, I considered bank erosion visual estimates as having low intra-annual observer variability.
Substrate composition	Low	Hannaford and Resh, 1995	**Metric**: Bottom substrate was rated from optimal (20 points, mixed rubble) stable to poor (5 points, uniform and unstable). **Results**: Bottom substrate ratings were distributed between two rating categories. **My interpretation**: Therefore, I considered the ratings of bottom substrate to have low intra-annual observer variability.
	Moderate	Wang et al., 1996	**Metric**: The percentage of silt, sand, fine gravel, coarse gravel, cobbles and boulders was visually estimated at four quadrats along a transect.

			Results: Wang et al. (1996) considered the precision of stream substrate estimates as moderate, as FMP were exceeded occasionally.
Fine sediment/ embeddedness	High	Hannaford and Resh, 1995	**Metric**: Embeddedness was rated from optimal (20 points, low fine sediment) to poor (5 points, high fine sediment). **Results**: Embeddedenss ratings were distributed across all four rating categories. **My interpretation**: Therefore, I assigned high intra-annual observer variability to embeddedness.
	Moderate	Clapcott, 2015	NA
	High	Wang et al., 1996	**Metric**: Embeddedness was defined as the "percentage of bank with bare soil along the transect: if the bank was not discernible, percentage of bare soil within 5 m of stream edge" and was "visually estimated along transect line" (Wang et al., 1996). **Results**: Estimates of embeddedness had relatively high observer variability.
Canopy cover	Moderate	Hannaford and Resh, 1995	**Metric**: Canopy cover (shading) was rated from optimal (20 points, mixed cover) to poor (5 points, uniform shading or light). **Results**: Canopy cover ratings were distributed across three rating categories. **My interpretation**: Therefore, I considered canopy cover ratings to have moderate intra-annual observer variability.
	Moderate	Clapcott, 2015	NA
Channel alternation	High	Hannaford and Resh, 1995	**Metric**: Channel alternation was rated from optimal (10 points, no channelisation) to poor (3 points, extensive channelisation). **Results**: Channel alternation ratings were distributed across all four rating categories. **My interpretation**: Therefore, I considered channel alternation ratings to have high intra-annual observer variability.
	Moderate	Clapcott, 2015	NA
Riparian vegetation cover	High	Hannaford and Resh, 1995	**Metric**: Streamside cover was rated from optimal (10 points, well developed) to poor (2 points, no vegetation). **Results**: Streamside cover ratings were distributed across all four rating categories. **My interpretation**: Therefore, I considered streamside cover ratings to have high intra-annual observer variability.
	Low	Clapcott, 2015	NA

Riparian vegetation type	High	Wang et al., 1996	**Metric:** Visual estimates of "percentage of land dominated by" woodland (trees > 3 m), shrub (trees < 3m), meadow ("grass and forbs with few woody plants, not subject to regular mowing or grazing") and residential ("land modified for human use") were taken "along the transect line extended 10 m from the water's edge on each bank" (Wang et al., 1996). **Results**: Estimates of bank vegetation-land use had relatively high observer variability.
	Low	Clapcott, 2015	NA
Riparian width	Moderate	Hannaford and Resh, 1995	**Metric**: Riparian zone (width) was rated from optimal (10 points, > 9.1 m wide) to poor (2 points, < 1.8m wide). **Results**: Riparian zone ratings were distributed across three rating categories. **My interpretation**: Therefore, I considered riparian zone ratings to have moderate intra-annual observer variability.
	Moderate	Clapcott, 2015	NA
Hydraulic heterogeneity	High	Clapcott, 2015	NA
Invertebrate habitat	Low	Clapcott, 2015	NA
Quantitative metrics			
Bank stability	High	Archer et al., 2004	**Metric**: Bank stability "Measured at 30-cm rectangular plots at each bank sampling location and calculated as the number of "stable" plots divided by the total number of plots in the reach" (Archer et al., 2004). **Results:** Percent variability for bank stability was 47.21%. **My interpretation**: The results of Archer et al. (2004) match the observer variability scale used in my study. Therefore, I considered bank stability measurements to have high intra-annual observer variability.
Fish cover/ bank undercut	Low	Newcomb et al., 2007	NA
	Low	Wang et al., 1996	**Metric**: Undercut bank "Banks overhanging water by at least 0.3 m. and no more than 0.1 m above water surface, at a point where water is at least 0.3 m deep: measured with a meter stick at transect and reported as a percentage" (Wang et al., 1996).

			Results: Undercut bank measurements had low observer variability, as FMP was not exceeded.
	Low	Archer et al., 2004	**Metric**: Undercut depth “Measured as the maximum distance from under bank to bank edge, average of measurements at 20 locations on both sides of the stream” (Archer et al., 2004). **Results:** Percent variability for undercut depth was 17.34%. **My interpretation**: The results of Archer et al. (2004) match the observer variability scale used in my study. Therefore, I considered bank undercut measurements to have low intra-annual observer variability.
Overhanging vegetation	Moderate	Wang et al., 1996	**Metric**: Overhanging vegetation “Thick vegetation overhanging water and meeting criteria for undercut bank cover: measured with a meter stick at transect and reported as a percentage” (Wang et al., 1996). **Results**: Overhanging vegetation measurement results at two stations were precise, while at one station, FMP was exceeded by 3 points. **My interpretation**: Therefore, I considered overhanging vegetation measurements to have moderate intra-annual observer variability.
Wetted width	Low	Newcomb et al., 2007	NA
	Low	Wang et al., 1996	**Metric**: Wetted width “Wetted width of the stream, perpendicular to the flow, at the existing surface: measured with a 50-m tape at transect” (Wang et al., 1996). **Result**: Wang et al. (1996) found that wetted width and stream depth measurements were the most precise measurements. **My interpretation**: Therefore, I considered wetted width measurements to have low intra-annual observer variability.
Bank-to-bank width	High	Newcomb et al., 2007	NA
Stream depth	Low	Newcomb et al., 2007	NA
	Low	Wang et al., 1996	**Metric**: Stream depth “Vertical distance from streambed to water surface: measured with a meter stick at 4 points along transect” (Wang et al., 1996).

			Result: Wang et al. (1996) found that wetted width and stream depth measurements were the most precise measurements. **My interpretation**: Therefore, I considered stream depth measurements to have low intra-annual observer variability.
Sediment depth	Moderate	Wang et al., 1996	**Metric:** Sediment depth "Depth of fine sediments (sand or silt) that overlie or compose the stream bed: measured with a meter stick at four quadrats along transect" (Wang et al., 1996). **Results**: Sediment depth measurement results at two stations were precise, while at one station, FMP was exceeded by 3 points. **My interpretation**: Therefore, I considered sediment depth measurements to have moderate intra-annual observer variability.
Substrate size/ composition	Low	Archer et al., 2004	**Metric**: D_{50} "Measured as the median diameter of a minimum of 100 particles sampled from three to four consecutive riffles" (Archer et al. 2004). **Results:** Percent variability for D_{50} was 9.56%. **My interpretation**: The results of Archer et al. (2004) match the observer variability scale used in my study. Therefore, I considered D_{50} measurements to have low intra-annual observer variability.