Freshwater Fish in the Waikato Region: An Evaluation of Novel Standardised Data and Drivers of Fish Distribution and Abundance

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

Masters of Science

in

Ecology

at Massey University, Manawatū,

New Zealand.

Paul Boyce

2016

Abstract

Native freshwater fish populations in New Zealand are in rapid decline and continue to be affected by multiple impacts. As fish populations contract and species become rarer, monitoring and prediction of species presence and abundance throughout New Zealand becomes an increasingly important management tool. Until recently, the majority of modelling analyses in New Zealand have used small or inconsistent datasets which limit analyses to presence/absence type assessments and can make even those analyses inaccurate. Using a novel dataset that was collected using a consistent set of sampling protocols collected by Waikato Regional Council (WRC) staff in the Waikato region, New Zealand, fish presence/absence and abundance data were analysed. Models were built to predict species distributions and abundance across the region and to identify key environmental drivers determining distribution and abundance. Model performance was also assessed in comparison to a commonly used freshwater fish dataset – the New Zealand Freshwater Fish Database (NZFFDB). Drivers of the distribution and abundance of all sampled species were identified and three species were investigated in further detail: longfin eels (*Anguilla dieffenbachii*); shortfin eels (*Anguilla australis*); and redfin bullies (*Gobiomorphus huttoni*). Abundance and distribution predictions were mapped throughout the Waikato region River Environment Classification (REC) waterways for these three species. An analysis of population demographics for redfin bully mean length comparing the east and west coasts of the Waikato region was also undertaken.

Sampling consistency (mean button time and mean area fished) within the WRC dataset was high and model performance was higher for the majority of species using the WRC dataset compared with NZFFDB data. Comparisons between reference and impact sites for the three species revealed significantly higher relative abundance of redfin bullies at reference sites, along with significantly larger longfin eels (subject to methodological bias). Shortfin eel relative abundance was significantly higher at impact sites. A greater proportion of very small and small shortfin eels were caught during sampling with longfin eels having relatively greater numbers in larger size classes. Substantial differences in the size of eels caught using different methodologies were also found with larger individuals caught using netting methods compared with electric fishing. Significant differences in mean length between years for redfin bullies were identified and a significant difference in mean length between east and west coast populations was found. This difference was also present within the NZFFDB dataset.

Distribution and population characteristics for the three species were examined in reference to land use and model identified drivers of variation. Distance, elevation, temperature, and slope frequently ranked high as drivers of native fish distribution and abundance. Longfin eels and redfin bullies appear to be excluded from areas of intensive agriculture throughout the central Waikato with redfin bullies in particular predominantly limited to areas of remaining indigenous forest cover. This pattern is mirrored by shortfin eels which have high predicted presence and relative abundance values in lowland agricultural areas. Models of native richness and abundance IBI scores also show this pattern of distribution across the Waikato region. Both native richness and abundance IBI scores are higher at the coast where intensive agriculture is absent and a greater proportion of reference sites are available. Scores consistently decrease moving inland towards lowland catchments at a rate that is likely higher than innate species distributions due to diadromy alone.

Predictive modelling for both abundance and presence data provided extensive mapping opportunities for waterways throughout the Waikato region in the River Environment Classification network. This tool used in conjunction with a robust dataset provided a versatile and accurate method for describing fish populations in the region. Many characteristics of the biology and life history of native species were able to be explored within this study, raising questions about recruitment and population biology of fish species, in particular redfin bullies. The importance of establishing consistent sampling protocols throughout New Zealand freshwater monitoring programmes is exemplified in the ability to uncover these characteristics accurately.

Acknowledgements

This thesis would not have been possible first and foremost without the work of the freshwater monitoring team at Waikato Regional Council. Dr. Bruno David, Callum Bourke, Mark Hamer, and many more staff and contributors have collected and continue to collect an immense and comprehensive dataset throughout the Waikato region. The time and effort that has gone into sampling and sample design of this dataset have provided an enormous amount of information about New Zealand native fish. I would like to further thank Dr. Bruno David for his input and guidance in analyses and writing without which the learning curve would have been substantially steeper. Thank you to my supervisor, Dr. Mike Joy for your patience and guidance throughout the project. Sticking with me as I learned has meant a lot and your dedication to freshwater conservation in New Zealand is infectious. Thank you to those who provided financial support which made this whole project possible and thank you to the staff and students of the Massey University Ecology department for your advice and help throughout the project. Finally, thank you to my family, friends, and partner who provided the support and reassurance I needed.

Additional notes and terminology

This thesis is presented as 5 chapters including an introduction chapter. Some repetition is present between chapters in regards to methodologies and dataset descriptions. Where referred to, abundance or count refers to relative abundance (fish/m^2). Random and impact are synonymous when describing random sites within the Waikato sampling network. All models within this thesis are cross-validated (cv) using data subsets during model construction (Elith, Leathwick, & Hastie, 2008). Terminology for ROC scores of presence/absence model discrimination (performance) categories is taken from Hosmer, Lemeshow, and Sturdivant (2013). Discrimination categories are as follows; outstanding (ROC ≥0.9); excellent (0.8 ≤ ROC <0.9); acceptable (0.7 ≤ ROC > 0.8). Waterways within the Waikato Regional Council monitoring network are 1 – 4th order and waterways 5th order and above are grey in modelling maps.

Waikato Regional Council is often abbreviated to WRC and the New Zealand Freshwater Fish Database to NZFFDB. River Environment Classification (REC) network sites used for environmental variable analysis were taken from the REC version 2 dataset available for download from the National Institute of Water and Atmospheric Research[1] (National Institute of Water and Atmospheric Research, 2013).

---

[1] http://www.niwa.co.nz/freshwater-and-estuaries/management-tools/river-environment-classification-0

TABLE OF CONTENTS

Chapter One

General Introduction

Chapter 1 General Introduction

This is an opportunity for greatness which has never been offered to any civilisation, any generation in any civilisation in human history before. To act as a generation to do the right thing. If we fail to receive that opportunity, to act on it, then my feeling is we will become the most vilified generation that's ever lived in human history.

- Dr Roger Payne, President, Ocean Alliance

Ecological conservation in New Zealand has a long history with active organisations and individuals working to establish nature reserves and protect biota for well over 100 years (Beattie, 2003). As development of natural areas in New Zealand continues to increase, rapid declines in the distribution and scale of New Zealand’s indigenous biodiversity are being seen. The decline of our terrestrial fauna, in particular birds, has been well documented as being largely a result of the introduction of mammalian predators closely following settlement (Bellingham et al., 2010; O'Donnell, Clapperton, & Monks, 2015). Declines in our freshwater biota have similarly been under pressure since settlement although declines in freshwater species are only recently beginning to be representatively documented (Weeks, Anderson-Lederer, Death, & Joy, 2015). Many of the issues affecting freshwater biodiversity in New Zealand are historic. For example, predatory exotic fish (salmonids) were introduced as game species by acclimatisation societies as early as 1867 (Townsend, 1996), and the prolific pest species *Gambusia affinis* was introduced in the 1930’s with multiple introductions following (Purcell, Ling, & Stockwell, 2012). Another international trend coinciding with freshwater biodiversity decline is increasing dam construction and hydrological alteration. The construction of dams and the associated changes in the hydrological regime consistently create negative effects on freshwater ecosystems (Jellyman & Harding, 2012; Millennium Ecosystem Assessment, 2005, p. 30-32). Migratory routes for diadromous species can be entirely interrupted and important upstream and downstream flow dynamics, to which New Zealand native fish have been shown to be adapted, are lost (Leprieur et al., 2006). New Zealand has a history of controversial dam construction for hydro-power and irrigation schemes (e.g. Manapouri, Mokihinui, and Ruataniwha dams) despite multiple campaigns created in an effort to save the associated ecosystems. Globally, dam construction continues to increase and projections for the coming decades indicate double the amount of existing dams will be built (Figure 1.1) (Zarfl, Lumsdon, Berlekamp, Tydecks, & Tockner, 2015). As a result of combinations of these impacts, the trend in freshwater biodiversity decline is occurring globally, and at present freshwater ecosystems have the highest proportion of species threatened with extinction of any ecosystem (Millennium Ecosystem Assessment, 2005, p. 4).

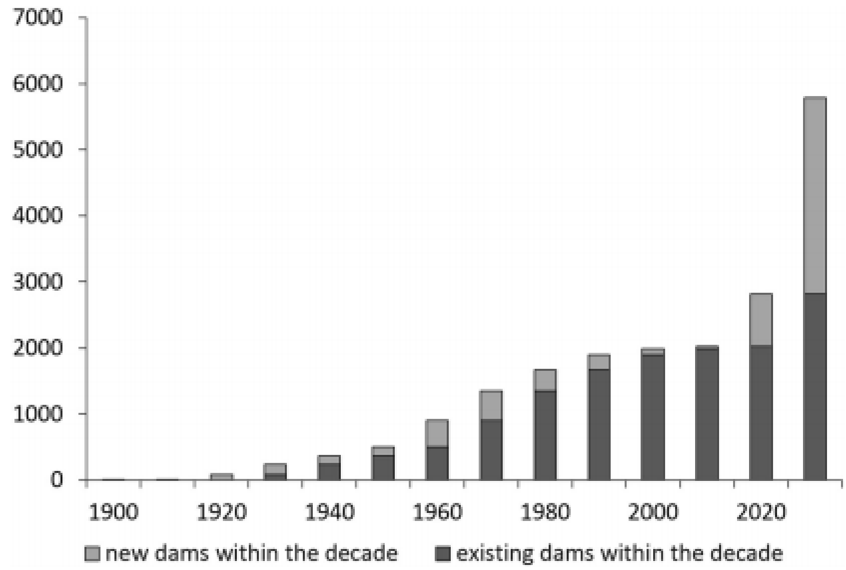


Figure 1.1 Changes in dam construction over the last century and projections into the future. Taken from Zarfl et al. (2015).

In recent decades, pressure on New Zealand native fish has come in many forms including habitat loss, commercial fishing, point source discharges from wastewater plants and factories, damming and hydrological alteration, and a suite of factors associated with agricultural development. The latter presents perhaps the greatest threat to freshwater ecosystems in recent New Zealand history (Stow, Maclean, & Holwell, 2014). Impacts from intensive agriculture include, nitrate leaching, phosphorous runoff, stock intrusion into waterways, bank erosion, sedimentation of waterways, hydrological alteration and habitat loss, and effluent discharges directly into waterways. Land continues to be cleared and converted to intensive farms and habitat and biodiversity loss is following accordingly (Clapcott et al., 2012; McDowall, 2006; McDowell, 2015; McDowell, Larned, & Houlbrooke, 2009).

This loss of biodiversity carries a number of implications for New Zealanders. Many species are caught for food either recreationally and/or commercially (e.g. whitebait and eel species) and many also have cultural importance for Māori. The loss of these species as a food resource has already been highlighted in significant drops in catch volumes of eels (Allibone et al., 2010; Boubee, Chisnall, Watene, Williams, & Roper, 2002) and in the fact that 4 of our 5 native species caught as whitebait are now threatened. Additionally, New Zealand's native fish are a diverse and specialised group. Their distribution and life history have been shaped by many factors including the unique physical environment New Zealand waterways create, and the often turbulent geological history of the country (McDowall, 1990, 2006). Population distributions can differ remarkably between and within species and these distributions are often related to the high proportion of diadromous life histories: diadromous populations occurring in relative proximity to coastal areas and non-diadromous species more limited to specific inland catchments. These distributions are themselves variable however, and diadromy may be facultative for some species (Closs, Smith, Barry, & Markwitz, 2003; Leathwick, Elith, Chadderton, Rowe, & Hastie, 2008; Ward, Northcote, & Boubee, 2005). Within diadromous species, precise migratory behaviours and breeding strategies have been identified (Closs, Hicks, & Jellyman, 2013; Jarvis & Closs, 2015), along with adaptive advantages and habitat selection in relation to frequently, naturally disturbed environments, afforded by both diadromous life histories and behavioural plasticity (David & Closs, 2002; Leathwick et al., 2008). There are species particularly adept at climbing such as the koaro (*Galaxias brevipinnis*), able to scale waterfalls; there are fast water adapted species such as the torrentfish and blue gill bully which have developed specific physical adaptations for fast water environments; and even aestivating mudfish (McDowall, 1990). In recent years, developments in taxonomic methodologies have discovered at least six species found within one previously thought single species of galaxiid, *G. vulgaris* with some of these species having extremely restricted geographical ranges (McDowall, 2010). This depth in the diversity and distribution of freshwater fish in New Zealand is becoming better understood, however much still needs to be learnt about our freshwater biota and how best to conserve and study it. Without a better understanding of species and their population characteristics, the high degree of endemic species with distinctive traits stands to be lost through ineffective management and monitoring (Allibone et al., 2010; McDowall, 1990, 2006).

Crucial to understanding how native fish populations are changing and functioning is quantifying native fish distribution and identifying differences in distributions within and between species. With respect to conservation, frequent advances in statistical methods, computing power, and sampling technology are enabling complex ecological assessments of species characteristics to be undertaken more rapidly and with far greater spatial coverage (Elith et al., 2006). Consequently, such advances should enable more effective assessment of freshwater species and ensure relevant information is used to inform appropriate conservation actions in the highest priority places. It is only with these accurate and consistent monitoring approaches that freshwater ecologists will be able to effectively describe New Zealands' native fish populations on a scale relevant to species conservation.

Although many indicators of water quality and biotic metrics have been developed and used in New Zealand, data collection protocols are less developed. These have hampered the use of data for important analyses of metrics such as relative abundance, and even accurate measures of presence/absence when collection methods are unknown. Through the special interest group SWIM (Surface Water Integrated Management) local council bodies within New Zealand have been tasked with testing standardised sampling protocols for wadeable streams in New Zealand based on a USEPA sampling methodology developed through the Environmental Monitoring and Assessment Programme (EMAP) (EPA, 1998). The approach to sampling in New Zealand has also taken into account the diverse range of fluvial and physical environments that are known to greatly affect native fish distributions (e.g. (Jowett & Richardson, 1996, 2003; Joy, Henderson, & Death, 2000; Leathwick et al., 2008; McDowall, 2006, 2007; McEwan & Joy, 2009). David et al. (2010), investigated further the fishing effort required to detect >90% of fish species present in wadeable streams in New Zealand. From this combination of sampling strategies, a standardised set of protocols was developed for freshwater fish data collection in New Zealand (Joy, David, & Lake, 2013). The data used in this thesis represent data collected to these protocols over the past 5 years by the Waikato Regional Council Freshwater Monitoring team. This is a novel dataset appropriate not only for relative abundance analyses but theoretically of comparative accuracy for other population analyses. Population differences between and within common New Zealand native species and key drivers of distribution and relative abundance are discussed. Additionally, closer examination of one species (redfin bully) is used as an example to examine geographical differences in the size of this species between the east and west coasts of the Waikato region. The analyses and discussion in this thesis go some way to outline what can be achieved with these recent developments and techniques in sampling and ecological science when used in a regional and national context in New Zealand.

Aim of research

The aim of this research was to determine environmental and land use drivers affecting fish populations in the Waikato region of New Zealand, and to develop accurate models to predict the presence/absence and relative abundance of selected species. Anecdotal differences in fish length of redfin bullies between the east and west coast of the Waikato were examined in detail to identify the size of differences, and potential drivers of this demographic variation. Analyses of fish populations in the Waikato were undertaken using a novel dataset to assess characteristics of several fish species in light of new, robustly collected data. Modelling and prediction of species distributions has been undertaken in New Zealand, however, generally only with presence/absence data extracted from the New Zealand Freshwater Fish Database (NZFFDB). In addition to the very coarse presence/absence assessment, effective analysis is also hampered by a lack of data collection consistency and methodology. A model is only as good as the data it is based on and theoretically a model built on a suitability robust dataset should provide meaningful predictions about the state of freshwater fish populations in New Zealand in relation to environmental variables. In this instance, it should provide more accurate identification of population differences and characteristics within the Waikato region. Assessments of improvements gained through use of the standardised protocols, and an assessment of the level of consistency routinely achievable through implementation of these protocols was undertaken to highlight both areas for potential improvement in freshwater monitoring in New Zealand, and to identify limitations of currently used datasets.

References

Allibone, R., David, B., Hitchmough, R., Jellyman, D., Ling, N., Ravenscroft, P., & Waters, J. (2010). Conservation status of New Zealand freshwater fish, 2009. *New Zealand Journal of Marine and Freshwater Research, 44*(4), 271-287. doi: 10.1080/00288330.2010.514346

Beattie, J. (2003). Environmental anxiety in New Zealand, 1840-1941: climate change, soil erosion, sand drift, flooding and forest conservation. *Environment and History, 9*(4), 379-392. doi: 10.3197/096734003129342881

Bellingham, P. J., Towns, D. R., Cameron, E. K., Davis, J. J., Wardle, D. A., Wilmshurst, J. M., & Mulder, C. P. H. (2010). New Zealand island restoration: seabirds, predators, and the importance of history. *New Zealand Journal of Ecology, 34*(1), 115-136.

Boubee, J., Chisnall, B., Watene, E., Williams, E., & Roper, D. (2002). Enhancement and management of eel fisheries affected by hydroelectric dams in New Zealand. In D. A. Dixon (Ed.), *Biology, Management, and Protection of Catadromous Eels* (Vol. 33, pp. 191-205). Bethesda: Amer Fisheries Soc.

Clapcott, J. E., Collier, K. J., Death, R. G., Goodwin, E. O., Harding, J. S., Kelly, D., . . . Young, R. G. (2012). Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity. *Freshwater Biology, 57*(1), 74-90. doi: 10.1111/j.1365-2427.2011.02696.x

Closs, G. P., Hicks, A. S., & Jellyman, P. G. (2013). Life histories of closely related amphidromous and non-migratory fish species: a trade-off between egg size and fecundity. *Freshwater Biology, 58*(6), 1162-1177. doi: 10.1111/fwb.12116

Closs, G. P., Smith, M., Barry, B., & Markwitz, A. (2003). Non-diadromous recruitment in coastal populations of common bully (Gobiomorphus cotidianus). *New Zealand Journal of Marine and Freshwater Research, 37*(2), 301-313.

David, B. O., & Closs, G. P. (2002). Behavior of a stream-dwelling fish before, during, and after high-discharge events. *Transactions of the American Fisheries Society, 131*(4), 762-771. doi: 10.1577/1548-8659(2002)131<0762:boasdf>2.0.co;2

David, B. O., Hamer, M. P., Collier, K. J., Lake, M. D., Surrey, G. M., McArthur, K., . . . Dale, M. (2010). A standardised sampling protocol for robust assessment of reach-scale fish community diversity in wadeable New Zealand streams. *New Zealand Journal of Marine and Freshwater Research, 44*(3), 177-187.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., . . . Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography, 29*(2), 129-151. doi: 10.1111/j.2006.0906-7590.04596.x

EPA (1998). *Environmental Monitoring and Assessment Program (EMAP): Research plan 1997* (EPA/620/R-98/002). Washington, DC: US Environmental Protection Agency.

Jarvis, M. G., & Closs, G. P. (2015). Larval drift of amphidromous Gobiomorphus spp. in a New Zealand coastal stream: a critical spatial and temporal window for protection. *New Zealand Journal of Marine and Freshwater Research, 49*(4), 439-447. doi: 10.1080/00288330.2015.1072569

Jellyman, P. G., & Harding, J. S. (2012). The role of dams in altering freshwater fish communities in New Zealand. *New Zealand Journal of Marine and Freshwater Research, 46*(4), 475-489. doi: 10.1080/00288330.2012.708664

Jowett, I. G., & Richardson, J. (1996). Distribution and abundance of freshwater fish in New Zealand rivers. *New Zealand Journal of Marine and Freshwater Research, 30*(2), 239-255.

Jowett, I. G., & Richardson, J. (2003). Fish communities in New Zealand rivers and their relationship to environmental variables. *New Zealand Journal of Marine and Freshwater Research, 37*(2), 347-366.

Joy, M. K., David, B., & Lake, M. (2013). *New Zealand Freshwater Fish Sampling Protocols*. Palmerston North, New Zealand: Massey University

Joy, M. K., Henderson, I. M., & Death, R. G. (2000). Diadromy and longitudinal patterns of upstream penetration of freshwater fish in Taranaki, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 34*(3), 531-543.

Leathwick, J. R., Elith, J., Chadderton, W. L., Rowe, D., & Hastie, T. (2008). Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *Journal of Biogeography, 35*(8), 1481-1497. doi: 10.1111/j.1365-2699.2008.01887.x

Leprieur, F., Hickey, M. A., Arbuckle, C. J., Closs, G. P., Brosse, S., & Townsend, C. R. (2006). Hydrological disturbance benefits a native fish at the expense of an exotic fish. *Journal of Applied Ecology, 43*(5), 930-939. doi: 10.1111/j.1365-2664.2006.01201.x

McDowall, R. M. (1990). *New Zealand freshwater fishes : a natural history and guide* (Rev. ed ed.). Wellington [N.Z.]: Heinemann Reed, MAF Publishing Group.

McDowall, R. M. (2006). Fish, fish habitats and fisheries in New Zealand. *Aquatic Ecosystem Health & Management, 9*(4), 391-405. doi: 10.1080/14634980601026352

McDowall, R. M. (2007). On amphidromy, a distinct form of diadromy in aquatic organisms. *Fish and Fisheries, 8*(1), 1-13. doi: 10.1111/j.1467-2979.2007.00232.x

McDowall, R. M. (2010). *New Zealand freshwater fishes. [electronic resource] : an historical and ecological biogeography*: Dordrecht ; New York : Springer, c2010.

McDowell, R. W. (2015). Relationship between Sediment Chemistry, Equilibrium Phosphorus Concentrations, and Phosphorus Concentrations at Baseflow in Rivers of the New Zealand National River Water Quality Network. *Journal of Environmental Quality, 44*(3), 921-929. doi: 10.2134/jeq2014.08.0362

McDowell, R. W., Larned, S. T., & Houlbrooke, D. J. (2009). Nitrogen and phosphorus in New Zealand streams and rivers: control and impact of eutrophication and the influence of land management. *New Zealand Journal of Marine and Freshwater Research, 43*(4), 985-995.

McEwan, A. J., & Joy, M. K. (2009). Differences in the distributions of freshwater fishes and decapod crustaceans in urban and forested streams in Auckland, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 43*(5), 1115-1120.

Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: synthesis*. Retreived from http://www.millenniumassessment.org/documents/document.356.aspx.pdf

O'Donnell, C. F. J., Clapperton, B. K., & Monks, J. M. (2015). Impacts of introduced mammalian predators on indigenous birds of freshwater wetlands in New Zealand. *New Zealand Journal of Ecology, 39*(1), 19-33.

Purcell, K. M., Ling, N., & Stockwell, C. A. (2012). Evaluation of the introduction history and genetic diversity of a serially introduced fish population in New Zealand. *Biological Invasions, 14*(10), 2057-2065. doi: 10.1007/s10530-012-0213-1

Stow, A., Maclean, N., & Holwell, G. I. (Eds.). (2014). *Austral ark : the state of wildlife in Australia and New Zealand*.

Townsend, C. R. (1996). Invasion biology and ecological impacts of brown trout Salmo trutta in New Zealand. *Biological Conservation, 78*(1-2), 13-22. doi: 10.1016/0006-3207(96)00014-6

Ward, F. J., Northcote, T. G., & Boubee, J. A. T. (2005). The New Zealand common smelt: biology and ecology. *Journal of Fish Biology, 66*(1), 1-32. doi: 10.1111/j.1095-8649.2004.00569.x

Weeks, E. S., Anderson-Lederer, R., Death, R. G., & Joy, M. K. (2015). *Are we ignoring the plight of New Zealand's freshwater species.* Society for Conservation Biology (Oceania), Sydney.

Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L., & Tockner, K. (2015). A global boom in hydropower dam construction. *Aquatic Sciences, 77*(1), 161-170. doi: 10.1007/s00027-014-0377-0

Chapter Two

Evaluation of Improvements in Analysis of Fish Communities Resulting from Standardised Sampling Methodologies in the Waikato Region, New Zealand

Chapter 2 Evaluation of Improvements in Analysis of Fish Communities Resulting from Standardised Sampling Methodologies in the Waikato Region, New Zealand

Introduction

Animals can appear ubiquitous and abundant despite having rapidly decreasing populations. The diversity of species and populations of single species are often spread over large spatial scales and can exist in varying degrees at different times of the year. What determines this variation is itself variable, and that variation does not remain constant (Glova, Jellyman, & Bennett, 2001; Jowett & Richardson, 1996; McDowell, 2015). Ultimately then, it can be very difficult to monitor or even obtain baseline information to describe animal distributions (Leathwick, Elith, Francis, Hastie, & Taylor, 2006). Undertaking research with aquatic species is also logistically challenging with observations and capture requiring different skills and often specialised equipment. Furthermore, species may be difficult to locate and some species can be highly cryptic (e.g. Raadik (2014)). Many records of aquatic populations are sporadic or at the least incomplete with inconsistent sampling, and ecology is often a multi-discipline approach to fill these gaps in sampling by defining and measuring distributions of undetected individuals or populations (Elith et al., 2006). Finally, although not always applicable or even achievable in ecological monitoring programmes, controlled and repeatable sampling remain crucial when evaluating perceived spatial and temporal changes in populations (Cohen, 1988). Effective management decisions are predicated on these accurate understandings of native fish population dynamics.

Consequently, when monitoring aquatic populations it becomes vitally important that sampling is consistent both with respect to method and temporal and spatial scale of assessment to minimise noise within data as much as possible. This affords researchers confidence that even though they are unlikely to be capturing the entire population, representative and relative comparisons may be made between years. This becomes even more important when a species is studied by different researchers and comparisons are intended to be made between studies (Meixner et al., 2013; Rekha et al., 2015). However, sampling consistency is hard to achieve between projects and often the objectives of sampling differ. For example, objectives determining sampling within this data set are to obtain effective reach-scale fish diversity and relative abundance measures. Reach scale and fishing methodologies between and even within most monitoring programmes in New Zealand however, do not result in data that can reliably make these types of assessments between waterways, over time restricting these types of analyses to a regional or even waterway/reach context where data was collected. Many of the trends and analyses undertaken on these datasets collected with non-standardized methodologies inherently contain substantial amounts of noise resulting from these differing approaches in data collection (Clark, 2004).

As freshwater monitoring programmes are often established by local government departments to monitor trends in water quality, nationally applied sampling guidelines should be established to monitor diversity and relative abundance in freshwater species (Joy, David, & Lake, 2013). As New Zealand does not have state governance additional to regional and national governance, as in larger countries, nationally implemented sampling measures would be comparatively easy to co-ordinate. This would ensure robust and accurate national and regional comparisons of freshwater fish diversity and abundance are able to be undertaken and effective management plans produced. That this is not the default situation is not surprising given the wide range of organisations sampling aquatic species in some form and generally by the comparative paucity of studies undertaken with freshwater species (Tonkin, Death, & Barquin, 2013). However, to accurately understand, manage and subsequently conserve New Zealand’s freshwater biodiversity these standardisations of sampling must be implemented.

Unique to the Waikato region is a freshwater monitoring programme conducted by Waikato Regional Council (WRC) staff that has collected data to a set of protocols yet to be introduced to council monitoring programmes on a national level. Research within the Waikato region has been pivotal in the development and production of the current New Zealand Freshwater Fish Sampling Protocols (Joy et al., 2013; David et al., 2010). These protocols are designed to minimise methodological variation such as fishing effort, fishing area, equipment used, and identification methodologies to improve the robustness of collected data: data which has been subsequently used in this thesis. Furthermore, accurate length measurements of fish caught provide information about recruitment and size interactions amongst populations of native fish generally unavailable on a wide spatial and temporal scale in New Zealand. Typical analyses of native fish data in New Zealand rely on the New Zealand Freshwater Fish Database (NZFFDB) – a large, partially open data resource held by the National Institute of Water and Atmospheric Research (NIWA). This has become an immense and valuable resource, however sampling effort and methodology are not necessarily entered with fish data. Consequently, analyses must be limited to presence/absence, and even analysis of presence/absence data is subject to limitations of sampling inconsistency: sampling effort and identification can and do vary between single data entries making analyses variable and inaccurate.

This chapter investigated the level of consistency achieved in sampling through application of standardised protocols in the Waikato region, New Zealand. Sampling metrics were analysed to assess consistency between years throughout a network of sites and subsequent analyses of fish species (chapters 3, 4, and 5) were undertaken to assess improvements in accuracy. Comparisons were also made using boosted regression tree models between the WRC dataset and the New Zealand Freshwater Fish Database (NZFFDB).

Methods

For the past 6 years the freshwater ecology team at Waikato Regional Council (WRC) has used techniques to collect data used in these analyses, that later became the basis of the New Zealand Freshwater Fish Sampling protocols (Joy et al., 2013). Although not all sites have a complete 6 year sampling history all sites have been sampled consistently and the majority have temporal dimensions of two or more years. Reference sites are sampled annually with a 3 year rotational sampling of random (“impact”) sites. Reference sites selected based on minimal anthropogenic catchment and waterway disturbance were established where possible (i.e. reference sites are rare), however random sites are spread throughout land use and ecosystem types (Figure 2.1).

Data collection

The dataset is composed of samples from reference sites and random sites. Network reference sites that are accessible are fished each year with random (impact) sites being fished on a rotating schedule. Fish samples are collected in conjunction with invertebrate and habitat sampling which comprises a total of one hundred and eighty sites, 60 of which are sampled each year, with each site being sampled once every three years. Of the 60 invertebrate and habitat sites sampled each year, 40-50 of these are able to be fished. Streams are sampled by either electric fishing or fyke nets and minnow traps with each reach sampled measured at 150m in length and divided into ten, 15m subreaches. Fishing method selection is based on method decision scoring sheet which indicates the most appropriate method for a site based on a range of site characteristics (e.g. conductivity, turbidity, and depth): detailed prescriptive protocols can be found in Joy et al. (2013). Count data and fish length data for a suite of species are collected during fishing events with all species except *Gambusia affinis* being counted in totality: counts of *Gambusia* are estimated into bins of 1-10, 10-100, 100-1000 due to the propensity for this species to often attain very high abundances (1000’s). With respect to measurement of fish length, the first 50 individuals of each species is measured and then an additional 10 fish of that species (first 10 captured) per subsequent subreach after that. In effect if a species is extremely abundant throughout the entire sampling reach, a total of 140 individuals will be measured for that species (50 in first subreach + 10x 9 subsequent subreaches). Table 2.1 outlines the sample structure of the WRC dataset.

Table 2.1 Outline of site types (reference and impact) and sampling methodology as sampled by sampling season in the Waikato network. Counts represent number of sites sampled in a given year in each category. EFM refers to electro-fishing method, and NET to net capture methods.

Site and Sample Type	**2009/10**	**2010/11**	**2011/12**	**2012/13**	**2013/14**	**2014/15**
Impact	0	0	15	39	52	44
EFM	0	0	15	39	44	30
NET	0	0	0	0	8	14
Reference	3	8	9	9	9	9
EFM	3	8	9	9	9	9
Total	3	8	24	48	61	53

The dataset was compiled originally as Microsoft Excel files with counts and lengths organised in separate worksheets. Data cleaning was required to correct mistakes (largely in assigned River Environment Classification (REC) network reach ID values – used for environmental variable analysis in chapters 4 & 5) and to create files appropriate for analyses in R (R Core Team, 2014). Additional information was provided for fishing sites including REC reach ID numbers and associated environmental data including conductivity, temperature, and water visibility measured at each sampling event. Figure 2.1 shows the sampled sites from the WRC network that were used in analyses. The sites in green (reference) are described in further in Table 3.1 in chapter 3.

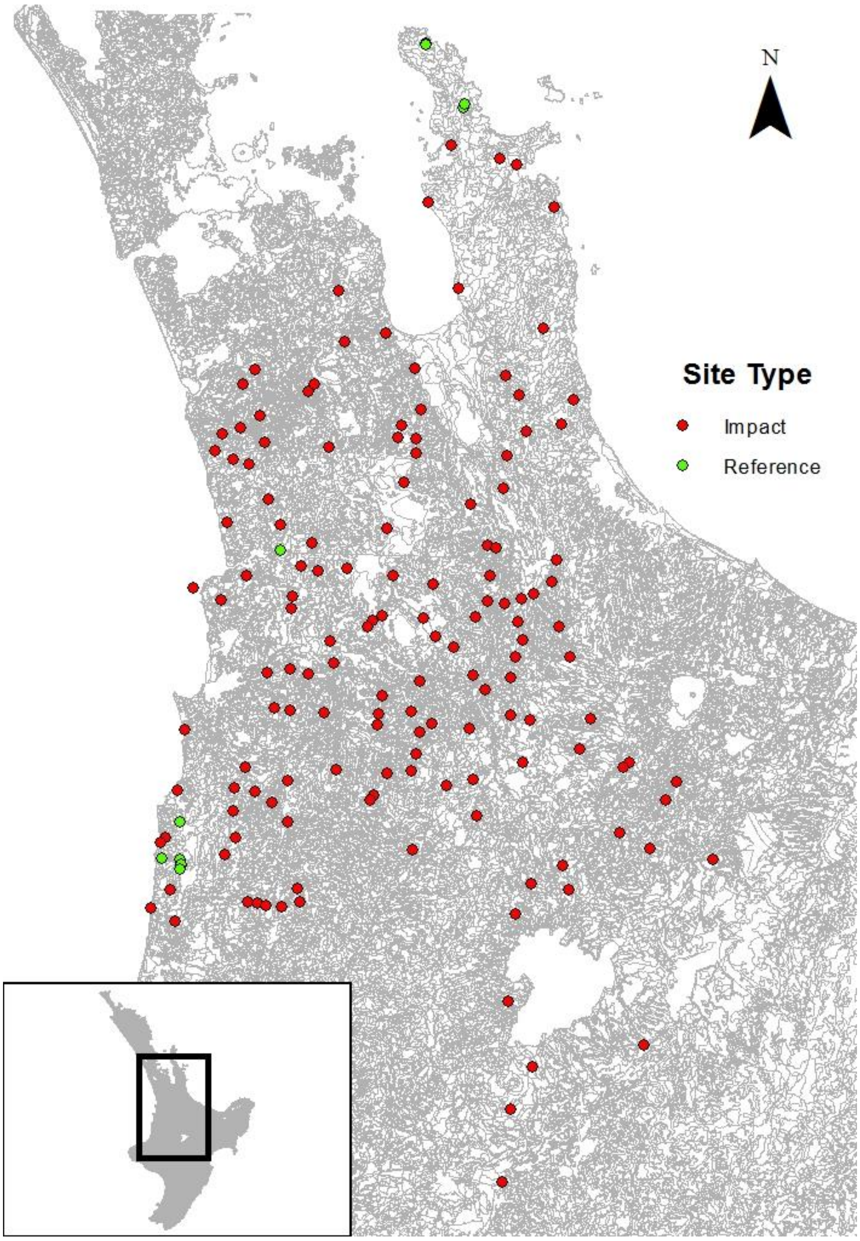


Figure 2.1 Waikato Regional Council Monitoring network sites. Sites were sampled from 2009-2014 by Waikato Regional Council Staff from the freshwater monitoring team. Sites in red are random (impact) sites sampled on a 3 year rotation. Green sites are annually sampled reference sites.

Data analysis

Mean button time (time electric fishing machine was in use) for electric fishing and mean area fished at reference sites were examined. Only reference sites were used as these have been sampled more than twice. As described above, most random sites have not yet been sampled twice. However, identical procedures are followed for both reference and random sites during sampling.

Fishing effort was also examined for NZFFDB data and this is determined as area fished *or* number of nets set in NZFFDB records however many entries do not have records of fishing effort. Mean fishing effort (area) from the NZFFDB data was therefore analysed for sites in the Waikato region fished using an electric fishing machine (EFM) for comparison with WRC sites fished using EFM. Only waterways that have been fished more than once and have records of fishing effort were selected for comparison from NZFFDB data.

Multi-variate analyses (boosted regression trees (BRT); described below) for three species well-represented in the dataset (*Anguilla dieffenbachii, Anguilla australis, Gobiomorphus huttoni*) were also compared between the NZFFDB and WRC built models for presence/absence data in the Waikato Region (decision processes for species selection are further outlined in chapter 3). Analyses of boosted regression tree differences were undertaken as these were built and subsequently used to map species distributions throughout the Waikato region (Chapter 4). Comparing model performance between the WRC dataset and the NZFFDB dataset was conducted to provide an indication of the differences in accuracy of predictive models built from datasets commonly used in New Zealand.

BRT models were also used to test the improvements in model performance when using presence/absence versus relative abundance data. Models were built using presence/absence or abundance data and then the ability to successfully detect whether a site was a reference or an impact site was assessed for each model.

Boosted Regression Trees (BRT)

Regression trees are a decision based statistical method of predicting response variables in reference to relationships with input variables. Trees are built by splitting all data into clusters reflecting dominant relationships with response variables. A constant is then fitted to each split point as a result of classification tree algorithms describing the most likely class resulting in that particular cluster. Mean responses are then fitted to all observations in that cluster with split points being chosen to minimise prediction errors. Boosting then provides a means to improve the accuracy of the model, by iteratively taking a random subset of the dataset, and sequentially improving model predictions on the hardest observations to predict. The final model is a combination of trees, describing a regression relationship with multiple terms (Elith, Leathwick, & Hastie, 2008). Code to produce boosted regression trees was sourced from Elith et al. (2008) and the rules of thumb outlined in their paper were used for tree complexity, bag fractions, and learning rates. As a recommended starting point all trees were built using a bag fraction of 0.75, a learning rate of 0.001, and a tree complexity of 5.

Results

Sampling consistency

Fishing effort has been consistent throughout the sampling process. Mean button time and the mean area fished for reference sites which have been sampled each year of monitoring are consistent throughout sampling years (Figures 2.2 and 2.3). Each site will have a unique button time due to varying widths, substrates, and morphologies of the waterway, however times are consistent with standard error of the mean less than 10 minutes for any of the sites. Further, mean area fished at each site varies little with the biggest site standard error less than 50m^2 (49.2 at site 474_5). Average standard error across these sites was 19.4 m^2.

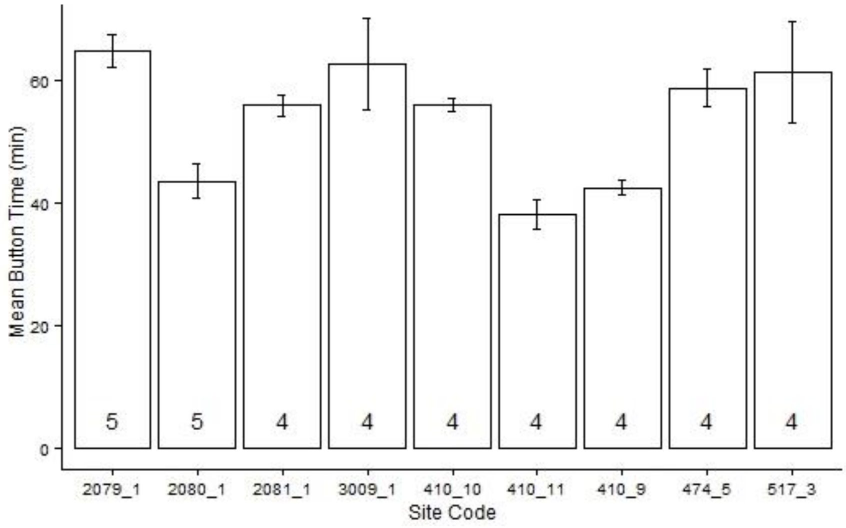


Figure 2.2 Mean button time for reference sites in the WRC monitoring network[2]. N values shown at base of bars represent the number of years each site has been sampled. Error bars are standard error of the mean.

---

[2] Site Codes are described in table 3.1, chapter 3.

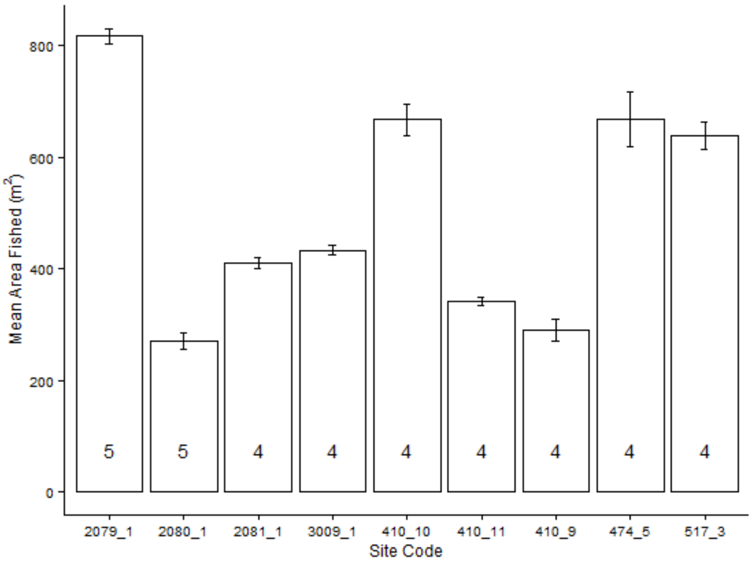


Figure 2.3 Mean area fished for reference sites in the WRC monitoring network. N values shown at base of bars represent the number of years each site has been sampled. Error bars are standard error of the mean.

Fishing effort (area) is more variable within NZFFDB (Figure 2.4) sites in comparison to WRC sites (Figure 2.3). Out of a potential 112 waterways which have been fished more than once using an electric fishing machine in the Waikato region over the same time period as the WRC dataset (2009-2014), 16 had records of fishing effort. Some sites have particularly high standard errors despite comparatively high N values (c.f. low N in WRC sites). Site 10 for example has a standard error of 242.5 m^2 (c.f. 49.2 m^2 for the most variable WRC site). Average standard error across these NZFFDB sites was 53.1 m2

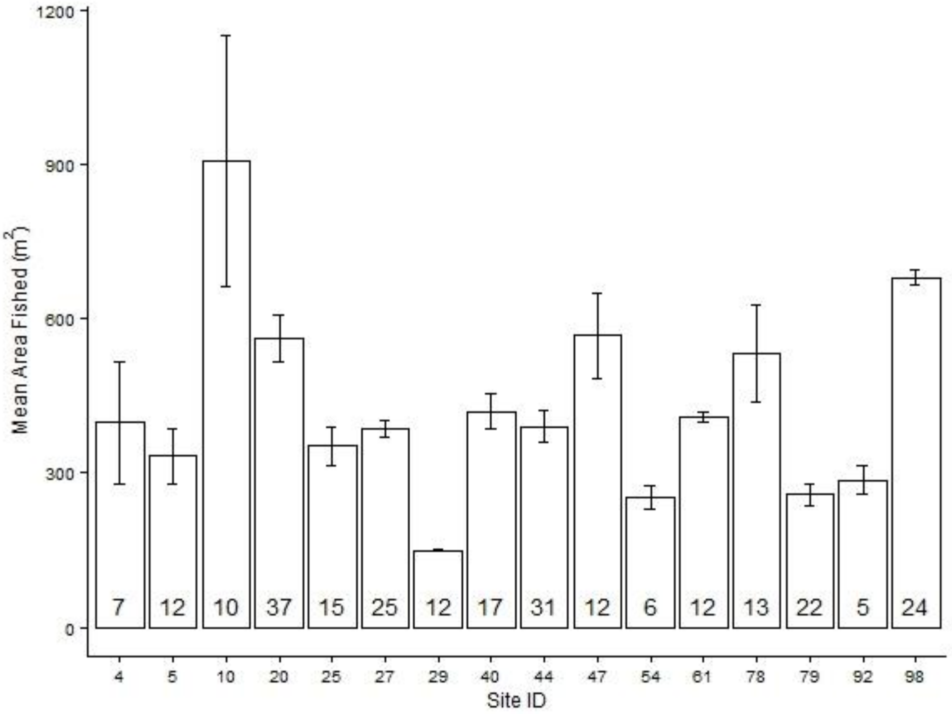


Figure 2.4 Mean area fished for EFM sites in the Waikato region from the NZFFDB. N values shown at base of bars represent the number of times each site has been sampled. Site IDs are described in appendix A, Table A-2. Error bars are standard error of the mean.

Comparison of predictive accuracy between models built from data collected using protocols and data collected not using protocols

In all species WRC data models performed better than those built with NZFFDB data from the Waikato region. ROC scores were higher and model deviance was lower for presence/absence analyses using WRC data compared with NZFFDB data (Table 2.2). For both redfin bullies and longfin eels, WRC data built models were “outstanding” and, the shortfin eel model was “excellent”. All NZFFDB models performed lower and were “acceptable” (Hosmer, Lemeshow, & Sturdivant, 2013). Similarly, relative abundance models all performed higher using WRC data than those built using NZFFDB data. Training and test data correlation was higher and model deviance was lower using WRC data compared with NZFFDB data (Table 2.3).

Table 2.2 Boosted regression tree (BRT) model outputs (ROC score and deviance) for presence/absence BRT simplified models using WRC data and NZFFDB data.

Species	**NZFFDB ROC Score**	**NZFFDB deviance**		**WRC ROC Score**	**WRC deviance**
Anguilla dieffenbachii	0.77	0.86		0.957	0.531
Anguilla australis	0.775	0.841		0.889	0.8
Gobiomorphus huttoni	0.78	0.931		0.986	0.269

Table 2.3 Boosted regression tree (BRT) model outputs (training and test data correlation and deviance) for relative abundance BRT simplified models using WRC data and NZFFDB data.

Species	**NZFFDB correlation**	**NZFFDB deviance**		**WRC correlation**	**WRC deviance**
Anguilla dieffenbachii	0.267	11.31		0.679	0.019
Anguilla australis	0.384	52.733		0.474	0.039
Gobiomorphus huttoni	0.382	50.019		0.694	0.051

Comparison between WRC models built using presence/absence and relative abundance data

The performance of models built to determine whether a site was an impact or a reference site was slightly higher for the model built using relative abundance data compared to presence/absence data. The ROC score was higher for relative abundance while model deviances were similar for both (Table 2.4)

Table 2.4 BRT model performance for WRC data built models to predict if sites are impact or reference sites. Models were built based on relative abundance or presence/absence data from the WRC dataset.

Model Data	**ROC score**	**Deviance**
Relative Abundance	0.995	0.29
Presence/Absence	0.977	0.28

Discussion

Predictive modelling using regression trees revealed that data collected using a set of protocols was more accurate than data collected without using a set of protocols, demonstrating the importance of using consistent and tested sampling methods. The Waikato Regional Council dataset was collected using a set of protocols which are effectively the New Zealand Freshwater Fish Sampling Protocols (Joy et al., 2013). These protocols were put in place to provide guidelines for freshwater ecologists when undertaking sampling or monitoring of fish species within New Zealand. The guidelines were intended to produce reliable datasets where actual differences in species populations can be examined while minimising the methodological sampling noise. As described, the National Institute of Water and Atmospheric Research (NIWA) holds a national dataset (the NZFFDB) consisting of thousands of entries of fish samples throughout the country with a time range of close to a century. This database is a great resource for examining historical fish presence, however there is comparatively little spatial or temporal consistency in sampling within any of the records in the dataset, and often there is no record of how a site may have been sampled. Further inconsistencies include records where no fish were caught during sampling not having been entered, or conversely, records where fish were caught also not being entered ultimately making the understanding of populations less clear. This high degree of variability begins to affect the level of accuracy of analyses and interpretations of fish populations.

Consistency of sampling is of course vitally important and a central tenant of any sampling design and this particularly comes in to play when making relative comparisons between years at a site for a given species. The level of consistency in sampling that can be achieved through implementing standardised sampling methods is high (Figures 2.2 and 2.3). The type of consistency typical in NZFFDB data (Figure 2.4), from which comparisons and analyses of New Zealand fish species have frequently been made both temporally and spatially, and in relation to environmental variables is comparatively low (e.g. Jowett & Richardson (2003)). Many species, such as eels, are already difficult to monitor in terms of total population as different cohorts within the population inhabit vastly different habitats (McDowall, 1990). For example, the large majority of small eels will remain in the sediment of a waterway to avoid predation (Glova, 2001). Netting methods will typically underestimate the numbers of these smaller individuals. Conversely, large eels will typically inhabit deep slow moving water, and areas with in-stream debris, for which electric fishing is often ineffective, underestimating large eels in the population (Baillie, Hicks, van den Heuvel, Kimberley, & Hogg, 2013; Chisnall & Hicks, 1993; Jellyman, Chisnall, Sykes, & Bonnett, 2002). The type of waterway and sampling methodology then can be strongly influential on population estimates. Multiple method use and varied fishing effort, as is generally the rule in the NZFFDB, begins to introduce additional types of unnecessary errors. Consequently, conclusions about fish species from the NZFFDB dataset may be inaccurate with little opportunity to determine how inaccurate.

Consistency in sampling effort and methods has been achieved with the above WRC dataset through strict governance of fishing area and time. A national dataset that was collected in the same manner as the WRC dataset would provide substantially improved predictions for species across the wide range of ecosystems present in New Zealand. Such a dataset would cover a far greater range of REC classes and waterway types than the WRC dataset has done. The reduction in noise around model building and predictions gained from consistent sample collection would allow for a far greater understanding of native fish populations at a national level.

Statistical methods will continue to improve accuracy and versatility of making predictions in species occurrence and abundance however models are only as accurate as the datasets they are based on (Clark, 2004; Cohen, 1988). Further, the opportunities for a greater detail of analyses becomes apparent when sampling error is minimised. For example, Indices of Biotic

Integrity (IBI) which are increasingly used to assess the health of waterways can offer a detailed picture of stream communities and health (Joy, 2005; Joy & Death, 2004; Stark & Maxted, 2007). Being able to use abundance IBIs produces far more accurate picture of the stream community, however this is predicated on reliable abundance data. This has been shown previously in New Zealand freshwater fish data where reduction in variance was observed when using abundance IBI calculations versus presence/absence IBI calculations for the same dataset (Joy, 2013).The NZFFDB provides an immense data resource with thousands of entries for fish samples which represent the sampling effort of many individuals, councils and/companies. It is a shame that the usefulness of such a resource is limited somewhat by variable sample collection and that even presence/absence analyses may be subject to sampling noise. Differences between councils and anthropogenic regions in regards to objectives for freshwater monitoring introduce further noise and to accurately monitor aspects such as relative abundance and reach scale diversity on a national or regional scale, protocols such as those used by Waikato Regional Council in collecting this dataset should be followed. These protocols have not only created a more accurate dataset, but have provided a wider range of metrics able to be analysed, that are important in understanding fish populations.

Models built using relative abundance data also proved to be more accurate than those built using presence/absence data from the WRC dataset and were better able to predict whether a site was an impact or reference site. This improvement again highlighted the benefits of collecting consistent abundance data. Model improvements were small although this may be related to a relatively small dataset in particular the small number of available reference sites. Assessing the performance of relative abundance versus presence/absence models in years to come following the collection of a larger dataset, may illustrate greater improvements in model performance with similar analyses when using relative abundance data.

In conclusion, consistency in data collection provides several benefits to understanding native fish populations. Freshwater ecologists can be confident that relative differences observed in populations between years are actual rather than due to sampling methodology. The improvements in model performance when using data from the WRC dataset versus those from the NZFFDB further demonstrate the types of gains that may be achieved by employing a consistent set of protocols. It is recommended, with demonstrations throughout this thesis, that protocols similar to these be adopted nationwide in all sampling efforts. This approach will provide robust and accurate datasets through which populations of native fish can be more effectively managed by utilising a more accurate understanding of population dynamics, seasonal changes, habitat preferences, and ecosystem interactions between and within species.

References

Baillie, B. R., Hicks, B. J., van den Heuvel, M. R., Kimberley, M. O., & Hogg, I. D. (2013). The effects of wood on stream habitat and native fish assemblages in New Zealand. *Ecology of Freshwater Fish, 22*(4), 553-566.

Chisnall, B. L., & Hicks, B. J. (1993). Age and growth of longfinned eels (anguilla-dieffenbachii) in pastoral and forested streams in the Waikato river basin, and in 2 hydroelectric lakes in the North Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 27*(3), 317-332.

Clark, M. (2004). *A first course in applied statistics : with applications in biology, business and the social sciences* (SprintPrint ed ed.). Auckland: Pearson Education New Zealand.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*: Hillsdale, N.J. : L. Erlbaum Associates, 1988, 2nd ed.

David, B. O., Hamer, M. P., Collier, K. J., Lake, M. D., Surrey, G. M., McArthur, K., . . . Dale, M. (2010). A standardised sampling protocol for robust assessment of reach-scale fish community diversity in wadeable New Zealand streams. *New Zealand Journal of Marine and Freshwater Research, 44*(3), 177-187.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., . . . Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography, 29*(2), 129-151. doi: 10.1111/j.2006.0906-7590.04596.x

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802-813. doi: 10.1111/j.1365-2656.2008.01390.x

Glova, G. J. (2001). Effects of the presence of subadult longfinned eels (Anguilla dieffenbachii) on cover preferences of juvenile eels (Anguilla spp.) in replicate channels. *New Zealand Journal of Marine and Freshwater Research, 35*(2), 221-233.

Glova, G. J., Jellyman, D. J., & Bennett, M. L. (2001). Spatiotemporal variation in the distribution of eel (Anguilla spp.) populations in three New Zealand lowland streams. *Ecology of Freshwater Fish, 10*(3), 147-153. doi: 10.1034/j.1600-0633.2001.100304.x

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*: Hoboken, New Jersey : Wiley, [2013]3rd ed.

Jellyman, D. J., Chisnall, B. L., Sykes, J. R. E., & Bonnett, M. L. (2002). Variability in spatial and temporal abundance of glass eels (Anguilla spp.) in New Zealand waterways. *New Zealand Journal of Marine and Freshwater Research, 36*(3), 511-517.

Jowett, I. G., & Richardson, J. (1996). Distribution and abundance of freshwater fish in New Zealand rivers. *New Zealand Journal of Marine and Freshwater Research, 30*(2), 239-255.

Jowett, I. G., & Richardson, J. (2003). Fish communities in New Zealand rivers and their relationship to environmental variables. *New Zealand Journal of Marine and Freshwater Research, 37*(2), 347-366.

Joy, M. K. (2005). *A Predictive Model of Fish Distribution and Index of Biotic Integrity (IBI) for Wadeable Streams in the Waikato Region*. Hamilton, New Zealand: Environment Waikato.

Joy, M. K. (2013). *Freshwater fish predictive modelling for bioassessment; A scoping study into fish bioassessment models as national indicators in New Zealand*. Wairesearch.

Joy, M. K., David, B., & Lake, M. (2013). *New Zealand Freshwater Fish Sampling Protocols*. Palmerston North, New Zealand: Massey University

Joy, M. K., & Death, R. G. (2004). Application of the index of biotic integrity methodology to New Zealand freshwater fish communities. *Environmental Management, 34*(3), 415-428. doi: 10.1007/s00267-004-0083-0

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series, 321*, 267-281. doi: 10.3354/meps321267

McDowall, R. M. (1990). *New Zealand freshwater fishes : a natural history and guide* (Rev. ed ed.). Wellington [N.Z.]: Heinemann Reed, MAF Publishing Group.

McDowell, R. W. (2015). Relationship between Sediment Chemistry, Equilibrium Phosphorus Concentrations, and Phosphorus Concentrations at Baseflow in Rivers of the New Zealand National River Water Quality Network. *Journal of Environmental Quality, 44*(3), 921-929. doi: 10.2134/jeq2014.08.0362

Meixner, M. D., Pinto, M. A., Bouga, M., Kryger, P., Ivanova, E., & Fuchs, S. (2013). Standard methods for characterising subspecies and ecotypes of Apis mellifera. *Journal of Apicultural Research, 52*(4), 5-5.

R Core Team. (2014). R: A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Raadik, T. A. (2014). Fifteen from one: a revision of the Galaxias olidus Gunther, 1866 complex (Teleostei, Galaxiidae) in south-eastern Australia recognises three previously described taxa and describes 12 new species. *Zootaxa, 3898*(1), 1-+.

Rekha, P. N., Gangadharan, R., Ravichandran, P., Mahalakshmi, P., Panigrahi, A., & Pillai, S. M. (2015). Assessment of impact of shrimp farming on coastal groundwater using Geographical Information System based Analytical Hierarchy Process. *Aquaculture, 448*, 491-506.

Stark, J. D., & Maxted, J. R. (2007). A biotic index for New Zealand's soft-bottomed streams. *New Zealand Journal of Marine and Freshwater Research, 41*(1), 43-61.

Tonkin, J. D., Death, R. G., & Barquin, J. (2013). Productivity-diversity relationships for stream invertebrates differ geographically. *Aquatic Ecology, 47*(1), 109-121.

Chapter Three

Population Characteristics of Fish Species in the Waikato Region – Analysis from Data Collected

Chapter 3 Population Characteristics of Fish Species in the Waikato Region – Analysis from Data Collected

Introduction

Assessing species populations over temporal and spatial scales requires consistent sampling and accurate datasets. The New Zealand Freshwater Fish Database (NZFFDB) is often used as a source for analysis of fish distributions in New Zealand. The NZFFDB is a collection of data sampled by a wide range of industries and researchers using a range of methods and techniques, resulting in highly variable sampling effort and ability. Such data may be acceptable for presence/absence analysis at a given site although estimating abundance and other metrics crucial to understanding populations is often inappropriate due to substantial sampling noise (Clark, 2004). Furthermore, presence/absence analyses must be undertaken with caution as sampling information is often lacking and can vary widely between data points within the NZFFDB. Through increasing awareness of the limitations of this dataset, the New Zealand Freshwater Fish Sampling Protocols (NZFFSP) have been developed with the intention of encouraging freshwater ecologists to collect data in a more regionally and nationally comparable way (Joy, David, & Lake, 2013). These protocols remain voluntary however the Waikato Regional Council (WRC) freshwater monitoring team, members of which helped develop the NZFFSP, have implemented these protocols since 2009 and now possess a detailed and consistent dataset composed of comparable fish length and abundance data. Analyses of freshwater fish from this dataset offer an improved and more accurate picture of fish populations and drivers of population dynamics in the Waikato region.

The WRC dataset presents an opportunity to gain a far greater understanding of species in the Waikato region and to examine in greater detail population characteristics, species richness, and biotic index scores. Additionally, species and environmental features of sites can be analysed and mapped to give an accurate picture of site characteristics within the monitoring network. With the added ability to interrogate accurate fish size and abundance data, dynamics of recruitment, inter and intra-specific interactions with age cohorts, and spatial and temporal variance in size and abundance characteristics can begin to be examined on a regional scale. For example, temporal changes in species relative abundance will provide much earlier signs of population contraction than presence/absence data can. Analysis of this abundance data may also help elucidate the drivers of population spatial contraction in changing environments (Stapley, Garcia, & Andrews, 2015). This study goes someway in exploring questions of population characteristics for New Zealand fish species within the Waikato region in relation to this novel Waikato dataset. A number of species are monitored in the WRC monitoring programme, however spatial coverage is variable for some species. Accordingly, investigations in this study were broad for the majority of species, and more detailed for those with appropriate representation. Key differences in species and population distributions of 3 of the most common species (*Anguilla dieffenbachii* – longfin eels*; Anguilla australis* – shortfin eels*;* and *Gobiomorphus huttoni* – redfin bullies) were examined in more detail and general descriptions of these species follows.

Redfin Bullies

Redfin bullies are small (typically 80-100mm) diadromous freshwater fish (McDowall, 1990). A benthic fish found in high quality coastal and near-coastal waterways in New Zealand, they are an amphidromous eleotrid with characteristic red stripes on the dorsal, anal, and tail fins (Tonkin, Wright, & David, 2012). Redfin bullies have a diurnal habit like many New Zealand fish returning to larger substrate areas during the day after increased nocturnal activity (McEwan & Joy, 2014). Records in the NZFFDB suggest this species is more common close to the coast but has been found as far as 250km inland (NIWA, 2015). The greatest distance inland redfin bullies are detected within the WRC dataset is approximately 68 km.

Longfin and shortfin eels

There are two main species of freshwater eel in New Zealand; the longfin eel and the shortfin eel. The longfin eel (*Anguilla dieffenbachii*) is locally common but in decline in New Zealand (Jellyman, 2007). Longfin eels are endemic to New Zealand and have a strong cultural significance with Māori as a traditional food source (Jellyman, Booker, & Watene, 2009). Females can grow up to 2 metres long and live as long as 100 years before migrating to sea to spawn and die. Their long growth time until sexual maturity and single spawning at the end of their life has made them vulnerable to overfishing predominantly with commercial harvesting. Longfin eels are also comparatively sensitive to pollution and along with the removal of instream habitat such as woody debris, longfin eels, particularly larger individuals are becoming less common (Allibone et al., 2010; Glova, Jellyman, & Bennett, 2001; Jellyman, 2007; McCleave & Jellyman, 2002).

The shortfin eel (*Anguilla australis*), is found in New Zealand, south-eastern Australia and the western Pacific Islands (McDowall, 2006). This species is generally smaller than the longfin eel with a smaller dorsal fin but otherwise looks similar to the longfin. The shortfin is often found in polluted or degraded waters and it has been suggested that its range and abundance has increased in some areas as habitat has become degraded, and observed in some studies, at the expense of longfin eels (Chisnall, 1994; Chisnall & Hicks, 1993; Jellyman, 1991; Jowett, Richardson, & Boubee, 2009). As the extent of quality habitat reduces then, and commercial fishing of eels continues, the distribution and abundance of longfin eels are likely to decrease while the distribution and abundance of shortfins are likely to increase.

In general, juveniles of both species prefer to reside deep within the streambed to avoid predation and cannibalism by larger conspecifics, while adults prefer more diverse habitat (e.g. cobbles, woody debris) in open water, particularly longfin eels (Baillie, Hicks, van den Heuvel, Kimberley, & Hogg, 2013; Glova, 2001; Jellyman & Chisnall, 1999).

Methods

Choice of species for analyses

There is a temptation to analyse and model all species populations within the WRC dataset to gain a picture of all native fish encountered in the region. However, factors of restricted range or abundance (anthropogenic or natural) must be considered when analysing any species in detail. As the temporal resolution of the random and reference site dataset improves, greater opportunities to investigate more species will arise, however, current limitations of spatial and temporal coverage of species in the WRC dataset must be established prior to detailed analyses. To establish these restrictions in the Waikato dataset, initial basic analyses were undertaken to determine how abundant and widespread each species was within the dataset.

There was a distinct bias in species representation within the WRC dataset with a few species represented far more than most others (Figure 3.1). In conjunction with Figure B-1 (appendix B) which illustrates total abundance of species by year, these Figures show each species representation in the dataset. A cut-off point of 5% network site coverage was chosen in that species that were present in less than 5% of the network sites were not investigated further. Species models (chapter 4) were developed for species above the 5% cut-off. As it was impractical to undertake a detailed investigation into each species within the timeframe of this thesis, and inappropriate for many species given the scarcity of spatial coverage, three common species were chosen for more detailed analysis within this chapter, and for the modelling chapter (chapter 4): these are; redfin bullies (*Gobiomorphus huttoni*) and the two eel species, longfin and shortfin eels (*Anguilla dieffenbachii* and *Anguilla australis* respectively).

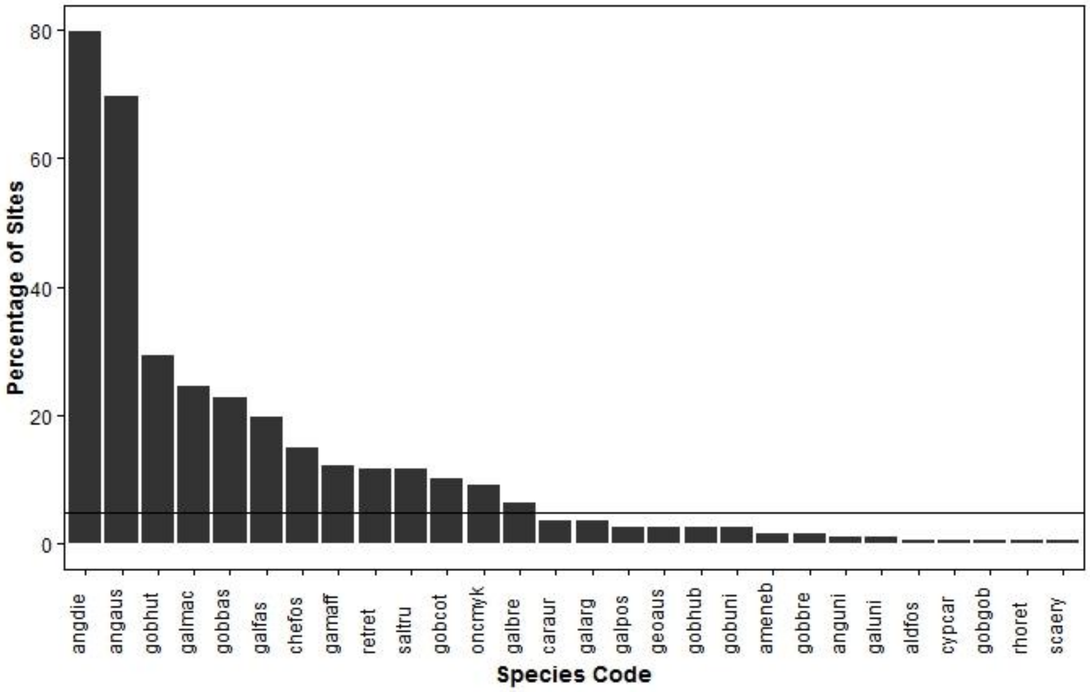


Figure 3.1 Percentage of sites in the WRC network each species was caught for all years with a 5% cut-off line shown. Percentages are calculated for all years and represent total coverage in the current WRC dataset. Commonality between years did not change greatly (Figure B-1 appendix B)[3].

Descriptive statistics of fish assemblages

Several analyses were undertaken to describe the dataset. General sample site and network characteristics were evaluated to determine similarities and differences between sites. Differences in altitude and distance, as well as indigenous forest cover between sites was a focus due to the strong influence of these variables on New Zealand fish biota (Jowett & Richardson, 2003; Leathwick, Elith, Chadderton, Rowe, & Hastie, 2008; McDowall, 1990). Reference sites, selected for their relative lack of modification (i.e., control sites), were described in greater detail to demonstrate their representative position in the North Island and to demonstrate the difficulty in finding representative reference sites for waterways in New Zealand.

Aspects of length characteristics in relation to capture methodologies were examined for eels as fishing method has been identified as a source of potential bias in New Zealand eel species previously (Jellyman & Graynoth, 2005). Other analyses included native richness and abundance IBI scores. Native richness in this instance was native species count with abundance IBI scores calculated based on an IBI model developed by Joy (2005). IBIs have been developed using presence/absence data and are used to varying degrees both in New Zealand and internationally (Joy & Death, 2004). Abundance IBIs have been shown to be more accurate than

---

[3] Table A-3, appendix A shows species codes and common names for each species

presence/absence IBIs however, robust datasets with consistent sampling of abundance data are needed (Joy, 2013). This IBI takes into account the relative and overall abundance of native species and elevation and distance from sea provides a weighted index score for a given waterway (i.e., a higher score indicates a more pristine waterway). The IBI includes trout (brown and rainbow trout) as “honorary native” species in the due to their ubiquity and need for high quality water (Joy, 2005). IBI scores vary from 0-120 with 120 being the highest possible score indicating a site with the highest possible fish assemblage integrity in the region.

Mapping of native richness and IBI scores

ARCGIS (ARCMAP 10.1) was used to map native richness and IBI scores throughout the WRC network sites.

Non-metric multi-dimensional scaling (NMDS)

NMDS provides a visualisation of data similarities and helps to interpret ecological data. Non-metric scaling uses the dissimilarity *ranks* of data points as opposed to the actual values of dissimilarity, and measures the lack of fit between dissimilarities in a *stress* value. This is an approach to visualising data with minimal assumptions and is suitable for data that does not follow typical distributions, which is characteristic of ecological data (Buja et al., 2008; Gu, Singh, & Townsend, 2015; Huene & Vega, 2015; Wellnitz, 2015). NMDS analysis was conducted on all sites to assess dissimilarity distances between them based on the relative abundance of all 28 species recorded during WRC sampling. Where sites had been sampled more than once, as in reference sites and the few impact sites sampled twice, mean relative abundance for a given site was calculated prior to NMDS analysis to avoid duplicates of individual sites within the matrix. Matrix axes were correlated against River Environment Classification (REC) environmental variables associated with each site to identify the variables highly correlated with the NMDS axes differentiating fish assemblage communities.

Detailed analyses of selected species

For the three species investigated in detail, one-way ANOVA analyses were undertaken to determine the significance of temporal trends and impact versus reference differences in mean length and relative abundance. Analysis of variance (ANOVA) methods to test mean values of each group using linear model functions were used (R Core Team, 2014). Following ANOVA analysis, post-hoc HSD (Tukeys’, honest significant difference) analyses were used to differentiate significant differences (Coolidge, 2013). This analysis reveals where significant differences lie by assigning a letter to groups with groups assigned the same letter representing non-significant differences. Power testing was also undertaken to assess the measure of the significance of that change or difference, and the likelihood of type I and type II error. The power of a statistical test, "…is the probability that it will yield statistically significant results" (Cohen, 1988). A full examination and example of a power analysis testing for similar data is appended (appendix C). Specifically section 3.2 of appendix C details the concept and importance of incorporating power.

Results

Site characteristics

The reference sites selected by WRC are typically much closer to the coast than impact sites and generally have a high percentage of indigenous forest cover (Table 3.1; Figure 3.2). Geographical descriptions of random (impact) sites are provided in appendix A although summary comparisons of impact versus reference sites are also shown here (Figures 3.2 and 3.3). Differences in altitude between site types are less apparent than distance although impact sites cover a larger range of altitude (Figure 3.3). Mean and median for distance from the sea for impact sites are 132.1 and 121.5 km, respectively. For altitude at impact sites, the mean and median are 149.7 and 97.5 metres above sea level respectively (c.f. reference sites; Table 3.1).

Table 3.1 Reference site characteristics from Waikato Regional Council (WRC) monitoring network. Location and environmental variables shown. Full list of site characteristics for random (impact) sites shown in appendix A (Table A-1).

Name	**Site Code**	**Distance from Sea (km)**	**Altitude (m above sea level)**	**Catchment Indigenous Forest Cover[4] (%)**
Manganui River	410_9	28	134	99
Manganui River	410_10	31.8	187	83
Manganui River	410_11	28.4	150	94
Mangatu Stm	474_5	3	15	95
Mataiterangi Stm	517_3	3.25	15	65
Maunganui Stream	2081_1	6.4	40	99
Paparahia Stm	3009_1	3.7	40	90
Stony Bay Stream 2nd order	2080_1	2.1	75	100
Stony Bay Stream 3rd order	2079_1	1.6	60	95

---

[4] Derived from River Environment Classification Network

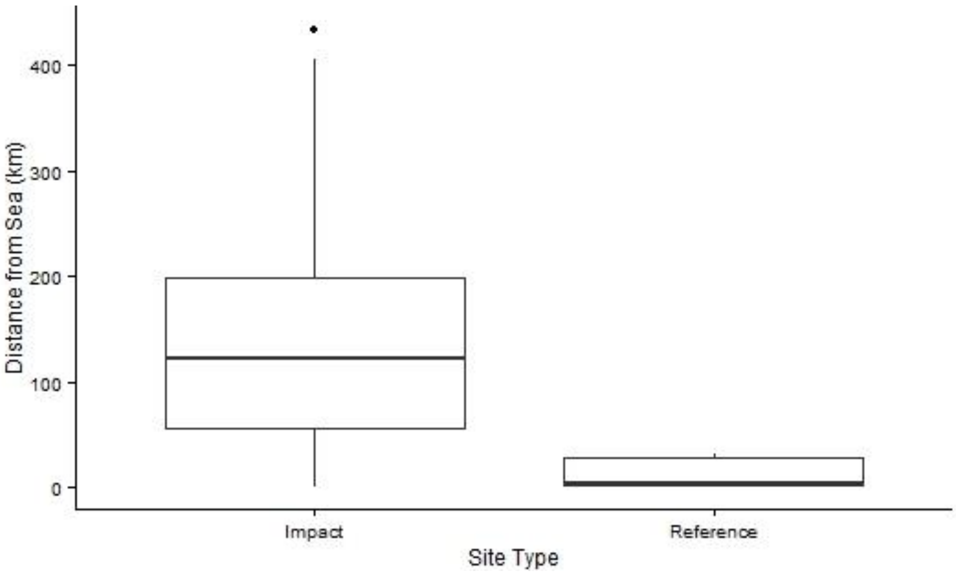


Figure 3.2 Boxplots comparing differences in distance from sea between reference and impact sites in the WRC sampling network[5]. Outliers are shown as points. Dark lines are showing median values.

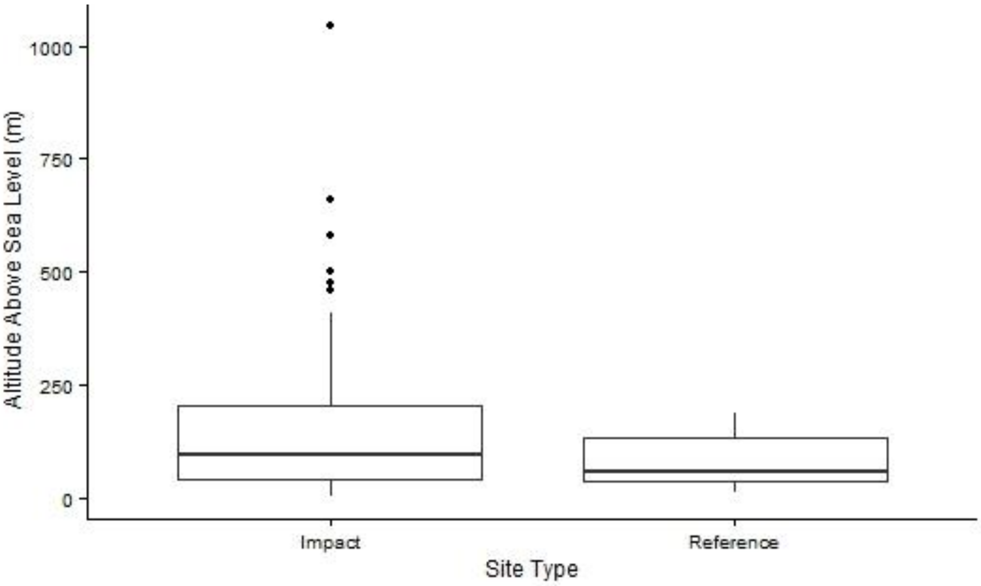


Figure 3.3 Boxplots comparing differences in altitude between reference and impact sites in the WRC sampling network. Outliers are shown as points. Dark lines are showing median values.

Non-Metric multi-dimensional scaling

NMDS analyses did not illustrate any obviously discernible groupings of sites in relation to relative abundance of species. Points were labelled with distance and altitude markers and sites did not show any strong groupings in regards to these variables (Figure 3.4). Sites that deviated

---

[5] Some of the clustering of reference sites closer to the coast is due to the Lake Karapiro hydro-dams removing the opportunity for reference sites above Lake Karapiro.

the most from others tended to have low or no counts of most species. The stress value for the NMDS ordination was > 0.21 indicating generally poor ordination, making interpretation of groupings dubious.

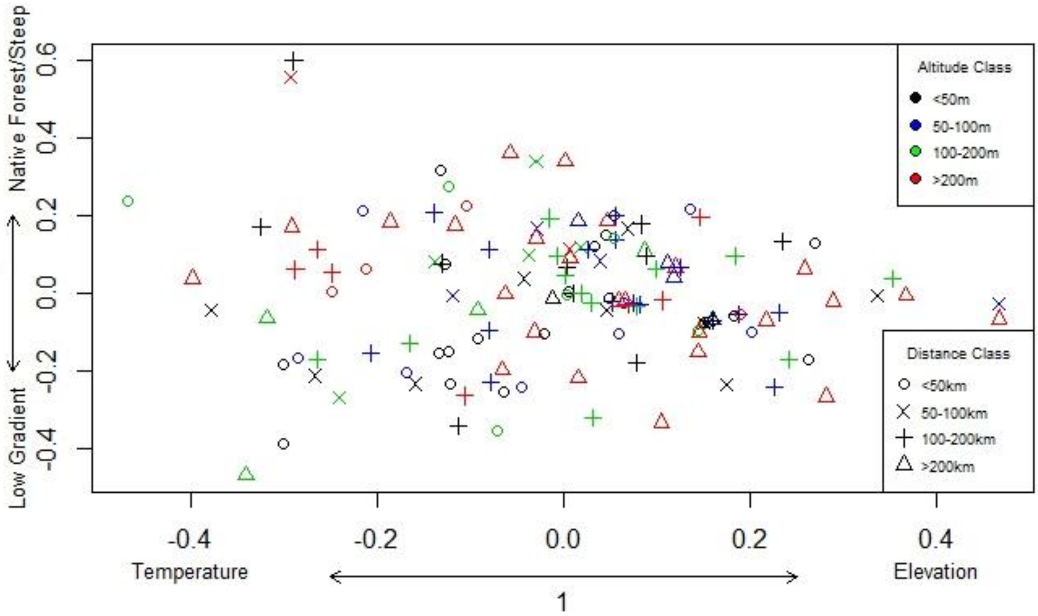


Figure 3.4 NMDS dissimilarity matrix for WRC network sampling sites based on relative abundances for all 28 species routinely recorded, plotted against axes 1 and 2 (stress = 0.21). For sites sampled more than once, average relative abundance for each species at the site was calculated to avoid duplicates. Sites were labelled for both altitude class (colour), and distance from sea class (symbol). Variables along axes are taken from Table B-3 and are most influential drivers of site position within the matrix.

Temperature, elevation, native forest cover and gradient had the strongest correlations with site position in the NMDS matrix. Specifically, altitude and proportion of catchment rain days had the highest correlations. Indigenous forest cover of both the segment and the upstream catchment also had high r values for axis 2. Current summertime equilibrium temperature and historic summertime equilibrium temperature feature high for axis 1 as do segment average January (warm) and June (cold) temperatures. Distance also had comparatively high r values in relation to the remaining variables for both axes (axis 1 = 0.513, axis 2 = 0.309) (Table B-3, appendix B).

Native richness and IBI maps

Native species richness decreases as distance from the coast increases (Figure 3.5). Areas of highest native richness values are along the Raglan coast and on the Coromandel Peninsula. Abundance IBI scores show a similar pattern to native species richness calculations with higher values towards the coast and a collection of high scores on the Coromandel peninsula (Figure 3.6). There are also higher scores around Lake Taupo which are not apparent in richness scores.

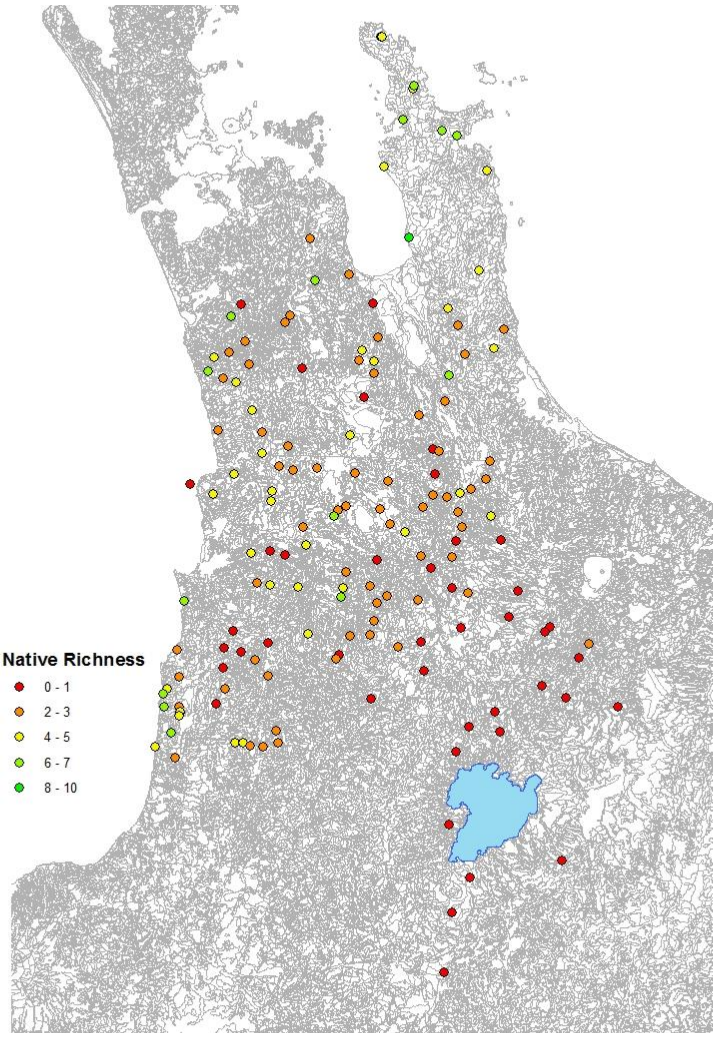


Figure 3.5 Native species richness (species count) at all sites sampled in the Waikato Regional Council monitoring network.

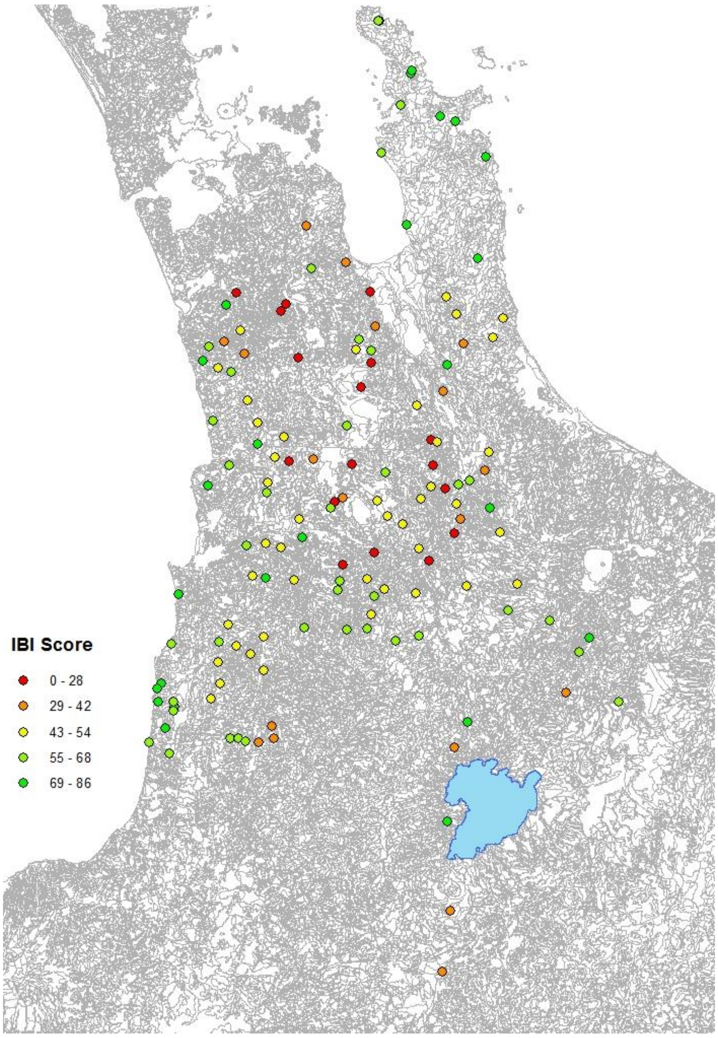


Figure 3.6 Abundance IBI scores (0-120 as calculated based on abundance IBI model derived from Joy (2005)) at all sites sampled in the Waikato Regional Council monitoring network. Lake Taupo in blue highlighting influence of trout dominated areas in high IBI score sites: trout are "honorary natives" in abundance IBI calculations (Joy, 2005).

Individual species characteristics

Eels

The mean relative abundance and lengths of eels caught in the Waikato region varied over the survey period 2009-2014 for the Waikato Regional Council dataset[6] (Figures 3.7 - 3.9). The most appropriate years to compare for impact sites are 2011 and 2014 as these represent the same sites sampled following a full sampling cycle. ANOVA analysis of both longfin and shortfin eel abundance over the sampling period was not significantly different between years for either impact or reference sites (Table 3.2). Mean relative abundance was not different between reference sites compared with impact sites for longfin eels. Mean relative abundance was particularly high in 2010 although this may be a reflection of a larger proportion of reference

---

[6] Data in this study are collected in wadeable streams which adds some bias towards smaller streams.

sites sampled that year (Figure 3.7). Mean relative abundance of shortfin eels was significantly higher at random sites compared with reference sites ($F_{1, 186}$=10.7, p=0.00128) (Figure 3.8).

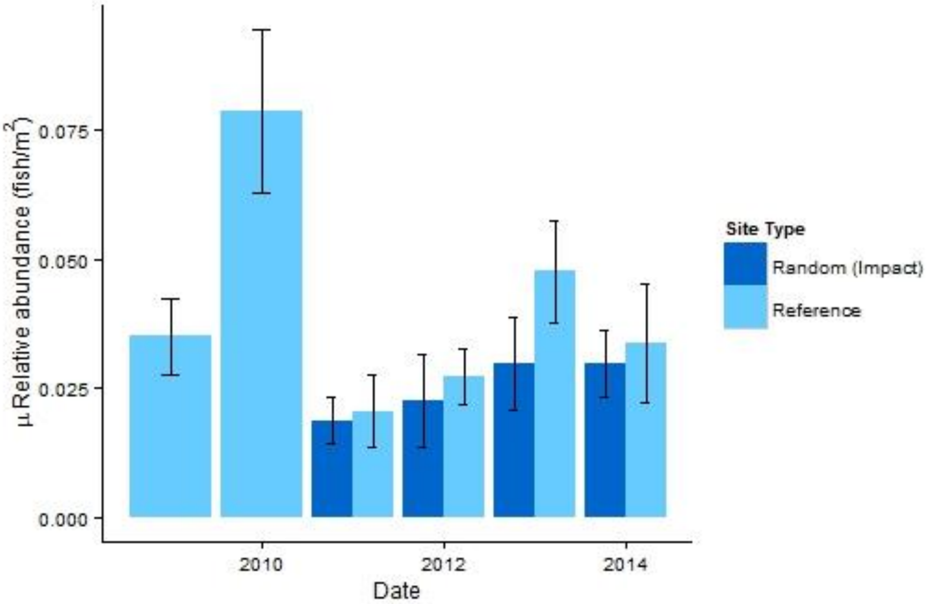


Figure 3.7 Bar plot for mean longfin eel (*Anguilla dieffenbachii*) relative abundance throughout sampling period. Bars are mean relative abundance of eels during each year of sampling for reference and random sites. Error bars are standard error of the mean.

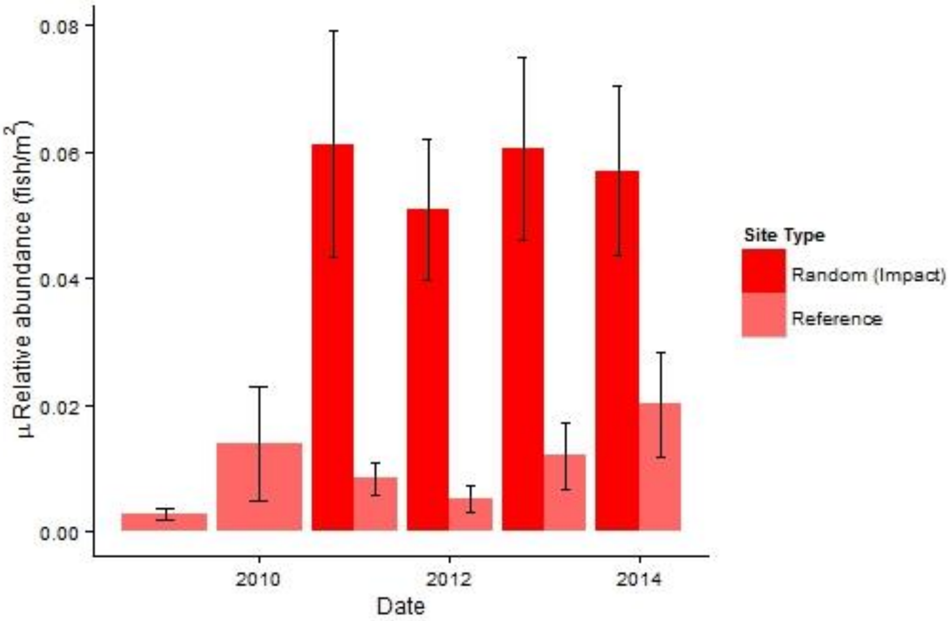


Figure 3.8 Bar plot for mean shortfin eel (*Anguilla australis*) relative abundance throughout sampling period 2009-2014. Bars are mean relative abundance of eels during each year of sampling for reference and random sites. Error bars are standard error of the mean.

The length of eel species was different at reference sites compared with impact sites. Mean length was higher at reference sites for longfin eels ($F_{1, 1668}=222.4$, $p<0.000$), while mean length was higher at random sites for shortfin eels ($F_{1, 2396}=129.6$, $p<0.000$). These differences however are subject to differences in capture methodology discussed further below.

There were differences between years for the length of longfin eels. ANOVA analysis identified significant differences between years for eels caught via EFM at impact sites ($F_{1, 885}=17.88$, $p<0.000$), and those caught via netting at impact sites ($F_{1, 146}=8.86$, $p=0.0034$). No reference sites were fished with netting methods and longfin eels caught via EFM at reference sites did not show a significant difference for length and date ($F_{1, 633}=0.15$, $p =0.70$). A significant difference was found for length and date with shortfin eels caught at impact sites via EFM ($F_{1, 2060}=53.12$, $p=<0.000$), however for EFM caught shortfin eels at reference sites there were no significant differences ($F_{1, 212}=1.62$, $p=0.20$). Again, no reference sites were fished with nets and shortfin eels caught at impact sites via netting showed no significant differences for mean length ($F_{1, 120}=3.42$, $p=0.067$) (Figure 3.9). The substantial differences in length found between net capture and electric fishing methodologies for both eel species are shown more concisely in Figure 3.10 with median length higher for net caught eels.

Abundance in various size classes also differed for both eel species with a predominance of very small shortfin eels, and small longfin eels in the dataset (Figures 3.11 and 3.12). Groupings of size classes for both the entire dataset and for a selection of 6 reference sites displayed variable size distributions for both species (size classes – 1 = <100mm, 2 = 100-300mm, 3 = 300-500mm, 4 = >500mm). Size class information in Figure 3.11 shows a much higher number of small shortfin eels (size class 1) than longfin eels and similar numbers between species of class 2 sized eels. The larger size classes (3 and 4) show higher numbers of longfin eels than shortfins with shortfins least represented in these two classes. Size class differences were explored further for reference sites and showed that for all sites there are far higher numbers of small shortfin eels than longfin eels except at the “Stoney Stream @ 3rd order” site for which small eel numbers are even (Figure 3.12). These size class figures show a predominance of eels <500mm for both species.

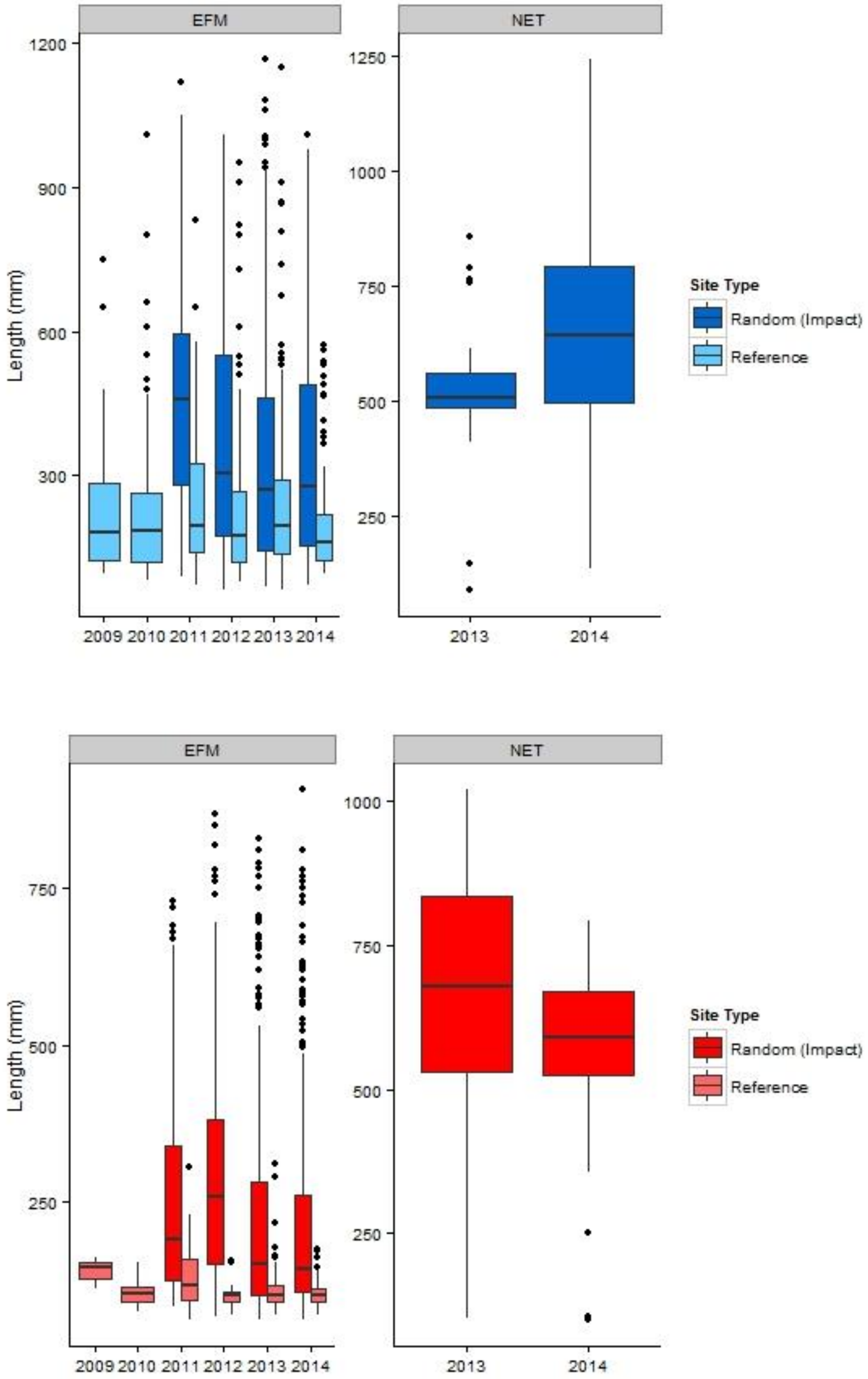


Figure 3.9 Longfin (top) and shortfin (bottom) lengths from the WRC monitoring network separated into different capture methodologies - electric fishing (EFM) and netting (NET) – and reference and random sites for the years of sampling period 2009-2014.

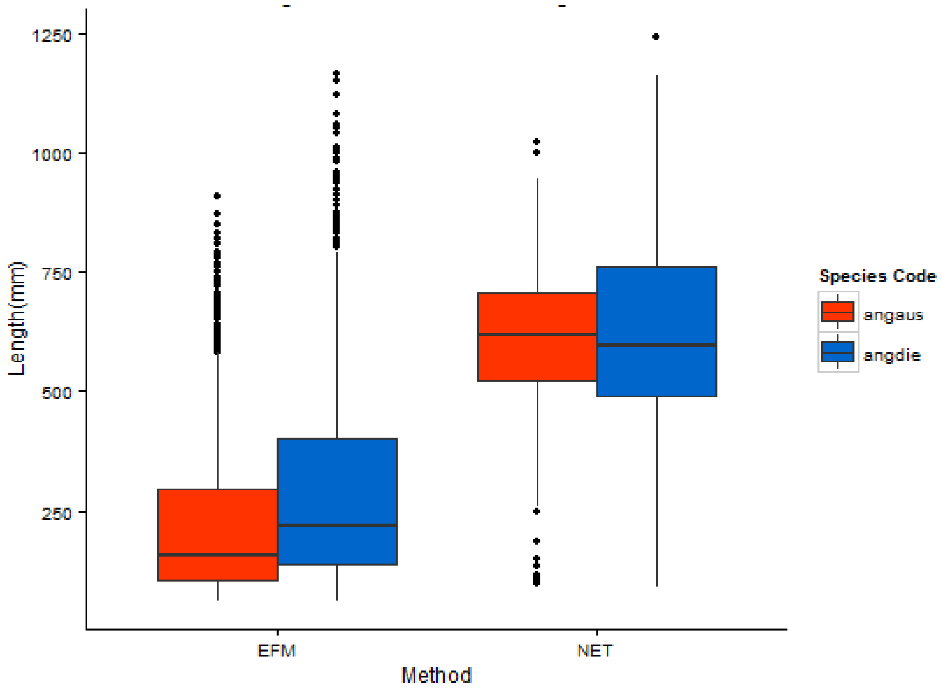


Figure 3.10 Comparison between lengths of longfin (angdie) and shortfin (angaus) eels using electro-fishing (EFM) and netting techniques (NET). Values are from WRC dataset from all sampled sites in sampling network from 2009-2014.

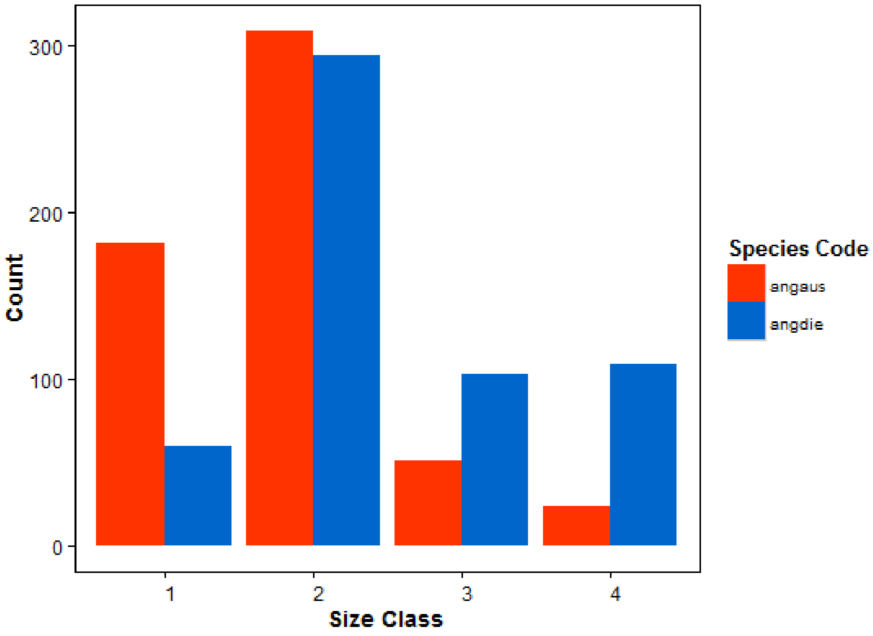


Figure 3.11 Size class counts for longfin (angdie) and shortfin (angaus) eels from WRC network. Size classes are as follows – 1 = <100mm, 2 = 100-300mm, 3 = 300-500mm, 4 = >500mm. Data represents complete current WRC dataset.

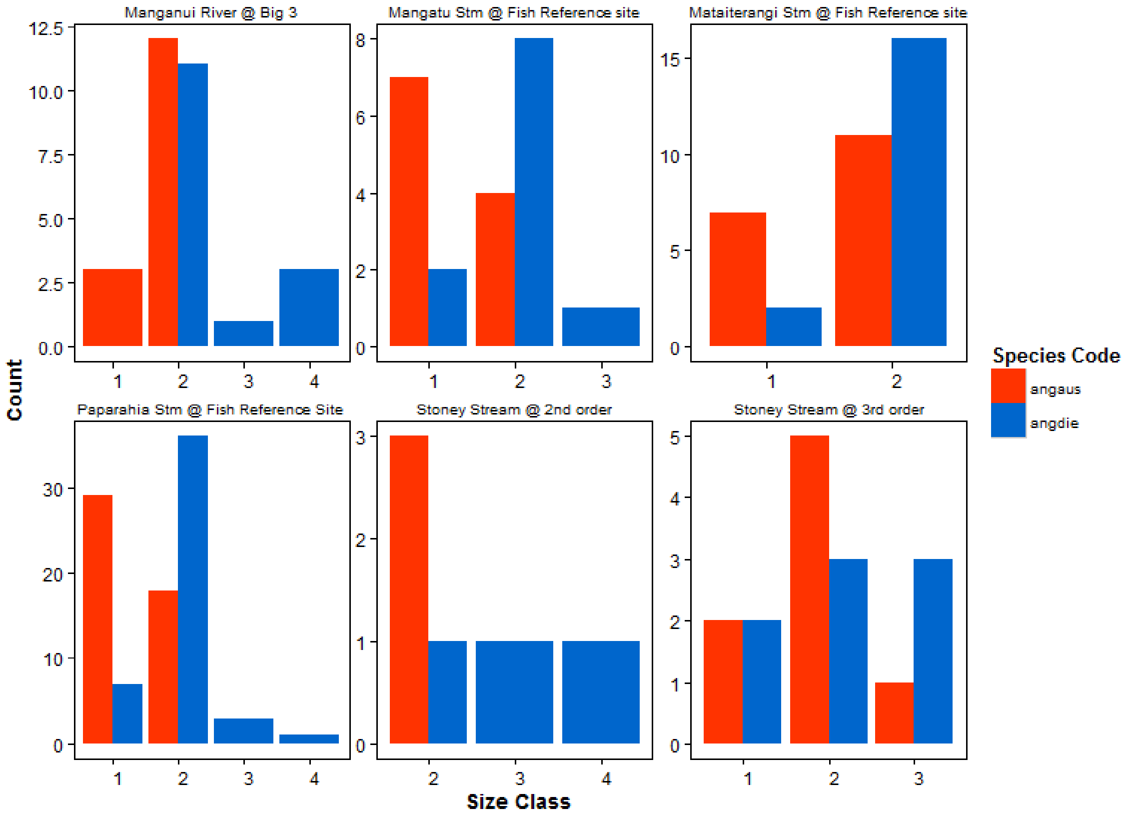


Figure 3.12 Eel size class counts for 6 of 9 reference sites in WRC network for longfin (angdie) and shortfin (angaus) eels. Reference sites absent here are due to very low overall eel counts at those sites. Size classes are as follows: 1 = <100mm, 2 = 100-300mm, 3 = 300-500mm, 4 = >500mm.

Redfin bullies

Mean relative abundance of redfin bullies was higher at reference sites than impact sites ($F_{1,\,186}$ =67.94, p<0.000) and mean length was also significantly higher at reference sites compared with impact sites ($F_{1,\,3811}$ =75.67, p<0.000) (Figures 3.13 and 3.14).

Sample date and abundance was not significantly different for either reference or random sites (reference: $F_{1,\,36}$=0.24, p=0.63; impact: $F_{1,\,148}$=0.19, p=0.66). Length data for redfin bullies reveals significant differences between years for both impact and reference sites, with greater mean length in earlier sample years (reference: $F_{1,\,2862}$=160.8, p<0.000; impact: $F_{1,\,947}$=106, p<0.000). An increase in mean length in 2010 corresponds with an increase in relative abundance of redfin bullies and may relate to the sampling of solely reference sites during this year. Sampling during 2009 was also solely reference sites however 2009 sites were on the east coast whereas 2010 sites were west coast. This characteristic of size differences is explored further in the redfin bully case study in Chapter 5.

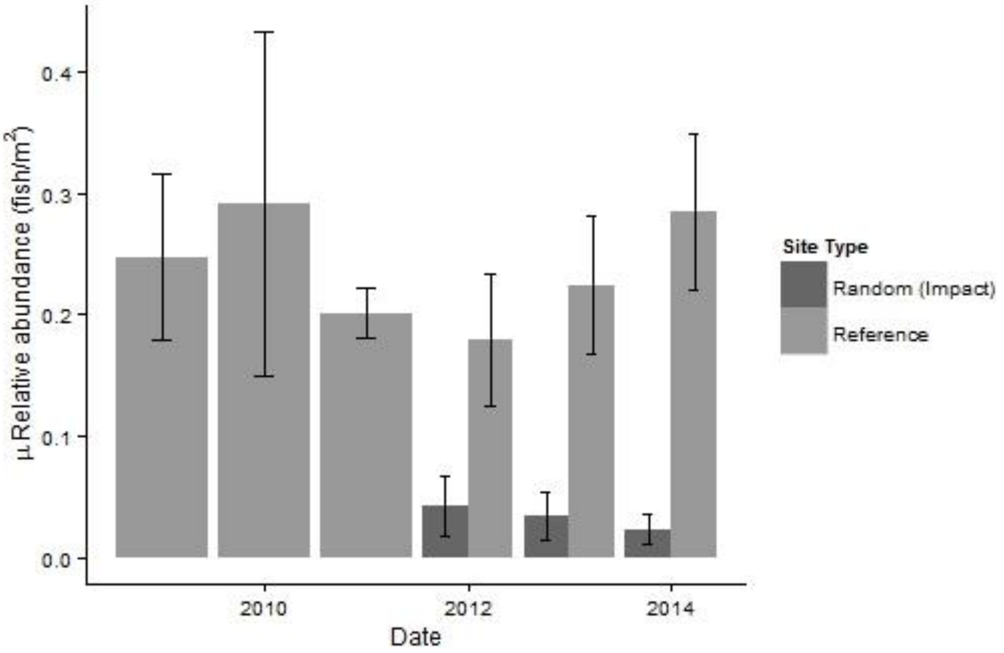


Figure 3.13 Mean relative abundance of redfin bullies from the WRC monitoring network over the period 2009-2014. Dark grey bars are random/impact sites (3 year rotation), light grey bars are reference sites (sampled annually). Error bars are standard error of the mean.

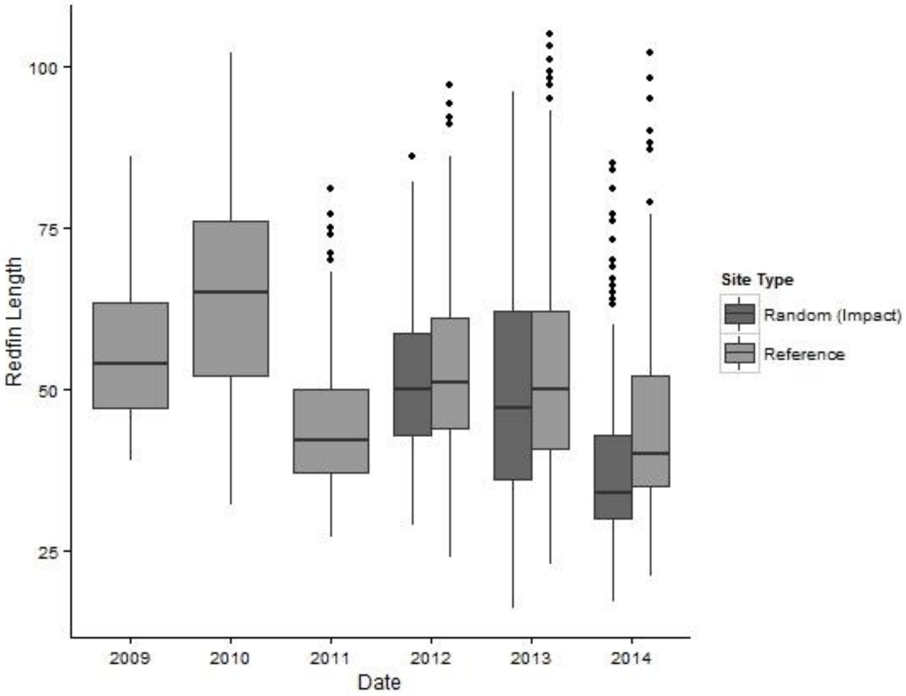


Figure 3.14 Redfin bully length from WRC sampling network for years 2009-2014. Light grey boxes are reference sites (sampled annually), dark grey boxes are random (impact) sites sampled on a 3 year rotation.

Yearly differences in species abundance and length at reference and impact sites

Length was significantly different between years for at least one site type for all species, and for all site types for redfin bullies. Summary results of ANOVA and power analyses are presented below in Table 3.2 for relative abundance, and Table 3.3 for mean length for the three species examined. Power values are high in all tests due to large sample size although no relative abundance measures were significantly different between years. Tukeys HSD tests were undertaken to isolate differences for mean length between years for eels and redfin bullies (Table 3.4). Shaded cells are years 2011 and 2014 which represent the same sites in the random site network and therefore a direct comparison. As described above all reference sites are sampled each year and are therefore comparable between years.

Mean redfin length at impact sites decreased and was lower in 2014 than previous years. For redfin mean length, all years were significantly different from each other except 2011 and 2014 for reference sites, and 2012 and 2013 for impact sites. These two years are significantly different from all others. For eel species at random sites, both species mean length is significantly lower in 2014 compared with 2011. At reference sites, mean length was not different for any years for longfin eels. For shortfin eels at reference sites, mean length in 2009 and 2011 – not significantly different from each other – was significantly higher than all other years. All other years are significantly different from each other (Table 3.4).

Table 3.2 ANOVA and power (1-β as percentage: see appendix C) results testing WRC species data between years 2009-2014 for relative abundance for three species: longfin eel; shortfin eel; and redfin bully. Species tests are divided by reference and random sites. Degrees of freedom stated in F statistic subscript long with significance (p).

Species: Test	**Site Type**	**Results**	**Power (%)**
Anguilla dieffenbachii : mean relative abundance	Reference	$F_{1,36}$=0.62, p=0.44	100
	Random	$F_{1,148}$=1.11, p=0.29	100
Anguilla australis: mean relative abundance	Reference	$F_{1,36}$=2.603, p=0.12,	100
	Random	$F_{1,148}$=0.8618, p=0.35	100
Gobiomorphus huttoni: mean relative abundance	Reference	$F_{1,36}$ =0.24, p=0.63;	72
	Random	$F_{1,148}$ =0.19, p=0.66	99

Table 3.3 ANOVA and power (1-β as percentage: see appendix C) results testing WRC species data between years 2009-2014 for mean length for three species: longfin eel; shortfin eel; and redfin bully. Species tests are divided by reference and random sites. Degrees of freedom stated in F statistic subscript long with significance (p).

Species: Test	**Site Type**	**Method**	**Results**	**Power (%)**
Anguilla dieffenbachii : mean length	Reference	EFM	$F_{1,633}$=0.15, p=0.70	100
	Random	EFM	$F_{1,885}$ =17.88, p<0.000	100
		NET	$F_{1,146}$=8.86, p=0.0034	100
Anguilla australis: mean length	Reference	EFM	$F_{1,212}$=1.62, p=0.20	100
	Random	EFM	$F_{1,2060}$=53.12, p<0.000	100
		NET	$F_{1,120}$=3.42, p=0.067	100
Gobiomorphus huttoni: mean length	Reference	EFM	$F_{1,2862}$ =160.8, p<0.000	100
	Random	EFM	$F_{1,947}$=106, p<0.000	100

Table 3.4 Tukeys HSD test for locations of significant differences between date groupings for length data. Significant differences are identified by group letter: years with different letters from one another are significantly different for mean length. Species data are divided into reference and impact sites and mean length for each year is shown. Shaded cells represent repeat impact site samples from years 2011 and 2014 and therefore a direct comparison. Reference sites are sampled annually and can be compared between each year.

Anguilla australis

Reference			**Random**					
EFM			**EFM**			**NET**		
Group	Year	Mean	Group	Year	Mean	Group	Year	Mean
a	2009	138.7	a	2012	291.4	a	2014	628.5
a	2011	136	b	2011	250.8	a	2013	557
ab	2013	112.2	c	2013	210.3			
ab	2014	103.9	c	2014	208.7			
ab	2010	103.4						
ab	2012	99.9						

Anguilla dieffenbachii

Reference			**Random**					
EFM			**EFM**			**NET**		
Group	Year	Mean	Group	Year	Mean	Group	Year	Mean
a	2011	255.6	a	2011	477.7	a	2014	646.8
a	2013	245.8	b	2012	380	b	2013	520.1
a	2012	226.4	b	2014	337.4			
a	2010	221.4	b	2013	333			
a	2009	220.5						
a	2014	198.3						

Gobiomorphus huttoni

Reference			**Random**		
Group	Year	Mean	Group	Year	Mean
a	2010	63.76	a	2012	51.42
b	2009	56.18	a	2013	48.93
bc	2012	53.26	b	2014	37.91
c	2013	52.89			
d	2011	44.61			
d	2014	43.71			

Discussion

Throughout the WRC monitoring network, reference sites are located close to the coast with inland lowland reference sites absent. The majority of lowland environments within New Zealand have been heavily modified and few areas of indigenous forest remain (Glova, Jellyman, & Bonnett, 1998; Graynoth, Francis, & Jellyman, 2008; Jowett & Richardson, 1996). Sites with suitable habitat for native species are restricted to these areas where modification is reduced. For comparatively pollution sensitive species like redfin bullies and longfin eels, these differences in available habitat were further shown with significantly higher numbers of redfin bullies at reference sites, and significantly larger longfin eels at reference sites. Native richness values are also higher closer to the coast as are abundance IBI values. Many diadromous species may be limited in their ability to penetrate inland either due to their size or strength or their ability to climb (McDowall, 1990). Although specialised modes of locomotion have evolved to accommodate habitats within New Zealand, migrating through large river systems, rapids, waterfalls, or any manmade obstacles would be difficult for some diadromous species, restricting them to areas closer to the coast (McDowall, 1993). As it becomes difficult to penetrate further inland, fewer species are likely to be found which may contribute to higher values of both native richness, for species entirely excluded from inland areas, and abundance IBI values in areas closer to the coast (Jowett & Richardson, 1996). Grouping all the network sites by relative abundance (Figure 3.4) further showed elevation, gradient, and indigenous forest as the most important variables determining site groupings.

There are many species, however, that would have little difficulty migrating large distances inland, yet native species richness values are much lower than would be expected throughout inland waterways. This is more likely to reflect the substantial land use change and native habitat loss through lowland catchments (Dodd et al., 2011; Jowett, Hayes, Deans, & Eldon, 1998; Walker, Price, & Rutledge, 2008). Lowland waterways in New Zealand have been heavily modified both physically and in terms of the fish community present. The contribution of these alterations to the decrease in native richness is likely to be far greater than inherent limitations of species migration abilities (McDowall, 1990, 2007). For example, the Mangauika stream on the west coast of the North Island has a comparatively high native diversity (4) and IBI value (72) despite being 165km inland and at an altitude of 75m. Native fish are certainly capable of migrating large distances given suitable habitat to migrate to and through, however, in most instances suitable habitat is no longer present (Weeks, Anderson-Lederer, Death, & Joy, 2015).

Species loss is particularly high within inland lowland areas where land clearance and development have taken place which introduce stresses novel to species such as nutrient increases, heavy metals loads, temperature changes, and sedimentation of instream habitats (Jowett & Richardson, 2003; McDowall & Taylor, 2000; Walker et al., 2008). Typical migration distances of those species that do migrate can be greater than 200km. Additionally, there are non-diadromous species (e.g. Crans' bullies, lake recruiting smelt, and some bully species) suggesting that native richness inland should not be solely limited by diadromy or migration ability (McDowall, 1993) and low inland native richness values (2-4) imply an effect on native richness separate from migration ability. Furthermore, though many species can migrate these distances inland, without sufficient annual recruits due to habitat or migration disruption, these inland areas may not remain colonised.

Specific pressures on eel species may also be reflected in variable populations. The commercial eel fishery for example may impact the two species differently. Larger eels are likely to be targeted which will potentially increase the pressure on longfin eels which typically grow larger and may make up relatively larger proportions of the eel catch (Chisnall, 1994; McDowall, 1990). When these longfin eels are removed from waterways, significant increases in shortfin recruitment have been seen previously (Glova, 2001). This may be occurring the Waikato region also: throughout the network there are far higher numbers of small shortfin eels than any other size class and in comparison to longfin eels. There was also a significant reduction in mean length for both eel species over the monitoring period. Although relative abundance of shortfin eels did not increase significantly over the monitoring period, it remains to be seen whether fishing pressure and habitat loss may be disproportionately affecting longfin eels. Further, the significantly higher relative abundance of shortfin eels at impact sites illustrates their ability to proliferate in disturbed habitats.

Their smaller size, and higher tolerance of polluted waterways, may be giving shortfin eels a recruitment advantage over longfins, which an abundance of smaller shortfin eels might suggest. Also conspicuous is the changing median length for shortfin eels (Figure 3.9) and mean length values for shortfin eels in 2013 and 2014 respectively were 231 and 228, down from 280mm in 2012. This may again illustrate a greater influx of smaller individuals (i.e. recruitment) to waterways sampled which has been identified occurring in streams where fishing pressure removed market size longfin *and* shortfin eels (Chisnall, 1994). Market demand, the availability of market size fish, and the number of eel fishers is likely to change from year to year which will invariably alter the effects on both eel species (Boubee, Chisnall, Watene, Williams, & Roper, 2002). In a 2005 study on the management of longfin eels it was noted that a 75% drop in numbers had occurred since the 1970s (Doole, 2005). The impact of fisheries is difficult to isolate without detailed records of eel catches (for which identification would need to be accurate), although fishing pressure additional to habitat loss has the potential to affect longfin eels to a greater extent given the observed ability for shortfin eels to move into and persist in degraded waterways (Chisnall, 1994; Chisnall & Hicks, 1993; Glova et al., 2001; Jellyman et al., 2009; Jowett et al., 2009).

It might be expected that as shortfin eel numbers increase there would be a corresponding decrease in longfin eel counts however, this may not be immediately evident for a number of reasons. Such changes may not be observed at all sites; the longevity of eels may predicate that drops in numbers of longfin eels would be seen over long time spans (e.g. a lack of recruitment once mature adults have migrated); or perhaps competition created from increasing shortfin numbers does not immediately impact large longfin eels. Competition from large longfin eels for example can prevent the influx of shortfin eels (Chisnall, 1994; Graynoth et al., 2008). However, recruitment sizes of eels can be highly variable particularly for longfin eels. This will ultimately introduce noise in length analysis specifically in respect to detecting trends in length changes for both species and requires further investigation (Francis & Jellyman, 1999).

Analysis of eel length in relation to capture method in this study reiterated known factors of methodological bias. It is likely that electric fishing underestimates large eels due to the habitat they typically inhabit. Greater numbers of large eels would typically inhabit complex and deeper waterways where electric fishing is less effective (Glova et al., 1998; Jellyman & Chisnall, 1999). Netting methods such as fyke nets, are typically used in these environments where electric fishing is not appropriate or effective. Due to the habitat smaller eels occupy, i.e. often in sediments or substrates avoiding predation, nets tend to underestimate counts of small eels (Figures 3.9-3.10). Differences between methods are less pronounced for count data indicating that consideration of the method used is important predominantly for length analyses.

At both reference and impact sites, redfin bully length decreased significantly over the survey period 2009-2014 (Table 3.4). This may reflect a general removal of suitable adult habitat and an influx of usually out competed juveniles however, this remains to be seen over coming monitoring years. These types of intra-specific interactions and habitat selection behaviours have been seen previously in redfin bullies and other small benthic New Zealand native fish. Recruitment and breeding can be highly variable in diadromous species, along with variable behaviours and habitat selection in relation to hydrological alterations (Closs, Smith, Barry, & Markwitz, 2003; David & Closs, 2002; David, Closs, Crow, & Hansen, 2007; Hansen & Closs, 2005;

Hansen & Closs, 2009; Leprieur et al., 2006; Vanderpham, Nakagawa, & Closs, 2013; Whitehead, David, & Closs, 2002). Conclusions here are likely to be premature, and further investigation is explored in chapter 5. Data from continued monitoring will also further help explain reductions in size of redfin bullies.

Conclusion

This study has highlighted differences in species distributions and characteristics throughout the Waikato region along with differences between impact and reference sites in the WRC network. Carefully collected data has provided substantially improved datasets able to analyse not only length data but also relative abundance. Size and abundance data collected during sampling have identified key differences between species in the Waikato region. Higher numbers of small shortfin than longfin eels are present in the Waikato and for both species mean length significantly decreased over the monitoring period. Greater numbers of redfin bullies were found at reference sites as well as larger longfin eels whereas shortfin eels were significantly more abundant at impact sites. Known differences in the size of eels caught using different methods were reaffirmed with larger eels caught using netting methods than electric fishing.

Patterns of native richness and IBI scores reflect both life history of diadromous species and well known impacts on New Zealand freshwater fish in general. Throughout developed and developing areas, native richness and IBI scores are low whilst in less modified areas scores are high. Extensive habitat loss and novel pressures have been introduced throughout the central Waikato through land use change and hydrological alteration. Metrics of native richness and IBI scores support site characteristics and distribution illustrating the vulnerability of native fish, and the impact continued modification will likely have. This is perhaps best illustrated through the lack of available reference sites in the entirety of the Waikato region, for which only nine have been able to be found. Sites that have been found are restricted to areas close to the coast. This paucity of pristine sites, the exclusion of species away from much of the central Waikato, and changing distributions of sensitive species such as longfin eels relative to those more tolerant of pollution (shortfin eels), illustrate the need for strong conservation measures for New Zealands' freshwater fish and their habitats.

References

Allibone, R., David, B., Hitchmough, R., Jellyman, D., Ling, N., Ravenscroft, P., & Waters, J. (2010). Conservation status of New Zealand freshwater fish, 2009. *New Zealand Journal of Marine and Freshwater Research, 44*(4), 271-287. doi: 10.1080/00288330.2010.514346

Baillie, B. R., Hicks, B. J., van den Heuvel, M. R., Kimberley, M. O., & Hogg, I. D. (2013). The effects of wood on stream habitat and native fish assemblages in New Zealand. *Ecology of Freshwater Fish, 22*(4), 553-566.

Boubee, J., Chisnall, B., Watene, E., Williams, E., & Roper, D. (2002). Enhancement and management of eel fisheries affected by hydroelectric dams in New Zealand. In D. A. Dixon (Ed.), *Biology, Management, and Protection of Catadromous Eels* (Vol. 33, pp. 191-205). Bethesda: Amer Fisheries Soc.

Buja, A., Swayne, D. F., Littman, M. L., Dean, N., Hofmann, H., & Chen, L. (2008). Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics, 17*(2), 444-472. doi: 10.1198/106186008x318440

Chisnall, B. L. (1994). *An unexploited mixed species eel stock (Anguilla australis and A. dieffenbachii) in a Waikato pastoral stream, and its modification by fishing pressure.* Wellington: Department of Conservation.

Chisnall, B. L., & Hicks, B. J. (1993). Age and growth of longfinned eels (anguilla-dieffenbachii) in pastoral and forested streams in the Waikato river basin, and in 2 hydroelectric lakes in the North Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 27*(3), 317-332.

Clark, M. (2004). *A first course in applied statistics : with applications in biology, business and the social sciences* (SprintPrint ed ed.). Auckland: Pearson Education New Zealand.

Closs, G. P., Smith, M., Barry, B., & Markwitz, A. (2003). Non-diadromous recruitment in coastal populations of common bully (Gobiomorphus cotidianus). *New Zealand Journal of Marine and Freshwater Research, 37*(2), 301-313.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*: Hillsdale, N.J. : L. Erlbaum Associates, 1988, 2nd ed.

Coolidge, F. L. (2013). *Statistics : a gentle introduction* (3rd ed ed.). Thousand Oaks, Calif.: SAGE Publications.

David, B. O., & Closs, G. P. (2002). Behavior of a stream-dwelling fish before, during, and after high-discharge events. *Transactions of the American Fisheries Society, 131*(4), 762-771. doi: 10.1577/1548-8659(2002)131<0762:boasdf>2.0.co;2

David, B. O., Closs, G. P., Crow, S. K., & Hansen, E. A. (2007). Is diel activity determined by social rank in a drift-feeding stream fish dominance hierarchy? *Animal Behaviour, 74*, 259-263. doi: 10.1016/j.anbehav.2006.08.015

Dodd, M., Barker, G., Burns, B., Didham, R., Innes, J., King, C., . . . Watts, C. (2011). Resilience of New Zealand indigenous forest fragments to impacts of livestock and pest mammals. *New Zealand Journal of Ecology, 35*(1), 83-95.

Doole, G. J. (2005). Optimal management of the New Zealand longfin eel (Anguilla dieffenbachii). *Australian Journal of Agricultural and Resource Economics, 49*(4), 395-411. doi: 10.1111/j.1467-8489.2005.00310.x

Francis, R., & Jellyman, D. J. (1999). Are mean size data adequate to monitor freshwater eel fisheries? *Marine and Freshwater Research, 50*(4), 355-366. doi: 10.1071/mf98148

Glova, G. J. (2001). Effects of the presence of subadult longfinned eels (Anguilla dieffenbachii) on cover preferences of juvenile eels (Anguilla spp.) in replicate channels. *New Zealand Journal of Marine and Freshwater Research, 35*(2), 221-233.

Glova, G. J., Jellyman, D. J., & Bennett, M. L. (2001). Spatiotemporal variation in the distribution of eel (Anguilla spp.) populations in three New Zealand lowland streams. *Ecology of Freshwater Fish, 10*(3), 147-153. doi: 10.1034/j.1600-0633.2001.100304.x

Glova, G. J., Jellyman, D. J., & Bonnett, M. L. (1998). Factors associated with the distribution and habitat of eels (Anguilla spp) in three New Zealand lowland streams. *New Zealand Journal of Marine and Freshwater Research, 32*(2), 255-269.

Graynoth, E., Francis, R. I. C. C., & Jellyman, D. J. (2008). Factors influencing juvenile eel (Anguilla spp.) survival in lowland New Zealand streams. *New Zealand Journal of Marine and Freshwater Research, 42*(2), 153-172.

Gu, H., Singh, A., & Townsend, P. A. (2015). Detection of gradients of forest composition in an urban area using imaging spectroscopy. *Remote Sensing of Environment, 167*, 168-180.

Hansen, E. A., & Closs, G. P. (2005). Diel activity and home range size in relation to food supply in a drift-feeding stream fish. *Behavioral Ecology, 16*(3), 640-648. doi: 10.1093/beheco/ari036

Hansen, E. A., & Closs, G. P. (2009). Long-term growth and movement in relation to food supply and social status in a stream fish. *Behavioral Ecology, 20*(3), 616-623. doi: 10.1093/beheco/arp039

Huene, M., & Vega, R. (2015). Spatial variation in the diet of Patagonotothen tessellata (Pisces, Nototheniidae) from the fjords and channels of southern Chilean Patagonia. *Polar Biology, 38*(10), 1613-1622.

Jellyman, D. J. (1991). Factors affecting the activity of 2 species of eel (anguilla-spp) in a small New Zealand lake. *Journal of Fish Biology, 39*(1), 7-14. doi: 10.1111/j.1095-8649.1991.tb04336.x

Jellyman, D. J. (2007). Status of New Zealand fresh-water eel stocks and management initiatives. *Ices Journal of Marine Science, 64*(7), 1379-1386. doi: 10.1093/icesjms/fsm073

Jellyman, D. J., Booker, D. J., & Watene, E. (2009). Recruitment of Anguilla spp. glass eels in the Waikato River, New Zealand. Evidence of declining migrations? *Journal of Fish Biology, 74*(9), 2014-2033.

Jellyman, D. J., & Chisnall, B. L. (1999). Habitat preferences of shortfinned eels (Anguilla australis), in two New Zealand lowland lakes. *New Zealand Journal of Marine and Freshwater Research, 33*(2), 233-248.

Jellyman, D. J., & Graynoth, E. (2005). The use of fyke nets as a quantitative capture technique for freshwater eels (Anguilla spp.) in rivers. *Fisheries Management and Ecology, 12*(4), 237-247. doi: 10.1111/j.1365-2400.2005.00445.x

Jowett, I. G., Hayes, J. W., Deans, N., & Eldon, G. A. (1998). Comparison of fish communities and abundance in unmodified streams of Kahurangi National Park with other areas of New Zealand. *New Zealand Journal of Marine and Freshwater Research, 32*(2), 307-322.

Jowett, I. G., & Richardson, J. (1996). Distribution and abundance of freshwater fish in New Zealand rivers. *New Zealand Journal of Marine and Freshwater Research, 30*(2), 239-255.

Jowett, I. G., & Richardson, J. (2003). Fish communities in New Zealand rivers and their relationship to environmental variables. *New Zealand Journal of Marine and Freshwater Research, 37*(2), 347-366.

Jowett, I. G., Richardson, J., & Boubee, J. A. T. (2009). Effects of riparian manipulation on stream communities in small streams: two case studies. *New Zealand Journal of Marine and Freshwater Research, 43*(3), 763-774.

Joy, M. K. (2005). *A Predictive Model of Fish Distribution and Index of Biotic Integrity (IBI) for Wadeable Streams in the Waikato Region*. Hamilton, New Zealand: Environment Waikato.

Joy, M. K. (2013). *Freshwater fish predictive modelling for bioassessment; A scoping study into fish bioassessment models as national indicators in New Zealand*. Wairesearch.

Joy, M. K., David, B., & Lake, M. (2013). *New Zealand Freshwater Fish Sampling Protocols*. Palmerston North, New Zealand: Massey University

Joy, M. K., & Death, R. G. (2004). Application of the index of biotic integrity methodology to New Zealand freshwater fish communities. *Environmental Management, 34*(3), 415-428. doi: 10.1007/s00267-004-0083-0

Leathwick, J. R., Elith, J., Chadderton, W. L., Rowe, D., & Hastie, T. (2008). Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *Journal of Biogeography, 35*(8), 1481-1497. doi: 10.1111/j.1365-2699.2008.01887.x

Leprieur, F., Hickey, M. A., Arbuckle, C. J., Closs, G. P., Brosse, S., & Townsend, C. R. (2006). Hydrological disturbance benefits a native fish at the expense of an exotic fish. *Journal of Applied Ecology, 43*(5), 930-939. doi: 10.1111/j.1365-2664.2006.01201.x

McCleave, J. D., & Jellyman, D. J. (2002). Discrimination of New Zealand stream waters by glass eels of Anguilla australis and Anguilla dieffenbachii. *Journal of Fish Biology, 61*(3), 785-800. doi: 10.1006/jfbi.2002.2108

McDowall, R. M. (1990). *New Zealand freshwater fishes : a natural history and guide* (Rev. ed ed.). Wellington [N.Z.]: Heinemann Reed, MAF Publishing Group.

McDowall, R. M. (1993). Implications of diadromy for the structuring and modeling of riverine fish communities in New Zealand. *New Zealand Journal of Marine and Freshwater Research, 27*(4), 453-462.

McDowall, R. M. (2006). Fish, fish habitats and fisheries in New Zealand. *Aquatic Ecosystem Health & Management, 9*(4), 391-405. doi: 10.1080/14634980601026352

McDowall, R. M. (2007). On amphidromy, a distinct form of diadromy in aquatic organisms. *Fish and Fisheries, 8*(1), 1-13. doi: 10.1111/j.1467-2979.2007.00232.x

McDowall, R. M., & Taylor, M. J. (2000). Environmental indicators of habitat quality in a migratory freshwater fish fauna. *Environmental Management, 25*(4), 357-374. doi: 10.1007/s002679910028

McEwan, A. J., & Joy, M. K. (2014). Habitat use of redfin bullies (Gobiomorphus huttoni) in a small upland stream in Manawatu, New Zealand. *Environmental Biology of Fishes, 97*(2), 121-132. doi: 10.1007/s10641-013-0131-9

NIWA. (2015). Redfin Bully. *Identification Guides and Fact Sheets.* Retrieved 2015 from https://www.niwa.co.nz/freshwater-and-estuaries/nzffd/NIWA-fish-atlas/fish-species/redfin_bully

R Core Team. (2014). R: A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Stapley, J., Garcia, M., & Andrews, R. M. (2015). Long-Term Data Reveal a Population Decline of the Tropical Lizard Anolis apletophallus, and a Negative Affect of El Nino Years on Population Growth Rate. *Plos One, 10*(2). doi: 10.1371/journal.pone.0115450

Tonkin, J. D., Wright, L. A. H., & David, B. O. (2012). Mussel Spat Ropes Assist Redfin Bully Gobiomorphus huttoni Passage through Experimental Culverts with Velocity Barriers. *Water, 4*(3), 683-689.

Vanderpham, J. P., Nakagawa, S., & Closs, G. P. (2013). Habitat-related patterns in phenotypic variation in a New Zealand freshwater generalist fish, and comparisons with a closely related specialist. *Freshwater Biology, 58*(2), 396-408. doi: 10.1111/fwb.12067

Walker, S., Price, R., & Rutledge, D. (2008). New Zealand's remaining indigenous cover: recent changes and biodiversity protection needs. *Science for Conservation*(284), 82 pp.-82 pp.

Weeks, E. S., Anderson-Lederer, R., Death, R. G., & Joy, M. K. (2015). *Are we ignoring the plight of New Zealand’s freshwater species.* Society for Conservation Biology (Oceania), Sydney.

Wellnitz, T. (2015). How do stream grazers partition their benthic habitat? *Hydrobiologia, 760*(1), 197-204.

Whitehead, A. L., David, B. O., & Closs, G. P. (2002). Ontogenetic shift in nocturnal microhabitat selection by giant kokopu in a New Zealand stream. *Journal of Fish Biology, 61*(6), 1373-1385. doi: 10.1006/jfbi.2002.2147.

Chapter Four

Predictive Modelling of Fish Populations using a Novel Dataset in the Waikato Region

Chapter 4 Predictive Modelling of Fish Populations using a Novel Dataset in the Waikato Region

Introduction

Accurately defining and understanding differences in species abundances and distributions is one of the fundamental challenges of ecological science (McGill et al., 2007). The Waikato region of New Zealand's North Island is a large area (25,000 Km^2) spanning the width of the country and encompassing a diversity of ecosystems and land uses. As with many parts of New Zealand, the region has been heavily modified including damming of several major river systems and substantial agricultural development in lowland areas (Allibone et al., 2010; Boubee, Chisnall, Watene, Williams, & Roper, 2002; McDowall, 2006; Weeks, Anderson-Lederer, Death, & Joy, 2015). Indigenous forest has been lost from much of the lowland areas with a few coastal and inaccessible areas remaining unaltered. Additionally, many invasive species now plague Waikato waterways (some of which are legally protected: salmonids) while a number of declining native species remain open to commercial and/or recreational harvest (Boubee et al., 2002; Chisnall, 1994; Jellyman, 2007; McDowall, 2006). Many New Zealand species are also diadromous, and have specialised life histories or locomotory habits that make them particularly vulnerable to impacts such as damming and habitat degradation (McDowall & Taylor, 2000).

Faced with the challenge of conserving our rapidly declining and vulnerable freshwater biota, freshwater ecologists have endeavoured to gain an accurate understanding of the characteristics of different species, and how these differ regionally and temporally in response to environmental variables. Working to fill the many gaps in our understanding of freshwater species is vital to conserve them. Data on freshwater fish has been collected and consolidated in New Zealand for close to 100 years in the New Zealand Freshwater Fish Database (NZFFDB, 2015). However, data in this large collection are difficult to utilise for detailed national or regional analyses largely due to the variable methodologies and objectives that have been used in their collection. Although analyses of these datasets have been undertaken, limitations of these types of data are inherent, and it has become clear that more detailed and consistent monitoring of populations is needed to effectively understand ecological diversity in New Zealand's freshwater biota (David et al., 2010; Mancini, Elsadek, & Madon, 2015; Reiss et al., 2015; Robinson et al., 2011). Over the past 5 years, Waikato Regional Council (WRC) freshwater monitoring staff have set about rectifying this within the Waikato region through implementation of standardised sampling throughout their monitoring network. The sampling protocols used include a standardised reach and sub-reach length, electric and net fishing methodologies ensuring consistent fishing effort, and consistent sample timing. This dataset should provide comparatively robust predictions and ecological conclusions when assessing the population dynamics of species within this region. Chapter 2 of this thesis outlines in detail the consistency in sampling effort achieved through implementation of these protocols. Furthermore, where records from databases such as NZFFDB can often only be used as presence/absence records, the WRC dataset used in these analyses can be used to assess relative abundance and length data in addition to presence/absence.

As species habitats are further fragmented and restricted, predicting the location and abundance of remaining populations becomes vital in conservation efforts (Guisan et al., 2013). To predict species distributions across large spatial areas, methods other than reconnaissance are required both for efficiency and coverage (Elith et al., 2006; Guisan & Thuiller, 2005). The predictive capacity that models provide for changing climates and future habitat characteristics, along with ecological community dynamics, are hugely beneficial in population management (Heino, 2013). Machine learning techniques provide a method to rapidly assess the relationships between predictor and response variables. Regression trees are a popular decision based classification method for response variables in relation to input variables. Boosting (iterative tree building) using regression trees with machine learning techniques has grown popular in the last decade and allows meaningful predictions and assessments to be made from sparse data while overcoming the inherent limitations of a single tree constructed from a dataset (Elith et al., 2006). Long term monitoring of populations where historic data are relied on, is often faced with limitations in consistency and spatial coverage. Being able to utilise techniques that can take fragmented datasets and make conclusions is a huge advantage in predictive modelling (Ramsey, Caley, & Robley, 2015). Models are limited by the accuracy of datasets upon which they are based and until now modelling approaches within New Zealand have been based on data from the NZFFDB, which comes with the inherent limitations of inconsistent sampling effort and methodology. With the availability of this novel WRC dataset which exhibits a high level of consistency and the ability to analyse and model a wide range of metrics, a greater understanding of fish populations in the Waikato region can be achieved.

This study begins to assess in detail species distributions and characteristics in the Waikato region using this dataset. Population distributions and abundance data are modelled with models used to predict presence and abundance across the Waikato region. Key environmental drivers of these distributions are also identified.

Methods

Boosted Regression Trees (BRT)

Boosted regression trees combine traditional statistical methods and machine learning techniques to produce a combination of multiple models. As described in Elith, Leathwick, & Hastie (2008), trees are built by splitting all data into clusters reflecting dominant relationships with response variables. A constant is then fitted to each split point as a result of classification tree algorithms describing the most likely class resulting in that particular cluster. Mean responses are then fitted to all observations in that cluster with split points being chosen to minimise prediction errors. Boosting then provides a means to improve the accuracy of the model by iteratively taking a random subset of the dataset, and sequentially improving model predictions on the hardest observations to predict. The final model is a combination of trees, describing a regression relationship with multiple terms (Elith et al., 2008). Models are then simplified by dropping the least important explanatory variables optimising predictive performance. The predictions that are made using this process are comparatively robust and take into account the multitude of variables that determine the outcome in the response variable.

Using BRT modelling, presence/absence and abundance data were analysed to predict where species are likely to occur and in what abundance within the Waikato region. The primary drivers of the presence and abundance of species were identified through fitted function analysis during model building processes. The drivers of presence of species provide an idea of the types of environments species would likely be found and how these might relate to life history traits or changing population dynamics. Precisely, the drivers that explain variation in presence are likely to be having a strong impact on the historical ability of a species to exist in an environment (Joseph, Field, Wilcox, & Possingham, 2006; Pearce & Boyce, 2006). This is in comparison to drivers that explain differences in relative abundance which potentially describe contemporary influences on species populations (Guisan et al., 2013; Leathwick, Elith, Chadderton, Rowe, & Hastie, 2008; Miller et al., 2014).

Finally, boosted regression trees were used first to analyse the characteristics of sampled populations and then to predict to the un-sampled sites within the River Environment Classification (REC) network. One hundred and eighty (180) wadeable sites are sampled by WRC staff, however, over 50,000 waterways are registered in the REC network in the Waikato region, 93% of which are wadeable sites (1-4th order). BRT models were used to predict to these REC waterways for both species presence and abundance in relation to the three species analysed in detail: *Anguilla dieffenbachii, Anguilla australis,* and *Gobiomorphus huttoni* (species description and rationale for choice described in chapter 3).

Code to produce boosted regression trees was sourced from Elith et al., (2008) and the rules of thumb outlined in their paper were used for tree complexity, bag fractions, and learning rates. As a recommended starting point, all trees were built using a bag fraction of 0.75, a learning rate of 0.001, and a tree complexity of 5.

Presence/absence and abundance analyses

For presence/absence analyses a *Bernoulli* distribution (family) was used representing a binary output for fitted values. The typical measure of presence/absence model performance is the cross validated ROC (receiver operating characteristic) score (equivalent to Area Under the Curve) which is a measure of the proportion of correct predictions made by the built model in relation to test data (Elith et al., 2008). Also included in model outputs below is the cross validated deviance. The particular value reported is the mean of the cross-validated deviances – a measure of how far the built tree lies from the observed or target value (i.e. smaller deviance is a better model) (Pearce & Boyce, 2006).

Worth noting for presence/absence models is that missing values or individuals in this case may not represent the absence of a species at a site, but simply that the individual was missed during sampling (e.g., (Boyce, 1945)). Determining the presence of an individual with absolute certainty is difficult without ensuring every individual within a reach or sampling event is caught which is *per se*, extremely difficult without undertaking depletion sampling. Assumptions for depletion sampling can be difficult to meet and undertaking this sampling with representative spatial coverage is generally financially and time prohibitive. The type of sampling used in the WRC dataset has been shown to be effective both in calculating numbers and presence of species (e.g. Graynoth, Bonnett, and Jellyman (2012)). Species detection using the methodologies in the WRC dataset is high (>90%) and it is assumed that absences of species at sites represent actual species absence (David et al., 2010).

For relative abundance data boosted regression trees were analysed with a *Laplace* distribution, with all other factors followed to guidelines outlined in Elith et al. (2008) as with presence/absence models. Points of difference when analysing model outputs using a Laplace distribution is the lack of a ROC (receiver operating characteristic) score which is provided in binary model outputs. This measures the rate of true positive results versus false positive results. As relative abundance data were analysed using a Laplace distribution, this no longer applies. For abundance model performance, mean cross-validated correlation and deviance, respectively, were used which provide a measure of model performance in relation to correlations between training and test data (Elith et al., 2006; Pearce & Boyce, 2006).

There were some species for which models could not be built. The BRT model for *Gambusia affinis* relative abundance did not work likely due to the high variation in numbers seen at a limited number of sites. For instance, many sites had no *Gambusia* at all while others had counts in the thousands. Evidently abundance estimates and model predictions were problematic for this species. The BRT models for *Galaxias brevipinnis, Gobiomorphus cotidianus,* and *Oncorhynchus mykiss* also did not work due to too few data points. *Gobiomorphus basalis, Retropinna retropinna,* and *Salmo trutta* models were the only species for which simplification procedures did not drop any variables for abundance model improvement.

Fish community analyses

Abundance IBI (Joy, 2013) and native richness scores were also modelled using boosted regression tree analysis. For both, a *Laplace* distribution was used during model building processes. Boosted regression tree outputs were further analysed to determine the most common predicted communities in the Waikato region and compared with raw data communities from the WRC dataset. This was done by using prediction vectors from all analysed species. Using presence/absence, a simple community analysis was undertaken to establish the ten most common communities predicted in the Waikato REC waterways. The presence of a species was taken as a probability of greater than 0.5 at a given waterway. The composition of the community then is composed of all species with a probability of presence greater than 0.5 at a given waterway as predicted by BRT models. There are several limitations to this community analysis, predominantly the coverage of some species distribution throughout the network and/or their general abundance (see Figure 3.4; chapter 3). However, community analysis was undertaken with the intention to provide a simple indication of the composition of waterways in the Waikato region.

Mapping out model predictions

ARCGIS (ARCMAP 10.1) was used to map predictions derived from BRT models to River Environment Classification network waterways in the Waikato region. Waterways in the WRC dataset are 1 – 4th order and grey sections in maps are waterways greater than 4th order and therefore not modelled. Following an initial map of redfin bully distribution, subsequent maps depict relative abundance to a 60km limit inland. Mapping throughout the Waikato is dubious due to sites greater distances inland having yet to sample any redfin bullies as noted.

Landcare Database 4 (LCDB4) data from Landcare Research[7] was mapped within the Waikato region outlining main categories of land use relevant to chapter discussions.

Results

Boosted regression tree results for species presence/absence data

Model performances were high (ROC score) for most species and in particular species with high spatial representation. Among highest were the three species investigated in detail, including the redfin bully model which had the highest ROC score (0.986; "outstanding"). Longfin eel (0.957; "outstanding") and shortfin eel (0.889; "excellent") both had relatively high model performance following simplification, along with more common species such as banded kokopu (*Galaxias fasciatus*; 0.935). Simplification of models improved the majority of ROC scores although not greatly. For presence/absence analysis, only the model for *Gobiomorphus cotidianus* did not suggest any improvement by dropping variables through the simplification process. High ROC scores and low deviance indicated built models predicted well against test data (Tables 4.1 and 4.2).

---

[7] http://www.landcareresearch.co.nz/resources/data

Table 4.1 Initial outputs for boosted regression tree presence/absence models built using all River Environment Classification (REC) variables for all species. Receiver Operating Characteristic (ROC) score and deviance shown [8].

Species	**ROC Score**	**Deviance**
Anguilla dieffenbachii	0.924	0.547
Anguilla australis	0.856	0.850
Cheimarrichthys fosteri	0.870	0.590
Galaxias brevipinnis	0.784	0.327
Galaxias fasciatus	0.935	0.470
Galaxias maculatus	0.933	0.611
Gambusia affinis	0.934	0.360
Gobiomorphus basalis	0.925	0.594
Gobiomorphus cotidianus	0.879	0.504
Gobiomorphus huttoni	0.984	0.281
Oncorhynchus mykiss	0.834	0.494
Retropinna retropinna	0.907	0.469
Salmo trutta	0.814	0.508

---

[8] REC variables described in Appendix B

Table 4.2 Presence/absence model outputs from Table 4.1 after dropping least useful variables through simplification function for improved (reduced) predictive deviance.

Species	**ROC Score**	**Estimated deviance**	**Number of Variables dropped**
Anguilla dieffenbachii	0.957	0.531	3
Anguilla australis	0.889	0.8	2
Cheimarrichthys fosteri	0.874	0.594	1
Galaxias brevipinnis	0.789	0.333	8
Galaxias fasciatus	0.931	0.46	7
Galaxias maculatus	0.914	0.617	10
Gambusia affinis	0.912	0.378	7
Gobiomorphus basalis	0.909	0.629	6
Gobiomorphus huttoni	0.986	0.269	4
Oncorhynchus mykiss	0.844	0.499	7
Retropinna retropinna	0.896	0.457	9
Salmo trutta	0.812	0.529	3

Drivers of fish distribution

Distance inland (the distance of the reach from the coast) features as the top determinant of presence for multiple diadromous species including redfin bullies, Crans' bullies and banded kokopu (Table 4.3). For redfin bullies and banded kokopu, model variation explained by distance inland was particularly high compared to other species models (40% and 55.7% respectively). Distance inland in all cases was a negative driver of species presence. Elevation and/or slope variables also featured frequently as presence determinants for many species including longfin and shortfin eels. Elevation was variable in terms of directional effect on presence: elevation was a negative driver for both eel species but was positive for Crans' bullies. Slope was a positive driver for most endemic species where present.

Table 4.3 Top 3 explanatory variables for presence/absence species models from Tables 4.1 and 4.2. Variables determined using fitted function analysis. +/- represents increasing or decreasing probability of occurrence with the given variable percentage value showing size of effect of variable in model processing.

Species	**Top 3 explanatory variables (in order)** [9]
Anguilla dieffenbachii	Upstream average slope (16.7% +); segment minimum elevation (9% -); upstream area (7.3% +)
Anguilla australis	Segment average elevation (11.6% -); altitude (9.4% -); upstream winter solar radiation (9.3% +)
Cheimarrichthys fosteri	Segment mean January air temperature (23.8% +); segment mean minimum June air temperature (16.9% +); % upstream indigenous forest (5.4% +)
Galaxias brevipinnis	Segment maximum slope (21.9% +); average catchment hardness (soil) (15.9% +); mean segment slope – DEM grid (13.8% +)
Galaxias fasciatus	Distance from sea (55.7% -); segment winter solar radiation (8% +); upstream runoff from scrub (7.6% -)
Galaxias maculatus	Altitude (23.1% -); segment upstream elevation (20.5% -); upstream area (7% +)
Gambusia affinis	Upstream catchment elevation (60.4% -); segment upstream elevation (6.7% -); upstream annual rain coefficient of variation (4.9% -)
Gobiomorphus basalis	Distance from sea (13.7% -); altitude (12.6% +); upstream mean slope (10.3% +)
Gobiomorphus cotidianus	Segment mean slope (17.3% -); segment mean elevation (12.1% -); % upstream pastoral cover (11.3% -)
Gobiomorphus huttoni	Distance from sea (40% -); segment mean minimum June air temperature (13.3% +); upstream mean slope (11.5% +)
Oncorhynchus mykiss	Upstream low flow (48.3% +); upstream area (8.7% +); upstream catchment hardness (soil) (5.1% -)

---

[9] Refer to Appendix B for variable descriptions – all applicable variables are flow weighted (Q).

Retropinna retropinna	Upstream area (22.2% +); altitude (13.7% -); segment winter equilibrium temperature (12.2% +)
Salmo trutta	% annual runoff from catchment <30° (21.6% -); % annual runoff from catchment >30° (14.7% +); % upstream indigenous forest (6.6% +)

As with presence/absence analyses, abundance models performed highest (both correlation and deviance) in more ubiquitous species such as longfin eels (0.702), inanga (0.601) and redfin bullies (0.743) (Table 4.4). Shortfin eel model performance was comparatively low (0.478). Following simplification of models, model performance, although similar dropped for all three focal species. Initial models were therefore used for both abundance driver fitted function analysis (Table 4.6) and subsequent prediction mapping.

Table 4.4 Initial boosted regression tree relative abundance models built using all River Environment Classification (REC) variables for species from Figure 3.1. Correlation of training and test data shown along with deviance [10].

Species	**Test and Training Data Correlation**	**Deviance**
Anguilla dieffenbachii	0.702	0.017
Anguilla australis	0.478	0.038
Cheimarrichthys fosteri	0.545	0.001
Galaxias fasciatus	0.523	0.006
Galaxias maculatus	0.601	0.052
Gobiomorphus basalis	0.565	0.105
Gobiomorphus huttoni	0.743	0.046
Retropinna retropinna	0.583	0.013
Salmo trutta	0.423	0.002

---

[10] REC variables described in Appendix B

Table 4.5 Relative abundance model outputs from Table 4.4 after dropping least useful variables through simplification function for improved (reduced) predictive deviance. Species not present from Table 4.4 are those where model building processes did not suggest improvement through simplification.

Species	**Test and Training Data Correlation**	**Deviance**	**Number of Variables dropped**
Anguilla dieffenbachii	0.679	0.019	1
Anguilla australis	0.474	0.039	7
Cheimarrichthys fosteri	0.667	0.001	3
Galaxias fasciatus	0.535	0.006	7
Galaxias maculatus	0.607	0.051	1
Gobiomorphus huttoni	0.694	0.051	8

Drivers of fish abundance

Distance from sea again featured multiple times as the largest determinant of species relative abundance (Table 4.6). This included negative effects on abundance for longfin eels, redfin bullies, banded kokopu, Crans' bullies, and a positive effect for torrentfish. Increasing segment length and decreasing catchment elevation were also positive determinants for longfin eel abundance while decreasing catchment elevation, increasing upstream calcium content and increasing scrub cover were the top determinants for shortfin eel abundance. For redfin bullies following distance from sea, increasing upstream catchment and segment indigenous forest cover were second and third in importance respectively. Indigenous forest cover of upstream catchment or segment respectively, also featured highly for both torrentfish and Crans' bullies. Elevation and temperature variables were common drivers amongst other species.

Table 4.6 Top 3 explanatory variables for relative abundance species models from Tables 4.4 and 4.5. Variables determined using fitted function analysis. +/- represents increasing or decreasing effect on relative abundance with percentage value showing size of effect of variable in model processing.

Species	**Top 3 explanatory variables (in order)**
Anguilla dieffenbachii	Distance from sea (7.8% -); segment length (4.1% +); upstream catchment elevation (3.9% -)
Anguilla australis	Upstream catchment elevation (28.2% -); upstream calcium content (7.1% +); % upstream scrub cover (5.9% +)
Cheimarrichthys fosteri	Distance from sea (50.1% +); segment average January air temperature (13.5% -); % upstream indigenous forest cover (10.6% +)
Galaxias fasciatus	Distance from sea (24.1% -); upstream annual rain coefficient of variation (13% +); segment pastoral cover (11% -)
Galaxias maculatus	Segment mean elevation (10% -); altitude (9.8% -); segment winter equilibrium temperature (9.2% +)
Gobiomorphus basalis	Segment indigenous forest cover (47.5% -); altitude (8% -); distance from sea (6.7% +)
Gobiomorphus huttoni	Distance from sea (22% -); segment indigenous forest cover (8.8% +); upstream indigenous forest (7.8% +)
Retropinna retropinna	Segment average elevation (33% -); segment maximum elevation (18.3% -); segment slope (12.6% -)
Salmo trutta	Upstream catchment hardness (soil) (58.9% +); annual runoff from catchment <30° (14.4% +); upstream low flow (6.1% +)

Mapping of model predictions

Predicted native richness was highest towards coastal areas particularly around the Raglan coast (Figure 4.1). Coromandel peninsula had the highest concentration of waterways with 6-8 species or more. Predicted native richness and actual native richness from the WRC dataset follow similar patterns with decreasing richness with distance from the coast. By removing non-diadromous Crans’ bullies from the native richness count, native richness is lower, particularly at distances less than 200km from the coast (Figure 4.2). Native richness approximately halves at a distance of 200km from the sea. The abundance IBI scores show a similar pattern to native richness calculations with higher values towards the coast and a collection of high scores on the Coromandel peninsula (Figure 4.3). A key difference for abundance IBI scores is a collection of waterways around Lake Taupo with high scores which were absent for native richness.

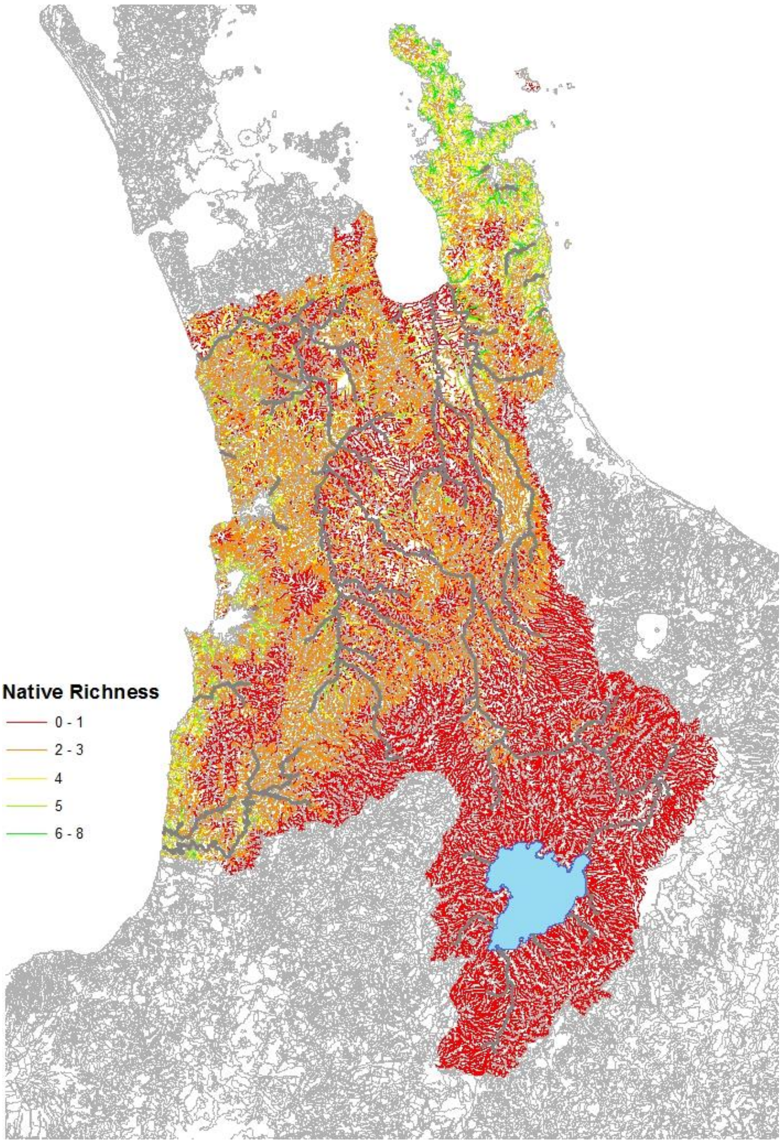


Figure 4.1 Predicted Native species richness for the Waikato region as determined by BRT models.

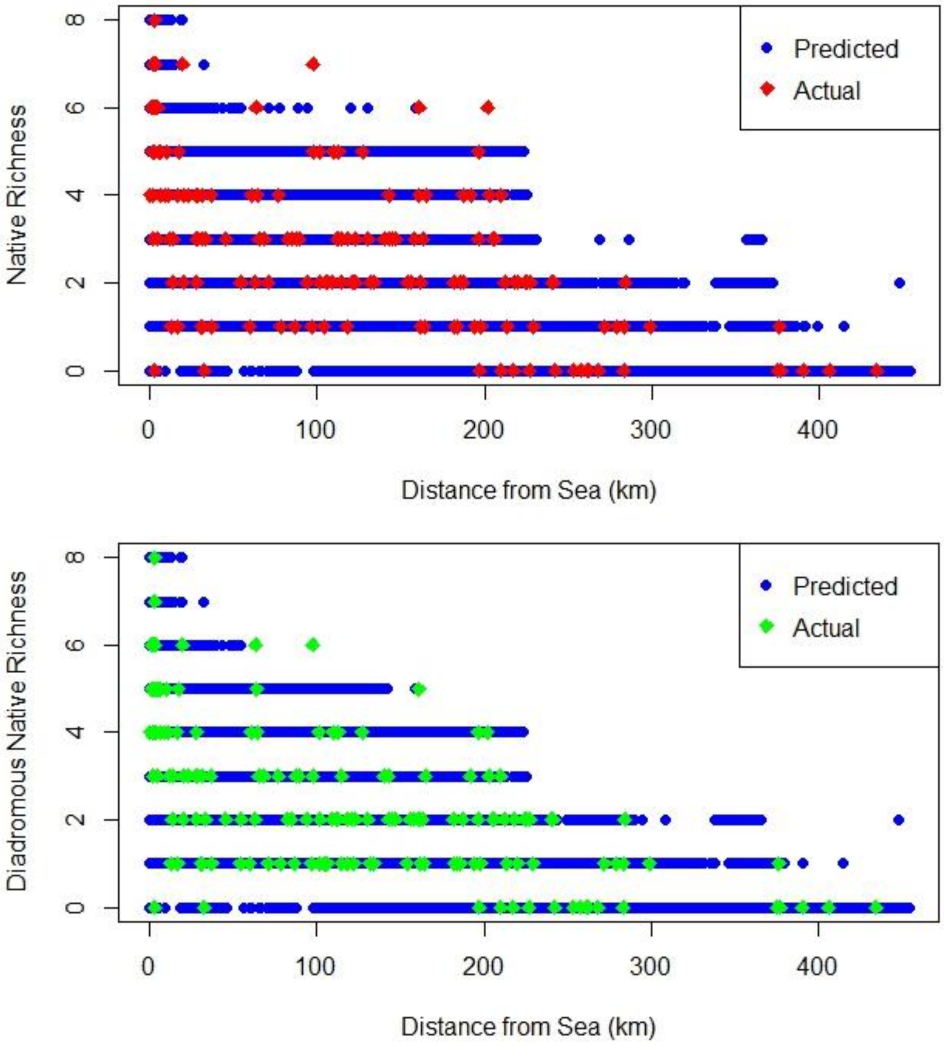


Figure 4.2 Native species richness values as predicted from BRT models built from Waikato Regional Council data, plotted against site distance from sea. Predictions from BRT models include all native species (top), and only diadromous native species (bottom) respectively. Actual total native species richness values (top; red) and diadromous native species richness values (bottom; green) observed in WRC sampling are overlaid on predicted values.

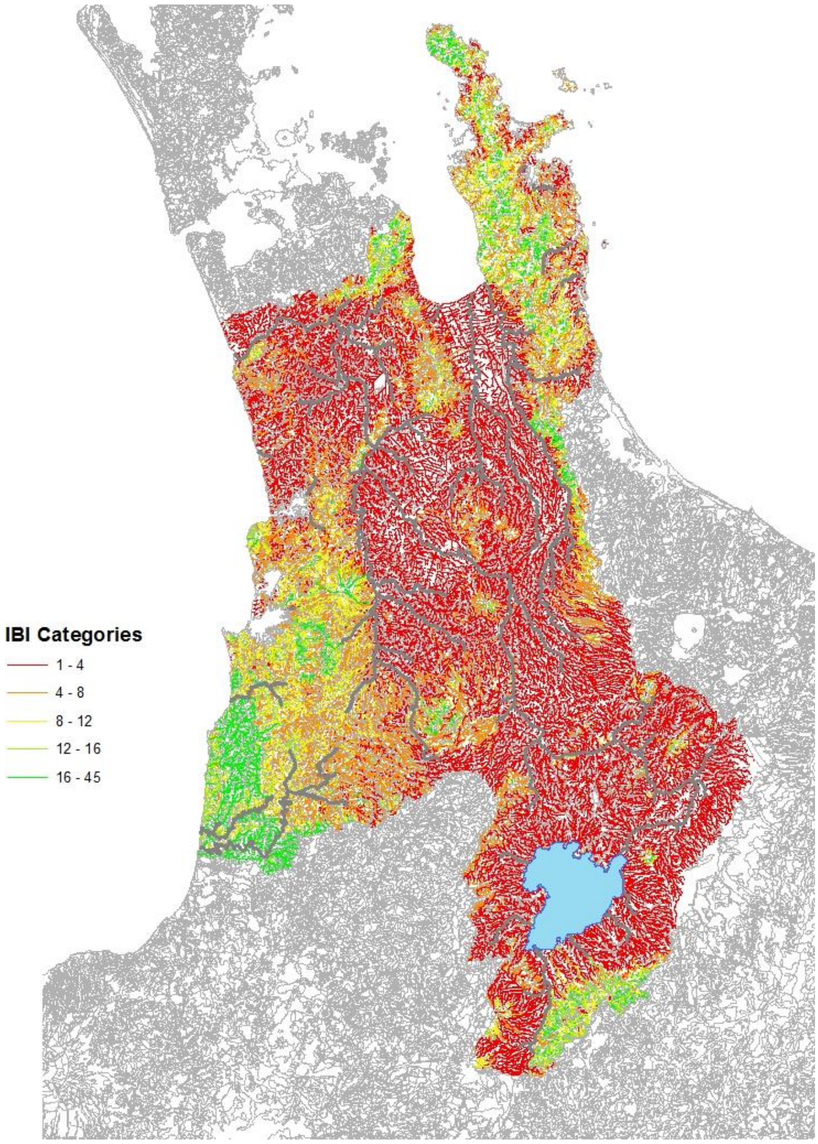


Figure 4.3 Abundance IBI predictions for the Waikato region River Environment Classification waterways as determined from boosted regression tree analysis.

Predictive fish presence maps

Presence probabilities for shortfin eels (*Anguilla australis*) and longfin eels (*Anguilla dieffenbachii*) respectively show distinct differences in distributions throughout the Waikato region (Figures 4.4 and 4.5). Shortfin eels have high presence probability over much of the central Waikato (Figure 4.4) while longfin eels look to be excluded to a greater extent from the central Waikato with higher presence probabilities mostly towards coastal areas (Figure 4.5).

Redfin bully distribution appears to be strongly distance limited with very low probability of presence inland (Figure 4.6). Highest probabilities are again found on the west coast of the Waikato region and throughout the Coromandel peninsula. As discussed in model outputs, distance is the most important indicator accounting for (40%) of model variation (Table 4.3).

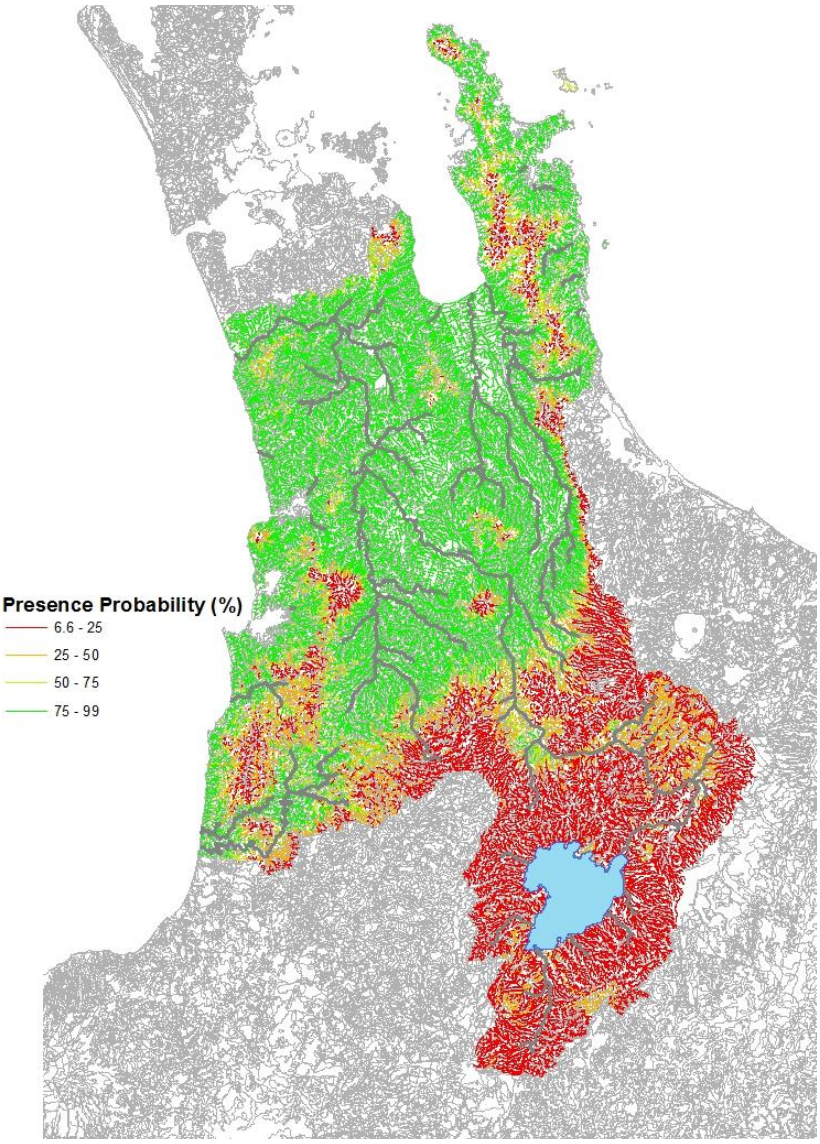


Figure 4.4 Probabilities of the presence of shortfin eels in the Waikato Region based on a BRT model built from presence/absence data. Green areas show high presence probability and red areas low presence probability. Model performance for shortfin eels was 0.889 (ROC score) and predictive deviance was 0.8.

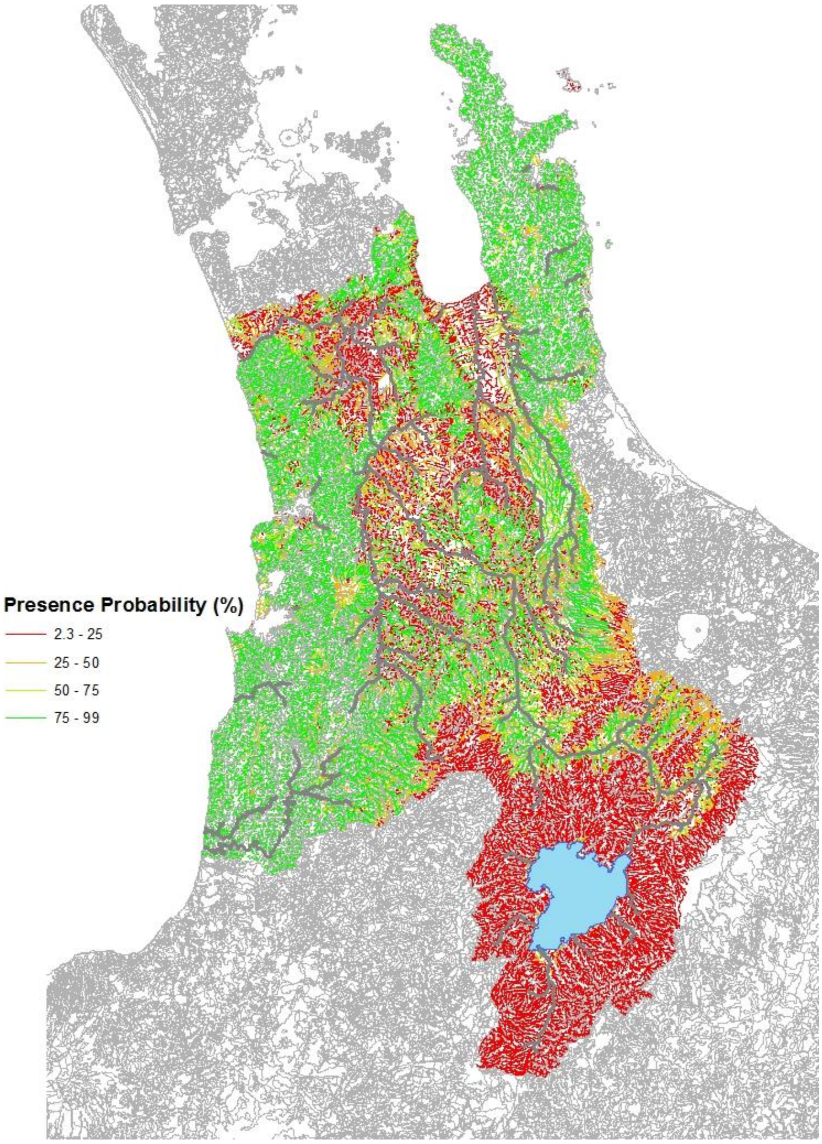


Figure 4.5 Presence probabilities for longfin eels in the Waikato Region based on BRT model built from presence/absence data. Green areas show high presence probability and red areas low presence probability. Model performance for longfin eels was 0.957 (ROC score) and predictive deviance was 0.531.

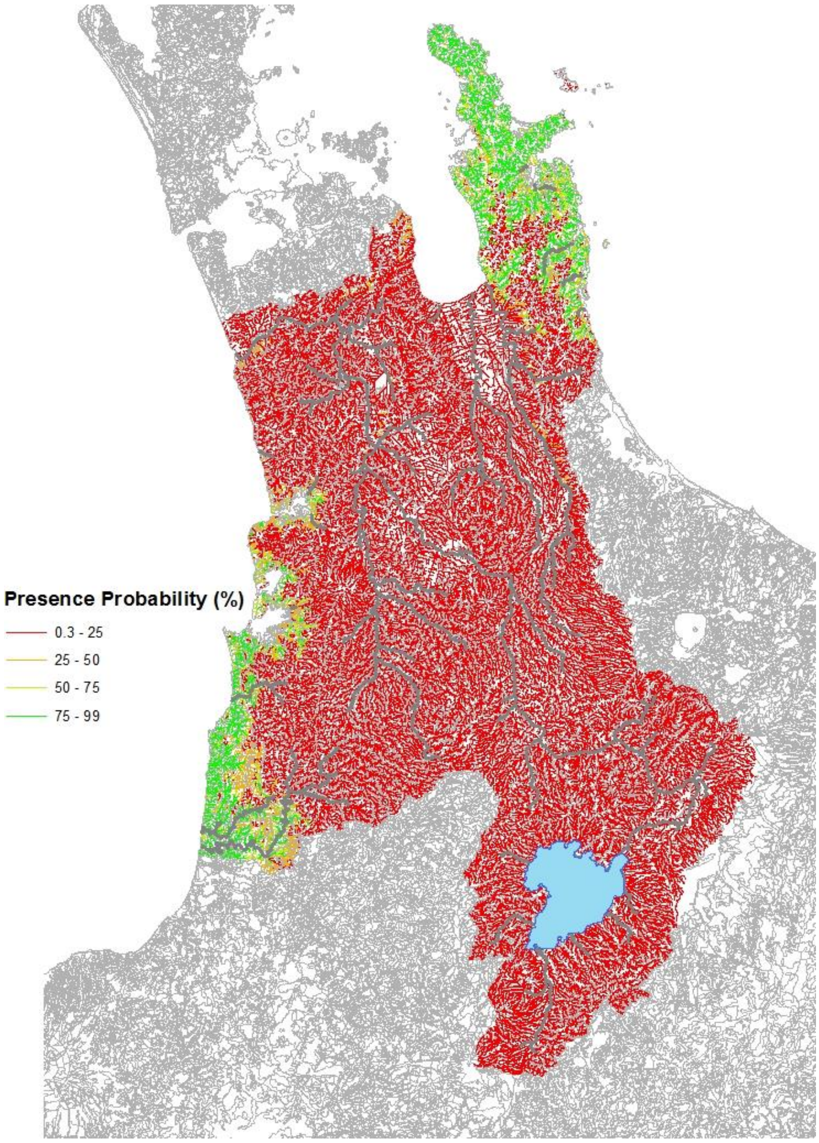


Figure 4.6 Presence probabilities for redfin bullies in the Waikato Region based on BRT model built from presence/absence data. Green areas show high presence probability and red areas low presence probability. Model performance for redfin bullies was 0.986 (ROC score) and predictive deviance was 0.269.

Predictive maps of fish abundance

Abundance predictions for the three key species show somewhat similar patterns as presence/absence predictions. For shortfin eels, highest abundance predictions occur throughout the central Waikato region with low numbers towards coastal areas and on the Coromandel peninsula. The opposite is true for longfin eels with lower numbers throughout the central Waikato and higher numbers particularly around the Raglan area. Both eel species have low numbers towards and around Lake Taupo (Figures 4.7 and 4.8). For redfin bullies, the abundance map shows a similar pattern to that for longfin eels with higher numbers around

Coromandel and Raglan, however the distribution of areas with increased abundance is restricted in comparison to areas closer to the coast (Figure 4.9).

Landcare Database land use classes for the Waikato region provides an insight into the drivers of abundance of these species (Figure 4.10). Shortfin eel abundance patterns in the Waikato appear to match closely with modified lowland areas: areas of exotic grassland around Thames and through the central Waikato shows high predicted relative abundance of shortfin eels. Conversely, the abundance predictions for longfin eels exhibit almost the opposite pattern where they tend have low numbers in these central lowland areas with higher relative abundance restricted to areas of indigenous forest. Higher numbers of redfin bullies are restricted to areas that are both close to the coast and those that are predominantly in indigenous forest. This is again supported by earlier analyses (Table 4.6) highlighting the REC variables most influential in determining redfin bully abundance. Decreasing distance from the sea, increasing segment indigenous forest cover, and increasing upstream indigenous forest cover were the top three drivers of redfin bully abundance. These characteristics are discussed further in relation to resident and juvenile redfin bullies in chapter 5.

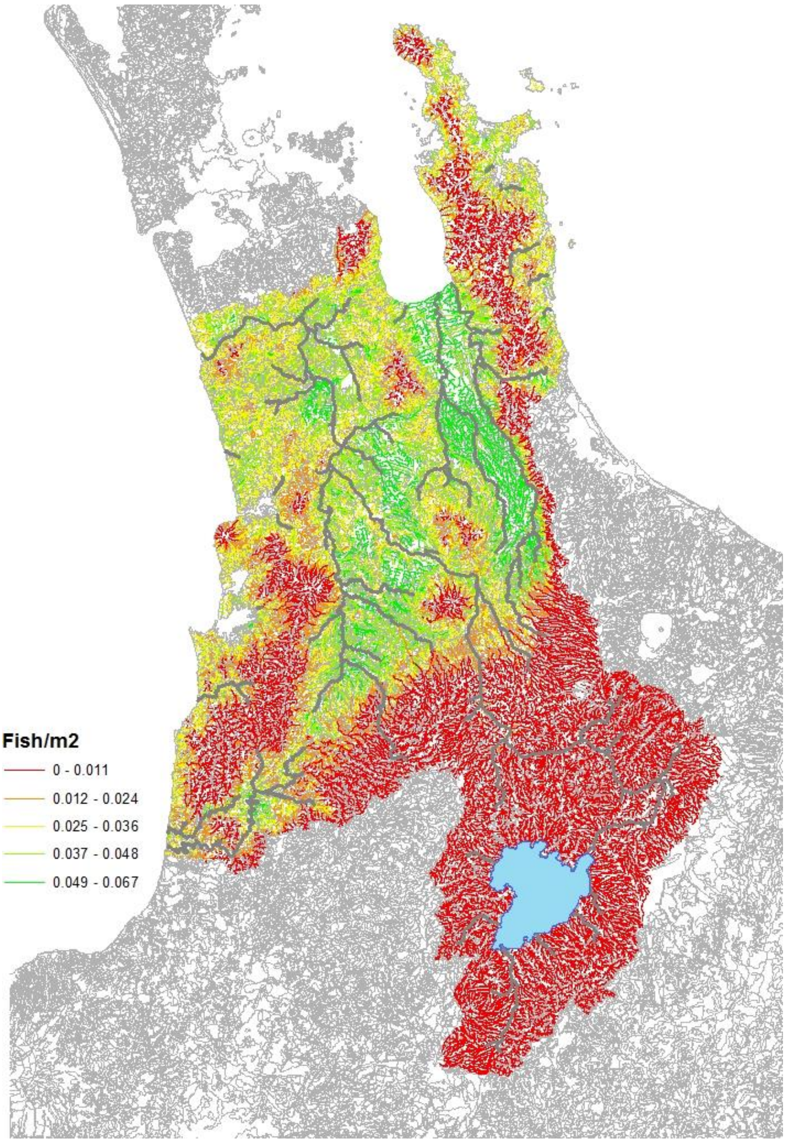


Figure 4.7 Relative abundance predictions for shortfin eels in the Waikato Region based on the BRT model built from WRC abundance data. Green areas show high relative abundance and red areas low relative abundance. Model performance for shortfin eel relative abundance was 0.474 and predictive deviance 0.03.

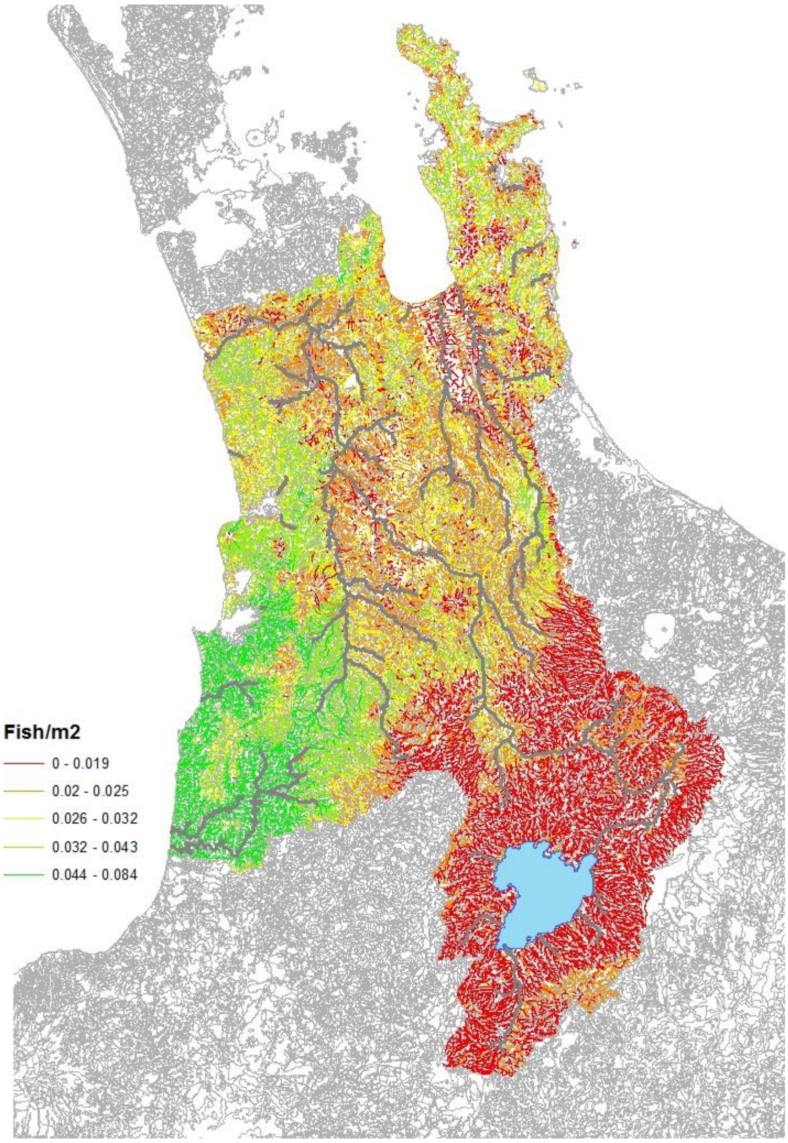


Figure 4.8 Relative abundance predictions for longfin eels in the Waikato Region based on the BRT model built from WRC abundance data. Green areas show high relative abundance and red areas low relative abundance. Model performance for longfin eel relative abundance was 0.679 and predictive deviance 0.019.

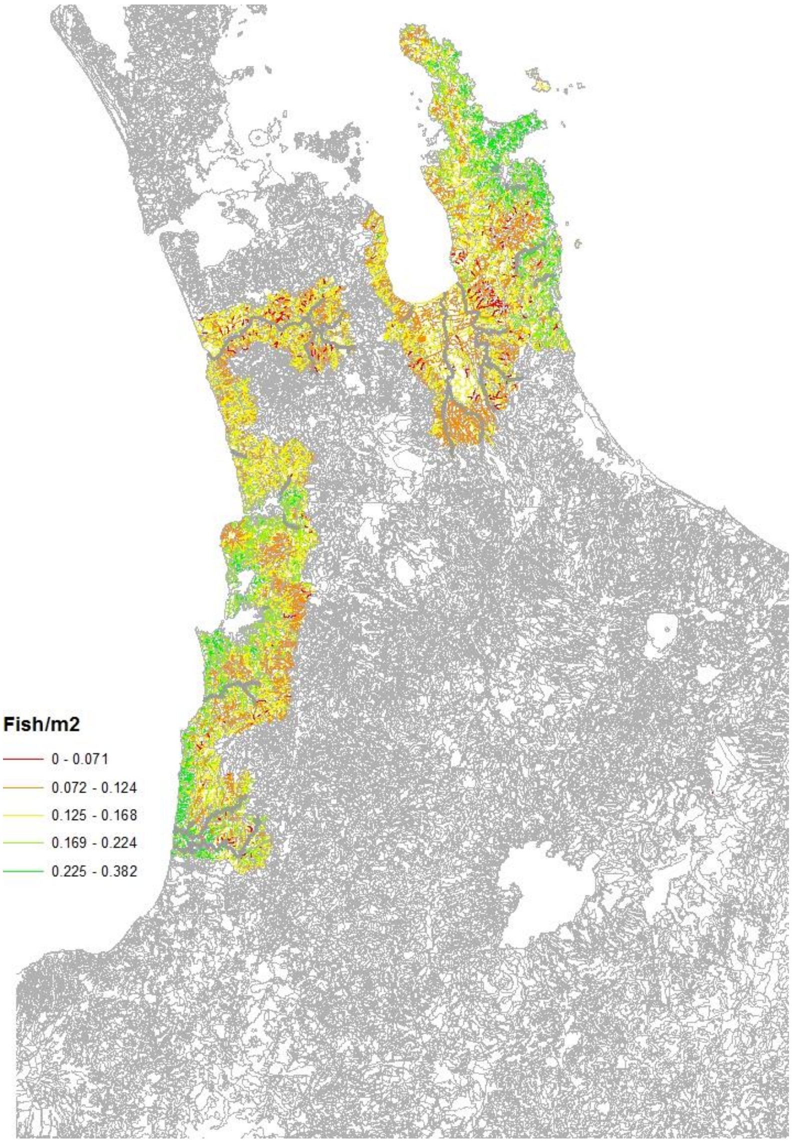


Figure 4.9 Relative abundance predictions for redfin bullies in the Waikato Region based on the BRT model built from WRC abundance data. Green areas show high relative abundance and red areas low relative abundance. Model performance for redfin bullies relative abundance was 0.694 (correlation) and predictive deviance 0.051. Model predictions are limited to 60km from the coast following initial analyses mapping results in Figure 4.6.

Landcare database 4 (LCDB4) land use categories for the Waikato region are dominated by high producing exotic pasture grassland and indigenous forest[11] (Figure 4.10).

---

[11] http://www.landcareresearch.co.nz/resources/data

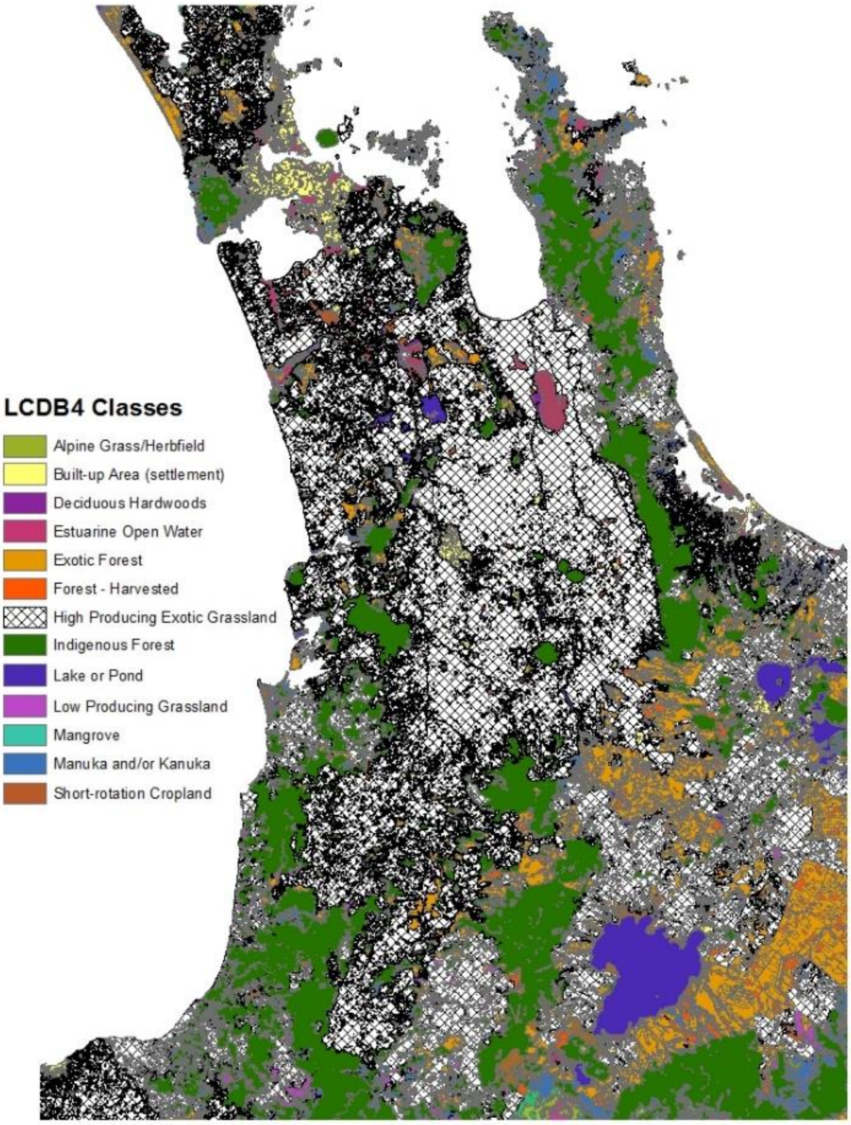


Figure 4.10 Landcare Database 4 (2012) categories for land use in the Waikato region. Main categories of note are *high producing exotic grassland*, and *indigenous forest*.

Predicted fish communities in the Waikato

Longfin eels, shortfin eels, and Crans' bullies are among the most common species in community analyses for both raw and predicted data (Tables 4.7 and 4.8). For raw data the most common community in the WRC dataset is shortfin eels, longfin eels, and Crans' bullies, followed by waterways with only the two eel species, followed by waterways with zero species. For predicted data, waterways with no species present are the most common, making up over a quarter of REC waterways in the Waikato. This is followed by waterways with only longfin and shortfin eels. Eels are present in all waterways with other common fish species present at lower proportions of modelled waterways throughout the Waikato region. Redfin bullies appear in the fifth and sixth most common communities for raw and predicted datasets, respectively.

That the largest proportion are those with no species may be a reflection of a paucity in habitat types sampled in relation to the entire network of REC waterways in the Waikato region (i.e., there are proportionally a large number of first order streams are not sampled given the equal 1st to 4th order stream sample design in the WRC network), or a function of the anthropogenic and/or natural hydrological alteration.

Table 4.7 Ten most common "communities" in the Waikato Regional Council monitoring network from raw presence/absence data.

Species[12]	**# of Occurrences**	**Percentage of Total REC waterways**
SF|LF|BK|In|RFB|BT	3	1.60
SF|LF|TF|Cran B|RT|BT	3	1.60
SF|LF|TF|BK|In|RFB	3	1.60
SF|Cran B	4	2.14
SF|LF|BK|RFB	5	2.67
SF|LF|TF|BK|RFB	5	2.67
LF	12	6.42
0	14	7.49
SF|LF	16	8.56
SF|LF|Cran B	18	9.63

---

[12] LF = longfin Eel (*Anguilla dieffenbachii*); SF = shortfin (*Anguilla australis*); BK = Banded Kokopu (*Galaxias fasciatus*); Gam = Gambusia (*Gambusia affinis*); Com B = Common bully (*Gobiomorphus cotidianus*); RFB = redfin bully (*Gobiomorphus huttoni*); Cran B = Crans' bully (*Gobiomorphus basalis*); BT = Brown Trout (*Salmo trutta*); In = Inanga (*Galaxias maculatus*); RT = Rainbow Trout (*Oncorhynchus mykiss*); TF = Torrentfish (*Cheimarrichthys* fosteri); 0 = No species present.

Table 4.8 Ten most common “communities” in the Waikato region as predicted by BRT models. Values are calculated from waterways from the River Environment Classification (REC) network within the Waikato region. Determination of presence in the waterway is based off 50% or higher predicted probability of presence for a given species. Number of occurrences and percentage of total waterways are calculated for each community type from model predictions.

Species	# of Occurrences	Percentage of Total REC waterways
LF|SF|BK	722	1.42
SF|Gam|Com B	774	1.53
LF|SF|Gam	1101	2.17
LF|SF|BK|RFB	1744	3.44
SF	2039	4.02
LF|SF|Cran B	2998	5.91
SF|Gam	4867	9.60
LF	5204	10.26
LF|SF	7031	13.86
0	13296	26.21

Discussion

Distance, elevation and slope were common model determinants among many native species for both presence and abundance reflecting the influence diadromy can have on native fish populations (McDowall, 1990, 2007). For common species, model performance was high and life history traits relating to diadromy and habitat were identified as primary determinants of distribution and abundance through model building processes. Between the three species investigated in detail, there are marked differences in probabilities of presence and relative abundance across the Waikato region.

Drivers of Fish Distribution

Distance from the coast was one of the most common drivers of fish presence identified in BRT models. Many native fish, particularly diadromous species, are distance limited and the pattern of decreasing richness with distance from the coast has been identified previously (Jowett & Richardson, 1996, 2003; Joy, Henderson, & Death, 2000; Leathwick et al., 2008; McDowall, 2006, 2007). This certainly seems to be the case for redfin bullies, a species for which diadromy is considered obligatory (McDowall, 1990, 2006). For many species this distance pattern likely reflects the propensity for a diadromous life history but in some cases inland lakes and waterways may facilitate non marine recruitment if appropriate larval rearing conditions exist (e.g. common smelt and common bullies) (Closs, Smith, Barry, & Markwitz, 2003; Ward, Northcote, & Boubee, 2005). The ability of many of New Zealand's diadromous species to penetrate inland is likely to be determined by connectivity and locomotory ability. As has been reported in previous studies (Jowett & Richardson, 1996, 2003; McDowall, 1990) and in the present study (Table 4.3) variables such as elevation, slope, and altitude consistently rank as the highest explanatory variables for species presence. For example, the highest predictor of presence for redfin bullies is decreasing distance from the sea, which itself accounts for 40% of the variation seen in the presence/absence predictions for this species. This is an untypically large amount of variation explained when considering factors driving most other species in this dataset and exemplifies the observation that this species are strongly distance limited. The predicted presence distributions illustrates this pattern well with areas of high probability of presence close to the coast versus areas indicating low presence probability in lowland areas (Figure 4.6). Interestingly, for another obligatorily diadromous species, torrentfish, there was a positive relationship with distance inland and probability of presence. In the Rakaia River in the South Island, large torrentfish were found to be throughout the length of waterways although smaller individuals were restricted to areas closer to the coast. Juveniles are restricted to the coast during larval stages meaning typically larger individuals are found further upstream where suitable habitat is available (Atkinson & Joy, 2009; McDowall, 2000). That data and analyses in this thesis found a greater likelihood of torrentfish further upstream and may reflect similar demographic distinctions in torrentfish populations in the Waikato region.

Both eel species distributions are determined strongly by altitude or elevation variables for presence probabilities. Elevation, slope, and altitude make up the top variables accounting for presence predictions for these two species with slope having a positive relationship with longfin eels perhaps highlighting their preference for riffle sites (Glova, Jellyman, & Bonnett, 1998; Jellyman & Chisnall, 1999), and altitude/elevation being a negative relationship with likelihood of presence for both species. Both of these drivers may be a reflection of a typically lowland habitat (Jellyman, Chisnall, Sykes, & Bonnett, 2002; McDowall, 1990). Presence distributions of eels shown may reflect land use within the region. These species may be innately limited to lowland areas which would be the majority of the central Waikato, but when land use change begins to alter lowland habitat, their likelihood of presence begins to be pushed to lowland areas closer to the coast (for longfin eels) or those altered habitats may concurrently create wider distributions for those that can utilise degraded environments (shortfin eels) (Jellyman, 2007; McCleave & Jellyman, 2002). Small longfin eels may be restricted to areas closer to the coast where during this life stage they are able to find shelter in bed sediments. Adult longfin habitat, being further upstream may be limited, restricting adults to areas closer to the coast. Shortfin eels may then be able to move further inland as a result of reduced competition from adult longfins (Chisnall, 1994), or by utilising more rapid inland transition periods when competing with recruiting young from both species, or simply by being the dominant species recruiting into waterways (August & Hicks, 2008). Additional factors may then begin to affect the size of eels when restricted to these coastal areas. Density-dependant growth has been identified in longfin eels for example, whereby eels confined to densely populated coastal areas through lack of adult habitat are likely to grow slowly, adding to the reduction in large adult longfin eels present further upstream (Beentjes & Jellyman, 2003).

Examining the drivers of presence for species outside of the three investigated in detail, a brief discussion of trout and smelt are below. The presence probability of the introduced species rainbow trout (*Oncorhynchus mykiss*) is strongly (48.8% +) determined by upstream low flow. This is a comparatively large species, requiring a substantially greater volume of water for consistent habitat. It is unsurprising then that, low flow conditions would be a strong predictor of the presence of this species and studies have established the strong impacts of flow on migration, recruitment and survival of trout species (Avery, Korman, & Persons, 2015; Hayes, Olsen, & Hay, 2010; Korman, Walters, Martell, Pine, & Dutterer, 2011; Kristensen et al., 2011). Perhaps in support of this is the following predictor of rainbow trout presence, upstream catchment area (+). This relates to the volume of water a waterway will receive. Combined, these two variables account for nearly 60% of the variation in the predicted presence of trout (it must be noted here however that the model for trout is based off comparatively poor spatial coverage in relation to three species above– approximately 20% of sites– Figure 3.4). Common smelt (*Retropinna retropinna*) share this variable as the largest explanatory variable for fish presence. This species tends to live at low gradients (McDowall, 1990) and although upstream catchment area might equate to volume of water a waterway will receive, the second largest predictor of presence is altitude. Fitted function analysis for this predictor was negative indicating a reduction in altitude increases the likelihood of smelt presence. This is perhaps expected as smelt are considered poor climbers and are commonly observed in lake populations and lowland areas (Northcote & Chapman, 1999).

Drivers of Fish Abundance

Catchment elevation, calcium concentrations, and upstream scrub cover were primary variables driving shortfin eel abundance. The top explanatory variable for shortfin eel abundance was a negative association with upstream catchment elevation again highlighting this species' predominance in typical lowland habitat. Although both longfin and shortfin eel species are capable of migrating long distances they typically live in lowland habitats and the shortfin eel specifically dominates lowland areas (Jellyman, 2007). Upstream calcium content was the second biggest driver of shortfin abundance, having a positive relationship with abundance. There are many potential interactions with the environment or the species that could result in calcium being identified as a driver of abundance. Nonetheless, several studies have examined the influence of calcium in salinity and its effect on waterway discrimination in *Anguilla* species (Arai, Chino, & Dung Quang, 2013; Gallego et al., 2014; J. Gu et al., 2015; Hamer et al., 2015; Pujolar et al., 2015). Calcium may be assisting shortfin eels in some way to discriminate waterways, selecting those with higher calcium content. Soil calcium content can be high in agricultural/pastoral systems with high rates of leaching observed under some soils for example (Early, Cameron, & Fraser, 1998; Giddens, Parfitt, & Percival, 1997; McIntosh, Hewitt, Giddens, & Taylor, 1997). There may be an interaction here for shortfin eels moving into pastoral catchments although the identification of segment or upstream pastoral cover during model building processes would be expected. Additionally, catchments with higher calcium content may be important in physiological processes as identified in other *Anguilla* studies, although this is hard to explore further without more detailed understanding of specific shortfin eel physiology in relation to calcium in the New Zealand environment. Finally, upstream scrub cover also had a positive relationship with shortfin eel abundance. Looking again at Figure 4.7, high shortfin eel abundance can be seen on the east coast of the Coromandel peninsula for which Manuka/Kanuka scrub land class is most prevalent (Figure 4.10). This may also be a reflection of a proliferation into modified catchments where scrub growth is often extensive following indigenous forest clearance (Dymond, Ausseil, Kirschbaum, Carswell, & Mason, 2013; Hughes & Quinn, 2014; Magesan, Wang, & Clinton, 2012; Scott et al., 2000). As a proxy for catchment modification then, upstream scrub cover may again highlight the influx of shortfin eels into environments where suitable adult habitat for longfin eels, and the associated competition has been removed (Chisnall, 1994; Graynoth, Francis, & Jellyman, 2008).

For longfin eels, the following variables were the top three explanatory variables driving longfin abundance – decreasing distance from the sea, increasing segment length, and decreasing catchment elevation. Distance from the sea may relate to longfin eels having a shorter migration period restricting them to areas closer to the coast (Todd, 1981). However, as many central lowland habitats have been heavily modified, the distance limitation may be more representative of available habitat only existing in areas less modified close the coast. The fact that shortfin eels are not driven in abundance by distance may also support this, with shortfin eels more likely to exploit degraded central lowland areas as discussed (Chisnall, 1994; Chisnall

& Hicks, 1993; Jellyman & Chisnall, 1999). Catchment elevation, as with shortfin eels, is expected given the species' typically lowland habitat. For segment length, previous studies have identified that longfin eels are better associated with riffle sites and areas of more diverse instream habitat (i.e., greater proportion of diverse instream habitat), than are shortfin eels which had no similar association (Glova et al., 1998; Jellyman & Chisnall, 1999). The positive relationship with segment length may reflect this tendency whereby waterways with increased length and potential diversity of habitat provide increased opportunities for longfin eels.

Redfin bully abundance was determined by decreasing distance from sea; increasing segment indigenous forest cover; and increasing upstream indigenous forest cover. Distance has been identified as a strong determinant of redfin bully distributions in New Zealand and it is unsurprising that this ranks as the most important driver of abundance (McDowall, 1990). Upstream indigenous forest cover may relate to available habitat within substrate in waterways, as a function of debris fall, as has been identified as important determinants of redfin bully populations (McEwan & Joy, 2014). Both segment and upstream indigenous forest cover however, rank in the top three variables driving redfin bully abundance and this would likely be a reflection of the need for high quality waterways in unmodified environments. Similarly, two other species had positive relationships with the extent of indigenous forest cover surrounding the waterway in their top three drivers (torrentfish and Crans' bullies), while banded kokopu had a negative relationship with segment pastoral cover. Other New Zealand studies have also found strong associations with catchment forest cover and the presence of koaro and banded kokopu (Eikaas & McIntosh, 2006; Eikaas, McIntosh, & Kliskey, 2005). All native species are likely to flourish more in high quality waterways but some species examined are perhaps more limited by distance and elevation characteristics due to locomotory limitations, or an increased ability to survive in sub-optimal waterways (e.g. eel species) (Chisnall & Hicks, 1993; McCleave & Jellyman, 2002; McDowall, 1990; McDowall & Taylor, 2000).

Smelt abundance was explained by the following variables, in order – decreasing segment average elevation, decreasing segment maximum elevation, and decreasing segment slope. These drivers all indicate a restriction for this species to low elevation waterways with minimal slope. Smelt are a pelagic species and as discussed can often become lake locked. They are typically found in waterways with low elevation and slope and although smelt can migrate long distances they are restricted in doing so in waterways with low slope (Northcote & Chapman, 1999; Ward et al., 2005). This is seen in Waikato waterways such as the Waipa River where without migration barriers and due to the river being of sufficiently low gradient, smelt migrate as far inland as Te Kuiti. Finally, for trout, increasing upstream catchment hardness (soil);

increasing annual runoff from catchment <30°, and increasing upstream low flow were top explanatory variables for relative abundance. As with presence/absence values, these variables seem likely to be related to flow and or discharge in the catchment, as has been identified as important for both recruiting young and adult fish (Avery et al., 2015; Elliot, 1994)[13].

Effects of land use on native fish

Habitat for native species throughout New Zealand has been reduced substantially through modification in land use (Figure 4.10). Throughout the central Waikato and spanning almost to the coast is high producing exotic grassland. These exotic grasslands consist of high density dairy herds with high fertiliser and feed inputs and are often located on fertile flood plains where native forest and riparian vegetation have been cleared to make way for pasture and stock watering (Aarons & Gourley, 2013; Doole & Romera, 2015; McDowell, 2015; Scarsbrook & Melland, 2015). For species such as redfin bullies, which require comparatively pristine water quality and available, appropriate bed substrate for daytime refuge, these types of modifications can be very effective at excluding populations (McEwan & Joy, 2014). Only the most resilient of species can survive in these environments and along with pressures of commercial and recreational fishing of juveniles (whitebait) and adult eels, recruitment and establishment of species into these environments would be very difficult. The decrease in native species coverage in response to increases in intensive agriculture has been seen in many studies internationally; including throughout New Zealand, and has again been highlighted in predictive mapping in this study (Alho, Reis, & Aquino, 2015; Allibone et al., 2010; Clapcott et al., 2012; dos Santos & Esteves, 2015; Hickford & Schiel, 2013; McDowall, 2006; Pease, Taylor, Winemiller, & King, 2015). Loss of instream habitat in these lowland areas is common in the Waikato and is particularly relevant for many areas in New Zealand. The Landcare Database[14] (LCDB4) land use change information details a loss in the Waikato region of 48ha of indigenous forest for example over the period 2008-2014. Over the same period there was an increase of 19810ha of high producing pasture and 753ha of low producing pasture. These trends in intensifying agriculture, in particular dairy intensification, were identified in 2009 in an Environment Waikato regional council technical report investigating the effect of increased agricultural development (Cameron, Barrett, Cochrane, & McNeill, 2009).

---

[13] It should be noted again that abundance conclusions regarding species should be considered in the light of Figure 3.1 which outlines the spatial and abundance coverage of all species. Those that are less represented may have less robust models.

[14] http://www.landcareresearch.co.nz/resources/data

Continued development of native fish habitat will undoubtedly reduce the range and populations of New Zealands' freshwater species. Community analyses of raw and predicted data show less than 10% of WRC monitored waterways and less than 6% of REC waterways in the Waikato region have at least 3 native species (Tables 4.7 and 4.8). Approximately 35% of communities in the Waikato REC waterway network that are predicted to have any species, are composed solely of our two eel species, and the invasive *Gambusia* (Table 4.8). Along with the multitude of studies already highlighting the degradation of freshwater habitat and the threats to freshwater biodiversity in New Zealand, this study has shown with comparatively accurate and robust data, that New Zealand fish communities are becoming composed of only a few of the most hardy species. Without improved protection and restoration of freshwater habitat, New Zealand stands to lose vulnerable and currently contracted populations of our unique freshwater fauna.

Observations from comparisons between WRC and NZFFDB datasets

A full discussion of the improvements afforded to model building through using standardised sampling protocols is provided in chapter 2; however, a brief discussion is provided here in relation to species distributions. When assessing the difference in model performance for the three main species discussed, all abundance models are lower using the New Zealand Freshwater Fish Database dataset than using the Waikato Regional Council dataset. This is likely to be the result of greater consistency in sampling providing less variation in sample data. Of particular note are the differences between longfin eel and redfin bully models both presence/absence and abundance between the NZFFDB dataset and the WRC dataset. For both species the difference between model performances for both presence/absence and abundance are far higher using the WRC dataset than using the NZFFDB dataset. This compares with a relatively lower difference for shortfin eels despite an almost even split in the number of eel records available in the NZFFDBB (shortfin eels = 46.5%, longfin eels = 53.5%). This may suggest a more variable distribution of shortfin eels resulting in less predictive performance. Alternatively, this may relate to a widespread dispersal of shortfin eels as discussed above. The increase in relative abundance of shortfin eels concurrent with the loss of native habitat for, and subsequent loss of, longfin eels is a pattern already observed widely in New Zealand (Baillie, Hicks, van den Heuvel, Kimberley, & Hogg, 2013; Jowett & Richardson, 2003) and an influx and dispersal of shortfin eels *en masse* might contribute to the comparative low predictability with shortfin eel models using Waikato Regional Council data (Chisnall, 1994; Graynoth et al., 2008).

Distributions of redfin bullies reflect a similar pattern to that of longfin eels. Endemic riffle dwellers in New Zealand have been shown to be in decline when looking at the NZFFDB although analyses of the NZFFDBB are subject to limitations already discussed (Allibone et al., 2010). Results from the WRC dataset however also show that, although redfin bullies appear to be strongly limited to areas within 60km of the coast, those areas where they are found also appear to be strongly linked to land use. The Coromandel peninsula for example was an area that featured high numbers of both longfin eels and redfin bullies. This area is predominantly indigenous forest as is the area south of Kawhia (Figure 4.10) where the numbers of redfin bullies are also high. Although there are some interesting differences in relation to length of redfin bullies between the east and west coasts (discussed in detail in chapter 5), it is likely that this species is heavily impacted by land use. As described already, redfin bullies require high quality water and bed substrate for daytime refuge. These two parameters alone in conjunction with distance can likely explain much of the predicted redfin bully distribution: areas surrounded by indigenous forest are known to have higher water quality and lower sediment inputs (Baillie & Neary, 2015). The high predictive performance in the WRC dataset for redfin bullies compared to all other species, both presence/absence and abundance metrics, may further reflect these factors whereby habitat sensitivity provides a stronger predictor of distribution and abundance. For example, redfin bullies are certainly more sensitive to sedimentation than are shortfin eels and require comparatively un-impacted habitat (McEwan & Joy, 2014; Rowe, Hicks, Smith, & Williams, 2009). In addition to distance limitations observed in redfin bullies, this increasing paucity of indigenous forest surrounding coastal waterways suggests little habitat may be left for redfin bullies and other New Zealand species requiring some level of quality habitat. Many studies have identified these effects and preventing further species loss and population contraction appears predicated on better management of land use (Clapcott et al., 2012; Jowett & Richardson, 2003; Rowe, Chisnall, Dean, & Richardson, 1999).

Conclusion

Species distributions in the Waikato region are closely related to life history traits such as diadromy and distance limitations, and to land use patterns particularly in the central Waikato. Distributions of the two eel species appear to be reverse images of each other for both presence and relative abundance with high numbers of shortfin eels, and low numbers of longfin eels located in the central Waikato. Similarly, for redfin bullies, high abundances and presence probabilities are restricted to areas of indigenous forest close to the coast. These distributions likely reflect habitat requirements for all species, and for eels in particular, may reflect changing distributions in response to both habitat loss and fishing pressure with shortfin eels perhaps better equipped to take advantage of these pressures. Implementation of sampling protocols has provided a robust and accurate dataset which can produce accurate models for both abundance and presence/absence. Known factors of distance and elevation limitations for diadromous species were confirmed to a high level of accuracy through model building processes along with identification of variables likely important in New Zealand fish biology.

References

Aarons, S. R., & Gourley, C. J. P. (2013). The role of riparian buffer management in reducing off-site impacts from grazed dairy systems. *Renewable Agriculture and Food Systems, 28*(1), 1-16. doi: 10.1017/s1742170511000548

Alho, C. J. R., Reis, R. E., & Aquino, P. P. U. (2015). Amazonian freshwater habitats experiencing environmental and socioeconomic threats affecting subsistence fisheries. *Ambio, 44*(5), 412-425. doi: 10.1007/s13280-014-0610-z

Allibone, R., David, B., Hitchmough, R., Jellyman, D., Ling, N., Ravenscroft, P., & Waters, J. (2010). Conservation status of New Zealand freshwater fish, 2009. *New Zealand Journal of Marine and Freshwater Research, 44*(4), 271-287. doi: 10.1080/00288330.2010.514346

Arai, T., Chino, N., & Dung Quang, L. (2013). Migration and habitat use of the tropical eels Anguilla marmorata and A. bicolor pacifica in Vietnam. *Aquatic Ecology, 47*(1), 57-65. doi: 10.1007/s10452-012-9424-x

Atkinson, N. K., & Joy, M. K. (2009). Longitudinal size distributions of bluegill bullies (Gobiomorphus hubbsi) and torrentfish (Cheimarrichthys fosteri) in two large New Zealand rivers. *New Zealand Journal of Marine and Freshwater Research, 43*(2), 643-651.

August, S. M., & Hicks, B. J. (2008). Water temperature and upstream migration of glass eels in New Zealand: implications of climate change. *Environmental Biology of Fishes, 81*(2), 195-205. doi: 10.1007/s10641-007-9191-z

Avery, L. A., Korman, J., & Persons, W. R. (2015). Effects of Increased Discharge on Spawning and Age-0 Recruitment of Rainbow Trout in the Colorado River at Lees Ferry, Arizona. *North American Journal of Fisheries Management, 35*(4), 671-680. doi: 10.1080/02755947.2015.1040560

Baillie, B. R., Hicks, B. J., van den Heuvel, M. R., Kimberley, M. O., & Hogg, I. D. (2013). The effects of wood on stream habitat and native fish assemblages in New Zealand. *Ecology of Freshwater Fish, 22*(4), 553-566.

Baillie, B. R., & Neary, D. G. (2015). Water quality in New Zealand's planted forests: a review. *New Zealand Journal of Forestry Science, 45*. doi: 10.1186/s40490-015-0040-0

Beentjes, M. P., & Jellyman, D. J. (2003). Enhanced growth of longfin eels, Anguilla dieffenbachii, transplanted into Lake Hawea, a high country lake in South Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 37*(1), 1-11.

Boubee, J., Chisnall, B., Watene, E., Williams, E., & Roper, D. (2002). Enhancement and management of eel fisheries affected by hydroelectric dams in New Zealand. In D. A. Dixon (Ed.), *Biology, Management, and Protection of Catadromous Eels* (Vol. 33, pp. 191-205). Bethesda: Amer Fisheries Soc.

Boyce, S. W. (1945). Statistical studies on New Zealand wheat trials. I. The efficiency of lattice designs. *New Zealand Jour Sci and Tech, 27 A*((3)), 270-275.

Cameron, M., Barrett, P., Cochrane, B., & McNeill, K. (2009). *Implications of Agricultural Change in the Waikato Region: Current Trends and Future Scenarios* (1505953). Hamilton East: Environment Waikato.

Chisnall, B. L. (1994). *An unexploited mixed species eel stock (Anguilla australis and A. dieffenbachii) in a Waikato pastoral stream, and its modification by fishing pressure.* Wellington: Department of Conservation.

Chisnall, B. L., & Hicks, B. J. (1993). Age and growth of longfinned eels (anguilla-dieffenbachii) in pastoral and forested streams in the Waikato river basin, and in 2 hydroelectric lakes in the North Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 27*(3), 317-332.

Clapcott, J. E., Collier, K. J., Death, R. G., Goodwin, E. O., Harding, J. S., Kelly, D., . . . Young, R. G. (2012). Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity. *Freshwater Biology, 57*(1), 74-90. doi: 10.1111/j.1365-2427.2011.02696.x

Closs, G. P., Smith, M., Barry, B., & Markwitz, A. (2003). Non-diadromous recruitment in coastal populations of common bully (Gobiomorphus cotidianus). *New Zealand Journal of Marine and Freshwater Research, 37*(2), 301-313.

David, B. O., Hamer, M. P., Collier, K. J., Lake, M. D., Surrey, G. M., McArthur, K., . . . Dale, M. (2010). A standardised sampling protocol for robust assessment of reach-scale fish community diversity in wadeable New Zealand streams. *New Zealand Journal of Marine and Freshwater Research, 44*(3), 177-187.

Doole, G. J., & Romera, A. J. (2015). Trade-offs between profit, production, and environmental footprint on pasture-based dairy farms in the Waikato region of New Zealand. *Agricultural Systems, 141*, 14-23. doi: 10.1016/j.agsy.2015.09.005

dos Santos, F. B., & Esteves, K. E. (2015). A Fish-Based Index of Biotic Integrity for the Assessment of Streams Located in a Sugarcane-Dominated Landscape in Southeastern Brazil. *Environmental Management, 56*(2), 532-548. doi: 10.1007/s00267-015-0516-y

Dymond, J. R., Ausseil, A. G. E., Kirschbaum, M. U. F., Carswell, F. E., & Mason, N. W. H. (2013). Opportunities for restoring indigenous forest in New Zealand. *Journal of the Royal Society of New Zealand, 43*(3), 141-153. doi: 10.1080/03036758.2012.736393

Early, M. S. B., Cameron, K. C., & Fraser, P. M. (1998). The fate of potassium, calcium, and magnesium in simulated urine patches on irrigated dairy pasture soil. *New Zealand Journal of Agricultural Research, 41*(1), 117-124.

Eikaas, H. S., & McIntosh, A. R. (2006). Habitat loss through disruption of constrained dispersal networks. *Ecological Applications, 16*(3), 987-998. doi: 10.1890/1051-0761(2006)016[0987:hltdoc]2.0.co;2

Eikaas, H. S., McIntosh, A. R., & Kliskey, A. D. (2005). Catchment- and site-scale influences of forest cover and longitudinal forest position on the distribution of a diadromous fish. *Freshwater Biology, 50*(3), 527-538. doi: 10.1111/j.1365-2427.2005.01337.x

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., . . . Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography, 29*(2), 129-151. doi: 10.1111/j.2006.0906-7590.04596.x

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802-813. doi: 10.1111/j.1365-2656.2008.01390.x

Elliot, J. M. (1994). *Quantitative ecology and the brown trout*. USA: Oxford University Press.

Gallego, V., Martinez-Pastor, F., Mazzeo, I., Penaranda, D. S., Herraez, M. P., Asturiano, J. F., & Perez, L. (2014). Intracellular changes in Ca2+, K+ and pH after sperm motility activation in the European eel (Anguilla anguilla): Preliminary results. *Aquaculture, 418*, 155-158. doi: 10.1016/j.aquaculture.2013.10.022

Giddens, K. M., Parfitt, R. L., & Percival, H. J. (1997). Comparison of some soil properties under Pinus radiata and improved pasture. *New Zealand Journal of Agricultural Research, 40*(3), 409-416.

Glova, G. J., Jellyman, D. J., & Bonnett, M. L. (1998). Factors associated with the distribution and habitat of eels (Anguilla spp) in three New Zealand lowland streams. *New Zealand Journal of Marine and Freshwater Research, 32*(2), 255-269.

Graynoth, E., Bonnett, M., & Jellyman, D. (2012). Estimation of native fish density in lowland streams by repeated electric fishing during the day and following night. *New Zealand Journal of Marine and Freshwater Research, 46*(2), 243-261. doi: 10.1080/00288330.2011.638646

Graynoth, E., Francis, R. I. C. C., & Jellyman, D. J. (2008). Factors influencing juvenile eel (Anguilla spp.) survival in lowland New Zealand streams. *New Zealand Journal of Marine and Freshwater Research, 42*(2), 153-172.

Gu, J., Li, J.-W., Tse, W. K.-F., Chan, T.-F., Lai, K.-P., & Wong, C. K.-C. (2015). Transcriptomic responses of corpuscle of Stannius gland of Japanese eels (Anguilla japonica) to Changes in Water Salinity. *Scientific Reports, 5*. doi: 10.1038/srep09836

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters, 8*(9), 993-1009. doi: 10.1111/j.1461-0248.2005.00792.x

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., . . . Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters, 16*(12), 1424-1435. doi: 10.1111/ele.12189

Hamer, P., Henderson, A., Hutchison, M., Kemp, J., Green, C., & Feutry, P. (2015). Atypical correlation of otolith strontium : calcium and barium : calcium across a marine-freshwater life history transition of a diadromous fish. *Marine and Freshwater Research, 66*(5), 411-419. doi: 10.1071/mf14001

Hayes, J. W., Olsen, D. A., & Hay, J. (2010). The influence of natural variation in discharge on juvenile brown trout population dynamics in a nursery tributary of the Motueka River, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 44*(4), 247-269. doi: 10.1080/00288330.2010.509905

Heino, J. (2013). The importance of metacommunity ecology for environmental assessment research in the freshwater realm. *Biological Reviews, 88*(1), 166-178. doi: 10.1111/j.1469-185X.2012.00244.x

Hickford, M. J. H., & Schiel, D. R. (2013). Artificial Spawning Habitats Improve Egg Production of a Declining Diadromous Fish, Galaxias maculatus (Jenyns, 1842). *Restoration Ecology, 21*(6), 686-694. doi: 10.1111/rec.12008

Hughes, A. O., & Quinn, J. M. (2014). Before and After Integrated Catchment Management in a Headwater Catchment: Changes in Water Quality. *Environmental Management, 54*(6), 1288-1305. doi: 10.1007/s00267-014-0369-9

Jellyman, D. J. (2007). Status of New Zealand fresh-water eel stocks and management initiatives. *Ices Journal of Marine Science, 64*(7), 1379-1386. doi: 10.1093/icesjms/fsm073

Jellyman, D. J., & Chisnall, B. L. (1999). Habitat preferences of shortfinned eels (Anguilla australis), in two New Zealand lowland lakes. *New Zealand Journal of Marine and Freshwater Research, 33*(2), 233-248.

Jellyman, D. J., Chisnall, B. L., Sykes, J. R. E., & Bonnett, M. L. (2002). Variability in spatial and temporal abundance of glass eels (Anguilla spp.) in New Zealand waterways. *New Zealand Journal of Marine and Freshwater Research, 36*(3), 511-517.

Joseph, L. N., Field, S. A., Wilcox, C., & Possingham, H. P. (2006). Presence-absence versus abundance data for monitoring threatened species. *Conservation Biology, 20*(6), 1679-1687. doi: 10.1111/j.1523-1739.2006.00529.x

Jowett, I. G., & Richardson, J. (1996). Distribution and abundance of freshwater fish in New Zealand rivers. *New Zealand Journal of Marine and Freshwater Research, 30*(2), 239-255.

Jowett, I. G., & Richardson, J. (2003). Fish communities in New Zealand rivers and their relationship to environmental variables. *New Zealand Journal of Marine and Freshwater Research, 37*(2), 347-366.

Joy, M. K. (2013). *Freshwater fish predictive modelling for bioassessment; A scoping study into fish bioassessment models as national indicators in New Zealand*. Wairesearch.

Joy, M. K., Henderson, I. M., & Death, R. G. (2000). Diadromy and longitudinal patterns of upstream penetration of freshwater fish in Taranaki, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 34*(3), 531-543.

Korman, J., Walters, C., Martell, S. J. D., Pine, W. E., III, & Dutterer, A. (2011). Effects of flow fluctuations on habitat use and survival of age-0 rainbow trout (Oncorhynchus mykiss) in a large, regulated river. *Canadian Journal of Fisheries and Aquatic Sciences, 68*(6), 1097-1109. doi: 10.1139/f2011-045

Kristensen, E. A., Closs, G. P., Olley, R., Kim, J., Reid, M., & Stirling, C. (2011). Determining the spatial distribution of spawning by anadromous and resident brown trout Salmo trutta L using strontium content of eggs collected from redds. *Ecology of Freshwater Fish, 20*(3), 377-383. doi: 10.1111/j.1600-0633.2010.00451.x

Leathwick, J. R., Elith, J., Chadderton, W. L., Rowe, D., & Hastie, T. (2008). Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *Journal of Biogeography, 35*(8), 1481-1497. doi: 10.1111/j.1365-2699.2008.01887.x

Magesan, G. N., Wang, H., & Clinton, P. W. (2012). Nitrogen cycling in gorse-dominated ecosystems in New Zealand. *New Zealand Journal of Ecology, 36*(1), 21-28.

Mancini, A., Elsadek, I., & Madon, B. (2015). When simple is better: Comparing two sampling methods to estimate green turtles abundance at coastal feeding grounds. *Journal of Experimental Marine Biology and Ecology, 465*, 113-120. doi: 10.1016/j.jembe.2015.01.004

McCleave, J. D., & Jellyman, D. J. (2002). Discrimination of New Zealand stream waters by glass eels of Anguilla australis and Anguilla dieffenbachii. *Journal of Fish Biology, 61*(3), 785-800. doi: 10.1006/jfbi.2002.2108

McDowall, R. M. (1990). *New Zealand freshwater fishes : a natural history and guide* (Rev. ed ed.). Wellington [N.Z.]: Heinemann Reed, MAF Publishing Group.

McDowall, R. M. (2000). Biogeography of the New Zealand torrentfish, Cheimarrichthys fosteri (Teleostei : Pinguipedidae): a distribution driven mostly by ecology and behaviour. *Environmental Biology of Fishes, 58*(2), 119-131. doi: 10.1023/a:1007666014842

McDowall, R. M. (2006). Fish, fish habitats and fisheries in New Zealand. *Aquatic Ecosystem Health & Management, 9*(4), 391-405. doi: 10.1080/14634980601026352

McDowall, R. M. (2007). On amphidromy, a distinct form of diadromy in aquatic organisms. *Fish and Fisheries, 8*(1), 1-13. doi: 10.1111/j.1467-2979.2007.00232.x

McDowall, R. M., & Taylor, M. J. (2000). Environmental indicators of habitat quality in a migratory freshwater fish fauna. *Environmental Management, 25*(4), 357-374. doi: 10.1007/s002679910028

McDowell, R. W. (2015). Relationship between Sediment Chemistry, Equilibrium Phosphorus Concentrations, and Phosphorus Concentrations at Baseflow in Rivers of the New Zealand National River Water Quality Network. *Journal of Environmental Quality, 44*(3), 921-929. doi: 10.2134/jeq2014.08.0362

McEwan, A. J., & Joy, M. K. (2014). Habitat use of redfin bullies (Gobiomorphus huttoni) in a small upland stream in Manawatu, New Zealand. *Environmental Biology of Fishes, 97*(2), 121-132. doi: 10.1007/s10641-013-0131-9

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., . . . White, E. P. (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters, 10*(10), 995-1015. doi: 10.1111/j.1461-0248.2007.01094.x

McIntosh, P. D., Hewitt, A. E., Giddens, K., & Taylor, M. D. (1997). Benchmark sites for assessing the chemical impacts of pastoral farming on loessial soils in southern New Zealand. *Agriculture Ecosystems & Environment, 65*(3), 267-280. doi: 10.1016/s0167-8809(97)00071-6

Miller, J. H., Behrensmeyer, A. K., Du, A., Lyons, S. K., Patterson, D., Toth, A., . . . Reed, D. (2014). Ecological fidelity of functional traits based on species presence-absence in a modern mammalian bone assemblage (Amboseli, Kenya). *Paleobiology, 40*(4), 560-583. doi: 10.1666/13062

Northcote, T. G., & Chapman, M. A. (1999). Dietary alterations in resident and migratory New Zealand common smelt (Retropinna retropinna) in lower Waikato lakes after two decades of habitat change. *New Zealand Journal of Marine and Freshwater Research, 33*(3), 425-436.

NZFFDB. (2015). New Zealand Freshwater Fish Database. Retrieved 2015, from https://www.niwa.co.nz/our-services/online-services/freshwater-fish-database

Pearce, J. L., & Boyce, M. S. (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology, 43*(3), 405-412. doi: 10.1111/j.1365-2664.2005.01112.x

Pease, A. A., Taylor, J. M., Winemiller, K. O., & King, R. S. (2015). Ecoregional, catchment, and reach-scale environmental factors shape functional-trait structure of stream fish assemblages. *Hydrobiologia, 753*(1), 265-283. doi: 10.1007/s10750-015-2235-z

Pujolar, J. M., Jacobsen, M. W., Bekkevold, D., Lobon-Cervia, J., Jonsson, B., Bernatchez, L., & Hansen, M. M. (2015). Signatures of natural selection between life cycle stages separated by metamorphosis in European eel. *BMC Genomics, 16*(600), (13 August 2015)-(2013 August 2015).

Ramsey, D. S. L., Caley, P. A., & Robley, A. (2015). Estimating Population Density From Presence-Absence Data Using a Spatially Explicit Model. *Journal of Wildlife Management, 79*(3), 491-499.

Reiss, H., Birchenough, S., Borja, A., Buhl-Mortensen, L., Craeymeersch, J., Dannheim, J., . . . Degraer, S. (2015). Benthos distribution modelling and its relevance for marine ecosystem management. *Ices Journal of Marine Science, 72*(2), 297-315. doi: 10.1093/icesjms/fsu107

Robinson, L. M., Elith, J., Hobday, A. J., Pearson, R. G., Kendall, B. E., Possingham, H. P., & Richardson, A. J. (2011). Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography, 20*(6), 789-802. doi: 10.1111/j.1466-8238.2010.00636.x

Rowe, D. K., Chisnall, B. L., Dean, T. L., & Richardson, J. (1999). Effects of land use on native fish communities in east coast streams of the North Island of New Zealand. *New Zealand Journal of Marine and Freshwater Research, 33*(1), 141-151.

Rowe, D. K., Hicks, M., Smith, J. P., & Williams, E. (2009). Lethal concentrations of suspended solids for common native fish species that are rare in New Zealand rivers with high suspended solids loads. *New Zealand Journal of Marine and Freshwater Research, 43*(5), 1029-1038.

Scarsbrook, M. R., & Melland, A. R. (2015). Dairying and water-quality issues in Australia and New Zealand. *Animal Production Science, 55*(7), 856-868. doi: 10.1071/an14878

Scott, N. A., White, J. D., Townsend, J. A., Whitehead, D., Leathwick, J. R., Hall, G. M. J., . . . Whaley, P. T. (2000). Carbon and nitrogen distribution and accumulation in a New Zealand scrubland ecosystem. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere, 30*(8), 1246-1255. doi: 10.1139/cjfr-30-8-1246

Todd, P. R. (1981). Timing and periodicity of migrating New Zealand fresh-water eels (anguilla spp). *New Zealand Journal of Marine and Freshwater Research, 15*(3), 225-235.

Ward, F. J., Northcote, T. G., & Boubee, J. A. T. (2005). The New Zealand common smelt: biology and ecology. *Journal of Fish Biology, 66*(1), 1-32. doi: 10.1111/j.1095-8649.2004.00569.x

Weeks, E. S., Anderson-Lederer, R., Death, R. G., & Joy, M. K. (2015). *Are we ignoring the plight of New Zealand's freshwater species.* Society for Conservation Biology (Oceania), Sydney.

Chapter Five

Spatial Variation in Redfin Bully Populations in the North Island, New Zealand; East versus West

Chapter 5 Spatial Variation in Redfin Bully Populations in the North Island, New Zealand; East versus West.

Introduction

The redfin bully (*Gobiomorphus huttoni*) is a benthic freshwater fish found in high quality waterways throughout coastal and inner margins of coastlines in New Zealand. They are an amphidromous eleotrid adapted to generally moderate to fast flowing waters (Vanderpham, Nakagawa, & Closs, 2013a, 2013b). Adult fish range in size from 50-100mm with characteristic red stripes on the dorsal, anal, and tail fins (McQueen, 2013; Tonkin, Wright, & David, 2012). They exhibit a diurnal habit like many New Zealand fish and they become more active at night with one study showing fish returning to larger substrate areas during the day (McEwan & Joy, 2014). There is still much to be learned about the life history and factors affecting the distribution and growth of redfin bullies in New Zealand. Little is known about factors affecting their colonisation and recruitment rates, or how these differ between regions. For instance, Waikato Regional Council aquatic ecologists noted that in two streams near Coromandel town where redfins are predicted to occur with high probability none were captured. Oddly, the streams appear to support high numbers of either Crans' or common bullies instead. Another interesting observation by this team was an apparent size discrepancy for redfin bullies between streams on the east coast compared to the west coast of the Waikato region with west coast individuals having a greater mean length. These differences were first noticed in the WRC monitoring programme between 2009 and 2014, although investigation into the scale and cause of differences has yet to be undertaken. Furthermore, on the east coast of the South Island, there appears to be an isolated population around Banks Peninsula, with distinct gaps north to Kaikoura and south to Oamaru (NZFFDB, 2015). Although these differences have been observed, current understanding of the biology and life history of redfin bullies offers little explanation for this anomaly. Ultimately, there are still large knowledge gaps regarding the life history of most freshwater fish in New Zealand, and in particular with the larval stages of migratory fish and their interaction with oceanic variables, which is hampering their effective management (Allibone et al., 2010; Jowett & Richardson, 1996, 2003; Leathwick, Elith, Chadderton, Rowe, & Hastie, 2008).

Until recently, and despite a substantial number of records available within the New Zealand Freshwater Fish Database (NZFFDB), consistently collated size and abundance data over a sufficient geographic area has not been available to robustly investigate these observations. Records within the NZFFDB are often lacking biometrics and have little or no information about sampling methodology. Analysis beyond presence/absence of species is therefore dubious. With the recent availability of reliable and detailed data through the Waikato Regional Council (WRC) monitoring programme it is now possible to interrogate and explore differences in Waikato redfin bully populations accurately. Exploring these length and abundance differences in redfin bullies is an important step in better understanding the populations that exist in different areas in the country and how these might relate to land use. The distribution, abundance, and length of redfin bullies - as a locally abundant benthic fish that requires appropriate bed substrate for habitat and egg laying (McDowall, 1990, 2006) - may even provide surrogate metrics to highlight changes in riverine processes such as sedimentation or hydraulic alteration. Through beginning to develop an understanding of spatial population variability of size and abundance distributions of redfin bullies, scientists and conservation managers will be better equipped to effectively manage populations of redfin bullies and perhaps begin to ask similar questions about other freshwater species.

The aim of this chapter was to investigate these differences in redfin bully population characteristics using the WRC dataset. Results from this dataset were also compared with results from the NZFFDB.

Methods

Data collection

A full discussion of sampling strategy and standardisation for WRC data has been provided in chapter 2 and as such, a brief reiteration is provided here.

Data collection for redfin bullies and other species is ongoing in the Waikato through the WRC freshwater monitoring team. Data within this study represents a period from 2009-2014. Sampling is undertaken using electro-fishing and the network of monitored sites includes 9 reference sites sampled annually and a rotation of impact or random sites sampled on a 3 year basis. Reaches are measured to 150m in length and divided into ten, 15m subreaches with all sampling methods followed to New Zealand Freshwater Fish Sampling Protocols (NZFFSP) (Joy, David, & Lake, 2013). Count data and fish length data for redfin bullies are collected during fishing events. With respect to measurement of fish length, the first 50 individuals are measured and then an additional 10 fish (first 10 captured) per subsequent subreach after that. In effect if redfin bullies are extremely abundant throughout the entire sampling reach, a total of 140 individuals will be measured (50 in first subreach + 10x 9 subsequent subreaches). A location map of reference and impact sites is provided in chapter 2 (Figure 2.1).

Analysis

All analyses were undertaken using R statistical software (R Core Team, 2014) and ANOVA analyses of mean relative abundance and length data were first undertaken to assess differences by year and site type (reference vs impact). Sites were then split into east and west localities and analysed using two sample t-tests, followed by separation and subsequent analysis of data into subsets containing juveniles and adult individuals, and then adults alone. This was undertaken in an effort to remove the noise introduced from individuals not resident in the waterway (i.e. differences in abundance resulting from differences in juvenile recruitment (e.g. (Maes, Stevens, & Ollevier, 2005; McDowall, 2007)). Sampling for sites in the WRC dataset occurs between December and April which will incorporate periods of recruitment and the presence of a much higher and much more variable proportion of small fish (Closs, Hicks, & Jellyman, 2013; McDowall, 2007). By removing individuals that are estimated to be juveniles a more accurate picture of the core population of redfin bullies is predicted to be seen. An arbitrary size cut-off of >50mm was chosen to split populations into juveniles and adults (this is somewhat based off discussion with WRC staff, however, juvenile size is likely to be variable within and between populations. Establishing data of age and size for redfin bullies is currently underway but as yet unavailable. As such, a >50mm cut-off was determined to be appropriate for a first investigation into population abundance characteristics). These two subsets were then compared for east and west localities again both for reference and impact sites over the monitoring period years. To examine these differences further the same plots were broken down by year (in the earlier years of sampling the whole network was not sampled each year, and therefore years from 2012 onwards are shown (e.g., some years do not have an east/west comparison)).

Analysis of the differences between east and west on a national scale was then undertaken using NZFFDB data for both count and length data. In regards to a comparison between the two datasets, not all records of redfin bullies in the NZFFDB have any length measurement and those that do measure a maximum length and a minimum length for the whole sampling event. There are obviously limitations in mean length comparisons then considering all individuals are measured in the WRC dataset. However, comparisons were made between mean maximum and mean minimum values for east and west locations in the WRC dataset to compare with NZFFDB data. Further, as with model building in chapters 2 and 3, it is not certain which sampling protocols were used and thus consistency between samples cannot be assumed. Nonetheless, differences were compared between east and west locations for the north and the south island for both maximum and minimum lengths.

Modelling and mapping

Using the WRC dataset, boosted regression tree (BRT) models were built to predict both mean length and relative abundance. Relative abundance was calculated using the same methodology as species models in chapter 4. Abundance was modelled here incorporating only individuals >50mm to investigate potential recruitment characteristics as discussed further below. Mean length was also modelled for this purpose and differences were mapped in a regional context using River Environment Classification (REC) environmental data as with chapter 4. This was done using mean lengths derived from subsets of data for both adult and adult and juvenile individuals. Boosted regression tree models were built for each mean length group using *Gaussian* family distribution modelling. Waterways within 60km from the coast were then selected from the REC network and the built models were used to predict the mean length of redfin bullies in each waterway within the Waikato region. Waterways within 60km were selected in an effort to focus on waterways where redfin bullies have a high known and predicted distribution (e.g. chapter 4). An abundance BRT model was also built for redfin bully distributions focusing on individuals greater than 50mm to compare with the abundance model in chapter 4 which was built using all redfin individuals (relative abundance data was modelled as per chapter 4).

Mapping of model predictions

ARCGIS (ARCMAP 10.1) was used to map modelled mean length and abundance predictions across the Waikato region.

Results

The length of redfin bullies varied over the monitoring period and was significantly higher at reference sites compared with impact sites ($F_{1, 3811} = 75.67$, $p<0.000$) (Figure 5.1). ANOVA results show mean relative redfin abundance was higher at reference sites than impact ($F_{1, 186}=87.66$, $p<0.000$) during WRC sampling seasons from 2009 – 2014. Also apparent is the large amount of variation in redfin bully abundance at reference sites, particularly in 2010. It should be noted that the sampling programme was in its infancy in the early years with a consistent network of sites yet to be sampled. This becomes more relevant in subsequent figures [15].

---

[15] Sampling in 2009 and 2010 reflects both a smaller number of samples and a higher proportion of reference sites – see Table 2.1, chapter 2.

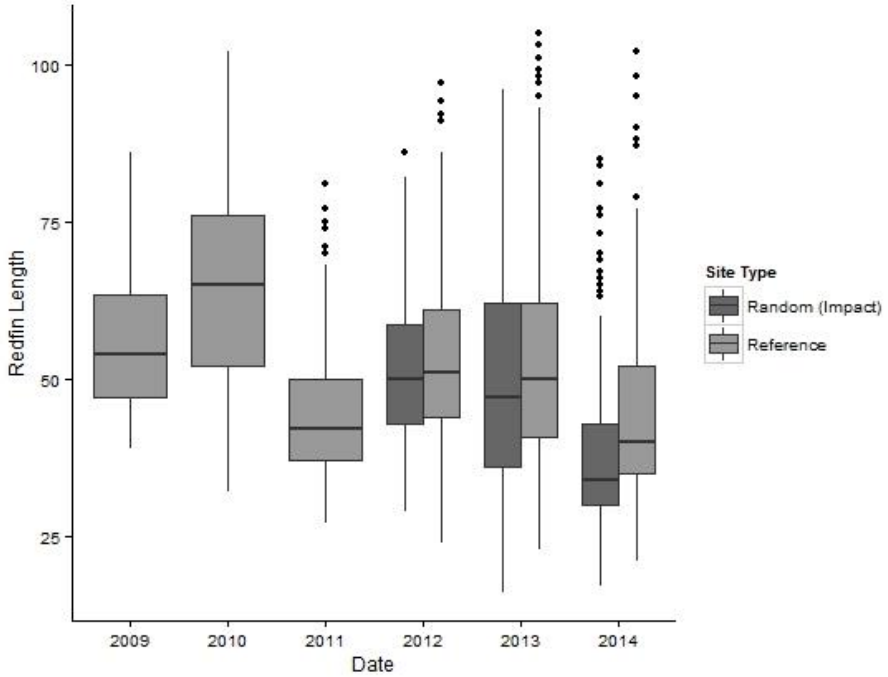


Figure 5.1 Redfin bully length during different years of sampling from 2009 – 2014 in the Waikato Regional Council (WRC) monitoring network. Length data is separated into reference and random sites. Outliers are shown as black dots and median values are centre blacks lines.

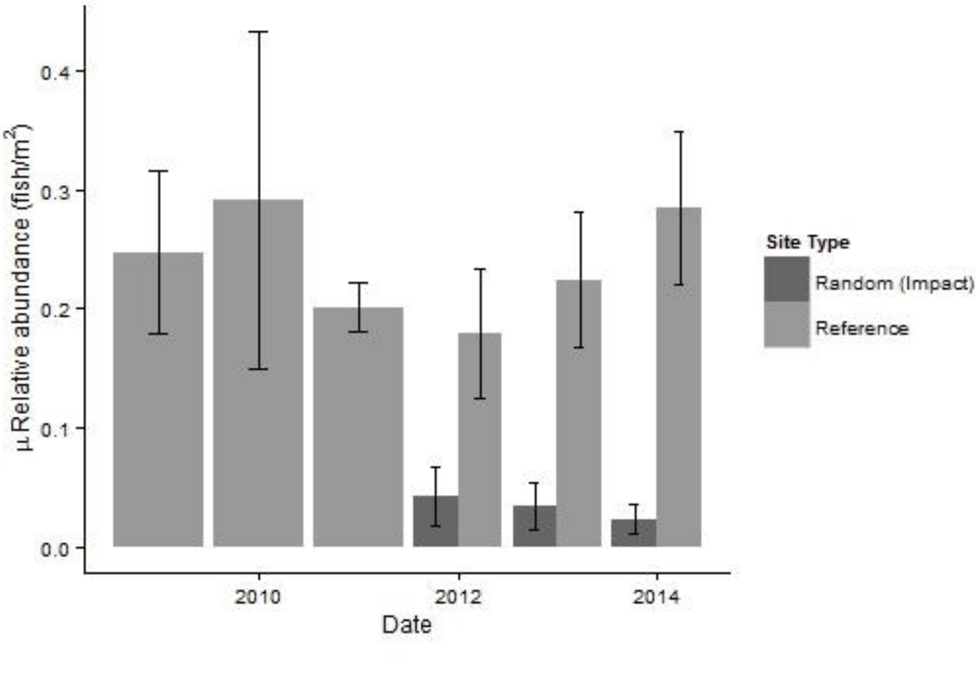


Figure 5.2 Redfin bully relative abundance during different years 2009 – 2014 in the WRC monitoring network. Sampling sites are divided into reference and random sites. Error bars are standard error of the mean.

Median length of redfin bullies analysed based on location (i.e. east vs west) for both populations was higher on the west coast than the east (Figures 5.3 and 5.4). The two plots represent analyses undertaken for all individuals (Figure 5.3) and for those above 50mm (Figure 5.4). Mean length at western sites is significantly higher than eastern sites for both all individuals (t_{3250} = -27.9427, p-value < 0.000) and those above 50mm (t_{1694} = -16.2271, p=<0.000).

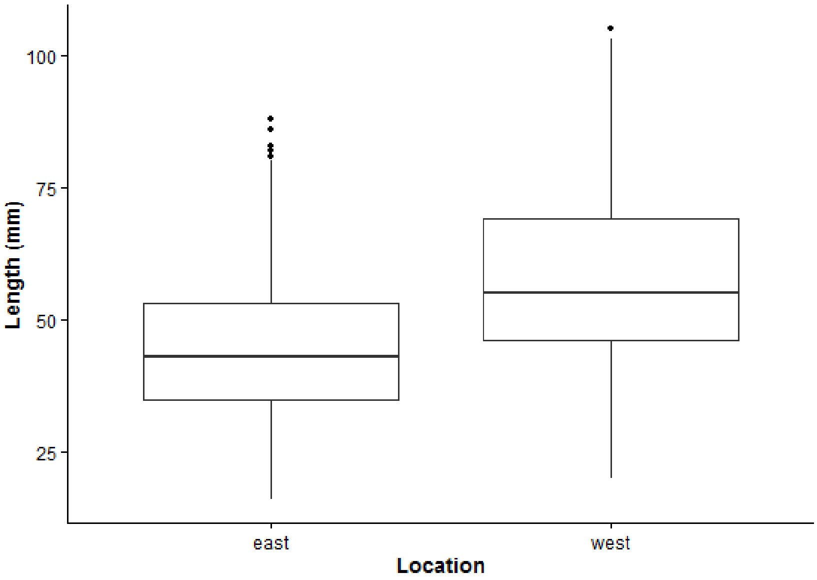


Figure 5.3 Length of redfin bullies from the WRC dataset grouped by geographical location in the Waikato region. Data in this analysis includes all individuals.

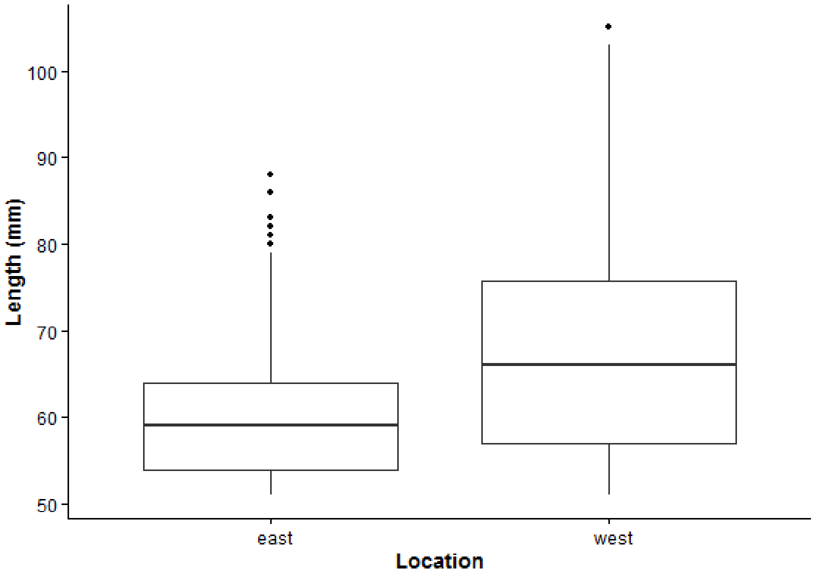


Figure 5.4 Length of redfin bullies from the WRC dataset in the Waikato region grouped by geographical location. This analysis excludes all individuals less than 50mm long.

Length of redfin bullies compared between western and eastern sites broken down by sample year show a similar pattern for 2013 with median length typically higher on the west coast than the east coast particularly for reference sites (Figure 5.5). When juveniles (<50mm) were removed some of these differences become more pronounced (e.g. 2012, 2014) (Figure 5.6).

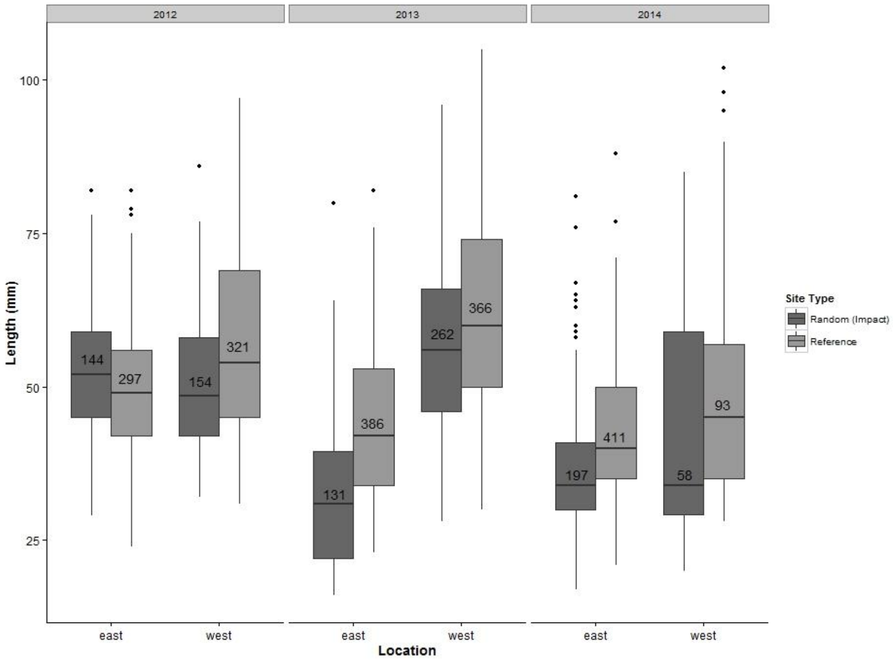


Figure 5.5 Length of redfin bullies in the Waikato as sampled by WRC, analysed by geographical location and year. This analysis includes all individuals from the WRC dataset. Numbers in boxes are N individuals for each group.

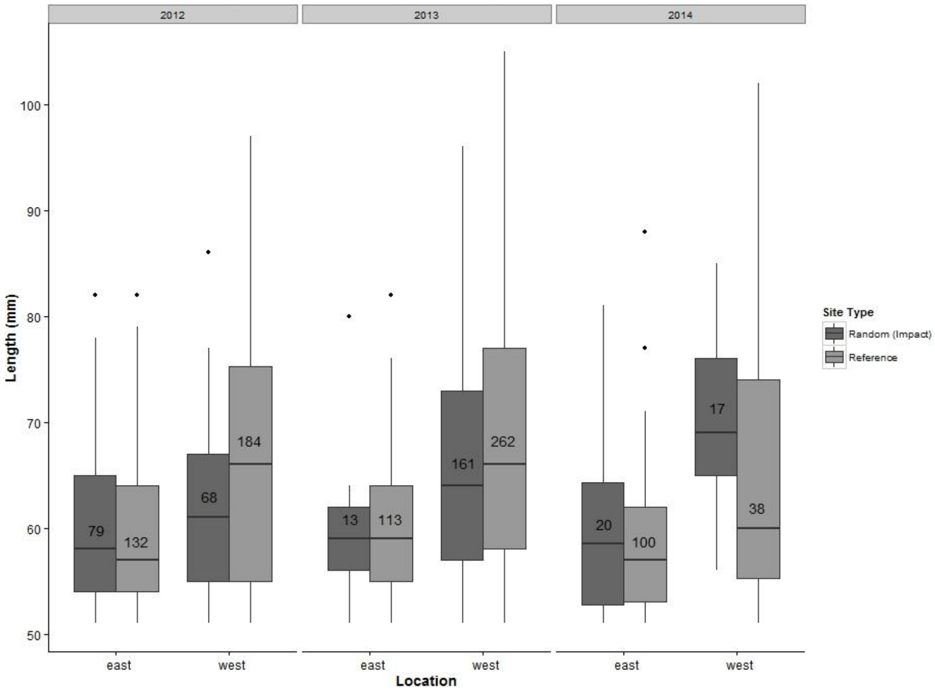


Figure 5.6 Length of redfin bullies >50mm long analysed by geographical location and year from the WRC dataset. Numbers in boxes are N individuals for each group.

WRC and NZFFDB dataset comparisons

The east to west length pattern is seen again in the NZFFDB database with mean maximum and mean minimum values both higher on the west of the North Island than the east (Figure 5.7). There is a difference between the South and North Island where both mean maximum and mean minimum values are higher within the eastern southern island than the western. There are fewer records with length measurements for redfin bullies on the east in the south island than all other groups (NorEas n = 421, NorWes n = 285, SouEas n = 24, SouWes n = 319) and this may be having some effect on the pattern seen, particularly as a large proportion (85%) of the total number of redfins were sampled at an isolated location south of Oamaru (Kakanui Stream). The number of individuals at each location somewhat reflects the same pattern with fewer individuals in the SouEas group (Figure 5.9). Although, mean maximum and mean minimum length (e.g. Figure 5.7) for the least sampled location (SouEas) are based on approximately 800 individuals from the NZFFDB data. In comparison with available NZFFDB metrics, WRC data was compared for both mean minimum and mean maximum length in the Waikato region (Figure 5.8). Patterns of these metrics matched NZFFDB data of mean maximum and mean minimum length.

It is important to note that mean minimum length of redfins is likely to be subject to some variation depending on the time of year as can be seen particularly on the east coast within WRC data for mean minimum length: standard error of the mean for mean minimum is much greater than mean maximum compared in both WRC and NZFFDB data. January samples would likely have a greater proportion of small individuals than would the rest of the year due to young of the year recruits (Hamer, 2007). Mean maximum size is less likely to be subject to within year variation. Additionally, the large percentage of individuals caught at an isolated location for SouEas individuals may be having an effect on mean calculations although in all instances the mean maximum and mean minimum values are used, and those used from the isolated location are within the interquartile ranges for the entire category.

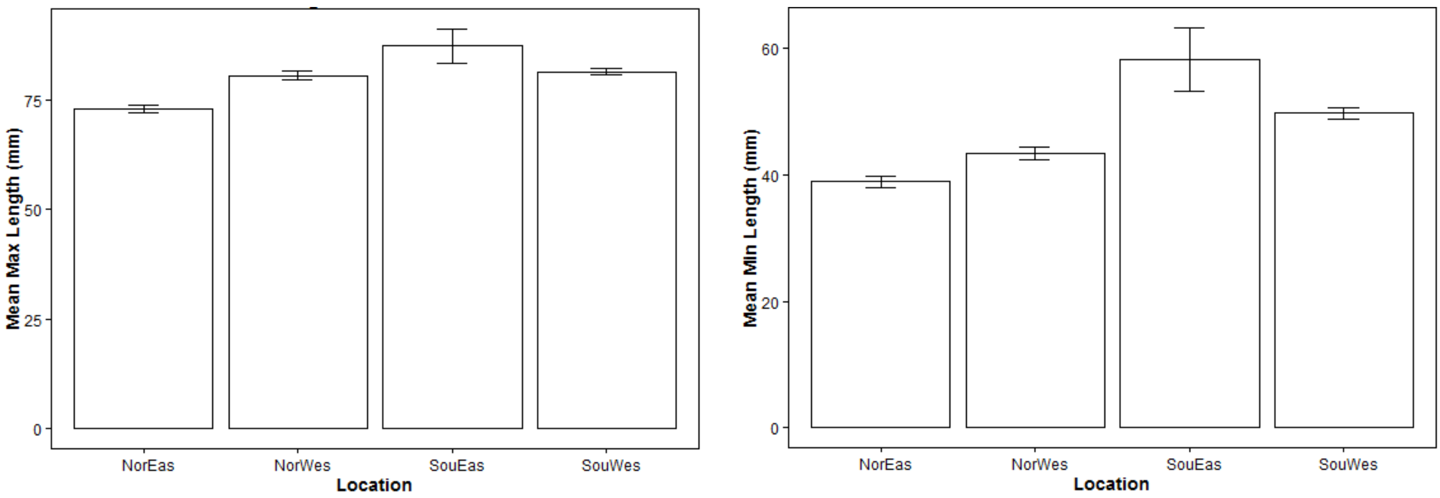


Figure 5.7 Mean maximum and mean minimum length by geographical location for redfin bullies in the NZFFDB dataset. Error bars are standard error of the mean. N values for the number of samples at a given geographical location are as follows: NorEas n = 421, NorWes n = 285, SouEas n = 24, SouWes n = 319.

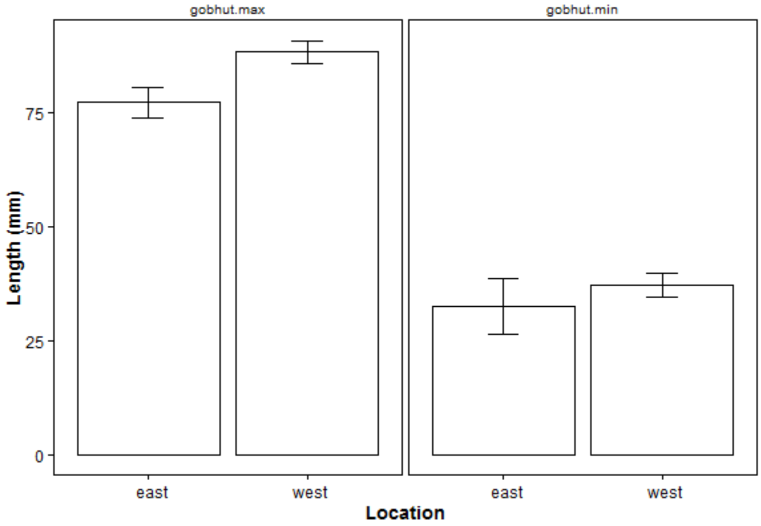


Figure 5.8 Mean maximum and mean minimum length by geographical location for WRC data redfin bullies. Error bars are standard error of the mean. N values for the number of individual fish in each grouping are as follows: N = 2174 (east), 2404 (west).

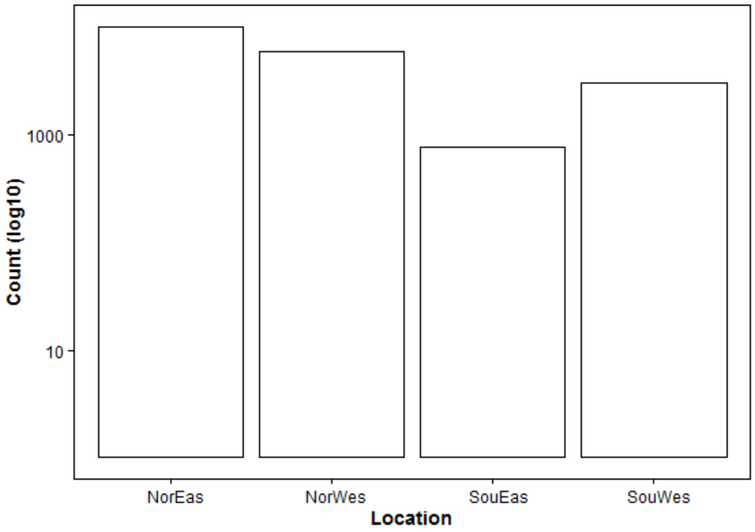


Figure 5.9 Count of individuals used in mean maximum and mean minimum length analysis by location for NZFFDB data.

Predictive modelling of redfin population demographics

The predictive model for abundance built using redfin bully individuals greater that 50mm performed better than the model including all individuals (Table 5.1; model simplification dropped 10 input variables). Model correlation and deviance were higher and lower respectively in the model built with only adult individuals compared with the model using all individuals (0.694 – cv correlation; 0.054 – cv deviance: Table 4.5, chapter 4). The top three drivers of abundance for the adult model were decreasing distance from sea, increasing segment indigenous forest cover, and increasing upstream indigenous forest cover (Table 5.2). These are the same drivers identified in the abundance model using all individuals (Chapter 4).

Table 5.1 Model outputs for relative abundance BRT model of redfin bullies using individuals >50mm. Both the initial model outputs using all River Environment Classification variables, and post simplification model outputs are stated. The simplified BRT model removed 10 variables [16].

Model	**Correlation**	**Deviance**
Gobiomorphus huttoni >50mm	0.626	0.035
Gobiomorphus huttoni >50mm – simplified (10)	0.736	0.033

---

[16] REC variables described in Appendix B

Table 5.2 Top explanatory variables for model predictions of redfin bully BRT model using individuals >50mm. Variables determined using fitted function analysis. +/- represents direction of variable effect with percentage value showing size of effect of variable in model processing.

Model Variable	**Effect Size and Direction**
Distance from sea	15.3% -
Segment indigenous forest cover	14.2% +
Upstream indigenous forest cover	9.2% +

Variables driving redfin mean length variation

The variables that best predicted redfin bully length were temperature and elevation (Table 5.3). In both models (all individuals; >50mm individuals) temperature variables were the top two explanatory variables for differences in mean length predictions (both followed by upstream catchment elevation). Cross validated correlation for simplified models using all individuals was 0.760 and deviance was 67.9. For the model including only individuals >50mm, correlation was 0.731, and deviance was 25.2: i.e., correlation values for both models were similar; however predictive deviance was lower in the >50mm model.

Table 5.3 Top 3 explanatory variables from BRT models for redfin bully mean length for both mean length models: all individuals and >50mm individuals from WRC dataset. Fitted function analysis showing direction of and effect size for model variables presented as -/+ and % respectively.

>50mm model	**All individuals model**
Segment historic winter equilibrium temperature - 43.1% -	Segment historic winter equilibrium temperature - 17.4% -
Upstream Average Cold temp - 21.6% -	Upstream Average Warm temp - 13.2% -
Upstream Catchment Elevation - 9.4% +	Upstream Catchment Elevation - 10.9% +

Mapped predictions

Modelled distributions of mean length in the Waikato region show a predominance of smaller individuals on the east coast particularly around the Coromandel peninsula (Figure 5.10). Areas of smaller individuals on the west coast appear around Raglan and Kawhia. There are a greater number of waterways on the east with smaller mean redfin length, particularly in the Firth of Thames close to Thames town and for the majority of the length of the eastern Coromandel peninsula. Mapped predictions from models built using only adult individuals sees these patterns change. The concentration of very small individuals around Thames disappears and the pattern of smaller individuals north of Raglan appears to strengthen. The apparent gradation of increasing lengths moving inland with intermediate sized individuals closer to the coast and larger individuals further inland, strengthens in the adult model (Figure 5.11).

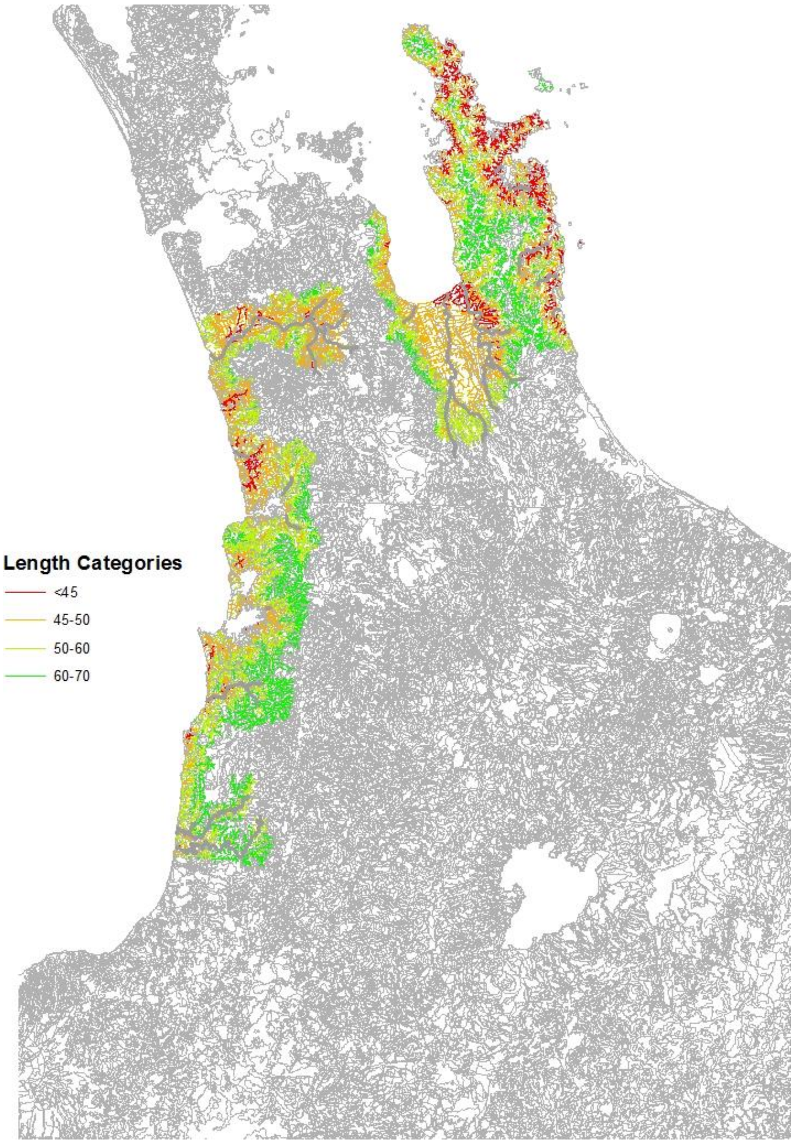


Figure 5.10 Mean length distribution of redfin bullies based on a boosted regression tree model built using all individuals from the WRC dataset. Green areas indicate larger mean length and red areas smaller mean length. Waterways mapped are those within 60km of the coast based on the strong distance limitation of redfin bullies described above and in detail in Chapter 4.

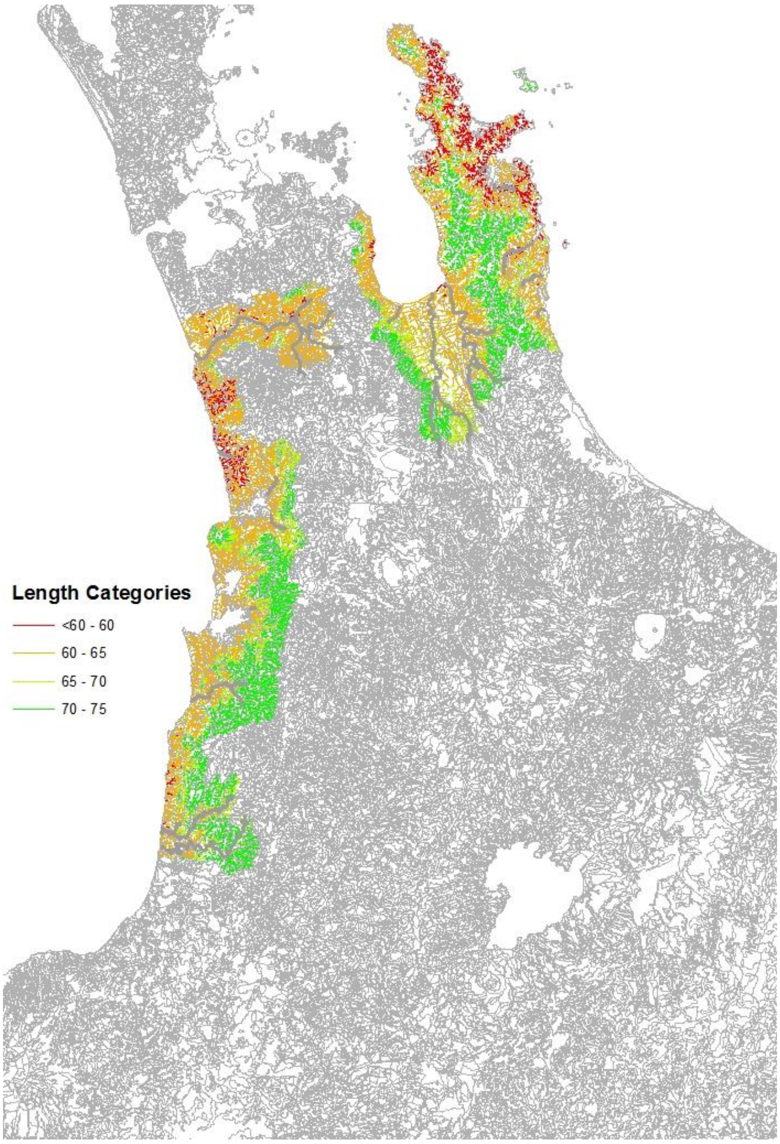


Figure 5.11 Mean length distribution of redfin bullies based on a BRT model built using only individuals >50mm from the WRC dataset. Green areas indicate larger mean length and red areas smaller mean length. Waterways mapped are those within 60km of the coast based on the strong distance limitation of redfin bullies described above and in detail in chapter 4.

Abundance predictions for the Waikato region based on the BRT model built from individuals greater than 50mm in length show a drop in total number of individuals predicted at sites (Figure 5.12). This is in comparison to Figure 4.9 (chapter 4) which shows abundance predictions based on the model built with all individuals. The highest abundance occurs on the Coromandel peninsula again with a decrease in numbers moving southward. This was also the case in mapped predictions using all individuals, however there were higher numbers predicted immediately south of Kawhia and Raglan and very few numbers predicted around the Firth of Thames. These points differ in Figure 5.12 where there are few numbers south of Kawhia and a relative increase in numbers around the Firth of Thames.

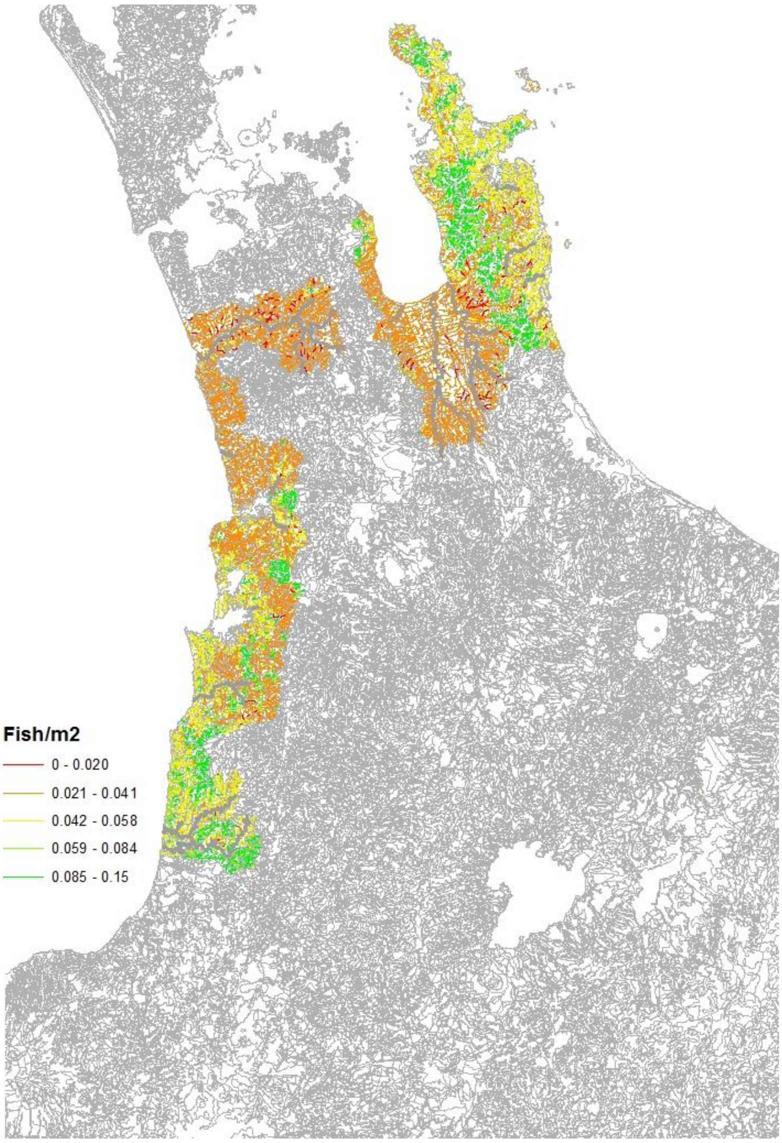


Figure 5.12 Relative abundance predictions based on the BRT model built with redfin bully individuals >50mm from the WRC dataset. Green areas show high relative abundance and red areas low relative abundance. Model performance for redfin bullies relative abundance using individuals >50mm was 0.736 (cross validated correlation) and predictive deviance 0.033 (cross validated deviance). This map is directly comparable with Figure 4.9 in chapter 4.

Discussion

There are significant differences in the size of redfin bullies between the east coast and the west coast of the Waikato region in the North Island. East coast populations are smaller than those on the west coast. This pattern was further observed in the NZFFDB data for the North Island, however the South Island appeared to reverse this pattern. In other New Zealand studies the growth of larval snapper populations has also been identified as faster on the west coast of the North Island compared with the east coast along with significant variation in fish length and habitat selection of juvenile snapper among harbours in the North Island (Crossland, 1981; Parsons et al., 2014). Although very little is known about larval biology of New Zealand freshwater fish species, there are a number of potential explanations for the differences in populations observed. Ocean currents could play a big role in the distribution of species in regards to both dispersal and nutrient distribution. For example, in their 2011 study regarding larval dispersal, Chiswell and Rickard identified 8 different coastal current systems surrounding the coast of New Zealand. A north eastern current (i.e., different from the east cape and south); a north west coastal current (Taranaki Bight/Auckland); a south island west coast current; and a south island east coast current were amongst those identified as different coastal current systems. Differences within these currents could provide some explanation to the difference in length distributions of redfin bullies in regards to larval dispersal, differences in recruitment, and/or the supply of nutrients although, little is known regarding their feeding and life history during this ocean going stage of their lives. For example, the current along the eastern coast of the South Island is relatively fast. It may be difficult for smaller individuals to be able to colonise the east coast waterways of the south island. Alternatively, this may indicate a greater upwelling of nutrients or those brought from sub-Antarctic currents providing a greater source of resources for productivity and ultimately growth of larvae in the waters directly adjacent to the south eastern coast (Chiswell, Bostock, Sutton, & Williams, 2015; Chiswell & Rickard, 2011; Waters, King, O'Loughlin, & Spencer, 2005). A similar pattern of larger fish on the east coast versus the west coast of the South Island was identified in sea perch (*Helicolenus percoides*) (Paul & Francis, 2002) and source populations for South Island east coast waterways of larval redfin bullies may be growing larger during ocean growing stages. Correcting for sampling bias and technique however would be important to accurately compare size distributions in the NZFFDB. Although this correction is not possible, (see description of NZFFDB) similar size distribution patterns were observed in the North Island for both NZFFDB and WRC datasets.

The interpretation of ocean currents can also be viewed in conjunction with variables used to build the boosted regression tree models. In both models (all individuals; >50mm individuals), temperature variables were the top two explanatory variables for differences in mean length predictions (both followed by upstream catchment elevation: Table 5.3). This suggests an interaction with temperature perhaps in terms of colonisation of waterways both close to the coast and as individuals move further inland. Different ocean currents around the coastlines have different temperatures and these temperatures will also differ throughout the year and during periods when larvae of redfin bullies are present (Chiswell et al., 2015; Cotroneo, Budillon, Fusco, & Spezie, 2013). Waterways within the country also have different temperatures and these different temperatures have been shown to affect life histories of New Zealand freshwater fish species (Bannon, 2006; Kearney, Jeffs, & Lee, 2008; August & Hicks, 2008). Within the waterways sampled in the WRC network where redfin bullies were found, median temperatures were almost 2 degrees lower on the west coast than the east coast (west = 15.7°C, east = 17.6°C). Differences of this size have been shown to affect a range of life history and growth traits in freshwater fish both in New Zealand and internationally (August & Hicks, 2008; Bannon, 2006; Elliot, 1994; Olsen, Tremblay, Clapcott, & Holmes, 2012).

The east coast of the South Island is influenced heavily by cold sub-Antarctic currents particularly in comparison to the west coast of the south island (Carter, McCave, & Williams, 2008). Fluctuations in temperature and coastal currents also differ greatly between the west and east coasts of the North Island. Larger individuals on the east coast of the south island may be a reflection of cold temperatures, whether it be an increase in upwelling of nutrients supplying a greater abundance of zooplankton and productivity in general, an adaptation to colder waterways through increased size (e.g. thermal inertia), or whether there is a more direct influence on growth of juvenile redfin bullies (Baltar, Stuck, Morales, & Currie, 2015; Bostock, Hayward, Neil, Sabaa, & Scott, 2015). Further, differences in larval growth in relation to temperatures and consequently distribution throughout New Zealand were observed in the sea urchin *Centrostephanus rodgersii* (Pecorino, Lamare, Barker, & Byrne, 2013).

The observed effect on larval growth of temperature highlights the possibility of a direct effect of temperature on the growth of larval redfin bullies during their ocean going life stage. Many studies have identified relationships between temperature and growth of larval species, in addition to migratory and feeding affects. Within New Zealand, studies of Gobiomorphus spp. and other freshwater species have identified many varying effects of temperature: migration differences in regards to thermal plumes with species avoiding high temperature waterways; optimal spawning, embryonic development and hatching temperatures; and swimming abilities at differing thermal optima (August & Hicks, 2008; Jellyman, Booker, & Watene, 2009; Richardson, Boubee, & West, 1994; Stancliff, Boubee, & Mitchell, 1989; Jellyman, 1977). The latter affect in particular has shown that, for post larval stage inanga (i.e. recruiting juveniles), swimming ability was highest at 9°C and decreased rapidly after 10°C (Bannon, 2006). Resistance to cold temperatures has also been shown to increase with size in some freshwater fish and perhaps smaller fish can only persist in warm temperatures, and therefore may be excluded from these regions where cold water dominates (Elliot, 1994). This would also potentially explain the areas around Thames and Raglan where inlets and harbours may have a higher average temperature allowing for the residence of smaller redfin bullies. Ultimately, there will likely be an optimal growth temperature for redfin bullies and exploring how varying combinations of temperature and nutrients optimises growth of redfins to varying degrees may better illustrate the role of temperature in any or all life stages of redfin bullies from spawning to juvenile recruitment. These energetics and developmental field studies, in relation to temperature are both logistically and resource demanding, particularly during larval periods. Nevertheless, national and international literature suggests that temperature is commonly a critical variable explaining a range of life history characteristics (e.g. behaviour, spatial and temporal distribution, and growth). The suggested interaction of redfin bully distribution with temperature as illustrated in the current data appears to be a fruitful area of future research for this species in New Zealand (Bannon, 2006; Elliot, 1994; George, Baldigo, Smith, McKeown, & Faulring, 2016; Parsons et al., 2014; Richardson, Boubee, & West, 1994; Rowe & Chisnall, 1996).

Reduced ocean currents and coastal conditions may also describe some of the size distributions seen in redfin bullies. Areas of smaller individuals aside from those on the coast for both Figures 5.10 and 5.11 are associated with relatively large inlets or sheltered areas (e.g. the Firth of Thames and Raglan). These may represent areas that are comparatively sheltered, creating a nursery effect and a source population of smaller individuals that can then move into the connecting waterways. This differs from the areas immediately associated with the coast which would offer comparatively less protection and perhaps therefore not have a source population of smaller individuals able to colonise the connecting waterways. This may also explain populations seen on the east coast, particularly the eastern Coromandel peninsula. Compared to the west coast area of the Waikato Region, coastal processes, currents and storm effects are different and subject to less intense weather processes which move predominantly from the west (Stanton, 1995; Stephens & Gorman, 2006; Sturman, McGowan, & Spronken-Smith, 1999). This may allow a greater proportion of juveniles to colonise the waterways immediately adjacent to the coast. Figure 5.8 shows greater standard error bars for lengths recorded on the east coast which may also reflect increased and variable juvenile recruitment on the east coast resulting in greater variation in redfin bully length. In these areas there is still a pattern of greater mean length further inland which may suggest larger individuals are able to migrate and remain further inland or recruitment into these waterways is lower enabling resident fish to grow larger. For example, in other small benthic New Zealand freshwater species, important intra-specific social interactions have been identified between larger (dominant) and small (sub-dominant) fish. Larger individuals can have significant effects on feeding of small fish, altering timing and duration of feeding. In addition to feeding, larger dominant fish have been identified as being better able to inhabit deeper slow-flowing areas of waterways through reduced risk of predation and increased growth in relation to resource availability (David, Closs, Crow, & Hansen, 2007; Hansen & Closs, 2005; Hansen & Closs, 2009; Whitehead, David, & Closs, 2002). Upstream areas that may exhibit preferable feeding habitat may be excluded to smaller individuals which are exposed to both intra-specific social competition and predatory effects. Therefore, near-coastal and/or harbour areas may ultimately be staging and growth locations where juveniles can safely (relatively) persist until of sufficient size to compete with conspecifics and better avoid predation. Alternatively, if east coast waterways exhibit extreme or less stable environments than the west coast, such that less sub-adult/adult habitat is available, recruiting individuals may be repeatedly thinned out during severe hydraulic conditions. This would alter the number and size of individuals that eventually make it into the core population, resulting in the lack of larger individuals seen in east coast areas in the Waikato region. As with many of these questions, further investigation is required to establish and subsequently measure with consistency, the causes for these differences in length. If these differences were related to local hydrological disturbance however, differences in length between populations of redfin bullies and potentially other benthic species may offer new ways to measure hydrological disturbance in New Zealand.

Finally, the difference seen in The Firth of Thames may be occurring due to the removal of smaller individuals during analysis making these sites relatively higher in abundance. That is, the number of small individuals at these sites is relatively few compared to the other waterways. When these individuals are removed, the relative abundance of larger individuals may become apparent. This is further supported when comparing the abundance maps (Figure 4.9 and Figure 5.12). By removing small individuals, relative abundance is higher inland rather than immediately on the coast for both areas of high redfin abundance (Coromandel Peninsula and south of Kawhia). Smaller recruiting individuals will be in areas closer to the coast and Figure 5.12 shows what is likely to be the core or resident population, being predominantly further inland. Again, this proximity of smaller individuals closer to the coast also highlights the potential influence differing coastal processes between the east and west coast are having on growth and distribution of recruiting redfin bullies.

Conclusion

This study has highlighted interesting differences in redfin bully size and abundance distributions throughout the Waikato region, and New Zealand in general. Ultimately, establishment of age and size relationships of redfin bullies must be undertaken in these populations to better confirm at what size and age redfin bullies begin to colonise waterways and to accurately determine what size constitutes an adult fish. Investigations in to oceanic life stages and ocean processes in areas of redfin bully populations is vital to better understand why these differences in size between the east and west coasts have arisen. Throughout model building processes, and highlighted in multiple national and international studies, is the effect temperature can have on a range of life history traits and behavioural characteristics in freshwater fish. It seems likely that temperature is having an important effect on the distribution and size of redfin bullies in New Zealand and further investigation into this interaction should prove productive. This study has both raised important questions and has begun to highlight valuable information about redfin bullies in New Zealand of relevance to management and conservation of freshwater fish.

References

Allibone, R., David, B., Hitchmough, R., Jellyman, D., Ling, N., Ravenscroft, P., & Waters, J. (2010). Conservation status of New Zealand freshwater fish, 2009. *New Zealand Journal of Marine and Freshwater Research, 44*(4), 271-287. doi: 10.1080/00288330.2010.514346

August, S. M., & Hicks, B. J. (2008). Water temperature and upstream migration of glass eels in New Zealand: implications of climate change. *Environmental Biology of Fishes, 81*(2), 195-205. doi: 10.1007/s10641-007-9191-z

Baltar, F., Stuck, E., Morales, S., & Currie, K. (2015). Bacterioplankton carbon cycling along the Subtropical Frontal Zone off New Zealand. *Progress in Oceanography, 135*, 168-175.

Bannon, H. J. (2006). *Effects of Water Quality Parameters on Prolonged Swimming Ability of Freshwater Fishes.* (Doctoral), University of Waikato, Hamilton.

Bostock, H. C., Hayward, B. W., Neil, H. L., Sabaa, A. T., & Scott, G. H. (2015). Changes in the position of the Subtropical Front south of New Zealand since the last glacial period. *Paleoceanography, 30*(7), 824-844.

Carter, L., McCave, I. N., & Williams, M. (2008). *Circulation and Water Masses of the Southern Ocean: a review.* Amsterdam: Elsevier.

Chiswell, S. M., Bostock, H. C., Sutton, P. J. H., & Williams, M. J. M. (2015). Physical oceanography of the deep seas around New Zealand: a review. *New Zealand Journal of Marine and Freshwater Research, 49*(2), 286-317.

Chiswell, S. M., & Rickard, G. J. (2011). Larval connectivity of harbours via ocean currents: A New Zealand study. *Continental Shelf Research, 31*(10), 1057-1074.

Closs, G. P., Hicks, A. S., & Jellyman, P. G. (2013). Life histories of closely related amphidromous and non-migratory fish species: a trade-off between egg size and fecundity. *Freshwater Biology, 58*(6), 1162-1177. doi: 10.1111/fwb.12116

Cotroneo, Y., Budillon, G., Fusco, G., & Spezie, G. (2013). Cold core eddies and fronts of the Antarctic Circumpolar Current south of New Zealand from in situ and satellite data. *Journal of Geophysical Research-Oceans, 118*(5), 2653-2666.

Crossland, J. (1981). *The Biology of the New Zealand Snapper* Wellington Ministry of Agriculture and Fisheries.

David, B. O., Closs, G. P., Crow, S. K., & Hansen, E. A. (2007). Is diel activity determined by social rank in a drift-feeding stream fish dominance hierarchy? *Animal Behaviour, 74*, 259-263. doi: 10.1016/j.anbehav.2006.08.015

Elliot, J. M. (1994). *Quantitative ecology and the brown trout*. USA: Oxford University Press.

George, S. D., Baldigo, B. P., Smith, M. J., McKeown, D. M., & Faulring, J. W. (2016). Variations in water temperature and implications for trout populations in the Upper Schoharie Creek and West Kill, New York, USA. *Journal of Freshwater Ecology, 31*(1), 93-108. doi: 10.1080/02705060.2015.1033769

Hamer, M. P. (2007). *The Freshwater Fish Spawning and Migration Calendar Report* (1105573). Hamilton East.

Hansen, E. A., & Closs, G. P. (2005). Diel activity and home range size in relation to food supply in a drift-feeding stream fish. *Behavioral Ecology, 16*(3), 640-648. doi: 10.1093/beheco/ari036

Hansen, E. A., & Closs, G. P. (2009). Long-term growth and movement in relation to food supply and social status in a stream fish. *Behavioral Ecology, 20*(3), 616-623. doi: 10.1093/beheco/arp039

Jellyman, D. J. (1977). Summer upstream migration of juvenile freshwater eels in New Zealand. *New Zealand Journal of Marine and Freshwater Research, 11*(1), 61-71.

Jellyman, D. J., Booker, D. J., & Watene, E. (2009). Recruitment of Anguilla spp. glass eels in the Waikato River, New Zealand. Evidence of declining migrations? *Journal of Fish Biology, 74*(9), 2014-2033.

Jowett, I. G., & Richardson, J. (1996). Distribution and abundance of freshwater fish in New Zealand rivers. *New Zealand Journal of Marine and Freshwater Research, 30*(2), 239-255.

Jowett, I. G., & Richardson, J. (2003). Fish communities in New Zealand rivers and their relationship to environmental variables. *New Zealand Journal of Marine and Freshwater Research, 37*(2), 347-366.

Joy, M. K., David, B., & Lake, M. (2013). *New Zealand Freshwater Fish Sampling Protocols*. Palmerston North, New Zealand: Massey University

Kearney, M., Jeffs, A., & Lee, P. (2008). Effects of salinity and temperature on the growth and survival of New Zealand shortfin, Anguilla australis, and longfin, A-dieffenbachii, glass eels. *Aquaculture Research, 39*(16), 1769-1777. doi: 10.1111/j.1365-2109.2008.02053.x

Leathwick, J. R., Elith, J., Chadderton, W. L., Rowe, D., & Hastie, T. (2008). Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *Journal of Biogeography, 35*(8), 1481-1497. doi: 10.1111/j.1365-2699.2008.01887.x

Maes, J., Stevens, M., & Ollevier, F. (2005). The composition and community structure of the ichthyofauna of the upper Scheldt estuary: synthesis of a 10-year data collection (1991-2001). *Journal of Applied Ichthyology, 21*(2), 86-93. doi: 10.1111/j.1439-0426.2004.00628.x

McDowall, R. M. (1990). *New Zealand freshwater fishes : a natural history and guide* (Rev. ed ed.). Wellington [N.Z.]: Heinemann Reed, MAF Publishing Group.

McDowall, R. M. (2006). Fish, fish habitats and fisheries in New Zealand. *Aquatic Ecosystem Health & Management, 9*(4), 391-405. doi: 10.1080/14634980601026352

McDowall, R. M. (2007). On amphidromy, a distinct form of diadromy in aquatic organisms. *Fish and Fisheries, 8*(1), 1-13. doi: 10.1111/j.1467-2979.2007.00232.x

McEwan, A. J., & Joy, M. K. (2014). Habitat use of redfin bullies (Gobiomorphus huttoni) in a small upland stream in Manawatu, New Zealand. *Environmental Biology of Fishes, 97*(2), 121-132. doi: 10.1007/s10641-013-0131-9

McQueen, S. (2013). *A photographic guide to freshwater fishes of New Zealand*. Auckland, New Zealand: New Holland Publishers.

NZFFDB. (2015). New Zealand Freshwater Fish Database. Retrieved 2015, from https://www.niwa.co.nz/our-services/online-services/freshwater-fish-database

Olsen, D. A., Tremblay, L., Clapcott, J., & Holmes, R. (2012). *Water Temperature Criteria for Native Aquatic Biota* (No. 2012/036).

Parsons, D. M., Sim-Smith, C. J., Cryer, M., Francis, M. P., Hartill, B., Jones, E. G., . . . Zeldis, J. (2014). Snapper (Chrysophrys auratus): a review of life history and key vulnerabilities in New Zealand. *New Zealand Journal of Marine and Freshwater Research, 48*(2), 256-283. doi: 10.1080/00288330.2014.892013

Paul, L. J., & Francis, M. P. (2002). *Estimates of age, growth, and mortality parameters of sea perch (Helicolenus percoides) off the east coast of the South Island, New Zealand* (SPE 2000/01). Hamilton, New Zealand: National Institute of Water and Atmospheric Research.

Pecorino, D., Lamare, M. D., Barker, M. F., & Byrne, M. (2013). How does embryonic and larval thermal tolerance contribute to the distribution of the sea urchin Centrostephanus rodgersii (Diadematidae) in New Zealand? *Journal of Experimental Marine Biology and Ecology, 445*, 120-128.

R Core Team. (2014). R: A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Richardson, J., Boubee, J. A. T., & West, D. W. (1994). Thermal tolerance and preference of some native New Zealand fresh-water fish. *New Zealand Journal of Marine and Freshwater Research, 28*(4), 399-407.

Rowe, D. K., & Chisnall, B. L. (1996). Size-related differences in diel feeding activity, prey selection and nocturnal migration strategy for the planktonic larvae of Gobiomorphus cotidianus in Lake Rotoiti (NI), New Zealand. *Archiv Fur Hydrobiologie, 135*(4), 485-497.

Stancliff A.G., Boubee J.A.T., & Mitchell C.P. (1989). *Cross-channel movement of small migratory fish and shrimps in the Waikato River near Huntly thermal power station. Prepared for Electricorp*.

Stanton, B. R. (1995). Sea-level variability on the west coast of New Zealand. *Journal of Physical Oceanography, 25*(6), 1265-1272. doi: 10.1175/1520-0485(1995)025<1265:slvotw>2.0.co;2

Stephens, S. A., & Gorman, R. M. (2006). Extreme wave predictions around New Zealand from hindcast data. *New Zealand Journal of Marine and Freshwater Research, 40*(3), 399-411.

Sturman, A. P., McGowan, H. A., & Spronken-Smith, R. A. (1999). Mesoscale and local climates in New Zealand. *Progress in Physical Geography, 23*(4), 611-635. doi: 10.1191/030913399677305474

Tonkin, J. D., Wright, L. A. H., & David, B. O. (2012). Mussel Spat Ropes Assist Redfin Bully Gobiomorphus huttoni Passage through Experimental Culverts with Velocity Barriers. *Water, 4*(3), 683-689.

Vanderpham, J. P., Nakagawa, S., & Closs, G. P. (2013a). Feeding ability of a fluvial habitat-specialist and habitat-generalist fish in turbulent and still conditions. *Ecology of Freshwater Fish, 22*(4), 596-606. doi: 10.1111/eff.12064

Vanderpham, J. P., Nakagawa, S., & Closs, G. P. (2013b). Habitat-related patterns in phenotypic variation in a New Zealand freshwater generalist fish, and comparisons with a closely related specialist. *Freshwater Biology, 58*(2), 396-408. doi: 10.1111/fwb.12067

Waters, J. M., King, T. M., O'Loughlin, P. M., & Spencer, H. G. (2005). Phylogeographical disjunction in abundant high-dispersal littoral gastropods. *Molecular Ecology, 14*(9), 2789-2802. doi: 10.1111/j.1365-294X.2005.02635.x

Whitehead, A. L., David, B. O., & Closs, G. P. (2002). Ontogenetic shift in nocturnal microhabitat selection by giant kokopu in a New Zealand stream. *Journal of Fish Biology, 61*(6), 1373-1385. doi: 10.1006/jfbi.2002.2147

Synthesis

Synthesis

A high level of consistency in sampling effort is achievable through the implementation of protocols in freshwater fish monitoring sampling design. Button time and area fished between years within sites was consistent in WRC monitoring with differences of less than 10 minutes between years achieved. The use of standardised protocols also provides important relative abundance and fish length metrics which provide far greater depth to population analysis. Data collected to consistent protocols produced better performing and more accurate models than New Zealand Freshwater Fish Database models. Models built from WRC data were “excellent” to “outstanding”, compared with “acceptable” models from NZFFDB datasets.

Throughout the WRC dataset, reference sites are much closer to the coast than are impact sites. Reference sites within the central Waikato lowland areas are effectively absent. Native richness and abundance IBI scores follow these patterns with higher scores with decreasing distance from the coast and in locations where indigenous forest and minimal habitat modification occur such as the Coromandel Peninsula and south of Raglan. These richness and IBI metrics reflect both modification in lowland areas and the diadromous life history of many native New Zealand species. Significantly more redfin bullies were found at these reference sites along with larger longfin eels, whereas significantly more shortfin eels were found at impact sites. The sensitivity of redfin bullies and longfin eels to pollution and their need for quality habitat is likely to be reflected in these significant differences, particularly in comparison to shortfin eels which appear to proliferating in modified waterways.

Drivers of abundance and distribution within BRT models again highlighted both characteristics of diadromy in native fish and habitat modification in the Waikato region. Decreasing distance from the sea, elevation, and slope were key drivers for many native species distributions and abundances along with indigenous forest cover and temperature. Shortfin eels are driven both in distribution and abundance by different factors and their lesser sensitivity to pollution. The predominance of small size class individuals, and their predicted presence and abundance distributions across the Waikato suggest shortfin populations may be expanding in distribution and recruitment may increase in response to further habitat modification and or loss.

Lengths of redfin bullies are significantly smaller on the east coast compared with the west coast of the Waikato region. This size discrepancy was also reflected to a lesser degree in NZFFDB data. This may be related to coastal process although factors of recruitment are likely to be playing a role. A predominance of smaller individuals in harbour or sheltered areas may suggest variable nursery or recruitment areas resulting in differences in source population size. In all cases, length differences are primarily driven by temperature variables suggesting differences in the physical environment of near coastal waterways is driving redfin bully population differences.

These are preliminary conclusions about both redfin bully and eel populations in the Waikato region. To further support these conclusions and develop testable hypotheses about the causes of these differences in populations, establishment of both the recruiting/adult redfin bully size differentiation is required, along with more investigation in to varying coastal processes at recruitment sites between the east and west coast of the Waikato region and perhaps New Zealand. Investigation into the physiological interaction of shortfin eels with reduced water quality in the New Zealand environment, in addition to further exploration of longfin/shortfin eel population interaction would better reveal causes of shortfin eel expansion and reveal how this expansion may relate to conservation of the longfin eel.

Finally, this study has highlighted the importance of implementing standardised protocols into freshwater fish monitoring programmes in New Zealand. To effectively conserve and manage our freshwater fish, having the ability to interrogate robust presence/absence data, and importantly length and relative abundance data is crucial to understanding and building a picture of fish populations. As our fish diversity and abundance continue to decline and habitat modification carries on, these accurate analyses and detailed understandings, afforded through sampling standardisation, will become increasingly important in providing some means to effectively conserve native fish populations.

Appendices

Appendix A: Supplementary site characteristics and codes for Waikato Regional Council impact sites and NZFFDB site names; descriptions of species codes from Chapter 3.

Table A-1 WRC sample network random (impact) site physical descriptions.

Name	**Site Code**	**Distance from Sea (km)**	**Altitude (m above sea level)**	**Catchment Indigenous Forest cover[17] (%)**
Akaterewa Stm	192	271.6	322	0.42
Awakino River	204	11.8	25	0
Bridal Creek	94	2	24	12
Coopers Creek	398	20.1	63	3
Hauoira Stm	81	161	137	48
Hauwai Stm	128	261	336	0.726
Homunga Stm	466	82.7	120	4
Huioteko Stm	311	106.3	170	69
Kaiwhio Stm	251	225.5	200	2
Kakahu Stm	113	143	95	64
Karapiro Stm	465	163	138	7
Karapiti Stm	176	284	313	4
Kihikihi Stm	232	109.2	321	1
Kiripaka Stream	317	123	99	0.917
Kiritehere Stm	24	5	140	35
Komakorau Stm	25	113	32	0
Kurere Stm	306	37	35	7
Maihiihi Stm	39	224	207	23
Maire Stream Unnamed	491	87.6	20.2	11
Mangaharakeke Stm Matangi	429	132	40	1
Mangahouhounui Stm Rangipo Prison	0	406	661	1
Mangakotukurua Stm	254	23.4	59	0

---

[17] River Environment Classification derived value

Mangakowhai Stream	344	118	157	9
Mangamaori Stm	105	113	40	20
Mangamauku Stm	61	156	104	58
Mangamawhitiwhiti Stream	340	390	397	0
Manganui River	332	19.8	54	64
Mangaohoi Stm	258	194	75	0
Mangaokewa Stream	327	258	410	16
Mangaonua Stm	157	145	62	19
Mangaorino Stm	485	210	54	1
Mangaorongo Stm Mahoenui	168	59.7	200	52
Mangaotama Stm	173	162	40	0
Mangaotama Stm Ohaupo	109	161	35	0.309
Mangapiko Stm Waiterimu	469	97.8	137	71
Mangapukatoa Stm	478	158	36	0
Mangaputa Stm	146	183.4	210	10
Mangarere Stm	436	37.3	285	1
Mangarewa Stm	15	210	339	0
Mangatea Stm	73	118	28	0
Mangatea Stream	9	121	190	87
Mangatete Stm	336	298.7	316	2
Mangati Stm Te Akau	190	14	30	0
Mangatia Stm	259	84.6	45	0.168
Mangatukituki Stm	103	218	142	0
Mangauika Stm	290	165.2	75	68
Mangauwhauwhi Stm	87	242	286	0
Mangawara Stm	437	127	62	40
Mangawhero Stm Cambridge	93	147	63	1
Mangawhero Stm Pirongia	2	186	60	0
Mangawhitikau Stm	244	229.2	178	12
Maramarua River	422	64	18	84

Matahana Stm	288	283	398	50
Mokauteure Stm	75	268	579	16
Mystery Creek Stm	29	140	52	2
Naike Stm	291	101.3	55	4
Oamaru Stm	286	196.6	120	48
Oamaru Stream	386	192	82	53
Ohote Stm	45	134	32	0
Okupata Stm	238	17	340	16
Omahu Stm Te Aroha	185	63.4	12	26
Onukutauira Stm	97	198	322	32
Opuatia Stm	474	97.6	59	2
Oraka Stm	65	164	179	3
Oraka Stm Trib	241	154.2	100	10
Otautora Stream Trib	353	187	215	63
Owera Stream	314	3.1	16	24
Owhatupuku Stm	477	203	50	0
Paraheka Stm Aria	12	61	38	0
Paraheka Stm Mangatutu	455	212.3	101	0
Paraunahi Stm	150	0.3	10	21
Piako River	473	86.8	48	0
Piakonui Stm	69	97	87	0
Pirorua Stm	110	20	245	27
Pirua Stm	156	376.6	662	92
Pitone Stream Fraser-Smith Rd	28	2	7	68
Pokaiwhenua Stm	127	217	367	0
Puniu River	279	240.2	500	55
Rahopakapaka Stm	464	283	383	0
Swampy Stream	363	253	476	7
Taiwawe Stream	394	3	25	0
Tarapatiki Stm	170	2.2	18	56

Tararu Stm	22	3	46	76
Taringapeka Stream	19	102	65	3
Tauraroa Stm	391	227.2	230	0
Te Hui Stm	271	213.6	315	30
Te Marama Stm	216	6	16	64
Timaru Stm	149	112.2	30	79
Toomeys Stm	483	45.5	175	55
Toreparu Stm	366	11.6	205	28
Tunaeke Stm	237	123.1	20	0
Unnamed Trib	56	94	229	0
Unnamed trib of Mangawhero Stream	365	206	120	8
Unnamed Tributary of Waikato R	320	279	320	0
Upper Waikato Stm	420	434	1045	0.156
Waiharakeke West Stm	301	68.3	20	0
Waihora Stm	219	375.1	458.6	0.46
Waihou River	305	123	43	0
Waikato River	302	32.6	101	2
Waikato River Trib	129	182	135	0
Waiohotu Stm	385	162	260	74
Waioraka Stm	273	196.3	239	0
Waipa Stm	17	197	185	0
Waipapa River Mangakino	95	227	321	56
Waipapa Stm Mokai	47	262	475	0
Waipapa StreamMere MereSpringhill	342	54.8	19	0
Wairere Stm	246	33.8	75	32
Waitanguru Stm	376	104.1	230	18
Waitetuna River	222	16.4	57	4
Waitoa River	261	109.9	70	1
Waitoa River Unnamed Trib	133	115	99	0
Waiwarawara Stm	406	5.2	78	56

Whakariawaka Stm	82	187	92	15
Whangarahi Stream	346	1.3	8	64
Wharekawa River	34	17.4	72	80
Whareroa Stm Aria	231	101.5	139	70
Whareroa Stm Taupo District	55	376	397	16
Wharewera Stm	72	31	220	46
Kaawa Stream	201_8	10	16.8	20
Mangaohae Stm	2064_1	31.6	176	53
Mangaoronga Stm	2067_3	202	46	14
Mangapapa Stm (Matamata)	433_13	115	159	6
Mangatawhiri River	459_21	71	102	59
Mangatutu Stm (Waikeria)	476_15	205	63	51
Mangawhero Stm (Matamata)	490_20	130.6	60	0.25
Mangawhero Stm (Matamata)	490_22	143	95	0.301
Mangoparo Stream	2093_1	105	154	0
Maukoro Canal Unnamed Trib @ NZR08704-438	2092_1	12.7	6.1	0.0367
Ohuka Stm	1405_2	5.1	36	17
Okupata Stream	635_4	9.4	45	27
Paraheka Stm (Aria)	723_4	68.6	55	12
Piakoiti Stm	751_10	89.4	57	6
Waikaretu Stm Trib	1404_14	13.6	96	4
Waikato River Trib	1132_92	78.5	18	0
Waikawau River	1134_2	2.7	15.1	77
Waimata Stm (Waihi)	1164_7	77	110	36
Waipa River	1191_61	220	95	46
Waitawheta River	1235_10	63.2	120	85
Waiwhero Stm (Waihou)	1262_18	65.2	20	0
Whangamarino River Trib	1294_8	54.4	19	0
Waiohipa Stream	360	1.4	15	26

Table A – 2 NZFFDB site ID description from Figure 2.4; Chapter 2.

Site Name	**Site ID**
Manganui River	40
Manganui River tributary	44
Mangaotama Stream	25
Mangatea Stream	54
Mataiterangi Stream	98
Maunganui Stream	27
Mokau River	78
Mokauiti Stream	10
Paraheka Stream	5
Stony Bay Creek	47
Stony Bay Creek tributary	20
Timaru Stream	29
Waikato River tributary	92
Wainui Stream	61
Waiwawa River tributary	79
Whareroa Stream	4

Table A-3 Species codes, common and Latin names from Figure 3.1; Chapter 3.

Species Code	Latin Name	Common Name
Gamaff	*Gambusia affinis*	Mosquito Fish
Gobhut	*Gobiomorphus huttoni*	redfin bully
Gobbas	*Gobiomorphus basilis*	Crans' bully
Angaus	*Anguilla australis*	shortfin Eel
Galmac	*Galaxias maculatus*	Inanga
Angdie	*Anguilla dieffenbachii*	longfin Eel
Gobcot	*Gobiomorphus cotidianus*	Common bully
Retret	*Retropinna retropinna*	Smelt
Gobbre	*Gobiomorphus breviceps*	Upland bully
Galfas	*Galaxias fasciatus*	Banded Kokopu
Oncmyk	*Oncorhynchus mykiss*	Rainbow Trout
Saltru	*Salmo trutta*	Brown Trout
Chefos	*Cheimarrichthys fosteri*	Torrent Fish
Ameneb	*Ameiurus nebulosus*	Brown Bullhead Catfish
Caraur	*Carassius auratus*	Goldfish
Galarg	*Galaxias argenteus*	Giant Kokopu
Gobuni	*Gobiomorphus unidentified*	N/A
Galbre	*Galaxias brevipinnis*	Koaro
Aldfos	*Aldrichetta forsteri*	Yellow Eyed Mullet
Geoaus	*Geotria australis*	Lamprey
Anguni	*Anguilla unidentified*	N/A
Gobhub	*Gobiomorphus hubbsi*	Bluegill bully
Galpos	*Galaxias postevicus*	Shortjaw Kokopu
Galuni	*Galaxias unidentified*	N/A
Cypcar	*Cyprinus carpio*	Common Carp
Scaery	*Scardinius erythrophthalmus*	Common Rudd
Gobgob	*Gobiomorphus gobioides*	Giant bully
Rhoret	*Rhombosolea retiaria*	Black Flounder

Appendix B: Description of code used in BRT model building processes; complete list of all network variables available from the River Environment Classification (REC) network dataset, and those used in model building in chapters 2, 4, and 5; NMDS dissimilarity matrix REC correlation values; species counts by year in WRC dataset.

Code Used for Boosted Regression Trees

Code to produce boosted regression trees was sourced from (Elith et al., 2008). The rules of thumb outlined in this paper were used for tree complexity, bag fractions, and learning rates. As a model starting point all trees were built using a bag fraction of 0.75, a learning rate of 0.001, and a tree complexity of 5. Predictions of species abundance and species absence/presence based on these built models were mapped across the Waikato region using REC data and waterway ID's downloaded from the Ministry for the Environment data service (https://data.mfe.govt.nz/). There are many variables available in REC datasets and not all are relevant to the modelling process. Obviously inappropriate variables were identified and excluded based on prior knowledge of the area (e.g. upstream glacial coverage not relevant in the Waikato region), while others were excluded due to a lack of variation (and irrelevance) in the variable, established during modelling. Tables B1 and B2 display all REC variables available in REC datasets and those used in building boosted regression trees. The REC user guide explains how each category is calculated in full detail (Snelder, Briggs, & Weatherhead, 2004)

Table B-1 All available REC variables[18].

Utility variables		
Description	**Variable Name**	**Units**
Network from node	Nzfnode	dimensionless
NZREACH	NZREACH	dimensionless
Network to node	Nztnode	dimensionless
Segment maximum elevation based on 30m DEM	segMaxElev_Grid	m
Segment minimum elevation based on 30m DEM	segMinElev_Grid	m
X coordinate of catchment centroid	usXcentroid	m
Y coordinate of catchment centroid	usYcentroid	m
Segment length	segLen	m

---

[18] Us = Upstream; Seg – Segment; Q – flow weighted

Euclidean length of segment	segEucLen	m
Stream order	SegOrder	dimensionless
Segment average elevation based on 30m DEM	segAveelev_Grid	m
Elevation of upstream end of segment (From REC)	segUpElev	m
Elevation of downstream end of segment (From REC)	segDownElev	m
Total catchment area	usArea	m^2

Downstream Variables		
Description	**Variable Name**	**Units**
Average slope of downstream network	dsAveSlope	ratio
Distance to coast from segment	dsDistToSea	m
Maximum slope of downstream segments	dsMaxSlope	ratio
Maximum of maximum downstream grid slope	dsMaxSlope_Grid	angle - degrees

Segment Variables 1 (Climate)		
Description	**Variable Name**	**Units**
Average within segment mean minimum June air temperature	segAveTCold	^{o}C*10
Average within segment mean January air temperature	segAveTWarm	^{o}C*10
Current summertime equilibrium temperature	segEquiTSum	^{o}C
Historic summertime equilibrium temperature	segEquiTSum_Hist	^{o}C
Current wintertime equilibrium temperature	segEquiTwin	^{o}C
Historic wintertime equilibrium temperature	segEquiTwin_Hist	^{o}C
Segment December solar radiation	segSolarRadSum	W/m^2
Segment June solar radiation	segSolarRadWin	W/m^2

Segment Variables 2 (Morphology)		
Description	**Variable Name**	**Units**
Maximum segment slope based on 30m DEM grid	segMaxSlope_Grid	angle
Segment sinuosity	segSinu	reachlen/euclen
Average segment slope	segSlope	ratio
Average Within Segment Slope based on 30m DEM grid	segSlope_Grid	angle
Estimate of historic segment land cover	segVeg_Hist	Dimensionless
Estimate of current segment shade	segShade	Dimensionless
Estimate of historic segment shade	segShade_Hist	Dimensionless

Segment Variables 3 (Land cover)		
Description	**Variable Name**	**Units**
% of riparian area in LCDB category (wetland)	segWetland	%
% of riparian area in LCDB category (urban)	segUrban	%
% of riparian area in LCDB category (tussock)	segTussock	%
% of riparian area in LCDB category (miscellaneous)	segMiscLandCover	%
% of riparian area in LCDB category (pastoral)	segPastoral	%
% of riparian area in LCDB category (scrub)	segScrub	%
% of riparian area in LCDB category (bare)	segBare	%
% of riparian area in LCDB category (exotic forest)	segExoticForest	%
% of riparian area in LCDB category (indigenous forest)	segIndigForest	%

Upstream Variables 1 (Climate/Flow)		
Description	**Variable Name**	**Units**
Coefficient of variation of annual catchment rainfall	usAnRainVar	mm
Runoff weighted coefficient of variation of annual catchment rainfall	usAnRainVar_Q	mm
Runoff weighted catchment average slope calculated for 30m DEM grid	usAveSlope_Q	angle - degrees
Mean minimum July air temperature	usAvTCold	^{o}C*10
Runoff weighted mean minimum July air temperature	usAvTCold_Q	^{o}C*10

Mean January air temperature	usAvTWarm	^{o}C*10
Runoff weighted Mean January air temperature	usAvTWarm_Q	^{o}C*10
Catchment rain days (greater than 10mm/month)	usRainDays10	mean # days/yr
Runoff weighted catchment rain days (greater than 10mm/month)	usRainDays10_Q	mean # days/yr
Catchment rain days (greater than 100mm/month)	usRainDays100	mean # days/yr
Runoff weighted catchment rain days (greater than 100mm/month)	usRainDays100_Q	mean # days/yr
Catchment rain days (greater than 15mm/month)	usRainDays15	mean # days/yr
Runoff weighted catchment rain days (greater than 15mm/month)	usRainDays15_Q	mean # days/yr
Catchment rain days (greater than 20mm/month)	usRainDays20	mean # days/yr
Runoff weighted catchment rain days (greater than 20mm/month)	usRainDays20_Q	mean # days/yr
Catchment rain days (greater than 200mm/month)	usRainDays200	mean # days/yr
Runoff weighted catchment rain days (greater than 200mm/month)	usRainDays200_Q	mean # days/yr
Catchment rain days (greater than 25mm/month)	usRainDays25	mean # days/yr
Runoff weighted catchment rain days (greater than 25mm/month)	usRainDays25_Q	mean # days/yr
Catchment rain days (greater than 50mm/month)	usRainDays50	mean # days/yr
Runoff weighted catchment rain days (greater than 50mm/month)	usRainDays50_Q	mean # days/yr
Total annual runoff volume	usFlow	mm*m^2/yr
Mean annual low flow	usLowFlow	l/s
Annual potential evapotranspiration of catchment	usPET	mm
Runoff weighted annual potential evapotranspiration of catchment	usPET_Q	mm
December catchment solar radiation	usSolarRadSum	W/m^2
Runoff weighted catchment December solar radiation	usSolarRadSum_Q	W/m^2
June catchment solar radiation	usSolarRadWin	W/m^2
Runoff weighted June catchment solar radiation	usSolarRadWin_Q	W/m^2

Upstream Variables 2 (Topography)		
Description	**Variable Name**	**Units**
Average slope of catchment calculated from 30m DEM grid	usAveSlope	angle - degree's
Average elevation in upstream catchment	usCatElev	m
Average elevation in upstream catchment flow weighted	usCatElev_Q	m
Lake index	usLake	Dimensionless
Proportion of catchment with slope >30° (steep)	usLowGrad	%/100
Proportion of catchment with slope <30° (not steep)	usSteep	%/100
% annual runoff volume from area of catchment with slope < 30°	usLowGrad_Q	%/100
% annual runoff volume from area of catchment with slope > 30°	usSteep_Q	%/100

Upstream Variables 3 (Geology)		
Description	**Variable Name**	**Units**
% of catchment in LRI category (alluvium)	usAlluvium	%/100
% of catchment annual runoff from LRI category (alluvium)	usAlluvium_Q	%/100
% of catchment in LRI category (glacial)	usGlacial	%/100
% of catchment annual runoff from LRI category (glacial)	usGlacial_Q	%/100
% of catchment in LRI category (peat)	usPeat	%/100
% of catchment annual runoff from LRI category (peat)	usPeat_Q	%/100
Catchment average of calcium	usCalc	Ordinal scale
Runoff weighted catchment average of calcium	usCalc_Q	Ordinal scale
Catchment average of hardness (induration)	usHard	Ordinal scale
Runoff weighted catchment average of hardness (induration)	usHard_Q	Ordinal scale
Catchment average of particle size	usParticalSize	Ordinal scale
Runoff weighted catchment average of particle size	usParticalSize_Q	Ordinal scale
Catchment average of phosphorous	usPhos	Ordinal scale

Runoff weighted catchment average of phosphorous	usPhos_Q	Ordinal scale

Upstream Variables 4 (Land Cover)		
Description	**Variable Name**	**Units**
% of catchment in LCDB category (bare ground)	usBare	%/100
% of annual runoff from LCDB category (bare)	usBare_Q	%/100
% of catchment in LCDB category (exotic forest)	usExoticForest	%/100
% of annual runoff from LCDB category (exotic forest)	usExoticForest_Q	%/100
% of catchment in LCDB category (indigenous forest)	usIndigForest	%/100
% of annual runoff from LCDB category (indigenous forest)	usIndigForest_Q	%/100
% of catchment in LCDB category (mangrove, riparian, willows, coastal sands)	usMangrove	%/100
% of annual runoff from LCDB category (mangrove, riparian, willows, coastal sands)	usMangrove_Q	%/100
% of catchment in LCDB category (other than category 1-9)	usMiscLandCover	%/100
% of annual runoff from LCDB category (other than category 1–9)	usMiscLandCover_Q	%/100
% of annual runoff from LCDB category (pastoral)	usPastoral_Q	%/100
% of catchment in LCDB category (pastoral)	usPastoral	%/100
% of catchment in LCDB category (scrub)	usScrub	%/100
% of annual runoff from LCDB category (scrub)	usScrub_Q	%/100
% of catchment in LCDB category (tussock)	usTussock	%/100
% of annual runoff from LCDB category (tussock)	usTussock_Q	%/100
% of catchment in LCDB category (urban)	usUrban	%/100
% of annual runoff from LCDB category (urban)	usUrban_Q	%/100
% of catchment in LCDB category (inland and coastal wetlands)	usWetland	%/100
% of annual runoff from LCDB category (wetlands)	usWetland_Q	%/100

Variables immediately excluded were those that weren't flow weighted (non-Q variables) for which both flow weighted and non- flow weighted values were available, and those which had no values throughout the waterways where original data were collected. Table A-2 shows variables that were included in model building and subsequent predictions.

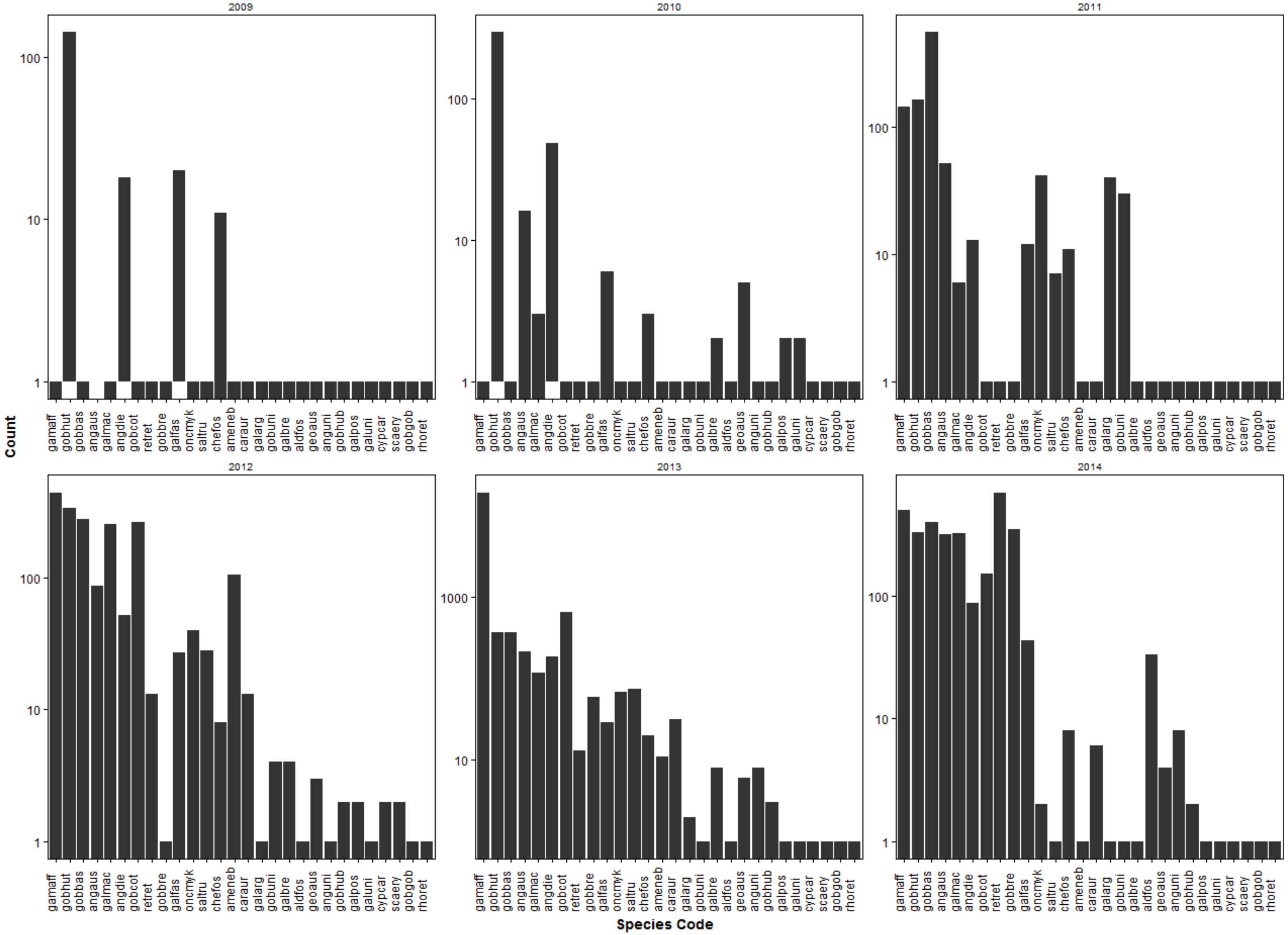
Table B-2 REC Variables used in Model Building and Predictions

Model REC Variables			
NZReach	segUrban	usLowFlow	usIndigForest_Q
segAveTCold	segPastoral	usPET_Q	usMiscLandCover_Q
segAveTWarm	segScrub	usSolarRadSum_Q	usPastoral_Q
segEquiTSum	segExoticForest	usSolarRadWin_Q	usScrub_Q
segEquiTSum_Hist	segIndigForest	usAveSlope	UpstreamVariables4_LandCover_NI_usSteep_Q
segEquiTWin	usAnRainVar_Q	usCatElev	usTussock_Q
segEquiTWin_Hist	usAveSlope_Q	usLake	usUrban_Q
segSolarRadSum	usAvTCold_Q	usLowGrad_Q	usWetland_Q
segSolarRadWin	usAvTWarm_Q	UpstreamVariables2_Topography_NI_usSteep_Q	segMaxElev
segMaxSlope_Grid	usRainDays10_Q	usAlluvium_Q	segMinElev
segSinu	usRainDays100_Q	usPeat_Q	segLen
segSlope	usRainDays15_Q	usCalc_Q	segEucLen
segSlope_Grid	usRainDays20_Q	usHard_Q	ORDER
segVeg_Hist	usRainDays200_Q	usParticleSize_Q	segAveelev
segShade	usRainDays25_Q	usPhos_Q	segUpElev
segShade_Hist	usRainDays50_Q	usBare_Q	segDownElev
segWetland	usFlow	usExoticForest_Q	usArea

Table B-3 Correlation co-efficients (r) for River Environment Classification (REC) variables of each network site plotted against respective NMDS dissimilarity matrix axes (1 and 2). Direction of relationship shown (+/-). Full list of variable descriptions is provided in appendix B.

Axis 1			**Axis 2**	
Variable	**r**		**Variable**	**r**
Winter equilibrium temperature	-0.638		Upstream average slope	-0.566
Altitude	0.628		Catchment average of hardness (induration)	-0.497
Historic winter equilibrium temperature	-0.607		Upstream catchment rain days >20mm/month	-0.489
Segment average elevation	0.600		Runoff weighted catchment average of particle size	-0.479
Segment minimum elevation	0.598		Upstream catchment rain days >15mm/month	-0.463
Segment downstream elevation	0.597		Upstream catchment rain days >25mm/month	-0.457
Segment maximum elevation	0.588		Upstream indigenous forest cover	-0.432
Segment upstream elevation	0.587		Segment indigenous forest cover	-0.426
Segment mean June temperature	-0.563		% annual runoff volume from area of catchment with slope < 30°	0.426
Segment mean Jan temperature	-0.540		Upstream steepness	-0.414

Figure B-1 Yearly total counts for each species in the WRC network from 2009-2014.



Appendix C: Power analysis of Department of Conservation freshwater fish dataset.

FRESHWATER TIER 1 EXPLORATORY ANALYSIS: FISH COMPONENT

- Prepared for

Department of Conservation

- June 2015

Quality Control Sheet

TITLE	FRESHWATER TIER 1 EXPLORATORY ANALYSIS: FISH COMPONENT
CLIENT	Department of Conservation
VERSION	Draft
ISSUE DATE	June 2015

DOCUMENT CONTRIBUTORS

Prepared by

SIGNATURE

Paul Boyce

Limitations:

This report has been prepared on the basis of information and data provided by the Department of Conservation. Analyses and interpretation have been completed on the data provided and undertaken assuming the data provided is correct and valid for the requested services.

Executive Summary

The Department of Conservation (DOC) is seeking to develop a monitoring programme aimed at assessing biodiversity on Public Conservation Land (PCL). DOC wishes to design a monitoring programme that will provide statistically valid information when detecting trends and changes in freshwater fish populations across the country. DOC commissioned the power analysis of current datasets to provide information relating to the design of future monitoring programmes in relation to Freshwater Environment New Zealand (FWENZ) class groupings and appropriate biodiversity metrics.

In order to assess the statistical validity of these datasets, power analyses were undertaken on current DOC held datasets in relation to PCL. Assessment of both structure and power of the PCL datasets were considered. Effect sizes and typical variation along with power values were attained for testing trends and changes in given metrics for FWENZ class groupings.

Power values were low and the current dataset is heavily skewed. Recommendations and suggested sampling strategies were developed in order to provide a more comprehensive dataset that would allow DOC to inform the design of future monitoring programmes in a more statistically valid manner.

Table of Contents

Table of Figures

Table of Tables

1.0 Introduction

DOC is seeking to investigate the capability of several datasets to be used in the creation of a national freshwater monitoring programme. Several indicators have been identified to inform monitoring programmes on the status of freshwater (Kelly, 2013). Three key ecosystems are being used to answer this question: river freshwater fish, river invertebrates, and inland wetlands. All three indicators and their associated datasets are being investigated individually in order to assess their suitability to determine the status of New Zealand's freshwater. Within this monitoring programme are questions about New Zealand's native fish, their status and potential trends observed on Public Conservation Land (PCL from here within). The following analysis and the associated report relates to freshwater fish.

In order to assess the suitability of the datasets available to DOC, statistical analyses were undertaken to establish how effectively the current data could answer questions about metrics of interest for fish populations. This is a crucial step when ensuring that future fish monitoring programmes are statistically robust. Natural variation in biological communities or individual species populations may be indistinguishable from sampling variation or from trends of interest (e.g. population declines). In order to design an effective monitoring programme then, it is vital to understand how sampling within the monitoring programme needs to be undertaken to collect statistically valid information that can detect relevant changes in biodiversity.

2.0 Scope of Work

The scope of works to be undertaken were outlined during preliminary discussions with DOC staff. The following paragraphs briefly outline the scope of works as discussed for this draft report.

2.1 Overview

The Department of Conservation (DOC from here within) has prepared a business case for the creation of a national freshwater monitoring programme based around the question: "what is the status and trend of native biodiversity in freshwater ecosystems on Public Conservation Land?" This report concerns components relating to fish. A component of the business case consists of statistical analyses to ensure that the design of the monitoring programme is statistically robust. In order to do this DOC wish to commission power analyses to be undertaken on datasets from various councils throughout the country.

2.2 The data

The following datasets have all been collected using the New Zealand Freshwater Fish Sampling Protocols - NZFWFSP (Joy, David, & Lake, 2013):

- Waikato Regional Council fish monitoring data
- Otago Regional Council fish monitoring data
- Greater Wellington Regional Council fish monitoring data
- Horizons Regional Council fish monitoring data
- Hawkes Bay Regional Council fish monitoring data
- Data from a recent DOC monitoring pilot in Northland

2.3 The tasks

The overall task has two parts:

Chapter 3 Part 1: assessment of the available datasets in terms of their ability to provide information adequate for us to design a statistically robust monitoring programme reporting on spatial and temporal trends in the nominated metrics for fish on PCL.

Rather than using geographic regions as spatial classifiers in the eventual monitoring program, we would like to be able to report on fish biodiversity in FWENZ classes (potentially stream orders). We would also like to be able to report on temporal trends (e.g. 5-10 years).

The results of part 1 will inform whether to proceed with either part 2A, part 2B, or a combination of both.

Table 4. Metrics which we are interested in reporting on for the future monitoring programme.

Measure	**Data collection protocol**	**Metrics**
1.2.3 Fisheries production (native fish community)	Joy et al. protocols (2013) **Scale applied:** 150m reach	1. Native species richness 2. An appropriate abundance metric 3. Other useful metrics
2.1.1 Occurrence of self-maintaining populations of new exotic fish species		1. Appropriate metric(s)
2.2.1 Distribution and abundance of exotic fish species		1. introduced taxa richness 2. nativeness

Chapter 4 Part 2A: If the available data is adequate for performing the analyses that we are after, then this component will consist of the performance of the analyses.

Chapter 5 Part 2B: If the available data is not adequate for performing the analyses that we are after, then this component will involve using research-based expert opinion to outline means by which we can design a statistically robust monitoring programme for the nominated metrics.

2.4 The outputs

A final report including:

1. A description of our assessment regarding the suitability of the available data for informing the design of future monitoring programme including areas where expert opinion may be necessary and why.
2. Recommendations for achieving robust statistical design of our future monitoring programme. This will involve outputs from power analyses where possible, expert opinion where not and may include recommendations for further data gathering if required.

3.0 Desktop Review

3.1 General

Power analysis is a crucial component of statistical testing. It provides a measure of the strength or power of data to provide statistically valid analyses and conclusions. In a broad sense, power testing provides a value representing a measure of the samples' ability to detect a specified difference or change: this difference or change is termed the *effect size*. Looking generally at the approach of statistical testing, there are different ways to interpret a power analysis. For example, the following questions could be asked in regards to statistical power: 1) "*how many samples do I need to collect, in order to be able to detect a given effect size, with a given level of power, testing at a given significance level?"*, or, 2) *"given the number of samples I have collected, what power do I have when detecting a given effect size, at a given significance level?"*. For completeness, the following could also be asked, *"what size effect can I detect, given the number of samples I have, and the amount of power and the level of significance I wish to test at?"* Power analyses can be undertaken to provide information for survey or experiment design, i.e. to provide a sample size required to meet the objectives of the study, or can be undertaken *post hoc* to inform on the power or statistical validity of an already collected dataset: the latter applying to this analysis.

Well established in statistical testing and study design is the importance of representative sampling: ensuring that enough, and representative samples are taken during a study to appropriately measure changes in a given metric for a particular grouping. In depth discussions around this aspect of sampling design are available in all standard statistical text books and further discussion is outside the scope of this report. However, in order to convey the analyses and broader concept of the following report, a discussion of effect size and power, and how these relate to hypothesis testing and statistical significance is warranted (section 3.2).

Finally, designing a survey able to detect and monitor the state of freshwater ecology requires selection of the strata for which changes will be monitored. There are several classification systems in New Zealand that could be used to group waterways. The River Environment Classification (REC) and Freshwater Environment New Zealand (FWENZ) classifications are two such grouping systems. There are also classification systems such as stream order which are perhaps wider recognised globally, if not between freshwater/hydrological disciplines. Within these classification systems various strata often exist that classify waterways into groups with increasing specificity.

As discussed in this document, the classification level chosen within a given classification system at which trends and changes in freshwater diversity will be monitored, will be a large determinant of statistical representativeness for the monitoring survey. While broader classifications will group more waterways together and provide larger sample sizes, the specificity with which any conclusions can be made will be reduced. As mentioned in Kelly, West, Robertson, Doehring, & Gansell (2013) there is a current lack of representativeness on PCL land for all existing monitoring networks. Consideration needs to be given then to the classification system used in the monitoring survey, the likelihood that this classification system will provide representative samples for river freshwater fish in New Zealand. For example, LevelOneAlpha groupings from the FWENZ classification system splits waterways in one of 20 types, nominally “A” to “T” (see Appendix D). In contrast, REC “Source-of-flow” classification would have 8 classifications ranging from “Glacial-Mountain” to “Regulated” source of flow. There are many different ways to classify waterways and each classification system may group a given waterway differently. Determining which classification system will be used then and maintaining consistency throughout monitoring, sampling and analyses is crucial.

Classification in this document is largely discussed in relation to LevelOneAlpha groupings upon following discussion with DOC staff.

3.2 Power, Hypothesis Testing, and Effect Size.

When conducting a study or analysis of existing data, often the ultimate aim is to detect the presence/absence of a change or difference in a specified metric, and to have some measure of the significance of that change or difference. Power of a statistical test then, “…is the probability that it will yield statistically significant results” (Cohen, 1988).

Hypothesis testing is the underlying principle to which the concept of power testing is applied. Often the question related to a given study is in the form of the null hypothesis; that the given metric being considered is equal to zero for the group or groups being tested. Determining there is a very small probability (usually <0.05) of falsely rejecting this null hypothesis, is deemed to be a significant result. The alternative hypothesis can take many forms where the given metric is not equal to zero (e.g. greater than, less than, greater or less than).

Significance values in relation to hypothesis testing, represent the likelihood of a *“type I”* error occurring: the probability of rejecting the null hypothesis when it is true (there is no effect). This is usually denoted as α. Therefore the commonly used significance level of $\alpha = 0.05$ is equivalent to saying there is a 5% possibility of committing a type I error. Conversely, the probability of accepting the null hypothesis, when it is false (i.e. there is an effect, and results say there is not), is termed a *“type II”* error and is symbolised as β. Statistical power = $1 - \beta$. To further illustrate the interaction between statistical power and significance, the following example provides some explanation. If a statistical test was undertaken with a 5% probability of committing a type I error, i.e., $\alpha = 0.05$, the *power* of this test will be less than if $\alpha = 0.1$. With a stricter level of significance required, there is a reduction in the power, or an increase in β error, i.e., the less risk of falsely detecting a given effect, the more risk there is in not detecting an effect when there is one. As sample size increases, the corresponding power values related to varying levels of significance testing converge (Figure. 1) and as such, careful consideration of the interaction between these two aspects of power testing is most important at smaller sample sizes (Osler, 2002).

This brief explanation outlines the need to balance these two aspects in the process of determining statistical power. The widely accepted significance value of $\alpha = 0.05$, may result in a meaningless level of power, and vice versa depending on the data (Cohen, 1988). Which of these errors is more important may have further influence on how these aspects are balanced (e.g., the probability of falsely detecting the presence of cancer, might be considered less consequential than falsely concluding there is *no* cancer, when there is). Generally in ecological studies, a beta value of 0.2 or 80% power is considered acceptable (Johnson, Barry, Ferguson, & Mueller, 2015). Therefore, in this case where DOC is hoping to design a monitoring survey for future work, establishing sample size estimates to satisfy and balance type I (α) and type II (β) criteria is the heart of the issue.

Determining a meaningful effect size is the final aspect of power testing. The effect size, is essentially the degree of change or difference that is being detected in the metric in question (Cohen, 1988). For example, detecting the expected difference in proportions of two samples – if the null hypothesis is that there is no difference, and the alternative hypothesis is that the difference is greater than zero, the effect size gives a measure of *how* different the two proportions are. In general - if the difference were 0.1, for example, P_1 = 2.5 – P_2 = 2.4, this would be a smaller effect size than P_1 = 2.5 – P_2 = 1.5 (this is often not a linear relationship however). Figure 1 illustrates the relationships between effect size, power, and sample size for a given significance level (α=0.05).

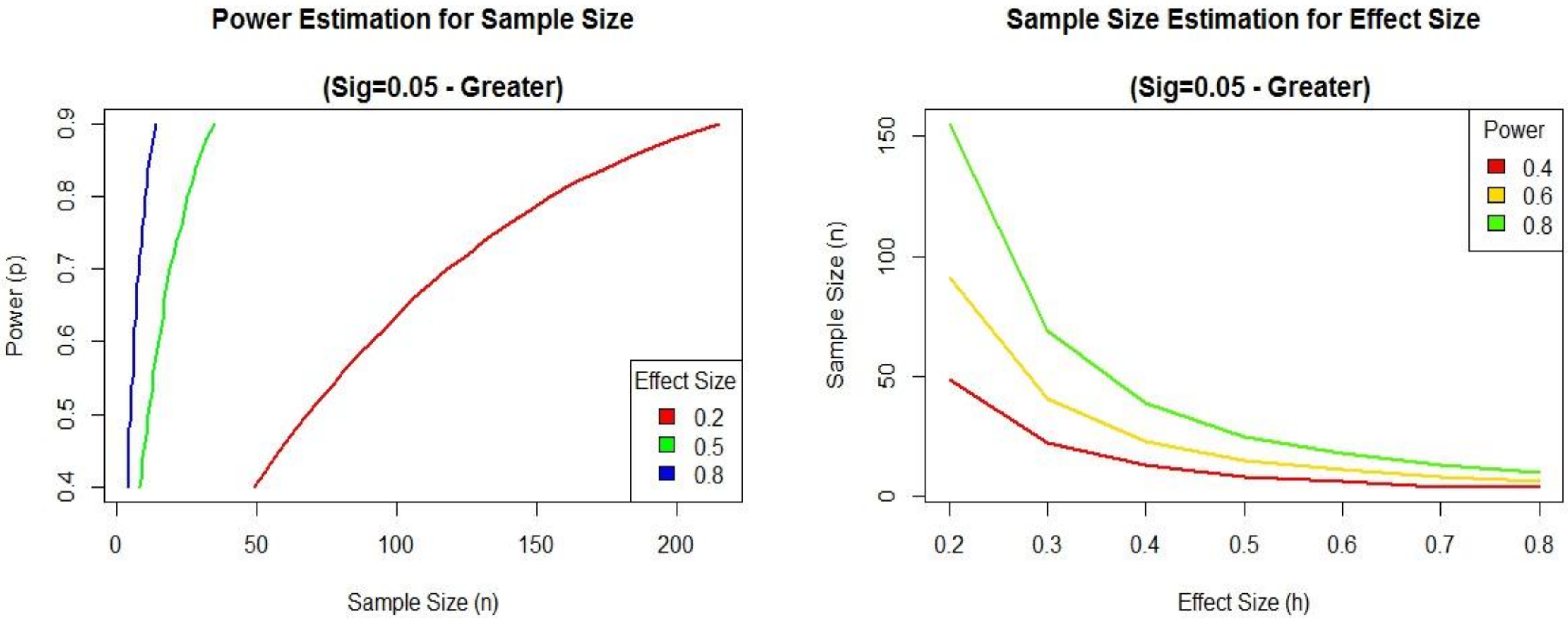


Figure 13: Representations of the relationship between key factors involved in power analyses.

3.3 Applicability to data

Given the characteristics of the dataset in question, the data will have varying abilities to detect varying effect sizes. Intuitively, a large effect size would be easier to detect than a small effect size, all other characteristics being equal. Of course, a large effect size it not necessarily meaningful or appropriate, and it is determining which effect size (usually a minimum; i.e., "which is the minimum effect size I can confidently predict with my data, over and above any "natural variation"?") is meaningful that is important in any power analysis (Emmerson & Raffaelli, 2000; Irvine, Dinger, & Sarr, 2011; Somerfield, Clarke, & Olsgard, 2002). In order to answer the underlying question directing these analyses in this particular case, an effect size must represent a larger change in a given metric, than would be seen with natural variation in populations.

4.0 Methodology and Statistical Analysis

All analyses were undertaken using Microsoft Excel and R statistical software. General power analysis and effect size concepts were initially taken from Cohen (1988) with ecological interpretation taken from the various references cited and from discussions with fellow ecologists. The LevelOneAlpha grouping used in most of the analyses describes waterways in ways such as "A-Lowland, low-gradient streams and rivers" or "G-Mid-elevation streams and rivers in dry inland areas". There are 20 groupings all together nominally "A"-"T". These are further broken down into a number of more detailed levels (see appendix D). As discussed throughout the document, discussions with DOC staff initially directed analyses towards these LevelOneAlpha groupings, and with analyses showing the paucity of sites at this level, analyses at more detailed levels (i.e., to group them into smaller samples) were deemed inappropriate.

Metric calculations for each site were as follows: native species richness was calculated as the count of native species; relative native species richness was calculated as the native species richness relative to the total species richness at the site; and relative abundance was calculated as the number of native individuals relative to the total number of individuals at a site. These metrics were chosen in an effort to provide simple measures of biodiversity following both review of literature on native and exotic species population dynamics, and discussions and review of proposed metrics to be included in monitoring programmes being developed by DOC, NIWA, and Regional Councils (e.g. (Braby & Somero, 2006; Leathwick, Elith, Francis, Hastie, & Taylor, 2006; Walsh et al., 2012), including others).

4.1 Determining Effect Sizes

Effect sizes for testing between mean species richness and proportional differences in relative native abundance and the proportion of native species to total species counts were calculated from principles outlined in Cohen (1988) and using the *pwr* package available in R.

Equation 1 Effect size formula for testing the difference between two means (e.g., species richness)

$$\frac{\mathbf{d} = \mu_1 - \mu_2}{\sigma}$$

Where **d** is the effect size, **μ** is mean, and **σ** is the pooled standard deviation (two tailed test: i.e. non-directional).

To estimate the effect size for the difference in two proportions an Arcsine Transformation of proportions is calculated:

Equation 2 Arcsin Transformation for calculating the effect size for the difference between two proportions.

$$\emptyset = 2\arcsin\sqrt{P}$$

Where $\emptyset$ is the arcsine transformation of the proportion P_i

From this equation, effect size h is calculated:

$h = \emptyset_1 - \emptyset_2$

The function "ES.h(P1, P2)" in R provides a means to calculate the effect size for two proportions.

In calculating the effect size for correlation coefficients, an arctanh transformation of correlation coefficients is performed:

Equation 3 Arctanh transformation for correlation co-efficient

$$z = \text{arctanh}(r) + r/2(n-1)$$

Again, the *pwr* package in R for determining power of correlation coefficients performs this transformation for given correlation coefficient.

Finally, although not used in R, the following equation applies to Table 4.4.1 in Cohen (1988) which relates to section 6.1 in this report where detecting differences between correlation coefficients is discussed:

Equation 4 Effect size for differences between correlation coefficients

$q = z_1 - z_2$

Where **q** is the effect size and **z** is the Fisher z transformation of r, i.e. $z = \text{arctanh } r$.

5.0 Analysis and Results

Full datasets of fishing events were provided as an initial look at the data available. Through further discussion, analyses were predominantly conducted on those sites on Public Conservation Land (PCL) for "LevelOneAlpha" stream orders. The following analyses and interpretation relate to those PCL sites.

A summary outlining the characteristics of the data provided for PCL sites follows. "LevelOneAlpha" and "LevelTwoAlpha" groupings of data: table 2. "LevelOneAlpha" groupings by council: table 3. Figures 2 and 3 also show the counts of each LevelOneAlpha class by region, and by year respectively for the current PCL dataset. Finally, for interest, the differences in native richness for LevelOneAlpha classifications between the regions is shown in figure 4.

Table 5. Number of PCL sites and their splits within the dataset.

Number of PCL Sites in Dataset	98
“LevelOneAlpha” Splits Within PCL Dataset	5: (A,C,D,G,H)
“LevelTwoAlpha” Splits Within PCL Dataset	17: (A1, A3, A4, C1, C10, C4, C5, C6, C7, C8, D1, D2, D4, G1, G2, H1, H6

Table 6. FWENZ LevelOneAlpha stream order groupings by council.

	A	**C**	**D**	**G**	**H**	**Total**
DOC	**2**	**9**				**11**
GWRC		**5**				**5**
HZNS	**1**	**28**		**5**	**3**	**37**
ORC		**4**	**9**	**7**		**20**
WRC	**1**	**23**		**1**		**25**
Total	**4**	**69**	**9**	**13**	**3**	**98**

Relating to the question raised in the description of analyses requested – “…how accurately can we report on “x” metrics at a range of sites?” – The following tables outline the statistical accuracy of reporting on given metrics for the “LevelOneAlpha” groupings described (table 4), and hypothetical statistical accuracy given particular sample and effect sizes being tested (table 5). Illustrations and tables relating to using different metrics on both PCL and NON- PCL are displayed in Appendix A. Stream orders “C” and “G” were chosen for the majority of analyses to be undertaken using an arbitrary sample size cut off of 10.

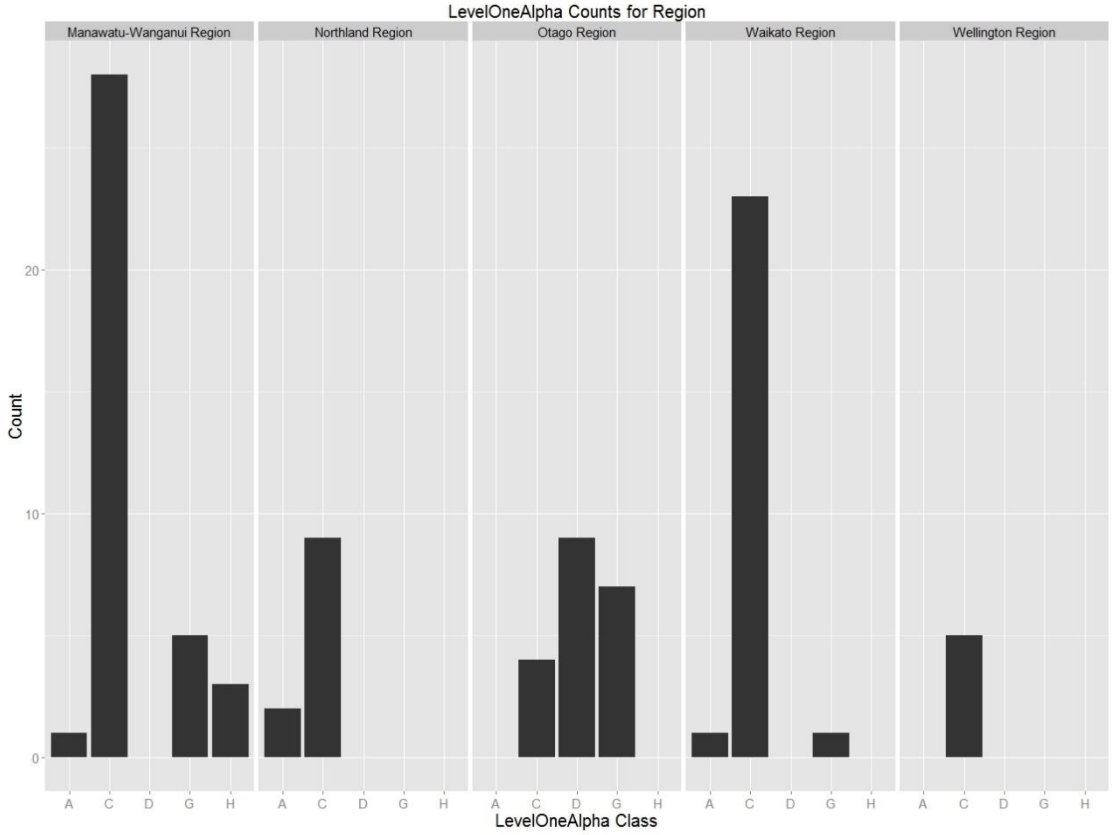


Figure 14: Total counts in PCL dataset for each LevelOneAlpha class waterway in the different regions.

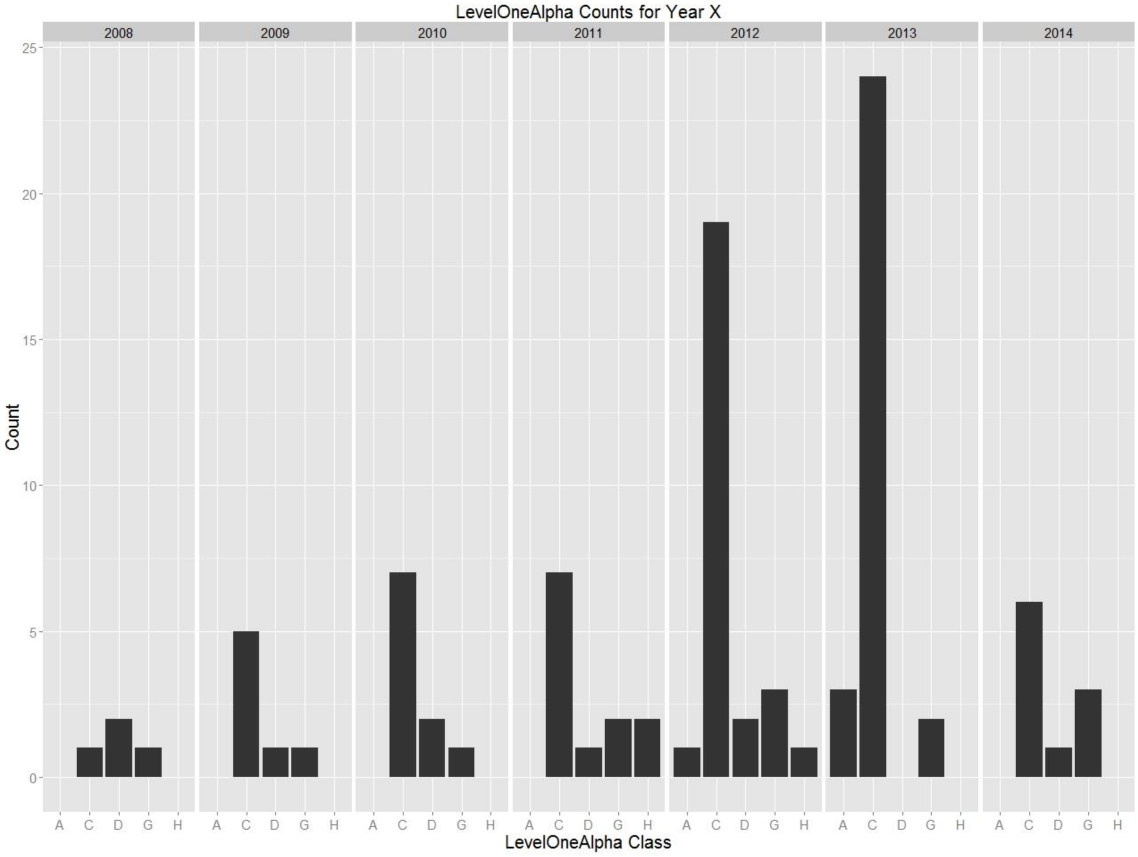


Figure 15: Total counts in PCL dataset for each LevelOneAlpha class waterway for the different years sampled.

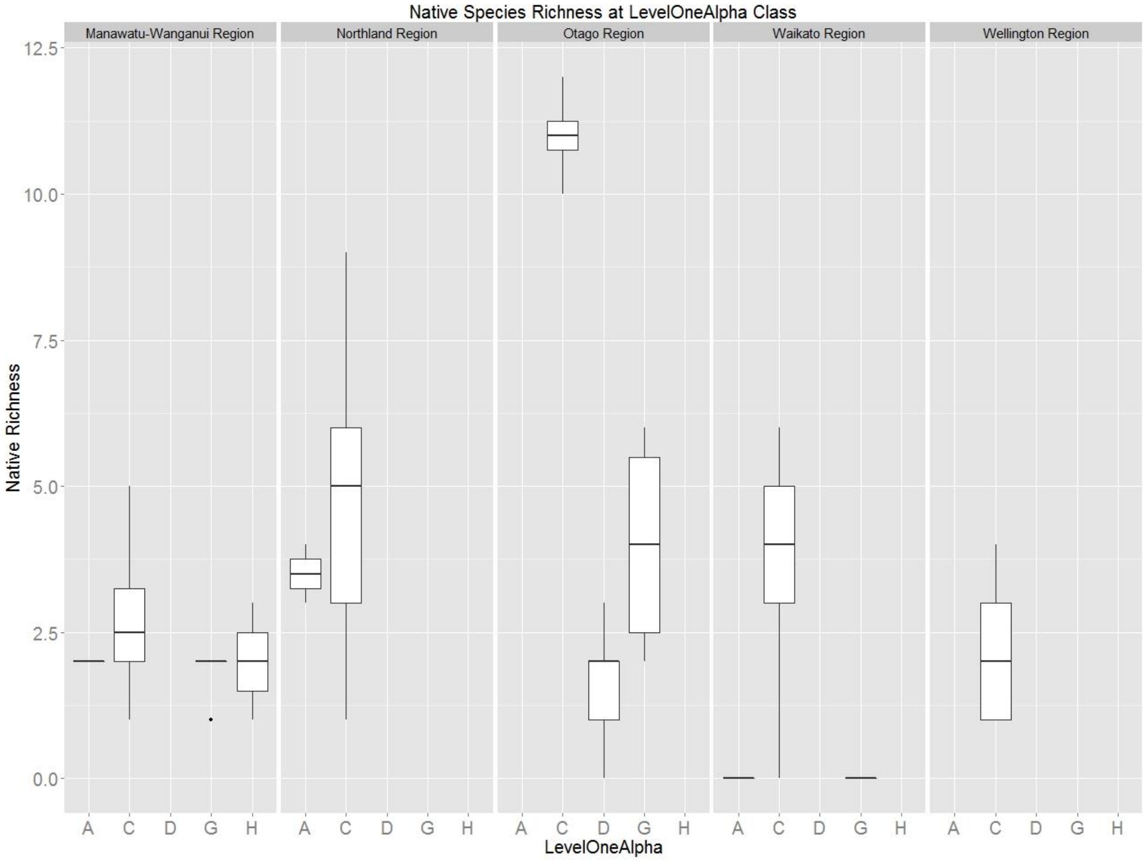


Figure 16: Native species richness in the PCL dataset for LevelOneAlpha classifications across regions.

Table 7. Power estimates for current PCL data tested at varying levels of significance (confidence).

Metric and application	**Effect Size**	**Significance level (Confidence) (α=0.05=95%)**	**Sample size/s (1n=150m reach)**	**Power (1-β)**	**Example**
Native species richness between (t-test) FWENZ groupings (“C” and “G”)	d = 0.42	95%	“C” = 69, “G” = 13	28%	Testing species richness differences between stream order classes
		90%	“C” = 69, “G” = 13	40%	
Proportion changes from monitoring start - finish relative native species richness within “C”.	h = 0.74	95%	2009 = 5, 2014 = 6	23%	Testing changes in the relative native species richness within a stream order class
		90%	2009 = 5, 2014 = 6	33%	
Trend in native species richness (annual mean) for “C” class stream orders	r = -0.57	95%	6	23%	Testing trends in relative native species richness over time
		90%	6	35%	

Table 8. Hypothetical power values for n=50 and n=100, tested at varying levels of significance (confidence) and sample size (all tested at "medium" conventional effect size (Cohen, 1988)).

Metric and application	**Statistical test**	**Significance level (Confidence)** **(α=0.05=95%)**	**Sample size (1n=150m reach)**	**Power (1-β)**	**Example**
Native species richness between FWENZ groupings (e.g. Stream Order)	t-test – difference in mean between sites	95%	50/100	70%/94%	Testing the difference between mean species richness for stream orders "C", and stream orders "D".
		90%	50/100	80%/97%	
Changes in relative native species richness between spatial groupings	Relative proportional change (arcsine transformation of proportion)	95%	50/100	71%/94%	Differences in relative native species richness between stream orders.
		90%	50/100	80%/97%	
Trends in species richness over time	Correlation coefficient (z transformation of correlation coefficient)	95%	5/10	8%/14%	Trends in species richness of a given stream order/class over time.
		90%	5/10	16%/22%	

5.1 Detecting spatial differences

5.1.1 Determining Effect Size

Effect sizes between metrics assessed in this power analysis were viewed in terms of – the maximum, median, mean values of differences between years at given sites and between grouped sites (e.g., "C" class). Specifically, inter-annual and spatial differences and their corresponding effect sizes, from individual sites, between individual sites, and from groupings of sites were assessed to establish typical ranges of effect size. There are a number of assumptions implied when basing natural changes off analyses like these such as the artificial groupings of streams which includes sites from different geographical locations (Leathwick, Elith, Francis, Hastie, & Taylor, 2006). However, based on differences identified in this analysis and consultation with the literature and fellow ecologists, the following effect sizes were determined to be reasonable as an estimate of variation exhibited on PCL sites with current datasets. This approach was favoured as the characteristics of the data in terms sample size, location, and areas of obvious skew, deemed testing for all between and grouped differences, or to take an average of differences inappropriate. Specifically, the following tables are demonstrating the sample sizes required to detect the effect sizes observed on PCL sites with the widely accepted power value of 0.8 (Johnson et al., 2015).

Table 6 shows ranges and effect sizes for spatial analysis.

Table 9. Typical ranges of differences and corresponding effect sizes observed from PCL sites (spatial).

	Species Richness	Relative Native Species Richness	Native Abundance
Range	1 - 2	0.1 - 0.2	0.1 - 0.2
Effect Size	0.4 - 0.8	0.1 - 0.3	0.1 - 0.4

Using the above data analysed from PCL sites, the following table (table 7) illustrates the required sample size to be able to detect changes in population dynamics of the same size observed on PCL sites but tested to a power of 0.8.

Table 10. Sample size required to detect effect sizes observed (table 6) between proportions within and between PCL sites (spatial) tested at power 0.8. Values are ranges relating to effect size ranges in table 6.

Confidence (sig.level)	Species Richness	Relative Native Species Richness	Native Abundance
0.05	100 - 25	1570 - 174	1570 - 98
0.1	78 - 20	1236 - 137	1236 - 77

5.2 Detecting temporal trends

Analyses were undertaken where possible to determine trends in FWENZ "LevelOneAlpha" grouped sites and within individual sites. This was done by assessing correlation coefficients by year for the metrics described. For "C" class sites on PCL land mean metrics were analysed as samples in the "C" class came from different geographical regions and sites within the "C" class were sampled varying numbers of times (Osler, 2002). Table 8 shows the typical correlation ranges observed and their corresponding effect size (r) values. Tables 9 and 10 show the corresponding sample sizes required at power 0.8 for given confidence levels.

Table 11. Typical correlations and corresponding effect sizes observed from PCL sites (temporal).

Effect	Species Richness	Relative Native Species Richness	Native Abundance
Individual site Effect Size	0.4	0.35	0.3
"C" class effect size	0.06	0.6	0.5

Table 12. Sample size required to detect typical trends observed within individual PCL sites (temporal). Tested at power = 0.8.

Confidence (sig.level)	Species Richness	Relative Native Species Richness	Native Abundance
0.05	45	60	84
0.1	36	48	66

Table 13. Sample size required to detect trends observed within "C" class PCL sites (temporal). Tested at power = 0.8.

Confidence (sig.level)	Species Richness	Relative Native Species Richness	Native Abundance
0.05	2177	18	28
0.1	1715	15	22

As can be seen with sample size estimates, seemingly minor changes in effect size produce large differences in n. This is due to non-linearity in the effect size r scale. This non-linearity of the r – scale can result in very large sample sizes required in order to satisfy a given power level, at a given significance level for increasingly small r values. This is also seen with small proportion effect sizes in table 7. Effect size in relation to proportions depends both how different two proportions are and where that difference lies (e.g. 0.9 and 0.8, versus 0.55 and 0.45). As such, appropriate interpretation and inference of effect size in relation to sampling strategy is crucial (Cohen, 1988).

6.0 Discussion

At the core of power analysis in this context of designing future monitoring programmes is establishing an appropriate sample size (Cohen, 1988; Johnson et al., 2015; Osler, 2002). The analyses above highlight this. As described in the introduction, multiple factors will affect the validity and the strength with which statistical conclusions can be made. Table 1 outlines the distribution of samples in relation to the FWENZ class splits. The majority of the sampled data lies within the class "C" (69 samples c.f. 13 in next closest - "G"). Although class "C" holds a relatively large number of samples, these are spread over a wide range of geographical sites (table 3) which carries its own assumptions. Therefore, despite having a seemingly large number of observations available, there are a number of inherent assumptions that must be accepted when conducting analyses within this class. For example, to say anything about class "C" streams in their entirety (not temporal), geographical variation and differences between years are not accounted for. Reasonable statements can still be made, for example, "Mean species richness in streams of order "C" has declined since monitoring began.", or "Class "C" streams have consistently had higher values than "G" class streams for relative native species richness values", or further still, "The proportion of native species in class "C" streams has declined significantly over 10 years of monitoring". All of these statements can be made and be statistically valid, with an understanding that stream order groupings are not natural, in terms of geographical differences (Irvine, Dinger, & Sarr, 2011).

Were the current data to be further divided, say into "LevelTwoAlpha" groupings, the ability to make statistically valid conclusions would be reduced as sample sizes for individual groupings are also reduced. As described above, given the dataset available for PCL sites (i.e., most classes having less than 10 sites) descriptive power analyses were undertaken largely on "C" class stream orders or by comparing "C" class streams to other classes. To reiterate the reasoning behind this, and tables 4 and 5 illustrate the power achieved by this grouping for given metrics and hypothetical power for increased sample sizes respectively. Furthermore, reporting on biodiversity using a classifier such as stream order or FWENZ class, means averaging the metric for each class and testing the number of sites required for each average metric: potentially losing isolated differences and trends at a geographical or regional scale. The paucity of regional coverage for FWENZ groupings highlighted in table 3, illustrates the likely issues when basing analyses of the current PCL dataset especially when talking about trends and differences at a national or even regional scale.

In an effort to provide meaningful inferences, analyses are discussed largely in terms of performance within class "C" groupings and within individual sites. In ecological systems, sample size needs to be able to detect differences of interest in populations or trends. As discussed above, what is defined as a meaningful difference is what is assessed by the ecologist to be meaningful in the context of the study (Cohen, 1988; Osler, 2002). This will vary depending on the practical application of the study outcomes and the constraints of the resources available. In a practical sense and in line with objectives of the future monitoring programme, a study should be designed to facilitate the detection of differences greater than natural variation that may be seen in the system. However, designing a survey that has the power to detect differences of a size that may be observed in "untreated" systems could be resource intensive. For example, if a sampling strategy was designed to detect differences of 1% in a population metric the number of samples needed to detect this difference would be large and probably prohibitive. This population/species/ecosystem may differ (inter-annually for example) by 5% "naturally", and sampling effort might be better directed at acquiring enough samples to detect this difference of 5% (of course a bigger difference: one that is more significant, is more easily detected).

"What is natural variation?" This is the next logical question. Intuitively, unaffected systems would be the best proxy in determining how populations naturally vary. In relation to sampling on PCL land, this concept may be applied by determining the difference seen between years for individual sites or classes of streams (or any given grouping) for PCL, assuming these systems are the least affected of the available data that can be used to make predictions (c.f., NON-PCL). This of course raises flags immediately as this is a rather substantial assumption and invasive species are present and anthropogenic influences are still realised at these sites. Were full datasets available of National Park population dynamics (or any given ecosystem where human influence was not a factor) for example, then this would likely be a better representation of expected natural population dynamics. Hence, care must be taken not to conclude too strongly that differences observed on PCL are reflective of complete community dynamics as even these sites will be affected to differing extents by anthropogenic factors. Also, the collected data from these sites may not encompass all population dynamics, and of course, however the available data were sampled may not capture changes that *can* occur. Nonetheless, this data *is* available and provides some idea of how populations are changing on PCL at some sites over a range of years. This information along with review of literature and prior ecological knowledge about what would be a change of interest given the objectives of the monitoring programme, can aid in determining sampling and data requirements to detect meaningful changes in biodiversity on PCL.

6.1 Sampling Strategy and Metrics

To reiterate the general concepts in terms of designing a future monitoring programme – power is dependent on sample size, effect size, variability in the response variable and the significance level being tested. When designing a survey programme, sample size is at the control of the tester, and in the same sense, the effect size, the significance level and power level criteria are also. Of course, variability in the response variable is the unknown. Designing a future monitoring survey then, is a matter of determining the sample size required to satisfy significance and power levels, for a given meaningful effect size (Cohen, 1988; Johnson et al., 2015).

Given the diversity and distribution of New Zealand fish populations, designing a sampling programme requires careful consideration when determining a meaningful effect. For example, an invasive species richness metric may not accurately reflect increases or decreases in the stream ecological health, given the varying distributions in invasive species. This may be especially important when sampling on PCL land. The distribution of species will differ greatly in lowland or enriched sites compared with those in comparatively uncontaminated PCL sites. Furthermore, substantial differences occur naturally such as native species richness declining as distance from sea and altitude increase (Joy & Death, 2001; Leathwick, Elith, Chadderton, Rowe, & Hastie, 2008; McEwan & Joy, 2009). Taking species richness as a metric then, lower native species richness naturally will alter the relative effect size that a survey might be expected to measure which, as discussed above, is intimately linked with the power of a given sample size to detect that difference. Also, species richness will of course miss the effects of relative abundance. For example, an upland stream may only one invasive species, with a number of native species. However, if the one invasive species dominate the proportion of fish in the stream, they are likely to reduce the population of natives through various effects, which will not be reflected in species richness values (McIntosh et al., 2010). Essentially, using species richness as metric can disproportionately weight any given individual organism under certain conditions. This has further flow on effects in relation to calculating an effect size for a given power and/or confidence level that is being tested to. A small number of individuals could make a relatively large difference when comparing samples over time or spatially with species richness (Emmerson & Raffaelli, 2000). Given the likely resource limitations, reporting on a metric which has consistency in terms of providing meaningful results is intuitively the better option.

Relative abundance (i.e. the number of native individuals as a proportion of total individuals observed at site) may be a more representative metric. Following the above example, this would provide a measure that would show an increase in the number of invasive species relative to the number of native species in a given observation. This may better highlight important changes in the population – e.g., "despite no change in species richness, there has been a drop of 30% in the number of native fish at site X". Deciding which metric is best used to analyse information collected is perhaps relegated to the post sampling efforts as the metrics (including species richness) detailed above are able to be calculated from the data routinely collected using the New Zealand Freshwater Fish Sampling Protocols. However, it is important to note that basing a future monitoring programme and sampling effort off effect sizes observed in a metric such as species richness could substantially affect the power of statistical analyses on other metrics. Changes in species richness, as discussed can be large with relatively small changes in numbers of fish. Basing the sampling effort then, in terms of sample sizes estimated in power analyses of species richness, may mean statistical power is low when assessing other metrics. Other metrics such as relative proportions for which effect sizes are likely to be smaller, requiring larger samples sizes, should therefore define the sampling effort. Defining precisely which metric or range of metrics that will be reported on will require engagement with experts prior to undertaking future analyses. Although this will unlikely alter the sampling strategy given a consistent approach in the NZFWFSP, this will ensure reporting following subsequent analyses is appropriately reflecting biodiversity, with the associated metrics being *able* to be reported on.

For temporal trends, tables 9 and 10 show how interpretation and judgement based on ecological knowledge is critical in order to design realistic sampling procedures (Cohen, 1988). For example, the r – value for the sole significant relationship seen on an individual PCL site is 0.96. The estimated sample size based on this r value for the same power and significance level is 4.5 (i.e. 4.5 years to test this relationship). Increasingly small r values and correlations are therefore not likely to provide meaningful information for designing a monitoring programme. As explained in section 4.0 this concept governed the reasoning behind effect size determination for most of the temporal trends (except "C" class).

Finally, examining the difference between two sites in terms of how they are trending is a likely analysis for future data. For example, which FWENZ class group of sites has declining fish populations? Which FWENZ grouping should we direct our management efforts towards in order to assess differences in trends of a size observed, to a power and significance desired? To highlight this - analysis assessing the power to test the largest difference observed between correlations detected in samples from PCL land returns a sample size of 11 (power = 0.8, q = 1.4, n = 11, significance level = 0.05). Explicitly, in order to detect the differences in temporal trends between PCL sites observed from current data, approximately 11 years of data for a given site would be required. Considering the sample sizes established from these analyses therefore provided a realistic interpretation of the power analysis in terms of recommendations for sample sizes. This and the above interpretation of the power analysis informed the following sections outlining how an appropriate dataset might be collected.

7.0 How appropriate data may be collected

In order to achieve the statistical power desired across the sites of interest, there are numerous ways that the required number of samples can be collected. Importantly, using consistent protocols, as has been advised is crucial to provide data with the comparability necessary to assess dynamics of populations. The metrics able to be tested using the NZFFSP, whether they be abundance, richness, or presence/absence can, in various ways be compared spatially, and temporally due to the fact they have been collected consistently. Consistency in sampling should be the ultimate target then, with PCL sites sampled the same number of times, the same way, for the same period prior to statistical analyses being undertaken (Jellyman, Chisnall, Sykes, & Bonnett, 2002).

There are clear biogeographical groupings in terms of species and community dynamics as has been documented (Jowett & Richardson, 1996, 2003; Leathwick et al., 2008; McEwan & Joy, 2009). It is recommended that sampling not be directed in terms of accounting for these differences and rather that consistent sampling techniques be applied across the survey areas. Predominantly, this will avoid introducing potential sampling biases and ending up with another skewed dataset. Using the NZFWFSP, *relative* abundances and proportions can be calculated from data collected and these provide comparable metrics spatially and temporally which will describe the population dynamics that the monitoring survey is intending to measure (Cohen, 1988; Osler, 2002).

The power achieved using the current data attained for PCL sites, testing at both a 90 and 95% confidence level is low (Tables 4 and 5). Therefore, attempting to design a monitoring survey off this data is inappropriate and data collection should be aimed at attaining a more balanced dataset. In order to achieve the sampling described, the appropriate approach may simply be to wait until samples of the appropriate dimensions are collected under the current monitoring programme or, alternatively, sampling may be directed at filling gaps in FWENZ class groupings and providing sample sizes of sufficient size for all spatial groupings. This of course depends on the ultimate objective of the monitoring programme (Oakley, Thomas, & Fancy, 2003). Relative abundance and proportional changes, provide a means to compare locations in like terms and to assess changes within a site on a consistent scale (c.f. simple species richness as discussed above). The assessment in this report provides direction about the scale of population fluctuations that might be expected to be seen and the level of sampling that would be required to detect these trends or differences between populations. As such, the approach of aiming to fill gaps in the spatial groupings described here, will provide a dataset with sufficient power to inform the design of a monitoring programme aiming to detect changes on PCL, in a statistically valid way. Filling the gaps in this manner however, may not be appropriate should the objective of the future monitoring programme differ from these spatial groupings.

7.1 Further considerations and applications

Trends and differences between years are likely to be more pronounced at individual sites rather than when sites are grouped together across geographical regions. As discussed, in order to assess a grouping such as

"C" class sites, this necessarily assumes sites are comparable or perhaps "groupable" in order to make comparisons, and individual site differences or trends will be diluted or lost in such groupings. Designing a monitoring survey aimed at sampling higher groupings (e.g. LevelOneAlpha vs individual sites) at a rate of 100 samples per group as power analyses suggest ("A" class = 100, "B" class = 100 … nth class) may not be appropriate to be able to test all lower levels of the hierarchy to sufficient detail. This concept is crucial in determining an effective sampling strategy and again relates to the exact objective of the programme. A suitable approach would likely be to establish the lowest denominator that a trend or comparison would want to be made at and to ensure that the sampling plan is sufficient to satisfy the statistical confidence sought. It would be optimal to establish a plan that could accurately detect differences at an individual site. Site characteristics are likely to be vastly different between sites and having the ability to decipher variation at individual level would provide some security when analysing trends in future datasets (Jellyman et al., 2002). The example below outlines this issue.

Were the objective of the monitoring programme be to monitor a particular LevelOneAlpha grouping then the approach would need to be directed at attaining sufficient data collected from this grouping. For example, "we would like to know the status of biodiversity in terms of LevelOneAlpha FWENZ classes in New Zealand". This could be realised in the following term: 100 sites from monitored regions per LevelOneAlpha: the above power analysis suggests this would be sufficient for all metrics and effect sizes observed. Aiming to monitor these same sites each year for a period of 10 years would also account for detecting expected temporal trends. This sampling effort could be divided amongst councils for which there are sampling programmes making the target more achievable. FWENZ class groupings could then be compared with sample sizes sufficient to detect a meaningful difference to the desired significance and power level. However, the same analyses could not then be applied to lower hierarchical levels. As such it is difficult to outline a specific monitoring programme without defined goals/objectives (i.e. precise goals for a given metric at a specific hierarchy level).

When confirming a temporal sample size, estimating from the post-hoc power analysis of PCL sites is problematic due to the fact that weak correlations and corresponding effect sizes are present in the sites for which correlation analysis is appropriate (i.e. those with multiple years = n>5). For example, in order to collect enough samples to test with a power of 80% for correlations observed for "C" class streams in regards to species richness would require over 2000 samples for a confidence level of 95% and over 1700 for a confidence of 90%. As discussed the relationship between effect size and sample size requirements when using the r scale is non-linear and n values greatly decrease when stronger correlations are detected. In PCL site ORC7240004 ("G") – with a correlation of r=0.96, n = 5 and 6 for 90% and 95% confidence respectively at 80% power. These n sizes decrease further when an established direction for hypothesis testing exists. To test the same parameters with a directional test (e.g. Ha = <0) n = 4.5 and 5.4 for 90% and 95% confidence respectively (differences in n are more meaningful in weaker correlations with larger n requirements).

Considering the above, a more appropriate way to look at the problem would be to establish sample sizes based on a range of r values that represent correlations of interest – i.e. to establish the number of sampling events required to detect meaningful biological relationships, should they be present (Irvine et al., 2011). To clarify, this differs from the approach for spatial analysis or within year differences at sites – where the observed variation can be assessed based on changes observed at PCL sites. Whereas, a strong negative or positive relationship in a given metric for a particular number of sites cannot be assessed if it is not present at a site. This is perhaps intuitive given PCL sites may represent more stable (in terms of non-significant correlations) populations (Allibone et al., 2010). This can be highlighted in the relationships that *were* significant in PCL data. For example, site ORC7240004 ("G") had a significant negative relationship in both native species richness and relative native species richness proportions over the 7 years of sampling. The *degree* of correlation at this site with this metric affords the analysis a high level of power (93% for native species richness; 98% for relative native species richness) despite a small sample size (7). However, in designing a sampling strategy, in order to detect weaker trends which are still ecologically important, a large sample size is necessary in order to maintain a reasonable power level with a given significance criterion. This is also illustrated in Cohen (1988): as sample size decreases, the correlation required to reject the null hypothesis (i.e. significance) increases. Or, in other words, in order to detect small, but still statistically significant trends, a large sample size is necessary. Considering this and the likely constraints – a realistic sampling size for temporal trends would be 10 years. Given this sample size, in order to detect a statistically significant result with this number of samples, for a given α of 0.05, and β of 0.8, a relationship would need to be greater than r = 0.55 (Table 3.3.2 (Cohen, 1988)). Analysis of PCL data suggest that PCL sites may typically exhibit r values of this size, and that trends of interest will be stronger and thus detectable at an appropriate level of statistical significance.

7.1.1 Settling on a sampling programme

To provide some definite description of proposed sampling, small and large effect sizes were chosen to illustrate a reasonable range of sample sizes for what could be expected. This differs from the typical ranges determined in tables 6 through 10 which were established to highlight the range of effect sizes seen in the current PCL data. The approach to use the small and large conventions from Cohen (1988) is in an effort to reflect what may be seen with a larger more consistent dataset. To suggest sample sizes for the range of effect sizes observed in the PCL was deemed inappropriate as this is unlikely to reflect the types of effect sizes that a complete dataset would produce due to substantial skew and scarcity of spatial and temporal coverage. Tables 11 and 12 show the power afforded by varying sample sizes.

Table 14. Power as a result of varying sample sizes using a conventional effect size range (spatial).

Metric	Effect Size Small Large	Power (%) for sample size N α =0.05(α=0.1)		
		N = 50	N = 75	N = 100
Species Richness	0.2	51 (63)	68(78)	80(88)
	0.8	98(99)	99(99)	99 (99)
Relative Native Species Richness Or Native Abundance	0.2	17(26)	23(34)	29(41)
	0.8	98(99)	99(99)	99(99)

Table 15. Power as a result of varying sample sizes using a conventional effect size range (temporal).

Metric	Effect Size Small Large	Power (%) for sample size N α =0.05(α=0.1)		
		N = 5	N = 7	N = 10
Temporal trends	0.5	14(24)	21(33)	32(45)
	0.9	62(74)	87(92)	98(99)

Power values for the small effect are low. As discussed in the section 5, the relationship between pairings of proportions and their corresponding effect sizes is not linear. There were effect sizes that were smaller than this “small” in the PCL dataset and it is probable that these small effect sizes are unlikely to be able to be detected with a reasonable monitoring programme. Furthermore, the large amount of variation in geographical distributions and natural population dynamics in native fish fauna make it difficult to group and assess biodiversity changes in populations in a comparable manner (Allibone et al., 2010; Jellyman et al., 2002; Leathwick et al., 2008). As the recommendations below suggest, it would be a pragmatic approach to collect a more consistent and comprehensive dataset covering these groupings to provide a more representative example of the likely spatial and temporal changes in natural populations on PCL than the current dataset suggests.

Temporal trends with much larger effect sizes than those suggested in Cohen (1988) were observed in the data – conventional correlation effect sizes as described and widely used in effect size interpretation, are determined in relation to behavioural science and are noted for being characteristically small. Considering this, small and large effect sizes for correlations in relation fish biodiversity trends were established based on interpretation of trends observed on PCL data with the current dataset as is advised to provide a more meaningful analysis (Cohen, 1988). The small value was taken from the analysis results above suggesting that and effect size would need to be greater than this value to provide a significant relationship given the likely number of samples taken. The large effect size was based on estimation of the range of larger correlations observed from current data. Working on the assumption that a given waterway would only be sampled once a year, the ability to detect trends of significance would require 10 years of sampling. Many trends of importance would be detectable at earlier stages (e.g. ORC7240004 ("G")), however a target of sampling for 10 years would afford the study programme the ability to detect trends that are significant as determined from current PCL data.

Finally then, to advise on the approach for future sampling based on the above information, sampling effort should be directed towards attaining a sample size of approximate 100 sites for a given spatial grouping, sampling for a period of 10 years. This would provide a dataset that would be able to assess most likely trends and changes in fish populations to a power and significance level that is appropriate. . To reiterate, the most conservative approach then would be to establish a dataset that had sample sizes consistent across these groupings and to then determine the statistical performance (e.g. n=100 for classes “A”, “B”, “C” …”*ith”*). This should not be difficult to establish for a nationwide dataset given council monitoring teams such as Waikato regional Council monitor at least 80 sites per year alone. Having a more balanced dataset, and then performing power analyses as has been done on the current dataset would provide a much better idea about how such sampling is performing in monitoring biodiversity at national and regional scales.

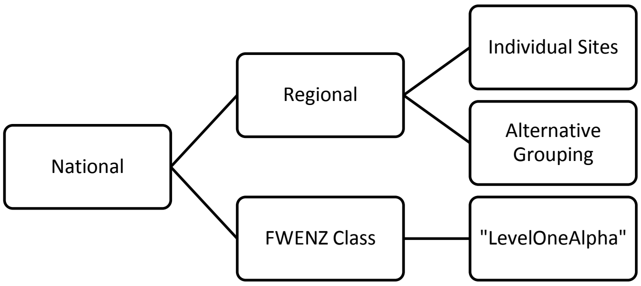
As described above it is crucial to define how biodiversity information is intended to be reported on in order to direct sampling. If FWENZ groupings are to be reported on nationwide then it must be appreciated that the same power for a given analysis and its criteria will not be achieved say, by region (e.g., Sample N for FWENZ grouping x, is made up of n samples from regions). In terms of differences that exist in the various ecosystems being sampled (wetlands, rivers, streams etc.) it is noted that again this is a vital component in determining what is ultimately to be reported on and which metric will reflect biodiversity in these ecosystems. Using metrics such as relative abundance and proportional changes makes temporal and between sites comparisons possible and meaningful.

8.0 Recommendations

The power analyses undertaken suggest that for the trends and effects seen currently on PCL, a sampling regime targeted to establish a dataset with consistent n (approximately 100 sites per grouping, monitored for 10 years). Some sites have been sampled close to this extent already and perhaps data exist further back at least for presence/absence data if not for abundance. An ultimate goal of ten years of data provides the means to detect the smallest of trends currently observed within PCL sites. Of course this does not exclude the possibility of analyses being undertaken each year with larger trends likely to be detected easily. In terms of the spatial aspects of sampling, attaining a dataset for a spatial grouping such as FWENZ LevelOneAlpha groupings, a minimum sample size of 100 for each grouping in a given year would be appropriate to detect effects established in the current PCL dataset but to an acceptable power level. This should not be a difficult task given the number of council parties involved (e.g., Waikato Regional Council currently monitor 80 per year). This value of 100 samples for a given grouping is based on the differences observed in LevelOneAlpha groupings and individual sites observed with the current PCL dataset. As such, this should be a sufficient sample size to detect changes of interest in these groupings. It is recommended however, that due to the large degree of variation in New Zealand native fish populations (Jellyman et al., 2002; Jowett & Richardson, 2003; McEwan & Joy, 2009) and the scarcity of PCL data with which these analyses were undertaken, that once a consistent dataset covering all groupings of interest with sample sizes discussed above that a further analysis of trends and variation on PCL sites be undertaken. Finally, in terms of spatial coverage, to which hierarchical level sampling is directed is vitally important in determining subsequent analyses. For example, to satisfy statistical power requirements for PCL with FWENZ LevelOneAlpha groupings as the structural determinant, this immediately reduces the level of confidence and the strength with which statements about New Zealand’s freshwater biodiversity can be made at lower hierarchical levels. Perhaps the most pragmatic approach would be to establish the lowest level at which analyses and statements about New Zealand’s freshwater want to be made and to establish a sample size based on recommendations above. Any analyses undertaken on higher hierarchical levels with a dataset collected this way will have sufficient statistical power for the objectives outlined.

Finally, in bullet point form:

- Establish which grouping or classification is to be reported on – e.g.,



Which level within this hierarchy that is wanted to be reported on will determine the ability to collect samples (or perhaps vice versa) sufficient to satisfy power and significance criteria for the effect sizes observed with current PCL datasets.

- Establish a sampling programme or collect available data to provide consistent coverage for the chosen classification/grouping in order to provide a dataset where variation and trends can be assessed in order design a monitoring programme that can detect these changes in biodiversity. For example, if “LevelOneAlpha” was determined to be the classification which would ultimately be reported on, direct sampling/data collection at providing a dataset with sample sizes consistent across the classifications of interest to the dimensions described above.
- Once the above dataset is available, assess the variation between classifications/groupings that will ultimately be reported on and develop a monitoring programme aimed at sampling these classifications as power analyses above suggest (adjusted for new data with consistent coverage of classifications).

9.0 Limiting Factors

Basing the preceding power analysis off a limited dataset creates some obvious limitations. Analyses and corresponding sample size and power estimates were based off differences seen in both PCL groupings at LevelOneAlpha groupings partly as these were suggested initially and because this coincidently provided a grouping with a reasonable number of samples to analyse. Due to the intention to monitor PCL sites analysis was also directed to the samples that currently lie on PCL. It was deemed reasonable to base power analyses and estimation of reasonable effect sizes from observed variation within PCL sites. However, due to the structure and sample size within PCL sites there will be some discrepancies – hence, the need to establish a more comprehensive dataset covering PCL sites prior to basing a future monitoring programme off the dataset.

10.0 References

Allibone, R., David, B., Hitchmough, R., Jellyman, D., Ling, N., Ravenscroft, P., & Waters, J. (2010). Conservation status of New Zealand freshwater fish, 2009. *New Zealand Journal of Marine and Freshwater Research, 44*(4), 271-287. doi: 10.1080/00288330.2010.514346

Braby, C. E., & Somero, G. N. (2006). Ecological gradients and relative abundance of native (Mytilus trossulus) and invasive (Mytilus galloprovincialis) blue mussels in the California hybrid zone. *Marine Biology, 148*(6), 1249-1262. doi: 10.1007/s00227-005-0177-0

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*: Hillsdale, N.J. : L. Erlbaum Associates, 1988. 2nd ed.

Emmerson, M. C., & Raffaelli, D. G. (2000). Detecting the effects of diversity on measures of ecosystem function: experimental design, null models and empirical observations. *Oikos, 91*(1), 195-203.

Irvine, K. M., Dinger, E. C., & Sarr, D. (2011). A power analysis for multivariate tests of temporal trend in species composition. *Ecology, 92*(10), 1879-1886.

Jellyman, D. J., Chisnall, B. L., Sykes, J. R. E., & Bonnett, M. L. (2002). Variability in spatial and temporal abundance of glass eels (Anguilla spp.) in New Zealand waterways. *New Zealand Journal of Marine and Freshwater Research, 36*(3), 511-517.

Johnson, P. C. D., Barry, S. J. E., Ferguson, H. M., & Mueller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution, 6*(2), 133-142. doi: 10.1111/2041-210x.12306

Jowett, I. G., & Richardson, J. (1996). Distribution and abundance of freshwater fish in New Zealand rivers. *New Zealand Journal of Marine and Freshwater Research, 30*(2), 239-255.

Jowett, I. G., & Richardson, J. (2003). Fish communities in New Zealand rivers and their relationship to environmental variables. *New Zealand Journal of Marine and Freshwater Research, 37*(2), 347-366.

Joy, M. K., David, B., Lake, M. . (2013). *New Zealand Freshwater Fish Sampling Protocols*. Palmerston North, New Zealand: Massey University

Joy, M. K., & Death, R. G. (2001). Control of freshwater fish and crayfish community structure in Taranaki, New Zealand: dams, diadromy or habitat structure? *Freshwater Biology, 46*(3), 417-429. doi: 10.1046/j.1365-2427.2001.00681.x

Kelly, D., West, D., Robertson, H., Doehring, K.A.M., and Gansell, O. (2013). *Development of a freshwater tier 1 biodiversity monitoring programme: scoping report 2013* (No. 2362).

Leathwick, J. R., Elith, J., Chadderton, W. L., Rowe, D., & Hastie, T. (2008). Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *Journal of Biogeography, 35*(8), 1481-1497. doi: 10.1111/j.1365-2699.2008.01887.x

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series, 321*, 267-281. doi: 10.3354/meps321267

McEwan, A. J., & Joy, M. K. (2009). Differences in the distributions of freshwater fishes and decapod crustaceans in urban and forested streams in Auckland, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 43*(5), 1115-1120.

McIntosh, A. R., McHugh, P. A., Dunn, N. R., Goodman, J. M., Howard, S. W., Jellyman, P. G., . . . Woodford, D. J. (2010). The impact of trout on galaxiid fishes in New Zealand. *New Zealand Journal of Ecology, 34*(1), 195-206.

Oakley, K. L., Thomas, L. P., & Fancy, S. G. (2003). Guidelines for long-term monitoring protocols. *Wildlife Society Bulletin, 31*(4), 1000-1003.

Osler, G. H. R. (2002). The Need for, and Use of Power Analysis in Tests of the Role of Species Richness in Ecosystem Processes. *Oikos, 96*(2), 383-385. doi: 10.2307/3547741

Somerfield, P. J., Clarke, K. R., & Olsgard, F. (2002). A comparison of the power of categorical and correlational tests applied to community ecology data from gradient studies. *Journal of Animal Ecology, 71*(4), 581-593. doi: 10.1046/j.1365-2656.2002.00624.x

Walsh, J. C., Venter, O., Watson, J. E. M., Fuller, R. A., Blackburn, T. M., & Possingham, H. P. (2012). Exotic species richness and native species endemism increase the impact of exotic species on islands. *Global Ecology and Biogeography, 21*(8), 841-850. doi: 10.1111/j.1466-8238.2011.00724.x

11.0 Appendices

11.1 Appendix A: Summary data for LevelOneAlpha groupings.

Table 16. PCL Site Summary Data

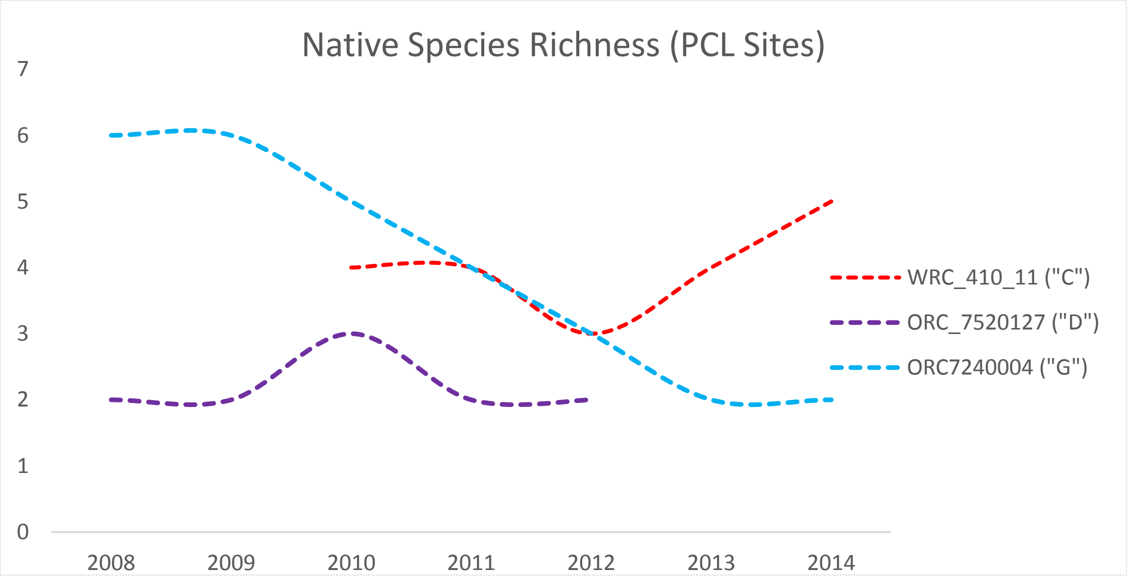
LevelOneAlpha class	Number of Samples	Mean Sp Richness	Mean Relative Species Richness	Native Abundance	Natives/m	Std Deviation Sp Richness
A	4	2.25	0.75	0.75	0.126666667	1.707825128
C	69	3.768115942	0.896325414	0.939977273	0.856231884	2.474249307
D	9	1.444444444	0.388888889	0.510430599	0.822962963	1.013793755
G	13	2.846153846	0.751098901	0.619192322	0.112820513	1.863963244
H	3	2	0.583333333	0.597787833	0.388888889	1

Table 17. NON-PCL Site Summary Data

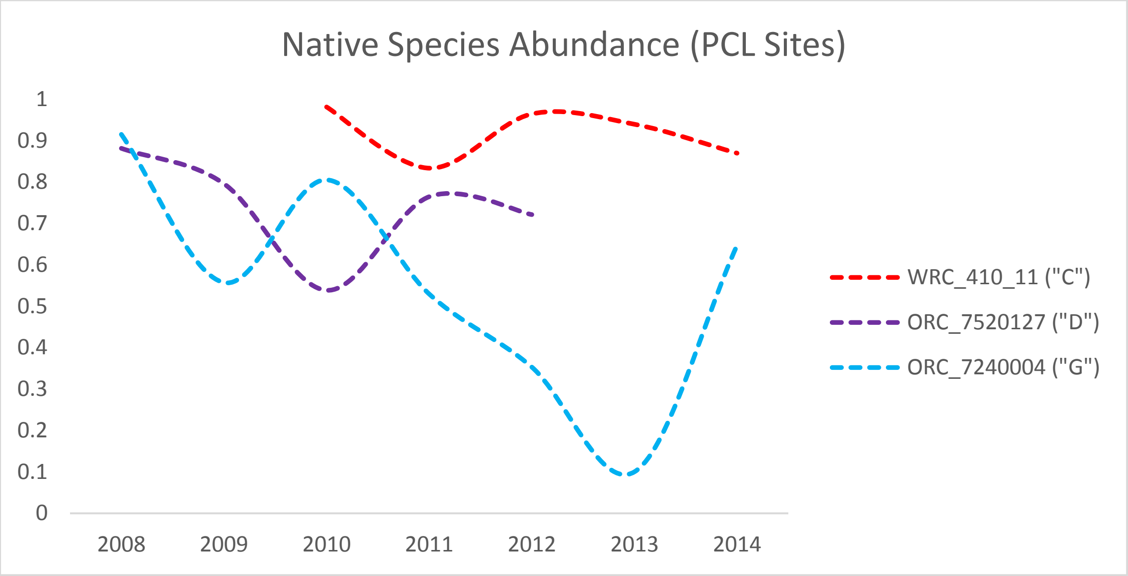
LevelOneAlpha class	Number of Samples	Mean Sp Richness	Mean Relative Species Richness	Native Abundance	Natives/m	Std Deviation Sp Richness
A	91	4.186813187	0.867569335	0.891090564	0.606300366	2.128703768
B	3	4	0.821428571	0.960605644	1.488888889	1
C	345	3.417391304	0.758189451	0.738774906	1.053487923	2.234699691
D	32	3.59375	0.829960317	0.862853151	0.800208333	2.38082901
G	84	3.142857143	0.82542517	0.840470568	0.78047619	2.146124875

11.2 Appendix B. Summary data and figures for sites which temporal trends were able to be calculated.

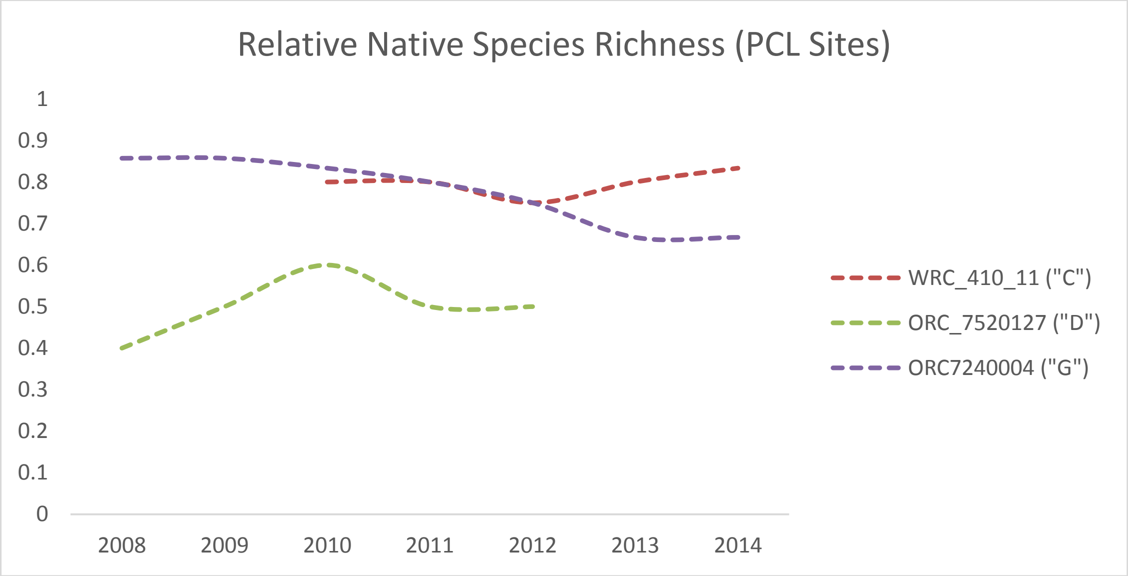
Native Species Richness (PCL sites)			
Year	**WRC_410_11 ("C")**	**ORC_7520127 ("D")**	**ORC7240004 ("G")**
2008		2	6
2009		2	6
2010	4	3	5
2011	4	2	4
2012	3	2	3
2013	4		2
2014	5		2
Correlation (r)	**0.447213595**	**0**	**-0.979957887**
Power (α = 0.05)	**12%**	**5%**	**93%**
Significance (α = 0.05 regression)	**0.450184856**	**1**	**0.000108052**



Native Species Abundance (PCL sites)			
Year	**WRC_410_11 ("C")**	**ORC_7520127 ("D")**	**ORC_7240004 ("G")**
2008		0.88125	0.914893617
2009		0.795698925	0.557142857
2010	0.981132075	0.538461538	0.805194805
2011	0.833333333	0.764397906	0.529411765
2012	0.964285714	0.721212121	0.352941176
2013	0.939393939		0.1
2014	0.869565217		0.647058824
Correlation (r)	**-0.291781006**	**-0.437109781**	**-0.611993102**
Power (α = 0.05)	**8%**	**12%**	**21%**
Significance (α = 0.05 regression)	**0.633833787**	**0.461723473**	**0.144129179**

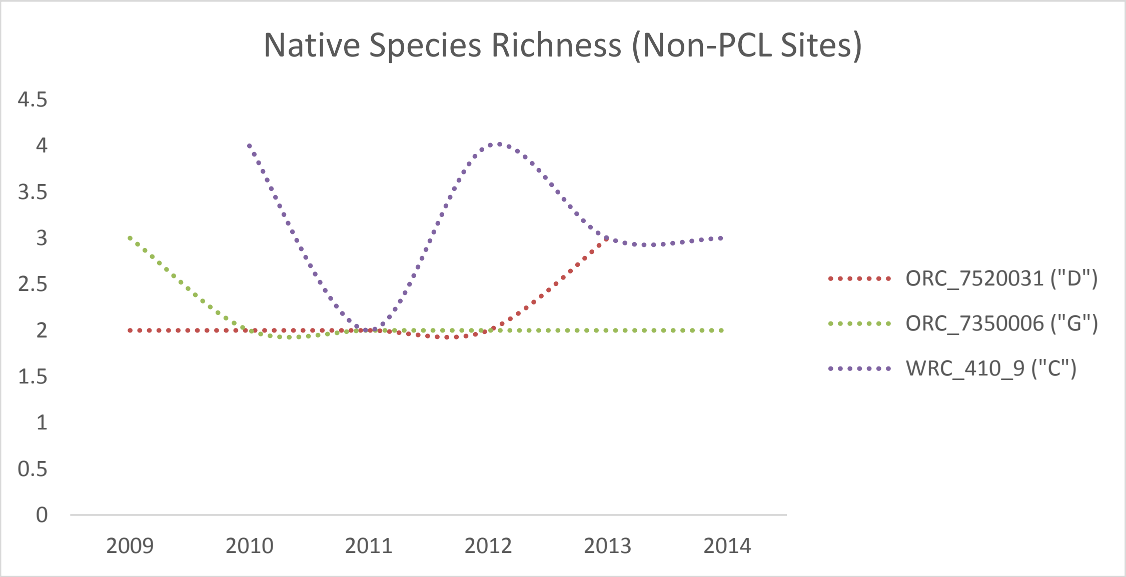


Relative Native Species Richness (PCL Sites)			
Year	**WRC_410_11 ("C")**	**ORC_7520127 ("D")**	**ORC7240004 ("G")**
2008		0.4	0.857142857
2009		0.5	0.857142857
2010	0.8	0.6	0.833333333
2011	0.8	0.5	0.8
2012	0.75	0.5	0.75
2013	0.8		0.666666667
2014	0.833333333		0.666666667
Correlation (r)	**0.353553391**	**0.447213595**	**-0.959482051**
Power (α = 0.05)	**9%**	**12%**	**98%**
Significance (regression)	**0.559404344**	**0.450184856**	**0.000620983**

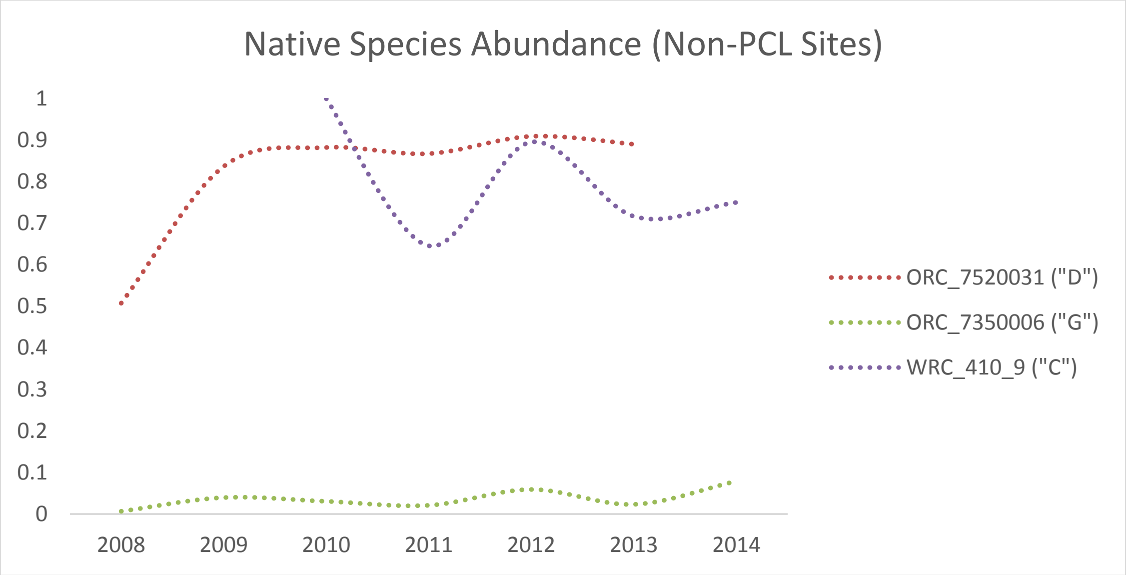


11.3 Appendix C: Summary data and figures for NON-PCL sites which temporal trends were able to be calculated.

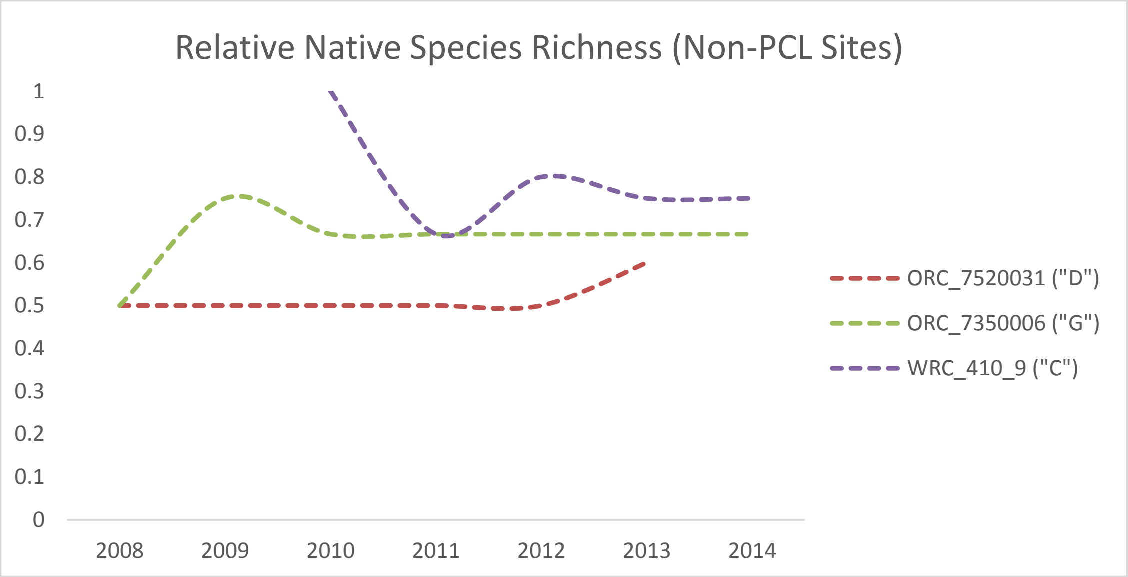
Native Species Richness (Non-PCL Sites)			
Year	**ORC_7520031 ("D")**	**ORC_7350006 ("G")**	**WRC_410_9 ("C")**
2008	2	1	
2009	2	3	
2010	2	2	4
2011	2	2	2
2012	2	2	4
2013	3	2	3
2014		2	3
Correlation (r)	**0.654653671**	**0.133630621**	**-0.188982237**
Power (α = 0.05)	**31%**	**6%**	**6%**
Significance (regression)	**0.158302423**	**0.775161696**	**0.760820376**



Native Species Abundance (Non-PCL Sites)			
Year	**ORC_7520031 ("D")**	**ORC_7350006 ("G")**	**WRC_410_9 ("C")**
2008	0.507352941	0.006329114	
2009	0.836653386	0.03930131	
2010	0.882352941	0.03030303	1
2011	0.8675	0.020689655	0.64556962
2012	0.909246575	0.058823529	0.895522388
2013	0.890140845	0.022988506	0.716666667
2014		0.078651685	0.75
Correlation (r)	**0.739898268**	**0.66459063**	**-0.47247679**
Power (α = 0.05)	**43%**	**41%**	**13%**
Significance (regression)	**0.092681047**	**0.103404334**	**0.421622611**



Relative Native Species Richness (Non-PCL Sites)			
Year	**ORC_7520031 ("D")**	**ORC_7350006 ("G")**	**WRC_410_9 ("C")**
2008	0.5	0.5	
2009	0.5	0.75	
2010	0.5	0.666666667	1
2011	0.5	0.666666667	0.666666667
2012	0.5	0.666666667	0.8
2013	0.6	0.666666667	0.75
2014		0.666666667	0.75
Correlation (r)	**0.654653671**	**0.34299717**	**-0.52681219**
Power (α = 0.05)	**31%**	**12%**	**15%**
Significance (regression)	**0.158302423**	**0.451349052**	**0.361708164**



11.4 Appendix D: LevelOneAlpha Groups Descriptions.

A	Lowland, low-gradient streams and rivers
B	Lowland, low-gradient streams draining peatlands
C	Lowland hill country gravel-bed streams
D	South Island low-elevation streams and rivers in dry inland areas
E	South Island low-elevation large rivers in dry, inland areas
F	Small tributaries on braided river floodplains
G	Mid-elevation streams and rivers in dry inland areas
H	Steep, mid-elevation streams and rivers in eastern areas
I	Rivers connecting mountains to sea in wet western areas
J	Mid-elevation headwater streams in wet western areas
K	Mid-elevation, glacially-influenced small rivers in Southern Alps
L	Mid-elevation, glacially-influenced streams in the Southern Alps
M	Glacially-influenced low-gradient rivers
N	Eastern, high-elevation steep headwater streams
O	Western, high-elevation steep headwater streams
P	High-elevation, very cold, steep headwater streams
Q	Very high-elevation, very cold, steep headwater streams
R	Very high-elevation, very cold headwater streams in cirque basins
S	Very high-elevation, very cold, steep headwater streams with glacial influence
T	Very high-elevation, very cold, steep headwater streams with strong glacial influence