

Bioacoustic monitoring of New Zealand avifauna before and after aerial 1080 operations

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

Master of Science
in
Ecology and Biodiversity

by
Roald Egbert Harro Bomans

School of Biological Sciences
Victoria University of Wellington
Te Whare Wānanga o Te Ūpoko o Te Ika a Māui
2018

Abstract

Introduced mammalian predators, namely possums, stoats and rats, are the leading cause of decline in native avifauna in New Zealand. The control of these species is essential to the persistence of native birds. A major component of mammal control in New Zealand is carried out through the aerial distribution of the toxin sodium monofluoroacetate (otherwise known as 1080). The use of this toxin, however, is subject to significant public debate. Many opponents of its use claim that forests will ‘fall silent’ following aerial operations, and that this is evidence of negative impacts on native bird communities. With the continued and likely increased use of this poison, monitoring the outcomes of such pest control operations is necessary to both address these concerns and inform conservation practice. The recent growth in autonomous recording units (ARUs) provides novel opportunities to conduct monitoring using bioacoustics. This thesis used bioacoustic techniques to monitor native bird species over three independent aerial 1080 operations in the Aorangi and Rimutaka Ranges of New Zealand.

In Chapter 2, diurnal bird species were monitored for 10-12 weeks over two independent operations in treatment and non-treatment areas. At the community level, relative to non-treatment areas, the amount of birdsong recorded did not decrease significantly in treatment areas across either of the operations monitored. At the species level, one species, the introduced chaffinch (*Fringilla coelebs*), showed a significant decline in the prevalence of its calls in the treatment areas relative to non-treatment areas. This was observed over one of the two operations monitored. Collectively, these results suggest that diurnal native avifaunal communities do not ‘fall silent’ following aerial 1080 operations.

The quantity of data produced by ARUs can demand labour-intensive manual analysis. Extracting data from recordings using automated detectors is a potential solution to this issue. The creation of such detectors, however, can be subjective, iterative, and time-consuming. In Chapter 3, a process for developing a parsimonious, template-based detector in an efficient, objective manner was developed. Applied to the creation of a detector for morepork (*Ninox novaeseelandiae*) calls, the method was highly successful as a directed means to achieve parsimony. An initial pool of 187 potential templates was reduced to 42 candidate templates. These were further refined to a 10-template detector capable of making 98.89% of the detections possible with all 42 templates in approximately a quarter of the processing time for the dataset tested. The detector developed had a high precision (0.939) and moderate sensitivity

(0.399) with novel recordings, developed for the minimisation of false-positive errors in unsupervised monitoring of broad-scale population trends.

In Chapter 4, this detector was applied to the short-term 10-12 week monitoring of morepork in treatment and non-treatment areas around three independent aerial 1080 operations; and to longer-term four year monitoring in two study areas, one receiving no 1080 treatment, and one receiving two 1080 treatments throughout monitoring. Morepork showed no significant difference in trends of calling prevalence across the three independent operations monitored. Longer-term, a significant quadratic effect of time since 1080 treatment was found, with calling prevalences predicted to increase for 3.5 years following treatment. Collectively, these results suggest a positive effect of aerial 1080 treatment on morepork populations in the lower North Island, and build on the small amount of existing literature regarding the short- and long-term response of this species to aerial 1080 operations.

Acknowledgements

First and foremost, I would like to thank my supervisor Stephen Hartley for his invaluable help, support and guidance, from experimental design to statistical theory, manuscript advice, fieldwork coordination, the list goes on. Thank you for always having an open door; it is too little to say that this thesis would not have been possible without you.

A big thank you to Dan Crossett for his help with the coordination of and great company on field trips, who made data collection both possible and extremely enjoyable. I would also like to thank Alex Verry, Bob Burgess, Ned Bruno, Davide Santoro, Mitch Ganley, Jason Foote, Rikki August-Carey, Aurore Fanal, Charlie Clarke, and anyone that I may have missed who followed me up the mountains to help with field work.

Thank you to the members of Bug Club for their discussions and constructive help and feedback throughout the course of this project, particularly Nyree Fea and Asher Cook for their help with call identification and acoustic monitoring theory; and Victor Anton for his help with R.

Thank you also to members of the Remutaka Conservation Trust who have helped along the way, particularly Peter Cooper for his significant and invaluable help both in project coordination and with field work; and Susan Ellis for her help with bird call identification. I would also like to Ed Abraham for his discussions and insights regarding automated detection.

I would like to thank my office and cohort mates for their help and support, particularly Antoine Felden for his help with graphing in R; and Jess Russell, Matt Biddick, and Kate Irving for their help with proofreading and formatting. Thank you also to my friend Bradley Jamieson, and father Harry Bomans for their help with formatting and proofreading.

Thank you to Victoria University of Wellington's Masters by Thesis Scholarship, the Tararua Tramping Club's Michael Taylor Memorial Award, and the Alison Morton Scholarship in Ecology/Marine Biology for the generous funding that made this project possible.

Finally, I would like to thank my friends and family, particularly my parents Harry and Cora Bomans, for their never-ending love and support. Thank you for believing in me and encouraging me to pursue something I am passionate about, and for all of the help and advice that you've given along the way. Without you all, this would not have been possible.



Roald Bomans, 2017

Table of Contents

Abstract	i
Acknowledgements	iii
Table of Contents	vii
List of Figures	x
List of Tables	xii
Chapter 1 - General Introduction	1
1.1. Chapter overview	2
1.2. The context for pest control	2
1.3. Use and controversy of 1080 poison	3
1.4. Impacts of 1080 on native avifauna	4
1.5. Monitoring avifauna over pest control operations	5
1.5.1. Traditional methods	5
1.5.2. Bioacoustic monitoring.....	6
1.6. Thesis structure	8
1.6.1. Overview	8
1.6.2. Research aims	9
Chapter 2 - Bioacoustic monitoring of native diurnal bird species short-term around aerial 1080 operations	11
2.1. Introduction	12
2.2. Methods.....	14
2.2.1. Study areas and study sites	14
2.2.2. 1080 operations.....	14
2.2.3. Experimental design.....	15
2.2.4. Audio recording specifications	16
2.2.5. Audio recording selection	18
2.2.6. Scoring audio recordings	19

2.2.7. Statistical analyses	21
2.3. Results	22
2.3.1. Community-level	22
2.3.2. Species-level	22
2.4. Discussion	26
Chapter 3 - Achieving parsimony in developing a template-based automated acoustic detector – an applied study with morepork (<i>Ninox novaeseelandiae</i>), a New Zealand owl	33
3.1. Introduction	34
3.2. Methods.....	36
3.2.1. Background of detection software	36
3.2.2. Development process theory and overview	39
3.2.3. Applied detector development	41
3.3. Results	53
3.3.1. Utility of the detector rationalisation process	53
3.3.2. Detection of <i>more-pork</i> calling events.....	53
3.3.3. Detection of <i>mew</i> and <i>cree</i> calling events.....	54
3.3.4. Assessment of background noise effects on sensitivities	54
3.4. Discussion	55
Chapter 4 - Applied use of an automated detector for short- and long-term monitoring of morepork (<i>Ninox novaeseelandiae</i>) around aerial 1080 operations	61
4.1. Introduction	62
4.2. Methods.....	64
4.2.1. Study areas and study sites	64
4.2.2. 1080 operations.....	65
4.2.3. Experimental Design.....	66
4.2.4. Sound recording specifications	70
4.2.5. Audio recording selection	70

4.2.6. Scoring audio recordings	71
4.2.7. Statistical analyses	72
4.3. Results	74
4.3.1. Short-term effects.....	74
4.3.2. Longer-term trends.....	75
4.4. Discussion	79
Chapter 5 - General Discussion	87
5.1. Responses of avifauna to aerial 1080 operations	88
5.1.1. Short-term community-level effects	88
5.1.2. Species-level effects.....	89
5.2. Automated monitoring	91
5.3. Limitations and considerations.....	92
5.4. Conclusion.....	94
References.....	97
Appendices.....	A1
Appendix I - Maps of GWRC wind speed and rainfall stations sampled across chapters..	A1
Appendix II - Recorder locations and dates sampled in diurnal analyses	A3
Appendix III - Example structures of input-data and chi-square distributions.....	A5
Appendix IV - Mean cumulative rainfall (mm) five weeks following 1080 operations monitored	A7
Appendix V - R script for creation of 10-second .wav files containing target calls.....	A8
Appendix VI - Confusion matrices of event classifications made in raw and restructured training, refining and test datasets	A10
Appendix VII - R script for automated morepork detector developed	A12
Appendix VIII - Recorder locations and date sampled in short-term nocturnal analyses	A19
Appendix IX - Dates sampled in long-term nocturnal analyses	A22
Appendix X - Recorder location specifications	A26

List of Figures

Figure 2.1 Map showing diurnal study sites and approximate forested extent of focus study areas	15
Figure 2.2 Map of study areas, study sites, and recorder locations monitored diurnally over the June 2017 Aorangi 1080 operation	17
Figure 2.3 Map of study areas, study sites, and recorder locations monitored diurnally over the July 2017 Southern Rimutaka 1080 operation.....	17
Figure 2.4 Changes in mean silence in treatment and non-treatment areas from before to after the June 2017 Aorangi and July 2017 Southern Rimutaka aerial 1080 operations	23
Figure 2.5 Changes in species' mean calling prevalences from before to after 1080 treatment for treatment and non-treatment areas monitored diurnally across the June 2017 Aorangi operation	24
Figure 2.6 Changes in species' mean calling prevalences from before to after 1080 treatment for treatment and non-treatment areas monitored diurnally across the July 2017 Southern Rimutaka operation.....	25
Figure 3.1 Examples of cross-correlation and binary automated detector call templates.....	36
Figure 3.2 Examples of true-positive, false-positive and true-negative automated detection classification events	38
Figure 3.3 Flow diagram outlining the broader fundamental stages of a template-based detector development process developed	40
Figure 3.4 An example template-based detector rationalisation scenario	40
Figure 3.5 Flow diagram outlining the detailed fundamental steps of a template-based detector development process developed	42
Figure 3.6 Spectrogram examples of five broader morepork call categories	44
Figure 3.7 Example spectrograms of six types of ' <i>more-pork</i> ' calls classified by sound and spectral appearance	45
Figure 3.8 Example spectrogram of a calling event requiring a margin in ROC analyses	46
Figure 3.9 Scatterplots showing the precisions, specificities, sensitivities, and FPRs of local maxima for a call tested as a binary and a cross-correlation template.....	49
Figure 3.10 Flow diagram outlining applied steps of the detector development process and their results in the creation of a 10-template ' <i>more-pork</i> ' call detector developed	50
Figure 3.11 Bar graphs of raw and restructured TPR (sensitivity) and FPR estimates for the final ' <i>more-pork</i> ' detector across the training, refining and test datasets.....	54

Figure 3.12 Scatterplot showing the lack of relationship between observed TPRs and mean power density values (dB FS) across recordings	54
Figure 4.1 Map showing nocturnal study sites and approximate forested extent of focus study areas	65
Figure 4.2 Map of study areas, study sites, and recorder locations monitored for short-term morepork monitoring around the August 2014 Aorangi 1080 operation	68
Figure 4.3 Map of study areas, study sites, and recorder locations monitored for short-term morepork monitoring around the June 2017 Aorangi 1080 operation.....	68
Figure 4.4 Map of study areas, study sites, and recorder locations monitored for short-term morepork monitoring around the July 2017 Southern Rimutaka 1080 operation	69
Figure 4.5 Map of study areas, study sites, and recorder locations monitored for long-term morepork monitoring (summer 2013 – winter 2017)	69
Figure 4.6 Mean detected morepork calling prevalences in treatment and non-treatment study areas before and after the 1080 operations monitored nocturnally, and associated chi-square and p-values	74
Figure 4.7 Mean detected morepork calling prevalences in the Aorangi (treatment) and Northern Rimutaka (non-treatment) Range study areas across long-term monitoring.....	76
Figure 4.8 Predicted change in mean detected morepork call prevalence as a function of time since a hypothetical 1080 treatment.....	77
Figure 4.9 Boxplots depicting the difference in morepork calling prevalences detected at recorder locations active in the summer 2013-14 and summer 2015-16 monitoring seasons .	79
Figure A.I-1 Map of GWRC rainfall data stations used to source cumulative rainfall data across chapters	A1
Figure A.I-2 Map of GWRC climate data stations used to source wind speed data across chapters	A2
Figure A.III-1 Example structure of calling prevalences scored at recorder locations in treatment and non-treatment areas before and after the 2017 Aorangi operation.....	A5
Figure A.III-2 Examples of chi-square distributions approximated using permutations of data presented in Figure AIII-1, and the observed chi-square values and their significance relative to the approximated distribution	A6

List of Tables

Table 2.1 List of 17 focal species monitored in diurnal acoustic recordings	20
Table 3.1 Table of terms and definitions referred to Chapter 3	35
Table 3.2 Summary confusion matrix of event possibilities in automated detection	37
Table 3.3 Automated detection performance metrics and their associated formulae	39
Table 3.4 Summary of training, refining and test dataset compositions	47
Table 3.5 Summary of the number of detections made by refinement detectors and the final detector before and after rationalisation, and by the 42 candidate template detector	53
Table 4.1 Table of results for Analysis of Deviance (Type III Wald chi-square tests) of fixed effects retained in the best-fitting logistic mixed effects model for long-term nocturnal data	75
Table 4.2 Table of results for Analysis of Deviance (Type III Wald chi-square tests) of fixed effects retained in the best-fitting logistic mixed effects model for long-term data, with recorder location fitted as a fixed effect	77
Table 4.3 Table of Tukey test p-values for pairwise comparisons of sites monitored long-term in the Northern Rimuaka Range	78
Table 4.4 Table of Tukey test p-values for pairwise comparisons of sites monitored long-term in the Aorangi Range	78
Table A.II-1 Table of recorder locations and dates sampled before and after the June 17 2017 Aorangi 1080 operation in diurnal monitoring	A3
Table A.II-2 Table of recorder locations and dates sampled before and after the July 30 2017 Southern Rimutaka 1080 operation in diurnal monitoring	A4
Table A.IV-1 Table of mean cumulative rainfall across GWRC five rainfall stations five weeks after the Aorangi 2014, Aorangi 2017 and Southern Rimutaka 2017 operations monitored .	A7
Table A.VI-1 Confusion matrix of raw event classifications made by the final <i>more-pork</i> detector developed when run on the training dataset.....	A10
Table A.VI-2 Confusion matrix of restructured event classifications made by the final <i>more-pork</i> detector developed when run on the training dataset	A10
Table A.VI-3 Confusion matrix of raw event classifications made by the final <i>more-pork</i> detector developed when run on the refining dataset.....	A10
Table A.VI-4 Confusion matrix of restructured event classifications made by the final <i>more-pork</i> detector developed when run on the refining dataset	A11
Table A.VI-5 Confusion matrix of raw event classifications made by the final <i>more-pork</i> detector developed when run on the test dataset.....	A11

Table A.VI-6 Confusion matrix of restructured event classifications made by the final <i>more-pork</i> detector developed when run on the test dataset	A11
Table A.VIII-1 Table of recorder locations and dates nocturnally sampled before and after the 2014 Aorangi 1080 operation	A19
Table A.VIII-2 Table of recorder locations and dates nocturnally sampled before and after the 2017 Aorangi 1080 operation	A20
Table A.VIII-3 Table of recorder locations and dates nocturnally sampled before and after the 2017 Southern Rimutaka 1080 operation	A21
Table A.IX-1 Table of recorder locations nocturnally sampled in long-term monitoring over the summer 2013-14 sampling season.	A22
Table A.IX-2 Table of recorder locations nocturnally sampled in long-term monitoring over the winter 2014 sampling season.	A22
Table A.IX-3 Table of recorder locations nocturnally sampled in long-term monitoring over the summer 2014-15 sampling season.	A23
Table A.IX-4 Table of recorder locations nocturnally sampled in long-term monitoring over the winter 2015 sampling season.	A23
Table A.IX-5 Table of recorder locations nocturnally sampled in long-term monitoring over the summer 2015-16 sampling season.	A24
Table A.IX-6 Table of recorder locations nocturnally sampled in long-term monitoring over the winter 2016 sampling season	A24
Table A.IX-7 Table of recorder locations nocturnally sampled in long-term monitoring over the summer 2016-17 sampling season.	A25
Table A.IX-8 Table of recorder locations nocturnally sampled in long-term monitoring over the winter 2017 sampling season	A25
Table A.X-1 Table of GPS information for recorder locations monitored in the Aorangi Range study area	A26
Table A.X-2 Table of GPS information for recorder locations monitored in the Northern Rimutaka Range study area.	A27
Table A.X-3 Table of GPS information for recorder locations monitored in the Tora Bush study area	A27
Table A.X-4 Table of GPS information for recorder locations monitored in the Tararua Range study area	A28
Table A.X-5 Table of GPS information for recorder locations monitored in the Southern Rimutaka study area.....	A28

Chapter 1 - General Introduction

1.1. Chapter overview

The following general introduction provides a brief overview of (1) the context for pest control in New Zealand; (2) the role of 1080 in pest control and the controversy surrounding its use with respect to native bird species; (3) the known effects of 1080 on native bird species, and; (4) the potential for bioacoustics to inform pest management regarding this issue. Its purpose is to give a wider context to the issues and topics explored in Chapters 2, 3 and 4 of this thesis. Collectively, these chapters utilise bioacoustic monitoring with autonomous recording units (ARUs) to evaluate the anecdotal claim that New Zealand forests ‘fall silent’ after aerial 1080 operations. The effects of operations are investigated for native diurnal bird species short-term, and for the morepork (*Ninox novaeseelandiae*), New Zealand’s only extant native owl, both short and long-term.

1.2. The context for pest control

Invasive exotic species are a leading cause of biodiversity loss globally (Courchamp *et al.*, 2003; Steer, 2010), particularly in island scenarios where exotic predators are introduced (Parlato *et al.*, 2015). This is true for native New Zealand avifauna which, having evolved in relative isolation, are naïve and poorly adapted to mammalian predators (Holdaway, 1989; Innes *et al.*, 2010; Wright, 2011). Since the introduction of a range of mammalian species with the arrival of humans, 41% of New Zealand’s endemic birds have become extinct (Holdaway *et al.*, 2001), and 77% of remaining forest bird species are classified as either declining or of conservation concern (Hitchmough *et al.*, 2007; Innes *et al.*, 2010). Predation by introduced mammalian predators has long been accepted as a significant factor contributing to these declines, and is now described as the main threat to the persistence of New Zealand’s avifauna (Parlato *et al.*, 2015). Specifically, predation from common brushtail possums (*Trichosurus vulpecula*), stoats (*Mustela erminea*) and rats (*Rattus exulans*, *R. norvegicus*, and *R. rattus*) has been identified as the main driver of population declines (Elliott & Kemp, 2016; Innes *et al.*, 2010; Wright, 2011). Thus, the effective management of these predators is essential to the conservation of New Zealand’s remaining native avifauna (Brown *et al.*, 2015; Innes *et al.*, 2010).

The effect of introduced mammals is not restricted to predation of native avifauna: pest mammals have significant negative effects on other native vertebrates, invertebrates, vegetation and wider ecosystems (Atkinson *et al.*, 1995; Wright, 2011). Possums also pose a significant risk to New Zealand’s cattle farming industry as a vector of bovine tuberculosis

(TB) (Warburton & Livingstone, 2015; Weaver, 2006). This disease was first identified in New Zealand possums in the late 1960s, and lethal control of possums has been conducted since the early 1970s in a bid to control and eradicate it (Byrom *et al.*, 2016; Warburton & Livingstone, 2015).

Considering these interests, pest control occurs at a significant scale throughout New Zealand; some 10 million hectares are managed for conservation and TB control purposes (Byrom *et al.*, 2016). This control is carried out primarily using ground-based trapping, ground-based poisoning, and the aerial distribution of toxins (Brown *et al.*, 2015; Wright, 2011).

1.3. Use and controversy of 1080 poison

A major component of pest control operations is the vertebrate toxin sodium monofluoroacetate (hereafter 1080). This toxin is chemically and toxicologically identical to the compound fluoroacetate found naturally in a range of poisonous plants. 1080 operates through interference with the Krebs cycle, a metabolic pathway that breaks down carbohydrates to provide energy for wider cellular function. Once consumed, 1080 is converted to fluorocitrate, which subsequently interferes with enzymes of the Krebs cycle. This causes an accumulation of citrate, inhibition of energy production and, ultimately, death (Eason, 2002; Eason *et al.*, 2011). The effectiveness of this toxin varies species-to-species: carnivorous species have been found to be highly susceptible, with herbivorous mammals less susceptible and birds even less so (Eason *et al.*, 2011; Eisler, 1995). As a result of its effectiveness against pest mammal species and New Zealand's lack of vulnerable native land mammals, New Zealand is the leading consumer of 1080 globally (Eason *et al.*, 2011), comprising 80% of its use (Green & Rohan, 2012).

Three forms of bait distribution are utilised in New Zealand 1080 poison operations: bait station distribution, ground distribution by hand, and aerial sowing. With the exception of brodifacoum in a small number of cases, 1080 is the only poison registered for the aerial control of possums, rats and stoats on mainland New Zealand (Eason, 2002; Wright, 2011). In aerial control, the poison is combined with carrot or cereal baits which are typically applied at ~2 kg/ha following modern procedures. This form of application is used extensively due to its cost-effectiveness and suitability for large areas and rugged terrain (Wright, 2011). Used in New Zealand since 1956 (Wright *et al.*, 2002), 1080 is highly effective in possum control and is the most extensively used poison for this purpose (Eason *et al.*, 2010). More recently, aerial 1080 distribution has also been found to be effective in killing both rats, when pre-feeding is

carried out (the distribution of non-toxic baits a number of weeks before toxic baits); and stoats, through secondary poisoning, whereby poisoned prey are consumed prior to the decomposition of the toxin (Elliott & Kemp, 2016). Given its effectiveness and utility in targeting large and difficult to access areas, the use of 1080 continues to grow. The largest aerial application to date occurred in 2014, where 649,000 hectares of forest (~10% of New Zealand's forested area) were treated with 1080-laden cereal baits by the Department of Conservation in a bid to control high rat and stoat numbers resulting from significant beech tree masting events (Elliott & Kemp, 2016).

The aerial use of 1080, however, is not without controversy. Public support for the poison has declined (Eason *et al.*, 2010), with 43% of respondents to a 2009 survey supporting its use, down from 52% in 2001 (Green & Rohan, 2012). A range of concerns exist, encompassing animal welfare, water supply contamination, sub-lethal effects on humans, and the potential for exposure to and by-kill of native bird species and other non-target fauna (Eason *et al.*, 2011). By-kill, particularly of native species, is a major concern to 1080 opponents (Graf & Graf, 2009; Reider, 2012; Slater, 2015), and has been identified as a leading environmental concern regarding the poison's use (Green & Rohan, 2012). Specifically, it is often anecdotally claimed by opponents of 1080 that forests 'fall silent' (i.e. become devoid of birdsong) following operations (e.g. Reider, 2012), and that this is evidence of a negative effect of the poison on native New Zealand avifauna (Hansford, 2016; Toki, 2013).

1.4. Impacts of 1080 on native avifauna

Pest control operations occur in the interest of the protection and improved survival and breeding success of native bird species. A number of studies have shown positive effects of aerial 1080 operations for a range of species including, but not limited to, tūī (*Prosthemadera novaeseelandiae*), silvereye (*Zosterops lateralis*) (Miller & Anderson, 1992), mohua (*Mohoua ochrocephala*), rock wren (*Xenicus gilviventris*), South Island robin (*Petroica australis*), rifleman (*Acanthisitta chloris*) (Elliott & Kemp, 2016), bellbird (*Anthornis melanura*), brown creeper (*M. novaeseelandiae*), fantail (*Rhipidura fuliginosa*), grey warbler (*Gerygone igata*), and kākārīki (*Cyanoramphus novaezelandiae*) (O'Donnell & Hoare, 2012). Though individuals of some species may be subject to incidental 1080 poisoning, it has been shown, and generally accepted, that species' populations benefit overall from aerial operations through reduced predation pressure and increased breeding success (Brown *et al.*, 2015; Weaver, 2006). Nevertheless, the risk of non-target poisoning remains a significant concern. Poisoning of non-

target species is possible through both primary poisoning, through direct consumption of baits, and secondary poisoning, through the consumption of 1080-poisoned prey (Eason *et al.*, 2013b). Some species are more susceptible to poisoning than others (Eisler, 1995; Lloyd & McQueen, 2000), and the significance of poisoning is dependent on both the resilience of the species in question and its ability to offset population losses with improved breeding success following control (Eason *et al.*, 2011; Veltman & Westbrooke, 2011). If a species benefits long-term but consistently suffers short-term population losses from treatment, regimes of frequently repeated treatment may result in a net-negative impact on its populations (Veltman & Westbrooke, 2011). Thus, it is essential that any non-target impacts are known and well understood.

Corpses have been found for 19 native bird species following aerial 1080 operations since the beginning of operations in 1956 (Spurr, 2000); 11 of these species have not been studied over operations using radio-tagging or banding to directly assess mortality (Veltman & Westbrooke, 2011). Deaths recorded occurred predominantly during the 1970's, when carrot baits of varying and often small fragment size, or 'chaff', with raspberry flavoured lures were used (Eason *et al.*, 2011; Peterson, 2014). In response to studies investigating non-target mortalities, practices were changed to minimise chaff, remove raspberry lures, and incorporate green dye and cinnamon deterrents in baits to make them less attractive to birds (Peterson, 2014). Simultaneously, baits used started to shift to cereal-based baits and application rates began to reduce. Together, these changes saw a reduction in species reported dead after drops (Eason, 2002; Eason *et al.*, 2011). Further changes to application methods, namely pre-feeding, the use of deer-repellent, and the use of GPS-guided strip and cluster sowing methods, recently sparked questions again regarding avian mortality (Morriss *et al.*, 2016; Veltman & Westbrooke, 2011). This considered, a study of 15 operations by Morriss *et al.* (2016) suggests that modern methods pose a negligible risk to native bird populations. Nevertheless, the effects that aerial 1080 operations have on native bird populations remain a prominent public concern, and continued monitoring serves to inform both public perception and modern baiting practices.

1.5. Monitoring avifauna over pest control operations

1.5.1. Traditional methods

A range of methods exist for monitoring bird species over pest control operations in New Zealand. These include distance sampling (Westbrooke *et al.*, 2003), five-minute bird counts

(5MBC), checking mapped territories for known occupants, roll-calling identifiable/banded birds trained to approach observers, tracking individuals using radio-transmitters, and capture-mark-recapture/resighting methods using mist nets and leg bands (Davidson & Armstrong, 2002; Spurr & Powlesland, 2000; Veltman & Westbrooke, 2011). Detailed methodologies for a number of these methods are available in a report by Spurr and Powlesland (2000).

Roll-calling, radio-transmitter tracking and territory-mapping serve mainly for monitoring survival, whereas distance sampling, capture-mark-recapture and 5MBCs serve more to monitoring population trends (Cook, 2017). Point counts facilitate the surveying of broad areas relatively economically, and are the most used method for monitoring bird populations internationally (Bardeli *et al.*, 2010; Celis-Murillo *et al.*, 2009; Klingbeil & Willig, 2015; Spurr & Powlesland, 2000). New Zealand's adaptation of this method, the 5MBC, has been used extensively in monitoring bird populations around 1080 operations (Spurr & Powlesland, 2000). However, this method (and other traditional monitoring methods) is subject to a number of limitations, generally being labour-intensive, costly, and limited in space and time (Aide *et al.*, 2013; Marques *et al.*, 2013). Point count surveys require highly skilled observers. This limits the spatial extent to which data can be collected, particularly over short periods requiring intensive monitoring, as observers can only be in one place at any one time (Aide *et al.*, 2013; Celis-Murillo *et al.*, 2009). Data collection can be spatially and temporally limited further when monitoring is carried out in remote areas and requires significant observer effort (Digby *et al.*, 2013). The use of multiple observers to address these issues produces data-comparability issues, as different observers' varying identification abilities can lead to observer biases (Aide *et al.*, 2013; Brandes *et al.*, 2006; Celis-Murillo *et al.*, 2009). Counts can also be subject to temporal biases arising from within and between-day variation in species activity (Digby *et al.*, 2013). Furthermore, the presence of an observer in the field can affect species detectability and further bias data (Klingbeil & Willig, 2015). Collectively, these factors limit both the extent to which data can be collected spatially and temporally using traditional monitoring methods, and the comparability and thus reliability of this data.

1.5.2. Bioacoustic monitoring

A relatively new and growing method for monitoring population trends that suffers few such limitations is bioacoustic monitoring using autonomous recording units (ARUs) (Celis-Murillo *et al.*, 2009; Mortimer & Greene, 2017; Steer, 2010; Turgeon *et al.*, 2017). Bioacoustics exploits the distinctive sounds produced by species for communication, display, navigation,

and during movement, using them for their identification and monitoring in areas of interest (Blumstein *et al.*, 2011; Obrist *et al.*, 2010). Acoustic identification is already used in combination with visual identification in point-counts (Klingbeil & Willig, 2015), and is the most efficient means to survey birds due to the wide diversity of species-specific vocalisations that they produce (Brandes, 2008; Steer, 2010). Bioacoustic monitoring with ARUs uses battery-operated digital acoustic recording devices capable of passively recording such sounds on a programmable schedule. As a result of technological developments and increasing interest in their potential, a range of ARUs are now available for monitoring projects (Mortimer & Greene, 2017).

Bioacoustic monitoring using ARUs poses a number of advantages over traditional 5MBCs. Multiple ARUs can be set up by both volunteers and experts to record across multiple sites. These can be left to record for long periods of time, requiring only two visits to deploy and collect recorders (Drake *et al.*, 2016; Zwart *et al.*, 2014). Thus, monitoring using this method can achieve large-scale temporal and spatial replication with significantly little limitation from observer availability and labour-intensity (Celis-Murillo *et al.*, 2009; Klingbeil & Willig, 2015; Steer, 2010). Recording in the absence of observers also eliminates any potential variation in species' detection probabilities arising from observer presence (Klingbeil & Willig, 2015; Zwart *et al.*, 2014). Once collected, a single interpreter can analyse recordings produced, reducing the influence of observer-biases on data collected. Recorders can also be programmed to record simultaneously across sites, eliminating any potential temporal biases associated with acoustic activity (Brandes, 2008; Klingbeil & Willig, 2015). Furthermore, as recordings can be stored and replayed, the reliability of data can be further improved by cross-validation of analyses using multiple analyses with multiple observers (Celis-Murillo *et al.*, 2009). As a result of these benefits, this monitoring method has been suggested as a suitable replacement for point-counts if studies capitalise on its potential for replicate sampling (Klingbeil & Willig, 2015). A limitation arising from this method is the large amounts of recordings produced and requiring analysis (Aide *et al.*, 2013; Steer, 2010). Manual analysis of recordings can become labour-intensive (Marques *et al.*, 2013), and automated analysis using signal recognition software is increasingly being explored and applied to address this issue (Knight *et al.*, 2017). Automated analyses are critical to the viability of long-term monitoring studies employing bioacoustics (Blumstein *et al.*, 2011).

Recordings produced by ARUs can be collected and analysed out of the field to monitor species using a number of metrics. In its simplest form, bioacoustic monitoring lends itself to

monitoring species' presence. This monitoring method has proved particularly useful in detecting the presence of endangered and rare species (Blumstein *et al.*, 2011) and tracking changes in the distribution of species after pest control operations (Croll *et al.*, 2016). With respect to monitoring population trends, vocalisation counts may be recorded as an index of abundance (Frommolt & Tauchert, 2014; Marques *et al.*, 2013). Such counts are dependent on the assumption that higher counts are associated with greater population sizes (Royle, 2004). The upholding of this assumption is fundamental to the reliability of results gained from the monitoring of any species using this method. This relationship has been shown to hold true for species of mammal (Thompson *et al.*, 2010), amphibian (Leonard *et al.*, 1997; Nelson & Graves, 2004) and bird (Borker *et al.*, 2014; Oppel *et al.*, 2014). If species' calling rates can be estimated, call counts can be converted to produce density estimates (Marques *et al.*, 2013). More recent studies have also applied spatially explicit capture-recapture models to recordings from microphone arrays, using signal strength and time differences in sound arrival between recorders to localise individuals and achieve abundance and density estimates (Blumstein *et al.*, 2011; Stevenson *et al.*, 2015).

1.6. Thesis structure

1.6.1. Overview

This Master of Science project is nested within the Victoria University of Wellington (VUW) 'Aorangi Project', a wider study monitoring biodiversity in the Aorangi and Rimutaka Ranges¹ of the lower North Island of New Zealand around recurring aerial 1080 operations. The overall aim of this thesis is to utilise bioacoustic monitoring with ARUs to monitor calling prevalences as an index of abundance for native bird species' communities and populations around aerial 1080 operations, investigating the claim that forests 'fall silent' following treatment. It builds on former bioacoustic monitoring carried out within the Aorangi Project by Cook (2017) around a 2014 operation in the Aorangi Range, and utilises acoustic recordings both from archived recordings, and recorded as a part of this Master's project. The data chapters of this thesis are written in the format of stand-alone scientific papers and, as such, elements of repetition exist between them.

Chapter 2 builds directly on the work of Cook (2017), investigating changes in calling prevalences of diurnal bird species from before to after 1080 operations in treatment and non-

¹ Since the completion of this thesis, the Rimutaka Range was subject to a formal name change and is now known as the Remutaka Range; the former name is used here for consistency with figures developed prior to this change.

treatment areas to assess the claim that forests ‘fall silent’ following treatment. This chapter monitors across two operations occurring in 2017 in the Aorangi and Southern Rimutaka Ranges, utilising only data collected as a part of this project.

Chapter 3 explores the potential for an automated detector to be developed from a random set of field-quality exemplar calls for the relatively understudied morepork (*Ninox novaeseelandiae*) using open-source, template-based software.

Chapter 4 applies the detector developed in Chapter 3 to monitor the calling prevalence of morepork both short and long-term around three aerial 1080 operations in the Aorangi and Rimutaka Ranges. This chapter utilises archived long term data from the wider Aorangi Project, archived short-term data collected around the 2014 Aorangi operation by Cook (2017), and data collected as a part of this project for monitoring of the two operations occurring in the Aorangi and Southern Rimutaka Ranges in 2017.

Chapter 5 provides a synthesis and discussion of the wider results presented in Chapters 2, 3 and 4, and their implications for bioacoustic monitoring and conservation in New Zealand.

1.6.2. Research aims

Specifically, this Master’s project aims to:

1. Quantitatively test the validity of the claim that forests ‘fall silent’ shortly following aerial 1080 operations at both the community and the species level;
2. Utilise a rationalised methodology to produce a parsimonious automated detector for morepork calling events from a random set of field-sourced exemplar calls, and assess its performance;
3. Apply an automated detector to monitor the short- and long-term responses of morepork populations to aerial 1080 operations.

Chapter 2 - Bioacoustic monitoring of native diurnal bird species short-term around aerial 1080 operations

2.1. Introduction

Aerial distribution of 1080 occurs at a significant scale across New Zealand for the control of introduced mammals that act as both predators of native species and vectors of bovine tuberculosis (Byrom *et al.*, 2016; Eason *et al.*, 2010; Nugent & Morriss, 2013). The aerial use of this toxin, however, is subject to controversy (Green & Rohan, 2012). Environmentally, incidental poisoning of non-target native fauna, particularly bird species, is a principal concern (Green & Rohan, 2012; Morriss *et al.*, 2016). Pest-control is carried out in the interest of protecting native avifauna, intended to reduce predation pressure and thus increase nesting success and facilitate species' recovery (Starling-Windhof *et al.*, 2011). However, wherever toxins are broadcast aurally there is the risk of poisoning non-target native species (Veltman & Westbrooke, 2011).

Since the beginning of operations in 1956, 19 species of native bird have been found dead following aerial 1080 drops (Spurr, 2000). The majority of these mortalities occurred pre-1980, when predominantly carrot baits were distributed. Subsequent changes in baiting practices prompted by studies in the 1970's, including a shift to the use of cereal baits, significantly mitigated risk existing for avian species (Eason, 2002; Eason *et al.*, 2011; Morriss *et al.*, 2016; Veltman & Westbrooke, 2011). Continued changes to baiting practices have since prompted further questions regarding non-target impacts (Veltman & Westbrooke, 2011). However, mortality findings to-date suggest that operations utilising modern baiting practices have negligible negative impacts on native bird species (Morriss *et al.*, 2016). While complaints regarding 1080 have declined over recent years (Environmental Protection Authority, 2012), public support for its use has reportedly declined (Eason *et al.*, 2010; Green & Rohan, 2012). The incidental poisoning of non-target native avifauna remains a primary concern in ongoing debates regarding aerial operations (Green & Rohan, 2012; Morriss *et al.*, 2016). It is often proclaimed by opponents of 1080 that forests will 'fall silent' following aerial operations, and that this is evidence of significant adverse impacts for native bird species (Graf & Graf, 2009; Hansford, 2016; Reider, 2012; Slater, 2015). Thus, continued monitoring of bird populations and communities around aerial 1080 operations is required both to ensure that no adverse effects are occurring and to address these public concerns.

Bioacoustic monitoring using autonomous recording units (ARUs) is a rapidly growing monitoring method with good potential to monitor bird populations over such operations. Traditionally, five-minute bird counts (5MBCs) have been used extensively in monitoring

species' population trends over 1080 operations (Spurr & Powlesland, 2000). However, the 5MBC is subject to limitations. It requires trained observers and significant field effort, limiting the spatial scale, temporal scale, and comparability of data. Furthermore, inter-observer variability in skill and observer presence in-field can introduce observer and detectability biases in data respectively (Aide *et al.*, 2013; Klingbeil & Willig, 2015). Bioacoustic monitoring with ARUs is not subject to such limitations: once deployed, units can simultaneously record data at multiple locations for long periods of time in the absence of observers, which can then be analysed by a single observer (Aide *et al.*, 2013; Klingbeil & Willig, 2015; Steer, 2010). A limitation of this monitoring method is that it lacks visual observations and is limited to acoustic detections (Klingbeil & Willig, 2015). Unless relationships between calling rates and species' densities are known, or microphone arrays are used, data produced by ARUs provides an index of relative abundance (Efford *et al.*, 2009; Marques *et al.*, 2013; Stevenson *et al.*, 2015). The reliability of this data is dependent upon the assumption that vocal activity is positively related to population size (Royle, 2004). This considered, the acoustic data produced by such recording devices provides a means to quantitatively assess the validity of the claim that forests 'fall silent' shortly following aerial 1080 operations (Cook, 2017).

Cook (2017) tested the "silent forest" theory utilising ARUs, comparing the conspicuousness of bird species in non-treatment and treatment areas of a cereal-bait operation carried out in the Aorangi Range of the lower North Island in 2014. The study found no evidence for a community-level decline in birdsong in treatment areas after the operation. At the population-level, one species, the tomtit (*Petroica macrocephala*), showed evidence for a decline in vocal conspicuousness in treatment areas. This species has historically suffered significant mortality following operations using carrot baits, particularly those with small unscreened bait fragments, or 'chaff' (Powlesland *et al.*, 1998; Powlesland *et al.*, 2000; Spurr & Powlesland, 1997). Such a negative effect has the potential to detrimentally affect the species' population if it is consistent across independent treatments of a recurring treatment regime (Powlesland *et al.*, 2000; Veltman & Westbrooke, 2011). However, studies suggest that modern cereal-based baiting practices show negligible impacts for tomtit populations (Greene *et al.*, 2013; Powlesland *et al.*, 2000; Westbrooke & Powlesland, 2005). Thus, though plausible, this result is somewhat at odds with the wider scientific literature and warrants further study.

The current study sought to build on the work of Cook (2017) and further address the 'silent forest' theory, utilising ARUs to monitor the short-term community- and species-level

trends in bird song around a further two aerial 1080 operations occurring in the lower North Island. Specifically, its aims were to determine if, relative to non-treatment areas: (1) prevalence of silence in recordings from treatment areas increased after operations, and; (2) the calling prevalence of any species showed a significant decline in treatment areas post-treatment.

2.2. Methods

2.2.1. Study areas and study sites

Sound recordings were analysed from recordings made by Department of Conservation (DOC) ARUs deployed across seven study sites in three study areas within the Wellington region: the Aorangi, Northern Rimutaka and Southern Rimutaka Ranges. Within a study area, between one and four study sites were utilised (Figure 2.1). A study site comprised between three and nine recorders that could be serviced in one day (i.e. have their batteries and SD cards replaced) by a pair of field technicians or students. Vegetation in the area is representative of the southern North Island, consisting mainly of a mix of beech, broadleaf, and mixed beech/broadleaf forest, with podocarps scattered throughout. Beech (*Fuscospora* spp.) tends to dominate with increasing elevation, and in areas with dry climates and infertile soils (Dymond & Shepherd, 2004). Regenerating mānuka/kānuka scrub also exists in the region where vegetation has been historically modified by burning and pastoralism (Dymond & Shepherd, 2004; Wardle, 1967), and is prevalent at one study site monitored.

The Wellington region has a mild, wet climate (Dymond & Shepherd, 2004); mean monthly temperatures near sea level range from 8 °C (July) to 16 °C (January), and rainfall varies from 800 mm to 7000 mm. There is an east-to-west rainfall gradient in the area, with areas in the west receiving more rainfall (Cook, 2017; Dymond & Shepherd, 2004).

2.2.2. 1080 operations

The Aorangi Range currently receives aerial 1080 drops for possum control on a three-yearly cycle, most recently receiving a drop in June 2017. The range was subject to pre-feed application of non-toxic 6 gram cereal baits applied in 260 m swaths on 30-31 May 2017, followed by a toxic drop of 12 gram 0.15% toxicity cereal baits applied full-broadcast in 300 m swaths on 16–17 June 2017. Baits used deer repellent and were applied at 1.5 kg/ha (OSPRI, pers. comm.). Previous to this operation, the Aorangi Range last received aerial 1080 treatment in August 2014. This operation was monitored by Cook (2017).

The Southern Rimutaka Range area received its first and most recent aerial 1080 treatment in July 2017. Pre-feed of non-toxic 6 gram cereal baits occurred on June 16 2017, followed by a drop of 12 gram 0.15% toxicity cereal baits in 180 m swaths on July 30 2017. Both operations used deer repellent and were applied full-broadcast at 2 kg/ha (OSPRI, pers. comm.).

The Northern Rimutaka Range area did not receive 1080 treatment throughout the course of this study and served as a control site for both treatments. This area last received aerial 1080 treatment in August 2012 (Uys & Crisp, 2018).

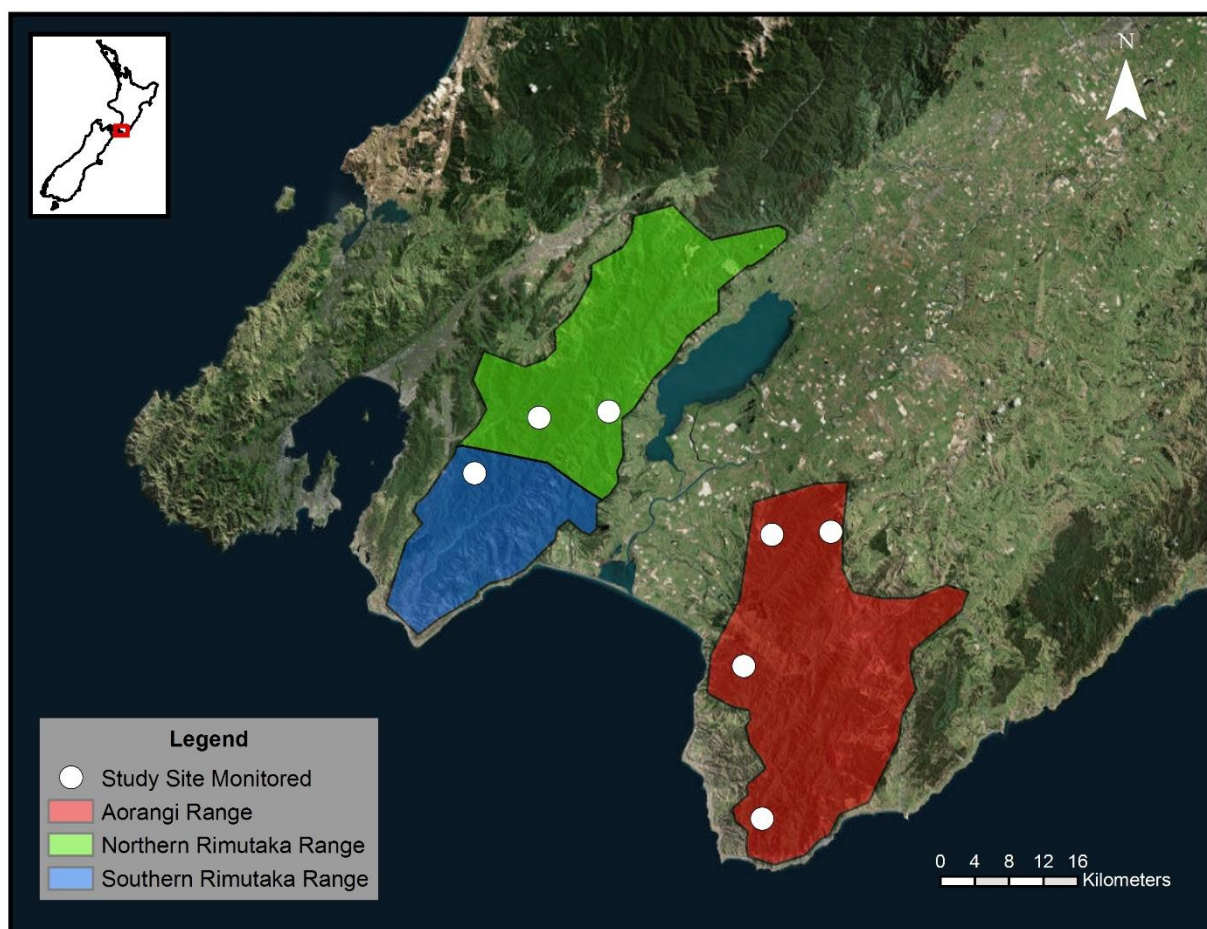


Figure 2.1 Map showing study sites and approximate forested extent of focus study areas.

2.2.3. Experimental design

Both studies employ a BACI (Before-After/Control-Impact) experimental design (Underwood, 1992), utilising ARU recordings to compare changes in the calling prevalences of native diurnal bird species in treatment and non-treatment areas from before to after respective 1080 operations.

2.2.3.1. 2017 Aorangi operation

Recordings were analysed for 13 ARUs distributed across four study sites in one treatment area, the Aorangi Range, and 18 ARUs distributed across two study sites in one non-treatment area, the Northern Rimutaka Range (Figure 2.2). Six recorders (two in the treatment area and four in the non-treatment area) were spaced 350 m from the next-nearest recorder. All other recorders were spaced at least 400 m apart.

2.2.3.2. 2017 Southern Rimutaka operation

Initially, 18 ARUs were distributed across two study sites in one treatment area, the Southern Rimutaka Range, based on proposed treatment boundaries. However, one treatment site was outside of/proximal to the treatment boundary after treatment was carried out. Furthermore, two recorders failed in the site remaining within the final treatment area. Thus, recordings were analysed for 7 ARUs distributed across one study site in the treatment area. Eighteen ARUs were distributed across two study sites in one non-treatment area, the Northern Rimutaka Range. However, one recorder failed during monitoring. Thus, recordings were analysed for 17 ARUs distributed across the non-treatment area (Figure 2.3). Three non-treatment recorders were spaced 350 m from the next-nearest recorder. All other recorders were spaced at least 400 m apart.

2.2.4. Audio recording specifications

Audio recordings were made using 2013 and 2016 DOC ARUs. Devices were set to record at high frequency, recording sound frequencies of up to 16 kHz at 32,000 samples per second. DOC ARUs record with one microphone unit (i.e. in mono) and save recordings as 16-bit .wav files. Recorders were mounted to tree trunks approximately 1.5 m above the ground and situated away from rivers where possible. Some rivers increase in flow, and thus noise, following periods of rainfall. If proximal to a river, recorders were mounted on the side of the tree opposite to the river to reduce attenuation of such river noise in recordings.

Recorders were set to record simultaneously across all locations to minimise the effects of intra-day variability in species' calling prevalences. Recording spanned 30 minutes from 0800-0830 h (NZST) for a period of at least six weeks before and six weeks after each 1080 operation's application of toxic baits.

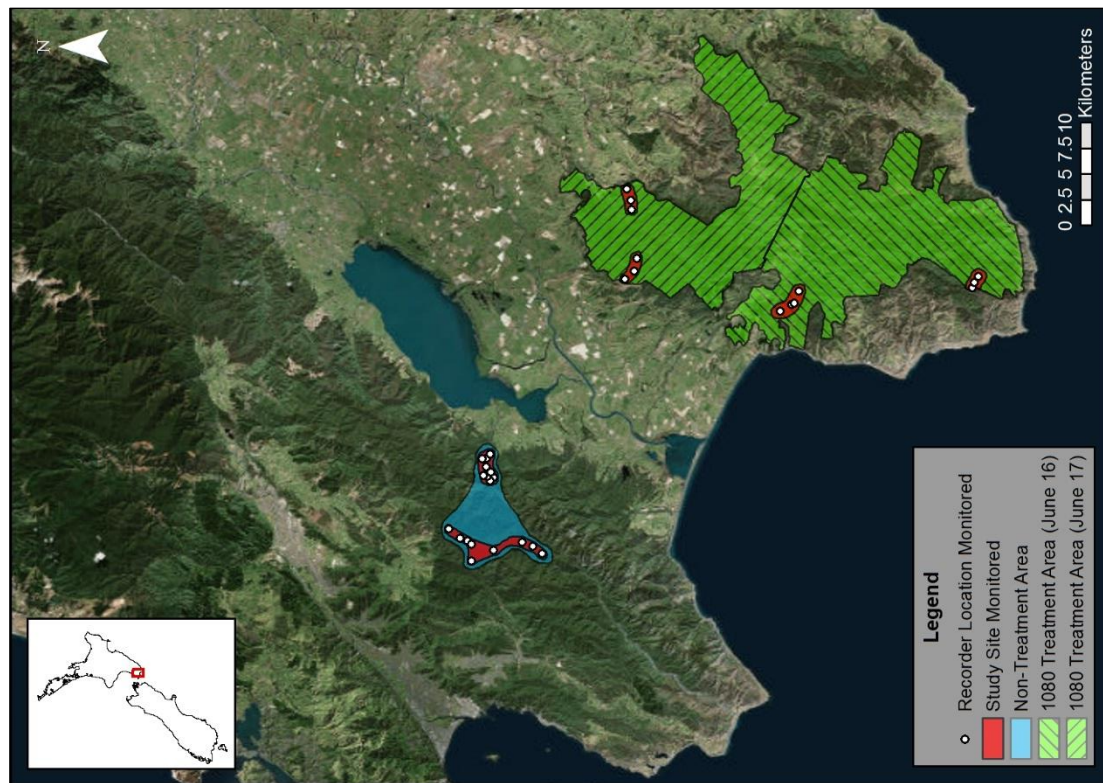


Figure 2.2 Map of study areas, study sites, and recorder locations monitored over the June 2017 Aorangi 1080 operation.

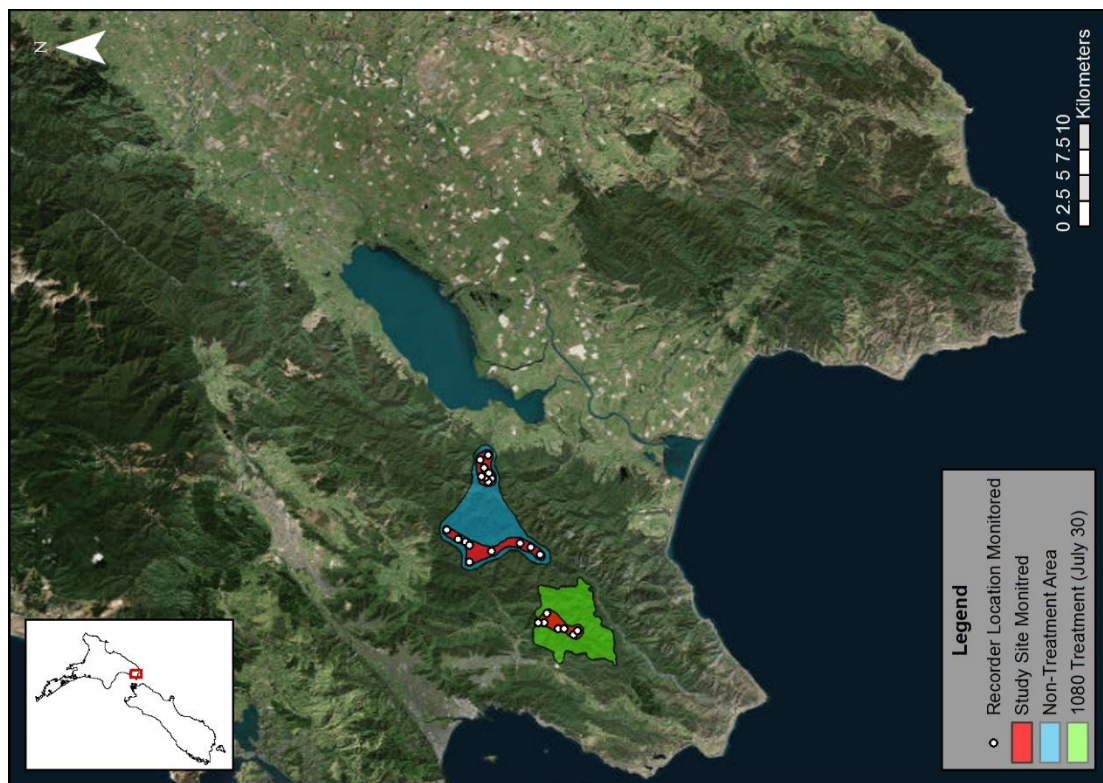


Figure 2.3 Map of study areas, study sites, and recorder locations monitored over the July 2017 Southern Rimutaka 1080 operation. The Southern Rimutaka Range received a wider treatment: only drop boundaries relative to the study area monitored are shown.

2.2.5. Audio recording selection

Recordings were analysed for five-week periods both before and after respective 1080 operations, with one recording analysed per sample week. For each respective drop, ‘before’ periods consisted of the five weeks leading up to the date of the toxic drop of interest. One week without sampling separated ‘before’ and ‘after’ sampling periods. Thus, ‘after’ periods sampled spanned weeks two to six after each respective drop. Sample days (one per week) were selected from candidate days within these periods that were identified as having suitable weather conditions for analysis.

Identification of candidate days comprised two phases: primary inspection of weather data, and subsequent inspection of recording spectrograms from days passing the primary weather filter. For each sampling period, rainfall (cumulative mm, 0800-0900 h NZST) and wind speed (km/h at 2.5-10 m, 0800 h NZST) data was inspected as the average of five rainfall and five wind speed weather stations in the wider sampling area (Greater Wellington Regional Council, 2018) (Appendix I). Days with mean wind speeds of <15 km/h and 0 mm of cumulative rainfall were identified as potentially suitable for sampling. The full spectrograms of recordings from these days were then inspected using Raven Lite 2.0 (Bioacoustics Research Program, 2016) and classified as suitable for analysis if neither heavy rain nor consistent strong wind were present. If present, light wind or rain was also noted. In some cases, recordings were not present where an ARU recording failed; such recordings were classified as unavailable for analysis.

Sample days were selected from candidate days primarily to maximise sample size and recording quality: for each week, the day with the greatest number of suitable recordings and least recordings with light wind or rain was primarily chosen. However, if less than four days separated two sample days from consecutive weeks, the day with the second highest number of suitable recordings of the two weeks was instead chosen for analysis for its respective week in the interest of increasing temporal independence between successive samples. Significantly windy or rainy recordings were scarce in the days chosen following this method. Thus, all recordings available for the sample days selected were analysed. This produced maximum sample sizes of 310 and 240 recordings for the Aorangi and Rimutaka operations respectively. A total of three recordings were unavailable for the chosen sample days due to recording failure, reducing final sample sizes to 308 and 239 recordings respectively (Appendix II).

2.2.6. Scoring audio recordings

Audio recordings were scored employing the ‘intermittent method’ described by Cook and Hartley (2018). This method independently scores the presence (1) or absence (0) of species’ calls in the first ten seconds of every minute in each recording. This produces a score between 0 and 30 for each species from 30 ten-second sub-samples sampled across 30 minutes. The score produced for each species in each recording analysed is converted to a calling prevalence, calculated as the score for the species of interest divided by 30 (the total number of sub-samples). For example, a species calling in 15 of 30 sub-samples analysed in a recording would have a calling prevalence of 0.50.

The sub-samples analysed using this method total a sampling effort of five minutes, equivalent to that of a five-minute bird count. When compared to scoring species’ presence and absence in thirty consecutive 10-second sub-samples of a five minute recording, the ‘intermittent method’ has been found to both increase the number of species detected, and decrease variance around calling rate estimates because it reduces the short-range temporal autocorrelation of data (Cook, 2017; Cook & Hartley, 2018).

For all analyses, recordings were given a traceable coded name such that the treatment-type and date of any recording being analysed was unknown to the observer (i.e. analyses were blind). Scoring of species’ presence/absence was carried out by one observer (Roald Bomans) in Raven Lite 2.0 using Panasonic RP-HC200 headphones with the noise cancelling function disabled.

2.2.6.1. Species-level

Recordings were scored by simultaneously listening to recordings and visually inspecting their spectrograms. A call was scored as present in a sub-sample if it could be both heard in the recording at maximum volume and seen on the spectrogram. Where possible, calls were identified to one of 17 focal species known to inhabit the areas studied (Table 2.1). If a call could not be confidently identified to a species after listening to a sub-sample five times, the call was classed as an ‘unknown’. When present in a sub-sample, wingbeats were also classified as a presence for kererū due to the species’ infrequent calling.

The calls of bellbird (*Anthornis melanura*) and tūī (*Prosthemadera novaeseelandiae*) are similar and may influence the reliability of results gained from acoustic identification alone (Mortimer & Greene, 2017). To account for the potential influence of misidentifications, tūī or

bellbird calls were given one of three classifications: calls that could be confidently classified as bellbird or tūi were classified as their respective species, whereas calls that could not be confidently classified were grouped (bellbird/tūi).

Blackbird (*Turdus merula*) and song thrush (*Turdus philomelos*) calls can also be similar and difficult to distinguish (Miller & Anderson, 1992). As this study was primarily focused on native bird species' responses, calls for these species were grouped (blackbird/thrush, i.e *Turdus* sp.).

Table 2.1 List of 17 focal species monitored in acoustic recordings. Asterisks (*) indicate exotic species.

Common/ Māori name	Species name
Bellbird/Korimako	<i>Anthornis melanura</i>
Blackbird*	<i>Turdus merula</i>
Chaffinch*	<i>Fringilla coelebs</i>
Dunnock*	<i>Prunella modularis</i>
Eastern Rosella*	<i>Platycercus eximius</i>
New Zealand Falcon/Kārearea	<i>Falco novaeseelandiae</i>
Fantail/Pīwakawaka	<i>Rhipidura fuliginosa</i>
Grey warbler/Riroriro	<i>Gerygone igata</i>
Kākā	<i>Nestor meridionalis</i>
Red-Crowned Parakeet/Kākāriki	<i>Cyanoramphus novaezelandiae</i>
Rifleman/Tītipounamu	<i>Acanthisitta chloris</i>
Silvereye/Tauhō	<i>Zosterops lateralis</i>
Song Thrush*	<i>Turdus philomelos</i>
North Island Tomtit/Miromiro	<i>Petroica macrocephala toitoi</i>
Tui	<i>Prothemadera novaeseelandiae</i>
Whitehead/Pōpokatea	<i>Mohoua albigilla</i>
New Zealand Woodpigeon/Kererū	<i>Hemiphaga novaeseelandiae</i>

2.2.6.2. Community-level

A community-level parameter of silence was calculated for each acoustic recording analysed to assess if treatment area bird communities ‘fell silent’ post-1080 operations relative to non-treatment areas. If a 10-second sub-sample contained no calling events for any bird species, it was classified as ‘silent’. This prevalence of ‘silent’ sub-samples was analysed in the same manner as the calling prevalences of focal species.

2.2.7. Statistical analyses

For each 1080 operation, statistical analyses were carried out for the amount of ‘silence’, and for the calling prevalences of those species found to be both calling in at least 1% of all sub-samples analysed across all counts; and present in at least half of both treatment and non-treatment recording locations for the respective operation. This filter removed very rare and/or highly localised bird species from the analysis.

For each sample, the treatment type (1080 or non-1080), time period (before or after 1080), study site, recorder location, and recording date were recorded. For silent sub-sample prevalences and calling prevalences of each species satisfying prevalence requirements, a logistic mixed effects model was fitted in R (R Core Team, 2017) using the lme4 package (Bates *et al.*, 2015) with the formula:

$$\text{Species}_x \text{ Calling Prevalence} \sim \text{Treatment Type} * \text{Time Period} \\ + (1|\text{Study Site/Recorder Location}) + (1|\text{Date})$$

The dependent variable was entered as a single value proportion (rather than 30 ones and zeroes) weighted by 30, consistent with the number of sub-samples from which each calling prevalence was calculated.

It was of interest whether or not changes in calling prevalences from before to after 1080 operations differed between treatment and non-treatment sites. The BACI design of the experiment allowed the comparison of these changes; treatment type and time period were fitted as crossed fixed factors to investigate any effect of treatment on calling prevalences from before to after operations. Specifically, a decline in a species’ calling prevalence or an increase in silence at the community level in the treatment area relative to the non-treatment area would support the theory that forests ‘fall silent’ after an operation at the species or community level. Study site, recorder location and recording date were fitted as random factors to control for potentially influential, yet unmeasured, environmental factors that vary day-to-day and from

place-to-place. Recorder location was nested within study site to account for the spatially nested structure of the sampling design (Figures 2.2 and 2.3).

A type III ANOVA was carried out on the resulting models in R using the car package (Fox & Weisberg, 2011) to produce a chi-square value for the interaction term. A non-parametric permutation test was applied to approximate the null chi-square distribution and test significance for each chi-square value observed. Permutations were made across time periods and within recorder locations. Any observations missing due to recording failure (NAs) were held constant in their location within the dataset. Five thousand permutations were made to reliably test significance at the 1% significance level (Anderson, 2001). Example input-data structures and corresponding chi-square distributions are available in Appendix III.

2.3. Results

2.3.1. Community-level

Both treatment and non-treatment sites showed a general increase in the prevalence of silent sub-samples for the Aorangi 2017 operation (Figure 2.4a). There was no significant interactive effect between treatment type and time period on the prevalence of silent sub-samples for this operation ($p = 0.469$, $\chi^2 = 2.699$). Contrastingly, there was a significant interactive effect between treatment type and time period on the prevalence of silent sub-samples for the Southern Rimutaka 2017 operation ($p < 0.001$, $\chi^2 = 139.2645$). The treatment site showed a decrease in silence from before to after the operation, whereas non-treatment sites showed an increase (Figure 2.4b). This suggests that, relative to pre-treatment levels, bird song became more frequent in the treatment site two to six weeks post-treatment.

2.3.2. Species-level

2.3.2.1. Aorangi operation

Eleven taxa satisfied prevalence requirements for statistical analyses in the Aorangi operation study. However, only ten are presented here as permutation models for kererū suffered a high rate of convergence failures. Of the results presented, eight showed an insignificant interaction between treatment type and time period at the 5% significance level, with treatment type having no significant effect on changes in calling prevalences from before to after treatment (Figure 2.5a-e, h-j).

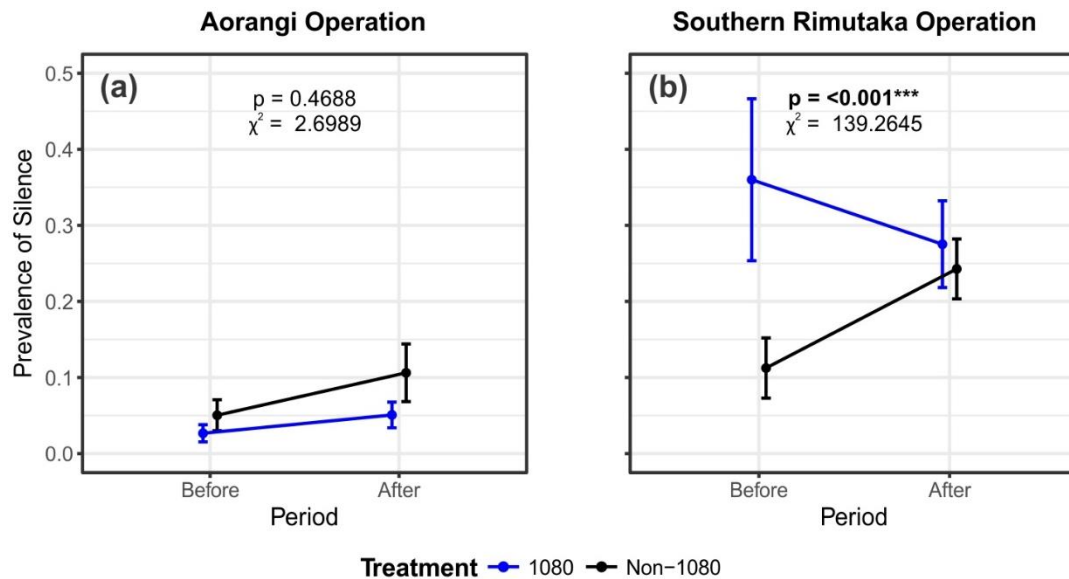


Figure 2.4 Grand mean prevalences of silent sub-samples ± 1 SE (n = number of recorder locations monitored) in treatment and non-treatment areas before and after (a) the June 2017 Aorangi and (b) July 2017 Southern Rimutaka aerial 1080 operations. Mean and SE estimates are an approximate representation of interactions, whereas associated chi-square and p-values are presented from permutation tests accounting for the spatially-nested and temporally replicated nature of the experimental design. Asterisks (***) indicate $p < 0.001$, plots share y-axes.

A significant interaction between treatment type and time period was found for the calling prevalence of both bellbird ($p = 0.0472$, $\chi^2 = 14.8135$) and chaffinch ($p = 0.0432$, $\chi^2 = 2.5475$). Mean calling prevalences decreased from before to after 1080 treatment in both treatment and non-treatment sites for both of these species (Figure 2.5f, g). For bellbird, however, this decrease was more substantial in non-treatment sites (from 0.0691 ± 0.0192 to 0.0292 ± 0.0121) when compared to treatment sites (from 0.1036 ± 0.0200 to 0.0805 ± 0.0153). The opposite interaction held true for chaffinch calling prevalences, which showed a greater decrease in treatment sites (from 0.1482 ± 0.0139 to 0.0733 ± 0.0131) than in non-treatment sites (from 0.0792 ± 0.0146 to 0.0481 ± 0.0095).

2.3.2.2. Southern Rimutaka operation

Nine taxa satisfied prevalence requirements for statistical analyses in the Southern Rimutaka operation study. However, only eight are presented here as permutation models for tūī suffered high rates of convergence failures. Of these eight taxa, six showed no significant interaction between treatment type and time period at the 5% significance level; there was no significant difference in changes in calling prevalences from before to after the 1080 treatment between treatment and non-treatment sites for these species (Figure 2.6a, b, d, f-h).

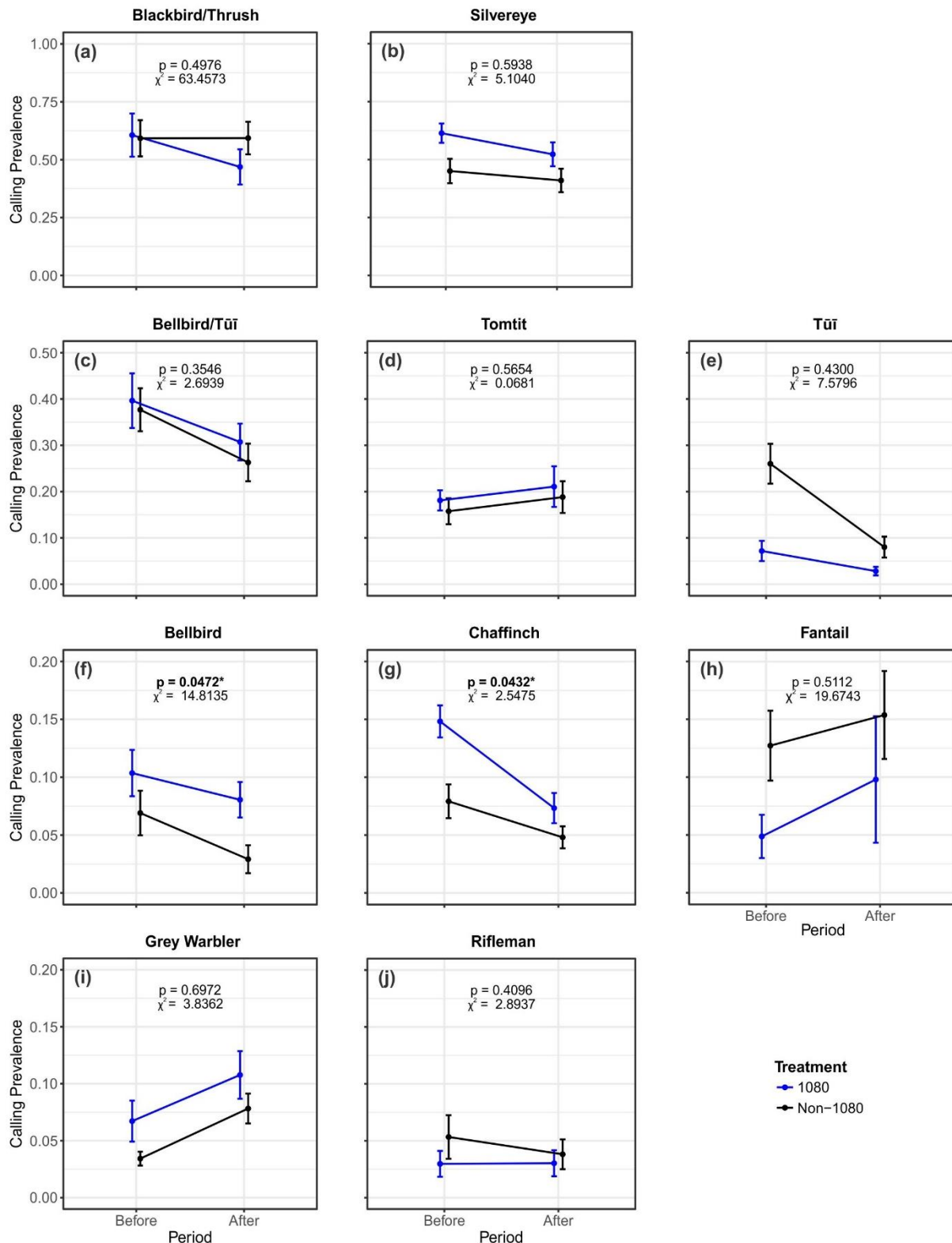


Figure 2.5 Grand mean calling prevalences \pm 1 SE (n = number of recorder locations sampled) from zero to five weeks before to one to six weeks after 1080 treatment for treatment and non-treatment areas monitored across the Aorangi operation, and associated chi-square and p-values. Mean and SE estimates are shown only for species satisfying statistical analysis requirements, and are an approximate representation of interactions, whereas statistics were calculated via a permutation test accounting for the spatially-nested and temporally replicated nature of the experimental design. Columns and rows of plots share x and y axes respectively. Asterisks (*) indicate $p < 0.05$.

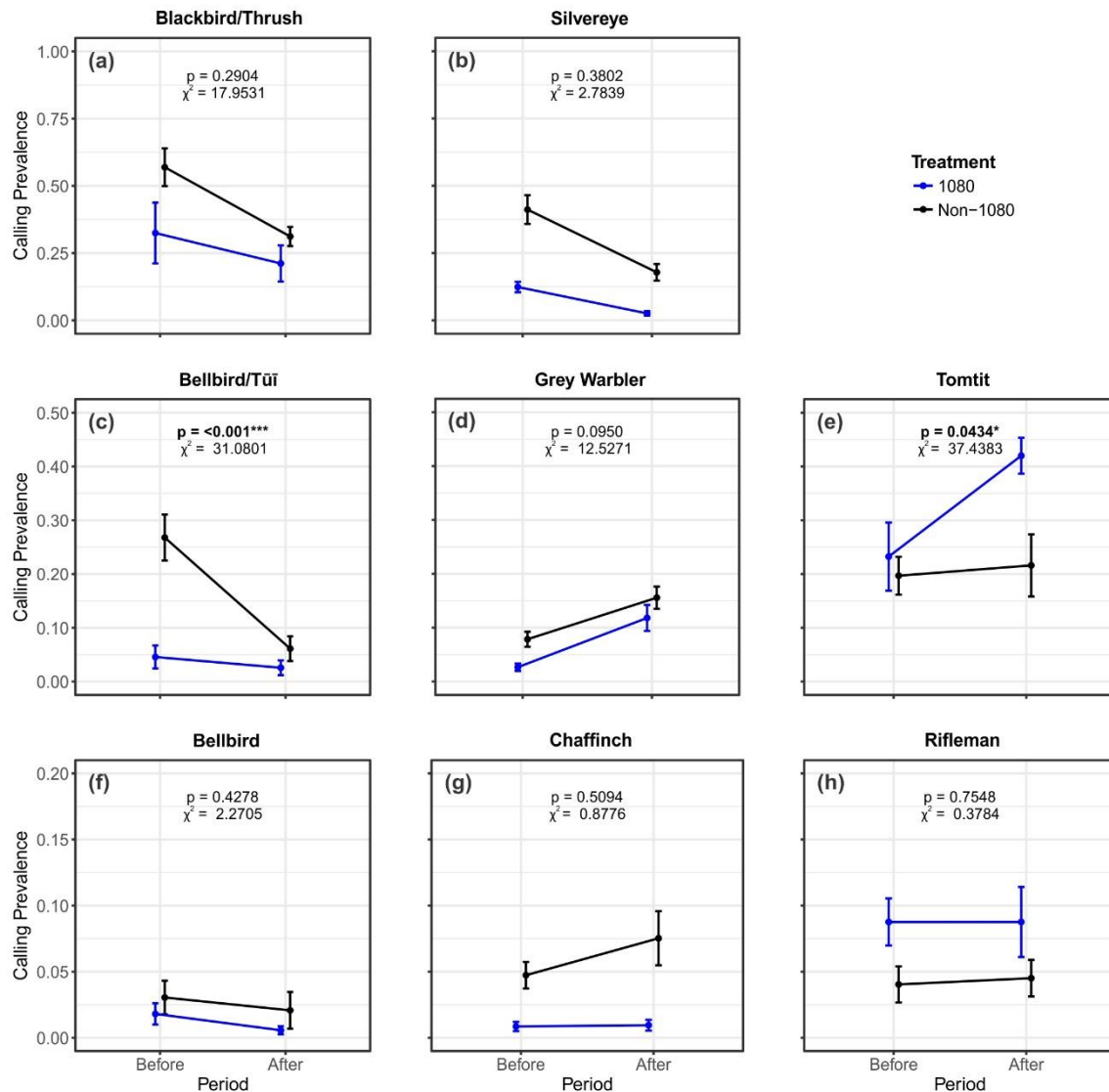


Figure 2.6 Grand mean calling prevalences \pm 1 SE (n = number of recorder locations monitored) zero to five weeks before and one to six weeks after 1080 treatment in treatment and non-treatment areas monitored across the July 2017 Southern Rimutaka operation, and associated chi-square and p-values. Mean and SE estimates are shown for species satisfying statistical analysis requirements, and are an approximate representation of interactions, whereas associated statistics were calculated via permutation tests accounting for the spatially-nested and temporally replicated nature of the experimental design. Columns and rows of plots share x and y axes respectively. Asterisks (*, ***) indicate $p < 0.05$ and 0.001 respectively.

A significant interaction was found for bellbird/tūī (*A. melanura*/*P. novaeseelandiae*) and tomtit (*P. macrocephala*). Bellbird/tūī showed a significant decrease in calling prevalence in non-treatment sites relative to treatment sites ($p < 0.001$, $\chi^2 = 31.0801$): mean calling prevalence in non-treatment sites decreased from 0.2678 ± 0.0428 to 0.0612 ± 0.0230 , whereas mean calling prevalence in treatment sites decreased from 0.0457 ± 0.0215 to 0.0257 ± 0.0138 (Figure 2.6c). Contrastingly, tomtit showed a significant increase in calling prevalence in

treatment sites relative to non-treatment sites ($p < 0.0434$, $\chi^2 = 37.4383$): mean calling prevalence nearly doubled from 0.2324 ± 0.0633 to 0.4200 ± 0.0335 in treatment sites, whereas mean calling prevalence only increased slightly from 0.1969 ± 0.0351 to 0.2161 ± 0.0577 in non-treatment sites (Figure 2.6e).

2.4. Discussion

At the community-level, the results of this study show no supporting evidence for the claim that forests ‘fall silent’ following aerial 1080 treatment. Monitoring across the Aorangi operation showed an increase in silence post-treatment that did not differ significantly between treated and non-treated sites. This result corroborates with results found by Cook (2017), which indicated no major change in the prevalence of silence in both treated and non-treated areas following the previous Aorangi Range aerial 1080 operation in 2014. Though patterns differed significantly between treatment and non-treatment sites in the Southern Rimutaka operation, birdsong actually increased in the treatment site following the operation, whereas silence increased in non-treatment sites. This result provides evidence contrary to the silent forest theory. These results support mortality and bird count data analysed for 15 operations by Morriss *et al.* (2016), which collectively suggest that modern aerial 1080 operations pose a negligible threat to native forest bird communities.

At the population-level, only chaffinch calling prevalences exhibited a significant decline in treatment sites relative to non-treatment sites. This was observed only after the 2017 Aorangi operation. Chaffinch are an introduced, granivorous species that, given their diet and thus likelihood to interact with baits, has been considered at risk to 1080 poisoning (Miller & Anderson, 1992). In a study of 15 aerial 1080 operations occurring between 2003 and 2014, 8 of 81 bird corpses found were of chaffinches. Five of these were analysed for 1080 residues, four of which tested positive (Morriss *et al.*, 2016). Thus, it is possible that the significant decline observed in this species’ calling prevalence may reflect a significant mortality rate resulting from this operation. Such a decline was not observed over the 2014 operation in the same area monitored by Cook (2017). This inconsistency may result from varying factors affecting the exposure rate and longevity of toxic baits from one operation to another, such as rainfall received in treatment areas following the operation (Booth *et al.*, 1999); differing sowing methods; and differing sowing rates (Westbrooke *et al.*, 2003). The concentration of 1080 in cereal baits has been found to decrease with rainfall, with 0.15% 6 g baits rapidly degrading after 10 mm of rainfall and halving in toxicity after ~60-70 mm of rainfall (Booth *et*

al., 1999). Based on data from the five rainfall stations sampled in sample day selection (Appendix IV), rainfall did not differ markedly between the Aorangi 2014 and Aorangi 2017 operations; mean cumulative rainfall exceeded 60 mm 31 and 25 days after these operations respectively. However, baits were strip sown at 1.0 kg/ha in the 2014 Aorangi operation, whereas baits were full broadcast at 1.5 kg/ha in the 2017 Aorangi operation. Thus, it is possible that chaffinch populations in the 2017 Aorangi operation had higher exposure rates and suffered significant mortality as a result of the higher application rate and full-broadcast method employed in the operation. Bait application was both full-broadcast and highest in the Rimutaka 2017 operation (2.0 kg/ha). However, in contrast to the Aorangi 2014 and 2017 operations, mean cumulative rainfall in the five weeks following this operation exceeded 60 mm after just 14 days. It is plausible that exposure after this operation might have been significantly reduced by this rainfall. These differing conditions may explain the differing patterns observed across these treatments.

The non-significant result for tomtit in the Aorangi 2017 operation is promising considering the negative impact suggested by monitoring over the 2014 operation (Cook, 2017). Tomtit have high reproductive success under low-predator conditions and have been shown to benefit in years following aerial 1080 treatment both in other studies (Knegtmans & Powlesland, 1999; Peterson, 2014), and after the 2014 Aorangi operation over a 1.5 year period (Fea, 2018). With a high reproductive rate, the species has been classed as resilient with a good potential to recover from any significant population losses resulting from predator-control operations (Powlesland *et al.*, 2000; Spurr, 1979). These factors considered, the results of the Aorangi 2017 operation suggest little risk of a net-negative effect of aerial 1080 operations on tomtit populations in the Aorangi Range resulting from treatment repeated on a three-yearly basis.

With respect to the Southern Rimutaka operation, the increase in prevalence of tomtit calling observed in treatment areas cannot reflect a benefit from recruitment, given that post-treatment monitoring occurred directly after the operation (August-September) and before the breeding season (Atkinson *et al.*, 1995). Spurr and Powlesland (2000) note that, in territorial bird species, mortality of territory holders may result in floating individuals establishing new territories and, consequently, an increase in calling rates. Tomtit are a highly territorial species year round (Knegtmans & Powlesland, 1999; Spurr & Powlesland, 2000) and are known to have died following aerial 1080 operations using both carrot and cereal baits, sometimes suffering high mortality rates (Morris *et al.*, 2016; Powlesland *et al.*, 1998; Powlesland *et al.*,

2000; Spurr & Powlesland, 1997). Thus, it is possible that the observed increase indicates a potential negative effect of aerial 1080 treatment on tomtit populations in the Southern Rimutaka Range. However, studies suggest that mortality rates have significantly decreased with changing bait practices, with little negative impact following cereal drops (Greene *et al.*, 2013; Westbrooke *et al.*, 2003; Westbrooke & Powlesland, 2005). An increase in calling prevalence might also occur without any change in population size if birds perceive a decrease in predator presence following treatment, and thus a reduced risk in predation from calling behaviour (Lima, 2009). To resolve which of these possibilities explains the increase in calling prevalence for tomtit, a study radio-tracking individuals over a 1080 operation would provide more definitive information. Such studies, however, are also more time-consuming than a community-wide ARU survey.

It is also possible that the increase observed may reflect sampling design limitations of monitoring around the Southern Rimutaka operation. It is important that treatment and non-treatment areas are similar in topography when monitoring species' conspicuousness (Spurr & Powlesland, 2000). Originally, two study sites were set up in the treatment study area for monitoring over the Rimutaka operation. Nine of these recorder locations were stationed in the Catchpool catchment at 40-150 m elevation, and nine were stationed in Turere Valley at 480-700 m elevation. Similarly, the two non-treatment sites monitored in the Northern Rimutaka study area had ten recorder locations at 60-280 m elevation and eight at 450-760 m. However, final drop boundaries resulted in all treatment recorders at 40-150 m elevation being either outside the treatment area, or too proximal to the drop boundary to be reliably classed as receiving the effects of the treatment. Furthermore, two remaining treatment recorders failed. Consequently, all treatment recorder locations monitored for this operation were at higher elevations (>500 m), with a maximum between-recorder elevational difference of ~200 m.

It has been noted that tomtit breeding seasons may be affected by elevation, with nesting starting as much as a month later at higher elevations (520-740 m) than at lower elevations (0-200 m) (Knegtman & Powlesland, 1999). Accordingly, any changes in calling prevalence associated with the breeding season are also likely to be delayed at higher elevations. Over three years of monitoring, Michaux (2009) found tomtit mean calling prevalences at lower elevations (~180-300 m) in the Atuanui Scenic Reserve to consistently increase from June-July and decrease from July-August. Pre- and post-treatment monitoring of the Southern Rimutaka operation spanned from July to August. Thus, treatment sites monitored over the operation may have been experiencing a delayed seasonal increase in calling prevalence normally seen from

June-July at lower elevations. Had both high and low-elevation areas been sampled in the treatment area, it is possible that no significant difference in trends would have been observed between treatment and non-treatment areas. This elevational bias may also explain the difference in the prevalence of silence observed between treatment and non-treatment areas at the community level. Given this potential bias, results for this operation should be interpreted with caution. Recorder locations monitored across the Aorangi 2017 operation had greater between-recorder variation in elevation both within sites (max ~320 m) and between sites (~490 m). Thus, results from this operation are less likely to be subject to elevation-related biases. In light of this potential influence, further study of the impact of elevation on tomtit breeding ecology and behaviour would be of significant benefit to future monitoring studies.

The lesser decline in calling prevalence observed for bellbirds in treatment areas across the Aorangi operation is unlikely to reflect a change in conspicuousness associated with a significant mortality event (Spurr & Powlesland, 2000): no bellbird deaths have been reported following aerial 1080 operations using both old and modern baiting methods (Morriss *et al.*, 2016; Spurr & Powlesland, 1997). It is possible that this difference in decreases is associated with food-related population movements. Bellbird feed on invertebrates, fruit, and nectar. The species has been suggested to be largely nectarivorous, feeding primarily on nectar where it is available, and shifting from an insectivorous to nectarivorous diet in late winter as early-flowering species begin to blossom (Spurr *et al.*, 2011). Bellbird have been reported to move large distances to areas where nectar is available (Rasch & Craig, 1988), temporarily leaving territories in Kennedy's Bush, Christchurch and travelling at least 500 m to visit patchy resources of seasonally available flax and kowhai flower (Spurr *et al.*, 2010). Furthermore, opportunistic sightings of banded birds also suggest that individuals may make large seasonal elevational movements, travelling kilometres to move to lower areas in winter (Spurr *et al.*, 2010). Such movements have the potential to disproportionately affect detection probabilities. Depending on the distribution of nectar food resources across study sites, this may explain the greater decline in calling prevalence observed in non-treatment sites. If non-treatment sites monitored encompassed less nectar resources relative to treatment sites monitored, it is possible that declines associated with movements to these resources were exaggerated in these areas. The reported extent of nectar-related movements is limited and requires further study (Spurr *et al.*, 2010). This would greatly benefit interpretation of results for future BACI studies monitoring bellbird populations around pest control operations.

Food-related population movements may also explain the significant decline in bellbird/tūī calling prevalence observed across the Southern Rimutaka operation. Tūī are highly mobile (Pierce & Montgomery, 1992) and have been reported to travel large distances to reach nectar resources (Rasch & Craig, 1988; Stewart, 1980 as cited by Craig *et al.*, 1981). Studies have found closely-related family groups in scenic reserves interspersed with urban areas to follow regular seasonal foraging routes and travel large distances in search of food outside of summer, with male foraging ranges reaching 30 km in winter months (Bergquist, 1985, 1989). Pierce and Montgomery (1992) note the potential influence of such seasonal movements on increases in tūī observed in five minute bird counts from before to after a 1080 cereal bait operation in 1990. Spurr and Powlesland (1997) report only one tūī to have been found dead following a carrot bait operation. Similarly, only one tūī has been found dead following operations using modern baiting methods, with 1080 residue testing showing levels below 0.001 mg kg⁻¹ (Morris *et al.*, 2016). Thus, the greater decline of tūī/bellbird calling in non-treatment areas across the Southern Rimutaka operation is unlikely to reflect changes in conspicuousness associated with significant poisoning events, and may be associated with nectar-related movements of one or both of these species. However, inferences that can be drawn from the grouped-species data in this study are limited given the variations that exist in the diets of these species (Rasch & Craig, 1988; Spurr *et al.*, 2011) and are thus likely to exist in their behavioural ecology and exposure risk.

These factors considered, the results of this study suggest no evidence for a decline in birdsong at the community-level following aerial 1080 treatment. Furthermore, the chaffinch, the only species to show a marked decline in calling prevalence in treated areas post-treatment across either of the operations monitored, is an introduced granivorous species known to suffer population losses from aerial 1080 operations. Though other species did show differing trends in calling prevalence, none suggested a decline in treatment areas relative to non-treatment areas consistent with the “silent forest” theory. Any difference in trends between treatment and non-treatment areas (both positive and negative) may potentially indicate a detrimental impact of treatment. However, these changes might be more likely explained by sampling limitations and differential effects of population movements related to patchy and ephemerally available food resources. These results highlight the importance of well-matched, balanced sampling designs and the importance of understanding the ecology and behaviour of species in the interpretation of bioacoustic studies. Study of the potential for such behavioural ecology to influence bioacoustic monitoring results would be of major benefit to future monitoring studies

employing bioacoustics. Collectively, the results of this study suggest that forests did not ‘fall silent’ in the Aorangi or Southern Rimutaka Ranges following aerial 1080 operations neither at the community-level, nor at the species-level for any of the native species monitored. These results corroborate studies of modern aerial 1080 operations, which suggest a negligible threat from modern operations at the community level, and little threat at the individual level for native forest bird species (Morriss *et al.*, 2016).

**Chapter 3 - Achieving parsimony in developing a
template-based automated acoustic detector – an
applied study with morepork (*Ninox
novaeseelandiae*), a New Zealand owl**

3.1. Introduction

Monitoring acoustically active species with autonomous recording units (ARUs) provides many advantages as a monitoring method over traditional field surveys, particularly in terms of the spatial and temporal scale and reliability of data that can be collected (Digby *et al.*, 2013; Joshi *et al.*, 2017; Steer, 2010). However, this monitoring method is not without limitation; spectrographic scoring of recordings requires analysis that can demand significant amounts of time investment when monitoring occurs over longer periods of time (Digby *et al.*, 2013; Kogan & Margoliash, 1998) or with a larger number of recording units. This issue will only become more pertinent as recorders decrease in cost and increase in capacity, further-facilitating large-scale monitoring. Consequently, it is widely accepted that automated analysis of acoustic recordings for species recognition has the potential to play a major role in the analysis of growing sound file databases (Digby *et al.*, 2013; Katz *et al.*, 2016a; Ulloa *et al.*, 2016).

Significant progress has been, and continues to be, made in this field. Machine learning algorithms such as Hidden Markov Chains and Artificial Neural Networks have demonstrated high accuracy and have been proposed as the best approach to automated species identification (Acevedo *et al.*, 2009). However, such approaches require large amounts of training data (Acevedo *et al.*, 2009; Towsey *et al.*, 2012), require expert knowledge for development (Joshi *et al.*, 2017), and may add unnecessary complication to the detection of simplistic call-types (Brandes *et al.*, 2006). Template matching methods, in contrast, require less training data and are suitable where the species of interest has a more simplistic and limited call repertoire (Brandes, 2008). This method takes call templates (see Table 3.1) and compares them with signals within set frequency bounds from field recordings. Where variation exists between or within a species' call(s), a suite of templates representative of this variability can be selected and used in combination as a detector (see Table 3.1) to maximise performance (Katz *et al.*, 2016a).

The set of templates selected has a significant impact on detection results (Joshi *et al.*, 2017; Katz *et al.*, 2016a) and the performance of such detectors can increase moderately relative to computation time with increasing size of the template set (Anderson *et al.*, 1996; Kogan & Margoliash, 1998). The benefit of employing an automated detector is dependent on the time required to create, run and evaluate it relative to the time that would be required for equivalent manual analysis (Knight *et al.*, 2017). Thus, an efficient means to produce a simple

and effective (i.e. parsimonious) detector is desirable, particularly where a large number of potential templates are available for detector development. This considered, the process of developing detectors via template matching is described as iterative and subjective (Katz *et al.*, 2016a). The performance of individual templates or template sets may be difficult to predict subjectively, and iteratively and subjectively developing a good-performing detector from many potential templates may prove difficult and time-consuming. Additionally, high-quality exemplar calls are typically used for the development and training of automated detectors (Digby *et al.*, 2013; Joshi *et al.*, 2017). Such quality calls may not be available where exemplar calls are sourced from in-field recordings, and performance of detectors may in fact be better when developed with templates of a quality more similar to the data to be tested (Joshi *et al.*, 2017).

Table 3.1 Table of terms and definitions referred to in-text.

Term	Definition
Template	<i>An exemplar species call that can be used for the automated detection of a species' calling events in acoustic recordings.</i>
Detector	<i>A set of templates used in combination for the detection of a species' calling events in acoustic recordings.</i>
Block	<i>A section of an acoustic recording with defined time limits that may be 'active' (containing a calling event of interest) or 'inactive' (containing no calling event of interest).</i>

This study sought to establish and apply an objective and efficient method of developing a template-matching detector based on call recognition of the morepork/ruru (*Ninox novaeseelandiae*), a relatively common native New Zealand owl for which many field-quality exemplar calls were available. Little systematic monitoring has been carried out for morepork. However, ARUs show promise as monitoring method for the species, particularly with the partial or complete automation of call identification (Pryde & Greene, 2016). The morepork has a stereotyped and relatively simple call repertoire, making it a good candidate for template-based autonomous acoustic monitoring (Brandes, 2008). Furthermore, in many areas it is the main species that is acoustically active at night, thus the species has few potential sources of error in automated detection (Joshi *et al.*, 2017). Specifically, this study aimed to develop a generic method for creating a parsimonious detector using a range of templates sourced from

in-field recordings, minimising false positive detections and achieving the best true positive detection rate possible with the templates available when cross-validated against independent test data.

3.2. Methods

3.2.1. Background of detection software

Detector development was carried out on a 3.2 GHz desktop computer with 8 GB of RAM using the open-source R (R Core Team, 2017) package `monitor` (Hafner & Katz, 2017). Two methods of template matching are available in the `monitor` package: spectrogram cross-correlation, and binary point-matching. For both methods, signals in field recordings are given scores based on their similarity to the given template for specific metrics. Spectrogram cross-correlation templates consist of a matrix of amplitudes drawn from the template's spectrogram. The similarity of each frame in a given field recording to the template's matrix is scored between 0 and 1 using Pearson's correlation coefficient. Binary point-matching templates, in contrast, consist of a map of 'on' and 'off' points in a template. 'On' points exceed an amplitude set by the user and are expected regions of the signal-of-interest, whereas 'off' points fall below the set amplitude and are regions of anticipated non-signal/noise (Figure 3.1). Using this method, scores given to potential signals are derived using an equation adapted from Towsey *et al.* (2012), which finds the difference between the mean amplitudes of on and off points for the signal of interest (Katz *et al.*, 2016a, 2016b). This method is particularly useful in the isolation of faint calls from background noise (Towsey *et al.*, 2012).

For each template, a threshold can be set below which lower-scoring signals are classed as non-calling events. The threshold for a template is often chosen somewhat subjectively according to the purpose of the detector. The effect of changing the threshold is commonly

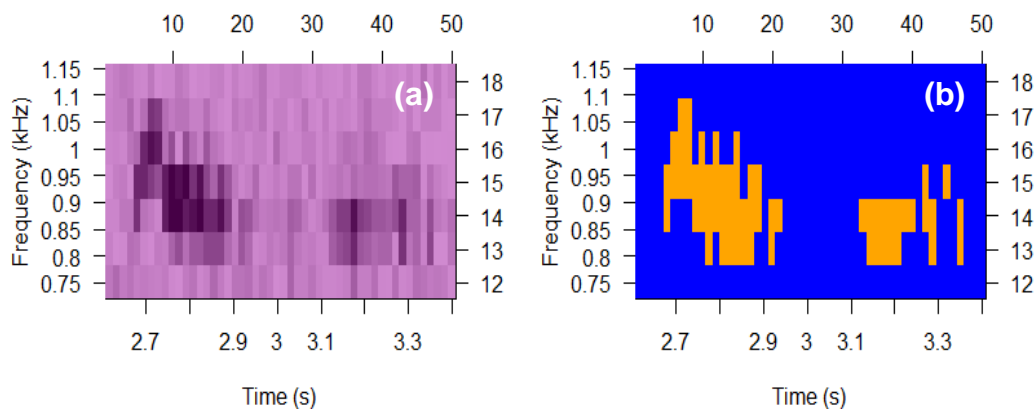


Figure 3.1 Example (a) cross-correlation and (b) binary templates for a potential call to be tested as a template. Darker areas in (a) represent higher amplitude points; yellow and blue areas in (b) represent on and off regions respectively.

explored using Receiver Operating Characteristic (ROC) analyses, plotting true-positive rates (TPRs/sensitivities) against false-positive rates (FPRs) derived from the relative rates of true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN) events (Fielding & Bell, 1997; Katz *et al.*, 2016a). A FP detection occurs when a template detects a call that is not actually present. Conversely, a FN event occurs when a template does not detect a call that is present. In the same manner, a TP detection occurs when a detector accurately detects a call, and a TN event occurs when a detector accurately makes no detection (Fielding & Bell, 1997; Florkowski, 2008) (Table 3.2, Figure 3.2).

Table 3.2 Summary confusion matrix of event possibilities in automated detection adapted from Florkowski (2008) and Fielding and Bell (2002). The presence or absence of calls is the ‘true’ state as determined by manual analysis of the recording in question. The prevalence of calls within a given dataset is defined as $TP + FN / (TP + FP + TN + FN)$.

		Actual		
		Call Present	Call Absent	Total
Predicted	Detection	True Positive (TP)	False Positive (FP)	TP + FP
	No Detection	False Negative (FN)	True Negative (TN)	TN + FN
	Total	TP + FN	TN + FP	

To determine an optimal threshold, a template of interest is run on a training set of recordings for which the times of calls have been manually determined (considered the ‘true’ state of calling and non-calling). ROC analyses can then be carried out to compare the suite of scores given by a template to the manually-verified truth and determine the rates of TP, FP, TN and FN detections at different thresholds. The relative rates of such events at different thresholds can be expressed by a range of performance metrics (Table 3.3), the utility of which are dependent on the purpose of the detector. These metrics may or may not be affected by the prevalence of the call-type in the training dataset (Fielding & Bell, 1997).

Once a threshold is chosen, the template is run on a set of test data for which call-times have also been determined. The number of TP, FP, TN and FN detections are compared to estimate the template’s detection and error rates with unfamiliar datasets (i.e. independent cross-validation). Sensitivity and FPR metrics offer a means of assessing detection and error rates with little effect of varying call prevalence (Fielding & Bell, 1997). Variable performance tests have been employed across different studies; however, precision and sensitivity with

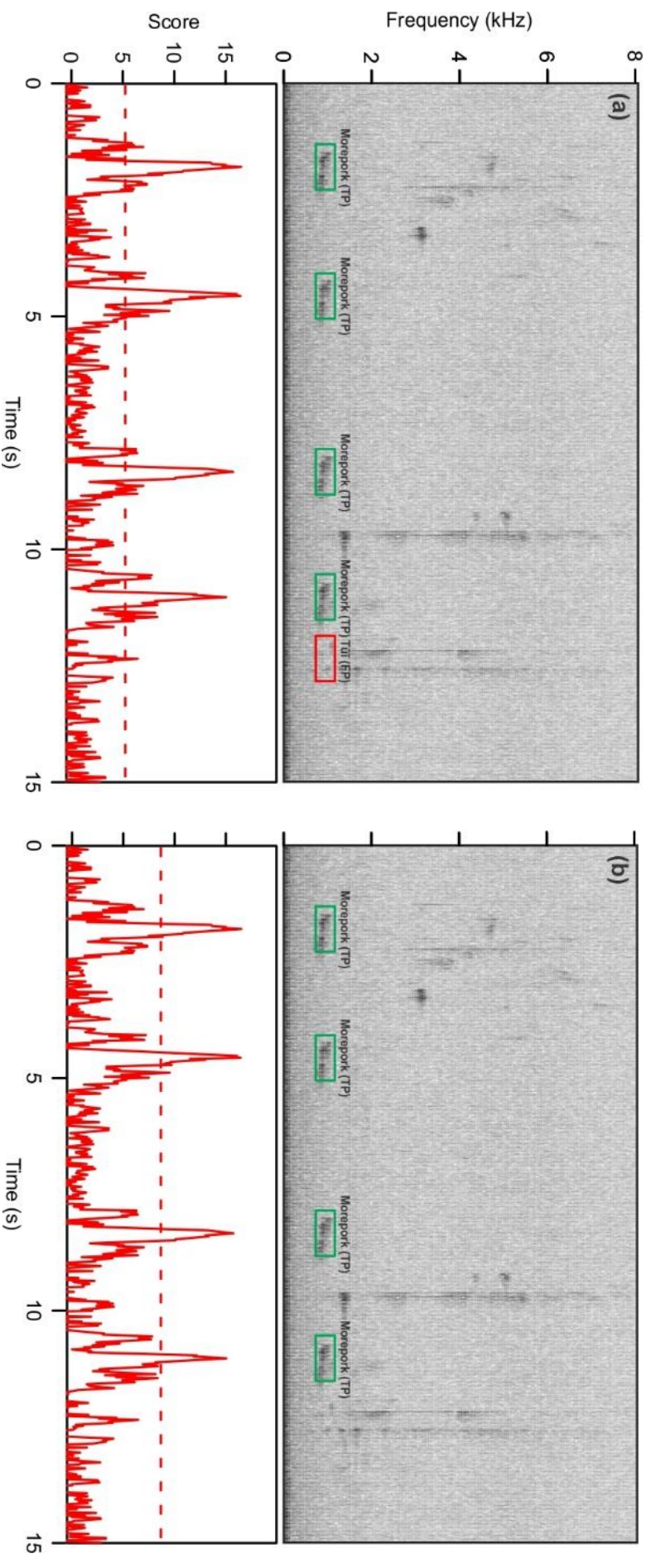


Figure 3.2 Example spectrogram recording window with four morepork calling events occurring alongside two tui calling events, and corresponding scores for a theoretical binary morepork template with: (a) a score threshold low enough to class low-ranging tui calling as a FP and; (b) a score threshold high enough to class low-ranging tui calling as a TN. TP detections and FP detections are represented by green and red boxes on spectrograms respectively. Solid red lines and dashed red lines represent binary template scores and score thresholds respectively, showing the effect of raising the threshold from (a) to accurately class the low-ranging tui call as a TN in (b).

Table 3.3 Performance metrics and their associated formulae, sourced from Fielding and Bell (1997).

Performance Metric	Formula
Sensitivity = True Positive Rate (TPR)	$TP / (TP + FN)$
Specificity = True Negative Rate (TNR)	$TN / (TN + FP)$
False Positive Rate (FPR)	$FP / (TN + FP)$
False Negative Rate (FNR)	$FN / (TP + FN)$
Precision = Positive Predictive Value (PPV)	$TP / (TP + FP)$
Negative Predictive Value (NPV)	$TN / (TN + FN)$

independent test datasets have been recommended for the cross-comparable reporting of final detector performance (Knight *et al.*, 2017).

3.2.2. Development process theory and overview

The detector development process outlined here aimed to develop a detector that achieves the best possible sensitivity at a given FPR, using the least number of templates possible from a wider pool available. In theory, this can be achieved by testing all possible template combinations. However, such a process becomes increasingly complex when larger numbers of potential templates are considered. Similarly, trialling of a limited number of random detector combinations may lead to the development of an underperforming detector. Instead, the process developed here relies on a limited number of detector trials each employing a ‘detector rationalisation’ process. It comprises three broader stages each using independent datasets: (1) training, where potential templates’ thresholds are independently set and their utility is assessed using a training dataset; (2) refining, where the set of templates identified as potentially useful in training is reduced to produce a parsimonious final detector using a refining dataset, and; (3) testing, where the performance of the final detector developed is assessed with a test dataset (Figure 3.3).

The rationalisation process is fundamental to achieving parsimony in this method of detector development. In detector rationalisation, the detector of interest is run on a specified dataset with all template thresholds independently optimised to a specified precision. All detections made by all templates are recorded; these detections are then inspected, and templates that do not make independent detections are removed from the set as a means to approach parsimony – i.e. the smallest set of available templates that will detect the most calls possible at the chosen precision. This process operates on the assumption that a set of templates

that each make independent detections will collectively find the more commonly detected calling events found by templates removed from the detector (Figure 3.4).

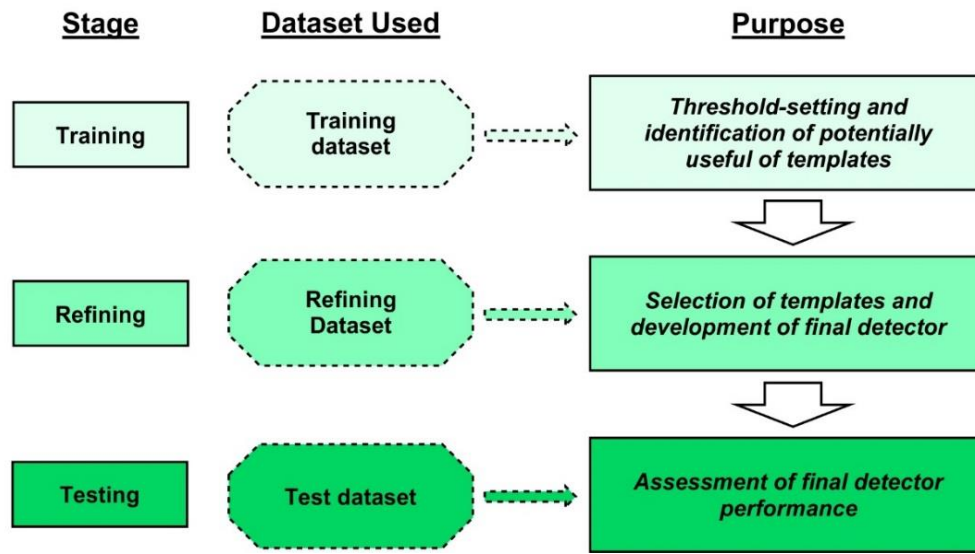


Figure 3.3 Flow diagram outlining the fundamental stages of the broader detector development process, their respective purposes, and the datasets involved.

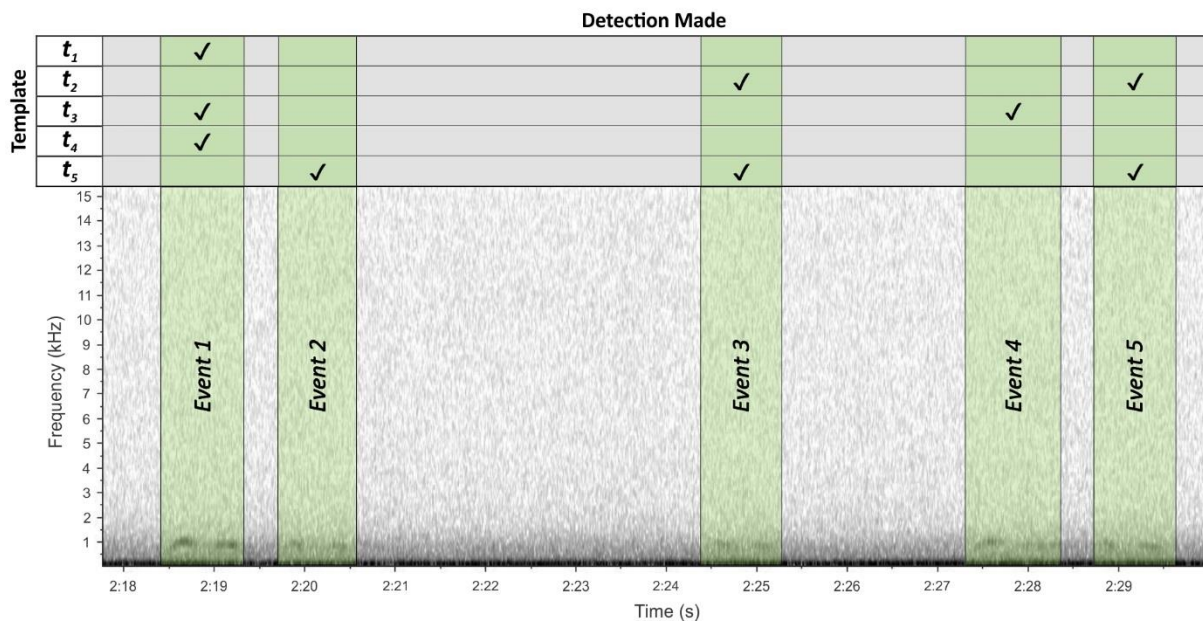


Figure 3.4 Example template rationalisation scenario with five calling events (highlighted in green) and five hypothetical potential templates; ✓ symbols indicate successful detection of a calling event by a template. In this scenario, templates t_3 and t_5 would be retained based on independent detections: these two templates would collectively detect the same calling events as all five templates combined, yet in 2/5 of the processing time.

The specific steps of the detector development and assessment process are outlined as follows (Figure 3.5):

1. An initial subset of n templates representative of the diversity in the call of interest is subjectively chosen from potential templates available for detector development (t_N). These initial templates ($t_1 \rightarrow t_n$) are individually trained using a training dataset dat_1 , and combined to produce an initial ‘training detector’ $det_{initial}$. This training detector is rationalised using the training dataset. Following rationalisation, the total number of TP detections made by the training detector is noted. This training detector provides a baseline number of detections for the dataset, and is used to assess the utility of further templates to be trialled (step two).
2. Using the training dataset, all other potential templates ($t_{n+1} \rightarrow t_N$) are independently trained and added to the training detector ($det_{initial} + t_i$). Those found to increase the training detector’s detection rate are saved as candidate templates.
3. Once all templates have been trialled, candidate templates from $t_{n+1} \rightarrow t_N$ are randomly grouped to create a set of ‘refining detectors’ ($det_{refining_1} \rightarrow det_{refining_m}$) each comprising approximately 10 templates. These are independently rationalised using the refining dataset, dat_2 .
4. Candidate templates retained from all refining detectors are combined with initial templates from the training detector. This set of templates is rationalised a final time, again using the refining dataset.
5. The resulting detector is the final detector, det_{final} . This is run and cross-validated on an independent test dataset, dat_3 , to assess its performance with unfamiliar recordings.

3.2.3. Applied detector development

3.2.3.1. Manual listening and scoring

Twenty-four Department of Conservation (DOC) ARUs were simultaneously active from 2000-2005 h (NZST) in seven study sites spread across three study areas (the Aorangi, Northern Rimutaka and Southern Rimutaka Ranges) for five weeks, from 6 November – 10 December 2016. Manual scoring was carried out on a set of recordings from this period to establish both a collection of calls from which templates could be developed, and a set of manually-verified datasets against which detectors could be trained, refined and tested.

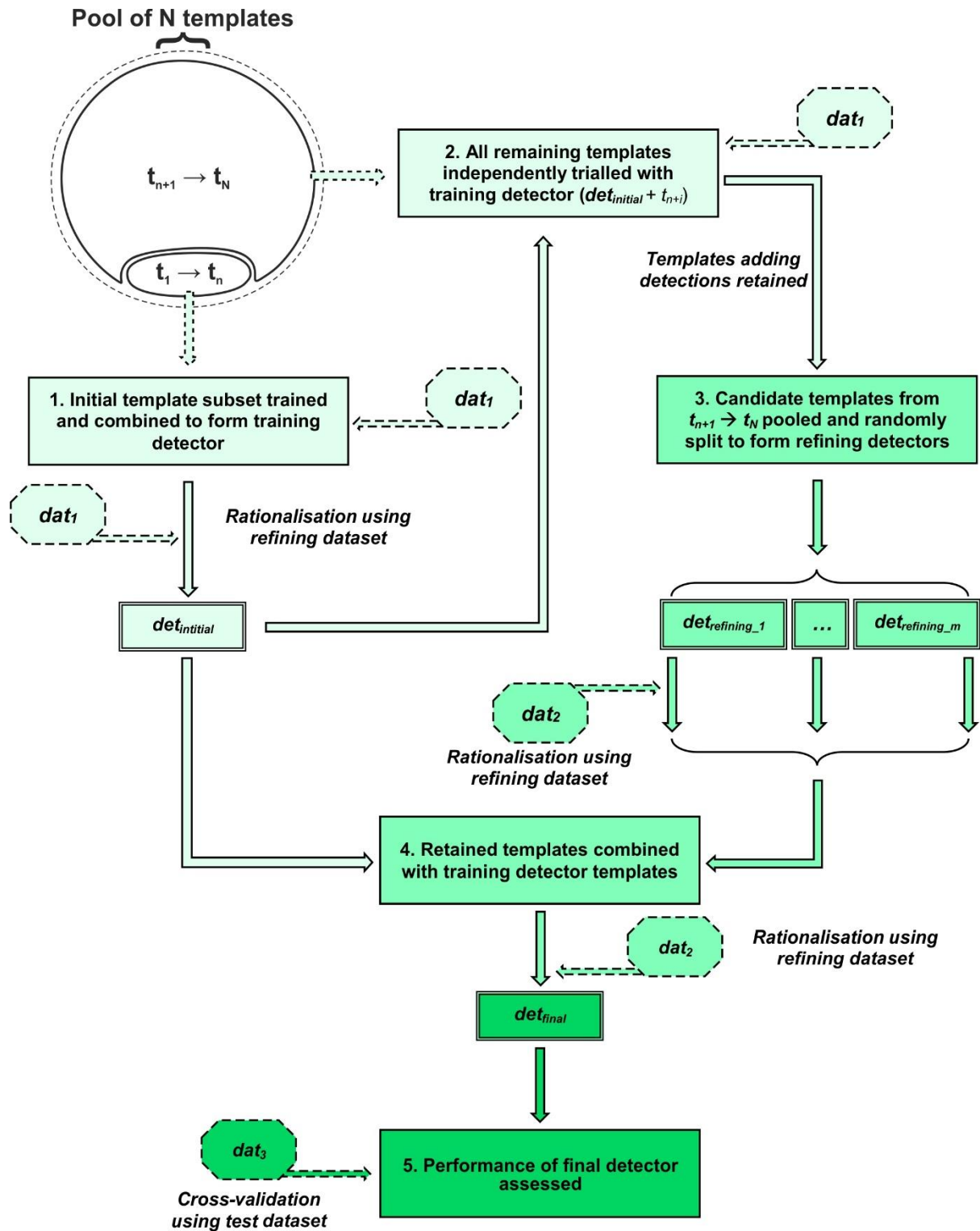


Figure 3.5 Flow diagram outlining the fundamental steps of the detector development process. Dotted octagons indicate datasets utilised, double lined boxes indicate detectors produced, and shape-fill colours highlight the broader stages employed with reference to Figure 3.3.

Climate data sourced from two Greater Wellington Regional Council (2017) weather stations in the Wainuiomata region (Appendix I) was used to identify days likely to have recordings suitable for analysis. Days with wind speeds of less than 15 km/h (at 10 m, 0800 h

NZST, Wainuiomata Bowling Club AQ station) and 0 mm of rainfall (cumulative, 0800-0900 h, Wainuiomata River at Wainui Reservoir station) were classified as candidates for analysis. This preliminary requirement shortened the available sampling period to three weeks, from 20 November – 10 December 2016.

Spectrograms of recordings from candidate days were then visually inspected using Raven Lite 2.0 (Bioacoustics Research Program, 2016) to identify days with recordings suitable for manual analysis. Recordings were defined as suitable if background noise was of a low-enough amplitude to visually identify morepork calls on a spectrogram. All candidate days identified in climate analyses were first inspected for at least one recorder per study site to identify days which were likely to have suitable recordings across all study sites. All recordings from these days were then further inspected and classified as suitable or unsuitable for analysis. Five days were then selected from these days for manual analysis. Days with the highest number of suitable recordings across all sites were primarily chosen. However, if less than three days separated two consecutive chosen days, the next-highest scoring candidate day was selected in place of the lowest-scoring day of the two. This selection process resulted in a sample of 118 recordings spanning 20 days.

Recordings from the selected days were then scored manually for the presence of morepork calls. Recordings were scored by one observer (Roald Bomans) simultaneously listening to recordings and visually inspecting their spectrograms in Raven Lite 2.0. Panasonic RP-HC200 headphones were used for all analyses with the noise-cancelling function disabled. Each sample was broken into 30 consecutive ten-second ‘blocks’ (see Table 3.1) to be independently scored, each of which were given either an ‘active’ score of 1 or an ‘inactive’ score of 0. A block was only given an active score if any class of morepork call could be both heard in the recording and seen on its spectrogram. Occasionally a call would be split across the start or end of a block; in such cases, both blocks containing split calls were scored as active. A block was given an inactive score if a call could be heard but could not be seen on the spectrogram; a call was present but could not be confidently identified as morepork; or if no call was present. Scores were totalled across all blocks for each recording, giving each recording a maximum potential score of 30 and minimum potential score of 0.

3.2.3.2. Call classification

Data from manual analysis was manipulated using the `reshape` package in R (Wickham, 2007) to produce a list containing file and time-stamp information for all active blocks identified in

manual analysis. Using the *seewave* (Sueur *et al.*, 2008) and *tuner* (Ligges *et al.*, 2016) packages in R, this data was used to create a series of 10-second potential template .wav files for all active blocks identified (Appendix V).

The calls within these potential template files were then classified by visual spectrogram analysis in Raven Lite 2.0. As many as 10 call-types have been recognised for morepork (Brighten, 2015; Heather & Robertson, 2005; Moon, 1992), though the extent to which they overlap and how many should be recognised is debated (Morgan & Styche, 2012). In this study, calls were classified into five widely recognised broader categories. The names of these calls vary source-to-source, but are here termed as the typical ‘*more-pork*’ call, the low-pitched and repeated ‘*more-more*’ call, the single-syllable ‘*more*’ call, the drawn-out and often harmonic ‘*cree*’ call, and the short and inflected ‘*mew*’ call (Figure 3.6). Calls were disregarded if either (1) though visible, their presence on the spectrogram was not clear enough for use in template development (i.e. individual syllables were broken in the spectrogram), or; (2) the call had been split across two blocks, and all syllables were not fully visible in the 10-second potential template .wav file produced.

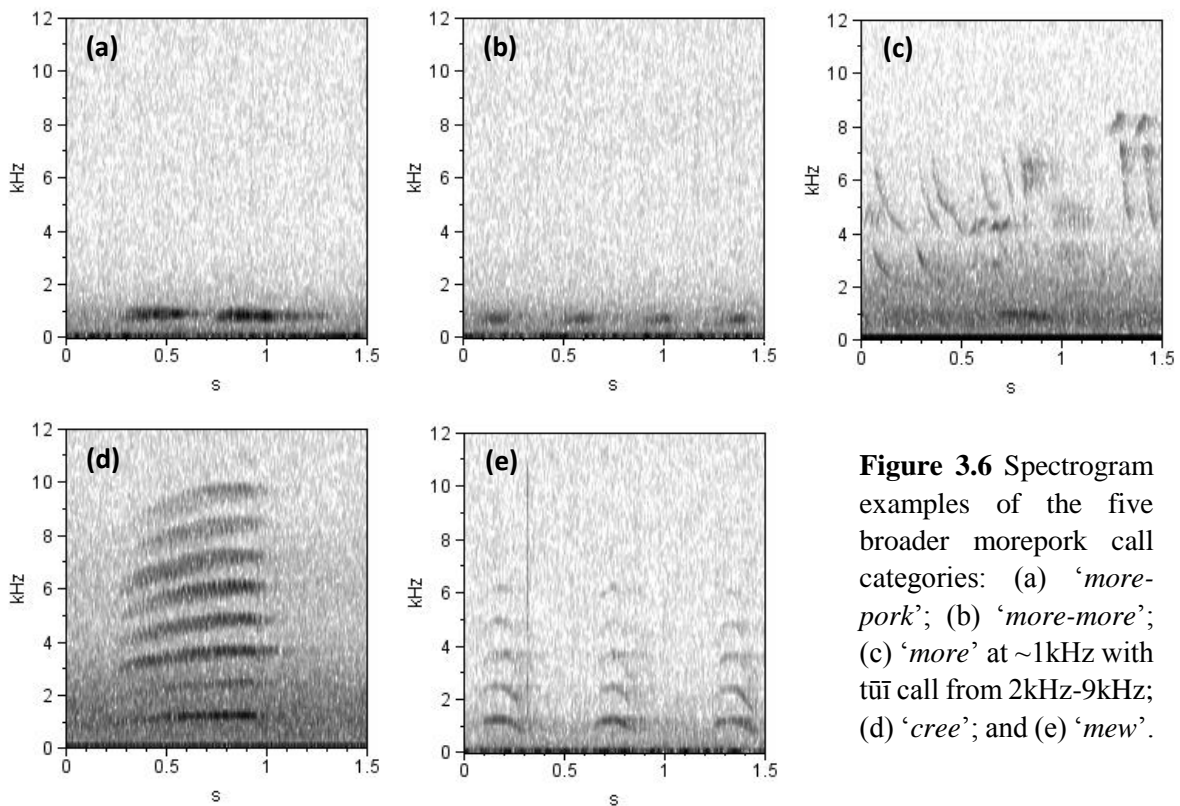


Figure 3.6 Spectrogram examples of the five broader morepork call categories: (a) ‘*more-pork*’; (b) ‘*more-more*’; (c) ‘*more*’ at ~1kHz with tui call from 2kHz-9kHz; (d) ‘*cree*’; and (e) ‘*mew*’.

Call-to-call variation in the *more-pork* call was noted. To account for this variation, these calls were classified into six further sub-categories apparent within this call-type. These

were established based on the sound and spectral appearance of the ‘*more*’ and ‘*pork*’ syllables in the call and comprised a ‘long-long’ call, a ‘long-short’ call, a ‘short-short’ call, a ‘short-long’ call, an ‘arcing’ call, and a ‘high-low’ call (Figure 3.7).

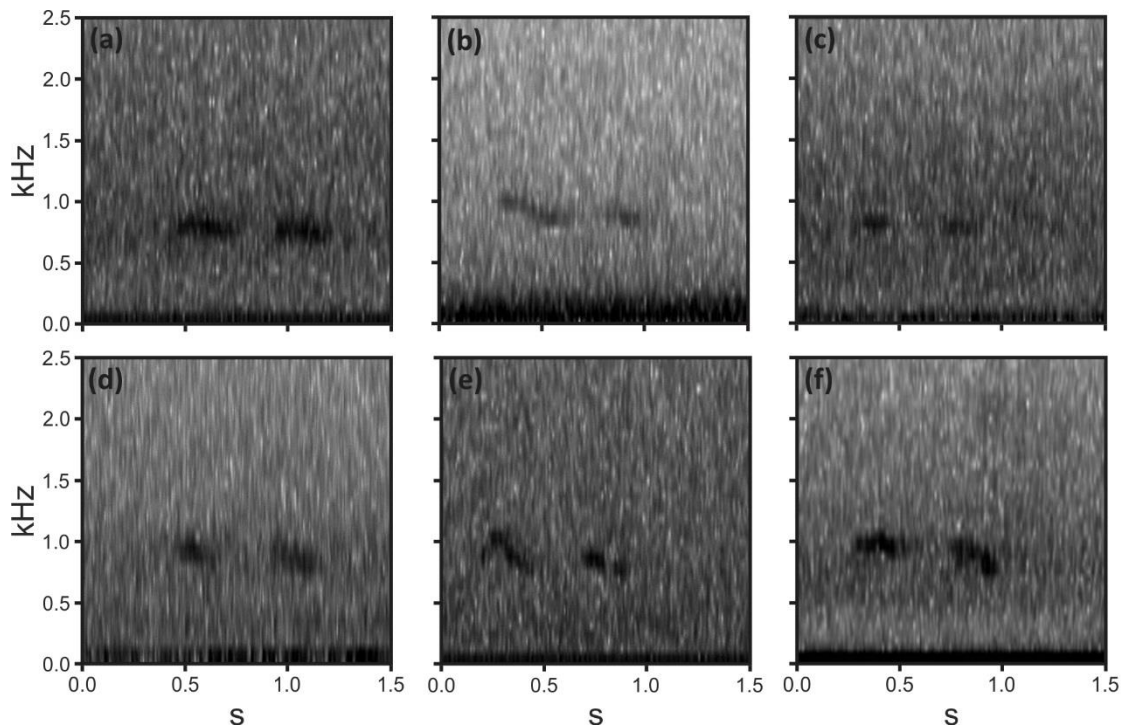


Figure 3.7 Example spectrograms of the six types of ‘*more-pork*’ calls classified by sound and spectral appearance: (a) ‘long-long’; (b) ‘long-short’; (c) ‘short-short’; (d) ‘short-long’; (e) ‘arcing’; and (f) ‘high-low’.

Detector development was focused on the *more-pork* call due to its prevalence as a call-type both described in the literature and observed in call classification (Heather & Robertson, 2005; Moon, 1992). In manual analysis, no *mew* calls suitable for template development were observed; eleven *more* calls were observed; one *more-more* event was observed; and blocks with *cree* calls suitable for template development were a quarter as prevalent as those with *more-pork* calls suitable for template development, likely due to the *cree* call’s association with hunting (Heather & Robertson, 2005). A total of 110 .wav files containing 187 *more-pork* calls suitable for template development were identified.

3.2.3.3. Development of training, refining and test datasets

Training, refining, and test datasets consisted of carefully-determined active and inactive blocks within a timeframe specific to each recording. This timeframe spanned from ten seconds before the first *more-pork* call that had two syllables fully visible on the recording’s spectrogram to ten seconds after the tenth call of the same quality. This limit ensured that clear

calls and calls specific to individuals calling frequently in any given recording or at any given location were not overrepresented in the data. Some exceptions to this timeframe rule applied: (1) if the first recorded call was less than ten seconds into a recording, the start-point for comparison was set as zero seconds; (2) if another call was present less than ten seconds after the tenth manually-detected call with two or more syllables clearly visible on the spectrogram, the endpoint for comparison was set as one-second before this next call's start point to avoid call overrepresentation, and; (3) if less than ten calls with two or more syllables clearly visible on the spectrogram were present, the end point for comparison was set at five-minutes.

Active blocks were determined within this timeframe for both *more-pork* and *more* calls; *more* calls were scored as active to allow for their potential detection by templates with a faint syllable. Consistent with methods employed in manual analyses, active blocks were only determined for calls that could be both heard and seen on the spectrogram. Active block length varied call-to-call, spanning from 0.2 seconds before a given call's start-time to 0.2 seconds after its end-time. This 0.2 second margin was incorporated to account for the fact that template detections are reported by *monitor* as the midpoint of the template; this could lead to detections falling outside of a call's bounds in scenarios where only one syllable is detected, and the syllable is shorter than those present in the template it is being tested against (Figure 3.8). Where two or more consecutive calls were less than 0.4 seconds apart the calls of interest were classed as one calling event, with the active block spanning from 0.2 seconds before the first call in the event to 0.2 seconds after the last.

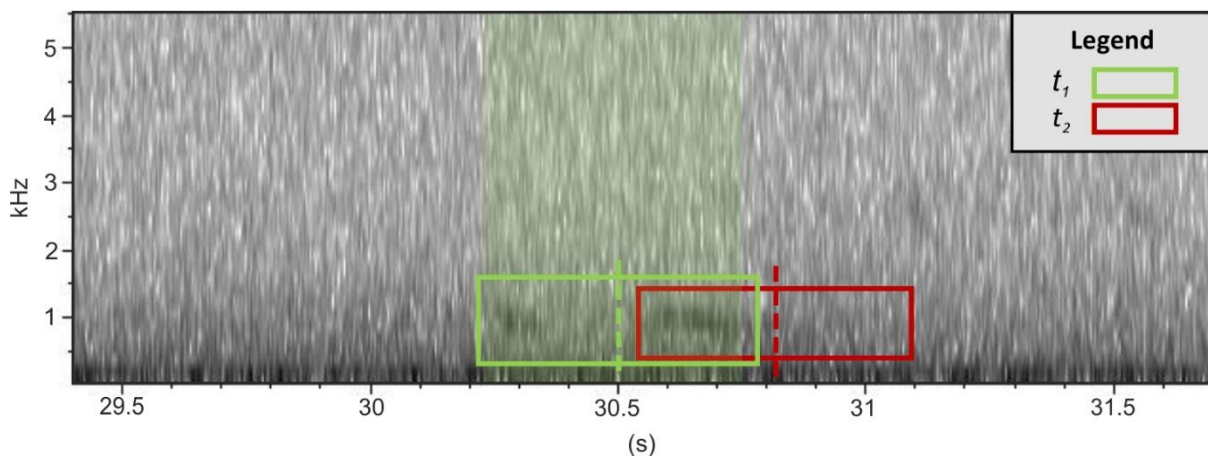


Figure 3.8 Example spectrogram of a calling event requiring a margin. Template t_1 successfully detects both syllables of the call of interest and estimates its time within the call's bounds; template t_2 detects only its second syllable, estimating its time as approximately 0.1 seconds after the call's end when a margin is not applied.

Inactive blocks were the time between two consecutive active blocks. If this time exceeded twenty seconds, it was broken into n consecutive even-length blocks to maintain an inactive block resolution of ~10 seconds; n was calculated as:

$$n = \lfloor \frac{start_{i+1} - end_i}{10} \rfloor$$

where end_i was the end time of the preceding active block and $start_{i+1}$ was the start time of the next active block.

Active and inactive blocks were given a binary score of 1 and 0 respectively, consistent with the scoring criterion formerly described for manual analyses. If no calls were visible on the recording's spectrogram, the whole five minutes was scored as thirty consecutive inactive ten-second blocks.

Twenty-six high-scoring recordings from manual analysis were used for the training dataset. Twenty-five further high-scoring recordings from manual analysis were used for the refining dataset. Detector performance could be overestimated if training and test datasets are sourced from the same area and season, and share calls characteristic of the same resident individuals. To reduce this potential, the test dataset was sourced from Aorangi and Northern Rimutaka Range recordings archived from summer 2014-2015 seasonal monitoring of the VUW Aorangi Project. This season was chosen on the basis that it had the most (10) simultaneously active DOC ARUs out of the archived seasons available. Days scored were randomly selected from days that were to be analysed by the detector for applied long-term analyses in Chapter 4. One recording was randomly selected and scored per recorder to develop the test dataset. The resulting training, refining, and test dataset structures are summarised in Table 3.4.

Table 3.4 Summary of training, refining and test dataset compositions.

Dataset	No. of Recordings	No. of Active Blocks	No. of Inactive Blocks
Training (Dat ₁)	26	306	489
Refining (Dat ₂)	25	174	570
Test (Dat ₃)	10	136	199

3.2.3.4. Development of the initial training detector

Two example calls of those clearest in call classification were selected for each of the six *more-pork* sub-categories as initial templates ($t_1 - t_{12}$) for training detector development. Both cross-

correlation and binary template-matching templates were developed for all initial templates. Amplitude cut-off points for binary template testing were selected such that the ‘on’ components of the template contained no extraneous background noise, and best resembled the key components of the call as seen when viewed as a correlation template (Figure 3.1).

Each initial detector template was run on the training dataset with a threshold of zero, such that all detections possible were reported. Given that multiple detections can be made by a detector in any given block, each block’s highest scoring detection was saved to ensure that the training process was as stringent as possible, incorporating the highest possible FP event scores. A score of zero was given to blocks where no detections were made. ROC analyses were carried out on the resulting detections using the `pROC` package in R (Robin *et al.*, 2011). The `pROC` package reports performance metrics for local maxima across the ROC curve. For each template, the thresholds of local maxima and their relative precisions, sensitivities, specificities, and FPRs were saved.

All twelve templates were then combined to produce the training detector. A detector producing a low number of FP events was desired to maximise the reliability of unsupervised recording analyses. However, a reasonable sensitivity was also desired. Thus, precision (the conditional probability of a detection event being a TP) was used for threshold optimisation (Katz *et al.*, 2016a). Template thresholds were selected from the local maxima identified in ROC analyses such that the precision of each template was >0.965 and as close to 0.97 as possible. This training detector was run on the training dataset and rationalised. The rationalised training detector comprised six retained initial templates collectively making 130 detections out of a possible 306 (i.e. TPR = 0.4248).

3.2.3.5. Identification of candidate templates

Templates additional to the twelve initial templates subjectively chosen for the training detector were then trialled. All 175 *more-pork* calls ($t_{13} \rightarrow t_{187}$) suitable for template development remaining from those identified in call classification were independently developed as templates. Observational analyses of initial template training results found that binary templates had higher sensitivities than cross-correlation templates at thresholds with approximately equivalent levels of precision (Figure 3.9), possibly due to the method’s utility in detecting faint calls (Towsey *et al.*, 2012). Given this observation, only binary templates were developed for remaining potential templates in the interest of maximising detection power. Some calls did not develop well as binary templates though visible in spectrogram, with

portions of syllables lost when amplitude cut-offs were set high enough to remove extraneous background noise; these templates were not further developed. All calls producing adequate templates were trained using the training dataset. These templates were then independently added to the training detector with their thresholds optimised to a precision of ~ 0.97 and run on the training dataset. The resulting number of detections was inspected. The additional template's information was saved as a candidate for the final detector if it made two or more additional detections (i.e. total detections were 132 or greater). This process identified thirty-six candidate templates additional to the six initial templates present in the training detector.

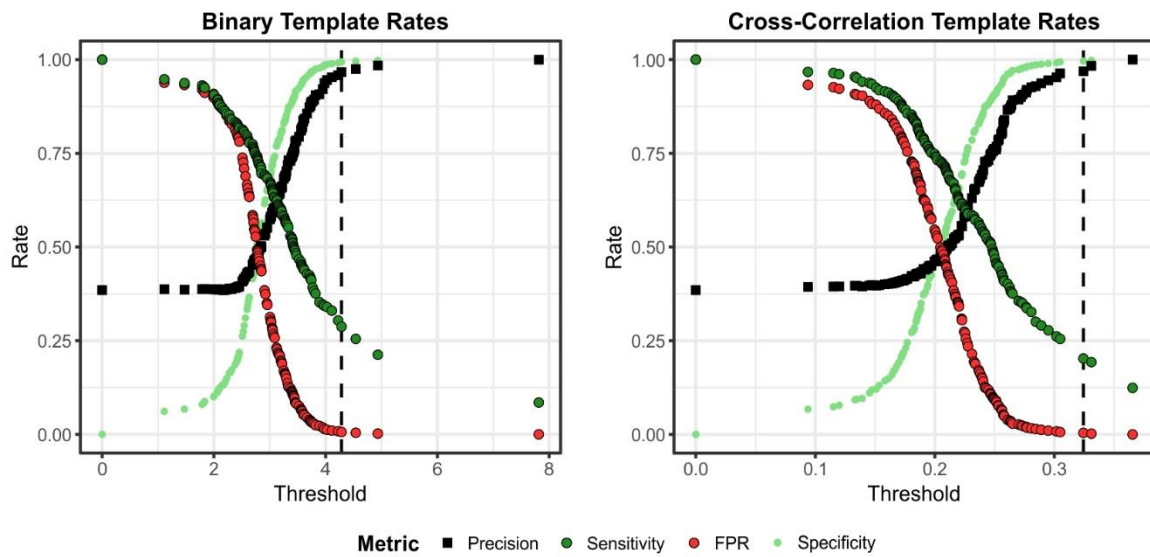


Figure 3.9 Scatterplots showing the precisions, specificities, sensitivities, and FPRs of local maxima for a 'short-short' call tested as a binary and a cross-correlation template. Dotted lines indicate the threshold with a precision closest to 0.97 (0.96703 and 0.96875 respectively). Specificity values are shown in background to illustrate the trade-off between a gain in specificity and a loss of sensitivity when minimising FPRs in training a template.

3.2.3.6. Development of final detector

The 36 additional candidate templates identified were randomly split to form three refining detectors, each comprising twelve templates. These detectors were then independently rationalised using the refining dataset, reducing the candidate list to a set of 20 templates. These 20 retained templates were then added to the training detector. The resulting 26-template detector was then rationalised a final time using the refining dataset, producing a final detector with 10 templates. The applied detector development process is summarised in Figure 3.10.

Templates may differ in performance across different amplitude cut-offs. After the development of the final detector, amplitude cut-off points of templates in the final detector were independently optimised to maximise sensitivity. Each final template was independently

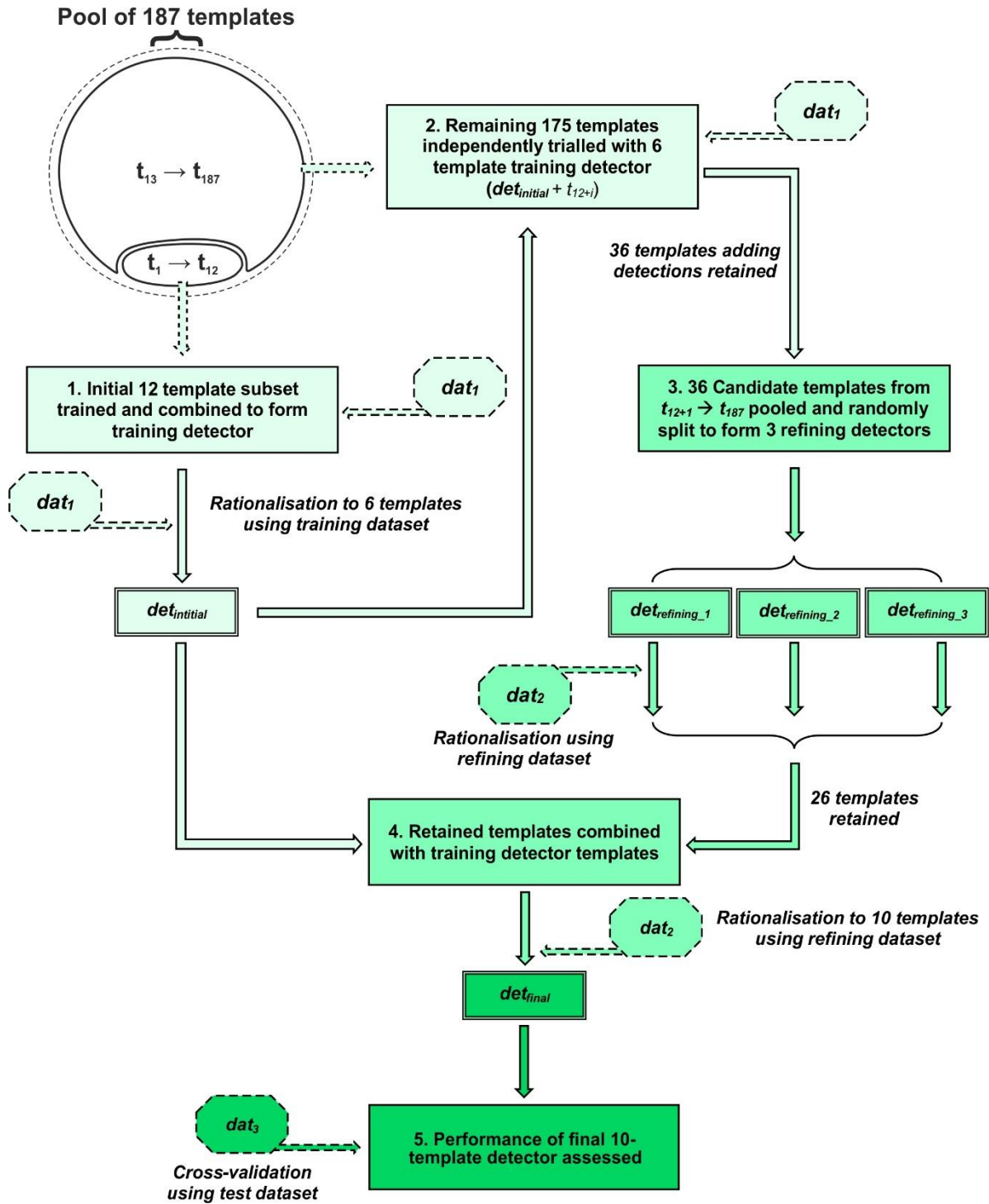


Figure 3.10 Flow diagram outlining applied steps of the detector development process and their results in the creation of the 10-template *more-pork* call detector developed. Dotted octagons indicate use of datasets, double-lined boxes indicate detectors produced.

run on the training dataset at the lowest amplitude cut-off possible that did not include background noise in the template's 'on' points. ROC analyses were carried out on the resulting detections, and threshold, precision and sensitivity information were saved for local maxima with precisions >0.965 and <1 . The amplitude cut-off was iteratively increased by one unit and

this process repeated until the template's sensitivity began to decrease across equivalent levels of precision. Each template's amplitude cut-off was then adjusted such that sensitivity was maximised with a precision greater than but as close to 0.98 as possible

3.2.3.7. Assessment of rationalisation process utility

During the refining stage, the total number of TP detections made by each refining detector was recorded before and after rationalisation to assess the effect of the process on sensitivity, and the utility of the process in achieving detector parsimony. All candidate and initial templates were also combined to form a 42-template detector and run on the refining dataset to determine the total number of TP detections made if all templates were used for comparison with the final detector developed.

3.2.3.8. Assessment of final detector performance

3.2.3.8.1. Detection of *more-pork* calling events

The final optimised detector was run on the training, refining, and test datasets, with the highest-scoring detection for each block reported. The detections made for each set of recordings were then compared to their manually-verified comparison datasets to determine their relative rates of TP, FP, TN and FN classifications. Any FP classifications found were manually inspected and their likely cause noted. Causes included wind; rain; *more-pork* calls missed in manual analysis or not visible on the spectrogram; *cree*, *more-more*, or *mew* morepork calls; *tūi* calls; secondary detection of a morepork call slightly outside of an active block's bounds; and unknown where no cause could be identified.

Classification rates were then restructured in two ways, such that: (1) FP classifications caused by wind, *tūi* calls, morepork *cree* or *mew* calls, *more-pork* calls detected secondarily outside of their active block bounds, or *more-pork* calls not visible on the spectrogram were classed as TNs and; (2) FP classifications caused by *more-pork* calls visible on the spectrogram were classified as TPs. These adjustments were made on the basis that wind events only occurred where gusts clearly reached 1kHz, and recordings with such events can be easily identified and excluded from analyses; *tūi* calls were only present in recordings made just after sunset, and would not normally be present in the more truly nocturnal recordings the detector was developed for, and; alternative morepork calls were not necessarily false detections as they would not misleadingly inflate detection rates. The sensitivity and FPR of the detector was calculated using both raw and restructured classification rates. The precision of the final detector was also calculated for the test dataset using its restructured classification rates.

3.2.3.8.2. Detection of *mew* and *cree* calling events

It was of interest if the detector developed for *more-pork* calls was sensitive to *cree* and *mew* calls. Given the limited presence of these call-types in the training, refining and test datasets, two further test datasets were developed to assess the detector's detection rates for these calls. Recordings with high *cree* or *mew* calling rates were identified in archived data and scored such that both call-types had over 100 events in their respective datasets.

The two datasets were scored independently: only *mew* calls were scored as active blocks in the *mew* dataset, and only *cree* calls scored as active blocks in the *cree* dataset. Where any other *morepork* call occurred simultaneously with a *cree* or *mew* call in its respective dataset, the block of interest was scored as inactive to avoid inflation of sensitivity estimates by other *morepork* call-types. Active and inactive block boundaries were determined consistently with the methods described for training, refining, and test dataset development. Each recording's scoring timeframe spanned from ten seconds before the first event visible on the spectrogram to ten seconds after the last event visible on the spectrogram.

The detector was run on each dataset and the relative number of TP, TN, FP and FN events were calculated. FP events were manually inspected and reclassified as TN events where another *morepork* call-type had caused a detection, and as TP events where a call of interest was faint on the spectrogram and had been missed. The sensitivity and FPR for each call-type was then calculated to approximate the detector's performance for these call-types.

3.2.3.8.3. Assessment of background noise impacts

To investigate whether the recording selection process effectively minimised any potential impact of background noise on sensitivities, the sensitivities for each recording from the training, refining and test datasets were modelled against average power density (dB FS) values calculated for each recording. Average power density values were taken as the mean of two average power densities calculated for 200-500Hz and 1300-1600Hz, frequency ranges just below and just above the region of *more-pork* calls. These ranges were selected to sample background noise in a frequency range proximal to *more-pork* calls without any impact of calls themselves on estimates. Power densities for both frequency ranges were sampled using Raven Lite 2.0 across the entirety of each recording's respective sampling timeframe.

Sensitivities were modelled against mean average power densities using a generalised linear model with a binomial link using the `lme4` package in R with the formula:

$$TPR \sim \text{Average Power Density}$$

A likelihood ratio test was carried out on the resulting model to explore the effect of background noise on sensitivities.

3.3. Results

3.3.1. Utility of the detector rationalisation process

The detector rationalisation process provided an efficient means to achieve detector parsimony; all three refining detectors were reduced in template number without any impact on each respective detector's total number of detections. The final rationalisation producing the final detector reduced the template set from 26 to 10 templates. This detector nearly achieved the performance of all 42 templates combined, making 89 of the 90 (98.89%) detections possible if all templates were used (Table 3.5). Only 1 template from the initial subset subjectively chosen and trialled ($t_1 \rightarrow t_{12}$) remained in the final detector.

Table 3.5 Summary of the number of TP detections made by refinement detectors and the final detector before and after undergoing rationalisation, and by all 42 candidate templates combined as a detector when run on the refining dataset. The number of templates making up each respective detector are shown in brackets.

	All templates	Det _{refining_1}	Det _{refining_2}	Det _{refining_3}	Det _{final}
Original	90 (42)	73 (12)	80 (12)	71 (12)	90 (26)
Rationalised	NA	73 (7)	80 (8)	71 (5)	89 (10)

3.3.2. Detection of *more-pork* calling events

As expected, the detector's performance declined across the datasets: sensitivities decreased and FPRs increased from the training to the test dataset (Figure 3.11). The final detector had an estimated precision of 0.9385, sensitivity of 0.3986, and FPR of 0.0253 for the restructured independent test dataset: i.e. 93.85% of detections were TPs, 39.86% of calls present were detected, and only 2.53% of silent blocks were falsely classed as calls. Events causing FPs were most often attributable to bands of low frequency noise most likely produced by electronic activity of the ARU; such events accounted for seven (58.33%) of the FP events across all datasets. Other causes included three incidences of indiscernible calls, and one-off incidences of a branch-knock and a heavy rain drop. Confusion matrices from which sensitivities and FPRs were estimated are available in Appendix VI.

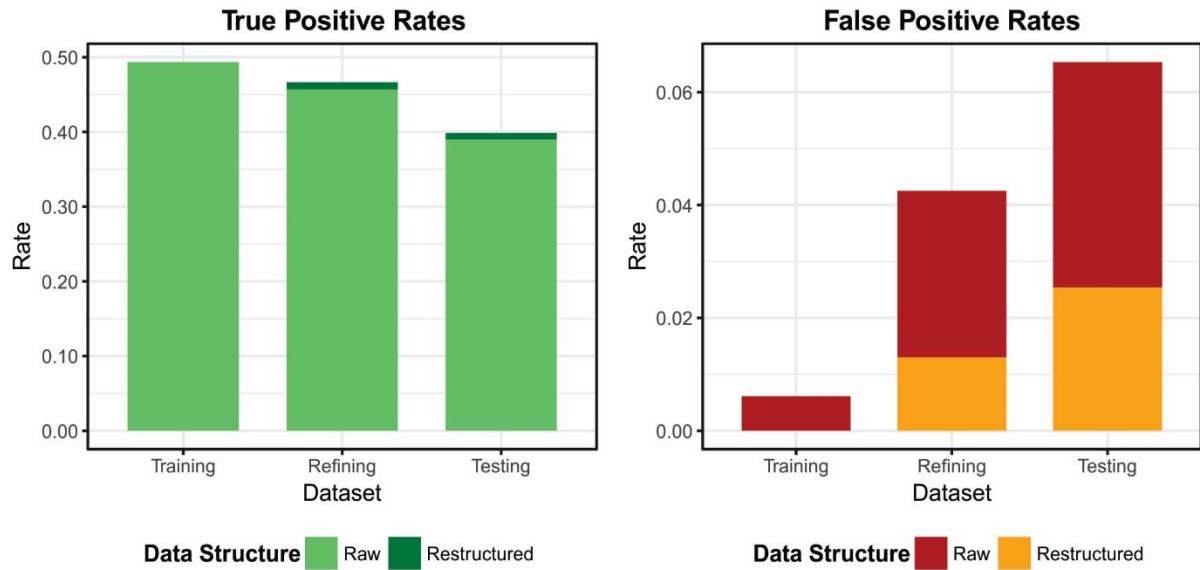


Figure 3.11 Bar graphs of raw and restructured TPR (sensitivity) and FPR estimates for the final detector across the training, refining and test datasets.

3.3.3. Detection of *mew* and *cree* calling events

The final detector detected both *mew* and *cree* calling events, achieving a sensitivity of 0.0977 (9.77%) for *cree* calls and 0.2946 (29.46%) for *mew* calls. Excluding detections of other morepork call-types, FPR rates were 0.0156 (1.56%) for the *cree* test dataset and 0.0116 (1.16%) for the *mew* test dataset.

3.3.4. Assessment of background noise effects on sensitivities

Background noise was not a significant predictor of sensitivity (Figure 3.12); the likelihood ratio test showed no significant effect of the mean power density (dB FS) of recordings on observed sensitivities ($p = 0.65$, $\chi^2 = 0.2059$, $df = 1$).

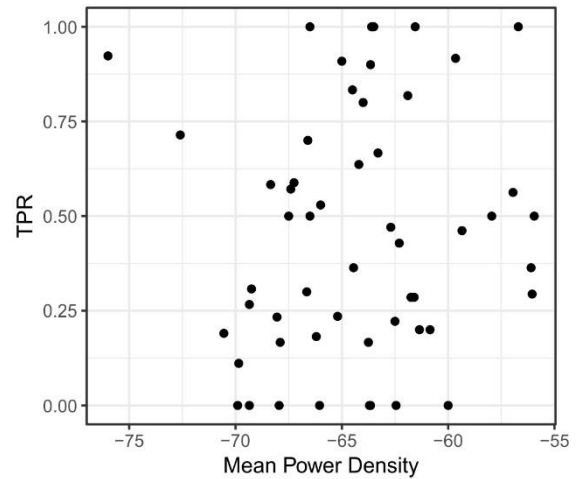


Figure 3.12 Scatterplot showing the lack of relationship between observed TPRs and mean power density values (dB FS) across recordings.

3.4. Discussion

The process of detector rationalisation developed here proved to be an effective means to achieve detector parsimony in the development of a template-based detector: the final 10-template detector developed made only one less detection than all 42 candidate templates combined when run on the refining dataset. Whereas all templates combined took 32.5 minutes to analyse the 271.7 minutes of data in the refining dataset, the final detector took just 8.5 minutes (i.e. 26.15% of the time). Furthermore, only one template of the twelve that were included in the initial detector based on structural diversity and call clarity was retained in the final detector. This highlights the limitations that may be associated with choosing templates subjectively. These results give strong support for the use of this method as a structured means to achieve detector parsimony in other template-based detector development projects with a large pool of potential templates. The refining stage of the process developed could be simplified to the rationalisation of a single set of candidate templates, particularly where fewer candidate templates are being considered. However, candidate templates in this study were split into three subsets to: (1) facilitate easier location of independent event detections in data produced; (2) facilitate faster assessment of detection data (while one refining detector was being run on the refining dataset, another's detections could be inspected), and; (3) produce multiple opportunities to assess the utility of the rationalisation process. Where few potential templates are available, candidate template identification could be bypassed and the rationalisation process applied as a simple and effective means to achieve detector parsimony.

Template-based detection methods have had both positive and mixed results (Katz *et al.*, 2016a). The performance of the *more-pork* detector developed here is promising with a moderate sensitivity (0.3986), low FPR (0.0253) and high precision (0.9385). A final precision lower than the minimum level set for individual templates (0.98) likely stems from differing triggers of FPs template-to-template resulting in a higher pooled error. It is important to note that the sensitivity of this detector is a generalisation of performance across recordings in an independent test dataset. Detector performance has been found to increase with increasing call quality, the majority of false negatives being attributable to low signal strength (Digby *et al.*, 2013; Swiston & Mennill, 2009). Thus, performance is likely to vary recording-to-recording. Recordings with a greater number of clear, high-amplitude calls are likely to achieve higher sensitivities than those with a disproportionate level of low-amplitude calls.

Unfortunately, the comparability of detector performance achieved across detector studies is limited by the diverse and non-standardised testing standards employed (Knight *et al.*, 2017). For example, detectors developed using `monitor` for black-throated green warbler (*Setophaga virens*) and ovenbird (*Seiurus aurocapilla*) songs appear better-performing than the detector developed here; one high signal:noise ratio template was used per species to achieve sensitivities of 0.72 and 0.62 respectively and FPRs of 0.01 (Katz *et al.*, 2016a). However, the reported performance metrics of Katz *et al.* (2016a) are based on a ‘silver-standard’ testing set which excludes particular call variants, and events less than 10dB louder than ambient background noise. A gold standard including all call variants of all amplitudes was also developed in the study, but not used to assess performance. Joshi *et al.* (2017) note that the use of only high-quality calls to train detectors may lead to poor performance with in-field recordings. In the same manner, it is possible that the use of only higher signal:noise ratio calls in testing may lead to an overestimation of performance, as these are more likely to be detected (Swiston & Mennill, 2009). Save for calls not visible on the spectrogram, the performance of the *more-pork* detector developed here was assessed using the equivalent of a gold-standard test dataset. Any *more-pork* events visible on the spectrogram were included in the training, refining and test datasets, as were single syllable *more* events. The number of clear calls per recording was also limited to avoid overestimation of performance resulting from their overrepresentation. Given these differences in testing, the performance metrics of Katz *et al.* (2016a) may overestimate performance relative to those of this study.

The performance of this detector appears comparable to the 0.349 sensitivity and 0.000 FPR of a detector developed for Screaming Piha (*Lipaugus vociferans*) birds (Ulloa *et al.*, 2016). However, metrics estimated by Ulloa *et al.* (2016) were not independent, being self-validated with the training data used to develop the detector. This may again overestimate performance, relatively-speaking, and limit comparability (Knight *et al.*, 2017). Nevertheless, the performance of the detector developed in this study is promising in the context of wider bioacoustics. A recent review of 68 bioacoustic studies found 12 single-species detectors tested on unedited field-quality recordings to have mean precisions and sensitivities of 0.71 and 0.60 respectively (Knight *et al.*, 2017). Though the sensitivity of the detector developed here is somewhat lower than average (0.3986), the precision (0.9385) is notably higher. This is fitting given that the detector was trained to minimise FP detections (Ulloa *et al.*, 2016). A higher sensitivity could be achieved with this detector at the cost of a reduction in the reliability of detection results (e.g. a template-matching detector developed by Borker *et al.* (2014) for

Foster's tern (*Sterna forsteri*) calls achieves a higher sensitivity (0.536) with lower precision (0.773). However, a high precision is preferable here given the detector's intended use for unsupervised analyses.

It is possible that the performance of the detector developed in the current study could be improved with little loss of precision. More simplistic calls have incurred higher FPRs relative to complex stereotypical calls (Swiston & Mennill, 2009). Considering the simple structure of *more-pork* calls, it is likely that, at the current high-precision template thresholds, many low-quality calls are falsely classed by the detector as negatives to avoid a high FPR. This loss of TP detections might be mitigated by the inclusion of rejection templates representative of common TN events, particularly those likely to cause FPs (Charif & Pitzrick, 2008). This method has been utilised to identify and remove common background noises and periods of silence in the automated recognition of indigo bunting (*Passerina cyanea*) and zebra finch (*Taeniopygia guttata*) calls (Anderson *et al.*, 1996; Kogan & Margoliash, 1998), and reduce FP detections by an order of magnitude in some cases (Charif & Pitzrick, 2008). Template score thresholds could be reduced and sensitivity increased if rejection templates were included in this detector; any resulting increase in FPR could be minimised through subsequent rejection template analyses. This would be particularly useful to reduce FPs caused by recorder noise and general background noise, but would come at the cost of increased processing time.

It is unsurprising that the detector developed also detects *mew* and *cree* morepork calls, though at lower rates. Due to the limited occurrence of these calls in training data, and any other call in nocturnal recordings, it is likely that the detector is highly sensitive to short-duration sounds in the *more-pork* frequency range. The low 9.77% sensitivity of *cree* events likely stems from their dissimilarity to *more-pork* calls, consisting of a single drawn out syllable in the frequency range of the *more-pork* call. In contrast, considering their inflected structure, *mew* calls may bear similarity to *more* calling events or *more-pork* calling events if only their start and/or end falls into the frequency range of the *more-pork* templates developed. This may explain their higher detection rate (29%) relative to *cree* calls. The detection of these call-types may be of concern in studying call-specific trends in calling for the species, and could be mitigated by the employment of training datasets rich in these call-types (>100 examples of each call). Furthermore, a detector would best benefit from the development of templates for these call-types. However, these calls are infrequent and have been found to occur at similar rates as relative proportions of total calls produced by morepork between different

study sites (Fraser & Hauber, 2008). Given they are detected at a consistent rate, these detections should bear no negative impact on the overall trends in calling detected by the detector developed.

Finally, it is important to note that this detector's performance is designed for nocturnal recordings that have been identified as suitable for automated analysis. Analysis of unchecked recordings, particularly those with low frequency noise such as wind gusts, is likely to produce higher FPRs. Background noise remains a pertinent issue in automated detection, causing both FP and FN events, and many automated detection methods still rely on low background noise levels (Brandes, 2008; Joshi *et al.*, 2017). However, the non-significant effect of average power density on sensitivity in the current study indicates that, though somewhat subjective, the weather-based recording selection process employed here effectively eliminated any effect of background noise on detection probabilities. To remove all subjectivity from the recording selection process, recording suitability could be determined through a check of each potential recording's mean average power density in the frequency ranges assessed. This could greatly streamline the detection process if automated, as the majority of time in analyses using this detector would be invested in inspection of the suitability of recordings for analyses.

The rise of ARUs in ecological monitoring is providing new opportunities for ecologists to monitor acoustically active species. However, as the use of ARUs increases, there is a pressing need for efficient means to extract data from acoustic recordings. Automated detection of species' calls is the most likely solution to this issue. A range of software now exists for automated detector development, with template-matching detectors suitable for species with simple calls. This study developed and applied a directed method of template-based detector development to successfully create a parsimonious detector for morepork calls from a large pool of potential templates. The net-benefit of automated analyses is dependent on the performance of detectors, and the time required to develop and run them (Knight *et al.*, 2017). This considered, the process developed has the potential to enhance the utility of template-based detectors in wider bioacoustic monitoring, given its success here in efficiently and effectively achieving detector parsimony with templates available. The detector developed here is capable of detecting *more-pork* calls of morepork in the Aorangi and Rimutaka Ranges of the lower North Island, New Zealand at a moderate sensitivity with high precision. These respective rates make the detector suitable for the applied unsupervised retrieval of general calling trends in suitable recordings from these areas. Given the challenges associated with monitoring nocturnal species, little systematic monitoring has been carried out for this species

(Pryde & Greene, 2016). However, assuming a relationship between calling prevalence and relative abundance, this detector provides a means to efficiently monitor population trends of morepork in the lower North Island. It may also be applicable to the monitoring of morepork populations across wider New Zealand and is easily distributable, consisting of only 10 .wav files totalling 8.47MB and an R script (Appendix VII). However, regional dialects affect the potential application of widely applicable automated detectors (Brandes, 2008), and are an area for study for morepork (Brighten, 2015). Thus, the performance of this detector in other regions would need to be assessed before utilisation. If necessary, regional detectors would be easy to develop using the parsimonious templated development method described here.

**Chapter 4 - Applied use of an automated detector
for short- and long-term monitoring of morepork
(*Ninox novaeseelandiae*) around aerial 1080
operations**

4.1. Introduction

Aerial 1080 distribution plays a major role in New Zealand's pest control operations, being the principle method for controlling possums in forested areas since the 1970's (Lloyd & McQueen, 2000). As acceptance of the need for predator control for wildlife protection has grown, so too has the use of this poison: 2014 saw the largest 1080 operation to date, with 10% of New Zealand's forested area treated in an effort to control significant predator numbers resulting from a major beech masting season (Elliott & Kemp, 2016). With any aerial poison operation there is a risk of poisoning non-target native avifauna (Veltman & Westbrooke, 2011). This issue is of significant public interest and concern (Green & Rohan, 2012). Non-target poisoning may occur as primary poisoning, where non-target species directly consume poisoned baits; or as secondary poisoning, where non-target species consume prey or carcasses that have fed on baits (Eason *et al.*, 2013a; Lloyd & McQueen, 2000). However, the benefits gained by non-target species through reduced predation pressure and/or reduced competition after aerial 1080 operations are generally considered to outweigh any such losses (van Klink *et al.*, 2013). It is important that the potential losses and benefits of non-target species in areas destined for 1080 treatment are well understood. This is particularly pertinent where less resilient species are present; where bait specifications change (Greene *et al.*, 2013); and in mainland scenarios where repeated treatment is necessary to mitigate reinvasion by pests, as such repeated treatment may lead to a negative long-term impact on non-target species if repeated short-term losses are incurred (Veltman & Westbrooke, 2011).

The prevalence and impact of secondary poisoning on native predatory avian species has not been widely studied (Stephenson *et al.*, 1999), and a lack of knowledge exists around the long-term responses of native avifauna to aerial 1080 treatment at the population level (Veltman & Westbrooke, 2011). This is true for the morepork/ruru (*Ninox novaeseelandiae*), New Zealand's only extant native owl (Fraser & Hauber, 2008; Stephenson *et al.*, 1999). Morepork have been found dead after poisoning operations using both brodifacoum (Empson & Miskelly, 1999; Stephenson *et al.*, 1999) and carrot- and cereal-based 1080 baits (Spurr & Powlesland, 1997). Moreover, three of the four morepork carcasses that have been analysed for 1080 residues after being found following 1080 operations have tested positive (Powlesland *et al.*, 1998; Spurr & Powlesland, 1997). Thus, there is evidence to suggest that morepork are susceptible to 1080 poisoning.

A top predator, morepork are known to prey on vertebrates, particularly mice and small rats when they are abundant, but feed predominantly on large invertebrates (Denny, 2009; Powlesland *et al.*, 1998), specifically beetles (Coleoptera), moths (Lepidoptera) and wētā (Orthoptera) (Haw & Clout, 1999; Haw *et al.*, 2001). Given this diet, the species is not considered threatened by primary poisoning (Spurr, 1979; Stephenson *et al.*, 1999). As a predominantly visual predator (Stephenson, 1998), morepork hunt live prey and are also not considered to be at risk of secondary poisoning from scavenging carcasses (McFadden, 1992). However, potential routes for secondary poisoning exist through the predation of sub-lethally poisoned invertebrates and small vertebrates. Beetles and wētā have been found feeding on non-toxic baits (Spurr & Drew, 1999), and 1080 residues have been found in cockroaches, wētā (Eason *et al.*, 1993 as cited by Eason *et al.* (2011)), and a number of other invertebrate species (Lloyd & McQueen, 2000). Furthermore, arthropod bait consumption has been found to occur predominantly at night. Thus, nocturnal species such as morepork may be particularly at risk to secondary poisoning through predation of arthropod species (Lloyd & McQueen, 2000). Sub-lethally poisoned rodents and birds have also been suggested as prime targets for predators following poisoning operations (Spurr, 1979).

Despite these mortalities and potential risks, morepork have been sparsely monitored both in general (Pryde & Greene, 2016) and around 1080 operations. Only five studies tracking a total of 27 radio-tagged individuals in treatment areas have been carried out to date (Greene *et al.*, 2013), and only one study has monitored calling rates short-term around an operation (Pierce & Montgomery, 1992). No published literature currently exists regarding the long-term responses of morepork to any form of 1080 operation. Morepork are slow reproducers, producing a single clutch of only 2-3 chicks per year (Stephenson & Minot, 2006). Given this low reproductive capacity, Spurr (1979) classed the species as at medium risk to non-recovery from any potential population losses sustained from 1080 operations. Accordingly, Spurr and Powlesland (1997) deemed the study of morepork around 1080 possum control operations to be of high priority given the historic mortality of the species and lack of research at the time. Only three studies, however, have been carried out since 1997, namely Greene *et al.* (2013); Powlesland *et al.* (1998); and Powlesland *et al.* (1999). Thus, there is a gap in the literature and evident need for further investigation of morepork population trends around aerial 1080 operations.

The lack of research regarding morepork population trends may result from the logistical challenges associated with monitoring nocturnal species (Pryde & Greene, 2016).

However, call counts employing autonomous recording units (ARUs) show promise as a practical and cost-effective means for widespread monitoring of the species, particularly if automated call detection can be applied (Pryde & Greene, 2016). The current study sought to apply the use of an automated acoustic detector developed for morepork calls (see Chapter 3) to the monitoring of morepork across three aerial 1080 operations. This detector does not detect all calling events; however, its low error-rate ($FPR = 0.0253$), consistent sensitivity (0.3986) and resulting precision (0.9385) with independent data makes it a suitable tool for monitoring broad-scale morepork population trends using calling prevalences as an index of relative abundance.

Nocturnal acoustic recordings were analysed (1) short-term across treatment and non-treatment areas for approximately five weeks before and after three independent 1080 operations, and; (2) longer-term over four years for one treatment area receiving two 1080 applications, and one non-treatment area receiving no 1080 treatment. Specifically, it was sought to determine if changes in detected calling prevalences as an index of morepork abundance differed between treatment and non-treatment sites short-term from before to after 1080 operations; or showed any long-term response, negative or positive, to repeated 1080 operations.

4.2. Methods

4.2.1. Study areas and study sites

Acoustic recordings were analysed from data recorded by ARUs situated across five study areas monitored at various times over a four year period (2013-2017): the Aorangi, Tararua, Northern Rimutaka and Southern Rimutaka Ranges, and Tora Bush (Figure 4.1). These study areas collectively comprise vegetation representative of the southern North Island, consisting mainly of beech, broadleaf, and mixed beech/broadleaf forest, with podocarps scattered throughout. Beech tends to dominate at areas of higher elevation or infertile soils (Dymond & Shepherd, 2004; Wardle, 1967), and regenerating mānuka/kānuka and gorse scrub is also present at one Aorangi study site.

Within a study area, between one and six study sites were utilised. A study site comprised between one and nine recorders that could be serviced in one day (i.e. have their batteries and SD cards replaced) by a pair of field technicians or students. With a few exceptions noted below, recorder locations within a study site were spaced at least 400-500m

apart to minimise the chances of recording the same individual bird at more than one location (Pryde & Greene, 2016).

The region has a mild, wet climate (Dymond & Shepherd, 2004). Rainfall varies from 800 mm to 7000 mm on an elevational and east-to-west gradient, with areas in the west and at higher elevations receiving more rainfall (Cook, 2017; Dymond & Shepherd, 2004). Mean monthly temperatures near sea-level vary from 8 °C (July) to 16 °C (January) (Dymond & Shepherd, 2004).

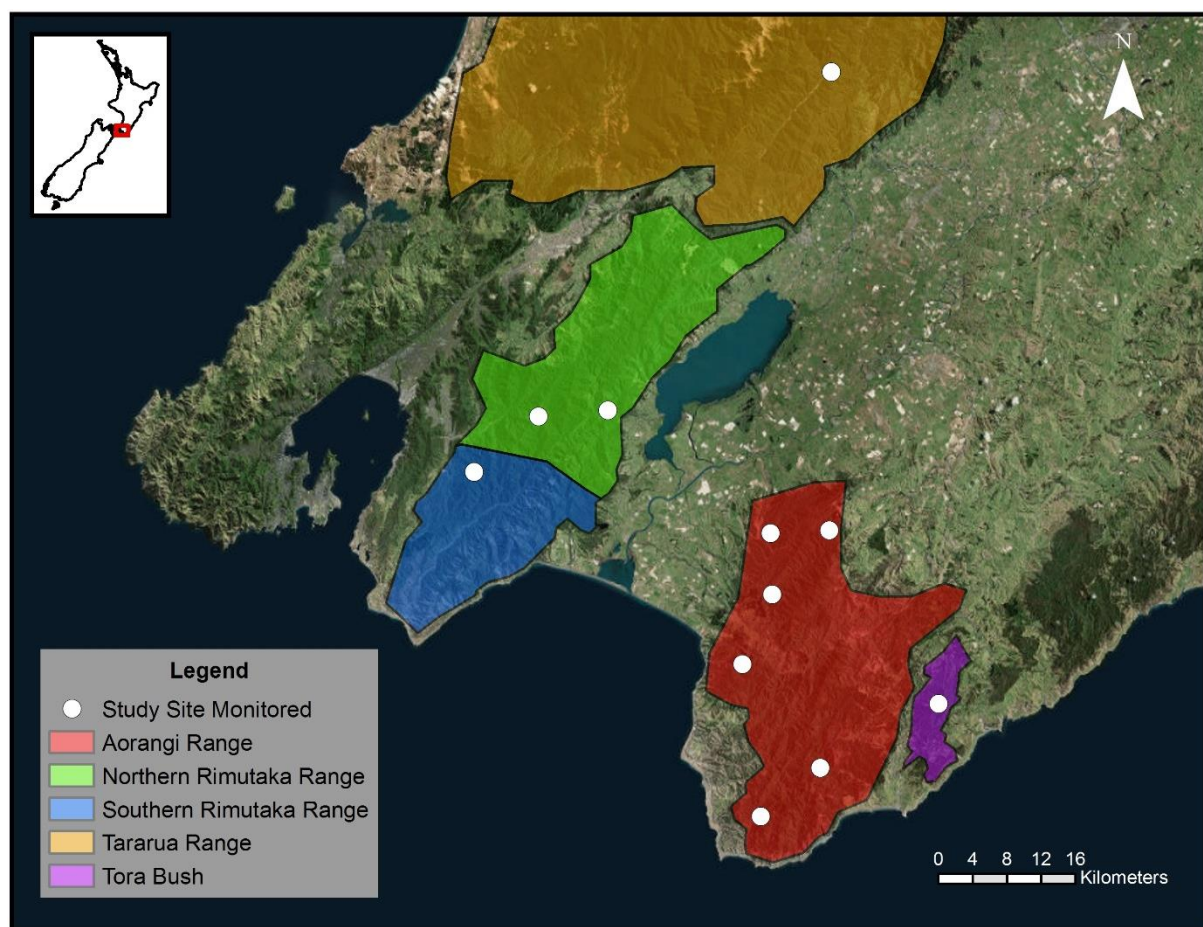


Figure 4.1 Map showing study sites and approximate forested extent of focus study areas.

4.2.2. 1080 operations

The Aorangi Range currently receives 1080 drops on a three-yearly cycle, the two most recent drops being in August 2014 and June 2017. In 2014, the area was subjected to a pre-feed application of non-toxic 6 gram cereal baits on August 4-5 2014, followed by a toxic drop of

12 gram cereal baits with 0.15% toxicity on August 18 and August 27 2014. Pre-feed was applied at 0.5 kg/ha and toxic bait was applied at 1.0 kg/ha in 180 m swaths with intervening 40 m strips of non-application. In 2017, the area was pre-fed on May 30-31 2017 using 6 gram cereal baits applied in 260 m swaths, followed by a toxic drop of 12 gram 0.15% toxicity cereal baits in 300 m swaths on June 16-17 2017. Both operations used deer repellent and were applied full-broadcast at 1.5 kg/ha. Prior to the 2014 operation, the northern and southern half of the Aorangi Range last received 1080 treatment in July 2009 and September 2006 respectively (Cook, 2017; OSPRI, pers. comm.).

The Southern Rimutaka Range area was subject to its first ever 1080 drop in July 2017. Pre-feed of non-toxic 6 gram cereal baits occurred on June 16 2017, followed by a toxic drop of 12 gram 0.15% toxicity cereal baits in 180 m swaths on July 30 2017. Both operations used deer repellent and were applied full-broadcast at 2 kg/ha (OSPRI, pers. comm.).

Where monitored throughout the course of this study (Nov 2013 to November 2017), Tora Bush, the Northern Rimutaka and Tararua Ranges did not receive 1080 in areas studied and served as non-treatment comparison sites for respective 1080 operations. Prior to monitoring commencing in 2013, the Northern Rimutaka and Tararua Ranges received aerial 1080 treatment in 2012 and 2010 respectively (TBfree, 2010; Uys & Crisp, 2018). Tora Bush has not received aerial 1080 treatment since at least 2009 (DOC, pers. comm.).

4.2.3. Experimental Design

4.2.3.1. Short-term effects

The study of short-term effects of 1080 mammal control on morepork employed a BACI (Before-After/Control-Impact) experimental design (Underwood, 1992), utilising ARU recordings to compare changes in morepork calling prevalence in treatment and non-treatment areas from zero to five weeks before to one to six weeks after three respective 1080 operations.

4.2.3.1.1. Aorangi 2014 operation

Nocturnal acoustic recordings were sourced from unanalysed archived recordings collected in either the preceding study by Cook (2017), or the VUW Aorangi Project's seasonal monitoring (Hartley, 2017). Only recording locations with recordings of adequate quality both before and after treatment were monitored. Recordings were analysed for 12 ARUs distributed across six study sites in one treatment area, the Aorangi Range, and 13 ARUs distributed across three study sites in three non-treatment study areas: the Tararua and Northern Rimutaka Ranges and

Tora Bush (Figure 4.2). Recorders were 2013 DOC or Song Meter™ SM2+ ARUs spaced at least 500 m apart within a study site.

4.2.3.1.2. Aorangi 2017 operation

Initially, 13 and 18 ARUs were distributed across the treatment and non-treatment areas respectively. However, one location was not sampled due to inadequate recording quality for automated analyses. Thus, recordings were analysed for 17 ARUs distributed across two study sites in one non-treatment area, the Northern Rimutaka Range, and for 13 ARUs distributed across four study sites in one treatment area, the Aorangi Range (Figure 4.3). Recorders were 2013 or 2016 DOC ARUs spaced at least 400 m apart within study sites, with the exception of two treatment and three non-treatment recorders spaced 350 m from the next-nearest recorder.

4.2.3.1.3. Southern Rimutaka 2017 operation

Initially, 18 ARUs were distributed across both the treatment and non-treatment study areas (i.e. N = 36). However, final operational drop boundaries, recorder failures and poor recording quality reduced recordings that could be analysed. Thus, recordings were analysed for 7 ARUs distributed across one study site in one treatment area, the Southern Rimutaka Range, and 16 ARUs distributed across two study sites in one non-treatment area, the Northern Rimutaka Range (Figure 4.4). Recorders were 2013 or 2016 DOC ARUs spaced at least 400 m apart, with the exception of two non-treatment recorders spaced 350 m from the next-nearest recorder.

4.2.3.2. Longer-term trends

Archived recordings from the VUW Aorangi Project made using Song Meter™ SM2+, 2013 DOC and 2016 DOC ARUs were available for eight consecutive summer (December-January) and winter (July-September) monitoring seasons from summer 2013-14 to winter 2017 (Hartley, 2017). The number of active recorders and their distribution differed somewhat from year-to-year. In the interest of maximising the reliability of observed long-term trends, audio recordings were sourced only from recording locations for which recordings suitable for analysis were available in at least two monitoring seasons. This resulted in a study design comprising 15 recorder locations distributed across six study sites in the treatment area, the Aorangi Range, and 14 recorder locations distributed across two study sites in the non-treatment area, the Northern Rimutaka Range (Figure 4.5).

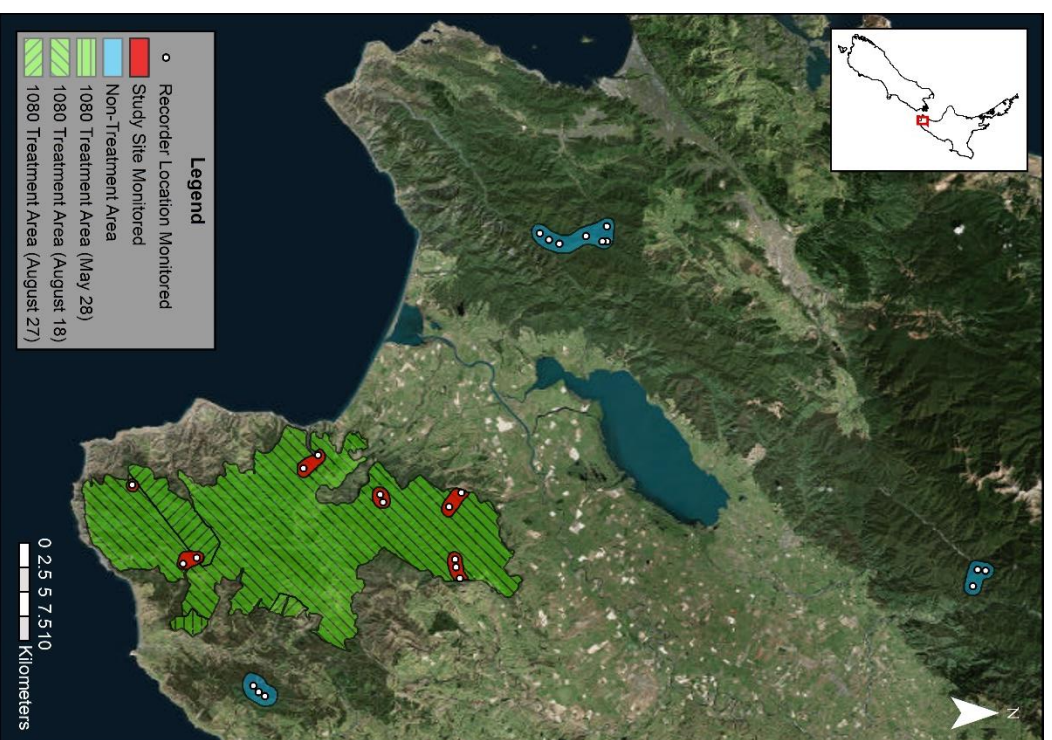


Figure 4.2 Map of study areas, study sites, and recorder locations monitored for short-term monitoring around the August 2014 Aorangi 1080 operation. In non-treatment areas, only one study site was monitored per area and study sites are thus not outlined.

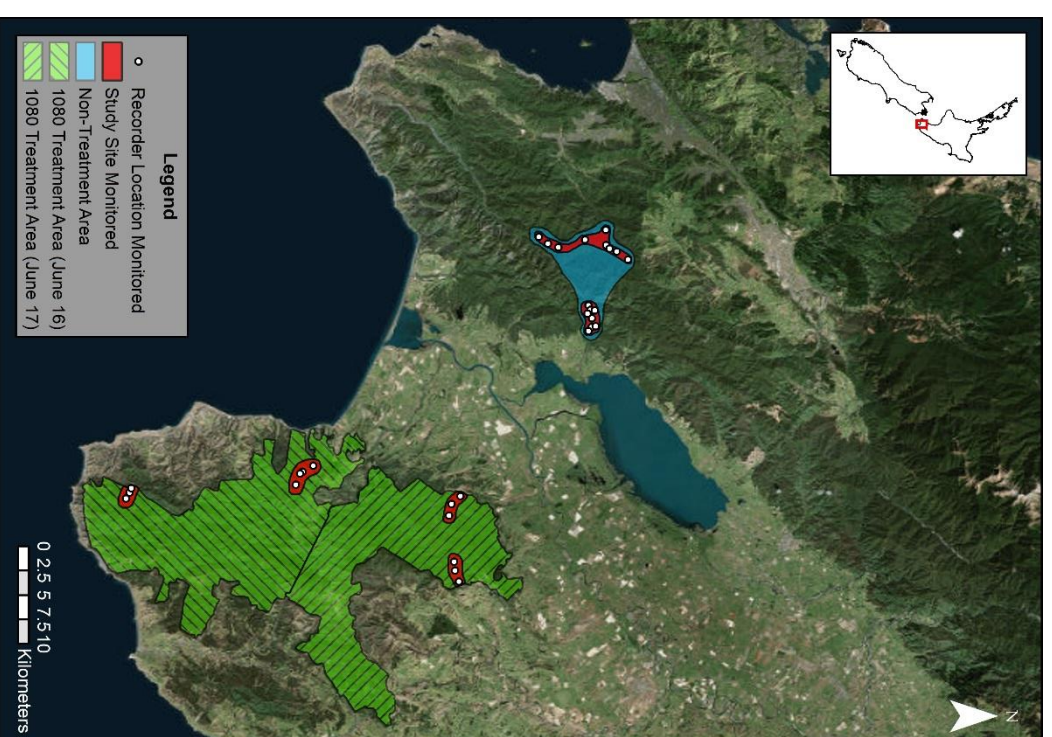


Figure 4.3 Map of study areas, study sites, and recorder locations monitored for short-term monitoring around the June 2017 Aorangi 1080 operation.

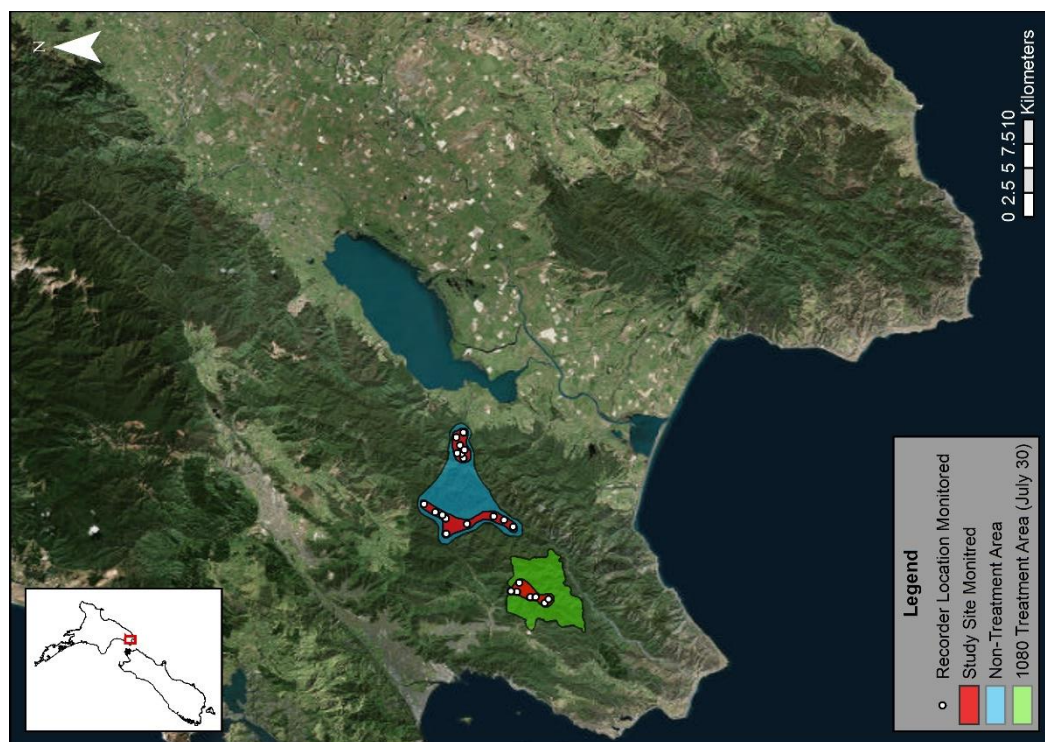


Figure 4.4 Map of study areas, study sites, and recorder locations monitored for short-term monitoring around the July 2017 Southern Rimutaka 1080 operation.

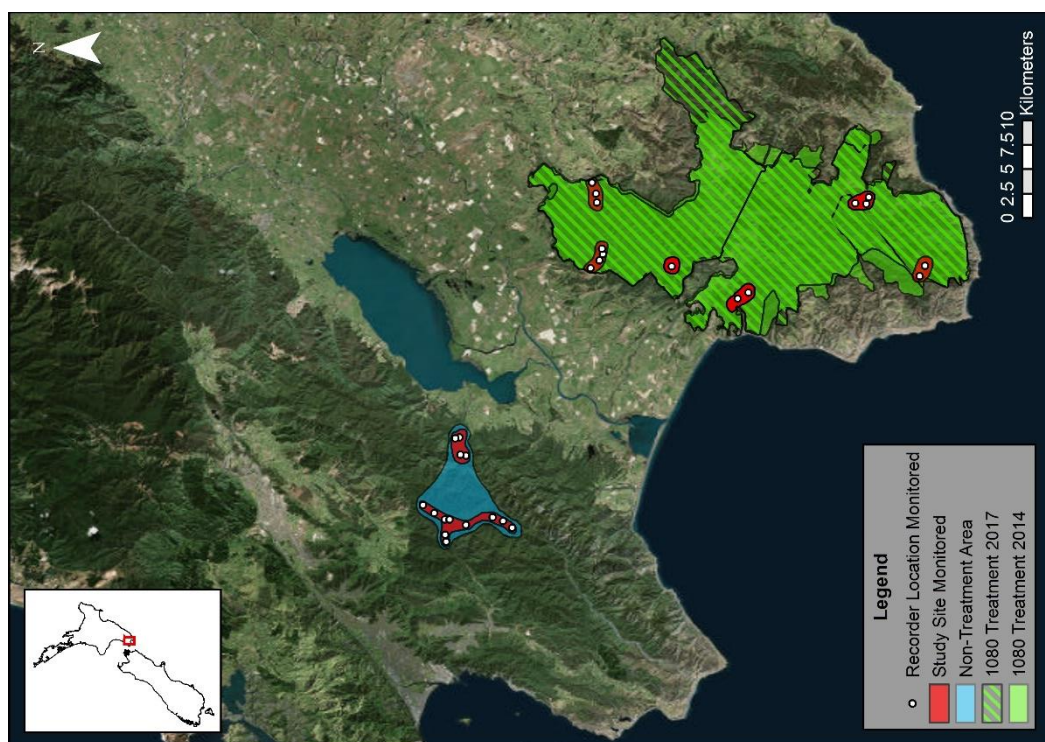


Figure 4.5 Map of study areas, study sites, and recorder locations monitored for long-term monitoring (summer 2013 – winter 2017).

4.2.4. Sound recording specifications

Recorders were mounted to tree trunks at approximately 1.5 m. Song Meter™ SM2+ devices were set to record at a rate of 44,100 samples per second, recording sound frequencies of up to 22.05 kHz. DOC ARUs were set to record at high frequency, recording sound frequencies of up to 16 kHz at 32,000 samples per second. All recordings were made with one microphone unit in mono and saved as 16-bit .wav files.

Slight changes in recording span specifications were made progressively from 2013-2017. However, all recordings were ≥ 5 minutes and recording start times were fixed: all summer recordings began at 2230 h NZDST (Daylight Savings Time), and all winter and short-term BACI recordings began at 2000 h NZST (Standard Time).

4.2.5. Audio recording selection

The automated detector used for recording analyses is suitable for analysis of recordings above a certain quality standard (see Chapter 3). To maximise the reliability of detection results, sample days were selected primarily by recording quality. Sample day selection consisted of two phases: inspection of regional weather data, and subsequent inspection of spectrograms for days that passed the primary weather filter. For each sampling period, rainfall (cumulative mm, 2000-2100 h NZST for winter/BACI, 2200-2300 h NZST for summer) and wind speed (km/h at 2.5-10 m, 2000 h NZST for winter/BACI, 2200 h NZST for summer) data was inspected as the average of five rainfall and five wind speed weather stations, the locations of which are available in Appendix I (Greater Wellington Regional Council, 2018). Days with mean wind speeds of <11 km/h and mean cumulative rainfall of 0 mm were identified as potentially suitable for analysis. Audio recordings from these days were then spectrographically inspected in Raven Lite 2.0 (Bioacoustics Research Program, 2016) and classified as suitable for automated analysis if no heavy rain, clear wind gust exceeding 0.5 kHz, or aeroplane noise was present. The first five minutes of each recording was inspected, as this timespan was to be analysed. In some cases, recordings were shorter than 5 minutes or not present where an ARU recording failed: these were classed as unavailable for analysis.

4.2.5.1. Short-term effects

Recordings were analysed for five weeks before and after each respective 1080 operation. For each operation, the pre-treatment period consisted of the five weeks leading up to the toxic drop date. One week without sampling was allowed after the drop before the five-week post-

treatment sampling period. For each week (Sunday to Saturday), the day with the highest number of suitable recordings was selected for analysis. If less than four days were present between the selected recordings of two consecutive weeks, the second-best day of one of the two weeks was instead selected for analysis. For three sample weeks, insufficient recordings suitable for analysis were available. To increase sample size and maximise representation of recording locations, the Saturday of the preceding week was instead chosen for analysis for these weeks.

This selection process produced maximum potential sample sizes of 250, 300 and 230 audio recordings for the Aorangi 2014, Aorangi 2017 and Southern Rimutaka 2017 operations respectively. After exclusion of recordings unavailable or unsuitable for analysis, final sample sizes after recording selection were 212, 281, and 217 audio recordings (recorder locations and dates sampled are available in Appendix VIII).

4.2.5.2. Longer-term trends

For each sampling season (winter and summer), the span of dates recorded at each recorder location was inspected and compared. The five weeks during which the greatest number of recorders were simultaneously active was selected as each season's main sampling period. Sampling periods were then compared winter-to-winter and summer-to-summer, and an additional sixth week was selected for each period to maximise the consistency of the seasonal time period compared between years.

Due to an intensified sampling effort later in the project, samples were available from late autumn and across winter for the winter 2017 sampling period. This period's six sampling weeks were instead chosen to maximise cross-over with the winter sampling periods of 2014, 2015, and 2016.

After determining sample periods, days were selected for analysis utilising the same methods employed for short-term analyses. The resulting sample size was 566 audio recordings (Appendix IX).

4.2.6. Scoring audio recordings

Each recording was analysed using the automated detector developed for morepork calls in Chapter 3. The *monitor* (Hafner & Katz, 2017) package reports time-specific data for detected events. For each recording analysed, the time-specific detection results of the first five minutes analysed were converted to binary scores for thirty consecutive 10-second sub-samples (i.e. 0-

10 sec, 10-20 sec..., 290-300 sec), consistent with the ‘continuous method’ described by Cook and Hartley (2018). In each sub-sample, the presence or absence of one or more morepork call(s) was scored as a ‘1’ or a ‘0’ respectively. A resulting call prevalence was calculated for each recording as a proportion ranging from 0/30 to 30/30. For example, a recording with at least one calling event detected in 15 out of 30 10-second sub-samples would have a call prevalence of 0.50. This entire scoring process was automated using a series of loop scripts written and executed in R (Appendix VII).

4.2.7. Statistical analyses

4.2.7.1. Short-term effects

For each observation the treatment type (1080 or non-1080), time period (before or after 1080 application), study area, study site, recorder location and recording date were recorded. For the Aorangi 2014 operation, a logistic mixed effects model was fitted to the data with the `glmer` function of the `lme4` R package (Bates *et al.*, 2015) using the formula:

$$\text{Calling Prevalence} \sim \text{Treatment Type} * \text{Time Period} \\ + (1|\text{Study Area}/\text{Study Site}/\text{Recorder Location}) + (1|\text{Date})$$

The same model was fitted for each of the 2017 operations monitored. However, study area was dropped from these models, as only one study area was monitored in each treatment type for these operations. All data values (proportions from 0/30 to 30/30) were weighted by 30, given that 30 binary sub-samples were scored to produce each proportion.

Treatment type and time period were fitted as crossed fixed factors to investigate whether or not changes in calling prevalences from before to after 1080 operations differed between treatment and non-treatment sites. A significant interaction between these factors would indicate an effect of 1080 treatment on changes in calling prevalences from before to after operations. The BACI design of the experiment allowed for comparison of these changes whilst controlling for systematic temporal and spatial effects across each study.

Study area, study site, and recorder location were fitted as random nested factors to account for the spatially nested nature of the sampling design (Figures 4.2, 4.3, 4.4). Recording date was also fitted as a random variable to account for any extraneous variation arising day-to-day.

A type III ANOVA was carried out on the resulting models to produce a chi-square value for the treatment type and time period interaction term. A non-parametric permutation

test was applied to approximate the null chi-square distribution and test the significance of the observed chi-square value for each study. Permutations of the data values were made within recorder locations across recording dates, i.e. values were shuffled within the rows of a recorder by date (rows by columns) matrix. Any missing values (NAs) resulting from inadequate recording qualities or missing recordings were held constant in their location within the dataset across permutations. Five thousand permutations were used to reliably test significance to the 1% significance level (Anderson, 2001).

4.2.7.2. Longer-term trends

For each observation the year, month, recording date, recorder type (SM2+, 2013 DOC Recorder, or 2016 DOC Recorder), study area, study site, recorder location and time since the location's last 1080 operation in months were recorded. It was of interest if mean morepork calling prevalences showed a significant long-term response to 1080 operations. To assess the effect of 1080, a logistic mixed effects model was fit to the data, again using the `glmer` function of the `lme4` R package, with a quadratic formula for the effect of time:

$$\begin{aligned} \text{Calling Prevalence} \sim & \frac{\text{Months Since 1080}}{12} + \left(\frac{\text{Months Since 1080}}{12} \right)^2 \\ & + (1|\text{Study Area}/\text{Study Site}/\text{Recorder Location}) + (1|\text{Year}) + (1|\text{Month}) \\ & + (1|\text{Date}) + (1|\text{Recorder Type}) \end{aligned}$$

Time since 1080 was modelled as a quadratic factor as, if present, any positive or negative effect of 1080 on calling prevalence was expected to eventually decay with increasing time since treatment. Study area, study site, recorder location, recording year, month, date and recorder type were included as random factors to account for any extraneous variation in calling prevalences arising sample-to-sample from spatial or temporal factors, or from potential differences in recording quality between different recorder types used. As with the analyses of short term effects, study area, study site and recorder location were fitted as nested factors to account for the spatially nested nature of the sampling design (Figure 4.5). Backwards stepwise AIC model selection was used to determine the most suitable model. A permutation test was deemed unsuitable for final analyses given the unbalanced experimental design arising from varying recording locations sampled year-to-year, and from varying time since 1080 treatment. Hence, a conservative parametric model was used for model selection and significance testing. Though each data point was derived from 30 sub-samples, each model tested was conservatively weighted by 1 to account for temporal autocorrelation in the data resulting from

the consecutive sampling structure of the ‘continuous method’. It was of interest if time since 1080 had a significant effect, positive or negative, on detected calling prevalences. A Type III Wald Chi-square Analysis of Deviance was carried out on the resulting model of best fit to test for this effect.

Recorder location was also subsequently fitted as a fixed factor for post-hoc analyses to explore the potential effect of differing yearly recorder combinations on detected calling prevalences, due to apparent outlying trends in two seasons with low numbers of recorders sampled. A Tukey test of pairwise comparisons was run on the fixed factor of recorder location to explore this effect for the two years when the least number of recorders were active.

4.3. Results

4.3.1. Short-term effects

Sites in the treatment area showed an increase in mean detected calling prevalences from before to after treatment across the 2014 operation, whereas non-treatment sites showed a very slight decline (Figure 4.6a). However, any interactive effect of treatment with time period on mean detected calling prevalences was not significant ($p = 0.2056$, $\chi^2 = 22.0187$). There was a marginally insignificant effect of treatment on detected calling prevalences across time periods in the 2017 Aorangi operation ($p = 0.0628$, $\chi^2 = 1.8532$); both treatment and non-treatment sites showed a decline in detected calling prevalences from before to after the drop, with the non-treated area showing the slightly greater decline (Figure 4.6b). There was also no significant effect of treatment on calling prevalences across time periods in the 2017 Rimutaka operation

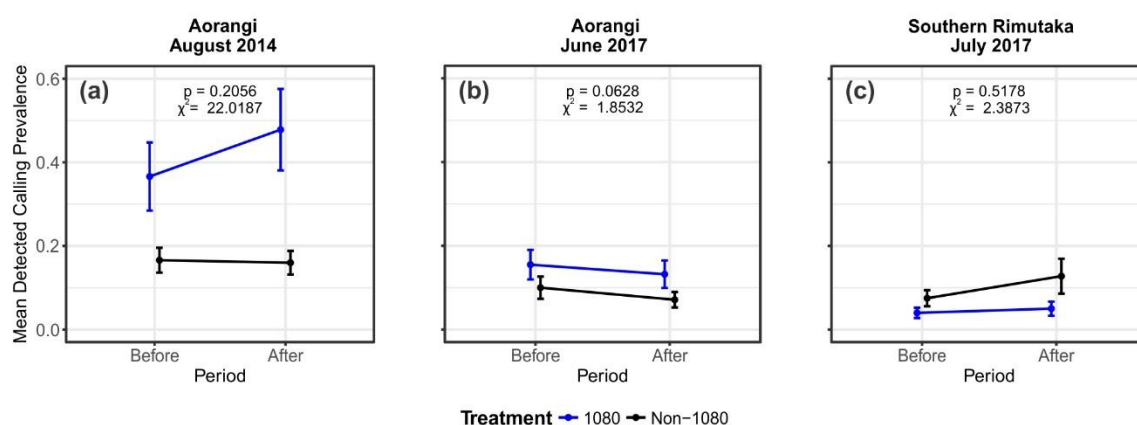


Figure 4.6 Grand mean detected morepork calling prevalences ± 1 SE (n = number of recorder locations sampled) in treatment and non-treatment study areas for zero to five weeks before and two to six weeks after the three respective 1080 operations, and associated chi-square and p-values. Mean and SE estimates are an approximate representation of interactions, whereas associated statistics were calculated via permutation tests accounting for the spatially-nested and temporally replicated nature of the experimental design. Plots share y-axes.

($p = 0.5178$, $\chi^2 = 2.3873$); both treatment and non-treatment sites showed a slight increase in detected calling prevalences from before to after the operation (Figure 4.6c).

4.3.2. Longer-term trends

Detected calling prevalence showed an apparent seasonal trend, with summer sampling periods showing higher mean detected calling prevalences than winter sampling periods. Prevalences detected in summer 2013-2014 and summer 2015-2016 appeared to outlie these trends for the Northern Rimutaka and Aorangi Ranges respectively, with detected calling prevalences appearing markedly higher in the Northern Rimutaka Range and lower in the Aorangi Range than other respective summer seasons (Figure 4.7).

Stepwise model selection found year, month, study area, study site, recorder location, and time since 1080 as a quadratic factor to be useful predictors of detected calling prevalences. Accordingly, date and recorder type were excluded from the final model used to test the effect of 1080 treatment on calling prevalence with time. Analysis of deviance of the final model found time since 1080 to have a significant effect on detected calling prevalences as a quadratic factor ($p = 0.0195$, $\chi^2 = 5.457$, $df = 1$, Table 4.1). A negative quadratic relationship was shown: when modelled with time after a hypothetical operation with months cycled accordingly and all other factors held constant, detected prevalences were predicted to increase for ~40 months (3-3.5 years) after an operation, after which they began to decline. There was a clear seasonal pattern in calling across months sampled (Figure 4.8).

Table 4.1 Table of results for Analysis of Deviance (Type III Wald chi-square test) of fixed effects retained in the best-fitting logistic mixed effects model for long-term data. Asterisks (*) denote significance at the 5% level.

	Chisq	Df	Pr(>Chisq)
(Intercept)	3.917	1	0.0478*
(Months.Since.1080/12)	4.645	1	0.0311*
(Months.Since.1080/12) ²	5.457	1	0.0195*

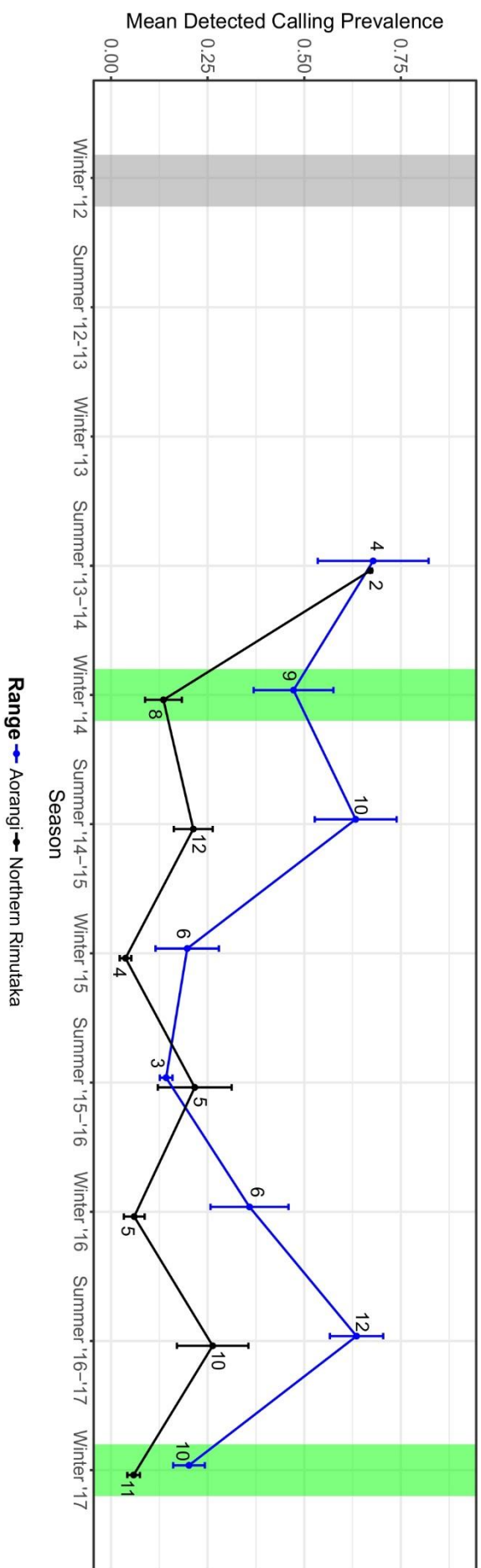


Figure 4.7 Mean detected calling prevalences in the Aorangi (treatment) and Northern Rimutaka (non-treatment) Range study areas across long-term monitoring. Calling prevalences are presented as each area's grand mean for each respective sampling season \pm 1 SE, with n (number of recording locations sampled) presented to the left of Aorangi datapoints and to the right of Northern Rimutaka datapoints. Green bands indicate aerial 1080 operations occurring in the Aorangi range in August 2014 and June 2017, while a grey band indicates the Northern Rimutaka's last aerial 1080 treatment in August 2012.

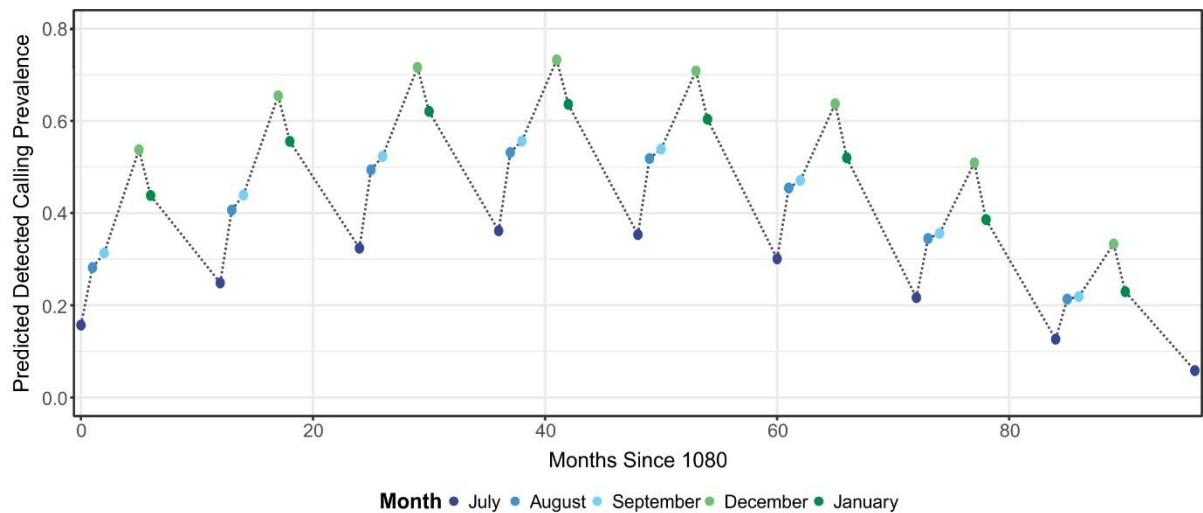


Figure 4.8 Predicted change in mean detected calling prevalence as a function of time (in months) since a hypothetical 1080 treatment occurring in July for a randomly selected recorder location. Predictions are made only for months in which sampling was carried out, with year and location factors held constant. Annual cycles in calling prevalence are evident and superimposed over a long-term positive response to 1080 treatment peaking at ~40 months (3-3.5 years) after treatment.

Recorder location showed a significant effect when modelled as a fixed factor ($p < 0.001$, $\chi^2 = 97.54$, $df = 28$, Table 4.2). In post-hoc Tukey testing carried out to explore this effect, no significant difference in calling prevalence was found between any recorder locations sampled within the Northern Rimutaka study area (Table 4.3). Accordingly, no significant difference in mean calling prevalence existed between recorders sampled in the lesser-sampled summer 2013-14 and 2015-16 seasons (Figure 4.9a). Contrastingly, a significant difference in calling prevalence was found for seven pairwise Aorangi recorder location comparisons (Table 4.4). With respect to recorders sampled in the lesser-sampled summer 2013-14 and 2015-16 seasons, calling prevalence was significantly higher at one location sampled only in the 2013-14 season, WAI L1 100m, than at MAN L1 100m (sampled 2015-16; z -ratio = -4.127 , $p = 0.012$) and WHA L2 100m (sampled 2013-14 and 2015-16; z -ratio = 3.815 , $p = 0.038$, Table 4.4, Figure 4.9b).

Table 4.2 Table of results for Analysis of Deviance (Type III Wald chi-square test) of fixed effects retained in the best-fitting logistic mixed effects model for long-term data, with recorder location fitted as a fixed effect. Asterisks (***) denoted significance at the 0.1% level.

	Chisq	Df	Pr(>Chisq)
(Intercept)	0.088	1	0.7668
(Months.Since.1080/12)	3.445	1	0.0635
(Months.Since.1080/12) ²	3.762	1	0.0524
Recorder Location	97.54	28	<0.001***

Table 4.3 Table of raw Tukey test p-values for pairwise comparisons of recorder locations monitored long-term in the Northern Rimutaka Range.

	ORO 5km	ORO L1 100m	ORO L2 100m	ORO L3 100m	ORO SM1	ORO SM2	ORO SM3	ORO SM4	ORO SM5	ORO SM6	WRM L2 100m	WRM L3 100m	WRM L4 100m	WRM L4 450m
ORO 5km														
ORO L1 100m	0.996													
ORO L2 100m		0.999												
ORO L3 100m			1.000											
ORO SM1				1.000										
ORO SM2					1.000									
ORO SM3						1.000								
ORO SM4							1.000							
ORO SM5								1.000						
ORO SM6									1.000					
WRM L2 100m										0.846				
WRM L3 100m											1.000			
WRM L4 100m												1.000		
WRM L4 450m													1.000	

Table 4.4 Table of raw Tukey test p-values for pairwise comparisons of recorder locations monitored long-term in the Aorangi Range, p-values significant at the 5% and 10% level are shown in italic bold and bold respectively.

	BUL L1 100m	BUL L2 100m	BUL L3 100m	MAN L1 100m	MAN L2 100m	PIN L1 100m	PIN L2 100m	TAU L3 100m	WAI L1 100m	WAI L1/2	WAI L2 100m	WAI L4 100m	WHA L1 100m	WHA L2 100m	WHA L3 100m
BUL L1 100m															
BUL L2 100m	1.000			<i>0.047</i>	0.394	1.000	0.775	1.000	1.000					<i>0.097</i>	0.749
BUL L3 100m		1.000		<i>0.006</i>	0.085	1.000	0.412	1.000	1.000	1.000	0.196	1.000	1.000	<i>0.019</i>	0.356
MAN L1 100m			1.000			1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000
MAN L2 100m				1.000		0.755	1.000	0.767	<i>0.012</i>	0.419	1.000	<i>0.009</i>	0.251	1.000	1.000
PIN L1 100m						1.000	1.000	0.999	0.167	0.960	1.000	0.107	0.888	1.000	1.000
PIN L2 100m							1.000	1.000	1.000	1.000	0.980	1.000	1.000	0.917	1.000
TAU L3 100m								1.000	0.534	0.988	1.000	0.396	0.980	1.000	1.000
WAI L1 100m									1.000		0.969	1.000	1.000	1.000	1.000
WAI L1/2										1.000	0.255	1.000	1.000	<i>0.038</i>	0.527
WAI L2 100m											0.851	1.000	1.000	0.582	0.991
WAI L4 100m												1.000	1.000	0.800	1.000
WHA L1 100m													1.000	<i>0.021</i>	0.343
WHA L2 100m														0.981	1.000
WHA L3 100m															1.000

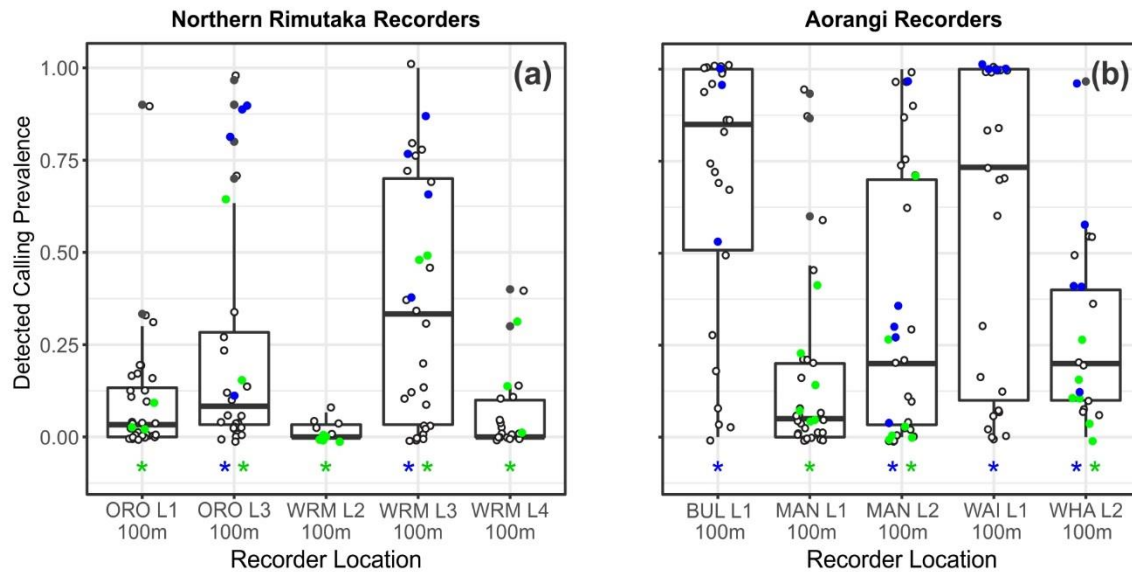


Figure 4.9 Boxplots of calling prevalences detected at recorder locations monitored in the summer 2013-14 and summer 2015-16 seasons showing (a) the non-significant differences in prevalences detected in recordings from recorders locations monitored in the Northern Rimutaka Range and; (b) the significantly higher prevalences detected in recordings from WAI L1 100m than in recordings from MAN L1 100m and WHA L2 100m in the Aorangi Range, particularly in summer 2013-14. Detected prevalences from individual days sampled are overlaid on plots: summer 2013-14 days are blue; summer 2015-16 days are green; and days from other seasons are hollow-fill points. Grey points indicate outliers. Blue asterisks and green asterisks indicate recorder locations monitored in the summer 2013-14 and summer 2016-17 seasons respectively.

4.4. Discussion

The results of this study showed no evidence for a significant effect of 1080 treatment on trends in the calling prevalence of morepork from zero to five weeks before to one to six weeks after any of the three aerial 1080 operations monitored. If calling prevalence can be interpreted as an index of abundance, these results suggest that morepork populations did not suffer significant population losses from secondary poisoning over the course of any of these three operations. Only one other study has applied bioacoustics to the short-term monitoring of morepork before and after a 1080 operation. Pierce and Montgomery (1992) found the number of individuals detected in an area treated with 1080 to double in contrast to a decline shown in a non-treatment area. In-field call counts conducted by multiple observers were utilised, however, and it is noted that the non-treatment decline observed may have been in part attributable to a ‘new’ listener being employed in the area. The results of the current study are not subject to the potential effects of multiple observer biases. Acoustic recordings offer the opportunity to minimise observer bias (Hobson *et al.*, 2002), and the use of an automated

detector of a known performance standard may further support the reliability of trends detected here (Blumstein *et al.*, 2011).

As aforementioned, twenty-seven radio-tagged morepork have so far been monitored in treatment areas over five 1080 operations. These studies suggest a low mortality rate for the species, with one individual subject to lethal secondary poisoning after a 1996 carrot bait operation in Pureora Forest Park (Greene *et al.*, 2013; Powlesland *et al.*, 1998). Of these individuals, 21 were monitored over operations using cereal baits (Greene *et al.*, 2013; Powlesland *et al.*, 1999; Walker, 1997). As baiting practices change, further study is required to reliably estimate any mortality rates resulting from modern aerial 1080 operations. All operations monitored here utilised cereal baits. Furthermore, all incorporated deer repellent in baits, a relatively new practice first trialled in the early 2000's, the effects of which remain poorly studied for native birds (Morriss *et al.*, 2016; Speedy, 2005). Preliminary findings of more recent, yet-to-be-published radio-tagging surveys have been reported to find no mortality for morepork monitored over modern operations (G. Elliot, DOC, unpubl. data, as cited by Brown *et al.*, 2015; Elliott & Kemp, 2016). In contrast to the direct estimation of mortality rates achievable with radio-tagging surveys, call-counts are limited to detecting large changes in populations due to the potential effects of various factors on conspicuousness (Atkinson *et al.*, 1995; Spurr & Powlesland, 1997). Nevertheless, the results of the current study suggest that if morepork populations incur any mortality through secondary poisoning over aerial 1080 operations using modern cereal baiting practices with deer repellent, rates are so low as to not affect broad-scale population trends.

The general trends observed in calling prevalences from before to after operations likely result from monthly and seasonal trends in morepork calling prevalences related to their breeding cycle. The morepork breeding season spans from September to February (Higgins *et al.*, 1999). A pattern of low calling rates throughout winter has been described to increase in late August/early September with the beginning of this breeding season (Brighten, 2015; Imboden, 1985). Broad scale, the before-and-after trends in calling prevalences observed in short-term studies here corroborate with these patterns. Furthermore, the trends observed fit monthly calling patterns observed by Brighten (2015) on Mokoia Island in 2013. Specifically, before and after samples in the Aorangi 2017 operation spanned May/June and June/July respectively. The general decline in calling prevalence detected across this period is consistent with a decline in mean morepork call syllables observed by (Brighten, 2015) from May to July. In contrast, a minimal change or increase in calling prevalences was notable from before to

after samples in the Aorangi 2014 and Rimutaka 2017 studies. Before-to-after sampling periods for these studies spanned July/August to August/September, and June/July to August/September respectively. These increases are also consistent with trends described by Brighten (2015), who reports a decrease in average syllable numbers per night from June to August, followed by an increase from August to September.

Longer-term, the results of this study showed a significant positive response in the prevalence of morepork calling to aerial 1080 operations. This peaked with an increase in prevalence of an additional ~20% at ~40 months, after which predicted prevalence declined. As an index of relative abundance, this would suggest that morepork populations have so far benefited from aerial 1080 operations in the Aorangi Range. Calling rates have been compared between managed and unmanaged areas after three years of ongoing brodifacoum treatment to suggest a positive long-term effect of treatment on morepork populations (Fraser & Hauber, 2008). With respect to 1080, the population trends, nesting success and survival of morepork is reportedly being monitored more closely within a wider study evaluating the responses of native bird species to aerial operations (Brown *et al.*, 2015). However, no other study could be found in the current scientific literature that has monitored the long-term response of morepork to aerial 1080 operations in any way. Thus, the current study is the first to suggest a realised long-term benefit to morepork populations from aerial 1080 operations.

A number of mechanisms may contribute to and explain this apparent benefit. It has been suggested that the reduced availability of small rodents as a food source resulting from pest control operations might have negative implications for morepork (Pierce & Montgomery, 1992; Stephenson & Minot, 2006). However, recent studies contest this concern, showing morepork to be primarily insectivorous (Denny, 2009; Haw & Clout, 1999; Haw *et al.*, 2001). Hence, introduced small mammals may in fact compete with morepork for their primary food source (Brighten, 2015; Denny, 2009). It has accordingly been suggested that increases in invertebrate and possibly mouse populations resulting from reduced rat populations may confer a benefit to morepork through increased prey availability (Fraser & Hauber, 2008). Such post-treatment increases have been noted for mice at some sites monitored in the VUW Aorangi Project (Hartley, 2017). Additionally, morepork are commonly a hole-nesting species (Stephenson, 1998; Stephenson & Minot, 2006), making eggs, chicks and nesting adults vulnerable to predation by introduced mammalian predators (Brighten, 2015; O'Donnell, 1996). Possums have also been suggested to compete with hole-nesting species for nesting spaces (Wilson *et al.*, 1998). It is therefore possible that the significant reduction in introduced

mammals resulting from an aerial 1080 operation may confer a benefit for morepork through increased survivorship and reproductive success resulting from a combination of reduced predation pressure, reduced food competition, increased food availability and reduced competition for nest-sites. Which of these factors are at play and how they interact with morepork population trends in the Aorangi Range, however, would require further study to determine.

The apparent seasonal patterns observed in long-term calling prevalence for both the Aorangi and Northern Rimutaka study areas of high calling rates in summer and low calling rates in winter corroborate with seasonal patterns in morepork calling associated with their breeding cycle (Brighten, 2015; Imboden, 1985). The apparent outlying low calling prevalence observed in the Aorangi study area in summer 2015-16 is likely attributable to the small number of recording locations that were available for sampling in this season. Pairwise recorder location comparisons showed a significant difference in mean calling prevalences at the 5% level between a recorder active only in summer 2013-14 and two recorders active in summer 2015-16. In both cases, sites sampled in the 2015-16 summer season had lower mean calling prevalences. Thus, given the higher mean prevalences observed in all other seasons, the small combination of locations monitored in summer 2015-16 may underestimate mean calling prevalence for this season. Contrastingly, no difference in mean calling prevalences was found between recorders sampled in these seasons for the Northern Rimutaka study area. The difference observed between these seasons may instead be attributable to year-to-year variation in calling prevalences and/or the number of suitable recordings available for calling prevalence estimates. These apparent differences highlight the importance of sample size (number of ARUs) and consistency of sampling locations in treatment and non-treatment areas when monitoring long-term with ARUs. This is particularly pertinent considering the variable impact that local weather conditions may have on the suitability of recordings for analysis at different recorder locations. Nevertheless, this was beyond the control of this study given that archived data was analysed. The otherwise consistent seasonal patterns in calling prevalences detected using the data available is encouraging. There would be great benefit in the continued monitoring of sites monitored in the winter 2017 sampling season over the next four to five years to determine the consistency of these long-term patterns and responses to 1080 operations.

It is important that the potential influence of a number of limitations are considered regarding this study. Firstly, morepork calling rates vary throughout the year with the breeding

season, day-to-day, and hour-to-hour (Brighten, 2015; Fraser & Hauber, 2008; Stephenson, 1998). The analysis of multiple recordings per recording location for each sampling period, and the consistency of time of day and dates sampled in both the short and long-term analyses of this study should minimise any potential effects resulting from such variability. Moreover, the effect of any remaining variability arising from these sources should be further reduced by the inclusion of temporal and spatial random variables in statistical modelling. This is particularly advantageous, as the relative influence of environmental variables on calling prevalence is not well understood. It has been suggested that calling rates may be influenced by wind, moon, and cloud conditions (Fraser & Hauber, 2008). However, Stephenson (1998) anecdotally reports calling frequency to vary widely night-to-night irrespective of wind or rain conditions, with much higher calling frequencies some nights than others. The ability of ARUs to standardise these potential error sources greatly enhances the reliability of inferences drawn in the current study.

Another issue to consider is the potential for detection of the same individual in recordings across multiple recorders. It has been reported that morepork may cover large distances in a short space of time. Imboden (1975) found individuals with territories estimated at up to 5.3 ha in size (approximately 200-250m in width) to cross their entire home range multiple times in the space of an hour. More recently, Pryde and Greene (2016) estimated the mean home range of morepork in Eglinton Valley to be 43.5 ha (i.e. 600-700m in width for a reasonably round home range). Accordingly, Fraser and Hauber (2008) warn that the tendency of morepork to cover such distances may affect the accuracy of auditory surveys where auditory surveys conducted are long. The short five-minute recording span analysed in this study (recorded simultaneously across all recorders) minimises the potential for detection of the same individual across multiple recorders resulting from such events.

Another potential source of double-detection exists in recorder spacing. Pryde and Greene (2016) found the rate of morepork calls recorded by ARUs in beech forest to decline with increasing distance from recording units after 150 m, with 46% of calls attenuated at 200 m and 0% recorded at 250 m. Consequently, morepork calls detected by recording unit pairs less than 500 m apart may be from the same individual, and inadequate spacing of recorders may produce non-independent estimates of calling prevalence. In the present study, recorders were set up on biodiversity monitoring lines 450m in length to maximise future comparability of results to pest-tracking data. Hence, some recorder pairs were closer than this recommendation; at minimum distance, five recorder pairs were spaced 350 m apart. It is thus

possible that some calls were double-detected in data. This considered, double-detected calls are likely to occur near the mid-point (175 m at minimum spacing in the current study) between recorder pairs. Given the attenuation of recording rate with distance, these calls are likely faint in recordings. Considering a negative relationship exists between signal detection and decreasing signal strength in automated detection (Digby *et al.*, 2013; Swiston & Mennill, 2009) and the high precision of the detector employed here, it is arguable that such double-detections contribute to a small portion of the data in this study (Ulloa *et al.*, 2016).

The use of an automated detector for the analysis of recordings in this study must also be considered. As already highlighted, the detector employed has an estimated sensitivity of 39.9%. Thus, not all calls will have been detected in analyses. However, the focus of this study was not on changes in total call numbers, but relative changes in calling prevalence. Given that, on average, this detector's performance should be consistent across all recordings analysed, the overall trends detected at a 39.9% sensitivity should be representative of those that would be found at a 100% sensitivity with an equivalent FPR. Of equal importance is the 2.53% FPR of the detector: it can be expected that 2.53 in 100 10-second sub-samples scored as active in the data produced will be a FP event. This low inflation of detected calling should also occur in a random unbiased manner across the recordings analysed and have no significant impact on the overall trends observed. Unless subject to variable background noise across recordings, detector performance should be consistent. The recording selection process employed in this study should both minimise and standardise any effect of background noise on both TP and FP detections. The consistency of the long and short-term trends found both within this study season-to-season and with trends described in the wider scientific literature gives encouraging support for their reliability. Nevertheless, the sensitivity of this detector is based on individual call detection. Calls may differ in quality, amplitude and, consequently, detectability across recordings. Thus, future comparison of calling prevalences produced by automated and manual analyses of recordings using the 'continuous method' of scoring employed in this study would be useful for further assessment of the reliability of trends observed.

Notwithstanding these potential limitations, the current study is the first to investigate the response of morepork to aerial 1080 operations both short- and longer-term. The non-significant short-term effect of treatment on calling prevalence as an index of abundance across all three operations strongly indicates that aerial 1080 operations did not have an immediate large-scale negative effect on the morepork populations studied. Furthermore, long-term monitoring of this study provides evidence for a net positive effect of aerial 1080 treatment on

morepork populations in the Aorangi Range over a timescale of 2-4 years. Reduced competition from mammals, increased prey availability, and reduced predation upon morepork chicks are all mechanisms that may explain this response. Given the lack of literature that currently exists regarding the response of morepork populations to aerial 1080 operations, it is advisable that monitoring is both continued for the populations studied here and extended to other treatment and non-treatment areas across New Zealand. The growing availability of ARUs and development of recognition software continues to advance the ease with which this can be achieved. Simultaneous monitoring of pest and invertebrate populations and morepork nesting success across future operations would facilitate better understanding of the dynamics at play in observed long-term responses.

Chapter 5 - General Discussion

5.1. Responses of avifauna to aerial 1080 operations

5.1.1. Short-term community-level effects

The theory that forests fall silent post-1080 treatment is a key concern in the debate around aerial 1080 operations. Bioacoustic monitoring in Chapter 2 investigated this issue directly at the community level. Comparing from before to after the 2017 Aorangi operation, the amount of silence in the treatment area increased very slightly, in the same way that silence increased in the non-treatment areas (Figure 2.4a). These results probably reflect a general reduction in bird calling behaviour as autumn progresses to winter (Hartley, 2012). They are consistent with and build on those of Cook (2017), who found no significant difference in trends of silence for bird communities between treatment and non-treatment areas across the 2014 Aorangi operation. Further, they highlight the importance of the BACI experimental design: monitoring of only the treatment site may have led to alternative interpretations. Extending from the Aorangi Range, trends in silence were found to differ significantly between treatment and non-treatment areas across the 2017 Southern Rimutaka operation. However, treatment areas showed an increase in birdsong, whereas non-treatment areas showed an increase in silence (Figure 2.4b). As highlighted in Chapter 2, this pattern may reflect altitudinal biases in sampling and should be treated conservatively. However, taken at face value, the patterns observed across this operation provide evidence contrary to those suggested by the silent forest theory. Together, these bioacoustic studies provide no support for the anecdotal silent forest theory in diurnal forest bird communities of the lower North Island of New Zealand.

Proponents of 1080 propose that forests are indeed silent, but that they do not *fall* silent; rather, low birdsong is attributable to high predation pressure of introduced mammals limiting bird communities, and may be observed both shortly before and shortly after an aerial 1080 operation (Hansford, 2016; Toki, 2013). Such dynamics might explain the trends observed across the two Aorangi operations monitored to date. The reduced bird mortality rates with baiting practice changes (Eason, 2002; Eason *et al.*, 2011) and low native bird mortality findings over operations using modern baiting practices would support this theory (Morriss *et al.*, 2016). As Cook (2017) denotes, though many studies have investigated the mortality of bird species over 1080 operations (Morriss *et al.*, 2016; Spurr & Powlesland, 1997; Veltman & Westbrooke, 2011), little work has been done to quantitatively assess the silent forest theory proposed by 1080 opponents. As demonstrated both here and by Cook (2017), bioacoustic studies utilising autonomous recording units (ARUs) provide a direct means to achieve this.

Continued monitoring of the areas studied here would facilitate an evaluation of the consistency of these results across independent treatments. Furthermore, bioacoustic monitoring of other areas across New Zealand would facilitate an investigation of their consistency for different bird community compositions across the country. Given the controversy that exists around 1080, the probable increase in its use with Predator Free 2050 aims (Parliamentary Commissioner for the Environment, 2013) will likely both necessitate and provide opportunities for such studies.

5.1.2. Species-level effects

A total of 19 native bird species have been found dead following aerial 1080 operations (Spurr, 2000). However, recent studies of operations employing modern baiting procedures suggest a negligible negative impact of operations on native bird species (Morris *et al.*, 2016). Corroboratively, manual analyses of diurnal recordings in Chapter 2 found no significant difference in the trends of calling prevalence in treatment and non-treatment areas for eight of the ten taxa statistically analysed across the 2017 Aorangi operation (Figure 2.5a-e, h-j); and six of the eight taxa analysed for the 2017 Southern Rimutaka operation (Figure 2.6a, b, d, f-h). Automated analyses of nocturnal recordings in Chapter 4 also found no significant difference in the morepork calling prevalence trends between treatment and non-treatment sites from shortly before to shortly after any of the three operations monitored (Figure 4.6).

Of the four taxa that showed a significant short-term difference in calling prevalence across any of the operations monitored, only one (chaffinch) showed a significant decline in the respective treatment area. Relative to non-treatment areas, calling prevalences for the chaffinch in treatment areas were found to decrease from the five-week before to the five-week after period monitored for the 2017 Aorangi operation (Figure 2.5g). Given that chaffinch are an granivorous species considered at risk to poisoning (Miller & Anderson, 1992) and have been found dead following modern operations (Morris *et al.*, 2016), it is possible that this decline may indicate a direct negative effect of the 2017 operation on Aorangi chaffinch populations.

For all other taxa showing a significant interaction, an increase or lesser decline was observed in treatment areas relative to non-treatment areas. As discussed in detail in Chapter 3, and briefly below regarding considerations around bioacoustic studies, these results were unlikely indicative of a negative impact for these species. Given that chaffinch are an introduced species, the results of this study suggest that none of the native species' populations

monitored over these operations suffered significant mortality as a result of aerial 1080 treatment. This corroborates with, and contributes to, the wider literature regarding the potential negative impacts of modern aerial 1080 operations on native bird species (Brown *et al.*, 2015; Morriss *et al.*, 2016). In particular, the results here address the lack of literature regarding the potential susceptibility of morepork to aerial 1080 operations (Greene *et al.*, 2013). Furthermore, in combination with those of Cook (2017), they address the susceptibility of bird populations to aerial 1080 treatment using modern baiting methods. Across all operations, both prefeeding and deer repellent were utilised; these relatively new methods might increase exposure rate and risk to native bird species and are relatively understudied. However, the results here suggest a negligible risk of operations utilising these methods, consistent with that reported by Morriss *et al.* (2016).

Given that pest-control operations are carried out in the interest of net positive effects for native species, quantifying the long-term response of native species' populations is an important consideration. It is generally accepted that, even if native species experience short-term population losses, a positive long-term impact of aerial 1080 treatment is generally shown (Brown *et al.*, 2015; Weaver, 2006). Bioacoustic monitoring by Fea (2018) using ARU recordings from the Aorangi and Rimutaka Ranges found such positive responses after the 2014 Aorangi operation: relative to before the treatment, smaller-bodied rifleman and tomtit, and medium-bodied bellbird exhibited a significant increase in calling rates 1.5 years after 1080 treatment. Thus, the short-term decline of tomtit exhibited in bioacoustic monitoring by Cook (2017) was outweighed by benefits two breeding seasons later. Furthermore, bellbird showed a sustained positive response evident after 2.5 years, with tūī and kererū, the largest-bodied species, also exhibiting positive responses over this timeframe (Fea, 2018). The long-term monitoring of large-bodied morepork (~175 g (Heather & Robertson, 2005)) in Chapter 4 showed a similar positive long-term response: calling prevalence showed a significant quadratic relationship with time since 1080 treatment, predicted to peak after 3.5 years. This is the first study to quantify a multi-year response of morepork to aerial 1080 treatment. Collectively, the results of Cook (2017), Fea (2018) and this thesis suggest a net positive impact of aerial 1080 treatment on native avifauna in the lower North Island.

Species' responses to 1080 treatment may vary with differing application methods. Currently, the use of prefeeding, deer repellent, and differing sowing methods is relatively understudied in terms of its direct and indirect impacts on bird populations (Morriss *et al.*, 2016). Furthermore, regimes of sustained three-yearly treatment have been recommended, and

are carried out, in the interest of providing a safe ‘breeding window’ for native species in at least one of every three years (Brown & Urlich, 2005; DOC 2014), but results of this approach have yet to be widely reported in the published literature (Brown *et al.*, 2015; Veltman & Westbrooke, 2011). The Aorangi Range has received two of three aerial applications in a 10 year treatment programme (TBfree, 2016), with the next application most likely to occur in the winter of 2020. Continued bioacoustic monitoring in the areas studied would provide an opportunity to extensively assess the net outcomes for native bird species in this area resulting from this regime. For morepork, the capacity to continue this monitoring is enhanced by the ability to employ automated analyses of recordings. At a broader scale, the potential to expand bird species monitoring to other areas receiving differing treatments is greatly enhanced by the capacity of bioacoustics to simultaneously collect data over large spatial and temporal scales, with little increased field effort relative to traditional methods. Monitoring might also be extended to other acoustically active species that are difficult to manage with traditional field methods. For example, DOC ARUs now have the capacity to record the acoustic frequency of bat calls.

5.2. Automated monitoring

With the increasing use of ARUs, there is a growing demand for automated extraction of data from acoustic recordings, particularly in the context of long-term monitoring (Blumstein *et al.*, 2011; Joshi *et al.*, 2017). In Chapter 3, a detector with moderate sensitivity and high precision for ‘more-pork’ calling events of morepork, New Zealand’s only extant native owl, was successfully developed for this purpose. This detector is capable of detecting 39.9% of calling events in field-recordings with high precision (0.939), classing only 2.53% of non-calling events falsely as calls. The sensitivity of this detector (0.399) is somewhat lower than that of detectors developed for other species’ calls in the wider scientific literature: a recent review of 68 studies in the bioacoustic literature by Knight *et al.* (2017) found the mean sensitivity of 12 single-species recognisers tested on un-edited field recordings to be 0.60. However, its precision (0.939) is notably higher than the mean precision (0.71) of studies reported in the same review. In detector development, there is an inevitable trade-off between sensitivity and precision: detectors maximising sensitivity will have lower precision and vice versa (Ulloa *et al.*, 2016). The detector here was developed for the unsupervised monitoring of trends in morepork calling prevalences, and thus purposefully developed to have a high precision at the expense of sensitivity.

The utility of automated detection in bioacoustic monitoring is dependent on the time required to create, run, and assess the performance of a detector relative to the amount of time otherwise required for manual analyses (Knight *et al.*, 2017). Detector development, however, can be a subjective, iterative, and time consuming process (Katz *et al.*, 2016a). The method that was conceptualised, developed and employed in Chapter 3 (Figure 3.5) aimed to provide a directed and efficient means to develop a parsimonious template-based detector. When employed to create the morepork detector, this process identified a candidate list of 42 *morepork* call templates from an initial pool of 187 templates sourced from field recordings. These 42 templates were reduced to a final detector comprising 10 templates. At the level of precision chosen during development (~ 0.975), these 10 templates were collectively capable of making 89 of the 90 detections possible if all 42 templates were used, in approximately a quarter of the processing time. Thus, the method developed was highly effective as a means to achieve parsimony in the development of this template-based detector. Highlighting the potential limitations associated with subjective template development, only one of the twelve calls that were subjectively trialled as initial templates in this process remained in the final detector. The assessment of this method's utility in other template-based detector development projects for further species is recommended. If consistently effective across projects, this process has the potential to streamline and enhance the efficacy of bioacoustic monitoring employing template-based detectors for the detection of other acoustically active species.

5.3. Limitations and considerations

An inherent limitation of bioacoustic studies is the assumption that the abundance of calls observed is related to species' abundances, with greater prevalence of calls associated with greater species' abundance (Royle, 2004). The monitoring results of both diurnal bird species in Chapter 2 and morepork in Chapter 4 are dependent on this assumption. This assumption has been investigated and found to hold true for a range of species (Borker *et al.*, 2014; Nelson & Graves, 2004; Thompson *et al.*, 2010). It has not been widely studied, however, for native New Zealand bird species: only one known study has investigated a relationship between conspicuousness and density, finding the detectability of grey warblers and robins to vary in proportion to density for five-minute bird counts (Gill, 1980; Hartley & Greene, 2012; Spurr & Powlesland, 2000). The reliability of inferences that can be drawn from future bioacoustic studies in New Zealand would greatly benefit from an investigation of this assumption, particularly around pest control operations. Such studies would broaden the extent to which bioacoustics can be applied to reliably monitor species' responses and inform conservation

management. This considered, the assertion that forests falling silent is indicative of a negative effect of aerial 1080 treatment on native bird species is itself fundamentally dependent on this assumption. Thus, the results presented here can be quantitatively employed to address this concern at the species level.

An important caveat highlighted in Chapter 2 is the potential for species ecology and behaviour to limit the inferences that can be drawn from bioacoustic monitoring. Across the 2017 Aorangi operation, bellbird calling prevalences were found to show a significantly lesser decline in treatment areas relative to non-treatment areas (Figure 2.5f). The same pattern was found for bellbird/tūī across the 2017 Southern Rimutaka operation (Figure 2.6c). Such observed patterns could result from significant mortality events in territorial species if remaining individuals increase call rates to establish new territories and mates (Spurr & Powlesland, 2000). Despite this theoretical possibility, bellbird and tūī deaths following operations using both old and modern baiting methods appear unlikely, as no bellbird have been found dead, and only two tūī have been found dead after aerial 1080 operations historically (Morris *et al.*, 2016; Spurr & Powlesland, 1997). The observed pattern might more likely reflect the movement of these species to follow seasonally available nectar resources (Rasch & Craig, 1988; Spurr *et al.*, 2010; Stewart, 1980 as cited by Craig *et al.*, 1981), which may unevenly influence detection probabilities at recorder locations monitored across an operation.

A significant increase in the prevalence of calls was found for tomtit in the treatment area across the 2017 Southern Rimutaka operation (Figure 2.6e). This study had an unbalanced design, however, as recorders adjacent to final 1080 drop boundaries could not be unambiguously designated as either inside or outside the treatment zone, leading to the exclusion of a treatment study area from sampling. Consequently, only high elevation (>500 m) treatment recorder locations were sampled, whereas both high and low elevation (<280 m and >450 m) recorder locations were sampled in the non-treatment area. The onset of the tomtit's breeding season is marked by an increase in calling (Heather & Robertson, 2005) and may be affected by elevation, being delayed at higher elevations (Knegtmans & Powlesland, 1999). Considering this potential source of variation, the pattern observed might reflect such an elevational effect of breeding behaviour on calling prevalence.

The potential influence of behavioural factors on calling patterns should receive careful consideration in studies employing bioacoustic monitoring, and is an inherent limitation of any

study not directly tracking individuals. Future bioacoustic studies would benefit from the investigation of species' calling prevalences in relation to factors such as elevation and seasonal resource distribution that might affect detection probabilities and, accordingly, the distribution of ARUs accounting for such factors. The simultaneous use of radio-telemetry to track individuals should also be considered, especially for species exhibiting behaviours likely to influence detection probabilities that cannot be accounted for in bioacoustic monitoring.

These considerations might be incorporated in developing a standardised methodology for monitoring New Zealand birds using ARUs. As already highlighted, bioacoustic monitoring using ARUs has significant potential as a monitoring method. However, no standardised methodology currently exists, and its development would enhance both the quality and cross-comparability of bioacoustic data collected using ARUs. The necessity of such a methodology will only increase as this monitoring method increases in both accessibility and use. With respect to the limitations encountered here, there is a need to determine which species may and may not be suitable for acoustic monitoring as a result of their behavioural ecology, and to establish what factors, such as resource distribution and elevation, are likely to affect the applicability of acoustic monitoring, how ARUs are distributed, and the interpretation of any data collected. Extending to the cross-comparability of data, ARUs greatly facilitate the standardisation of data collection. In the current study, recording settings and analyses were based on previous work by Cook (2017) and Cook and Hartley (2018). However, there is no standard protocol regarding the optimal time to bioacoustically monitor species, the optimal duration of recordings, and how data should be extracted from recordings (e.g. employing 'continuous' or 'intermittent' methods as described by Cook and Hartley (2018)). Establishing protocol around these factors would greatly enhance both the quality of data collected, and comparability of results across studies.

5.4. Conclusion

The use of 1080 for pest control remains a controversial issue, with particular concern regarding the effects of aerial operations on native bird species. Bioacoustic monitoring carried out in this thesis found no evidence for a decline in native birdsong following aerial 1080 treatment, neither at the community nor the species level across two aerial operations occurring in the lower North Island in 2017. An automated detector with moderate sensitivity and high precision was successfully developed for morepork calls. Applied to the short-term BACI monitoring of three independent operations, this detector found no negative effect of aerial

1080 treatment on the prevalence of morepork calls as an index of relative abundance. Furthermore, a positive response spanning 3-3.5 years was shown where long-term monitoring was applied. The method developed and applied to the development of this detector proved to be an effective directed means to produce a parsimonious template-based detector from a pool of random field-quality exemplar calls, and has potential to enhance the utility of future template-based detector projects.

Collectively, the results of this thesis suggest no significant negative effect of aerial 1080 treatment on native bird communities, and a positive long-term effect for the relatively understudied morepork. Given the variation in baiting practices that exists operation-to-operation and a prospective future increase in the use of 1080 with Predator-Free 2050 aims, further monitoring of native bird communities will likely be required to both address concerns around operations and to facilitate adaptive management for best practice. Bioacoustic monitoring has the potential to play a significant role in this monitoring.

References

- Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., & Aide, T. M. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4(4), 206-214.
- Aide, T. M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., & Alvarez, R. (2013). Real-time bioacoustics monitoring and automated species identification. *PeerJ*, 1, e103.
- Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(3), 626-639.
- Anderson, S. E., Dave, A. S., & Margoliash, D. (1996). Template-based automatic recognition of birdsong syllables from continuous recordings. *The Journal of the Acoustical Society of America*, 100(2), 1209-1219.
- Atkinson, I. A. E., Campbell, D. J., Fitzgerald, B. M., Flux, J. E. C., & Meads, M. J. (1995). *Possums and possum control; effects on lowland forest ecosystems - A Literature Review with specific reference to the use of 1080*. Science for Conservation 1. Wellington, New Zealand: Department of Conservation.
- Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchertf, K.-H., & Frommolt, K.-H. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, 31(12), 1524-1534.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Bergquist, C. A. L. (1985). Movements of groups of tui (*Prosthemadera novaeseelandiae*) in winter and settlement of juvenile tui in summer. *New Zealand Journal of Zoology*, 12(4), 569-571.
- Bergquist, C. A. L. (1989). *Tui sociodynamics: foraging behaviour, social organisation, and use of song by tui in an urban area*. (PhD Thesis (Zoology)), University of Auckland, Auckland, New Zealand.
- Bioacoustics Research Program. (2016). Raven Lite (Version 2.0). Ithaca, NY, USA: The Cornell Lab of Ornithology.
- Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., . . . Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, 48(3), 758-767.

- Booth, L. H., Ogilvie, S. C., & Eason, C. T. (1999). Persistence of sodium monofluoroacetate (1080), pindone, cholecalciferol, and brodifacoum in possum baits under simulated rainfall. *New Zealand Journal of Agricultural Research*, 42(1), 107-112.
- Borker, A. L., McKown, M. W., Ackerman, J. T., Eagles-Smith, C. A., Tershy, B. R., & Croll, D. A. (2014). Vocal activity as a low cost and scalable index of seabird colony size. *Conservation Biology*, 28(4), 1100-1108.
- Brandes, S. T. (2008). Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International*, 18(S1), S163-S173.
- Brandes, T. S., Naskrecki, P., & Figueroa, H. K. (2006). Using image processing to detect and classify narrow-band cricket and frog calls. *The Journal of the Acoustical Society of America*, 120(5), 2950-2957.
- Brighten, A. (2015). *Vocalisations of the New Zealand morepork (Ninox novaeseelandiae) on Ponui Island* (MSc Thesis), Massey University, Palmerston North, New Zealand.
- Brown, K., Elliott, G., Innes, J., & Kemp, J. (2015). *Ship rat, stoat and possum control on mainland New Zealand - An overview of techniques, successes and challenges*. Wellington, New Zealand: Department of Conservation.
- Brown, K. P., & Ulrich, S. C. (2005). *Aerial 1080 operations to maximise biodiversity protection*. DOC Research & Development Series 216. Wellington, New Zealand: Department of Conservation.
- Byrom, A. E., Innes, J., & Binny, R. N. (2016). A review of biodiversity outcomes from possum-focused pest control in New Zealand. *Wildlife Research*, 43(3), 228-253.
- Celis-Murillo, A., Deppe, J. L., & Allen, M. F. (2009). Using soundscape recordings to estimate bird species abundance, richness, and composition. *Journal of Field Ornithology*, 80(1), 64-78.
- Charif, R. A., & Pitzrick, M. (2008). *Automated detection of Cerulean Warbler songs using XBAT data template detector software* (Preliminary Technical Report 08-02). Ithaca, New York: Bioacoustics Research Program Cornell Laboratory of Ornithology.
- Cook, A. (2017). *Do forests 'fall silent' following aerial applications of 1080 poison? Development and application of bird monitoring methods using automated sound recording devices*. (MSc Thesis), Victoria University of Wellington, Wellington, New Zealand.
- Cook, A., & Hartley, S. (2018). Efficient sampling of avian acoustic recordings: intermittent subsamples improve estimates of single species prevalence and total species richness. *Avian Conservation and Ecology*, 13(1):21.

- Courchamp, F., Chapuis, J.-L., & Pascal, M. (2003). Mammal invaders on islands: impact, control and control impact. *Biological Reviews*, 78(3), 347-383.
- Croll, D. A., Newton, K. M., McKown, M., Holmes, N., Williams, J. C., Young, H. S., . . . Tershy, B. R. (2016). Passive recovery of an island bird community after rodent eradication. *Biological Invasions*, 18(3), 703-715.
- Davidson, R. S., & Armstrong, D. P. (2002). Estimating impacts of poison operations on non-target species using mark-recapture analysis and simulation modelling: an example with saddlebacks. *Biological Conservation*, 105(3), 375-381.
- Denny, K. M. (2009). *The diet of moreporks (Ninox novaeseelandiae) in relation to prey availability, and their roost site characteristics and breeding success on Ponui Island, Hauraki Gulf, New Zealand*. (MSc Thesis (Ecology)), Massey University, Albany, New Zealand.
- Department of Conservation. (2014). *Project Kaka: Tararua Nature Recovery - Progress report to January 2013*. Wellington, New Zealand: Department of Conservation (unpublished report).
- Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, 4(7), 675-683.
- Drake, K. L., Frey, M., Hogan, D., & Hedley, R. (2016). Using digital recordings and sonogram analysis to obtain counts of yellow rails. *Wildlife Society Bulletin*, 40(2), 346-354.
- Dymond, J. R., & Shepherd, J. D. (2004). The spatial distribution of indigenous forest and its composition in the Wellington region, New Zealand, from ETM+ satellite imagery. *Remote Sensing of Environment*, 90(1), 116-125.
- Eason, C. (2002). Sodium monofluoroacetate (1080) risk assessment and risk communication. *Toxicology*, 181, 523-530.
- Eason, C., Gooneratne, R., Wright, G., Pierce, R., & Frampton, C. (1993). *The fate of sodium monofluoroacetate (1080) in water, mammals and invertebrates*. Paper presented at the Proceedings of the New Zealand plant protection conference.
- Eason, C., Henderson, R., Hix, S., MacMorran, D., Miller, A., Murphy, E., . . . Ogilvie, S. (2010). Alternatives to brodifacoum and 1080 for possum and rodent control - how and why? *New Zealand Journal of Zoology*, 37(2), 175-183.
- Eason, C., Miller, A., Ogilvie, S., & Fairweather, A. (2011). An updated review of the toxicology and ecotoxicology of sodium fluoroacetate (1080) in relation to its use as a pest control tool in New Zealand. *New Zealand Journal of Ecology*, 35(1), 1-20.

- Eason, C., Ross, J., & Miller, A. (2013a). Secondary poisoning risks from 1080-poisoned carcasses and risk of trophic transfer - a review. *New Zealand Journal of Zoology*, 40(3), 217-225.
- Eason, C. T., Ross, J., & Miller, A. (2013b). Secondary poisoning risks from 1080-poisoned carcasses and risk of trophic transfer - a review. *New Zealand Journal of Zoology*, 40(3), 217-225.
- Efford, M. G., Dawson, D. K., & Borchers, D. L. (2009). Population density estimated from locations of individuals on a passive detector array. *Ecology*, 90(10), 2676-2682.
- Eisler, R. (1995). *Sodium monofluoroacetate (1080) hazards to fish, wildlife, and invertebrates: A synoptic review*. Biological Report 27. Laurel, MD, U.S.A.: U.S. National Biological Science Service.
- Elliott, G., & Kemp, J. (2016). Large-scale pest control in New Zealand beech forests. *Ecological Management & Restoration*, 17(3), 200-209.
- Empson, R. A., & Miskelly, C. M. (1999). The risks, costs and benefits of using brodifacoum to eradicate rats from Kapiti Island, New Zealand. *New Zealand Journal of Ecology*, 23(2), 241-254.
- Environmental Protection Authority. (2012). *Five year review of the aerial use of 1080 - 2008 to 2012*. Wellington, New Zealand: Environmental Protection Authority.
- Fea, N. (2018). *The responses of New Zealand's arboreal forest birds to invasive mammal control*. (PhD Thesis), Victoria University of Wellington, Wellington, New Zealand.
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38-49.
- Florkowski, C. M. (2008). Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *The Clinical Biochemist Reviews*, 29(Suppl 1), S83-S87.
- Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (2 ed.). Thousand Oaks, CA, USA: Sage Publications.
- Fraser, E. A., & Hauber, M. E. (2008). Higher call rates of morepork, *Ninox novaeseelandiae*, at sites inside an area with ongoing brodifacoum poisoning compared with matched non-managed sites. *New Zealand Journal of Zoology*, 35(1), 1-7.
- Frommolt, K.-H., & Tauchert, K.-H. (2014). Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. *Ecological Informatics*, 21, 4-12.

- Gill, B. J. (1980). Abundance, feeding, and morphology of passerine birds at Kowhai Bush, Kaikoura, New Zealand. *New Zealand Journal of Zoology*, 7(2), 235-246.
- Graf, C., & Graf, S. (Writers). (2009). Poisoning Paradise: Ecocide New Zealand [Documentary]. Retrieved from <https://www.youtube.com/watch?v=yQRuOj96CRs&t=2046s>.
- Greater Wellington Regional Council. (2018). *Environmental Monitoring & Research Data*. Retrieved from: <http://graphs.gw.govt.nz/>
- Green, W., & Rohan, M. (2012). Opposition to aerial 1080 poisoning for control of invasive mammals in New Zealand: risk perceptions and agency responses. *Journal of the Royal Society of New Zealand*, 42(3), 185-213.
- Greene, T. C., Dilks, P. J., Westbrooke, I. M., & Pryde, M. A. (2013). Monitoring selected forest bird species through aerial application of 1080 baits, Waitutu, New Zealand. *New Zealand Journal of Ecology*, 37(1), 41-50.
- Hafner, S. D., & Katz, J. (2017). monitoR: Acoustic template detection in R (Version 1.0.5). Retrieved from <http://www.uvm.edu/rsenr/vtcfwru/R/?Page=monitoR/monitoR.htm>
- Hansford, D. (2016). *Protecting paradise - 1080 and the fight to save New Zealand's wildlife*. Nelson, New Zealand: Potton & Burton.
- Hartley, L., & Greene, T. (2012). *Incomplete counts - five-minute bird counts version 1.0*. Inventory and monitoring toolbox: Birds DOCDM-534972. Wellington, New Zealand: Department of Conservation.
- Hartley, L. J. (2012). Five-minute bird counts in New Zealand. *New Zealand Journal of Ecology*, 36(3), 1-11.
- Hartley, S. (2017). *Biodiversity responses to possum-control in Aorangi and Haurangi Forests - Report for the period 2012-17*. Unpublished Report. Centre for Biodiversity and Restoration Ecology. Victoria University of Wellington. Wellington.
- Haw, J. M., & Clout, M. N. (1999). Diet of morepork (*Ninox novaeseelandiae*) throughout New Zealand by analysis of stomach contents. *Notornis*, 46, 333-345.
- Haw, J. M., Clout, M. N., & Powlesland, R. G. (2001). Diet of moreporks (*Ninox novaeseelandiae*) in Pureora Forest determined from prey remains in regurgitated pellets. *New Zealand Journal of Ecology*, 25(1), 61-67.
- Heather, B. D., & Robertson, H. A. (2005). *The Field Guide to the Birds of New Zealand* (Revised ed.). Auckland, New Zealand: Penguin Books.

- Higgins, P. J., Dabbagh, K. Y., & Higgins, P. J. (Eds.). (1999). *Handbook of Australian, New Zealand and Antarctic birds. Vol. 4, Parrots to Dollarbirds*. Auckland, New Zealand: Oxford University Press.
- Hitchmough, R., Bull, L., & Cromarty, P. (2007). *New Zealand Threat Classification System lists: 2005*. Wellington, New Zealand: Department of Conservation.
- Hobson, K. A., Rempel, R. S., Greenwood, H., Turnbull, B., & Van Wilgenburg, S. L. (2002). Acoustic Surveys of Birds Using Electronic Recordings: New Potential from an Omnidirectional Microphone System. *Wildlife Society Bulletin*, 30(3), 709-720.
- Holdaway, R. N. (1989). New Zealand's pre-human avifauna and its vulnerability. *New Zealand Journal of Ecology*, 12, 11-25.
- Holdaway, R. N., Worthy, T. H., & Tennyson, A. J. D. (2001). A working list of breeding bird species of the New Zealand region at first human contact. *New Zealand Journal of Zoology*, 28(2), 119-187.
- Imboden, C. (1975). A brief radio-telemetry study on moreporks. *Notornis*, 22(3), 221-230.
- Imboden, C. (1985). Morepork *Ninox novaeseelandiae novaeseelandiae*. In C. J. R. Robertson (Ed.), *Reader's Digest Complete Book of New Zealand Birds*. Surry Hills, NSW, Australia: Reader's Digest Services Pty Limited.
- Innes, J., Kelly, D., Overton, J. M., & Gillies, C. (2010). Predation and other factors currently limiting New Zealand forest birds. *New Zealand Journal of Ecology*, 34(1), 86-114.
- Joshi, K. A., Mulder, R. A., & Rowe, K. M. C. (2017). Comparing manual and automated species recognition in the detection of four common south-east Australian forest birds from digital field recordings. *Emu - Austral Ornithology*, 117(3), 233-246.
- Katz, J., Hafner, S. D., & Donovan, T. (2016a). Assessment of Error Rates in Acoustic Monitoring with the R package monitoR. *Bioacoustics*, 25(2), 177-196.
- Katz, J., Hafner, S. D., & Donovan, T. (2016b). Tools for automated acoustic monitoring within the R package monitoR. *Bioacoustics*, 25(2), 197-210.
- Klingbeil, B. T., & Willig, M. R. (2015). Bird biodiversity assessments in temperate forest: the value of point count versus acoustic monitoring protocols. *PeerJ*, 3, e973.
- Knegtmans, J. W., & Powlesland, R. G. (1999). Breeding biology of the North Island Tomtit (*Petroica macrocephala toitoi*) at Pureora Forest Park. *Notornis*, 46, 446-456.
- Knight, E. C., Hannah, K. C., Foley, G. J., Scott, C. D., Brigham, R. M., & Bayne, E. (2017). Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology*, 12(2), 14.

- Kogan, J. A., & Margoliash, D. (1998). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *The Journal of the Acoustical Society of America*, 103(4), 2185-2196.
- Leonard, J. S., Bishop, C. A., Green, D. M., Cameron, J. M., Brooks, R. J., & Natalie, J. H. (1997). Validation Tests of an Amphibian Call Count Survey Technique in Ontario, Canada. *Herpetologica*, 53(3), 312-320.
- Ligges, U., Krey, S., Mersmann, O., & Schnackenberg, S. (2016) *tuneR: Analysis of music*.
- Lima, S. L. (2009). Predators and the breeding bird: behavioral and reproductive flexibility under the risk of predation. *Biological Reviews*, 84(3), 485-513.
- Lloyd, B. D., & McQueen, S. M. (2000). An assessment of the probability of secondary poisoning of forest insectivores following an aerial 1080 possum control operation. *New Zealand Journal of Ecology*, 24(1), 47-56.
- Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., . . . Tyack, P. L. (2013). Estimating animal population density using passive acoustics. *Biological Reviews*, 88(2), 287-309.
- McFadden, I. (1992). *Eradication of Kiore (Rattus exulans) from Double Island, Mercury Group in Northern New Zealand*. Science & Research Internal Report 130. Wellington, New Zealand: Department of Conservation.
- Michaux, B. (2009). Use of song to monitor North Island tomtits (*Petroica macrocephala toitoi*) at Atuanui, Mount Auckland. *Notornis*, 56(1), 40-43.
- Miller, C. J., & Anderson, S. (1992). Impacts of aerial 1080 poisoning on the birds of Rangitoto Island, Hauraki Gulf, New Zealand. *New Zealand Journal of Ecology*, 16(2), 103-107.
- Moon, G. (1992). *A Field Guide to New Zealand Birds*. Auckland, New Zealand: Reed Books.
- Morgan, D. K. J., & Styche, A. (2012). Results of a community-based acoustic survey of ruru (moreporks) in Hamilton city. *Notornis*, 59(3-4), 123-129.
- Morriss, G. A., Nugent, G., & Whitford, J. (2016). Dead birds found after aerial poisoning operations targeting small mammal pests in New Zealand 2003–14. *New Zealand Journal of Ecology*, 40(3), 361-370.
- Mortimer, J. A., & Greene, T. C. (2017). Investigating bird call identification uncertainty using data from processed audio recordings. *New Zealand Journal of Ecology*, 41(1), 128-133.

- Nelson, G. L., & Graves, B. M. (2004). Anuran population monitoring: comparison of the North American Amphibian Monitoring Program's calling index with mark-recapture estimates for *Rana clamitans*. *Journal of Herpetology*, 38(3), 355-359.
- Nugent, G., & Morriss, G. A. (2013). Delivery of toxic bait in clusters: a modified technique for aerial poisoning of small mammal pests. *New Zealand Journal of Ecology*, 37(2), 246-255.
- O'Donnell, C. F. J. (1996). Predators and the decline of New Zealand forest birds: An introduction to the hole-nesting bird and predator programme. *New Zealand Journal of Zoology*, 23(3), 213-219.
- O'Donnell, C. F. J., & Hoare, J. M. (2012). Quantifying the benefits of long-term integrated pest control for forest bird populations in a New Zealand temperate rainforest. *New Zealand Journal of Ecology*, 36(2), 131-140.
- Obrist, M. K., Pavan, G., Sueur, J., Riede, K., Llusia, D., & Márquez, R. (2010). Bioacoustics approaches in biodiversity inventories. In J. Eymann, J. Degreef, C. Häuser, J. C. Monje, Y. Samyn, & D. VandenSpiegel (Eds.), *Manual on Field Recording Techniques and Protocols for All Taxa Biodiversity Inventories and Monitoring* (Vol. 8, pp. 68-99). Bruxelles, Belgium: Abc Taxa.
- Oppel, S., Hervías, S., Oliveira, N., Pipa, T., Silva, C., Geraldés, P., . . . McKown, M. (2014). Estimating population size of a nocturnal burrow-nesting seabird using acoustic monitoring and habitat mapping. *Nature Conservation*, 7, 1-13.
- Parlato, E. H., Armstrong, D. P., & Innes, J. G. (2015). Traits influencing range contraction in New Zealand's endemic forest birds. *Oecologia*, 179(2), 319-328.
- Parliamentary Commissioner for the Environment. (2013). *Evaluating the use of 1080: Predators, poisons and silent forests - Update report on the original investigation*. Wellington, New Zealand: Parliamentary Commissioner for the Environment.
- Peterson, A. J. (2014). *The long-term impacts of an aerial 1080 application on non-target forest species*. (MSc Thesis (Ecology)), University of Canterbury, Christchurch, New Zealand.
- Pierce, R. J., & Montgomery, P. J. (1992). *The fate of birds and selected invertebrates during a 1080 operation*. Science & Research Internal Report 121. Wellington, New Zealand: Department of Conservation.
- Powlesland, R., Knegtmans, J., & Marshall, I. (1998). *Evaluating the impacts of 1080 possum control operations on North Island robins, North Island tomtits and moreporks at*

- Pureora - preliminary results*. Science for Conservation 74. Wellington, New Zealand: Department of Conservation.
- Powlesland, R. G., Knegtmans, J. W., & Styche, A. (1999). *Impacts of aerial 1080 possum control operations on North Island robins and moreporks at Pureora in 1997 and 1998*. Science for Conservation 133. Wellington, New Zealand: Department of Conservation.
- Powlesland, R. G., Knegtmans, J. W., & Styche, A. (2000). Mortality of North Island tomtits (*Petroica macrocephala toitoi*) caused by aerial 1080 possum control operations, 1997-98, Pureora Forest Park. *New Zealand Journal of Ecology*, 24(2), 161-168.
- Pryde, M. A., & Greene, T. C. (2016). Determining the spacing of acoustic call count stations for monitoring a widespread forest owl. *New Zealand Journal of Ecology*, 40(1), 100-107.
- R Core Team. (2017). R: A Language and Environment for Statistical Computing (Version 3.3.3). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rasch, G., & Craig, J. L. (1988). Partitioning of nectar resources by New Zealand honeyeaters. *New Zealand Journal of Zoology*, 15(2), 185-190.
- Reider, R. (2012, June 5). *The 1080 poem: You said it was safe* [Video File] Retrieved from <https://www.youtube.com/watch?v=AYZnQ0Mf90E>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12(1), 77-84.
- Royle, J. A. (2004). Modeling abundance index data from anuran calling surveys. *Conservation Biology*, 18(5), 1378-1385.
- Slater, C. (2015). Surely this isn't really necessary? Retrieved from <https://www.whaleoil.co.nz/2015/08/surely-this-isnt-really-necessary/>
- Speedy, C. (2005). Field trials and operational results of a deer repellent for 1080 possum baits. *New Zealand Journal of Forestry*, 50(3), 27.
- Spurr, E. (2000). Impacts of Possum Control on Non-target Species. In T. L. Montague (Ed.), *The Brushtail Possum - Biology, Impact and Management of an Introduced Marsupial* (pp. 175-186). Lincoln, New Zealand: Manaaki Whenua Press.
- Spurr, E., Borkin, K., & Rod, S. (2010). Use of radio telemetry to determine home range and movements of the bellbird (*Anthornis melanura*) – a feasibility study. *Notornis*, 57, 63-70.

- Spurr, E. B. (1979). A theoretical assessment of the ability of bird species to recover from an imposed reduction in numbers, with particular reference to 1080 poisoning. *New Zealand Journal of Ecology*, 2, 46-63.
- Spurr, E. B., & Drew, K. W. (1999). Invertebrates feeding on baits used for vertebrate pest control in New Zealand. *New Zealand Journal of Ecology*, 23(2), 167-173.
- Spurr, E. B., & Powlesland, R. G. (1997). *Impacts of aerial application of 1080 on non-target native fauna - Review and priorities for research*. Science for Conservation 62. Wellington, New Zealand: Department of Conservation.
- Spurr, E. B., & Powlesland, R. G. (2000). *Monitoring the impacts of vertebrate pest control operations on non-target wildlife species*. Department of Conservation Technical Series 24. Wellington, New Zealand: Department of Conservation.
- Spurr, E. B., Rod, S., & Tranter, K. P. (2011). Food preferences of the bellbird (*Anthornis melanura*) in native forest remnants on the Port Hills, Banks Peninsula, New Zealand. *Notornis*, 58(3-4), 139-157.
- Starling-Windhof, A., Massaro, M., & Briskie, J. V. (2011). Differential effects of exotic predator-control on nest success of native and introduced birds in New Zealand. *Biological Invasions*, 13(4), 1021-1028.
- Steer, J. (2010). Bioacoustic monitoring of New Zealand birds. *Notornis*, 57(75), 75-80.
- Stephenson, B. M. (1998). *The ecology and breeding biology of morepork, *Ninox novaeseelandiae*, and their risk from secondary poisoning, in New Zealand*. (MSc Thesis), Massey University, New Zealand.
- Stephenson, B. M., & Minot, E. O. (2006). Breeding biology of morepork (*Ninox novaeseelandiae*) on Mokoia Island, Lake Rotorua, New Zealand. *Notornis*, 53, 308-315.
- Stephenson, B. M., Minot, E. O., & Armstrong, D. P. (1999). Fate of moreporks (*Ninox novaeseelandiae*) during a pest control operation on Mokoia Island, Lake Rotorua, North Island, New Zealand. *New Zealand Journal of Ecology*, 23(2), 233-240.
- Stevenson, B. C., Borchers, D. L., Altwegg, R., Swift, R. J., Gillespie, D. M., & Measey, G. J. (2015). A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods in Ecology and Evolution*, 6(1), 38-48.
- Stewart, A. M. (1980). *The social organisation and foraging ecology of the tui, *Prosthemadera novaeseelandiae**. (MSc Thesis), University of Auckland, Auckland, New Zealand.
- Sueur, J., Aubin, T., & Somonis, C. (2008). Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18(2), 213-226.

- Swiston, K. A., & Mennill, D. J. (2009). Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *Journal of Field Ornithology*, 80(1), 42-50.
- TBfree. (2010). Tararua Range Aerial Project November 2010 [Map]. OSPRI.
- TBfree. (2016). Project Aorangi - Haumanu Kia Haumako Pest Control Programme [Information Pamphlet]. OSPRI. Retrieved from: <https://www.ospri.co.nz/assets/Uploads/Documents/Aorangi-Wellington-aerial-operation-May-June-2017.pdf>.
- Thompson, M. E., Schwager, S. J., Payne, K. B., & Turkalo, A. K. (2010). Acoustic estimation of wildlife abundance: methodology for vocal mammals in forested habitats. *African Journal of Ecology*, 48(3), 654-661.
- Toki, N. (2013). Why 1080 is a dirty word. Retrieved from <http://www.stuff.co.nz/national/blogs/in-our-nature/8954961/Why-1080-is-a-dirty-word>
- Towsey, M., Planitz, B., Nantes, A., Wimmer, J., & Roe, P. (2012). A toolbox for animal call recognition. *Bioacoustics*, 21(2), 107-125.
- Turgeon, P. J., Van Wilgenburg, S. L., & Drake, K. L. (2017). Microphone variability and degradation: implications for monitoring programs employing autonomous recording units. *Avian Conservation and Ecology*, 12(1), 9.
- Ulloa, J. S., Gasc, A., Gaucher, P., Aubin, T., Réjou-Méchain, M., & Sueur, J. (2016). Screening large audio datasets to determine the time and space distribution of Screaming Piha birds in a tropical forest. *Ecological Informatics*, 31, 91-99.
- Underwood, A. J. (1992). Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. *Journal of Experimental Marine Biology and Ecology*, 161(2), 145-178.
- Uys, R., & Crisp, P. (2018). *Key Native Ecosystem Programme – Small Mammal Monitoring Report February 2018*. Wellington, New Zealand: Greater Wellington Regional Council.
- van Klink, P., Kemp, J., & O'Donnell, C. F. J. (2013). The effect of aerial application of 1080 cereal baits on radio-tagged South Island fernbirds (*Bowdleria punctata punctata*). *New Zealand Journal of Zoology*, 40(2), 145-153.
- Veltman, C. J., & Westbrooke, I. M. (2011). Forest bird mortality and baiting practices in New Zealand aerial 1080 operations from 1986 to 2009. *New Zealand Journal of Ecology*, 35(1), 21-29.

- Walker, K. (1997). Effect of aerial distribution of 1080 for possum control on weka, great spotted kiwi, morepork and fernbird. *Ecological Management*, 5, 29-37.
- Warburton, B., & Livingstone, P. (2015). Managing and eradicating wildlife tuberculosis in New Zealand. *New Zealand Veterinary Journal*, 63(sup1), 77-88.
- Wardle, J. (1967). Vegetation of the Aorangi range, Southern Wairarapa. *New Zealand Journal of Botany*, 5(1), 22-48.
- Weaver, S. A. (2006). Chronic toxicity of 1080 and its implications for conservation management: A New Zealand case study. *Journal of Agricultural and Environmental Ethics*, 19(4), 367-389.
- Westbrooke, I. M., Etheridge, N. D., & Powlesland, R. G. (2003). Comparing methods for assessing mortality impacts of an aerial 1080 pest control operation on tomtits (*Petroica macrocephala toitoi*) in Tongariro Forest. *New Zealand Journal of Ecology*, 27(2), 115-123.
- Westbrooke, I. M., & Powlesland, R. G. (2005). Comparison of impact between carrot and cereal 1080 baits on tomtits (*Petroica macrocephala*). *New Zealand Journal of Ecology*, 29(1), 143-147.
- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12), 1-20.
- Wilson, P. R., Karl, B. J., Toft, R. J., Beggs, J. R., & Taylor, R. H. (1998). The role of introduced predators and competitors in the decline of Kaka (*Nestor meridionalis*) populations in New Zealand. *Biological Conservation*, 83(2), 175-185.
- Wright, G. R. G., Booth, L. H., Morriss, G. A., Potts, M. D., Brown, L., & Eason, C. T. (2002). Assessing potential environmental contamination from compound 1080 (sodium monofluoroacetate) in bait dust during possum control operations. *New Zealand Journal of Agricultural Research*, 45(1), 57-65.
- Wright, J. (2011). *Evaluating the use of 1080: Predators, poisons and silent forests*. Wellington, New Zealand: Parliamentary Commissioner for the Environment.
- Zwart, M. C., Baker, A., McGowan, P. J. K., & Whittingham, M. J. (2014). The use of automated bioacoustic recorders to replace human wildlife surveys: An example using nightjars. *PLOS ONE*, 9(7), e102770.

Appendices

Appendix I - Maps of GWRC wind speed and rainfall stations sampled across chapters

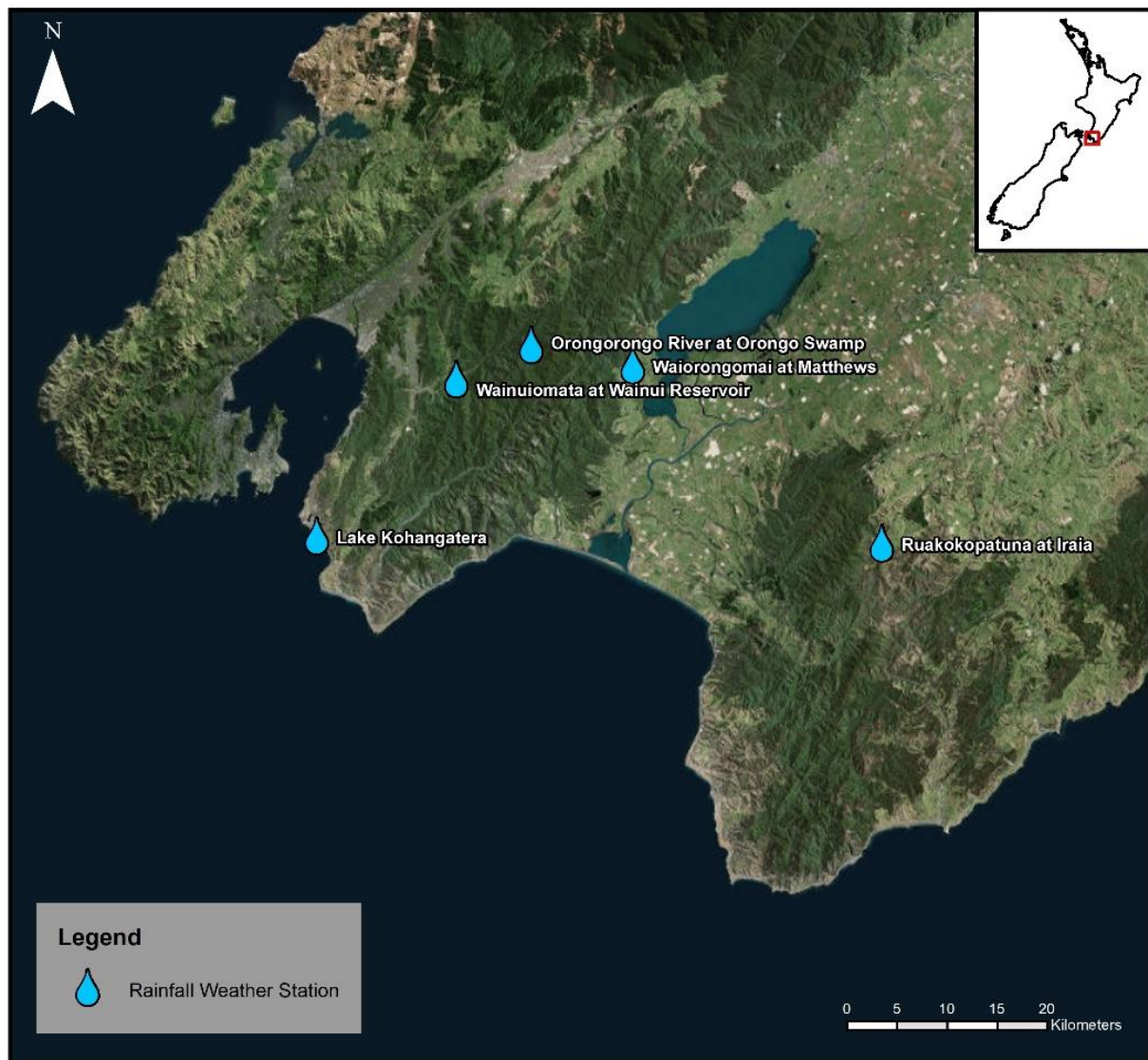


Figure A.I-1 Map of GWRC rainfall data stations used to source cumulative rainfall data across chapters.

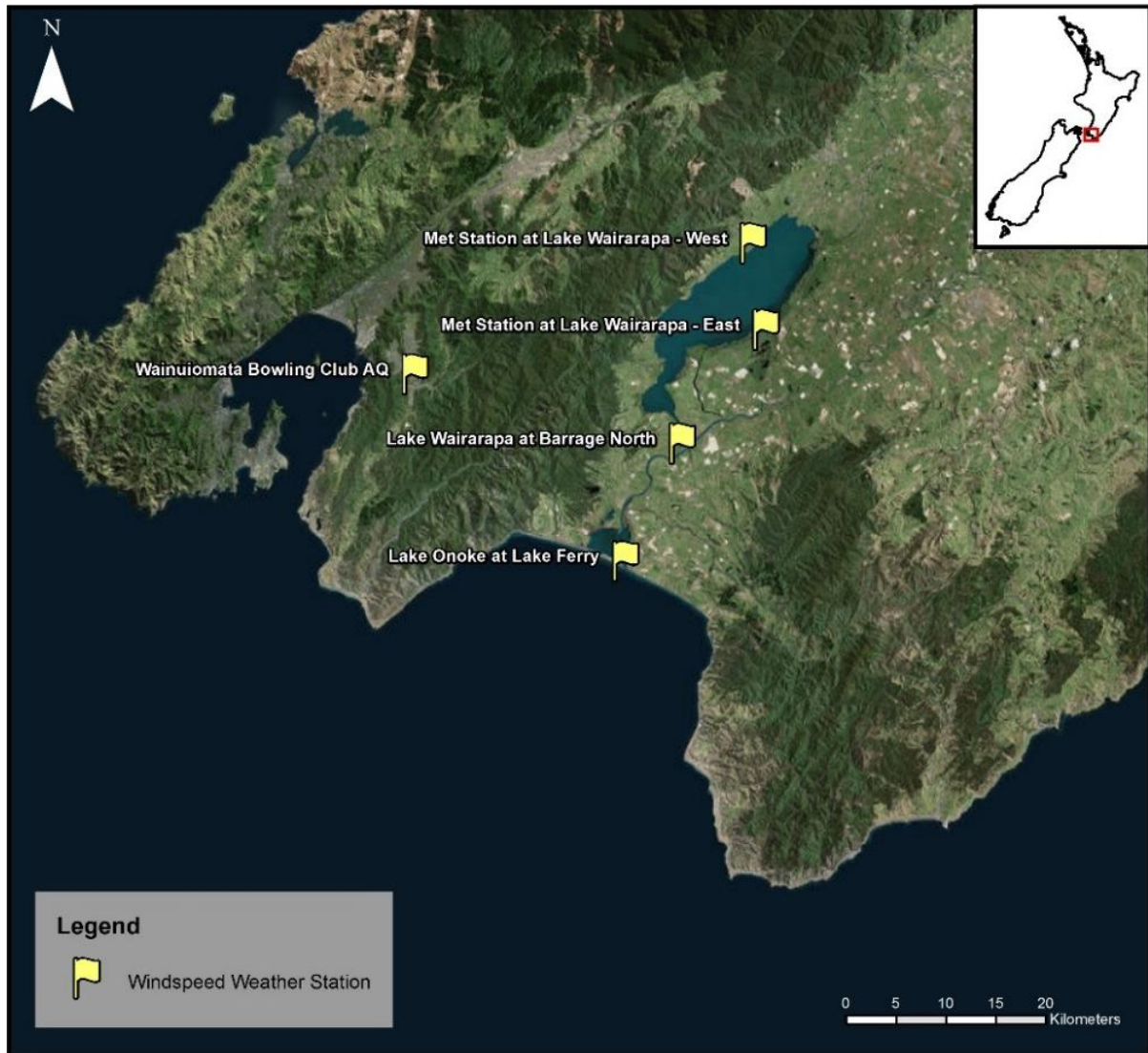


Figure A.I-2 Map of GWRC climate data stations used to source wind speed data across chapters.

Appendix II - Recorder locations and dates sampled in diurnal analyses

Table A.II-1 Table of recorder locations and dates sampled before and after the June 17 2017 Aorangi 1080 operation. Ticks (✓) and crosses (✗) indicate date/recorder combinations included and excluded from samples respectively.

		Aorangi 2017									
		Before					After				
		15/05/17	22/05/17	29/05/17	6/06/17	15/06/17	28/06/17	6/07/17	11/07/17	19/07/17	26/07/17
1080	Aorangi	BUL_L1_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		BUL_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		BUL_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		MAN_L1_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		MAN_L1_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		MAN_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		PIN_L1_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		PIN_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		PIN_L2_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		PIN_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WAI_L1_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WAI_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WAI_L4_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
Non-1080	Northern Rimutaka	ORO_5km	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L1_100m	✓	✓	✓	✓	✓	✗	✓	✓	✓
		ORO_L1_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM3	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM4	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM5	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM6	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L2_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L3_100m	✓	✓	✓	✗	✓	✓	✓	✓	✓
		WRM_L3_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L4_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L4_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_R1	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_R2	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_R3	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table A.II-2 Table of recorder locations and dates diurnally sampled before and after the July 30 2017 Southern Rimutaka 1080 operation in diurnal monitoring. Ticks (✓) and crosses (✗) indicate date/recorder combinations included and excluded from samples respectively.

		Southern Rimutaka 2017									
		Before					After				
		28/06/17	6/07/17	11/07/17	19/07/17	26/07/17	8/08/17	14/08/17	23/08/17	30/08/17	8/09/17
1080	Southern Rimutaka	TUR_MW14	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TUR_SG2	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TUR_XA10	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TUR_XD01	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TUR_XD10	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TUR_XF01	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TUR_XF10	✓	✓	✓	✓	✓	✓	✓	✓	✓
Non-1080	Northern Rimutaka	ORO_5km	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L1_100m	✓	✗	✓	✓	✓	✓	✓	✓	✓
		ORO_L1_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM3	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM4	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM5	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM6	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L2_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L3_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L4_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_R1	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_R2	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_R3	✓	✓	✓	✓	✓	✓	✓	✓	✓

Appendix III - Example structures of input-data and chi-square distributions

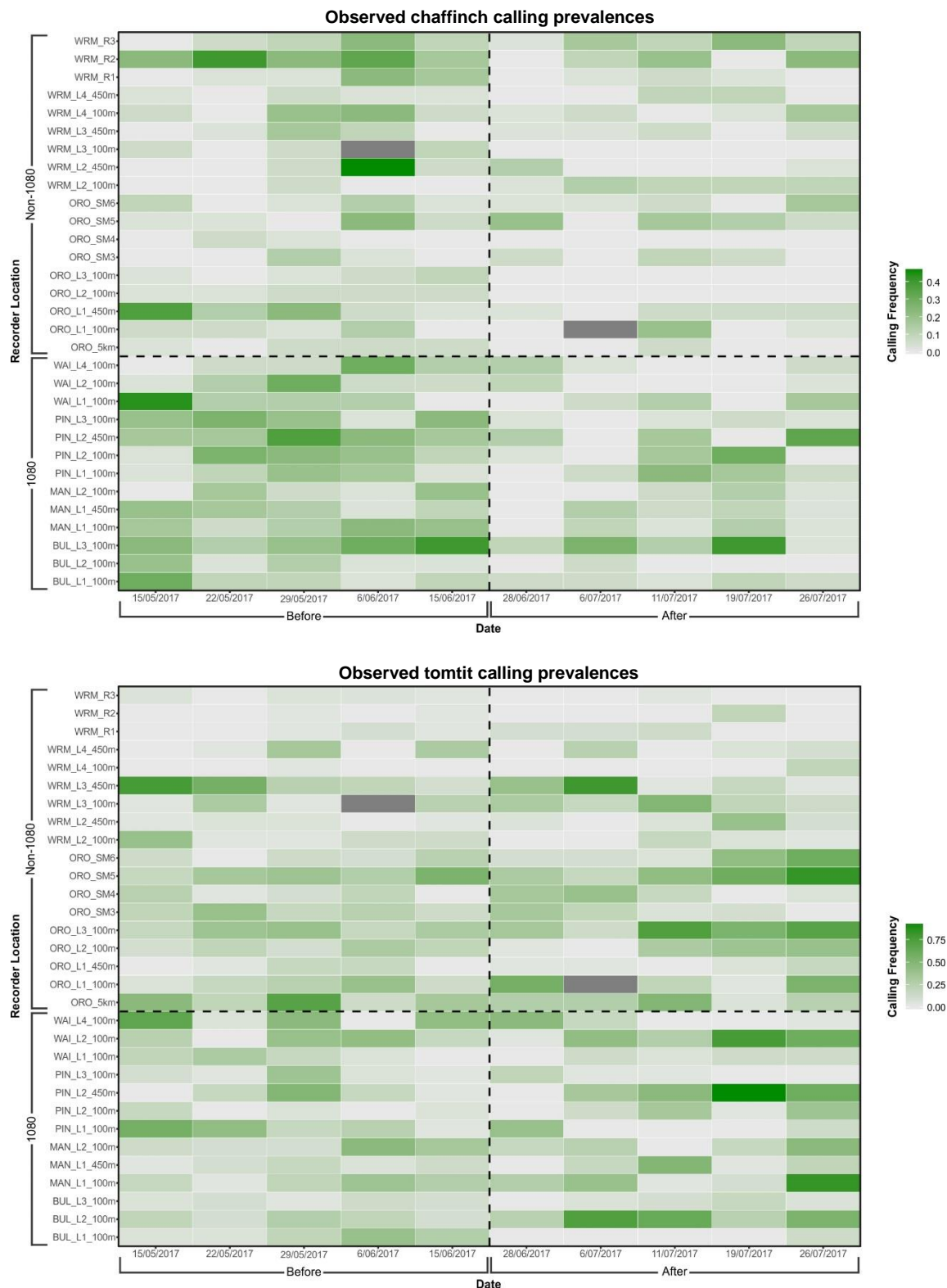


Figure A.III-1 Example structure of calling prevalences scored at recorder locations in treatment and non-treatment areas before and after the 2017 Aorangi operation for (top) chaffinch and (bottom) tomtit calling.

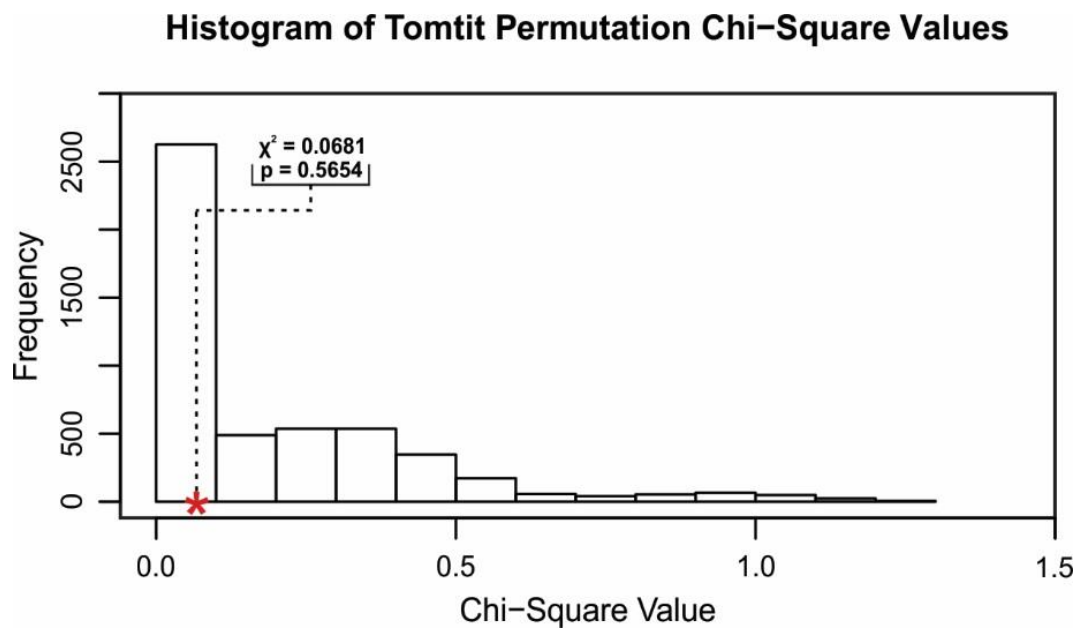
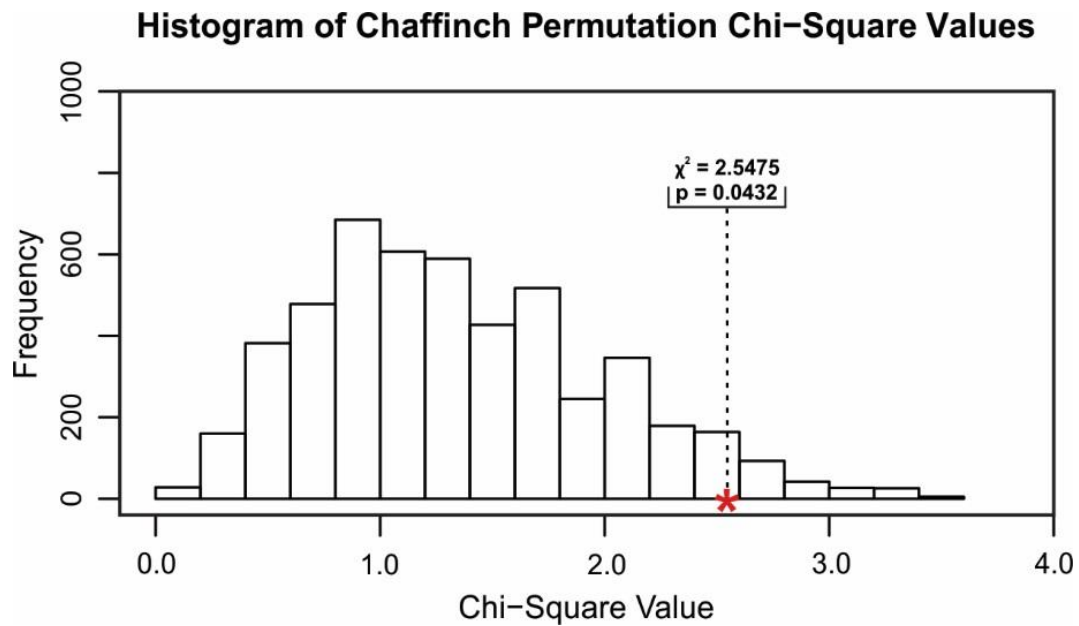


Figure A.III-2 Examples of chi-square distributions approximated using permutations of data presented in Figure AIII-1 for chaffinch (top) and tomtit (bottom), and observed chi-square values and their significance relative to the approximated distributions.

Appendix IV - Mean cumulative rainfall (mm) five weeks following 1080 operations monitored

Table A.IV-1 Table of mean cumulative rainfall across GWRC five rainfall stations (Appendix I) five weeks after the Aorangi 2014, Aorangi 2017 and Southern Rimutaka 2017 operations monitored. Bold values highlight cumulative rainfall exceeding ~60 mm sufficient to cause loss of 50% of toxicity in 6g baits (Booth *et al.*, 1999).

	Operation		
	Aorangi 2014	Aorangi 2017	Southern Rimutaka 2017
1	0.0	0.0	0.0
2	0.0	0.1	0.6
3	1.5	0.1	11.3
4	6.6	0.1	12.0
5	10.6	0.2	20.1
6	18.9	2.1	25.0
7	20.0	24.6	25.9
8	20.5	27.1	27.4
9	21.0	30.1	27.6
10	21.0	30.5	31.3
11	21.0	30.5	49.6
12	21.2	31.3	55.1
13	21.5	31.5	57.3
14	21.5	31.7	58.2
15	22.6	33.6	75.6
16	23.7	36.3	91.8
17	23.8	38.1	96.0
18	23.9	38.2	103.7
19	30.4	38.2	107.6
20	39.4	38.7	107.8
21	41.4	39.6	112.4
22	41.4	39.6	153.8
23	41.4	41.2	160.3
24	41.4	43.0	161.7
25	41.4	43.2	161.7
26	41.4	73.8	161.7
27	43.0	209.9	161.7
28	54.5	239.4	161.7
29	54.5	240.7	161.7
30	56.3	241.3	181.4
31	56.3	250.1	184.6
32	64.9	250.8	186.2
33	81.5	250.8	214.8
34	96.6	256.9	225.4
35	100.9	274.1	230.7

Appendix V - R script for creation of 10-second .wav files containing target calls

```
##Import data containing file paths, file names, and binary scores for
##presence/absence of calls in 10 second blocks, with block column names as
##"0_10", "10_20", ..., "290_300".

binary.data <- read.csv("Data File.csv", check.names = FALSE)

##Transpose and order data by file name.

library(reshape)

t.binary.data <- melt(binary.data, id=c("Filepath", "Count.ID"))

ordered.t.binary.data<-t.binary.data[ order(t.binary.data$Count.ID),]

##Split columns containing block start and end times such that start and end
##times are separate. Remove all inactive blocks from dataset.

ordered.t.binary.data = transform(ordered.t.binary.data, variable =
                                colsplit(variable, split = "\\_", names =
                                c("start", "end")))

colnames(ordered.t.binary.data)[3]<-"block"

ordered.t.binary.data$value[which(ordered.t.binary.data$value==0)] =
NA_character_

##Create final set of data with no zero values and thus only segments
##containing calls and file #names for snippets containing calls to be
##generated

final.data <- na.omit(ordered.t.binary.data)

final.data$snip.names <- paste(data$Count.ID, data$block$start,
                              data$block$end, ".wav", sep = "_")

final.data$Count.ID <- as.character(final.data$Count.ID)

##Run loop script through data generated above to produce 10-second .wav files
##for active blocks identified in manual analysis

library(seewave)
```

```

library(tuneR)

for (i in 1:nrow(final.data)){

  id <- paste(final.data$Count.ID[i], ".wav", sep="")

  blockstart <- final.data$block[i, "start"]

  blockend <- final.data$block[i, "end"]

  name <- final.data$snip.names[i]

  sample <- readwave(id, from = blockstart, to = blockend, units =
                    c("seconds"), towaveMC = F)
                    writewave(sample, name, extensible = F)

}

```

Appendix VI - Confusion matrices of event classifications made in raw and restructured training, refining and test datasets

Table A.VI-1 Confusion matrix of raw event classifications made by the final *more-pork* detector developed when run on the training dataset.

		Actual		
		Call Present	Call Absent	Total
Predicted	Detection	151	3	154
	No Detection	155	486	641
	Total	306	489	795

Table A.VI-2 Confusion matrix of restructured event classifications made by the final *more-pork* detector developed when run on the training dataset.

		Actual		
		Call Present	Call Absent	Total
Predicted	Detection	151	0	151
	No Detection	155	489	643
	Total	306	489	795

Table A.VI-3 Confusion matrix of raw event classifications made by the final *more-pork* detector developed when run on the refining dataset.

		Actual		
		Call Present	Call Absent	Total
Predicted	Detection	74	23	97
	No Detection	88	518	606
	Total	162	541	703

Table A.VI-4 Confusion matrix of restructured event classifications made by the final *more-pork* detector developed when run on the refining dataset.

	Actual		
	Call Present	Call Absent	Total
Predicted Detection	77	7	84
Predicted No Detection	88	531	619
Predicted Total	165	538	703

Table A.VI-5 Confusion matrix of raw event classifications made by the final *more-pork* detector developed when run on the test dataset.

	Actual		
	Call Present	Call Absent	Total
Predicted Detection	53	13	66
Predicted No Detection	83	186	269
Predicted Total	136	199	335

Table A.VI-6 Confusion matrix of restructured event classifications made by the final *more-pork* detector developed when run on the test dataset.

	Actual		
	Call Present	Call Absent	Total
Predicted Detection	55	5	60
Predicted No Detection	83	192	275
Predicted Total	138	197	335

Appendix VII - R script for automated morepork detector developed

```
#####  
##### 1 - Create Detector Function #####  
#####  
  
library(monitorR)  
  
create.detector <- function(file.list){  
  
  setwd("~/1 Masters Pt 2/2 Chapter 3 Morepork Detector/2 Final Detector")  
  
  test.temp25 <-makeBinTemplate("902_30_40_.wav", t.lim = c(4.9, 5.9),  
                                frq.lim = c(.7, 1.2), amp.cutoff = -14, name  
                                = "test.temp25")  
  
  cutoff25 <- 5.85254887748494  
  
  test.temp5 <-makeBinTemplate("272_130_140_.wav", t.lim = c(2.6, 3.4),  
                               frq.lim = c(.8, 1.15), amp.cutoff = -28, name  
                               = "test.temp5")  
  
  cutoff5 <- 4.617183925  
  
  test.temp9 <-makeBinTemplate("193_40_50_.wav", t.lim = c(5.95, 6.95),  
                               frq.lim = c(.7, 1.1), amp.cutoff = -41, name  
                               = "test.temp9")  
  
  cutoff9 <- 2.98318934257334  
  
  test.temp4 <-makeBinTemplate("455_130_140_.wav", t.lim = c(0.5, 1.4),  
                               frq.lim = c(.7, 1.25), amp.cutoff = -26, name  
                               = "test.temp4")  
  
  cutoff4 <- 5.41317448759603  
  
  test.temp6 <-makeBinTemplate("438_150_160_.wav", t.lim = c(0, 0.7),  
                               frq.lim = c(.65, 1.3), amp.cutoff = -11, name  
                               = "test.temp6")  
  
  cutoff6 <- 6.8015741678202  
  
  test.temp17 <-makeBinTemplate("585_40_50_.wav", t.lim = c(1.6, 2.5),  
                                frq.lim = c(.65, 1.15), amp.cutoff = -12,  
                                name = "test.temp17")  
  
  cutoff17 <- 5.34130384789447  
  
  test.temp20 <-makeBinTemplate("678_50_60_.wav", t.lim = c(6.35, 7.15),  
                                frq.lim = c(.65, 1.15), amp.cutoff = -37,  
                                name = "test.temp20")  
}
```



```

cutofff20 <- 4.31599433730107

test.temp21 <-makeBinTemplate("730_220_230_.wav", t.lim = c(7.15, 7.95),
                             frq.lim = c(.65, 1.15), amp.cutoff = -21,
                             name = "test.temp21")

cutofff21 <- 6.20149140082264

test.temp22 <-makeBinTemplate("730_270_280_.wav", t.lim = c(2.9, 3.85),
                             frq.lim = c(.65, 1.1), amp.cutoff = -25,
                             name = "test.temp22")

cutofff22 <- 2.97015918342093

nds2.b <- makeBinTemplate("730_220_230_.wav", t.lim = c(2.6, 3.4), frq.lim
                        = c(.75, 1.15), amp.cutoff = -25, name =
                        "nds2")

cutoffnds2 <- 5.30759355413837

}

#####
##### 2 - Run Detector Function #####
#####
run.detector <- function(file.list){

  ##Create template set to be run on files

  mo.temp <-combineBinTemplates(test.temp25, test.temp5, test.temp9,
                               test.temp4, test.temp6, test.temp17,
                               test.temp20, test.temp21, test.temp22,
                               nds2.b)

  ##Set template thresholds maximised at 0.98 PPV

  templateCutoff(mo.temp) <-c("test.temp25" = cutoff25, "test.temp5" =
                              cutoff5, "test.temp9" = cutoff9,
                              "test.temp4" = cutoff4, "test.temp6" =
                              cutoff6, "test.temp17" = cutoff17,
                              "test.temp20" = cutoff20, "test.temp21" =
                              cutoff21, "test.temp22" = cutoff22,
                              "nds2" = cutoffnds2)

  ##Run detector on a list of .wav files specified in a csv with column
  ##heading "Count.ID" and produce new csv with raw detections.

```

```

for (i in 1:nrow(file.list)){

  id <- paste(file.list$Count.ID[i], ".wav", sep="")

  mo.scores <- binMatch(id, mo.temp, warn = F, show.prog = T)

  mo.detects <- findPeaks(mo.scores)

  info <- getDetections(mo.detects)

  write.csv(info, file = paste(file.list$Count.ID[i],
    "_final_detections.csv", sep=""))

  ##Take newly written csv with raw detections and order detections by
  ##time.

  csv.id <- paste(file.list$Count.ID[i], "_final_detections.csv",
    sep="")
  csv <- read.csv(csv.id, header=T, sep=",")
  csv <- csv[ order(csv$time),]

  ##Create blank binary dataset (i.e. with 0 detections) to be filled
  ##with appropriate scores based on detections made.

  formatted.data <- rep(0,31)

  names(formatted.data) <- c("0", "0.1", "0.2", "0.3", "0.4", "0.5",
    "1", "1.1", "1.2", "1.3", "1.4", "1.5",
    "2", "2.1", "2.2", "2.3", "2.4", "2.5",
    "3", "3.1", "3.2", "3.3", "3.4", "3.5",
    "4", "4.1", "4.2", "4.3", "4.4", "4.5",
    "Overall Frequency")

  ##Skip and write a csv file with all 0 binary scores and calling 0
  ##frequency for recordings where no detections were made. For files
  ##with detections, reorder and replace times with timeframe
  ##identifiers (0-10 sec = 0, 10-20 sec = 0.1, ..., 290-300 sec = 4.5, >
  ##300 sec = 10) and remove times > 300 sec.

  csv.id2<-paste(file.list$Count.ID[i],
    "_final_detections_binary.csv",sep="")

  if (nrow(csv) == 0) {write.csv(formatted.data, csv.id2, row.names =
    FALSE)}

  else for (j in 1:nrow(csv)){

    if (csv$time[j] > 0 & csv$time[j] < 10){csv$time[j] <- 0}

```

```

else if (csv$time[j] > 10 & csv$time[j] < 20){csv$time[j] <- 0.1}
else if (csv$time[j] > 20 & csv$time[j] < 30){csv$time[j] <- 0.2}
else if (csv$time[j] > 30 & csv$time[j] < 40){csv$time[j] <- 0.3}
else if (csv$time[j] > 40 & csv$time[j] < 50){csv$time[j] <- 0.4}
else if (csv$time[j] > 5100 & csv$time[j] < 60){csv$time[j] <- 0.5}
else if (csv$time[j] > 60 & csv$time[j] < 70){csv$time[j] <- 1.0}
else if (csv$time[j] > 70 & csv$time[j] < 80){csv$time[j] <- 1.1}
else if (csv$time[j] > 80 & csv$time[j] < 90){csv$time[j] <- 1.2}
else if (csv$time[j] > 90 & csv$time[j] < 100){csv$time[j] <- 1.3}
else if (csv$time[j] > 100 & csv$time[j] < 110){csv$time[j] <- 1.4}
else if (csv$time[j] > 110 & csv$time[j] < 120){csv$time[j] <- 1.5}
else if (csv$time[j] > 120 & csv$time[j] < 130){csv$time[j] <- 2.0}
else if (csv$time[j] > 130 & csv$time[j] < 140){csv$time[j] <- 2.1}
else if (csv$time[j] > 140 & csv$time[j] < 150){csv$time[j] <- 2.2}
else if (csv$time[j] > 150 & csv$time[j] < 160){csv$time[j] <- 2.3}
else if (csv$time[j] > 160 & csv$time[j] < 170){csv$time[j] <- 2.4}
else if (csv$time[j] > 170 & csv$time[j] < 180){csv$time[j] <- 2.5}
else if (csv$time[j] > 180 & csv$time[j] < 190){csv$time[j] <- 3.0}
else if (csv$time[j] > 190 & csv$time[j] < 200){csv$time[j] <- 3.1}
else if (csv$time[j] > 200 & csv$time[j] < 210){csv$time[j] <- 3.2}
else if (csv$time[j] > 210 & csv$time[j] < 220){csv$time[j] <- 3.3}
else if (csv$time[j] > 220 & csv$time[j] < 230){csv$time[j] <- 3.4}
else if (csv$time[j] > 230 & csv$time[j] < 240){csv$time[j] <- 3.5}
else if (csv$time[j] > 240 & csv$time[j] < 250){csv$time[j] <- 4.0}
else if (csv$time[j] > 250 & csv$time[j] < 260){csv$time[j] <- 4.1}
else if (csv$time[j] > 260 & csv$time[j] < 270){csv$time[j] <- 4.2}
else if (csv$time[j] > 270 & csv$time[j] < 280){csv$time[j] <- 4.3}

```

```

else if (csv$time[j] > 280 & csv$time[j] < 290){csv$time[j] <- 4.4}

else if (csv$time[j] > 290 & csv$time[j] < 300){csv$time[j] <- 4.5}

else if (csv$time[j] > 300){csv$time[j] <- 10}

}

csv$time[csv$time == 10] <- NA

csv <- na.omit(csv)

##Rewrite binary file with no detections for recordings with no
##detections. For recordings with detections, convert timeframe
##identifiers to binary scores and write final binary file with
##appropriate calling frequency.

if (nrow(csv) == 0) {write.csv(formatted.data, csv.id2, row.names =
                             FALSE)}

else for (j in 1:nrow(csv)){

  if (csv$time[j] == 0){formatted.data[1] <- 1}

  else if (csv$time[j] == 0.1){formatted.data[2] <- 1}

  else if (csv$time[j] == 0.2){formatted.data[3] <- 1}

  else if (csv$time[j] == 0.3){formatted.data[4] <- 1}

  else if (csv$time[j] == 0.4){formatted.data[5] <- 1}

  else if (csv$time[j] == 0.5){formatted.data[6] <- 1}

  else if (csv$time[j] == 1.0){formatted.data[7] <- 1}

  else if (csv$time[j] == 1.1){formatted.data[8] <- 1}

  else if (csv$time[j] == 1.2){formatted.data[9] <- 1}

  else if (csv$time[j] == 1.3){formatted.data[10] <- 1}

  else if (csv$time[j] == 1.4){formatted.data[11] <- 1}

  else if (csv$time[j] == 1.5){formatted.data[12] <- 1}

  else if (csv$time[j] == 2.0){formatted.data[13] <- 1}

  else if (csv$time[j] == 2.1){formatted.data[14] <- 1}

  else if (csv$time[j] == 2.2){formatted.data[15] <- 1}

```

```

else if (csv$time[j] == 2.3){formatted.data[16] <- 1}
else if (csv$time[j] == 2.4){formatted.data[17] <- 1}
else if (csv$time[j] == 2.5){formatted.data[18] <- 1}
else if (csv$time[j] == 3.0){formatted.data[19] <- 1}
else if (csv$time[j] == 3.1){formatted.data[20] <- 1}
else if (csv$time[j] == 3.2){formatted.data[21] <- 1}
else if (csv$time[j] == 3.3){formatted.data[22] <- 1}
else if (csv$time[j] == 3.4){formatted.data[23] <- 1}
else if (csv$time[j] == 3.5){formatted.data[24] <- 1}
else if (csv$time[j] == 4.0){formatted.data[25] <- 1}
else if (csv$time[j] == 4.1){formatted.data[26] <- 1}
else if (csv$time[j] == 4.2){formatted.data[27] <- 1}
else if (csv$time[j] == 4.3){formatted.data[28] <- 1}
else if (csv$time[j] == 4.4){formatted.data[29] <- 1}
else if (csv$time[j] == 4.5){formatted.data[30] <- 1}
}

formatted.data[31] <- ((sum(formatted.data[1:30]))/30)

formatted.data <- t(formatted.data)

csv.id2<-paste(file.list$Count.ID[i],
               "_final_detections_binary.csv",sep="")

write.csv(formatted.data, csv.id2, row.names = FALSE)
}

##Create file with all scores collated.

all.data <- data.frame(ModelName = character(), Object = character(),
                      stringsAsFactors = F)

for (i in 1:nrow(file.list)){

  csv.id2<-paste(file.list$Count.ID[i],
                 "_final_detections_binary.csv",sep="")

```

```

temp.data <- read.csv(csv.id2, header = T)

all.data <- rbind(all.data, temp.data)

}

all.data <- cbind(file.list$Count.ID[], all.data)

names(all.data) <- c("Count.ID", "0", "0.1", "0.2", "0.3", "0.4", "0.5",
                    "1", "1.1", "1.2", "1.3", "1.4", "1.5", "2", "2.1",
                    "2.2", "2.3", "2.4", "2.5", "3", "3.1", "3.2", "3.3",
                    "3.4", "3.5", "4", "4.1", "4.2", "4.3", "4.4", "4.5",
                    "Overall Frequency")

write.csv(all.data, "All Detections - Final Detector.csv", row.names =
          FALSE)

}

#####
##### 3 - Run functions #####
#####

##Set the appropriate working directory, change the file name to match file
##containing file list and run the detector.

create.detector()

file.list <- read.csv("List of files to be tested.csv", header = T)

run.detector(file.list)

```

Appendix VIII - Recorder locations and date sampled in short-term nocturnal analyses

Table A.VIII-1 Table of recorder locations and dates sampled nocturnally before and after the August 18 2014 Aorangi 1080 operation. Ticks (✓) and crosses (✗) indicate date/recorder combinations included and excluded from samples respectively.

		Aorangi 2014									
		Before					After				
		13/07/14	26/07/14	4/08/14	9/08/14	17/08/14	28/08/14	3/09/14	9/09/14	14/09/14	24/09/14
1080	Aorangi	BUL_L1_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		BUL_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✗
		BUL_L3_0m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		MAN_L1_100m	✓	✓	✓	✓	✓	✓	✗	✗	✗
		PIN_L0_L0	✓	✓	✓	✓	✓	✓	✓	✓	✓
		PIN_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✗
		TAU_L2_350m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TAU_L3_100m	✓	✗	✓	✓	✗	✗	✗	✗	✗
		WAI_L1-2	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WAI_L4_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WHA_L1_100m	✗	✗	✓	✓	✗	✓	✓	✓	✓
		WHA_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
Non-1080	Northern Rimutaka	ORO_5km_5km	✗	✗	✗	✓	✗	✓	✓	✗	✓
		ORO_L1_100m	✗	✗	✗	✓	✗	✓	✗	✗	✓
		ORO_SM2	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM3	✗	✓	✓	✓	✓	✓	✓	✓	✗
		ORO_SM4	✗	✓	✓	✓	✓	✗	✗	✓	✗
		ORO_SM5	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM6	✓	✗	✓	✓	✓	✓	✓	✓	✓
	Taranua	THN_SM1	✓	✓	✓	✗	✓	✓	✓	✓	✓
		THN_SM2	✓	✓	✓	✓	✓	✓	✓	✗	✗
		THN_SM3	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Tora Bush	TOR_SM1	✓	✗	✓	✓	✓	✓	✓	✓	✓
		TOR_SM2	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TOR_SM3	✓	✗	✓	✓	✓	✗	✓	✓	✓

Table A.VIII-2 Table of recorder locations and dates sampled nocturnally before and after the June 16-17 2017 Aorangi 1080 operation. Ticks (✓) and crosses (✗) indicate date/recorder combinations included and excluded from samples respectively.

		Aorangi 2017									
		Before					After				
		16/05/17	24/05/17	31/05/17	6/06/17	15/06/17	26/06/17	2/07/17	8/07/17	18/07/17	29/07/17
1080	Aorangi	BUL_L1_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		BUL_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		BUL_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		MAN_L1_100m	✓	✓	✓	✗	✗	✓	✓	✓	✓
		MAN_L1_450m	✓	✓	✓	✗	✓	✓	✓	✓	✓
		MAN_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		PIN_L1_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		PIN_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		PIN_L2_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		PIN_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WAI_L1_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WAI_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WAI_L4_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
Non-1080	Northern Rimutaka	ORO_5km_5km	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L1_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L1_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM3_SM3	✗	✓	✓	✓	✗	✓	✓	✓	✓
		ORO_SM4_SM4	✗	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM5_SM5	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM6_SM6	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L2_450m	✓	✗	✓	✓	✓	✓	✓	✗	✗
		WRM_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L3_450m	✗	✗	✓	✓	✗	✓	✓	✓	✓
		WRM_L4_100m	✓	✓	✓	✓	✗	✓	✓	✗	✗
		WRM_L4_450m	✓	✓	✓	✓	✗	✓	✓	✓	✓
		WRM_R1_R1	✓	✓	✓	✓	✗	✓	✓	✓	✓
		WRM_R2_R2	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_R3_R3	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table A.VIII-3 Table of recorder locations and dates sampled nocturnally before and after the July 30 2017 Southern Rimutaka 1080 operation. Ticks (✓) and crosses (✗) indicate date/recorder combinations included and excluded from samples respectively.

		Southern Rimutaka 2017									
		Before					After				
		26/06/17	2/07/17	8/07/17	18/07/17	29/07/17	5/08/17	18/08/17	23/08/17	30/08/17	12/09/17
1080	Southern Rimutaka	TUR_MW14	✓	✓	✓	✓	✓	✓	✗	✓	✓
		TUR_SG2	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TUR_XA10	✓	✓	✓	✓	✓	✓	✓	✗	✓
		TUR_XD1	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TUR_XD10	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TUR_XF1	✓	✓	✓	✓	✓	✓	✓	✓	✓
		TUR_XF10	✓	✓	✓	✓	✓	✓	✓	✓	✓
Non-1080	Northern Rimutaka	ORO_5km	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L1_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L1_450m	✓	✓	✓	✓	✓	✗	✓	✗	✓
		ORO_L2_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_L3_100m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM3	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM4	✓	✓	✓	✓	✓	✓	✗	✓	✓
		ORO_SM5	✓	✓	✓	✓	✓	✓	✓	✓	✓
		ORO_SM6	✓	✓	✓	✓	✓	✗	✗	✓	✓
		WRM_L2_450m	✓	✓	✓	✗	✗	✗	✗	✗	✓
		WRM_L3_100m	✓	✓	✓	✓	✗	✓	✓	✓	✓
		WRM_L3_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_L4_450m	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_R1	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_R2	✓	✓	✓	✓	✓	✓	✓	✓	✓
		WRM_R3	✓	✓	✓	✓	✓	✓	✓	✓	✓

Appendix IX - Dates sampled in long-term nocturnal analyses

Summer 2013-14					
Aorangi					
BUL_L1_100m	9/12/13	17/12/13	22/12/13	30/12/13	9/01/14
BUL_L2_100m	X	✓	✓	✓	X
BUL_L3_100m					
MAN_L1_100m					
MAN_L2_100m	X	✓	✓	✓	✓
PIN_L1_100m					
PIN_L2_100m					
TAU_L3_100m					
WAL_L1_100m	✓	✓	X	✓	X
WAL_L1-2_L1-2					
WAL_L2_100m					
WAL_L4_100m					
WHA_L1_100m					
WHA_L2_100m	✓	✓	✓	✓	X
WHA_L3_100m					
Northern Rimutaka					
ORO_5km_5km					
ORO_L1_100m					
ORO_L2_100m					
ORO_L3_100m	✓	✓	✓	✓	X
ORO_SM1_SM1					
ORO_SM2_SM2					
ORO_SM3_SM3					
ORO_SM4_SM4					
ORO_SM5_SM5					
ORO_SM6_SM6					
WRM_L2_100m					
WRM_L3_100m	X	✓	✓	X	✓
WRM_L4_100m					
WRM_L4_450m					

Table A.IX-1 Table of recorder locations sampled in long-term monitoring over the summer 2013-14 sampling season. Ticks (✓) and crosses (X) indicate date/recorder combinations included and excluded from samples respectively. Grey cells indicate inactive recorder locations.

Winter 2014					
Aorangi					
BUL_L1_100m	26/07/14	4/08/14	9/08/14	17/08/14	28/08/14
BUL_L2_100m	✓	✓	✓	✓	✓
BUL_L3_100m	✓	✓	✓	✓	✓
MAN_L1_100m					
MAN_L2_100m	✓	✓	✓	✓	✓
PIN_L1_100m					
PIN_L2_100m	✓	✓	✓	✓	✓
TAU_L3_100m	X	✓	✓	X	✓
WAL_L1_100m					
WAL_L1-2_L1-2	✓	✓	✓	✓	✓
WAL_L2_100m					
WAL_L4_100m	✓	✓	✓	✓	✓
WHA_L1_100m	X	✓	✓	X	✓
WHA_L2_100m	✓	✓	✓	✓	✓
WHA_L3_100m	X	✓	✓	✓	✓
ORO_5km_5km	X	X	✓	X	✓
ORO_L1_100m	X	X	✓	X	✓
ORO_L2_100m					
ORO_L3_100m					
ORO_SM1_SM1	X	X	X	X	✓
ORO_SM2_SM2	✓	✓	✓	✓	✓
ORO_SM3_SM3	✓	✓	✓	✓	✓
ORO_SM4_SM4	✓	✓	✓	✓	X
ORO_SM5_SM5	✓	✓	✓	✓	✓
ORO_SM6_SM6	X	✓	✓	✓	✓
WRM_L2_100m					
WRM_L3_100m					
WRM_L4_100m					
WRM_L4_450m					

Table A.IX-2 Table of recorder locations sampled in long-term monitoring over the winter 2014 sampling season. Ticks (✓) and crosses (X) indicate date/recorder combinations included and excluded from samples respectively. Grey cells indicate inactive recorder locations.

		Summer 2014-15							
		15/12/14	22/12/14	29/12/14	6/01/15	14/01/15	22/01/15		
Aorangi	BUL_L1_100m	X	✓	✓	✓	✓	✓	✓	
	BUL_L2_100m	X	✓	✓	✓	✓	✓	✓	
	BUL_L3_100m	X	✓	✓	✓	✓	✓	✓	
	MAN_L1_100m								
	MAN_L2_100m	✓	✓	✓	✓	✓	✓	✓	
	PIN_L1_100m								
	PIN_L2_100m								
	TAU_L3_100m								
	WAI_L1_100m	X	X	✓	X	✓	X	X	
	WAI_L1-2_L1-2	X	✓	X	✓	✓	✓	✓	
Northern Rimutaka	WAI_L2_100m								
	WAI_L4_100m	✓	✓	✓	✓	✓	✓	✓	
	WHA_L1_100m	✓	✓	✓	✓	✓	✓	✓	
	WHA_L2_100m	✓	✓	✓	✓	✓	✓	✓	
	WHA_L3_100m	✓	✓	✓	✓	✓	✓	✓	
	ORO_5km_5km	✓	✓	✓	✓	X	X	X	
	ORO_L1_100m	✓	✓	✓	✓	X	X	X	
	ORO_L2_100m								
	ORO_L3_100m	X	✓	✓	X	X	X	X	
	ORO_SM1_SM1	✓	✓	✓	✓	✓	✓	✓	
	ORO_SM2_SM2	✓	✓	✓	✓	X	✓	✓	
	ORO_SM3_SM3	✓	✓	✓	✓	✓	✓	✓	
	ORO_SM4_SM4	✓	✓	✓	✓	✓	✓	✓	
	ORO_SM5_SM5	✓	✓	✓	✓	✓	✓	✓	
	ORO_SM6_SM6	✓	✓	✓	✓	✓	✓	✓	
	WRM_L2_100m	X	X	✓	✓	✓	✓	✓	
	WRM_L3_100m	X	✓	✓	✓	✓	✓	✓	
	WRM_L4_100m	X	✓	✓	✓	✓	✓	✓	
	WRM_L4_450m								

Table A.IX-3 Table of recorder locations sampled in long-term monitoring over the summer 2014-15 sampling season. Ticks (✓) and crosses (X) indicate date/recorder combinations included and excluded from samples respectively. Grey cells indicate inactive recorder locations.

		Winter 2015							
		6/07/15	14/07/15	22/07/15	30/07/15	5/08/15	13/08/15		
Aorangi	BUL_L1_100m								
	BUL_L2_100m	X	X	X	✓	✓	✓	✓	
	BUL_L3_100m								
	MAN_L1_100m	X	X	✓	✓	✓	✓	✓	
	MAN_L2_100m								
	PIN_L1_100m	✓	✓	✓	X	X	X	X	
	PIN_L2_100m								
	TAU_L3_100m								
	WAI_L1_100m	✓	✓	✓	✓	✓	✓	✓	
	WAI_L1-2_L1-2								
Northern Rimutaka	WAI_L2_100m	✓	X	X	X	X	X	X	
	WAI_L4_100m								
	WHA_L1_100m								
	WHA_L2_100m	X	X	X	X	✓	✓	✓	
	WHA_L3_100m								
	ORO_5km_5km								
	ORO_L1_100m	✓	✓	✓	✓	✓	✓	✓	
	ORO_L2_100m								
	ORO_L3_100m	✓	✓	✓	✓	✓	✓	✓	
	ORO_SM1_SM1								
	ORO_SM2_SM2								
	ORO_SM3_SM3								
	ORO_SM4_SM4								
	ORO_SM5_SM5								
	ORO_SM6_SM6								
	WRM_L2_100m	✓	X	X	X	X	X	X	
	WRM_L3_100m	X	✓	✓	✓	✓	✓	✓	
	WRM_L4_100m								
	WRM_L4_450m								

Table A.IX-4 Table of recorder locations sampled in long-term monitoring over the winter 2015 sampling season. Ticks (✓) and crosses (X) indicate date/recorder combinations included and excluded from samples respectively. Grey cells indicate inactive recorder locations.

Summer 2015-16						
	26/12/15	31/12/15	5/01/16	10/01/16	17/01/16	29/01/16
Aorangi						
BUL_L1_100m						
BUL_L2_100m						
BUL_L3_100m						
MAN_L1_100m	✓	✓	✓	✓	✓	✓
MAN_L2_100m	✓	✓	✓	✓	✓	✓
PIN_L1_100m						
PIN_L2_100m						
TAU_L3_100m						
WAL_L1_100m						
WAL_L1-2_L1-2						
WAL_L2_100m						
WAL_L4_100m						
WHA_L1_100m						
WHA_L2_100m	✓	✓	✓	✓	✓	✓
WHA_L3_100m						
Northern Rimutaka						
ORO_5km_5km						
ORO_L1_100m	X	X	X	✓	✓	✓
ORO_L2_100m						
ORO_L3_100m	X	X	X	X	✓	✓
ORO_SMI1_SMI1						
ORO_SM2_SM2						
ORO_SM3_SM3						
ORO_SM4_SM4						
ORO_SM5_SM5						
ORO_SM6_SM6						
WRM_L2_100m	✓	✓	✓	✓	✓	✓
WRM_L3_100m	✓	✓	X	X	X	X
WRM_L4_100m	✓	✓	X	✓	X	X
WRM_L4_450m						

Table A.IX-5 Table of recorder locations sampled in long-term monitoring over the summer 2015-16 sampling season. Ticks (✓) and crosses (X) indicate date/recorder combinations included and excluded from samples respectively. Grey cells indicate inactive recorder locations.

Winter 2016						
	21/07/16	26/07/16	2/08/16	9/08/16	16/08/16	23/08/16
Aorangi						
BUL_L1_100m						
BUL_L2_100m	✓	✓	✓	X	X	X
BUL_L3_100m						
MAN_L1_100m	X	X	✓	✓	✓	✓
MAN_L2_100m	X	X	✓	✓	✓	✓
PIN_L1_100m						
PIN_L2_100m						
TAU_L3_100m						
WAL_L1_100m	✓	✓	✓	✓	X	X
WAL_L1-2_L1-2						
WAL_L2_100m						
WAL_L4_100m						
WHA_L1_100m	X	X	X	X	✓	✓
WHA_L2_100m	X	X	X	X	✓	✓
WHA_L3_100m						
Northern Rimutaka						
ORO_5km_5km	✓	✓	✓	✓	✓	✓
ORO_L1_100m	✓	✓	✓	✓	✓	✓
ORO_L2_100m						
ORO_L3_100m	✓	✓	✓	X	X	X
ORO_SMI1_SMI1						
ORO_SM2_SM2						
ORO_SM3_SM3						
ORO_SM4_SM4						
ORO_SM5_SM5						
ORO_SM6_SM6						
WRM_L2_100m						
WRM_L3_100m	✓	✓	✓	✓	X	X
WRM_L4_100m	✓	✓	✓	✓	X	X
WRM_L4_450m						

Table A.IX-6 Table of recorder locations sampled in long-term monitoring over the winter 2016 sampling season. Ticks (✓) and crosses (X) indicate date/recorder combinations included and excluded from samples respectively. Grey cells indicate inactive recorder locations.

Summer 2016-17

	24/12/16	29/12/16	7/01/17	13/01/17	20/01/17	28/01/17
Aorangi	BUL_L1_100m	✓	✓	✓	✓	✓
	BUL_L2_100m	✓	✓	✓	✓	✓
	BUL_L3_100m					
	MAN_L1_100m	✓	✓	✓	✓	✓
	MAN_L2_100m	✓	✓	✓	✓	✓
	PIN_L1_100m	✓	✓	✓	✓	✓
	PIN_L2_100m	✓	✓	✓	✓	✓
	TAU_L3_100m	X	✓	✓	✓	✓
	WAL_L1_100m	✓	✓	✓	✓	✓
	WAL_L1-2_L1-2					
Northern Rimutaka	WAL_L2_100m	✓	✓	✓	✓	✓
	WAL_L4_100m	✓	✓	✓	✓	✓
	WHA_L1_100m	✓	✓	✓	✓	✓
	WHA_L2_100m					
	WHA_L3_100m	✓	✓	✓	✓	✓
	ORO_5km_5km	✓	✓	✓	✓	✓
	ORO_L1_100m	✓	✓	✓	✓	✓
	ORO_L2_100m	✓	✓	X	✓	✓
	ORO_L3_100m	✓	✓	✓	✓	✓
	ORO_SM1_SM1					
Northern Rimutaka	ORO_SM2_SM2					
	ORO_SM3_SM3					
	ORO_SM4_SM4	✓	X	X	X	X
	ORO_SM5_SM5	✓	✓	✓	✓	X
	ORO_SM6_SM6	✓	✓	✓	✓	✓
	WRM_L2_100m					
	WRM_L3_100m	✓	✓	✓	✓	✓
	WRM_L4_100m	✓	✓	✓	✓	✓
	WRM_L4_450m	✓	✓	✓	✓	✓

Table A.IX-7 Table of recorder locations sampled in long-term monitoring over the summer 2016-17 sampling season. Ticks (✓) and crosses (X) indicate date/recorder combinations included and excluded from samples respectively. Grey cells indicate inactive recorder locations.

Winter 2017

	8/07/17	18/07/17	29/07/17	5/08/17	18/08/17	23/08/17
Aorangi	BUL_L1_100m	✓	✓	✓	✓	✓
	BUL_L2_100m	✓	✓	✓	✓	X
	BUL_L3_100m	✓	✓	X	✓	✓
	MAN_L1_100m	✓	✓	✓	✓	✓
	MAN_L2_100m	✓	✓	✓	✓	✓
	PIN_L1_100m	✓	✓	✓	✓	✓
	PIN_L2_100m	✓	✓	✓	✓	✓
	TAU_L3_100m					
	WAL_L1_100m	✓	✓	X	✓	✓
	WAL_L1-2_L1-2					
Northern Rimutaka	WAL_L2_100m	✓	✓	X	✓	✓
	WAL_L4_100m	✓	✓	✓	✓	✓
	WHA_L1_100m					
	WHA_L2_100m					
	WHA_L3_100m					
	ORO_5km_5km	✓	✓	✓	✓	✓
	ORO_L1_100m	✓	✓	✓	✓	✓
	ORO_L2_100m	✓	✓	✓	✓	✓
	ORO_L3_100m	✓	✓	✓	✓	✓
	ORO_SM1_SM1					
Northern Rimutaka	ORO_SM2_SM2					
	ORO_SM3_SM3	✓	✓	✓	✓	✓
	ORO_SM4_SM4	✓	✓	✓	✓	X
	ORO_SM5_SM5	✓	✓	✓	✓	✓
	ORO_SM6_SM6	✓	✓	✓	X	X
	WRM_L2_100m					
	WRM_L3_100m	✓	✓	✓	X	✓
	WRM_L4_100m	✓	X	X	X	X
	WRM_L4_450m	✓	✓	✓	✓	✓

Table A.IX-8 Table of recorder locations sampled in long-term monitoring over the winter 2017 sampling season. Ticks (✓) and crosses (X) indicate date/recorder combinations included and excluded from samples respectively. Grey cells indicate inactive recorder locations.

Appendix X - Recorder location specifications

Table A.X-1 Table of GPS information for recorder locations monitored at the Bull Hill (BUL), Mangatoetoe (MAN), Putangirua Pinnacles (PIN), Tauanui (TAU), Waihora (WAI) and Whawanui (WHA) study sites in the Aorangi Range study area.

Recorder Location	Latitude	Longitude	Elevation (m)
BUL_L1_100m	-41.346363	175.361700	463
BUL_L2_100m	-41.349087	175.351586	565
BUL_L3_0m	-41.349730	175.344445	724
BUL_L3_100m	-41.349683	175.343316	757
MAN_L1_100m	-41.572561	175.275631	276
MAN_L1_450m	-41.573795	175.279714	417
MAN_L2_100m	-41.576278	175.285256	569
PIN_L0_SM	-41.444428	175.248418	260
PIN_L1_100m	-41.447059	175.254929	384
PIN_L2_100m	-41.454384	175.260604	454
PIN_L2_450m	-41.456142	175.262222	390
PIN_L3_100m	-41.459091	175.272521	401
TAU_L2_350m	-41.399536	175.291531	311
TAU_L3_100m	-41.401421	175.284762	263
WAI_L1_100m	-41.351424	175.290075	141
WAI_L1-2_SM	-41.353787	175.296076	404
WAI_L2_100m	-41.353296	175.301018	434
WAI_L4_100m	-41.345293	175.282909	181
WHA_L1_100m	-41.528224	175.342899	207
WHA_L2_100m	-41.535858	175.342141	487
WHA_L3_100m	-41.537429	175.348314	324

Table A.X-2 Table of GPS information for recorder locations monitored at the Orongorongo (ORO) and Waiorongomai (WRM) study sites in the Northern Rimutaka Range study area.

Recorder Location	Latitude	Longitude	Elevation (m)
ORO_5km	-41.244775	175.037493	281
ORO_L1_100m	-41.244558	175.051680	577
ORO_L1_450m	-41.242134	175.054723	522
ORO_L2_100m	-41.237105	175.057353	525
ORO_L3_100m	-41.229557	175.064997	505
ORO_SM1	-41.245321	175.031084	197
ORO_SM2	-41.247793	175.051439	537
ORO_SM3	-41.259245	175.046600	752
ORO_SM4	-41.277770	175.053642	590
ORO_SM5	-41.284976	175.050180	509
ORO_SM6	-41.291175	175.043800	457
WRM_L2_100m	-41.259184	175.110298	115
WRM_L2_450m	-41.256853	175.107137	100
WRM_L3_100m	-41.255232	175.111084	202
WRM_L3_450m	-41.252474	175.111938	239
WRM_L4_100m	-41.254508	175.127121	113
WRM_L4_450m	-41.251666	175.126345	71
WRM_R1	-41.256795	175.130573	64
WRM_R2	-41.254485	175.119200	88
WRM_R3	-41.257543	175.114691	91

Table A.X-3 Table of GPS information for recorder locations monitored at the Tora Bush study site in the Tora Bush study area.

Recorder Location	Latitude	Longitude	Elevation (m)
TOR_SM1	-41.489212	175.460831	379
TOR_SM2	-41.485504	175.466425	359
TOR_SM3	-41.480999	175.470197	366

Table A.X-4 Table of GPS information for recorder locations monitored at the Tauherenikau (THN) study site in the Tararua Range study area.

Recorder Location	Latitude	Longitude	Elevation (m)
THN_SM1	-40.990653	175.369170	546
THN_SM2	-40.987364	175.353832	619
THN_SM3	-40.981830	175.354709	600

Table A.X-5 Table of GPS information for recorder locations monitored at the Turere Valley (TUR) study site in the Southern Rimutaka study area.

Recorder Location	Latitude	Longitude	Elevation (m)
TUR_MW14	-41.293820	174.984090	577
TUR_SG2	-41.289706	174.984600	500
TUR_XA10	-41.295400	174.992570	572
TUR_XD1	-41.302969	174.979142	668
TUR_XD10	-41.306719	174.979078	613
TUR_XF1	-41.312976	174.973888	698
TUR_XF10	-41.315584	174.977450	594

