Notornis, 2019, Vol. 66: 109–128
0029-4470 © The Ornithological Society of New Zealand Inc.

Effectiveness and efficiency of avian species detection: a comparison between field observers and automatic recording devices

JAMES A.J. MORTIMER*
TERRY C. GREENE
Department of Conservation, Level 3, Grand Central, 161 Cashel Street, Christchurch 8011, New Zealand

PAUL VAN DAM-BATES
University of St Andrews, Buchanan Gardens, St Andrews, KY16 9LX, United Kingdom

IAN M. WESTBROOKE
Department of Conservation, Level 3, Grand Central, 161 Cashel Street, Christchurch 8011, New Zealand

Abstract: The monitoring of animal populations is essential for reporting on the state of the environment, with birds often used as indicators of ecosystem health. Traditionally, bird monitoring has been done by field observers; however, there has been recent interest in use of automatic recording devices (ARDs) as an alternative. A monitoring programme managed by the New Zealand Department of Conservation (DOC), used observers and ARDs concurrently for three survey seasons, providing the opportunity to compare results in terms of effectiveness and efficiency. The difference in species-richness estimates from the two methods was small, with the observer method detecting slightly higher numbers of species in all habitat types. Detection probabilities for individual species, derived from occupancy analysis, were similar between methods, with a few exceptions: bellbird (*Anthornis melanura*), brown creeper (*Mohoua novaeseelandiae*), tūī (*Prosthemadera novaeseelandiae*), North/South Island robin (*Petroica longipes/australis*), and rifleman (*Acanthisitta chloris*). Bellbird and rifleman had a higher probability of being detected by ARDs, whilst the remainder were more likely to be detected by observers. Differences in detection probability may be due to identification confusion in the case of bellbird and tūī, and observer ability to detect and identify birds visually for brown creeper and North/South Island robin. The relationship between indices of abundance from the observer and ARD methods varied between species and habitat types. These inconsistencies suggested that the ARD results did not correlate closely with observed abundance, which may limit the ARD method to provision of confirmed presence data. Observer counts proved to be more time-efficient given present levels of processing technology, mainly due to the longer processing time required for ARD recordings. However higher numbers of people were required for observer counts, which may be problematic when there is a shortage of appropriately skilled observers at the required time of year.

Mortimer, J.A.J.; Greene, T.C.; Van Dam-Bates, P.; Westbrooke, I.M. 2019. Effectiveness and efficiency of avian species detection: a comparison between field observers and automatic recording devices. *Notornis 66(3)*: 109–128.

Key words: bird monitoring, detection probability, species-richness, acoustic recording, indices of abundance

INTRODUCTION

There are strong social, economic, and ethical drivers, both national and international, behind the development of biodiversity monitoring systems which enable measurement of biodiversity trends and the impacts of management for reporting on the state of the environment (Allen *et al*. 2003; Lee *et al*. 2005). The National Biodiversity Monitoring and Reporting System, administered by the New

---

Received 23 June 2018; accepted 15 June 2019
*Correspondence: *jmortimer@doc.govt.nz*

Zealand Department of Conservation (DOC), was recently developed and implemented to provide a multi-tiered monitoring framework to enable reporting on national trends in biodiversity (referred to as "Tier 1"), effectiveness and impacts of management ("Tier 2"), and research/long-term monitoring objectives ("Tier 3"; Allen *et al*. 2009). Monitoring populations of common and widespread bird species is one of the key measures of the programme, as birds can be useful environmental indicators (Bibby *et al*. 2000). This is particularly the case when species diversity, distribution and abundance, species-habitat relationships, and responses to environmental change or management can be determined (Simons *et al*. 2007).

Implementing a national monitoring programme requires considerable resources, especially the mobilisation of appropriately-skilled staff and sufficient finances, both of which may be limited. The challenge is, therefore, to maximise data collection efficiency wherever possible, without compromising data quality. Use of novel technologies has the potential to reduce time and staffing requirements and improve efficiency. Recent technological developments have produced tools with the potential to augment or even replace the more traditional field observer approaches to ecological monitoring (e.g. Xie *et al*. 2008; Nagendra *et al*. 2013). Use of automatic recording devices (ARDs) has increased dramatically in recent years (Brandes 2008; Steer 2010; Frick 2013; Cook & Hartley 2018), with ever-more elaborate systems enabling, for example, monitoring of complete habitats via use of sensor networks (Szewczyk *et al*. 2004).

In bird monitoring, ARDs are gaining popularity as an increasing range of devices has become available (Brandes 2008; Frick 2013). Upon initiation of the Tier 1 Monitoring Programme, DOC recognised an opportunity to test the relative effectiveness for species detection and efficiency between field observer and ARD-based bird monitoring techniques, and to inform decisions on long-term choice of survey method.

Several previous studies have compared species detection abilities of observers and ARDs. Haselmeyer & Quinn (2000) found that the two methods were overall equally effective at detecting species-richness, although ARDs detected more species when richness was high and observers were more effective for rarely-heard species. Likewise, Sedlácek *et al*. (2015) found that the ARD and observer methods provided similar estimates for species-richness, abundance, and community composition. Wimmer *et al*. (2013) showed that ARDs were able to detect a higher number of species than observers, whereas in some other studies the observer method was more effective (Hutto & Stutzman 2009; Leach *et al*. 2016; Stewart & Hasenbank 2018). Holmes *et al*. (2014) found ARDs to be the most time-efficient method; however, Hutto & Stutzman (2009) found ARDs to be less time-efficient than observers. All these studies focused on a limited number of sites in their respective countries and were not part of a national monitoring programme. To our knowledge, there have been three studies to date which compared observers and ARDs within New Zealand (Digby *et al*. 2013; Stewart & Hasenbank 2018; Bombaci & Pejchar 2019). However, the comparisons were for single species and/or at only one or few geographic locations. This study compared data collected simultaneously by field observers and ARDs for a national monitoring programme on Public Conservation Lands (PCL), and simultaneously assessed the effectiveness of the two methods in terms of species detection, estimation of abundance, and efficient use of resources.

MATERIALS AND METHODS

Field sampling protocols

The Tier 1 Monitoring Programme was initiated in 2011 and is based upon a randomly-placed 8 km grid, covering mainland New Zealand and offshore islands. A total of 1,354 randomly-selected grid intersection points within Public Conservation Lands (which extend to one third of New Zealand's land area; DOC 2015), were used to determine sampling locations. Locations are sampled on a five-year cycle, meaning approximately 270 are sampled each survey season (October to March inclusive).

Sampling locations were established and measured as per the methods described in MacLeod *et al*. (2012) and Mortimer & Greene (2017). At each sampling location there were 5 bird count stations, spaced approximately 200 m apart (Fig. 1). A single ARD was deployed at each bird count station, set to record concurrently with the field observer bird count.

ARDs were set to record continuously for one nocturnal time-period (2000 h – 0600 h; New Zealand daylight saving time, GMT + 13 hrs) and one diurnal time-period (0700 h – 1300 h). The ARDs were developed and designed by DOC, each incorporating 4 x wm61a electrets microphones in parallel with a foam 'pop filter' and custom-made low noise pre-amplifier with a DSP anti-aliasing filter. Recordings were stored on Secure Digital (SD) memory card as a series of compressed 32 kHz, 16-bit audio files in waveform audio format (.WAV file extension), with a bit-rate of 512 kbps, each approximately 15 minutes in length.

A 5-minute bird count (an index of relative abundance, not adjusted for detection probability) was completed at each station, using standardised

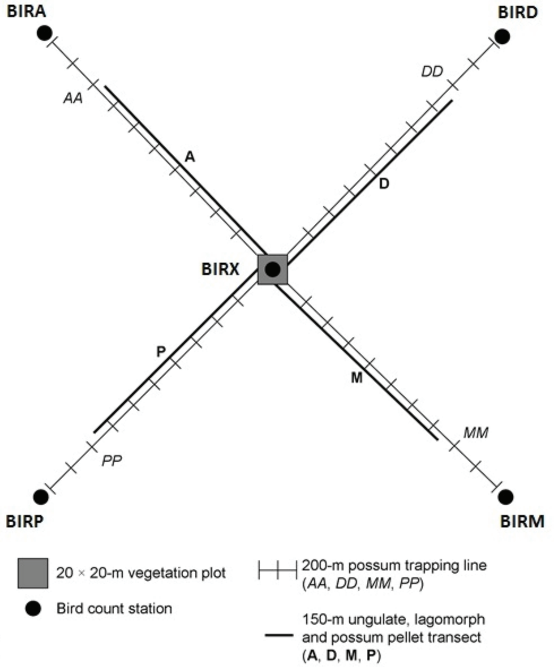

Figure 1. Tier 1 Monitoring Programme sampling location design, showing locations of bird count stations (BIRA, BIRD, BIRM, BIRP, and BIRX).

methods adapted from Dawson and Bull (1975). Although all birds seen or heard were noted, whether these observations were aural or visual was not recorded. To enable comparison between observer counts and processed ARD recordings, the precise start of each count was identified by a field observer clearly vocalising the start of the 5-minute bird count, to effectively synchronise both methods. Counts began not less than 1 hour after official sunrise and were completed by 1300 h. If time allowed, two 5-minute bird counts were completed at each station, with a minimum of 60 minutes between counts at the same station. A single 5-minute count period per station was used in the comparison between ARD and observer methods (usually the first period, unless in the event of adverse weather or ARD failure, in which case the second period was used). In addition to number and species of birds, observers recorded a range of environmental details as categorical covariates (i.e. temperature, sun, precipitation, wind, and noise; see Appendix 1). In practice, many sampling locations had fewer than five stations due to abandonment on safety grounds or excessive environmental noise (e.g. rivers, etc.). In addition, some recordings were excluded from processing, due to excessive noise from wind, rain, or other environmental sources (e.g. invertebrates). In cases where data for a 5-minute period were excluded for one method, the corresponding data for the other method in the same time period were also excluded (i.e. all 5-minute periods included in the analysis had data from the field observer and ARD). In summary, 47 sampling locations had 5 stations for which observations were recorded using both methods, 70 sampling locations had 4 stations, 92 sampling locations had 3 stations, 131 sampling locations had 2 stations, and 58 sampling locations had 1 station. During the first three seasons of the Tier 1 Monitoring Programme (2011–12, 2012–13, and 2013–14), diurnal ARD recordings were processed from 65, 88, and 245 sampling locations respectively, resulting in a total of 1,112 5-minute periods from 398 sampling locations with a field observer 5-minute bird count and a corresponding processed ARD 5-minute period (Fig. 2). The first two survey seasons had a reduced number of sampling locations surveyed due to phased implementation of the programme.

Processing of ARD recordings

A single 5-minute period was processed for each

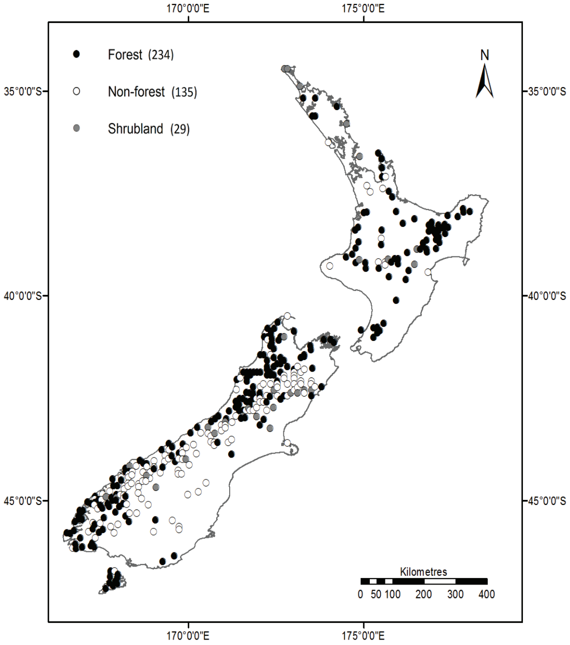

Figure 2. Tier 1 sampling locations across New Zealand at which observer counts were completed and corresponding ARD recordings processed, with habitat type (figures in brackets show number of sampling locations per habitat).

bird count station at each sampling location. ARD recordings were processed manually by experienced ornithologists, using the custom-designed Freebird call analysis software, version 1.1.6.4 (Freebird 2013). This generated sonograms from the recordings and allowed audio playback for species identification. Identified calls were tagged by drawing a box around the appropriate part of the sonogram and labelling it with the species name from a drop-down list. The processor identified and tagged presence of each species within 10-second blocks (30 blocks per 5-minute period). A species would not be tagged more than once in the same 10-second block, irrespective of the number of calls. If a single call spanned multiple 10-second blocks, then the species was tagged as present in each block. To limit effects of fatigue, processors were advised not to spend more than 25 hours per week on processing. Upon completion of processing a 5-minute period, the results were exported in comma separated values (CSV) format and later aggregated for analysis.

Data analyses

Each sampling location was assigned to the broad habitat type corresponding to assessment of the 20 x 20 m vegetation survey plot (see Fig. 1): forest (234 sampling locations), non-forest (135 sampling locations), or shrubland (29 sampling locations). This classification was used to divide the sampling locations by habitat for graphical presentation of results. For statistical analyses, each bird count station was assigned a habitat type (forest, non-forest, or shrubland), using Land Cover Database (LCDB) classifications (Thompson *et al*. 2003): see Appendix 2. For occupancy analysis this allowed the inclusion of habitat as a survey covariate in the estimation of detection probabilities for each survey method. Analyses were performed using the R statistical software (version 3.1.2; R Core Team 2014) except occupancy estimates, which used program PRESENCE (version 10.5; Hines 2006). Comparisons between observer-based counts and ARD recordings were made using a range of metrics, namely species-richness, detection probability, and indices of abundance.

Relative effectiveness at detecting species-richness was assessed by summing the total number of species recorded for each 5-minute period, for each survey method, then plotting the observer species-richness against ARD species-richness, with loess curves to model relationships. The data were then modelled using a generalised linear mixed model (GLMM) with a Poisson distribution, with observer species-richness as the response variable and ARD species-richness, habitat, wind, and noise as explanatory variables. Sampling location was included as a random effect, to account for the lack of independence of 5-minute count periods from the same sampling locations. Prior to modelling, ARD species-richness was normalized by subtracting the mean, and then dividing by the standard deviation (to produce a Z value). There were four candidate models, which included various combinations of explanatory variables (Table 1). The models were fitted with a unique intercept and slope (against the ARD species-richness) for each habitat type and the model with the lowest Akaike Information Criterion (AIC; Burnham & Anderson 2002) value was selected. The coefficient estimates indicate, on a logarithmic scale, the effect size of the explanatory variables on the response variable. Diagnostic plots were used to test validity of the model, following Zuur *et al*. (2013), and the data were tested for over-dispersion.

Occupancy analysis (MacKenzie *et al*. 2018) was carried out for the top 16 species (those recorded at the highest number of sampling locations). For each 5-minute count period, each species was assigned a 1 or 0 to indicate whether or not it was detected by each method. This resulted in four possible ‘detection histories’: 00 (not detected by either method); 10 (detected by the observer method but not the ARD method); 01 (not detected by the observer method but detected by the ARD method); and 11 (detected by both methods). From this, we were able to calculate the detection probability for each species in a 5-minute count period. For species that occur on only the North or only the South Island, data were included only for sampling locations from the relevant island. A single-season multi-method analysis approach was used, which accounts for the lack of independence between detection methods for each sampling occasion (Nichols *et al*. 2008). Eight biologically plausible *a priori* candidate models were considered, which included various combinations of what were considered to be the most important survey covariates, i.e. habitat, wind, and noise (Table 2; Appendix 1; Robbins 1981; Pacifici *et al*. 2008). All models assumed psi (Ψ; the probability that a site is occupied by the species) and theta (Θ; the probability that individuals are available for detection using a method, given presence) were constant, and p (the probability of detecting the species using a method in a survey) remained constant through time. Since we were only interested in p, there was no reason to allow psi or theta to vary by inclusion of covariates. Model fit was assessed using AIC (Burnham & Anderson 2002).

To compare our ability to measure changes in abundance, an index of relative abundance was created for each method, again only for the top 16 species. For the observer counts, this was simply the raw count of individuals of a species recorded

Table 1. Candidate models for GLMM analysis of species-richness estimates from the observer and ARD methods. OBS = observer species-richness; ARD = ARD species-richness; hab = habitat type (forest, non-forest, shrubland); w = wind (0–3); n = noise (0–2); sl = sampling location (included as a random effect). The '/' indicates that each model was fitted with a unique intercept and slope (against the ARD species-richness) for each habitat type.

Model	Model definition	No. of fixed effects
SR1	OBS ~ hab/ARD + w + n + (sl)	11
SR2	OBS ~ hab/ARD + w + (sl)	9
SR3	OBS ~ hab/ARD + n + (sl)	7
SR4	OBS ~ hab/ARD + (sl)	6

Table 2. *A priori* candidate models for individual species detection. Ψ = psi (the probability that a site is occupied by the species); Θ = theta (the probability that individuals are available for detection using a method, given presence); *p* = probability of detection; h = habitat type (forest, non-forest, shrubland); m = method (observer, ARD); w = wind (0–3); n = noise (0–2); K = the number of parameters in the model.

Model	Model definition	K	Model	Model definition	K
SD1	Ψ(.), Θ(.), *p*(m)	4	SD5	Ψ(.), Θ(.), *p*(m+w+n)	11
SD2	Ψ(.), Θ(.), *p*(m+h)	7	SD6	Ψ(.), Θ(.), *p*(m+h+w)	11
SD3	Ψ(.), Θ(.), *p*(m+w)	8	SD7	Ψ(.), Θ(.), *p*(m+h+n)	10
SD4	Ψ(.), Θ(.), *p*(m+n)	7	SD8	Ψ(.), Θ(.), *p*(m+h+w+n)	14

during each 5-minute bird count. For ARDs, the index was a count of the number of 10-second blocks in which a species was recorded for each 5-minute period, which was essentially an index of how frequently a bird species vocalised. This was named the 'acoustic prevalence index' (API), a term used by Cook & Hartley (2018) and calculated in a similar manner. The aim was to carry out an exploratory analysis of relationships between the observer index and API, identifying any that were consistent and predictable. Indices from the two methods were plotted against each other, with loess curves to model relationships for each habitat. Indices from the two methods were then compared for each species in each habitat type using GLMMs with a Poisson distribution. Prior to modelling, the API values were normalized by subtracting the mean, and then dividing by the standard deviation (to produce Z values). The models were specified using the same method as for species-richness analysis, but with observer index as the response variable and API as an explanatory variable.

A comparison of method efficiency was based upon estimates of mean time spent per sampling location for each method, effectively a proxy for cost. Detailed information was not available for all sampling locations. However, estimates could be calculated from observer field diaries and notes from ARD processing and data entry personnel. The ARD processing time estimates included time for processing of files (manual identification of bird calls) and data/file management. Travel time to and from the sampling location was excluded, as this would be identical regardless of which method was employed. The number of skilled individuals employed and number of 5-minute counts completed/5-minute ARD periods processed per person were also compared between methods, to provide insight into staff resource requirements.

RESULTS

Detection of species-richness

The number of species detected varied considerably between 5-minute periods for both methods (Fig. 3). From a total of 398 sampling locations, there were 93 where both methods detected no species. Interestingly, there were 57 sampling locations at which the observer method detected at least one species and the ARD method detected no species (mostly in non-forest habitats). Conversely, at three sampling locations the ARD method detected at least one species whilst the observer method detected none (all non-forest). Mean species-richness per 5-minute period was comparable for forest (observer = 5.03, 95% CI [4.88, 5.18]; ARD = 5.10, 95% CI [4.93, 5.28]) and shrubland (observer = 5.34, 95% CI [4.83, 5.85]; ARD = 4.80, 95% CI [4.22, 5.39]), with mean species-richness for non-forest being somewhat lower (observer = 2.77, 95% CI [2.55, 2.99]; ARD = 2.21, 95% CI [2.01, 2.41]). The loess model curves indicated a linear positive relationship between indices from the two methods, in all three habitat types (Fig. 3). There were two competing candidate GLMM models: SR2 and SR4 (Table 1), which achieved very similar delta AIC values (SR2 = 0, SR4 = 0.263). Model selection suggested that wind could be included as an explanatory variable; however, the effects of wind were small (Table 3). The existing data did not support the inclusion of noise into the model. The forest habitat type had the largest effect on species-richness, closely followed by shrubland (i.e. species-richness was generally highest in these habitat types; Table 3). ARD species-richness had an effect on observer species-richness, which means that for each increase in ARD species-richness, there was a slightly larger increase in mean observer species-richness. The effect of ARD species-richness was largest in non-forest (Table 3). Diagnostic plots did not indicate any problems with the model and the data were not over-dispersed.

A species accumulation curve (Southwood & Henderson 2000) was produced to compare the mean cumulative number of species detected with each successive bird count station surveyed, for both methods (Fig. 4). Significantly higher numbers of species were detected with increasing numbers of stations surveyed ($F(4, 2218) = 68.116$, $p < 0.001$). However, the results showed no significant difference in number of species detected by the two methods ($F(1, 2218) = 0.017$, $p = 0.897$).

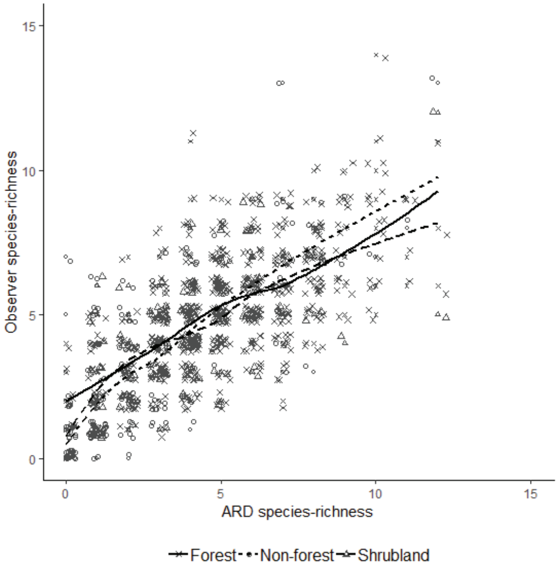

Figure 3. Comparison of species-richness detection per 5-minute period, for the observer and ARD methods. Loess curves show the relationship between indices from the two methods for each habitat type. Individual data points have been displayed using jittering to make those with the same values visible.

Detection of individual species

Occupancy analysis for individual species revealed that habitat was an important factor influencing detection probability for most species. Wind and/or noise were important for all but three species (Table 4).

Values for p (the probability of detecting the species using a method in a survey), where wind

Table 3. Results of GLMM analysis (model SR2) to test for relationships between species-richness estimates from the observer method (response variable) and ARD method, wind and habitat type (explanatory variables). ARD species-richness values were normalized by subtracting the mean, and then dividing by the standard deviation. The estimate for forest is not shown as this was the reference habitat type to which non-forest and shrubland were compared. The '/' indicates that the model was fitted with a unique intercept and slope (against the ARD species-richness) for that habitat type.

Variable	Estimate	Standard error	p value
(Intercept)	1.511	0.024	<0.001
Non-forest	-0.398	0.044	<0.001
Shrubland	-0.080	0.056	0.149
Wind 1	-0.068	0.040	0.094
Wind 2	-0.068	0.065	0.029
Wind 3	-0.178	0.096	0.064
Forest/ARD	0.285	0.021	<0.001
Non-forest/ARD	0.501	0.026	<0.001
Shrubland/ARD	0.339	0.049	<0.001

and noise variables, if included in the selected model for that species, were both equal to zero, were compared (Fig. 5; for effects of other wind and noise values see Appendix 3). Where habitat was included in the model, probability of detection was generally highest in forest or shrubland habitats and lowest in non-forest. However, common redpoll (*Carduelis flammea*) detection probability was higher in non-forest and shrubland, whilst chaffinch (*Fringilla coelebs*) detection probabilities were similar across all three habitat types. Habitat was not included in the selected models for brown creeper, New Zealand fantail (*Rhipidura fuliginosa*), song thrush (*Turdus philomelos*), and whitehead (*Mohoua albicilla*). Probability of detection for individual species was similar between methods with the exceptions of (in all habitat types) brown creeper, and (in forest only) North/South Island robin (*Petroica longipes/australis*), bellbird (*Anthornis melanura*), and to a lesser extent tūī (*Prosthemadera novaeseelandiae*) and rifleman (*Acanthisitta chloris*). Bellbird and rifleman had a higher probability of being detected by ARDs, whilst the remainder were more likely to be detected by observers.

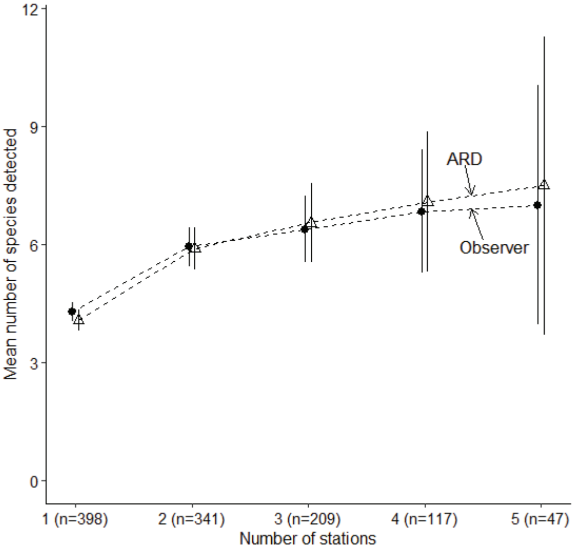

Figure 4. Mean cumulative number of species detected (±95% confidence intervals) by observers and ARDs with each successive bird count station surveyed/processed.

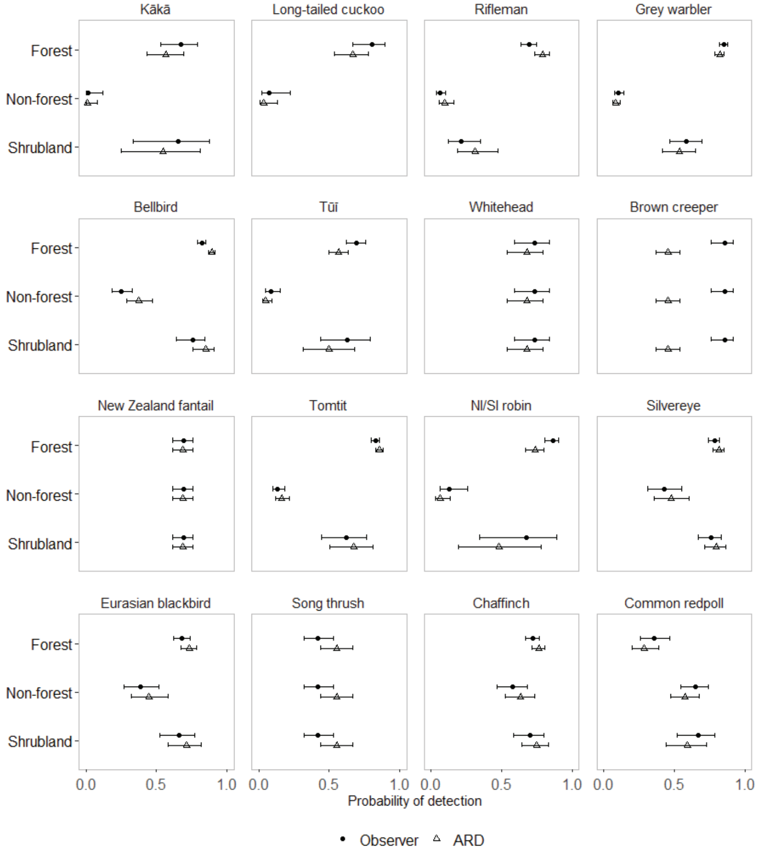

Figure 5. Probability of species detection (±95% confidence intervals), estimated using a single-season multi-method occupancy model (MacKenzie *et al.* 2006), for the observer and ARD survey methods within each habitat type.

Table 4. Occupancy analysis model selected for each species (Ψ = psi; Θ = theta; *p* = probability of detection; h = habitat type (forest, non-forest, shrubland); m = method (observer, ARD); w = wind (0–3); n = noise (0–2). Odds ratios provide an indication of which method was more effective at detection (1 = both methods equally effective; <1 = observer method more effective; >1 = ARD method more effective).

Species	Sampling locations	Model	Odds ratio
Kākā (*Nestor meridionalis*)	398	Ψ(.), Θ (.),*p*(m+h+n)	0.637
Long-tailed cuckoo (*Eudynamys taitensis*)	398	Ψ(.), Θ (.),*p*(m+h)	0.492
Rifleman (*Acanthisitta chloris*)	398	Ψ(.), Θ (.),*p*(m+h+w)	1.670
Grey warbler (*Gerygone igata*)	398	Ψ(.), Θ (.),*p*(m+h+w)	0.819
Bellbird (*Anthornis melanura*)	398	Ψ(.), Θ (.),*p*(m+h+w+n)	1.820
Tūī (*Prosthemadera novaeseelandiae*)	398	Ψ(.), Θ (.),*p*(m+h+n)	0.585
Whitehead (*Mohoua albicilla*)	93	Ψ(.), Θ (.),*p*(m+n)	0.778
Brown creeper (*Mohoua novaeseelandiae*)	305	Ψ(.), Θ (.),*p*(m+w)	0.140
New Zealand fantail (*Rhipidura fuliginosa*)	398	Ψ(.), Θ (.),*p*(m+w)	0.981
Tomtit (*Petroica macrocephala*)	398	Ψ(.), Θ (.),*p*(m+h+w+n)	1.268
North/South Island robin (*Petroica longipes/australis*)	398	Ψ(.), Θ (.),*p*(m+h+w)	0.446
Silvereye (*Zosterops lateralis*)	398	Ψ(.), Θ (.),*p*(m+h+n)	1.227
Eurasian blackbird (*Turdus merula*)	398	Ψ(.), Θ (.),*p*(m+h+w+n)	1.297
Song thrush (*Turdus philomelos*)	398	Ψ(.), Θ (.),*p*(m+n)	1.717
Chaffinch (*Fringilla coelebs*)	398	Ψ(.), Θ (.),*p*(m+h)	1.266
Common redpoll (*Carduelis flammea*)	398	Ψ(.), Θ (.),*p*(m+h)	0.731

Indices of abundance

Comparison of indices of abundance from observer and ARD methods (for the top 16 species) revealed varied strengths of relationship between indices for different species, and the nature of the relationship often changed for different index values (i.e. did not produce a straight line; Fig. 6). These loess curves also indicated differences between habitat types for many species, although in some cases this may be due to small sample sizes for non-forest and shrubland habitats (Table 5). Results of GLMM analysis also revealed inconsistent relationships between API and observer indices for different species, with the effect size often varying in different habitat types (Table 6). In forest, the effect size ranged from 0.018 (kākā; *Nestor meridionalis*) to 0.303 (rifleman); in non-forest it ranged from -0.262 (rifleman) to 0.339 (chaffinch); and for shrubland the range was -0.070 (whitehead) to 0.571 (rifleman). Although for some species the effect sizes were similar for different habitat types (e.g. kākā: 0.018 for forest, 0.019 for shrubland), for most species they were quite variable (e.g. rifleman: 0.303 for forest, -0.262 for non-forest, 0.571 for shrubland). Effects of wind and noise on indices were variable and not consistent between species (Appendix 4).

Method efficiency

The observer-based count method was clearly more time-efficient than the ARD method, mainly due to the longer processing time requirements of the latter (Table 7). It was estimated that approximately 30 minutes was required to process each ARD 5-minute period. In a scenario where 10 bird counts were completed for a sampling location (2 rounds of 5 counts), an estimated 24.5% additional time would be required for the ARD method, compared to observers. In a scenario where only 5 bird counts were complete (1 round of counts), this increased to an estimated 92.5% additional time required for the ARD method.

The number of skilled ornithologists required to carry out the field observations was much higher than that required to process the ARD recordings. A total of 55 field observers were employed to carry out observer counts over the three survey seasons, completing a mean of 20.22 (± 2.78 SE) counts per person. It is worth noting that 44% of these completed less than 10 counts each (less than 1% of the total). ARD processing employed 13 people, who processed a mean of 150.54 (± 22.70 SE) 5-minute periods per person. When split by survey season a similar pattern was observed, with the ARD method requiring fewer people, each processing a larger proportion of 5-minute periods (Table 8).

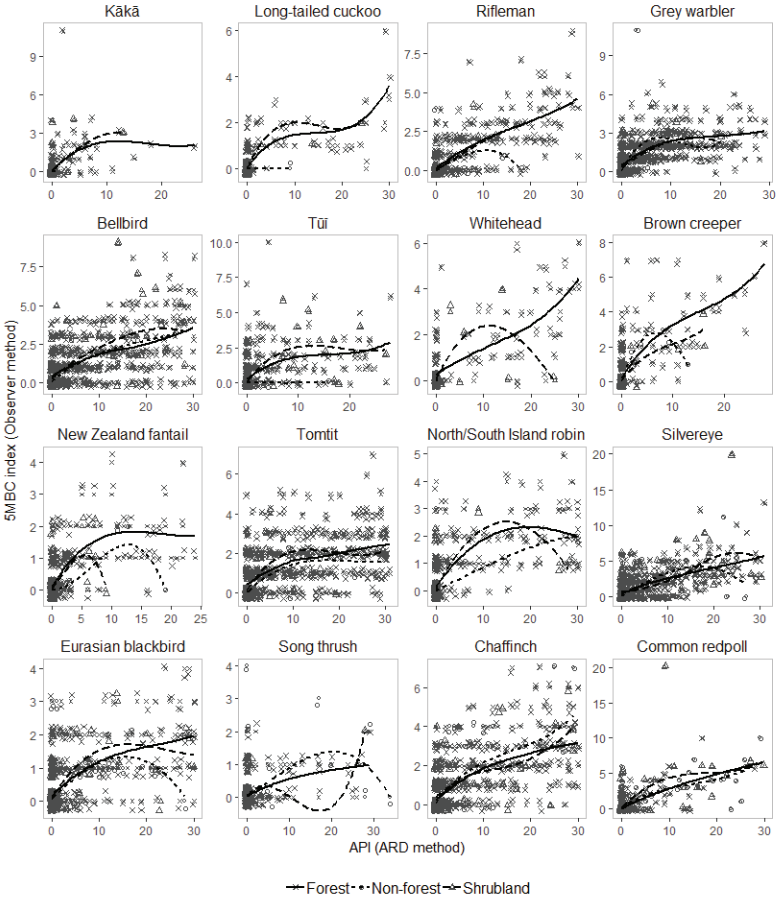

Figure 6. Comparison of observer index of abundance and Acoustic Prevalence Index (API) for the 16 most frequently-occurring species. Loess curves show the relationship between indices from the two methods for each habitat type. Individual data points have been displayed using jittering to make those with the same values visible.

Table 5. The number of 5-minute count periods in which each species was detected by at least one method (observer or ARD), for each habitat type.

Species	Forest	Non-forest	Shrubland	All habitats
Kākā (*Nestor meridionalis*)	65	1	9	75
Long-tailed cuckoo (*Eudynamys taitensis*)	63	3	4	70
Rifleman (*Acanthisitta chloris*)	270	13	14	297
Grey warbler (*Gerygone igata*)	513	38	56	607
Bellbird (*Anthornis melanura*)	503	62	53	618
Tūī (*Prosthemadera novaeseelandiae*)	223	13	23	259
Whitehead (*Mohoua albicilla*)	73	0	7	80
Brown creeper (*Mohoua novaeseelandiae*)	115	15	11	141
New Zealand fantail (*Rhipidura fuliginosa*)	181	5	15	201
Tomtit (*Petroica macrocephala*)	523	36	41	600
North/South Island robin (*Petroica longipes/australis*)	181	7	9	197
Silvereye (*Zosterops lateralis*)	399	71	69	539
Eurasian blackbird (*Turdus merula*)	257	51	37	345
Song thrush (*Turdus philomelos*)	71	26	11	108
Chaffinch (*Fringilla coelebs*)	366	85	56	507
Common redpoll (*Carduelis flammea*)	74	79	36	189

Table 6. Results of GLMM analysis to test for relationships between API (ARD) and observer 5-minute bird count (OBS) indices of abundance, in the three habitat types (hab). API values were normalized by subtracting the mean, and then dividing by the standard deviation. Estimates presented are for effects of ARD on OBS in each habitat type (each fitted with its own unique intercept and slope). Wind (w) or noise (n) were included where these improved model fit and sampling location (sl) was included as a random effect. For grey warbler, kākā, long-tailed cuckoo, and tūī, non-forest data were excluded due to small samples sizes causing problems with the models. There were no non-forest data for whitehead.

Species	**Model definition**	**Habitat type**	**Estimate**	**Standard error**
Kākā (*Nestor meridionalis*)	OBS ~ hab / ARD + w + n + (sl)	Forest	0.018	0.034
		Shrubland	0.019	0.095
Long-tailed cuckoo (*Eudynamys taitensis*)	OBS ~ hab / ARD + w + (sl)	Forest	0.076	0.031
		Shrubland	0.096	0.148
Rifleman (*Acanthisitta chloris*)	OBS ~ hab / ARD + n + (sl)	Forest	0.303	0.030
		Non-forest	-0.262	0.352
		Shrubland	0.571	0.233
Grey warbler (*Gerygone igata*)	OBS ~ hab / ARD + n + (sl)	Forest	0.216	0.025
		Shrubland	0.253	0.109
Bellbird (*Anthornis melanura*)	OBS ~ hab / ARD + w + (sl)	Forest	0.281	0.035
		Non-forest	0.332	0.142
		Shrubland	0.177	0.096
Tūī (*Prosthemadera novaeseelandiae*)	OBS ~ hab / ARD + n + (sl)	Forest	0.041	0.033
		Shrubland	0.099	0.083
Whitehead (*Mohoua albicilla*)	OBS ~ hab / ARD + w + (sl)	Forest	0.141	0.033
		Shrubland	-0.070	0.179
Brown creeper (*Mohoua novaeseelandiae*)	OBS ~ hab / ARD + n + (sl)	Forest	0.135	0.020
		Non-forest	0.005	0.128
		Shrubland	0.156	0.092
New Zealand fantail (*Rhipidura fuliginosa*)	OBS ~ hab / ARD + n + (sl)	Forest	0.106	0.031
		Non-forest	0.173	0.246
		Shrubland	-0.071	0.262
Tomtit (*Petroica macrocephala*)	OBS ~ hab / ARD + n + (sl)	Forest	0.255	0.032
		Non-forest	0.078	0.169
		Shrubland	0.100	0.154
North / South Island robin (*Petroica longipes/australis*)	OBS ~ hab / ARD + n + (sl)	Forest	0.144	0.032
		Non-forest	0.050	0.159
		Shrubland	-0.075	0.148
Silvereye (*Zosterops lateralis*)	OBS ~ hab / ARD + w + (sl)	Forest	0.267	0.031
		Non-forest	0.205	0.080
		Shrubland	0.294	0.063
Eurasian blackbird (*Turdus merula*)	OBS ~ hab / ARD + n + (sl)	Forest	0.183	0.036
		Non-forest	-0.011	0.130
		Shrubland	0.156	0.102
Song thrush (*Turdus philomelos*)	OBS ~ hab / ARD + n + (sl)	Forest	0.024	0.064
		Non-forest	-0.007	0.073
		Shrubland	0.094	0.092
Chaffinch (*Fringilla coelebs*)	OBS ~ hab / ARD + n + (sl)	Forest	0.290	0.034
		Non-forest	0.339	0.078
		Shrubland	0.228	0.087
Common redpoll (*Carduelis flammea*)	OBS ~ hab / ARD + w + (sl)	Forest	0.223	0.059
		Non-forest	0.184	0.038
		Shrubland	0.173	0.047

Table 7. Comparison of estimated time required per sampling location for observer and ARD methods, for two scenarios: (1) 10 bird counts completed; (2) 5 bird counts completed.

Bird counts completed	Method	Estimated time required (minutes)		
		Field	Data processing	Total
10	ARD	360	300	660
10	Observer	440	90	530
5	ARD	360	150	510
5	Observer	220	45	265

Table 8. Number of people employed and effort per person (mean number/percentage of counts completed) for observer (OBS) and ARD methods.

	2011–12		2012–13		2013–14		All seasons	
	ARD	OBS	ARD	OBS	ARD	OBS	ARD	OBS
No. of people	6	20	9	26	11	28	13	55
Mean no. counts/person	74.7	11.6	47.2	12.7	98.5	19.7	150.5	20.2
Mean % counts/person	16.7	2.2	11.1	4.0	9.1	3.7	7.7	1.9

DISCUSSION

Detection of species-richness

The loess model curves indicate that in general, as species-richness recorded by the ARD method increased, so did species-richness recorded by the observer method (or vice-versa). Clearly there is much variation, with the ARD method recording more species at some sampling locations and the observer method recording more species at others; however, the overall relationship appears to be reasonably consistent between habitat types. The number of non-forest sampling locations at which the observer method detected species and the ARD method did not, suggests that the observer method may have an advantage in this habitat type. Results from GLMM analysis indicated some differences between species-richness estimates for the two methods in different habitat types, although effect sizes were small. For all habitat types, the observer tended to record higher species-richness. The effect, though small, was highest for non-forest habitats. These results conflict with findings from some previous studies, which revealed that either there were no significant differences in ability to detect species-richness (Haselmayer & Quinn 2000; Celis-Murillo *et al*. 2009; Celis-Murillo *et al*. 2012) or that ARDs detected more species (Wimmer *et al*. 2013). Hutto & Stutzman (2009), however, found that observers detected a higher number of species, which was more consistent with our results. The differing conclusions of these studies may be influenced by variations in sample design, survey methods, location, environments sampled and/or species present; however, some considerations may be generally applicable. One of the main advantages of ARDs is that they produce a permanent record which can be reviewed multiple times (Haselmayer & Quinn 2000), whereas a field observer has only one chance to identify and record all species. At a location with high species-richness, this may give the ARD method an advantage. Haselmayer & Quinn (2000) found that data from audio recordings detected more species than field observers for sites with high species-richness, which was explained by having the ability to listen repeatedly to the ARD recordings, whereas observers can be over-whelmed in a field situation. An obvious advantage of the observer method, however, is the ability to detect species visually. In particular, this would potentially enable the observer to detect more species in open habitats (such as non-forest) or where there are species present that vocalise infrequently.

In this study, it could be that the visual detection advantage of observers has enabled them to detect higher numbers of species, especially in non-forest habitats which are generally more open environments. Perhaps the ability to visually detect species outweighed the ARD advantage of being able to repeatedly listen to recordings. In a study by Celis-Murillo *et al*. (2012), some species were detected more often by observers in pasture and coastal scrub, where greater visibility was thought to improve the likelihood of visual detection. In forest, however, where detections of these species were mostly auditory, the ARD method was at least as effective as the observer method. In this study, the observer advantage of being able to detect species visually may have been reduced at sampling locations in forest and shrubland habitats.

The cumulative mean number of species detected did not significantly differ between methods, suggesting that in this respect the two methods were equally effective and that the same number of stations (and therefore counts) were required to achieve similar species-richness detection. The cumulative mean number of species detected increased significantly for both methods, as the number of bird count stations increased, further reinforcing the importance of completing multiple bird counts at a location to maximise detection probability (MacKenzie & Royle 2005). The upward slope of the species accumulation curve between 4 and 5 bird count stations (Fig. 4) suggests that 5 stations may not be sufficient to detect all species present, and that to achieve this, further stations (bird counts) might be required. However, consideration must be given as to whether additional effort at the sampling location would justify the cost (Southwood & Henderson 2000; MacLeod *et al*. 2012).

Detection of individual species

Occupancy modelling revealed that habitat influenced detection probabilities for most species and wind and noise were often important. Most species included in this analysis were generally considered forest species, and therefore were more likely to be present in forest or shrubland habitats; not surprisingly probability of detection was highest in these habitats. For chaffinch, the difference was less pronounced, which may reflect its more generalist habitat requirements. It is perhaps less clear why habitat did not influence detection probabilities sufficiently to be included in the models selected for brown creeper, New Zealand fantail, and whitehead, as these species would normally occur in forest and shrubland rather than non-forest habitats. It is not surprising that wind and noise affected detection probabilities; previous studies have demonstrated that noise can have a negative effect (Simons *et al*. 2007; Pacifici *et al*. 2008), whilst wind can have the additional effect of influencing bird behaviour thereby reducing call frequency (O'Connor & Hicks 1980). As shown in Appendix 3, however, increasing wind and noise did not always appear to have a corresponding negative effect on detection probability. For the maximum wind value of 3, detection probability was higher when compared to wind value 2 for some species (e.g. bellbird and brown creeper). Similarly, a wind value of 0 had an apparent negative effect on detection probability for grey warbler (*Gerygone igata*), whereas a wind value of 1 had a positive effect. Some exceptions to the general pattern were also noted for noise, e.g. for Eurasian blackbird (*Turdus merula*) and kākā. The reasons for this are not obvious; however, smaller sample sizes for higher wind and noise values potentially resulting in unreliable results could partly explain these findings. It is also worth noting that the conditions during which 5-minute bird counts were conducted were, to a certain degree, self-censoring. That is, the observer was more likely to abandon the count in strong winds or when noise levels were high. This would also apply to ARDs, as recordings with high levels of noise would most likely be excluded from processing, hence the relatively small sample sizes for high wind and noise values.

Species detection probabilities, when compared between methods, were in most cases similar irrespective of habitat type. There were, however, a few instances (mostly in forest) where detection probabilities were different between the two methods, namely: bellbird, brown creeper, North Island/South Island robin and, to a lesser extent, tūī and rifleman (Fig. 5). Detection probability for bellbird was higher for the ARD method, whereas conversely for tūī it was higher for the observer method. There is no obvious explanation for these differences, unless it is related to species identification error, as both methods rely on manual identification of calls by observers or processors to generate data. There were 26 5-minute periods in which the ARD method detected bellbird only and the observer method detected tūī only, and a further 9 5-minute periods in which the ARD method detected tūī only and the observer method detected bellbird only. This suggests some degree of identification confusion between these two species, which could explain apparent differences in detection probability. Mortimer & Greene (2017) have also demonstrated that bellbird and tūī were frequently confused when the same ARD recordings were processed by two independent processors, due to these two species having similar-sounding calls. The apparent difference in detection probability for rifleman (in forest) is less obvious. One possibility, however, is that their high frequency calls (around 7–12 kHz; Mortimer 2013) can be missed by field observers, whereas the use of sonograms when processing ARD recordings could give a visual clue to the presence of this species.

Identification issues could potentially partly explain the difference in detection probabilities for brown creeper, as their calls can sound similar to, and therefore have the potential to be confused with yellowhead (mohua; *Mohoua ochrocephala*; Falla *et al*. 1966; Higgins & Peter 2002). It is unlikely, however, that this was a major cause for differences in detection probabilities, as Mortimer *et al.* (2019) established that confusion with mohua or other species was only occasional, and mohua has a much more restricted geographic range compared to brown creeper (Robertson *et al.* 2007). The potential for misidentification does, however, emphasise the need to quantify error rates so that these can be accounted for in analyses (Mortimer & Greene 2017).

Another possible explanation is that a large proportion of brown creepers were detected visually by observers, and consequently not detected by ARDs. ARD processors were unlikely to fail to detect brown creepers in large single-species flocks, as they have a tendency to keep in almost constant vocal contact with one another (Dean 1990). When in smaller groups, however, they can often be silent for many minutes (Henderson 1977; Cunningham 1985). Brown creepers also commonly occur in mixed species flocks consisting of silvereye (*Zosterops lateralis*), New Zealand fantail, grey warbler, parakeets (*Cyanoramphus* spp.), chaffinch and/or common redpoll (Henderson 1977; Dean 1990; Heather & Robertson 2000; Higgins & Peter 2002). On occasions when multiple species are calling simultaneously, it may be difficult to reliably pick out brown creeper calls. Under these circumstances, the visual advantage of the observer may enable them to identify brown creepers more often than ARD processors.

The ability of observers to detect birds visually is the most likely explanation for the higher detection probability of North/South Island robin, for the observer method. Vocalisations of this species are generally loud and distinctive and therefore we would expect both methods to be effective at detecting when birds are calling. However, the robin's habit of foraging close to the observer, in a quiet and unobtrusive manner (Higgins & Peter 2002), may result in many visual-only detections from observers, missed by the ARD method. This could be tested by re-analysing the data, including only aural records from field observers. However, since the mode of detection (aural or visual) was not recorded for the Tier 1 Monitoring Programme bird counts, this was not possible.

Another possible explanation for differences in species detection is that distance and/or frequency ranges vary between observers and ARDs. There will be some within-method variability, as both methods are affected by observer/processor hearing ability, which will vary between individuals and is affected by age and gender (Pearson *et al*. 1995). In some instances (e.g. rifleman), lack of agreement may be influenced by an inability to hear bird calls at higher frequencies. It is possible, however, that detection also varies between methods, due to limitations of human hearing and ARD technology. Although Celis-Murillo *et al*. (2009) concluded that their observers and ARDs had similar auditory ranges and consequently sampled equal areas, a comparison of detection ranges for Tier 1 Monitoring ARDs and observers has not been undertaken. Pryde & Greene (2016) tested ARD detection range for morepork (*Ninox novaeseelandiae*), but did not compare this to human observers. Never-the-less, we propose that detection ranges are likely to be similar for both methods, because if they were not then we would expect to observe a systematic difference in detection probability for all species. Environmental effects, such as sound attenuation, weather, and noise, could also affect detection (Morton 1975); however, since the two methods relied on human hearing for species detection (and the recordings were not noise-filtered or altered in any way), we would expect these effects to influence both methods and not result in significantly different detection differences.

Indices of abundance

Loess curves and results of GLMM analysis indicated considerable variation in strength and nature of relationship between indices from the two methods, both for different species and different habitat types for the same species. These results suggest, therefore, that ARDs were limited to detecting presence. The ARD index was a measure of call activity, which would be influenced by many factors including weather conditions (Keast 1994), presence of an observer (Gutzwiller *et al*. 1994; McShea & Rappole 1997), habitat, and species density (McShea & Rappole 1997). To our knowledge, there are no detailed studies of factors influencing call activity specifically for New Zealand birds; however, according to Dowding (2012), call rates are influenced by a number of factors, and this would make it extremely difficult to relate these to abundance. From ARD data, we would not be able to ascertain if multiple calls were from different birds or a single individual. In contrast, a field observer can estimate, using direction and distance, an approximation of the actual number of birds calling. In the absence of data for actual numbers of birds present, our study makes the assumption that the observer method produces an index which is representative of species abundance. We recognise, however, that this may not be the case, as such an index is not adjusted for detection probability and can be affected by a wide range of variables including time of year, time of day, habitat, weather, environmental noise, and observer ability (Dawson 1981). Although observer indices of this kind have received considerable criticism from a number of sources (e.g. Farnsworth *et al*. 2002; MacKenzie & Kendall 2002; Rosenstock *et al*. 2002; Buckland 2006), several studies have demonstrated their ability to detect actual changes in species abundance (Murphy & Kelly 2001; Elliott *et al*. 2010; Greene & Pryde 2012). It is beyond the scope of this study to assess the accuracy of unadjusted indices of abundance. However, we believe that despite their limitations, indices of abundance from the observer method can be used to indicate major changes in relative abundance, and therefore a relationship (or lack of) between this and the ARD method index may provide an indication of whether or not the latter also has potential to show population trends.

An alternative approach to estimating species abundance using ARDs could be via use of ARD arrays (Dawson & Efford 2009; Efford *et al*. 2009). This technique estimates density from the spatial pattern of detection, using signal strength to improve precision. This approach shows promise and investigation into potential incorporation of this into DOC's monitoring programme is currently underway.

Method efficiency

Estimates of time taken per sampling location for the Tier 1 Monitoring Programme suggested that the observer method was most efficient, largely due to the additional time required for processing of ARD recordings. ARDs also required some additional field time, since two visits to each bird count station were required (one for deployment and a second for retrieval), whereas observer counts could be completed with a single visit to each station. The disparity in processing times resulted from the different processes involved to produce data in digital format for the two methods. For observer count data, it was simply a matter of manually entering data into a custom database with built-in data validation checks, which could be done relatively quickly. The ARD recordings, however, required careful listening, often multiple times, and recording presence of species in each 10-second block. Essentially the difference was that the majority of the identification work using observers was done whilst in the field, whereas with ARDs it was done upon return to the office. The field worker heard the 5-minute period once only, whilst the ARD processor could listen to the recorded 5-minute period as many times as they felt necessary. The ARD method would gain considerable efficiency if manual processing was replaced by automated call recognition. In 2014, DOC investigated the potential of recurrent neural network (RNN) techniques for automated call recognition of morepork, kiwi (*Apteryx* spp.), and weka (*Gallirallus australis*; Bagnall & Abraham 2014). Unfortunately, this has so far proved to be unreliable for Tier 1 bird data, with too many false positives and false negatives (*unpubl. data*). There has also been much recent research by others in this area, exploring various techniques (e.g. Chou *et al*. 2008; Bardeli *et al*. 2010; Chu & Blumstein 2011; Lopes *et al*. 2011; Towsey *et al*. 2012; Lasseck *et al*. 2018; Priyadarshani *et al*. 2018). DOC is continuing to explore possibilities in this area; however, at present automated call recognition is not sufficiently developed for incorporation into the Tier 1 Monitoring Programme.

Although the observer method appeared to be the most time-efficient for the Tier 1 Monitoring Programme, this may not necessarily apply to other projects as it will depend largely upon study objectives and design. Hobson *et al*. (2002) estimated that for their purposes the use of automated recording devices and associated manual processing would be more cost-effective than employing specialist ornithologists to carry out field surveys, whilst Wimmer *et al*. (2013) found use of field observers to be more time-efficient than ARDs. With this in mind, we recommend that for any monitoring programme in which efficiency is paramount, a pilot study is completed to test how potential methods perform within the proposed sampling design. Choice of method must also consider the data outputs and their ability to address the objectives of the programme (e.g. to determine species presence or estimate abundance). Use of ARDs, whilst potentially less efficient, could provide an alternative solution in circumstances where employing traditional observer-based methods is challenging, such as monitoring of nocturnal birds or across a large spatial scale. Another important consideration is cost of equipment (for example ARDs, processing software, and file storage). In addition to the initial cost, there will also be periodic repair, upgrade and/or replacement costs. These costs were not included in this study; however, they could be critical to the choice of method when finances are limited.

The high proportion of people who completed <10 observer counts reflects the difficulties the Tier 1 Monitoring Programme experienced in sourcing experienced ornithologists in the long-term. This was particularly the case for the first few survey seasons, although more recently it has been less of an issue, with most people employed for a full survey season (*unpubl. data*). Where presence-only data are sufficient, this difference in staff requirements suggests that the ARD method may be desirable in situations where skilled field observers are in short-supply or not available during the field season – an advantage of ARDs noted by other researchers (Hobson *et al*. 2002; Celis-Murillo *et al*. 2009). Further efficiencies could be made for observer counts by using electronic data capture devices, reducing the need for data entry time (van Tamelen 2004). The potential of such tools is currently being investigated for the Tier 1 Monitoring Programme.

Comparison of ARDs and observers, specifically the similar detection probabilities, reduced time-efficiency of ARDs and inability of ARDs to record abundance (and associated trends), resulted in the decision to discontinue processing of diurnal ARD recordings from the 2014–15 survey season onwards. Processing of nocturnal recordings continued, since there was no other source of data for this time period (i.e. no night-time observer counts). Diurnal recordings will continue to be collected, so that should techniques in automatic bird call recognition become sufficiently advanced, these recordings can be processed retrospectively. We must stress that in our study we have compared data from observers to that from a single type of ARD only, and that different results may be obtained from other ARDs.

Conclusions

The results show that for species-richness detection, observers generally recorded more species than ARDs; however, the overall differences were small. Detection probabilities for individual species were similar between methods for most species. Exceptions were probably due to either identification confusion or species behaviour resulting in visual-only detection by observers. The results suggested that ARDs may be limited to the collection of presence data, whereas observer counts could also monitor abundance (via indices or estimates of density). This may change with further technological developments, for example through the use of ARD arrays to calculate density. In this study, observer counts proved to be more time-efficient, mainly due to the relatively long processing time required for ARD recordings. Potential future technological developments in automated species identification could significantly reduce processing times, however. Higher numbers of skilled people were required for observer counts, which may make them less suitable in the absence of a dedicated team and if there is a shortage of appropriately skilled field observers during the survey season.

ACKNOWLEDGEMENTS

Many thanks to all the Tier 1 Monitoring Programme operations staff and in particular those who deployed/collected ARDs and completed the 5-minute bird counts, of which there are too many to name individually. Likewise, thanks go to those people who spent many hours processing ARD recordings: Marion Rhodes, Robyn Blyth, Mitch Bartlett, Miles Burford, Iris Broekema, Sabrina Leucht, Jason van de Wetering, Maddie van de Wetering, Rowan Hindmarsh-Walls, Tracey Dearlove, Ralph Powlesland, Rose Collen, and Kerry Weston. Thanks also to Elaine Wright for reviewing the draft manuscript and providing helpful comments.

LITERATURE CITED

Allen, R.B.; Bellingham, P.J.; Forsyth, D.M.; MacLeod, C.J.; Wright, E.F. 2009. *Implementing an inventory and monitoring programme for the Department of Conservation's Natural Heritage Management System.* Landcare Research contract report: LC0809/154. Wellington, Department of Conservation.

Allen, R.B.; Bellingham, P.J.; Wiser, S.K. 2003. Developing a forest biodiversity monitoring approach for New Zealand. *New Zealand Journal of Ecology 27*: 207–220.

Bagnall, D.; Abraham, E. 2014. *Identifying nocturnal bird calls*. Presentation to Department of Conservation, 28 February 2014. https://speakerdeck.com/edwardabraham/identifying-nocturnal-bird-calls (accessed June 2016).

Bardeli, R.; Wolff, D.; Kurth, F.; Koch, M; Tauchert, K.H.; Frommolt, K.H. 2010. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters 31*: 1524–1534.

Bibby, C.J.; Burgess, N.D.; Hill, D.A.; Mustoe, S.H. 2000. *Bird census techniques*. London, Academic Press.

Bombaci, S.P; Pejchar, L. 2019. Using paired acoustic sampling to enhance population monitoring of New Zealand's forest birds. *New Zealand Journal of Ecology 43*: 1–11.

Brandes, T.S. 2008. Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International 18*: 163–173.

Buckland, S.T. 2006. Point-transect surveys for songbirds: robust methodologies. *The Auk* 123: 345–357.

Burnham, K.P.; Anderson, D.R. 2002. *Model selection and multi-model inference: a practical information-theoretic approach (2nd ed.).* New York, Springer-Verlag.

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*: 64–78.

Celis-Murillo, A.; Deppe, J.L.; Ward, M.P. 2012. Effectiveness and utility of acoustic recordings for surveying tropical birds. *Journal of Field Ornithology 83*: 166–179.

Chou, C.H.; Liu, P.H.; Cai, B. 2008. On the studies of syllable segmentation and improving MFCCs for automatic birdsong recognition. In: *Asia-Pacific Services Computing Conference, 2008*. APSCC '08. IEEE.

Chu, W.; Blumstein, D.T. 2011. Noise robust bird song detection using syllable pattern-based hidden Markov models. *International Conference on Acoustics, Speech and Signal Processing, ICASSP 2011*. Prague, Czech Republic, IEEE.

Cook, A.; Hartley, S. 2018. Efficient samplings of avian acoustic recordings: intermittent subsamples improve estimates of single species prevalence and total species richness. *Avian Conservation and Ecology 13(1)*: 21. doi: 10.5751/ACE-01221-130121

Cunningham, J.B. 1985. *Breeding ecology, social organization and communicatory behaviour of the brown creeper* (Finschia novaeseelandiae). Unpublished PhD thesis, University of Canterbury, New Zealand.

Dawson, D.G. 1981. Counting birds for a relative measure (index) of density. *Studies in Avian Biology 6*: 12–16.

Dawson, D.G.; Bull, P.C. 1975. Counting birds in New Zealand forests. *Notornis 22*: 101–109.

Dawson, D.K.; Efford, M.G. 2009. Bird population density estimated from acoustic signals. *Journal of Applied Ecology 46*: 1201–1209.

Dean, S. 1990. Composition and seasonality of mixed-species flocks of insectivorous birds. *Notornis 37*: 27–36.

Department of Conservation 2015. *Annual report for the year ended 30 June 2015.* Wellington, Department of Conservation.

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*: 675–683.

Dowding, J. 2012. Introduction to bird monitoring. *Inventory and monitoring toolbox: birds.* Wellington, Department of Conservation.

Efford, M.G.; Dawson, D.K.; Borchers, D.L. 2009. Population density estimated from locations of individuals on a passive detector array. *Ecology 90*: 2676–2682.

Elliott, G.P.; Wilson, P.R.; Taylor, R.H.; Beggs, J.R. 2010. Declines in common, widespread native birds in a mature temperate forest. *Biological Conservation 143*: 2119–2126.

Falla, R.A.; Sibson, R.B.; Turbott, E.G.; Talbot-Kelly, C. 1966. *A field guide to the birds of New Zealand and outlying islands.* Auckland, Collins.

Farnsworth, G.L.; Pollock, K.H.; Nichols, J.D.; Simons, T.R.; Hines, J.E.; Sauer, J.R. 2002. A removal model for estimating detection probabilities from point-count surveys. *The Auk 119*: 414–425.

Freebird 2013. *Freebird: fast and easy to use bird call analysis.* http://www.freebird.co.nz (accessed January 2015).

Frick, W.F. 2013. Acoustic monitoring of bats, considerations of options for long-term monitoring. *Therya 4*: 69–78.

Greene, T.C.; Pryde, M.A. 2012. Three population estimation methods compared for a known South Island robin population in Fiordland, New Zealand. *New Zealand Journal of Ecology 36*: 1–13.

Gutzwiller, K.J.; Wiedenmann, R.T.; Clements, K.L.; Anderson, S.H. 1994. Effects of human intrusion on song occurrence and singing consistency in subalpine birds. *The Auk 111*: 28–37.

Haselmayer, J.; Quinn, J.S. 2000. A comparison of point counts and sound recording as bird survey methods in Amazonian southeast Peru. *The Condor 102*: 887–893.

Heather, B.D.; Robertson, H.A. 2000. *The field guide to the birds of New Zealand. Revised edition.* Auckland, Viking/Penguin Books.

Henderson, N.M. 1977. Autumn and winter flocking behaviour of the brown creeper (*Finschia novaeseelandiae*). *Mauri Ora 5*: 75–88.

Higgins, P.J.; Peter, J.M. (eds) 2002. *Handbook of Australian, New Zealand and Antarctic Birds. Volume 6: Pardalotes to shrike-thrushes.* Melbourne, Oxford University Press.

Hines, J.E. 2006. *PRESENCE - software to estimate patch occupancy and related parameters.* USGS-PWRC. www.mbr-pwrc.usgs.gov/software.presence.html (accessed April 2016).

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*: 709–720.

Holmes, S.B.; McIlwrick, K.A.; Venier, L.A. 2014. Using automated sound recording and analysis to detect bird species-at-risk in southwestern Ontario woodlands. *Wildlife Society Bulletin 38*: 591–598.

Hutto, R.L.; Stutzman, R.J. 2009. Humans versus autonomous recording units: a comparison of point-count results. *Journal of Field Ornithology 80*: 387–398.

Keast, A. 1994. Temporal vocalization patterns in members of a eucalypt forest bird community: the effects of weather on song production. *Emu 94*: 172–180.

Lasseck, M.; Frommolt, K.H.; Koch, M. 2018. Acoustic pattern recognition and its application on long-term bird monitoring. *2018 Ecoacoustics Congress, June 24–28, Brisbane, Australia.* International Society of Ecoacoustics, Queensland University of Technology/ Griffith University, Brisbane, p. 26.

Leach, E.C.; Burwell, C.J.; Ashton, L.A.; Jones, D.N.; Kitching, R.L. 2016. Comparison of point counts and automated acoustic monitoring: detecting birds in a rainforest biodiversity survey. *Emu 116*: 305–309.

Lee, W.; McGlone, M.; Wright, E. 2005. *Biodiversity inventory and monitoring: a review of national and international systems and a proposed framework for future biodiversity monitoring by the Department of Conservation.* Unpublished report. Lincoln, Landcare Research Ltd.

Lopes, M.T.; Gioppo, L.L.; Higushi, T.T.; Kaestner, C.A.; Silla Jr, C.N.; Koerich, A.L. 2011. Automatic bird species identification for large number of species. *International Symposium on Multimedia, ISM 2011*. Dana Point, California, USA, IEEE.

MacKenzie, D.I.; Kendall, W.L. 2002. How should detection probability be incorporated into estimates of relative abundance? *Ecology 83*: 2387–2393.

MacKenzie, D.I.; Nichols, J.D.; Royle, J.A.; Pollock, K.H.; Bailey, L.L.; Hines, A.E. 2018. *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Burlington, Elsevier.

MacKenzie, D.I.; Royle, J.A. 2005. Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology 42*: 1105–1114.

MacLeod, C.J.; Greene, T.C.; MacKenzie, D.I.; Allen, R.B. 2012. Monitoring widespread and common bird species on New Zealand’s conservation lands: a pilot study. *New Zealand Journal of Ecology 36*: 1–12.

McShea, W.J.; Rappole, J.H. 1997. Variable song rates in three species of passerines and implications for estimating bird populations. *Journal of Field Ornithology 68*: 367–375.

Mortimer, J.A.J. 2013. Alternative methods for detection of the South Island rifleman (*Acanthisitta chloris chloris*). *Notornis 60*: 319–321.

Mortimer, J.A.J.; Greene, T.C. 2017. Investigating bird call identification uncertainty using data from processed audio recordings. *New Zealand Journal of Ecology 41*: 126–133.

Mortimer, J.A.J.; Greene, T.C.; Mortimer, S.P. 2019. Assessing bird vocalisation identification accuracy using a computer-based quiz. *New Zealand Journal of Zoology 46*: 201–224.

Morton, E.S. 1975. Ecological sources of selection on avian sounds. *American Naturalist 109*: 17–34.

Murphy, D.J.; Kelly, D. 2001. Scarce or distracted? Bellbird (*Anthornis melanura*) foraging and diet in an area of inadequate mistletoe pollination. *New Zealand Journal of Ecology 25*: 69–81.

Nagendra, H.; Lucas, R.; Honrado, J.P.; Jongman, R.H.; Tarantino, C.; Adamo, M.; Mairota, P. 2013. Remote sensing for conservation monitoring: assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators 33*: 45–59.

Nichols, J.D.; Bailey, L.L.; O’Connell, A.F.; Talancy, N.W.; Campbell Grant, E.H.; Gilbert, A.T.; Annand, E.M.; Husband, T.P.; Hines, J.E. 2008. Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology 45*: 1321–1329.

O’Connor, R.J.; Hicks, R.K. 1980. The influence of weather conditions on the detection of birds during Common Birds Census fieldwork. *Bird Study 27*: 137–151.

Pacifici, K.; Simons, T.R.; Pollock, K.H. 2008. Effects of vegetation and background noise on the detection process in auditory avian point-count surveys. *The Auk 125*: 600–607.

Pearson, J.D.; Morrell, C.H.; Gordon□Salant, S.; Brant, L.J.; Metter, E.J.; Klein, L.L.; Fozard, J.L. 1995. Gender differences in a longitudinal study of age-associated hearing loss. *The Journal of the Acoustical Society of America 97*: 1196–1205.

Priyadarshani, N.; Marsland, S.; Castro, I; Pryde, M. 2018. Automated detection of kiwi (*Apteryx* spp.) from continuous field acoustic recordings – New Zealand Tier 1 data analysis. *2018 Ecoacoustics Congress, June 24–28 Brisbane, Australia.* International Society of Ecoacoustics, Queensland University of Technology / Griffith University, Brisbane, p. 36.

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*: 100–107.

R Core Team 2014. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Version 3.1.2. www.R-project.org (accessed January 2016).

Robbins, C.S. 1981. Bird activity levels related to weather. *Studies in Avian Biology 6*: 301–310.

Robertson, C.J.R.; Hyvönen, P.; Fraser, M.J.; Pickard, C.R. 2007. *Atlas of Bird Distribution in New Zealand 1999–2004*. Wellington, The Ornithological Society of New Zealand.

Rosenstock, S.S.; Anderson, D.R.; Giesen, K.M.; Leukering, T.; Carter, M.F. 2002. Landbird counting techniques: current practices and an alternative. *The Auk 119*: 46–53.

Sedlácek, O.; Vokurková, J.; Ferenc, M.; Djomo, E.N.; Albrecht, T.; Horák, D. 2015. A comparison of point counts with a new acoustic sampling method: a case study of a bird community from the montane forests of Mount Cameroon. *Ostrich 86*: 213–220.

Simons, T.R.; Alldredge, M.W.; Pollock, K.H.; Wettroth, J.M. 2007. Experimental analysis of the auditory detection process on avian point counts. *The Auk 124*: 986–999.

Southwood, T.R.E.; Henderson, P.A. 2000. *Ecological methods*. Oxford, Blackwell Science Ltd.

Steer, J. 2010. Bioacoustic monitoring of New Zealand birds. *Notornis 57*: 75–80.

Stewart, P.; Hasenbank, M. 2018. North Island brown kiwi (*Apteryx mantelli*) monitoring at Whenuakite: trend comparison of observer and acoustic recorder collected call counts. *Notornis 65*: 179–186.

Szewczyk, R.; Osterweil, E.; Polastre, J.; Hamilton, M.; Mainwaring, A.; Estrin, D. 2004.

Habitat monitoring with sensor networks. *Communications of the ACM 47*: 34–40.

Thompson, S.; Gruner, I.; Gapare, N. 2003. *New Zealand land cover database version 2. Illustrated guide to target classes.* Wellington, Ministry for the Environment.

Towsey, M.; Planitz, B.; Nantes, A.; Wimmer, J.; Roe, P. 2012. A toolbox for animal call recognition. *Bioacoustics 21*: 107–125.

van Tamelen, P.G. 2004. A comparison of obtaining field data using electronic and written methods. *Fisheries Research 69*: 123–130.

Wimmer, J.; Towsey, M.; Roe, P.; Williamson, I. 2013. Sampling environmental acoustic recordings to determine bird species richness. *Ecological Applications 23*: 1419–1428.

Xie, Y.; Sha, Z.; Yu, M. 2008. Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology 1*: 9–23.

Zuur, A.F.; Hilbe, J.; Ieno, E.N. 2013. *A beginner's guide to GLM and GLMM with R: a frequentist and Bayesian perspective for ecologists.* Newburgh, Highland Statistics.

Appendix 1. Environmental variables recorded during the observer 5-minute bird counts.

Variable	Score	Definition
Temperature	1	<0°C
	2	0–5°C
	3	6–10°C
	4	11–15°C
	5	16–22°C
	6	>22°C
Sun	0–5	Approximate number of minutes sun overhead
Precipitation (type)	M	Mist
	R	Rain
	H	Hail
	S	Snow
Precipitation (value)	0	None
	1	Dripping foliage
	2	Drizzle
	3	Light
	4	Moderate
	5	Heavy
Wind	0	Leaves still or move without noise
	1	Leaves rustle
	2	Leaves and branchlets in constant motion
	3	Branches or trees sway
Noise	0	Not important
	1	Moderate
	2	Loud

Appendix 2. Land Cover Database (LCDB) classifications (Thompson *et al*. 2003) included in each habitat type assigned to bird count stations, used in species occupancy and GLMM analyses.

Habitat type	LCDB classifications included
Forest	Broadleaved Indigenous Hardwoods; Deciduous Hardwoods; Exotic Forest; Indigenous Forest
Non-forest	Alpine Grass/Herbfield; Depleted Grassland; Fernland; Flaxland; Gravel or Rock; Herbaceous Freshwater Vegetation; Herbaceous Saline Vegetation; High Producing Exotic Grassland; Lake or Pond; Landslide; Low Producing Grassland; Permanent Snow and Ice; River; Sand or Gravel; Short-rotation Cropland; Tall Tussock Grassland
Shrubland	Gorse and/or Broom; Manuka and/or Kanuka; Matagouri or Grey Scrub; Mixed Exotic Shrubland; Sub Alpine Shrubland

Appendix 3. Effects of wind and noise on detection probabilities (untransformed estimates of coefficients); *n* = the number of 5-minute count periods with each wind or noise value; NA = not applicable (i.e. the covariate was not included in the model).

Species	**Survey covariate estimate ± SE$_{mean}$**						
	Wind				**Noise**		
	0	**1**	**2**	**3**	**0**	**1**	**2**
	n = 401	**n = 572**	**n = 83**	**n = 38**	**n = 722**	**n =302**	**n = 60**
Kākā (*Nestor meridionalis*)	NA	NA	NA	NA	-0.4619 ± 0.0004	-1.0008 ± 0.0005	1.4628 ± 0.0006
Long-tailed cuckoo (*Eudynamys taitensis*)	NA	NA	NA	NA	NA	NA	NA
Rifleman (*Acanthisitta chloris*)	0.4758 ± 0.0003	-0.0158 ± 0.0002	0.3530 ± 0.0003	-0.8130 ± 0.0003	NA	NA	NA
Grey warbler (*Gerygone igata*)	-0.2044 ± 0.0002	0.5006 ± 0.0002	0.1107 ± 0.0002	-0.4069 ± 0.0002	NA	NA	NA
Bellbird (*Anthornis melanura*)	0.4206 ± 0.0002	-0.1517 ± 0.0002	-0.9432 ± 0.0003	0.6742 ± 0.0002	0.8434 ± 0.0002	-0.1018 ± 0.0001	-0.7418 ± 0.0002
Tūī (*Prosthemadera novaeseelandiae*)	NA	NA	NA	NA	0.6399 ± 0.0002	-0.1890 ± 0.0002	-0.4509 ± 0.0002
Whitehead (*Mohoua albicilla*)	1.6778 ± NA	1.4974 ± NA	-0.0846 ± NA	0.8415 ± NA	NA	NA	NA
Brown creeper (*Mohoua novaeseelandiae*)	1.0065 ± 0.0004	-0.0468 ± 0.0002	-1.4787 ± 0.0005	0.5189 ± 0.0003	NA	NA	NA
New Zealand fantail (*Rhipidura fuliginosa*)	0.4352 ± 0.0005	0.0254 ± 0.0004	0.5409 ± 0.0005	-1.20013 ± 0.0007	NA	NA	NA
Tomtit (*Petroica macrocephala*)	0.3286 ± 0.0002	0.2301 ± 0.0002	0.2798 ± 0.0002	-0.8387 ± 0.0003	0.3139 ± 0.0002	0.0236 ± 0.0001	-0.3374 ± 0.0002
North/South Island robin (*Petroica longipes/australis*)	0.9766 ± 0.0005	0.3451 ± 0.0004	-0.4244 ± 0.0004	-0.8976 ± 0.0005	NA	NA	NA
Silvereye (*Zosterops lateralis*)	NA	NA	NA	NA	0.2424 ± 0.0002	-0.2035 ± 0.0002	-0.0389 ± 0.0001
Eurasian blackbird (*Turdus merula*)	0.1305 ± 0.0002	0.4580 ± 0.0003	-0.6860 ± 0.0003	0.0972 ± 0.0002	0.4866 ± 0.0002	-0.1446 ± 0.0002	0.3420 ± 0.0002
Song thrush (*Turdus philomelos*)	NA	NA	NA	NA	0.8386 ± 0.0004	-0.2526 ± 0.0003	-0.5860 ± 0.0004
Chaffinch (*Fringilla coelebs*)	NA	NA	NA	NA	NA	NA	NA
Common redpoll (*Carduelis flammea*)	NA	NA	NA	NA	NA	NA	NA

Appendix 4. Effects of wind and noise on observer method index of abundance (untransformed estimates of coefficients); *n* = the number of 5-minute count periods with each wind or noise value; NA = not applicable (i.e. the covariate was not included in the model). There are no estimates for wind or noise value 0, as these were used as the reference categories by the models (REF).

Species	**Survey covariate estimate $\pm SE_{mean}$**						
	Wind				**Noise**		
	0	**1**	**2**	**3**	**0**	**1**	**2**
	n = 401	**n = 572**	**n = 83**	**n = 38**	**n = 722**	**n =302**	**n = 60**
Kākā (*Nestor meridionalis*)	REF	0.1603 ± 0.4023	-0.3158 ± 0.6018	-0.5432 ± 0.5049	REF	-0.2282 ± 0.2847	1.1005 ± 0.2909
Long-tailed cuckoo (*Eudynamys taitensis*)	REF	-0.1779 ± 0.3064	0.7887 ± 0.4124	0.4352 ± 0.4786	NA	NA	NA
Rifleman (*Acanthisitta chloris*)	NA	NA	NA	NA	REF	0.2271 ± 0.1127	0.3087 ± 0.2089
Grey warbler (*Gerygone igata*)	NA	NA	NA	NA	REF	0.0985 ± 0.0728	0.1314 ± 0.1218
Bellbird (*Anthornis melanura*)	REF	-0.1492 ± 0.1072	-0.4282 ± 0.2287	-0.0611 ± 0.2405	NA	NA	NA
Tūī (*Prosthemadera novaeseelandiae*)	NA	NA	NA	NA	REF	-0.1515 ± 0.1534	0.2596 ± 0.2591
Whitehead (*Mohoua albicilla*)	REF	0.2973 ± 0.1975	-0.0398 ± 0.3304	-0.2968 ± 0.6307	NA	NA	NA
Brown creeper (*Mohoua novaeseelandiae*)	NA	NA	NA	NA	REF	0.0044 ± 0.1751	-0.0662 ± 0.2571
New Zealand fantail (*Rhipidura fuliginosa*)	NA	NA	NA	NA	REF	0.2369 ± 0.1584	0.2137 ± 0.3043
Tomtit (*Petroica macrocephala*)	NA	NA	NA	NA	REF	-0.0651 ± 0.0798	-0.1208 ± 0.1614
North/South Island robin (*Petroica longipes/australis*)	NA	NA	NA	NA	REF	0.1589 ± 0.1458	0.3038 ± 0.2282
Silvereye (*Zosterops lateralis*)	REF	-0.1772 ± 0.0972	0.0264 ± 0.1448	-0.0671 ± 0.2483	NA	NA	NA
Eurasian blackbird (*Turdus merula*)	NA	NA	NA	NA	REF	0.2109 ± 0.1202	-0.0399 ± 0.2699
Song thrush (*Turdus philomelos*)	NA	NA	NA	NA	REF	-0.1024 ± 0.3463	-0.6268 ±1.0363
Chaffinch (*Fringilla coelebs*)	NA	NA	NA	NA	REF	0.0424 ± 0.0929	-0.0176 ± 0.1889
Common redpoll (*Carduelis flammea*)	REF	0.0951 ± 0.1707	-0.4263 ± 0.3116	-0.2345 ± 0.4128	NA	NA	NA