*Notornis, 2018, Vol. 65: 179–186*
0029-4470 

# North Island brown kiwi (*Apteryx mantelli*) monitoring at Whenuakite: Trend comparison of observer and acoustic recorder collected call counts

PATRICK STEWART*
Soundcounts, PO Box 21, Coromandel 3506, New Zealand

MARC HASENBANK
Independent Researcher, 46 Borich Road, Sunnyvale, Auckland 0612, New Zealand

**Abstract:** Observer call count surveys are utilised throughout New Zealand to monitor kiwi populations. The development of affordable autonomous acoustic recorders by the Department of Conservation has enabled the collection of large quantities of digital data. Utilising call count data from the North Island brown kiwi (*Apteryx mantelli*) monitoring programme at Whenuakite from the 2010 and 2015 survey periods, a retrospective comparison between data collected by human observers and acoustic recorders was undertaken. Both survey methods indicated an increase in the number of kiwi calls per hour between the 2010 and 2015 surveys. The overall ratio of the number of calls per hour detected by acoustic recorders to those detected by human observers was 1:1.52. Results from the occupancy modelling indicated that the average detection probability for human observers was almost twice as high as that for acoustic recorders. Furthermore, increasing the number of sites for monitoring kiwi populations improved the associated level of precision of the derived occupancy probability estimates. Adjusting the survey design to the underlying characteristics of the kiwi population are therefore important to gain reliable estimates of their population trajectory.

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(4)*: 179–186.

**Key words:** acoustic recorders, call count survey, detection probability, sampling precision, applied monitoring

## INTRODUCTION

The use of acoustic recorders in biodiversity surveys has created new opportunities for long-term studies of bird populations. Acoustic recorders have been shown to reduce observer bias (Rosenstock *et al*. 2002; Hutto & Stutzman 2009), and to avoid disturbance effects often associated with the presence of human observers (Alldredge *et al*. 2007). In addition, acoustic recorders are a cost-effective sampling method that can be deployed in difficult to access regions over long periods of time (Hutto & Stutzman 2009; Steer 2010). The latter also has the benefit of reducing sampling time-related bias by allowing data collection over a wider range of time periods (Diefenbach *et al*. 2007). However, acoustic recorders may not be as sensitive as human observers over larger distances (Hutto & Stutzman 2009). Surveys of North Island brown kiwi (*Apteryx mantelli*) and southern brown kiwi (*Apteryx australis*) populations using acoustic recorders have respectively been undertaken within Tongariro Forest Park in the Central North Island (Guillotel *et al*. 2015) and Sinbad Gully in Fiordland (Loe & Smart 2016). However, apart from a study on little spotted kiwi (*Apteryx owenii*) call counts by Digby *et al*. (2013), there are no published data regarding the outcome of human observer versus acoustic recorder efforts for monitoring kiwi populations. In their study, Digby *et al*. (2013) found that acoustic

*Received 13 March 2018; accepted 6 July 2018*
*Correspondence: *paddy@soundcounts.com*

recorders could detect a similar proportion of the total number of calls compared to human observers (up to 80% compared to 94% respectively).

Our study utilised call count data from the Whenuakite North Island brown kiwi monitoring programme to compare the underlying linear trend in the number of calls detected in the data collected either by human observers or by acoustic recorders. In addition, the respective derived probabilities for a site being identified as being occupied by kiwi by either data collection method are compared. Furthermore, this study aims to provide guidance on selecting an appropriate number of sites and repeat surveys for kiwi monitoring based on different simulation scenarios.

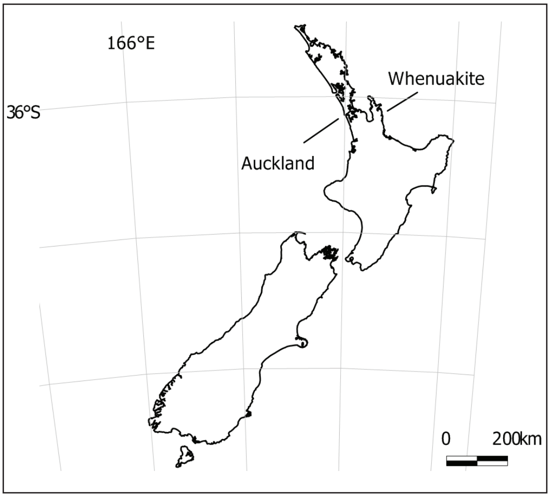


**Figure 1.** Map showing location of the Whenuakite study site. Graticule grid lines represent 4° latitude/longitude intervals.

## MATERIALS & METHODS

### Study area and field methods

The study area is situated within the Tairua Ecological District between Tairua and Whitianga (Fig. 1), on the east coast of the Coromandel Peninsula (36°56′ S, 175°50′ E), New Zealand. Historic land use practices have led to much of the indigenous vegetation becoming modified. A diverse range of secondary forest and induced scrublands cover steep hillsides. There are also some areas of farmed pasture where slope angles are gentler (Kessels *et al*. 2010). Remnant broadleaved associations on the coast grade to regenerating conifer forest on the landward side of the dividing coastal ridge, which reaches 311 metres a.s.l. Elements of primary lowland forest remain in the more inaccessible areas. As part of the 2,700 ha Whenuakite Kiwi Care project, distribution surveys have counted the number of calling kiwi from 24 permanently marked sites on 4 occasions between 2001 and 2015. These surveys have shown that kiwi densities increased fourfold between 2001 and 2015, with adult birds distributed evenly throughout the treatment area (Stewart *et al*. 2015).

This indirect comparison study was conducted retrospectively with the regular kiwi monitoring programme at Whenuakite during the 2010 and 2015 survey periods. Calls from 5 of the survey sites were chosen based on previous years presence of kiwi within the area, and to mirror population densities at sites where human observers were located. This allowed for a sufficiently high number of calls to be available for recording during the survey periods that were comparable between data collection methods. All survey sites chosen were located within indigenous forest habitat and at least 1 km apart. Autonomous acoustic recorders (ARs) developed by the Department of Conservation were used during this study (version B 2). ARs were deployed at the same sites as observers; however, they were programmed to operate at times when observers were absent to avoid creating an overlap between human observers and ARs. ARs were positioned on small tree branches approximately 1.5 m above ground level. Recordings were made in mono and were digitised at 16 kHz, 16-bit precision.

Human observers collected call count data following methods outlined by Robertson & Colbourne (2003) over 3 nights from the periods 5–19 May 2010, and from 17 April to 28 May 2015. ARs were deployed for 20 consecutive nights from 14 June to 3 July 2010, and 18 consecutive nights from 16 April to 3 May 2015. Kiwi call count studies by Colbourne & Digby (2016) suggest that a small amount of variation in call rates can be attributed to differences in kiwi call activity during the different sampling months mentioned above. As these sampling months fell within the breeding season for brown kiwi, call activity was generally higher than during off-breeding season (Colbourne & Digby 2016). However, most of the variation is likely to stem from nightly fluctuations in call rates (Colbourne & Digby 2016). No human observer surveys were conducted during extreme weather, such as heavy rain or strong wind, or during full-moon periods. While extreme weather may affect a human observer's ability to detect kiwi call counts, kiwi are also known to be less likely to call during times of full moon (Colbourne & Digby 2016). In contrast, ARs continued data collection throughout their deployment period regardless of environmental conditions; hereby only the first 10 nights of readable data from each AR were subsequently inspected for kiwi calls using Raven

Pro 1.5© (Bioacoustics Research Program 2014) with a 512-sample Hann window and 15.6 Hz resolution. Limiting the data to 10 nights of readable recordings allowed for an equal number of good quality recordings to be used for each survey period. Call count data collection started 45 minutes after sunset, and human observers collected call count data for 1 hour during each survey night (3 hours/site/year). For the purpose of comparing the 2 data collection methods statistically, only AR data for the first hour post survey start (sunset plus 45 minutes) were used in the subsequent call count analysis (10 hours/site/year).

All data collected by human observers and ARs were completed during the same breeding season of the respective years.

**Data analysis and simulation**

Linear mixed-effects models (LMEs) were fitted to the call count data to test whether the average call count for human observer and AR data changed at a comparable magnitude between the 2 survey years. Inferences were made using an information theoretic approach, where multiple models were compared based on their relative AICc weights (Akaike 1974; Burnham & Anderson 2002). In contrast to the regular AIC value, the AICc value is corrected for sample size and the number of model parameters, therefore providing a better measure of model fit when comparing several similar models (Burnham & Anderson 2002). AICc weights are a relative measure used to compare similar models, whereby higher AICc weights indicate what combination of variables, of those tested, is better suited to explain the observed call count pattern (Burnham & Anderson 2002). The multi-model inference compared 5 different models (Table 1), all of which included the number of calls per hour, based on the first hour of the nightly sampling period, as response variable. In addition, all five models included a random effect for the survey year, which accounted for the repeated measures structure of the data. In comparison, model 1 included an interaction term between the sampling year and data collection method, as well as separate fixed effects for sampling site and month of year. The interaction term was included to compare the regression slopes predicted by the model for the change in call counts between survey years and data collection methods. Model 2 included all fixed effects of model 1 apart from month of year, while model 3 included the month of year parameter but not the fixed effect for sampling site. Furthermore, model 4 only included the interaction term between survey year and data collection method. In addition, a fifth intercept-only model was included as part of the multi-model inference.

The multi-model inference indicated model 2 to be the statistically best supported model. Based on that result, multiple comparisons between the different levels of the site fixed effect were performed using Tukey contrasts. To determine whether an effect was statistically significant, a bootstrap 95% confidence interval was computed for a given model parameter. An overlap of the 95% confidence interval with zero indicated that the model fit predicted the parameter to be statistically non-significant.

In a second step, the nightly call counts were converted into binary detection data (detection, non-detection) and fitted to a static, single-season occupancy model. This occupancy model was used to predict separate detection probabilities for both ARs and human observers for each survey season respectively. The occupancy model was fitted using Bayesian modelling approach as described by Kéry & Royle (2016). For this purpose, the sampled population of kiwi was assumed to be closed for the duration of the sampling period, which allowed for the simultaneous estimation of detection and occupancy probabilities for the respective survey periods and sampling methods (MacKenzie *et al*. 2002). The Bayesian modelling used (1) a state model to describe occupancy (z) at a particular site (i), and (2) an observation model to describe observations (y) made at a given site (i) and sampling night (j) based on the occupancy state at that particular site:

$z_i \sim \text{Bernoulli}(\psi)$

$y_{ij} \mid z_i \sim \text{Bernoulli}(z_i p)$

Both, the occupancy probability ($\psi$) and the probability of observation (p) were modelled as Bernoulli distributions using uninformative priors for the purpose of this study.

To provide some perspective on the number of sampling sites and number of survey nights required to achieve reliable estimates of occupancy probability, the above Bayesian modelling approach was used to simulate data assuming an average occupancy probability ($\psi$) of 0.8, and a set of different detection probabilities (p): 0.1, 0.5, and 0.9. Simulations were run for 5, 20, and 100 sampling sites, each scenario running over 5, 20, 50 survey nights and 1,000 iterations, respectively. Based on the estimated occupancy probabilities the root mean squared error (RMSE) was calculated for each scenario. The lower the RMSEs the more similar the estimated occupancy probabilities were to the actual occupancy probability used in a particular scenario. Hereby, a RMSE of 0.1 or below was taken as a threshold for adequate estimated precision. This level of precision is generally regarded as adequate in the current related literature for estimating occupancy probabilities (MacKenzie & Royle 2005; Guillera-Arroita *et al*. 2010). Notable here is that all simulations were run assuming nightly call counts to only being performed during 1 hour. The predictions should therefore be taken with some caution when comparing them to methodology that uses data collected over several hours per night, or where kiwi population density and structure vary from those at Whenuakite.

All statistical data analysis and simulation was conducted in R 3.4.0 (R Core Team 2017) using the following packages and their dependencies: lme4 (Bates *et al*. 2015), merTools (Knowles & Frederick 2016) and boot (Canty & Ripley 2017) for fitting and summarizing linear mixed effects models, plyr and ggplot2 (Wickham 2009, 2011) for data summary and visualization. The R package jagsUI (Kellner 2016) was used together with JAGS 4.2.0 (Plummer 2003) to conduct the Bayesian analysis of the occupancy model.

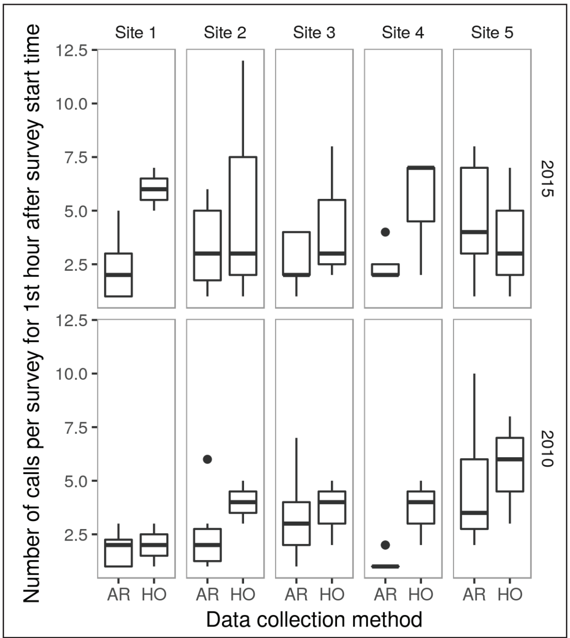


**Figure 2.** Box and whisker plot for the first-hour call counts from acoustic recorder (AR) and human observer (HO) collected data for the 5 different sites during the 2010 and 2015 call count surveys. Lower and upper limit of box indicating the range between 1st and 3rd quartiles, with centred bold line indicating the median of the data; Lower and upper whiskers indicating the minimum and maximum respectively, with points beyond whiskers indicating potential outliers (data points that lie beyond the ± 1.5 Inter Quartile Range).

## RESULTS

Kiwi calls were detected by human observers and ARs at all sites during both survey seasons (Fig. 2). The number of calls per hour detected by human observers was generally higher than that detected by ARs and subsequent analysis at the same site during the first hour of recording (Fig. 2). The overall ratio of the number of calls per hour detected by ARs (during first hour of recording) to that detected by human observers was 1:1.52.

The multi-model inference indicated that the combination of parameters fitted to model 2 performed best in explaining the observed call count pattern, followed closely by model 1 (Table 2). The fixed effects common to both models were the interaction term between sampling year and data collection method, as well as the site the data were collected. In addition, model 1 also included the month of year parameter; however, as indicated by the slightly lower AICc weight, the month of year parameter did not add significantly to explaining the observed call count pattern. Rather, the extra number of model parameters in model 1 compared to model 2 meant that the AICc weight was comparably lower. Furthermore, none of the parameter combinations included in models 3, 4 or 5 provided a good explanation of the observed call count pattern. From this it is possible to infer that the main variation in call counts is due to variations between sites, rather than between different survey months.

A more detailed examination of the model 2 predictions supported the finding that human observers generally detected a higher number of calls during the first hour post survey start (sunset + 45 minutes) than ARs did at a given monitoring site (beta-estimate: 1.88, 95% CI: 0.67, 3.02). Furthermore, the model predicted overall differences in call counts between sites, particularly between (i) site 5 and site 1, and (ii) site 5 and site 4 (Tukey Contrasts: (i) beta-estimate: 1.95, 95% CIs: 0.43, 3.46; (ii) beta-estimate: 2.11, 95% CIs: 0.50, 3.71). Neither the survey year, nor the interaction term between survey year and data collection method were predicted to be statistically significant (year: beta-estimate: -0.31, 95% CIs: -1.13, 0.55; year:type: beta-estimate: -0.82, 95% CIs: -2.46, 0.76).

The occupancy model indicated that the average detection probability for human observers was about twice that of ARs (Table 3). Predictions for the detection probabilities for ARs and human observers stayed constant between the 2010 and 2015 survey seasons. In contrast, the model estimated the same probability of occupancy for both datasets. Similarly, high estimates for occupancy probability for the 2 different data collection methods may be due to the aforementioned ubiquitous spread of calls across all sites during both survey seasons.

For the simulated scenarios, the RMSE decreased with increasing number of sites and survey nights (Fig. 3). In these simulations, the number of sites had the greatest effect on RMSE, while the increasing number of survey nights produced a smaller decrease in RMSE. These results are highly dependent on the underlying detection probability. Scenarios with a low detection probability (0.1) required a higher number of survey nights to reduce their associated RMSE than when the detection probability was high (0.9).

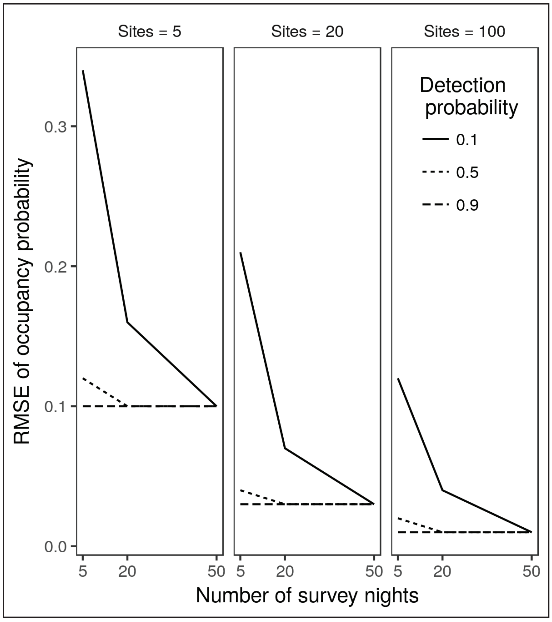


**Figure 3.** Comparison between number of sampling nights and the associated root mean square error (RMSE) estimates for occupancy probability, in relation to different number of survey sites. The estimates for RMSE gained from simulation are based on an occupancy probability of 0.8, and a detection probability of 0.1, 0.5 and 0.9. An RMSE of 0.1 or below is generally regarded as an adequate level of precision in the current literature (MacKenzie & Royle 2005; Guillera-Arroita *et al.* 2010).

**Table 1.** Models used as part of the multi-model inference. The intercept only is denoted with ~1 as fixed effect.

| Model | Response | Fixed effects | Random effects (slope) |
|---|---|---|---|
| 1 | Call count | Year * Type + Site + Month | Year |
| 2 | Call count | Year * Type + Site | Year |
| 3 | Call count | Year * Type + Month | Year |
| 4 | Call count | Year * Type | Year |
| 5 | Call count | ~ 1 | Year |

**Table 2.** Multi-model inference: * denotes an interaction term, while K refers to the number of model parameters, and AICc to the sample size corrected AIC value.

| Model | Fixed effects | Fixed effects | K | neg log -likelihood | AICc | Delta AICc | AICc weight |
|---|---|---|---|---|---|---|---|
| 1 | Year * Type + Site + Month | Year * Type + Site + Month | 11 | -243.63 | 511.78 | 0.60 | 0.42 |
| 2 | Year * Type + Site | Year * Type + Site | 10 | -244.55 | 511.18 | 0.00 | 0.56 |
| 3 | Year * Type + Month | Year * Type + Month | 7 | -25212 | 519.27 | 8.10 | 0.01 |
| 4 | Year * Type | Year * Type | 6 | -252.72 | 518.21 | 7.03 | 0.02 |
| 5 | ~ 1 | ~ 1 | 3 | -258.65 | 523.51 | 12.33 | 0.00 |

**Table 3.** Results from occupancy model for detection (p) and occupancy (ψ) probabilities for different sampling methods and survey years, respectively. Estimates are based on a set of 5 sampling sites. Acoustic recorder estimates based on data collected during first hour of recording only.

| Method | Year | Survey nights per site | Detection Probability (p) Mean | P Lower 95% CI | P Upper 95% CI | Occupancy probability (ψ) Mean | psi Lower 95% CI | psi Upper 95% CI |
|---|---|---|---|---|---|---|---|---|
| Acoustic recorder | 2010 | 10 | 0.42 | 0.29 | 0.56 | 0.86 | 0.54 | >0.99 |
| | 2015 | 10 | 0.48 | 0.35 | 0.62 | 0.86 | 0.54 | >0.99 |
| Observer | 2010 | 3 | 0.94 | 0.79 | >0.99 | 0.86 | 0.54 | >0.99 |
| | 2015 | 3 | 0.94 | 0.80 | >0.99 | 0.86 | 0.54 | >0.99 |

## DISCUSSION

While the call count data collected by human observers and ARs followed a similar pattern, human observers routinely recorded a higher number of calls during the first hour post survey start (sunset + 45 minutes) than ARs (an exception was site 5 in 2015 where AR median calls/hour was higher compared to observer median calls/hour, Fig. 2). Also this study accounted for shifts in call rates during different periods of the night by filtering both observer and AR data to only include those calls recorded during the first hour post survey start time, with no adjustment for potential differences in detection range for human observers or ARs being made. Digby *et al*. (2013) determined that simulated little spotted kiwi calls of both sexes were reliably detected from spectrogram inspection to at least 400 m, while human observers have been found to detect kiwi calls well beyond that distance. By potentially missing more distant kiwi calls, the ARs may have failed to identify the presence of kiwi at certain sites during the 2 sampling periods. In contrast, Stewart & Hasenbank (2012), and Zwart *et al*. (2014) demonstrated that ARs can provide similar results to, or even outperform human observers in detecting bird calls under certain circumstances. While no comparative study was available on the performance of different AR models used by this, or Zwart *et al*.'s study, the different outcomes in terms of sampling method may indicate that the effective sampling range of ARs depends on a variety of factors. These may include spectral analysis techniques, the sensitivity of the microphone and hardware used, the species monitored, the background noise level at the time of sampling, or the presence of acoustic barriers, such as tree trunks or steep hillsides, between the AR and the calling individual (Digby *et al*. 2013; Pryde & Greene 2016).

Furthermore, following the results from the multi-model inference the differences in month during which the different surveys were conducted did not affect the call count in a statistically significant manner. As no information on individual observers was available as part of this study, the effects of an individual observer's experience and ability to detect kiwi calls could not be tested. However, differences between individual observer ability to detect kiwi calls are likely to have contributed to the overall variance present in observer call counts. While observers in this study were experienced in detecting kiwi calls, and thus reducing the possibility of false positives, variation in their performance may stem from differences in their ability to detect faint calls, or to distinguish between individuals when multiple kiwi call at the same time. Spectral analysis of the sound files may also be affected by observer variation. While no observer bias during spectral analysis was assessed, this potential issue was minimised by having the same experienced observer utilising the same software settings for both years' data analysis. No measure of identifying false negatives for either data collected in the field or subsequent analysis was available as part of this study.

While some variation in the number of calls recorded per hour was found between sites during both survey seasons, kiwi were found to be present at all sites most of the time. This translated into a relatively high predicted occupancy probability of over 0.8, with a moderate to high estimate for the detection probability for human observers (0.94), and a low to moderate estimate for detection probability for ARs (0.42 and 0.48 for the 2010 and 2015 survey seasons respectively). In regards to the per site sampling effort, the number of repeated samplings met the suggested minimum requirements for human observers and ARs proposed by MacKenzie & Royle (2005): a site with a probability of occupancy of 0.8 should be sampled at least 2 times when the probability of detection is 0.9 or greater, or at least 4 times when detection probability is 0.5 or greater. In terms of the number of sampling sites, Guillera-Arroita & Lahoz-Monfort (2012) found that with decreasing detection probability the number of sampling sites required to gather reliable information on site occupancy increases. Likewise, in scenarios with rare or cryptic species, the lower occupancy and detection probabilities may make it necessary to increase the number of sampling sites to achieve the same estimator quality (MacKenzie & Royle 2005; Guillera-Arroita *et al*. 2010). Therefore, an increased survey effort may be required when surveying relict populations across large landscapes.

Based on the simulations conducted as part of this study, the predicted occupancy probabilities for data collected by human observers and ARs in the field were close to, or slightly above, the RMSE threshold of 0.1 for what is considered adequate sampling precision of the underlying site occupancy in the relevant literature (MacKenzie & Royle 2005; Guillera-Arroita *et al*. 2010). Looking at ways to improve the sampling precision in this applied setting, based on predictions from simulated scenarios, increasing the number of survey sites (e.g. from 5 to 20 sites for ARs) would improve the RMSE below the 0.1 threshold for both human observer and AR collected data (refer to Fig. 3). Increasing the number of sampling nights, however, would provide only a small gain in sampling precision of the underlying site occupancy for either human observers or ARs in this applied setting. It is important to note that this interpretation is based on a scenario that only uses data collected for 1 hour post sunset per night, as data collection over additional hours may yield higher nightly call counts that may increase the overall probability of detection and predicted occupancy probability.

Developing a survey design that takes into account the characteristics of the to-be-surveyed kiwi population is therefore important. Factors to consider at the design stage should include: expected distribution of population across landscape, the number of ARs required to cover a certain area, selection of sites offering similar sampling coverage, as well as the spacing between recorders. The latter is important to prevent double counting of calls by different ARs (pseudo-replication), and where subsequent analysis of call counts does not allow for filtering of replicate recordings. Furthermore, the survey design should also evaluate the number of hours of recording during each sampling night, as well as the number of sampling nights required to adequately estimate the probability of occupancy for certain sites.

## ACKNOWLEDGEMENTS

Observer surveys were funded by the Whenuakite Kiwi Care Group. Field data collected by: Carol Nanning, Hamish Kendal, Wendy Hare, Murray Blake, Nigel Milius, Tony Harrison, Peter Novis, Sally Armstrong, and Patrick Stewart. Thanks to landowners and other Whenuakite locals who helped make the survey a success over the years: The Hinds Family of Oneroa, Merata Mita, Roger Elliot from Tapuaetahi Reserves Ltd, Malcolm McFarlane, Derek Boyd, Wallace Family, Lewis Price, K J Clark, K M Schelde-Jense, and Kate Purdon. Stu Cockburn kindly loaned Department of Conservation autonomous recorders in 2010. Thanks to Dr Hilary Miller for proofreading, Derek Christie for comments, and reviewers Andrew Digby and Lee Shapiro with editorial help on a draft of the manuscript.

## LITERATURE CITED

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*: 716–723.

Alldredge, M.W.; Pollock, K.; Simons, T.; Collazo, J.; Shriner, S.; Johnson, D. 2007. Time-of-detection method for estimating abundance from point-count surveys. *The Auk 124*: 653–664.

Bates, D.; Mächler, M.; Bolker, B.; Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*: 1–48.

Burnham, K.P.; Anderson, D.R. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. New York, Springer.

Bioacoustics Research Program 2014. *Raven Pro: Interactive sound analysis software (Version 1.5).* New York, The Cornell Lab of Ornithology.

Canty, A.; Ripley, B.D. 2017. *Boot: Bootstrap R (S-Plus) functions*. R package version 1.3-19.

Colbourne, R.; Digby, A. 2016. Call rate behaviour of brown kiwi (*Apteryx mantelli*) and great spotted kiwi (*A. haastii*) in relation to temporal and environmental parameters. *DOC Research & Development Series No. 348.* Wellington, Department of Conservation.

Diefenbach, D.; Marshall, M.; Mattice, J. 2007. Incorporating availability for detection in estimates of bird abundance. *The Auk 124*: 96–106.

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.

Guillera-Arroita, G.; Lahoz-Monfort, J.J. 2012. Designing studies to detect differences in species occupancy: power analysis under imperfect detection. *Methods in Ecology and Evolution 3*: 860–869.

Guillera-Arroita, G.; Ridout, M.S.; Morgan, B.J. 2010. Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution 1*: 131–139.

Guillotel, J.; Potae, R.; Hayward, J.; Scrimgeour, J. 2015. *Tongariro Forest Kiwi Sanctuary annual report July 2014 – June 2015*. Tongariro District Office, Department of Conservation.

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.

Kellner, K. 2016. *jagsUI: A wrapper around "Rjags" to streamline "JAGS" analyses*. R package version 1.4.4.

Kéry, M.; Royle, J.A. 2016. *Applied hierarchical modeling in ecology - Analysis of distribution, abundances and species richness in R and BUGS. Volume 1.* London, Academic Press.

Kessels, G., Deichmann, B., Kendal, H., Stewart, P., Clark, R., Robb, M., & Hermans, A. 2010: *Significant natural areas of the Thames-Coromandel district: Terrestrial and wetland ecosystems*. Environment Waikato Technical Report 2010/36. Environment Waikato, Hamilton.

Knowles, J.E.; Frederick, C. 2016. *merTools: Tools for analyzing mixed effect regression models*. R package version 0.3.0.

Loe, E.; Smart, A. 2016. *The Sinbad Sanctuary project - Sinbad Gully, Milford Sound - 2015/16 annual report*. Fiordland District Office, Department of Conservation.

MacKenzie, D.I.; Nichols, J.D.; Lachman, G.B.; Droege, S.; Royle, J.A.; Langtimm, C.A. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology 83*: 2248–2255.

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

Plummer, M. 2003. *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*. Proceedings of DSC International Workshop, Vienna.

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. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robertson, H.; Colbourne, R. 2003. *Kiwi (*Apteryx *spp.) best practice manual*. Wellington, Department of Conservation.

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

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

Stewart, P.; Hasenbank, M.; Hare, W.; Armstrong, S.; Harrison, T. 2015. *The response of kiwi to predator control and advocacy, Whenuakite 2001–2015*. Whenuakite Kiwi Care Group, Whitianga.

Stewart, P.; Hasenbank, M. 2012. *Moehau Kiwi Sanctuary: Call count survey and analysis 2002–2012*. Thames, Department of Conservation.

Wickham, H. 2009. *ggplot2: Elegant graphics for data analysis*. New York, Springer-Verlag.

Wickham, H. 2011. The split-apply-combine strategy for data analysis. *Journal of Statistical Software 40*: 1–29.

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*: e102770. https://doi.org/10.1371/journal.pone.0102770