

VOLUME 13, ISSUE 1, ARTICLE 21

Cook, A., and S. Hartley. 2018. Efficient sampling of avian acoustic recordings: intermittent subsamples improve estimates of single species prevalence and total species richness. *Avian Conservation and Ecology* 13(1):21. https://doi.org/10.5751/ACE-01221-130121 Copyright © 2018 by the author(s). Published here under license by the Resilience Alliance.

Methodology

Efficient sampling of avian acoustic recordings: intermittent subsamples improve estimates of single species prevalence and total species richness

Asher Cook¹ and Stephen Hartley¹

¹Centre for Biodiversity and Restoration Ecology, Victoria University of Wellington

ABSTRACT. Automated sound recording devices have become an important monitoring tool used to estimate species richness and abundance of birds in a variety of ecological and conservation studies. The prevalence of calls detected in a specific time period can be used as an index of relative abundance, to compare between populations. However, the statistical power to infer true differences in abundance between populations is low when detections are highly aggregated in time leading to high variance between samples from the same population. Here, we used two different sampling methods, and used the data from each to calculate species richness and acoustic prevalence of nine bird taxa from a total of 50 sound recordings. The first method simulated typical monitoring techniques used by observers in the field by using a continuous five-minute section of the recording. The second method used the first 10 seconds of each minute to create a composite recording, also of five minutes total duration. There was no difference in the mean prevalence index between methods. The intermittent samples, however, produced prevalence indices with a lower standard deviation (mean difference = 19 %), detected 26% more species per five-minute sample and required 60% less total listening time to detect as many species as the continuous method. The only cost of subsampling from a long recording is the extra digital memory and battery life required to obtain the recordings in the first place. Given that these costs are minor, the intermittent method holds much promise because it detects species more efficiently and provides greater power to detect differences in a species' relative abundance, which in turn should allow for better-informed management regarding population status and trends.

Échantillonnage efficace d'enregistrements d'oiseaux : le sous-échantillonnage intermittent améliore les estimations d'occurrence pour une espèce et de la richesse spécifique totale

RÉSUMÉ. Les appareils d'enregistrement automatiques sont devenus des outils de suivi importants pour estimer la richessse spécifique et le nombre d'oiseaux dans le cadre de diverses recherches en écologie et conservation. La fréquence des chants détectés dans une période spécifique peut servir d'indice d'abondance relative, afin de comparer des populations. Toutefois, la puissance statistique à révéler les vraies différences d'abondance entre des populations est faible lorsque les détections sont très rapprochées dans le temps, ce qui engendre une grande variance entre les échantillons d'une même population. Dans la présente étude, nous avons utilisé deux méthodes différentes d'échantillonnage et les données provenant de celles-ci pour calculer la richesse spécifique et l'occurrence de neuf taxons d'oiseaux à partir de 50 enregistrements. La première méthode simulait la technique habituellement utilisée par les observateurs sur le terrain, soit l'écoute d'une section de 5 minutes continues de l'enregistrement. La seconde méthode reposait sur les 10 premières secondes de chaque minute afin de créer un enregistrement composite, également d'une durée de 5 minutes au total. Il n'y avait pas de différence de l'indice moyen d'occurrence entre les deux méthodes. Toutefois, les échantillons intermittents ont produit des indices d'occurrence avec un écart type plus faible (différence moyenne = 19 %), ont détecté 26 % plus d'espèces par 5 minutes d'échantillon et ont requis 60 % moins de temps d'écoute pour détecter le même nombre d'espèces qu'avec la méthode continue. Le seul coût du souséchantillonnage d'un long enregistrement est le besoin plus grand de mémoire numérique et une plus grande longévité des piles pour obtenir l'enregistrement au départ. Comme ces coûts sont mineurs, la méthode intermittente est prometteuse puisqu'elle permet de détecter les espèces plus efficacement et est plus puissante pour détecter les différences d'abondance relative d'une espèce, ce qui en retour permet une gestion plus éclairée de l'état des populations et des tendances.

Key Words: autonomous recording units; avian; bioacoustics; detection rates; efficiency; experimental design; monitoring; statistical power; temporal autocorrelation

INTRODUCTION

Monitoring changes in biological populations is fundamental to a wide range of conservation and environmental management activities (Heywood 1995, Lindenmayer et al. 2012). Effective monitoring allows population trends to be identified and, in turn, appropriate management may be implemented to prevent species' decline and/or extinction (Lindenmayer and Likens 2009). Birds are a particularly popular and important taxon in many monitoring programs because they are sensitive and highly apparent indicators of environmental condition (Temple and Wiens 1989, Bibby 1999, Gregory and van Strien 2010).

Address of Correspondent: Stephen Hartley, School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington, New Zealand, 6140, stephen.hartley@vuw.ac.nz

Most bird monitoring is undertaken in the field using techniques such as point and transect counts (Ralph et al. 1995, Bibby 1999), distance sampling (Buckland et al. 2005) and capture-markrecapture (White and Burnham 1999). However, in recent years electronic bioacoustic methods, which utilize automated sound recording devices, have become an increasingly popular technique (Acevedo and Villanueva-Rivera 2006, Bardeli et al. 2010). Birds are especially appropriate for bioacoustic monitoring because they frequently communicate via conspicuous species-specific vocalizations.

Given the relative ease of collecting recordings, bioacoustic monitoring projects often end up with overwhelmingly large data sets (Rempel et al. 2005). Automated identification and scoring of species from these recordings can struggle with acoustically complex field recordings, resulting in high false-positive and falsenegative rates (Swiston and Mennill 2009, Venier et al. 2017). Manual listening by trained operators can generate accurate results, however considerable time and effort is required to process large sets of recordings. For this reason, if recordings are to be scored manually, it is beneficial to sample efficiently from the large amount of data potentially available.

Bioacoustic bird surveys typically make use of measures of vocal activity or call rates, which are presumed to be monotonically related to the underlying abundance of the species being sampled (Borker et al. 2014). Call rates (or vocal abundance) can be repeatedly measured across time and space to detect changes in population status. However, bird calls are generally highly contagious in time and space, i.e., periods of high vocal activity are often clumped together, followed by long periods of low vocal activity, meaning that estimates of mean call rates and related indices are highly variable and thus associated with high levels of uncertainty (Thompson 2002). The statistical power of ensuing models will tend to be low, making it difficult to detect differences and trends with high confidence. Given this scenario, it is important to devise sampling protocols that reduce variability in call rate indices, especially when attempting to answer management questions that rely on accurate estimates of population status (Nichols and Williams 2006).

Another common aim of bioacoustics monitoring is to estimate the species richness of avian communities (Wimmer et al. 2013, Towsey et al. 2014). Here too, efficient means of detecting less common species are needed (Wimmer et al. 2013).

Traditional monitoring techniques used by observers in the field are generally constrained to short continuous sampling periods, such as 5–10 minutes for point counts (Ralph et al. 1995, Hartley 2012). Monitoring projects obtaining data from automated sound recording devices have often used analogous sampling timeframes (Haselmayer and Quinn 2000), however this may be underexploiting the value of bioacoustics recordings because automated devices are not subject to the same logistical limitations and constraints as point counts involving in situ human observers.

Sound recorders are typically left in one location, and may be programmed to collect several hours of sound data daily over a period of weeks or months (Thompson et al. 2017). The data may subsequently be split into a large number of smaller temporal sample units. Having taken this step, it should be expected that there will be some degree of positive temporal autocorrelation in the data, meaning that two sample periods recorded close to each other in time are less independent than two periods recorded further apart in time (Bence 1995). Recordings taken on different days are likely to be more independent of each other than successive recordings made on the same day. Because birds move across landscapes at a range of spatial scales, two 10-second samples recorded 10 minutes apart are more likely to provide different information than two 10-second samples recorded backto-back.

In this study, we collected 30 minutes of acoustic recordings from forest sites in the lower North Island of New Zealand. We listened to the recordings using two different subsampling strategies to generate estimates of (i) species-level acoustic prevalence and (ii) community-level species richness. The first method mimicked approaches used by observers in the field by noting the prevalence of calls from a continuous five-minute sample of the original recording. The second method used the first 10 seconds of every minute to create a five-minute sample of 30 discontinuous snippets. Our aim was to examine the effect of temporal sample structure on the data and to empirically determine whether intermittent sampling reduced the variability of prevalence estimates and/or increased the number of species detected relative to a continuous five-minute sample.

We predicted that there would be no difference in mean call rates between methods, and that indices of prevalence generated using the intermittent sampling method would be less autocorrelated and therefore less variable, i.e., with lower standard deviation, than those generated by the continuous method. We also predicted that the intermittent sample would accumulate detections of new species more rapidly than the continuous method because it samples from a larger extent of time, albeit at the same total intensity as the continuous method.

METHODS

Study sites

Acoustic recordings were taken daily from 24 automated recording devices active between June 2014 and October 2014; the recording devices were spread across 10 forest sites in the lower North Island of New Zealand (Fig. A1.1). Within sites, recording devices were separated by a minimum of 500 m and remained in fixed positions for the duration of the study. The sites covered a range of forest types typical of the lower North Island including southern beech (*Fuscospora* spp. and *Lophozonia* spp.), mixed broadleaf, regenerating manuka-kanuka (*Leptospermum scoparium - Kunzea ericoides*), and mixed podocarp-broadleaf forest.

Acoustic recordings

The automated recording devices used in this study were the commercially available Song MeterTM (Model SM2+, Wildlife Acoustics Inc., Massachusetts, USA). Recordings were made at a rate of 44,100 samples per second (so that bird vocalizations of up to 22.05 kHz could be distinguished), with a single microphone per device. Recorders were mounted on tree trunks approximately 1.5 m above ground level and were programmed to record synchronously for 30 minutes every day from 0800–0830 (approximately 1–2 hours after dawn) for a period of at least 120 days starting in June 2014. Each 30-minute recording was saved as a 16-bit PCM uncompressed .WAV file.

To avoid sound recordings that were heavily affected by background environmental noise (primarily wind and rain) daily climate information was collected from three weather stations that were spread across the study area (Fig. A1.1). Only recordings from days that met the following set of weather criteria were considered for further investigation: 24 hr average wind speed < 10 km/hr, 24 hr max wind gust < 40 km/hr, and 24 hr rainfall total < 2 mm for two out of three of the meteorological stations. A random sample of 50 sound files was then selected from the pool of acoustic recordings that passed the weather criteria. Twenty-two different devices were represented in this random sample.

Protocol for manual listening

From each 30-minute recording we took two samples (each totalling five minutes in duration). The first sampling method, referred to here as the "continuous method," mimics Dawson and Bull's (1975) five-minute bird count, which is New Zealand's standard point-count monitoring technique used by observers in the field, by taking a continuous five-minute sample of the original 30-minute recording. The five minutes were extracted from a random starting point that was generated between the 1st and 25th minute of the original recording. The second method, referred to here as the "intermittent method," uses the first 10 seconds of each and every minute to create a composite five-minute recording (Fig. 1).

Scoring of bird calls was undertaken by manually listening to recordings while simultaneously looking at the associated spectrogram, using Raven 1.4 software (Bioacoustics Research Program 2011). Panasonic OverEar (noise control") RP-HC200 headphones were used to listen to the recording. For both sampling approaches, the presence/absence of bird taxa was scored across a set of thirty 10-second snippets or subsamples (Mortimer and Greene 2017). From these data, the number of taxa recorded per five-minute sample (as well as per 10-second subsample) was also calculated.

A taxon was scored as present within a subsample if its call met the following criteria: (i) it could be heard when the track was played at maximum volume, (ii) it was visually apparent on the spectrogram, and (iii) it could be confidently identified to species level (with two exceptions described below). The characteristic loud wingbeats of the kererū (*Hemiphaga novaeseelandiae*), a largely nonvocal pigeon, were also used to identify the presence of this species.

Some species have similar calls that can be easily confused, particularly when birds are some distance from the recording device. Therefore, the honeyeaters $T\bar{u}\bar{i}$ (*Prosthemadera novaeseelandiae*) and Bellbird (*Anthornis melanura*; Family: Meliphagidae), were scored as an aggregate taxon, as were Blackbird (*Turdus merula*) and Song Thrush (*T. philomelos*; Family: Turidade), in order to streamline the identification process and to minimize false-positive detections (Mortimer and Greene 2017). For the sake of brevity we refer hereafter to analysis of species' call rates, although recognizing that two of the "species" are actually higher order taxa.

Data analysis

We used a binary detection method per 10-second subsample, rather than counting the number of calls, because this overcomes problems associated with determining when one call starts and another finishes. This differs from standard field-based methods, where attempts are made to count the number of individual birds present based on visual and directional cues available to the observer. Our primary variable of interest was the number (or proportion) of 10-second subsamples (out of 30) in which a species was detected. We call this the acoustic prevalence index (API). It differs from many site-occupancy approaches (e.g., MacKenzie et al. 2002) because we are not estimating the probability that a particular site is occupied, but the average prevalence of calls at each site. For widespread, mobile species passing a static recorder, this finer-scale index is expected to be related to relative abundance, possibly via a Poisson relationship (Hartley 1998), and should be a more sensitive indicator of population change than coarsegrained, site-level occupancy. A mean and standard deviation for the prevalence index was calculated for each species, for each method, using the 50 five-minute samples.

Fig. 1. Schematic representation of a hypothetical 30-minute acoustic recording (top row) and the two subsampling methods. Each bar represents a single 10-second section of recording; black bars show intervals where a bird call was detected and white rectangles show where a bird call was not detected. Sample A, using the continuous sampling method, generated an overly high estimate of the underlying call rate because the sample happened to fall within a period of high vocal activity. Sample B sampled a period of low vocal activity, therefore the same method generated a prevalence index of just 0.03 (1/30). The noncontinuous method, Sample C, estimated a prevalence index of 0.20 (6/30) which was closer to the true prevalence from the entire 30 minute recording (0.19 or 34/180).



The autocorrelation structure of the five-minute samples was measured in two ways: first by counting the number of "runs" within each sample. A run is defined as a series of consecutive 10-second subsamples in which a species is detected as present. Separate runs must be separated by at least one 10-second subsample of "absence." A run may consist of a single 10-second presence. Secondly, we calculated the temporal autocorrelation coefficient of each sample for time lags 1–10, using the "acf" function in the R package "stats," and then averaged this across the samples of each method, for each species. Only samples with greater

than one or less than 29 presences (out of 30) were included because recordings with zero or one presence or absence lacked sufficient structure and produced trivial results. Note that lags 1, 2, 3 ... 10 in the intermittent sample are derived from subsamples taken from the 6th, 12th, 18th ... 60th real-time interval because the intermittent sample uses every sixth subsample of the continuous recording. In other words, the interval between lags in the continuous sample is 10 seconds and the interval between lags in the intermittent samples is one minute (Fig. 1).

Comparison of methods

Statistical analyses were undertaken on species that were deemed to have been detected frequently enough to allow for robust analysis. In this case, species detected across $\geq 50\%$ of sites and present in at least 1% of all 10-second subsamples.

Mean values of the acoustic prevalence index were compared between the two methods using a paired Student's t-test, with species as replicates. To determine whether the intermittent method produced less variable index values than the continuous method, we applied a paired Student's t-test on the standard deviations of each species (n = 9). We tested coefficient of variance (CV) values in the same way, as an alternative measure of variability (CV = standard deviation / mean).

To test for differences in the temporal structure of each species' presence and absence recorded by each subsampling protocol the mean number of runs was compared between methods as was the mean autocorrelation coefficient (at lag = 1) using Student's t-test.

We tested whether the total number of species detected per five minute sample differed between methods using a Wilcoxon signed-rank test with continuity correction; a nonparametric test, because the species richness data was non-normal (Skewness = 0.42, Kurtosis = 2.64, Shapiro-Wilk test for normality: W = 0.94, P = 0.0003). Rarefaction (with 100 randomisations) was used to compare differences in species' cumulative detection rates between the two sampling methods, using the program EstimateS v9 (Colwell 2013). All other statistical analyses were conducted using R (R Core Team 2015), and all tests were conducted at $\alpha = 0.05$.

RESULTS

Study species

A total of 25 bird taxa were detected across all recordings (Table A1.1). The following nine taxa were recorded at sufficient frequency to be included in a comparison of subsampling methods: Blackbird/Song Thrush, Chaffinch (*Fringilla coelebs*), Fantail (*Rhipifura fuliginosa*), Grey Warbler (*Greygone igata*), Rifleman (*Acanthisitta chloris*), Silvereye (*Zosterops lateralis*), Tomtit (*Petroica macrocephala*), Tūī/Bellbird, and Whitehead (*Mohoua albicilla*).

Mean and standard deviation of observed call rates

There was no significant difference in the mean acoustic prevalence index measured by the two sampling methods (intermittent = 0.092 vs continuous = 0.102; t_8 = -0.96, P = 0.37; Fig. 2a). The mean standard deviation (SD) of the index, however, was on average 19% lower using the intermittent method

compared to the continuous method (0.150 vs 0.185; $t_8 = -2.54$, P = 0.035; Fig. 2b). The coefficient of variation values were also significantly lower when using the intermittent method (1.62 vs 1.89; $t_8 = -2.87$, P = 0.021). The degree to which the intermittent method reduced standard deviation varied among species; for Silvereye and Tūī /Bellbird SD was reduced by more than 50%, whereas for Grey Warblers the reduction was just 3%, and one species out of nine (Rifleman) showed a small increase (Fig. 2b).

Fig. 2. Comparison of the mean and SD of the acoustic prevalence index obtained from continuous and intermittent subsampling methods for nine bird taxa: (a) mean acoustic prevalence index, (b) standard deviation of the index. NS = nosignificant difference in mean of the prevalence index between methods. * = a significant reduction in standard deviation using the intermittent method (P = 0.035). The continuous method used a continuous five-minute subset of 30 minute recordings, the intermittent method used the first 10 seconds of each minute to create a composite five-minute sample from the original recording (n = 50 recordings). Species codes: a =Blackbird (Turdus merula) / Song Thrush (T. philomelos), b = Chaffinch (Fringilla coelebs), c = Fantail (Rhipifura fuliginosa), d = Grey Warbler (Grevgone igata), e = Rifleman (Acanthisitta *chloris*), f = Silvereye (*Zosterops lateralis* $), <math>g = T\bar{u}\bar{i}$ (Prosthemadera novaeseelandiae) / Bellbird (Anthornis *melanura*), h = Tomtit (*Petroica macrocephala*), and i = Whitehead (Mohoua albicilla). Dot and error bar in margins = interspecific mean \pm SE.



Temporal structure of samples

Seven out of nine bird species showed an increase in the mean number of runs per sample using the intermittent subsampling scheme compared to the continuous method (e.g., Fig 3). In most cases the differences were minor (less than 10%), but for two species the difference was large (20–50% increase) and significant (paired Student's t-test: Chaffinch: $t_{49} = 2.02$, P = 0.049; Silvereye $t_{49} = 3.07$, P = 0.003; Table 1). This indicates that the intermittent method reduced the aggregation of detections, while making virtually no difference to the mean call rate of either species.

Table 1. Comparison of the temporal structure of calls for different bird species sampled according to two different schemes (i) cont. = a continuous five minutes divided into 10-second subsamples and (ii) intermittent = the first 10 seconds of each minute from 30 minutes of recording. In both cases, 30 subsamples are scored for presence/absence per sample (n = 50). Diff. = differences, where, \approx approximately equal (within 10% of the lower value), > continuous method at least 10% greater than intermittent method, < continuous method at least 10% less than the intermittent method. * indicate significant differences where P < 0.05, ** indicate significant differences at P < 0.005. Mean number of runs were compared with a paired Student's t-test, autocorrelation (lag 1) was compared with an unpaired two sample t-test because of an unequal number of trivial (nonempty) samples per method.

		mean no. of runs			mean autocorrelation (r) lag = 1		
Taxon	cont.	diff.	intermittent	cont.	diff.	intermittent	
Blackbird/Song thrush	4.08	*	4.10	0.082	<	0.172	
(Turdus merula)/(T. philomelos)							
Chaffinch	3.50	<	4.24*	0.183	>	0.149	
(Fringilla coelebs)							
Fantail	2.84	≈	2.70	0.115	>	0.094	
(Rhipifura fuliginosa)							
Grey Warbler	4.22	≈	4.46	0.000	>	-0.027	
(Greygone igata)							
Rifleman	1.66	<	2.14	0.048	<	0.113	
(Acanthisitta chloris)							
Silvereye	2.62	<	3.94**	0.273	>	0.068**	
(Zosterops lateralis)							
Tomtit	2.78	≈	2.80	0.193	<	0.393	
(Petroica macrocephala)							
Tūī/Bellbird	6.82	~	6.22	0.028	<	0.155*	
(Prosthemadera novaeseelandiae)/(Anthornis melanura)							
Whitehead	2.88	~	3.00	0.090	<	0.143	
(Mohoua albicilla)							

Species' records of presences and absences at a 10-second resolution were generally weakly and nonsignificantly positively autocorrelated at lag one and then dropped to near zero autocorrelation by lags 2–4 regardless of the subsampling scheme (Fig. A1.2). Nonetheless, comparing between methods there were two cases of significant differences in autocorrelation at lag = 1 (which in the case of the intermittent method is equivalent to lag = 6 of a continuous subsample): Silvereye records became significantly less autocorrelated when collected via the intermittent method (two sample t-test: $t_{30} = 3.00$, P = 0.005) whereas the Tūī/Bellbird taxon became more autocorrelated ($t_{63} = 2.18$, P = 0.033; Table 1).

Species detection

A significantly greater number of species were detected per five minutes of listening using the intermittent method (5.08 ± 0.51) compared to the continuous five-minute method (4.04 \pm 0.44; Wilcoxon signed rank test: V = 32.5, n = 50, P < 0.0001) (Fig. 4a), despite the fact that detection rates at the level of the 10second subsample were slightly higher across the set of 50 continuous samples (0.92 per 10 seconds) compared to the intermittent samples (0.82 detections per 10 seconds). The greater level of within-sample species richness achieved by the intermittent method was also reflected in greater gamma diversity: across the entire set of 50 samples 24 species were detected in total using the intermittent samples compared to 20 using the continuous samples. Rarefaction curves show that, when using the intermittent method, an average of only 20 samples (40% of the total listening time) were required to detect the same number of species as 50 samples from the continuous method (Fig. 4b).

DISCUSSION

Surveys of bird communities remain one of the most widely employed approaches for monitoring biota and assessing the ecological health of ecosystems, with many schemes dating back over half a century (e.g., Sauer et al. 2017). The increasing availability of automated bird recorders provides a number of advantages over traditional field-intensive methods, but brings with it new challenges and opportunities related to the extraction of biologically meaningful data. Here, we found that an intermittent subsampling strategy, e.g., subsampling the recording using the first 10 seconds of every minute, significantly decreased variability in our index of acoustic prevalence, without any significant effect on the value of its mean. Intermittent sampling also led to more rapid detection of a wider variety of species, thus proving it to be a more efficient scheme for assessing the avifauna of an area.

Reduction of variability between samples

The average reduction in standard deviation achieved with intermittent sampling, across all species, was 19%. This translates to a 19% reduction in standard errors around the mean and an increase in statistical power equivalent to increasing a survey's sample size (*n*) by a factor of 1.52, i.e., $(1-0.19)^{-2}$, assuming samples are truly independent of one another. Given call rates (and other similar measures of vocal activity) are often used as indices of abundance (Sepúlveda et al. 2006, Borker et al. 2014, Oppel et al. 2014) greater statistical power will translate into an enhanced ability to detect differences in abundance between two different places or across different years. This is of particular importance for scientists attempting to understand ecological processes and conservation practitioners whose management

Fig. 3. Example of the presence/absence data from two species (a) Silvereye (*Zosterops lateralis*), above and (b) Tomtit (*Petroica macrocephala*) below, collected by the two subsampling methods: continuous samples (left) and intermittent samples (right). Intervals along the x-axis are thirty 10-second subsamples with black shading indicating a recorded presence and white an absence, as illustrated in Figure 1. Fifty files (rows) were analyzed for each species using each method. Most species' samples showed an increase in the number of short run lengths using the intermittent method, although the silvereye was the only species to exhibit a significant decrease in autocorrelation (lag = 1) between continuous and intermittent sampling.



decisions rely on precise and reliable monitoring programs (Nichols and Williams 2006, Guillera-Arroita and Lahoz-Monfort 2012).

There was considerable variation between species in the degree to which the intermittent method reduced variability of the prevalence index; both Silvereye and Tūī/Bellbird scores experienced an approximate 50% reduction in standard deviation, whereas Grey Warbler experienced just a 3% reduction. Silvereye also showed the greatest increase in number of runs when moving from the continuous to the intermittent samples, and had the highest level of temporal autocorrelation in the continuous sample. This suggests that the level to which the intermittent method reduces variability of the prevalence index is likely to depend on the unique calling characteristics and/or movement patterns of the species. Silvereye tend to move around the forest

in small flocks during winter (Heather and Robertson 1996), hence periods of intense and almost continuous "chattering" are punctuated by relatively long periods of absence when the flock has moved on (*personal observation*). This explains why the continuous method is more susceptible to recording both high values (when a flock is present) and low values (if the continuous five-minute recording spans a period of temporary absence). Because the intermittent method subsamples sparsely from a longer time period it is less likely to be influenced by aggregated calling events and thus the prevalence values are less variable and individually closer to the grand mean (Fig. 1).

Even for nonflocking species we expected a considerable level of temporal clustering in calls. Quite often the same individual bird will repeat a series of songs (typically 2–10 seconds each) in quick succession for up to a minute, before an extended period of silence

Fig. 4. Mean number of species detected by continuous and intermittent subsampling of acoustic recordings: (a) mean species richness (\pm SE) of a single five-minute sample, (b) rarefaction curves of mean species richness (\pm 95% CI [dotted lines]) accumulated over 1 to 50 samples. The dashed horizontal line shows the total number of species detected across all 50 counts using the continuous method. The continuous method used continuous five-minute samples randomly selected from 30-minute acoustic recordings (n = 50); the intermittent method used the first 10 seconds of each minute to create a composite five minute sample.



(of several minutes or more) before starting up again (Hartshorne 1956). Additionally, the call of one individual can trigger response calls of other nearby conspecifics (McGregor 1991). Both of these behaviors will tend to increase the clustering of calls in a recording, such that moving from a continuous to an intermittent sample should reduce levels of temporal aggregation. This shift in structure was evident from the increased fragmentation of "runs" in intermittent samples, although somewhat surprisingly it was not reflected by a consistent change in the autocorrelation function. Only one species, Silvereye, showed a significant decrease in measured autocorrelation in the intermittent samples. This is likely due to the continuous samples generally starting off with what was already a fairly low level of autocorrelation (r <0.2), hence the correlation statistic was not a sensitive measure of the change in structure. Other indices that measure gap structure, e.g., lacunarity, or spectral methods, may prove to be more sensitive to changes in sample method, but they were not explored here because they generally require time series of at least 200 sample units to be meaningful (Halley et al. 2004). Instead, we concentrated on how the sampling method affected the empirical measures that most ecologists seek to report: the mean and standard error of an abundance index and the species richness of a community.

Studies in meteorology have confirmed that increasing the interval between subsamples taken from an autocorrelated time series improves parameter estimates of the underlying properties (Ramírez and Carta 2005). Similar results have been found when sampling spatially aggregated populations, in which systematic surveys (with regular gaps between samples) produce more efficient estimates of abundance, with greater precision and lower bias, than random or adaptive sampling schemes (Kimura and Somerton 2006, Mier and Picquelle 2008). In general, it has been noted that "For constant effort, experimental designs that have more, smaller blocks, broadly spread across the experimental area, lead to tests that have more power in the presence of spatial autocorrelation" (Legendre et al. 2004: 3212). Positive autocorrelation in a time series is intimately linked to the concept of an effective sample size (n_{eff}) for a dataset, according to the formula $n_{\text{eff}} = n_{\text{s}}(1-r)/(1+r)$, where n_{s} is the actual number of samples and r is the lag 1 autocorrelation (Bartlett 1935, Santer et al. 2000). In the case of the Tomtit, the effective sample size of the continuous sample (r = 0.273) would reduce to 57% of the actual sample size while for the less autocorrelated intermittent sample (r = 0.068) n_{eff} is 87% of n_{s} . Analysis of song structure and the duration of silences at the macroscale of 10-1000 seconds has rarely been studied, although some early observations are given in Hartshorne (1956) and Nice (1931). Further research in this area, potentially drawing on the literature of Markov Chain Monte Carlo (MCMC) sampling and thinning of autocorrelated series (e.g., Kass et al. 1998, Link and Eaton 2010), could shed light on the optimal grain and intervals for subsampling aimed at detecting particular species.

Increased probability of species detections

The increased efficiency of the intermittent subsampling method was even more dramatic when estimates of species richness were compared. Within a five-minute equivalent sample the intermittent method detected an average of 5.08 species compared to 4.04 from the continuous samples (a 25% increase). Accumulating detections over multiple samples for the intermittent method would, on average, detect 20 species from 20 samples, whereas the continuous method required 50 samples to reach the same species richness: a 2.5-fold increase in listening time. In other words, the extra effort required to detect increasingly rare species was substantially less using the intermittent subsampling method. These increases in sampling efficiency were apparent even though there was no appreciable difference in the mean call rate of common species between methods.

Estimating species richness is an important aspect of biodiversity studies and conservation management (Rahbek and Graves 2001, Gotelli and Colwell 2011). Given that resources for conservation are limited, scientists and conservation managers aim to develop sampling protocols that maximize species detections while minimizing the costs and effort required (Wimmer et al. 2013). In this case, the intermittent subsampling method achieved exactly that. The more efficient nature of intermittent subsampling methods means they will be more cost-effective when undertaking rapid biodiversity inventories (e.g., Sueur et al. 2008) or indeed for any study estimating species richness.

In biodiversity monitoring, failing to detect a species when it is actually present within a targeted habitat, i.e., a false negative, is a common yet problematic issue (MacKenzie et al. 2002); even low rates of false negatives potentially influence ecological inferences (Tyre et al. 2003). Intermittent or diffuse subsampling reduces the likelihood of false negatives through improved rates of species' detections. Thompson et al. (2017) recommend that having more recordings of shorter duration (1–4 minutes) is more efficient than fewer recordings of 10 minutes duration, whereas La and Nudds (2016) noted that a set of 10-minute samples collected from different times of the day recorded a greater species richness than the same number of samples taken consistently at the same time of day. Our method of intermittent subsampling extends these ideas to a much finer temporal resolution.

Furthermore, several studies have found that manual scoring of acoustic data from automated sound recording devices yields higher estimates of species richness than traditional field-based observer methods (Wimmer et al. 2013, La and Nudds 2016).

Costs and benefits of intermittent sampling with digital recorders

One of the advantages of electronic bioacoustic monitoring is that data can be collected for relatively long periods of time, simultaneously across multiple locations, compared to in situ point counts, which generally last only 5–10 minutes before the observer moves to a new station (Ralph et al. 1995), and the data may be archived for future quality assurance. A potential advantage of field-based observers, however, is that they are better able to sample a large number of locations over the course of a few days, i.e., the surveys can have greater spatial replication, whereas a limited number of static digital recorders tend to allow for greater temporal replication (Klingbeil and Willig 2015). As the cost of recorders continues to fall, the financial constraints on sample number will also fall.

One of the costs of intermittent subsampling methods compared to continuous samples of the same length, is that the automated recording devices may need to record for longer periods in order to obtain the sample, resulting in more rapid draining of batteries and saturation of the data storage capacity. This cost can be eliminated, however, if recording devices can be programmed to record in bursts of specified length and frequency, e.g., a duty cycle, rather than continuously. Some recording devices, such as Song MeterTM Model SM2+ have this capability, although others do not.

In conclusion, we found that intermittent subsampling of acoustic recordings significantly reduces the standard deviation of a prevalence index, and increases the probability of species' detection in five minute's worth of listening, thereby dramatically increasing the rate at which the total number of species accumulates over the entire dataset. Given the benefits of increased statistical power and the relatively minor cost of subsampling from long recordings, we propose that when the time available for listening to sound files is limited, intermittent subsampling methods, such as the method illustrated here, should be adopted over continuous subsampling methods.

AUTHOR CONTRIBUTIONS

SH conceived the ideas and designed methodology; AC conducted field work and manual listening; AC led the data analysis; AC and SH shared the writing of the manuscript and production of figures. Both authors contributed critically to the drafts and gave final approval for publication.

Responses to this article can be read online at: http://www.ace-eco.org/issues/responses.php/1221

Acknowledgments:

We thank Adrian Pike for field assistance and TbFree NZ for support towards field expenses. AC was supported by a Victoria University of Wellington scholarship and a grant from the Centre for Biodiversity and Restoration Ecology. Access to field sites was approved under Department of Conservation research permit 35226-RES.

LITERATURE CITED

Acevedo, M. A., and L. J. Villanueva-Rivera. 2006. From the field: using automated digital recording systems as effective tools for the monitoring of birds and amphibians. *Wildlife Society Bulletin* 34(1):211-214. http://dx.doi.org/10.2193/0091-7648(2006)34[211: UADRSA]2.0.CO;2

Bardeli, R., D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt. 2010. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters* 31(12):1524-1534. http://dx.doi. org/10.1016/j.patrec.2009.09.014

Bartlett, M. S. 1935. Some aspects of the time-correlation problem in regard to tests of significance. *Journal of the Royal Statistical Society* 98:536-543. http://dx.doi.org/10.2307/2342284

Bence, J. R. 1995. Analysis of short time series: correcting for autocorrelation. *Ecology* 76(2):628-639. http://dx.doi. org/10.2307/1941218

Bibby, C. J. 1999. Making the most of birds as environmental indicators. *Ostrich* 70(1):81-88. http://dx.doi.org/10.1080/00306-525.1999.9639752

Bioacoustics Research Program. 2011. *Raven Pro: interactive sound analysis software (Version 1.4)*. Cornell Lab of Ornithology, Ithaca, New York, USA.

Borker, A. L., M. W. McKown, J. T. Ackerman, C. A. Eagles-Smith, B. R. Tershy, and D. A. Croll. 2014. Vocal activity as a low cost and scalable index of seabird colony size. *Conservation Biology* 28(4):1100-1108. http://dx.doi.org/10.1111/cobi.12264

Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 2005. *Distance sampling: estimating abundance of biological populations.* Wiley Online Library. http://dx.doi.org/10.1002/0470011815. b2a16019

Colwell, R. K. 2013. *EstimateS: Statistical estimation of species richness and shared species from samples. (Version 9).*

Dawson, D., and P. Bull. 1975. Counting birds in New Zealand forests. *Notornis* 22(2):101-109.

Gotelli, N. J., and R. K. Colwell. 2011. Estimating species richness. Pages 39-54 in A. E. Magurran and B. J. McGill, editors. *Biological diversity: frontiers in measurement and assessment.* Oxford University Press, Oxford, UK.

Gregory, R. D., and A. van Strien. 2010. Wild bird indicators: using composite population trends of birds as measures of environmental health. *Ornithological Science* 9(1):3-22. http://dx. doi.org/10.2326/osj.9.3

Guillera-Arroita, G., and J. J. Lahoz-Monfort. 2012. Designing studies to detect differences in species occupancy: power analysis under imperfect detection. *Methods in Ecology and Evolution* 3:860-869. http://dx.doi.org/10.1111/j.2041-210X.2012.00225.x

Halley, J. M., S. Hartley, A. S. Kallimanis, W. E. Kunin, J. J. Lennon, and S. P. Sgardelis. 2004. Uses and abuses of fractal

methodology in ecology. *Ecology Letters* 7(3):254-271. http://dx. doi.org/10.1111/j.1461-0248.2004.00568.x

Hartley, S. 1998. A positive relationship between local abundance and regional occupancy is almost inevitable (but not all positive relationships are the same). *Journal of Animal Ecology* 67 (6):992-994. http://dx.doi.org/10.1046/j.1365-2656.1998.6760992. x

Hartley, L. J. 2012. Five-minute bird counts in New Zealand. *New Zealand Journal of Ecology* 36(3):1-11.

Hartshorne, C. 1956. The monotony-threshold in singing birds. *Auk* 73(2):176-192. http://dx.doi.org/10.2307/4081470

Haselmayer, J., and J. S. Quinn. 2000. A comparison of point counts and sound recording as bird survey methods in Amazonian southeast Peru. *Condor* 102(4):887-893. http://dx.doi.org/10.1650/0010-5422(2000)102[0887:ACOPCA]2.0.CO;2

Heather, B. D., and H. A. Robertson. 1996. *The field guide to the birds of New Zealand*. Viking, Auckland, New Zealand.

Heywood, V. 1995. *The global biodiversity assessment*. United Nations Environment Programme. Cambridge University Press, Cambridge, UK.

Kass, R. E., B. P. Carlin, A. Gelman, and R. M. Neal. 1998. Markov Chain Monte Carlo in practice: a roundtable discussion. *American Statistician* 52(2):93-100.

Kimura, D. K., and D. A. Somerton. 2006. Review of statistical aspects of survey sampling for marine fisheries. *Reviews in Fisheries Science* 14:3:245-283. http://dx.doi.org/10.1080/10641-260600621761

Klingbeil, B. T., and M. R. Willig. 2015. Bird biodiversity assessments in temperate forest: the value of point count versus acoustic monitoring protocols. *PeerJ* 3:e973 http://dx.doi.org/10.7717/peerj.973

La, V. T., and T. D. Nudds. 2016. Estimation of avian species richness: biases in morning surveys and efficient sampling from acoustic recordings. *Ecosphere* 7(4):e01294. http://dx.doi.org/10.1002/ecs2.1294

Legendre, P., M. R. T. Dale, M.-J. Fortin, P. Casgrain, and J. Gurevitch. 2004. Effects of spatial structures on the results of field experiments. *Ecology* 85(12):3202-3214. http://dx.doi. org/10.1890/03-0677

Lindenmayer, D. B., P. Gibbons, M. Bourke, M. Burgman, C. R. Dickman, S. Ferrier, J. Fitzsimons, D. Freudenberger, S. T. Garnett, C. Groves, R. Hobbs, R. T. Kingsford, C. Krebs, S. Legge, A. J. Lowe, R. Mclean, J. Montambault, H. Possingham, J. Radford, D. Robinson, L. Smallbone, D. Thomas, T. Varcoe, M. Vardon, G. Wardle, J. Woinarski, and A. Zerger. 2012. Improving biodiversity monitoring. *Austral Ecology* 37(3):285-294. http://dx.doi.org/10.1111/j.1442-9993.2011.02314.x

Lindenmayer, D. B., and G. E. Likens. 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in Ecology & Evolution* 24(9):482-486. http://dx.doi.org/10.1016/ j.tree.2009.03.005

Link, W. A., and M. J. Eaton. 2012. On thinning of chains in MCMC. *Methods in Ecology and Evolution* 3:112-115. http://dx. doi.org/10.1111/j.2041-210X.2011.00131.x

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. Andrew Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83(8):2248-2255. http://dx.doi.org/10.1890/0012-9658 (2002)083[2248:ESORWD]2.0.CO;2

McGregor, P. K. 1991. The singer and the song: on the receiving end of bird song. *Biological Reviews* 66:57-81. http://dx.doi. org/10.1111/j.1469-185X.1991.tb01135.x

Mier, K. L. and S. J. Picquelle. 2008. Estimating abundance of spatially aggregated populations: comparing adaptive sampling with other survey designs. *Canadian Journal of Fisheries and Aquatic Sciences* 65(2):176-197. http://dx.doi.org/10.1139/ f07-138

Mortimer, J. A. J., and T. C. Greene. 2017. Investigating bird call identification uncertainty using data from processed audio recordings. *New Zealand Journal of Ecology* 41(1):126-133. http://dx.doi.org/10.20417/nzjecol.41.10

Nice, M. M. 1931. A study of two nests of the ovenbird. *Auk* 48 (2):215-228. http://dx.doi.org/10.2307/4076789

Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. *Trends in Ecology & Evolution* 21(12):668-673. http://dx.doi.org/10.1016/j.tree.2006.08.007

Oppel, S., S. Hervías, N. Oliveira, T. Pipa, C. Silva, P. Geraldes, M. Goh, E. Immler, and M. McKown. 2014. Estimating population size of a nocturnal burrow-nesting seabird using acoustic monitoring and habitat mapping. *Nature Conservation* 7:1-13. http://dx.doi.org/10.3897/natureconservation.7.6890

R Core Team. 2015. *R: A language and environment for statistical computing (Version 3.2.3).* R Foundation for Statistical Computing. [online] URL: https://www.R-project.org/

Rahbek, C., and G. R. Graves. 2001. Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences* 98(8):4534-4539. http://dx.doi.org/10.1073/pnas.071034898

Ralph, C. J., J. R. Sauer, and S. Droege. 1995. *Monitoring bird populations by point counts*. General Technical Report PSW-GTR-149. U.S. Forest Service, Pacific Southwest Research Station, Albany, California, USA. http://dx.doi.org/10.2737/PSW-GTR-149

Ramírez, P., and J. A. Carta. 2005. Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: a case study. *Energy Conversion and Management* 46:2419-2438 http://dx.doi. org/10.1016/j.enconman.2004.11.004

Rempel, R. S., K. A. Hobson, G. Holborn, S. L. Van Wilgenburg, and J. Elliott. 2005. Bioacoustic monitoring of forest songbirds: interpreter variability and effects of configuration and digital processing methods in the laboratory. *Journal of Field Ornithology* 76(1):1-11. http://dx.doi.org/10.1648/0273-8570-76.1.1

Santer, B. D., T. M. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker and K. E. Taylor. 2000. Statistical significance of trends and trend differences in layer-average temperature time series. *Journal of Geophysical Research* 105:7337-7356. http://dx.doi.org/10.1029/1999JD901105

Sauer, J. R., D. K. Niven, J. E. Hines, D. J. Ziolkowski, Jr, K. L. Pardieck, J. E. Fallon, and W. A. Link. 2017. *The North American breeding bird survey, results and analysis 1966 - 2015.* Version 2.07.2017 USGS Patuxent Wildlife Research Center, Laurel, Maryland, USA.

Sepúlveda, M., F. Hernández, D. G. Hewitt, W. P. Kuvlesky, G. Waggerman, and R. L. Bingham. 2006. Evaluation of auditory counts for estimating breeding populations of White-winged Doves. *Journal of Wildlife Management* 70:1393-1402. http://dx. doi.org/10.2193/0022-541X(2006)70[1393:EOACFE]2.0.CO;2

Sueur J., S. Pavoine, O. Hamerlynck, and S. Duvail. 2008. Rapid acoustic survey for biodiversity appraisal. *PLoS ONE* 3(12): e4065. https://doi.org/10.1371/journal.pone.0004065

Swiston, K. A., and D. J. Mennill. 2009. Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *Journal of Field Ornithology* 80(1):42-50. http://dx. doi.org/10.1111/j.1557-9263.2009.00204.x

Temple, S. A., and J. A. Wiens. 1989. Bird populations and environmental changes: can birds be bio-indicators. *American Birds* 43(2):260-270.

Thompson, S. K. 2002. *Sampling*. Second edition. John Wiley & Sons, New York, New York, USA.

Thompson, S. J., C. M. Handel, and L. B. Mcnew. 2017. Autonomous acoustic recorders reveal complex patterns in avian detection probability. *Journal of Wildlife Management* 81:1228-1241. http://dx.doi.org/10.1002/jwmg.21285

Towsey, M., J. Wimmer, I. Williamson, and P. Roe. 2014. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. *Ecological Informatics* 21:110-119. http://dx.doi.org/10.1016/j.ecoinf.2013.11.007

Tyre, A. J., B. Tenhumberg, S. A. Field, D. Niejalke, K. Parris, and H. P. Possingham. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications* 13(6):1790-1801. http://dx.doi. org/10.1890/02-5078

Venier, L. A., M. J. Mazerolle, A. Rodgers, K. A. McIlwrick, S. Holmes, and D. Thompson. 2017. Comparison of semiautomated bird song recognition with manual detection of recorded bird song samples. *Avian Conservation and Ecology* 12(2):2. http://dx. doi.org/10.5751/ACE-01029-120202

White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. *Bird study* 46 (S1):S120-S139. http://dx.doi.org/10.1080/00063659909477239

Wimmer, J., M. Towsey, P. Roe, and I. Williamson. 2013. Sampling environmental acoustic recordings to determine bird species richness. *Ecological Applications* 23(6):1419-1428. http://dx.doi. org/10.1890/12-2088.1

Editor-in-Chief: Keith A.Hobson Subject Editor: Steven L.Van Wilgenburg



Sponsored by the Society of Canadian Ornithologists and Bird Studies Canada Parrainée par la Société des ornithologistes du Canada et Études d'oiseaux Canada



APPENDIX 1

	continuous sample set			intermittent sample set			
Taxon	Detected	Mean call rate	No. of samples where detected	Detected	Mean call rate	No. of samples where detected	
Australian Magpie (Gymnorhina tibicen)	\checkmark	0.03	10	\checkmark	0.04	10	
Blackbird (<i>Turdus</i> <i>merula</i>) / Song Thrush (<i>Turdus philomelos</i>)	~	0.17	20	\checkmark	0.16	23	
Bellbird (Anthornis melanura melanura) / Tūī (Prosthemadera novaeseelanidae)	✓	0.26	33	✓	0.17	37	
Chaffinch (Fringilla coelebs)	√	0.11	16	√	0.12	23	
Common Starling (Sturnus vulgaris)	\checkmark	0.02	5	\checkmark	0.03	6	
Dunnock (Prunella modularis)	\checkmark	0.03	5	\checkmark	0.01	3	
Eastern Rosella (Platycercus eximius)	\checkmark	0.01	6	\checkmark	0.02	5	
European Goldfinch (Carduelis carduelis)	\checkmark	0.003	4	\checkmark	0.01	6	
Fantail (<i>Rhipifura fuliginosa</i>)	\checkmark	0.06	16	\checkmark	0.04	14	
Greenfinch (<i>Carduelis chloris</i>)	\checkmark	0.029	4	\checkmark	0.029	7	
Grey Warbler (Greygone igata)	√	0.08	24	√	0.08	25	
House Sparrow (Passer domesticus)	~	0.01	1	~	0.001	1	
Kererū (Hemiphaga novaeseelandiae)	1	0.01	6	1	0.01	7	

Table A1.1. List of the species detected from the 50 continuous and intermittent samples.

Kākāriki (<i>Cyanoramphus</i> spp.)	\checkmark	0.001	1	~	0.002	3
Sacred Kingfisher (<i>Todiramphus</i> sanctus)	\checkmark	0.01	2	~	0.01	3
Rifleman (Acanthisitta chloris)	\checkmark	0.02	6	1	0.03	9
Silvereye (Zosterops lateralis)	\checkmark	0.07	15	1	0.07	30
Tomtit (<i>Petroica</i> <i>macrocephala</i>)	\checkmark	0.09	14	1	0.08	22
Whitehead (Mohoua albicilla)	\checkmark	0.07	13	1	0.07	13
White-faced Heron (Egretta novaehollindiae)	\checkmark	0.002	1	X	0	0
Kākā (Nestor meridionalis)	Х	0	0	1	0.001	1
Paradise Duck (<i>Tadorna variegate</i>)	Х	0	0	1	0.002	2
Skylark (Alauda arvensis)	Х	0	0	1	0.02	1
Spur-winged Plover (Vanellus miles)	Х	0	0	1	0.001	1
Yellowhammer (Emberiza citronella)	Х	0	0	1	0.005	2



Figure A1.1. Location of sound recording devices across the lower North Island of New Zealand.



Figure A1.2: Mean temporal autocorrelation (r) for lags 1 to 10, for nine bird taxa sampled with two different presence-absence manual bioacoustic scoring methods. The continuous sampling method (red line) used a chronologically continuous five-minute subsets of 30-minute audio recordings, the intermittent method (blue line) used the first 10 seconds of every minute to create a non-continuous five minute subset of the original recoding. In both methods presence/absence was scored in ten-second blocks. Audio recordings were taken from 10 forest sites across the lower North Island, New Zealand. Only recordings with greater than 1 and less than 29 presences (out of 30) were included in the calculation of autocorrelation.