Please cite the Published Version

ORIGINAL RESEARCH

Optimizing tropical forest bird surveys using passive acoustic monitoring and high temporal resolution sampling

Oliver C. Metcalf[1] (iD), Jos Barlow[2,3,4], Stuart Marsden[1], Nárgila Gomes de Moura[5],
Erika Berenguer[3,6], Joice Ferreira[7,8,9] & Alexander C. Lees[1,5]

[1]Division of Biology and Conservation Ecology, Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK
[2]Departamento de Biologia, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil
[3]Lancaster Environment Centre, Lancaster University, Lancaster, Lancashire, UK
[4]MCTIC/Museu Paraense Emílio Goeldi, Belém, Pará, Brazil
[5]Cornell Lab of Ornithology, Cornell University, Ithaca NY, 14850, USA
[6]Environmental Change Institute, University of Oxford, Oxford OX1 3QY, UK
[7]Programa de Pós-Graduação em Ciências Ambientais, Universidade Federal do Pará/Embrapa Amazônia Oriental/Museu Paraense Emílio Goeldi, Belém PA, 66075-110, Brazil
[8]Programa de Pós-Graduação em Ecologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém PA, 66075-110, Brazil
[9]Laboratório de Análises de Sistemas Sustentáveis, Empresa Brasileira de Pesquisa Agropecuária da Amazônia Oriental, Belém PA, 66095-903, Brazil

Abstract

Estimation of avian biodiversity is a cornerstone measure of ecosystem condition. Surveys conducted using autonomous recorders are often more efficient at estimating diversity than traditional point-count surveys. However, there is limited research into the optimal temporal resolution for sampling—the trade-off between the number of samples and sample duration when sampling a survey window with a fixed survey effort—despite autonomous recorders allowing easy repeat sampling compared to traditional survey methods. We assess whether the additional temporal coverage from high temporal resolution (HTR) sampling, consisting of 240 15-s samples spread randomly across a survey window detects higher alpha and gamma diversity than low temporal resolution (LTR) sampling of four 15-min samples at the same locations. We do so using an acoustic dataset collected from 29 locations in a region of very high avian biodiversity—the eastern Brazilian Amazon. We find HTR sampling outperforms LTR sampling in every metric considered, with HTR sampling predicted to detect approximately 50% higher alpha diversity, and 10% higher gamma diversity. This effect is primarily driven by increased coverage of variation in detectability across the morning, with the earliest period containing a distinct community that is often under sampled using LTR sampling. LTR sampling produced almost four times as many false absences for species presence. Additionally, LTR sampling incorrectly found 70 species (34%) at only a single forest type when they were in fact present in multiple forest types, while the use of HTR sampling reduced this to just two species (0.9%). When considering multiple independent detections of species, HTR sampling detected three times more uncommon species than LTR sampling. We conclude that high temporal resolution sampling of passive-acoustic monitoring-based surveys should be considered the primary method for estimating the species richness of bird communities in tropical forests.

Introduction

Estimation of avian biodiversity is a cornerstone measure of ecosystem condition. Reliable detection, identification and counting of birds can, however, be difficult in high biodiversity environments like tropical forests (Robinson et al., 2018) and accumulating sufficient inventory completeness can be challenging (Karr, 1981; Robinson et al., 2000; Terborgh et al., 1990). Point counts are well established as a standard survey technique for obtaining measures of bird species richness, abundance and population density, particularly in forest habitats (Bibby et al., 2000). Now that affordable and reliable passive-acoustic monitoring (PAM) equipment has become available (Gibb et al., 2019), autonomously recorded surveys—in which recording units are left to document soundscapes over extended periods—are emerging as a supplement or alternative to traditional field-conducted point counts (Shonfield & Bayne, 2017). A recent review found that over equivalent time periods recorder-based surveys detect an average of 11% more species than traditional point counts with field-based observers, hereafter 'traditional surveys', albeit often with slightly different species composition—alongside a host of other advantages (Darras et al., 2019).

Passive acoustic monitoring enables high temporal resolution sampling through taking a large number of non-consecutive short samples spread across the survey window with no additional logistical costs. For instance in a survey window consisting of 1 h, a low temporal resolution sampling approach may be to take a single 5-min sample (300 s), necessarily clumped together at a single point in time. High temporal resolution (HTR) sampling would instead take twenty 15-s samples (also totalling 300 s), spread across the whole hour. This approach can be extended to survey windows of a morning, multiple days, seasons or years. Despite this potential benefit, very few studies have investigated the effect of using high temporal resolution sampling (sample durations of <1 min across the sort of periods traditionally associated with species richness assessment—days or weeks, as opposed to months, seasons or years). There are good theoretical reasons to believe sampling at very high temporal resolution across a broader overall time period may produce better estimates of species richness and more accurate species inventories.

Estimating species richness depends on two factors: availability and detectability (Kéry & Schmidt, 2008). The number of species available for detection over time (e.g. the number of species close enough to the recorder to be heard), varies as species move—for instance the number of available species would be much greater if a large mixed-species flock entered the detection space of the recorders. The distance over which a species is available

for detection also varies, impacted by the amplitude and acoustic frequency of the call and a range of environmental factors (Darras et al., 2016; Yip et al., 2017). The detectability of each species (e.g. whether an individual of the species makes an identifiable sound during the survey) is the probability of recording the species when it is available, and is dependent on calling frequency. This is influenced by the abundance and call rate of the available species. In the tropics, hyper-diverse avian communities have high variation in abundance, with a small number of commoner species and a long tail of rarer species (Robinson et al., 2000; Terborgh et al., 1990). Call rates can vary by many orders of magnitude—for example Screaming Pihas *Lipaugus vociferans* may vocalize for 77% of the time between 06:45 and 17:15 (Snow, 1961), while Variegated Antpittas *Grallaria varia* have been shown to sing only twice in 50 days (Jirinec et al., 2018). The more this variation is temporally structured, the more beneficial the additional temporal spread gained of HTR sampling will be.

Some temporal structuring of species availability and detectability is well known—traditional point counts are often conducted in the 2 h following sunrise targeting the periods with most vocal activity (Bibby et al., 2000; La & Nudds, 2016; Lynch, 1995; Venier et al., 2012). However, there is also likely to be fine-scale variation in the proportion of the total species pool that is available and detectable within that period (Verner & Ritter, 1986), meaning that distributing a higher number of samples across the period increases the likelihood of some samples coinciding with periods in which a high proportion of the total species pool is detectable (Fig. 1). Furthermore, high temporal resolution also increases the probability of detection of species that only vocalize within strict temporal niches, or are only detectable at certain periods (Gil & Llusia, 2020). For example forest falcons *Micrastur* spp., only reliably vocalize before and around sunrise (Fjeldså et al., 2020), when a low proportion of the total species pool is detectable. Other species may have habitual movements that make them only available for detection during narrow windows. With low temporal resolution sampling, it may be possible to sample during one or several of these availability windows if they are known; but this would reduce the capacity to sample at times with a high proportion of the species pool available. At the scale of a whole survey period, this variation could also be structured by feeding opportunities or weather events between days, or for longer survey periods, seasonal variation caused by, for example migration or breeding activity. By investing effort in many shorter samples distributed across the survey period, it ought to be possible to obtain a more representative sample.
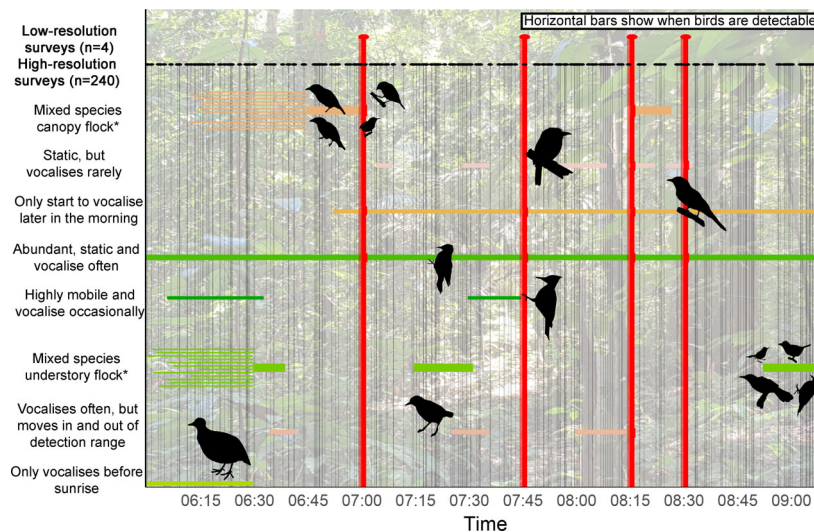
Figure 1. Theoretical model of high and low temporal resolution sampling regimes over one morning in the tropics. Black vertical lines show 240 1-s samples (illustrated at double width), while red vertical lines show four 1-min samples. This represents a similar contrast in resolution to four 15-min samples and 240 15-s samples over a 15-day survey season, as compared in this paper. The y-axis shows a non-exhaustive selection of behaviours that impact detection probability. Mixed flocks shown both prior to and after formation. Patterns of bird behaviour affecting detectability are hypothetical. Bird silhouettes from www.phylopic.org.

A number of studies, both traditional and PAM based, indicate that using a higher temporal resolution for sampling allows detection of a higher number of species compared to lower resolution samples (Bayne et al., 2017; Cook & Hartley, 2018; Fuller & Langslow, 1984; Klingbeil & Willig, 2015; Smith et al., 2020; Wimmer et al., 2013). However, these studies were conducted predominantly in temperate forests or arid systems in regions of relatively low species richness, and none used a minimum sample duration of less than 1 min. Additionally, it has been suggested that shorter survey durations may reduce detection probability, increase false negatives and produce inferior estimates of species diversity (Sugai et al 2020).

We tested the impact of using high temporal resolution (HTR)—240 15-s samples at each site—and low temporal resolution (LTR) sampling—four 15-min samples—on species detection, without increasing total sampling effort. We used an acoustic dataset collected between June and August 2018 in eastern Amazonia We compared the results between sampling resolutions to answer the following questions: (1) does HTR sampling result in higher estimates of species richness and a faster species accumulation, and (2) if so, is the increase in a benefit driven by the better capture of variation in species richness across the dawn chorus or by day-to-day variation? (3) Does HTR sampling increase the number of false absences and falsely unique occurrences? Finally, (4) is HTR sampling more efficient at detecting rare species?

Materials and Methods

Data collection

We collected acoustic data from 29 of the survey transects of the Sustainable Amazon Network (Gardner et al. 2013) distributed across an area of approximately 1 million ha of the eastern Brazilian Amazon in the municipalities of Santarém, Belterra and Mojuí dos Campos (latitude ~ −3.046, longitude −54.947 WGS 84), hereafter 'Santarém' in the Brazilian state of Pará. Survey points were located halfway along permanent 300-m forest transects. All transects were located in non-seasonally inundated 'Terra firme' forest and distributed across a human-disturbance gradient, comprising seven forest classes. To minimize spatial correlation, survey points were separated by a minimum distance of 2 km.

All recordings were made between 12th June 2018 and 16th August 2018, outside of the peak period for bird breeding (Kirwan, 2009) which commences with the onset of the rainy season in November, and across a period in which detectability and community composition should be relatively constant. Recordings at each survey point were made over one or two recording periods, with each recording period varying in length between 3 and 22 days for logistical reasons. Recordings were made for between 6 and 22 days (mean 16 days ±3.4 (SD)) at each location, meaning that temporal resolution of the samples varied between sites, but not

relatively between the sampling methods. Full details of recording periods and sampling resolutions for each location are given in SOM Appendix S1 and S2.

We installed Frontier Labs Bioacoustic Recording Units with a 16-bit 44.1 kHz sampling rate each survey point. Recorders were placed in trees at a height of 7–10 m, with the microphone facing downward, 10–20 m from the transect to reduce the chance of recorder theft. Recording units were positioned to avoid sound being blocked by overhanging branches. Frontier Labs microphones have 80 dB signal to noise ratio and 14dBA self-noise, a fixed gain pre-amp of 20dB, a flat frequency response (±2dB) from 80Hz to 20kHz and an 80 Hz high-pass filter to filter out low-frequency wind noise (Frontier Labs, 2015). All files were recorded continuously in wav format.

The continuous acoustic recordings were randomly and independently subsampled twice. In the first subsample (hereafter 'LTR samples'), survey periods were 15 min in duration, and four periods were extracted per survey point, totalling 1 h of data from each transect. Across all survey points, there were a total of 116 LTR samples. We used 15-min durations as it is a commonly used point-count duration in tropical forests (Robinson et al., 2018), and as previous traditional surveys from the same location have used this survey duration (e.g. Moura et al., 2013). The second subsample (hereafter 'HTR samples') again independently sampled 1 h of recordings from each survey point, but this time in the form of 240 15 s periods, totalling 6960 samples across all transects. The selection of 15 s durations for HTR sampling is primarily a trade-off between the highest possible resolution, the associated increase in effort during analysis through the increasing number of files and the number of complete versus truncated vocalizations, which can be difficult or impossible to identify without a longer recording. Further considerations include minimizing bird movement in and out of the detection space of recorders and 15 s spectrograms can easily be displayed on a standard monitor at a resolution where vocalizations can be visually recognized. All samples for both survey methods were taken in a two-and-a-half-hour period starting 30 min before sunrise, which has been shown to be the most effective period for estimating species richness with PAM surveys (Wimmer et al., 2013). Subsampling was not stratified within that period, but LTR samples commenced on the hour, or 15, 30 or 45 min past the hour, to avoid overlapping samples. Audio containing heavy rainfall was removed manually prior to initial sampling for LTR sampling, while rainfall in HTR samples was detected using the hardRain package in R (Metcalf et al., 2020) and removed.

Analysis

The audio samples were analysed manually, through visually inspecting spectrograms generated in Raven Pro (Center for Conservation Bioacoustics, 2019) at the default settings, and listening to the recordings. All identifiable avian vocalizations were assigned to species by a highly experienced ornithologist (NGM, for survey experience in the region, see Moura et al., (2013), and Moura et al., (2016)). All vocalizations that could only be determined to family level were discarded. During analysis, it was apparent that 343 of the 6960 HTR samples fell during periods of rain intense enough to significantly inhibit bird vocalization activity and/or detection. These were removed from consideration but not replaced, leaving 6617 samples and uneven sample size (see SOM Appendix 1). Consequently, for each survey point, we calculated both observed species richness and rarefied species richness for 45 min of sample effort to account for the uneven total sampling effort across methods, using the iNext package in R (Hsieh et al., 2020), but patterns and results were similar to observed species richness, so only observed species richness is considered hereafter.

We compared alpha and gamma diversity metrics between the two sampling methods. First, we modelled species richness against sampling resolution at each survey point using linear-mixed effect models in the lmerTest package (Kuznetsova et al., 2017) with a Gaussian error structure. We used survey point nested within the forest disturbance class as a random effect. We also calculated total species richness across all survey points (gamma diversity). For a repeat of this analysis including rarefied species richness, and data from traditional point-counts conducted in 2016, see SOM Appendix 3. To address whether the use of HTR sampling accrued species richness at a faster rate than LTR sampling, we constructed sample-based species accumulation curves for each survey method, interpolating for 20 h of sampling effort using the iNext package.

We investigated the relative impact of increased temporal coverage within each morning and across survey days. We partitioned the effect of within-morning and day-to-day variation in detectability using six variables related to the temporal coverage of the two sampling methods. Two variables measured coverage across the morning, in which 06:00–08:30 was split into 1-min ($n = 150$) and 30-min ($n = 5$) intervals and the percentage of intervals sampled by either method across all days of recording at each survey point was calculated. The proportion of 1-min intervals sampled is intended to represent stochastic or unstructured variation in detection rates across the morning, whilst 30-min intervals capture temporal trends across the morning. We produced similar measurements of day-to-day coverage, calculating the proportion of

surveyed dates covered and the percentage of quintiles of the survey dates covered, treating survey days as continuous. We also included two measures of the evenness of the sampling, the mean and maximum difference between Julian dates sampled in days. We used a Lasso-type modelling approach with L1 penalized estimation in the glmmLasso package (Groll, 2017) to ascertain variable importance. All variables were scaled by subtraction of the variable mean prior to analysis, to aid in comparison. We used survey point as a random effect and a Poisson distribution. We tested lambda values between zero and 100, and used Bayesian Information Criterion score (Kass & Raftery, 1995) to select the optimal value.

In addition, we examined structure in species communities across the morning and across the survey season, to see whether this could be driving differences in detection rates with increased temporal coverage with HTR sampling. This was done by conducting a non-metric multidimensional scaling analysis on the species communities detected by HTR sampling in each of the 30-min intervals between 06:00 and 08:30 using the vegan package (Oksanen et al., 2019), and each of five evenly spaced periods of 13 days between the first survey on 13th June 2018 and the last on 16th August 2018. We conducted a pairwise Permutational Multivariate Analysis of Variance (PERMANOVA) using the Jaccard method from the pairwiseAdonis (Martinez Arbizu, 2020) package to test for significant differences between species communities at each 30-min interval and each quintile of the survey season.

To explore the difference in sites at which HTR or LTR sampling detected species. We summed the number of survey points at which each species was detected and compared both methods using a Wilcoxon signed ranks test. We also calculated the number of species falsely found to be absent per survey point. A species was determined to be falsely absent if it was undetected at a location by one temporal resolution of sampling but detected at the same location by the converse resolution. In addition, we looked at extreme cases of false absences, in which species were detected at only a single survey point by a sampling resolution, but were actually detected at other locations by the converse method (hereafter 'false uniqueness'), something that is likely to be highly detrimental to the accuracy of habitat modelling in particular (Gu and Swihart, 2004). As most analyses of this type are directed at the habitat level we analysed this at the scale of forest class, and calculate the proportion of the total species richness of each forest class that was determined to be falsely unique species. The seven forest classes are: undisturbed forest (five survey points), selectively logged forest (four survey points), secondary forest—forest recovering from complete historical clearance sensu Putz

and Redford, (2010) (three points), and four categories of burnt forest. The four burnt categories were categorized dependent on whether they burnt during the extensive El Niño-induced fires in 2015 and whether they have been selectively logged, with all logging occurring prior to 2015. The categories are; burned in 2015 but never logged (five points), logged and burned prior to 2015 (four points), logged and burned in 2015 (five survey points) and logged and burned both before 2015 and in 2015 (three survey points).

To test if HTR sampling detected a higher proportion of rare species, we compared the relative abundance of species detected by both methods using chi-squared tests. We designated each species as common, fairly common or uncommon, using the Parker et al., (1996) Neotropical bird trait database. Species marked as intermediate between two abundance classes in Parker were assumed to belong to the rarer class, categories marked as uncertain were assumed to be correct, and we combined the categories of uncommon, patchily distributed and rare. Species nomenclature was aligned to the taxonomy of Piacentini et al., (2015). We also tested whether HTR sampling detected each rare species more often. To emphasize the importance of a high number of independent detections of rare species, we also compared the proportion of species that were detected from a minimum of two transects and with >10 total detections (hereafter 'multiple independent detections').

Results

Species richness

We detected higher alpha and gamma diversity (Fig. 2A) using HTR sampling. The mixed-effects model predicted that HTR sampling detects 22.9 ± 3.7 (SE) more species per survey point than LTR, with HTR detecting 66.4 per point and LTR sampling detecting 43.5 species per point. In total, we detected 245 species; 224 species using HTR sampling with a median of 4.0 ± 0.0 (SE) species and 204 species using LTR sampling with a median of 19.5 ± 0.7 species per sample. HTR sampling detected 41 species undetected in LTR samples across the landscape, twice as many as LTR sampling which found 21 species not detected by HTR sampling. We found that for sample-based rarefaction/extrapolation by sample method (Fig. 2B), HTR sampling led to steep increases in species accumulation up to around 4 h of sampling effort, with 176 ± 2 (SE) species detected, and then attenuated, with species accumulation continuing up to 20 h. In contrast, LTR sampling showed a shallower curve, in which the accumulation did not slow as quickly. LTR sampling detects lower species richness at all quantities of sampling
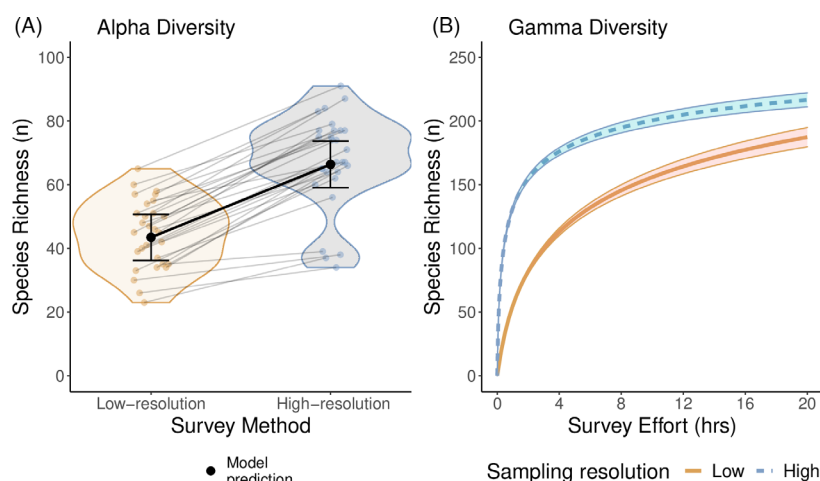
Figure 2. (A) Comparison of the species richness detected at each of 29 survey points employing either low temporal resolution samples, comprised of four 15-min periods, or high temporal resolution surveys of 240 15-s periods. (B) Sample-based species accumulation curves for the two sampling methods, showing interpolated predictions up to 20 h of sampling effort.

effort and was predicted to detect 187 ± 8 (SE) species after 20 h of sampling effort, compared to 217 ± 5 species by HTR sampling. HTR sampling was predicted to take just 11 hrs 23 mins to achieve the same species total as LTR did in all surveys (204 species, 29 h).

When looking at the impact of HTR sampling increasing coverage across different temporal scales, we found that increased coverage across the morning had the greatest effect. The Lasso model showed that only the proportion of 30-min intervals covered ($p \leq 0.01$) had a significant impact on species richness (Table 1). This indicates that the difference in detected species richness between the two methods is driven by high community turnover across the morning, which is better detected with HTR sampling.

The ordination analysis shows in more detail the importance of increased coverage across the morning. The ordination of the five 30-min periods (stress = 0.21) shows clear temporal gradation along MDS2 (Fig. 3A), with the earliest period occupying the highest values and the latest periods the lowest. The first 30-min interval, 06:00-06:29, is both the most distinct from other periods, and also has the most internal variation in species community. The PERMANOVA results indicated this period to be significantly different from all other periods (all Bonferroni adjusted p-values = 0.01). After this time period, the distinction between the community decreases, although the PERMANOVA analysis still found significant differences between 06:30–06:59 and 07:30–07:59, and between 06:30–06:59 and 08:00–08:29 (Bonferroni adjusted p-values of 0.03 and 0.01 respectively). In contrast the ordination of seasonal variation (stress = 0.18) shows that there is very little community turnover

Table 1. Coefficients for the fixed effects of an L1-Penalized estimation model showing variable importance

Variable	Estimate	z value	p value
(Intercept)	4.0 ± 0.02	222.12	<0.01
30-min intervals sampled (%)	0.14 ± 0.05	2.79	<0.01
Surveyed dates sampled (%)	0.08 ± 0.06	1.39	0.16
Surveyed date quintiles sampled (%)	0.02 ± 0.02	0.78	0.44
Mean difference between sampled dates	0 ± 0.03	0.09	0.93

The following variables have coefficients collapsed to zero; Maximum difference between sampled dates (days), Proportion of 1-min intervals sampled (%).

between the start and end of the survey season (Fig. 3B), with each of the quintiles overlapping and the PERMANOVA analysis showing no significant differences between any of the groups (Bonferroni adjusted p-value >0.05).

Species were detected at more survey points using HTR sampling, a median of 8 ± 0.57 to 4 ± 0.47 (V = 976, $P < 0.001$; Fig. 4A). Consequently, we found that the higher detection frequency of HTR had a striking effect on the number of false negatives, with LTR sampling producing 927 false absences compared to just 263 for HTR sampling. Every survey point had fewer false absences with HTR sampling than LTR (Fig. 4B). In the most extreme case, 46 of 96 (48%) species were missed with LTR samples, but only nine (9%) were missed with HTR sampling. This pattern was also apparent when looking at false uniqueness (Fig. 4C). There were only two species that HTR sampling wrongly identified as unique to a
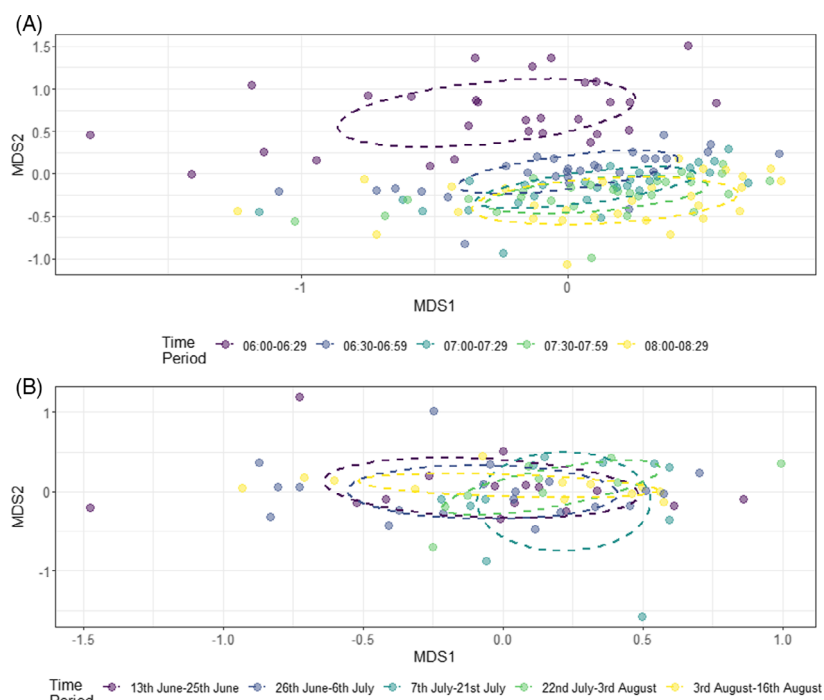
Figure 3. Non-metric multidimensional scaling analysis on the species communities detected by HTR sampling. Fig. 3A shows species communities from 30-min intervals across the dawn chorus, while Fig. 3B shows communities from evenly spaced periods across the survey season. Points represent the species community at a single survey point and time interval, while size of the point represents estimated species richness based on 50 HTR samples. Ellipses show the standard deviations for each interval.

forest class, compared to 70 by LTR sampling. One class, forest logged and burned in 2015, had an exceptionally high error rate using LTR samples, with 25 species or 21% of the total detected species at that class being wrongly detected as unique—something that could be highly misleading in habitat or distribution modelling.

HTR sampling detected a mean 10% ±0.7 (SD) more species for common, fairly common and uncommon birds. However, both sampling methods detected a remarkably similar proportion of common, fairly common and uncommon species (Fig. 5). When only considering multiple independent detections of species (10+ total detections and detected at two or more locations), HTR sampling detected substantially more species than LTR sampling, with the largest difference being for uncommon species for which HTR sampling detected nearly three times as many species ($n = 13$ and 38 respectively). Furthermore, the number of uncommon species detected as a proportion of all species detected multiple times independently declined for LTR sampling (28% to 18%) but stayed relatively stable for HTR (29–25%). When analysing LTR sampling, the proportion of uncommon species in the total species pool declined from 28% for all species detected, to 18% when considering only multiple independent detections. For HTR, the detection of uncommon species remained similar, regardless of the abundance metric used - 29–25%.

Discussion

We found that high temporal resolution (HTR) sampling outperformed low temporal resolution sampling in every metric considered, often by a substantial margin. In particular, HTR sampling recorded just over 50% more species at each location, as well as finding substantially higher gamma diversity across the entire survey. This is similar to results reported by Bayne et al (2017), that found after 1000 min of sampling effort, 1-min samples had detected twice the species richness of 10-min samples. HTR sampling also produces far fewer false negatives for species presence, and wrongly identifies far fewer species as restricted to a single forest class. Additionally, the proportion of rare species independently detected multiple times declines far less rapidly with HTR sampling than with LTR sampling, and this robustness to rarer species is highly advantageous in surveying bird communities where there is a disproportionately high number of rare species, particularly in the tropics (Robinson & Curtis, 2020).

We demonstrate that in this study the difference in species detections between the sampling methods is largely driven by increased temporal coverage across the morning. In particular, HTR sampling at every site included sampling during the period with a highly distinct community between 06:00 and 06:29, in which
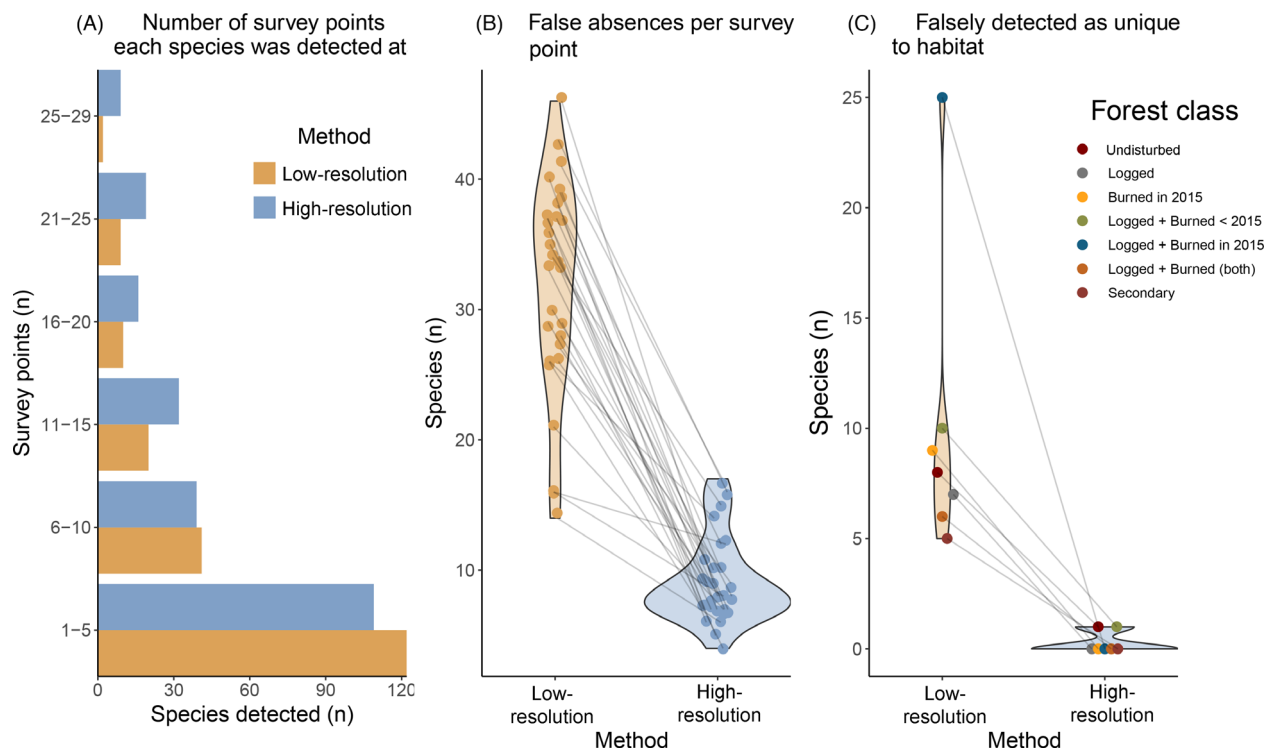
Figure 4. Species presence by survey point. (A) The number of survey points at which each species was detected. (B) The number of species falsely identified as absent per survey point. (C) the number of species wrongly identified as unique to each forest class.
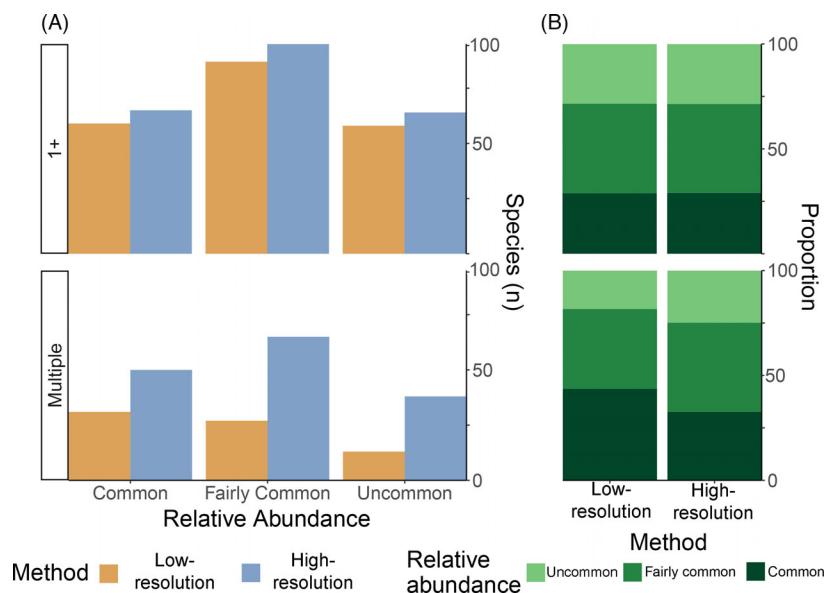
Figure 5. The proportion of common, fairly common and uncommon species detected using both high and low temporal resolution sampling methods.

nocturnal and crepuscular species such as nightjars, owls and tinamous are most likely to be recorded, and forest falcons (Micrastur spp.) most often vocalize (Fjeldså

et al., 2020). Whereas previous studies have suggested optimal sampling strategies that focus on periods of peak activity (Smith et al., 2020), these results suggest

including less species rich, but distinct periods are a vital part of any optimal sampling strategy.

However, community turnover across the morning period is only a subset of the variation across the entire diel cycle. Had we extended the sampling window across the full 24 h of each day we would expect to see an even greater positive benefit for HTR sampling, with several studies having highlighted the distinct communities that do not receive adequate coverage when surveys are solely conducted in the mornings (Araújo et al., 2020; La & Nudds, 2016; Wimmer et al., 2013). This principle can be further extended to other, longer, cycles that impact species availability and detection such as the lunar or annual cycles. Although we saw little evidence of community turnover across our short (66 days) field season, this is unlikely to remain the case had we surveyed over a longer period that included seasonal changes in weather, food resource and arrival/departure of migratory species (Thompson et al., 2017).

The potential to extend sampling over long periods highlights necessary trade-offs between total sampling effort, sampling density and spread. Here, our strategy of random sampling is unlikely to be optimal—either for this study or over longer periods. HTR sampling allows a high proportion of samples to be taken at periods of high availability and detectability—around dawn, for instance or the onset of the breeding season (Pieretti et al., 2015), and fewer samples at periods with unique community composition such as the dusk chorus (Farina & Gage, 2017; Wimmer et al., 2013). The lasso model did not show a significant effect for the proportion of days surveyed, suggesting that density of sampling is less important than increasing temporal coverage. This is in keeping with findings elsewhere that showed the majority of species are detected on the first day when temporal coverage is high (Wimmer et al., 2013). However, this could be caused by a high sampling rate for both methods leading to a level of redundancy—100% of days with HTR sampling and always greater than 15% with LTR. It is plausible that at much lower levels of coverage, species detections will decline as redundancy in coverage declines and small amounts of day-to-day variation are not accounted for.

While traditional point counts (with supplementary recording) may offer some advantages in areas with a high proportion of species only detectable by sight, LTR sampling offers few benefits over HTR sampling except efficiency in analysis. HTR sampling can take substantially longer to analyse than LTR sampling due to the 'overhead costs'—recording metadata and results for each sample, increased data management time caused simply by the increase in the number of samples. It is also worth noting that the manual analysis of acoustic data requires highly skilled ornithologists for the identification process, and accurate interpretation of results requires intimate knowledge of the ecosystems from which the data are collected. It is therefore likely that HTR sampling will be most successfully used by those who have spent a considerable amount of time in the field gaining that knowledge, and should not be seen as an alternative to doing so.

Conclusion

Much recent research on PAM surveys has focussed on automated methods (Stowell & Sueur, 2020) but it is likely to be years or even decades before off-the-shelf, highly accurate classification methods are available for the world's most speciose regions (Gibb et al., 2019; Priyadarshani et al., 2018; Sugai et al., 2019). In this context, the improved efficacy of high temporal resolution sampling for species inventories and estimation of species richness represent a significant benefit.

A combination of traditional and autonomous survey techniques should still be considered the gold standard for conducting bird species inventories (Robinson & Curtis, 2020), however, if only a single survey method is to be used, repeated HTR sampling is likely to be the most effective. There is strong evidence that surveys conducted on lower resolution samples from PAM surveys outperform human observations for bird inventories (Darras et al., 2019), suggesting that autonomous surveys should be used preferentially or in combination with traditional point-count surveys. We believe that HTR sampling from PAM surveys should be considered the primary method for sampling bird communities in tropical forests most of the time, exceptions being when autonomous recordings are not possible, for example if equipment cost is too high, when estimates of abundance are of higher priority than estimates of species richness. The degree of advantage, and hence the benefits of intelligent sampling design, depend on the degree to which each of diurnal, daily and seasonal variation in detectability of individual species plays out. Given these additional benefits of HTR sampling, we believe that within tropical forest environments manually conducted point counts should mainly be employed as a supplement to HTR sampling.

Acknowledgements

in Brazil, and to Jack Shutt and Christian Devenish for advice on statistical analysis. Fieldwork in Brazil was supported by research grants ECOFOR (NE/K016431/1), and AFIRE (NE/P004512/1), PELD-RAS (CNPq/CAPES/ PELD 441659/2016-0) and the BNP Paribas Foundation's Climate & Biodiversity Initiative.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contributions

OCM designed the study, undertook data collection, undertook statistical analysis and drafted the paper. ACL, JB and SM made significant contributions to the study conception and design, aided in statistical analysis and made substantial revisions to paper drafts. NGM undertook all processing of the audio data. EB and JF contributed to the study design, collection of the data and made substantial revisions to paper drafts.

Data Availability Statement

We will make the species presence data (per survey and survey method) available in the Dryad data repository on acceptance of this paper.

REFERENCES

Bayne, E., Knaggs, M. & Solymos, P. (2017) How to Most Effectively Use Autonomous Recording Units When Data are Processed by Human Listeners, Bioacoustic Unit, University of Alberta and Alberta Biodiversity Monitoring Institute.

Bibby, C.J., Burgess, N.D. & Hill, D.A. (2000) *Bird census techniques*, Revised edition. London, UK: Academic Press.

Center for Conservation Bioacoustics, (2019). Raven Pro: Interactive Sound Analysis Software (Version 1.4) [Computer software]

Cook, A. & Hartley, S. (2018) Efficient sampling of avian acoustic recordings: Intermittent subsamples improve estimates of single species prevalence and total species richness. *Avian Conserv. Ecol.*, **13**, https://doi.org/10.5751/ ACE-01221-130121.

Darras, K., Batáry, P., Furnas, B.J., Grass, I., Mulyani, Y.A. & Tscharntke, T. (2019) Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide. *Ecological Applications*, **29** (6), e01954. https://doi.org/10.1002/eap.1954.

Darras, K., Pütz, P., Fahrurrozi, R.K. & Tscharntke, T. (2016) Measuring sound detection spaces for acoustic animal sampling and monitoring. *Biological Conservation*, **201**, 29– 37. https://doi.org/10.1016/j.biocon.2016.06.021.

de Araújo, C.B., Jardim, M., Saturnino, N.D.S.F., Rosa, G.M., Lima, M.R. & dos Anjos, L. (2020) The optimal listening period for an effective assessment of bird richness and composition: a case study of Neotropical forest. *Journal of Ornithology*, **1**, 3. https://doi.org/10.1007/s10336-020-01812- 6.

de Piacentini, V.Q., Aleixo, A., Agne, C.E., Maurício, G.N., Pacheco, J.F., Bravo, G.A. et al. (2015) Annotated checklist of the birds of Brazil by the Brazilian Ornithological Records Committee / Lista comentada das aves do Brasil pelo Comitê Brasileiro de Registros Ornitológicos. *Revista Brasileira de Ornitologia*, **23**(2), 91–298.

Farina, A. & Gage, S.H. (2017) *Ecoacoustics: the ecological role of sounds*. Ecoacoustics: The Ecological Role of Sounds. https://doi.org/10.1002/9781119230724.

Fjeldså, J., Marks, J.S. & Sharpe, C.J. (2020) Cryptic Forest-Falcon (Micrastur mintoni), version 1.0. In Birds of the World (J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, and E. de Juana, Editors). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.cryfof1.01

Fuller, R.J. & Langslow, D.R. (1984) Estimating numbers of birds by point counts: how long should counts last? *Bird Study*, **31**, 195–202. https://doi.org/10.1080/ 00063658409476841.

Gardner, T.A., Ferreira, J., Barlow, J., Lees, A.C., Parry, L., Vieira, I.C.G. et al. (2013) A social and ecological assessment of tropical land uses at multiple scales: the Sustainable Amazon Network. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**(1619), 20120166. https://doi.org/10.1098/rstb.2012.0166.

Gibb, R., Browning, E., Glover-Kapfer, P. & Jones, K.E. (2019) Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, **10**(2), 169–185. https://doi.org/10.1111/2041- 210X.13101.

Gil, D. & Llusia, D. (2020) The bird dawn chorus revisited. In: Aubin, T. & Mathevon, N. (Eds.), *Coding strategies in vertebrate acoustic communication*. Springer, Cham, pp. 45– 90. https://doi.org/10.1007/978-3-030-39200-0_3

Groll, A. (2017) glmmLasso: Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation. R package version 1.5.1. https://CRAN.R-project.org/package=glmmLasso

Gu, W. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, **116**, 195–203. https://doi. org/10.1016/S0006-3207(03)00190-3.

Hsieh, T.C., Ma, K.H. & Chao, A. (2020) iNEXT: iNterpolation and EXTrapolation for species diversity. R package version 2.0.20 URL: http://chao.stat.nthu.edu.tw/ wordpress/software-download/.

Jirinec, V., Elizondo, E., Rutt, C.L., Elisa Elizondo, B.C., Cameron Rutt, B.L. & Philip Stouffer, B.C. (2018) Space use, diurnal movement, and roosting of a Variegated Antpitta (Grallaria varia) in Central Amazonia.

Karr, J. (1981) Surveying birds in the tropics. *Studies in Avian Biology*, **6**, 548–553.

Kass, R.E. & Raftery, A.E. (1995) Bayes factors. *Journal of American Statistical Association*, **90**, 773–795. https://doi.org/10.1080/01621459.1995.10476572.

Kéry, M. & Schmidt, B.R. (2008) Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, **9**, 207–216. https://doi.org/10.1556/ComEc.9.2008.2.10.

Kirwan, G.M. (2009) Notes on the breeding ecology and seasonality of some Brazilian birds. *Revista Brasileira de Ornitologia*, **17**, 121–136.

Klingbeil, B.T. & Willig, M.R. (2015) Bird biodiversity assessments in temperate forest: The value of point count versus acoustic monitoring protocols. *PeerJ*, **2015**, e973. https://doi.org/10.7717/peerj.97.

Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B. (2017) lmerTest Package: tests in linear mixed effects models. *Journal of Statistical Software*, **82**(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

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

Labs, F. (2015) Bioacoustic Audio Recorder User Guide. Accessed at: https://frontierlabs.com.au/.

Lynch, J.F. (1995) Effects of Point Count Duration, Time-of-Day, and Aural Stimuli on Detectability of Migratory and Resident Bird Species in Quintana Roo, Mexico. Ralph, C. John; Sauer, John R.; Droege, Sam, Tech. Ed. 1995. Monit. bird Popul. by point counts. Gen. Tech. Rep. PSW-GTR-149. Albany, CA U.S. Dep. Agric. For. Serv. Pacific Southwest Res. Stn. p. 1-6 149.

Martinez Arbizu, P. (2020) pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4

Metcalf, O.C., Lees, A.C., Barlow, J., Marsden, S.J. & Devenish, C. (2020) hardRain: An R package for quick, automated rainfall detection in ecoacoustic datasets using a threshold-based approach. *Ecological Indicators*, **109**, 105793. https://doi.org/10.1016/j.ecolind.2019.105793.

Moura, N.G., Lees, A.C., Aleixo, A., Barlow, J., Berenguer, E., Ferreira, J. et al. (2016) Idiosyncratic responses of Amazonian birds to primary forest disturbance. *Oecologia*, **180**, 903–916. https://doi.org/10.1007/s00442-015-3495-z.

Moura, N.G., Lees, A.C., Andretti, C.B., Davis, B.J.W., Solar, R.R.C., Aleixo, A. et al. (2013) Avian biodiversity in multiple-use landscapes of the Brazilian Amazon. *Biological Conservation*, **167**, 339–348. https://doi.org/10.1016/j.biocon.2013.08.023.

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Parker, T.A. III, Stotz, D.F. & Fitzpatrick, J.W. (1996). Ecological and distributional databases. In: Stotz, D.F., Fitzpatrick, J.W., Parker, T.A. & Moskovits, D.K. III (Eds.), *Neotropical birds, ecology and conservation*. University of Chicago Press, Chicago, Illinois, pp. 113–460.

Pieretti, N., Duarte, M.H.L., Sousa-Lima, R.S., Rodrigues, M., Young, R.J. & Farina, A. (2015) Determining temporal sampling schemes for passive acoustic studies in different tropical ecosystems. *Tropical Conservation Science*, **8**, 215–234. https://doi.org/10.1177/194008291500800117.

Priyadarshani, N., Marsland, S. & Castro, I. (2018) Automated birdsong recognition in complex acoustic environments: a review. *Journal of Avian Biology*, **49**(5), jav-01447. https://doi.org/10.1111/jav.01447.

Putz, F.E. & Redford, K.H. (2010) The importance of defining "Forest": Tropical forest degradation, deforestation, long-term phase shifts, and further transitions. *Biotropica*, **42**, 10–20. https://doi.org/10.1111/j.1744-7429.2009.00567.x.

Robinson, W.D., Brawn, J.D. & Robinson, S.K. (2000) Forest bird community structure in central Panama: Influence of spatial scale and biogeography. *Ecological Monographs*, **70**, 209–235.

Robinson, W.D. & Curtis, J.R. (2020) Creating benchmark measurements of tropical forest bird communities in large plots. *Condor*, **122**(3), duaa015. https://doi.org/10.1093/condor/duaa015.

Robinson, W.D., Lees, A.C. & Blake, J.G. (2018) Surveying tropical birds is much harder than you think: a primer of best practices. *Biotropica*, **50**(6), 846–849. https://doi.org/10.1111/btp.12608.

Shonfield, J. & Bayne, E.M. (2017) Autonomous recording units in avian ecological research: current use and future applications. *Avian Conservation and Ecology*, **12**(1), art14. https://doi.org/10.5751/ACE-00974-120114.

Smith, D.G., Truskinger, A., Roe, P. & Watson, D.M. (2020) Do acoustically detectable species reflect overall diversity? A case study from Australia's arid zone. Remote Sens. *Ecological Conservation*, **6**(3), 286–300. https://doi.org/10.1002/rse2.173.

Snow, B.K. (1961) Notes on the behavior of three Cotingidae. *The Auk*, **78**, 150–161. https://doi.org/10.2307/4082129.

Stowell, D. & Sueur, J. (2020) Ecoacoustics: acoustic sensing for biodiversity monitoring at scale. Remote Sens. *Ecological Conservation*, **6**(3), 217–219. https://doi.org/10.1002/rse2.174.

Sugai, L.S.M., Desjonquères, C., Silva, T.S.F. & Llusia, D. (2020) A roadmap for survey designs in terrestrial acoustic monitoring. *Remote Sensing in Ecology and Conservation*, **6**(3), 220–235. https://doi.org/10.1002/rse2.131.

Sugai, L.S.M., Silva, T.S.F., Ribeiro, J.W. & Llusia, D. (2019) Terrestrial passive acoustic monitoring: review and

perspectives. *BioScience*, **69**(1), 15–25. https://doi.org/10.1093/biosci/biy147.

Terborgh, J., Robinson, S.K., Parker, T.A., Munn, C.A. & Pierpont, N. (1990) Structure and organization of an Amazonian forest bird community. *Ecological Monographs*, **60**, 213–238. https://doi.org/10.2307/1943045.

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

Venier, L.A., Holmes, S.B., Holborn, G.W., Mcilwrick, K.A. & Brown, G. (2012) Evaluation of an automated recording device for monitoring forest birds. *Wildlife Society Bulletin*, **36**, 30–39. https://doi.org/10.1002/wsb.88.

Verner, J. & Ritter, L.V. (1986) Hourly variation in morning point counts of birds. *The Auk*, **103**, 117–124. https://doi.org/10.1093/AUK/103.1.117.

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

Yip, D.A., Leston, L., Bayne, E.M., Sólymos, P. & Grover, A. (2017) Experimentally derived detection distances from audio recordings and human observers enable integrated analysis of point count data. *Avian Conservation and Ecology*, **12**, 73–84. https://doi.org/10.5751/ace-00997-120111.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Material