Received: 16 November 2020 | Accepted: 20 January 2021
DOI: 10.1111/2041-210X.13571

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Survey coverage, recording duration and community composition affect observed species richness in passive acoustic surveys

Connor M. Wood[1] | Stefan Kahl[1] | Philip Chaon[2] | M. Zachariah Peery[3] | Holger Klinck[1]

[1]Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, USA

[2]San Jose State Research Foundation, San Jose, CA, USA

[3]Department of Forest and Wildlife Ecology, University of Wisconsin – Madison, Madison, WI, USA

Correspondence
Connor M. Wood
Email: cmw289@cornell.edu

Funding information
Cornell Lab of Ornithology; California Department of Forestry and Fire Protection; U.S. Forest Service Region 5

Handling Editor: Camille Desjonquères

Abstract

1. Bioacoustic assessments of species richness are rapidly becoming attainable, but uncertainty regarding the optimal acoustic survey design remains. Selecting the duration of recording and the number of recording units are critical decisions, and we used both simulated and empirical data to quantify the trade-offs those choices present.
2. We evaluated the performance of 30 hypothetical acoustic survey designs (e.g. continuous recording, every other 5 min, etc.). Simulated bird species' ($n \leq 60$) abundance across the study area, probability of daily availability and time-dependent probability of vocal activity varied randomly within ranges of realistic values. Field data, collected in central New York, USA (747 hr) and in the northern Sierra Nevada, USA (1,090 hr), was analysed with a novel machine-learning algorithm, BirdNET. All three datasets were subsampled at 5-min intervals, observed species richness was compared across survey designs, and detection probability was calculated for each species.
3. Observed species richness increased with survey coverage (number of recording units) and with recording duration in all three datasets. The impact of differences in survey coverage decreased as recording duration decreased. Species' detection probabilities were negatively affected by reducing the number of days of recording and by reducing the daily recording duration. The more rare species a community had, the more species richness was underestimated as survey coverage decreased. Rarefaction curves indicated that increasing recording time has diminishing marginal utility but that the asymptote varies among communities. The cost per species observed decreased with increasing recording duration.
4. Discontinuous and reduced-coverage sampling may still yield fairly accurate assessments of biodiversity but reducing recording duration or coverage will result in different species remaining undetected. Whether the performance of a study design is 'good' or 'bad' depends on researchers' constraints and scientific questions to be answered. More hardware and longer recording durations are not always better, but we caution researchers against doing the bare minimum required for their present needs without pressing financial reasons to do so.

KEYWORDS
artificial intelligence, avian diversity, bioacoustics, bird community, convolutional neural network, ecoacoustics, machine learning, point count

1 | INTRODUCTION

Passive acoustic monitoring (PAM) is rapidly becoming a widespread and effective avian survey technique (Darras et al., 2018; Shonfield & Bayne, 2017; Wood, Popescu, et al., 2019). Data collected with autonomous recording units (ARUs) consistently yield diversity estimates comparable to those generated by trained observers (Darras et al., 2018), and falling hardware costs are making them widely accessible (e.g. Hill et al., 2018). Efficient signal classification, the process of identifying animal vocalizations in long-term soundscape recordings and attributing them to the correct species, remains a challenge when the volume of data is high, the soundscape is complex (e.g. low signal-to-noise ratio, many overlapping vocalizations, etc.), or when researchers desire community-level data. Yet recent advances in machine learning, particularly convolutional neural networks, are enabling the rapid identification of hundreds of species from audio archives (Kahl et al., 2021; LeCun et al., 2015). One likely result of readily accessible hardware and improving software is the broader use of bioacoustics as a research tool for community ecology (Wood et al., 2019). A critical first step in community studies – and indeed in ecology more broadly – is measuring species richness. Sampling curves – both accumulation and rarefaction – are long-standing tools for this task (Gotelli & Colwell, 2001; Soberon & Llorente, 1993; Thompson & Withers, 2003). In principle, observed species richness increases with survey effort towards an asymptote of true species richness (Soberon & Llorente, 1993). Bioacoustic survey effort can vary substantively in both spatial extent and recording duration (Sugai et al., 2020), which presents a fundamental question: how many ARUs are needed, and how long do they need to record to accurately and efficiently measure avian diversity?

Determining an appropriate level of survey effort entails trade-offs between the resolution and accuracy of the data and the costs and logistical complexity of collecting it. The number of ARUs required will be affected by the biology of the focal species and the research goals, among other factors. For example, landscape-scale studies of species or guilds with large home ranges will require more survey locations (and thus ARUs and/or logistical effort) than fine-scale work of species with small home ranges (e.g. compare Wood et al., 2020 and Campos-Cerqueira & Aide, 2016). Researchers interested in measuring biodiversity change through time should conduct power analyses because starting population sizes, detection probabilities (which themselves are affected by the recording duration), the magnitude of population change researchers wish to detect, and the time horizon can affect statistical power (Wood, Popescu, et al., 2019). Similarly, the appropriate recording duration will also depend on the species and application. Species that make frequent, loud, stereotyped vocalizations are more likely to be detected quickly than relatively cryptic species, though this relationship is also affected by home range size and thus the probability of a target species being in range of an ARU at any given time (e.g. compare Wood et al., 2020 and Furnas, 2020). Using PAM to collect detection/non-detection data for occupancy analyses will allow for shorter recording durations because such models require only one true positive observation per survey. In contrast, behavioural data such as variation in vocal activity rates among sites is improved by extending recording durations (Pérez-Granados et al., 2019; Wood et al., 2020). However, empirically based guidelines for determining the appropriate level of survey effort to measure the species richness of bird communities remain scarce.

Recent work supports the intuitive conclusion that more ARUs recording for longer durations will be more likely to record all vocally active species in a given area (de Camargo et al., 2019). However, buying ARUs, deploying them, and analysing and storing large quantities of audio data is expensive and time consuming. Therefore, designing studies that are sufficient to yield diversity estimates that are suitable for researchers' needs will help researchers make the best use of available resources. We addressed this issue by comparing known species richness to that observed under different combinations of recording durations and sampling intensities using simulated data and two field studies. For the simulations, we based species' abundance, daily availability and patterns of vocal activity on empirically estimated values from bird communities globally. For the field data, we applied a novel machine learning algorithm capable of identifying >90% of North American bird species (Kahl et al., 2021) to two PAM datasets collected in spring 2018. Insights from these three cases may help researchers optimize acoustic survey designs to accurately assess species richness without excess survey effort, thus increasing the efficiency of PAM programs and conservation more broadly.

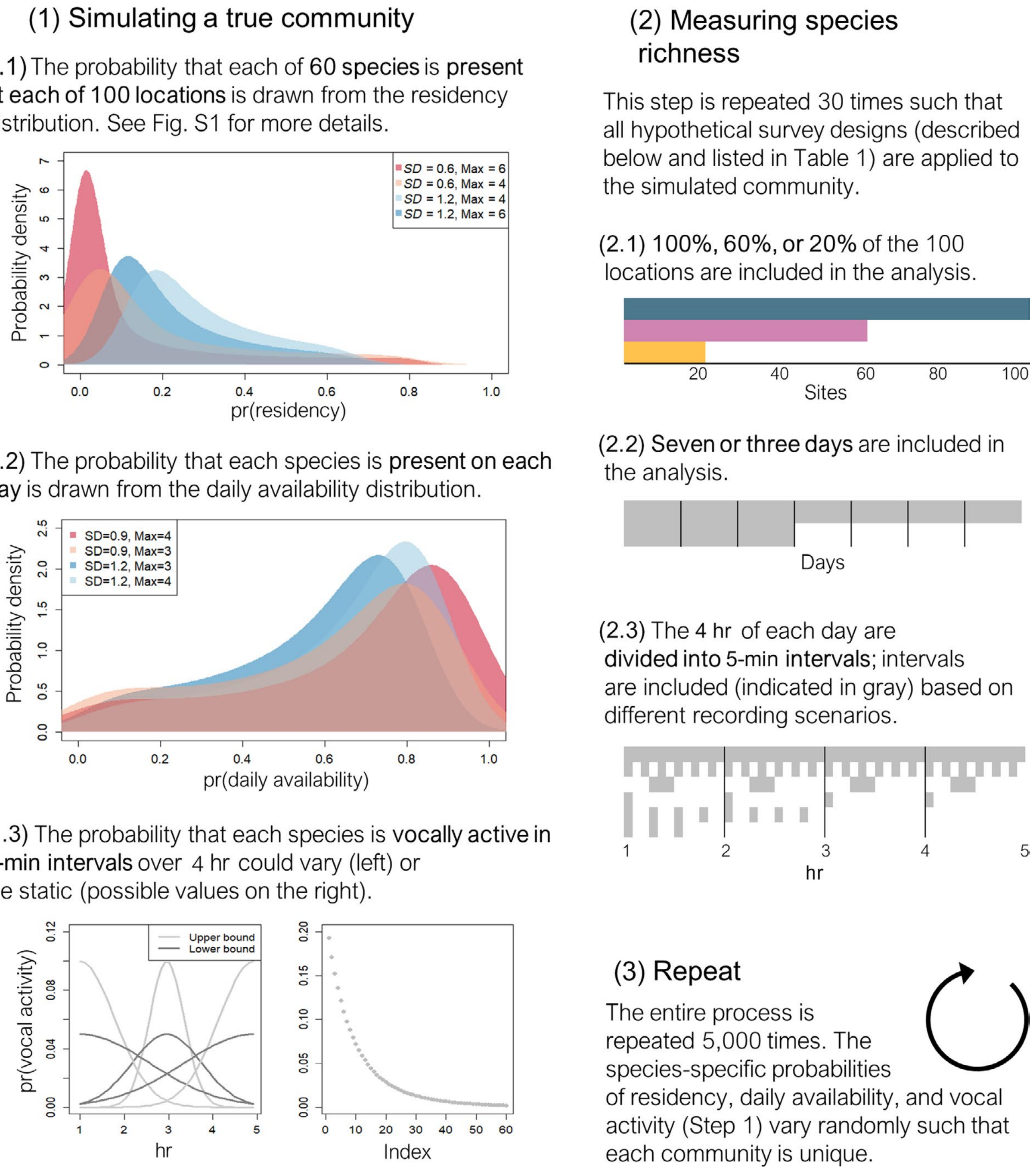
2 | MATERIALS AND METHODS

2.1 | Survey design scenarios

We evaluated the efficacy of 10 recording duration scenarios and three ARU deployment scenarios (Table 1, Figure 1). We considered 7- and 3-day surveys for each of four recording settings: continuous recording for 4 hr (4 hr total/day), every other 5 min for 4 hr (2 hr total/day), second 15 min of every hour for 4 hr (1 hr total/day), and the first 5 min of each of 4 hr (20 min total/day). We also considered two 1-day surveys representing audio-only point count transects: recording for 5 of every 15 min for a total of either 120 or 45 min. The 4-hr baseline reflects the fact that diurnal avian species are most vocally active in the hour before and 3 hr following sunrise

TABLE 1 The hypothetical survey design choices that were considered in the simulations and that were applied to the field data when applicable. Each day had a total of 4 hr of simulated avian vocal activity

Dataset	Recording durations	Days (total hours)		Deployment
All three	Continuous	7 (28)	3 (12)	100%, 60%, and 20% of locations
	Every other 5 min	7 (14)	3 (6)	
	Second 15 min/hr	7 (7)	3 (3)	
	First 5 min/hr	7 (2.33)	3 (1)	
Simulation only	Five/15 min for 2 hr	1 (0.67)		
	Five/15 min for 0.75 hr	1 (0.25)		


(1) Simulating a true community

(1.1) The probability that each of **60 species** is **present at each of 100 locations** is drawn from the residency distribution. See Fig. S1 for more details.

(1.2) The probability that each species is **present on each day** is drawn from the daily availability distribution.

(1.3) The probability that each species is **vocally active in 5-min intervals** over 4 hr could vary (left) or be static (possible values on the right).

(2) Measuring species richness

This step is repeated 30 times such that all hypothetical survey designs (described below and listed in Table 1) are applied to the simulated community.

(2.1) **100%, 60%, or 20%** of the 100 locations are included in the analysis.

(2.2) **Seven or three days** are included in the analysis.

(2.3) The **4 hr** of each day are **divided into 5-min intervals**; intervals are included (indicated in gray) based on different recording scenarios.

(3) Repeat

The entire process is repeated 5,000 times. The species-specific probabilities of residency, daily availability, and vocal activity (Step 1) vary randomly such that each community is unique.

FIGURE 1 A conceptual overview of the three steps of the analyses. A true bird community was simulated (1), its species richness was measured according to various survey designs (2), and the process was repeated (3). For the two empirical datasets, >1,800 hr of empirical data replaced step 1, and the subsampling process (step 3) was only repeated 500 times (de Araújo et al., 2020; Robbins, 1981; Wimmer et al., 2013). These design choices assume a breeding season research effort in which measurements of species richness are the primary interest; studies with a phenological emphasis – either diel or seasonal – are likely to face very different temporal constraints.

We also considered scenarios in which 100%, 60% and 20% of the study area was sampled (Figure 1). The simulation assumed no intraspecific competition (i.e. the probability of a species being present at a given site was not affected by its presence at other sites), so the spatial grain of the simulation is equal to the largest home range of the species in the regional pool. For example, 60% coverage means that ARUs were deployed in 60% of the possible home ranges of that largest-bodied species. That less than 60% of the home ranges of the smallest-bodied species would be sampled under such a scenario is also true of real-world acoustic surveys and is a separate issue that is beyond the scope of this paper (i.e. without a potentially complex hierarchical survey design, the proportion of the population that is surveyed will vary based on home range size and territoriality).

2.2 | Simulations

2.2.1 | Simulating communities and species

In each of 5,000 iterations, we simulated communities of up to 60 species surveyed at up to 100 sites for a maximum of 7 days for up to 4 hr per day (Figure 1). There were three components of each simulated bird community: determining species' abundance across the study area, determining species' probability of daily availability, and determining species' vocal activity patterns (whether and how vocal activity varied).

First, we determined whether each of up to 60 species (*n*) was a resident at each of the 100 sites. Species abundance curves generally conform to a lognormal distribution, though some are log-left-skewed (Gaston & Fuller, 2007; McGill, 2003); that is, there are many rare species and few common species in any given community. Therefore, the probability that each species was a resident at any location was derived from a lognormal distribution with a mean = 0 and standard deviation (*SD*) that varied randomly between [0.6–1.2] among iterations. We extracted 72 probabilities (i.e. 1.2 × *n*) from the complement of the cumulative probability distribution function of that distribution in the range 0.6–[4–6]; drawing from a truncated range of values eliminated that possibility of ubiquitous or extremely rare species (i.e. pr(site residency) >0.95 or <0.01) (Figure 1, 1.1; see Figure S1 for further detail). From that resulting vector of probabilities, we randomly drew *n* values without replacement to determine the range of probabilities of site residency for each community (i.e. iteration of the simulation).

Second, we determined the probability of daily availability for each species at the recording sites. The probability that an individual is within the listening range of an ARU within its home range on any given day is inversely related to its home range size. Home range size scales with body size (Haskell et al., 2002), and body size in birds follows a right-skewed log distribution (Blackburn & Gaston, 1994). Therefore, the probability that a resident species was present on a given day was derived from a left-skewed lognormal distribution with a mean = 0 and a randomly selected *SD* that varied between [0.9–1.2] among iterations. As before, we drew 1.2 × *n* values from 0.06–[4–6] of the complement of the cumulative distribution function of that distribution and sampled them without replacement to determine the distribution of daily availability probabilities for each community (Figure 1, 1.2). Each species' probability of daily availability was constant across days within in each iteration, and whether a species was available for detection at a site at which it was a resident was a Bernoulli trial based on that probability. Importantly, the volume (i.e. 'loudness' or amplitude) and frequency (i.e. 'pitch') of a species vocalization will also affect their daily availability. Vocalizations that are louder and of a lower frequency will propagate farther, thus effectively making the species in range of the recording unit more often. In this way, although the geographic home range of a species and its vocal characteristics may not be correlated, they are confounded, both in the simulation and in real life.

Third, each species was assigned a probability of vocal activity for each 5-min interval of the day. Species have different intrinsic vocal activity rates; even when they are present, they may not vocalize or their vocalizations may not be recorded with sufficient clarity to be correctly classified. Avian vocal activity peaks within 4 hr of dawn; within that period, some species' vocal activity peaks early, fewer species' activity peaks later in that period, and many species have a uniform – generally low or moderate – vocal activity in that period (Robbins, 1981). We allowed the strength and prevalence of these patterns to vary among species and among communities (i.e. iterations of the simulation). In each iteration, species would have a 10%–30% chance of being an early caller and a 5%–15% chance of being either a middle or late caller; the specific value for each type was a random draw from a uniform distribution and was fixed in each iteration. There was thus a 40%–80% chance that species had uniform vocal activity. Variable activity rates were drawn from the probability density functions of normal curves with a mean = 0 and *SD* that varied randomly [4–8] among species and iterations; early-peaking species sampled that curve from [0–20], mid-peaking species from [−20–20] and late-peaking species from [−20–0] (Figure 1, 1.3). The probability of vocal activity per interval of uniformly active species was drawn from the complement of the cumulative probability function of a lognormal distribution with a mean = 0, *SD* = 0.8 between the values [2, 10] (Figure 1, 1.3). If a resident species was present on a given day, its observed vocal activity in each interval was a Bernoulli trial based on its assigned probability of vocal activity for that interval.

2.2.2 | Outputs and assumptions

The outcome of each iteration of the simulation was a bird community whose species abundance distribution, daily availability distribution and the probability of vocal activity at different times were stochastically generated within a range of realistic values. At each of the 100 sites, species richness was recorded in 336 intervals (7 days × 4 hr × twelve 5-min intervals/hr). We then measured species richness at different spatial and temporal resolutions corresponding with the 30 hypothetical study designs (see Section 2.1; Table 1; Figure 1, 2.1-3), yielding sample-based rarefaction curves (Gotelli & Colwell, 2001). We also calculated each species' detection probability (p, the probability that it was observed given that it was present) by calculating the average of the encounter histories at all sites at which it was present. Using p, which treated each interval as a secondary sampling period, we could calculate the seasonal detection probability (p^*) for each species, or the probability that it was detected at least once throughout the survey ($p^* = 1 - (1 - p)^{\text{number of intervals}}$) under scenarios with varying numbers of intervals. Whether a resident species was available for detection on a given day and whether it was vocally active were confounded, but the two processes are also confounded in real-world studies. The simulation code is publicly available (Supporting Information) and was written for program R (R Core Development Team, 2014). Users can easily modify the number of sites, species, days, hours and recording intervals, as well as the community composition and vocal behaviour of the species.

We made several simplifying assumptions. Species' probability of residency was constant across sites in each iteration of the simulation, meaning that the species present at each site would be a random sample of the overall community. In ecological terms, it suggests a lack of environmental heterogeneity. We assumed that each species' vocal activity pattern was constant across days within each iteration. Thus, events like rain, high wind or high humidity, all of which are known to affect species vocal activity and sound propagation (and thus ARU performance), were not included. In areas where these and other similar events are common, researchers should consider modifying the code to include some probability of a day-long, community-wide reduction in the probability of vocal activity or simply extending the real-world duration of ARU deployments to buffer against them. We assumed site closure during the 7-day simulation, meaning that individuals did not die or emigrate, though the probability of resident species' daily availability did vary. Finally, we did not explicitly incorporate variation in habitat, which means that the results are most applicable to studies focused on a single ecologically cohesive area. The probability of site residency varied among species (though it was constant across sites for each species), which allows for the possibility of environmental heterogeneity and habitat specialization within the given region.

2.3 | Case studies and machine learning

In both case studies, we deployed ARUs (Swift recorder, Cornell Lab of Ornithology Center for Conservation Bioacoustics, Ithaca, NY, USA) that recorded with one omni-directional microphone. We then extracted bird species identifications from the raw audio data with BirdNET, a novel machine learning algorithm capable of identifying 984 North American and European bird species by sound (Kahl et al., 2021), and used the resulting biodiversity data to evaluate the performance of the recording scenarios described above (Section 2.1; Table 1).

BirdNET was trained on ~1.5 million 3-s audio clips extracted from focal recordings (i.e. recordings made with directional microphones that typically have a high signal-to-noise ratio) provided by two major community collections, Xeno-canto and the Macaulay Library. Prior to applying BirdNET to our data, we used several hours of fully annotated data from each study area to optimize program settings (primarily the shape of the sigmoid curve used in the activation function) to maximize precision. When we applied BirdNET to the full passively recorded audio datasets, the continuous recordings were split into consecutive 3-s chunks to match the input size of the BirdNET model. The output feature vector for every chunk contained confidence scores for all species that are known to occur at the recording location. In a post-processing step, these scores were smoothed and pooled with a moving mean exponential average with a width of three chunks (i.e. 9 s) for each species. This reduced false negatives that can arise when vocalizations are cropped during the 'chunking' process, while also filtering out one-off high-confidence false positives. BirdNET achieves an overall precision of about ~0.8 for foreground species in focal recordings (Kahl et al., 2021). However, scores drop considerably when applied to omnidirectional soundscape data, which often have poor signal to noise ratio relative to focal recordings. Therefore, we eliminated all detections that had an average confidence score below 0.5 such that the data used for the species richness analyses contained only highly likely true positives (Figure S2). The threshold of 0.5 was selected to maximize precision after a review of ~100 randomly selected hours of results that had been validated by expert birders (i.e. all of BirdNET's classifications were marked as true or false). This process of expert validation also revealed that false positives were extremely rare (<2% of all detections), indicating that our restrictive post-processing scheme maximized precision sufficiently for reliable analyses. Reduced recall is a common consequence of increased precision, and the increased false negatives can be accounted for by explicitly modelling imperfect detection, which is why we conducted the detection probability analyses described above.

2.3.1 | Central New York

We deployed 28 ARUs in the 0.9-km^2 Sapsucker Woods Sanctuary in Ithaca, New York, USA; units were deployed such that uniform coverage of the study area was achieved and all units were ≥250 m apart. Audio was recorded at a sample rate of 48 kHz and with 16 bits resolution from 05:00 to 09:00 on May 29 to June 4, 2018 (sunrise occurred at 05:30), which yielded 747 hr of data (37 files were truncated during battery checks and other maintenance). The study area contained stands of mixed hardwoods (sugar maple *Acer saccharum*, red maple *A. rubrum* and beech *Fagus grandifolia*), conifers (white pine *Pinus strobus* and eastern hemlock *Tsuga canadensis*), a seasonal marsh, and a pond; understory vegetation varied from open stands of mature maples to dense shrubs. After post-processing the BirdNET results, we had 118,868 detections (Figure S2 shows the distribution of detections before and after filtering based on the confidence score threshold). Sapsucker Woods is an eBird hotspot, so we downloaded user-submitted data from the same period and compared the species lists generated by (a) PAM and BirdNET and (b) volunteer birders.

2.3.2 | Northern Sierra Nevada

We conducted passive acoustic surveys in the Lassen and Plumas National Forests in May – August of 2018. Survey grid cells (4 km^2) were randomly selected from a ~6,000-km^2 area, ARUs were deployed at acoustically advantageous locations (e.g. ridges rather than gullies) within those cells, and we selected 38 ARUs from across the entire area such that units were ≥7 km apart. Audio was recorded a sample rate of 32 kHz and with 16 bits resolution from 04:00 to 08:00 for 5–7 days between May 9 and June 10 (sunrise was roughly 05:35–05:50 during that time), which yielded 1,090 hr of audio. The study area was dominated by montane Sierran mixed conifer forest (white fir *Abies concolor*, Douglas-fir *Pseudotsuga menziesii*, ponderosa pine *P. ponderosa*, sugar pine *P. lambertiana*, incense-cedar *Calocedrus decurrens* and California black oak *Quercus kelloggii*); understory vegetation ranged from very open stands of mature spruce to dense woody shrubbery. After post-processing the BirdNET results, we had 114,011 detections.

2.4 | Analysis

With the simulated data and two field datasets, we compared the differences between observed species richness among survey designs. With 5,000 data points in each pairwise comparison of simulated data, standard probabilistic tests (e.g. one-way ANOVAs and Tukey's HSD) found differences in mean species richness of just three species (of a possible 60) between groups to be significant ($\alpha = 0.05$). We felt that this arbitrary – albeit widely accepted – standard was not informative; whether the difference between the two study designs is 'significant' depends on the application and the user's tolerance for type I errors. Therefore, for each study design, we reported the mean number of species observed, mean percent of the community missed, and the pairwise differences in the number of species missed between study designs. We subsampled the field data 500 times to determine observed species richness at 60% and 20% coverage. We excluded the acoustic point count transects from our comparisons among simulated survey designs (yielding 24 survey designs). We compared the acoustic point count transect scenarios ($n = 6$) to each other and to the full recording scenario (4 hr/day for 7 days).

In reporting species richness at the scale of the entire study, we are measuring γ-diversity, the overall species pool, rather than α-diversity, or species richness observed at any one ARU. However, because we did not simulate habitat differences among sites, β-diversity (i.e. γ/α) is expected to be low and to vary randomly among sampling locations. Therefore, the findings of the 100% coverage scenarios are transferrable to measuring α-diversity.

We conducted three additional analyses with the simulated data. First, we compared the range of seasonal detection probabilities for each community (i.e. iteration of the simulation) across survey designs to assess how survey design choices negatively affected detection relative to maximum survey coverage (i.e. recording 4 hr/day for 7 days). Second, we conducted a sensitivity analysis using linear regression to determine whether simulation parameters defining species probability of residency (i.e. abundance; standard deviation and maximum quantile; see Section 2.2.1) affected the difference between true and observed species richness. Third, we calculated the approximate cost per species observed using three basic assumptions. First, we assumed that audio was recorded at a sample rate of 32 kHz and was stored in .flac format, yielding 100 MB/hr. Second, we assumed that data were stored on the cloud for $0.023 USD/GB (https://aws.amazon.com/s3/pricing/; accessed 17 Dec. 2020). Although external hard drives or network-assisted storage (NAS) devices could also be used, neither are as reliable as cloud-based options. Third, we assumed that it costs $25 USD per 10 ARUs deployed, though we recognize that these costs could vary by orders of magnitude. For example, the Sierra Nevada data (collected across ~6,000 km^2) required field technicians, rental vehicles and field housing, while the Central New York data (collected across 0.9 km^2) could be collected by existing personnel without vehicles.

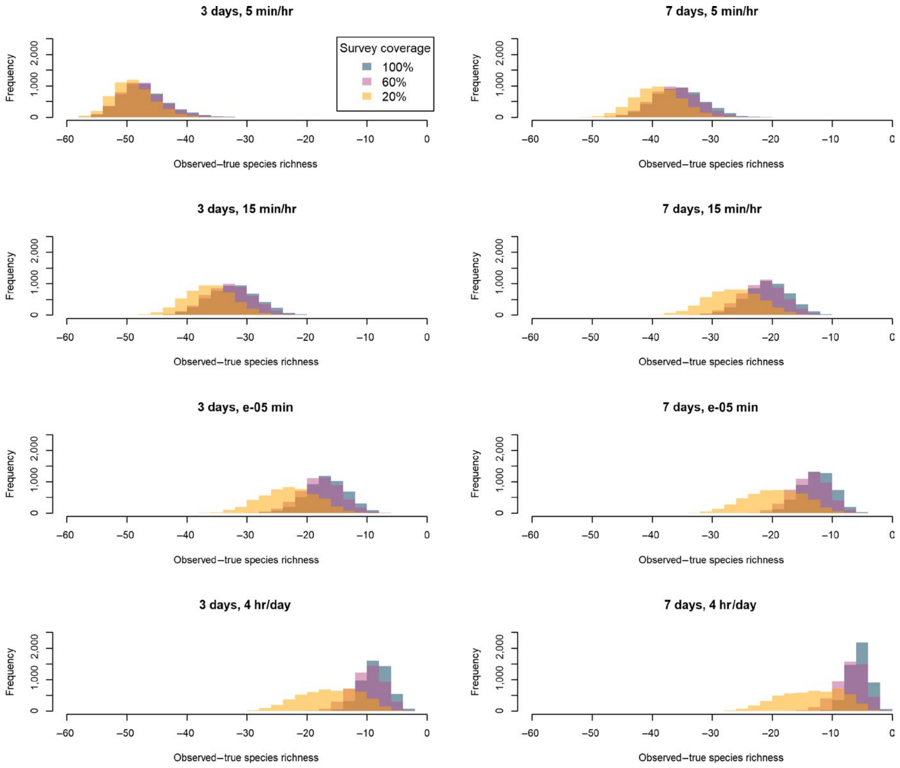
3 | RESULTS

3.1 | Simulations

Increasing the number of days on which recording occurred, increasing the recording time per day, and increasing the number of ARUs all increased the observed species richness (Figures 2 and 3, Table 2). However, these factors contributed unequally: observed species richness decreased substantially as recording duration decreased (Table 2) but decreased only slightly between 60% and 100% survey coverage (Figure 2). At full survey coverage, the full recording duration (7 days, 4 hr/day; 28 total hours of recording) missed an average of five species (an average of 9% of the community) (Table 2). Reducing recording time by 50% (every other 5 min over 7 days) resulted in an average of 3.2 additional species not being detected. In contrast, reducing the recording time by 63% (4 hr/day for just 3 days) resulted in an average of seven additional species not being detected (Table 2). Recording for just fiv5e minutes per hour for 3 days yielded 1 hr of audio data and documented just 20% of the community (Table 2).

Species' detection probabilities were negatively affected by reducing the number of days of recording and by reducing the daily recording duration (Figure 4). In the baseline scenario of full recording (7 days, 4 hr/day) seasonal detection probabilities are uniformly

FIGURE 2 Differences between observed and true species richness in simulated bird communities ($n = 5{,}000$) based on different hypothetical acoustic survey designs (see Table 1)



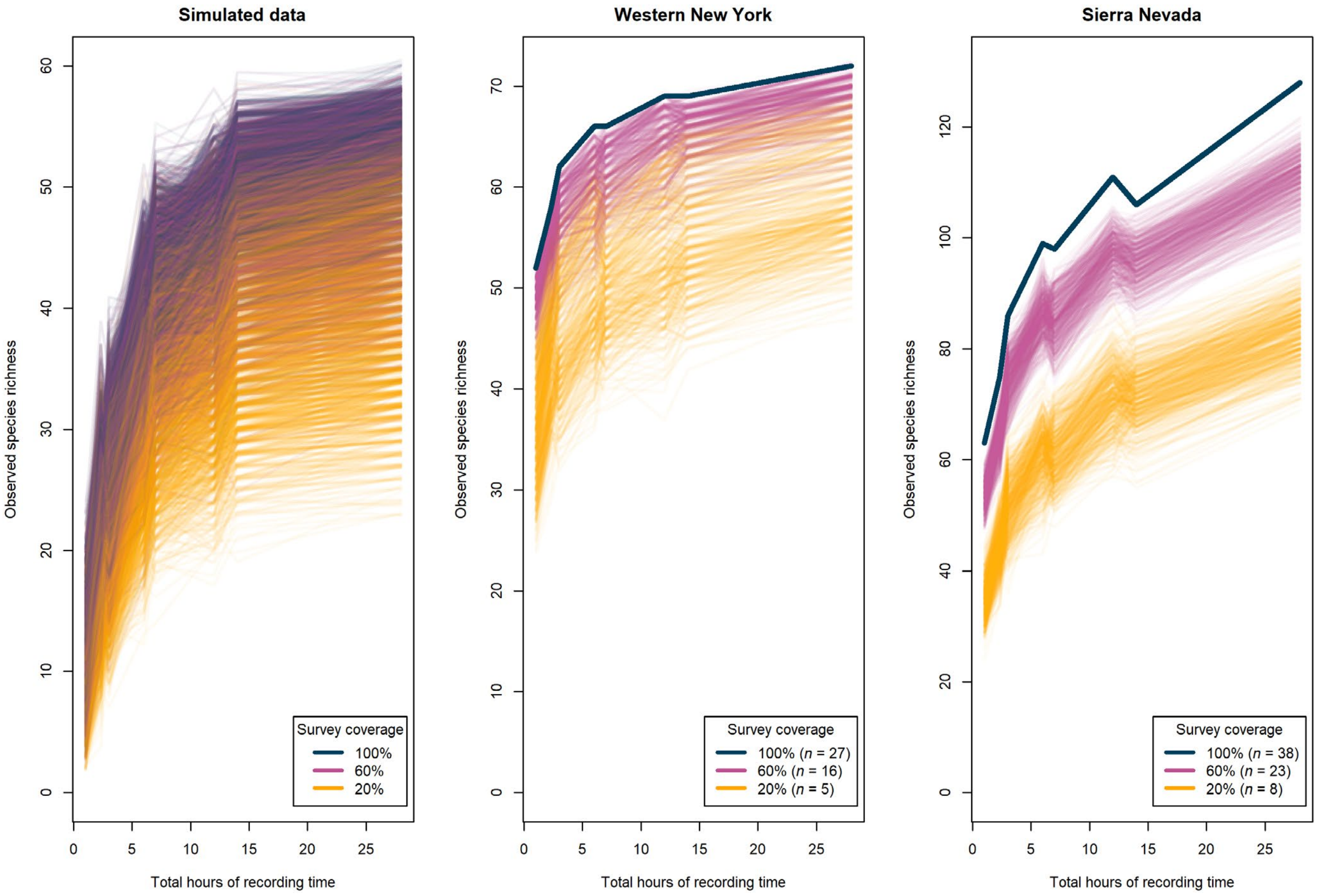


FIGURE 3 Observed species richness increased unevenly as total recording duration increased (see Table 1 for recording scenarios). All sites ($n = 100$) in the simulations ($n = 5{,}000$) were equivalent and there were no associations among species. The Sapsucker Woods surveys covered 0.9 km^2; the Sierra Nevada surveys covered ~6,000 km^2. The 60% and 20% coverage scenarios of the two field datasets were achieved by randomly sampling the full dataset 500 times without replacement

TABLE 2 Differences (row − column) in observed species richness between survey designs (simulated data, $n = 5{,}000$ iterations; 100% survey coverage). 'e-o' stands for every-other; cell colour reflects performance (red is worse, blue is better) [Correction note added on 09 March, 2021, after first online publication: The data in Table 2 has been amended]

	3 days, 5 min/hr	7 days, 5 min/hr	3 days, 15 min/hr	3 days, e-o 5 min	7 days, 15 min/hr	3 days, 4 hrs/day	7 days, e-o 5 min	7 days, 4 hrs/day
Total recording hours	1	2.33	3	6	7	12	14	28
Mean species observed	11.8	23.5	26.7	38.3	42.2	46.4	50.3	53.4
Mean percent of community missed	80%	60%	54%	34%	28%	21%	14%	9%
7 days, 4 hrs/day	−41.6	−30.0	−26.7	−15.1	−11.2	−7.0	−3.1	
7 days, e-o 5 min	−38.5	−26.8	−23.6	−12.0	−8.0	−3.9		
3 days, 4 hrs/day	−34.6	−23.0	−19.7	−8.1	−4.2			
7 days, 15 min/hr	−30.4	−18.8	−15.5	−4.0				
3 days, e-o 5 min	−26.5	−14.8	−11.6					
3 days, 15 min/hr	−14.9	−3.3						
7 days, 5 min/hr	−11.6							

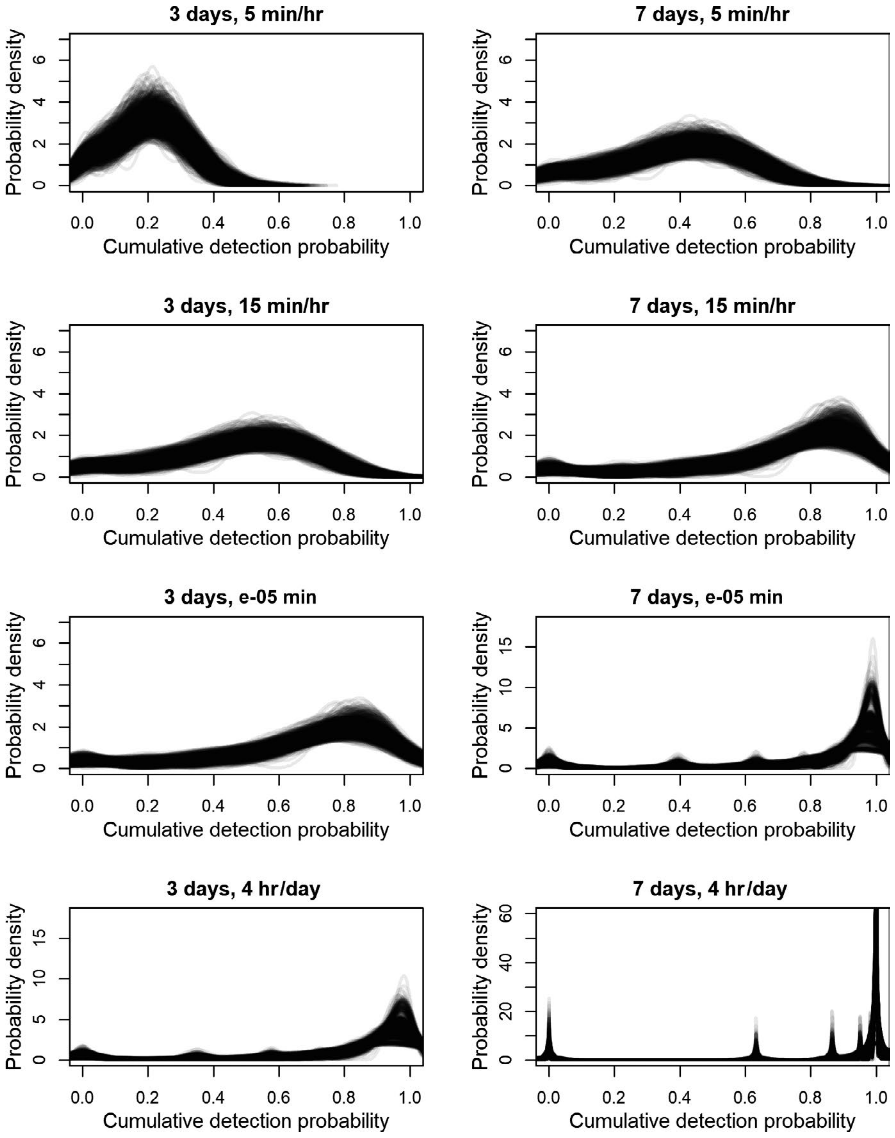


FIGURE 4 Cumulative detection probability (p^*), the probability of observing a species at least once given that it is present, increased with the number of hours of recording per day and the number of days of recording. Each line represents the distribution of detection probabilities in a simulated community ($n = 5{,}000$). Note that the y-axis scale changes in the bottom two rows

high (>0.9; Figure 4, bottom right); departures from this distribution represent avoidable decreases in detection probabilities.

Acoustic point count transects (surveying 5 of every 15 min) underestimated overall species richness by 86%–95% (Table S1). Survey coverage of the landscape influenced observed species richness less than the amount of recording time. Conducting longer transects (eight surveys over two hours) over just 20% of the study area yielded more than twice as many species as conducting shorter transects (three surveys over 45 min) across the entire study area.

Community structure influenced study design performance. The more rare species a community had, the more species richness was underestimated as survey coverage decreased. At full coverage, the difference in observed species richness between a community with the fewest possible rare species (low $SD_{\text{residency}}$ and high maximum quantile; see Section 2.2.1 and Figure 1) and the most was just one species (observed richness = −4.85 − 1.17 × *SD* + 0.13 × max.Q, $F = 44.3$, $df = 2$ and 4,997, $p < 0.001$). At 20% coverage, the difference in observed species richness between a community with the fewest and most rare species was −18.5 species (observed richness = −20.08 + 22.22 × *SD* − 2.56 × max.Q, $F = 4{,}944.3$, $df = 2$ and 4,997, $p < 0.001$). The $SD_{\text{residencyy}}$ value explained 58% of the variation in the observed species richness data at 20% coverage and full recording duration. The measurable effects of daily availability on study design performance were much less substantial because daily availability was confounded with but independent of vocal activity (see Section 2.2). At the full recording duration, the difference in observed species richness between a community with more and fewer large-bodied species (maximum daily availability quantile; see Section 2.2.1 and Figure 1) was 1.3 species (observed richness = −9.67 + 1.34 × max.Q; $F = 261.1$, $df = 1$ and 4,998, $p < 0.001$). At a substantially reduced recording duration (3 days, 15 min/hr/day), the difference in observed species richness between a community with the more and less rare species was 2.4 species (observed richness = −40.10 + 2.43 × max.Q; $F = 158.2$, $df = 1$ and 4,998, $p < 0.001$).

The most important pattern was that the cost per species observed decreased as the recording duration increased (Table S2), indicating that when the cost of deploying hardware is fixed and the cost of data storage is very low, false negatives (missed detections) accrued faster than the cost savings of generating less audio. Cost per species observed also decreased with survey coverage, but this finding is more sensitive to the assumptions of the cost calculations and the simulation (discussed in Section 4.1).

3.2 | Empirical data from Central New York and the Sierra Nevada

We detected 71 species in the Sapsucker Woods data and 129 species in the Sierra Nevada data (Table S3). Both datasets yielded the same general pattern as the simulated data: observed species richness decreased as total recording duration decreased (Tables S4 and S5). However, there were two consistent exceptions not observed in the simulated data: (a) recording every other 5 min over 4 hr for 7 days (14 hr) yielded fewer observed species than did recording for 4 hr for 3 days (12 hr), and (b) recording the second 15 min of each of 4 hr for 7 days (7 hr) yielded fewer observed species than did recording every other 5 min over 4 hr for 3 days (6 hr) (Figure 3). In both of those cases, more continuous recording over fewer days yielded more accurate results than slightly more recording time distributed across more days.

The rarefaction curves in Sapsucker Woods appeared to be approaching an asymptote by 28 hr of recording; the Sierra Nevada rarefaction curves were still increasing – albeit more slowly – by 28 hr of recording time (Figure 3). The two studies also differed in the importance of survey coverage. A 20% reduction in survey coverage in Sapsucker Woods led to relatively minimal changes in observed species richness (a loss of ~5 species, or about 9.7% of the community), indicating that the data provided by those recording units had been fairly redundant; observed species richness decreased more substantially – but variably – when survey coverage was reduced by 40%, indicating that α-diversity varied among those units and at some was fairly close to γ-diversity (Figure 3). Reducing survey coverage by 20% and 40% in the Sierra Nevada resulted in fairly consistent, substantial reductions in observed species richness (~13%–15% of the community with each reduction; Figure 3), indicating that α-diversity at any given ARU was moderate relative to γ-diversity and that β-diversity was more uniform among ARUs than at Sapsucker Woods.

In contrast to the simulated data, both empirical datasets generally showed slight increases in performance for 3 days/continuous (12 hr total) and 3 days/every-other 5 min (6 hr) scenarios compared to the 7 days/every other 5 min (14 hr) and 15 min per hour/7 days (7 hr) scenarios, respectively (Figure 3). This suggests that our simulation slightly overestimated the prevalence of species with low probabilities of daily availability and slightly underestimated the prevalence of species with very low probabilities of vocal activity.

eBird users in Sapsucker Woods reported 73 species, compared to the 71 identified by PAM and BirdNET (Table S3); a total of 90 species were observed. Predictable and widely reported biases (Darras et al., 2018) were present in both lists. For example, at least one Turkey Vulture *Cathartes aura* and Barred Owl *Strix varia* were reported by eBird users, but the former species is extremely quiet and the latter species' vocal activity is concentrated outside the sampling period, while common resident species (House Sparrow *Passer domesticus* and Rock Pigeon *Columbia livia* were not reported by eBird users.

4 | DISCUSSION

Determining the optimal passive acoustic survey design has been a persistent question in ecology and conservation (Balantic & Donovan, 2019; Cook & Hartley, 2018; Sugai et al., 2020; Wimmer et al., 2013), in part because buying equipment, conducing fieldwork and storing and processing many terabytes of audio can cost tens of thousands of dollars annually. Yet the proliferation of acoustic data and the emergence of machine learning algorithms capable of efficiently extracting community-level data (e.g. Kahl et al., 2021) has added urgency to this issue because novel ecological insights are tantalizingly attainable (de Camargo et al., 2019) and conservation challenges are increasingly complex and spatially extensive (Wood & Jones, 2019). Comparing species richness estimates among survey designs may also be useful when comparing or integrating results collected under different sampling schemes. These simulations and case studies show that discontinuous sampling may still yield fairly accurate assessments of biodiversity: a 50% reduction in recording duration may result in a 5%–17% decrease in the proportion of the observed community (Table 2, Tables S4 and S5). Yet, the spatial aspects of a study can substantially influence such outcomes.

Both sampling less of the landscape and decreasing recording durations decrease the number of species that are observed. Yet those choices will likely result in different species remaining undetected. Rabinowitz (1981) articulated seven ways in which species could be 'rare', and the classification is informative here. Species with small geographic ranges and/or narrow habitat requirements are more likely to be missed if fewer units are deployed. This is illustrated generally in the difference observed species richness between survey coverages (100%, 60% or 20% of the study area) when the same recording duration is used (Figures 2 and 3). Specifically, in the sensitivity of the results to the *SD* of the residency parameter in the simulations (Figure 1, 1.1) – the more communities were composed of rare species, the more biodiversity was underestimated in low survey coverage designs. In contrast, habitat generalists with low population densities, which often have large home ranges, are more likely to be missed if the number of recording days is reduced, regardless of survey coverage (Figure 2, Table 2). Reducing the daily recording durations will result in a bias against cryptic species with relatively low vocal activity. Further analyses could address some of these challenges explicitly by incorporating environmental heterogeneity into comparable power analyses, including differences in vocal activity rates and species composition driven by anthropogenic disturbances. Seasonal variability in vocal activity could be a very important factor in determining recording times for PAM projects that span many weeks or months, and species composition and vocal activity rates could also differ markedly in tropical ecosystems.

The simulated acoustic point count scenarios represent a potential application of mobile technology to citizen science projects. BirdNET has been released as a mobile app that currently supports opportunistic reports of almost one thousand North American and European bird species. Yet, such presence-only data present more analytical challenges than detection/non-detection data. Though the simulations show that acoustic point count transects substantially underestimate species richness (Table S1), such reports would enable much more robust analyses. Thus, the non-detections of many species would still be valuable, though the data would be more suitable, at least initially, for population trend analyses of more common species.

4.1 | How many units do I need?

Greater environmental heterogeneity in the study area will generally lead to greater β-diversity and thus increase the importance of surveying more of the landscape to measure γ-diversity accurately. Differences in survey coverage decrease in importance as the recording duration decreases because many species are already not being detected (Figure 2). The contrasting patterns among the rarefaction curves for Sapsucker Woods (0.9 km^2) and the Sierra Nevada (~6,000 km^2) illustrated this principle: reducing survey coverage affected observed species richness much more substantially in the Sierra Nevada, a vastly larger study area entailing much more variation in habitat. Basic principles of biogeography influence this outcome as well: more species were observed in the larger study area (MacArthur & Wilson, 1967). Thus, the larger the study area, the more reductions in survey coverage will affect the results. Of course, surveying larger areas requires more hardware (and results in more audio data requiring storage), but efficiencies can be achieved by rotating ARUs across the landscape (Wood et al., 2019). Moving ARUs back and forth between two sets of locations effectively doubles the amount of area surveyed, but it entails potentially significant logistical challenges. Allowing detection probability to vary with time could help account for bias that may be introduced if species' vocal activity changes dramatically between deployments, as could leaving some units in place for the entire time. Increasing microphone gain could increase the effective survey range of the ARUs, thus increasing species' probability of daily availability (see Section 2.2.1) and the overall amount of area surveyed, but the spatial benefits would be dwarfed by those of implementing a full or partial hardware rotation.

While 20% survey coverage yielded the lowest cost per (simulated) species observed, this finding is sensitive to logistical costs and environmental heterogeneity. As the cost per ARU deployed decreases, greater survey coverage will be increasingly efficient; those marginal costs will vary at different rates for different studies. Greater heterogeneity would likely lead to greater penalties (in terms of species missed) for reducing survey coverage, which would increasingly cancel out cost savings. Finally, cost savings are meaningless if research objectives are compromised; researchers should refer to the extensive literature on spatial design when determining survey coverage.

4.2 | How long should my units record?

The rarefaction curves indicate that increasing recording time has diminishing marginal utility but that the asymptote varies with the community (Figure 3). Results from the field data indicated that increased sampling over fewer days is better than slightly more time distributed over more days. This suggests that the results are more influenced by species with a high probability of daily availability but a low probability of vocal activity. However, contrasting results have been found when the maximum recording time is very short (30 min/day) (Cook & Hartley, 2018). Although such technology is not prevalent, researchers using ARUs that can be reprogrammed daily (e.g. via a cellular network) can further optimize their sampling by modifying recording times in response to factors like wind and precipitation (Balantic & Donovan, 2019). Although reducing recording time can reduce data storage costs (albeit minimally; Table S2) and data processing costs, reductions will eventually result in avoidable decreases in detection probabilities (Figure 4), a finding that is broadly supported by other studies (Sugai et al., 2020). Decreasing detection probability will reduce statistical power to detect population changes (Wood, Popescu, et al., 2019). Furthermore, reducing recording days will constrain researchers' ability to implement more nuanced analyses of the data, such as using vocal activity rate to estimate population density (Pérez-Granados et al., 2019). We recommend that researchers start with long recordings and assess rarefaction curves based on recording time before implementing highly reduced recording schemes. If software limitations or processing costs preclude the analysis of long recordings, researchers should maximize recall (rather than precision, as we did) to offset some of the missed detections. They should also consider implementing continuous recording and only analysing a subset of the data – the additional audio can be archived and analysed later. The biggest test of whether recording durations are too long is whether researchers can process all the audio that has been collected in one season (however long that may be) before the next begins.

5 | CONCLUSIONS

Whether the performance of a study design is 'good' or 'bad' depends on researchers' constraints and scientific questions to be answered: documenting 20% of the avian community at locations across the landscape in just three days for a fraction of the cost of conducting point counts may be considered a success if funds for fieldwork and data storage are limited and bioacoustic technology to extract community data are lacking. Conversely, detecting 86% of the community in 7 days could be disappointing if 91% of the community could have been documented in the same period when analysing twice as much audio data is not a limiting factor. Researchers developing acoustic monitoring programs must therefore balance scarce conservation resources, computational limitations and information quality. More hardware and longer recording durations are not always better, but we caution researchers against doing the bare minimum required for their present needs without pressing computational and financial reasons to do so (also see Sugai et al., 2020). Passive acoustic surveys yield a uniquely flexible permanent record of a place and time, and the value of such data may not be fully appreciated until unforeseen challenges emerge.

ACKNOWLEDGEMENTS

We thank Michaela Gustafson for leadership in the field, Chris Pelkie for data management assistance, Camila Gomez for translating the abstract, and two reviewers whose input improved this paper. Funding was provided by the U.S. Forest Service Region 5, the California Department of Fish and Wildlife and the Cornell Lab of Ornithology. The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTIONS

C.M.W. conceived the ideas, supervised data collection, conducted the analyses, and led the writing of the manuscript; S.K. created necessary software and contributed to the analyses; P.C. contributed to the analyses; M.Z.P. contributed to data collection; H.K. contributed to software development and supervised data collection. All authors contributed critically to the drafts and gave final approval for publication.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13571.

DATA AVAILABILITY STATEMENT

Simulation code is publicly available at Zenodo: http://doi.org/10.5281/zenodo.4193364 (Wood, 2020). The field data used in these analyses are also available at Zenodo: https://doi.org/10.5281/zenodo.4477478 (Wood, 2021).

ORCID

Connor M. Wood https://orcid.org/0000-0002-0235-5214
Stefan Kahl https://orcid.org/0000-0002-2411-8877
Holger Klinck https://orcid.org/0000-0003-1078-7268

REFERENCES

Balantic, C., & Donovan, T. (2019). Temporally adaptive acoustic sampling to maximize detection across a suite of focal wildlife species. *Ecology and Evolution*, *9*, 10582–10600. https://doi.org/10.1002/ece3.5579

Blackburn, T. M., & Gaston, K. J. (1994). The distribution of body sizes of the world's bird species. *Oikos*, *70*, 127–130. https://doi.org/10.2307/3545707

Campos-Cerqueira, M., & Aide, T. M. (2016). Improving distribution data of threatened species by combining acoustic monitoring and occupancy modelling. *Methods in Ecology and Evolution*, *7*, 1340–1348. https://doi.org/10.1111/2041-210X.12599

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

Darras, K., Batáry, P., Furnas, B., Celis-Murillo, A., Wilgenburg, S. L. V., Mulyani, Y. A., & Tscharntke, T. (2018). Comparing the sampling performance of sound recorders versus point counts in bird surveys: A meta-analysis. *Journal of Applied Ecology*, *55*(6), 2575–2586. https://doi.org/10.1111/1365-2664.13229

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*, *162*(1), 303–306. https://doi.org/10.1007/s10336-020-01812-6

de Camargo, U., Roslin, T., & Ovaskainen, O. (2019). Spatio-temporal scaling of biodiversity in acoustic tropical bird communities. *Ecography*, *42*, 1936–1947. https://doi.org/10.1111/ecog.04544

Furnas, B. J. (2020). Rapid and varied responses of songbirds to climate change in California coniferous forests. *Biological Conservation*, *241*, 108347.

Gaston, K., & Fuller, R. (2007). Commonness, population depletion and conservation biology. *Trends in Ecology & Evolution*, *23*, 14–19. https://doi.org/10.1016/j.tree.2007.11.001

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, *4*, 379–391. https://doi.org/10.1046/j.1461-0248.2001.00230.x

Haskell, J. P., Ritchie, M. E., & Olff, H. (2002). Fractal geometry predicts varying body size scaling relationships for mammal and bird home ranges. *Nature*, *418*, 527–530.

Hill, A. P., Prince, P., Piña Covarrubias, E., Doncaster, C. P., Snaddon, J. L., & Rogers, A. (2018). AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolution*, *9*, 1199–1211.

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, *61*, 101236.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. Princeton University Press.

McGill, B. J. (2003). Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? *Ecology Letters*, *6*, 766–773.

Pérez-Granados, C., Gómez-Catasús, J., Bustillo-de la Rosa, D., Barrero, A., Reverter, M., & Traba, J. (2019). Effort needed to accurately estimate Vocal Activity Rate index using acoustic monitoring: A case study with a dawn-time singing passerine. *Ecological Indicators*, *107*, 105608.

R Core Development Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rabinowitz, D. (1981). Seven forms of rarity. In H. Synge (Ed.), *Biological aspects of rare plant conservation* (pp. 205–217). John Wiley & Sons.

Robbins, C. S. (1981). Effect of time of day on bird activity. *Studies in Avian Biology*, *6*, 275–286.

Shonfield, J., & Bayne, E. M. (2017). Autonomous recording units in avian ecological research: Current use and future applications. *Avian Conservation and Ecology*, *12*, 14.

Soberon, J. M., & Llorente, J. B. (1993). The use of species accumulation functions for the prediction of species richness. *Conservation Biology*, *7*, 480–488.

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*, 220–235.

Thompson, G. G., & Withers, P. C. (2003). Effect of species richness and relative abundance on the shape of the species accumulation curve. *Austral Ecology*, *28*, 355–360.

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

Wood, C. M. (2020). R code: Acoustic survey design for species richness. *Zenodo*, https://doi.org/10.5281/zenodo.4193364

Wood, C. M. (2021). Field Data: Acoustic survey design for species richness. *Zenodo*, https://doi.org/10.5281/zenodo.4477478

Wood, C. M., Gutiérrez, R. J., Keane, J. J., & Peery, M. Z. (2019). Early detection of rapid Barred Owl population growth within the range of the California Spotted Owl advises the Precautionary Principle. *The Condor*, *122*, 1–10.

Wood, C. M., Gutiérrez, R. J., & Peery, M. Z. (2019). Acoustic monitoring reveals a diverse forest owl community, illustrating its potential for basic and applied ecology. *Ecology*, *100*, e02764.

Wood, C. M., & Jones, G. M. (2019). Framing management of social-ecological systems in terms of the cost of failure: The Sierra Nevada, USA as a case study. *Environmental Research Letters*, *14*, 105004.

Wood, C. M., Klinck, H., Gustafson, M., Keane, J. J., Sawyer, S. C., Gutiérrez, R. J., & Peery, M. Z. (2020). Using the ecological significance of animal vocalizations to improve inference in acoustic monitoring programs. *Conservation Biology*, *35*(1), 336–345. https://doi.org/10.1111/cobi.13516

Wood, C. M., Popescu, V. D., Klinck, H., Keane, J. J., Gutiérrez, R. J., Sawyer, S. C., & Peery, M. Z. (2019). Detecting small changes in populations at landscape scales: A bioacoustic site-occupancy framework. *Ecological Indicators*, *98*, 492–507.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Wood CM, Kahl S, Chaon P, Peery MZ, Klinck H. Survey coverage, recording duration and community composition affect observed species richness in passive acoustic surveys. *Methods Ecol Evol*. 2021;12:885–896. https://doi.org/10.1111/2041-210X.13571