

Pairing a user-friendly machine-learning animal sound detector with passive acoustic surveys for occupancy modeling of an endangered primate

Connor M. Wood¹  | Alicia Barceinas Cruz^{2,3}  | Stefan Kahl¹

¹K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, New York, USA

²Nelson Institute for Environmental Studies, Department of Geography, University of Wisconsin-Madison, Madison, Wisconsin, USA

³Natura y Ecosistemas Mexicanos AC, Mexico City, Mexico

Correspondence

Connor M. Wood, K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, 159 Sapsucker Woods Rd, Ithaca, NY 14850, USA.
Email: cmw289@cornell.edu

Abstract

Population declines and range contractions due to habitat loss are pervasive among nonhuman primates, with 60% of species threatened with extinction. However, the extensive vocal activity displayed by many primates makes them excellent candidates for passive acoustic surveys. Passive acoustic survey data is increasingly being used to support occupancy models, which have proven to be an efficient means of estimating both population trends and distributions. Passive acoustic surveys can be conducted relatively quickly and at broad scales, but efficient audio data processing has long proven elusive. The machine learning algorithm BirdNET was originally developed for birds but was recently expanded to include nonavian taxa. We demonstrate that BirdNET can accurately and efficiently identify an endangered primate, the Yucatán black howler monkey (*Alouatta pigra*), by sound in passive acoustic survey data (collected in southeastern Chiapas, Mexico), enabling us to use a single-season occupancy model to inform further survey efforts. Importantly, we also generated data on up to 286 co-occurring bird species, demonstrating the value of integrated animal sound classification tools for biodiversity surveys. BirdNET is freely available, requires no computer science expertise to use, and can readily be expanded to include more species (e.g., its species list recently tripled to >3000), suggesting that passive acoustic surveys, and thus occupancy modeling, for primate conservation could rapidly become much more accessible. Importantly, the long history of bioacoustics in primate research has yielded a wealth of information about their vocal behavior, which can facilitate appropriate survey design and data interpretation.

KEYWORDS

artificial intelligence, BirdNET, Chiapas, fragmentation, Mexico, monitoring, tropical forest

1 | INTRODUCTION

Assessments of range contractions and population declines feature prominently in the IUCN's Red List criteria (IUCN, 2012), and occupancy models have proven to be an efficient way to evaluate these population parameters, particularly at broad spatial scales (MacKenzie et al., 2017). Seventy-five percent of nonhuman primate species (hereafter "primates") are declining and 60% are threatened with extinction (Estrada et al., 2017), suggesting that there is an urgent need for population assessments. Fortunately, the extensive—and often well-studied—vocal activity of many primates makes them ideal candidates for passive acoustic surveys (e.g., Clink & Klinck, 2021, 2021; Do Nascimento et al., 2021; Dufourq et al., 2021; Piel et al., 2022; Zambolli et al., 2022), which are rapidly gaining traction as a means of generating the detection/non-detection data required for occupancy models (e.g., Campos-Cerqueira & Aide, 2016; Wood et al., 2019). Facilitating the implementation of occupancy-oriented passive acoustic monitoring (PAM) could be beneficial to primate conservation globally.

Occupancy models enable researchers to account for imperfect detection, or the failure to observe animals that are actually present. Imperfect detection is expressed as the "detection probability" (p), or the probability of detecting a species given that it is present, and is used to correct for underestimates of a population ("occupancy", ψ) (MacKenzie et al., 2002; Tyre et al., 2003). Critically, sites must be surveyed multiple times (called secondary sampling periods) to estimate the detection probability. Passive acoustic surveys require visiting sites twice: once to deploy a recording unit and once to retrieve it, but the intervening time can be partitioned into multiple secondary sampling periods. In contrast, human-based surveys require one visit to a site per survey, and the surveys themselves tend to be quite short (e.g., Neilson et al., 2013; Vu et al., 2020) both applied occupancy models to primate data. Conducting three or four surveys is common, but in rugged tropical forests, conducting three or four visits to a site for a total of 45–60 min of active observation is likely to be less efficient than a passive acoustic survey approach in which two visits to a site can yield weeks or even months of continuous observation. Thus, the upfront cost of passive acoustic survey devices can be offset by substantial logistical savings, making bioacoustics a particularly appealing approach to surveying species in remote areas (Wood et al., 2023).

Yet throughout the field of bioacoustics, the scarcity of broadly available and easy-to-use sound detection and classification tools has consistently been a limiting factor. Collecting tens of thousands of hours of audio is simple compared to the task of efficiently identifying relevant sounds in that audio. Researchers interested in using passive acoustic surveys generally face the choice of manually reviewing their audio data, which imposes severe limits on the quantity of data that can be processed, or developing their own detectors, which requires a nontrivial investment of time and effort. Machine learning algorithms have rapidly become a popular tool for automated sound identification, with deep convolutional neural networks (CNN) proving particularly effective. However, even

implementing a pretrained machine learning algorithm (e.g., Kong et al., 2020) can require substantial computer science expertise.

Researchers have combined passive acoustic surveys, machine learning sound identification tools, and occupancy modeling for primate research (Heinicke et al., 2015; Kalan et al., 2015; Ravaglia et al., 2023), but this promising workflow has seen limited growth—possibly because many machine learning tools are species-specific and require substantive training to implement. Thus, tools created by one research team may be difficult for other groups to use. We present an example of an efficient and easy-to-use primate vocalization detector which can readily be employed to analyze the large audio data sets generated by large-scale population monitoring programs, and has the potential to include many species. In the context of urgent threats to primate habitat and the potential for occupancy-oriented acoustic monitoring to provide statistically powerful assessments of species trends and habitat associations (Wood et al., 2019), such a framework may facilitate much needed conservation research.

2 | DESCRIPTION

The BirdNET algorithm (Kahl et al., 2021), employs a CNN architecture for identifying acoustic events by leveraging visual patterns in spectrograms. Specifically, the CNN processes 3-second audio snippets sampled at 48 kHz (via upsampling input audio if needed) and transforms them into linear-scale spectrograms, from which high-level features are extracted using learned filter operations. A classification layer subsequently assigns logit values for each of a fixed set of classes. A sigmoid activation function is then applied to the classification layer's output and transforming logits into confidence scores [0-1], representing BirdNET's confidence that a class is in the input audio. BirdNET is capable of detecting multiple acoustic events in a single audio snippet, making it well-suited for multilabel tasks. The neural network was trained on a data set comprising approximately 10 million audio samples, with an emphasis on robustness to diverse noise sources and nonevents. BirdNET has several features that may enable it to facilitate passive acoustic surveys of primates. First, BirdNET has proven highly scalable. As the name suggests, it was initially developed for avian applications, and the first version could identify 984 North American and European birds by sound alone. A recent expansion increased its capabilities to over 3000 species, including some amphibians and mammals, notably the Yucatán black howler monkey (*Alouatta pigra*). Thus, the computational framework has proven quite capable of accommodating many new acoustic event classes, indicating that additional training data from primates of interest could be used to further expand its capabilities. Second, BirdNET is freely available and can be implemented with a simple point-and-click graphical user interface (<https://github.com/kahst/BirdNET-Analyzer>). Using BirdNET requires no expertise in computer science or machine learning, although users are generally limited to the existing species list (but see "Comparison and Critique" and McGinn et al., 2023).

We demonstrate that audio data from passive acoustic surveys can be analyzed with BirdNET to detect an endangered primate and that these results can subsequently be used for occupancy modeling, yielding results that can inform subsequent population survey efforts. We also show that the BirdNET approach enables unified biodiversity surveys, a major advantage over single-species vocalization detection tools. Rapidly generating the data needed to robustly estimate the distribution and, over time, population trends of acoustically active primates in an occupancy modeling framework can be a powerful tool for conservation.

3 | EXAMPLE

3.1 | Study area and acoustic surveys

The Yucatan black howler monkey (*A. pigra*) is an endangered primate (Cortes-Ortiz et al., 2020) endemic to the Maya Forest in southeastern Mexico, northern Guatemala, and Belize (Rylands et al., 2006). Studies in the Lacandona region in Chiapas, MX have shown that this primate is sensitive to landscape spatial changes as population decreases are associated with reductions in forest patch size and increases in patch isolation (Arce-Peña et al., 2019; Arroyo-Rodríguez et al., 2013).

We deployed autonomous recording units (ARUs Swift Recorder, Pelican case edition; Cornell Lab of Ornithology) in six locations, two

in southern Tabasco and four in eastern Chiapas, MX in July 2018. The ARUs had one internal, omnidirectional microphone set to record at a sample rate of 32 kHz, gain of +35.0 dB, and bit depth of 16 from 4:00–10:00 to 16:00–22:00 each day for 2–5 days (the recorders can last much longer, but logistical constraints precluded longer deployments). In total we collected and analyzed 163 h of audio across 20 cumulative survey days. Passive acoustic survey locations were chosen to reflect different disturbance and fragmentation levels of tropical rainforest habitat inside and three nearby federal Protected Areas: Cañón del Usumacinta, Montes Azules, and Yaxchilán (Figure 1). These areas have been described as human-modified forest landscapes (Arroyo-Rodríguez et al., 2017), as forests have experienced rapid transformation to give way to agricultural lands, chiefly cattle pastures and palm oil plantations (Gallardo-Cruz et al., 2021), as a result of directed human settlement policies that started in the 1970s (De Vos, 2002). The ARUs were placed in the interior of patches of tropical rainforest habitat where previous efforts of community-based wildlife monitoring had detected the presence of mammals including but not limited to primates. ARUS were placed at least 2000 m apart; howler monkey vocalizations can be audible at 1000 m (Schön, 1971; Schön Ybarra, 1986) so the assumption that survey sites are independent was reasonable.

All survey efforts were entirely passive; at no point did we interact with nonhuman primates. This work complied with the Mexican legal requirements for primate research and the American Society of Primatologists Principles for Ethical Treatment of Nonhuman Primates.

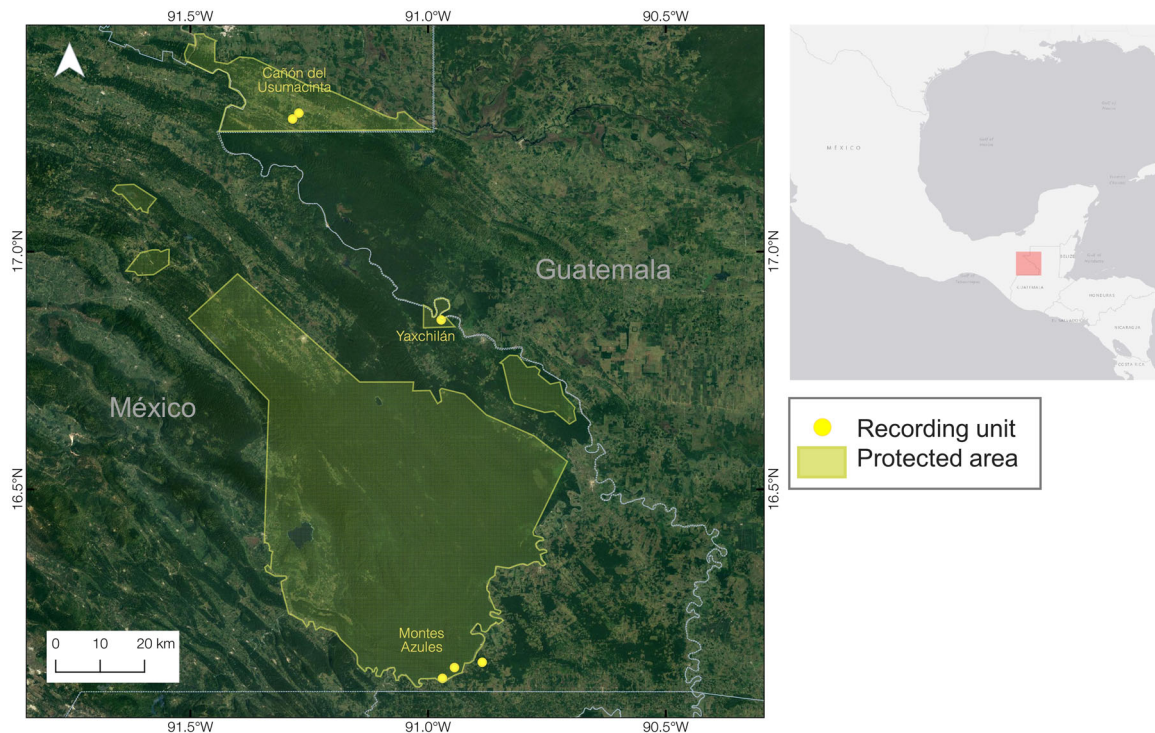


FIGURE 1 Six survey sites in Chiapas, Mexico. Locations were chosen to reflect different disturbance and fragmentation levels of tropical rainforest habitat inside and nearby federal Protected Areas, with recording units placed in the interior of patches of tropical rainforest habitat where previous efforts of community-based wildlife monitoring had detected the presence of mammals including but not limited to primates.

These surveys and subsequent analyses were conducted only after receiving explicit permission to do so from local communities.

3.2 | Audio analysis with BirdNET

We analyzed the resulting audio with BirdNET. We used a species list that included the Yucatán black howler monkey and 286 birds known to be present in our study area, but we focused our analysis on the howler monkey results. Including nontarget species does not influence BirdNET's performance, but we included them to illustrate the multispecies potential of combining passive acoustic surveys with BirdNET. Although BirdNET can generate bird-only species lists automatically based on geographic coordinates provided by the user, we recommend that users review the species list provided on GitHub and manually create a list based on the regional species pool (<https://github.com/kahst/BirdNET-Analyzer>; 26 languages, including English and Spanish).

Briefly, BirdNET breaks input audio into 3-second chunks and generates a prediction for each listed species for each chunk. BirdNET can be configured to analyze audio with up to 50% overlap between adjacent chunks to mitigate the possibility that target signals are not identified because they have been truncated. However, consistent with other research on this species and other howler monkeys (Van Belle et al., 2013; Pérez-Granados & Schuchmann, 2021), we found that when howler monkeys were vocally active, they produced many sequential vocalizations such that even if one call was truncated and thus not correctly identified, other calls were successfully identified. Predictions with a negligible likelihood of being correct are automatically filtered out, and remaining predictions receive a "confidence score" from 0.0 to 1.0. Although higher confidence scores correspond with a greater chance that the prediction is correct, there is no direct connection to the probability that a prediction is correct. Users could simply begin reviewing predictions in descending score order, but a more systematic approach is to develop probabilistic scores.

As described above, BirdNET provides "confidence scores" that resemble probabilities insofar as they range [0-1], but which are merely a unitless quantification of BirdNET's confidence in its predictions. Higher scores are more likely to be correct, but the specifics of that relationship are species-specific and unknown. Fortunately, confidence scores can be translated into probabilities: the user manually reviews a random selection of predictions and then uses logistic regression to establish a probabilistic relationship between the binary outcome (correct or incorrect) of a prediction and its prediction score. Thus, a user could set a threshold, such as 80% accuracy, and review all predictions above that level.

To structure our review of BirdNET's howler monkey predictions and gain a detailed understanding of detector performance, we developed probabilistic scores. First, after the initial BirdNET analysis, we used the BirdNET "segments.py" script (<https://github.com/kahst/BirdNET-Analyzer>) to randomly select 200 howler monkey predictions across a broad range of scores (0.10–1.0) and another 200 predictions from a higher score range (0.85–1.0). Next, CW

manually validated all these predictions in Raven Pro 1.6 (K. Lisa Yang Center for Conservation Bioacoustics, 2017), classifying each as correct or incorrect.

We then back-transformed the confidence score into its original logit scale to use as an independent variable ($\text{logitscore} = \ln\left(\frac{\text{confidencescore}}{1 - \text{confidencescore}}\right)$) in the logistic regression analysis. We used the binary outcome of all 400 validated predictions (correct or incorrect) as a response variable and the logit-scale prediction score as an independent variable. There was strong support in the data for a positive relationship between score and the probability that a prediction was correct ($\text{AIC}_{\text{score model}} \ll \text{AIC}_{\text{null model}}$; intercept = -2.54 , $\beta_{\text{score}} = 0.55$ [$\text{SE}_{\beta} = 0.06$; $p < 0.001$]; Figure 2).

Based on the logistic regression analysis, we found that even at the highest scores, accuracy was 0.79, meaning that BirdNET predictions above that score had a 79% chance of being correct. If false-positive rates of <1%–2% cannot be achieved, substantial bias can be introduced to occupancy models, meaning that researchers should either use a model that explicitly accounts for false positives (Clare et al., 2021) or manually validate detections above a given threshold to ensure that only true positives are included in subsequent analyses. Other machine learning-based primate vocalization detectors have also suggested manual validation as an important post-processing step (Kalan et al., 2015) (as opposed to using an occupancy model that can account for false positives), and, for a large-scale PAM project for a threatened bird, even processing >100,000 h of audio and validating >10,000 high-confidence predictions proved sufficiently efficient that monitoring could be conducted annually without falling behind (Wood et al., 2020).

Increasing the score threshold to review putative detections generally results in increased precision (fewer false positives) and

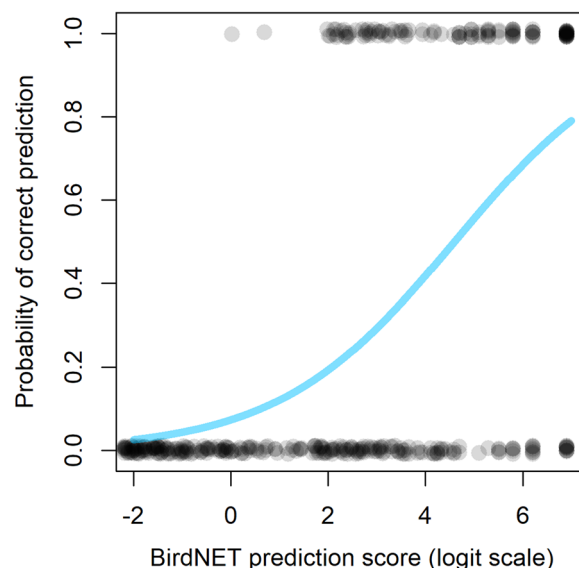


FIGURE 2 The probability of a correct BirdNET prediction (validated predictions shown in gray) increased with prediction score. Highest possible accuracy, estimated via logistic regression (blue line), was 0.79, meaning that BirdNET predictions at or above that score had a 79% chance of being correct.

decreased recall (more false negatives, or missed detections). For animals with extensive vocal activity, including many primates, low recall is not necessarily a problem for occupancy modeling. First, missing three out of five vocalizations, for example, nonetheless means that the animal has been “detected.” Second, occupancy models were explicitly designed to account for “missed detections” such that failing to observe an animal even when it is present can be accounted for statistically (Mackenzie et al., 2002; Tyre et al., 2003).

3.3 | Howler monkey results and an example occupancy model

We manually reviewed Howler monkey predictions with a $\geq 70\%$ chance of being correct. Yucatán black howler monkeys were recorded and correctly identified at all six locations, and on 14 of the 20 survey days. At some sites, howler monkey vocalizations were quite close to the recording site, at others, vocalizations were distant, and at least one site, vocalizations of both distant and foreground individuals were detected by BirdNET (Figure 3).

Importantly, recall (the proportion of vocalizations that are recorded and identified) exceeded 80% across all sites. For any acoustic signal, amplitude decreases with distance to recorder and higher-frequency sounds attenuate more rapidly than lower-frequency ones, such that both the signal-to-noise ratio of distant signals and their fidelity to the original signal are reduced. Nonetheless, BirdNET prediction scores were robust to howler monkey distance from an ARU, with many distant vocalizations scoring just as high as foreground vocalizations (e.g., Figure 3).

With just six sites, we did not have enough data to test for the influence of any covariates on occupancy (ψ) or detection (p), and chose to fit a null model in which both parameters (ψ and p) were uniform across sites. Because we had manually confirmed that howler monkeys were recorded at all six sites, we knew that occupancy was indeed uniform and, of course, that the monkeys

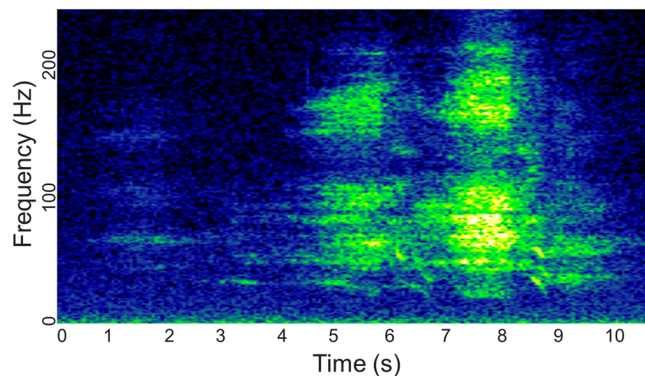


FIGURE 3 BirdNET identified both the distant (0.5–2.5 s) and foreground (5–9 s) Yucatán black Howler monkey vocalizations in this spectrogram with extremely high confidence; compare to the long calls described by (Briseño-Jaramillo et al., 2017). The audio used for that spectrogram is available in the Supporting Information: Material.

were present at all locations. However, an estimate of detection would be quite valuable, as it would enable us to design future howler monkey surveys in this area, a priority for local communities. Therefore, with the code below, we used the package *unmarked* (Fiske & Chandler, 2011) in program R (R Core Development Team, 2020), to fit a single-season, single-species occupancy model. There is extensive documentation online for *unmarked* and other occupancy modeling tools.

```
> library(unmarked)
> # manually create our encounter histories
> monkey_data <- data.frame(
  rbind(
    c(0, 1, 1, NA, NA), # site 1 encounter history
    c(0, 1, NA, NA, NA), # site 2 encounter history
    c(0, 1, 1, 0, 1), # site 3 encounter history
    c(1, 1, 0, NA, NA), # site 4 encounter history
    c(0, 1, 1, 1, NA), # site 5 encounter history
    c(0, 1, 1, NA, NA) # site 6 encounter history
  ),
  row.names = c("site1", "site2", "site3", "site4",
    "site5", "site6")
  # neither row nor column names are needed for the
  # encounter histories, but they can help you stay
  # organized)
> # note: in practice, reading in a.csv containing the
  # encounter histories is much simpler
> # convert the data frame to a singleseason occupancy
  # model format. You can add covariates at this stage
> occ_input <- unmarked::unmarkedFrameOccu
  (monkey_data, siteCovs = NULL, obsCovs = NULL)
> # fit a singleseason, singlespecies occupancy model
> occ_mod <- unmarked::occu(~1~1,
  knownOcc = seq(1,6,1),
  starts = c(1, 0),
  occ_input)
> # ~1~1 are the detection and occupancy formulas,
  # respectively.
> # (cont.) Both are uniform here, but either can be
  # allowed to vary with site or observation
  # covariates
> # knownOcc tells the model that we know all six sites
  # are occupied. This command is not always needed.
> # starts tells the model to start by assuming those
  # occupancy and detection values.
> # (cont.) Confusingly, starts asks for occupancy
  # then detection, in contrast to the initial
  # formula.
> # (cont.) If you don't know a good starting point,
  # which we don't for detection, use 0.
> occ_mod # view the results.
> # Note: we get the warning "Hessian is singular",
  # which occurs when there is perfect fit, such as
  # 100% occupancy
> unmarked::backTransform(occ_mod, 'state') # view
  the occupancy estimate
> unmarked::backTransform(occ_mod, 'det') # view
  the detection estimate
```

As expected, our model generated a warning because all sites were occupied (and with just six sites, the model has very little data with which to estimate any parameter values). But in our case, we are more interested in the detection estimate, which was 0.65. Detection probability is measured at the level of a secondary sampling period, in this case, at least 4 h of passive acoustic survey effort at either dawn or dusk. This means that if for each day of passive acoustic survey effort, we have a 65% chance of detecting a Yucatán black howler monkey if one is present. An important extension of detection is the seasonal detection probability (p^*), or the probability that an animal is observed over the entire duration of the sampling season:

$$p^* = 1 - (1 - p)^n,$$

where p is the detection probability and n is the number of secondary sampling periods. Thus, the probability of observing a howler monkey with a 3-day passive acoustic survey effort would be 0.96; adding a fourth day to the survey would increase p^* to 0.98. High p is positively related to the statistical power to detect population changes, a valuable attribute in the context of endangered species (Wood et al., 2019).

Conducting accurate broad-scale surveys for the Yucatán black howler monkey could therefore be accomplished with a relatively small up-front investment in recording units, as long as personnel were available to deploy and retrieve ARUs. For example, if 10 ARUs were deployed for 4-day periods, with two additional days required to deploy and retrieve the units, 50 locations could be surveyed in 30 days, with just a 2% chance of failing to detect Yucatán black howler monkeys if they were present at one of those sites. Information about space-, time-, and behavior-specific vocalizations (Van Belle et al., 2013; Briseño-Jaramillo et al., 2017) could inform minimum recording unit spacing. While occupancy estimates (ψ) would be important for population assessments, spatial variation in detection (p) could be quite informative because it is influenced by animal behavior (i.e., the choice of whether to vocalize). For example, Yucatán black howler monkeys can survive and reproduce in forest fragments, but form smaller troops than in intact forests (Estrada et al., 2002), a difference that could result in altered patterns of intragroup and intergroup vocal activity. Habitat-driven differences in vocal activity, in turn, could be manifested as model support for a “fragmentation” covariate of p . Furthermore, although howler monkeys are vocally active year-round, seasonal variation in vocal activity rates has been documented (Chiarello, 1995; Pérez-Granados & Schuchmann, 2021), suggesting that the detection probability we observed in July may be higher or lower than in other seasons. In summary, our small case study demonstrates the viability of combining passive acoustic surveys, BirdNET-based audio data analysis, and occupancy modeling for Yucatán black howler monkey population studies, and illustrates the potential for future work.

4 | COMPARISON AND CRITIQUE

Combining PAM and occupancy modeling has substantial potential to advance primatology (Piel et al., 2022), and global primate population declines (Estrada et al., 2017) suggest that monitoring efforts capable

of informing conservation are urgently needed. Though the use of occupancy modeling in primatology lags behind other disciplines of wildlife ecology (Piel et al., 2022), it is gaining traction (e.g., Almeida-Rocha et al., 2020; Keane et al., 2012; Vu et al., 2020). Combined PAM and occupancy studies are more scarce (e.g., Crunchant et al., 2020), though we are not the first to propose the PAM to machine learning to occupancy workflow (Kalan et al., 2015). Yet the accessibility of BirdNET as well as its proven ability to readily accommodate new sound classes in its identification repertoire, as well as the growing global capacity for bioacoustic research, suggest that rapid growth in primate population monitoring may be possible.

With both a graphical user interface and command line implementation options (<https://github.com/kahst/BirdNET-Analyzer>), BirdNET is a relatively user-friendly tool. However, classification performance is known to vary among species, and careful testing of the performance of newly added phylogenetic classes (e.g., Mammalia) is warranted for. Indeed, the global performance metrics, which are dominated by birds, may not be particularly informative. We found that BirdNET achieved high recall in our data set, meaning that most howler monkey vocalizations were identified as such. Furthermore, it was able to identify vocalizations produced near to and far from the ARU with high confidence (e.g., Figure 3). With an estimated maximum accuracy of 0.79 (Figure 2), however, reducing false positive rates would be beneficial. Improving the quality of training data, likely by increasing the quantity of tightly cropped recordings that contain only the target vocalization, as well as examples of nontarget soundscape examples, should increase precision (reduce false positives).

A lack of training data for additional primate species is currently the biggest limitation of BirdNET. At present, the Yucatán black howler monkey is the only primate in BirdNET. Incorporating this species proved to be a straightforward extension of standard, periodic improvements to the model, which suggests that training BirdNET on additional primate sounds would be similarly simple. There are at least three solutions to the challenge of “missing species.” First, researchers can use BirdNET feature embeddings as a means of querying the audio for sounds of interest (see case study two in McGinn et al., 2023). In short, CNNs like BirdNET can be easily modified to provide their feature embeddings, which are vectors of numbers that characterize an input, in this case a 3-second chunk of audio (for BirdNET, feature embeddings can be generated using the “embeddings.py” script available on GitHub). Feature embeddings thus allow researchers to compare the multivariate distance among segments of audio, and this approach has been used to successfully identify target sounds in large audio datasets (McGinn et al., 2023). Thus, with just one example of a target sound, such as the vocalization of a primate that BirdNET has not been trained to identify, researchers could search for other such sounds in their data. Second, rapid advances in transfer learning suggest that researchers could use a new update to the BirdNET graphical user interface (available on GitHub) to retrain BirdNET to detect a sound of interest. (Transfer learning is the application of a pretrained classifier to the identification of a new, previously unknown class, such as another

primate in this case.) Thus, researchers may not need to wait for updated versions of BirdNET or resort to multivariate analyses to use BirdNET to identify species that are not already in the classifier but for which they have training data. Third, we encourage researchers with labelled audio samples of a species of interest to make annotated audio data publicly available with archives such as xeno-canto (<https://xeno-canto.org/>) or ecoSound-web (Darras et al., 2023). Many primates have diverse vocal repertoires and although BirdNET has proven capable of identifying many bird species on the basis of multiple very different sounds, prioritizing stereotyped vocalizations used for long-range communication is likely to be the most efficient path forward.

Other detectors may be more suitable for primate detection than BirdNET, and several are available (e.g., Clink & Klinck, 2019; Dufourq et al., 2021; Heinicke et al., 2015). The potentially substantial acoustic repertoires of some primates may warrant customized signal detection and classification solutions. However, there is increasing evidence that the architecture of CNNs may be less important to performance than post-processing steps (Kahl et al., 2017). Thus, we encourage researchers to conduct detailed analyses of detector prediction performance, potentially by expanding the logistic regression approach we demonstrated to include covariates that could affect performance. If the detector recognizes multiple vocalization types, prediction accuracy could differ among types; accuracy could also be influenced by seasonal or spatial variation in the soundscape. In these or other cases, the logistic regression equation might not be a simple univariate model relating outcome to score, but one that relates outcome to score, vocalization type, and habitat, for example. Importantly, occupancy modeling was specifically designed to account for missed observations. An animal may be silent during a passive acoustic survey or a detector may fail to identify a target vocalization; these possibilities are not particularly problematic in an occupancy modeling framework, and their cumulative effect can be measured directly via detection probability.

Previous research on primate vocal activity can help researchers optimize study design, including the duration of secondary sampling periods and daily recording schedules. For example, vocal activity rates have been well-studied for many species of howler monkey (Van Belle et al., 2013; Chiarello, 1995; Do Nascimento et al., 2021; Pérez-Granados & Schuchmann, 2021), as well as gibbons (Cl et al., 2020), lemurs (Batist et al., 2022), and others. Understanding temporal variation in vocal activity enables researchers to make informed decisions about the appropriate duration of their secondary sampling periods (i.e., the intervals that constitute the repeated visits to a site, whether they are discrete deployments or subsets of a single longer deployment). Simulation-based power analyses based on realistic vocal activity rates and population parameters and on preliminary field data can further improve study design (Wood et al., 2021).

In the context of conducting broad-scale passive acoustic surveys to assess primate distribution, habitat associations, and, over time, population trends, efficient conversion of raw audio to detection/non-detection data is essential. Yet the process of defining

“detections” can be complex (Wood & Peery, 2022). Recording a species just once may be an inappropriate definition, particularly because high-amplitude, low-frequency primate vocalizations can travel many hundreds of meters. Indeed, in our data, vocal activity at was equally extensive at two sites, but was quite distant from one ARU and quite close to the other. Moreover, there may be substantial variation in the levels of acoustic activity among “occupied” sites, which may be a function of the distance between an ARU and a hotspot of primate activity. Fortunately, passively recorded audio can contain vastly more information about a species than simply whether an individual was present and vocally active at a given place and time.

A wide range of techniques may help researchers convert primate vocalizations located in passively recorded audio into ecologically informative inputs for subsequent applications. An amplitude-based criterion could be used to count sites as “occupied” only if vocalizations are relatively close to the ARU, though even with data about source volume, distance could be difficult to estimate because of the multitude of factors that can affect sound propagation. Moving beyond binary site classifications (detected or not detected) is also possible with multistate occupancy models (Nichols et al., 2007), an approach that can improve the level of ecological detail obtained from passive acoustic datasets (Reid et al., 2021; Wood et al., 2020). The “state” of a site at which the focal species is present could be determined by the number of individuals for particularly social primates, something that can be determined acoustically (Torti et al., 2018), or by vocalizations that are unique to sex or age class. Feature embeddings can help distinguish vocalization types within species. McGinn et al. (2023) demonstrated that adult and juvenile vocalizations of the same species could be accurately grouped using multivariate distance metrics derived from BirdNET's feature embeddings.

Importantly, passive acoustic surveys designed for occupancy-oriented population modeling by no means preclude analyses of animal behavior, including intraspecific communication and group size (e.g., Torti et al., 2018) and interspecific interactions (Wood et al., 2020). Even in our relatively small data set, we detected examples of a howler monkey far from the ARU calling and one much closer responding (Figure 3). Follow-up surveys to provide even a general estimate detection range would be a valuable piece of future monitoring for this species. Population density estimation may also be a valuable objective, and passive acoustic surveys have proven to be a viable means of conducting such work (Markolf et al., 2022; Stevenson et al., 2021). With a multispecies detector like BirdNET, interspecific acoustic interactions among primates (and other acoustically active species) could also be studied. Perhaps the most detailed piece of information is individual identity, and acoustic-based identification of primates has been successful for some species (Clink & Klinck, 2021). Individual identification could improve occupancy modeling, for example, by helping researchers avoid treating two sites as occupied when really one individual was recorded at both, but it could also help monitoring efforts move beyond occupancy modeling entirely, instead implementing acoustic-based mark-recapture monitoring or even localized censuses.

Ongoing threats to tropical forests are driven by complex, rapidly changing interactions between local and global economic forces, changing temperatures, precipitation patterns, and fire regimes and associated feedback loops, and even political dynamics (Austin et al., 2017; Brancalion et al., 2020; Ruggiero et al., 2021; Seymour & Harris, 2019). Obtaining or maintaining an understanding of species responses to these changes is likely to require broad-scale survey efforts, ideally ones capable of yielding data on multiple species. Although primates are particularly threatened (Estrada et al., 2017), tropical biodiversity as a whole is at risk as well (Symes et al., 2018). The combination of passive acoustic surveys and multi-species sound identification tools like BirdNET may support holistic conservation efforts for vocally active species. For example, we scanned our audio data for howler monkeys, and for 286 birds known to be present in our study area. Results of interest (primate vocalizations) were partitioned and analyzed in depth, while putative detections of birds (and possibly amphibians and insects) can be reviewed later or shared with collaborators. The ability to generate taxonomically broad data with a single sound identification tool is an ancillary benefit of further developing BirdNET for primate research, albeit an important one. In the Sierra Nevada, USA, concerns about the status of the charismatic spotted owl (*Strix occidentalis*) justified investment in a large-scale PAM program that now enables population monitoring of the entire avian community; as flagship species, primates may have a similar incidental umbrella effect. With careful survey design and continued development of machine learning sound identification tools, PAM projects designed to provide broad-scale population data about primates can also provide equally rigorous population estimates about many other species.

AUTHOR CONTRIBUTIONS

Connor M. Wood: conceptualization (lead), data curation (equal), formal analysis (lead), methodology (equal), visualization (equal), writing—original draft (lead), writing—review and editing (equal). **Alicia Barceinas Cruz:** conceptualization (supporting), data curation (equal), funding acquisition (lead), methodology (equal), visualization (equal), writing—original draft (supporting), writing—review and editing (equal). **Stefan Kahl:** conceptualization (supporting), software (lead), writing—review and editing (equal).

ACKNOWLEDGMENTS

We thank the communities of El Bejucal, Frontera Corozal, and Playón de la Gloria for granting us access to their lands to conduct acoustic surveys and sharing their knowledge to choose survey sites. Lázaro Chavarría, Caralampio Jiménez, Mario Lombera, Jesús Narvaez Pascual, Leonora Torres Knoop, Paulina Arroyo, and Rodrigo León guided us to get to survey sites. We also thank personnel of the National Commission of Natural Protected Areas, Estación Chajul, and Mujeres, Organización y Territorio AC for facilitating our access to sites. We also thank Dena Clink for feedback that improved this manuscript. The BirdNET project was made possible by the generous support of Jake Holshuh.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

No data beyond the script in the manuscript are publicly available without the permission of the communities in which the audio was recorded; however, the authors welcome inquiries (e.g., bird results).

ORCID

Connor M. Wood  <http://orcid.org/0000-0002-0235-5214>

Alicia Barceinas Cruz  <https://orcid.org/0000-0001-8636-2341>

REFERENCES

- Almeida-Rocha, J. M., Peres, C. A., Monsalvo, J. A. B., & Oliveira, L. D. C. (2020). Habitat determinants of golden-headed lion tamarin (*Leontopithecus chrysomelas*) occupancy of cacao agroforests: Gloomy conservation prospects for management intensification. *American Journal of Primatology*, 82, e23179.
- Arce-Peña, N. P., Arroyo-Rodríguez, V., Dias, P. A. D., Franch-Pardo, I., & Andresen, E. (2019). Linking changes in landscape structure to population changes of an endangered primate. *Landscape Ecology*, 34, 2687–2701.
- Arroyo-Rodríguez, V., González-Perez, I. M., Garmendia, A., Solà, M., & Estrada, A. (2013). The relative impact of forest patch and landscape attributes on black howler monkey populations in the fragmented Lacandona rainforest, Mexico. *Landscape Ecology*, 28, 1717–1727.
- Arroyo-Rodríguez, V., Melo, F. P. L., Martínez-Ramos, M., Bongers, F., Chazdon, R. L., Meave, J. A., Norden, N., Santos, B. A., Leal, I. R., & Tabarelli, M. (2017). Multiple successional pathways in human-modified tropical landscapes: New insights from forest succession, forest fragmentation and landscape ecology research. *Biological Reviews*, 92, 326–340.
- Austin, K. G., González-Roglich, M., Schaffer-Smith, D., Schwantes, A. M., & Swenson, J. J. (2017). Trends in size of tropical deforestation events signal increasing dominance of industrial-scale drivers. *Environmental Research Letters*, 12, 054009.
- Batist, C. H., Razafindraibe, M. N., Randriamanantena, F., & Baden, A. L. (2022). Factors affecting call usage in wild black-and-white ruffed lemurs (*Varecia variegata*) at Mangevo, Ranomafana National Park. *Primates*, 63, 79–91.
- Van Belle, S., Estrada, A., & Garber, P. A. (2013). Spatial and diurnal distribution of loud calling in Black Howlers (*Alouatta pigra*). *International Journal of Primatology*, 34, 1209–1224.
- Brancalion, P. H. S., Broadbent, E. N., de-Miguel, S., Cardil, A., Rosa, M. R., Almeida, C. T., Almeida, D. R. A., Chakravarty, S., Zhou, M., Gamarra, J. G. P., Liang, J., Cruzeilles, R., Hérault, B., Aragão, L. E. O. C., Silva, C. A., & Almeyda-Zambrano, A. M. (2020). Emerging threats linking tropical deforestation and the COVID-19 pandemic. *Perspectives in Ecology and Conservation*, 18, 243–246.
- Briseño-Jaramillo, M., Biquand, V., Estrada, A., & Lemasson, A. (2017). Vocal repertoire of free-ranging black howler monkeys' (*Alouatta pigra*): Call types, contexts, and sex-related contributions. *American Journal of Primatology*, 79, e22630.
- 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.
- Chiarello, A. G. (1995). Role of loud calls in brown howlers, *Alouatta fusca*. *American Journal of Primatology*, 36, 213–222.
- Clare, J. D. J., Townsend, P. A., & Zuckerberg, B. (2021). Generalized model-based solutions to false-positive error in species detection/nondetection data. *Ecology*, 102, e03241.

- Clink, D. J., Hamid Ahmad, A., & Klinck, H. (2020). Gibbons 'aren't singing in the rain: Presence and amount of rainfall influences ape calling behavior in Sabah, Malaysia. *Scientific Reports*, 10, 1282.
- Clink, D. J., & Klinck, H. (2019). GIBBONFINDER: An R package for the detection and classification of acoustic signals. <https://arxiv.org/abs/1906.02572v2>
- Clink, D. J., & Klinck, H. (2021). Unsupervised acoustic classification of individual gibbon females and the implications for passive acoustic monitoring. *Methods in Ecology and Evolution*, 12, 328–341.
- Cortes-Ortiz, L., Rosales-Meda, L. K., Marsh, & Mittermeier, R. A. (2020). *Alouatta pigra*. IUCN.
- Crunchant, A.-S., Borchers, D., Kühl, H., & Piel, A. (2020). Listening and watching: Do camera traps or acoustic sensors more efficiently detect wild chimpanzees in an open habitat? *Methods in Ecology and Evolution*, 11, 542–552.
- Darras, K. F. A., Pérez, N., Mauladi, Dilong, L., Hanf-Dressler, T., Markolf, M., & Wanger, T. C. (2023, March 1). ecoSound-web: An open-source, online platform for ecoacoustics [version 2; peer review: 2 approved]. *F1000Research*, 9, 1224. <https://doi.org/10.12688/f1000research.26369.2>
- Dufourq, E., Durbach, I., Hansford, J. P., Hoepfner, A., Ma, H., Bryant, J. V., Stender, C. S., Li, W., Liu, Z., Chen, Q., Zhou, Z., & Turvey, S. T. (2021). Automated detection of Hainan gibbon calls for passive acoustic monitoring. *Remote Sensing in Ecology and Conservation*, 7, 475–487.
- Estrada, A., Garber, P. A., Rylands, A. B., Roos, C., Fernandez-Duque, E., Di Fiore, A., Nekaris, K. A.-I., Nijman, V., Heymann, E. W., Lambert, J. E., Rovero, F., Barelli, C., Setchell, J. M., Gillespie, T. R., Mittermeier, R. A., Arregoitia, L. V., de Guinea, M., Gouveia, S., Dobrovolski, R., ... Li, B. (2017). Impending extinction crisis of the world's primates: Why primates matter. *Science Advances*, 3, e1600946.
- Estrada, A., Mendoza, A., Castellanos, L., Pacheco, R., Van Belle, S., García, Y., & Muñoz, D. (2002). Population of the black howler monkey (*Alouatta pigra*) in a fragmented landscape in Palenque, Chiapas, Mexico: A. Pigrain forest fragments. *American Journal of Primatology*, 58, 45–55.
- Fiske, I., & Chandler, R. (2011). Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, 43, 1–23.
- Gallardo-Cruz, J. A., Peralta-Carreta, C., Solórzano, J. V., Fernández-Montes de Oca, A. I., Nava, L. F., Kauffer, E., & Carabias, J. (2021). Deforestation and trends of change in protected areas of the Usumacinta River basin (2000–2018), Mexico and Guatemala. *Regional Environmental Change*, 21, 97.
- Heinicke, S., Kalan, A. K., Wagner, O. J. J., Mundry, R., Lukashovich, H., & Kühl, H. S. (2015). Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods in Ecology and Evolution*, 6, 753–763.
- IUCN. (2012). *IUCN Red list categories and criteria: Version 3.1* (p. iv + 32). IUCN.
- Kahl, S., Wilhelm-Stein, T., Hussein, H., Klinck, H., Ritter, M., & Eibl, M. (2017). Large-scale bird sound classification using convolutional neural networks (pp. 1–14). BirdCLEF.
- Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61, 101236.
- Kalan, A. K., Mundry, R., Wagner, O. J. J., Heinicke, S., Boesch, C., & Kühl, H. S. (2015). Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecological Indicators*, 54, 217–226.
- Keane, A., Hobinjatovo, T., Razafimanahaka, H. J., Jenkins, R. K. B., & Jones, J. P. G. (2012). The potential of occupancy modelling as a tool for monitoring wild primate populations. *Animal Conservation*, 15, 457–465.
- K. Lisa Yang Center for Conservation Bioacoustics. (2017). *Raven Pro: Interactive Sound Analysis Software*. Cornell Lab of Ornithology.
- Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., & Plumbley, M. D. (2020). PANNs: Large-scale pretrained audio neural networks for audio pattern recognition. *arXiv*. <https://arxiv.org/abs/1912.10211>
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.
- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L., & Hines, J. E. (2017). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence* (2nd ed.). Academic Press.
- Markolf, M., Zinowsky, M., Keller, J. K., Borys, J., Cillov, A., & Schülke, O. (2022). Toward passive acoustic monitoring of lemurs: Using an affordable open-source system to monitor phaner vocal activity and density. *International Journal of Primatology*, 43, 409–433.
- McGinn, K., Kahl, S., Peery, M. Z., Klinck, H., & Wood, C. M. (2023). Feature embeddings from the BirdNET algorithm provide insights into avian ecology. *Ecological Informatics*, 74, 101995.
- Do Nascimento, L. A., Pérez-Granados, C., & Beard, K. H. (2021). Passive acoustic monitoring and automatic detection of diel patterns and acoustic structure of Howler Monkey Roars. *Diversity*, 13, 566.
- Neilson, E., Nijman, V., & Nekaris, K. A. I. (2013). Conservation assessments of arboreal mammals in difficult terrain: Occupancy modeling of pileated gibbons (*Hylobates pileatus*). *International Journal of Primatology*, 34, 823–835.
- Nichols, J. D., Hines, J. E., MacKenzie, D. I., Seamans, M. E., & Gutiérrez, R. J. (2007). Occupancy estimation and modeling with multiple states and state uncertainty. *Ecology*, 88, 1395–1400.
- Pérez-Granados, C., & Schuchmann, K.-L. (2021). Passive acoustic monitoring of the diel and annual vocal behavior of the black and gold Howler monkey. *American Journal of Primatology*, 83, e23241.
- Piel, A. K., Crunchant, A., Knot, I. E., Chalmers, C., Fergus, P., Mulero-Pázmány, M., & Wich, S. A. (2022). Noninvasive technologies for primate conservation in the 21st century. *International Journal of Primatology*, 43, 133–167.
- R Core Development Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Ravaglia, D., Ferrario, V., De Gregorio, C., Carugati, F., Raimondi, T., Cristiano, W., Torti, V., Hardenberg, A. V., Ratsimbazafy, J., Valente, D., Giacoma, C., & Gamba, M. (2023). There you are! automated detection of Indris' Songs on features extracted from passive acoustic recordings. *Animals: An Open Access Journal from MDPI*, 13, 241.
- Reid, D. S., Wood, C. M., Whitmore, S. A., Berigan, W. J., Keane, J. J., Sawyer, S. C., Shaklee, P. A., Kramer, H. A., Kelly, K. G., Reiss, A., Kryshak, N., Gutiérrez, R. J., Klinck, H., & Peery, M. Z. (2021). Noisy neighbors and reticent residents: Distinguishing resident from non-resident individuals to improve passive acoustic monitoring. *Global Ecology and Conservation*, 28, e01710.
- Ruggiero, P. G. C., Pfaff, A., Nichols, E., Rosa, M., & Metzger, J. P. (2021). Election cycles affect deforestation within Brazil's Atlantic Forest. *Conservation Letters*, 14, e12818.
- Rylands, A. B., Groves, C. P., Mittermeier, R. A., Cortés-Ortiz, L., & Hines, J. J. H. (2006). Taxonomy and Distributions of Mesoamerican Primates. In A. Estrada, P. A. Garber, M. S. M. Pavelka, & L. Luecke (Eds.), *New Perspectives in the Study of Mesoamerican Primates: Distribution, Ecology, Behavior, and Conservation* (pp. 29–79). Springer US.
- Schön, M. A. (1971). The anatomy of the resonating mechanism in howling monkeys. *Folia Primatologica*. *International Journal of Primatology*, 15, 117–132.
- Schön Ybarra, M. A. (1986). Loud calls of adult male red howling monkeys (*Alouatta seniculus*). *Folia Primatologica*, 47, 204–216.
- Seymour, F., & Harris, N. L. (2019). Reducing tropical deforestation. *Science*, 365, 756–757.

- Stevenson, B. C., Dam-Bates, P., Young, C. K. Y., & Measey, J. (2021). A spatial capture–recapture model to estimate call rate and population density from passive acoustic surveys. *Methods in Ecology and Evolution*, 12, 432–442.
- Symes, W. S., Edwards, D. P., Miettinen, J., Rheindt, F. E., & Carrasco, L. R. (2018). Combined impacts of deforestation and wildlife trade on tropical biodiversity are severely underestimated. *Nature Communications*, 9, 4052.
- Torti, V., Valente, D., De Gregorio, C., Comazzi, C., Miarsetsoa, L., Ratsimbazafy, J., Giacoma, C., & Gamba, M. (2018). Call and be counted! Can we reliably estimate the number of callers in the indri's (*Indri indri*) song? *PLoS One*, 13, e0201664.
- Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications*, 13, 1790–1801.
- De Vos, J. (2002). *Una tierra para sembrar sueños. Historia reciente de la Selva Lacandona, 1950-2000*. Fondo de Cultura Económica.
- Vu, T. T., Hoa Anh, N. Q., Rawson, B. M., Tran, D. V., Nguyen, H. T., & Van, T. N. (2020). Monitoring occurrence, extinction, and colonization probabilities for gibbon populations. *American Journal of Primatology*, 82, e23171.
- Wood, C. M., Champion, J., Brown, C., Brommelsiek, W., Laredo, I., Rogers, R., & Chaopricha, P. (2023). *Challenges and opportunities for bioacoustics in the study of rare species in remote environments*. Conservation Science and Practice, e12941.
- Wood, C. M., Kahl, S., Chaon, P., Peery, M. Z., & Klinck, H. (2021). Survey coverage, recording duration and community composition affect observed species richness in passive acoustic surveys. *Methods in Ecology and Evolution*, 12, 885–896.
- 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, 336–345.
- Wood, C. M., & Peery, M. Z. (2022). What does 'occupancy' mean in passive acoustic surveys? *Ibis*, 164, 1295–1300.
- 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.
- Zambolli, A. H., Manzano, M. C. R., Honda, L. K., Rezende, G. C., & Culot, L. (2022). Performance of autonomous recorders to detect a cryptic and endangered primate species, the black lion-tamarin (*Leontopithecus chrysopygus*). *American Journal of Primatology*, 85, e23454.

SUPPORTING INFORMATION

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

How to cite this article: Wood, C. M., Barceinas Cruz, A., & Kahl, S. (2023). Pairing a user-friendly machine-learning animal sound detector with passive acoustic surveys for occupancy modeling of an endangered primate. *American Journal of Primatology*, e23507.

<https://doi.org/10.1002/ajp.23507>