

# Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring



Ammie K. Kalan<sup>a,\*</sup>, Roger Mundry<sup>a</sup>, Oliver J.J. Wagner<sup>a</sup>, Stefanie Heinicke<sup>a</sup>, Christophe Boesch<sup>a</sup>, Hjalmar S. Kühl<sup>a,b</sup>

<sup>a</sup> Max Planck Institute for Evolutionary Anthropology, Department of Primatology, Deutscher Platz 6, 04103 Leipzig, Germany

<sup>b</sup> German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

## ARTICLE INFO

### Article history:

Received 20 October 2014

Received in revised form

17 December 2014

Accepted 6 February 2015

### Keywords:

Acoustic detection

*Cercopithecus diana*

*Colobus polykomos*

Loud calls

Occurrence probability

*Pan troglodytes*

## ABSTRACT

Recent advancements in technology have made possible the use of novel, cost-efficient biomonitoring techniques which facilitate monitoring animal populations at larger spatial and temporal scales. Here, we investigated using passive acoustic monitoring (PAM) for wild primate populations living in the forest of Taï National Park, Côte d'Ivoire. We assessed the potential of using a customized algorithm for the automated detection of multiple primate species to obtain reliable estimates of species occurrence from acoustic data. First, we applied the algorithm on continuous rainforest recordings collected using autonomous recording units (ARUs) to detect and classify three sound signals: chimpanzee buttress drumming, and the loud calls of the diana and king colobus monkey. Using an occupancy modelling approach we then investigated to what extent the automated, probabilistic output needs to be listened to, and thus manually cleaned, by a human expert, to approach occupancy probabilities derived from ARU data fully verified by a human. To do this we explored the robustness of occupancy probability estimates by simulating ARU datasets with various degrees of cleaning for false positives and false negative detections. We further validated the approach by comparing it to data collected by human observers on point transects located within the same study area. Our study demonstrates that occurrence estimates from ARU data, combined with automated processing methods such as our algorithm, can provide results comparable to data collected by humans and require less effort. We show that occupancy probabilities are quite robust to cleaning effort, particularly when occurrence is high, and suggest that for some species even naïve occupancy, as derived from ARU data without any cleaning, could provide a quick and reliable indicator to guide monitoring efforts. We found detection probabilities to be most influenced by time of day for chimpanzee drums while temperature and, likely, poaching pressure, affected detection of diana monkey loud calls. None of the covariates investigated appeared to have strongly affected king colobus loud call detection. Finally, we conclude that the semi-automated approach presented here could be used as an early-warning system for poaching activity and suggest additional techniques for improving its performance.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

As concern for the current biodiversity crisis grows and the loss of species continues, conservationists and researchers struggle to find practical solutions for regularly monitoring wild animal populations that may be under threat or especially vulnerable to human disturbance. Recent technological advances have opened up the domain of biomonitoring to new methods and tools that have great

potential for real-time applications such as early-warning systems in areas where there is human–wildlife conflict (Zeppelzauer et al., 2014); or to alert field workers of increased anthropogenic stress to populations (Wrege et al., 2010; Penone et al., 2013), including poaching pressure (Wrege et al., 2012) or changes in populations due to natural causes such as disease and dispersal. In the last decade, autonomous audio–visual devices have gained popularity due to their ability to non-invasively collect objective data on wild animals with minimal human disturbance (Kühl and Burghardt, 2013). However, the increased popularity of autonomous recording units (ARUs), especially their potential for longitudinal monitoring with relatively little additional cost, has created the need for

\* Corresponding author. Tel.: +49 341 3550 205.

E-mail address: [ammie.kalan@eva.mpg.de](mailto:ammie.kalan@eva.mpg.de) (A.K. Kalan).

more sophisticated automated methods of handling the large audio datasets they produce (Blumstein et al., 2011; Wimmer et al., 2013; Kühl and Burghardt, 2013).

Passive acoustic monitoring (PAM) is a non-invasive method for surveying wild animals using remote acoustic technologies such as microphone arrays, hydrophones, or other autonomous recording devices (Blumstein et al., 2011; Marques et al., 2013). PAM has already proven useful for monitoring taxa at large geographic scales or species that are rare or elusive, particularly for birds (Swiston and Mennill, 2009; Celis-Murillo et al., 2012; Oppel et al., 2014), cetaceans (Mellinger et al., 2007; Van Parijs et al., 2009), insects (Riede, 1998; Penone et al., 2013), bats (Walters et al., 2012), and even forest elephants (Wrege et al., 2012). With over 50% of all primate species threatened by extinction (Mittermeier et al., 2009), methods which aid in long-term monitoring of primate populations would be a useful asset for conservationists working to protect some of the last strongholds of threatened wild populations. This was our principal objective in applying PAM for primates, to contribute to the research on automated biomonitoring techniques in hopes of improving the limited temporal and spatial scope of existing methods.

Automated methods for capturing sounds of interest from continuous audio recordings have proven problematic with many acoustic monitoring studies still using human experts to detect calls of interest in the data (Wrege et al., 2012; Wimmer et al., 2013; Penone et al., 2013). This is largely because of the variability in the signal-to-noise ratio of sounds captured on ARUs due to two principal reasons: distance of the source animal from the device and the variation in background noise levels in the environment. Despite these difficulties, recent studies have demonstrated considerable advancement in the automated detection of species-specific calls from continuous field recordings made in the field for birds (Digby et al., 2013), insects (Chesmore and Ohya, 2004), elephants (Zeppelzauer et al., 2014), cetaceans (Mellinger and Clark, 2000), and bats (Walters et al., 2012). Additionally, an impressive multi-species system for a tropical rainforest was recently proposed (Aide et al., 2013). However, it is often unclear whether the same data was used for training and testing which can considerably inflate performance statistics when conducting algorithm evaluations.

The rainforest provides a particularly challenging environment for passive acoustics. The background noise level is on average high and varies dramatically over the course of a day (Waser and Waser, 1977), of which the consequence is a high false positive and false negative rate. This is the central impediment to developing species-specific algorithms for automated sound detection since human experts are much better at discriminating between a call of interest from high background noise (Blumstein et al., 2011). However, the error associated with human detection and classification is rarely quantified although it can considerably bias results and lead to inappropriate inferences (Guschanski et al., 2009; Miller et al., 2011; Kühl and Burghardt, 2013; Digby et al., 2013). At the very least then, automated approaches allow for this error in detection and classification to be a known quantity, which can then be accounted for statistically when drawing inference from the data. Still, it remains that, due to the complex soundscape found in terrestrial ecosystems, few customized automated acoustic approaches exist for land mammals. Nevertheless automated systems have the potential to provide real-time data on animal populations and thus would be an invaluable tool for wildlife managers (Blumstein et al., 2011; Aide et al., 2013).

Occupancy probability is the probability that a species is present at a site and is the most basic unit of information field practitioners can use to study and monitor wild populations (MacKenzie et al., 2006). It requires fewer assumptions than estimating abundance when using indirect cues such as vocalizations, and addresses the central problem in PAM of failing to detect a species when

it is actually present, also known as imperfect detection probability (MacKenzie et al., 2002). To our knowledge, this is the first study to apply an automated detection and classification algorithm specifically aimed at primate vocalizations to determine occupancy probabilities of multiple species. Bat researchers, for which there exists commercially available software for automated species detection and recognition of calls, have similarly estimated occupancy probability for conservation implications using automated data collection (Yates and Muzika, 2006; Hein et al., 2009). For PAM to be applicable as a primate biomonitoring tool in the field, we created a system where the input data were continuous ARU recordings of the forest, and the output was a short audio clip associated with a probabilistic vocal class assignment; thereby not including any human intervention until the moment of validation and interpretation (for a detailed description of the algorithm see Heinicke et al., in press).

Our primary aim was to examine to what extent the automated algorithm output must be manually validated or ‘cleaned’ by a human expert in order to provide reliable information about primate species occurrence that could be used to guide field practitioners. Specifically, we investigated whether occupancy probabilities derived using an occupancy modelling framework improved naïve occupancy estimates for the ARU data, and compared this to human point transect data from the same area. Additionally, we assessed the influence of environmental covariates on vocalization detection probability at ARUs. Moreover, we investigated how robust our occupancy probability estimates were to using the automated output by simulating datasets with varying degrees of cleaning for false positives and false negatives in the ARU data. Finally, we interpret our findings to provide recommendations on how the current method could be applied as an early-warning system, and suggest improvements on how to enhance algorithm performance and optimize data collection for future surveys.

## 2. Materials and methods

### 2.1. Study site and data collection

The study was conducted in the rainforest of the Taï National Park, Côte d'Ivoire (for details see Boesch and Boesch-Achermann, 2000). From November 2011 through May 2012, 20 Wildlife Acoustics SM2+ ARUs were systematically placed within an area of 35–45 km<sup>2</sup> (Fig. 1) which encompassed the territories of two habituated chimpanzee (*Pan troglodytes verus*) communities, south and east group (Kouakou et al., 2011). There are also eight diurnal monkey species inhabiting Taï forest, including the diana monkey (*Cercopithecus diana*), red colobus (*Procolobus badius*), and the black and white colobus, also known as the king colobus (*Colobus polykomos*) (McGraw and Zuberbühler, 2007). Of the nine primate species present at Taï, we focussed on only a subset of three species (chimpanzee, diana monkey, and king colobus) due to limitations in development (i.e., lack of sufficient training data for the other primates), and performance of the automated sound recognition algorithm for primate vocalizations used in this study (also see Section 2.1.2). The Taï chimpanzees regularly engage in hunting red colobus monkeys, the most frequently hunted prey item, as well as king colobus, the second most hunted prey item (Boesch and Boesch-Achermann, 2000; McGraw and Zuberbühler, 2007). The chimpanzee communities were 26.5 km<sup>2</sup> for the south group and 30.1 km<sup>2</sup> for the east group (Kouakou et al., 2011) with 25 and 34 individuals (including dependent offspring), respectively, at the end of the study period. The monkeys are also territorial but have overlapping home ranges

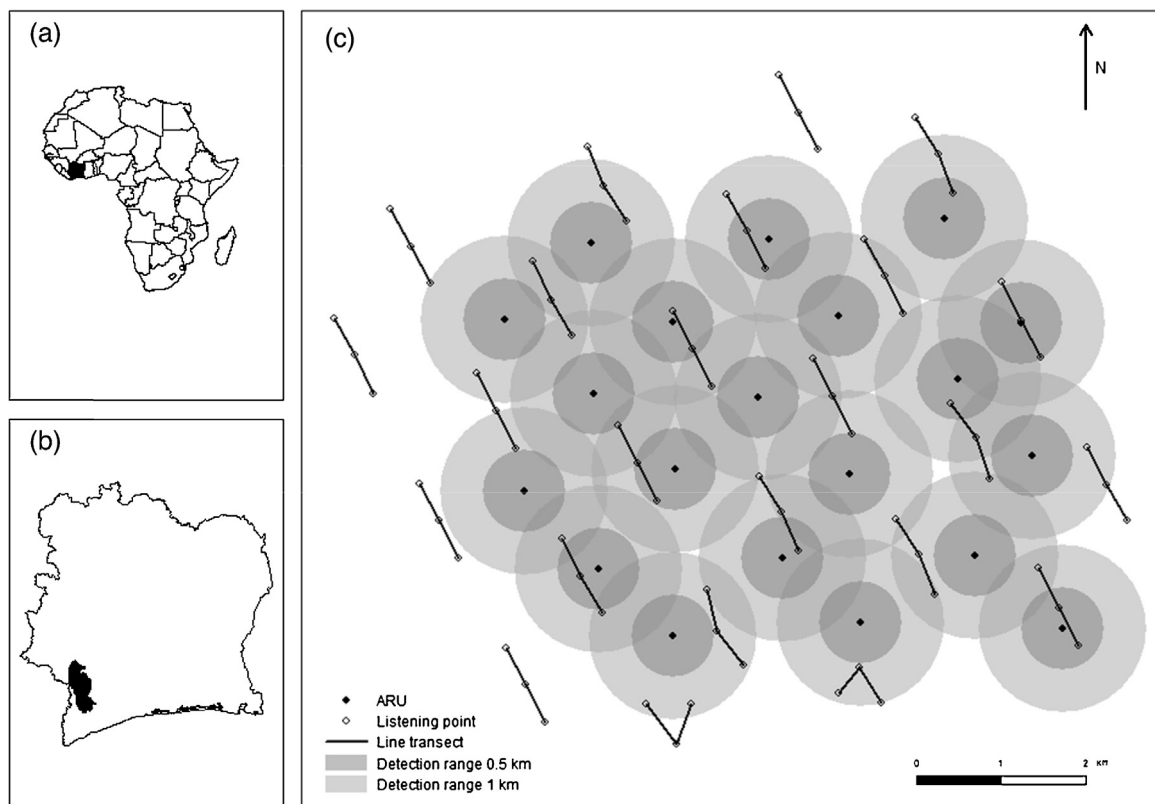


Fig. 1. Map of the study area located in Côte d'Ivoire (a) within the Taï National Park (b) as well as the arrangement of ARUs and point transects (c).

of on average  $0.63 \text{ km}^2$  for diana monkey and  $0.78 \text{ km}^2$  for king colobus (McGraw and Zuberbühler, 2007). All three primate species are considered threatened (IUCN, 2014).

#### 2.1.1. ARU data

During a pilot study in 2010 we determined that ARUs should be spaced at a minimum of one kilometre apart from each other to ensure that a call, for any of the primate species considered, detected on one device would not be detected on another. ARUs were therefore spaced 1.2–2 km apart. Only drumming sounds by chimpanzees had the potential to reach more than one ARU since drums have a maximum detection distance of up to 1 km by ARUs (Fig. 1) as well as human ears (Arcadi et al., 1998). Recordings were made in stereo at 16 kHz sampling rate and 16 bits/s and saved onto 32GB SD memory cards in uncompressed wav format. The sensitivity of microphones was calibrated upon deployment to ensure consistency across devices. ARUs were programmed to record for 30 min on the hour from 6 am to 5 pm. At times ARUs stopped recording due to high humidity or damage to the hardware caused by hard rain, tree falls and insects. In total, we collected 12 851 h of audio data during the entire seven-month study period (Fig. 2).

#### 2.1.2. Algorithm for detection of primate sounds

We developed an algorithm to automatically detect and classify chimpanzee drumming and vocalizations, king colobus and diana monkey loud calls, and the red colobus contact call (Heinicke et al., in press). The algorithm was trained using a total of 72 h of annotated sounds which were 351 chimpanzee drums, 228 chimpanzee vocalizations, 223 diana monkey loud calls, 61 king colobus loud calls, and 158 red colobus contact calls. There was also an explicit background class used for training which encompassed any sounds not of interest, including other species, thunder and airplanes. The algorithm was developed in collaboration with the Fraunhofer IDMT using their methods for pattern recognition,

feature extraction and sound segmentation (Heinicke et al., in press). The automated algorithm consisted of multiple steps: (1) sound source separation, (2) a low pass filter of 400 Hz for chimpanzee drum sounds, (3) feature extraction for signal classification in combination with (4) a sound segmentation approach to identify the onset and ends of sounds of interest (for full details see Heinicke et al., in press). All parts of the algorithm were implemented using MATLAB (MathWorks, 2011).

A validation study carried out on a subset of 179 ARU-hours, which were not used for the training of the algorithm, showed that the algorithm performed best for the loud calls of king colobus and diana monkeys, followed by chimpanzee drumming (Heinicke et al., in press). Therefore, this study focused on those three species which held the greatest potential for PAM. The algorithm works by scanning the audio data using 30 ms window frames to determine onsets and endpoints of the targeted sounds while simultaneously assigning a probabilistic classification to each frame. Neighbouring frames of the same class and signal are merged together to form a segment or a single detection (Heinicke et al., in press).

We used the same 179 h ARU dataset for this study with algorithm parameters set at a 10% output rate and 10-segment limit.<sup>1</sup> These 179 h were composed of 358 files of 30 min long ARU recordings randomly selected across the seven months. We strived to obtain a representative dataset of ARU recordings by balancing across devices for time of day (one of three blocks: 6–9; 10–13; 14–17) while taking one 30 min file from each ARU once every seven days giving a total of 21 sampling days per device. This ensured temporally and spatially balanced sampling effort across

<sup>1</sup> Output rate corresponds to the proportion of detections permitted as background class (e.g. 10% output means 90% of all detections are classified as background) and the segment limit refers to the restriction of the number of consecutive detections classified as the same sound (Heinicke et al., in press).

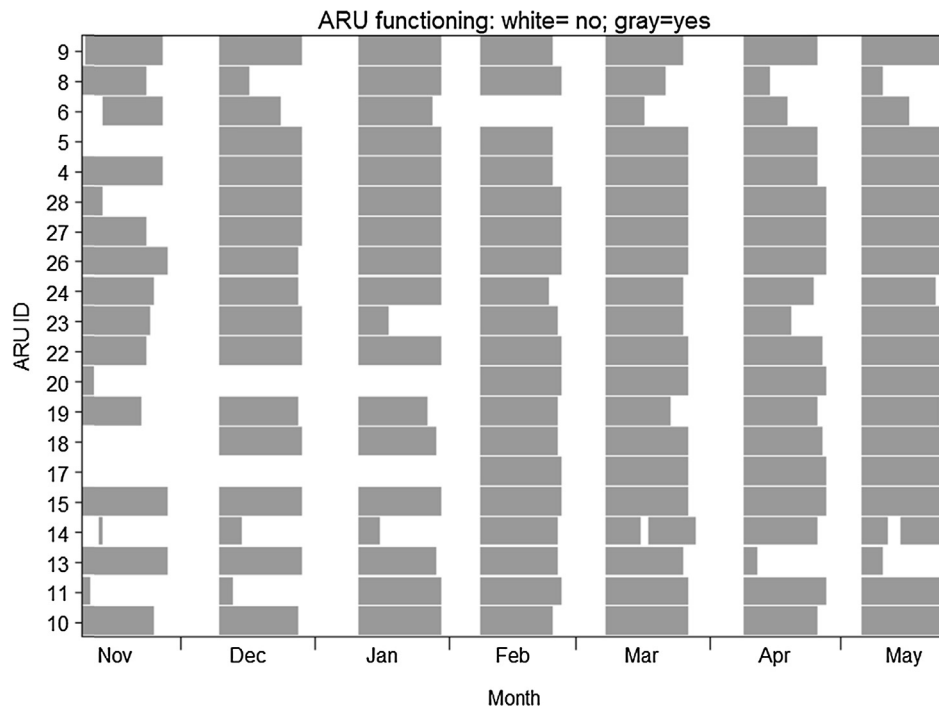


Fig. 2. Overview of the time span of recordings for the 20 ARUs across the 7 month study period ( $N = 12\,851$  h).

the study period. Additionally, files selected for neighbouring ARUs were not of the same hour to avoid detecting the same chimpanzee drum on more than one device.

### 2.1.3. Human transect data

Every year, a large mammal census is conducted across the Tai National Park as part of an on-going biomonitoring programme which uses a systematic transect design and trained surveyors. Twenty-four of the 1 km long biomonitoring transects overlapped with the ARU grid (Fig. 1). The 24 transects were similarly spaced as the 20 ARUs, approximately 1–1.8 km apart from one another and were surveyed once in 2011 and again in 2012. Point transects were conducted at the beginning, middle and end of each transect where a team, of on average three experts, stopped for 30 min to listen for primate vocalizations as well as sightings. Overall this amounted to 207 person-hours of listening effort (1.5 h per transect  $\times$  3 surveyors  $\times$  24 transects  $\times$  2 visits (minus 9 h because two transects were visited only once)). This is similar to the 179 ARU-hours. The two datasets were also collected during overlapping study periods.

## 2.2. Occupancy modelling

Occupancy is defined as the probability ( $\psi$ ) with which a species occurs in a study area (MacKenzie et al., 2006). Occupancy modelling is based on the detection history of a species over repeated visits to multiple sites to provide an estimate of the proportion of sites occupied by a species in the area sampled (MacKenzie et al., 2002). Detection probability ( $p$ ), the ability to detect a species correctly when it is present, is expected to vary due to environmental factors, the behaviour of a species, and observer bias. Occupancy models provide an estimate of detection probability to correct for underestimating occurrence when detection did not occur, which is essential when using indirect indices for species detection such as vocalizations. Thus, occupancy models provide estimates of two parameters of interest: detection probability ( $p$ ) and occupancy probability ( $\psi$ ), along with estimates for any covariates included

to account for heterogeneity in probabilities across sampling sites and observation periods (MacKenzie et al., 2006).

The major assumptions to fit a single-season occupancy model are a closed population during the study period, which is given for territorial primates, and that detection at one site is independent from detection at another. Due to the small home ranges of the monkeys, individuals of the same group would rarely be heard at more than one device. With respect to the chimpanzees we assumed that their movements were more or less random throughout their large territories and therefore occupancy would remain unbiased to any particular ARU (MacKenzie et al., 2006). Since we evaluated this method using ARU data collected in an area for which we know these primates exist, albeit in varying densities and distributions, we predicted that occupancy probability should generally be high for all primates considered but detection probability was expected to vary due to species-specific differences in behavioural ecology, call rate, and signal amplitude.

### 2.2.1. Datasets and definitions

We calculated occupancy probability for the ARU data using two levels of human validation effort: a half-cleaned and a fully cleaned dataset. The half-cleaned ARU dataset was constructed as follows: a human expert listened to all audio clips classified by the algorithm that contained one of the three primate species and marked them as a true positive (TP) when the algorithm correctly identified a call or a false positive (FP) when the algorithm was wrong. Importantly, such a half-cleaned dataset is an easily obtained method for improving the automated data output because it takes little effort to verify 30 ms long audio clips and therefore offers important benefits in time-saving (Swiston and Mennill, 2009). For comparison, we included a fully cleaned ARU dataset, where a human expert has listened to all 179 h of continuous ARU recordings, thereby accounting for calls missed by the algorithm or false negatives (FN). Although this is practically not a feasible option we used it in this study to determine how well occupancy probabilities estimated from a half-cleaned dataset approximate a fully cleaned, human verified dataset. Moreover, we compared estimates of occurrence



probability derived from occupancy models with the naïve occupancy probability for the fully cleaned ARU and human point transect data to determine which estimate(s) was more accurate for each species. The naïve occupancy probability assumes no variation in detection probability (MacKenzie et al., 2006) and is equivalent to the number of sites (ARUs ( $N=20$ ) or point transects ( $N=24$ )) where at least one positive detection occurred for a species out of the total number of sites sampled during the study period.

### 2.2.2. Creating species detection histories for each ARU

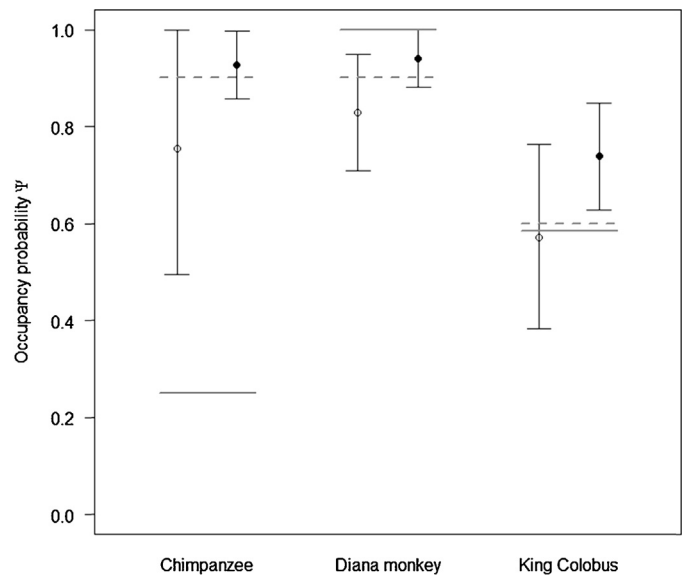
Detection histories were created separately for each species, for each ARU, across the 21 sampling days. Every 30 min file was considered one sampling visit at the ARU. We assigned a TP, FP or FN to the algorithm's detection for each species. Only one type of detection was permitted for each species per file. If more than one type of detection occurred in a file for a particular species, the detection was assigned according to what was most informative practically (i.e., priorities were TP > FP > FN). For a half-cleaned dataset, TP files were assigned a positive detection or a '1' and all FP files were corrected to non-detection or a '0'. For a fully cleaned dataset both TP and corrected FN files were assigned a '1' and all FP files corrected to '0'.

### 2.2.3. Model development

Variation in detection and occupancy probability were modelled using covariates. Site-level covariates included the dummy-coded variables ARU position: core ( $N=9$ ) or periphery ( $N=11$ ) of a chimpanzee territory and in which chimpanzee group the ARU was located: south group ( $N=11$ ), east group ( $N=6$ ) or the border between the two ( $N=3$ ). Data for assigning site-level covariates came from previous studies of the Tai chimpanzees (Herbinger et al., 2001; Kouakou et al., 2011). Observation-level covariates included daily mean temperature, daily rainfall, and the hour the recording was made. Position was a predictor for occupancy probability because chimpanzee behaviour can be different in the core versus the periphery of the territory (Herbinger et al., 2001; Wilson et al., 2007) and this could indirectly affect the occurrence of the other two species via predator–prey dynamics or resource competition. Similarly, group was a predictor for detection probability since variation in chimpanzee density might influence chimpanzee drumming rates which could indirectly influence the vocal behaviour of the monkeys (i.e., call less when chimpanzees are present to avoid being detected by them). Temperature, rainfall, hour and hour<sup>2</sup> were also modelled on detection probability to account for diurnal patterns in vocalization activity (Sekulic, 1982) and the effects of abiotic factors on call propagation through a rain-forest (Waser and Waser, 1977). Due to our relatively low sample size of TPs we could not fit models with interactions, although we expect these variables to interact in nature, and could therefore draw inference at the level of main effects only. With the six predictor variables we constructed 48 possible candidate models; one with each possible combination of the predictors (models including hour<sup>2</sup> without hour were not included).

### 2.2.4. Model implementation

We fit single-season occupancy models on the ARU data, separately for each species, using the 'occu' function of the package "Unmarked" (Fiske and Chandler, 2011) in R version 3.0.2 (R Core Team 2013). All observation-level covariates were z-transformed to a mean of 0 and standard deviation of 1 before running the models. We ran all 48 models on a half-cleaned and fully cleaned ARU dataset separately for each species and in all cases all models converged. As recommended by MacKenzie et al., 2006, we tested the goodness of fit by comparing the Chisq statistic for the full model to the respective reference distributions derived using the 'parboot' function in "Unmarked" (Fiske and Chandler, 2011) using 1000



**Fig. 3.** Occupancy probability ( $\psi$ ) estimates ( $\pm$ SE) for the best model (see Table 1) for each species using a half-cleaned ARU dataset (open circle) and a fully cleaned ARU dataset (filled circle) compared to the naïve ARU occupancy probability (dashed horizontal line) and the naïve occupancy probability for the human point transect data (solid horizontal line) in the same study area.

simulations. For all species there was no indication of a lack of fit (all  $P > 0.05$ ).

For each dataset, and for each species, the candidate models were ranked according to Akaike's Information Criterion (AIC) where we considered top-ranked models as those with  $\Delta AIC < 2$  (Burnham and Anderson, 2002). For the half-cleaned and fully cleaned datasets we also report the best models ( $\Delta AIC = 0$ ). Note that 'occu' models produce estimates and standard errors for both occupancy and detection probability on the logit scale. Hence, estimates and their respective SE were transformed to the original scale using the functions 'backTransform' and 'predict' of the package "Unmarked" (Fiske and Chandler, 2011). We reported summed Akaike weights ( $\sum w_{AIC}$ ) to assess the amount of relative support of each covariate on either detection or occupancy probability.

### 2.3. Evaluating the amount of human validation necessary

We tested varying degrees of cleaning to determine how much human validation effort should be invested to approximate occupancy estimates of a fully cleaned ARU dataset. For each species, a proportion of FPs and FNs, ranging from 0 to 100% (by increments of 2–4 FN files and 7–10 FP files), were randomly chosen to be corrected (repeated 1000 times for each species) to generate datasets representing various efforts of manual cleaning which were then analyzed using 'occu' for all 48 candidate models (as described above). Again, models were ranked according to their AIC to calculate model-averaged estimates of occupancy probability.

## 3. Results

The naïve occupancy from the ARU data and the human point transect data were similar for all species except the chimpanzee (Fig. 3). The transect data gave a naïve occupancy probability for chimpanzees which was considerably lower than that from ARUs. For all species, except the king colobus, the naïve ARU occupancy was within the occupancy probability estimate of the best model for half-cleaned and fully cleaned ARU datasets. Only for the king colobus we observed that failing to account for detection probability could lead to an underestimate of 'true' occupancy (the fully

**Table 1**  
The best model, according to AIC, for each species, using the half-cleaned and fully cleaned ARU dataset ( $\psi$ , occupancy probability;  $p$ , detection probability; wAIC, Akaike weight; NPar, number of parameters; SE, standard error).

Species	Data cleaning	Best model	wAIC	NPar	Occupancy $\psi$	SE ( $\psi$ )	Detection $p$	SE ( $p$ )
Chimpanzee	Half	$\psi(\cdot) p(\text{Hour} + \text{Rain})$	0.109	4	0.755	0.26	0.035	0.02
Chimpanzee	Full	$\psi(\cdot) p(\text{Hour} + \text{Hour}^2)$	0.133	4	0.927	0.07	0.138	0.03
Diana monkey	Half	$\psi(\cdot) p(\text{Hour} + \text{Hour}^2 + \text{Temp} + \text{Group})$	0.171	7	0.829	0.12	0.101	0.03
Diana monkey	Full	$\psi(\cdot) p(\text{Temp} + \text{Group})$	0.245	5	0.940	0.06	0.176	0.02
King colobus	Half	$\psi(\cdot) p(\text{Hour})$	0.136	3	0.572	0.19	0.061	0.02
King colobus	Full	$\psi(\text{Position}) p(\cdot)$	0.137	3	0.738	0.11	0.098	0.02

Notes: Probabilities were calculated while fixing covariates at their mean value. See also Tables S1 and S2 for detailed results for all models and covariates for each dataset.

cleaned ARU data; Fig. 3). Moreover, a fully cleaned dataset gave more precise occupancy probability estimates (i.e., narrower SEs) for all species. The best fitting model for each species was different although best fit models generally had small Akaike weights (Table 1) suggesting there are potentially additional covariates one could consider. Additionally, we assessed the robustness of our derived occupancy probability estimates and found that using subsets of the full dataset gave estimates ( $\pm$ SE) very similar to the full, 7 month dataset, but in many cases we had trouble fitting models with a lower number of sampling days due to the reduction in sample size (Fig. S3).

For two of the three species the null model was not in the top-ranked models ( $\Delta\text{AIC} < 2$ ; Burnham and Anderson, 2002), namely chimpanzee drums and diana monkey loud calls, while only for diana monkeys it was not in the 95% model confidence set (Mundry, 2011; Table S1). The model-averaged estimates for all predictors for each species and dataset are provided in Table S2. Time of day, or hour, appeared to be the most influential covariate on detection probability for half-cleaned datasets for all species (Table S4). Also rain reduced the algorithm's performance for detecting chimpanzee drums which was likely due to rain masking the low frequency drum sound (Table S4). Meanwhile, detection probability modelled using a fully cleaned dataset, more closely reflected the vocal behaviour of the species since it was not limited by the ability of the algorithm to detect calls. Hour obtained the most substantial support for the detection probability of chimpanzee drums whilst temperature and group had strong influences on diana monkey loud call detection probability (Table 2). Detection probability was highest in the morning for chimpanzee drums (Fig. 4a). For diana monkey loud calls detection probability increased with

increasing daily mean temperature and was generally higher in the south group and the border between the two chimpanzee groups (Fig. 4b and c). Position did not obtain much support for affecting occupancy probability for either the chimpanzee or diana monkey.

The evaluation of various degrees of cleaning for FPs and FNs demonstrated that occupancy probability estimates appeared to be quite robust (Fig. 5). However, cleaning for FPs alone considerably lowered precision and led to an underestimate of occupancy, which was especially detrimental with respect to the chimpanzee and diana monkey but less so for the king colobus. Alternatively, cleaning for only FNs had little to no effect for all species. If no cleaning was done the occupancy estimates for chimpanzees and diana monkeys were very close to the fully cleaned occupancy estimate which was likely a consequence of 'true' occupancy being very high for both these species. For king colobus, an uncleaned dataset clearly overestimated occupancy which may be a result of king colobus having the most unbalanced ratio of FPs to FNs in the data, with 10 times as many FPs as FNs (chimpanzees: 1.5; diana monkey: 3.7).

## 4. Discussion

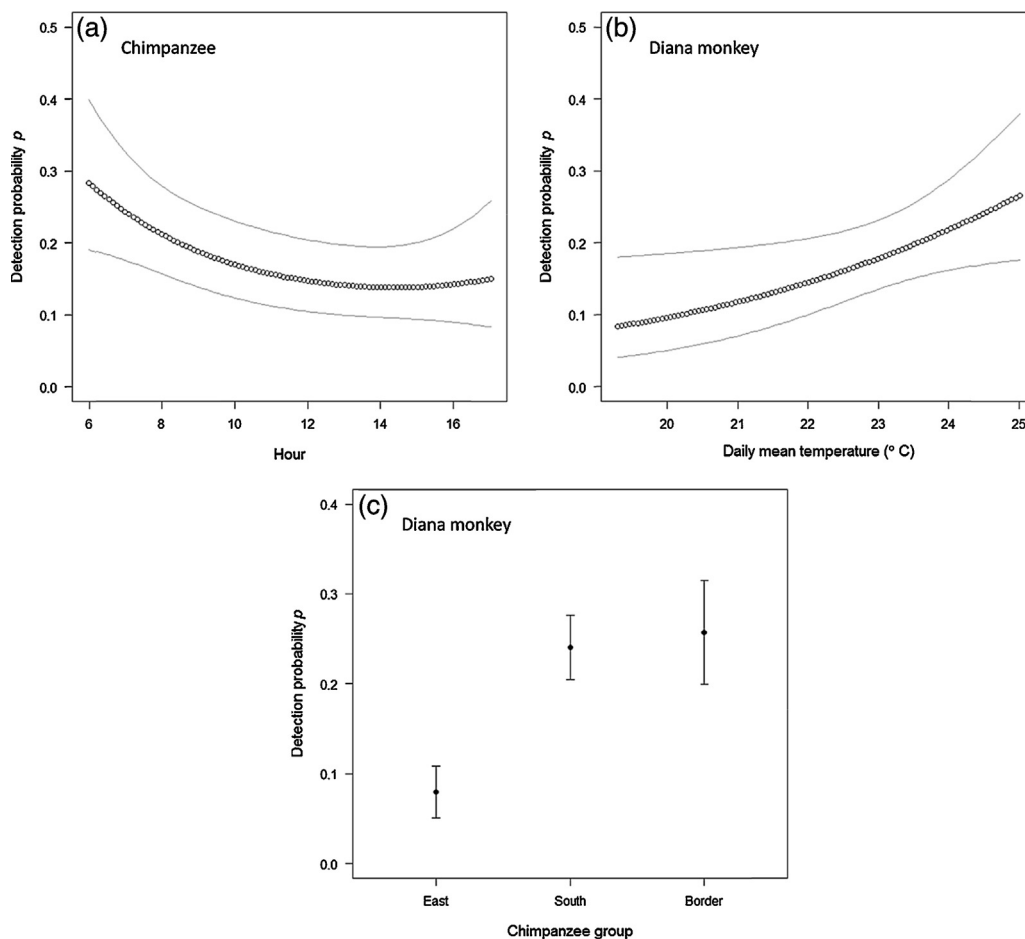
### 4.1. Occupancy probabilities approximated naïve estimates from ARUs and point transects

Generally, ARU data gave comparable occupancy probabilities (naïve and model estimates) as human point transect data but were more accurate for chimpanzees. The human point transect data would have considerably underestimated the presence of resident chimpanzees in the study area thereby highlighting the necessity for longitudinal data when surveying species with large home ranges. From a PAM perspective, this may be important when attempting to monitor species with seasonally low call rates or species that adjust vocal behaviour in response to human presence, such as chimpanzees (Hicks et al., 2013). PAM has already been shown to be an effective alternative or complementary method when surveying cryptic or rare bird species (Celis-Murillo et al., 2012; Oppel et al., 2014), and this study shows that it may be a suitable method for primates. However, with the need for further research on data mining and extraction algorithms for the automatic recognition of primate sounds, we do not suggest that PAM can replace traditional survey methods but should be used to complement and improve the accuracy of existing approaches. Although line transects are currently the most widely used method for primate surveys, data from PAM can also be used to ascertain primate occurrence and could even help to guide transect surveys to better allocate resources on the ground. Moreover, PAM data can be included as a predictor of primate presence to improve density distribution and abundance models. The major benefits of PAM arise from its power as a non-invasive, cost-efficient, longitudinal data collection tool that can be objectively and repeatedly analyzed (Blumstein et al., 2011; Kühl and Burghardt, 2013).

Our study suggests that accounting for detection probability using occupancy models for ARU data becomes more pertinent

**Table 2**  
Summed Akaike weights ( $\sum \text{wAIC}$ ) for each covariate in models of detection and occupancy probability for each species using the fully cleaned ARU datasets.

Parameter	Covariate	Summed Akaike weights
<i>Chimpanzee drum</i>		
Occupancy $\psi$	Position	0.28
Detection $p$	Hour <sup>2</sup>	0.57
	<b>Hour</b>	<b>0.92</b>
	Temp	0.35
	Rain	0.27
	Group	0.27
<i>Diana monkey loud call</i>		
Occupancy $\psi$	Position	0.35
Detection $p$	Hour <sup>2</sup>	0.12
	Hour	0.40
	<b>Temp</b>	<b>0.87</b>
	Rain	0.27
	<b>Group</b>	<b>0.99</b>
<i>King colobus loud call</i>		
Occupancy $\psi$	<b>Position</b>	<b>0.68</b>
Detection $p$	Hour <sup>2</sup>	0.19
	Hour	0.50
	Temp	0.30
	Rain	0.28
	Group	0.18



**Fig. 4.** Effect of the most influential covariates (model-averaged estimates) of detection probability ( $p$ ) for the fully cleaned ARU data: (A) hour on chimpanzee drums, (B) daily mean temperature ( $^{\circ}\text{C}$ ), and (C) chimpanzee group on diana monkey loud calls. The grey lines and bars represent 95% CI (see also Table 2). Plots were created using the 'predict' function in "Unmarked" (Fiske and Chandler, 2011) which calculates model-averaged predictions of detection probability for every value of the covariate while holding all other covariates in the models constant at their mean value.

when 'true' occupancy is relatively low (e.g., king colobus which vocalize least often and are more cryptic than diana monkeys). Otherwise, naïve ARU occupancy probability approximated estimates derived from occupancy models which corrected for detection. We suggest this is likely a result of the quality of the ARU data, particularly its longitudinal temporal coverage at a single location which reduces the chance of failing to detect a species if it is actually present. This has important practical implications since it implies that long-term monitoring of naïve ARU occupancy probability for some species could provide repeated occupancy estimates over a long period of time which allows for inferring population trends (Kühl and Burghardt, 2013).

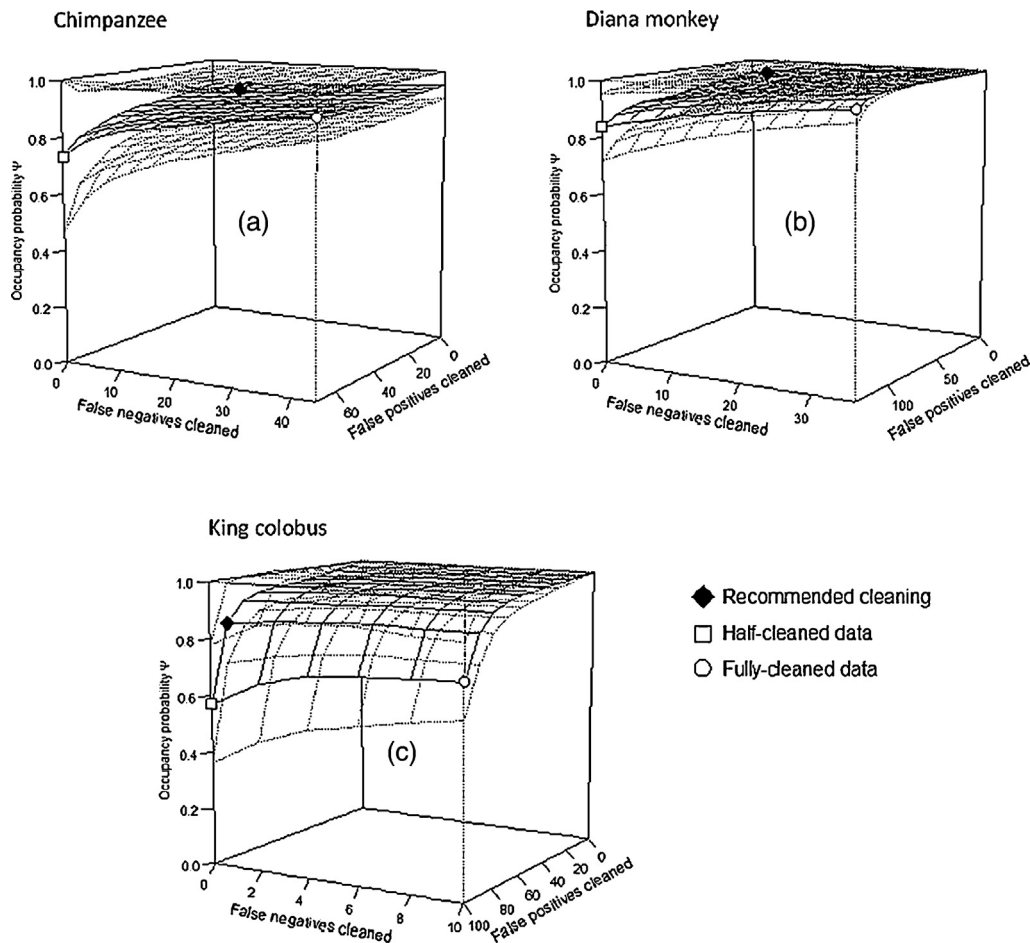
#### 4.2. PAM offers benefits in time-saving although some cleaning is needed

The method requires some human intervention otherwise occupancy probabilities would be overestimated since the automated algorithm as it stands now has such a high rate of FPs for all species. Our results demonstrate that the current method requires some degree of cleaning to provide reliable occupancy information but this depends on species-specific performance of the algorithm. In particular, the ratio of FPs to FNs appeared to affect the degree to which one needs to clean the automated output, and this could be inferred by practitioners using the performance statistics of the algorithm (e.g., percent recall and precision for a call type; Table S6). For example, in this study recall was high for king colobus

where one should clean 90% of the FPs (Fig. 5c). This translates to an effort of only 23 min (4 FP detections can be validated in a minute). The recommended cleaning level for diana monkey would require approximately 40 h to verify 8% of the FPs and 11% of the FNs (almost 20 h of ARU data) whereas for chimpanzees one needs 65 h (18% FNs and 9% FPs; Fig. 5a and b). This is still less human effort compared to walking the 24 line transects in this study which requires, at minimum, 12 days (half a day to complete a 1 km transect).

#### 4.3. Effects of covariates on call detection

Using occupancy modelling we were able to investigate how particular covariates affected call detection probability for the various species. This can improve future survey designs to maximize positive detections when logistics might be difficult, especially in remote areas. The influence of covariates for the king colobus lacked support in comparison to the null model (Table S1) therefore we do not speculate on their potential effect (Mundry, 2011). Overall, the ability for the algorithm to detect calls was clearly affected by the time of day (Table S4) which was likely a function of the ambient noise level in the rainforest as well as diurnal vocalization patterns of the species (Waser and Waser, 1977). With respect to detection probability of chimpanzee drums, previous studies have shown that drumming functions to coordinate group travel (Arcadi et al., 1998), and that drum production often peaks in the morning (unpublished data; Fig. S5) which is consistent with the pattern



**Fig. 5.** Model-averaged occupancy probability ( $\psi$ ) estimates ( $\pm$ SE) for the ARU datasets created by cleaning various combinations of percentages of false positives (FP) and false negatives (FN;  $N=1000$  simulations per species). The actual number of FPs and FNs for each species are displayed on the axes. Recommended cleaning levels (chimpanzee: FP=7, FN=8; diana monkey: FP=10, FN=4; king colobus: FP=90, FN=0) are based on minimizing time effort for validation while obtaining an occupancy probability within the standard error of a fully cleaned dataset.

we observed for drum detection (Fig. 4a). Surprisingly, we did not observe any effect of position on occupancy probability. We suspect that this may be due to recent shifts in the home ranges of the two habituated chimpanzee groups. However, the null model for chimpanzees did have some support ( $\Delta$ AIC 2–7; Burnham et al., 2011); therefore we caution that our interpretation is preliminary and the effects of covariates on chimpanzee drum detection should be further investigated.

Diana monkey detection probability was strongly affected by temperature and group. The diana monkey loud call is produced in instances of alarm, but it is also produced in response to other stimuli, such that there appears to be no clear pattern to its production (Zuberbühler et al., 1997). Thus it is not surprising that diana monkey loud calls were not a function of time of day but instead detection increased as average daily temperature increased (Fig. 4b). Theoretically, an increase in temperature favours call propagation but the relationship between sound attenuation and temperature is complex (Waser and Waser, 1977) and there could be indirect ways for temperature to promote loud call production, such as changes in activity patterns (Sekulic, 1982). The other covariate to affect diana monkey call detection was group, whereby detection probability was lowest in the east territory (Fig. 4c). There may be a lower density of diana monkeys in the east; although we find this unlikely since biomonitoring surveys consistently found the greatest monkey density in the research areas (Campbell et al., 2011; N'goran et al., 2012). Alternatively, during this study period, we observed repeated instances of poaching activity in the east

territory (*personal observations*); therefore, diana monkeys may have adapted their vocal behaviour in response to the threat of poaching. Unfortunately, we did not have detailed data to include this as a predictor in our models. However, previous studies at Taï demonstrated that the diana monkey is indeed flexible in its vocal behaviour, becoming quieter when poaching pressure becomes more intense (Kone and Refisch, 2007). The authors suggested that diana monkey vocal behaviour could be a reliable indicator of poaching pressure in Taï National Park, whereby longitudinal monitoring of call rates could guide management interventions. We advocate the use of the semi-automated method presented here for such a purpose. ARUs could monitor diana monkey loud calls and provide occupancy probability estimates that are more or less continuously updated, working as an early-warning system. Changes in call rate and latency in calling events could also be used as early-warning statistics. Additionally, commercial technologies for detecting gun shots (ShotSpotter: <http://www.shotspotter.com/>) and recent advancements for detecting illegal logging activity in real-time (Rainforest Connection: <https://rfcx.org/>) already exist and could be adopted into a PAM framework to further aid in protecting wild populations.

#### 4.4. Future applications

We have presented an application of a semi-automated method for detecting the presence of multiple primate species using PAM in a rainforest. Although some data cleaning is still required, for



some species, namely the diana monkey and king colobus, the system provides reliable occurrence data on wild populations and is ready for software deployment provided a user-friendly interface can be incorporated. However, there are continued improvements to be made, such as integrating self-learning techniques to continuously re-train the algorithm and determine optimum precision and recall for a given species. The algorithm could also be fine-tuned to provide equal rates of FP and FN detections to allow for more accurate inferences requiring less cleaning effort. In the field, surveys can be designed to maximize TP detections while minimizing FPs and FNs but this requires careful consideration of species-specific vocal behaviour (Fig. S5). Additionally, one can improve the automated output by using data mining and post-processing techniques (Table S6). Hence, there remain many avenues of research to improve the current system in order to reach a fully automated method. We stress that this will only be possible with continued collaborations amongst field researchers, animal vocalization experts and sound engineers. We hope this study motivates further research on the potential for PAM in applied primate conservation.

## Acknowledgements

This study was conducted as part of the SAISBECO project, a collaboration funded by the Pact for Research and Innovation between the Max Planck Society and the Fraunhofer-Gesellschaft. Specifically, we thank the Fraunhofer IDMT, particularly Hannah Lukashevich for support with algorithm development. We thank the Ministère de la Recherche Scientifique, the Ministère de l'Environnement et des Eaux et Forêts and the Office Ivoirien des Parcs et Réserves of Côte d'Ivoire for permissions to work in the country and Taï National Park. We gratefully acknowledge the logistical support from the Centre Suisse de Recherche Scientifique and the Wild Chimpanzee Foundation. A special thank you to the Taï Chimpanzee Project, especially field assistants and Roman Wittig. Additional thanks to the Robert Bosch Foundation and we also thank two referees for helpful comments that improved this manuscript.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ecolind.2015.02.023>.

## References

- Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1, e103.
- Arcadi, A.C., Robert, D., Boesch, C., 1998. Buttress drumming by wild chimpanzees: temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates* 39, 505–518.
- Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J.L., Krakauer, A.H., Clark, C., Cortopassi, K.A., Hanser, S.F., McCowan, B., Ali, A.M., Kirschel, A.N.G., 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *J. Appl. Ecol.* 48, 758–767.
- Boesch, C., Boesch-Achermann, H., 2000. *The Chimpanzees of the Taï Forest: Behavioural Ecology and Evolution*. Oxford University Press, Oxford, UK.
- Burnham, K.P., Anderson, D.R., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, NY.
- Burnham, K.P., Anderson, D.R., Huyvaert, K.P., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* 65, 23–35.
- Campbell, G., Kühl, H., Diarrassouba, A., N'Goran, P.K., Boesch, C., 2011. Long-term research sites as refugia for threatened and over-harvested species. *Biol. Lett.* <http://dx.doi.org/10.1098/rsbl.2011.0155>.
- Celis-Murillo, A., Deppe, J.L., Ward, M.P., 2012. Effectiveness and utility of acoustic recordings for surveying tropical birds. *J. Field Ornithol.* 83, 166–179.
- Chesmore, E.d., Ohya, E., 2004. Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition. *Bull. Entomol. Res.* 94, 319–330.
- Digby, A., Towsey, M., Bell, B.D., Teal, P.D., 2013. A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods Ecol. Evol.* 4, 675–683.
- Fiske, I.J., Chandler, R.B., 2011. *Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance*. *J. Stat. Softw.* 43, 1–23.
- Guschanski, K., Vigilant, L., McNeillage, A., Gray, M., Kagoda, E., Robbins, M.M., 2009. Counting elusive animals: comparing field and genetic census of the entire mountain gorilla population of Bwindi Impenetrable National Park, Uganda. *Biol. Conserv.* 142, 290–300.
- Hein, C.D., Castleberry, S.B., Miller, K.V., 2009. Site-occupancy of bats in relation to forested corridors. *For. Ecol. Manage.* 257, 1200–1207.
- Heinicke, S., Kalan, A.K., Wagner, O.J.J., Mundry, R., Lukashevich, H., Kühl, H.S., 2015. Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods Ecol. Evol.* (in press).
- Herbinger, I., Boesch, C., Rothe, H., 2001. Territory characteristics among three neighboring chimpanzee communities in the Taï National Park, Côte d'Ivoire. *Int. J. Primatol.* 22, 143–167.
- Hicks, T.C., Roessingh, P., Menken, S.B.J., 2013. Impact of humans on long-distance communication behaviour of eastern chimpanzees (*Pan troglodytes schweinfurthii*) in the northern Democratic Republic of the Congo. *Folia Primatol. (Basel)* 84, 135–156.
- IUCN, 2014. IUCN Red List of Threatened Species. [www.iucnredlist.org](http://www.iucnredlist.org)
- Kone, I., Refisch, J., 2007. Can monkey behavior be used as an indicator for poaching pressure? A case study of the Diana guenon (*Cercopithecus diana*) and the western red colobus (*Procolobus badius*) in the Taï National Park, Côte d'Ivoire. In: McGraw, W.S., Zuberbühler, K., Noë, R. (Eds.), *Monkeys of the Taï Forest: An African Primate Community*. Cambridge University Press, Cambridge, UK, pp. 257–289.
- Kouakou, C.Y., Boesch, C., Kühl, H.S., 2011. Identifying hotspots of chimpanzee group activity from transect surveys in Taï National Park, Côte d'Ivoire. *J. Trop. Ecol.* 27, 621–630.
- Kühl, H.S., Burghardt, T., 2013. Animal biometrics: quantifying and detecting phenotypic appearance. *Trends Ecol. Evol.* 28, 432–441.
- 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.L., Hines, J.E., 2006. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Academic Press, Burlington, MA.
- Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti, D.J., Harris, D., Tyack, P.L., 2013. Estimating animal population density using passive acoustics. *Biol. Rev.* 88, 287–309.
- MathWorks, 2011. *MATLAB 7.13*. The MathWorks Inc., Natick.
- McGraw, W.S., Zuberbühler, K., 2007. The monkeys of the Taï forest: an introduction. In: McGraw, W.S., Zuberbühler, K., Noë, R. (Eds.), *Monkeys of the Taï Forest: An African Primate Community*. Cambridge University Press, Cambridge, UK, pp. 1–48.
- Mellinger, D.K., Clark, C.W., 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. *J. Acoust. Soc. Am.* 107, 3518–3529.
- Mellinger, D.K., Stafford, K.M., Moore, S.E., Dziak, R.P., Matsumoto, H., 2007. An overview of fixed passive acoustic observation methods for cetaceans. *Oceanography* 20, 36–45.
- Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., Bailey, L.L., Weir, L.A., 2011. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology* 92, 1422–1428.
- Mittermeier, R.A., Wallis, J., Rylands, A.B., Ganzhorn, J.U., Oates, J.F., Williamson, E.A., Palacios, E., Heymann, E.W., Kierulff, M.C.M., Yongcheng, L., Supriatna, J., Roos, C., Walker, S., Cortés-Ortiz, L., Schwitzer, C., 2009. Primates in peril: the world's 25 most endangered primates 2008–2010. *Primate Conserv.* 24, 1–57.
- Mundry, R., 2011. Issues in information theory-based statistical inference—a commentary from a frequentist's perspective. *Behav. Ecol. Sociobiol.* 65, 57–68.
- N'goran, P.K., Boesch, C., Mundry, R., N'goran, E.K., Herbinger, I., Yapi, F.A., Kühl, H.S., 2012. Hunting, law enforcement, and African primate conservation. *Conserv. Biol.* 26, 565–571.
- Oppel, S., Hervias, S., Oliveira, N., Pipa, T., Silva, C., Geraldes, P., Goh, M., Immler, E., McKown, M., 2014. Estimating population size of a nocturnal burrow-nesting seabird using acoustic monitoring and habitat mapping. *Nat. Conserv.* 7, 1–13.
- Van Parijs, S.M., Clark, C.W., Sousa-Lima, R.S., Parks, S.E., Rankin, S., Risch, D., van, I.C., Opzeeland, 2009. Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales. *Mar. Ecol. Prog. Ser.* 395, 21–36.
- Penone, C., Le Viol, I., Pellissier, V., Julien, J.-F., Bas, Y., Kerbiriou, C., 2013. Use of large-scale acoustic monitoring to assess anthropogenic pressures on Orthoptera communities. *Conserv. Biol.* 27, 979–987.
- Riede, K., 1998. Acoustic monitoring of Orthoptera and its potential for conservation. *J. Insect Conserv.* 2, 217–223.
- Sekulic, R., 1982. Daily and seasonal patterns of roaring and spacing in four red howler *Alouatta seniculus* troops. *Folia Primatol. (Basel)* 39, 22–48.
- Swiston, K.A., Mennill, D.J., 2009. Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *J. Field Ornithol.* 80, 42–50.
- Walters, C.L., Freeman, R., Collen, A., Dietz, C., Brock Fenton, M., Jones, G., Obrist, M.K., Puechmaille, S.J., Sattler, T., Siemers, B.M., Parsons, S., Jones, K.E., 2012. A continental-scale tool for acoustic identification of European bats. *J. Appl. Ecol.* 49, 1064–1074.

- Waser, P.M., Waser, M.S., 1977. Experimental studies of primate vocalization: specializations for long-distance propagation. *Z. Tierpsychol.* 43, 239–263.
- Wilson, M.L., Hauser, M.D., Wrangham, R.W., 2007. Chimpanzees (*Pan troglodytes*) modify grouping and vocal behaviour in response to location-specific risk. *Behaviour* 144, 1621–1653.
- Wimmer, J., Towsey, M., Roe, P., Williamson, I., 2013. Sampling environmental acoustic recordings to determine bird species richness. *Ecol. Appl.* 23, 1419–1428.
- Wrege, P.H., Rowland, E.D., Bout, N., Doukaga, M., 2012. Opening a larger window onto forest elephant ecology. *Afr. J. Ecol.* 50, 176–183.
- Wrege, P.H., Rowland, E.D., Thompson, B.G., Batruch, N., 2010. Use of acoustic tools to reveal otherwise cryptic responses of forest elephants to oil exploration. *Conserv. Biol.* 24, 1578–1585.
- Yates, M.D., Muzika, R.M., 2006. Effect of forest structure and fragmentation on site occupancy of bat species in Missouri Ozark Forests. *J. Wildl. Manage.* 70, 1238–1248.
- Zeppelzauer, M., Hensman, S., Stoeger, A.S., 2014. Towards an automated acoustic detection system for free-ranging elephants. *Bioacoustics*, 1–17.
- Zuberbühler, K., Noë, R., Seyfarth, R.M., 1997. Diana monkey long-distance calls: messages for conspecifics and predators. *Anim. Behav.* 53, 589–604.