Int J Primatol (2014) 35:841–858
DOI 10.1007/s10764-014-9761-9

Primates and Cameras

Noninvasive Sampling to Make Population-Level Inferences While Accounting for Imperfect Detection

Brian D. Gerber · Perry J. Williams ·
Larissa L. Bailey

Received: 3 May 2013 / Accepted: 17 October 2013 / Published online: 4 April 2014

Abstract Field-based primate studies often make population inferences using count-based indices (e.g., individuals/plot) or distance sampling; the first does not account for the probability of detection and thus can be biased, while the second requires large sample sizes to obtain precise estimates, which is difficult for many primate studies. We discuss photographic sampling and occupancy modeling to correct for imperfect detection when estimating system states and dynamics at the landscape level, specifically in relation to primate ecology. We highlight the flexibility of the occupancy framework and its many applications to studying low-density primate populations or species that are difficult to detect. We discuss relevant sampling and estimation procedures with special attention to data collection via photographic sampling. To provide tangible meaning to terminology and clarify subtleties, we use illustrative examples. Photographic sampling can have many advantages over observer-based sampling, especially when studying rare or elusive species. Combining photographic sampling with an occupancy framework allows inference to larger scales than is common in primate studies, addresses uncertainty due to the observation process, and allows researchers to examine questions of how landscape-level anthropogenic changes affect primate distributions.

Keywords Camera trap · Detection · Modeling · Occupancy · Primate

---

Electronic supplementary material The online version of this article (doi:10.1007/s10764-014-9761-9) contains supplementary material, which is available to authorized users.

B. D. Gerber (✉) · P. J. Williams
Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado 80523-1484, USA
e-mail: bgerber@colostate.edu

L. L. Bailey
Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado 80523-1484, USA

Introduction

Population studies of wild primates are logistically challenging and often costly. Estimating population size of primates can be difficult, as species often live in remote areas, where complex vegetation structure causes poor visibility and species are cryptic, highly mobile, and clustered into groups. Given the high costs and logistical limitations of rigorous primate density or abundance estimation, landscape-level inference is a challenge. As a result, primate studies often use count-based indices as a relative measure of abundance, and then apply these values to make inference to the landscape level (e.g., individuals/plot, individuals/transect; Ross and Reeve 2003). An inherent assumption of most abundance indices is that they are proportional to the true abundance (increasing or decreasing together), and that the proportion does not change across space, time, and sample units (Anderson 2001; Nichols 1992). Thus, indices do not account for varying detection probabilities due to a wide range of conditions. Though indices may be appealing because they are often less expensive and easier to implement than population state variables that account for detection, they can be significantly biased, leading to erroneous conclusions about a population's status (Nichols 1992). Depending on the direction (+ or –) and magnitude of bias, conclusions can lead to a species not receiving conservation action when needed, or misappropriated when unneeded actions are directed to a stable population of a difficult to detect species (Guschanski *et al.* 2009; MacKenzie *et al.* 2006; Zhan *et al.* 2006).

Line transect methods are also used to estimate absolute primate density (Buckland *et al.* 2010a). Many of these commonly used techniques, however, do not follow important survey design principles, and/or the analyses suffer from a lack of statistical rigor (see Buckland *et al.* 2010a, b). Even when idealized survey procedures are possible and a rigorous distance sampling approach is used, density estimation of rare or difficult to detect species (such as some primates) leads to considerable parameter uncertainty (poor precision). Common reasons for this include small sample bias in the detection function (Buckland *et al.* 2001) and uncertainty in group size (Buckland *et al.* 2010a). This lack of precision reduces the power to detect changes in population size (Plumptre 2000), limiting the use of distance sampling methods in studies of many primate populations. Additional challenges of primate detections via line transects occur when populations are heavily hunted, the species is nocturnal, or when a population occurs across difficult terrain, such as mountainous areas (Fashing and Cords 2000).

Without rigorous landscape scale studies, important spatial trends and effects on populations that are occurring at the landscape level may be overlooked. Because primate habitat is being lost globally (Chapman and Peres 2001), with remnant populations being isolated in fragmented and low quality habitat (Cowlishaw and Dunbar 2000), empirical studies are needed at larger spatial scales (Harcourt and Doherty 2005) than are often achievable with primate populations using traditional techniques. Instead of focusing on absolute abundance or density, we consider another fundamental component of animal ecology, the pattern of species distribution (Andrewartha and Birch 1954), as a primary state variable of interest.

In this paper, we discuss 1) the advantages of using cameras to sample primate species, 2) the use of occupancy models to make inference about species distribution patterns, 3) relevant study design principles, 4) the relationship between occupancy and abundance, 5) occupancy model extensions, and 6) modeling implementation. The approach outlined here allows researchers to noninvasively collect data on rare and difficult to detect primate species, explicitly deal with imperfect detection when sampling primates, incorporate spatial variability that is applicable for landscape scale inference, and enable researchers to ask a diversity of important ecological and conservation driven questions. Occupancy and photographic sampling have been widely adopted in ecological studies (MacKenzie *et al.* 2006; O'Connell *et al.* 2010), in combination or separately. Numerous primate studies have used photographic sampling (Bezerra et al., this issue, Boyer and Pruetz, this issue), but few have embraced the occupancy framework. We are aware of only three field-based (i.e., direct sampling) primate studies that use the occupancy framework (Baker *et al.* 2011; Guillera-Arroita *et al.* 2010; Keane *et al.* 2012) and one non-field-based (i.e., indirect sampling) study (Karanth *et al.* 2010).

Advantages of Cameras When Sampling Primates

Photographic sampling is expanding into many ecological studies and monitoring programs (Kucera *et al.* 2010; Martins *et al.* 2007). Camera traps are being used to investigate species abundance, density, distribution, richness, predator–prey interactions, habitat selection, and animal behavior (O'Connell *et al.* 2010; this special issue).

There are many advantages to using cameras for sampling wild animal populations (Kays and Slauson 2008; O'Connell *et al.* 2010). Perhaps their most appealing quality is that the animal and the observer do not have to be at the same place at the same time, which may be especially important for primate species that are rare or difficult to detect via human observers (e.g., cryptic, nocturnal). Reliable cameras can operate continuously for weeks to months in a variety of habitats, including subtropical rainforests, and can be simultaneously deployed across large spatial scales (Larney, E., *pers. comm.*). The analogous simultaneous sampling via observers may be costly and is often logistically unfeasible, especially in remote areas where travel is difficult. In addition, photographic sampling reduces, or eliminates, interactions between observers and primates, possibly reducing stress or behavioral changes, such as avoidance of people by hunted populations.

Cameras may be optimally placed within selected sample units to maximize detection probability in occupied areas. This flexibility allows investigators to target species that are terrestrial, arboreal, or both. Moreover, the photographs themselves provide physical evidence of the presence of a species at an exact spatial location, and are important in documenting and understanding long-term change in species distributions. These are many of the reasons why photographic sampling is so prevalent in carnivore studies (Gerber *et al.* 2010; Kays and Slauson 2008). Still, there are a number of considerations that investigators must address before employing cameras as the main sampling method in their study. For instance, the type of environment (e.g., tropical forest, high mountain top), quality of battery, and storage card are very important in a camera's ability to operate continuously at a variety of locations. Deciding where to place cameras, both in terms of defining and selecting sample units/sites and selecting camera locations within these units/sites, will influence the inference that is being sought (see Study Design Principles section).

We focus on the static or single-season occupancy model (*sensu* MacKenzie *et al.* 2002) to estimate distributional patterns and investigate species–habitat relationships. We briefly discuss two additional occupancy models that could be useful in studying primates. There are, however, many other applications of occupancy models, such as species interactions (Bailey *et al.* 2009), meta-population dynamics (MacKenzie *et al.* 2003), disease mapping (Abad-Franch *et al.* 2010), invasive species mapping (Gormley *et al.* 2011), and large-scale community monitoring (Ahumada *et al.* 2011).

A Static Occupancy Model to Make Inference About Species Distribution

In this section, we outline the essential elements of the static occupancy model when used with photographic sampling data. We 1) discuss the idea of occupancy and detection probability, 2) define the statistical model, 3) consider information-theoretics and multiple models, 4) demonstrate how camera malfunctions can be accommodated, 5) define occupancy model assumptions, and 6) address the issue of false positives.

Occupancy and the Need for Detection Probability

Occupancy is defined as the probability that a site or patch is occupied by a target species during a specified time period, often referred to as a season (MacKenzie *et al.* 2006). Generally, a site or sampling unit is defined as any patch of habitat that is either naturally occurring (e.g., ponds or habitat fragments) or arbitrarily defined by the investigator (e.g., sampling block or a camera site). Robust inference about species occurrence requires the consideration of two main sources of variation, spatial variation in occurrence and spatial-temporal variation in detection (Thompson 2004).

Historically, species occurrence was referred to as presence/absence, where spatial variation was often considered, but not detection variation (Vojta 2005). There was no formal way of separating a site that had no individuals of a species (true absence) and a site that contained individuals but the observer did not detect them (non-detection). When a site is truly occupied and there are no detections, this is referred to as a false-absence or false-zero. An occupancy analysis addresses the variation in detection by using a probability model. This model is used to estimate the probability that a site is occupied from a series of detections/non-detections, the characteristics of the site, and the detection histories of the species at all other sites. Without considering how detection varies, estimates of occurrence can be drastically biased (MacKenzie *et al.* 2006).

Consider the scenario in which species occurrence at (or use of) a camera site depends on some habitat feature, such as the size of the habitat fragment within which

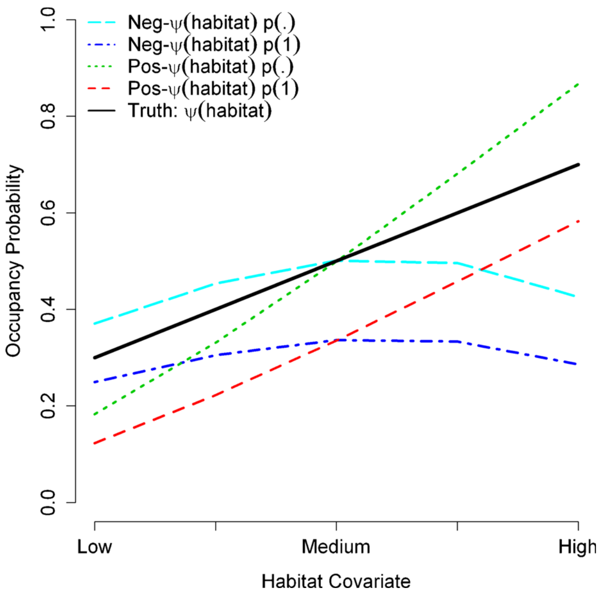

Fig. 1 Demonstration of the bias in the probability of species occupancy when detection is assumed to be perfect p(1) or constant p(.), when occupancy is positively related to a habitat feature: Neg = negative relationship between detection and habitat covariate. Pos = positive relationship between detection and habitat covariate.

the site is located. Suppose this same habitat feature also affects the probability of detecting the species at occupied sites, making it harder or easier to detect the species. If we falsely assume detection is perfect or constant across all sites and surveys, our inference about the relationship between species occupancy and the habitat feature can vary dramatically; all estimates will be biased (Fig. 1). For example, assume true occupancy is positively related to canopy cover, whereas detection is negatively related; more sites with high canopy cover will be occupied, but we will be less likely to detect the species at these sites compared to occupied sites with less canopy cover. Without considering this variability and assuming detection is perfect, occupancy will be negatively biased with the greatest bias at high canopy cover values. In other words, where we are most likely to not detect the species is where there is the greatest discrepancy between true and naïve occupancy. Alternatively, if detection is positively related to canopy cover (higher canopy cover → higher detection probability) and we assume detection is constant, naïve occupancy will be negatively biased at low canopy cover and positively biased at high canopy cover and the relationship between species occurrence and canopy cover is masked (Fig. 1).

Occupancy is naturally viewed hierarchically, where there is an observation process that may imperfectly represent the ecological reality (Fig. 2). To separate non-detections and true absences, the occupancy framework uses repeat evaluations of a species as either detected (1) or not detected (0) at each site. When using cameras, these evaluations are repeat surveys over time, defined by discrete sampling occasions. To illustrate, consider a population of 18 sites where 8 sites are

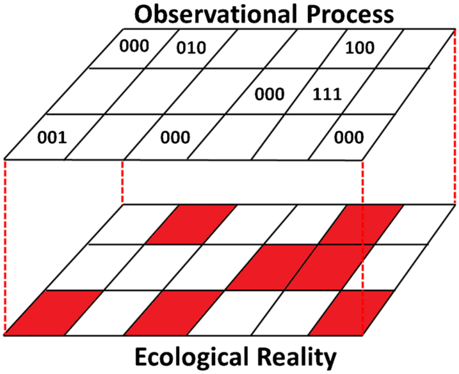

Fig. 2 Hierarchical observation and ecological process, where occupancy of a grid cell can be observed imperfectly. The population of interest is 18 sites, of which 8 were randomly selected and sampled on 3 occasions (cameras placed for 3 wk). On each occasion, the species was either detected or not detected (1 or 0, respectively). The bottom grid cells that are filled in indicate the site is occupied by the species.

randomly selected for sampling brown mouse lemur (*Microcebus rufus*; Fig. 2). We place cameras within each of these sites and sample for 3 wk. We then record detection/non-detection information, where one occasion equals 1 wk. As such, for each sampling occasion and camera site, we would mark a 1 if there was one or more mouse lemur photograph and a zero if there were no mouse lemur photographs. We end up with a matrix (referred to as a detection history) that is 8 (sites) × 3 (occasions). Because we are concerned only with species occurrence, the number of photographs above one within an occasion is irrelevant; there is no difference in one photograph of a single mouse lemur and many photographs of one or more mouse lemurs. Photographic sampling studies often define occasions as days or weeks; we discuss this further in the Study Design Principles section. In this example, we are using such few sites for simplicity only; the number of sites in a real camera-occupancy study should likely be many more (see Study Design Principles section).

Defining the Model

To make formal inference about our unknown quantities of interest, mouse lemur occupancy and detection probability, we use a probability model. Following conventional notation, we define ψ_i as the probability of mouse lemur occupancy at site i and p_{ik} as the probability of detection at site i on occasion k, given the site is occupied. Using a detection history (h_i) in which mouse lemurs were not photographed during the first and second occasions, but are successfully photographed during the third occasion (see the lower left site in Fig. 2), h_i = 001, and the probability statement is written as:

$$Pr(h_i = 001) = \psi_i(1 - p_{i1})(1 - p_{i2})(p_{i3}) \tag{1}$$

We know that this site is occupied by mouse lemurs based on the detection on occasion 3; the probability of this event is ψ_i. Notice that all possible events are included in this statement; we read the probability statement as a joint product of occupancy (ψ_i) AND no detection on occasion 1 ($1 - p_{i1}$) AND 2 ($1 - p_{i2}$), AND detection on occasion three (p_{i3}). Now, let's consider two sites where the mouse lemur was not detected, but truly occupied one of these sites (see lower right site and the upper left site of Fig. 2). The probability statement associated with the detection history observed at both sites is:

$$Pr(h_i = 000) = \psi_i(1 - p_{i1})(1 - p_{i2})(1 - p_{i3}) + (1 - \psi_i) \tag{2}$$

At these camera sites, our data cannot directly indicate whether both, one, or neither of these sites is truly occupied. The first possibility is stated by the terms left of the addition sign, where mouse lemurs may have been present (ψ_i), but not detected over the three sampling occasions $[(1 - p_{i1})(1 - p_{i2})(1 - p_{i3})]$. The right of the addition sign is the other possibility that mouse lemurs simply did not occupy the site $(1 - \psi_i)$; there is no need for detection parameters, as the species was not present and available to be detected. The probability statement is inclusive of all possible outcomes, where the uncertainty in occupancy is mathematically represented (i.e., OR) as a sum of two possibilities.

Assigning probability statements to each site's detection history permits the estimation of our unknown parameters of interest (ψ_i and p_{ik}) through maximum likelihood methods. Maximum likelihood is often done by numerical optimization routines that maximize the joint density (i.e., likelihood) of all our unknown parameters, given our data. Maximum likelihood estimation is a common method for statistical inference owing to desirable statistical properties under asymptotic conditions; we encourage interested readers to read more on likelihood theory and its applications (Hilborn and Mangel 1997). We can define the likelihood of our unknown parameters, given our data (detection histories), as

$$Likelihood(\psi_i, p_{ik}|h_1, h_2, h_3, \ldots, h_N) = \prod_{i=1}^{N} Pr(h_i) \tag{3}$$

However, while defined generally, this model is over-parameterized with as many unknown parameters as there is sample size; the model has zero degrees of freedom and cannot estimate all parameters separately. To estimate the parameters, it is necessary to reduce the dimension of this model. To do this, an investigator considers the response variables to have a functional relationship with a limited number of characteristics at each site or sampling occasion. The characteristics of the site or sampling occasion are used as covariate information. This covariate information (x_{ij} or x_{ijk} for $i = 1, \ldots, N$ sites, and $j = 1, \ldots, M$ covariates in the model, and $k = 1, \ldots, T$ surveys at each site) can be introduced using a logistic regression model, where each α_j and β_{ij} are the $j = 0, \ldots, M + 1$ unknown "effects" on occupancy or detection probability, respectively, for each independent covariate. The model can be expressed as:

$$\begin{aligned} \text{logit}(\psi_i) &= \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots \\ \text{logit}(p_{ik}) &= \beta_0 + \beta_1 x_{ik1} + \beta_2 x_{ik2} + \ldots \end{aligned} \tag{4}$$

The logit ($\log_e(\theta/1 - \theta)$) is used as a link function that takes a probability from a scale of [0,1] and projects it on a line taking values from $-\infty$ to ∞. Final estimates are back-transformed using the inverse-logit ($exp(\theta)/1 + exp(\theta)$), so that occupancy and detection probabilities are properly supported between 0 and 1. By incorporating auxiliary information into the modeling process, we can rigorously formulate biologically driven hypotheses in a statistical framework. Consider the hypothesis that brown mouse lemur occupancy varies owing to differences among sites in unlogged and logged forest, as well as to variation in fruit abundance at each site (ESM1). As such, x_{i1} is a binary covariate to indicate site i is either an unlogged or logged forest site and x_{i2} is a continuous covariate, representing the abundance of fruit for site i; α_0 is the intercept, α_1 is the effect difference between unlogged and logged forest, and α_2 is the effect of variability in fruit abundance. This model considers fruit abundance to vary additively (on the logit scale) between unlogged and logged sites. To consider fruit abundance to vary completely differently between our forest sites, we need to include an interaction covariate, x_3, that is the product of the two other covariates ($x_{i3} = x_{i1} * x_{i2}$).

We can also define the occupancy model with a hierarchical model formulation, where an observational model is conditional on a true spatial process model (ESM2). Such a model can be naturally fit using a Bayesian framework, where inference is achieved by sampling full conditional posterior distributions of unknown parameters using Markov chain Monte Carlo methods (Royle and Dorazio 2008). This approach has advantages when there is prior information to incorporate, it is preferred to treat parameters as random (i.e., random effect), and there are multiple hierarchal structures considered, such as multispecies modeling (Burton *et al.* 2012).

Information Theoretics and Multiple Models

Within a likelihood framework, multiple models are usually fit to the data to evaluate competing hypotheses involving covariates (independent explanatory variables) thought to influence species occurrence or detection probability. Evaluating the relative influence of these predetermined variables on occupancy probabilities is often the main objective of a given study. Following from the mouse lemur example above, it could be important to differentiate the relative influence forest type (logged vs. unlogged) and fruit abundance have on their occurrence. This could aid in forest logging practices; if fruit abundance has more influence on mouse lemur occupancy than general forest type, it may be possible to achieve high mouse lemur occupancy in a logged forest by maintaining a certain fruit abundance, possibly through a shelterwood cutting plan.

Information-theoretic approaches are used to determine the relative weight of evidence for each plausible model, or hypothesis (Burnham and Anderson 2002). From these weights, the relative importance of each independent variable can be calculated by summing the weights of models that contain each variable (cumulative model weights); however, care should be taken to ensure that each variable being compared occurs in the same number of models (i.e., a balanced model set; Burnham and Anderson 2002; Doherty *et al.* 2012; Lukacs *et al.* 2010). Importantly, model selection uncertainty can be incorporated and accommodated via model averaging, where parameter estimates from each model are weighted based on the model's relative support to yield estimates that are less biased compared to any single model's estimates (Burnham and Anderson 2002; Doherty *et al.* 2012).

Camera Malfunctions and Missed Survey Occasions

Despite best efforts, it is common for sites to not be surveyed on some occasions. A camera may malfunction, deplete its battery power, or fill its memory card or film cartridge. Determining whether an occasion has not been surveyed will depend on the ability of the researcher to determine when a camera malfunctioned. This could be done by routine checking of the camera, or by programming the camera to actively take one or more photographs during each occasion. Cameras that lacked photographs at these preset times could be considered inoperable during the entire occasion.

Ignoring these missing occasions in the probability model (i.e., treating them as non-detections) will lower detection probability and induce heterogeneity among sites and sampling occasions. Because each sampling occasion is treated as independent, we can simply drop the missed occasion from a detection history's probability statement and thus no information enters into the model likelihood for that sampling occasion. For example, if we set up a camera site 1 wk (one occasion) later than all others, we would simply rewrite the probability statement of Eq. 1 as

$$Pr(h_i = -01) = \psi_i(1 - p_{i2})(p_{i3}) \tag{5}$$

Though still statistically rigorous, this allows unequal sampling across sites, something that may be planned or accidental. If cameras are partially operational during a long occasion, such as a week, it may be a poor choice to ignore any detection data. Instead, we can use the detection/non-detection information for these occasions, but also include a sampling-level covariate that corresponds to the amount of time (e.g., days, hours) that the camera site was actually operational.

Model Assumptions

The static or single-season occupancy model is based on four major assumptions: 1) sites are closed to changes to occupancy status over the designated sampling period or season (closure assumption), 2) the probability of occupancy is constant across sites or otherwise appropriately modeled with covariates such that there is no unaccounted for heterogeneity, 3) the probability of detection is constant across sites and sampling occasions or otherwise modeled, and 4) detections and thus detection histories of sites are independent.

When sampling primates with cameras, it is especially important to consider the closure assumption because individuals are often highly mobile and may violate this assumption. Simply, no primate will either truly occupy or not occupy a single camera site for the duration of the study because the size of the site is much smaller than most primates' home range. Instead, a species will likely be near the camera some of the time and not at other times. Obviously, when members of the species are not near the camera, the species is not available for detection. This process does not necessarily violate the closure assumption; if the species availability at a camera location is a random process, such that during each occasion the species has a nonzero probability of being near the camera, the closure assumption is not violated. However, occupancy will reflect the probability a site is used during the study period, rather than physical presence the entire time (MacKenzie and Royle 2005; MacKenzie *et al.* 2006). Note that if the site size is poorly defined, such as when an individual is attracted to a site that it would not normally use owing to placement of bait or lure, the ability to make inference to the species' natural use of sites is jeopardized (Efford and Dawson 2012). For further discussion on the closure assumption, see "Defining the Season" in the Study Design Principles section below.

To meet assumptions 2 and 3 above, one must thoroughly consider the ecology of the primate species being studied. Expert knowledge and published literature is helpful in deciding on relevant covariates to include in models. The best way to meet assumption 4 is to follow study the deign principles outlined below and not purposively place camera sites. For a more detailed account of the assumptions, see Mackenzie *et al.* (2006).

False Positives

We have so far considered only when a primate species goes undetected at an occupied site. Now we consider when a species is erroneously detected at a site when it is truly absent (false-positive). This can occur when species are misidentified as a result of blurry or poor quality photographs. If false-positives occur and cannot be resolved following strict design protocols, modeling approaches are available that combine data sources without false-positive errors to correct a more extensive dataset where false-positives are likely (Miller *et al.* 2011; Royle and Link 2006). However, if there are relatively few poor quality photographs in which the species is not known exactly, it may be more appropriate to ignore these photographs and consider them as non-detections. If these were true detections, ignoring photographs will lower the detection probability and likely reduce the precision of the occupancy estimates, while including them when the species was incorrectly identified will often introduce bias (Miller *et al.* 2011; Royle and Link 2006). Ideally, all photographs with primate detections will be identified to the species without error.

Study Design Principles

Study design is a critical aspect of most studies, including occupancy studies, but is often overlooked. In this section, we 1) outline the general ideas of using a sample to make inferences about a larger population, 2) address how to randomize site selection, and 3) discuss the importance of defining the season of an occupancy study.

Using a Sample to Make Inference to a Population

A statistical population is a set of elements that we want to make statements about. Generally, the scope of an occupancy study involves a large spatial scale that spans multiple home ranges of the primate of interest. At this spatial scale, to study all of the elements in a population (i.e., a census of sites of interest) is physically and fiscally prohibitive. For example, if there is interest in the use of naturally occurring salt/mineral deposits by a primate species, it is unlikely that all deposits can be sampled. As an alternative we select a sample from the population, then use the sample to make statements about the population (i.e., statistical inference). Using a sample has several appealing characteristics. Mainly, it is faster, cheaper, easier, and, when implemented correctly, is mathematically justifiable. Identifying what the population of interest is depends on your ecological question, while selecting the sample will ultimately dictate where cameras are placed. Once a sample of sites is identified, it can be a challenge to secure a camera to achieve an ideal angle to sample the species' use of a site (e.g., in a canopy pointing down, on a tree bole pointing parallel to the ground); this will largely depend on the mode of locomotion of the primate.

Making inference from a sample to a population requires that the sample represents the population of interest. If the sample is not representative of the population, the conclusions about the sample are still valid, but when making inference to the population the conclusions will be wrong. For example, if camera sites are selected based on observed primate use of salt/mineral deposits, the sample may be informative regarding those sites, but may be unrepresentative of the population of salt/mineral deposits. The paradox in applied statistical inference is that because we do not know the characteristics of the population, it is impossible to say whether our sample is representative of it. The best we can do is take actions that maximize the probability that our sample will be representative, which we can do by selecting our sample randomly. A random or probabilistic sample is a fundamental concept in design-based inference (Cochran 1977; Thompson 2012). Occupancy estimation, in contrast to design-based inference, is inherently model based (see Gregoire 1998 for a comparison of design- and model-based inference). However, most of the concepts that rely on randomization in design-based inference also apply to model-based inference.

First, randomization reduces the chance that variables that are not explicitly modeled or accounted for will have an effect on the results of the study. That is, it reduces the chance that confounding variables will affect the results when compared to a nonrandom study. Second, randomization increases the probability that the range of values in the population will be represented in the sample; following from the mineral deposits example, the selection of camera sites based on past use is unlikely to capture the full range of use, thus possibly missing the explanatory power of why that may be. If the functional relationship between the response variables (occupancy and detection probability) and predictor variables is unknown, then observing how the response changes as a function of a range of predictor variables helps understand that relationship. Third, a random selection helps prevent against intentional or unintentional bias imposed by the investigator that might occur when selecting a nonrandom sample. The investigator-imposed bias increases the probability that the sample is not representative of the population.

Randomizing Site Selection for Occupancy Studies

Recall that in occupancy analysis, the spatial distribution of the species is the state variable of interest. One method for random sampling for an occupancy study is to first define the population as some spatial boundary. Then divide the boundary into smaller subunits (called sites) that are capable of being sampled with camera traps. Finally, randomly select the sites to be sampled. The number of sites included affects the range of values in the population that are represented in the sample, the ability to identify the functional relationship between the response variable and the predictor variables, and the ability to detect changes in species distributions if those changes occur. Generally, more sites are better. We recommend researchers explore study design scenarios using different numbers of sampling sites when considering statistical power and to discriminate among alternative hypotheses (i.e., models). Briefly, to explore study design options, one would first consider realistic variation in occupancy and detection for the study of interest (via expert opinion or published literature), then simulate data sets with varying number of sites and occasions, and lastly fit models to the data to observe characteristics that are important, such as bias, precision, and power to discriminate among alternative models. Several articles have addressed power calculations and optimal sample size selection for occupancy studies and we direct readers to those articles (Bailey *et al.* 2007; MacKenzie and Royle 2005). In addition, if a study objective is to contrast the occupancy in two or more different types of habitat, then it is essential that the sample contain a suitable number of samples from each type (this is a model-based analogue to the stratified random sample in design-based inference).

Defining the Season

The definition and duration of the sampling period will depend on a study's objectives. If interest in primate occurrence is during a species' mating season, then this is the period during which sampling should occur. The length of sampling should also be considered with respect to the closure assumption (site occupancy status does not change over the sampling period). Maintaining a relatively short season will help ensure this assumption is not violated. It is the definition of the season that determines the validity of the closure assumption.

Typically, it is best to choose a period when the species spatial distribution is relatively stable (i.e., avoid migratory or dispersal periods). Within that study period, you can use what is known about the species ecology to help define an appropriate survey occasion. For example, if it is known that the primate typically rotates among 8–10 feeding locations within a week, and you have placed a camera at one randomly chosen tree within a grid cell size that corresponds to the species spatial range, then it makes sense to define your occasion as a week, because if the species is using the grid cell, it is likely to visit the tree at least once within a given occasion. However, if the species moves out of the area during the middle of the study period and is no longer available for detection (i.e., the probability of detection is zero) than the closure assumption is violated.

A combination of the study period and the species' ecology helps define an appropriate sampling occasion. It is also important to define occasions such that detection probabilities are not so low they cause estimation problems (Mackenzie *et al.* 2002). Generally, detection probabilities per occasions lower than 0.10 are prone to cause difficulties in estimating parameters. Estimating lower per occasion detection probabilities is possible, if there are many occasions and little variation in detection probability among occasions. When using cameras to sample highly mobile species, it is often convenient to define an occasion as a day or week, so as to achieve a moderate detection probability. Using a day as an occasion is often appropriate because it matches with the circadian rhythm of many species' activity. However, if daily detection probability is very low, numerical convergence of the maximum likelihood could fail, and thus necessitate redefining an occasion to a longer period of time.

Let us consider three different studies with very different definitions for a site and season. In the first study, the objective is to understand whether primate occupancy of certain habitat patches is driven by highly clumped food during the dry season. We define the geographic area in which we would like to make inference, identify all potential sites, and then deploy cameras randomly within patches that have and do not have this food resource. We define the season as a 2-wk period during the middle of the dry season where the effect of the food is potentially strongest and where it is reasonable to assume that species ranging is stable, thus meeting the closure assumption. If the closure assumption is unlikely to hold, but the species availability can be considered a random process, we may want to consider occupancy as the "probability of use."

In the second study, the objective is to determine habitat fragment use of a primate species during cyclones. The researcher identifies a population of fragments of interest, randomly selects a subset of these fragments to sample, and deploys one or more cameras in each fragment. Before a cyclone reaches the fragments, the researcher activates the cameras. To estimate occupancy at the fragment level, we could combine the detection histories of all cameras within each fragment, thus appropriately treating the fragment as the site.

Lastly, in the third study, the objective is to determine occupancy of a terrestrial primate species throughout a national park that is divided into 10 management zones based on ecoregions. Each ecoregion is managed differently; thus the occurrence of the primate will affect future management decisions differently throughout the park. We do not believe the occupancy of these ecoregions to vary within the wet season, which is 5 mo, as the primate maintains a stable home range. We have enough cameras to sample two ecoregions simultaneously. Thus, in each month of the wet season, we select two ecoregions, randomly select sites, and deploy cameras.

Note that the scale and season of occupancy for each study is different and not comparable, and determined by the objective. In the first study, the season is relatively short, to determine a possible driver of distributional patterns by a resource at a local level. In the second study, the season is defined by a cyclone event, and in the third study, the season is long to consider relatively stable landscape-scale distributional patterns. Also, in the first and third studies, we are able to define how many sites to sample by using as many cameras as available, while in the second study, sites are defined by the availability of fragments in the study area. In the first two studies, our objectives were driven by ecological relationships, while the third study was motivated by future decisions that may affect occurrence at a landscape scale throughout a national park.

The Relationship Between Occupancy and Abundance

There is a natural tendency to relate occupancy and abundance, however, this needs to be done with some caution. Occupancy, as described here, is simply a discretization of a species abundance distribution across the landscape, $\psi_i = Pr\,(N_i > 0)$, where N_i is the local abundance at site i (Royle *et al.* 2005). Occupied sites may have a single individual or many individuals, but the occupancy state is the same for all these sites. Unique cases may exist for low-density, highly territorial species, where occupied sites would be expected to contain only a single or pair of individuals. In these cases the relationship between occupancy and abundance may approach equity (MacKenzie and Nichols 2004), but these cases are likely rare in primate studies where groups of varying size travel together.

It is often expected that local abundance has a strong effect on detection probability at occupied sites (Royle and Nichols 2003). Simply, it is likely to be easier to detect the species at a site with high density than a low density. If we consider a site as a habitat fragment, we are more likely to detect a primate species living in a group of 30 than a group of 2. In these instances, an explicit relationship between the two quantities can be expressed as $p_i = 1 - (1 - r)^{N_i}$, where p_i is the probability of detecting one or more individual of the target species at an occupied site i (i.e., the usual species detection probability), r is the probability that a given individual is detected, and N_i is the number of individuals present at site i. If investigators believe that variation in local abundance is a primary source of heterogeneity in detection probability that cannot be accounted for with habitat variables, then models that explicitly account for the relationship between detection probability and local abundance should be employed to reduce bias in occupancy estimates (Royle and Nichols 2003). However, it is always the case that there will be covariates the investigator will not be able to model that affect detection. For example, the investigator may not have the data, or is unaware of the importance of certain covariates. The solution for this problem goes back to random sampling, which minimizes the probability that confounding variables affect the estimates of detection.

Occupancy Model Extensions

For some studies, static occupancy of a site may not be a primary interest. There may be a greater interest in the factors that affect occupancy dynamics and thus changes in species distributions. MacKenzie *et al.* (2003) present a multiseason model that focuses on the dynamic processes of extinction and colonization at sites, which are the primary determinants of future occupancy (Fig. 3). Extinction and colonization may be influenced and thus modeled by habitat features or changes over time (e.g., seasonal or annual changes). For example, instead of modeling primate occupancy as a function of some previous disturbance, we can draw direct inference about how this disturbance influences changes in occupancy across seasons. The multiseason model links single seasons together, where sites are still surveyed multiple times within a season, when occupancy is considered static, to correct for imperfect detection.

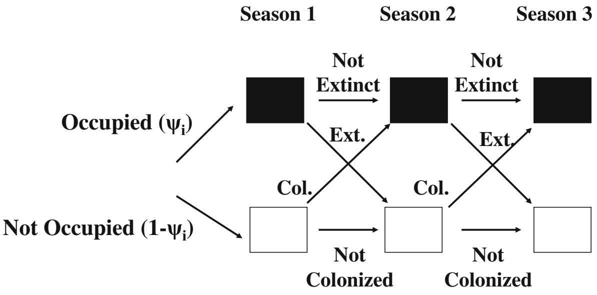

Fig. 3 Diagram of changes in a sites species occupancy caused by extinction and colonization across multiple seasons. Col = colonized. Ext = extinct.

Another occupancy model of particular use to primate studies using cameras is a multiscale model. Following from the cyclone example above, when we collapse detection histories from multiple cameras and treat the fragment as the site, we lose any inference of "use" at the camera site. There may be interest in how a species is using specific habitats within a fragment. Instead of collapsing detection histories, we can explicitly estimate occupancy at the fragment and "use" or availability at the camera site scale using a multiscale occupancy model (Nichols *et al.* 2008). An additional advantage to this model is that multiple methods can easily be combined if they are independent of each other, for example, using observer area searches and cameras. This allows a direct comparison in the detection efficiency of different methods, as well as a way to reduce potential heterogeneity in detection; some individuals of a species may be more difficult to detect with certain methods.

Modeling Implementation and Conclusion

Both camera traps and occupancy could be of great use in the study and conservation of wild primates. One recent study used cameras to validate previously unconfirmed sightings of the critically endangered greater bamboo lemur (*Prolemur simus*) in Madagascar (Olson *et al.* 2012). Such work helps to increase the known distribution of this rare species, although species presence is less helpful in understanding distributional patterns of why a species occurs where it does. By following study design and occupancy modeling procedures outlined here, cameras can be used to understand the many factors that likely determine greater bamboo lemur occurrence on the landscape. This knowledge is critical for successful long-term conservation actions to restore this unique species to its former population and range.

A number of free software packages are available to implement occupancy models discussed here, as well as many others. Some packages are used in annual training workshops (http://www.phidot.org/forum/). Two packages with a graphical user interface and considerable documentation are programs MARK (White and Burnham 1999) and PRESENCE (Hines 2013). For users of the R programming language (R Core Team 2013), the package "unmarked" includes occupancy models (Fiske and Chandler 2011). Bayesian inference can be done directly through OpenBugs (Lunn *et al.* 2009), where pseudo-code is used directly to define the model (Royle and Dorazio 2008) or alternatively indirectly through PRESENCE. See ESM3 for links to documentation and software.

Camera traps and occupancy are being used throughout the world to better understand rare and elusive species in the hopes of implementing more effective conservation strategies. Primate conservation could benefit equally from the many important insights that could be gained when using camera traps in an occupancy framework. We hope this paper will encourage primatologists to add these methods to their tool kits.

Acknowledgments We thank Marni LaFleur and Chia Tan for inviting us to contribute to this special issue. We are also grateful to Marni LaFleur, Joanna Setchell, and two anonymous reviewers for useful comments that helped improve this manuscript.

References

Abad-Franch, F., Ferraz, G., Campos, C., Palomeque, F., Grijalva, M., Aguilar, H., Miles, M. (2010). Modeling disease vector occurrence when detection is imperfect: Infestation of Amazonian palm trees by triatomine bugs at three spatial scales. *PLoS Neglected Tropical Diseases*, *4*, e620.

Ahumada, J.A., Silva, C.E.F., Gajapersad, K., Hallam, C., Hurtado, J., Martin, E. (2011). Community structure and diversity of tropical forest mammals: Data from a global camera trap network. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *4*, 2703–2711.

Anderson, D.R. (2001). The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin*, *29*, 1294–1297.

Andrewartha, H., & Birch, L. (1954). In *The distribution and abundance of animals*. Chicago: University of Chicago Press.

Bailey, L., Hines, J., Nichols, J., MacKenzie, D. (2007). Sampling design trade-offs in occupancy studies with imperfect detection: Examples and software. *Ecological Applications*, *17*, 281–290.

Bailey, L., Reid, J., Forsman, E., Nichols, J. (2009). Modeling co-occurrence of northern spotted and barred owls: Accounting for detection probability differences. *Biological Conservation*, *142*, 2983–2989.

Baker, L., Arnold, T., Olubode, O., Garshelis, D. (2011). Considerations for using occupancy surveys to monitor forest primates: A case study with sclaters monkey (*Cercopithecus sclateri*). *Population Ecology*, *53*, 549–561.

Brus, D., & De Gruijter, J. (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*, *80*, 1–44.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L. (2001). In *Introduction to distance sampling*. Oxford: Oxford University Press.

Buckland, S.T., Plumptre, A.J., Thomas, L., Rexstad, E.A. (2010a). Line transect sampling of primates: Can animal-to-observer distance methods work? *International Journal of Primatology*, *31*, 485–499.

Buckland, S.T., Plumptre, A.J., Thomas, L., Rexstad, E.A. (2010b). Design and analysis of line transect surveys for primates. *International Journal of Primatology*, *31*, 833–847.

Burnham, K.P., & Anderson, D.R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach*. New York: Springer.

Burton, A.C., Sam, M.K., Balangtaa, C., Brashares, J.S. (2012). Hierarchical multi-species modeling of carnivore responses to hunting, habitat and prey in a West African protected area. *PloS one*, *7*, e38007.

Chapman, C.A., & Peres, C.A. (2001). Primate conservation in the new millennium: The role of scientists. *Evolutionary Anthropology*, *10*, 16–33.

Cochran, W.G. (1977). In *Sampling techniques*. New York: Wiley.

Cowlishaw, G., & Dunbar, R. (2000). *Primate conservation biology*. Chicago: University of Chicago Press.

Doherty, P.F., White, G.C., Burnham, K.P. (2012). Comparison of model building and selection strategies. *Journal of Ornithology*, *152*, 1–7.

Efford, M., & Dawson, D. (2012). Occupancy in continuous habitat. *Ecosphere*, *3*, art32.

Fashing, P.J., & Cords, M. (2000). Diurnal primate densities and biomass in the Kakamega Forest: An evaluation of census methods and a comparison with other forests. *American Journal of Primatology*, *50*, 139–152.

Fiske, I., & Chandler, R. (2011). Unmarked: An r package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, *43*, 123.

Gerber, B., Karpanty, S., Crawford, C., Kotschwar, M., Randrianantenaina, J. (2010). An assessment of carnivore relative abundance and density in the eastern rainforests of Madagascar using remotely-triggered camera traps. *Oryx*, *44*, 219–222.

Gerber, B., Karpanty, S., Randrianantenaina, J. (2012). The impact of forest logging and fragmentation on carnivore species composition, density and occupancy in madagascars rainforests. *Oryx*, *46*, 414–422.

Gormley, A., Forsyth, D., Griffioen, P., Lindeman, M., Ramsey, D., Scroggie, M., Woodford, L. (2011). Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology*, *48*, 2534.

Gregoire, T. (1998). Design-based and model-based inference in survey sampling: Appreciating the difference. *Canadian Journal of Forest Research*, *28*, 1429–1447.

Guillera-Arroita, G., Lahoz-Monfort, J., Milner-Gulland, E., Young, R., Nicholson, E. (2010). Using occupancy as a state variable for monitoring the critically endangered Alaotran gentle lemur *Hapalemur alaotrensis*. *Endangered Species Research*, *11*, 157–166.

Guschanski, K., Vigilant, L., McNeilage, A., Gray, M., Kagoda, E., Robbins, M.M. (2009). Counting elusive animals: Comparing eld and genetic census of the entire mountain gorilla population of Bwindi Impenetrable National Park, Uganda. *Biological Conservation*, *142*, 290–300.

Harcourt, A., & Doherty, D. (2005). Species-area relationships of primates in tropical forest fragments: A global analysis. *Journal of Applied Ecology*, *42*, 630–637.

Hilborn, R., & Mangel, M. (1997). *The ecological detective: Confronting models with data*. Princeton, NJ: Princeton University Press.

Hines, J.E. (2013). Presence 5.5. Software to estimate patch occupancy and related parameters. USGS-PWRC. http://www.mbr-pwrc.usgs.gov/software/presence.shtml.

IUCN (2012). *The iucn red list of threatened species*. Version 20122. http://www.iucnredlist.org. Downloaded on 17 October 2012.

Karanth, K., Nichols, J., Hines, J. (2010). Occurrence and distribution of indian primates. *Biological Conservation*, *143*, 2891–2899.

Kays, R., & Slauson, K. (2008). Remote cameras. In R.A. Long, P. MacKay, W.J. Zielinski, J. Ray (Eds.), *Noninvasive survey methods for carnivores* (pp. 110–140). Washington, DC: Island Press.

Keane, A., Hobinjatovo, T., Razafimanahaka, H., Jenkins, R., Jones, J. (2012). The potential of occupancy modelling as a tool for monitoring wild primate populations. *Animal Conservation*, *15*, 457–465.

Kendall, W., & White, G. (2009). A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. *Journal of Applied Ecology*, *46*, 1182–1188.

Kucera, T., & Barrett, R. (2010). A history of camera trapping. In: A.F.O'Connell, J. D. Nichols, K. U. Karanth (Eds), *Camera traps in animal ecology: Methods and analyses* (pp. 9–26). New York: Springer.

Little, R.J. (2004). To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association*, *99*, 546–556.

Lukacs, P.M., Burnham, K.P., Anderson, D.R. (2010). Model selection bias and freedmans paradox. *Annals of the Institute of Statistical Mathematics*, *62*, 117–125.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N. (2009). The bugs project: Evolution, critique and future directions. *Statistics in Medicine*, *28*, 3049–3067.

MacKenzie, D., & Nichols, J. (2004). Occupancy as a surrogate for abundance estimation. *Animal Biodiversity and Conservation*, *27*, 461–467.

MacKenzie, D., Nichols, J., Hines, J., Knutson, M., Franklin, A. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, *84*, 2200–2207.

MacKenzie, D., Nichols, J., Lachman, G., Droege, S., Royle, J.A., Langtimm, C. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*, 2248–2255.

MacKenzie, D., & Royle, J. (2005). Designing occupancy studies: General advice and allocating survey effort. *Journal of Applied Ecology*, *42*, 1105–1114.

MacKenzie, D., Nichols, J., Royle, J., Pollock, K., Bailey, L., Hines, J. (2006). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. New York: Academic Press.

Martins, S.S., Sanderson, J.G., e Silva-Júnior, J.S. (2007). Monitoring mammals in the Caxiuanã National Forest, Brazil-first results from the tropical ecology, assessment and monitoring (TEAM) program. *Vertebrate Conservation and Biodiversity, 16*, 31–44.

Miller, D., Nichols, J., McClintock, B., Grant, E., Bailey, L., Weir, L. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology*, *92*, 1422–1428.

Nichols, J. (1992). Capture-recapture models. *BioScience*, *42*, 94–102.

Nichols, J., Bailey, L., Talancy, N., Campbell, G., Gilbert, A., Annand, E. *et al.* (2008). Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, *45*, 1321–1329.

O'Connell, A., Nichols, J., Karanth, K. (2010). *Camera traps in animal ecology: Methods and analyses*. In: New York: Springer.

O'Connell, A., Talancy, N., Bailey, L., Sauer, J., Cook, R., Gilbert, A. (2006). Estimating site occupancy and detection probability parameters for meso-and large mammals in a coastal ecosystem. *Journal of Wildlife Management*, *70*, 1625–1633.

Olson, E., Marsh, R., Bovard, B., Randrianarimanana, H., Ravaloharimanitra, M., Ratsimbazafy, J., King, T. (2012). Arboreal camera trapping for the critically endangered greater bamboo lemur *Prolemur simus*. *Oryx*, *46*, 593–597.

Plumptre, A.J. (2000). Monitoring mammal populations with line transect techniques in african forests. *Journal of Applied Ecology*, *37*, 356–368.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ross, C., & Reeve, N. (2003). Survey and census methods: Population distribution and density. Field and laboratory methods in primatology. In J.M. Setchell & D.J. Curtis (Eds.), *Field and laboratory methods in primatology: A practical guide*, (pp. 90–109). Cambridge, U.K.: Cambridge University Press.

Royle, J., & Dorazio, R. (2008). *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities*. New York: Academic Press.

Royle, J., & Link, W. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, *87*, 835–841.

Royle, J.A., & Nichols, J.D. (2003). Estimating abundance from repeated presence-absence data or point counts. *Ecology*, *84*, 777–790.

Royle, J.A., Nichols, J.D., Kèry, M. (2005). Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, *110*, 353–359.

Šarndal, C.E., Thomsen, I., Hoem, J.M., Lindley, D.V., Barndorff-Nielsen, O., Dalenius, T. (1978). Design-based and model-based inference in survey sampling [with discussion and reply]. *Scandinavian Journal of Statistics*, *5*, 27–52.

Thompson, S. (2012). *Sampling*. Hoboken, NJ: Wiley.

Thompson, W. (2004). *Sampling rare or elusive species: Concepts, designs, and techniques for estimating population parameters*. Washington, D.C.: Island Press.

Vojta, C.D. (2005). Old dog, new tricks: Innovations with presence-absence information. *Journal of Wildlife Management*, *69*, 845–848.

White, G., & Burnham, K. (1999). Program mark: Survival estimation from populations of marked animals. *Bird Study*, *46*, 120–139.

Wiewel, A.S., Adams, A.A.Y., Rodda, G.H. (2009). Evaluating abundance estimate precision and the assumptions of a count-based index for small mammals. *The Journal of Wildlife Management*, *73*, 761–771.

Zhan, X., Li, M., Zhang, Z., Goossens, B., Chen, Y., Wang, H., Bruford, M., Wei, F. (2006). Molecular censusing doubles giant panda population estimate in a key nature reserve. *Current Biology*, *16*, R451–R452.