COMMUNITY ECOLOGY 9(2): 207-216, 2008
1585-8553/$20.00 © Akadémiai Kiadó, Budapest
DOI: 10.1556/ComEc.9.2008.2.10

Imperfect detection and its consequences for monitoring for conservation

M. Kéry[1] and B. R. Schmidt[2,3]

[1] *Corresponding author. Swiss Ornithological Institute, 6204 Sempach, Switzerland. Tel.: +41 41 462 97 93, Fax: +41 41 462 97 10, E-mail: marc.kery@vogelwarte.ch*
[2] *Zoologisches Institut, Universität Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland*
[3] *KARCH, Passage Maximilien-de-Meuron 6, 2000 Neuchâtel, Switzerland*

Keywords: Abundance, Availability, Biodiversity, Detectability, Distribution, Inventory, Monitoring, Occupancy, Sampling, Survey.

Abstract: Biodiversity monitoring is important to identify conservation needs and test the efficacy of management actions. Variants of "abundance" (*N*) are among the most widely monitored quantities, e.g., (true) abundance, number of occupied sites (distribution, occupancy) or species richness. We propose a sampling-based view of monitoring that clearly acknowledges two sampling processes involved when monitoring *N*. First, measurements from the surveyed sample area are generalized to a larger area, hence the importance of a probability sample. Second, even within sampled areas only a sample of units (individuals, occupied sites, species) is counted owing to imperfect detectability *p*. If $p < 1$, counts are random variables and their expectation $E(n)$ is related to *N* via the relationship $E(n) = N*p$. Whenever $p < 1$, counts vary even under identical conditions and underestimate *N*, and patterns in counts confound patterns in *N* with those in *p*. In addition, part of the population *N* may be unavailable for detection, e.g., temporarily outside the sampled quadrat, underground or for another reason not exposed to sampling; hence a more general way of describing a count is $E(n) = N*a*p$, where *a* is availability probability and *p* detection, given availability. We give two examples of monitoring schemes that highlight the importance of explicitly accounting for availability and detectability. In the Swiss reptile Red List update, the widespread and abundant slow worm (*Anguis fragilis*) was recorded in only 22.1% of all sampled quadrats. Only an analysis that accounted for both availability and detectability gave realistic estimates of the species' distribution. Among 128 bird species monitored in the Swiss breeding bird survey, detection in occupied 1 km^2 quadrats averaged only 64% and varied tremendously by species (3–99 %); hence observed distributions greatly underestimated range sizes and should not be compared among species. We believe that monitoring design and analyses should properly account for these two sampling processes to enable valid inferences about biodiversity. We argue for a more rigorous approach to both monitoring design and analysis to obtain the best possible information about the state of nature. An explicit recognition of, and proper accounting for, the two sampling processes involved in most monitoring programs will go a long way towards this goal.

Nomenclature: AERC TAC's Taxonomic Recommendations (2003).

Abbreviations: AIC – Akaike's Information Criterion, MHB – Monitoring Häufige Brutvögel.

Introduction

Monitoring of biodiversity for conservation is important to identify conservation needs and to test the effectiveness of management actions. Some of the most frequently used measures of biodiversity are three variants of "abundance" ***N*****:** the number of individuals (abundance), the number or proportion of occupied spatial units (sites; distribution), and the number of species (species richness). An important reason for why the three variants of abundance *N* are so popular in monitoring and ecology alike is their conceptual simplicity; it is the number of reasonably well-defined units such as individuals, occupied sites, and species.

However, despite its conceptual simplicity, the quantification of *N* in practice can be challenging for two reasons (Nichols et al. 2008): first, the area about which inference is desired is typically much larger than the area that can actually be examined. One main task of a well-designed monitoring program is to ensure that the characteristics of the larger area can be estimated from the smaller sample. Second, even at sampled places not all units of *N* will usually be observed. That is, only a fraction of all individuals or species present at each site will be detected. Similarly, the occupancy status of a site will be imperfectly assessed, i.e., part of a species' distribution, may go undiscovered.

Here, we describe an explicit sampling-based view of "abundance" monitoring that clearly acknowledges two stochastic processes involved when monitoring abundance *N* (we use the term "abundance" interchangeably for all three varieties of *N*) at more than a single site: first, a spatial sam-

ple, and second, a sample of all individuals present in the spatial sample is observed. In the next section, we expand briefly on this, describe the conceptual relationship between a count and true N via detectability p, which may include a component for availability (Pollock et al. 2004, Nichols et al. 2008), and explain why explicit accounting for p is important in monitoring and ecology. We describe the methodological framework of site-occupancy models (MacKenzie et al. 2002, 2003), a statistical method to estimate species distributions free of distorting effects of imperfect detectability. We then present two case studies of national monitoring schemes that highlight the importance of accounting for detectability. Throughout the article we use the term "monitoring" in a general sense and do not distinguish between single inventories (surveys) and inventories that are repeated over time (i.e., monitoring in the strict sense) since the issues we deal with have relevance for both.

Monitoring abundance as a two-stage sampling process

A sample of space

Typical monitoring programs want to make inference about large areas, such as national parks, countries or perhaps even an entire continent. Usually, therefore, the area of inference is larger than the examined area and the studied area forms a sample of the former; see circles within the larger shaded (statistical) population in Fig. 1. One main design task is to lay out the spatial sample (sites) in such a way that what is measured in them properly reflects the same quantity in the larger area, and that estimates be as unbiased and accurate, while presumably also as cheap as possible (Thompson 2002, Williams et al. 2002, Thompson 2004).

Usually, a spatial probability sample, i.e., typically one that is random of some kind, e.g., completely random, stratified random or adaptive (Thompson 2002), is essential for making valid inference from the set of selected sites to the entire population of sites. Deviations from this ideal may be unavoidable in many cases, and very few monitoring programs are based on a truly random spatial sample, rather, "convenience" sampling is common (Anderson 2001). However, it is important to recognize that this creates a danger of losing the ability to generalize to the entire area about which inference is desired, unless selection bias can be modelled (Little 2004).

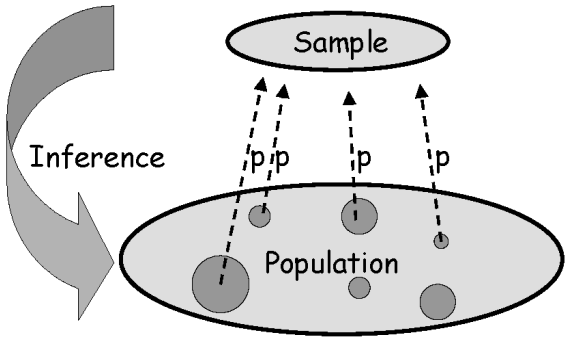

Figure 1. A monitoring program for 'abundance' conceptualized as a two-stage sampling process; first, a spatial sample (gray circles) is taken within the wider area about which inference is desired; second, only a fraction (i.e., a sample of individuals) of those present in the spatial sampling units are detected; even none are detected in some units. The quantity measured could be abundance, occupancy status (distribution) or number of species (species richness).

A sample of individuals, occupied areas or species

The sample of individuals or species counted or of detection/nondetection observations is obtained by surveying each unit in the spatial sample. This second step of sampling yields the observations called "sample" in Fig. 1. It includes another stochastic process: typically not every individual or species is detected nor is occurrence of a species determined with certainty at each sampled site. Instead, only a fraction p of all individuals, occupied patches or species, will be detected, where p is called detectability.

Although $p = 1$ is possible in principle, it will be a rare event. Whenever detection is uncertain ($p < 1$), counts of individuals are random variables and not fixed quantities. Flipping a coin is the best analogy of the counting process: each individual, occupied patch or species present flips a coin to determine whether it is detected (i.e., counted) or not. The coin is loaded and can have any probability p of heads between 0 and 1. This has important consequences for the resulting counts or detection/nondetection observations, since counts may vary, even under identical conditions, from one observation to the next, and they can only be described "on average".

Assuming detections are independent and absent double counts or misidentifications, the counted number n among N can be described by a binomial distribution, $n \sim Bin(N,p)$, where p is the probability that a member of N is detected and appears in the count. In the case of species distributions, detection/nondetection observations in each spatial unit are Bernoulli trials (a binomial distribution with $N = 1$); in an occupied patch at least one member of its inhabitants is recorded with probability p. The average or expected value of a count of individuals or species over repeated realizations of the count random variable is

$$E(n) = N * p \quad (1)$$

hence, the expected count $E(n)$ equals the product of true "abundance" N and detectability p. For example, when there are $N = 16$ greenfinches and detectability p for each one of them is 0.6, counts will average 9.6, may (rarely) be zero or 16 and more frequently any number in between. Given a count n and an estimate $\hat{p}$ of detectability p, an estimate of population size can be obtained by rearranging eq. 1: $\hat{N} = n / \hat{p}$ is the canonical estimator underlying almost all methods of population size estimation (Williams et al. 2002). Similarly, for detection/nondetection (distribution) data

$$E(\eta) = \psi * p \quad (2)$$

i.e., the average observed occurrence at a site $E(\eta)$ is the product of true distribution (expressed as site occupancy probability ψ) and detectability p.

Whenever $p < 1$, counts will be smaller than true abundance N. Similarly, patterns in observed counts or distributions will confound patterns in N and patterns in detectability p. Hence, an observed pattern over time in abundance or species richness (i.e., a trend T) based on counts at two time points is

$$E(T_{obs}) \approx E(n_2 / n_1) \approx \frac{N_2 * p_2}{N_1 * p_1} \quad (3)$$

In contrast, the true trend (or more generally, relative abundance; MacKenzie and Kendall 2002) is $E(T_{true}) = N_2/N_1$. A trend estimate based on observed counts will therefore only be unbiased for the true trend if $p_1 = p_2$.

This is illustrated using simulated data in Fig. 2, where a species with a shrinking (case A) and one with a stable distribution (case B) is shown. Detectability of an occupied patch increases, perhaps as a result of increased numbers of better observers or improved survey methods. In 267 sampled quadrats, this leads to wrong diagnoses: a true distributional trend is masked (case A) and an artefactual trend diagnosed (case B). Many scenarios are possible; depending on patterns in the distribution and p.

Conceptually, detectability can be partitioned into components (Marsh and Sinclair 1989, Pollock et al. 2004, Nichols et al. 2008), among them availability for detection (or exposure to sampling), and detection, given availability. Frequently, part of the sampled population may be unavailable for detection in principle; an individual may be temporarily in a part of its activity range outside the sampled quadrat, a bird in an acoustic survey may not sing during a 10 min sampling interval, a marine mammal in a visual surface survey may be diving or a salamander underground. A more general form of eq. 1 is therefore

$$E(n) = N*a*p \quad (4)$$

where a is the probability that a member of N is detectable in principle (availability) and p the probability to detect it, given availability. There is a certain symmetry in availability probability; one may say that population size is reduced (i.e., $N' = N*a$), and each member of N is detectable with probability p, or equivalently, that each member of N is detectable with net detectability $a*p$. To determine whether one might be estimating N or N' it is helpful to consider the time scale at which availability might be changing. If availability changes daily, and repeated surveys are only conducted on a single day, then one is likely estimating N', but if surveys are conducted on different days then one might be closer to estimating N. A key consideration then is which quantity is of more biological interest. Thus, in any particular monitoring

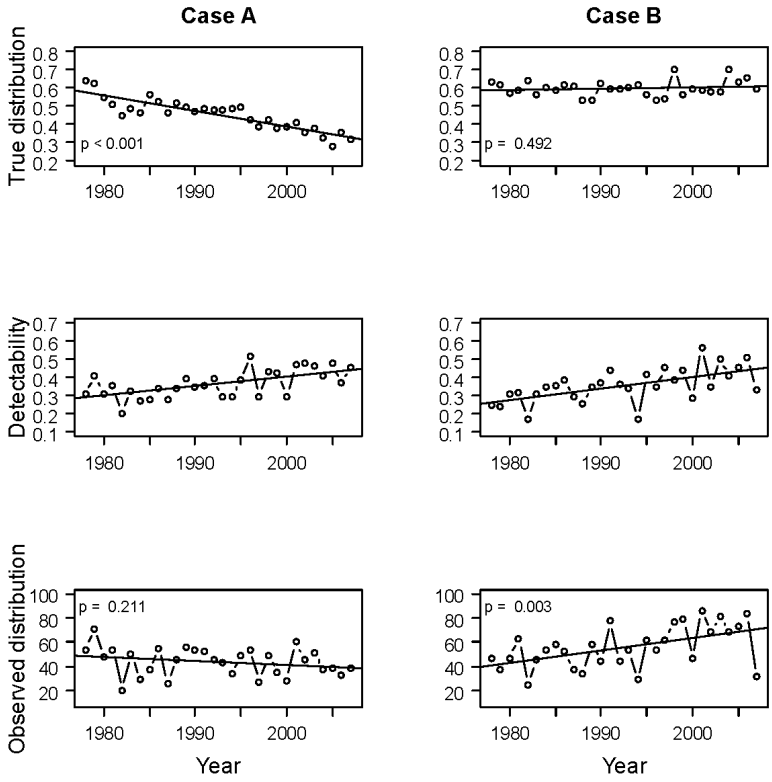

Figure 2. Risks of ignoring detectability in monitoring and ecology as illustrated by simulated distribution assessments. In A, a truly shrinking distribution would not be diagnosed correctly because of an increase in detectability. In B, a stable distribution would be diagnosed wrongly as expanding. Insets give p-values of test for a linear trend over time.

program involving counts, it is important to know what part of the population one is actually sampling. Often, only net detectability is estimable and N is estimated directly, while sometimes, only N' can be estimated, and then, to estimate N, additional information on availability is required (Kendall 1999, Pollock et al. 2004).

Attitudes towards detectability

Three attitudes towards possible complications engendered by the fact of $p < 1$ are observed: ignorance, standardisation, and estimation (Schmidt 2004, Mazerolle et al. 2007). Proponents of ignorance assume (usually tacitly) $p = 1$ or at least that the expectation $E(p)$ is constant across all dimensions of desired comparisons such as time, space, species or habitats. Typically, this strong assumption of constant is neither tested nor even stated explicitly, although the assumptions of perfect or constant have probably been proven wrong whenever they were tested, e.g., in plants (Kéry et al. 2006), fish (Royle and Dorazio 2006) or frogs (Schmidt 2005).

The standardisation attitude is based on the fact that if all that matter are patterns (e.g., a species becomes "more or less common" over time, or is "more common in habitat A than in habitat B"), then p is irrelevant as long as it stays constant over dimensions of comparison (i.e., $E(p)$ = constant). Hence, if one is only interested in temporal trends, p may be < 1 or even vary among years or habitats, but there must not be an annual trend in p. Variation around a constant expectation over time will just add noise into the observed counts, but no bias relative to the true population trajectory (Bart et al. 2004).

Standardisation is important in monitoring because the more factors causing variation in p are eliminated, the better the ability to discern real patterns in N. This is widely recognized and most monitoring programs contain elements of standardisation. This may be called standardisation at the design and execution stage. Unfortunately, analysis of monitoring programs where standardized methods were employed show that detectabilities were not constant (Schmidt 2005, Hochachka and Fiedler 2008).

There is also standardisation at the analysis stage: when confounding factors suspected to affect counts or detection/non-detection observations via their effect on p are measured during surveys, their effects can be eliminated using a statistical model. This is important and has been applied to correct for variation in p caused by observer identity and effort in a continent-wide avian monitoring program (Link and Sauer 2007). A crucial caveat, though, is that only covariates that are uncorrelated with abundance itself must be used (Nichols et al. 2008).

Whenever there is an interest in true abundance rather than relative abundance or when 'dangerous' patterns in detectability are suspected to be present, there is no way around estimating p to correct observed counts or distributions for imperfect detection. There is now an armada of methods available for that (Buckland et al. 2001, Williams et al. 2002, MacKenzie et al. 2006). Here, we give two examples of site-occupancy models (MacKenzie et al. 2002, 2003) that fully correct for p and estimate the desired population quantity in an unbiased way. Site-occupancy models are particularly useful for many monitoring programs; therefore, we briefly introduce them first.

Primer on site-occupancy modeling of species distributions

The site-occupancy model is a species distribution model that accounts for possible nondetection error. The main goal of the site-occupancy model is estimation of the proportion of sampling units occupied by a species and to learn about factors that determine species distributions and the factors that affect nondetection (MacKenzie et al. 2006).

The sampling unit ("site") definition is up to the researcher. It may be a woodlot, 1 km^2 quadrat, pond, etc. The proportion of occupied such units is indeed a frequently used and useful metric in many monitoring programs.

Only detection/nondetection data need to be collected across a large number of sites, and these are usually easier and cheaper to collect than abundance data in large-scale monitoring programs (Weber et al. 2004, Joseph et al. 2006). Key element of the model is allowance for imperfect detections and for that, it requires replicate observations of occurrence in at least some sites. Site-occupancy models are really just two coupled logistic regressions where one binomial distribution describes the true distribution and the other the observations, conditional on occurrence. Data used are 'detection/nondetection histories', i.e., survey results from repeat visits to the same sites. Assume a site is visited three times during a season (within which occupancy state is assumed constant); one resulting 'detection history' may be "detected – not detected – detected". Hence, the species is present but imperfectly detected. From elementary probability, MacKenzie et al. (2002) define the probability of this history as $\psi p_1 (1 - p_2) p_3$, where ψ is probability of site occupancy and p_1 is detectability at occasion i. In its multi-year version, the site occupancy model also yields estimates of colonisation and extinction rates (MacKenzie et al. 2003).

For both cases, the combined probability of all detection histories and their frequencies together form the likelihood of the observed data and standard statistical techniques can be used to obtain maximum likelihood parameter estimates. All four parameter types can be modelled as functions of covariates (MacKenzie et al. 2002, 2003).

Site-occupancy models make the following key assumptions: (i) colonisation and extinction only take place between seasons (the closure assumption, see case study 2 for how closure problems may be relaxed), (ii) species are not falsely recorded when absent (but see Royle and Link 2006), (iii) all sites have the same probability of occupancy, extinction, colonisation and detection, except insofar as differences can be modeled by covariates or using random effects (Royle 2006).

Table 1. Model selection results for slow worm (*Anguis fragilis*) survey data (Case study 1). *K* is the number of parameters in the model and (.) denotes a constant parameter (i.e., no covariate). AIC of the best model was 472.89. Based on 1000 bootstraps, the model fit the data well ($\chi^2 = 16.01$, $P = 0.32$). Parameter estimates are given in the main text.

Model	K	ΔAIC	AIC weight
Ψ(1/mean survey duration) p(survey duration)	4	0.00	0.57
Ψ(.) p(survey duration)	3	1.09	0.33
Ψ(1/mean survey duration) p(.)	3	3.65	0.09
$\psi(.)p(.)$	2	13.07	<0.01

Case Study 1: Slow worm distribution in Switzerland

Our first example illustrates a case where the "availability" component of detection can be teased apart from "detection, given availability." We analyse a survey aimed at estimating the distribution occupancy of a fossorial legless lizard, the slow worm (*Anguis fragilis*), as part of a national multi-species survey that provided the basis for assessing the Red List status of Swiss reptiles (Monney and Meyer 2005). We randomly selected 294 1-km^2 quadrats and surveyed each 1–3 times from April-October in either 2003 or 2004. Field workers were free to choose the day of visit and survey duration (mean survey duration per visit = 218 min, SD = 107 min). Herpetologists preferentially searched areas appearing most suitable for reptiles, using visual encounter surveys with occasionally cover objects turned. For financial reasons, no attempt was made to use cover boards or other objects that increase detectability of reptiles (e.g., Reading 1997). For full details of methods, see Monney and Meyer (2005).

Slow worms were detected in 65/294 (22.1%) of quadrats, a very low proportion given the Swiss herpetologists' consensus that this species is the most abundant and widespread among Swiss reptiles (Hofer et al. 2001). To estimate the Swiss distribution corrected for all components of detectability, we used PRESENCE 2 to fit site-occupancy models to the survey data.

First, we allowed for imperfect detection but assumed probabilities of occupancy and detection were constant (model $\psi(.)p(.)$, Table 1); estimated occurrence probability ($\hat{\psi}$) was 38.3% (SE 6.0) and detection probability ($\hat{p}$) 33.9% (SE 5.4). Thus, accounting for imperfect detection almost doubled the estimated size of Swiss slow worm distribution relative to the "naive" estimate of 22.1%. Nevertheless, this improved estimate was still low in comparison with the experts' view. Since the assumption of a constant probability of detection across sites, observers, months, etc. was deemed implausible, in a next step, we included promising covariates that might explain some of this variation. For simplicity, we considered only one obvious covariate: survey duration.

The ensuing model was better supported by the data based on Akaike's information criterion (AIC, Burnham and Anderson 2002) than a model with constant detectability (Table 1). Detectability increased with increasing survey duration. However, detectability was still well below 100% even after 600 minutes of searching (Fig. 3a), and estimated occupancy ($\hat{\psi}$) was 56.3% (SE 57.1), higher than before, but still lower than expected.

Thus, in a next step we checked whether accounting for non-availability of slow worms may yield more realistic distribution estimates. Slow worms spend most of their time underground and few are visible at the surface. Thus, a large area has to be searched until one encounters a slow worm that is at the surface. Site occupancy models do not (yet) include a separate parameter to estimate availability. However, Royle et al. (2007: section "Modeling spatial coverage bias") suggest a way forward in this situation if a covariate is available that is informative on the spatial sampling intensity; they used transect route length. Transect length was not available for the Swiss reptiles but survey duration was. Greater survey duration implies a larger area searched. Therefore, route length and survey duration both describe availability in a similar way and we used average survey duration per site as a covariate on occupancy. Survey duration likely affects the number of slow worms exposed to sampling via the effectively sampled area and all else equal, larger areas are more likely to be occupied than smaller areas. Assuming diminishing returns of survey duration (i.e., increasingly redundant searches), the reciprocal of survey duration is the appropriate form of the covariate and the intercept of the linear model that relates occupancy probability to inverse survey duration can be interpreted as the estimated occupancy in an exhaustively sampled 1 km^2 quadrat (see Royle et al. 2007, for discussion).

This model was slightly worse in terms of AIC than the model that included survey duration as a covariate for detectability only (Table 1). However, survey duration may simultaneously affect availability and detectability, and indeed, a model where survey duration was included as a covariate on both, taking account of (spatial) availability and detectability, given availability was best supported by the data (Table 1). There was a positive effect of survey duration

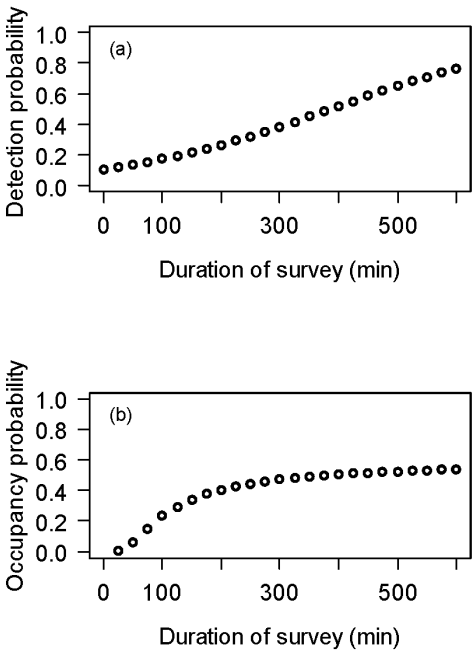

Figure 3. (**a**) Relationship between survey duration and detectability of slow worms (*Anguis fragilis*). (**b**) Effect of availability for detection on site occupancy. Availability was expressed as inverse survey duration and used as a covariate for occupancy.

as a surrogate for availability (i.e., effective sampling area) on site occupancy (Fig. 3b) and estimated site occupancy $\hat{\psi}$ reached an asymptote at ~ 60% after >1000 min average survey duration. Our analysis shows that inference regarding Swiss slow worm distribution differed markedly among models and that the most realistic distribution estimate was obtained with multiple components of detection taken account of.

Case Study 2: Distribution of Swiss breeding birds

Our second example illustrates a case where the "availability" component of detection cannot be teased apart from "detection given availability" and therefore, net detection is estimated. Hence, for species with large activity ranges, the proportion of quadrats used is estimated rather than the proportion of quadrats inhabited permanently (see discussion for more on this important distinction). 'Monitoring Häufige Brutvögel' (MHB; Schmid et al. 2004), the Swiss national breeding bird survey, has been conducted annually since 1999 and aims at assessing distribution and abundance of breeding birds. A random sample of 267 1-km^2 quadrats was laid out as a grid (Kéry and Schmid 2006) and each quadrat was surveyed three times per breeding season (15 April–15 July) by volunteers using territory mapping. Quadrats above the timberline were surveyed only twice. Surveys followed a quadrat-specific transect route averaging 5.1 km (range 1.2–9.4 km) that attempts to cover as much area as possible. Mean duration of single surveys in 2001–2004 was 229 min (range 60–427); first to third survey dates averaged 8 May, 28 May, and 8 June. Each survey aims to map as many species and individuals as possible out of about 150 targeted breeding bird species. Maps from all surveys in one season are later overlaid and putative territories determined based on known species-specific territory size and clustered observations. For our current analysis, territory counts were quantised, so for each species, quadrat and year, we obtained a detection history.

To characterise each species in terms of the difficulty with which it is found in MHB, we estimated a single average detectability for each by fitting a simple dynamic site-occupancy model (MacKenzie et al. 2003, Royle and Kéry 2007) to all MHB data from 266 quadrats surveyed 2001–2004. For some late-arriving migrants in our analysis, within-season occupancy status may not be constant over all three surveys. Therefore, data from surveys taking place before species-specific threshold dates, when the bulk of the population of a species had arrived, were turned into missing values, i.e., not included in the analysis (MacKenzie et al. 2002).

We used WinBUGS (Spiegelhalter et al. 2003) to fit the model to all four years of data for the 143 species detected in at least one year (see Royle and Kéry (2007) for WinBUGS code). We ran three Markov chains with random initial values for 2000 iterations each, discarded 1000 as burn-in and thinned the remainder by 2, which resulted in 1500 iterations for inference. Judging by a Gelman-Rubin statistic of <1.1 for all parameters (Gelman and Hill 2007), this was sufficient to obtain convergence for all but 19 species. For them, we ran up to 60,000 iterations, with a burnin of 20,000. We excluded from analysis 15 species which were not detected in ≥ 1 quadrats in all four years.

Among the remaining 128 species, detection in occupied quadrats was far from certain. On average, Swiss bird species were detected with probability of only 64% (Fig. 4a). Species differed tremendously in detectability. The least detectable species, *Pernis apivorus*, was detected in occupied quadrats in barely 3% of surveys, while the easiest species to detect, *Fringilla coelebs* and *Passer hispaniolensis*, had an estimated p of essentially 1 (see Appendix). One consequence of is that the distribution of the vast majority of species is underestimated. This is shown in Fig. 4b, where the relative underestimation of distribution is shown as a percentage of the estimated true occupancy. Obviously, the distribution of more elusive species was more strongly underestimated. For about half of Swiss bird species, with $p <\approx 0.7$, range size in MHB would be clearly underestimated when based on raw detections only. When the objective of a monitoring program is accurate assessment of species distribution, an alternative to estimation would be to either increase survey duration or survey number until every species is detected when present with some predetermined level of confidence, e.g., 95%. The combined probability to detect a species after n surveys is given by $1–(1–p)^n$ (Kéry 2002), where p in our case is detectability estimated under the site-occupancy model. Plotting this for some numbers of surveys (Fig. 4c) is extremely useful both for retrospectively assessing the sensitivity of a monitor-

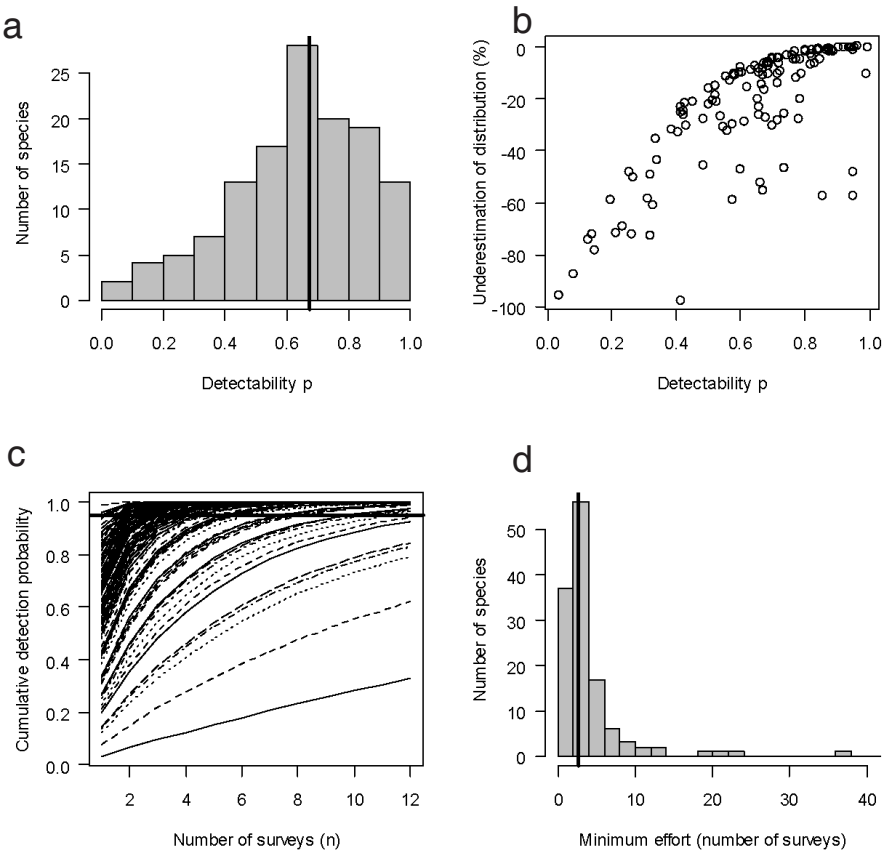

Figure 4. (**a**) Detectability (per single survey) of occupied 1km^2 quadrats for 128 bird species in the Swiss breeding bird survey MHB 2001–2004. Vertical line denotes median. (**b**) Relative underestimation of distribution (in %, relative to the estimated true distribution) for 128 species when imperfect detection is not accounted for (relative to the estimated true distribution) in relation to estimated detectability. (**c**) Sensitivity curves for occupied quadrats in 128 species showing the probability to detect a species at least once during n surveys. Horizontal black line indicates 'almost certain' (95%) detection. (**d**) Minimum number of surveys required for 'almost certain' detection for 127 bird species (one with 91 not shown). Vertical line is median of 2.7.

ing program and as a prospective tool for survey design. Obviously, for most species repeated surveys are required for 'almost certain' probability of detection (median 2.7; Fig. 4d).

Discussion

We have described a formal sampling-based view of the monitoring of "true" abundance, species distributions and species richness that explicitly recognizes the two sampling (i.e., stochastic) processes involved when monitoring animals and plants at more than a single site. The first process is a spatial sample and the second, a sample of individuals, occupied sites or species. To draw valid inference on the state of abundance, or (temporal or spatial) patterns in abundance, we need to select our spatial sample randomly and control or estimate detectability. Hardly anything of this individually is really novel (e.g., Thompson 2002, Williams et al. 2002, Thompson 2004, Royle and Dorazio 2006, Nichols et al. 2008) and yet, very few monitoring programs are based on a decent spatial probability sample nor do they account for possible "dangerous" patterns in detectability in a satisfactory way. We fear that this may in many cases seriously impair their usefulness.

In our case studies, we have focused on the issue of detectability. Our ability to detect individuals, populations and species is arguably always imperfect and this affects our ability to accurately describe the state of nature. In addition, it has perhaps not been sufficiently widely recognized that detection consists of several components (Pollock et al. 2004, Nichols et al. 2008). Our examples showed clearly that analyses of distribution data need to account for imperfect detection, sometimes even for multiple components of detection. The same holds true for related parameters such as abundance or species richness. The key insight is that observed counts are not the same as true population size or species richness, and observed distributions (=detection/nondetection observations) are not the same as true distributions (=presence/absence state); instead, they are linked with each other by (components of) detectability. Extra-information is required to partition variation in counts into its components due to true *N* and detectability.

An emerging and very powerful paradigm for the analysis of animal population data is that of hierarchical models (e.g., Royle and Dorazio 2006, 2008). One reason why these models are so useful for population analysis is that they simply replicate the hierarchical genesis of essentially all field data on animals and plants: the first level is the unobserved or only partially observed true state (e.g., abundance or occurrence) and the second is the observation process, which in a way acts like a filter or a dirty lens through which we obtain our observations.

The components of detection, availability and detectability, are easy to distinguish in theory but less so in practice or

detection may need to be split into further components still (Johnson 2008, Nichols et al. 2008). A case in point is the slow worm study where survey duration affected both availability and (narrow-sense) detectability. Despite such ambiguity, it is important to think carefully about the two (or more) components of imperfect detection when designing a monitoring program or when analysing data. The interpretation of the estimated quantity (*N* or proportion of area occupied) depends on the components of detection involved (Kendall 1999, Nichols et al. 2008). Kendall's (1999) assessment of the performance of closed population estimators for *N* serves as an example. Without mortality or migration, the Lincoln estimator for *N* (Williams et al. 2002) estimates the number of individuals within the study area. However, if there are births/immigration into the study area or deaths/emigration from the study area, it estimates the size of a superpopulation instead, that is, the number of all animals using the study area at least some time during the entire sampling period.

While detectability is relatively well understood because of decades of research and development of methods to estimate it, the understanding of availability varies widely. The need to account for availability is well understood in, e.g., distance sampling of marine mammals (Buckland et al. 2001) but poorly understood, nor even recognized as a potential problem, in many other fields and taxonomic groups. This is evidenced by the fact that many estimation models do not include an explicit parameter for availability; Pollock's robust design (Williams et al. 2002) is a notable exception.

How can we recognize incomplete availability and how can we deal with it? In general, when availability is <1, there is a mismatch in temporal or spatial terms between the scale at which a population is sampled and the desired scope of inference. In some cases, non-availability for detection is obvious and it is straightforward to deal with the problem. For example, if one wishes to know the number of butterfly species in an area and surveys are only made during May and June, then species that fly later in the season will be missed. Hence, the total number of species – the quantity of interest – remains unknown. The solution is clear: Survey the area for butterflies throughout the entire activity season and use a model that accounts for asynchronous presence at the sampling site (e.g., Condit et al. 2007, Kéry et al. in press). Another availability issue may arise from a mismatch between the spatial unit of inference and the scale of the field sampling. For example, in the Swiss breeding bird survey, spatial sampling units are 1 km^2 quadrats, but field workers often search only a fraction of the quadrat and consequently many individuals within the desired spatial scope of inference are not actually exposed to sampling. Three solutions to deal with the problem are: i) reduce the quadrat size or ii) use a covariate informative about availability (such as route length, e.g., Royle and Dorazio 2006; the latter is the approach taken in the slow worm analysis) and iii) design the survey in such a way that availability is not a problem. In general, it is clear that one always needs to think clearly about how the biology of the species, the sampling process and the model used interact to yield an estimate and how that estimate needs to be interpreted.

A simple rule of thumb that may help to recognize when availability is an issue (see also Royle and Dorazio 2006) is this: if a sampling covariate affects abundance or distribution (occupancy), then availability is probably less than 1. Accounting for them in the analysis will adjust estimates for low availability. Sampling covariates include route length (Royle and Dorazio 2006, Royle et al. 2007), survey duration (as in the slow worm example), weather conditions and observer effects. They describe the sampling process rather than a species' biology. For example, route length does not affect the true number of individuals within a sampling quadrat but instead affects the number of individuals an observer can detect at all.

In our Swiss bird example, we estimate net detectability and therefore correct our estimates of distribution for imperfect detection in the broad sense, i.e., including availability. We do not have the additional information required to partition detection into its two components. This would also give us the actual proportions of the area that are occupied at each particular survey among the three. Many species (e.g., large raptors) have activity ranges larger than our sample quadrats, so to be detected on any one visit, they must be (a) in that part of their range covered by a sampling quadrat (the availability part of detection) and (b) given that they are, they must be detected (which is narrow-sense detectability). Given the temporal frame of our bird sampling, what we estimate as "p" is in fact the combination of $p*a$ in eq. 4. Therefore, for species with activity ranges larger than a sampling quadrat, our estimate of occupancy must be interpreted not necessarily as "permanent occupancy" but rather as "use" sometime during the sampled time period (see also MacKenzie 2005, p. 850).

In summary, we believe that a more rigorous approach to both the design of monitoring programs, and their analysis, will help to obtain the best possible and least ambiguous information about the state of nature. A clear recognition of, and proper accounting for, the two sampling processes involved in all monitoring programs for 'abundance', and for the components of detection, will go a long way towards this goal.

Acknowledgements: We thank J.-C. Monney and A. Meyer for organising the Red List survey and allowing us to use the slow worm data. We also thank the volunteers who conduct field work in the Swiss breeding bird survey MHB as well as H. Schmid and G. Häfliger for providing the bird data. D.I. MacKenzie, V. Keller, T. Roth, A. Royle, H. Schmid, D. Weber, N. Zbinden and an anonymous referee commented on the manuscript. BRS was supported by the "MAVA Fondation pour la nature".

References

AERC TAC's Taxonomic Recommendations. 2003. http://www.aerc.eu/aerc_tac.htm.

Anderson, D.R. 2001. The need to get the basics right in wildlife field studies. *Wildlife Soc. Bull.* 29:1294-1297.

Bart, J., S. Droege, P. Geissler, B. Peterjohn and C.J. Ralph. 2004. Density estimation in wildlife surveys. *Wildlife Soc. Bull.* 32:1242-1247.

Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D. L. Borchers and L. Thomas. 2001. *Introduction to Distance Sampling*. Oxford University Press, Oxford.

Burnham, K.P. and D. R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information Theoretic Approach.* 2 ed. Springer, New York.

Condit, R., B.J. Le Boeuf, P. A. Morris and M. Sylvan. 2007. Estimating population size in asynchronous aggregations: A Bayesian approach and test with elephant seal censuses. *Marine Mammal Sci.* 23:834-855.

Gelman, A. and J. Hill. 2007. *Data Analysis Using Regression and Multilevel/hierarchical Models*. Cambridge University Press, New York.

Hochachka, W.M. and W. Fiedler. 2008. Trends in trappability and stopover duration can confound interpretations of population trajectories from long-term migration ringing studies. *J. Ornithol.* 149: 375-391.

Hofer, U., J.-C. Monney and G. Dušej. 2001. *I rettili della Svizzera: Distribuzione, habitat, protezione*. Birkhäuser Verlag, Basel.

Johnson, D.H. 2008. In defense of indices. *J. Wildlife Manage.* 72: 857-868.

Joseph, L.N., S.A. Field, C. Wilcox and H.P. Possingham. 2006. Presence-absence versus abundance data for monitoring threatened species. *Conserv. Biol.* 20:1679-1687.

Kendall, W.L. 1999. Robustness of closed capture-recapture methods to violations of the closure assumption. *Ecology* 80:2517-2525.

Kéry, M. 2002. Inferring the absence of a species – A case study of snakes. *J. Wildlife Manage.* 66:330-338.

Kéry, M., J.A. Royle, M. Plattner and R.M. Dorazio. in press. Species richness and occupancy estimation in communities subject to temporary emigration. *Ecology*.

Kéry, M. and H. Schmid. 2006. Estimating species richness: Calibrating a large avian monitoring program. *J. App. Ecol.* 43:101-110.

Kéry, M., J.H. Spillmann, C. Truong and R. Holderegger. 2006. How biased are estimates of extinction probability in revisitation studies ? *J. Ecol.* 94:980-986.

Link, W.A. and J. R. Sauer. 2007. Seasonal components of avian population change: Joint analysis of two large-scale monitoring programs. *Ecology* 88:49-55.

Little, R.J. 2004. To model or not to model ? Competing modes of inference in finite population sampling. *J. Amer. Stat. Assoc.* 99:546-556.

MacKenzie, D.I. 2005. What are the issues with presence-absence data for wildlife managers? *J. Wildlife Manage.* 69:849-860.

MacKenzie, D.I. and W.L. Kendall. 2002. How should detection probability be incorporated into estimates of relative abundance? *Ecology* 83:2387-2393.

MacKenzie, D.I., J.D. Nichols, J.E. Hines, M.G. Knutson and A.D. Franklin. 2003. Estimating site occupancy, colonization and local extinction when a species is detected imperfectly. *Ecology* 84:2200-2207.

MacKenzie, D.I., J.D. Nichols, G.B. Lachman, S. Droege, J.A. Royle and C.A. Langtimm. 2002. Estimating site occupancy rates when detection probability rates are less than one. *Ecology* 83: 2248-2255.

MacKenzie, D.I., J.D. Nichols, J.A. Royle, K.H. Pollock, J.E. Hines and L.L. Bailey. 2006. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence* Elsevier, San Diego.

Marsh, H., and D. F. Sinclair. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. *J. Wildlife Manage.* 53:1017-1024.

Mazerolle, M.J., L.L. Bailey, W.L. Kendall, J.A. Royle, S.J. Converse and J.D. Nichols. 2007. Making great leaps forward: Accounting for detectability in herpetological field studies. *J. Herpetol.* 41:672-689.

Monney, J.-C. and A. Meyer. 2005. *Rote Liste der gefährdeten Reptilien der Schweiz.* Bundesamt für Umwelt, Wald und Landschaft, Bern, and Koordinationsstelle für Amphibien- und Reptilienschutz in der Schweiz, Bern.

Nichols, J.D., L. Thomas. and P.B. Conn. 2008. Inferences about landbird abundance from count data: recent advances and future directions. *Environmental and Ecological Statistics*, in press.

Pollock, K.H., H. Marsh, L.L. Bailey, G.L. Farnsworth, T.R. Simons and M.W. Alldredge. 2004. Separating components of detection probability in abundance estimation: An overview with diverse examples. In: W.L. Thompson (ed.), *Sampling Rare or Elusive Species*. Island Press, Washington, USA. pp. 43-58.

Reading, C.J. 1997. A proposed standard method for surveying reptiles on dry lowland heath. *J. App. Ecol.* 34:1057-1069.

Royle, J.A. 2006. Site occupancy models with heterogeneous detection probabilities. *Biometrics* 62:97-102.

Royle, J.A. and R. M. Dorazio. 2006. Hierarchical models of animal abundance and occurrence. *J. Agric. Biol. Environ. Stat.* 11:249-263.

Royle, J.A. and R. M. Dorazio. 2008. *Hierarchical Modeling and Inference in Ecology.* Academic Press, Amsterdam.

Royle, J.A. and M. Kéry. 2007. A Bayesian state-space formulation of dynamic occupancy models. *Ecology* 88:1813-1823.

Royle, J.A., M. Kéry, R. Gautier and H. Schmid. 2007. Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecol. Monogr.* 77:465-481.

Royle, J.A. and W.A. Link. 2006. Generalized occupancy models allowing false positive and false negative errors. *Ecology* 87:835-841.

Schmid, H., N. Zbinden V. and Keller. 2004. *Überwachung der Bestandsentwicklung häufiger Brutvögel in der Schweiz.* Swiss Ornithological Institute, Sempach, Switzerland.

Schmidt, B.R. 2004. Declining amphibian populations: The pitfalls of count data in the study of diversity, distributions, dynamics, and demography. *Herpetol. J.* 14:167-174.

Schmidt, B.R. 2005. Monitoring the distribution of pond-breeding amphibians when species are detected imperfectly. *Aquatic Conservation: Marine and Freshwater Ecosystems* 15:681-692.

Spiegelhalter, D., A. Thomas and N.G. Best. 2003. *WinBUGS User Manual, Version 1.4.*, MCR Biostatistics Unit, Cambridge.

Thompson, S.K. 2002. *Sampling*. Wiley, New York.

Thompson, W.L. 2004. *Sampling Rare and Elusive Species*. Island Press, Washington.

Weber, D., U. Hintermann and A. Zangger. 2004. Scale and trends in species richness: considerations for monitoring biological diversity for political purposes. *Global Ecol. Biogeogr.* 13:97-104.

Williams, B.K., J.D. Nichols and M.J. Conroy. 2002. *Analysis and Management of Animal Populations*. Academic Press, San Diego.

Appendix

Observed occupancy and estimated detectability of an occupied quadrat and estimated occupancy under a dynamic site-occupancy model for 128 Swiss breeding bird species in the Swiss breeding bird survey MHB 2001–2004. Downloadable from the publisher's web site at www.akademiai.com.

Received January 30, 2008
Revised June 14, 2008
Accepted October 18, 2008