



Observers are a key source of detection heterogeneity and biased occupancy estimates in species monitoring

Benedikt R. Schmidt^{a,b,*}, Sam S. Cruickshank^{c,1}, Christoph Bühler^d, Ariel Bergamini^c

^a info fauna karch, Bâtiment G, Bellevaux 51, 2000 Neuchâtel, Switzerland

^b Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

^c Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111, CH-8093 Birmensdorf, Switzerland

^d Hintermann & Weber AG, Austrasse 2a, CH-4153 Reinach, BL, Switzerland

ARTICLE INFO

Keywords:

Amphibian
Citizen-science
Detection probability
Heterogeneity
Monitoring
Observers
Occupancy
Surveys
Volunteer monitoring

ABSTRACT

Reliable assessments of population status and trends underpin conservation management efforts but are complicated by the fact that imperfect detection is ubiquitous in monitoring data. We explore the most commonly considered variables believed to influence detection probabilities, quantifying how they influence detectability and assessing how occupancy rates are impacted when a variable is ignored. To do so, we used data from two multi-species amphibian monitoring programmes, collected by volunteers and professional surveyors.

Our results suggest that although detection rates varied substantially in relation to commonly considered factors such as seasonal and annual effects, ignoring these factors in the analysis of monitoring data had negligible effect on estimated occupancy rates. Variation among surveyors in detection probabilities turned out to be most important. It was high and failing to account for it led to occupancy being underestimated. Importantly, we identified that heterogeneity among observers was as high for professional surveyors as for volunteers, highlighting that this issue is not restricted to citizen-science monitoring.

Occupancy modelling has greatly improved the reliability of inference from species monitoring data, yet capturing the relevant sources of variation remains a challenge. Our results highlight that variation among surveyors is a key source of heterogeneity, and that this issue is just as pertinent to data collected by experts as by volunteers. Detection heterogeneity should be accounted for when analysing monitoring data. Furthermore, efforts to increase training of field crews and collecting data to quantify differences between observer abilities are important to avoid biased inference resulting from unmodelled observer differences.

1. Introduction

Obtaining robust knowledge on species abundance and distribution, two important essential biodiversity variables, is one of the fundamental challenges in ecology and conservation (Noss, 1990; Pollock et al., 2002; Jetz et al., 2019). Unfortunately, imperfect detection of species hampers the estimation of distribution and abundance. Imperfect detection has long been recognised as ubiquitous in monitoring and failing to account for detectability when assessing population status or trends can lead to bias (Preston, 1979; Pollock et al., 2002; Nichols and Williams, 2006) which can readily impact upon conservation management (Kéry and Schmidt, 2008; Beaudrot et al., 2016; Cruickshank et al., 2016). In recognition of this issue, monitoring programmes have adapted and now

increasingly collect the data required to estimate detection probabilities and account for imperfect detection (Nichols and Williams, 2006; Ficetola, 2015; Guillera-Aroita, 2017). Many large-scale monitoring programs focus upon occupancy as a state variable because repeated detection/non-detection data are relatively straightforward to collect (Kéry and Schmidt, 2008; Sewell et al., 2010; Powney et al., 2019) and such data can be used to estimate occupancy probabilities adjusted for imperfect detection using occupancy models (MacKenzie et al., 2002; Guillera-Aroita, 2017). However, even if detection probabilities are estimated and can be modelled using suitable covariates in occupancy models, occupancy estimates can be biased if model assumptions are not met. An important assumption of such models is that factors that may affect detection are explicitly accounted for. In particular, there should

* Corresponding author at: info fauna karch, Bâtiment G, Bellevaux 51, 2000 Neuchâtel, Switzerland.

E-mail address: benedikt.schmidt@ieu.uzh.ch (B.R. Schmidt).

¹ Current address: Swiss National Park, Runatsch 124, Chasté Planta-Wildenberg, 7530 Zernez, Switzerland.

be no unmodelled heterogeneity in detection probabilities (MacKenzie et al., 2002; Guillera-Aroita, 2017; Altwegg and Nichols, 2019).

Many factors may affect detection of species and lead to spatial and temporal heterogeneity in detection, including species characteristics (e.g. vocalisation behaviour or cryptic patterning), spatial variation (e.g. habitat characteristics and local abundance), survey methodology (e.g. survey effort, observer skills) or temporal variation (e.g., phenology, weather, or time of day) (Guillera-Aroita, 2017). Appreciating the relative importance of different sources of detection heterogeneity is important because not all sources of variation are expected to influence and bias occupancy estimates in a similar way. Ignoring some important sources of detection heterogeneity can result in strongly biased occupancy estimates, whereas the effects of ignoring others may be negligible. For example, several studies demonstrated that although detectability of amphibians was strongly affected by temperature during the survey, occupancy estimates were almost identical between models which included or excluded these effects (MacKenzie et al., 2002; Schmidt, 2005). In contrast, Kéry (2004) showed that estimates of extinction probability of plant populations were biased when the effect of habitat type upon detectability was ignored.

Inference from monitoring data will be most reliable and useful if the most problematic sources of detection heterogeneity are identified and incorporated into analyses. We undertook a review of 40 articles (full details of methods and references are provided in the supplementary text SI1) to describe which explanatory variables are commonly used to model detectability. The review revealed strong differences in the frequency of different explanatory variables used in occupancy studies. Of a sample of 40 studies, variables describing temporal variation (e.g. Julian date, year) were most common and were considered in 27 studies. Weather conditions were used in 20 studies, and habitat variables in 13. Nine studies considered the effects of survey methodology and effort upon detection probabilities. The effects of abundance and variation in observer skills were each included by only one study in our sample. Given that it has long been known that observers vary in their abilities to detect amphibians and other species (Kendall et al., 1996; Link and Sauer, 1998; Cunningham et al., 1999; Genet and Sargent, 2003; Weir et al., 2005; Fitzpatrick et al., 2009; Kéry et al., 2009; Dickinson et al., 2010; Farmer et al., 2012; Lardner et al., 2015; Austen et al., 2016; Casula et al., 2017), it is surprising that observer effects were so rarely considered. For example, Kendall et al. (1996) showed that removing an observer's first year of observation decreased estimates of population trend. Thus, occupancy estimates in many studies not accounting for variation in observer skills may be biased (Royle, 2006).

In this study, we use empirical data from two monitoring datasets, one collected by volunteers and one by professional surveyors using comparable field methods, to explore variation in the detection probabilities of 16 amphibian species in relation to survey-, site- and observer-specific characteristics and the effect of the choice of variables explaining detection probability on occupancy estimates. Specifically, we modelled detection in relation to annual and seasonal variation, differences in habitat and in abundance, differences among observers (including volunteers vs. professional surveyors), between observation types (aural vs. visual), and in relation to the presence of false-positive errors. We characterise patterns of detection and assess the effects of ignoring sources of detection heterogeneity by comparing occupancy estimates between models accounting for and ignoring such heterogeneity. Our primary goal was to compare the relative importance of factors that may affect occupancy estimation.

2. Methods

2.1. Monitoring data

We used data from two Swiss amphibian monitoring programmes to explore different sources of detection heterogeneity in species detection data. Both programmes conducted surveys in the same time period, and

we focus on data from 15 pond-breeding species and the *Pelodytes* genus, which contains multiple hybridising and invasive species which are challenging to identify to species level in the field (Schmidt, 1993; Dubey et al., 2014).

The monitoring programme “Monitoring the Effectiveness of Habitat Conservation in Switzerland” (<https://biotopschutz.wsl.ch/en>; hereafter, we use the German acronym WBS) is an ongoing programme tasked with monitoring the conservation status of habitats of national importance across Switzerland (Bergamini et al., 2019). Monitoring was carried out by professional herpetologists at 240 amphibian breeding sites. Breeding sites were either single ponds or pond clusters (for an example, see map of an amphibian breeding site in Fig. 1 of Siffert et al., 2022). Sites were selected from the national inventory of amphibian breeding sites of national importance (Borgula et al., 1994) using a stratified random selection process to ensure all biogeographic regions within Switzerland and all pond-breeding amphibian species were represented in a sufficiently large sample of sites. The mean site size is 24.7 ± 52.3 (mean \pm s.d.) hectares and includes the ponds and some surrounding terrestrial habitat (Borgula et al., 1994). Forty sites are surveyed in each year in a rotating-panel design such that all sites are surveyed in a 6-year period (McDonald, 2003). We used data from 2011 to 2016 in which 49 surveyors each surveyed on average 5.2 (range 1–36) breeding sites. Four monthly surveys were carried out at each site during the amphibian breeding season (March–June), in which the surveyor recorded all species observed within a 1-hour period. Additional details of the survey methodology are given in Bergamini et al. (2019) and Cruickshank et al. (2020) and supplementary text SI7.

The second dataset is from a volunteer-based monitoring programme which was initiated in the Swiss canton of Aargau in 1999 (Meier and Schelbert, 1999; hereafter, we call this programme “AMA” (amphibian monitoring Aargau)). It covers all known amphibian breeding sites within species-rich areas in the canton (see the supplement to Moor et al., 2022 for a map). We used data from the 587 amphibian breeding sites located within 10 areas designated as priority areas for amphibians (Meier and Schelbert, 1999). Each year, two or three of these priority areas are surveyed in a rotating-panel design and almost all known breeding sites within these areas are visited, such that all 10 priority areas are comprehensively surveyed in a 5-year period. We focused on data for the years 2011 to 2015, during which 115 volunteers surveyed on average 5.1 (range 1–20) amphibian breeding sites. Three surveys were carried out during the amphibian breeding season (April–July). Volunteers were allowed to participate in the surveys if they had previously undertaken similar survey work, carried out a formal amphibian identification course, or were students committed to learning amphibian species through self-study. Additional details of survey methodology are given in Schmidt (2005), Cruickshank et al. (2019), Moor et al. (2022) and supplementary text SI7. Not all 16 species native to Switzerland and included in the WBS monitoring programme are present within the canton of Aargau; consequently, this dataset focusses on only the 12 species which were present within this region.

Comparable survey protocols (time-constrained visual and aural surveys) were used in both monitoring programmes, with observers instructed to report the presence of species based on detections of any life-stage (eggs/larvae/juvenile/adult) or the detection of calls. Surveyors were also instructed to count the number of individuals whenever possible (e.g., egg masses, adults, calling males). In WBS, visual detections were reported separately from detections of calls, allowing distinction between these two observation types. In both datasets, all visits to a site within a given year were carried out by the same surveyor.

The monitoring programmes focussed on the detection of species within breeding ponds. However, differences in breeding phenology mean that not every species would have been present in the ponds during every survey (i.e. available for detection; Kéry and Schmidt, 2008; Nichols et al., 2009). We therefore carried out species-specific dataset filtering prior to analysis in order to assure population closure as recommended by MacKenzie et al. (2002). Within each dataset and

year, we identified the first and last date that each species was reported and considered these dates the start and end of the breeding season for that species (MacKenzie et al., 2002). Surveys carried out outside of the breeding season for each species were excluded from the analyses. The mean effective number of surveys available per species and monitoring programme is given in SI2. The AMA monitoring data are available from the WSL repository “envidat” (Moor et al., 2021).

2.2. Data analysis

We used Bayesian occupancy modelling (MacKenzie et al., 2002; Royle and Kéry 2007; Kéry and Schaub 2011). The assumptions of occupancy models are: (1) occupancy status at each site does not change over the survey season such that sites are “closed” to changes in occupancy; (2) the probability of occupancy is constant across sites, or differences in occupancy probability are modelled using covariates; (3) the probability of detection is constant across all sites and surveys or is a function of site-survey covariates; there is no unmodelled heterogeneity in detection probabilities; and (4) detection of species and detection histories at each location are independent (MacKenzie et al., 2016, p. 104). The goal of the analysis was to examine heterogeneity in detection probabilities in relation to sources of heterogeneity and to estimate occupancy rates. Models were fitted independently to each of the two datasets and species. As our focus was on the effect of detection heterogeneity on occupancy rates, we used an intercept-only model (i.e. no covariates) to estimate a constant occupancy probability for all sites, varying only the specification of the detection process. We considered seven factors which might induce detection heterogeneity, allowing detection probabilities to vary in relation to: i) year, ii) seasonality (Julian date), iii) habitat area, iv) abundance, v) heterogeneity among observers, vi) differences in observation type (aural or visual), and vii) the presence of false-positive errors. Weather data were not available. We compared the occupancy estimates for each model to a null model which assumed constant detection across all sites. Furthermore, to test the impact of species availability on detectability (sensu Kéry and Schmidt, 2008; Nichols et al., 2009), we also ran a null model upon the datasets prior to data filtering.

To examine annual differences in detection, we treated year as a categorical factor influencing detection probability. Seasonal and habitat area effects were modelled using a linear and quadratic effect of Julian date or habitat area upon detection probability, respectively. Priors for the regression coefficients were specified as Uniform (−10,10) for analysis of habitat area effects, and Normal (0,0.01) for the analysis using Julian dates.

We fitted observer identity as a random effect to examine differences among surveyors in detection probability (within data sets, i.e., within the AMA and within the WBS data set), using the vague prior specifications of Stolen et al. (2019). As can be seen in the code in the supplement SI3, the prior distribution for the surveyor effect was Normal(0, σ^2) where σ was given a truncated half-normal(0, 1/2.252, 7) prior distribution (Stolen et al., 2019, following Gelman et al., 2008). To compare observer heterogeneity in visual and aural detections, we applied the same approach to the WBS dataset in which these two detection types were separately recorded. For the analysis of visual detections and call surveys, the dataset was split up. Any observations where $N(\text{calls}) > 0$ counted as a detection of calls, and all observations with $N(\text{seen}) > 0$ as a detection of observed animals. This means that a single survey can count as a detection for both classes simultaneously. Abundance-induced variation in detection probabilities (Dorazio, 2007; Tanadini and Schmidt, 2011) was assessed using the model of Royle and Nichols (2003), with λ (the parameter for the average abundance) modelled using a Uniform (0,10) prior. In the Royle-Nichols model, the estimated detection probability is for an individual at a site (in contrast to the detection of the species at a site, as in all other models).

Finally, we examined the impact of false-positive detections (Royle and Link, 2006) upon detection and occupancy rates using a single-

season version of the false-positive occupancy model as presented in Cruickshank et al. (2019). This model uses informative Beta (2,1) and Beta (1,2) priors for the true-detection probability and false-positive error rates (a similar approach can be found in Griffin et al., 2020), respectively, and works well as long as false positive detection probabilities are smaller than 0.15. For all other models, vague priors were used for all parameters. All model code and prior specifications are given in SI3. Occupancy parameters were given a Uniform (0,1) prior in all models except the Royle-Nichols model, which derives the occupancy probability from the proportion of sites where the abundance is non-zero. All models were run in JAGS (Plummer, 2003) using the R-package jagsUI (Kellner, 2019) using a burn-in of 5000, 3 chains, and thinning 1 in 5, with 10,000 iterations for all models except for those examining observer heterogeneity or false-positive observations; these models required 50,000 iterations to ensure convergence. Convergence was assessed using the Brooks-Gelman-Rubin statistic (Roy, 2020).

3. Results

3.1. No variation in detection probability

Models with a constant detection probability, i.e. null models, produced occupancy estimates which were generally higher in WBS than in the AMA programme (which had a smaller geographic extent), although there were some species where occupancy rates were similar in both datasets (e.g., *B. variegata*, *E. calamita*, *Pelophylax* species). Detection probabilities for null models were mostly high. Across species, they ranged from 0.44 to 0.81 in AMA and from 0.44 to 0.75 in WBS. In general, detection probabilities were lower for newts than for anurans (Table S4.1). As such, the probability of a population going undetected after several visits was low (Fig. S5.1, S5.2). The occupancy rate of the null models differed only slightly from the naïve occupancy rate (sensu MacKenzie et al., 2002; i.e. assuming perfect detection; Table S4.1).

3.2. Availability for detection

Although our data filtering process (to ensure population closure) removed visits for all species which were deemed outside of the species' breeding season, occupancy probabilities were largely unchanged when we ran models on the unfiltered data (Fig. 1, Table S4.1). Detection probabilities, however, were reduced strongly without filtering for five species (*E. calamita*, *H. arborea*, *H. intermedia*, *T. cristatus*, and *T. carnifex*) and declined slightly for the remaining species within the WBS programme. In contrast, for the AMA monitoring programme the filtering step had a negligible effect on detection probabilities, perhaps because visits were done later in the year and over a shorter period of time (Table S4.1).

3.3. Temporal effects

Detection probabilities often differed substantially between years (Fig. S5.3; it should be noted that due to the rotating panel design in both datasets different sites were visited in different years). The mean range of detection probabilities across the 5-year period of Aargau monitoring was 0.43 ± 0.21 (average across species \pm SD), which was much higher than the range seen in the 6 years of the WBS programme (0.31 ± 0.16).

When detection probabilities were modelled as a function of Julian date, detection tended to increase through the breeding season within the WBS programme, whereas the AMA programme showed detection probabilities that remained constant or declined slightly over this period (Fig. S5.4). The uncertainty around these trends (i.e. the width of 95 % credible intervals [CRI]) was comparable between datasets; in both cases the CRI were sufficiently high for rarer species such as *P. ridibundus*, *T. cristatus*, *L. vulgaris*, and *E. calamita* that it was not possible to determine whether seasonal trends in detection were present.

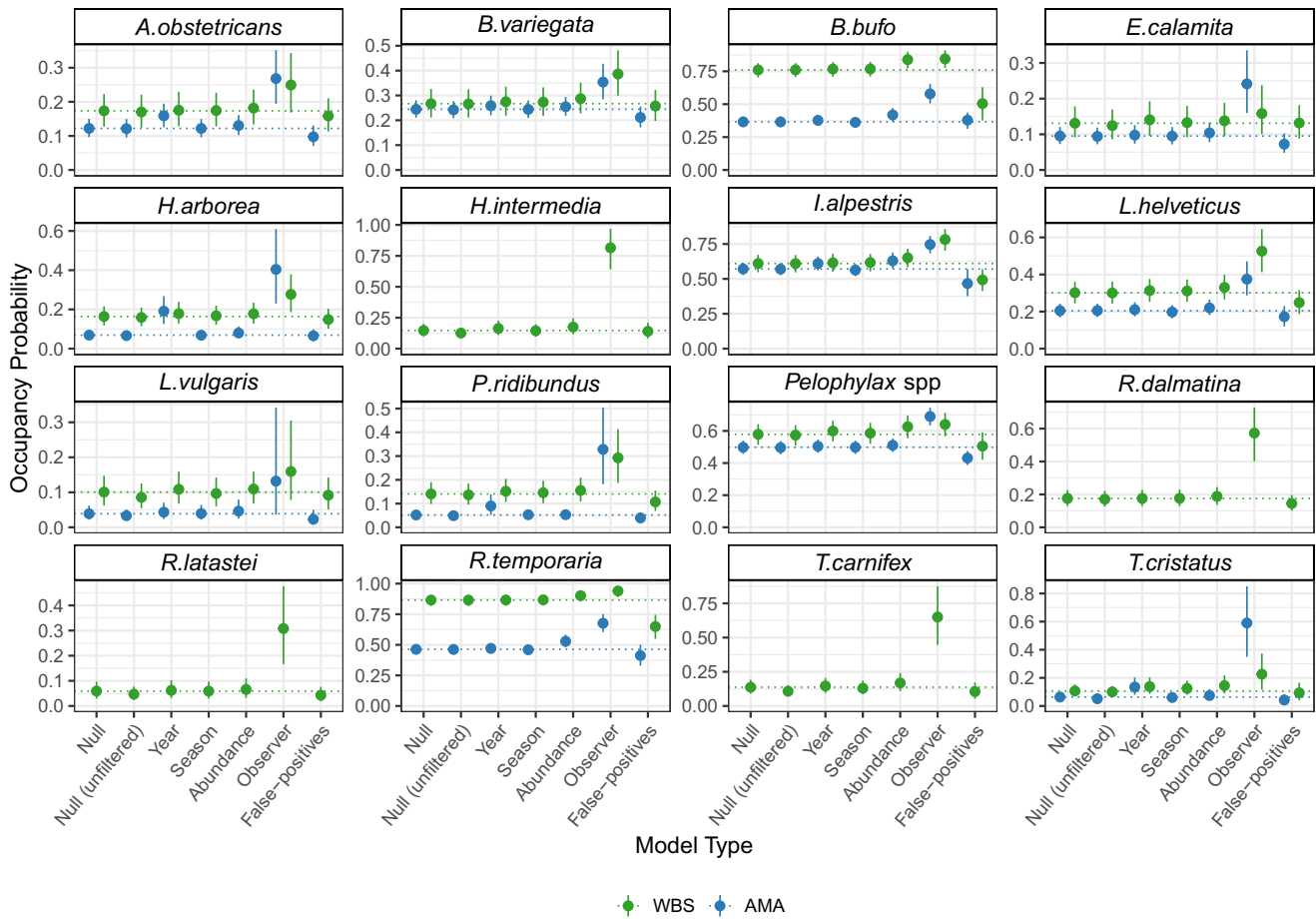


Fig. 1. Estimates of occupancy probability based on models that included different covariates for detection probability. Posterior means and 95 % credible intervals for the two monitoring programmes AMA and WBS are shown. *Hyla intermedia*, *Rana latastei* and *Triturus carnifex* do not occur in the canton of Aargau and thus in the AMA programme.

While mean detectability varied among years, the effect of Julian date was largely consistent across years and species (Fig. S5.5).

Despite sometimes substantial differences in detection probability both among and within years, accounting for this variation had only negligible effect on estimated occupancy probabilities (Fig. 1, Table S4.1), with the exception of *H. arborea* and *T. cristatus*, for which occupancy probabilities within the AMA programme were substantially elevated when accounting for annual fluctuations.

3.4. Habitat area

The area of habitat was only available for a subset of the WBS sites; we therefore carried out this analysis on the WBS sites for which area data were available. Detection probabilities for the majority of species were unaffected by the amount of habitat area (Fig. S5.6), despite the fact that all surveys were of fixed duration irrespective of site size. Detection probabilities for a number of species decreased or increased above a habitat area of 100 ha; an area that only few sites exceeded. Occupancy probabilities remained unchanged by the inclusion of a habitat area effect on occupancy (Fig. 1, Table S4.1).

3.5. Abundance

Estimates of detection estimated from the Royle-Nichols occupancy model differ from those of the other models presented in that they represent the per-individual detection probability. Per-individual detection probabilities ranged from 0.29 (*T. cristatus* in WBS) to 0.77 (*H. arborea* in WBS) and were higher in the WBS programme (mean \pm

SD = 0.62 ± 0.15) than in the AMA programme (0.55 ± 0.12 , Table S4.1). Abundance estimates at occupied sites were low (Fig. S5.7), which is further emphasised by the fact that for many species, the per-individual detection probabilities were not much lower than species-level detection probabilities (Fig. 1, Table S4.1). Occupancy estimates from this model were in most cases comparable with estimates from the null models, although accounting for abundance-induced heterogeneity in detection probabilities resulted in a substantially increased estimate of occupancy in one species, *H. intermedia* (Fig. 1, Table S4.1). There was a strong positive correlation between species occupancy probabilities and mean estimated abundance at occupied sites ($r_{27} = 0.93$, $p < 0.001$).

3.6. Observer effects

Detection probabilities were extremely heterogeneous among observers for all species in both datasets (Fig. 2). In the WBS monitoring programme, the range between the lowest and highest detection probabilities varied from 0.05 for *T. carnifex* to 0.91 for *B. variegata* (mean \pm SD across all species: 0.70 ± 0.23). The range was slightly higher in the AMA volunteer programme; observers were most consistent at detecting *L. vulgaris* (range = 0.57) and least consistent with *Pelophylax* species (range = 0.96) with the mean range of detection probabilities across all species of 0.80 ± 0.10 . Furthermore, the distribution of observer detection probabilities was more evenly spread for the professional programme, whereas the volunteer programme tended to be more skewed (Fig. S5.8) and characterised by a large number of observers with relatively low probabilities and a small number of observers with detection rates as high (or higher) than the highest rates attained by

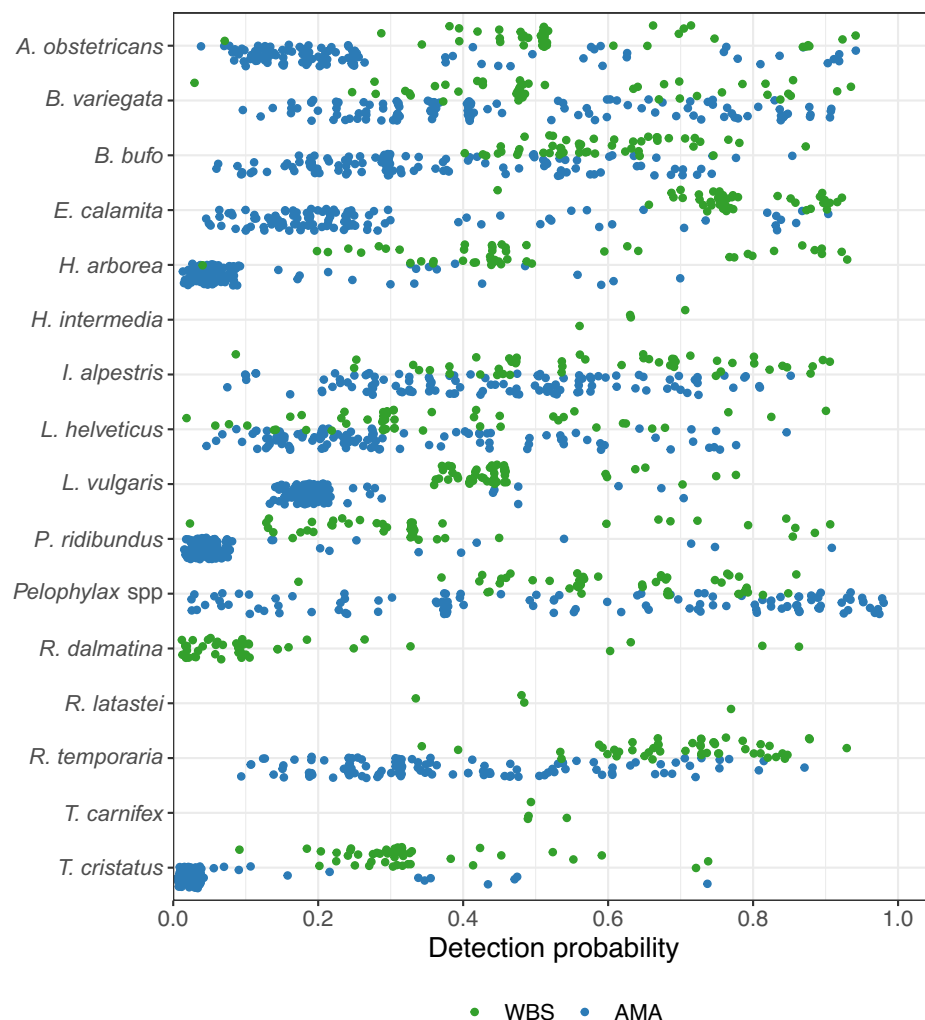


Fig. 2. Detection probabilities of individual observers. Blue points represent detection probabilities attained by each volunteer surveyor in the AMA dataset, and green points represent professional surveyors in the WBS dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

professionals. Consequently, the mean detection probability of all observers was higher within the WBS programme (mean across all species: 0.43 ± 0.18) than in the AMA programme (0.30 ± 0.16).

Allowing for variation among observers led to substantially higher occupancy estimates for all species as well as lower precision (Fig. 1, Table S4.1), with greater increases seen within the AMA volunteer programme than in the WBS programme. Estimated occupancy probabilities were $322 \pm 257\%$ and $179 \pm 106\%$ higher than the corresponding null models for the two monitoring programmes, respectively.

3.7. Detection type

Within WBS, distinctions were made between visual and aural species detections. When averaged across species and observers, detection probabilities of anurans were similar between these two detection types; the detection probability of calls was 0.30 ± 0.18 and for visual observations 0.29 ± 0.13 . There was substantial variation among observers with a small number of observers having high call detection probabilities. Notable exceptions to this pattern are *H. arborea*, *A. obstetricans* and *Pelophylax* species, which all had a more evenly spread distribution of call detection probabilities (Fig. 3). Similar patterns of variation within and among species were found in the detection rates by visual encounters. The estimated occupancy probabilities were higher than those predicted by null models, and in most cases were comparable with the more general models which allowed for heterogeneity among observers

in relation to the detection of any sign of the species (Fig. 1, Table S4.1). One notable exception is *H. intermedia*, a species with loud calls; the occupancy rate estimated from visual observations of the species was extremely low, reflecting the fact that this species is rarely seen during surveys.

3.8. False positives

Allowing for some detections to be false-positives led to high estimates of true-detection probability (Fig. S5.9). False-positive errors occurred at negligible rates for all but the most common species (*B. bufo*, *I. alpestris*, *R. temporaria*, *Pelophylax* species) but also in *T. carnifex* and *H. intermedia*, two species which occur only in southern Switzerland. For these species false positive errors were much more prevalent in the WBS monitoring programme. The probability of a professional surveyor making a false-positive observation of *R. temporaria* during a survey at a site unoccupied by the species was as high as 25 %; in contrast, the highest rate for volunteers was 7 % for *I. alpestris*. Accounting for false-positive observations led to a reduction in occupancy probabilities: species occupancy probabilities were only $85 \pm 10\%$ the rates estimated in the null models within the professional monitoring programme, and $82 \pm 12\%$ for the volunteer programme.

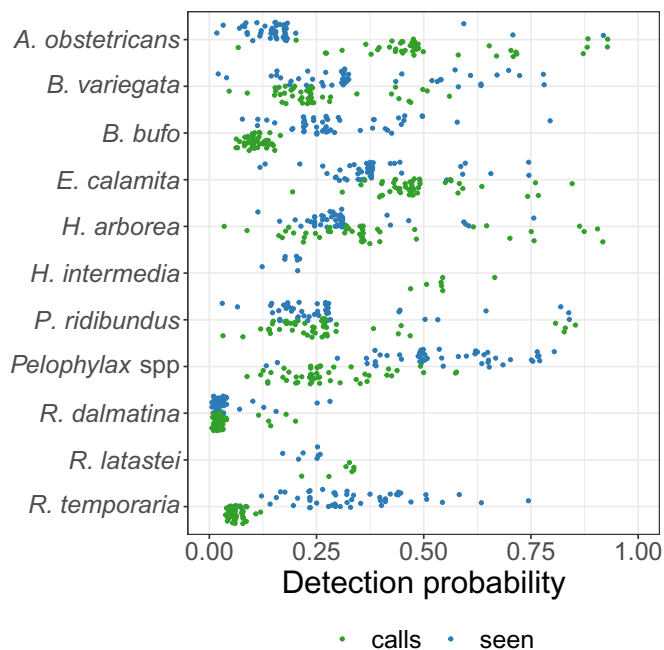


Fig. 3. Differences in detection probabilities within the WBS monitoring program between detections made through calls, and visual observation of the species. Orange points represent detection probabilities for calls attained by each professional surveyor, and green denote visual detection probabilities. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Ensuring that monitoring programmes provide reliable information to inform conservation management requires an understanding of the factors which determine how effectively species can be detected. Our results demonstrate that species detection probabilities are affected by a broad range of factors which vary between surveys, between sites, between methods, and between observers (Fig. 2–3, Fig. S5.3–S5.6). However, the consequences of accounting for these sources of variation during data analysis varied substantially. Even though detection probabilities were strongly influenced by commonly considered sources of variation, ignoring some sources of variation had little impact on estimated occupancy probabilities (e.g., year effects). In contrast, failing to account for observer-induced detection heterogeneity, i.e., differences in the abilities of surveyors to detect species, led to serious underestimation of species occupancy probabilities, as predicted by statistical theory (Royle and Nichols, 2003; Royle, 2006; Dorazio, 2007; Miller et al., 2015).

Several covariates which are commonly used to model detection probability, such as season and year, had strong effects on detectability whereas others, such as habitat area or method, had smaller effects on detectability. All these variables had in common that models that did or did not include these covariates produced very similar estimates of occupancy, independent of their effect on detectability (Fig. 1). This suggests that these variables, although commonly reported in scientific publications (see Supplement S2), may not be the most important variables to be used in the analysis of data from monitoring programmes if the goal is to estimate occupancy. In the data sets that we analysed, average detectability was high and sites were visited multiple times such that cumulative detection probabilities were high (Fig. S5.1 and S5.2). In other data sets where cumulative detection probabilities are lower, these covariates may turn out to be of greater importance than in this study.

The closure assumption is a key assumption in occupancy modelling (MacKenzie et al., 2002) because occupancy status must not change

during the time when surveys are done. This implies that the species has to be available for detection during the surveys (Kéry and Schmidt, 2008). We used the filtering approach recommended by MacKenzie et al. (2002) to truncate the data sets in such a way that the species were always available for detection. Surprisingly, and in contrast to previous studies (Royle and Dorazio, 2006; Kéry and Schmidt, 2008), non-availability for detection had negligible effects on occupancy estimates; for reasons unknown to us, a notable exception was the common frog *Rana temporaria*. In WBS, most species were available during most of the period during which surveys were undertaken (first as adults, then eggs, larvae and metamorphic juveniles) but AMA begins later such that the early-breeding species are generally only observable as tadpoles. Additionally, detection probabilities were high when species were available for detection. This may have negated the effects of non-availability on estimates of occupancy. Whether this is true in other systems remains to be evaluated.

The phenology of species is related to the closure assumption as species may not always be available for detection and detection probabilities may vary seasonally (Schmidt, 2005; Weir et al., 2005; Canessa et al., 2012). Therefore, monitoring is often planned to coincide with important phenological events such as breeding activity. The number of individuals present within sites is likely to vary over such periods, and consequently the detectability of the species will also vary (Tanadini and Schmidt, 2011; McCarthy et al., 2013). This consideration likely motivates the high proportion of studies which model seasonal variations in detection probabilities (e.g., Weir et al., 2005; Supplement S1), however, our results demonstrate that the presence of such variation need not have strong effects on occupancy estimates.

While many covariates for detection probabilities had negligible effects on occupancy estimates, allowing for false positives had strong effects in some species. For some of the common species, occupancy estimates derived from these models were much lower than those derived from models which did not allow for false positives. It is therefore important to train field crews to only report observations where species detection is unambiguous. Alternatively, they might flag an observation as uncertain and data analysts may either model certain and uncertain detection separately (Molinari-Jobin et al., 2012) or they may not use uncertain observations. Cruickshank et al. (2019) suggested that some observations which appear to be false positives may arise from spatiotemporal patterns of habitat use. For example, if adults are present at the ponds for only a short time (e.g., in explosive breeders such as *Rana temporaria* and *Bufo bufo* or species that switch ponds within a season, such as *Hyla arborea*; this would be a violation of the closure assumption) and tadpoles are difficult to observe, then one may get detection histories which have only a single observation. A false positive model may then erroneously treat such an observation as a false positive, especially if detection probability in wetlands which are used during the entire breeding season (when surveys are conducted) is high (see Sutherland et al., 2013).

The results clearly show that among-observer heterogeneity in detection probability had the greatest effect on occupancy estimates of all tested covariates. Accounting for the large differences in detection probability between observers in our study led to substantial changes in estimated occupancy probabilities. Many studies have previously identified differences among observers participating in monitoring datasets (Sauer et al., 1994; Cunningham et al., 1999; Kéry and Plattner, 2007; Kéry et al., 2009; Lardner et al., 2015; Casula et al., 2017) and failing to explicitly model such heterogeneity has long been recognised as a potentially serious problem in wildlife studies and monitoring data (Burnham and Overton, 1978; Link and Sauer, 1998; Royle, 2006). When multiple surveys are carried out at a site, the probability of a species remaining undetected generally falls rapidly with an increasing number of site visits (Pellet and Schmidt, 2005; Canessa et al., 2012; Cruickshank et al., 2016), and consequently most sites at which a string of non-detections occur will be inferred to be unoccupied. However, if there is detection heterogeneity, the least-effective observers will

generate strings of non-detections while effective observers will generate strings of detections. Detection heterogeneity leads to an overestimation of detection probability and as a consequence to an underestimation of occupancy (this is easy to verify using simulation; see Supplement 6). Thus, explicitly accounting for observer differences in detection probability will lead both to higher estimates of occupancy, and a higher degree of uncertainty surrounding the estimates (Royle and Nichols, 2003; Royle, 2006; Miller et al., 2015; Altwegg and Nichols, 2019). The latter happens because the precision (or uncertainty) of an occupancy estimate depends, among other things, on the estimate of detection probability. If detection probability is high, then the precision of the occupancy estimate will be high. If there is detection heterogeneity, then detection probability will be overestimated and the occupancy will be more precise than when one accounts for detection heterogeneity. Thus, unmodelled detection heterogeneity leads to biased occupancy estimates and the estimates have little uncertainty, meaning that one will have too much confidence in them. Thus, it is important to account for differences among observers in the analysis of species occurrence data (Johnston et al., 2023).

Species detection probabilities were overall lower within the volunteer-collected dataset than that collected by professional surveyors. Nevertheless, at least some volunteers had very high detection probabilities (Figs. 2, 3). This adds further support to the growing evidence that volunteers can carry out monitoring as reliably as expert surveyors (Genet and Sargent, 2003; Szabo et al., 2012; Aceves-Bueno et al., 2017; Callaghan et al., 2020). Furthermore, for most species there were volunteer surveyors who achieved detection probabilities at least as high as those of the best-performing professionals. Such patterns reflect the fact that participants in volunteer monitoring come from a wide range of backgrounds and many will likely have extensive knowledge of the species in question (some are professional biologists; Groom et al., 2017) but new participants in the monitoring programmes may have lower detection probabilities (Kéry et al., 2009).

Accounting for differences between observers not only influenced the estimated occupancy rate, but also decreased the precision of these estimates. In our analysis, we used random effects to model differences. However, the precision of occupancy estimates is likely to increase if the abilities of surveyors can be accurately quantified and incorporated as an explanatory variable into models; for example, through qualitative assessments of experience (Genet and Sargent, 2003; Weir et al., 2005; Kéry et al., 2009) or the use of detection test scores (McClintock et al., 2010; Miller et al., 2015) which are then used as covariates to describe surveyor-specific detection probabilities. However, the success of these variables in predicting observer detection probabilities has thus far been mixed (Genet and Sargent, 2003; Lotz and Allen, 2007; Kéry et al., 2009; McClintock et al., 2010; Miller et al., 2015). Furthermore, evidence suggests that self-reported measures of experience or confidence relate poorly to true performance, and so should be avoided if possible (Farmer et al., 2012; Austen et al., 2018). Nonetheless, monitoring programmes should increase efforts to better quantify and model differences among observers. Furthermore, monitoring programmes should invest into the training of field crews and should perhaps switch to more reliable methods. For example, capturing individuals may lead to more reliable detections than visual detections from a distance. It might also be feasible to ask observers with little experience to visit sites more often than experienced surveyors (Barata et al., 2017), but this increases the costs of the monitoring programme. Training and methodological changes will undoubtedly increase the precision and reliability of occupancy estimates derived from such studies. It may nevertheless be best to model both known observer differences (e.g., new vs experienced observers) and unstructured heterogeneity (i.e., using a random effect for observer into the models).

Designing monitoring programmes that can effectively deal with imperfect detection is an ongoing challenge, motivated by the fact that failing to disentangle biological and observation effects can readily impact species management (Tingley and Beissinger, 2009; Cruickshank

et al., 2016). Monitoring programmes often collect a broad suite of variables in order to account for factors influencing detectability, yet our results suggest that including these variables in analyses may have little to no influence on estimates of occupancy. However, differences among observers are an underappreciated form of heterogeneity in detection probability, and accounting for observer effects leads to substantial changes in occupancy estimates. Our literature survey (Supplement S1) suggests that inter-observer differences are not often accounted for, despite the existence of a large body of literature exploring inter-observer differences in volunteer datasets (Sauer et al., 1994; Cunningham et al., 1999; Weir et al., 2005; McClintock et al., 2010; Dennett et al., 2018; Johnston et al., 2018). We do not want to suggest that commonly used covariates for detection should be abandoned but rather that models for detectability of occupancy models should include observer effects. Ignoring detection heterogeneity in monitoring programmes can have unwanted consequences for conservation because ignoring heterogeneity in detection probability can lead to the underestimation of occupancy probabilities.

CRediT authorship contribution statement

Benedikt R. Schmidt: Conceptualization, Investigation, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Sam S. Cruickshank:** Conceptualization, Formal analysis, Investigation, Data curation, Writing – review & editing, Visualization. **Christoph Bühler:** Investigation, Data curation, Writing – review & editing, Project administration. **Ariel Bergamini:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data of one monitoring programme (AMA) can be found in the envdat repository. Code can be found in the supplement.

Acknowledgements

We thank the Abteilung Landschaft und Gewässer of canton Aargau for allowing us to use their amphibian monitoring dataset and Helen Moor and the reviewers for comments on the manuscript. This study would not have been possible without the participation of the many surveyors who contribute to amphibian monitoring in Switzerland. This project was funded by the Swiss Federal Office for the environment (grants: 00.5062.PZ / 67CC7844A and 00.5040.PZ / R242-1677).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.biocon.2023.110102>.

References

- Aceves-Bueno, E., Adeleye, A.S., Feraud, M., Huang, Y., Tao, M., Yang, Y., Anderson, S. E., 2017. The accuracy of citizen science data: a quantitative review. *Bull. Ecol. Soc. Am.* 98, 278–290.
- Altwegg, R., Nichols, J.D., 2019. Occupancy models for citizen-science data. *Methods Ecol. Evol.* 10, 8–21.
- Austen, G.E., Bindemann, M., Griffiths, R.A., Roberts, D.L., 2016. Species identification by experts and non-experts: comparing images from field guides. *Sci. Rep.* 6, 33634.
- Austen, G.E., Bindemann, M., Griffiths, R.A., Roberts, D.L., 2018. Species identification by conservation practitioners using online images: accuracy and agreement between experts. *PeerJ* 2018, e4157.

- Barata, I.M., Griffiths, R.A., Ridout, M.S., 2017. The power of monitoring: optimizing survey designs to detect occupancy changes in a rare amphibian population. *Sci. Rep.* 7, 16491.
- Beaudrot, L., Ahumada, J.A., O'Brien, T., Alvarez-Loayza, P., Boekee, K., Campos-Arceiz, A., Eichberg, D., Espinosa, S., Fergaus, E., Fletcher, C., Gajapersad, K., Hallam, C., Hurtado, S., Jansen, P.A., Kumar, A., Larney, E., Lima, M.G.M., Mahony, C., Martin, E.H., McWilliam, A., Mugerwa, B., Ndongou-Hockemba, M., Razafimahaimodison, J.C., Romero-Saltos, H., Rovero, S., Salvador, J., Santos, F., Sheil, D., Spironello, W.R., Willig, M.R., Winarni, N.L., Zvoleff, A., Andelman, S.J., 2016. Standardized assessment of biodiversity trends in tropical forest protected areas: the end is not in sight. *PLoS Biol.* 14, e1002357 <https://doi.org/10.1371/journal.pbio.1002357>.
- Bergamini, A., Ginzler, C., Schmidt, B.R., Bedolla, A., Boch, S., Ecker, K., Graf, U., Küchler, H., Küchler, M., Dosch, O., Holderegger, R., 2019. Zustand und Entwicklung der Biotope von nationaler Bedeutung: Resultate 2011–2017 der Wirkungskontrolle Biotopschutz Schweiz. *WSL Ber.* 85, 1–104.
- Borgula, A., Fallot, P., Rysler, J., 1994. Inventar der Amphibienlaichgebiete von nationaler Bedeutung (Schlussbericht). Schriftenreihe Umwelt Nr. 233, Bern.
- Burnham, K.P., Overton, W.S., 1978. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* 65, 625–633.
- Callaghan, C.T., Roberts, J.D., Poore, A.G.B., Alford, R.A., Cogger, H., Rowley, J.J.L., 2020. Citizen science data accurately predicts expert-derived species richness at a continental scale when sampling thresholds are met. *Biodiv. Conserv.* 29, 1323–1337.
- Canessa, S., Heard, G.W., Parris, K.M., McCarthy, M.A., 2012. Integrating variability in detection probabilities when designing wildlife surveys: a case study of amphibians from South-Eastern Australia. *Biodiv. Conserv.* 21, 729–744.
- Casula, P., Vignoli, L., Luiselli, L., Lecis, R., 2017. Local abundance and observer's identity affect visual detectability of Sardinian mountain newts. *Herpetol. J.* 27, 258–265.
- Cruckshank, S.S., Ozgul, A., Zumbach, S., Schmidt, B.R., 2016. Quantifying population declines based on presence-only records for red-list assessments. *Conserv. Biol.* 30, 1112–1121.
- Cruckshank, S.S., Bühler, C., Schmidt, B.R., 2019. Quantifying data quality in a citizen science monitoring program: false negatives, false positives and occupancy trends. *Conserv. Sci. Pract.* 1, e54.
- Cruckshank, S.S., Schmidt, B.R., Ginzler, C., Bergamini, A., 2020. Local habitat measures derived from aerial pictures are not a strong predictor of amphibian occurrence and abundance. *Basic Appl. Ecol.* 45, 51–61.
- Cunningham, R.B., Lindenmayer, D.B., Nix, H.A., Lindenmayer, B.D., 1999. Quantifying observer heterogeneity in bird counts. *Austral. Ecol.* 24, 270–277.
- Dennett, J.M., Gould, A.J., MacDonald, S.E., Nielsen, S.E., 2018. Investigating detection success: lessons from trials using decoy rare plants. *Plant Ecol.* 219, 577–589.
- Dickinson, J.L., Zuckerman, B., Bonter, D.N., 2010. Citizen science as an ecological research tool: challenges and benefits. *Annu. Rev. Ecol. Evol. Syst.* 41, 149–172.
- Dorazio, R.M., 2007. On the choice of statistical models for estimating occurrence and extinction from animal surveys. *Ecology* 88, 2773–2782.
- Dubey, S., Leuenberger, J., Perrin, N., 2014. Multiple origins of invasive and “native” water frogs (*Pelophylax* spp.) in Switzerland. *Biol. J. Linn. Soc.* 112, 442–449.
- Farmer, R.G., Leonard, M.L., Horn, A.G., 2012. Observer effects and avian call-count survey quality: rare-species biases and overconfidence. *Auk* 129, 76–86.
- Ficetola, G.F., 2015. Habitat conservation research for amphibians: methodological improvements and thematic shifts. *Biodiv. Conserv.* 24, 1293–1310.
- Fitzpatrick, M.C., Preisser, E.L., Ellison, A.M., Elkington, J.S., 2009. Observer bias and the detection of low-density populations. *Ecol. App.* 23, 515–522.
- Gelman, A., Jakulin, A., Pittau, M.G., Su, Y.-S., 2008. A weakly informative default prior distribution for logistic and other regression models. *Ann. App. Stat.* 2, 1360–1383.
- Genet, K.S., Sargent, L.G., 2003. Evaluation of methods and data quality from a volunteer-based amphibian call survey. *Wildl. Soc. Bull.* 31, 703–714.
- Griffin, J.E., Matechou, E., Buxton, A.S., Bormpoudakis, D., Griffiths, R.A., 2020. Modelling environmental DNA data: bayesian variable selection accounting for false positive and false negative errors. *J. Roy. Stat. Soc. Series C: App. Stat.* 69, 377–392.
- Groom, Q., Weatherdon, L., Geijzenroder, I.R., 2017. Is citizen science an open science in the case of biodiversity observations? *J. App. Ecol.* 54, 612–617.
- Guillera-Aroita, G., 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* 40, 281–295.
- Jetz, W., McGeoch, M.A., Guralnick, R., Ferrier, S., Beck, J., Costello, M.J., Fernand, M., Geller, G.N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F.E., Pereira, H.M., Regan, E.C., Schmeller, D.S., Turak, E., 2019. Essential biodiversity variables for mapping and monitoring species populations. *Nat. Ecol. Evol.* 3, 539–551.
- Johnston, A., Fink, D., Hochachka, W.M., Kelling, S., 2018. Estimates of observer expertise improve species distributions from citizen science data. *Methods Ecol. Evol.* 9, 88–97.
- Johnston, A., Matechou, E., Dennis, E.B., 2023. Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods Ecol. Evol.* 14, 103–116.
- Kellner, K., 2019. jagsUI: A Wrapper Around “rjags” to Streamline “JAGS” Analyses: R package version 1.5.1. <https://CRAN.R-project.org>.
- Kendall, W.L., Peterjohn, B.G., Sauer, J.R., 1996. First-time observer effects in the North American breeding bird survey. *Auk* 113, 823–829.
- Kéry, M., 2004. Extinction rate estimates for plant populations in revisitation studies: importance of detectability. *Conserv. Biol.* 18, 570–574.
- Kéry, M., Plattner, M., 2007. Species richness estimation and determinants of species detectability in butterfly monitoring programmes. *Ecol. Entomol.* 32, 53–61.
- Kéry, M., Schaub, M., 2011. Bayesian Population Analysis Using WinBUGS. Academic Press, Waltham.
- Kéry, M., Schmidt, B.R., 2008. Imperfect detection and its consequences for monitoring for conservation. *Comm. Ecol.* 9, 207–216.
- Kéry, M., Dorazio, R.M., Soldaat, L., van Strien, A.J., Zuiderwijk, A., Royle, J.A., 2009. Trend estimation in populations with imperfect detection. *J. App. Ecol.* 46, 1163–1172.
- Lardner, B., Rodda, G.H., Adams, A.A.Y., Savidge, J.A., Reed, R.N., 2015. Detection rates of geckos in visual surveys: turning confounding variables into useful knowledge. *J. Herpetol.* 49, 522–532.
- Link, W.A., Sauer, J.R., 1998. Estimating population change from count data: application to the North American breeding bird survey. *Ecol. App.* 8, 258–268.
- Lotz, A., Allen, C.R., 2007. Observer bias in anuran call surveys. *J. Wildl. Manag.* 71, 675–679.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A., Langtimm, C.A., 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83, 2248–2255.
- MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E., 2016. Occupancy Estimation and Modeling. Elsevier, Amsterdam.
- McCarthy, M.A., Moore, J.L., Morris, W.K., Parris, K.M., Garrard, G.E., Vesk, P.A., Rumpff, L., Giljohann, K.M., Camac, J.S., Bau, S.S., Friend, T., Harrison, B., Yue, B., 2013. The influence of abundance on detectability. *Oikos* 122, 717–726.
- McClintock, B.T., Bailey, L.L., Pollock, K.H., Simons, T.R., 2010. Experimental investigation of observation error in anuran call surveys. *J. Wildl. Manag.* 74, 1882–1893.
- McDonald, T.L., 2003. Review of environmental monitoring methods: survey designs. *Environ. Monit. Assess.* 85, 277–292.
- Meier, C., Schelbert, B., 1999. Amphibienschutzkonzept Kanton Aargau. *Mitt. aarg. Naturf. Ges.* 35, 41–70.
- Miller, D.A.W., Bailey, L.L., Grant, E.H.C., McClintock, B.T., Weir, L.A., Simons, T.R., 2015. Performance of species occurrence estimators when basic assumptions are not met: a test using field data where true occupancy status is known. *Methods Ecol. Evol.* 6, 557–565.
- Molinari-Jobin, A., Kéry, M., Marboutin, E., Molinari, P., Koren, I., Fuxjäger, C., Breitenmoser-Würsten, C., Wolf, S., Fasel, M., Kos, I., Wolf, M., Breitenmoser, U., 2012. Monitoring in the presence of species misidentification: the case of the Eurasian lynx in the Alps. *Anim. Conserv.* 15, 266–273.
- Moor, H., Bühler, C., Bergamini, A., Vorburger, C., Holderegger, R., Schmidt, B.R., Egger, S., 2021. Amphibian observation and pond data (Aargau, Switzerland). *EnviDat*. <https://doi.org/10.16904/enviDat.270>. Deposited 23 December 2021.
- Moor, H., Bergamini, A., Vorburger, C., Holderegger, R., Bühler, C., Egger, S., Schmidt, B.R., 2022. Bending the curve: simple but massive conservation action leads to landscape-scale recovery of amphibians. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2123070119.
- Nichols, J.D., Williams, B.K., 2006. Monitoring for conservation. *Trends Ecol. Evol.* 21, 668–673.
- Nichols, J.D., Thomas, L., Conn, P.B., 2009. Inferences about landbird abundance from countdata: recent advances and future directions. In: Thomson, D.L., Cooch, E.G., Conroy, M.J. (Eds.), *Modeling Demographic Processes in Marked Populations*. Springer, Boston, pp. 201–235.
- Noss, R.F., 1990. Indicators for monitoring biodiversity: a hierarchical approach. *Cons. Biol.* 4, 355–364.
- Pellet, J., Schmidt, B.R., 2005. Monitoring distributions using call surveys: estimating site occupancy, detection probabilities and inferring absence. *Biol. Conserv.* 123, 27–35.
- Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 124, pp. 1–10.
- Pollock, K.H., Nichols, J.D., Simons, T.R., Farnsworth, G.L., Bailey, L.L., Sauer, J.R., 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13, 105–119.
- Powney, G.D., Carvell, C., Edwards, M., Morris, R.K.A., Roy, H.E., Woodcock, B.E., Isaac, N.J.B., 2019. Widespread losses of pollinating insects in Britain. *Nat. Comm.* 10, 1018. <https://doi.org/10.1038/s41467-019-08974-9>.
- Preston, F.W., 1979. The invisible birds. *Ecology* 60, 451–454.
- Roy, V., 2020. Convergence diagnostics for Markov Chain Monte Carlo. *Ann. Rev. Stat. App.* 7, 387–412.
- Royle, J.A., 2006. Site occupancy models with heterogeneous detection probabilities. *Biometrics* 62, 97–102.
- Royle, J.A., Dorazio, R.M., 2006. Hierarchical models of animal abundance and occurrence. *J. Agric. Biol. Env. Stat.* 11, 249–263.
- Royle, J.A., Kéry, M., 2007. A Bayesian state-space formulation of dynamic occupancy models. *Ecology* 88, 1813–1823.
- Royle, J.A., Link, W.A., 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.
- Sauer, J.R., Peterjohn, B.G., Link, W.A., 1994. Observer differences in the North American breeding bird survey. *Auk* 111, 50–62.
- Schmidt, B.R., 1993. Are hybridogenetic frogs cyclical parthenogens? *Trends Ecol. Evol.* 8, 271–273.
- Schmidt, B.R., 2005. Monitoring the distribution of pond-breeding amphibians when species are detected imperfectly. *Aquat. Conserv.* 15, 681–692.
- Sewell, D., Beebe, T.J.C., Griffiths, R.A., 2010. Optimising biodiversity assessments by volunteers: the application of occupancy modelling to large-scale amphibian surveys. *Biol. Conserv.* 143, 2102–2110.

- Siffert, O., Pellet, J., Ramseier, P., Tobler, U., Bergamini, A., Schmidt, B.R., 2022. Where land and water meet: making amphibian breeding sites attractive for amphibians. *Diversity* 14, 834.
- Stolen, E.D., Oddy, D.M., Gann, S.L., Holloway-Adkins, K.G., Legare, S.A., Weiss, S.K., Breininger, D.R., 2019. Accounting for heterogeneity in false-positive detection rate in southeastern beach mouse habitat occupancy models. *Ecosphere* 10, e02893.
- Sutherland, C., Elston, D.A., Lambin, X., 2013. Accounting for false positive detection error induced by transients. *Wildl. Res.* 40, 490–498.
- Szabo, J.K., Fuller, R.A., Possingham, H.P., 2012. A comparison of estimates of relative abundance from a weakly structured mass-participation bird atlas survey and a robustly designed monitoring scheme. *Ibis* 154, 468–479.
- Tanadini, L.G., Schmidt, B.R., 2011. Population size influences amphibian detection probability: implications for biodiversity monitoring programs. *PLoS ONE* 6, e28244.
- Tingley, M.W., Beissinger, S.R., 2009. Detecting range shifts from historical species occurrences: new perspectives on old data. *Trends Ecol. Evol.* 24, 625–633.
- Weir, L.A., Royle, J.A., Nanjappa, P., Jung, R.E., 2005. Modelling anuran detection and site occupancy on north american amphibian monitoring program (NAAMP) routes in Maryland. *J. Herpetol.* 39, 627–639.