Monitoring distributions using call surveys: estimating site occupancy, detection probabilities and inferring absence

Jérôme Pellet [a,*], Benedikt R. Schmidt [b,c,1]

[a] *Laboratory for Conservation Biology, Institute of Ecology, Department of Ecology and Evolution, Biology Building, University of Lausanne, CH-1015 Lausanne, Vaud, Switzerland*
[b] *Zoologisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[c] *KARCH, Naturhistorisches Museum, Bernastrasse 15, CH-3005 Bern, Switzerland*

Abstract

Monitoring programs serve to track changes in the distribution and abundance of species. A major problem with most monitoring programs is that species detection is imperfect and some populations are inevitably missed. Therefore, in most monitoring programs the true distribution of a species will be underestimated. Here, we report a field test of the reliability and performance of a commonly used method to monitor the distribution of amphibians (anuran call surveys). We surveyed the distribution of four anuran species in western Switzerland, and estimated detection probabilities to account for imperfect species detection and used these estimates to adjust our estimate of site occupancy (i.e., distribution). Next, we assessed how detection probabilities were affected by weather and how site occupancy was affected by site specific covariates. For one species (*Hyla arborea*), call surveys proved efficient in determining the regional distribution with only few site visits because detection probabilities were relatively high. The call surveys apparently missed many populations of another common species (*Bufo calamita*) because detection probabilities were lower. Two other species (*Bombina variegata* and *Alytes obstetricans*) were uncommon and strong inference from the analysis is not possible. Thus, multispecies surveys may be inefficient for rare species. Estimates of detection probabilities were used to calculate how many site visits are necessary to infer the absence of a species with some predetermined statistical certainty. The implications of "false absences" are important in ecology as they are known to bias usual habitat suitability models and overestimate extinction/colonization events in metapopulations. Large-scale monitoring programs would benefit from the application of an estimation-based approach to monitoring the distribution of species.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: Amphibians; Anuran call survey; Detection probability; Distribution; Site occupancy; Monitoring; *Bufo calamita*; *Hyla arborea*; *Alytes obstetricans*; *Bombina variegata*

1. Introduction

Monitoring programs increasingly are used to assess trends in species abundance, distribution, and biodiver-sity (Gibbs et al., 1998; Hintermann et al., 2000; Yoc-coz et al., 2001). Monitoring data are essential to identify key issues for policy and management goals, such as assessing priorities for conservation and land use, for environmental impact assessment, and for informing managers, policy-makers, and the general public about the state of nature (Stork and Samways, 1995). Two elements of a monitoring program are cru-cial to ensure that the goals can be achieved: clear specification of objectives and the collection of data

* Corresponding author. Tel.: +41 0 21 692 41 76; fax: +41 0 21 692 41 65.

E-mail addresses: jerome.pellet@unil.ch (J. Pellet), bschmidt@zool. unizh.ch, benedikt.schmidt@unine.ch (B.R. Schmidt).

[1] Tel.: +41 0 1 635 49 84 (Zürich)/+41 0 31 350 74 55 (Bern); fax: +41 0 1 635 68 21.

from which reliable inference can be made (Thompson et al., 1998; Yoccoz et al., 2001; Pollock et al., 2002; Karanth et al., 2003).

It is often very difficult, if not impossible, to detect all individuals, populations, or species during a monitoring program (Preston, 1979; Nichols and Conroy, 1996; Yoccoz et al., 2001; Pollock et al., 2002; Williams et al., 2002). Hence, inference from such data will bias estimates of population and community parameters, wildlife-habitat models and metapopulation models (Yoccoz et al., 2001; Pollock et al., 2002; Karanth et al., 2003; Schmidt, 2003; Tyre et al., 2003; Gu and Swihart, 2004; Kéry and Schmid, 2004). The problem can be stated simply using the formula

$$E(C) = Np, \tag{1}$$

where $E(C)$ is the expected value of a count (e.g., the number of ponds where a species was heard), N is the true value of the population parameter of interest (e.g., the number of sites occupied by a species) and p is a detection probability (Nichols, 1992). Because p is generally smaller than one and often variable, analysis of trends based upon these data is difficult at best (Yoccoz et al., 2001). In a trend analysis based on counts C, one assumes tacitly that detection probabilities p are constant. Although most monitoring uses standard methods, detection probabilities are never constant (MacKenzie and Kendall, 2002). A related problem is that one can detect the presence of individuals, populations, or species, but one can never be certain that individuals, populations, or species are truly absent rather than simply undetected (McArdle, 1990; Solow, 1993; Reed, 1996; MacKenzie et al., 2002; Kéry, 2002). The perfect monitoring program would identify sites where a species is present and those where it is truly absent. The estimation of detection probabilities can improve monitoring programs (Yoccoz et al., 2001; Pollock et al., 2002). Here, we address these issues using an example from an anuran call survey.

Amphibians are often considered reliable indicators of ecosystem health (Welsh and Droege, 2001). Moreover, a large number of amphibian species are declining or at risk of extinction because of human activity (Cooke, 1972; Blaustein and Wake, 1990; Alford and Richards, 1999; Houlahan et al., 2000; Collins and Storfer, 2003). This has lead to the creation of monitoring to determine the status of species and to improve conservation and management (e.g., Mossman et al., 1998; Dodd, 2003; Buckley and Beebee, 2004). Monitoring the distribution and abundance of anuran amphibians often involves registering the calling activity during the breeding season. Such call surveys are used to monitor both distribution and abundance (e.g., Lepage et al., 1997; Mossman et al., 1998;

Hemesath, 1998; Stevens et al., 2002; Buckley and Beebee, 2004). Although calling facilitates detection of anurans, the frequency and intensity of calls may be influenced by date and time of day, survey length, observer experience, and other factors (e.g., Shirose et al., 1997; Bridges and Dorcas, 2000; Crouch and Paton, 2002; Genet and Sargent, 2003). The typical protocol of a study to assess the reliability of an anuran call survey is to visit a site many times or for much longer time period than is typical for a standard anuran call survey. Subsequently, the time until all species present at a site are detected is measured. Such information can be used to estimate the optimal allocation of survey effort (monitoring of birds is often also based on vocalizations and the same problems apply, e.g., Bart and Schoultz, 1984; Thompson et al., 1998; Nebel and McCaffery, 2003). Ideally, surveys should aim for high detection probabilities at peak calling periods. However, some uncertainty in the proportion of species that have been detected, and in the timing of peak calling, inevitably will remain (MacKenzie et al., 2002).

Here, we present the results of a survey to determine the regional distribution of four anuran species. We estimated detection probabilities from repeated site visits and adjusted estimates of species' distributions accordingly (MacKenzie et al., 2002, 2003; Royle and Nichols, 2003; Tyre et al., 2003; Wintle et al., 2004). We then used the detection probabilities to determine the number of visits for adequate statistical power to reliably determine site absences. Taken together, these two elements of information are essential to design a monitoring program from which reliable inference is possible. Moreover, the method for inferring absence could be used in environmental impact assessment to shift the burden of proof. One may demand that a developer provides evidence that a species is absent from a site rather than providing evidence that a species is present.

The four anuran species we studied are prolonged breeders that differ strongly in their calls (Nöllert and Nöllert, 1992). Two species, the natterjack toad *Bufo calamita* and the tree frog *Hyla arborea*, have very loud calls that can be heard more than a kilometre away in good conditions. The calls of the other two species, the yellow-bellied toad *Bombina variegata* and the midwife toad *Alytes obstetricans*, are rather soft and have a low (*Bombina*) or high (*Alytes*) pitch. The four species often occur together in the same sites (Grossenbacher, 1988) and thus are suitable for comparative analysis. Here, we apply a rigorous statistical approach to estimate detection probabilities and site occupancy based on mark-recapture theory (MacKenzie et al., 2002) that proved to be very useful for the analysis of monitoring data.

2. Methods

2.1. Study area and field work

During spring, 2002, we surveyed 27 ponds in western Switzerland, between the villages of Allaman and Bière (Lat. N46°30′; Long. E6°25′). The study area covers a total of 396 km^2 (see Pellet and Neet (2001) and Pellet et al. (in press) for further information). Surveys were conducted for 15 min, beginning at sundown, on 32 nights from March 27 to June 18. Each site was visited on average 3.7 times (range 1–17) during the breeding season. Pond shores were walked systematically, and calling activity of all four species was recorded as present (1) or absent (0). Weather data were provided by Meteo-Swiss from the Changins weather station, 16 km from the centre of the study area. RAIN was computed as the sum of rainfall (in mm) during the day of the survey while TEMP was calculated as the mean temperature (in °C) from 3 measurements taken during the day of the survey (morning, noon and evening). During the days that preceded nights of fieldwork, average rainfall was 10.3 mm (median 0.05 mm, range 0–90.8 mm). Temperature during field work was on average 13.1 °C (median 12.8 °C, range 4.8–26.8 °C). RAIN and TEMP were weakly correlated ($R^2 = 0.093$, $P = 0.045$). Conditions during field work were not the same for the four species because of variation in breeding phenology (Table 1).

2.2. Data analysis

We used the mark-recapture-like approach of MacKenzie et al. (2002) as implemented in program PRESENCE (available for download from http://www.proteus.co.nz/) to estimate the proportion of sites occupied by each species. These models assume that a distribution is "closed" within a season, i.e., there are neither colonisations nor extinctions. Thus, for each species, we restricted analyses to the time span during which that species was active and calling in at least one site (MacKenzie et al., 2002). For *Hyla* and *Alytes*, all 32 survey nights were retained while only 28 and 14 nights were included in the analyses for *Bufo* and *Bombina*, respectively.

We defined a small set of a priori models we believed might explain site occupancy of the four species. First, we used the continuous covariates "distance to nearest

road" (DIST2ROAD). Previous analysis showed this variable was strongly correlated with tree frog presence and absence (Pellet et al., in press). Second, we used the categorical variable NATURAL, which describes whether a site was man-made (e.g., gravel pits), or natural or originally man-made, but left undisturbed for many years. Some species are almost completely restricted to man-made habitat (Grossenbacher, 1988). The third model assumed no effect of DIST2ROAD or NATURAL. Mean DIST2ROAD was 0.111 km (SE 0.070; $n = 14$) for natural habitats and 0.084 km (SE 0.054; $n = 13$) for gravel pits (two-sample t-test, $P = 0.277$).

To model detection probability, we developed three models. In the first model, we assumed that detection probability was constant (p). In the second and third, we assumed that detection probabilities were affected by temperature during the day (TEMP) or by the amount of rainfall (RAIN), respectively. Both temperature and rainfall affect anuran calling and hence detection probabilities (e.g., Blankenhorn, 1972). The three models for occurrence and detection were combined to yield a total of nine candidate models.

Since the data set was small, we kept the candidate models simple (Anderson and Burnham, 2002). The pattern of species distribution is probably best described by models with multiple covariates and some relationships may be non-linear (e.g., Austin, 2002; Knapp et al., 2003), but such complex models require larger data sets (i.e., more sites surveyed).

We used Akaike's Information Criterion (AIC) to rank models and to calculate Akaike weights (Burnham and Anderson, 2002). Akaike weights are equivalent to Bayesian posterior model probabilities and indicate the relative support of a model (Burnham and Anderson, 2002; Wintle et al., 2003). The sum of the Akaike weights across all candidate models is one. Akaike weights were then used to calculate model-averaged parameter estimates (proportion of sites occupied, detection probabilities) and confidence intervals. Model-averaged confidence intervals take model selection uncertainty into account (Burnham and Anderson, 2002; Wintle et al., 2003).

Estimates of detection probabilities can be used to estimate, with a specified degree of confidence, the number of visits necessary to assert that a species is truly absent from a site (McArdle, 1990; Reed, 1996; Kéry,

Table 1
Weather conditions during field work

Species	Temperature (°C)				Rainfall (mm)			
	Mean	Median	Minimum	Maximum	Mean	Median	Minimum	Maximum
Hyla arborea/Alytes obstetricans	13.2	12.9	4.8	26.8	10.4	0.1	0.0	90.8
Bufo calamita	13.8	13.9	5.7	26.8	11.9	0.5	0.0	90.8
Bombina variegata	16.6	16.0	12.0	26.8	18.8	2.8	0.0	90.8

Conditions were the same for *H. arborea* and *A. obstetricans*.

2002). Assuming that visits are comparable and independent, the probability F of not seeing a species after N visits with detection probability p is

$$F = (1 - p)^N.$$

If we want to be 95% confident that a species is absent, then $F = 0.05$. The above equation can be solved for the N_{min}, the minimum number of visits necessary to be 95% certain that a species is absent,

$$N_{min} = \frac{\log(0.05)}{\log(1 - p)}.$$

3. Results

The difference between the naïve proportion of ponds occupied (naïve in the sense that detection probability is assumed to be 1; i.e., the number of ponds where the species was detected), and the estimated proportion of ponds occupied varied considerably among species after accounting for imperfect detection (Table 2). Also, the models that explained the data best varied among species, but the estimated proportion of ponds occupied was consistent among models (Table 2).

3.1. Hyla arborea

This species was heard at 18 of 27 sites. Thus, the naïve proportion of sites used by this species is 18/27 = 0.67. During the period when *Hyla* was active, the sites were visited an average of 3.7 times (median 2, range 1–17). Tree frogs, if present, were detected an average of 3.3 times (median 2, range 1–15). Detection probability varied strongly with temperature (Fig. 1). The sum of Akaike weights of all three models including TEMP as a covariate for detection probability is 0.961. *Hyla* was detected more often on during warm nights (Fig. 1). In contrast, none of the factors thought to explain the proportion of ponds occupied was strongly supported by our data. All $\hat{\psi}$ (the estimated proportion of sites occupied) were within the range of less than 1 SE. The model averaged estimated proportion of ponds was slightly higher than the naïve proportion, but within the range of the standard error. With the small number of visits to each site, $\hat{\psi}$ tends to be positively biased (MacKenzie et al., 2002). Moreover, because we visited 27 ponds, naïve proportions can only take values which are a manifold of 1/27 (e.g., 0.666, 0.703, 0.740,...). Thus, some difference between the naïve and estimated proportion of sites occupied is expected. Therefore, we believe that we found all sites where *Hyla* breeds. This result is not surprising given the weather conditions under which field work was done and the high average detection probability ($\hat{p} = 0.743$).

Three visits at an average temperature of 13 °C were required to be 95% certain that a species was absent. However, since detection probability varied strongly with temperature (Fig. 1), the minimum number of visits declined with increasing temperature (Fig. 2).

3.2. Bufo calamita

This species was heard in 10 of 27 sites. Thus, the naïve proportion of sites used by this species is 0.37. During the period when *Bufo* was active, the sites were visited an average of 3.3 times (median 2, range 1–14) and *Bufo*, if present, detected an average of 2.0 times (median 1.5, range 1–5). Model selection was dominated by the effect of DIST2ROAD. Models including DIST2-ROAD had a joint Akaike weight of 0.628, indicating the species was more likely to occur further from roads. In contrast to tree frogs, there were no clear predictors of detection probabilities. All $\hat{\psi}$ were within the range of less than 1 SE. The model averaged estimated proportion of ponds occupied was much higher than the naïve proportion, suggesting that in many sites (23%) where the species occurs it was not detected. The estimated mean detection probability was 0.442. Thus, a minimum of 6 visits were required to be 95% certain the species was absent.

3.3. Bombina variegata

This species was heard in 3 of 27 sites. Thus, the naïve proportion of sites used by this species is 0.11. During the activity period of *Bombina*, the sites were visited an average of 1.2 times (median 1, range 1–4); the activity period of this species started in May whereas all other species were already active in late March or April, thus reducing the dataset to only 14 survey nights. *Bombina*, if present, was detected an average of 1.7 times (median 1, range 1–3). The model selection yielded no clear pattern because the models did not converge to good parameter estimates and the variance-covariance matrix could not be calculated in PRESENCE. We present the results of the PRESENCE analysis in Table 2, but refrain from an interpretation. For example, some estimates of detection probabilities were 0, which suggests that the entire analysis is probably of questionable value.

3.4. Alytes obstetricans

This species was heard in 3 of 27 sites. Thus, the naïve proportion of sites used by this species is 0.11. During the activity period of *Alytes*, the sites were visited an average of 3.7 time (median 2, range 1–17) and *Alytes*, if present, detected an average of 4.33 times (median 6, range 1–6). Model selection suggests that *Alytes* occurs more often in anthropogenic habitats (gravel pits and

Table 2
Model selection and parameter estimation of site occupancy

Model	ΔAIC	w	$\hat{\psi}$	SE(ψ)	Difference		$\hat{p}$
					Absolute	Proportional	
Hyla arborea (naïve ψ = 0.667)							
ψ(.)p(TEMP)	0.000	0.476	0.730	0.100	0.06	−0.09	0.742
ψ(NATURAL)p(TEMP)	0.913	0.302	0.717	0.097	0.05	−0.07	0.745
ψ(DIST2ROAD)p(TEMP)	1.913	0.183	0.732	0.100	0.06	−0.09	0.741
ψ(.)p(.)	7.181	0.013	0.740	0.101	0.07	−0.10	0.713
ψ(NATURAL)p(.)	7.940	0.009	0.729	0.098	0.06	−0.09	0.715
ψ(.)p(RAIN)	8.913	0.006	0.740	0.102	0.07	−0.10	0.716
ψ(DIST2ROAD)p(.)	9.088	0.005	0.742	0.102	0.08	−0.10	0.712
ψ(NATURAL)p(RAIN)	9.679	0.004	0.729	0.099	0.06	−0.08	0.719
ψ(DIST2ROAD)p(RAIN)	10.811	0.002	0.742	0.102	0.08	−0.10	0.716
Model averaged			*0.726*	*0.099*			*0.742*
Bufo calamita (naïve ψ = 0.370)							
ψ(DIST2ROAD)p(.)	0.000	0.224	0.482	0.116	0.11	−0.23	0.439
ψ(DIST2ROAD)p(TEMP)	0.078	0.215	0.478	0.114	0.11	−0.23	0.464
ψ(DIST2ROAD)p(RAIN)	0.340	0.189	0.484	0.115	0.11	−0.24	0.425
ψ(.)p(.)	1.768	0.093	0.501	0.133	0.13	−0.26	0.425
ψ(.)p(RAIN)	1.931	0.085	0.501	0.129	0.13	−0.26	0.413
ψ(.)p(TEMP)	1.951	0.084	0.491	0.128	0.12	−0.25	0.452
ψ(NATURAL)p(.)	3.474	0.039	0.516	0.141	0.15	−0.28	0.419
ψ(NATURAL)p(TEMP)	3.684	0.035	0.503	0.135	0.13	−0.26	0.447
ψ(NATURAL)p(RAIN)	3.722	0.035	0.511	0.135	0.14	−0.28	0.409
Model averaged			*0.489*	*0.122*			*0.438*
Bombina variegata (naïve ψ = 0.111)							
ψ(DIST2ROAD)p(TEMP)	0.000	0.146	0.248	0.134	0.14	−0.55	0.000
ψ(NATURAL)p(TEMP)	0.041	0.143	0.265	0.159	0.15	−0.58	0.000
ψ(DIST2ROAD)p(.)	0.249	0.129	0.241	0.130	0.13	−0.54	0.421
ψ(NATURAL)p(.)	0.284	0.127	0.263	0.171	0.15	−0.58	0.384
ψ(NATURAL)p(RAIN)	0.386	0.121	0.234	0.122	0.12	−0.53	0.001
ψ(DIST2ROAD)p(RAIN)	0.406	0.119	0.224	0.109	0.11	−0.50	0.840
ψ(.)p(TEMP)	1.134	0.083	0.335	0.220	0.22	−0.67	0.000
ψ(.)p(.)	1.449	0.071	0.320	0.217	0.21	−0.65	0.382
ψ(.)p(RAIN)	1.747	0.061	0.264	0.143	0.15	−0.58	0.943
Model averaged			*0.260*	*0.154*			*0.288*
Alytes obstetricans (naïve ψ = 0.111)							
ψ(NATURAL)p(TEMP)	0.000	0.281	0.130	0.067	0.02	−0.14	0.574
ψ(NATURAL)p(RAIN)	0.744	0.194	0.132	0.068	0.02	−0.16	0.566
ψ(NATURAL)p(.)	0.950	0.175	0.127	0.066	0.02	−0.13	0.569
ψ(.)p(TEMP)	1.790	0.115	0.142	0.078	0.03	−0.22	0.566
ψ(.)p(RAIN)	2.752	0.071	0.139	0.077	0.03	−0.20	0.561
ψ(.)p(.)	2.769	0.070	0.138	0.076	0.03	−0.20	0.562
ψ(DIST2ROAD)p(TEMP)	3.789	0.042	0.142	0.079	0.03	−0.22	0.566
ψ(DIST2ROAD)p(RAIN)	4.747	0.026	0.139	0.077	0.03	−0.20	0.561
ψ(DIST2ROAD)p(.)	4.769	0.026	0.138	0.077	0.03	−0.20	0.562
Model averaged			*0.133*	*0.071*			*0.568*

ΔAIC is the difference between the model with the lowest AIC and the given model, w is the Akaike weight, ψ is the estimated proportion of sites occupied and the SE(ψ) the standard error thereof, absolute difference is ($\hat{\psi}$ − naïve ψ) and relative difference (1 − (naïve$\psi/\hat{\psi}$)), and $\hat{p}$ is the estimated detection probability.

alike), but this results is based on only three sites where the species is present.

There was no clear pattern which covariates would best predict detection probabilities. Again, we cannot make any conclusive statement regarding the importance of the covariates. The $\hat{\psi}$ were within the narrowest range of all four species. The model averaged estimated proportion of ponds is slightly higher than the naïve proportion, but within the range of the standard error. As with *Hyla*, we believe that we found all the sites where *Alytes* occurs. Given the conditions under which field work was done, the estimated mean detection probability was 0.570. Thus, a minimum of four visits were required to be 95% certain the species was absent.

Most information on detection probabilities is obtained from sites where the species is detected at least
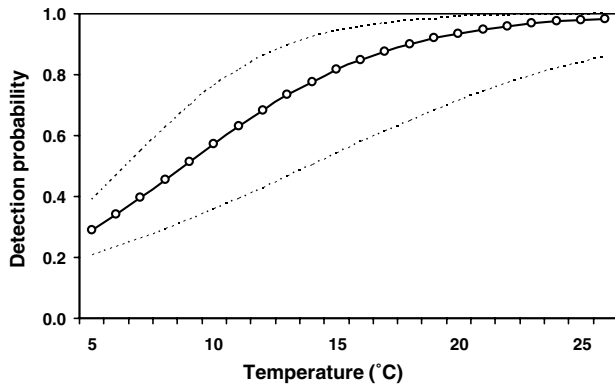
Fig. 1. The relationship between temperature and detection probability for the treefrog (±1 SE), *H. arborea*. Estimation of the detection probability is based on model (ψ(.), p(TEMP)) which has an Akaike weight $w = 0.645$.
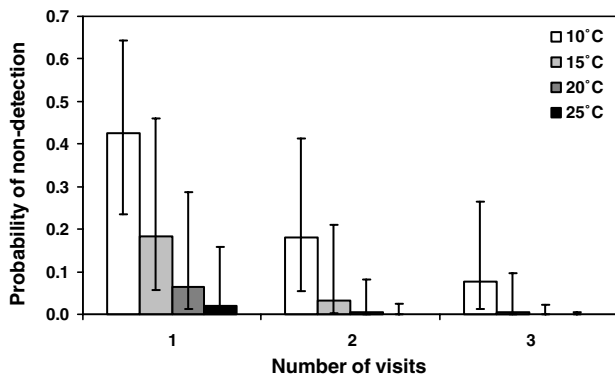

Fig. 2. The relationship between temperature during the survey and the probability that treefrogs, *H. arborea*, have not yet been detected during the current and previous visits to a site. The probabilities of non-detection were calculated using the relationship between temperature and detection probability shown in Fig. 1.

once (B.A. Wintle, personal communication). Because *Alytes* was detected at only three sites, a careful interpretation of detection probabilities and the minimum number of visits to infer absence is warranted.

4. Discussion

4.1. The utility of anuran call surveys

The probability of detection varied strongly among species. The good news is that it is possible to detect all the sites where a species is present using call surveys with relatively few visits (e.g., *Hyla*, *Alytes*). The bad news is that in the same survey many populations of other species were overlooked (*Bufo*, *Bombina*). The statistical analysis revealed that many sites with *Bufo* were apparently missed. This may not be surprising, since six visits were necessary to infer absence of this species with 95% certainty, and sites were visited an average of only

3.3 times. Thus, even when using standardized methods, simple unadjusted counts alone cannot reliably be used to determine species distributions, but estimates of detection probabilities are needed. While we could model the covariates of site occupancy for three species (*Hyla*, *Bufo*, *Alytes*), the analysis is only reliable for two species (*Hyla* and *Bufo*). *Alytes* was detected at only three sites and more sites are needed to model the relationship between the distribution and habitat variables reliably (Harrell, 2001).

We expected that species detection would be correlated with call volume. This was not the case. Instead, overall calling behaviour, not the call per se, was most important. Both *Alytes* and *Hyla* call relatively constantly during the breeding season whereas *Bufo* calls infrequently. This is most obvious at the most frequently visited site, where both *Hyla* and *Bufo* occur. This site was visited 13 times during the breeding season of *Bufo*. *Hyla* was heard calling 13 times while *Bufo* was heard only five times. Large *Bufo* choruses only occur after heavy rains have filled temporary pools. Nonetheless, candidate models that included rain were not well supported by the data, probably because of the intrinsic temporal structure of breeding activity in this species (Sinsch, 1988). Additionally, only a subset of the males call during a night (Arak, 1988) and the duration of calling activity in a given night may vary from 20–500 min (Blankenhorn, 1972). Knowledge of the factors that affect detection probabilities can help to design efficient surveys. For example, the relationship between temperature and detection probability displayed (Fig. 1), may be used to determine ideal conditions for anuran call surveys. However, the explanatory power of covariates varied among species. Temperature clearly predicted detection probability of tree frogs (Fig. 1) whereas no environmental variable significantly predicted detection probability for other species. This implies that optimal survey conditions may vary among species.

4.2. Estimating the proportion of sites occupied by a species and inferring absence

We used the models of MacKenzie et al. (2002) to estimate site occupancy and reliably infer species absence. For three species (*Hyla*, *Bufo*, and *Alytes*) the analysis worked well. For the fourth species, *Bombina*, the analysis was not satisfactory, because of sample size limitations. This shows that when multiple species are surveyed in parallel, reliable inference may only be possible for the commoner species which are present at a sufficiently large number of sites. Rare species may require surveys that are designed specifically for these species.

With relatively few visits to the sites, standard errors were quite narrow and in the range expected from the simulation study of MacKenzie et al. (2002). Coefficients

of variation depended strongly on $\hat{\psi}$: CV was lowest for *Hyla* (13%), intermediate for *Bufo* (25%), and highest for *Alytes* (54%). The simulations of MacKenzie et al. (2002) lead us to expect that adding more visits to sites would not reduce standard errors (and hence, CV).

The statistical methods we used assume that there are no false positives, i.e., species should not be recorded as present when they are in fact absent. Given our experience with the species, we believe that this assumption was met in our analyses. However, Genet and Sargent (2003) found that volunteers falsely recorded species at sites where they were not present. The mark-recapture approach that we have used can deal with false negatives (i.e., species present but not detected) but not with false positives. Future site occupancy models should allow for false positives.

Because detection probabilities were relatively high (Table 2), it is possible to infer species absence with relatively few site visits. If detection probabilities are much lower, then financial constraints on the monitoring program may preclude a sufficient number of site visits to infer species absence. However, if a development project threatens a site where an endangered species occurs, it would still be possible to reverse the burden of proof and demand that the development company provide evidence that a species is absent from a particular site with a statistical certainty (Dayton, 1998; Kéry, 2002).

The detection probabilities we estimated are based on a single season of field work. Before they are used to design a monitoring program, variation in detection probabilities among years should be assessed. MacKenzie et al. (2003) found that detection probabilities of the salamander *Ambystoma tigrinum* varied strongly between two years ($\hat{p} = 0.27$ and 0.65, respectively). Consequently, the number of visits necessary to be 95% certain that *Ambystoma tigrinum* was absent from a site was 10 and 3, respectively.

4.3. The importance of an estimation-based approach to monitoring

With multiple visits to the ponds, it was possible to detect all populations (or rather choruses) of two out of four species, namely *Hyla* and *Alytes*. If conditions for field work are selected appropriately (e.g., temperature >20 °C), the entire regional distribution of *Hyla* might be determined with >95% confidence using a single visit to each site (Fig. 2) without a need for statistical analysis. Nevertheless, we believe that an estimation-based approach, such as MacKenzie et al. (2002) models that we used, is superior to other approaches, for several reasons. First, the same survey may yield good data for some species whereas populations of other species remain undetected (*Bufo*, *Bombina*; see also Heimbucher, 1986). Here, it is important to estimate detection probabilities and adjust the counts accordingly. Second, an estimation-based approach also gives a standard error or confidence interval. This allows gauging of how much confidence is warranted in the results of the survey. Third, the estimate and confidence interval of site occupancy are independent of the method used. Thus, in a long-term monitoring program, different methods could be used in different years. It is also possible to use slightly different protocols (e.g., time spent at a site) or even different methods at different sites and incorporate such differences as covariates in the analysis. Whereas the counts are likely to differ, the estimates should not.

A statistical approach to the analysis of monitoring and survey data is important. However, it cannot replace a thorough knowledge of the natural history, calling behaviour, and phenology of the species (e.g., Shirose et al., 1997; Bridges and Dorcas, 2000; Crouch and Paton, 2002). Such knowledge and statistical analysis should go hand-in-hand as they can strengthen each other.

4.4. Implications for large-scale monitoring programs

Large-scale monitoring programs are often expensive. They should provide good quality data and allow the inference they were designed for (Yoccoz et al., 2001; Pollock et al., 2002; Kéry and Schmid, 2004). Estimating population sizes at multiple locations using mark-recapture methods is often prohibitively expensive (but see Royle, 2004; Dodd and Dorazio, 2004). Monitoring presence-absence of species is easier and cheaper and a commonly used approach to monitoring (Thompson et al., 1998; Campbell et al., 2002). If costs of multiple visits are high, then detection probabilities may only be estimated at a subset of the sites (Pollock et al., 2002). However, during most monitoring programs and surveys, sites are often visited multiple times during a season anyway, so basing inference on estimation methods (e.g., MacKenzie et al., 2002; Royle and Nichols, 2003; Tyre et al., 2003) comes at no great cost. Estimation approaches also provide a means of involving volunteers (Lepage et al., 1997; Mossman et al., 1998; Genet and Sargent, 2003; Kéry and Schmid, 2004) without suffering substantial bias in inference.

References

Alford, R.A., Richards, S.J., 1999. Global amphibian declines: a problem in applied ecology. Annual Reviews of Ecology and Systematics 30, 133–165.

Anderson, D.R., Burnham, K.P., 2002. Avoiding pitfalls when using information-theoretic methods. Journal of Wildlife Management 66, 912–918.

Arak, A., 1988. Callers and satellites in the natterjack toad: evolutionary stable decision rules. Animal Behaviour 36, 416–432.

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling 157, 101–118.

Bart, J., Schoultz, J.D., 1984. Reliability of singing bird surveys: changes in observer efficiency with avian density. AUK 101, 307–318.

Blankenhorn, H.-J., 1972. Meteorogical variables affecting onset and duration of calling in *Hyla arborea* and *Bufo calamita* Laur. Oecologia 9, 223–234.

Blaustein, A.R., Wake, D.B., 1990. Declining amphibian populations – a global phenomenon. Trends in Ecology and Evolution 5, 203–204.

Bridges, A.S., Dorcas, M.E., 2000. Temporal variation in anuran calling behaviour: implications for surveys and monitoring programs. COPEIA 2000, 587–592.

Buckley, J., Beebee, T.J.C., 2004. Monitoring the conservation status of an endangered amphibian: the natterjack toad *Bufo calamita* in Britain. Animal Conservation 7, 221–228.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multi-model Inference. Springer, Berlin.

Campbell, S.P., Clark, J.A., Crampton, L.H., Guerry, A.D., Hatch, L.T., Hosseini, P.R., Lawler, J.J., O'Connor, R.J., 2002. As assessment of monitoring efforts in endangered species recovery plans. Ecological Applications 12, 674–681.

Collins, J.P., Storfer, A., 2003. Global amphibian declines: sorting the hypotheses. Diversity and Distributions 9, 89–98.

Cooke, A.S., 1972. Indications of recent change in status in the British Isles of the frog (*Rana temporaria*) and the toad (*Bufo bufo*). Journal of Zoology 167, 161–178.

Crouch, W.B., Paton, P.W.C., 2002. Assessing the use of call surveys to monitor breeding anurans in Rhode Island. Journal of Herpetology 36, 185–192.

Dayton, P.K., 1998. Reversal of the burden of proof in fisheries management. Science 279, 821–828.

Dodd, C.K., 2003. Monitoring amphibians in Great Smoky Mountains National Park. US Geological Survey Circular 1258, Tallahassee.

Dodd, C.K., Dorazio, R.M., 2004. Using point counts to simultaneously estimate abundance and detection probabilities in a salamander community. Herpetologica 60, 68–78.

Genet, K.S., Sargent, L.G., 2003. Evaluation of methods and data quality from a volunteer-based amphibian call survey. Wildlife Society Bulletin 31, 703–714.

Gibbs, J.P., Droege, S., Eagle, P., 1998. Monitoring populations of plants and animals. BioScience 48, 935–940.

Grossenbacher, K., 1988. Verbreitungsatlas der Amphibien der Schweiz. Documenta faunistica helvetiae 7, 1–207.

Gu, W., Swihart, R.K., 2004. Absent or undetected? Effects of non-detection of species occurrences on wildlife-habitat models. Biological Conservation 116, 195–203.

Harrell, F.E., 2001. Regression Modelling Strategies: with Application to Linear Models, Logistic Regression, and Survival Analysis. Springer, New York.

Heimbucher, D., 1986. Gelbbauchunke (*Bombina variegata*) und Feuersalamander (*Salamandra salamandra*) – Problemarten bei Amphibienkartierungen. Schriftenreihe des Bayerischen Landesamts für Umweltschutz 73, 185–190.

Hemesath, L.M., 1998. Iowa's frog and toad survey, 1991–1994. In: Lannoo, M.J. (Ed.), Status and Conservation of Midwestern Amphibians. University of Iowa Press, Iowa City, pp. 206–216.

Hintermann, U., Weber, D., Zangger, A., 2000. Biodiversity monitoring in Switzerland. Schriftenreihe Landschaftspflege und Naturschutz 62, 47–58.

Houlahan, J.E., Findlay, C.S., Schmidt, B.R., Meyer, A.H., Kuzmin, S.L., 2000. Quantitative evidence for global amphibian population declines. Nature 404, 752–755.

Karanth, K.U., Nichols, J.D., Seidensticker, J., Dinerstein, E., Smith, J.L.D., McDougal, C., Johnsingh, A.J.T., Chundawat, R.S., Thapar, V., 2003. Science deficiency in conservation practice: the monitoring of tiger populations in India. Animal Conservation 6, 141–146.

Kéry, M., 2002. Inferring the absence of a species – a case study of snakes. Journal of Wildlife Management 66, 330–338.

Kéry, M., Schmid, H., 2004. Monitoring programs need to take into account imperfect species detectability. Basic and Applied Ecology 5, 65–73.

Knapp, R.A., Matthews, K.R., Preisler, H.K., Jellison, R., 2003. Developing probabilistic models to predict amphibian site occupancy in a patchy landscape. Ecological Applications 13, 1069–1082.

Lepage, M., Courtois, R., Daigle, C., Matte, S., 1997. Surveying calling anurans in Québec using volunteers. In: Green, D.M. (Ed.), Amphibians in Decline: Canadian Studies of a Global Problem, Herpetological Conservation, 1, pp. 128–140.

McArdle, B.H., 1990. When are rare species not there?. OIKOS 57, 276–277.

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

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

Mossman, M.J., Hartman, L.M., Hay, R., Sauer, J.R., Dhuey, B.J., 1998. Monitoring long-term trends in Wisconsin frog and toad populations. In: Lannoo, M.J. (Ed.), Status and Conservation of Midwestern Amphibians. University of Iowa Press, Iowa City, pp. 169–198.

Nebel, S., McCaffery, B.J., 2003. Vocalization activity of breeding shorebirds: documentation of its seasonal decline and applications for breeding bird surveys. Canadian Journal of Zoology 81, 1702–1708.

Nichols, J.D., 1992. Capture–recapture models: using marked animals to study population dynamics. BioScience 42, 94–102.

Nichols, J.D., Conroy, M.J., 1996. Techniques for estimating abundance and species richness. In: Wilson, D.E., Cole, F.R., Nichols, J.D., Rudran, R., Foster, M.S. (Eds.), Measuring and Monitoring Biological Diversity: Standard Methods for Mammals. Smithsonian Institution Press, Washington, pp. 177–234.

Nöllert, A., Nöllert, C., 1992. Die Amphibien Europas: Bestimmung, Gefährdung, Schutz. Franck-Kosmos, Stuttgart.

Pellet, J., Neet, C., 2001. La Rainette verte (*Hyla arborea*; Anura) dans le canton de Vaud: un état des lieux. Bulletin de la Société vaudoise des Sciences Naturelles 87, 287–303.

Pellet, J., Guisan, A., Perrin, N., 2004. A concentric analysis of the impact of urbanization on the threatened European tree frog in an agricultural landscape. Conservation Biology, in press.

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.

Preston, F.W., 1979. The invisible birds. Ecology 60, 451–454.

Reed, J.M., 1996. Using statistical probability to increase confidence of inferring species extinction. Conservation Biology 10, 1283–1285.

Royle, J.A., 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60, 108–115.

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

Schmidt, B.R., 2003. Count data, detection probabilities, and the demography, dynamics, distribution, and decline of amphibians. Comptes Rendues Biologies 326, S119–S124.

Shirose, L.J., Bishop, A., Green, D.M., MacDonald, C.J., Brooks, R.J., Helferty, N.J., 1997. Validation tests of an amphibian call count survey technique in Ontario, Canada. Herpetologica 53, 312–320.

Sinsch, U., 1988. Temporal spacing of breeding activity in the natterjack toad, *Bufo calamita*. Oecologia 76, 399–407.

Solow, A.R., 1993. Inferring extinction from sighting data. Ecology 74, 962–964.

Stevens, C.E., Diamond, A.W., Gabor, T.S., 2002. Anuran call survey on small wetlands in Prince Edward Island, Canada restored by dredging of sediments. Wetlands 22, 90–99.

Stork, N.E., Samways, M.J., 1995. Inventorying and monitoring. In: Heywood, V.H., Watson, R.T. (Eds.), Global Biodiversity Assessment. Cambridge University Press, Cambridge, pp. 453–543.

Thompson, W.L., White, G.C., Gowan, C., 1998. Monitoring Vertebrate Populations. Academic Press, San Diego, California, USA.

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K., Possingham, H.P., 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecological Applications 13, 1790–1801.

Welsh Jr., H.H., Droege, S., 2001. A case for using plethodontid salamanders for monitoring biodiversity and ecosystem integrity of North American forests. Conservation Biology 15, 558–569.

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

Wintle, B.A., McCarthy, M.A., Volinsky, C.T., Kavanagh, R.P., 2003. The use of Bayesian model averaging to better represent uncertainty in ecological models. Conservation Biology 17, 1579–1590.

Wintle, B.A., McCarthy, M.A., Parris, K.M., Burgman, M.A., 2004. Precision and bias of methods for estimating point survey detection probabilities. Ecological Applications 14, 703–712.

Yoccoz, N.G., Nichols, J.D., Boulinier, T., 2001. Monitoring of biological diversity in space and time. Trends in Ecology and Evolution 16, 446–453.