

Effect of sampling effort and species detectability on volunteer based anuran monitoring programs

Shane R. de Solla^a, Leonard J. Shirose^b, Kim J. Fernie^{a,*}, Glenn C. Barrett^a,
Chris S. Brousseau^b, Christine A. Bishop^c

^a Canadian Wildlife Service, Canada Centre for Inland Waters, P.O. Box 5050, Burlington, ON, Canada L8P 2A6

^b Ontario Ministry of Natural Resources, Peterborough, ON, Canada

^c Canadian Wildlife Service, 5421 Robertson Road, Delta, BC, Canada V4K 2N2

Received 7 January 2004; received in revised form 1 June 2004; accepted 7 June 2004

Abstract

A central question to the usefulness of surveys for estimating population trends is whether the surveys adequately represent the biological communities that are being monitored. There are two approaches to address this issue; the first is to statistically adjust site occupancy using species detection probabilities, and the second is to determine the minimum sampling effort required to adequately represent the communities. We focused on the latter approach, using data from two anuran monitoring programs, the Ontario Backyard Frog survey (1992–2001) and the Ontario Bait Frog survey (2001–2003). We determined the minimum sampling effort required to adequately represent the anuran community, and demonstrated the pitfalls of examining population trends when detection probabilities are less than one and vary either temporally or spatially. We found that approximately 12 and 24 randomly sampled nights were required to detect 80% and 90% of species richness, respectively. Detection probabilities varied among species, and were highest for spring peepers (*Pseudacris crucifer*) and green frogs (*Rana clamitans*), and lowest for wood frogs (*Rana sylvatica*) and northern leopard frogs (*Rana pipiens*). Variation in detection probabilities is likely associated with species specific differences in calling strategies and breeding periods. Anuran monitoring programs that use three stratified sampling periods require site occupancies to be adjusted using detection probabilities in order to adequately represent anuran communities. Failing to account for differences in detection probabilities may result in false significant differences in site occupancy rates. The issues raised here are not limited to monitoring anurans, but are suitable for surveys of many taxa.

Crown Copyright © 2004 Published by Elsevier Ltd. All rights reserved.

Keywords: Detection probability; Call surveys; Monitoring programs; Presence–absence

1. Introduction

Reports of amphibian declines (e.g. Pechman et al., 1991; Wake, 1991) have persuaded both Canadian and American government agencies to implement programs to monitor amphibian populations. In recent years, geographically extensive anuran surveys have been used to determine if amphibian populations are declining (Bishop et al., 1997; Lepage et al., 1997). Many of these

surveys are based primarily or solely upon volunteers to collect the data, and involve monitoring the calls of adult male anurans. Although much of the focus has centred on the use of these surveys for site occupancy, calling codes (an index of population size) and counts of calling males (a direct measure of population size) are generally also collected. Currently, many volunteer-based acoustic anuran surveys are active in Canada and the United States, most of which are funded, run by, or modeled after the North American Amphibian Monitoring Program (NAAMP), Environment Canada, and Frogwatch USA. Call surveys are also used in non-volunteer programs

* Corresponding author. Fax: +1-905-336-6434.

E-mail address: Kim.Fernie@ec.gc.ca (K.J. Fernie).

(e.g., Amphibian Research and Monitoring Initiative [ARMI]), as well as university funded projects.

Is the sampling effort of anuran call surveys sufficient to represent the anuran community? This question is central to the utility of acoustic surveys for monitoring population trends; unfortunately, instances of insufficient or inappropriate sampling effort have been well documented. Generally, monitoring programs are designed with the intent of sampling during peak calling times for anurans; nevertheless, success has been mixed. Bridges and Dorcas (2000) found that the typical timing of anuran surveys in South Carolina occurred when southern leopard frogs (*Rana sphenoccephala*) had a small probability of calling. Crouch and Paton (2002) found that not all anuran species were detected using acoustic surveys, and that four sampling periods were necessary to monitor the seven species that were detected. Many acoustic survey programs have three sampling periods, not four, although the NAAMP protocol includes three seasonal, but also allows for an optional earlier sampling period to target wood frogs. Paszkowski et al. (2002) found that sampling effort of acoustic surveys was too low when they evaluated different survey techniques, as they sampled after the peak calling times for some species. Detection probabilities for many species can be low, which may underestimate site occupancy (Storfer, 2003). However, probability detection can be incorporated into models estimating temporal changes in anuran occupancy rates (MacKenzie et al., 2002). Although the problem of call detection is becoming well recognized (Bailey et al., 2002; Genet and Sargent, 2003; Weir et al., 2003), it is doubtful that monitoring programs have been designed with respect to this issue.

MacKenzie et al. (2002) described an approach to compare temporal changes in site occupancy when the probability of detection is as low as 30%. Currently, at least two amphibian monitoring programs are using this approach to examine differences in amphibian occupancy in light of suboptimal detection probabilities (Genet and Sargent, 2003; Weir et al., 2003). While we have no objections with statistically adjusting the data to account for detection probabilities of less than one, knowledge of detection rates does not aid in determining whether or not a species is present at any particular sites. Surveys can only accurately determine species presence or absence by assuring that the sampling effort is sufficiently high so that nondetection ceases to be a major issue.

We evaluated the sampling effort of two acoustic anuran monitoring programs: the Ontario Backyard Frog survey (1992–2001) by Environment Canada, and the Bait Frog (northern leopard frog; *Rana pipiens*) survey, 2001–2003 by the Ontario Ministry of Natural Resources (OMNR). Previous work has shown that volunteer observers generally correctly identify species by their call (Genet and Sargent, 2003; Shirose et al., 1997). Call in-

tensity and count data are generally unsuitable for monitoring small areas or routes or for rarer species, because the combination of small sample size with large sampling error would reduce the precision of estimates of relative abundance (Shirose et al., *In review*), and so reduce power to unsuitable levels. Volunteers also had difficulty in estimating calling intensity (Genet and Sargent, 2003). Thus, we focused only on species richness and species detectability, specifically, we determined the sampling effort required for the Ontario Backyard Frog survey to adequately estimate species richness, and the relative detectability of different species. Lastly, we demonstrate using the Bait Frog survey of the OMNR how differences in detectability may give false significant differences in site occupancy among years.

2. Methods

2.1. Survey methods

The Ontario Backyard Frog survey is an atypical anuran monitoring program; volunteers surveyed only one location that they selected, usually near their residence, which was potentially sampled daily throughout the spring and summer months. All volunteers were familiarized with anuran calls through the use of audiotapes. Volunteers listened for three minutes, usually just after sunset, and recorded the species calling, the calling code (see below), the number of individuals calling, as well as precipitation and wind speed (Beaufort Wind scale). The calling code was rated from zero to three; code zero was recorded when no calling was detected, code one was recorded when each frog could be counted separately, and when the frog calls did not start at the same time as a conspecific; code two was recorded when few males were calling simultaneously, but calling was not continuous and individual males could still be counted; code three was recorded when calls overlapped and were continuous, and individuals could not be counted. For our analysis, we selected 37 locations in which the volunteers started sampling by early April, and continued throughout the spring and early summer, and had a large sampling effort (mean number of sampling nights per volunteer = 78.4 nights). We assumed that each volunteer had a detection probability of one for the entire survey.

The OMNR Bait Frog survey was designed primarily to monitor northern leopard frogs to manage their use as bait for fishing. Trained observers familiar with anuran calls conducted the surveys over three years. Sites were randomly selected from habitat suitable for northern leopard frogs. The eastern portion of Ontario was divided into six regions (four in eastern Ontario, two in central Ontario; for more details, see Shirose, 2001).

Two surveys were performed: nocturnal acoustic surveys, and daytime acoustic and visual surveys. Nocturnal call surveys were generally performed once per season, on warm, damp evenings, with wind less than 20–30 km/h, between half an hour after sunset and midnight. After a one minute quiet period, observers listened for five minutes, and temperature, precipitation, wind speed, and the presence or absence of frog calls were noted. If northern leopard frogs were not heard, a second acoustic survey was performed during the day, followed by a visual survey. Approximately 100 m of shoreline on either side of the site was visually surveyed, first in the water followed by a land search. A site was considered occupied if there was chorusing activity, or if mature individuals or eggs were found. The presence of immature individuals only was not considered sufficient indication of an established population or colonization event. The timing of the surveys was tailored specifically for monitoring northern leopard frogs, and were conducted April 20 through May 9, 2001; April 13 through May 4, 2002; and April 20 through May 6, 2003. We assumed that the combined visual and acoustic surveys had a detection probability of one for northern leopard frogs.

2.2. Sampling effort

Unfortunately, it is impossible to directly determine if the sampling effort by volunteers in the Backyard Survey was sufficient to adequately represent the species richness of the anuran community. However, resampling techniques can be used to generate rarefaction or accumulation curves (Gotelli and Colwell, 2001), and may indicate if the sampling effort is sufficient to represent the species richness of the anuran community. Resampling was used to estimate the relationship between the number of species heard and the number of backyard surveys taken from the original sample (see next section). Due to the relative rarity of some species, only American toads (*Bufo americanus*), chorus frogs (*Pseudacris triseriata* and *Pseudacris maculata*), spring peepers (*Pseudacris crucifer*), gray treefrogs (*Hyla versicolor*), wood frogs (*Rana sylvatica*), northern leopard frogs, green frogs (*Rana clamitans*), and American bullfrogs (*Rana catesbeiana*) were included in the analysis.

2.3. Species richness

Data from the Backyard Frog survey were resampled to determine if the sampling events were sufficient to obtain a good estimate of the species richness, and to determine if there were appreciable differences among locations. A species presence–absence matrix was made, in which the columns represented each species, and the rows were defined as sampling nights. A value of one

for presence or a value of zero for absence was given to each cell. Backyard Frog survey locations were resampled with a sample size of 1, 2, 3, 5, 10, 15, 20, 30, 40 and 50 sampling nights. For each sample size (1, 2, 3, 5, ..., 50), sampling events were randomly selected 1000 times each (with replacement) from the total number of sampling events. The number of species heard was calculated for each resample, and then divided by the total number of species recorded, thereby converting species richness of each resample to the proportion of the species found of the total species richness. A factorial ANOVA was used to compare the proportion of species richness found among locations and among sample sizes. As most anuran monitoring programs use stratified sampling, where three nights are sampled separated by at least 15 nights, we used resampling to emulate stratified three-night sampling. The methodology was the same as above, except three nights were selected only if separated by 15 days.

2.4. Species detectability

Data from the Backyard Frog survey were resampled to determine which species were most prone to under-representation when sampled infrequently. As previously described (see previous section), we created a species presence–absence matrix. Nights were randomly sampled n times, with replacement, with a sample size of 1, 2, 3, 5, 10, 15, 20, 30, 40 and 50 sampling nights. For each subsample, the proportion of trials in which each species was found was calculated. The random sampling was repeated 1000 times for each sample size. A repeated measures ANOVA was used to compare the proportion of trials in which the species was present, treating each subset as the repeated measure, and including species and location as main factors. Greenhouse–Geisser adjustments (Greenhouse and Geisser, 1959) were used to adjust the degrees of freedom because the different subsets ($n = 1, 2, 3, \dots, 50$) were not independent from one another. Tukey HSD tests were used for multiple comparisons following the ANOVAs. Only locations that had at least one observation of a species were included in the analysis.

2.5. Effect of unequal detection probabilities

The OMNR Bait Frog survey was used to evaluate if visual surveys improved estimates of site occupancy rates, and, the effects of detection probabilities on determining in site occupancy. To investigate whether the sampling method combining nocturnal acoustic and diurnal visual and acoustic surveys was superior to nocturnal acoustic surveys alone, a sign test was used to contrast the site occupancy of northern leopard frogs surveyed with each of the two methods. Kruskal–Wallis one-way analyses of variance were used to test for

independence of year and site occupancy (proportion of sites at which a species was detected) within each region for each species. In regions where site occupancy and year were not found to be independent, Mann–Whitney U tests were used to test for differences between each possible pair of years. Kruskal–Wallis one-way analysis of variance was applied to the northern leopard frog data to test for independence of presence and region with all years combined. The analyses were repeated within each region using only the data from nocturnal acoustic surveys, and the results were compared to the results using the combined acoustic and visual surveys. For all tests, differences were considered statistically significant at $p < 0.05$. Statistica 5.5 (StatSoft, 2000) and Systat 9 (SPSS, 1998) were used for statistical analyses and MathCad 6.0 (1996) was used for resampling techniques.

3. Results

3.1. Species richness (Backyard Frog survey)

The proportion of species found varied among locations ($F_{[31,60]} = 30.02$, $p < 0.0001$) and the number of survey nights ($F_{[9,60]} = 959.45$, $p < 0.0001$). With only one sampling night, the average species richness was only 25.1% of the total, but reached an average of 80% of the total species richness by approximately 12 sampling nights, and 90% by approximately 24 sampling nights (Fig. 1). However, there was no interaction between the location and sample size ($F_{[279,60]} = 0.0014$, $p = 0.9958$), indicating that the relationship between sample size and the proportion of species found was not influenced by the survey locations.

Species richness differed between the stratified three-night sampling and the random three-night sampling techniques. Species richness was higher for the stratified three-night sampling (58.2% vs. 49.9% of total species richness; $F_{[1,12]} = 5.98$, $p = 0.0011$; Fig. 2), and survey location influenced the proportion of species found ($F_{[31,12]} = 31.65$, $p = 0.0001$). There was no interaction between location and sampling regime ($F_{[31,12]} = 0.59$, $p = 0.8857$). The mean species richness was lower for the stratified three-night sampling compared to five nights or more of random sampling (vs. five-night random sampling; $F_{[1,12]} = 9.38$, $p = 0.0098$). Thus, although stratified sampling appears to be more efficient than random sampling when only a few nights are sampled, three nights of stratified sampling underestimate species richness.

3.2. Species detectability (Backyard Frog survey)

As expected, the assumption of sphericity was violated (Mauchly sphericity test, $\chi^2_{[44]} = 4243$, $p < 0.0001$), so the Greenhouse–Geisser adjustment was used. The number of trials (out of 1000) in which a species was found increased with the number of sampling nights ($F_{[0.17,262.9]} = 918.9$, $p < 0.0001$; Fig. 3(a)–(c)), but differed among each subsample of sampling nights, excepting 30, 40 and 50 nights, which were similar to each other (Tukey HSD test). The number of trials in which a species was found varied among species ($F_{[1.8,360.0]} = 1675$, $p < 0.0001$; Fig. 3(a)–(c)). Spring peepers were found significantly more frequently (89.8%) than all other species except green frogs (78.6%), while green frogs were found more frequently than wood frogs and northern leopard frogs (65.5% and 59.9%, respectively; Tukey HSD test). There were

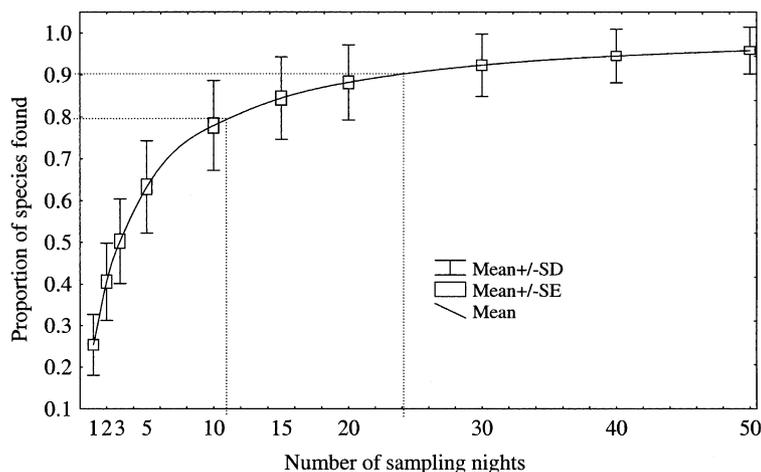


Fig. 1. Mean and standard error of the proportion of the total species found relative to the number of sampling nights, using data from the Ontario Backyard Frog survey. For each volunteer, the sampling nights were randomly sampled 1, 2, 3, 5, 10, 15, 20, 30, 40 and 50 times, with replacement from the total number of nights. The random sampling was repeated 1000 times. Species richness for each subsample was calculated, and expressed as a proportion of the species richness of the total number of nights.

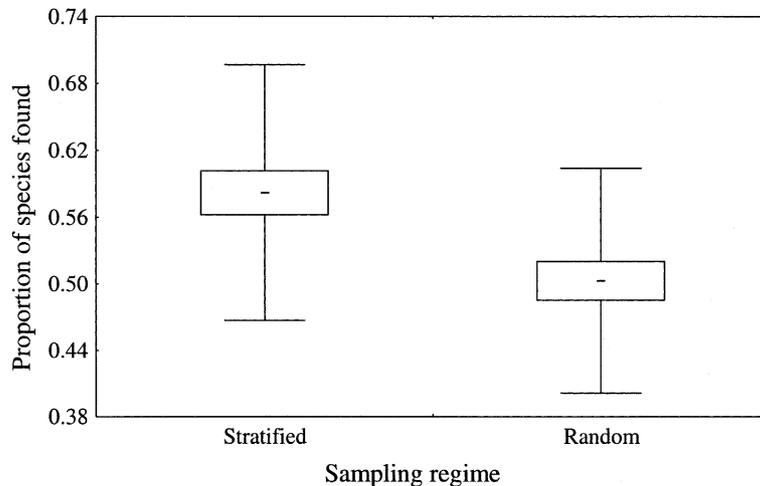


Fig. 2. Boxplots of mean, SE, SD of the proportion of species found, using stratified and random three-night sampling.

no differences among any other species. There was an interaction between species and the number of sampling nights, indicating that the relationship between the number of sampling nights and number of trials in which a species was found varied among species ($F_{[12.7, 360.0]} = 15.41$, $p < 0.0001$; Fig. 3(a)–(c)). Spring peepers were generally the easiest to detect of all the other species regardless of the number of sampling nights.

The proportion of trials in which a species was found did not differ among survey locations ($F_{[31, 187]} = 1.12$, $p = 0.2298$), nor was there an interaction between survey location and the number of sampling nights ($F_{[46.6, 280.9]} = 0.63$, $p = 0.9718$).

3.3. Species detectability (Bait Frog survey)

Based on the assumption that the detection of northern leopard frogs using the combined acoustic and visual surveys was one, the detectability of northern leopard frogs using only the nocturnal acoustic surveys was calculated as: $(\text{number of occupied sites}_{\text{nocturnal acoustic}} \times 100 / \text{number of occupied sites}_{\text{nocturnal acoustic} + \text{diurnal acoustic and visual}})$. The detectability of northern leopard frogs using only the acoustic surveys was variable, but underestimated site occupancy in almost all cases (Table 1). Site occupancy at each survey location based on only the nocturnal acoustic surveys was significantly different from estimates using the combined diurnal and nocturnal acoustic surveys (sign test, $p < 0.0005$). In the majority of cases where northern leopard frogs were detected with diurnal surveys, but not with nocturnal acoustic surveys, the animals were detected by their calls; diurnal calling activity was frequently more intense than nocturnal calling activity when nights were cool. Estimates of site occupancy from nocturnal acoustic surveys alone were substantially different from estimates using both nocturnal and diurnal surveys (Table 1). Detectability

of northern leopard frogs varied among years for central west Ontario (Pearson $\chi^2 = 9.6$, $p = 0.0080$). Out of seven comparisons among years (one for each location in Ontario), there was agreement between the two survey designs for only three comparisons, and they gave different statistical results for four comparisons (Table 1). In three of the four cases where there was a disagreement in statistical results, the acoustic-only survey failed to find a difference. However, in Central-west Ontario, there was no significant difference in prevalence among years using the combined acoustic and visual surveys, but there was a significant difference in prevalence using the acoustic-only survey (Table 1). This represents a type I error, as the acoustic-only survey should not have found a significant difference. Interestingly, there was a positive correlation between the detectability of the acoustic-only survey with site occupancy rates (Spearman rank correlation, $r = 0.47$, $t_{19} = 2.31$, $p = 0.0324$). This suggests that the estimated site occupancy rates of the acoustic-only survey were a function of the probability of detecting northern leopard frogs.

4. Discussion

Detectability affects estimates of anuran species richness and site occupancy for most, and potentially all, anuran acoustic surveys, and sampling effort is clearly an issue for numerous taxa regardless of the survey method used (Metcalf-Smith et al., 2000; Patton et al., 2000; Walther and Martin, 2001). The Backyard Frog survey likely produces higher species detection probabilities compared to other acoustic surveys, as some Backyard volunteers sample for more than 100 nights in a given year (Canadian Wildlife Service, unpublished data). Few other anuran monitoring programs that use volunteers approach this level of sampling effort, but generally

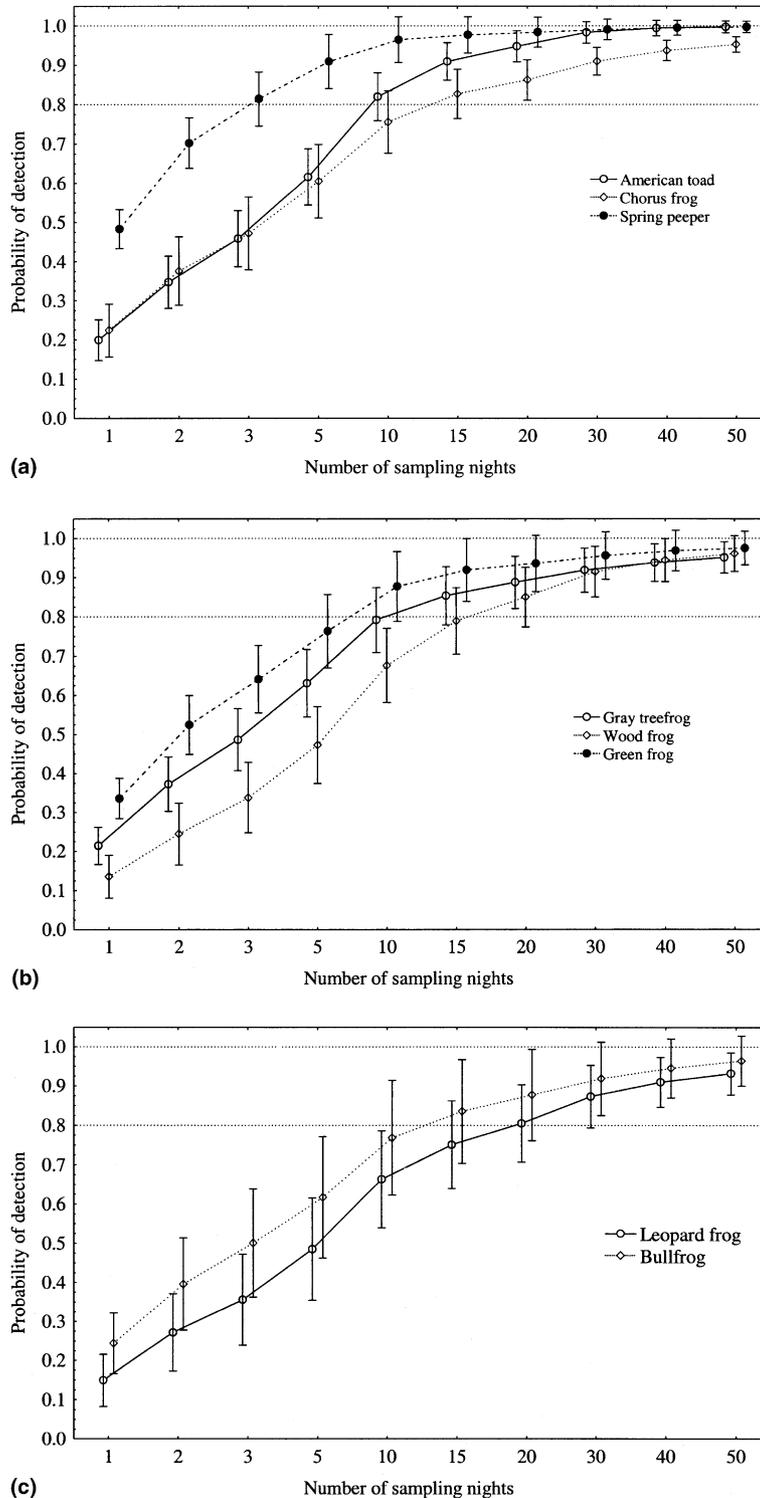


Fig. 3. (a–c) Relationship between number of nights sampled and mean (SE) probability of detection; (a) american toads, chorus frogs, and spring peepers; (b) gray treefrog, wood frog, and green frog; (c) northern leopard frog and American bullfrog

use three nights of sampling, albeit separated by approximately 15 nights between sampling events so as to maximize detection probabilities. Nevertheless, our data suggest that three nights of stratified sampling are not sufficient to reliably detect all anuran species, even after

excluding species that are rare or difficult to detect (e.g. Fowler's toad [*Bufo fowleri*], pickerel frog [*Rana palustris*]). Data presented from the NAAMP are consistent with our conclusions (Genet and Sargent, 2003; Weir et al., 2003).

Table 1
Comparison of nocturnal acoustic-only surveys with nocturnal acoustic + diurnal acoustic and visual surveys for *R. pipiens* from the Ontario Bait Frog survey

location	2001			2002			2003			p_a^a p_{av}
	Acoustic + visual prevalence (%)	Acoustic alone prevalence (%)	detectability (%) ^b	Acoustic + visual prevalence (%)	Acoustic alone prevalence (%)	Detectability(%)	Acoustic + visual prevalence (%)	Acoustic alone prevalence (%)	detectability(%)	
Region 1	100 (30/30)	80.0 (24/30)	80.0 (24/30)	80.0 (24/30)	60.0 (18/30)	75.0 (25/28)	93.3 (18/24)	83.3 (28/30)	89.3 (25/30)	0.083 0.022
Region 2	76.7 (23/30)	70.0 (21/30)	91.3 (21/23)	83.3 (25/30)	73.3 (22/30)	88.0 (22/25)	53.3 (16/30)	46.7 (14/30)	87.5 (14/16)	0.067 0.028
Region 3	80.6 (25/31)	64.5 (20/31)	80.0 (20/25)	77.4 (24/31)	74.2 (23/31)	95.8 (23/24)	67.7 (21/31)	61.3 (19/31)	90.5 (19/21)	0.537 0.476
Region 4	90.0 (28/31)	80.6 (25/31)	89.3 (25/28)	58.1 (18/31)	45.2 (14/31)	77.8 (14/18)	93.5 (29/31)	87.1 (27/31)	93.1 (27/29)	0.001 <0.0005
Central Ontario*	71 (22/31)	41.9 (13/31)	59.1 (13/22)	25.8 (8/31)	19.4 (6/31)	75.0 (6/8)	64.5 (20/31)	58.1 (18/31)	90.0 (18/20)	0.008 0.001
Central-east	80.0 (16/20)	60.0 (12/20)	75.0 (12/16)	30.0 (6/20)	30.0 (6/20)	100.0 (6/6)	75.0 (15/20)	65.0 (13/20)	86.7 (13/15)	0.059 0.002
Central-west	54.5 (6/11)	9.1 (1/11)	16.7 (1/6)	18.2 (2/11)	0.0 (0/11)	0.0 (0/2)	45.5 (5/11)	45.5 (5/11)	100.0 (5/5)	0.016 0.202

The data is separated for six regions in Ontario; Central Ontario is the combined Central-east and Central-west locations.

^a p -values for Kruskal–Wallis One-way Analysis of *Prevalence* vs. *Year* based on nocturnal acoustic surveys alone (p_a) and combined nocturnal acoustic + diurnal acoustic and visual surveys (p_{av}).

^b detectability of nocturnal acoustic surveys alone calculated as: $\text{number of occupied sites}_{\text{nocturnal acoustic}} \times 100 / \text{number of occupied sites}_{\text{nocturnal acoustic} + \text{diurnal acoustic and visual}}$.

Detection probabilities varied among species. Not surprisingly, spring peepers were relatively easy to detect even with little sampling effort, whereas wood frogs and northern leopard frogs were the most difficult to detect. Wood frogs are explosive breeders and call for only a short time, thus may be missed by call surveys that use few sampling nights. Crouch and Paton (2000), for example, found that wood frogs were likely to be missed in acoustic surveys in New England at breeding ponds. The calls of northern leopard frogs do not carry well over distance, and are easily masked by other calling frogs. Detection probabilities of chorus frogs were similar to the majority of species, but may have been overestimated. Spring peepers were the only species that had substantial error in inter-observer agreement in identification (Shirose et al., 1997), and were almost always misidentified as chorus frogs because of the similarity between the calls of chorus frogs and the aggressive call of peepers (Shirose, unpublished data). Considering the high detection probabilities of spring peepers, this misidentification may erroneously increase the probability of detecting chorus frogs. Chorus frogs can easily be missed using nocturnal acoustic surveys because they often call during the day (Shirose, L.; Schueller, F; unpublished data). Although we did not evaluate these species, both mink frogs (*Rana septentrionalis*) and pickerel frogs have proved to be difficult to detect in volunteer calling surveys in Quebec (Lepage et al., 1997).

One of the main objectives of anuran monitoring programs is to detect temporal changes in population status. Unfortunately, temporal or spatial differences in detectability may falsely mask a real population change, or falsely produce one. We used data from the Bait Frog survey to demonstrate both scenarios. There were significant numbers of sites where northern leopard frogs were detected with the combined acoustic and visual surveys but were not detected by the nocturnal acoustic-only surveys. All of the data from the nocturnal acoustic-only survey was also included in the combined acoustic and visual data set, and should have higher precision and thus statistical power. Hyde and Simons (2001) argued that heterogeneity of detectability reduces power to detect either spatial or temporal population trends. We found that the nocturnal acoustic-only protocol failed to detect significant temporal differences in site occupancies. However, while there was a general reduction in power due to suboptimal detection probabilities (Table 1), we also found that variation in detectability increased the likelihood of finding false differences. It is disconcerting that more than half the time the two survey designs gave different statistical results. Data on chorus frogs collected during the Bait Frog survey found that although there were no significant differences in prevalence among years in occupancy rates, the acoustic-only survey and the combined surveys had trends in different directions, which corresponded to

differences in detectability among years (Shirose and Brousseau, unpublished data). Unless the detectability of species with nocturnal acoustic-only surveys is constant across years, the estimates of population trends will have a bias of unknown size and direction.

Detectability varies among species, thus the effect of detection will depend on the calling characteristics of the species. Detectability problems may be expected with species having a very brief calling period (e.g., wood frogs), and species that tend not to form dense choruses, and tend to call sporadically and/or at relatively low volume (e.g., northern leopard frogs). Difficulty in detecting certain species has been a problem with other call surveys (Bridges and Dorcas, 2000; Crouch and Paton, 2002). MacKenzie and Kendall (2002) argued that we should assume that detection probabilities are different unless otherwise demonstrated.

Call surveys are not the only survey method affected by suboptimal detection probabilities. Gibbons et al. (1997) demonstrated that the measurement of species diversity of reptiles and amphibians is strongly dependent on sampling effort, based upon four decades of research at the Savannah River Ecology Laboratory. Over a 40-year period, amphibian and reptile diversity was estimated using nearly every viable surveying method, yet even after four decades of sampling, species richness remained a function of sampling effort (Gibbons et al., 1997). Relative sampling efficiency is an important factor influencing estimates of both species presence and species richness for insects, fish, and birds (Bayley and Peterson, 2001; Boulinier et al., 1998; Brown and Boyce, 1998). Sampling artifacts are likely a universal phenomenon affecting many survey programs.

A method to estimate detection probabilities or sampling effort should be included in survey methodologies (MacKenzie and Kendall, 2002). Surveys with relatively little sampling effort require extrapolation to estimate species richness. Here, we used a bootstrapping approach to model sampling effort. Both bootstrapping and rarefaction can be used for extrapolation of species richness using smaller sample sizes, but they cannot be used for extrapolation outside of the sample sizes used (Tipper, 1979). Both asymptotic and non-asymptotic estimators of species richness have been used with some success particularly with highly diverse communities. However, these methods can be risky and there is little reason to use them when sampling effort is sufficiently high to approach an asymptote (Gotelli and Colwell, 2001), and when there are relatively few frog species in a region.

Our study indicates that volunteer-based anuran monitoring programs and other population surveys would benefit from a number of methodological and/or statistical changes. An increase in sampling effort by the use of numerous sampling periods and/or the incorporation of detection probabilities into statistical

models would increase the accuracy of estimated population trends. An increase in sampling effort by surveyors is preferable as it reduces the likelihood of missing species. While changes in sampling protocol may reduce the number of volunteers, and at first seem to result in incompatible data, the use of likelihood-based statistical methods would overcome the issue of incompatible data and allow for temporal and/or spatial comparisons in site occupancy rates after adjusting for detection probabilities. Another benefit of using likelihood-based statistical methods is that covariates such as temperature or habitat may be incorporated into the model (MacKenzie et al., 2002). However, this approach does not determine which sites are occupied, and unequal detectability may bias habitat use inferences (Gu and Swihart, 2003; Pert et al., 1997). Finally, our study indicates that monitoring selected anurans, particularly ranids, would benefit from using a combination of diurnal visual and acoustic surveys to more accurately monitor populations.

Acknowledgements

Many volunteers were invaluable for the collection of the Ontario Backyard Frog survey data, and we thank them for their time and diligence. Michael Berrill, Stephen Hecnar, and Frederick Schueler gave valuable advice in designing the Ontario Bait Frog survey, and Tana McDaniel and Neil Osborne aided in collecting the data.

References

- Bailey, L.L., Simons, T.R., Pollock, K.H., 2002. Estimating temporary emigration and detectability in plethodon salamanders using Pollock's robust design. Society for Conservation Biology 16th Annual Meeting July 14–July 19 2002 co-hosted by DICE and the British Ecological Society.
- Bayley, P.B., Peterson, J.T., 2001. An approach to estimate probability of presence and richness of fish species. *Transactions of the American Fisheries Society* 130, 620–633.
- Bishop, C.A., Pettit, K.E., Gartshore, M.E., MacLeod, D.A. 1997. Extensive monitoring of anuran populations using call counts and road transects in Ontario (1992 to 1993). In: Green D.M., (Ed.), *Amphibians in Decline: Canadian Studies of a Global Problem*, Herpetological Conservation Volume 1. Society for the Study of Amphibians and Reptiles, St. Louis, Missouri, USA, Pp. 149–160.
- Boulinier, T., Nichols, J.D., Sauer, J.R., Hines, J.E., Pollock, K.H., 1998. Estimating species richness: The importance of heterogeneity in species detectability. *Ecology* 79, 1018–1028.
- Bridges, A.S., Dorcas, M.E., 2000. Temporal variation in anuran calling behavior: implications for surveys and monitoring programs. *Copeia* 2000, 587–592.
- Brown, J.A., Boyce, M.S., 1998. Line transect sampling of Karner blue butterflies (*Lycaeides melissa samuelis*). *Environmental and Ecological Statistics* 5, 81–91.
- Crouch III, W.B., Paton, P.W.C., 2000. Using egg-mass counts to monitor wood frog populations. *Wildlife Society Bulletin* 28, 895–901.
- Crouch III, 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.
- Genet, K., Sargent, L.G., 2003. Evaluation of methods and data quality from a volunteer-based amphibian call survey. *Wildlife Society Bulletin* 31, 703–714.
- Gibbons, J.W., Burke, V.J., Lovich, J.E., Semlitsch, R.D., Tuberville, T.D., Brodie, J.R., Greene, J.L., Niewiarowski, P.H., Whiteman, H.H., Scott, D.E., Pechman, J.H.K., Harrison, C.R., Bennett, S.H., Krenz, J.D., Mills, M.S., Buhlmann, K.A., Lee, J.R., Seigel, R.A., Tucker, A.D., Mills, T.M., Lamb, T., Dorcas, M.E., Congdon, J.D., Smith, M.H., Nelson, D.H., Dietsch, M.B., Hanlin, H.G., Ott, J.A., Karapatakis, D.J., 1997. Perceptions of species abundance, distribution, and diversity: lessons from four decades of sampling on a government-managed reserve. *Environmental Management* 21, 259–268.
- Gotelli, N.J., Colwell, R.K., 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4, 379–391.
- Greenhouse, S.W., Geisser, S., 1959. On methods in the analysis of profile data. *Psychometrika* 24, 95–112.
- Gu, W., Swihart, R.K., 2003. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation* 116, 195–203.
- Hyde, E.J., Simons, T.R., 2001. Sampling plethodontid salamanders: sources of variability. *Journal of Wildlife Management* 65, 624–632.
- Lepage, M., Courois, R., Daigle, C. 1997. Surveying calling anurans in Québec using volunteers. In: Green D.M. (Ed.), *Amphibians in Decline: Canadian Studies of a Global Problem*, Herpetological Conservation Volume 1. Society for the Study of Amphibians and Reptiles, St. Louis, Missouri, USA, Pp. 128–140.
- 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.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.
- Metcalfe-Smith, J.L., Di Maio, J., Staton, S.K., Mackie, G.L., 2000. Effect of sampling effort on the efficiency of the timed search method for sampling freshwater mussel communities. *Journal of the North American Benthological Society* 19, 725–732.
- Paszkowski, C.A., Scrimgeour, G., Gingras, B.A., Kendall, S., 2002. A comparison of techniques for assessing amphibian assemblages on streams in the western boreal forest. *Canadian Field-Naturalist* 116, 116–119.
- Patton, T.M., Hubert, W.A., Rahel, F.J., Gerow, K.G., 2000. Effort needed to estimate species richness in small streams on the great plains in Wyoming. *North American Journal of Fisheries Management* 20, 394–398.
- Pechman, J.H.K., Scott, D.E., Semlitsch, R.D., Caldwell, J.P., Vitt, L.J., Gibbons, J.W., 1991. Declining amphibian populations: the problem of separating human impacts from natural fluctuations. *Science* 253, 892–895.
- Pert, E.J., Deinstadt, J.M., Erman, D.C., 1997. Evidence of differential detectability of lotic adult rainbow and brown trout: implications for habitat use inference from direct-underwater observation. *Journal of Freshwater Ecology* 12, 359–365.
- Shirose, L., 2001. Monitoring bait-frog (northern leopard frog, *R. pipiens*) population status in Ontario. Unpublished report prepared for Fisheries Section, Ontario Ministry of Natural Resources.
- Shirose, L.J., Bishop, C.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.

- Shirose, L.J., Bishop, C.A., MacDonald, C.J., Judge, K.A., Brooks, R.J. In review. Anuran call surveys: is calling activity an index of abundance?.
- SPSS, Inc., 1998. SYSTAT 9.0 for Windows. Chicago.
- StatSoft, Inc., 2000. STATISTICA for Windows [Computer program manual]. Tulsa, OK 74104.
- Storfer, A., 2003. Amphibian declines: future directions. *Diversity and distributions* 9, 151–163.
- Tipper, J.C., 1979. Rarefaction and rarefaction – the use and abuse of a method in paleoecology. *Paleobiology* 5, 423–434.
- Wake, D.B., 1991. Declining amphibian populations. *Science* 253, 860.
- Walther, B.A., Martin, J.L., 2001. Species richness estimation of bird communities: how to control for sampling effort?. *Ibis* 143, 413–419.
- Weir, L.A., Royle, J.A., Jung, R., Nanjappa, P., Funk, L., Faust, S., Schneider, S., Erb, L., Hildebrand, W., 2003. Assessing site occupancy rates, species detectability and anuran-habitat relationships along North American amphibian monitoring program calling routes in Maryland. Annual Meeting of the Society for Conservation Biology, June 28, 2003 – July 2, 2003, Duluth, MN, USA.