Monitoring plan to detect trends in occupancy of Illinois chorus frogs (*Pseudacris streckeri illinoensis*)

Bradley J. Cosentino[1]

[1]Department of Biology, Hobart and William Smith Colleges, Geneva, NY, USA

Prepared for: Illinois Department of Natural Resources

© 2004 Stanley Trauth

Please cite as: Cosentino, B.J. 2014. Monitoring plan to detect trends in occupancy of Illinois chorus frogs (*Pseudacris streckeri illinoensis*). Illinois Department of Natural Resources, Springfield, Illinois, USA.

Executive summary

The Illinois chorus frog (ICF) has a narrow geographic distribution and is vulnerable to the loss and modification of wetlands and sand prairie habitat. Local extinctions have been documented range-wide, and ICF is listed as a threatened species in Illinois. A statewide monitoring program is needed to 1) determine the spatial extent of ICF's distribution (hereafter "occupancy"), 2) determine whether occupancy is declining, and 3) make informed decisions about conservation management and legal status.

When designing a monitoring program, wildlife biologists need to make decisions about how to allocate limited resources in order to maximize the likelihood of detecting declines. In this report, I propose methodology for conducting field surveys and conducting statistical analyses to detect declines in occupancy of ICF in Illinois. I used four years of pilot data and rigorous statistical analyses to make the following core recommendations:

- A total of 75–90 permanent sampling sites should be established in Illinois, and sites should be surveyed twice each year. This level of survey effort is required to obtain precise estimates of occupancy and to maximize the probability of detecting a 30-50% decline in occupancy over time.

- Because the probability of detecting ICF during surveys is low during drought years, sites should be sampled up to three times during drought years.

- Sites should be established throughout the range of suitable habitat in Illinois in order to make generalizations about the status of ICF at a statewide level.

- Detecting occupancy declines of 10-20% is unlikely, so these levels should not be used as thresholds for management or legal decisions.

- Although declines greater than 20% are worrisome, smaller year-to-year declines may reflect random changes in environmental conditions rather than true trends. Decisions about management or legal status should consider the consistency of trends over longer periods of time.

- Decisions about optimal allocation of survey effort are based on estimates on occupancy and the probability of detecting ICF during surveys. Occupancy and detection probability estimates should be evaluated every three years to ensure optimal survey effort.

1. Introduction

The Illinois chorus frog (*Pseudacris streckeri illinoensis*, hereafter "*P. illinoensis*") requires sandy soils in upland areas and ephemeral and fishless wetlands for reproduction (Brown and Rose 1988). Cultivation has led to the loss and fragmentation of wetland and sand prairie habitats, and local extinctions of ICF have been documented in Illinois, Arkansas, and Missouri (e.g., Tucker 1998, Trauth et al. 2006). However empirical evidence of a decline is lacking because of insufficient duration, spatial extent, and design of previous studies. In Illinois, *P. illinoensis* is listed as threatened, and it is listed as a "Critical Species" in the Illinois Wildlife Action Plan (Illinois Department of Natural Resources 2005). Applications of a long-term, large-scale monitoring program include making informed decisions about the legal status of *P. illinoensis* and evaluating responses to implementation of the Wildlife Action Plan.

Monitoring programs for pond-breeding amphibians can be focused at multiple spatial scales. Monitoring abundance within breeding ponds can provide insight into how deterministic (e.g., habitat loss and fragmentation) and stochastic factors affect local population dynamics. However, local population dynamics often depend on the dynamics of nearby populations due to dispersal (Sjögren-Gulve 1994, Cosentino et al. 2011), meaning that clusters of breeding ponds may not exhibit demographic or statistical independence (Petranka et al. 2004). Furthermore, pond-breeding amphibians have been shown to persist in regional metapopulations despite extinction-colonization dynamics at the pond scale (e.g., Werner et al. 2007, Cosentino et al. 2011). These observations suggest that monitoring and conservation approaches for pond-breeding amphibians should be considered at landscape scales. One effective approach to monitoring at landscape scales is to examine change in a species' spatial distribution over time (e.g., Weir et al. 2009, Sewell et al. 2012, Adams et al. 2013). The overarching objective of this report is to propose a monitoring plan that would allow the Illinois Department of Natural Resources (IDNR) to detect trends in occupancy for *P. illinoensis* at a statewide level.

1.1. Sampling and analytical approach

Occupancy is defined as the proportion of sample units occupied by a focal species, and occupancy is estimated by detecting or failing to detect the focal species within each sample unit. The primary goal of a monitoring program for *P. illinoensis* would be to examine whether occupancy varies over time. The main challenge to detecting occupancy trends is differentiating a true trend from changes in occupancy driven by sampling error (Gibbs et al. 1998). There are two important sources of sampling error that can make it difficult to detect true trends in occupancy.

First, occupancy varies spatially, but we are usually restricted from sampling all possible areas that could be occupied by the focal species (Bailey and Nichols 2010). Instead, a random sample of sites must be selected, and the sample should be representative of all potential areas to which inferences will be made (i.e., the statistical population). For long-term monitoring programs, surveyors may sample the same sites every year (i.e., standard design) or randomly sample new sites each year (i.e., rotating panel design; Urquhart and Kincaid 1999). However, a rotating panel design has not been shown to increase precision of trend estimates relative to a standard design, and occupancy trends inferred from a rotating panel design could be due to the confounding effect of sampling different sites in different years (MacKenzie 2005, Bailey et al. 2007).

Second, sites may be recorded as unoccupied even if the focal species is present because the surveyor's ability to detect the species is imperfect. Failing to consider imperfect detection probability can lead to occupancy estimates that are negatively biased (MacKenzie et al. 2006). With respect to monitoring, occupancy trends may be induced by trends in detection probability over time if the surveyor's ability to detect the focal species increases or decreases. Furthermore, inferences about habitat factors affecting occupancy probability may be wrong if detection probability varies spatially. Unbiased estimates of occupancy can be generated by using models to estimate occupancy probability after explicitly accounting for imperfect detection (MacKenzie et al. 2002, 2003).

Occupancy models that account for imperfect detection require that multiple surveys be conducted at each site within a short enough timeframe (i.e., a season) so that the site is closed to changes in occupancy status. Replicate surveys are used to estimate detection probability (p) and occupancy probability (ψ) simultaneously. When a site is surveyed multiple times, the probability of detecting the focal species at least once at a site ($p*$) can be quantified as

$$p* = 1 - (1 - p)^K ,$$

where p is the probability of detection during a single survey and K is the number of surveys. For example, if a site was surveyed three times and the probability of detecting the organism during a single survey was 0.5, the probability of detecting the focal species at least once during the three surveys is

$$p* = 1 - (1 - 0.5)^3 = 0.875 .$$

If detection probability during a single survey is known, the number of surveys needed to detect a species with a given degree of confidence can be estimated as

$$K = \frac{\log(1 - p*)}{\log(1 - p)}$$

(Pellet and Schmidt 2005, Sewell et al. 2012).

To estimate occupancy and detection probabilities within a single season, detection histories (h) are first constructed for each site where '1' represents the species being detected during a survey and '0' represents the species not being detected during a survey. For example, a site surveyed three times with the detection history '010' means that the species was not detected during the first survey, detected during the second survey, and not detected during the third survey. A maximum likelihood approach is used to determine occupancy and detection estimates given the detection histories for all sites:

$$L(\psi, p \middle| h) = \left[\psi^{S_D} p^d (1-p)^{KS_{D-d}}\right]\left[1-\psi p^*\right]^{S-S_D},$$

where S is the number of sites surveyed, S_D is the number of sites where the species was detected, and d is the total number of detections in the detection histories. Analytical solutions for the maximum likelihood estimates are reported in MacKenzie et al. (2006, p. 95). The single-season occupancy model with constant occupancy and detection parameters can be extended to model heterogeneity in occupancy and detection probabilities among sites or over time (MacKenzie et al. 2002). Additionally, multiple seasons of occupancy data can be included in a single model by estimating colonization and extinction probabilities (MacKenzie et al. 2003).

In the context of occupancy modeling, there are multiple statistical tests to test for temporal changes in occupancy. First, a multi-season model can be used to infer a decline by comparing estimated extinction and colonization probabilities, comparing models with constant or year-specific variation in occupancy probabilities, or using a regression approach to test for a trend in seasonal occupancy estimates derived from the multi-season model (e.g., Weir et al. 2009). Second, a Wald test or likelihood ratio test can be used to test for a difference in estimated occupancy between two time periods. For example, one could compare occupancy between the first and last year of a monitoring program to test for a decline.

The advantage of using a multi-season model to infer trends is that 1) there is no decision about which two seasons to compare when the number of seasons is >2, whereas a decision must be made about which occupancy estimates to compare when using a Wald test, 2) occupancy data for >2 seasons can be used directly in a single analysis, and 3) the multi-season model can account for Markovian dependence in occupancy status at sites (MacKenzie et al. 2006). Unfortunately, little has been published on how to design studies to maximize power to detect trends using multi-season models (e.g., Mattfeldt et al. 2009). Conversely, study design and power analyses have been explicitly examined for comparing occupancy estimates between seasons, so I used this approach for *P. illinoensis*. Importantly, Guillera-Arroita and Lahoz-Monfort (2012) showed 1) the power of a Wald test to detect occupancy differences between seasons was not sensitive to assuming Markovian dependence, and 2) the power to detect the difference in occupancy between the first and last seasons of a monitoring program was similar to the power of regression to detect a trend in occupancy estimates for >2 seasons (Guillera-

Arroita and Lahoz-Monfort 2012). Thus, power analyses to compare occupancy between two time periods may also be applicable to regression approaches for detecting trends.

A primary consideration when designing a monitoring program should be to maximize the precision of occupancy estimates (i.e., minimize the variance) because precision affects the power to detect differences in occupancy across seasons (Guillera-Arroita and Lahoz-Monfort 2012). Previous studies have shown that the magnitude of occupancy and detection probabilities affects the optimal number of repeated surveys to conduct at each site in order to 1) maximize precision of occupancy estimates and 2) achieve a given level of power to detect differences in occupancy across seasons (MacKenzie and Royle 2005, Bailey et al. 2007, Guillera-Arroita et al. 2010). Increasing survey replication increases the probability of detecting a species at least once at a site, but increasing survey replication usually comes at a cost to the number of sites that can be sampled (assuming the total number of surveys that can be conducted in a season is limited). Estimates of occupancy and detection probabilities from pilot data are helpful for guiding decisions about how to allocate survey effort in order to maximize precision of the occupancy estimator and the ability to detect changes in occupancy between time periods.

1.2. Objectives

From 2011-2014, IDNR biologists surveyed 30 sites for occupancy of *P. illinoensis* in multiple regions of Illinois using auditory chorus surveys. Sites were surveyed multiple times, which facilitates the estimation of occupancy probabilities that account for imperfect detection. I used occupancy models to accomplish the following objectives:

1) Evaluate factors that affect detection probability during surveys (e.g., date of survey, temperature, wind) in order to make recommendations for how to conduct surveys in a way that maximizes detection of *P. illinoensis*.

2) Assess whether occupancy probability varies among regions or depends on the amount of habitat (e.g., ponds, sandy/hydric soils) available.

3) Estimate average occupancy and detection probabilities for *P. illinoensis.*

4) Use average occupancy and detection probabilities to determine the optimal number of surveys to conduct at each site in order to maximize the precision of occupancy estimates.

5) Determine how the power to detect a decline in occupancy depends on the sampling effort, magnitude of decline, levels of detection probability, and levels of statistical significance.

6) Make recommendations about designing a monitoring program to detect changes in *P. illinoensis* occupancy over time.

2. Methods

2.1. Study area and sampling design

The statistical population of interest for monitoring is suitable habitat within the current geographic range of *P. illinoensis*. A GIS-based habitat model was generated for *P. illinoensis* based on its current range and presence of hydric soils, sandy soils, and wetlands for breeding (Hinz et al. 2011). The resulting area of suitable habitat consists of five regions: 1) Mason and Menard Counties (M-M), 2) Cass, Morgan, and Scott Counties (C-M-S), 3) Madison County, 4) Monroe County, and 5) Alexander County.

Three regions (M-M, C-M-S, and Alexander) were chosen for a pilot study from 2011-2014. Sites were defined as 1 x 1-mile plots (hereafter "sections") that included suitable habitat identified by the habitat model (Beltz 1993). A pool of sections was created within suitable habitat, and ten sections were randomly selected within each of the three regions for surveys. Sections in all regions were surveyed in 2011-2012, with the exception being that only 8 of 10 sections were surveyed in C-M-S in 2011 (Table 1). M-M and Alexander were not surveyed in 2013, and M-M was not surveyed in 2014.

Each section was surveyed 1–3 times between February and April in each year (Table 1), and sections were assumed to be closed to changes in occupancy status within years. During each survey, observers drove along roads along the perimeter of each section and made multiple stops to listen for chorusing *P. illinoensis* within the section. There was a single observer in each region. In addition to recording whether *P. illinoensis* was heard, observers recorded the date and time of the survey, number of stops during the survey (listening posts), temperature, and wind. Wind was recorded as an ordinal variable from 1 (light breeze) to 4 (moderate breeze).

2.2. Occupancy modeling

First, I used data from repeated surveys in all regions and years in a multi-season occupancy model to examine factors that affected detection probability and to estimate average detection probability during a survey (MacKenzie et al. 2003). I built eight models with different combinations of covariates while holding initial occupancy probability (ψ), colonization probability (γ), and extinction probability (ε) constant. Potential detection covariates included year, observer/region, Julian date, time of survey, number of listening posts, temperature, and wind. Because a single observer was assigned to each region, observer and region were confounded. Models included a null model (intercept only), a model with an effect of year, and six models with additive effects of year and one additional covariate. Year was included in all models except the null model because exploratory analyses revealed strong variation in detection probability among years.

Next, I used single-season occupancy models for 2011 and 2012 to examine how region and number of habitats affected occupancy probability and to estimate average occupancy probability in each year (MacKenzie et al. 2002). I also modeled variation in detection probability using the same covariates used for the multi-season model. I restricted this analysis to the 2011 and 2012 datasets because too few sections were sampled to model variation in occupancy in 2013 and 2014 (Table 1). The model set for each year consisted of 21 models, including a single null model (intercept only for occupancy and detection probabilities), six models that included a single detection covariate while holding occupancy constant, a single model including an effect of region on occupancy while holding detection constant, six models that included an effect of region on occupancy and a single detection covariate, a single model including an effect of habitats on occupancy while holding detection constant, and six models that included an effect of habitats on occupancy and a single detection covariate. Number of habitats was the number of GIS polygons in a section that represented wetlands, hydric soils, or sandy soils. Models with effects of observer/region and wind on detection did not converge for the 2011 model set (6 total models), and a model with an effect of observer/region on detection and an effect of region on occupancy did not converge for the 2012 model set.

All occupancy models were fit in R (R Core Team 2012) with the package *unmarked* (Fiske and Chandler 2011). The Akaike Information Criterion (AIC) was used to evaluate the relative support of each model in each model set (Burnham and Anderson 2002). Models were considered competitive when the difference between model AIC and AIC of the most-supported model (ΔAIC) was $\leq$2 (Burnham and Anderson 2002). Estimates of average occupancy and detection probabilities in each model set were calculated by first predicting occupancy and detection using each model while holding all covariates at their mean. Predictions from each model were then model-averaged across all models in the model set using Akaike weights. Unconditional standard errors were quantified to account for both sampling variance and model selection uncertainty (Burnham and Anderson 2002).

2.3. Survey design and power analyses

For all analyses used to make recommendations on survey design, I assumed a standard sampling design (S sections each sampled K times in each year) where occupancy probability was set to 0.5 and detection probability was evaluated at 0.6, 0.7, 0.8, and 0.9. Occupancy and detection probabilities were based on estimates of average occupancy probability in 2011 and 2012 and yearly variation in detection probability from 2011–2014 (see Results and Discussion). I considered scenarios in which fixed total survey effort (i.e., TS, the maximum number of total surveys that can be conducted in a year) was 60, 90, 120, 150, or 180. Survey effort can be allocated in various ways under a given TS. For example, $TS = 60$ when 30 sections are each sampled two times and when 20 sections are each sampled three times. My goal was to determine optimal allocation of survey effort to minimize variance of the occupancy estimator and to maximize power to detect differences in occupancy between time periods.

First, I determined the optimal number of surveys to conduct at each section in order to minimize the asymptotic variance of the occupancy probability estimate in a single season:

$$Var(\hat{\psi}) = \frac{\psi}{S}\left[(1-\psi) + \frac{1-p^*}{p^* - Kp(1-p)^{K-1}}\right] = \frac{\psi}{S}(1-\psi+F)$$

(MacKenzie and Royle 2005, Guillera-Arroita et al. 2010). I found that the optimal survey effort was either two or three surveys per section when $\psi = 0.5$ and $p = 0.6$–0.9. However, asymptotic approximations of occupancy and detection estimates assume large sample sizes (Bailey et al. 2007), and Guillera-Arroita et al. (2010) showed that estimates are biased when sample size is small. Thus, I used simulations to estimate the root mean-squared error (RMSE) for occupancy probability for each combination of parameter estimates. I considered survey design parameters where *TS* was 60–180 and *K* was 2, 3, or 4 assuming different combinations of occupancy and detection probabilities. Simulations for each combination of design parameters consisted of 50,000 iterations. Simulations were conducted in R using code provided by Guillera-Arroita et al. (2010).

Second, I conducted a power analysis to detect a proportional decline in occupancy probability between two time periods using a Wald test, which was shown to be a more powerful test than a likelihood ratio test (Guillera-Arroita and Lahoz-Monfort 2012). The null hypothesis for the Wald test is that there is no difference in occupancy probability between time periods. Two possible errors can be made when testing for differences in occupancy between time periods: 1) A difference in occupancy is detected when in fact there is no difference in occupancy between time periods (Type I error), and 2) No difference in occupancy is detected when in fact there is a difference in occupancy between time periods (Type II error). The significance value, α, represents the probability of committing a Type I error, and β represents the probability of committing a Type II error. Thus, power (1 – β) is the probability of detecting a difference in occupancy probability given there is a true difference of a given magnitude.

I examined how different levels of survey effort ($K = 2$ or 3), detection probability ($p = 0.6$-0.9), and significance ($\alpha = 0.05$ or 0.10) affected the power to detect declines in occupancy of 10-50%. Power was quantified by simulating occupancy data (Guillera-Arroita and Lahoz-Monfort 2012), and initial occupancy probability was assumed to be 0.5 (consistent with the occupancy estimate in 2012; see Results and Discussion). I ran 1000 simulations per scenario. For each simulation, occupancy was calculated using maximum likelihood, the standard error was quantified using the delta method, and the Wald statistic was quantified as

$$\frac{\hat{\psi}_1 - \hat{\psi}_2}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}},$$

where $\hat{\psi}_1$ and $\hat{\psi}_2$ are the estimates of occupancy in time periods one and two, and $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the estimated variances for occupancy in time periods one and two, respectively. Power was the percentage of simulations for which the Wald statistic was greater than 1.96 for $\alpha = 0.05$ or 1.645 for $\alpha = 0.10$. Simulations and power analyses were conducted in R using code provided by Guillera-Arroita and Lahoz-Monfort (2012).

3. Results and Discussion

Pseudacris illinoensis was detected at 15 of 28 sections in 2011 (naïve occupancy = 0.54), 12 of 30 sections in 2012 (naïve occupancy = 0.40), 8 of 10 sections in 2013 (naïve occupancy = 0.80), and 11 of 20 sections in 2014 (naïve occupancy = 0.55). Data from 2011 and 2012 are most comparable because they were the only two years in which sections were sampled in all regions.

3.1. Multi-season occupancy model for 2011-2014

Based on a multi-season occupancy model, model-averaged estimates were 0.56 (SE = 0.09) for initial occupancy probability, 0.12 (SE = 0.08) for colonization probability, 0.15 (SE = 0.07) for extinction probability, and 0.77 (SE = 0.05) for detection probability. Average detection probability was quite high compared to detection probability estimates for other pond-breeding amphibians from the North American Amphibian Monitoring Program (e.g., Weir et al. 2005, Cosentino et al., unpublished manuscript). However, there was some temporal variation in detection. The most-supported model of detection included effects of year and date (Table 2). Average detection probability was greatest in 2011 and lowest in 2012, whereas detection estimates in 2013 and 2014 were closer to the average (Fig. 1). The low detection probability in 2012 may have been due to drought conditions (IDNR 2013) causing wetlands to be dry, assuming fewer males chorus during drought compared to years with normal precipitation.

Detection probability was negatively related to Julian date (Fig. 2; beta = -0.76, SE = 0.40). Detection probability was greatest in February and early March and lowest in late March and April. These results indicate that surveys should generally be completed by mid-April, particularly in drought years (see 2012, Fig. 2).

Temperature, number of listening posts, wind, time of survey, and observer/region did not have consistent effects on detection probability across years. However, some of these variables did have marginal effects on detection probability in 2011 or 2012 (discussed below).

3.2. Occupancy modeling in 2011 and 2012

In 2011, the model-averaged parameter estimates were 0.97 (SE = 0.04) for detection probability and 0.56 (SE = 0.10) for occupancy probability. In 2012, the model averaged parameter estimates were 0.63 (SE = 0.19) for detection probability and 0.50 (SE = 0.17) for occupancy probability. In 2011, the estimate of occupancy probability after accounting for imperfect detection was only marginally different from naïve occupancy because detection probability approached unity. In 2012, estimated occupancy probability was 25% greater than naïve occupancy because detection probability was low. This temporal variation in detection probability illustrates why it is critical to account for imperfect detection when estimating occupancy. Naïve occupancy showed a much steeper decline from 2011 to 2012 compared to the occupancy estimate that accounted for imperfect detection.

The most-supported models for 2011 and 2012 suggested that occupancy probability varied among regions (Tables 3, 4). Occupancy was greatest in Alexander and lowest in M-M (Figs. 3, 4). Models with an effect of the number of habitats on occupancy had competitive support in both years. Occupancy probability was positively related to the number of habitats, but the effect was slightly stronger in 2011 (Fig. 5; beta from most-supported model with habitats = 1.57, SE = 0.76) than in 2012 (Fig. 6; beta = 0.74, SE = 0.61). I refrained from assessing how the number of each habitat type affected occupancy because the number of models in the model set was already high relative to the number of sections sampled.

It is important to note that the metric of habitat area (i.e., number of habitat polygons within each section) was relatively coarse. I recommend using a GIS to quantify the total area of polygons representing ponds, hydric soils, and sandy soils. In addition to examining effects of habitat area on occupancy, further studies should also address the effects of habitat quality (e.g., hydroperiod), habitat configuration (e.g., connectivity among habitat polygons) and other landscape features (e.g., road density) on occupancy. Habitat area, quality, and configuration can both be important for affecting occupancy and extinction-colonization dynamics of pond-breeding amphibians (e.g., Cosentino et al. 2011), and road disturbance has been shown to consistently constrain the distribution of anurans (Cosentino et al., unpublished manuscript).

There was some evidence for marginal variation in detection probability among sections and over time in 2011 and 2012. In 2011, models with number of listening posts, date of survey, temperature, and time of survey had competitive support. Based on beta estimates from the most-supported model with each detection covariate, detection probability was related negatively to number of listening posts (beta = -1.14, SE = 0.95), date of survey (beta = -0.96, SE = 1.05) and time of survey (beta = -0.71, SE = 1.22) and positively to temperature (beta = 1.39, SE = 1.90). In 2012, models with competitive support included wind, observer/region, temperature, time, date, and number of listening posts. Detection probability was lowest in M-M and greatest in Alexander (Fig. 7). Based on beta estimates from the most-supported model with each detection covariate, detection probability was related negatively to wind (Fig. 8; beta = -0.64, SE = 0.51), time of survey (beta = -0.30, SE = 0.45), date of survey (beta = -0.16, SE = 0.38), and positively to temperature (Fig. 9; beta = 0.61, SE = 0.49) and number of listening posts (beta = 0.11, SE = 0.33).

Out of all the detection covariates in 2011 and 2012, detection was most strongly related to observer/region in 2012. This inference is based on a comparison of model support for detection covariates in models with a consistent covariate for occupancy probability (i.e., number of habitats). Other detection covariates had only very weak effects, and some covariates likely represent pretending variables (e.g., date of survey and number of listening posts in 2012; Anderson 2008). The negative effect of the number of listening posts on detection on 2011 is unusual, but it is important to remember that detection was extremely high at all sections in 2011 regardless of the number of listening posts. Overall, the most consistent driver of variation in detection probability across all years was date of survey (Table 2).

It is important to note that the precision of occupancy and detection estimates varied strongly between 2011 and 2012. The lower precision in 2012 than in 2011 represented strong model selection uncertainty, meaning there was considerable variation in average occupancy and detection estimates among models. The variation in average occupancy and detection probabilities was mainly associated with models that included an effect of observer/region on detection probability. For these models, detection estimates were extremely low in M-M, which resulted in occupancy estimates that were very high. Occupancy estimates are known to be biased high when detection probability is low and when sample size is low (Guillera-Arroita et al. 2010). Increasing the number of sections sampled within each region should reduce this bias.

3.3. Optimal survey design during a single season

Three patterns emerged from an evaluation of the optimal number of surveys to perform at each section during a season. First, the optimal number of surveys to minimize variance of the occupancy estimator depended on detection probability. In years when detection probability was average to high ($p \geq 0.7$), conducting two surveys at each section generally minimized the variance of the occupancy estimate (Table 5). When $p = 0.6$, the optimal number of surveys was three. This makes intuitive sense, as it takes more surveys to determine the occupancy status of a section with a high degree of confidence when detection probability is low compared to when detection probability is high. These results are generally consistent across all levels of survey effort evaluated. Second, the variance of the occupancy estimator decreased as total survey effort increased regardless of the number of repeated surveys or detection probability (Table 5). This was expected because survey effort is allocated into sampling additional sections as total survey effort increases. Third, for a given survey effort and number of repeated surveys, variance of the occupancy estimator decreased as detection probability increased. However, it is important to note this decrease was marginal when the number of repeated surveys was four.

If a primary goal of monitoring is to minimize the variance of the occupancy estimator, I recommend sections be surveyed 2-3 times per year. When detection probability is expected to be average to high (≥0.70), I recommend surveying sections only twice each season, placing more effort into surveying additional sections rather than surveying each section more than twice for a fixed level of survey effort. When detection probability is expected to be low (e.g., during drought years), I recommend that all sections be surveyed three times without reducing the number of sections surveyed. However, this requires increasing total survey effort during years when detection probability is low. If total survey effort cannot be increased, I recommend one of two options.

First, a removal design can be used in which all sections are surveyed until *P. illinoensis* is detected, up to a maximum of three repeat surveys at each section. Once *P. illinoensis* is detected at a section, no more surveys are conducted at the section. The advantage of this approach is that no more effort is put into surveying a section that is known to be occupied, and the effort saved can be used to sample other sections up to three times. A disadvantage is that detection probability cannot be modeled as survey-specific under a complete removal design (MacKenzie et al. 2006). However, detection can still be modeled using covariates (e.g., Julian date). MacKenzie and Royle (2005) note that a hybrid design may be used in which some sections are surveyed using a standard design and some sections are surveyed using a removal design. If a hybrid approach is used, sections should be randomly assigned to the type of design.

Second, sections can continue to be surveyed only twice. When detection probability is low (p = 0.6) and survey effort is high (TS > 120), the results of the simulations show that RMSE is ~8% greater when K = 2 compared to when K = 3. If some loss of precision is acceptable, it may be easiest to use a consistent survey design even when detection probability is low. This approach may also be the most practical solution because 1) it can be difficult to reliably estimate detection probability before or during a sampling season in order to make decisions about sample design, and 2) coordinating substantial changes in survey design from year-to-year may not be logistically feasible.

3.4. Power analysis

The power to detect a proportional decline in occupancy between two time periods increased as detection probability increased from 0.6 to 0.9 (Figs. 10, 11). These results were consistent at different levels of total survey effort, number of repeat surveys, and for different levels of proportional decline in occupancy. Power was generally similar for K = 2 and K = 3 when detection probability was 0.6, but conducting two surveys always resulted in the greatest power when detection probability was ≥ 0.7 (Figs. 10, 11) . This corroborates the recommendations for single-season study design by showing that conducting two surveys per section is as good or better for detecting a decline in occupancy as three surveys per section.

Power to detect a decline increased as the proportional decline in occupancy increased (Figs. 10-11). Power to detect a 10-30% decline in occupancy between two time periods was generally low regardless of total survey effort, number of repeat surveys, and detection probability (Figs. 10, 11). Increasing power to detect declines ≤ 30% would take substantial effort. For example, assume p = 0.8 and α = 0.05. It would take around 1750 sections each surveyed twice per year for power to detect a 10% decline to reach 0.8 (approximated by simulation), which is a commonly used target in monitoring programs (Gibbs et al. 1998; Guillera-Arroita and Lahoz-Monfort et al. 2012). Under the same scenario, it would take around 186 sections to detect a 30% decline. This is consistent with a previous simulation study showing monitoring programs that use occupancy data typically have low power to detect declines of <20-50% (Strayer 1999).

As expected, power increased when the significance value increased (Figs. 12, 13). Thus, fewer sections would be required to detect a given level of decline if a greater risk of Type I error is acceptable (Figs. 12, 13).

4. **Recommendations for a long-term monitoring plan**

4.1. Number of sections and surveys at each section

I recommend a standard sampling design in which 75–90 sections are each sampled twice during each year. Conducting two surveys at each section per year generally maximizes precision of the occupancy estimator and power to detect a decline in occupancy. If each section is surveyed twice each year, survey effort is 150–180 total surveys each year. Surveying 75–90 sections twice per year leads to a reasonably high probability of detecting 40-50% declines when α = 0.05 (Table 6). See section 4.7 for a discussion of achieving greater power by raising the risk of Type I error.

If possible, sections should be surveyed three times each during drought years or more generally when detection probability is expected to be low. If surveying sections three times each compromises the number of sections that can be visited, I recommend a removal design at all sections for a maximum of three surveys. Alternatively, all sections can be surveyed twice even in drought years if it is more logistically feasible to use a consistent sampling design and a moderate loss of precision of the occupancy estimator is acceptable.

4.2. Selecting sections

If the goal is to determine the statewide status of *P. illinoensis* in suitable habitat, I recommend sampling sections within each of the five regions in which suitable habitat occurs. To ensure adequate representation of each region, a stratified random sampling design should be used to select sampling units, where each stratum represents a region. The number of sections per region should be allocated proportionally to the amount of suitable habitat available within the region, although I recommend a minimum of 5–10 sections per region to facilitate modeling regional variation in occupancy. Sections within each region should be selected randomly. Because *P. illinoensis* chorus can be heard from up to 1.3 miles (Brown and Rose 1988), I recommend that the borders of sections be separated by ≥ 1.3 miles. Ensuring sections are separated by this minimum distance will help to maximize biological and statistical independence among sections, which is an assumption of occupancy models.

4.3. Defining a season

A season is defined as the calling season of *P. illinoensis* within a single year. Surveys should be initiated as soon as possible in late winter when males begin chorusing. Detection probability was greatest early in the season compared to late in the season. I recommend surveys be completed by mid-April, with the possible exception being in years when *P. illinoensis* becomes active unusually late.

4.4. Preventing heterogeneity in detection probability

To ensure statistical independence among surveys at a section (an assumption of occupancy models), only a single survey should be conducted at a section during a given night. To prevent introducing heterogeneity in detection probability among sections, all sections should be surveyed a single time before a second survey at a section begins. If possible, observers should be rotated among sections so that no section is surveyed by only one observer. If this is not possible, observers should survey a minimum of 10 sections to facilitate modeling variation in detection probability, and the total number of observers should be kept to a minimum.

4.5. Maximizing detection probability during surveys

I recommend surveys begin at least 30 minutes after sunset and end by midnight. Peak chorusing ends by midnight for most anurans (Bowers et al. 1998). At each listening post within a section, observers should listen until *P. illinoensis* is heard or for a standard five minutes. Observers should record the number of minutes spent at each listening post. At the first and last survey within each section, I recommend surveyors record a variety of abiotic conditions that are known to affect anuran calling behavior (Dorcas et al. 2010): air temperature, humidity, wind speed, and presence of moonlight. At each listening post, I recommend surveyors record the number of cars that passed by during the survey, and whether or not anthropogenic noise interfered with detecting anurans during the survey. If possible, observers should also record whether standing water is visible from each listening post. These variables can be used as covariates to model variation in detection probability. A data sheet is included at the end of this report.

Surveys should not be conducted under certain conditions when detection of anurans is expected to be minimal: when ambient temperature is too cold (e.g., <42 F, the minimum temperature observed during the pilot seasons), there is heavy rainfall, or wind is above 4 using the Beaufort Wind Code (Dorcas et al. 2010).

4.6. Statistical test for detecting differences in occupancy between time periods

I recommend using a Wald test on the probability scale to test the null hypothesis that occupancy probability is not different between time periods. A Wald test on the probability scale was shown to be more powerful test than the Wald test on the logistic scale and a likelihood-ratio test (Guillera-Arroita and Lahoz-Monfort 2012). Occupancy probabilities can be considered to be significantly different between time periods if

$$\frac{|\hat{\psi}_1 - \hat{\psi}_2|}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} > z_{\alpha/2},$$

where $\hat{\psi}_1$ and $\hat{\psi}_2$ are the estimates of occupancy in time periods one and two, respectively, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the estimated variances for occupancy in time periods one and two, respectively, and $z_{\alpha/2}$ is the critical value for the standard normal distribution for a given α (Guillera-Arroita and Lahoz-Monfort 2012). Testing for differences in occupancy probability can be done for any pair of seasonal occupancy estimates.

For example, the following equation shows the test statistic for a difference in occupancy probability between 2011 and 2012:

$$\frac{|0.56 - 0.50|}{\sqrt{0.10^2 + 0.17^2}} = 0.304$$

The *P*-value for a Wald test with a test statistic of 0.304 is 0.76, so we fail to reject the null hypothesis that occupancy probability was different between 2011 and 2012 at a significance value of 0.05.

Alternative statistical approaches include using a multi-season occupancy model to estimate and compare extinction and colonization probabilities, test for year-specific variation in occupancy, or use a regression approach to test for a trend in seasonal occupancy estimates (e.g., Weir et al. 2009). When making conclusions about changes in occupancy across seasons, it may be wise to assess the consistency of results from multiple analytical approaches.

4.7. Effect size, statistical significance, and listing decisions

Decisions to upgrade or downgrade the listing of *P. illinoensis* could be informed by the magnitude of occupancy changes over time. Under the recommended design, detecting a change in occupancy of 10-30% is unlikely when α = 0.05. Power could be increased by increasing the number of sections surveyed (which would increase precision of the occupancy estimator) or by decreasing the significance value (α). In the context of monitoring, Type II error is typically of more concern than Type I error because the price of missing a decline in occupancy or abundance can mean extinction (Gibbs et al. 1998, Field et al. 2005, Guillera-Arroita and Lahoz-Monfort 2012). In the spirit of the precautionary principle, the significance value could be relaxed to greater values, which leads to an increase in power (see Table 6 for examples using the recommended design). If $\alpha = 0.20$ under the current design, there is a reasonably high probability of detecting a decline of 30%, and a strong possibility of detecting declines of 40-50% (Table 6). The risk is that there is a 0.20 probability of concluding there is a decline when in fact there is not a decline.

Although declines of 30-50% are high, it is important to remember that smaller declines over short time periods may reflect stochasticity rather than true trends. It may be wise to consider the consistency of trends over long periods of time (e.g., 3-5 years) rather than year-to-year changes.

4.8. Adapting this monitoring protocol

Analyses on optimal allocation of survey effort and power are exploratory, and the results depend strongly on assumptions about parameter values and target effect sizes. The values of occupancy and detection probability used in this report are based on pilot data, but survey design and power analyses assume occupancy and detection are constant over time. I recommend conducting occupancy analyses at three-year intervals to ensure average detection probabilities are consistent with the detection probabilities observed during the pilot season. Survey allocation may need to be adapted if occupancy and detection probabilities change drastically.

5. Literature Cited

Adams, M.J., D.A.W. Miller, E. Muths, P.S. Corn, E.H.C. Grant, L.L. Bailey, G.M. Fellers, R.N. Fisher, W.J. Sadinski, H. Waddle, and S.C. Walls. 2013. Trends in amphibian occupancy in the United States. PLoS One 8(5):e64347.

Anderson, D.R. 2008. Model Based Inference in the Life Sciences: A Primer on Evidence. Springer, New York, New York, USA.

Bailey, L.L., and J.D. Nichols. 2010. Capture-mark-recapture, removal sampling, and occupancy models. In Dodd, Jr., C.K., editor, Amphibian Ecology and Conservation: A Handbook of Techniques. Oxford University Press, Oxford, UK.

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

Beltz, E. 1993. Distribution and status of the Illinois chorus frog, *Pseudacris streckeri illinoensis*, in Cass, Menard, Morgan, and Scott Counties of west-central Illinois. Illinois Department of Conservation, Springfield, Illinois, USA.

Bowers, D.G., D.E. Anderson, and N.H. Euliss, Jr. 1998. Anurans as indicator of wetland condition in the Prairie Pothole Region of North Dakota: an environmental monitoring and assessment program pilot project. In Lannoo, M., editor. Status and Conservation of Midwestern Amphibians. University of Iowa Press, Iowa City, Iowa., USA.

Brown, L.E., and G.B. Rose. 1988. Distribution, habitat, and calling season of the Illinois Chorus Frog (*Pseudacris streckeri illinoensis*) along the lower Illinois River. Illinois Natural History Survey Biological Notes 132.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-theoretic Approach. Second Edition. Springer, New York, New York USA.

Cosentino, B.J., R. L. Schooley, and C. A. Phillips. 2011. Spatial connectivity moderates the effect of predatory fish on salamander metapopulation dynamics. Ecosphere 2(8):art95.

Cosentino, B.J., D.M. Marsh, K.S. Jones, J.J. Apodaca, C. Bates, J. Beach, K.H. Beard, K. Becklin, J.M. Bell, C. Crockett, G. Fawson, J. Fjelsted, E.A. Forys, K.S. Genet, M. Grover, J. Holmes, K. Indeck, N.E. Karraker, E. Kilpatrick, T.A. Langen, S.G. Mugel, A. Molina, J.R. Vonesh, R. Weaver, and A. Willey. Roads negatively affect frog and toad distributions throughout the eastern and central United States. Unpublished manuscript.

Dorcas, M.E., S.J. Price, S.C. Walls, and W.J. Barichivich. 2010. Auditory monitoring of anuran populations. In Dodd, Jr., C.K., editor. Amphibian Ecology and Conservation: A Handbook of Techniques. Oxford University Press, Oxford, UK.

Field, S.A., A.J. Tyre, and H.P. Possingham. 2005. Optimizing allocation of monitoring effort under economic and observational constraints. Journal of Wildlife Management 69:473–482.

Fiske, I., and R.B. Chandler. 2011. *unmarked*: An R package for fitting hierarchical models of wildlife occurrence and abundance. Journal of Statistical Software 43:1–23.

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

Guillera-Arroita, G., and J.J. Lahoz-Monfort. 2012. Designing studies to detect differences in species occupancy: power analysis under imperfect detection. Methods in Ecology and Evolution 3:860–869.

Guillera-Arroita, G., M.S. Ridout, and B.J.T. Morgan. 2010. Design of occupancy studies with imperfect detection. Methods in Ecology and Evolution 1:131–139.

Hinz, L., A. Hulin, and A.M. Holtrop. 2011. Using a landscape approach to refine conservation plans for Illinois Chorus Frog (*Pseudacris streckeri illinoensis*). Illinois Department of Natural Resources, Springfield, IL.

Illinois Department of Natural Resources. 2005. Illinois comprehensive wildlife conservation plan-strategy. Illinois Department of Natural Resources, Springfield, Illinois, USA.

Illinois Department of Natural Resources. 2013. The drought of 2012: a report of the Governor's drought response task force. Illinois Department of Natural Resources, Springfield, Illinois, USA.

MacKenzie, D.I. 2005. What are the issues with presence-absence data for wildlife managers? Journal of Wildlife Management 69:849–860.

MacKenzie, D.I., and J.A. Royle. 2005. Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology 42:1105–1114.

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

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

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Persistence. Academic Press, Burlington, Massachusetts, USA.

Mattfeldt, S.D., L.L. Bailey, and E.H.C. Grant. 2009. Monitoring multiple species: estimating state variables and exploring the efficacy of a monitoring program. Biological Conservation 142:720–737.

Pellet, J., and B.R. Schmidt. 2005. Monitoring distributions using call surveys: estimating site occupancy, detection probabilities and inferring absence. Biological Conservation 123:27–35.

Petranka, J.W., C.K. Smith, and A.F. Scott. 2004. Identifying the minimal demographic unit for monitoring pond-breeding amphibians. Ecological Applications 14:1065–1078.

R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sewell, D., G. Guillera-Arroita, R.A. Griffiths, and T.J.C. Beebee. 2012. When is a species declining? Optimizing survey effort to detect population changes in reptiles. PLoS One 7(8):e43387.

Sjögren Gulve, P. 1994. Distribution and extinction patterns within a northern metapopulation of the pool frog, *Rana lessonae*. Ecology 75:1357–1367.

Strayer, D.L. 1999. Statistical power of presence-absence data to detect population declines. Conservation Biology 13:1034–1038.

Trauth, J.B., S.E. Trauth, and R.L. Johnson. 2006. Best management practices and drought combine to silence the Illinois chorus frog in Arkansas. Wildlife Society Bulletin 34:514–518.

Tucker, J.K. 1998. Status of the Illinois chorus frog in Madison County, Illinois. In Lannoo, M.J., editor. Status and Conservation of Midwestern Amphibians. University of Iowa Press, Iowa City, Iowa, USA.

Urquhart, N.S., and T.M. Kincaid. 1999. Trend detection in repeated surveys of ecological responses. Journal of Agricultural, Biological and Environmental Statistics 4:404–414.

Weir, L.A., J.A. Royle, P. Nanjappa, and R.E. Jung. 2005. Modeling anuran detection and site occupancy on North American Amphibian Monitoring Program (NAAMP) routes in Maryland. Journal of Herpetology 39:627–639.

Weir, L.A., I.J. Fiske, and J.A. Royle. 2009. Trends in occupancy from northeastern states of the North American Amphibian Monitoring Program. Herpetological Conservation and Biology 4:389–402.

Werner, E. E., K. L. Yurewicz, D. K. Skelly, and R. A. Relyea. 2007. Turnover in an amphibian metacommunity: the role of local and regional factors. Oikos 116:1713–1725.

Table 1. Number of sections and surveys for *Pseudacris illinoensis* in three regions of Illinois from 2011-2014.

Region	Surveys	2011	2012	2013	2014
Alexander	1	0	0	0	0
	2	10	10	0	10
	3	0	0	0	0
C-M-S	1	2	0	0	0
	2	6	10	10	10
	3	0	0	0	0
M-M	1	1	0	0	0
	2	1	0	0	0
	3	8	10	0	0
	Total Sections	28	30	10	20

Table 2. Model selection statistics and estimates for models of initial occupancy (ψ), colonization (γ), extinction (ε), and detection (ρ) probabilities in 2011-2014 for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Main effects of detection probability include year, number of listening posts, date of survey, time of survey, temperature, wind, and observer/region. Dot (.) represents constant parameter estimates across space and time. Occupancy, colonization, and extinction probabilities were held constant in all models. Model selection statistics include difference between model AIC and AIC for the best model (ΔAIC), Akaike weight (ω), log likelihood (*LL*), and number of parameters (*K*). Model-averaged estimates were 0.56 (SE = 0.09) for initial occupancy, 0.12 (SE = 0.08) for colonization, 0.15 (SE = 0.07) for extinction, and 0.77 (SE = 0.05) for detection probability.

ψ	γ	ε	ρ	ΔAIC	ω	*LL*	*K*	$\hat{\psi}$ (*SE*)	$\hat{\gamma}$ (*SE*)	$\hat{\varepsilon}$ (*SE*)	$\hat{\rho}$ (*SE*)
.	.	.	Year + Date	0.00	0.461	-80.83	8	0.56 (0.09)	0.12 (0.08)	0.14 (0.06)	0.76 (0.05)
.	.	.	Year	2.55	0.129	-83.10	8	0.56 (0.09)	0.13 (0.08)	0.16 (0.07)	0.79 (0.05)
.	.	.	Year + Time	2.70	0.119	-82.18	7	0.56 (0.09)	0.12 (0.08)	0.15 (0.07)	0.79 (0.05)
.	.	.	Year + Observer/Region	3.11	0.097	-81.38	9	0.56 (0.09)	0.15 (0.10)	0.15 (0.06)	0.72 (0.05)
.	.	.	Year + Temp	3.57	0.077	-82.61	8	0.56 (0.09)	0.12 (0.08)	0.16 (0.07)	0.80 (0.05)
.	.	.	Year + Posts	4.49	0.049	-83.07	8	0.56 (0.09)	0.13 (0.08)	0.16 (0.07)	0.79 (0.05)
.	.	.	Year + Wind	4.51	0.048	-83.08	8	0.56 (0.09)	0.13 (0.08)	0.16 (0.07)	0.79 (0.05)
.	.	.	.	6.48	0.018	-88.07	4	0.55 (0.09)	0.17 (0.08)	0.21 (0.08)	0.83 (0.05)

Table 3. Model selection statistics and estimates for models of occupancy (ψ) and detection (ρ) probabilities in 2011 for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Main effects of occupancy probability include region and number of habitats within sections. Main effects of detection probability include number of listening posts, date of survey, time of survey, and temperature. Dot (.) represents constant parameters across space and time. Model selection statistics include difference between model AIC and AIC for the best model (ΔAIC), Akaike weight (ω), log-likelihood (*LL*), and number of parameters (*K*). Model-averaged parameter estimates were 0.56 (SE = 0.10) for occupancy and 0.97 (SE = 0.05) for detection.

ψ	ρ	ΔAIC	ω	*LL*	*K*	$\hat{\psi}$ (*SE*)	$\hat{\rho}$ (*SE*)
Region	.	0.00	0.187	-18.06	4	0.54 (0.08)	0.96 (0.04)
Region	Posts	0.18	0.171	-17.15	5	0.54 (0.08)	0.99 (0.02)
Region	Date	0.96	0.116	-17.54	5	0.54 (0.08)	0.97 (0.04)
Region	Temp	1.26	0.1	-17.69	5	0.54 (0.08)	0.96 (0.04)
Habitat	.	1.43	0.092	-19.78	3	0.60 (0.12)	0.96 (0.04)
Habitat	Posts	1.52	0.088	-18.82	5	0.60 (0.12)	0.99 (0.02)
Region	Time	1.64	0.083	-17.88	4	0.54 (0.08)	0.98 (0.04)
Habitat	Date	2.51	0.053	-19.32	4	0.60 (0.12)	0.97 (0.03)
Habitat	Temp	2.60	0.051	-19.36	4	0.60 (0.12)	0.96 (0.04)
Habitat	Time	3.11	0.04	-19.62	4	0.60 (0.12)	0.98 (0.03)
.	.	7.07	0.005	-23.60	2	0.54 (0.09)	0.96 (0.04)
.	Posts	7.25	0.005	-22.69	3	0.54 (0.09)	0.99 (0.02)
.	Date	7.99	0.003	-23.06	3	0.54 (0.10)	0.97 (0.04)
.	Temp	8.31	0.003	-23.22	3	0.54 (0.09)	0.96 (0.04)
.	Time	8.68	0.002	-23.41	3	0.54 (0.10)	0.98 (0.04)

Table 4. Model selection statistics and estimates for models of occupancy (ψ) and detection (ρ) probabilities in 2012 for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Main effects of occupancy probability include region and number of habitats within sections. Main effects of detection probability include number of listening posts, date of survey, time of survey, temperature, wind, and observer. Dot (.) represents constant parameters across space and time. Model selection statistics include difference between model AIC and AIC for the best model (ΔAIC), Akaike weight (ω), log-likelihood (*LL*), and number of parameters (*K*). Model-averaged parameter estimates were 0.50 (SE = 0.17) for occupancy and 0.63 (SE = 0.19) for detection.

ψ	ρ	ΔAIC	ω	*LL*	*K*	$\hat{\psi}$ (*SE*)	$\hat{\rho}$ (*SE*)
Region	.	0.00	0.175	-27.30	4	0.44 (0.08)	0.69 (0.12)
Region	Wind	0.35	0.147	-26.47	5	0.42 (0.08)	0.75 (0.11)
.	Observer/Region	0.36	0.146	-27.48	4	0.76 (0.18)	0.35 (0.07)
Region	Temp	0.53	0.134	-26.56	5	0.46 (0.09)	0.73 (0.12)
Habitat	Observer/Region	1.08	0.102	-26.84	5	0.59 (0.24)	0.43 (0.16)
Region	Time	1.57	0.080	-27.08	5	0.43 (0.08)	0.70 (0.12)
Region	Date	1.82	0.071	-27.21	5	0.44 (0.08)	0.67 (0.14)
Region	Posts	1.88	0.068	-27.24	5	0.44 (0.08)	0.68 (0.13)
Habitat	Wind	4.60	0.018	-29.60	4	0.41 (0.10)	0.76 (0.10)
Habitat	.	4.75	0.016	-30.68	3	0.42 (0.11)	0.73 (0.10)
Habitat	Time	5.87	0.009	-30.23	4	0.43 (0.11)	0.71 (0.11)
Habitat	Temp	6.10	0.008	-30.35	4	0.42 (0.11)	0.77 (0.11)
Habitat	Date	6.39	0.007	-30.49	4	0.43 (0.11)	0.68 (0.14)
Habitat	Posts	6.64	0.006	-30.62	4	0.42 (0.11)	0.71 (0.11)
.	Wind	8.14	0.003	-32.37	3	0.41 (0.09)	0.75 (0.11)
.	.	8.31	0.003	-33.46	2	0.42 (0.10)	0.72 (0.11)
.	Time	9.19	0.002	-32.89	3	0.43 (0.10)	0.70 (0.12)
.	Date	9.77	0.001	-33.18	3	0.44 (0.11)	0.65 (0.16)
.	Temp	9.78	0.001	-33.19	3	0.42 (0.09)	0.76 (0.11)
.	Posts	10.11	0.001	-33.35	3	0.43 (0.10)	0.70 (0.13)

Table 5. Root mean-squared errors (RMSE) for estimates of occupancy probability (ψ) under different study designs assuming $\psi = 0.5$ and multiple probabilities of detection ($\rho = 0.6–0.9$). Five levels of total number of surveys conducted in a season (TS) and three levels of repeated surveys conducted per section (K) were considered. RMSE was estimated with simulation using 50,000 iterations.

		K		
Parameter estimates	TS	2	3	4
$\psi = 0.5$, $\rho = 0.6$	60	0.153	0.128	0.134
	90	0.120	0.103	0.111
	120	0.099	0.088	0.094
	150	0.086	0.079	0.085
	180	0.077	0.071	0.077
$\psi = 0.5$, $\rho = 0.7$	60	0.118	0.117	0.130
	90	0.092	0.094	0.107
	120	0.078	0.082	0.092
	150	0.070	0.073	0.083
	180	0.062	0.067	0.075
$\psi = 0.5$, $\rho = 0.8$	60	0.099	0.113	0.130
	90	0.080	0.092	0.107
	120	0.069	0.079	0.091
	150	0.062	0.071	0.082
	180	0.057	0.065	0.074
$\psi = 0.5$, $\rho = 0.9$	60	0.093	0.112	0.129
	90	0.075	0.092	0.107
	120	0.066	0.079	0.091
	150	0.058	0.071	0.082
	180	0.053	0.065	0.075

Table 6. Power to detect a proportional decline (R) in occupancy probability (ψ) under the proposed sampling design (S = 75-90, K = 2) assuming ψ = 0.5 and an average detection probability (p = 0.77). Multiple significance levels were considered (0.05, 0.10, 0.20).

		Proportional decline				
S	α	10%	20%	30%	40%	50%
75	0.05	0.08	0.20	0.43	0.65	0.85
	0.10	0.16	0.31	0.53	0.75	0.92
	0.20	0.26	0.43	0.67	0.85	0.96
90	0.05	0.11	0.23	0.47	0.73	0.91
	0.10	0.17	0.36	0.61	0.83	0.94
	0.20	0.28	0.48	0.74	0.88	0.97

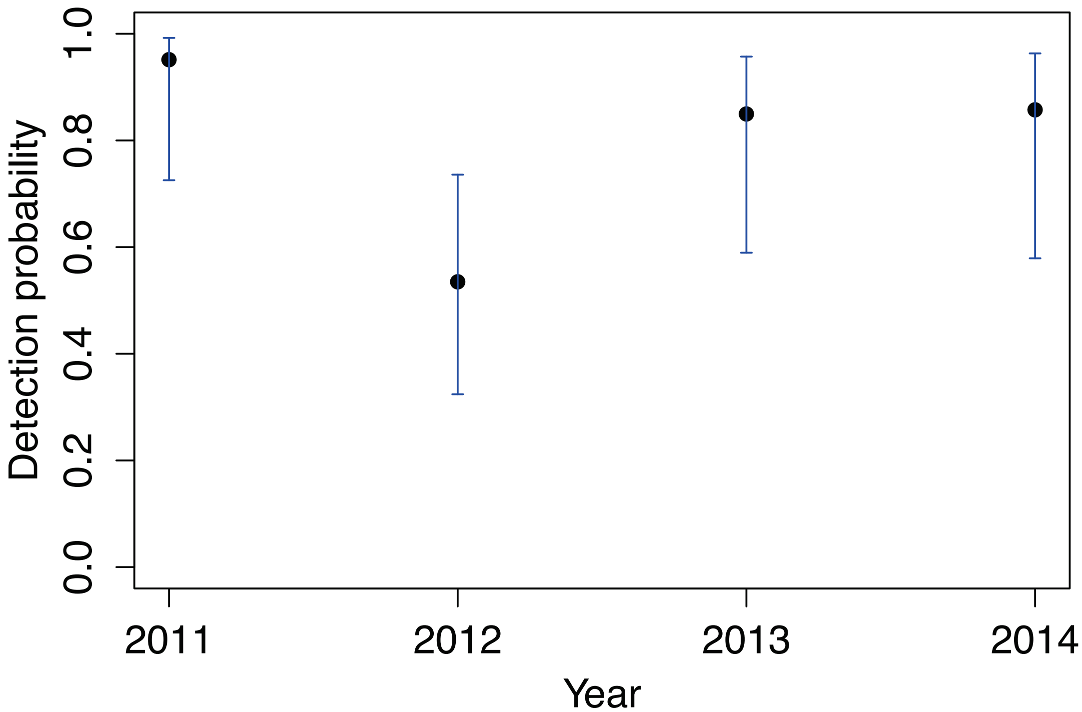

Figure 1. Predicted detection probability (filled circle) from 2011–2014 for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Error bars represent 95% confidence intervals.

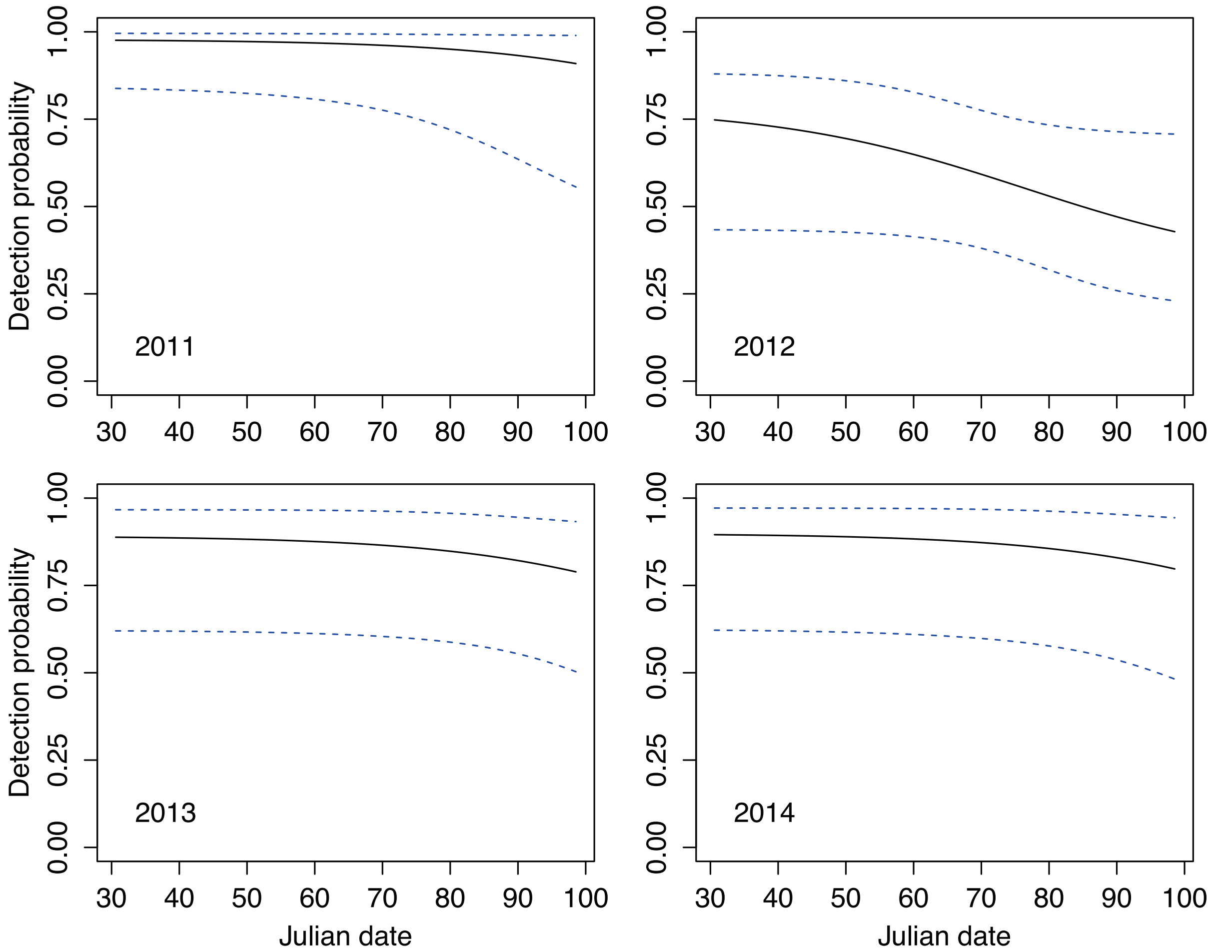

Figure 2. Relationship of predicted detection probability (solid lines) to Julian date from 2011–2014 for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Dotted lines represent 95% confidence intervals.

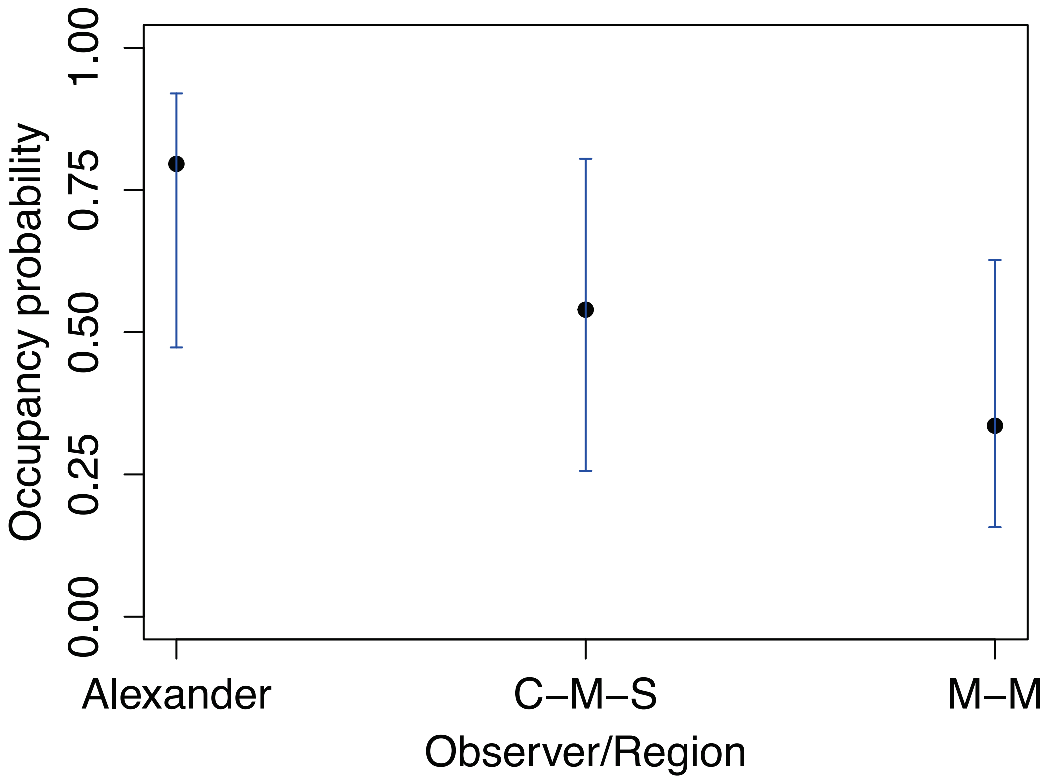

Figure 3. Relationship of predicted occupancy probability (filled circle) in 2011 to region for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Error bars represent 95% confidence intervals.

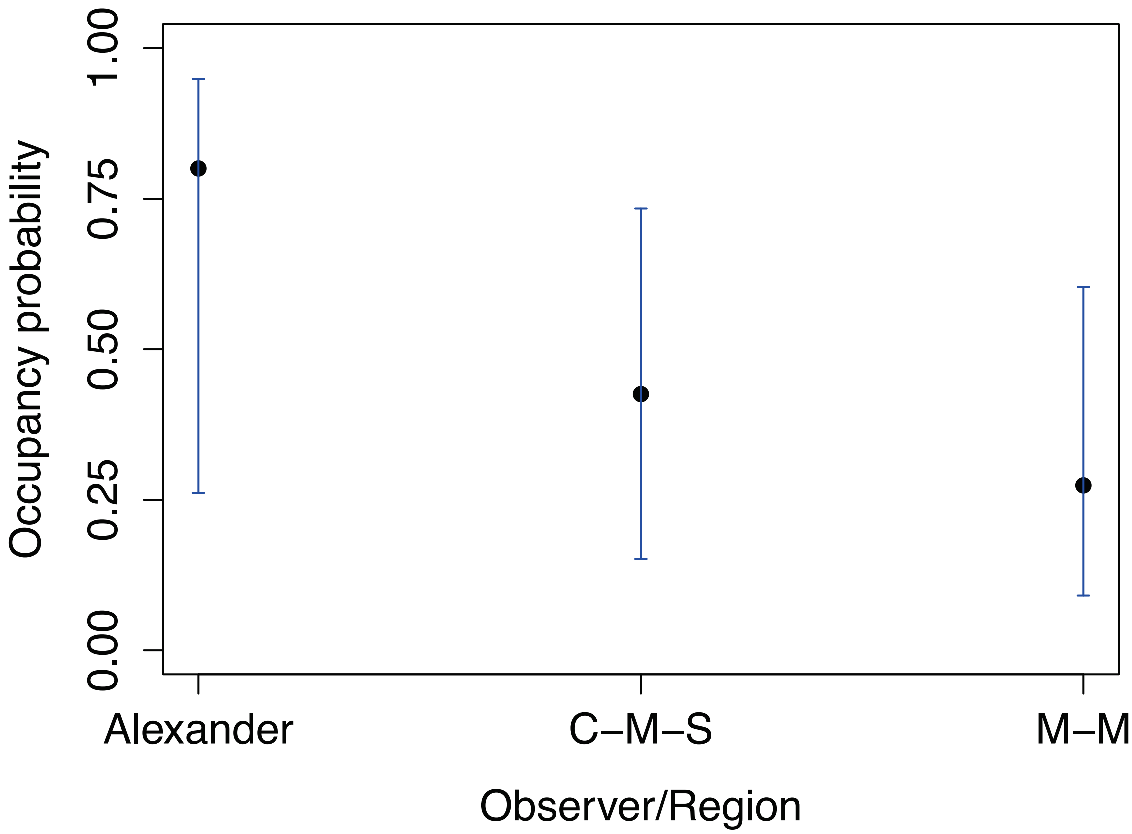

Figure 4. Relationship of predicted occupancy probability (filled circle) in 2012 to region for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Error bars represent 95% confidence intervals.

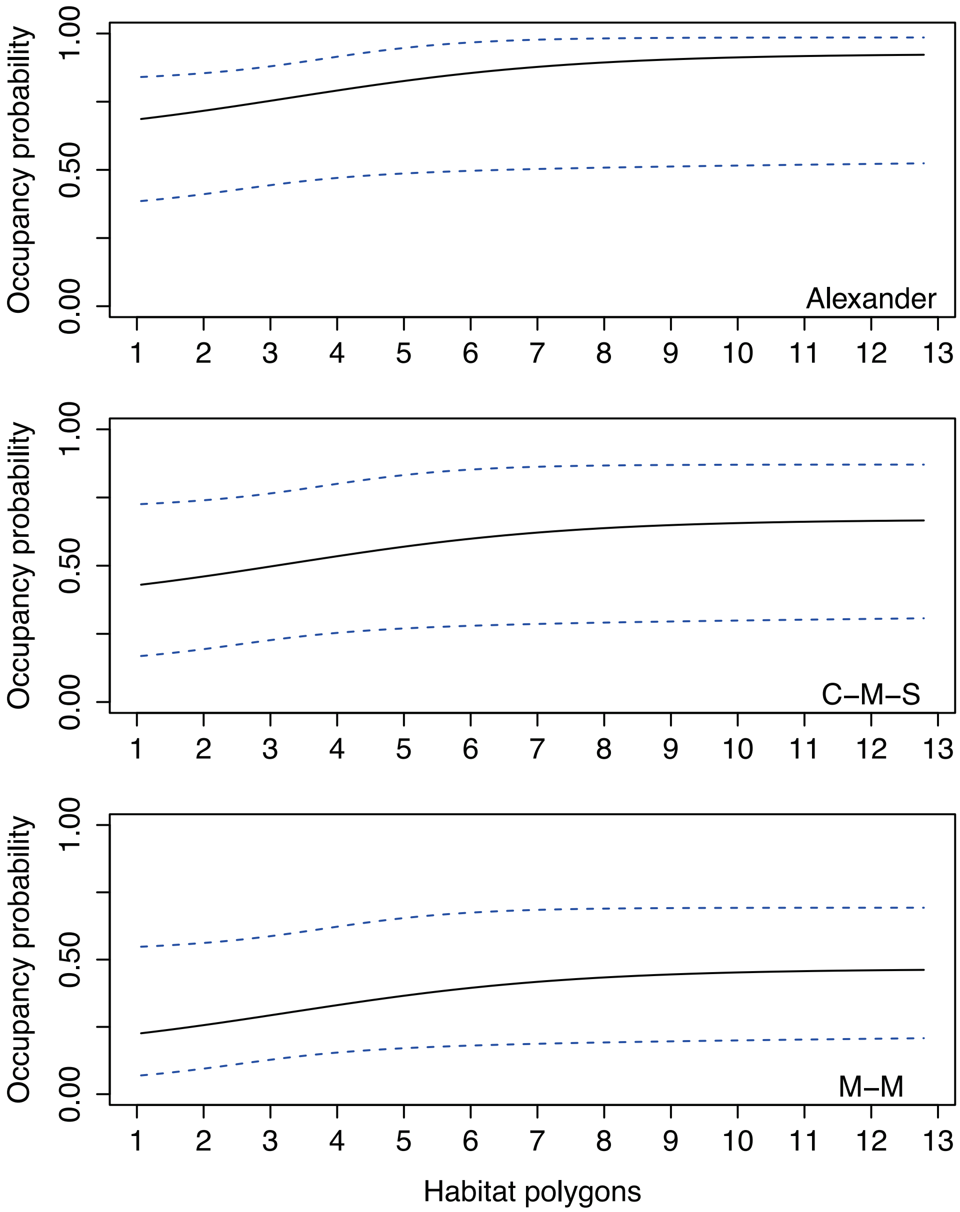

Figure 5. Relationship of predicted occupancy probability (solid lines) in 2011 to amount of habitat in each region for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Habitat represents the number of GIS polygons with ponds, sandy soils, or hydric soils within a section. Dotted lines represent 95% confidence intervals.

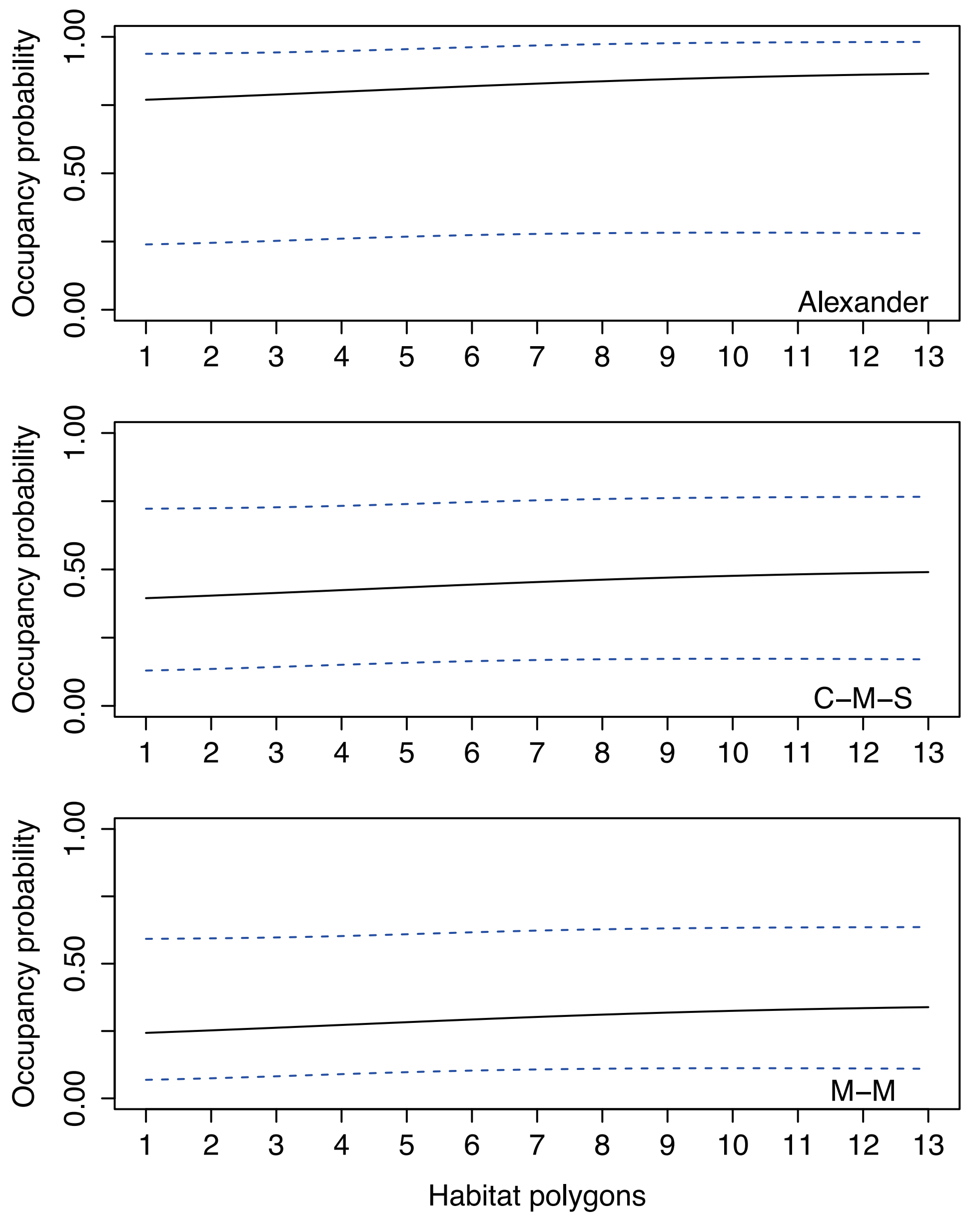

Figure 6. Relationship of predicted occupancy probability (solid lines) in 2012 to amount of habitat in each region for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Habitat represents the number of GIS polygons with ponds, sandy soils, or hydric soils within a section. Dotted lines represent 95% confidence intervals.

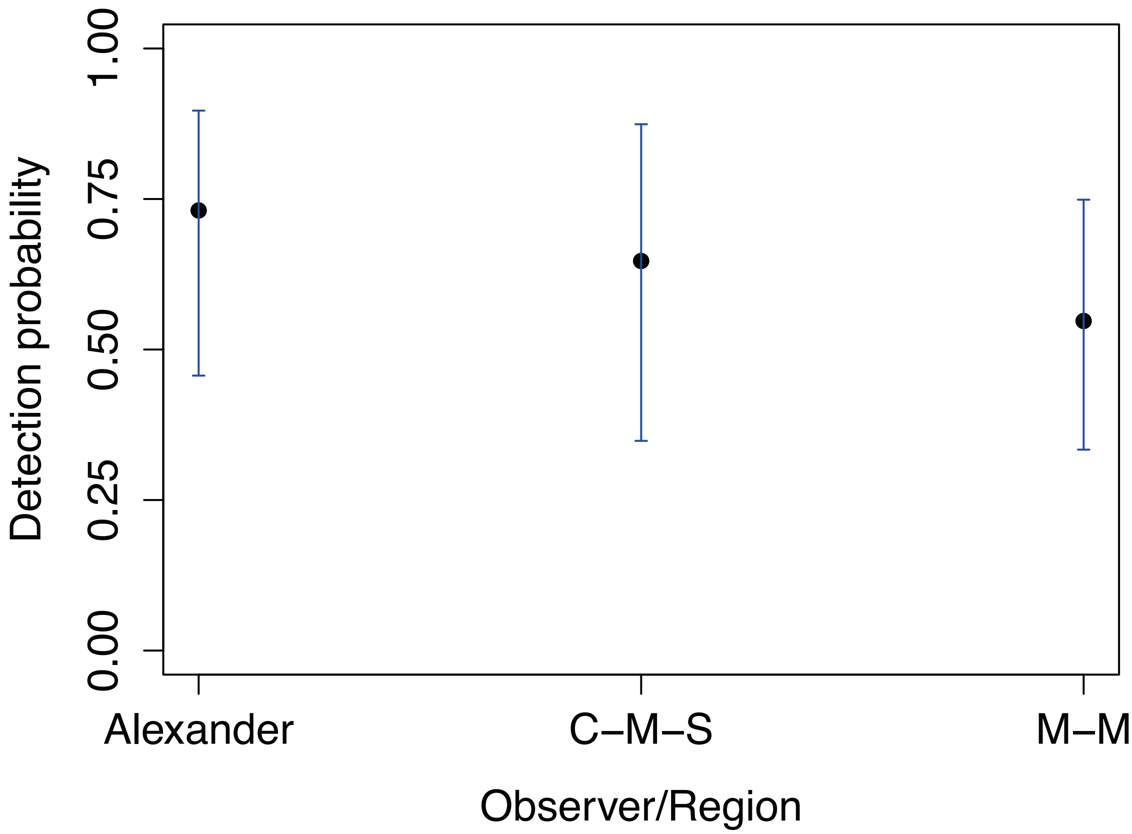

Figure 7. Relationship of predicted detection probability (filled circles) to observer and region in 2012 for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Observer was confounded with region. Error bars represent 95% confidence intervals.

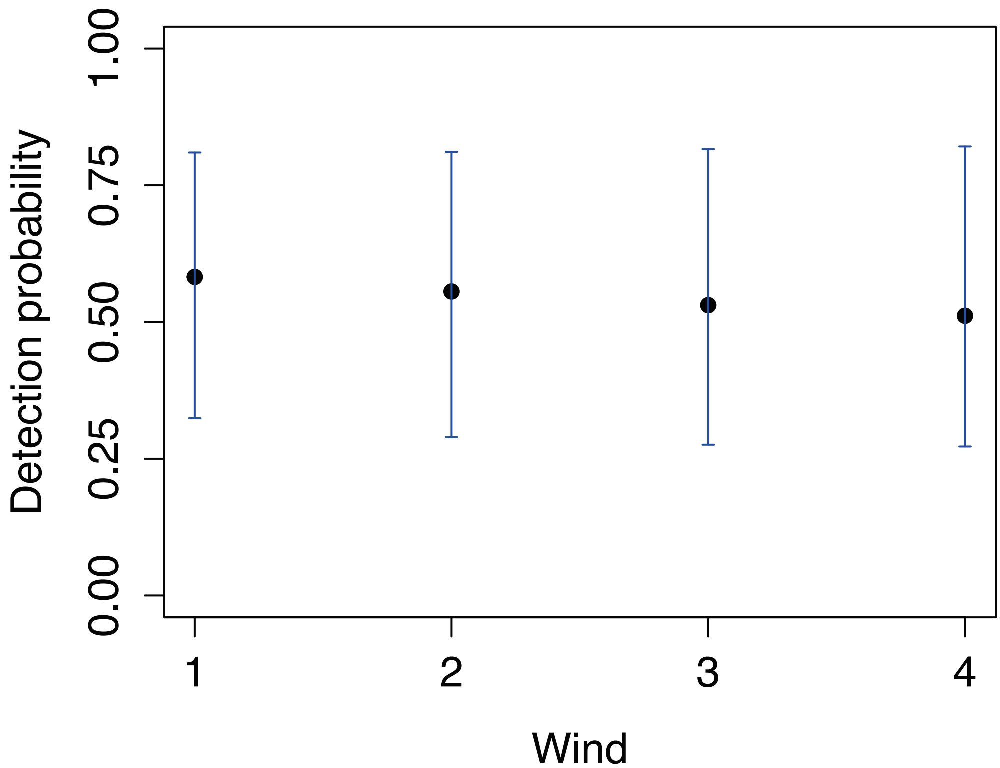

Figure 8. Relationship of predicted detection probability (filled circles) to wind during survey in 2012 for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Wind is on an ordinal scale from 1 (light wind) to 4 (moderate breeze). Error bars represent 95% confidence intervals.

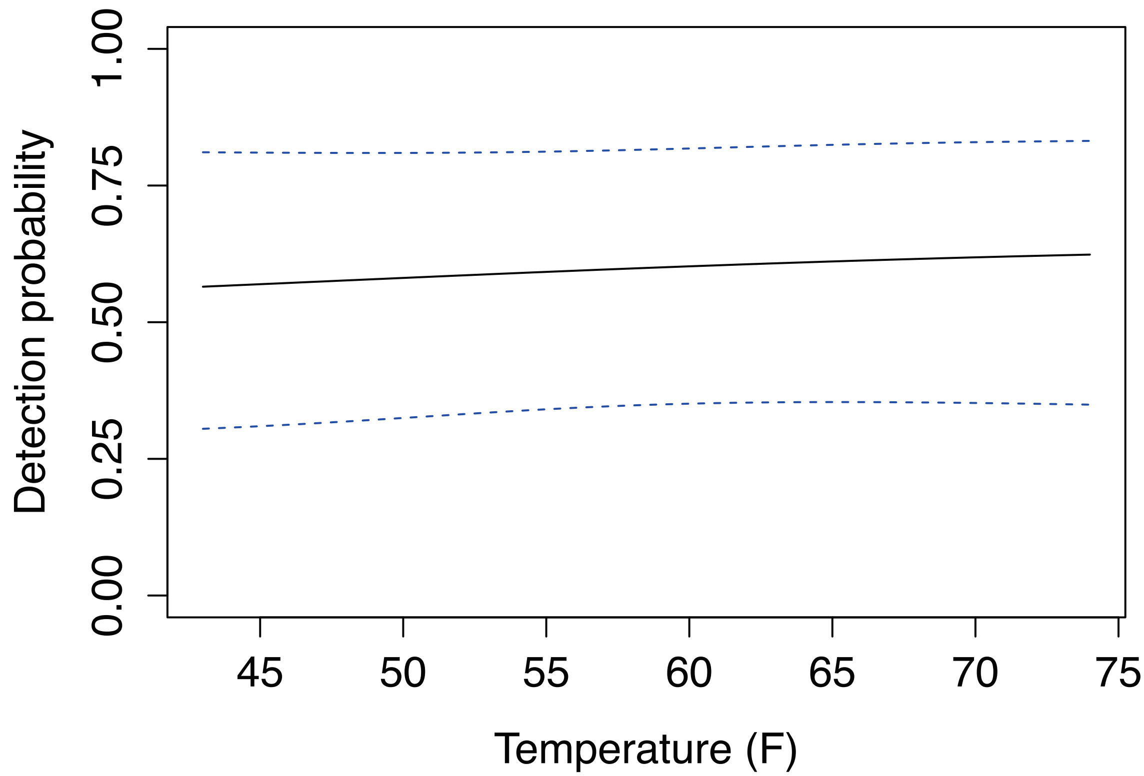

Figure 9. Relationship of predicted detection probability (solid line) in 2012 to temperature during survey for Illinois chorus frog (*Pseudacris illinoensis*) in Illinois. Dotted lines represent 95% confidence interval.

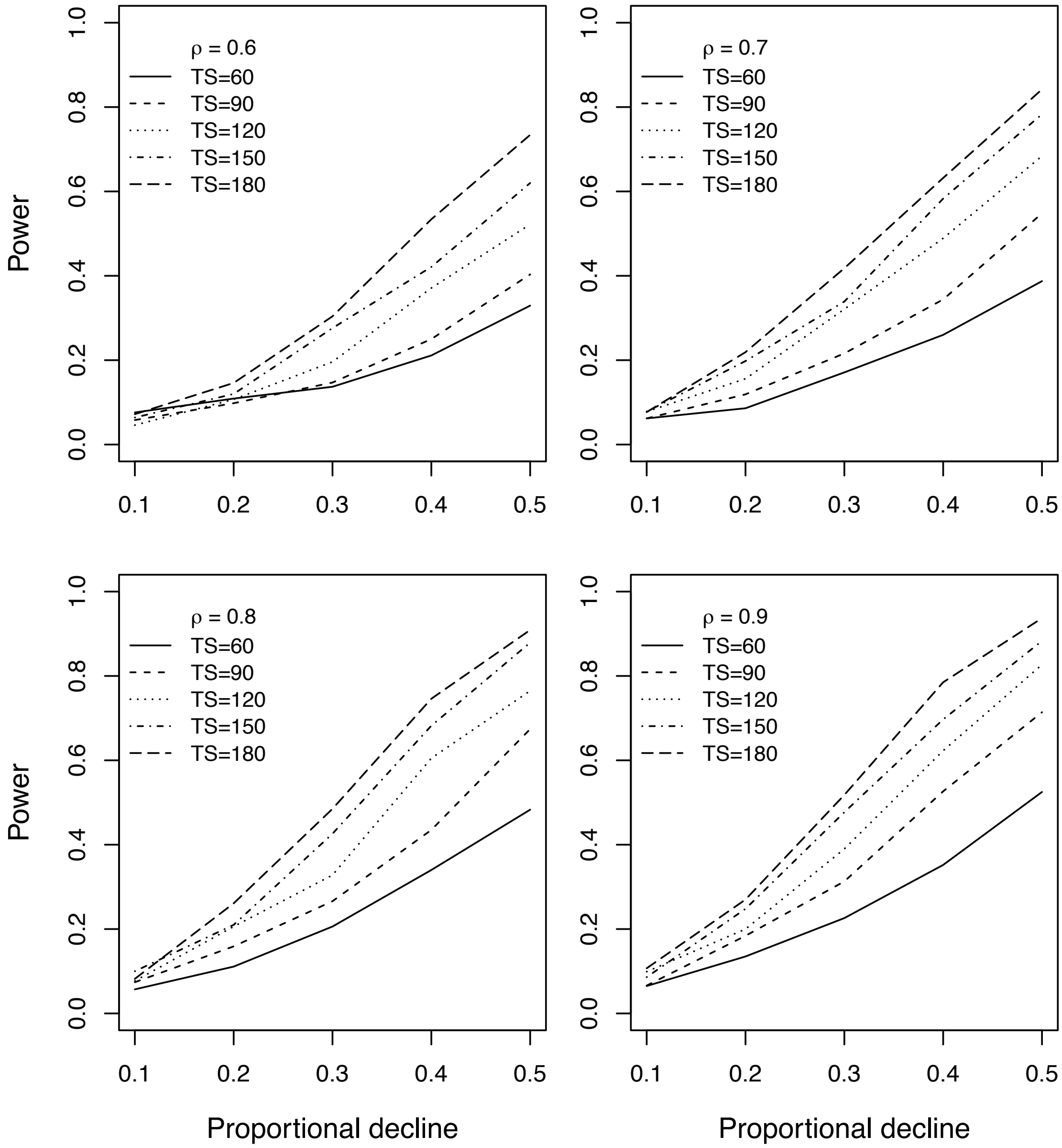

Figure 10. Power to detect a proportional decline (R = 0.1-0.5) of occupancy probability when two repeat surveys are conducted at each site assuming different levels of detection probability (p = 0.6-0.9) and total survey effort (TS = 60-180). Power was quantified using simulations. Initial occupancy probability was 0.5 in all simulations, and the significance level was set to 0.05.

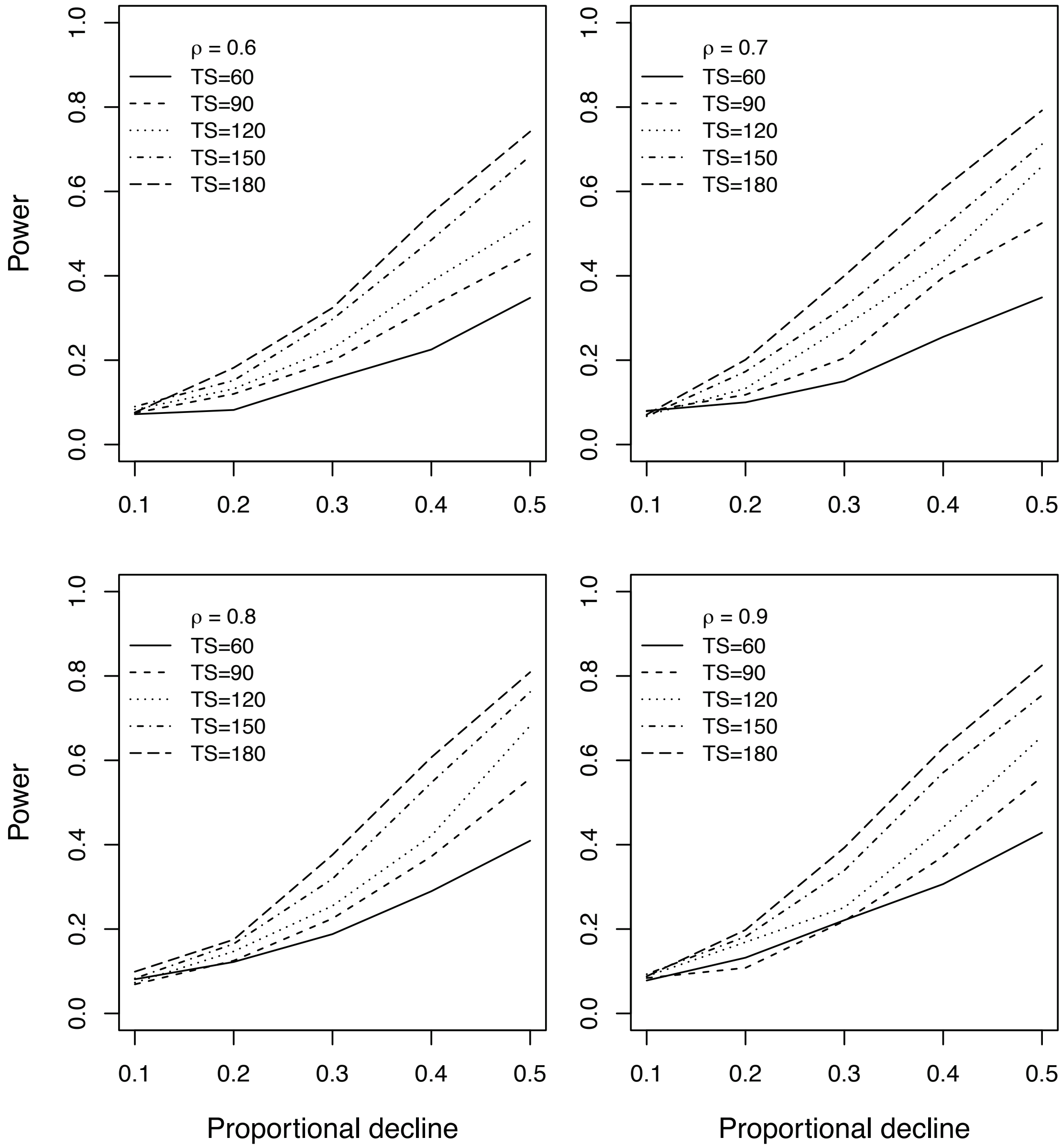

Figure 11. Power to detect a proportional decline (R = 0.1-0.5) of occupancy probability when three repeat surveys are conducted at each site assuming different levels of detection probability (*p* = 0.6-0.9) and total survey effort (TS = 60-180). Power was quantified using simulations. Initial occupancy probability was 0.5 in all simulations, and the significance level was set to 0.05.

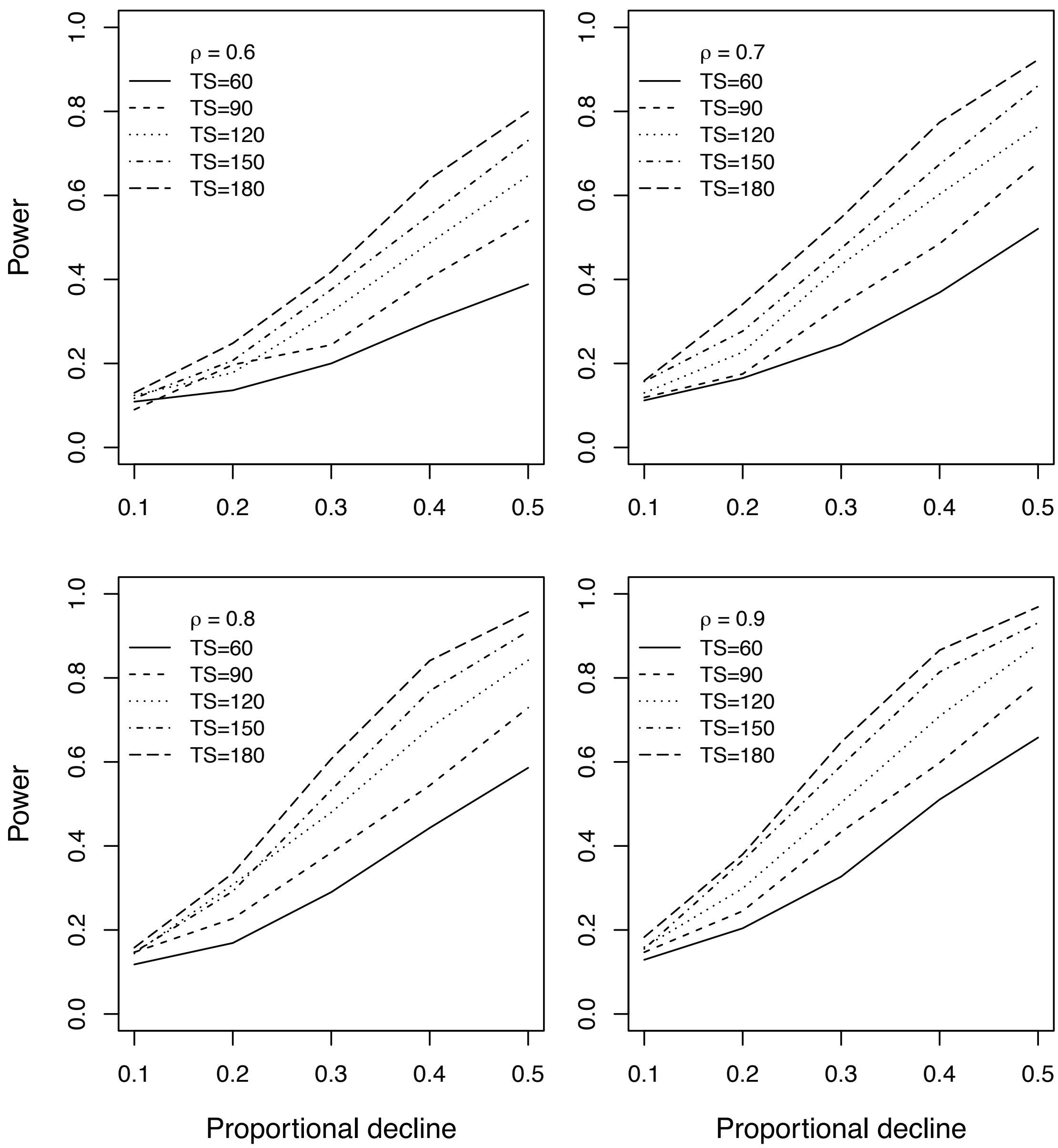

Figure 12. Power to detect a proportional decline (R = 0.1-0.5) of occupancy probability when two repeat surveys are conducted at each site assuming different levels of detection probability (p = 0.6-0.9) and total survey effort (TS = 60-180). Power was quantified using simulations. Initial occupancy probability was 0.5 in all simulations, and the significance level was set to 0.10.

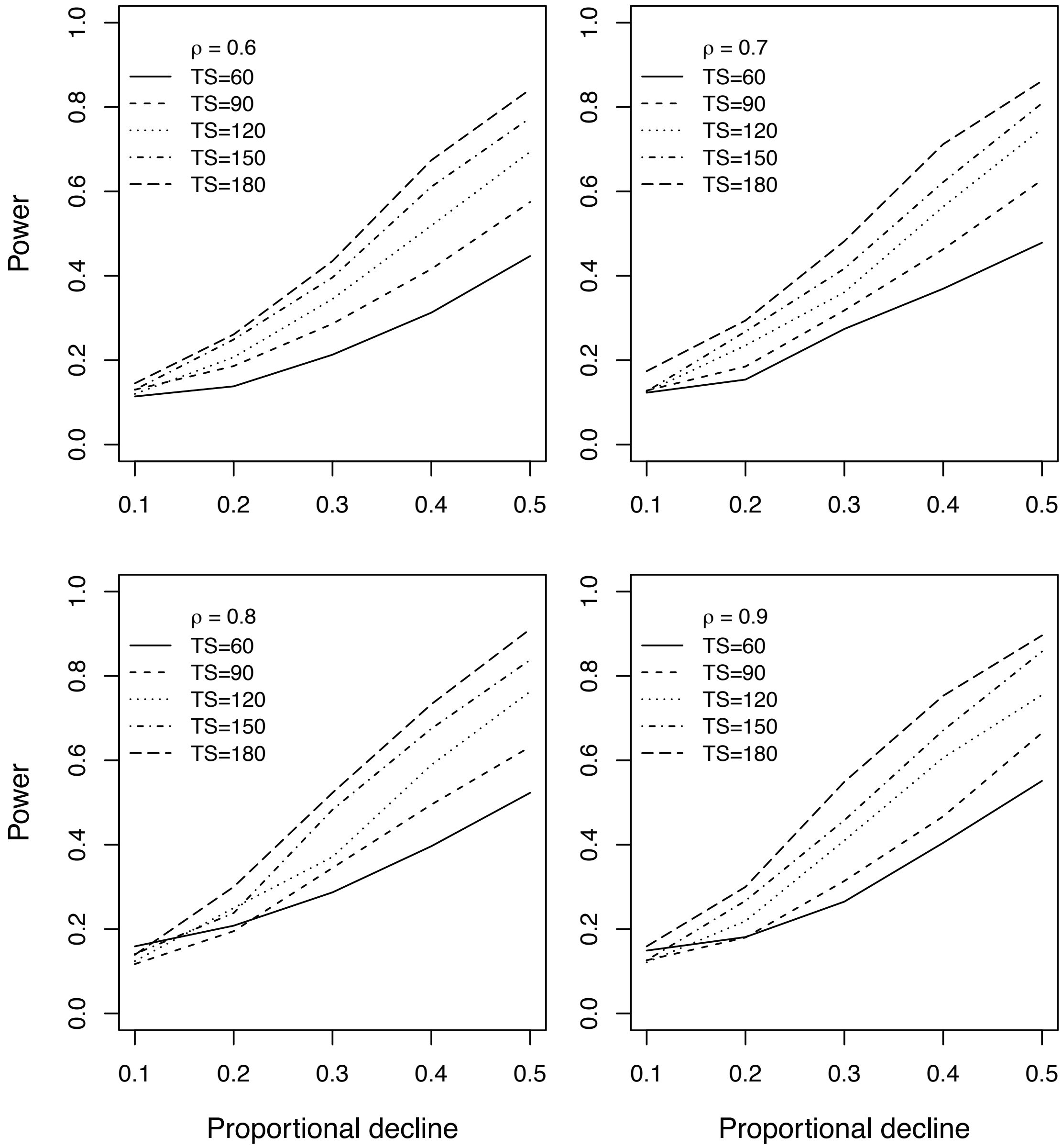

Figure 13. Power to detect a proportional decline (R = 0.1-0.5) of occupancy probability when three repeat surveys are conducted at each site assuming different levels of detection probability (p = 0.6-0.9) and total survey effort (TS = 60-180). Power was quantified using simulations. Initial occupancy probability was 0.5 in all simulations, and the significance level was set to 0.10.

Illinois Chorus Frog (*Pseudacris streckeri illinoensis*) Monitoring Datasheet

Date:	Observer:	Region:	Section:

Environmental variables at start and end of survey

Start	Time:	Air T (C or F):	Humidity (%):	Wind*: 1☐ 2☐ 3☐ 4☐ 5☐	Moon visible: Y☐ N☐
	Precipitation: ☐ None ☐ Light ☐ Mod ☐ Heavy		Other notes:		
End	Time:	Air T (C or F):	Humidity (%):	Wind*: 1☐ 2☐ 3☐ 4☐ 5☐	Moon visible: Y☐ N☐
	Precipitation: ☐ None ☐ Light ☐ Mod ☐ Heavy		Other notes:		

*Wind codes: 1 = calm (<1 mph), 2= light air (1-3 mph), 2 = light breeze (4-7 mph), 3 = gentle breeze (8-12 mph), 4 = moderate breeze (13-18 mph), 5 = fresh breeze (>18 mph; Do not conduct survey.)

Chorus survey data

Site data	**Listening post**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Minutes										
Polygon										
Time										
Noise (y/n)										
Water (y/n)										
# Cars										
Species	**Listening post**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
ICF										

Amphibian calling index:
0 = species not heard
1 = distinct calls of individuals can be counted; no overlapping calls
2 = individuals can be distinguished; some overlapping calls
3 = full chorus; constant and overlapping calls