Ecology, 91(2), 2010, pp. 341–346

Reports

Improved estimation of site occupancy using penalized likelihood

Monica Moreno[1] and Subhash R. Lele

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1 Canada

Abstract. When detection or occupancy probability is small or when the number of sites and number of visits per site is small, maximum likelihood estimators (MLE) of site occupancy parameters have large biases, are numerically unstable, and the corresponding confidence intervals have smaller than nominal coverage. We propose an alternative method of estimation, based on penalized likelihood. This method is numerically stable, the estimators have smaller mean square error than the MLE, and associated confidence intervals have close to nominal coverage.

Key words: penalized likelihood; species occurrence; zero inflated binomial.

Introduction

For decades, ecologists have measured the relative abundance of wildlife species to address various applied and theoretical ecological questions. However, the conclusions drawn by past studies that have used relative abundance measures are being increasingly questioned. Recognition that animals may be present at a sampling location but not detected has led to concerns that changes in relative abundance demonstrated in past studies may be due to changes in detectability and not the actual changes in animal abundance. The risk of such potential biases has led to an explosion of methods to account for imperfect detectability when estimating animal abundance or occupancy. One of the most common methods being used today by ecologists to account for imperfect detection when estimating animal occupancy rates is that of MacKenzie et al. (2002). The approach used by Mackenzie et al. (2002) requires that observers visit the same site multiple times. The rationale for this sampling approach is that by visiting the same site multiple times the factors influencing detection error can be separated from ecological factors influencing occupancy rate. A greater number of visits to a site will always improve the ability of occupancy models to estimate detectability as long as the surveys can be done over a relatively short period of time. The requirement of a short period between surveys is because occupancy models assume sites are "closed populations," meaning individuals are not born, do not die, or move out of or into a sampling area during the period of observation. In what circumstances this assumption is met is not clear for many species. This can create problems in designing sampling programs and modeling of subsequent data. This situation leads to an inherent trade-off for ecologists, as they have to balance the number of sites they visit with the number of revisits to the same sites often with insufficient a priori information to decide the appropriate balance. An additional problem in occupancy modeling that many ecologists are not familiar with is that the likelihood function used by MacKenzie et al. (2002) is ill behaved in several practical circumstances. For example, when the detection or occupancy probability is small, or when number of sites and number of visits per site is small, the maximum likelihood estimators (MLE) of site occupancy parameters tend to have large biases, are numerically unstable, and the corresponding confidence intervals tend to have smaller than nominal coverage. Thus, when ecologists choose a sampling design that has insufficient revisits or insufficient number of sites, they are left in a quandary as to how to analyze their data because the likelihood method gives very unstable and hence unusable results. In such situations, most ecologists have either ignored the problem and presented occupancy results as if they are robust estimates, or ignored the detection error and reverted to standard binary regression. We propose an alternative method of estimation, based on the penalized likelihood, that is numerically stable, leads to estimators that have smaller mean squared error than the MLE, and confidence intervals with coverage close to the nominal coverage. Our approach is intended to improve on that of Mackenzie et al. (2002) by allowing ecologists to utilize methods that account for detection error but that are more robust to issues caused by having limited number of visits or samples.

Likelihood and Penalized Likelihood Function

Let us assume that there are n study sites. The method proposed by MacKenzie et al. (2002) assumes that each site is visited k times. The number of visits need not be the same for every site. The visits are assumed to be independent of each other, and the sites are assumed to be independent of each other. Furthermore, it is assumed that a site that is occupied remains occupied

Manuscript received 18 June 2009; revised 10 September 2009; accepted 21 September 2009. Corresponding Editor: K. Newman.

[1] E-mail: mmoreno@ualberta.ca

throughout the study, and similarly the vacant sites remain vacant throughout the study.

Let ψ_i denote the probability of occupancy for the ith site. Let p_{ij} denote the probability of detecting the target species, given that it is present, for the ith site on the jth visit. These may depend on habitat characteristics and the conditions pertaining to the time of the visit. Let Y_{ij} denote the observed status, observed to be occupied (1) or observed to be unoccupied (0), of the ith site on the jth visit. A typical observation corresponding to a site is a sequence of 0's and 1's. The likelihood function in this general case is

$$L(\psi_i, p_{ij}/y_{ij}) = \prod_{i=1}^{n}\left[\psi_i\left(\prod_{j=1}^{k}(p_{ij})^{y_{ij}}(1-p_{ij})^{1-y_{ij}}\right) + (1-\psi_i)I(y_{i\cdot}=0)\right] \quad (1)$$

where $y_{i\cdot} = \Sigma_{j=1}^{k}\, y_{ij}$ and $I(\cdot)$ is an indicator function that is equal to one if $y_{i\cdot} = 0$ and 0 otherwise. If the probability of occupancy and probability of detection is constant, that is $\psi_i = \psi$ and $p_{ij} = p$, then the likelihood function reduces to

$$L(\psi, p/y_{i\cdot}) = \prod_{i=1}^{n}\left[\psi\left(\binom{k}{y_{i\cdot}}p^{y_{i\cdot}}(1-p)^{k-y_{i\cdot}}\right) + (1-\psi)I(y_{i\cdot}=0)\right]. \quad (2)$$

The results in Table 1 illustrate that the estimators based on maximizing this likelihood function can be quite unstable, with large biases, large standard errors, and incorrect coverage for the confidence intervals for common types of biological data.

On the other hand, suppose we ignore the detection error and base our estimation of the occupancy probability using $\tilde{Y}_i = \max_j\{Y_{ij}\}$. The likelihood function in this case is simply $\Pi_{i=1}^{n}(\psi_i)^{\tilde{Y}_i}(1-\psi_i)^{1-\tilde{Y}_i}$. The "naïve" estimator is obtained by maximizing this likelihood function. It can be seen from Table 1 that this estimator, for a small number of visits, can have large negative bias, but is extremely stable with small standard errors. If the number of visits to a site is large, the naïve estimator is a reasonable estimator assuming the assumption of closure is met.

The penalized likelihood that we propose combines the theoretical correctness of the MLE with the stability of the naïve estimator by considering

$$\ell_{\mathrm{P}}(\psi, p) = \left[\sum_{i=1}^{n}\ln\left(\psi\binom{k}{y_{i\cdot}}p^{y_{i\cdot}}(1-p)^{k-y_{i\cdot}} + (1-\psi)I(y_{i\cdot}=0)\right)\right] - \left[\lambda(k, p_0, n)f(\psi - \hat{\psi}_{\text{naïve}})\right]. \quad (3)$$

A complete explanation of how the penalty term was derived is presented in Appendix A. Roughly speaking, we shrink the MLE toward the naïve estimator ($\hat{\psi}_{\text{naïve}}$). The shrinkage factor, $\lambda(k, p_0, n)$, is determined by the number of visits, number of sites, and initial estimates of the average detection and occupancy probabilities.

Consider the regression setting where the occupancy probability for site i is related to the covariates x_1, x_2, $\ldots$, x_g,using the logistic link as

$$\psi_i = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_g x_{gi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_g x_{gi})}.$$

Similarly, the detection probability for the ith site in the jth survey is related to the covariates z_1, z_2, $\ldots$, z_m using the logistic link as

$$p_{ij} = \frac{\exp(\delta_0 + \delta_1 z_{1ij} + \cdots + \delta_m z_{mij})}{1 + \exp(\delta_0 + \delta_1 z_{1ij} + \cdots + \delta_m z_{mij})}.$$

In this case, the penalized likelihood estimator is computed as follows.

Step 1: Obtain the MLE for the detection parameters, δ_0, δ_1, $\ldots$, δ_m, their variances var($\hat{\hat{\delta}}_0$), var($\hat{\hat{\delta}}_1$), $\ldots$, var($\hat{\hat{\delta}}_m$), and the mean probability of detection, $\bar{\hat{p}}_0$, where

$$\bar{\hat{p}}_0 = \frac{1}{n \times k}\sum_{i=1}^{n}\sum_{j=1}^{k}\hat{p}_{ij}.$$

Step 2: Obtain the naïve estimator of the occupancy parameters $\tilde{\beta}_0$, $\tilde{\beta}_1$, $\ldots$, $\tilde{\beta}_{\mathrm{p}}$ and the mean estimated occupancy $\bar{\hat{\psi}}_{\text{naïve}} = 1/n\ \Sigma_{i=1}^{n}\ \hat{\psi}_i$ where

$$\hat{\psi}\mathrm{i} = \frac{\exp\left(\hat{\tilde{\beta}}_0 + \hat{\tilde{\beta}}_1 x_{1i} + \cdots + \hat{\tilde{\beta}}_g x_{gi}\right)}{1 + \exp\left(\hat{\tilde{\beta}}_0 + \hat{\tilde{\beta}}_1 x_{1i} + \cdots + \hat{\tilde{\beta}}_g x_{gi}\right)}.$$

Step 3: Maximize the penalized likelihood function using the following:

$$\ell_{\mathrm{P}} = \log\left[\prod_{i=1}^{n}\psi_i\left(\prod_{j=1}^{k}(p_{ij})^{y_{ij}}(1-p_{ij})^{1-y_{ij}}\right) + (1-\psi_i)I(y_{i\cdot}=0)\right] - \left[\lambda\left(k, \bar{\hat{p}}_0, n\right) \times \left(\sum_{i=0}^{g}\left|\beta_i - \tilde{\beta}_i\right|\right)\right] \quad (4)$$

where

$$\lambda(k, \bar{\hat{p}}_0, n) = \sqrt{\sum_{i=0}^{m}\mathrm{var}(\delta_i)} \times \left(1 - (1-\bar{\hat{p}}_0)^k\right)\left(1 - \bar{\hat{\psi}}_{\text{naïve}}\right).$$

Notice that as the number of sites or number of visits increases, the likelihood function is well behaved, and hence the penalty function is forced to converge to zero. If the detection probability is large, the naïve estimator

Table 1. Summary of the simulation results for 100 simulated data sets with 30 sites, two surveys, and constant probability of occupancy across the sites when the probability of detection is the same across the sites and surveys.

True probability of detection	True mean occupancy	MLE				Naïve				MPLE			
		Mean	Median	SE	MSE	Mean	Median	SE	MSE	Mean	Median	SE	MSE
0.10	0.29	0.73	1.00	0.44	0.38	0.06	0.03	0.05	0.06	0.30	0.16	0.26	0.06
	0.50	0.82	1.00	0.35	0.23	0.10	0.10	0.06	0.17	0.46	0.46	0.26	0.07
	0.80	0.85	1.00	0.31	0.10	0.15	0.17	0.06	0.43	0.63	0.73	0.26	0.10
0.30	0.30	0.66	0.92	0.37	0.27	0.16	0.17	0.07	0.02	0.45	0.41	0.27	0.10
	0.49	0.63	0.53	0.27	0.10	0.25	0.23	0.07	0.06	0.52	0.45	0.23	0.06
	0.79	0.80	0.83	0.21	0.04	0.40	0.40	0.08	0.16	0.71	0.73	0.20	0.04

Notes: The MLE are obtained by maximizing the likelihood of the zero inflated binomial, while the naïve estimates are obtained assuming that the probability of detection is 1. The MPLE combines the theoretical correctness of the MLE with the stability of the naïve by means of the penalized likelihood. Observe that MLE (maximum likelihood estimation) overestimates the occupancy, whereas MPLE (maximum penalized likelihood estimation) is nearly unbiased with smaller standard errors.

is a good estimator, and hence we can rely on it, and the shrinkage factor can be large. On the other hand, if the occupancy probability is large, the MLE usually is stable and hence shrinkage factor is small. An R program to obtain these estimators is *available online*.[2]

Simulation Analysis

A simulation study was conducted to compare the performance of the maximum likelihood estimator (MLE) and maximum penalized likelihood (MPLE) when the number of sites is small and the minimum number of surveys is conducted ($k = 2$ visits per location). We considered two settings: first we assumed that the probability of occupancy is the same for all sites, and that the probability of detection is constant for all sites and surveys, and, second, we assumed that the probability of occupancy and detection depend on some habitat and other exogenous covariates.

For the first case, we chose three different values for the probability of occupancy (0.80, 0.50, 0.30) and two different values for the probability of detection (0.30, 0.10). For the second case, the probabilities of occupancy and detection were assumed to depend on the covariates x_1, x_2, z_1, and z_2 according to the logistic link:

$$\psi_i = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})}$$

and

$$p_{ij} = \frac{\exp(\delta_0 + \delta_1 z_{1ij} + \delta_2 z_{2ij})}{1 + \exp(\delta_0 + \delta_1 z_{1ij} + \delta_2 z_{2ij})}.$$

One of the covariates was chosen to be discrete (habitat type = 0,1) and one was continuous. The values of the covariates were generated using $x_{1i} \sim$ Normal(2,1), $x_{2i} \sim$ Bernoulli(0.55), $z_{1ij} \sim$ Normal(1,1.5) and $z_{2ij} \sim$ Bernoulli(0.65), and the values of the parameters β_0, β_1, and β_2 were selected such that the mean probability of occupancy was 0.34 and the mean probability of detection was 0.15. These cases cover a broad range of possibilities for the site occupancy studies and allow us to evaluate the performance of the penalized likelihood method in situations where due to time, effort, and cost limitations the number of surveys and visited sites are small. For each combination of parameters, 100 data sets were generated and the ML, naïve, and MPL estimates were obtained. Subsequently, 100 bootstrap samples were obtained. The sites were sampled randomly with replacement to obtain bootstrap samples. The observations within a site were not sampled. These bootstrap samples were used to estimate the standard error and a confidence interval for the parameters of interest on each data set (see Appendix B for detailed algorithm).

Table 1 presents the simulation results for the first case. In general the MLE tends to overestimate the probability of occupancy; in some cases mean estimated values are more than twice the true occupancy. However, when the probability of occupancy is large, 0.80, the mean of the MLE is close to the true value even if the probability of detection is small. Additionally, if the probability of detection is small (0.10), no matter how large the probability of occupancy is, 50% of the times the estimated occupancy probability obtained by the MLE is 1. When the probability of occupancy and detection is low, 45% of the times the Quasi-Newton optimization method used in the package Presence (*available online*),[3] provides Fisher information matrices that are singular. This means that standard errors and confidence intervals based on the inverse of the Fisher information matrix are inappropriate in most small-data situations. We used a more stable method of optimization, data cloning (Lele et al. 2007), combined with bootstrap to obtain point estimates and confidence intervals, respectively.

The mean and median of the MPLE are closer to the true values than MLE and provide almost unbiased estimators in every case, except when the probability of occupancy is large. Notice that the bias of the naïve estimates is large for large values of the probability of occupancy (Table 1); consequently, the MPLE may be

[2] ⟨http://www.abmi.ca/abmi/home/home.jsp⟩

[3] ⟨http://www.proteus.co.nz/home.html⟩

TABLE 2. Coverage (the percentage of confidence intervals that contain the true value of the parameter), mean, and median length of the 90% confidence intervals under constant probability of occupancy and detection.

True probability of detection	True mean occupancy	MLE			Naïve			MPLE		
		Coverage (%)	Mean	Median	Coverage (%)	Mean	Median	Coverage (%)	Mean	Median
0.10	0.29	60	0.60	1.00	6	0.11	0.10	72	0.50	0.61
	0.50	53	0.52	0.86	0	0.16	0.17	86	0.63	0.69
	0.80	31	0.29	0.00	0	0.20	0.20	74	0.56	0.60
0.30	0.30	57	0.54	0.79	34	0.20	0.20	75	0.62	0.69
	0.49	74	0.59	0.74	15	0.24	0.24	84	0.63	0.70
	0.79	83	0.53	0.22	2	0.27	0.03	87	0.54	0.18

Note: Observe that the coverage for MLE-based confidence intervals is smaller than nominal coverage, whereas MPLE-based confidence intervals have better coverage and are shorter.

biased when the probability of occupancy is large. Nevertheless, the standard errors of the MPLE are smaller than for the MLE in every case. As observed in Table 2, the coverage of the bootstrap confidence intervals based on MPLE is closer to the nominal coverage than the ML-based bootstrap intervals.

Table 3 presents the results of the simulation study under the regression setting, where the probabilities depend on covariates. In this case, there are two quantities of interest, the regression coefficients, which measure covariate effects, and the average occupancy. For the latter, the simulation results show that the MLE and the MPLE perform equally well, with the MLE slightly overestimating and the MPLE slightly underestimating average occupancy. For the regression coefficients, the MLE are extremely unstable. This is because the likelihood function is fairly flat. The means of the MLE are substantially different from the medians, and consequently the standard errors are quite large. This indicates that, with a small number of sites and/or visits, it is highly likely that one might obtain MLE that are practically useless. On the other hand, the mean and the median of the MPLE are close to each other, indicating numerical stability, and are also closer to the true values of the parameters than the ML estimates (Table 3). The standard errors for the MPLE are also substantially smaller than those of MLE. Finally, the lengths of the confidence intervals obtained by the MPLE are at least 10 times shorter than the ones obtained by the ML method (Table 3). More importantly, except for the intercept parameter, the actual coverage of the MPLE-based confidence intervals is closer to the nominal coverage than for MLE-based confidence intervals. This means that the effect of the covariates is better estimated by the MPLE than by the MLE. In summary, simulation results show that, for those cases where MLE fails, the MPLE works extremely well; at the same time, when MLE does work well, MPLE works equally well.

Example Data Analysis

We illustrate the MPLE method using two occupancy studies: one for the Blue Ridge two-lined salamander (*Eurycea wilderae*) and the other for the Black-capped Chickadee (*Poecile atricapillus*).

The Blue-Ridge two-lined salamander study was conducted in the Great Smoky Mountains National Park during 2001. The data are available as part of the software Presence (see footnote 3). The data consist of 39 sites visited once every two weeks for a total of five surveys. There are no covariates available for the occupancy and detection probability models. Hence we use the simple model of constant probability of

TABLE 3. Summary results of the estimated parameters for the occupancy for 100 simulated data sets, with $n = 100$, two surveys, and two covariates for the occupancy.

Parameter	True value	Mean estimate	Median estimate	SE	Coverage of 90% CI	Mean length of 90% CI	Median length of 90% CI
MLE							
β_0	0.500	46.932	1.039	108.374	0.940	232.763	199.244
β_1	−1.000	−32.389	−1.460	64.890	0.870	164.670	122.346
β_2	1.200	29.344	2.004	65.356	0.850	214.047	163.653
Mean occupancy	0.340	0.373	0.334	0.169	0.960	0.474	0.476
MPLE							
β_0	0.500	−0.371	0.148	7.748	0.770	17.423	19.531
β_1	−1.000	−1.564	−0.961	1.893	0.930	5.266	4.069
β_2	1.200	3.187	1.417	5.552	0.910	15.867	18.546
Mean occupancy	0.340	0.314	0.277	0.172	0.900	0.453	0.445

Notes: Observe that MLE estimates are biased and have large standard errors. The MPLE estimates are almost unbiased with substantially smaller standard errors. Moreover, the MPLE-based confidence intervals for the parameters are 10 times shorter without sacrificing the coverage.

TABLE 4. Summary results of the estimated probability of occupancy and its standard error for the Blue-Ridge two-lined salamander (*Eurycea wilderae*) data.

	MLE				MPLE			
Surveys	$\hat{\psi}$	$\widehat{SE}$	90% CI		$\hat{\psi}$	$\widehat{SE}$	90% CI	
1, 2, 3, 4, 5	0.59	0.15	0.43	1.00	0.59	0.15	0.43	0.94
1, 2	0.31	0.35	0.12	1.00	0.24	0.27	0.10	0.88
1, 3	0.54	0.24	0.32	1.00	0.48	0.22	0.28	0.95
1, 4	0.31	0.19	0.16	0.78	0.29	0.14	0.13	0.57
1, 5	0.64	0.28	0.26	1.00	0.48	0.25	0.23	0.95
2, 3	0.92	0.23	0.32	1.00	0.65	0.23	0.30	0.97
2, 4	0.78	0.25	0.31	1.00	0.57	0.24	0.28	0.95
2, 5	1.00	0.00	1.00	1.00	0.92	0.07	0.79	0.96
3, 4	0.46	0.16	0.27	0.82	0.44	0.15	0.25	0.71
3, 5	0.72	0.23	0.37	1.00	0.62	0.22	0.35	0.98
4, 5	0.31	0.15	0.18	0.56	0.30	0.11	0.18	0.49

occupancy and constant probability of detection. The goal of the analysis is to compare the estimated occupancy obtained by the MLE and MPLE and their confidence intervals and standard errors under various scenarios.

We first present the analysis for MLE and MPLE using all five visits. The results presented in Table 4 show that the MLE and MPLE are quite similar, although standard errors and confidence intervals based on MPLE are somewhat shorter than for MLE. This is to be expected because when the number of visits is large, the penalty function is small and MPLE and MLE are similar. Next we consider the possibility of only two visits. There are 10 such combinations possible. In Table 4, we present the estimated occupancy obtained by using the ML and MPL estimators for every possible pair of visits. Notice that in every case the standard errors of the MPL estimator are smaller than the ones obtained by the ML estimator. The bootstrap confidence intervals based on the MPLE are shorter than those based on the MLE estimators. MPLE, thus, provides a more precise representation of the occupancy than the MLE. It is also interesting to note that the inferences from different pairs of surveys vary substantially from each other, with occupancy estimates ranging from 0.24 to 0.92. This suggests that perhaps the validity of this assumption of a closed population during the time of the study is questionable.

The second example corresponds to an occupancy study that was conducted on lands managed by Millar Western Forest Products in western Alberta from 2000 to 2002 (E. Bayne, *unpublished manuscript*). For illustrative purposes, only the data for the Black-capped Chickadee (BCHH) are used. The data were collected over a period of three years. Each year, 40 sites were visited once every week starting on 15 May and ending on 28 July. Two different observers, randomly assigned to the sites, were used. The purpose of the analysis is to determine whether or not there was a trend for the occupancy of the BCCH over the three years of the

TABLE 5. Estimated parameters, 90% confidence intervals, and standard errors for the occupancy and detection model of the Black-capped Chickadee.

	MLE		MPLE	
Parameter	Estimate (90% CL)	SE	Estimate (90% CL)	SE
Occupancy model				
Intercept	8.785 (2.524, 9.680)	2.126	3.223 (1.654, 4.240)	0.789
Year 2	−9.882 (−10.734, −3.624)	2.150	−4.324 (−5.587, −2.664)	0.858
Year 3	−10.310 (−11.503, 4.272)	2.113	−4.746 (−5.994, −3.249)	1.397
Detection model				
Intercept	−1.506 (−1.766, −1.222)	0.167	−1.475 (−1.724, −1.191)	0.164
Mean conditional occupancy given detection history	0.476 (0.444, 0.524)	0.026	0.464 (0.419, 0.510)	0.028
Detection probability	0.182 (0.146, 0.228)	0.025	0.186 (0.151, 0.233)	0.025

Notes: The standard errors and confidence intervals were estimated using 200 bootstrap samples. Notice that the confidence intervals for the occupancy model parameters obtained by the MPLE are ~60% shorter than the confidence intervals obtained by the MLE.

Reports

study. The covariates for the occupancy model correspond to the year of the survey, the reference (year 1) being year 2000. For the detection probability model covariates such as the observer, Julian date and the time of the survey were tried. Because they turned out to be nonsignificant, a constant detection model was fitted.

Table 5 presents the MLE and MPLE of the parameters for this model. The standard errors, as well as the 90% confidence limits, were calculated using 200 bootstrap samples. It was found that the standard errors for the occupancy model provided by the MLE were substantially larger than the ones obtained by the MPLE, and that the MPLE's confidence intervals were shorter than the ones obtained by the MLE. On the other hand, the standard errors and confidence intervals for the detection model were almost the same for both the MLE and the MPLE. Using the MPLE estimates it can be concluded that there was a decreasing trend for the occupancy of the BCCH. During the first year the estimated mean occupancy was ~0.9616, dropping for the second year to 0.2495 and decreasing further for the last year to 0.179. These data are part of a large ecological study of how forest density affects occupancy. Nearly 50% of the trees were removed from the area between year 1 and year 2. The drop in occupancy is the likely outcome of such a change in the forest density. A full analysis of the experiment is beyond the scope of this paper, but is underway for publication (E. Bayne, *personal communication*).

Discussion

If the number of sites and number of surveys are small as compared to the complexity of the model, namely the number of covariates, estimation of the parameters for a site occupancy model using MLE can be unstable. Moreover, the estimated standard error obtained by the MLE, based on the inverse of the Fisher information, is also unstable. On the other hand, the penalized likelihood estimators have better statistical properties with smaller mean squared error and bootstrap confidence interval coverage closer to the nominal coverage than for the ML estimator. Furthermore, the estimates for the occupancy model obtained by the MPLE are somewhat conservative, while the estimates obtained by the MLE are optimistic. We feel that from a conservation and management perspective, it is better to provide stable, albeit somewhat conservative, estimates of occupancy than wildly unstable, optimistic estimates of occupancy. We also found that the bootstrap confidence intervals, but not the standard errors, for both MLE and MPLE tend to be stable, and hence we recommend their use in practice. From a practical perspective, the use of penalized likelihood estimators can lead to a substantial reduction in the required number of surveys and sites. This ultimately can lead to a substantial reduction in the cost of implementing such surveys.

Acknowledgments

This research was supported by a grant from the National Science and Engineering Research Council (NSERC) of Canada and a fellowship from the Alberta Biodiversity Monitoring Institute (ABMI). We are grateful to Erin Bayne, Brian Gray, Jim Schieck, Peter Solymos, Daryl MacKenzie, Jim Hines, and the anonymous referee for their comments and suggestions.

Literature Cited

Lele, S., B. Dennis, and F. Lutscher. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecology Letters 10:551–563.

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.

APPENDIX A

A detailed explanation of how the penalty function was derived (*Ecological Archives* E091-025-A1).

APPENDIX B

An algorithm to obtain the bootstrap confidence intervals (*Ecological Archives* E091-025-A2).