

Open Population Abundance Estimation Using Repeated Presence-Absence Observations: With or Without the Robust Design

When animal monitoring efforts are concerned with estimating the total abundance of a species that is imperfectly detected, they typically collect at minimum point counts of the number of individuals observed in defined sites within the study area. Then, by repeating this process for multiple seasons, the repeated point counts provide enough information to enable estimation of both detection probability and seasonal abundance, along with parameters controlling emigration and immigration. However, it is much easier to record only the detected occupancy status of each site, since surveying in each site can conclude after the first individual is observed and it alleviates concern over double-counting individuals. Royle and Nichols (2003, *Ecology*, **84**, 777–790) propose a method that utilizes the fact that the probability of detecting a site as occupied increases with site abundance to obtain an estimate of total abundance using presence-absence observations, as long as all observations are obtained within a single season, so that site abundances can be assumed constant. Here, we propose two models that accommodate multiple seasons of presence-absence observations: one for the case of $J > 1$ sampling occasion per season, and another for the case of $J = 1$. A simulation study demonstrates that each model outperforms the other in its respective setting. The proposed models are then applied to six seasons of presence-absence observations (with $J = 1$) for five avian species detected in a Breeding Bird Survey route in western Oregon. Supplemental materials, including R code and a tutorial for implementing the proposed models, are available online.

Key words: Animal monitoring; Breeding Bird Survey; N -Mixture Models; Occupancy.

1 INTRODUCTION

A primary goal of many animal monitoring programs is to track trends in abundance over time. Abundance and trend estimates can be readily obtained from raw point counts of any species that is detected with perfect probability, such as trees or conspicuous, slow moving animals (He and Gaston 2000). The collection of point counts involves an observer counting the number of animals of the target population detected in a defined area (site) within a specified time period of observation, perhaps two or three minutes. However, since most species are imperfectly detected, the raw point count will tend to be less than the actual abundance. Therefore, additional information must be collected beyond the point count to estimate the detection probability in addition to the abundance (Thompson 1992; Yoccoz et al. 2001).

One way to estimate both the abundance and detection probability is to replicate the collection of point counts over both space and time, surveying R well-defined (and independent) sites J times during each of T seasons. This is often less burdensome for animal monitoring programs than requiring the collection of information beyond the counts (Kéry 2008), such as the capture histories of marked individuals or the distance from observer to detected individuals (Pollock et al. 2002).

For the case of $T = 1$, every sample is obtained during the same season and the population is assumed closed to births, deaths, immigration and emigration throughout the study. In such cases with $T = 1$ and $J > 1$, an N -mixture model proposed by Royle (2004) yields an accurate estimate of both total abundance and detection probability. When $T > 1$, population closure is not assumed between seasons. In these cases, for either $J = 1$ or $J > 1$ the model proposed by Dail and Madsen (2011) provides estimates of total abundance during each season, as well as detection probability and parameters controlling animal dynamics.

However, since less time and effort are required to establish the presence of a species in a site than is required to count all the animals detected on the site, it is more efficient to simply record the presence of an animal species on a site rather than to record the count of all animals detected. This enables a study to include more sites and to conduct more surveys per site (Joseph et al. 2006). In keeping with the literature in this field, observations of this type are referred to here as presence-

absence data, although perhaps more accurately they could be called detection-nondetection data.

An illustration of data that typically arise from presence-absence surveys is given Table 1, containing annual presence-absence and point count observations for the House Finch (*Carpodacus mexicanus*) from five of 50 sites along a Breeding Bird Survey (BBS) route in western Oregon. These data illustrate the connection between point counts and presence-absence observations, where the presence-absence observations are right-truncated at the value 1.

Since point counts contain more information than presence-absence data, the main advantage gained by collecting presence-absence data instead of point counts is that more surveys can be conducted using the same amount of time and resources. In this example, the number of surveys is fixed and equal for both data types ($R = 50, J = 1, T = 6$), so the method utilizing the point counts (Dail and Madsen 2011) will be used as a standard with which to compare proposed methods utilizing only the presence-absence observations.

[Table 1 about here.]

The predominant model for obtaining abundance estimates from presence-absence data is proposed by Royle and Nichols (2003), and is valid only for $T = 1$ and $J > 1$. Their model (henceforth the “RN” model) utilizes the fact that during occasion j , the probability of obtaining a detection on a site i , denoted $p_{i,j}$, will typically increase with increased animal abundance on a site (Hartley 1998). Specifically, if individual animal detections on a site are assumed independent with common probability $r_{i,j}$, the relationship between site abundance (denoted N_i) and $p_{i,j}$ is

$$p_{i,j} = 1 - (1 - r_{i,j})^{N_i}. \quad (1.1)$$

However, there are other ways that heterogeneity in p can be induced (e.g., beta mixtures or normal mixtures), and Royle (2006) demonstrates that these competing models are indistinguishable from data – the likelihoods for different models can be identical. Therefore, estimating abundance from presence-absence data via the RN model relies heavily upon the assumption that any heterogeneity in p not explained by the inclusion of covariates is entirely accounted for by

differing site abundance levels through (1.1).

Other choices for models connecting occupancy to abundance have been proposed, for example, by Sileshi et al. (2006), and detailed knowledge of the biological system under study would presumably lead to the proper choice. This contrasts the other main assumption inherent in the RN model, that the study population is closed (i.e., $T = 1$). To our knowledge, no alternative has yet been proposed to allow for multiple seasons ($T > 1$) or to test the closure assumption of the RN model.

This is important because models built upon the closure assumption yield notoriously poor estimates when closure is violated (Kendall 1999; Rota et al. 2009; Dail and Madsen 2011). Furthermore, Rota et al. (2009) demonstrate the closure assumption is violated in several studies of duration shorter than a single breeding season.

The dual restrictions of Equation (1.1) and $T = 1$ notwithstanding, the RN model is widely used to obtain abundance estimates in ecological studies that yield presence-absence observations (e.g., Dreitz, Lukacs, and Knopf 2006; Hood and Dinsmore 2007; Tipton 2007). Presumably, in these and other cases the choice to model heterogeneity in p using Equation (1.1) was indeed a modeling choice, since there are viable alternatives such as those proposed by Sileshi et al. (2006). However, once the choice was made to utilize the RN model and Equation (1.1), each study by necessity could only analyze a single season of observations.

The purpose of this paper is to provide models that use multiple seasons ($T > 1$) of presence-absence observations to form abundance estimates, building off of the assumption that heterogeneity in p is explained via Equation (1.1). We propose two models, one for the case $J > 1$ and another for the case $J = 1$.

When $T > 1$ and $J > 1$, multiple surveys are obtained each season during which population closure can be assumed. This is referred to as the “robust design,” developed by Pollock (1982) initially for capture-recapture studies but subsequently extended to other designs, including presence-absence studies as considered here. In this setting, the RN model could be utilized separately for each season, but this approach would not provide inference between seasons, such

as the estimation of trend. Our proposed model for this setting is an extension of the robust design point count model of Dail and Madsen (2011) to the presence-absence setting. The observations obtained within population closure provide information useful for estimating p , which then allows parameters controlling population dynamics to be estimated accurately.

When $T > 1$ and $J = 1$, the study lacks the robust design. In this setting, use of the RN model implies population closure, which is not a valid assumption and results in highly biased abundance estimates. Although the robust design is widely used in presence-absence studies due to the information it provides for estimating p (e.g., Rizkalla and Swihart 2006; MacKenzie et al. 2010; Long et al. 2010), it can be difficult to replicate the data collection process in a short enough time span to ensure population closure and so the robust design is not always implemented (Schaub, Gimenez, Schmidt, and Pradel 2004). An example of one such study lacking the robust design is presented in Table 1.

A recent proposed modeling framework accommodating $J = 1$ is given by Dail and Madsen (2012). Their model extends the site occupancy modeling framework of MacKenzie et al. (2003) by conditioning the likelihood of observed data on the number of occupied sites during each season, denoted $O_{\cdot,t}$. The marginal likelihood of observed data is then obtained by multiplying this conditional likelihood by the prior for the $O_{\cdot,t}$ values and then summing out $O_{\cdot,t}$. Our proposed model for this setting utilizes the Dail and Madsen (2012) model to obtain abundance estimates via arguments connecting presence-absence observations to abundances, such as the one given in Equation (1.1).

The outline of the article is as follows. Section 2 briefly describes abundance models for closed populations ($T = 1$), both for when the data are point counts (Section 2.1) and when the data are presence-absence observations (Section 2.2). The proposed models for open populations ($T > 1$) are then described in Section 3, starting with settings with the robust design (Section 3.1) and then proceeding to settings without the robust design (Section 3.2). Next, the performance of the proposed models is compared in a simulation study presented in Section 4. Then, in Section 5 the proposed models are applied to presence-absence observations of 5 different bird species

encountered in a BBS route consisting of six annual surveys in western Oregon. And finally, the article concludes with a discussion in Section 6.

2 ABUNDANCE MODELS FOR CLOSED POPULATIONS

2.1 POINT COUNT MODEL

For closed populations we have $T = 1$, so the observed point count (denoted $n_{i,j}$) is indexed by site ($i = 1, \dots, R$) and sampling occasion ($j = 1, \dots, J$). Although the number of sampling occasions J can in general depend on site, for clarity here we assume J is constant so the full point count data can be compiled into a (R by J) matrix, denoted \mathbf{n} .

Animal detections in a site are assumed independent, and all animals present in a site i during any occasion j , denoted N_i , are assumed to have the same detection probability, denoted $r_{i,j}$. Furthermore, it is assumed that errors of detection are only made by omission (i.e., no individual is misidentified or counted twice).

Therefore, conditional on N_i and $r_{i,j}$, $n_{i,j}$ is a binomial random variable with (unknown) index N_i and success probability $r_{i,j}$. Letting \mathbf{N} represent the vector of site abundances, assuming that sites are independent then gives

$$L(\mathbf{N}, \mathbf{r} | \mathbf{n}) = \prod_{i=1}^R \left\{ \prod_{j=1}^J \binom{N_i}{n_{i,j}} r_{i,j}^{n_{i,j}} (1 - r_{i,j})^{N_i - n_{i,j}} \right\}. \quad (2.1)$$

In the modeling framework proposed by Royle (2004), the conditional likelihood of Equation (2.1) is then multiplied by a prior likelihood for N_i , denoted $f(N_i)$, chosen to represent the distribution of animals to sites. This prior is typically a parametric model such as the Poisson or negative binomial, with the former accommodating animals distributed uniformly in space in a Poisson point process and the latter allowing over-dispersion such as might be present if animals cluster together. Both models can be parameterized to have mean λ , with over-dispersion parameter α for

the negative binomial.

Then, the site abundance N_i is removed via summation, leaving the marginal likelihood of the observed counts:

$$L(\mathbf{r}, \boldsymbol{\lambda}, \boldsymbol{\alpha} | \mathbf{n}) = \prod_{i=1}^R \left[\sum_{N_i=\max_t n_{it}}^{\infty} \left\{ \prod_{j=1}^J \binom{N_i}{n_{i,j}} r_{i,j}^{n_{i,j}} (1-r_{i,j})^{N_i-n_{i,j}} \right\} f(N_i; \boldsymbol{\lambda}, \boldsymbol{\alpha}) \right]. \quad (2.2)$$

In practice, the infinite summation over N_i is approximated using a finite bound K , chosen to be large enough to have negligible impact on MLEs obtained by numerically maximizing (2.2). Additionally, more parsimonious models can be obtained by either assuming r to be constant or by accounting for differences among $r_{i,j}$ values with a logit-linear model with q measured covariates:

$$\log \left(\frac{r_{i,j}}{1-r_{i,j}} \right) = \beta_0 + \beta_1 Z_{1,i,j} + \cdots + \beta_q Z_{q,i,j}. \quad (2.3)$$

The estimate of total abundance, denoted N , is then obtained as $R\hat{\lambda}$.

2.2 PRESENCE-ABSENCE MODEL

In presence-absence studies only the overall detection or non-detection of animals is collected during each survey, rather than the count of the number of individual animals detected. In this way, with $w_{i,j}$ denoting the observed occupancy status of site i during sampling occasion j , we have $w_{i,j} = I(n_{i,j} \geq 1)$, where $I(\cdot)$ is the indicator function.

The abundance model proposed by Royle and Nichols (2003) (referred to earlier as the RN model) combines Equation (1.1) with the modeling framework described for point counts in Section 2.1. This results in the following likelihood:

$$L(\mathbf{r}, \boldsymbol{\lambda}, \boldsymbol{\alpha} | \mathbf{w}) = \prod_{i=1}^R \left\{ \sum_{N_i=\max_j w_{i,j}}^{\infty} \left(\left[\prod_{j=1}^J \left\{ 1 - (1-r_{i,j})^{N_i} \right\}^{w_{i,j}} (1-r_{i,j})^{N_i(1-w_{i,j})} \right] f(N_i; \boldsymbol{\lambda}, \boldsymbol{\alpha}) \right) \right\} \quad (2.4)$$

where $f(N_i)$ is the prior likelihood for N_i (typically either Poisson or negative binomial) as de-

scribed in Section 2.1. Here again, numerical maximization of (2.4) is achieved by approximating the infinite summation with summation to a fixed finite K , covariates can be included for $r_{i,j}$ using Equation (2.3), and the estimate of total abundance is $R\hat{\lambda}$.

3 ABUNDANCE MODELS FOR OPEN POPULATIONS

3.1 PRESENCE-ABSENCE MODEL – WITH THE ROBUST DESIGN

Settings with the robust design have multiple seasons ($T > 1$) and multiple survey occasions within season ($J > 1$). Here, to simplify notation we assume that both T and J are constant for all sites i . Thus, our presence-absence observations are indexed as $w_{i,t,j}$ with $i = 1, \dots, R$, $t = 1, \dots, T$, and $j = 1, \dots, J_t$, and we can compile these observations into a matrix \mathbf{w} with R rows and $\sum_t J_t$ columns.

The proposed abundance model for robust design presence-absence observations is constructed by modifying the point count model of Dail and Madsen (2011) to instead consider occupancy data. This is a parallel construction to that described in Section 2, where the RN model is a presence-absence modification of the Royle (2004) point count model.

Point count model. The point count model of Dail and Madsen (2011) (henceforth the “DM” model) generalizes the Royle (2004) model to include multiple seasons ($T > 1$). As before with the Royle (2004) model, the marginal likelihood of the observed data is again formed by multiplying the conditional likelihood of the observed point counts given the site abundances by a prior likelihood for site abundances and then summing out the site abundances.

The main difference here is the prior for site abundances incorporates the entire vector $\{N_{i,1}, N_{i,2}, \dots, N_{i,T}\}$, as opposed to a single value N_i . This is greatly simplified by assuming that site abundances have the Markov property, where $N_{i,t}$ depends on the abundance history at that site

only through $N_{i,t-1}$. With this, we have

$$g(N_{i,1}, \dots, N_{i,T}) = g_1(N_{i,1}) \prod_{t=2}^T g_2(N_{i,t} | N_{i,t-1}). \quad (3.1)$$

It is natural to allow g_1 in Equation (3.1) to be the same as the prior chosen for f in Section 2.2 (typically either Poisson or negative binomial), since both are models for initial distribution of animals to sites.

Then, g_2 is formed by decomposing $N_{i,t} | N_{i,t-1}$ into the number of animals surviving from season $t-1$ and the number of new arrivals since season $t-1$, denoted $S_{i,t}$ and $G_{i,t}$, respectively. Then, $S_{i,t}$ is modeled as a binomial with index $N_{i,t-1}$ and success probability ω ; $G_{i,t}$ is modeled as a Poisson random variable with rate γ . The discrete convolution then gives:

$$g_2(N_{i,t} | N_{i,t-1}) = \sum_{c \in C} \binom{N_{i,t-1}}{c} \omega^c (1-\omega)^{N_{i,t-1}-c} \frac{e^{-\gamma} \gamma^{N_{i,t}-c}}{(N_{i,t}-c)!}, \quad (3.2)$$

where $C = \{0, \dots, \min(N_{i,t}, N_{i,t-1})\}$.

We therefore have

$$\begin{aligned} L(\mathbf{r}, \lambda, \alpha, \gamma, \omega | \mathbf{n}) &= \prod_{i=1}^R \left(\sum_{N_{i,1}=\max_j n_{i,1,j}}^{\infty} \dots \sum_{N_{i,T}=\max_j n_{i,T,j}}^{\infty} \left[\left\{ \prod_{t=1}^T \prod_{j=1}^{J_t} \binom{N_{i,t}}{n_{i,t,j}} r_{i,t,j}^{n_{i,t,j}} (1-r_{i,t,j})^{N_{i,t}-n_{i,t,j}} \right\} \right. \right. \\ &\quad \left. \left. \times g_1(N_{i,1}; \lambda, \alpha) \prod_{t=2}^T g_2(N_{i,t} | N_{i,t-1}; \gamma, \omega) \right] \right) \end{aligned} \quad (3.3)$$

Here, as in Section 2, the numerical optimization of Equation (3.3) proceeds by approximating the infinite summations by summations up to a large fixed value K , and covariates can be included for $r_{i,t,j}$ using Equation (2.3). Estimates of total abundance for each season, denoted $N_{\cdot,t}$, can then be

computed recursively using the MLEs:

$$\begin{aligned}\hat{N}_{\cdot,1} &= R\hat{\lambda} \\ \hat{N}_{\cdot,t} &= \hat{\omega}\hat{N}_{\cdot,t-1} + R\hat{\gamma}\end{aligned}\tag{3.4}$$

Presence-absence model. Equation (3.3) can be modified to allow presence-absence observations instead of point counts by utilizing Equation (1.1) and adjusting the lower limits of the summations. This results in the following marginal likelihood:

$$\begin{aligned}L(\mathbf{r}, \lambda, \alpha, \gamma, \omega | \mathbf{w}) &= \prod_{i=1}^R \left\{ \sum_{N_{i,1}=\max_j w_{i,1,j}}^{\infty} \cdots \sum_{N_{i,T}=\max_j w_{i,T,j}}^{\infty} \left(\left[\prod_{t=1}^T \prod_{j=1}^{J_t} \left\{ 1 - (1 - r_{i,t,j})^{N_{i,t}} \right\}^{w_{i,t,j}} \right. \right. \right. \\ &\quad \left. \left. \times (1 - r_{i,t,j})^{N_{i,t}(1-w_{i,t,j})} \right] g_1(N_{i,1}; \lambda, \alpha) \prod_{t=2}^T g_2(N_{i,t} | N_{i,t-1}; \gamma, \omega) \right) \right\}\end{aligned}\tag{3.5}$$

As is the case for the point count model above, to numerically optimize Equation (3.5) the infinite summations are replaced by summations to K , covariates can be included for $r_{i,t,j}$, and estimates of $N_{\cdot,t}$ (denoted $\hat{N}_{\cdot,t}^{(\text{DM})}$ for the DM model) are computed recursively using the MLEs and Equation (3.4).

3.2 PRESENCE-ABSENCE MODEL – WITHOUT THE ROBUST DESIGN

Dail and Madsen (2012) propose a modeling framework (henceforth the “PA” model) that uses presence-absence observations collected without the robust design to estimate the site occupancy rate for each season. Here, we extend the PA model to provide abundance estimates, utilizing arguments similar to that given in Equation (1.1). First, though, we briefly describe the PA model.

Site occupancy estimation via the PA model. The basic construction of the PA model is similar to that of the DM model. Both models utilize nuisance parameters, obtaining the marginal likelihood of the observed data by first conditioning on them, then multiplying by their joint prior, and

finally removing them via summation. Whereas the nuisance parameters used in DM model are site abundances N_t , in the PA model they are the number of occupied sites each season, denoted $O_{\cdot,t}$ where $O_{\cdot,t} = \sum_i O_{i,t}$.

First we consider the prior for $O_{\cdot,1}, \dots, O_{\cdot,T}$, denoted h . As is the case for the prior of N_1, \dots, N_T in the DM model given in Equation (3.1), the Markov assumption allows the prior for h to be written as

$$h(O_{\cdot,1}, \dots, O_{\cdot,T}) = h_1(O_{\cdot,1}) \prod_{t=2}^T h_2(O_{\cdot,t} | O_{\cdot,t-1}). \quad (3.6)$$

Letting Ψ_1 denote the probability a site is occupied during the first season, assuming site independence then allows h_1 to be modeled as a binomial with known index R and success Ψ_1 :

$$h_1 = \binom{R}{O_{\cdot,1}} \Psi_1^{O_{\cdot,1}} (1 - \Psi_1)^{R - O_{\cdot,1}} \quad (3.7)$$

Next, h_2 of Equation (3.6) is formed similarly to g_2 in Section 3.1, decomposing $O_{\cdot,t} | O_{\cdot,t-1}$ into the number of sites remaining occupied since $t-1$ (denoted S_t) and newly occupied sites (denoted G_t). Here, though, since the total number of sites is fixed at R , it is useful to model both S_t and G_t are modeled as binomial random variables. With τ_t and ε_t denoting site colonization and extinction probabilities, respectively, S_t has index $O_{\cdot,t-1}$ and survival probability $1 - \varepsilon_t$, and G_t has index $R - O_{\cdot,t-1}$ and probability τ_t . Assuming independence of S_t and G_t , the discrete convolution then gives:

$$\begin{aligned} h_2(O_{\cdot,t} | O_{\cdot,t-1}) &= \sum_{c=c_1}^{c_2} \binom{O_{\cdot,t-1}}{c} (1 - \varepsilon_t)^c \varepsilon_t^{O_{\cdot,t-1}-c} \\ &\quad \times \binom{R - O_{\cdot,t-1}}{O_{\cdot,t} - c} \tau_t^{O_{\cdot,t}-c} (1 - \tau_t)^{R - O_{\cdot,t-1} - O_{\cdot,t} + c}, \end{aligned} \quad (3.8)$$

where $c_1 = \max(0, O_{\cdot,t} + O_{\cdot,t-1} - R)$ and $c_2 = \min(O_{\cdot,t}, O_{\cdot,t-1})$.

Now we consider the conditional likelihood of \mathbf{w} given $O_{\cdot,1}, \dots, O_{\cdot,T}$; it is helpful to denote this likelihood as L_1 . To calculate L_1 , we first obtain the probability of a site i being occupied during season t given its entire detection history (through season t), denoted $\pi_{i,t}$. By assumption,

when $w_{i,t} = 1$ we have $\pi_{i,t} = 1$. When $w_{i,t} = 0$, Bayes' Formula yields:

$$\pi_{i,1} = \frac{(1 - p_1)\Psi_1}{(1 - p_1)\Psi_1 + (1 - \Psi_1)}.$$

With $\phi_{i,t} = \pi_{i,t-1}(1 - \varepsilon_t) + (1 - \pi_{i,t-1})\tau_t$, applying Bayes' Formula again yields:

$$P(O_{i,t} = 1 | w_{i,t} = 0, \pi_{i,t-1}) = \frac{(1 - p_1)\phi_t}{(1 - p_1)\phi_t + (1 - \phi_t)},$$

so $\pi_{i,t}$ can be calculated recursively for any detection history.

Next, $\pi_{i,t}^*$ is obtained by conditioning the R -length vector π_t on $O_{\cdot,t}$. Writing this as the joint probability $(O_{i,t} = 1, O_{\cdot,t} = k)$ divided by the marginal probability $(O_{\cdot,t} = k)$, we have:

$$\pi_{i,t}^*(k) = \frac{\pi_{i,t} \sum_{m \in \Theta_{(i)}} \left(\prod_{c=1}^{k-1} \pi_{m_c,t} \right) \{ \prod_{c=k}^{R-1} (1 - \pi_{m_c,t}) \}}{\sum_{m \in \Theta} \left(\prod_{c=1}^k \pi_{m_c,t} \right) \{ \prod_{c=k+1}^R (1 - \pi_{m_c,t}) \}}, \quad (3.9)$$

where Θ denotes all permutations of $(1, \dots, R)$, $\Theta_{(i)}$ denotes all permutations of $(1, \dots, i-1, i+1, \dots, R)$, and $\pi_{i,t}^*(k)$ denotes $P(O_{i,t} = 1 | w_1, \dots, w_t, O_{\cdot,t} = k)$.

Let $w_{(t)}$ denote the column vector of observed occupancy status for all sites during season t . We next calculate $P(O_{i,t} = 1 | w_{(1)}, \dots, w_{(t-1)}, O_{\cdot,t-1})$, denoted $\theta_{i,t}$, using $\theta_{i,t} = \pi_{i,t}^*(1 - \varepsilon_t) + (1 - \pi_{i,t}^*)\tau_t$. And finally, we obtain $\theta_{i,t}^*$ by conditioning $\theta_{i,t}$ on $O_{\cdot,t}$ using Equation (3.9), giving the probability a site i is occupied in season t given its detection history through season $t-1$ and the quantities $O_{\cdot,t}$ and $O_{\cdot,t-1}$.

We now proceed with the calculation of L_1 . The Markov assumption allows L_1 to be written as a product:

$$L_1 = P(w_{(1)} | O_{\cdot,1}) \prod_{t=2}^T P(w_{(t)} | w_{(1)}, \dots, w_{(t-1)}, O_{\cdot,t-1}, O_{\cdot,t}). \quad (3.10)$$

Since $P(w_{(t)} | w_{(1)}, \dots, w_{(t-1)}, O_{\cdot,t-1}, O_{\cdot,t})$ in Equation (3.10) and $\theta_{i,t}^*$ condition on the same quan-

tities, we can write the former as $P(\mathbf{w}_{(t-1)} | \theta_{i,t}^*)$. Then, we have

$$P(\mathbf{w}_{(t-1)} | \theta_{i,t}^*) = \prod_{i=1}^R \left\{ (\theta_{i,t}^* p_t)^{w_{i,t}} (1 - \theta_{i,t}^* p_t)^{1-w_{i,t}} \right\}, \quad (3.11)$$

since $w_{i,t} = 1$ only when a site is occupied and a detection is recorded given occupancy, which occur with probabilities $\theta_{i,t}^*$ and p_t , respectively.

Finally, $P(\mathbf{w}_{(1)} | O_{\cdot,1})$ can be written as the product of $P(w_{\cdot,1} | O_{\cdot,1})$ and $P(\mathbf{w}_{(1)} | w_{\cdot,1}, O_{\cdot,1})$. The former is a binomial probability with index $O_{\cdot,1}$ and success probability p_1 , and the latter is the probability of observing a specific ordering of occupied sites:

$$P(\mathbf{w}_{(1)} | O_{\cdot,1}) = \binom{O_{\cdot,1}}{w_{\cdot,1}} p_1^{w_{\cdot,1}} (1 - p_1)^{O_{\cdot,1} - w_{\cdot,1}} \left(\frac{R}{O_{\cdot,1}} \right)^{-1}. \quad (3.12)$$

Therefore, the PA model likelihood is

$$L(\mathbf{p}, \boldsymbol{\tau}, \boldsymbol{\varepsilon}, \Psi_1 | \mathbf{w}) = \sum_{O_{\cdot,1}=w_{\cdot,1}}^R \cdots \sum_{O_{\cdot,T}=w_{\cdot,T}}^R L_1 \left\{ h_1(O_{\cdot,1}) \prod_{t=2}^T h_2(O_{\cdot,t} | O_{\cdot,t-1}) \right\}, \quad (3.13)$$

where h_1 and h_2 are given in Equations (3.7) and (3.8), and L_1 is obtained using Equations (3.10), (3.11), and (3.12). Here again, as in Sections 2.1, 2.2, and 3.1, covariates can be included for \mathbf{p} , $\boldsymbol{\tau}$, and $\boldsymbol{\varepsilon}$ using Equation (1.1). Numerically maximizing Equation (3.13) yields the MLEs $(\hat{\mathbf{p}}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\varepsilon}}, \hat{\Psi}_1)$, which can then be used to obtain an estimate for site occupancy rate during each season via the recursive formula:

$$\Psi_t = \Psi_{t-1}(1 - \varepsilon_{t-1}) + (1 - \Psi_{t-1})\tau_{t-1}. \quad (3.14)$$

Abundance estimation. However, our goal here is to estimate abundance rather than site occupancy rate. Here we describe a method for obtaining seasonal estimates for total abundance using MLEs obtained maximizing Equation (3.13).

In an argument similar to that used to construct Equation (1.1) tying individual detections (r)

to occupancy detections (p), we note that an unoccupied site becomes colonized when at least one individual enters, and local extinction occurs when every individual in a site fails to survive. Thus, we have:

$$\tau_t = 1 - \exp(-\gamma_t) \quad (3.15)$$

$$\varepsilon_t = (1 - \omega_t)^{N_{i,t-1}}, \quad (3.16)$$

where γ_t and ω_t are the individual arrival rate the survival probability in the Dail and Madsen (2011) model from Section 3.1. Equations (3.15) and (3.16) can then be solved for $\hat{\gamma}_t$ and $\hat{\omega}_t$ and incorporated into $g_2(u, v)$ of Equation (3.2). This yields $\hat{P}(N_{i,t} = v | N_{i,t-1} = u)$, which is helpful to denote $\widehat{\mathcal{P}}_{t(u,v)}$. Letting K represent an upper bound for both $N_{i,t}$ and $N_{i,t-1}$ (as discussed at the end of Section 3.1), we can use the $\widehat{\mathcal{P}}_{t(u,v)}$ values to compile the transition probability matrix $\widehat{\mathcal{P}}_t$.

We assume a negative binomial prior is chosen for $g_1(N_{i,1})$, since this also accommodates the Poisson distribution as the over-dispersion parameter $\alpha \rightarrow \infty$. Thus we have:

$$P_{\lambda,\alpha}(N_{i,1} = m) = \frac{\lambda^m \Gamma(\alpha + m)}{m! \Gamma(\alpha) (\alpha + \lambda)^m} \frac{1}{(1 + \frac{\lambda}{\alpha})^\alpha}. \quad (3.17)$$

Then, setting $P_{\lambda,\alpha}(N_{i,1} = 0)$ equal to $1 - \hat{\Psi}_1$, from Equation (3.17) we obtain:

$$\lambda(\alpha) = \alpha(1 - \hat{\Psi}_1)^{-\frac{1}{\alpha}}.$$

Utilizing a discrete convolution then provides another estimate of site occupancy rate, complementing that of Equation (3.14):

$$\widetilde{\Psi}_t(\alpha) = 1 - \sum_{m=0}^K \left\{ \widehat{\mathcal{P}}_{t(m,0)} P_{(\lambda(\alpha), \alpha)}(N_{i,1} = m) \right\}$$

Let $\tilde{\alpha}$ denote the $\alpha > 0$ that minimizes $\sum_t |\hat{\Psi}_t - \widetilde{\Psi}_t|$. Then for $t = 1$ we have $\tilde{P}(N_{i,t} = v) =$

$P_{\lambda(\tilde{\alpha}), \tilde{\alpha}}(N_{i,1} = m)$, and for $t = 2, \dots, T$ we have:

$$\tilde{P}(N_{i,t} = v) = \sum_{m=0}^K \left[\left\{ \prod_{t'=1}^{t-1} \mathcal{P}_{t'} \right\}_{(m,v)} P_{\lambda(\tilde{\alpha}), \tilde{\alpha}}(N_{i,1} = m) \right].$$

Finally, we obtain the PA estimate for total abundance during each season using;

$$\widehat{N}_{\cdot,t}^{(\text{PA})} = R \sum_{v=0}^K v \tilde{P}(N_{i,t} = v).$$

4 SIMULATION STUDY

A simulation study was performed to compare the estimates of total abundance during each season obtained using the DM model and the PA model. Note that while in Section 3 the DM model was developed for settings with the robust design ($J > 1$) and the PA model was developed for cases lacking the robust design ($J = 1$), they can both be employed in the respective opposite setting to yield valid estimates. The purpose of this simulation study is to verify the optimality of using the PA model for $J = 1$ and the DM model for $J > 1$, and to quantify the advantages of doing so.

A total of 36 settings were considered, chosen to represent both a range of typical scenarios and the observed data analyzed in Section 5. For each setting, 500 point count data sets were generated according to the Dail and Madsen (2011) model (henceforth the “PC” model) assuming a Poisson prior for initial abundance g_1 . For use in the PA and DM models, the point counts were then reduced to occupancy observations using $w_{i,t,j} = I(n_{i,t,j} \geq 1)$. Abundance estimates using the point counts with the PC model were also obtained for each generated data set, thus providing a baseline to compare the PA and DM models against.

All 36 settings had $R = 50$ sites and $T = 6$ seasons, with either $J = 4$ or $J = 1$ sampling occasion per season. Eight cases of population dynamics were considered for each level of J , crossing $\gamma = (0.3 \text{ or } 1.0)$; $\omega = (0.4 \text{ or } 0.7)$; and $\lambda = (0.5 \text{ or } 2.0)$. Additionally, two cases of population closure were considered for $J = 1$, with $\gamma = 0$ and $\omega = 1$, and either $\lambda = 0.5$ or 2.0 . Note that

the cases with population closure can be equivalently described as coming from a single season ($T = 1$) with $J = 6$. Finally, for each of these 18 cases, individual detection probability was set to either $r = 0.2$ or 0.5 .

For each generated data set, the mean absolute error (MAE) between estimated and actual total abundance was recorded for each model, with $\text{MAE} = \frac{1}{RT} \sum_t |N_{\cdot,t} - \hat{N}_{\cdot,t}|$. The models were fit using the `optim` procedure in the software package R (R Development Core Team 2009). The numerical solution appeared stable in all cases.

4.1 SIMULATION RESULTS

Simulation results giving the mean and median MAE for all three models (PC, DM, and PA) for the settings with $r = 0.5$ are summarized in Table 2; results for the settings with $r = 0.2$ are similar and are provided in Web Table 1 of the Online Supplement. Note that low median and mean values for MAE indicate better model performance. Additionally, for the settings with $r = 0.5$, $\lambda = 2.0$, and $\gamma = 1$, the smoothed Monte Carlo sampling distributions of MAE for all three models are presented in Figures 1. The other scenarios investigated here yielded similar sampling distributions, as shown for the cases with $r = 0.5$, $\lambda = 1$ or 0.5 , and $\gamma = 1$ or 0.3 in Web Figures 1, 2, and 3 of the Online Supplement.

[Table 2 about here.]

[Figure 1 about here.]

These results show that when the data are collected according to the robust design ($J = 4$), on average the DM model yields estimates of $N_{\cdot,t}$ that are closer to the true values than does the PA model. This ordering is reversed when the data are collected without the robust design ($J = 1$). These orderings are true for every open population setting considered here, with the median MAE using the DM model between 9% and 58% lower when $J = 4$ and between 8% and 173% higher when $J = 1$.

The PC model uniformly provides estimates better than those of either the DM and PA model. This is not surprising, since the PC model considers the generated point counts, as opposed to less

informative occupancy observations considered by the DM and PA models. Consequently, the PC model yields the lowest median MAE by between 3% and 43% when $J = 4$ and between 4% and 78% when $J = 1$.

Although there is separation in mean and median MAE between the PA and DM models, Figure 1 and Web Figures 1, 2, and 3 show that there is substantial overlap between the overall MAE sampling distributions. Therefore, on any given data set, regardless of whether $J = 1$ or $J > 1$ it is possible to have either the DM or PA model provide the lower MAE value. For each of the 32 open population settings considered here, we recorded the proportion of generated data sets (out of 500) with larger MAE values from DM model than from the PA model. Box plots of these proportions, grouped by J , are presented in Figure 2.

[Figure 2 about here.]

For $J = 4$, the DM model yields a lower MAE value than the PA model for more than 70% of generated data sets in 15 of 16 settings, with the exception being $(\lambda = 0.5, \gamma = 1, \omega = 0.4, r = 0.5)$ with 57%. The PA model does not fare as well in the settings with $J = 1$ as the DM does in settings with $J = 4$, however. For $J = 1$, the PA model yields a lower MAE value than the DM model for more than 50% of generated data sets in all but two settings: $(\lambda = 2.0, \gamma = 1, \omega = 0.7, r = 0.5)$ with 49.6% and $(\lambda = 2.0, \gamma = 0.3, \omega = 0.7, r = 0.2)$ with 46%.

5 APPLICATION TO AVIAN PRESENCE-ABSENCE DATA

One of the most comprehensive bird monitoring programs is the North American Breeding Bird Survey (BBS), consisting of more than 4100 routes spread throughout the United States and Canada. Each route has 50 sites spread one-half mile apart, where on a single day each year a skilled volunteer will count the number of every bird species detected within a three minute time interval. This has occurred on an annual basis at many routes since 1966. More information on the

BBS is provided in Robbins, Bystrak, and Geissler (1986).

Here, we consider six seasons of observations from a BBS route in western Oregon, 2004 through 2009. Note that with a single survey occasion per season, the data setting is $T = 6, J = 1$. In total, during these six seasons there were 79 different species detected along this route. Here we restrict our attention to five of these species, selected for having sparse observed counts with at least one detection per year, and to represent a range of detection probabilities. These five species are the Western Wood-Pewee (*Contopus sordidulus*), House Finch (*Carpodacus mexicanus*), American Goldfinch (*Spinus tristis*), Song Sparrow (*Melospiza melodia*), and Spotted Towhee (*Pipilo maculatus*). Observations for the last five sites for the House Finch are given in Table 1.

Since the BBS takes precautions against double counting or misidentifications, the assumptions of site independence and independent detections within a site seem plausible, as does the assumption that $n_{i,t}$ is a binomial with index $N_{i,t}$ and detection probability r . Therefore, it is justifiable to use the PC model to obtain estimates of seasonal abundance.

We fit the PC models with either a negative binomial (λ, α) or Poisson (λ) prior for g_1 . Since these priors converge as $\alpha \rightarrow \infty$, we use the modified asymptotic likelihood ratio test of Self and Liang (1987, case 5) with level 0.05 to select between them. Due to a lack of measured covariates and issues for non-identifiability when parameters are allowed to vary seasonally, we assume constant r, γ, ω and λ . Despite the simplicity of these models, the parametric bootstrap procedure (Dixon 2002) gave no evidence of lack of fit for any of the five species (p-values 0.35, 0.18, 0.21, 0.41, and 0.37, respectively).

Although point counts are collected in the BBS, the focus here is on models that consider presence-absence observations. Therefore, occupancy data are obtained using $w_{i,t} = I(n_{i,t} \geq 1)$ and used to fit the DM and PA models. Since the PC model considers data with more information, it is used here as a benchmark to compare the DM and PA models against.

The DM model relies on the assumption of independent detections of individuals in a site to use Equation (1.1) to obtain abundance estimates from occupancy observations. Otherwise, the DM model is selected in the same way as the PC model, with the modified likelihood ratio test of

Self and Liang (1987) used to choose between the negative binomial and Poisson prior for g_1 , with all parameters constant.

The PA model relies on the assumption that individuals emigrate immigrate independently. It uses a negative binomial prior for the initial distribution of individuals to sites, since this accommodates the Poisson prior in the limit as $\alpha \rightarrow \infty$. Here, as with the PC and DM models, we assume all the PA model parameters are constant.

For the PC and DM models, results from the likelihood ratio test for g_1 and estimates and asymptotic 95% confidence intervals for λ , α , γ , ω , and r are provided in Web Tables 2 and 3. Estimates and asymptotic 95% confidence intervals for Ψ_1 , τ , ϵ , and p for the PA model are given in Web Table 4. As expected, the five avian species considered here represent a range of detection probabilities, with \hat{r} ranging from 0.21 to 0.96 for the PC model. In addition, there is a range of population dynamics represented, with the PC model estimating high turnover for the Goldfinch ($\gamma = 1.16$, $\omega = 0.11$) and low turnover for the Wood-Pewee ($\gamma = 0.06$, $\omega = 0.97$).

Estimates of seasonal abundance $N_{\cdot,t}$ for all three models are provided in Table 3; bootstrap 95% confidence intervals are provided in Web Table 5. Also included in Table 3 is the mean absolute difference (MAD) between the abundance estimates using the PC model and those of either the DM or PA model, where MAD for the DM model is $\frac{1}{300}|\hat{N}_{\cdot,t}^{PC} - \hat{N}_{\cdot,t}^{DM}|$.

[Table 3 about here.]

The MAD column of Table 3 indicates that for three out of five species, the estimates of $N_{\cdot,t}$ closest (in MAD) to those of the PC model are provided by the PA model. This is consistent with the results of the simulation study in Section 4: the results here (with PA better in 60% of cases) falls in the second quartile of the box plot with $J = 1$ in Figure 2. Furthermore, the DM model gives median MAD 238% higher than that of the PA model (1.69 and 0.50, respectively). While this is outside the range of 8% to 173% given for the simulation results in Section 4.1, it is consistent in the direction with the PA model giving better seasonal abundance estimates than the DM model in the absence of the robust design ($J = 1$).

6 DISCUSSION

Both proposed models DM and PA exploit the observation that occupancy detections tend to occur with greater probability in sites where animals are located in high density than in sites where they are sparsely located. The specific choice to model this relationship with Equation (1.1) is the basis for the DM model, with similar arguments related to individual immigration and emigration serving as the basis for the PA model. Here we re-emphasize that the choice to employ Equation (1.1) is indeed a choice, as alternative relationships between individual detection probability and occupancy are available (Sileshi et al. 2006). Although a comparison between these competing models is beyond the scope of this paper, Royle (2006) demonstrates that such comparisons are likely futile, as data alone cannot identify between the competing models. Therefore, the decision to utilize either the PA or the DM model must be grounded in a biological argument justifying the choice to relate occupancy to abundance using Equation (1.1).

Once this choice is made, we have demonstrated here that the DM model performs best in cases with the robust design ($J > 1$ sampling occasion per season) and the PA model is best in cases lacking the robust design ($J = 1$). However, these models rely on two further assumptions: (1) individuals enter and leave sites independently, and (2) sites are independent. These are briefly investigated below.

First, we note that the assumption that individuals enter and leave sites independently is vital to the PA model, as it forms the basis for connecting occupancy to abundance. In the DM model, it is important for $S_{i,t}$ to be binomial and $G_{i,t}$ to be Poisson. A framework that relaxes this assumption was proposed by Royle (2008), where instead of detecting individual animals, clusters of animals are detected and counted. Work is ongoing to extend this framework to accommodate repeated presence-absence data.

Second, the likelihoods for the both the DM and PA models, respectively Equations (3.5) and (3.13), rely on the assumption that sites are independent. This assumption is less restrictive than it may appear, though, since site-specific covariates can be included in the models through Equation (2.3). However, if sites are adjacent, individuals leaving one must enter another, and this dependent

structure is unlikely to be fully accounted for by a covariate effect. One possible solution being investigated is the use of Bayesian hierarchical models, combining the spatial models developed by Royle et al. (2007) with the dynamic presence-absence modeling framework of Royle and Kéry (2007).

6.1 DISCUSSION OF SIMULATION RESULTS

Results from the simulation study show that the DM model yields better abundance estimates than the PA model when $J = 4$, with the PA model outperforming the DM model when $J = 1$. Further study is needed to compare the models in settings with J between 1 and 4, such as $J = 2$ or a combination where $J = 2$ for some seasons and $J = 1$ for the rest. Since the DM model is constructed to utilize the robust design to estimate r , it is likely that any design with $J > 1$ will favor the DM model. However, this should be tested, and investigations seeking to quantify the direction and magnitude of difference between the DM and PA models in these settings will be useful for both designing optimal future studies and choosing the proper modeling framework with which to analyze data arising from such studies.

Here, we investigated a number of situations typical of moderately abundant populations. In these settings, therefore, the total number of occupancy detections never approached the total number of sampling occasions (RTJ). If this were to happen, neither the PA nor the DM models would be identifiable, as the information contained in w is partly derived from the changes between detected occupancy status within seasons, between seasons, and between sites. In such cases, only a framework that considers point counts (such as the PC model) can be used to estimate seasonal abundance.

A possible criticism for this simulation study is that the data were generated only according to the PC model. Since the DM and PC models are so similar, it is therefore not surprising that the DM model performs well for the robust design when r is readily estimated. However, the PC model is quite flexible in modeling different forms of population dynamics, such as the autoregressive model described in Dail and Madsen (2011). Furthermore, since these different population dynamics

models only involve γ and ω , they are all available to the DM model as well. It is due to space considerations that these different approaches are not described and investigated here; although some attention is given to them in Dail and Madsen (2011), more extensive investigations are warranted.

6.2 DISCUSSION OF THE APPLICATION

Estimates of seasonal abundance were obtained for five species of birds detected along a transect of the BBS in western Oregon between 2004 and 2009. This was done using each model, with the results for the DM and PA models compared to that from the PC model. The PA model yielded closer estimates for three of the five species, supporting the simulation study for the case $J = 1$.

Although detecting and estimating a trend in abundance is beyond the scope of this paper, it is interesting to note that for four of the species considered, both the DM and PA models estimate the trend in the same general direction (increasing, flat, decreasing) as the PC model. Further study is warranted to investigate the ability of the DM and PA models to accurately detect and estimate trend in abundance.

Due to a lack of available data, no settings with $J > 1$ were considered here. A spatially replicated study collecting point counts (or, even better, individual mark recapture histories) via the robust design would be valuable to verify or refute the findings of the simulation study for the case $J > 1$.

ONLINE RESOURCES

Web_Appendix.pdf: Tables and Figures referenced in Sections 4.1 and 5.

Web_Supplement_Files.R: R functions to fit PA, DM and PC models obtain seasonal abundance estimates.

Tutorial.txt: Explanation of how to utilize the above mentioned R functions to fit each model and obtain estimates of seasonal abundances.

REFERENCES

- Dail, D. and Madsen, L. (2011), “Models for estimating abundance from repeated counts of an open metapopulation,” *Biometrics*, 67, 577–587.
- (2012), “Estimating open population site occupancy rates from presence-absence data lacking the robust design,” *Biometrics*, in press.
- Dixon, P. M. (2002), “Bootstrap resampling,” *Encyclopedia of Environmetrics*, 1, 212–220.
- Dreitz, V. J., Lukacs, P. M., and Knopf, F. L. (2006), “Monitoring low density avian populations: an example using mountain plovers,” *The Condor*, 108, 700–706.
- Hartley, S. (1998), “A positive relationship between local abundance and regional occupancy is almost inevitable (but not all positive relationships are the same),” *Journal of Animal Ecology*, 67, 992–994.
- He, F. and Gaston, K. J. (2000), “Occupancy-abundance relationships and sampling scales,” *Ecography*, 23, 503–511.
- Hood, S. L. and Dinsmore, S. J. (2007), “Abundance of Snowy and Wilson’s Plovers in the lower Laguna Madre region of Texas,” *Journal of Field Ornithology*, 78, 362–368.
- Joseph, L. N., Field, S. A., Wilcox, C., and Possingham, H. P. (2006), “Presence-absence versus abundance data for monitoring threatened species,” *Conservation Biology*, 20, 1679–1687.
- Kendall, W. L. (1999), “Robustness of closed capture-recapture methods to violations of the closure assumption,” *Ecology*, 80, 2517–2525.
- Kéry, M. (2008), “Estimating abundance from bird counts: binomial mixture models uncover complex covariate relationships,” *Auk*, 125, 336–345.
- Long, R., Donovan, T., MacKay, P., Zielinski, W., and Buzas, J. (2010), “Predicting carnivore occurrence with noninvasive surveys and occupancy modeling,” *Landscape Ecology*, 1–14, URL <http://dx.doi.org/10.1007/s10980-010-9547-1>.
- MacKenzie, D., Seamans, M., Gutierrez, R., and Nichols, J. (2010), “Investigating the population dynamics of california spotted owls without marked individuals,” *Journal of Ornithology*, 1–8,

URL <http://dx.doi.org/10.1007/s10336-010-0544-6>. 10.1007/s10336-010-0544-6.

- MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., and Franklin, A. B. (2003), “Estimating site occupancy, colonization and local extinction probabilities when a species is not detected with certainty,” *Ecology*, 84, 2200–2207.
- Pollock, K. H. (1982), “A capture-recapture design robust to unequal probability of capture,” *Journal of Wildlife Management*, 46, 757–760.
- Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L., and Sauer, J. R. (2002), “Large scale wildlife monitoring studies: Statistical methods for design and analysis,” *Environmetrics*, 13, 105–119.
- R Development Core Team (2009), *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- Rizkalla, C. E. and Swihart, R. K. (2006), “Community structure and differential responses of aquatic turtles to agriculturally induced habitat fragmentation,” *Landscape Ecology*, 21, 1361–1375.
- Robbins, C. S., Bystrak, D. A., and Geissler, P. H. (1986), *The Breeding Bird Survey: Its First Fifteen Years (1965-1979)*, vol. 157, Washington, D.C.: USDOI, Fish and Wildlife Service Resource Publication.
- Rota, C. T., Fletcher, J., Robert J., Dorazio, R. M., and Betts, M. G. (2009), “Occupancy estimation and the closure assumption,” *Journal of Applied Ecology*, 46, 1173–1181.
- Royle, J. A. (2004), “N-mixture models for estimating population size from spatially replicated counts,” *Biometrics*, 60, 108–115.
- (2006), “Site occupancy models with heterogeneous detection probabilities,” *Biometrics*, 62, 97–102.
- (2008), “Hierarchical modeling of cluster size in wildlife surveys,” *Journal of Agricultural, Biological, and Environmental Statistics*, 13, 23–36.

- Royle, J. A. and Kéry, M. (2007), “A Bayesian state-space formulation of dynamic occupancy models,” *Ecology*, 88, 1813–1823.
- Royle, J. A., Kéry, M., Gautier, R., and Schmid, H. (2007), “Hierarchical spatial models of abundance and occurrence from imperfect survey data,” *Ecological Monographs*, 77, 465–481.
- Royle, J. A. and Nichols, J. D. (2003), “Estimating abundance from repeated presence-absence data or point counts,” *Ecology*, 84, 777–790.
- Schaub, M., Gimenez, O., Schmidt, B. R., and Pradel, R. (2004), “Estimating survival and temporary emigration in the multistate capture-recapture framework,” *Ecology*, 85, 2107–2113.
- Self, S. G. and Liang, K.-Y. (1987), “Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions,” *Journal of the American Statistical Association*, 82, 605–610.
- Sileshi, G., Hailu, G., and Mafongoya, P. (2006), “Occupancy-abundance models for predicting densities of three leaf beetles damaging the multipurpose tree *Sebania sesban* in eastern and southern Africa,” *Bulletin of Entomological Research*, 96, 61–69.
- Thompson, S. K. (1992), *Sampling*, New York: John Wiley.
- Tipton, H. C. (2007), “Occupancy, abundance, and density of Colorado breeding grassland birds: estimation and habitat correlations,” Dissertation, Colorado State University.
- Yoccoz, N. G., Nichols, J. D., and Boulinier, T. (2001), “Monitoring of biological diversity in space and time,” *Trends in Ecology and Evolution*, 16, 446–453.

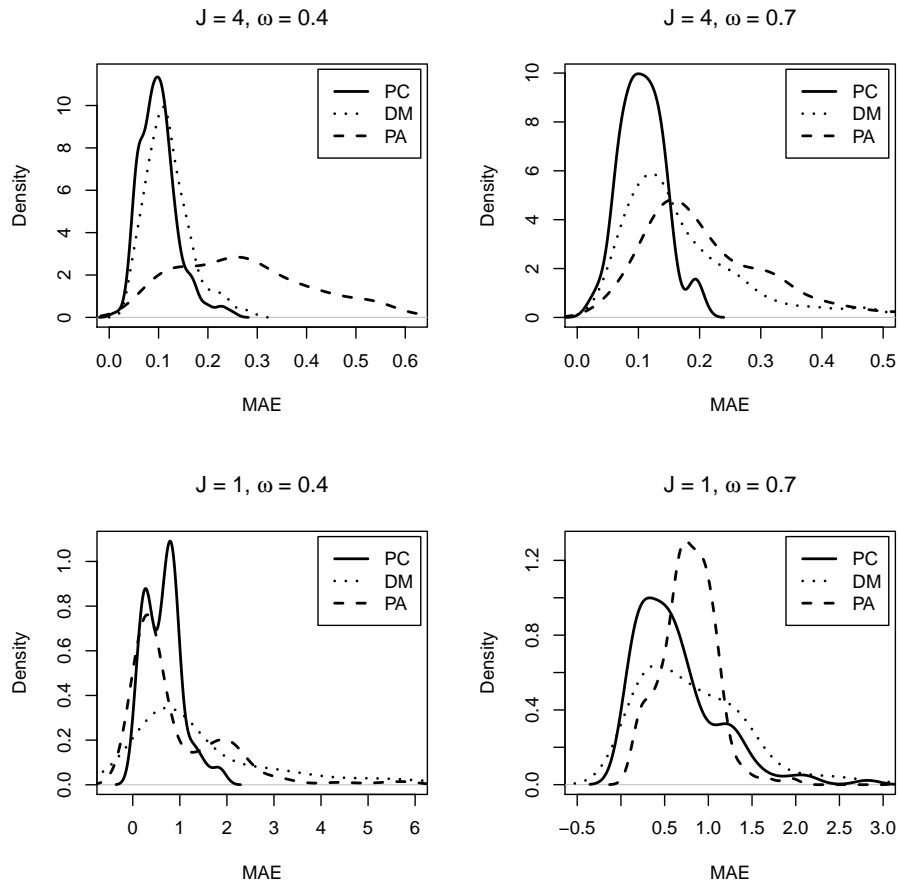


Figure 1. Smoothed Monte Carlo sampling distributions for MAE in $\widehat{N}_{i,t}$ using the models PC, DM and PA, from 500 data sets generated with $N = 50$ sites, $T = 6$ survey periods, initial abundance rate $\lambda = 2.0$, arrival rate $\gamma = 1$, and detection probability $r = 0.50$. The apparent positive density for negative MAE is an artifact of kernel smoothing.

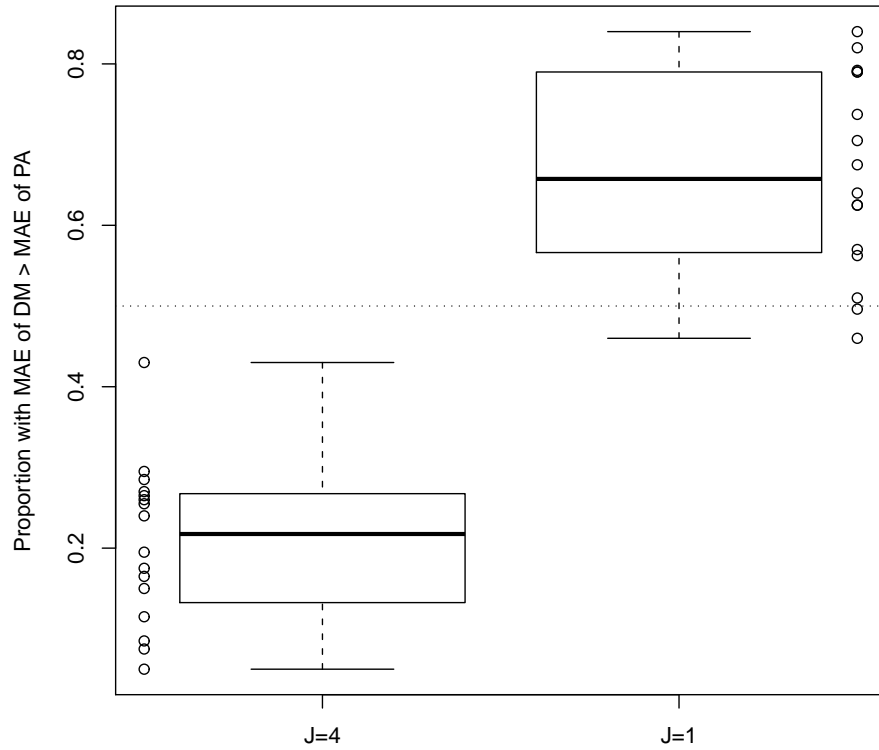


Figure 2. Box plots of the proportion of data sets with MAE in $\hat{N}_{\cdot,t}$ larger for DM model than PA model. J gives the number of samples per site per period; only scenarios with dynamic populations are considered here, so each group contains 16 points, representing 16 different scenarios of 500 data sets each.

Table 1. Yearly presence-absence data for the House Finch from the last five sites (stops) along a BBS route in Oregon; the observed point counts are in parenthesis.

	Survey year					
	2004	2005	2006	2007	2008	2009
Site 46	1 (1)	0 (0)	1 (1)	1 (0)	0 (0)	1 (1)
Site 47	0 (0)	0 (0)	0 (0)	1 (1)	0 (0)	1 (1)
Site 48	1 (4)	1 (3)	1 (4)	1 (2)	1 (2)	1 (1)
Site 49	0 (0)	0 (0)	1 (1)	0 (0)	1 (2)	1 (1)
Site 50	1 (2)	0 (0)	0 (0)	0 (0)	1 (2)	0 (0)

Table 2. Summaries of the sampling distributions of mean absolute error (MAE) in estimated total seasonal abundances, with estimates provided by the PC, DM, or PA model. Sampling distributions are based on 500 simulated data sets with $R = 50$ sites and $T = 6$ seasons. Results are shown here for cases with individual detection probability $r = 0.50$.

J	γ	ω	λ	MAE \hat{N}					
				Mean			Median		
				DM	PA	PC	DM	PA	PC
4	1.0	0.4	0.5	0.12	0.27	0.10	0.11	0.26	0.10
4	1.0	0.4	2.0	0.17	0.38	0.12	0.15	0.32	0.11
4	1.0	0.7	0.5	0.17	0.21	0.11	0.14	0.19	0.11
4	1.0	0.7	2.0	0.28	0.42	0.13	0.21	0.33	0.12
4	0.3	0.4	0.5	0.06	0.09	0.06	0.06	0.08	0.05
4	0.3	0.4	2.0	0.09	0.18	0.07	0.09	0.17	0.07
4	0.3	0.7	0.5	0.07	0.10	0.06	0.07	0.09	0.06
4	0.3	0.7	2.0	0.13	0.22	0.08	0.12	0.21	0.08
1	1.0	0.4	0.5	4.32	0.64	0.50	0.65	0.24	0.22
1	1.0	0.4	2.0	5.36	1.00	0.65	0.93	0.42	0.37
1	1.0	0.7	0.5	1.16	0.49	0.45	0.52	0.48	0.41
1	1.0	0.7	2.0	1.41	0.82	0.64	0.84	0.78	0.53
1	0.3	0.4	0.5	2.21	1.46	0.91	1.82	1.39	0.49
1	0.3	0.4	2.0	1.75	0.97	0.36	1.36	1.25	0.27
1	0.3	0.7	0.5	0.25	0.15	0.14	0.17	0.14	0.13
1	0.3	0.7	2.0	0.65	0.51	0.30	0.52	0.39	0.21
1	0.0	1.0	0.5	0.05	0.07	0.04	0.05	0.07	0.03
1	0.0	1.0	2.0	0.27	0.51	0.14	0.22	0.54	0.12

Table 3. Estimated total seasonal abundance ($\hat{N}_{\cdot,t}$) for five species encountered along a BBS route. “MAD PC” denotes the mean absolute difference in seasonal abundance estimates from the given model to the PC model.

Species	Model	$\hat{N}_{\cdot,1}$	$\hat{N}_{\cdot,2}$	$\hat{N}_{\cdot,3}$	$\hat{N}_{\cdot,4}$	$\hat{N}_{\cdot,5}$	$\hat{N}_{\cdot,6}$	MAD PC
Wood-Pewee	PC	92.02	91.65	91.30	90.96	90.63	90.31	.
Wood-Pewee	DM	74.80	70.90	67.25	63.83	60.63	57.63	0.51
Wood-Pewee	PA	85.56	59.81	48.89	42.60	38.44	35.45	0.79
Finch	PC	29.02	27.36	26.95	26.85	26.82	26.82	.
Finch	DM	33.13	33.76	34.15	34.38	34.53	34.62	0.14
Finch	PA	30.35	29.58	29.99	30.35	30.58	30.71	0.06
Goldfinch	PC	45.18	62.53	64.36	64.55	64.57	64.58	.
Goldfinch	DM	156.94	151.72	146.97	142.65	138.73	135.17	1.69
Goldfinch	PA	120.85	82.90	72.02	67.48	65.29	64.15	0.36
Sparrow	PC	119.48	108.16	99.38	92.59	87.32	83.24	.
Sparrow	DM	609.11	536.94	477.01	427.25	385.94	351.64	7.33
Sparrow	PA	98.13	73.41	68.47	67.04	66.55	66.35	0.50
Towhee	PC	136.64	152.49	168.34	184.18	200.03	215.88	.
Towhee	DM	54.70	60.96	67.23	73.49	79.75	86.01	2.12
Towhee	PA	43.89	47.89	51.88	55.61	58.98	61.99	2.46