Supplementary materials for this article are available at https://doi.org/10.1007/s13253-021-00484-w.

Improving Wildlife Population Inference Using Aerial Imagery and Entity Resolution

Xinyi LU, Mevin B. HOOTEN, Andee KAPLAN, Jamie N. WOMBLE, and Michael R. BOWER

Recent technological advancements have seen a rapid growth in the use of imagery data to estimate the abundance and spatial distribution of animal populations. However, the value of imagery data may not be fully exploited under traditional analytical frameworks. We developed a method that leverages aerial imagery data for population modeling through entity resolution, a technique that stochastically links the same individual across multiple images. Resolving duplicate individuals in overlapping images that are distorted requires realigning observed point patterns optimally; however, popular machine learning algorithms for image stitching do not often account for alignment uncertainty. Moreover, duplicated individuals can provide insight about detection probability when overlaps are viewed as replicate surveys. Our model resolves individual identities by linking observed locations to latent activity centers and estimates total population as informed by the linkage structure. We developed a hierarchical framework to achieve entity resolution and abundance estimation cohesively, thereby avoiding single-direction error propagation that is common in two-stage models. We illustrate our method through simulation and a case study using aerial images of sea otters in Glacier Bay, Alaska.

Supplementary materials accompanying this paper appear on-line

Key Words: Bayesian; Data augmentation; Hierarchical model; Spatial capture-recapture.

1. INTRODUCTION

Aerial surveys are widely used to provide abundance information about terrestrial and marine species (Caughley 1974; Ver Hoef 2014). Compared to traditional observer-based surveys, imagery surveys have the advantage of reducing risk for observers and providing a permanent record that can be independently verified (Buckland et al. 2012). In addition to

X. Lu (✉)· A. Kaplan, Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA (E-mail: *xinyi.lu@colostate.edu*). M. B. Hooten · M. R. Bower, Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, TX 78705, USA. J. M. Womble · M. R. Bower, Southeast Alaska Inventory and Monitoring Network, National Park Service, 3100 National Park Rd, Juneau, AK 99801, USA. J. M. Womble, Glacier Bay Field Station, National Park Service, 3100 National Park Rd, Juneau, AK 99801, USA.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 27, Number 2, Pages 364–381
https://doi.org/10.1007/s13253-021-00484-w

population counts, the imagery data (often referred to as photographs; Fig. 3 in Supplementary Material) provide individual-level information such as color, size, and location, which can be leveraged to identify animals without marking them (Williams et al. 2020). This gain in information leads to more reliable modeling of population abundance than using count data only (Dennis et al. 2015; Barker et al. 2018; Ketz et al. 2019). In what follows, we describe a Bayesian hierarchical model to identify unique individuals in overlapping images and estimate population size under a unified framework. We apply our model to analyze aerial imagery data of sea otters (*Enhydra lutris kenyoni*) in Glacier Bay, Alaska. During a survey, images are acquired at a regular time interval with overlapping regions in the direction of aircraft movement as it flies along transects that are systematically placed across the Glacier Bay. Sea otters in the images are located and counted by trained observers after the survey. Past studies using these data have either discarded overlapping images to meet the independent count assumption of binomial models (Lu et al. 2019), or treated counts from overlapping regions as temporal replicates in N-mixture models (Williams et al. 2017). We demonstrate the advantages of our method over the previously described methods in simulation.

The information we use to resolve individual identities are the observed locations of individuals in a sequence of images. However, individual positions may be distorted when the aircraft deviates from its scheduled trajectory due to a variety of reasons that can influence altitude and aircraft position, resulting in an artificial transformation of the image footprints. Further, micro-movement of sea otters and locating uncertainty during laboratory processing make exact matching of observed locations in overlapping regions nearly impossible. There exists a rich literature on image stitching where the common objective is to optimally combine a sequence of overlapping images into a composite image by minimizing a loss function (Levin et al. 2004; Szeliski 2006; Brown and Lowe 2015; Gross and Heumann 2016). However, optimization-based image stitching algorithms do not usually provide uncertainty about the stitching process and are seldom integrated into other models to provide additional learning about the system. On the other hand, the statistical literature associated with entity resolution, also known as record linkage when the objective is to merge multiple data files (in our case, images) in the absence of unique identifiers (in our case, individual tags, for example), may provide a theoretical basis for uncertainty quantification. We incorporate uncertainty in the record linkage process into a capture-recapture model for abundance estimation.

Traditional approaches to record linkage compare similarities between pairs of records from which matching decisions are made (Fellegi and Sunter 1969; Jaro 1989; Winkler 1995). Larsen and Rubin (2001) presented record linkage as a mixture of linkage probabilities between a model for probable links and a model for probable nonlinks. Fortini et al. (2001), McGlincy (2004), and Larsen (2004) developed the Bayesian approaches based on the same idea. However, comparison-based approaches are largely infeasible computationally, even when the number of possible links is moderately large (Winkler 2006). One way to reduce the computation cost of record linkage is by "blocking," where records partitioned into different blocks are considered nonlinks *a priori* (Christen 2011; Steorts et al. 2014). Alternatively, record linkage can be presented as the clustering of observed records by unobserved identities (Copas and Hilton 1990; Tancredi and Liseo 2011; Liseo and Tan-

credi 2011; Steorts et al. 2015; Tancredi et al. 2018). Each latent identity has a "true" value and the associated records are modeled as stochastic distortions from the truth. Steorts et al. (2015) introduced the graphical record linkage model by representing the linkage structure as a bipartite graph between observed records and latent identities. By comparing records to latent identities instead of each other, the computation time to link d data files with a maximum of n records per file can be substantially reduced from $\mathcal{O}\left(n^d\right)$ to $\mathcal{O}(dn)$. One distinction between the graphical record linkage model and other non-parametric clustering methods such as Dirichlet process models and Pitman-Yor process models is that the latter often assume linear growth of cluster size with the size of data (Wallach et al. 2010; Betancourt et al. 2016), whereas in record linkage problems, co-referent clusters tend to stay small even when the number of records grows. Following Liseo and Tancredi (2011) and Steorts et al. (2015), we made use of a bivariate Gaussian model conditional on the latent truths to identify unique individuals in the imagery data.

The output of a record linkage model can be used to learn about population size. When uncertainty exists in linkage structure, record linkage and size estimation are often regarded as two separate stages (LaPorte et al. 1993; Anderson and Fienberg 1999; Lum et al. 2013). Sadinle (2018) proposed using "linkage-averaging" to transfer linkage uncertainty as quantified by Bayesian posterior samples into the subsequent stage of population size estimation. Although linkage-averaging facilitates model exploration by allowing the combination of different record linkage models with population models, any bias in the record linkage stage will propagate into the size estimation stage regardless of model choice (Tancredi and Liseo 2011). Our hierarchical framework naturally relates entity resolution and abundance estimation as one generative process, thereby allowing information exchange and feedback between these two model objectives. Other unified modeling approaches exist, including those presented by Link et al. (2009) and Wright et al. (2009) that incorporate misidentification into capture-recapture models by sampling from latent multinomial distributions, the hierarchical record linkage models proposed by Tancredi and Liseo (2011) and Liseo and Tancredi (2011) that reflect capture-recapture dynamics through latent matching matrices, and the latent Poisson process model proposed by Green and Mardia (2005) to align partially labeled protein structures. We propose a novel framework that combines a record linkage model and a spatial capture-recapture model (Royle and Young 2008) to align distorted animal locations and to account for heterogeneity in detection probability due to temporally changing survey units.

We present our hierarchical record linkage model in Sect. 2. In Sect. 3, we illustrate the model through simulation and a case study using aerial photographs of sea otters in Glacier Bay, Alaska. Finally in Sect. 4, we discuss possible extensions and broader applications of our model.

2. MODEL

2.1. DATA MODEL

Consider a sequence of T images with n_t observed individuals in image t, for $t = 1, \ldots, T$ (see Fig. 2 in Supplementary Material, for example). Let $\boldsymbol{y}_{i,t}$ be a two-dimensional

vector of latitude and longitude denoting the observed location of the ith individual in image t, and let $\boldsymbol{u}_{i,t}$ denote the true location of that individual. Distortion in $\boldsymbol{y}_{i,t}$ occurs in laboratory processing when the image footprint, $\mathcal{F}_t$, is artificially scaled and rotated to fit in a template, $\mathcal{Q}_t$, assuming the aircraft trajectory follows a fixed height and orientation. The image centers (latitude, longitude) were recorded by a GPS device on the aircraft in real time and are reliable to represent the truth. Using the known image center $\boldsymbol{\mu}_t$ as a reference point, we connect the distorted displacement of the observed location from the image center to that of the true location, $\boldsymbol{u}_{i,t}$, from the image center as

$$\boldsymbol{y}_{i,t} - \boldsymbol{\mu}_t = (1 + c_t)\, \boldsymbol{R}\,(\theta_t)\left(\boldsymbol{u}_{i,t} - \boldsymbol{\mu}_t\right), \tag{1}$$

where the counterclockwise rotation matrix is given by

$$\boldsymbol{R}\,(\theta_t) = \begin{pmatrix} \cos\theta_t & -\sin\theta_t \\ \sin\theta_t & \cos\theta_t \end{pmatrix}.$$

The scaling parameter, c_t, and the rotation parameter, θ_t, are modeled using basis function regression to ensure smoothness and flexibility in the aircraft trajectory (Hefley et al. 2017). We specify

$$\begin{aligned} c_t &= \boldsymbol{w}(t)'\boldsymbol{\alpha}, \\ \theta_t &= \boldsymbol{v}(t)'\boldsymbol{\beta}, \end{aligned} \tag{2}$$

where $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$ are the basis functions evaluated at time t for scaling and rotation, respectively. Due to unknown distortion, the true image footprints are also unknown, and we model the four vertices of the rectangular image footprint $\mathcal{F}_t$ through a georectification process from the known template $\mathcal{Q}_t$,

$$\boldsymbol{v}_{j,t} - \boldsymbol{\mu}_t = \frac{1}{(1 + c_t)}\boldsymbol{R}\,(-\theta_t)\left(\boldsymbol{v}^*_{j,t} - \boldsymbol{\mu}_t\right), \quad j = 1, 2, 3, 4, \tag{3}$$

where $\boldsymbol{v}_{j,t}$ denote the vertices of $\mathcal{F}_t$ and $\boldsymbol{v}^*_{j,t}$ denote the vertices of $\mathcal{Q}_t$.

We assume every observed individual has a latent identity, $\lambda_{i,t}$, that may be shared across images but not within the same image. The true locations, $\boldsymbol{u}_{i,t}$, are modeled as Gaussian conditioned on a transient activity center associated with the latent identity, $\boldsymbol{s}_{\lambda_{i,t}}$, and movement uncertainty, $\sigma^2_u \boldsymbol{I}$, such that

$$\boldsymbol{u}_{i,t} | \boldsymbol{s}_{\lambda_{i,t}}, \sigma^2_u \sim \mathrm{N}\left(\boldsymbol{s}_{\lambda_{i,t}}, \sigma^2_u \boldsymbol{I}\right).$$

Subsequently, the conditional distributions of the observed locations are expressed as follows,

$$\boldsymbol{y}_{i,t} | \boldsymbol{s}_{\lambda_{i,t}}, \sigma^2_u, c_t, \theta_t \sim \mathrm{N}\left(\boldsymbol{\mu}_t + (1 + c_t)\, \boldsymbol{R}\,(\theta_t)\left(\boldsymbol{s}_{\lambda_{i,t}} - \boldsymbol{\mu}_t\right), \sigma^2_u (1 + c_t)^2\, \boldsymbol{R}\,(\theta_t)\, \boldsymbol{R}^T\,(\theta_t)\right). \tag{4}$$

We note that reliable inference from our model depends on a small σ_u^2 relative to the amount of distortion due to scaling and rotation, and we return to this concept in Sect. 3.1. Based on the latent identities, the data model in (4) allows us to minimize the Procrustes distance (Dryden and Mardia 1998) between configurations of points in the overlapping regions, and the process model that we describe in what follows enables inference about the latent identities.

2.2. PROCESS MODEL

We adopt a parameter expanded data augmentation approach (Royle 2009; Royle and Dorazio 2012) and assume there is a super-population of size M much greater than the total number of observations in a study domain, $\mathcal{D}$, that contains the union of all image footprints. Each individual in the super-population has a binary variable, z_m, representing whether the individual belongs to the population being sampled, where $z_m \sim \text{Bern}(\psi)$ for $m = 1, \ldots, M$. Conditional on the latent identities of the observed individuals, the augmented data are a zero-inflated version of the capture history. The prior specification on the zero-inflation parameter, ψ, along with the super-population size, M, implicitly suggests a prior for the unknown population size, N (Royle et al. 2007).

We let $\boldsymbol{\lambda}_t$ denote the vector of latent identities indexed by m for the observed individuals in image t. A plausible configuration of $\boldsymbol{\lambda}_t$ must satisfy two conditions: (a) there are no duplicate identities, and (b) any identity in $\boldsymbol{\lambda}_t$ must be detectable at time t. Otherwise the probability of observing $\boldsymbol{\lambda}_t$ is zero. Each individual in the super-population is associated with an activity center, $\boldsymbol{s}_m$. We let the activity centers be uniformly distributed in the study domain *a priori*. We require that an individual is detectable at time t if and only if it is a member of the population being sampled ($z_m = 1$) and its realized location is inside the image footprint at time t ($\boldsymbol{u}_{m,t} \in \mathcal{F}_t$). In the spatial capture-recapture model by Royle and Young (2008), realized locations are fully augmented for all individuals in the super-population and unobserved $\boldsymbol{u}_{m,t}$ are treated as missing data (the model does not account for measurement error so the observations are the realized locations). However, when the observed individuals are unidentified, accounting for missingness becomes challenging. Therefore, we integrate $\boldsymbol{u}_{m,t}$ from the process model by letting $p_{m,t}$ denote the probability that $\boldsymbol{u}_{m,t}$ falls in $\mathcal{F}_t$ conditional on $\mathcal{F}_t$, the activity center $\boldsymbol{s}_m$, and the movement process variance σ_u^2, such that

$$p_{m,t} = \mathbb{P}\left(\boldsymbol{u}_{m,t} \in \mathcal{F}_t \middle| \mathcal{F}_t, \boldsymbol{s}_m, \sigma_u^2\right) = \int_{\mathcal{F}_t} \frac{1}{2\pi\sigma_u^2} \exp\left(-\frac{(\boldsymbol{u}-\boldsymbol{s}_m)'(\boldsymbol{u}-\boldsymbol{s}_m)}{2\sigma_u^2}\right) \mathrm{d}\boldsymbol{u}. \quad (5)$$

Let p_0 denote the baseline detection probability (e.g., sea otter detectability due to diving behaviors). Then we have

$$\mathbb{P}\left(\lambda_{i,t} = m \middle| z_m, \boldsymbol{s}_m, \mathcal{F}_t, \sigma_u^2, p_0\right) = \begin{cases} p_0 \times p_{m,t}, & \text{if } z_m = 1; \\ 0, & \text{otherwise.} \end{cases}$$

Assuming the individuals are independently detected, the probability of observing $\boldsymbol{\lambda}_t$ is as follows,

$$\mathbb{P}\left(\boldsymbol{\lambda}_t \middle| \{z_m\}_{m=1}^{M}, \{\boldsymbol{s}_m\}_{m=1}^{M}, \sigma_u^2, \mathcal{F}_t, p_0\right) = \frac{1}{n_t!} \prod_{m:z_m=1} \left\{p_0 p_{m,t} \mathbb{I}\left(m \in \boldsymbol{\lambda}_t\right) + \left(1 - p_0 p_{m,t}\right) \mathbb{I}\left(m \notin \boldsymbol{\lambda}_t\right)\right\}, \tag{6}$$

where the factor of $\frac{1}{n_t!}$ indicates that all permutations of $\boldsymbol{\lambda}_t$ are equally likely *a priori*. The process model induces regularization on the number of unique latent identities by controlling the number of activity centers in an image that belong to the population being sampled. When the super-population is much larger than the total number of observed individuals, under-linkage is likely when each observation seeks its own activity center. However, to infer that a pair of observations in the overlapping region corresponds to different activity centers is to infer that each individual is detected once between two consecutive visits. Such inference, along with any extra activity center in the image that remain undetected, will be penalized by a high detection probability in the model for $\boldsymbol{\lambda}_{i,t}$. The process model thereby motivates linkage between observed locations that are spatially proximal.

2.3. PARAMETER MODEL

We used an informative inverse-gamma prior for σ_u^2 because we have specific knowledge about the extent of sea otter movement that is physically possible between consecutive images (Williams 1989). We imposed a penalization on the second derivatives of the fitted B-splines through the prior variances of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The penalty parameters were selected by cross-validation (Wahba 1978; Wood et al. 2016). We specified $\psi \sim \text{Beta}(0.001, 1)$ to approximate a scale prior for N ($[N] \propto 1/N$, Link, 2013), and we centered the prior for p_0 at 0.75 based on a prior data analysis and as suggested in past studies (Williams et al. 2017; Lu et al. 2019). A full description of prior distributions can be found in Appendix A.

The joint posterior distribution associated with our model is

$$\begin{aligned}\left[\{\boldsymbol{\lambda}_t\}_{t=1}^{T}, \{\boldsymbol{s}_m\}_{m=1}^{M}, \sigma_u^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, \{z_m\}_{m=1}^{M}, p_0, \psi \middle| \boldsymbol{Y}\right] \propto & \prod_{t=1}^{T}\prod_{i=1}^{n_t}\left[\boldsymbol{y}_{i,t} \middle| \boldsymbol{s}_{\lambda_{i,t}}, \sigma_u^2, \boldsymbol{\alpha}, \boldsymbol{\beta}\right] \times \prod_{m=1}^{M}[\boldsymbol{s}_m] \\ & \times \prod_{t=1}^{T}\left[\boldsymbol{\lambda}_t \middle| \{z_m\}_{m=1}^{M}, \{\boldsymbol{s}_m\}_{m=1}^{M}, \sigma_u^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, p_0\right] \times [p_0] \\ & \times \prod_{m=1}^{M}[z_m|\psi] \times [\psi] \times \left[\sigma_u^2\right] \times [\boldsymbol{\alpha}] \times [\boldsymbol{\beta}].\end{aligned}$$

The distortion parameters, c_t and θ_t, in Eq. 4 and the image footprint, $\mathcal{F}_t$, in Eq. 6 are deterministic functions of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and are therefore replaced by the basis function coefficients in the above expression. We implemented our model using MCMC and provide a full description of the algorithm in Supplementary Appendix A.

3. APPLICATION

3.1. SIMULATION

We simulated a population of $N = 200$ individuals and sampled their activity centers $\boldsymbol{s}_m$, for $m = 1, \ldots, N$, uniformly from a 100 m $\times$ 2000 m study domain, $\mathcal{D}$, to emulate the population intensity in the case study. For $t = 1, \ldots, T$, $T = 50$, we sampled realized locations $\boldsymbol{u}_{m,t} \sim \mathrm{N}\left(\boldsymbol{s}_m, \sigma_u^2 \boldsymbol{I}\right)$. We let σ_u^2 be 0.25 based on the estimated maximum underwater speed of sea otters according to empirical studies (Williams 1989). For the measurement process, we set the image centers to be equally spaced between $\boldsymbol{\mu}_1 = (50, 50)'$ and $\boldsymbol{\mu}_{50} = (1950, 50)'$ and let the footprint template at time t, $\mathcal{Q}_t$, be a 58 m $\times$ 58 m square centered at $\boldsymbol{\mu}_t$ and parallel to the horizontal axis. We generated distortion parameters from cubic B-splines with coefficients $\boldsymbol{\alpha} = \boldsymbol{\beta} = (0.1, 0.2, 0.1, -0.1, -0.2)'$ and obtained the image footprint $\mathcal{F}_t$ using Eq. 3. When $\boldsymbol{u}_{m,t} \in \mathcal{F}_t$, we detect individual m at time t with probability $p_0 = 0.75$ as informed by past studies (Williams et al. 2017; Lu et al. 2019). We recorded the distorted locations of the detected individuals by Eq. 1. Figure 1 illustrates the simulated image footprints and the true locations as well as the corresponding footprint templates and the observed locations.

In our implementation of the model, we let the super-population be of size $M = 3000$. We ran the MCMC algorithm in R version 3.0.2 (R Core Team 2019) for 15,000 iterations. Our algorithm took 2.5 hours on a 2.5 GHz Intel Core i5 processor. Advanced sampling strategies like the split-merge Metropolis-Hastings updates on the latent identities can be used to expedite computation (Jain and Neal 2004), and parallel computing techniques like recursive Bayesian methods can improve statistical scalability in future implementations (Hooten et al. 2021). We used a burn-in of 5000 iterations and obtained posterior realizations of population size as a derived quantity from the remaining $K = 10000$ posterior samples of z_m as

$$N^{(k)} = \sum_{m=1}^{M} z_m^{(k)}, \ k = 1, \ldots, K. \tag{7}$$

We obtained posterior realizations of the number of unique individuals from all observations by counting unique labels in the posterior samples of $\{\boldsymbol{\lambda}_t\}_{t=1}^T$ as

$$N_0^{(k)} = \left| \left\{ \boldsymbol{\lambda}_t^{(k)} \right\}_{t=1}^{T} \right|, \ k = 1, \ldots, K. \tag{8}$$

Our model captured the true parameters within their respective 95% credible intervals, which we summarize in Supplementary Table 1.

Two recent studies have used aerial photographs to estimate sea otter abundance in Glacier Bay, Alaska. Lu et al. (2019) proposed a nonlinear reaction-diffusion process model for population intensity, but used only every other image in accordance with the assumptions of their model. An arbitrary selection of images to use for data analysis may lead to bias in abundance estimation, especially if population intensity is spatially heterogeneous. Although our method is not directly comparable to that of Lu et al. (2019), we can compare the

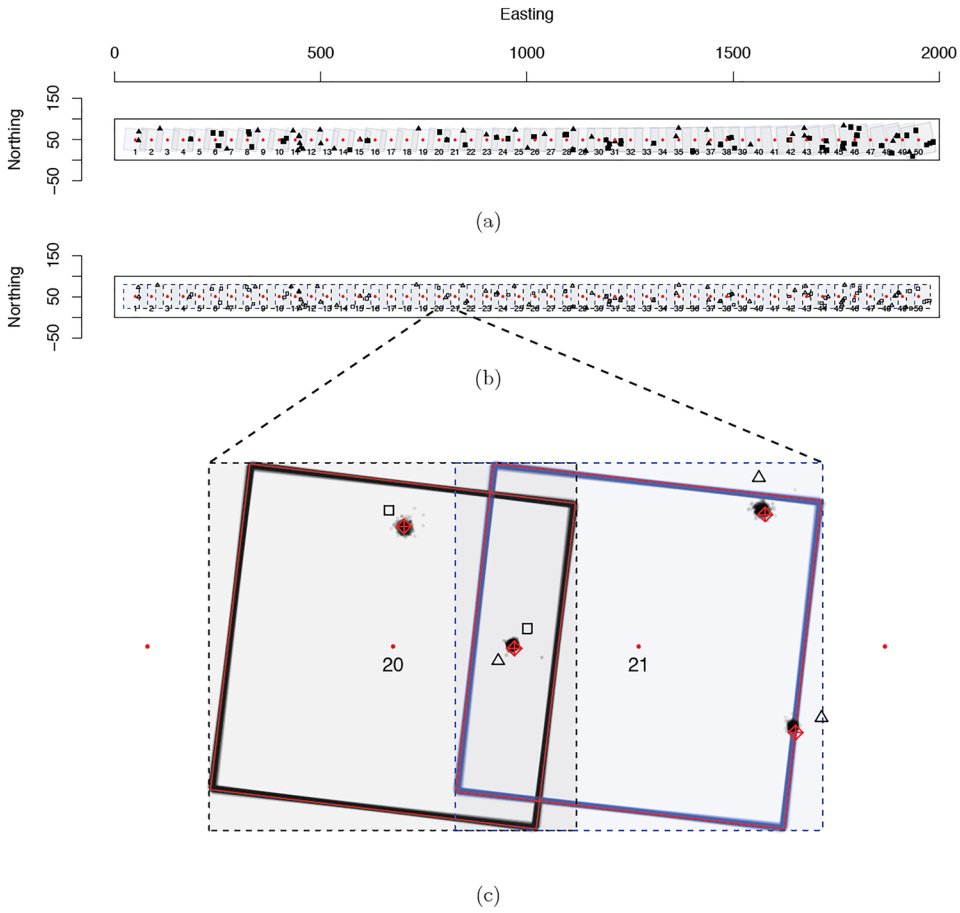
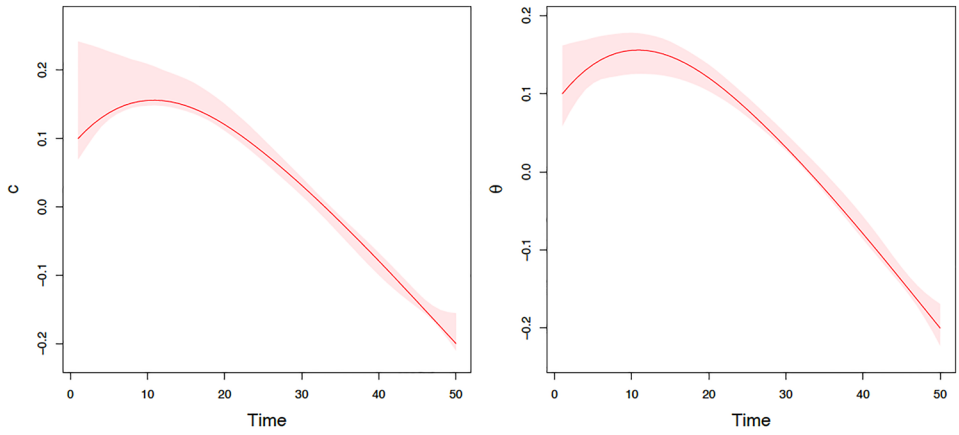


Figure 1. **a** Simulated image footprints, $\mathcal{F}_t$, overlaid with true locations, $\boldsymbol{u}_{i,t}$. The time-indexed points represent image centers, $\boldsymbol{\mu}_t$. True locations are marked with "■" in even images and "▲" in odd images. The largest rectangle containing all images is the study domain, $\mathcal{D}$; **b** simulated footprint templates (dashed rectangles), $\mathcal{Q}_t$, overlaid with observed locations, $\boldsymbol{y}_{i,t}$, that are marked with "□" in even images and "△" in odd images; **c** a focused illustration on observed images 20 and 21 (dashed rectangles), overlaid with posterior samples of image footprints (solid rectangles) and activity centers (points) with their truths (crossed diamonds) .

estimated number of unique individuals in all images because it refers to the observed abundance. We observed 111 locations from all images in the above simulation, which correspond to 90 unique individuals in truth. Our model estimated a posterior mean of 90 unique individuals. However, the number of unique individuals from counting all the odd images is 60, and the number of unique individuals from counting all the even images is 51. Discarding half of the images led to inconsistent and insufficient counts, whereas we improved abundance estimates by accounting for duplicate individuals in overlapping regions.

Williams et al. (2017) proposed an N-mixture model where the counts from overlapping images are considered temporal replicates. The model divides images into mutually exclusive regions of overlaps and non-overlaps and denotes $y(A_i, j)$ as the count from the jth overlap of region A_i, such that $\cup_{i=1}^{n} A_i = \cup_{t=1}^{T} \mathcal{F}_t$ and $A_i \cap A_j = \emptyset$, $i \neq j$. Under the



(a) (b)

Figure 2. **a** Estimated point-wise 95% credible interval overlaid with the truth for the scaling parameter, $\boldsymbol{c}$; **b** estimated point-wise 95% credible interval overlaid with the truth for the rotation parameter, $\boldsymbol{\theta}$.

assumption of homogeneous detection probability p_0 and population intensity η, the counts are modeled by

$$\begin{aligned} y\left(A_i,\, j\right) &\sim \text{Binom}\left(N\left(A_i\right),\, p_0\right) \\ N\left(A_i\right) &\sim \text{Pois}\left(\eta|A_i|\right), \end{aligned} \tag{9}$$

where $|A_i|$ is the area of A_i. Although Williams et al. (2017) accounted for heterogeneity in p_0 and η based on spatial covariates, we fit the homogeneous version of their model in Eq. 9 to our simulated data using an MCMC algorithm. A posterior realization of population size is obtained as a derived quantity by $N^{(k)} = \eta^{(k)}|\mathcal{D}|$, for $k = 1, \ldots, K$ MCMC iterations. The estimated posterior mean abundance was 178 with a 95% credible interval (142, 216). Both the method by Williams et al. (2017) and our method were able to recover the true population abundance in simulation; nonetheless, our process model distinguished sea otter detectability due to diving from that due to temporary emigration from the image footprints (Kendall et al. 1997). In addition to abundance estimation, our estimated activity centers can be used to inform spatial heterogeneity of the population intensity and our estimated distortion parameters can be used to reconstruct the aircraft trajectory via trigonometric projections.

To evaluate our model performance in linkage estimation, we used false discovery rate (FDR) and false negative rate (FNR) as recommended by Steorts (2015) to account for the large number of non-links in our application. There are four possible results when comparing the estimated linkage and the truth:

1. True positive (TP): two individuals have the same latent identity in both the estimation and the truth;
2. False positive (FP): two individuals are estimated to have the same latent identity when they are actually different;

Table 1. True parameter values and marginal posterior means (95% credible intervals) for population size (N) and the number of unique individuals (N_0) under different simulated movement uncertainties (σ_u^2)

σ_u^2	N	Posterior mean (95% CI)	N_0	Posterior mean (95% CI)
0.25	200	192 (160, 233)	91	90 (90, 91)
1	200	191 (155, 241)	87	88 (87, 91)
2.5	200	251 (196, 327)	91	95 (93, 97)
10	200	364 (216, 668)	88	100 (93, 108)

3. True negative (TN): two individuals have different latent identities in both the estimation and the truth;

4. False negative (FN): two individuals are estimated to have different latent identities when they are actually the same.

We computed the posterior mean FDR and FNR as $\mathbb{E}[\text{FDR}|\boldsymbol{Y}] = \frac{1}{K}\sum_{k=1}^{K}\frac{\text{FP}^{(k)}}{\text{FP}^{(k)}+\text{TP}^{(k)}} = 0.000001$ and $\mathbb{E}[\text{FNR}|\boldsymbol{Y}] = \frac{1}{K}\sum_{k=1}^{K}\frac{\text{FN}^{(k)}}{\text{FN}^{(k)}+\text{TP}^{(k)}} = 0.001$ based on the model fit to the simulated data.

Posterior realizations of the scaling and rotation parameters, $\boldsymbol{c}$ and $\boldsymbol{\theta}$, were obtained as derived quantities using Eq. 2, and posterior realizations of the image footprints, $\mathcal{F}_t$, were obtained as derived quantities using Eq. 3. Figure 1 illustrates posterior samples of the image footprints and the activity centers overlaid with their truth for a subset of images (images 20 and 21). Our model performed well, linking observations that correspond to the same individual in the overlapping region and correctly estimating their activity centers despite distortion. Figure 2 demonstrates the point-wise 95% credible intervals for $\boldsymbol{c}$ and $\boldsymbol{\theta}$. The point-wise 95% credible intervals contained the true simulated values for both parameters.

Lastly, we demonstrate our model inference using simulated data with increasing levels of individual movement, σ_u^2, while the other parameters are held constant. We note that any σ_u^2 significantly larger than 0.25 would be unrealistic for sea otters in southeastern Alaska; nonetheless, the following demonstration serves to emphasize a viable condition for applying our method in other survey scenarios. Table 1 summarizes the posterior distributions of the number of unique individuals and population size along with their truth under different values of σ_u^2. The true numbers of unique individuals vary between simulations due to dependence on σ_u^2. The estimated 95% credible interval for N_0 widens as individual movement increases, indicating less certainty in the linkage process. Consequently, the estimated 95% credible interval for N expands along with σ_u^2, and the credible intervals when $\sigma_u^2 = 10$ did not contain the respective truths for N_0 or N.

3.2. CASE STUDY

Sea otter populations have undergone significant fluctuations throughout their range over the past two centuries (Jameson et al. 1982). After being hunted to near extinction during the maritime fur trade, sea otter populations have recovered in many areas due to a combination of conservation efforts, translocations, and environmental changes (Larson

Table 2. Marginal posterior means and 95% credible intervals for the case study

Parameter	Posterior mean (95% CI)
p_0	0.64 (0.52, 0.75)
ψ	0.19 (0.15, 0.24)
N	566 (453, 704)
σ_u^2	0.45 (0.34, 0.59)

et al. 2014; Eisaguirre et al. 2021). Monitoring sea otter colonization in Glacier Bay provides important insight into the ability of a keystone species to recover from near extirpation and to understand their role in structuring the nearshore food web in Glacier Bay (Williams et al. 2019). From 1993 to 2012, observer-based aerial surveys were conducted from small single-engine aircraft along systematic transects (Esslinger et al. 2015). Beginning in 2017, aerial photographic methods (Womble et al. 2018) were conducted using model-based optimized surveys (Williams et al. 2019). Aerial photographic images were post-processed by trained observers that counted the number of sea otters in each image. For our case study, we analyzed a sequence of 20 consecutive images with a total of 151 observations (see Fig. 3 in Supplementary Material for an example of real images). Sea otter locations were recorded for 60 m $\times$ 90 m footprint templates with their long sides perpendicular to the direction of aircraft movement (vectors connecting consecutive image centers). Figure 3 illustrates the observed locations overlaid with footprint templates. Our study domain was a 300 m $\times$ 1000 m rectangular region containing all image footprints as shown in Fig. 3.

We let the super-population be of size $M = 3000$ and ran the MCMC algorithm for 15000 iterations with a burn-in of 5000 iterations. Table 2 summarizes the marginal posterior distributions from the case study. Our estimated detection probability agrees with the estimates from previous studies (Williams et al. 2017; Lu et al. 2019). Posterior realizations of population size were obtained as derived quantities by Eq. 7.

Our model inferred a posterior mean of 125 unique individuals among the 151 observations using Eq. 8. We illustrate posterior samples of the image footprints, $\mathcal{F}_t$, obtained as derived quantities using Eq. 3 and posterior samples of the true locations, $\boldsymbol{u}_{i,t}$, obtained as derived quantities using Eq. 1 for a subset of images ($t = 8, 9$) in Fig. 3. Using our model, we estimated counterclockwise rotation as the distortion process for both images, thereby linking the two pairs of observations in the overlapping region. The posterior mean linkage probability was 0.98 for observation pairs (a, h) and (b, i), and all other observation pairs have less than 0.02 posterior mean linkage probabilities. Figure 4 demonstrates the point-wise 95% credible intervals overlaid with the posterior means for scaling and rotation, respectively.

4. DISCUSSION

We presented a novel method to perform entity resolution and population size estimation using individual locations obtained from aerial imagery data of sea otters. We coupled

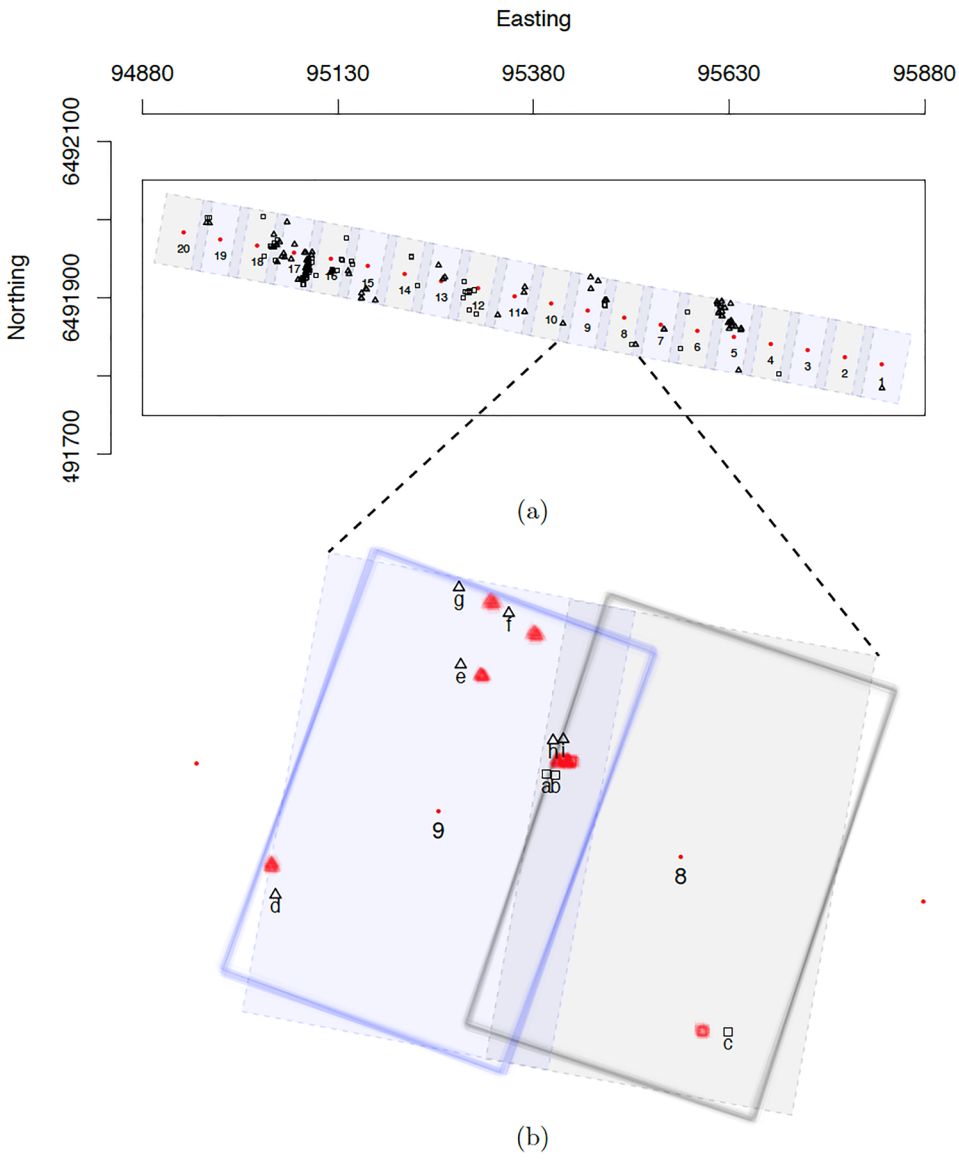


Figure 3. **a** Observations from the case study. Footprint templates (dashed rectangles), $\mathcal{Q}_t$, are overlaid with observed locations, $\boldsymbol{y}_{i,t}$, that are marked with "□" in even images and "△" in odd images. The time-indexed points represent image centers, $\boldsymbol{\mu}_t$. The largest rectangle containing all templates is the study domain, $\mathcal{D}$; **b** a focused illustration on observed images 8 and 9 (dashed rectangles), overlaid with posterior samples of image footprints (solid rectangles), $\mathcal{F}_t$. Observed locations are indexed by letters, and posterior samples of true locations, $\boldsymbol{u}_{i,t}$, are shown in solid squares in image 8 and solid triangles in image 9 .

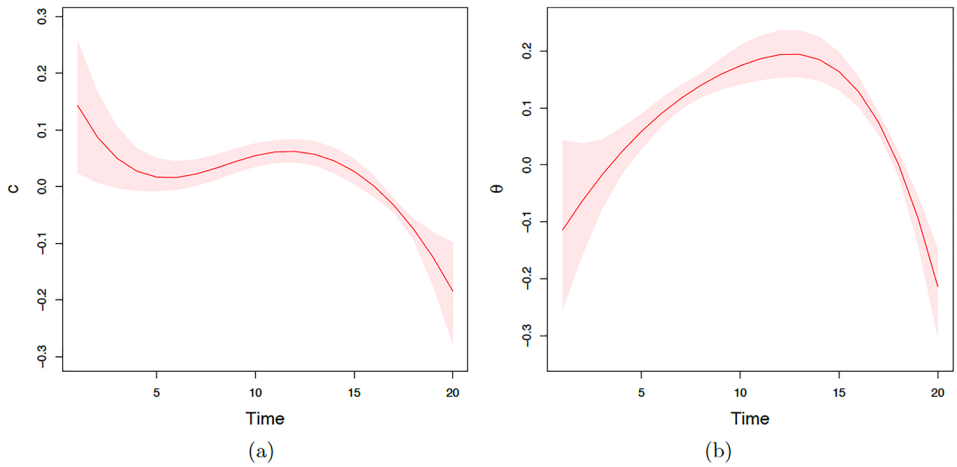


Figure 4. **a** Estimated posterior mean and point-wise 95% credible interval for the scaling parameter, $\boldsymbol{c}$; **b** estimated posterior mean and point-wise 95% credible interval for the rotation parameter, $\boldsymbol{\theta}$.

record linkage and capture-recapture models to accommodate important features of aerial imagery data. Our unified framework allows information exchange and uncertainty propagation between the estimation of linkage structure and population abundance, and our model is adequate for both inferential tasks.

Record linkage models are often sensitive to parameters that control linkage probability. In a sensitivity analysis for the graphical record linkage model, Steorts (2015) showed that linkage inference is only reliable when a very precise prior is used on the parameter for distortion probability. In the Bayesian alignment model, Green and Mardia (2005) advised that informative priors be used for parameters that dictate matching tendency. In our model, linkage of observed locations is motivated by their proximity in Euclidean distance to latent activity centers. Therefore, as expected, our model is sensitive to σ_u^2, the parameter controlling movement. Reliable inference requires that animal movement between consecutive detections be small relative to distortion, otherwise the model would struggle to identify unique individuals using locations only. Fortunately, much is known about movement characteristics of many species and this information can be used to specify an informative prior for σ_u^2. During aerial surveys in Glacier Bay, Alaska, the time lapse between consecutive images is so brief (1 second) that sea otter movement is significantly limited by their physical capability, thus we specified the prior for σ_u^2 such that movement distance between consecutive images was less than a meter (Williams 1989). We provide details for a sensitivity analysis of prior distributions on σ_u^2 in Supplementary Appendix B. In addition to limited individual movement, our method could potentially benefit from more overlapping regions and higher population intensities because they provide more instances for linkage. On the other hand, our method may be hindered by extensive distortions in the image footprints or highly clustered populations where distances between the true activity centers are closer than σ_u.

Although our method is designed to link observed individuals and estimate population size simultaneously, it can be useful even when the objective is only one of the two. The

output of a record linkage model provides insight about the number of unique individuals observed at least once, and abundance estimation requires only the additional subset of population that is not observed. Population models can be used to provide prior information about the total number of latent individuals in a graphical record linkage model (Tancredi et al. 2018), a parameter that has also proven to be influential for inference (Steorts 2015). Detection mechanisms can guide learning about the number of times an individual's record is observed: High detection probabilities indicate frequent observations of an individual, thereby promoting linkage; low detection probabilities indicate few observations of the individual, thereby proposing new latent identities. Our model assumptions can be generalized to account for more complicated monitoring situations. For example, hypergeometric models may be used in place of binomial models in capture-recapture studies when individual detections are correlated due to sampling without replacement from a finite population (Darroch 1958; Link et al. 2009; Tancredi and Liseo 2011). We may also model heterogeneity in p_0 to account for factors such as animal diving in response to aircraft disturbance and survey conditions that affect the backdrop (e.g., kelp, sun angle, sea state).

The use of observed locations in our model helped us better understand the spatial heterogeneity in population intensity. Under a uniform prior on $\boldsymbol{s}_m$, the variation in population intensity is implicitly reflected through the estimated activity centers. A natural extension to our method is to model the spatial distribution of activity centers explicitly (Efford 2004, 2011; Brost et al. 2017, 2020). We could account for heterogeneity in the distribution of activity centers using a species distribution model (SDM; e.g., Hefley and Hooten, 2016). An SDM is often specified as a spatial point process model, which, in our case, could take the form

$$[\boldsymbol{s}_m|\boldsymbol{x}(\boldsymbol{s}_m)] = \frac{\exp\left(\boldsymbol{x}(\boldsymbol{s}_m)'\boldsymbol{\beta}\right)}{\int_{\mathcal{D}} \exp\left(\boldsymbol{x}(\boldsymbol{s})'\boldsymbol{\beta}\right) d\boldsymbol{s}},$$

for $m = 1, \ldots, M$, where $\boldsymbol{x}(\boldsymbol{s}_m)$ denotes the vector of spatial covariates at $\boldsymbol{s}_m$ and $\boldsymbol{\beta}$ denotes the associated coefficients. Alternatively, we could attribute heterogeneity in the distribution of activity centers to the interaction among individuals which could be modeled mechanistically (e.g., Scharf et al., 2016).

Although our model was designed for aerial imagery data from sea otter population surveys in Glacier Bay, Alaska, our framework can be adapted for a variety of applications that involve intersecting fields of observation (Borchers et al. 2020). Our method is also useful for aligning unlabeled point patterns with consistent measurement error, such as reconstruction of a three-dimensional object from two-dimensional views (Ourselin et al. 2001; Rezende et al. 2016) and reconstruction of a movement trajectory using multiple snapshots (Ando 1991; Du et al. 2016).

ACKNOWLEDGEMENTS

This research was funded by NSF DMS 1614392 and NPS P16AC01524 and P19AC00063. Research and monitoring were conducted under US Fish & Wildlife Service Scientific Research Permit #MA14762C-0 and NPS Scientific Research Permit GLBA-2016-SCI-022. Any use of trade, firm, or product names is for descriptive

purposes only and does not imply endorsement by the US Government. We appreciate the assistance of Dennis Lozier and Louise Taylor with sea otter surveys and image processing.

[Received December 2020. Revised October 2021. Accepted December 2021. Published Online January 2022.]

APPENDIX A: PRIOR DISTRIBUTIONS

$$
\begin{aligned}
p_0 &\sim \text{Beta}\,(3, 1)\,, \\
\psi &\sim \text{Beta}\,(0.001, 1)\,, \\
\sigma_u^2 &\sim \text{IG}\,(100, 25)\,, \\
\boldsymbol{s}_m &\sim \text{Unif}(\mathcal{D}),\ m = 1, \ldots, M, \\
\boldsymbol{\alpha} &\sim \text{N}\,(\mathbf{0}, 0.001\boldsymbol{R} + 0.01\boldsymbol{I})\,, \\
\boldsymbol{\beta} &\sim \text{N}\,(\mathbf{0}, 0.001\boldsymbol{R} + 0.01\boldsymbol{I})\,,
\end{aligned}
$$

where $\boldsymbol{R} = \left(\boldsymbol{D}_2^{-}\right)' \boldsymbol{D}_2$ and $\boldsymbol{D}_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix}$.

REFERENCES

Anderson M, Fienberg SE (1999) Who counts? The politics of census-taking in contemporary America. Russell Sage Foundation

Ando H (1991) Dynamic reconstruction of 3d structure and 3d motion. In: Proceedings of the IEEE workshop on visual motion, pp 101–102

Barker RJ, Schofield MR, Link WA, Sauer JR (2018) On the reliability of N-mixture models for count data. Biometrics 74:369–377

Betancourt B, Zanella G, Miller JW, Wallach H, Zaidi A, Steorts RC (2016) Flexible models for microclustering with application to entity resolution. In Advances in neural information processing systems, pp 1417–1425

Borchers DL, Nightingale P, Stevenson BC, Fewster RM (2020) A latent capture history model for digital aerial surveys. Biometrics

Brost BM, Hooten MB, Small RJ (2017) Leveraging constraints and biotelemetry data to pinpoint repetitively used spatial features. Ecology 98(1):12–20

Brost BM, Hooten MB, Small RJ (2020) Model-based clustering reveals patterns in central place use of a marine top predator. Ecosphere 11:e03123

Brown M, Lowe DG (2015) Automatic panoramic image stitching using invariant features. Int J Comput Vision 74:59–73

Buckland ST, Burt ML, Rexstad EA, Mellor M, Williams AE, Woodward R (2012) Aerial surveys of seabirds: the advent of digital methods. J Appl Ecol 49:960–967

Caughley G (1974) Bias in aerial survey. J Wildl Manag 38(4):921–933

Christen P (2011) A survey of indexing techniques for scalable record linkage and deduplication. IEEE Trans Knowl Data Eng 24:1537–1555

Copas J, Hilton F (1990) Record linkage: statistical models for matching computer records. J R Stat Soc A Stat Soc 153:287–312

Darroch JN (1958) The multiple capture census i. Estimation of a closed population. Biometrika 45:343–358

Dennis EB, Morgan BJT, Ridout MS (2015) Computational aspects of N-mixture models. Biometrics 71:237–246

Dryden IL, Mardia KV (1998) Statistical analysis of shape. Wiley

Du Y, Wong Y, Liu Y, Han F, Gui Y, Wang Z, Kankanhalli M, Geng W (2016) Marker-less 3d human motion capture with monocular image sequence and height-maps. In: European conference on computer vision. Springer, pp 20–36

Efford M (2004) Density estimation in live-trapping studies. Oikos 106:598–610

Efford MG (2011) Estimation of population density by spatially explicit capture-recapture analysis of data from area searches. Ecology 92:2202–2207

Eisaguirre JM, Williams PJ, Lu X, Kissling ML, Beatty WS, Esslinger GG, Womble JN, Hooten MB (2021) Diffusion modeling reveals effects of multiple release sites and human activity on a recolonizing apex predator. Mov Ecol 9:34

Esslinger GG, Esler D, Howlin S, Starcevich L (2015) Monitoring population status of sea otters (Enhydra lutris) in Glacier Bay National Park and Preserve, Alaska: options and considerations. US Department of the Interior, US Geological Survey

Fellegi IP, Sunter AB (1969) A theory for record linkage. J Am Stat Assoc 64:1183–1210

Fortini M, Liseo B, Nuccitelli A, Scanu M (2001) On Bayesian record linkage. Res Official Stat 4:185–198

Green PJ, Mardia K (2005) Bayesian alignment using hierarchical models, with application in protein bioinformatics. Biometrika 93(2):235–254

Gross JW, Heumann BW (2016) A statistical examination of image stitching software packages for use with unmanned aerial systems. Photogr Eng Remote Sens 82(6):419–425

Hefley TJ, Broms KM, Brost BM, Buderman FE, Kay SL, Scharf HR, Tipton JR, Williams PJ, Hooten MB (2017) The basis function approach to modeling autocorrelation in ecological data. Ecology 98:632–646

Hefley TJ, Hooten MB (2016) Hierarchical species distribution models. Curr Lands Ecol Rep 1:87–97

Hooten MB, Johnson DS, Brost BM (2021) Making recursive Bayesian inference accessible. Am Stat 75:185–194

Jain S, Neal RM (2004) A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. J Comput Graph Stat 13:158–182

Jameson RJ, Kenyon KW, Johnson AM, Wight HM (1982) History and status of translocated sea otter populations in North America. Wildl Soc Bull 10(2):100–107

Jaro MA (1989) Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. J Am Stat Assoc 84:414–420

Kendall WL, Nichols JD, Hines JE (1997) Estimating temporary emigration using capture-recapture data with Pollock's robust design. Ecology 78:563–578

Ketz AC, Johnson TL, Hooten MB, Hobbs NT (2019) A hierarchical Bayesian approach for handling missing classification data. Ecol Evol 9(6):3130–3140

LaPorte RE, McCarty D, Bruno G, Tajima N, Baba S (1993) Counting diabetes in the next millennium: application of capture-recapture technology. Diabetes Care 16:528–534

Larsen MD (2004) Record linkage using finite mixture models. An essential journey with donald Rubin's statistical family, applied Bayesian modeling and causal inference from incomplete-data perspectives, pp 309–318

Larsen MD, Rubin DB (2001) Iterative automated record linkage using mixture models. J Am Stat Assoc 96:32–41

Larson SE, Bodkin JL, VanBlaricom GR (2014) Sea Otter conservation. Academic Press

Levin A, Zomet A, Peleg S, Weiss Y (2004) Seamless image stitching in the gradient domain. In: European conference on computer vision. Springer, pp 377–389

Link WA (2013) A cautionary note on the discrete uniform prior for the binomal N. Ecology 94(10):2173–2179

Link WA, Yoshizaki J, Bailey LL, Pollock KH (2009) Uncovering a latent multinomial: analysis of mark-recapture data with misidentification. Biometrics 66:178–185

Liseo B, Tancredi A (2011) Bayesian estimation of population size via linkage of multivariate normal data sets. J Official Stat 27:491

Lu X, Williams PJ, Hooten MB, Powell JA, Womble JN, Bower MR (2019) Nonlinear reaction-diffusion process models improve inference for population dynamics. Environmetrics 31(3):e2604

Lum K, Price ME, Banks D (2013) Applications of multiple systems estimation in human rights research. Am Stat 67:191–200

McGlincy MH (2004) A Bayesian record linkage methodology for multiple imputation of missing links. In: ASA proceedings of the joint statistical meetings. American Statistical Association, Alexandria, VA, pp 4001–4008

Ourselin S, Roche A, Subsol G, Pennec X, Ayache N (2001) Reconstructing a 3D structure from serial histological sections. Image Vis Comput 19:25–31

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rezende DJ, Eslami SA, Mohamed S, Battaglia P, Jaderberg M, Heess N (2016) Unsupervised learning of 3d structure from images. In: Advances in neural information processing systems, pp 4996–5004

Royle JA (2009) Analysis of capture-recapture models with individual covariates using data augmentation. Biometrics 65:267–274

Royle JA, Dorazio R (2012) Parameter-expanded data augmentation for Bayesian analysis of capture-recapture models. J Ornithol 152:521–537

Royle JA, Dorazio RM, Link WA (2007) Analysis of multinomial models with unknown index using data augmentation. J Comput Graph Stat 16:67–85

Royle JA, Young KV (2008) A hierarchical model for spatial capture-recapture data. Ecology 89(8):2281–2289

Sadinle M (2018) Bayesian propagation of record linkage uncertainty into population size estimation of human rights violations. Ann Appl Stat 12(2):1013–1038

Scharf HR, Hooten MB, Fosdick BK, Johnson DS, London JM, Durban JW et al (2016) Dynamic social networks based on movement. Ann Appl Stat 10:2182–2202

Steorts RC (2015) Entity resolution with empirically motivated priors. Bayesian Anal 10(4):849–875

Steorts RC, Hall R, Fienberg SE (2015) A Bayesian approach to graphical record linkage and deduplication. J Am Stat Assoc 111:1660–1672

Steorts RC, Ventura SL, Sadinle M, Fienberg SE (2014) A comparison of blocking methods for record linkage. International conference on privacy in statistical databases. Springer, Cham, pp 253–268

Szeliski R (2006) Image alignment and stitching: a tutorial. Found Trends® Comput Graph Vis 2:1–104

Tancredi A, Liseo B (2011) A hierarchical Bayesian approach to record linkage and population size problems. Ann Appl Stat 5(2B):1553–1585

Tancredi A, Steorts R, Liseo B et al (2018) A unified framework for de-duplication and population size estimation. Bayesian Anal 15(2):633–682

Ver Hoef JM (2014) Aerial survey data. Statistics Reference Online, Wiley StatsRef

Wahba G (1978) Improper priors, spline smoothing and the problem of guarding against model errors in regression. J Roy Stat Soc Ser B 40(3):364–372

Wallach H, Jensen S, Dicker L, Heller K (2010) An alternative prior process for nonparametric Bayesian clustering. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp 892–899

Williams PJ, Hooten MB, Esslinger GG, Womble JN, Bodkin JL, Bower MR (2019) The rise of an apex predator following deglaciation. Divers Distrib 25:895–908

Williams PJ, Hooten MB, Womble JN, Bower MR (2017) Estimating occupancy and abundance using aerial images with imperfect detection. Methods Ecol Evol 8:1679–1689

Williams PJ, Schroeder C, Jackson P (2020) Estimating reproduction and survival of unmarked juveniles using aerial images and marked adults. J Agric Biol Environ Stat 25:133–147

Williams TM (1989) Swimming by sea otters: adaptations for low energetic cost locomotion. J Comp Physiol A 164(6):815–824

Winkler WE (1995) Matching and record linkage. Bus Surv Methods 1:355–384

Winkler WE (2006) Overview of record linkage and current research directions. In: Bureau of the Census

Womble J, Williams P, Johnson W, Taylor-Thomas L, Bower M (2018) Sea otter monitoring protocol for Glacier Bay National Park, Alaska: Version SO-2017.1. Natural Resource Report NPS/SEAN/NRR—2018/1762, National Park Service, Fort Collins, Colorado

Wood SN, Pya N, Safken B (2016) Smoothing parameter and model selection for general smooth models. J Am Stat Assoc 111(516):1548–1563

Wright JA, Baker RJ, Schofield MR, Frantz AC, Byrom AE, Gleeson DM (2009) Incorporating genotype uncertainty into mark-recapture-type models for estimating abundance using DNA samples. Biometrics 65:833–840