

Integrated modeling of bilateral photo-identification data in mark–recapture analyses

BRETT T. MCCLINTOCK,^{1,4} PAUL B. CONN,¹ ROBERT S. ALONSO,^{2,3} AND KEVIN R. CROOKS²

¹National Marine Mammal Laboratory, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, 7600 Sand Point Way NE, Seattle, Washington 98115 USA

²Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado 80523 USA

³U.S. Geological Survey, Western Ecological Research Center, Thousand Oaks, California 91360 USA

Abstract. When natural marks provide sufficient resolution to identify individual animals, noninvasive sampling using cameras has a number of distinct advantages relative to “traditional” mark–recapture methods. However, analyses from photo-identification records often pose additional challenges. For example, it is often unclear how to link left- and right-side photos to the same individual, and previous studies have primarily used data from just one side for statistical inference. Here we describe how a recently developed statistical method can be adapted for integrated mark–recapture analyses using bilateral photo-identification records. The approach works by assuming that the true encounter history for each animal is a latent (unobserved) realization from a multinomial distribution. Based on the type of photo encounter (e.g., right, left, or both sides), the recorded (observed) encounter histories can only arise from certain combinations of these latent histories. In this manner, the approach properly accounts for uncertainty about the true number of distinct animals observed in the study. Using a Markov chain Monte Carlo sampling procedure, we conduct a small simulation study to show that this approach has reasonable properties and outperforms other methods. We further illustrate our approach by estimating population size from bobcat photo-identification records. Although motivated by bilateral photo-identification records, we note that the proposed methodology can be used to combine and jointly analyze other types of mark–recapture data (e.g., photo and DNA records).

Key words: abundance; latent multinomial; mark–resight; noninvasive capture–recapture; photo-identification; remote camera trapping.

INTRODUCTION

Ecologists are increasingly using mark–recapture methods to analyze photo-identification records, with numerous applications to sharks (e.g., Holmberg et al. 2008), reptiles (e.g., Nair et al. 2012), marine mammals (e.g., Wilson et al. 1999, Calambokidis and Barlow 2004, Mackey et al. 2008), and felids (e.g., Karanth and Nichols 1998, Soisalo and Cavalcanti 2006). When individuals are identifiable by natural or artificial marks, such experiments have the potential to provide detailed estimates of key demographic parameters such as abundance, survival, and recruitment, and are therefore extremely useful for informing management and testing ecological or evolutionary theory.

Using photo-identification instead of “traditional” mark–recapture (i.e., physically capturing and marking animals) has at least two distinct advantages. First, less effort is required to physically capture, sedate, and mark animals. Second, because sampling is noninvasive, the potential for harming animals is eliminated. The latter

point is important because invasive marking is often precluded for certain protected species, or is morally or politically unpalatable. Physical capture and tagging has also been shown to decrease survival and/or return rates for a number of taxa (e.g., Samuel et al. 1990), suggesting that parameter estimates may be biased under certain types of physical marking schemes.

Unfortunately, use of photo-identification records in mark–recapture analyses is not entirely without its problems. For instance, when natural markings are bilaterally asymmetrical, matching photographs to individuals can be difficult when investigators are routinely able to photograph only one side of an animal. The typical approach in the literature when confronted with this situation is to conduct separate analyses with left-sided and right-sided photographs and compare the results (e.g., Wilson et al. 1999, Berrow et al. 2012, Nair et al. 2012). However, this is not as efficient as conducting an integrated analysis with both types of encounters, given that sample sizes may be considerably reduced, resulting in lower precision (e.g., Madon et al. 2011). Corkrey et al. (2008) proposed a novel approach for modeling left- and right-sided encounters of dolphins in the context of state-space models, and this approach has recently been extended by Hiby et al. (2013).

Manuscript received 18 September 2012; revised 22 February 2013; accepted 25 February 2013. Corresponding Editor: E. G. Cooch.

⁴ E-mail: brett.mcclintock@noaa.gov

However, Corkrey et al. (2008) make a number of assumptions that will often be violated, including (1) the probability of being encountered on both sides during the same sampling occasion is the product of two independent events (encountered on left side, encountered on right side) and (2) records of individuals only seen on the left side are independent of those only seen on the right side. While the former assumption may be reasonable under certain sampling scenarios, the latter seems unlikely to hold in practical applications.

Here, we present an approach for simultaneously modeling bilateral photo-identification records in the context of mark-recapture studies. The approach relies heavily on recent statistical methodology introduced by Link et al. (2010), in which the true encounter history of an individual is treated as a realization from a latent (unobserved) multinomial process. The observed data (right-sided, left-sided, or both-sided detections) are then dependent on the true underlying (but unobserved) encounter histories. In this manner, left-sided detection histories are allowed to be matched with right-sided detections histories from the same individual while properly accounting for uncertainty about the true number of distinct animals observed in the study.

In the sections that follow, we first review recently proposed statistical methods and illustrate how modeling bilateral photographic records fits into this framework. Second, we adapt the approach of Link et al. (2010) to the analysis of bilateral photographic encounters, including methods needed to set up an efficient Markov chain Monte Carlo (MCMC) sampler for posterior simulation. We then conduct a small simulation study to examine possible biases and anticipated efficiency gains when our integrated approach is compared with the usual approach of analyzing encounters from one side only and with the approach of Corkrey et al. (2008). Finally, we apply our integrated framework to estimate population size from bilateral photo-identification records of bobcats in southern California.

METHODS

Detailed problem description

Consider a “typical” mark-recapture experiment where sampling is conducted over T sampling occasions and the identity of each animal is known with certainty when it is observed. When encounters are simple binary responses and $T = 2$, there are three possible encounter histories: 01 (encountered on the second occasion but not the first), 10 (encountered on the first occasion but not the second), and 11 (encountered on both occasions). It is common to assume that encounter history \mathbf{H}_i ($i = 1, \dots, 2^T - 1$) is a realization of a multinomial process with probability $\pi_i = \Pr(\mathbf{H}_i | \boldsymbol{\theta})$ that is a function of unknown parameters $\boldsymbol{\theta}$ (for example, $\boldsymbol{\theta}$ might consist of survival and detection probabilities). In this case, the number of uniquely identifiable animals (n) is known with certainty and is treated as a constant in statistical likelihood formulations (e.g., Williams et al. 2002).

When conditioning on first capture, the model likelihood is

$$[\mathbf{x} | \boldsymbol{\theta}] \propto \prod_{i=1}^{2^T-1} \pi_i^{x_i} \quad (1)$$

where x_i is the frequency for encounter history \mathbf{H}_i , such that $\sum_i x_i = n$. Conducting maximum likelihood estimation based on Eq. 1 is a common approach to making inferences about demographic parameters from mark-recapture studies (note that additional modifications are needed to incorporate individual heterogeneity or make inferences about individuals that are never encountered; e.g., for estimating population size).

In contrast to the preceding scenario, now consider the situation in which individuals are encountered via photographs and marks are bilaterally asymmetrical. We will consider two photo-sampling scenarios, where recorded (observed) encounter histories may include detections of (1) only the left or right side (LR data type) or (2) the left, right, or both sides (LRB data type). We denote a non-encounter with a 0, an encounter on the left side only as L, a right-sided encounter as R, and an encounter on both sides as B. Depending on the design of camera surveys, B encounters may be recorded from a single image (where both sides of an individual are visible), a pair of synchronized images, or multiple images (assuming both sides of an individual are known; see *Example: Bobcats in southern California*) within a sampling occasion. We assume all B encounters are observed for the LRB data type (but see *Discussion* for a relaxation of this assumption).

One-sided analyses are common for bilateral encounter data when researchers are unable to match opposite sides, but are able to accurately match left sides to left sides and right sides to right sides. Under these conditions, attempts to combine left- and right-sided encounters into a single analysis are problematic because detected individuals can yield a number of possible recorded histories, depending on whether B encounters are observable. For example, when presented with the recorded histories L0 and 0R, we do not know whether these observations arose from the same animal seen on both occasions (latent history LR), or whether it was indeed two different animals each seen on one occasion (latent histories L0 and 0R). Similarly, when B detections are not observed, the recorded histories L0 and R0 could arise from the latent history B0 or from the latent histories L0 and R0 (Table 1). The fundamental issue here is that there is no longer a one-to-one mapping of the number of unique animals encountered with the number of observed encounter histories. As such, it is no longer technically correct to apply Eq. 1 for statistical inference.

Fortunately, Link et al. (2010) proposed an approach for analyzing mark-recapture data under such a scenario. However, their focus was on individual misidentification from photographs (due to observer

TABLE 1. Latent histories (ω) and recorded histories ($\tilde{\omega}$) from left- and right-side encounters only (LR) or from left-, right-, and both-side encounters (LRB) with $T = 2$.

i	Latent history (ω_i)	$\Pr(\omega_i \theta)$	LR data type		j	Recorded history ($\tilde{\omega}_j$)	LRB data type		k	Recorded history ($\tilde{\omega}_k$)
			Contributed records (j from i)				Contributed records (k from i)			
1	00	$q_1 q_2$		1	0L		1	0L
2	0L	$q_1 p_2 \delta^L$	1.....		2	0R	1.....		2	0R
3	0R	$q_1 p_2 \delta^R$	2.....		3	L0	2.....		3	0B
4	0B	$q_1 p_2 \delta^B$	12....		4	LL	3.....		4	L0
5	L0	$p_1 \delta^L q_2$...3...		5	R0	4.....		5	LL
6	LL	$p_1 \delta^L p_2 \delta^L$...4..		6	RR	5.....		6	LB
7	LR	$p_1 \delta^L p_2 \delta^R$...23...				2.4.....		7	R0
8	LB	$p_1 \delta^L p_2 \delta^B$	2.4..				6.....		8	RR
9	R0	$p_1 \delta^R q_2$...5.				7.....		9	RB
10	RL	$p_1 \delta^R p_2 \delta^L$	1...5.				1...7.....		10	B0
11	RR	$p_1 \delta^R p_2 \delta^R$...6				8.....		11	BL
12	RB	$p_1 \delta^R p_2 \delta^B$	1...6				9.....		12	BR
13	B0	$p_1 \delta^B q_2$...3.5.				10.....		13	BB
14	BL	$p_1 \delta^B p_2 \delta^L$...45.				11.....			
15	BR	$p_1 \delta^B p_2 \delta^R$...3..6				12.....			
16	BB	$p_1 \delta^B p_2 \delta^B$...4.6				13.....			

Notes: Contributed records columns show the recorded histories arising from specific latent histories for the LR and LRB data types. For example, latent history 16, $\omega_{16} = \text{BB}$, gives rise to recorded histories $j = 4$ and $j = 6$, $\tilde{\omega}_4 = \text{LL}$ and $\tilde{\omega}_6 = \text{RR}$, for the LR data type. For the LRB data type, $\omega_{16} = \text{BB}$ spawns only recorded history $k = 13$, $\tilde{\omega}_{13} = \text{BB}$. A zero represents a non-encounter at time t . For illustration, we include the probability of each latent history, $\Pr(\omega | \theta)$, for the closed population abundance model M_t , allowing time variation in detection probability p_t ($t = 1, \dots, T$), where $q_t = 1 - p_t$. For the closed population abundance model M_t with $T = 2$, $\theta = (p_1, p_2, \delta^L, \delta^R, N)$, where $\delta^B = 1 - \delta^L - \delta^R$ and N is population abundance. The indices j and k refer to specific recorded histories for the LR and LRB data types, respectively. Other variables are defined in Table 2. This presentation mirrors that of Link et al. (2010).

error) or genetic material (due to genotyping error). Our task for the next section is to adapt their approach to the problem of bilateral photographs.

Statistical methods

Generally speaking, we shall attempt to analyze bilateral photo-identification mark-recapture records using a “complete data likelihood” (Dempster et al. 1977), where the goal is to model the data we “wish we had” instead of the messy data that we do have. We start by writing down this likelihood in a generic sense before describing its individual components. Following Link et al. (2010), we write

$$[\mathbf{f} | \boldsymbol{\theta}] = \sum_{\mathbf{x}} [\mathbf{f} | \mathbf{x}, \boldsymbol{\theta}] [\mathbf{x} | \boldsymbol{\theta}] \quad (2)$$

where \mathbf{f} is the vector of recorded (observed) encounter history frequencies (i.e., the messy data) and \mathbf{x} is a vector of latent (unobserved) encounter history frequencies (Table 2). Writing the complete data likelihood in this manner emphasizes two points. First, the complete data likelihood is a product of one of a number of mark-recapture likelihoods, $[\mathbf{x} | \boldsymbol{\theta}]$, and $[\mathbf{f} | \mathbf{x}, \boldsymbol{\theta}]$, which describes the process by which the latent histories are translated into recorded histories. The second point in writing the likelihood in this manner is to emphasize that the likelihood may only be evaluated by solving a multi-dimensional summation over the (unobservable) latent history frequencies, thus making maximum likelihood estimation difficult. We avert this problem by adopting a Bayesian perspective and conducting posterior simulation MCMC; however, several developments are needed first.

For convenience, we will refer to the latent and recorded histories using indices. Designating an L encounter at time t as $\varepsilon_t = 1$, an R encounter as $\varepsilon_t = 2$, and a B encounter as $\varepsilon_t = 3$, the latent histories $\omega_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{iT})$ for $i = 1, \dots, 4^T$ are uniquely indexed by

$$i = 1 + \sum_{t=1}^T \varepsilon_t 4^{T-t}. \quad (3)$$

For example, ω_{20} corresponds to latent history L0B when $T = 3$. Indices for the recorded histories preserve the same ordering as for the latent histories in Eq. 3, but are renumbered after any unobservable recorded histories are removed. For example, for the LR data type when $T = 2$, the $2^{T+1} - 2 = 6$ observable recorded histories $\tilde{\omega}_j \in \{0L, 0R, L0, LL, R0, RR\}$ are indexed by $j = 1, \dots, 6$, respectively. For the LRB data type when $T = 2$, the histories 00, LR, and RL are not observable, so the $4^T - 3^T + 2^{T+1} - 2 = 13$ observable recorded histories maintain the same relative ordering as in Eq. 3 and are indexed by $k = 1, \dots, 13$ (Table 1).

Given a set of latent history frequencies \mathbf{x} , any mark-recapture likelihood $[\mathbf{x} | \boldsymbol{\theta}]$ in Eq. 2 can be modified from the literature to accommodate bilateral encounters. This is accomplished by incorporating several additional parameters, all conditional on detection: the probability of a left-sided encounter (δ^L), the probability of a right-sided encounter (δ^R), and the probability of a both-sided encounter ($\delta^B = 1 - \delta^L - \delta^R$). For example, the classic closed population model allowing temporal variation in detection probability (model M_t ; Otis et al. 1978) estimates population size (N) and the probability of

TABLE 2. Definitions of parameters, latent variables, data, and modeling constructs used in the latent multinomial mark–recapture model for bilateral photo-identification data.

Item	Definition
Parameters	
p_t	Probability that an individual is encountered at time t
δ^R	Probability that an individual, encountered at time t , is photographed on the right side only
δ^L	Probability that an individual, encountered at time t , is photographed on the left side only
δ^B	Probability that an individual, encountered at time t , is photographed on both sides.
Latent variables	
x_i	Latent (unobserved) frequency of encounter history ω_i . Note that \mathbf{x} denotes a column vector of such frequencies [e.g., $(x_1, x_2, \dots, x_{4^T})'$ in the case of abundance estimation]
Data	
T	Number of sampling occasions
f_j	Frequency for recorded (observed) encounter history $\tilde{\omega}_j$. Note that \mathbf{f} denotes a column vector of such frequencies [e.g., $(f_1, f_2, \dots, f_{2^{T+1}-2} - 2)'$ for the LR data type]
Modeling constructs	
ω_i	Vector identifying the i th latent encounter history; $\omega_i = \{\omega_{i1}, \omega_{i2}, \dots, \omega_{iT}\}$
ω_{it}	Encounter type represented by the i th latent encounter history at time t ; $\omega_{it} = 0$ represents a non-encounter, $\omega_{it} = L$ a left-sided encounter, $\omega_{it} = R$ a right-sided encounter, and $\omega_{it} = B$ an encounter on both sides
$\tilde{\omega}_j$	Vector identifying the j th recorded encounter history. The tilde designates that the history is observed (rather than latent)
$\tilde{\omega}_{jt}$	Observation type represented by the j th recorded encounter history at time t ; $\tilde{\omega}_{jt} = 0$ represents no observation, $\tilde{\omega}_{jt} = L$ a left-sided observation, and $\tilde{\omega}_{jt} = R$ a right-sided observation. For the LRB data type, $\tilde{\omega}_{jt} = B$ represents an observation on both sides

Note: Symbols in boldface type represent collections (vectors) of parameters.

detection at time t (p_t , $t = 1, \dots, T$). This model can be extended to accommodate the probability of a left-, right-, or both-sided encounter, given detection, by incorporating the bilateral encounter probabilities $\delta = (\delta^L, \delta^R, \delta^B)$ into the model likelihood:

$$[\mathbf{x}|\mathbf{p}, \delta, N] = \frac{N!}{4^T} \prod_{i=1}^{4^T} \pi_i^{x_i} \quad (4)$$

where π_i gives the probability of latent encounter history ω_i ,

$$\pi_i = \prod_{t=1}^T p_t^{I(\omega_{it} \neq 0)} (1 - p_t)^{I(\omega_{it} = 0)} (\delta^L)^{I(\omega_{it} = L)} (\delta^R)^{I(\omega_{it} = R)} (\delta^B)^{I(\omega_{it} = B)} \quad (5)$$

and $I(y)$ is an indicator function taking the value 1 when argument y is true and 0 otherwise. Conditional on first detection, the classic Cormack-Jolly-Seber model (e.g., Williams et al. 2002) for survival probability from time t to $t + 1$ (ϕ_t) can be similarly modified to accommodate bilateral encounters:

$$[\mathbf{x}|\mathbf{p}, \delta, \phi] \propto \prod_{i=2}^{4^T} \pi_i^{x_i} \quad (6)$$

where

$$\pi_i = \prod_{t=2}^T \left\{ \prod_{j=t}^T \left[\phi_{j-1} p_j^* \prod_{k=t}^{j-1} \phi_{k-1} (1 - p_k) \right]^{m_{ij}} \times \left[1 - \sum_{j=t}^T \phi_{j-1} p_j^* \prod_{k=t}^{j-1} \phi_{k-1} (1 - p_k) \right]^{m_{i0}} \right\} \quad (7)$$

$$p_t^* = p_t (\delta^L)^{I(\omega_{it}=L)} (\delta^R)^{I(\omega_{it}=R)} (\delta^B)^{I(\omega_{it}=B)} \quad (8)$$

and where $m_{ij} = 1$ if latent history ω_i includes a detection at time $t - 1$ with the next detection at time $j \geq t$ (otherwise $m_{ij} = 0$), and $m_{i0} = 1$ if latent history ω_i includes a detection at time $t - 1$ with no subsequent detection (otherwise $m_{i0} = 0$). While modifying mark–recapture models this way is straightforward, the real challenge presented by bilateral encounters is that we do not observe the latent history frequencies (\mathbf{x}) and instead only observe the recorded history frequencies (\mathbf{f}). We therefore modify the approach of Link et al. (2010) to map latent histories to the recorded histories and evaluate Eq. 2.

To implement the method of Link et al. (2010), it is necessary to construct a matrix \mathbf{A} , such that $\mathbf{f} = \mathbf{A}'\mathbf{x}$. The matrix \mathbf{A} formally describes the relationship between recorded and latent encounter histories. In general, the dimension of this matrix (as well as the column vector \mathbf{x}) depends on whether or not one is interested in animals that were never encountered (i.e., all zero encounter histories). Inference about all zero histories is necessary when estimating abundance (as in Eq. 4), but not when conducting analyses that condition on first capture (e.g., when estimating survival as in Eq. 6). We shall proceed by assuming that abundance is indeed a parameter of interest in constructing the \mathbf{A} matrix, understanding that if one is not interested in unobserved animals then one may simply delete the first row and column of \mathbf{A} and the first element of \mathbf{x} .

The \mathbf{A} matrix consists of 4^T rows, where the i th row indicates which recorded histories are generated by latent history ω_i for $i = 1, \dots, 4^T$. The \mathbf{A} matrix therefore

consists of $2^{T+1} - 2$ columns for the LR data type and $4^T - 3^T + 2^{T+1} - 2$ columns for the LRB data type. Each entry in \mathbf{A} is either a 1 or 0, where entry $A_{ij} = 1$ if latent history ω_i generates recorded history $\tilde{\omega}_j$, and $A_{ij} = 0$ otherwise.

Intuition about how \mathbf{A} can be constructed is perhaps best provided through a simple example. Specifically, let us return to the two sample ($T = 2$) mark-recapture experiment conceptualized in Table 1. Using the contributed records column for the LR data type in Table 1, we can construct the corresponding $4^T \times (2^{T+1} - 2)$ \mathbf{A} matrix by replacing each dot (.) with a 0 and any other entry with a 1:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}. \quad (9)$$

For example, the fourth row in Eq. 9 indicates that latent history $\omega_4 = 0\text{B}$ gives rise to recorded histories $\tilde{\omega}_1 = 0\text{L}$ and $\tilde{\omega}_2 = 0\text{R}$. Similarly, the $4^T \times (4^T - 3^T + 2^{T+1} - 2)$ \mathbf{A} matrix for the LRB data type can be constructed from the corresponding contributed records column in Table 1 by replacing each dot (.) with a zero and any other entry with a one. As the dimensions of the \mathbf{A} matrix increase exponentially with the number of sampling occasions, it should be apparent that an algorithm for automating construction of this matrix is desirable. We provide computer code (written in R; R Development Core Team 2007) in the Supplement.

Once the \mathbf{A} matrix is defined, the key step toward constructing an efficient MCMC sampler for the summation over \mathbf{x} in Eq. 2 is to determine a basis for the null space of \mathbf{A}' (Link et al. 2010), which involves solving the system of equations $\mathbf{A}'\mathbf{x} = \mathbf{0}$. In our example with $T = 2$ (Table 1), one such basis for the LRB data type is the set of $r = 3$ column vectors $(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)', (0, 0, -1, 0, -1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)',$ and $(0, -1, 0, 0, 0, 0, 0, 0, -1, 1, 0, 0, 0, 0, 0, 0)'$. That is, if we start with any one solution \mathbf{x}_0 satisfying $\mathbf{A}'\mathbf{x}_0 = \mathbf{f}$, then any \mathbf{x} that is a linear combination of \mathbf{x}_0 and the basis vectors will be a solution to $\mathbf{A}'\mathbf{x} = \mathbf{f}$. For the LR ($r = 4^T - 2^{T+1} + 2$) and LRB ($r = 3^T - 2^{T+1} + 2$) data types, a trivial (but useful) solution \mathbf{x}_0 can be obtained by assuming a one-to-one mapping of the

recorded histories to the latent histories. Rewriting the corresponding component of Eq. 2,

$$[\mathbf{f} | \mathbf{x}, \boldsymbol{\theta}] = \mathbf{I}(\mathbf{A}'\mathbf{x} = \mathbf{f}) \quad (10)$$

such that

$$[\mathbf{f} | \boldsymbol{\theta}] = \sum_{\mathbf{x}: \mathbf{A}'\mathbf{x} = \mathbf{f}} [\mathbf{x} | \boldsymbol{\theta}] \quad (11)$$

the r basis vectors and \mathbf{x}_0 can thus be used to generate efficient candidate proposal values for \mathbf{x} when evaluating the summation in Eq. 11 using MCMC sampling (see Appendix A). Code to automate selection of a set of basis vectors for bilateral photograph analyses is provided in the Supplement.

SIMULATION STUDY

We conducted a small simulation study to compare the relative bias and precision of our integrated approach to several analysis alternatives. We focused our efforts on the closed population mark-recapture model M_t (Otis et al. 1978), which estimates abundance assuming time-varying detection probabilities, but again note that our proposed approach may be used by modifying other mark-recapture models accordingly. The mathematical form for the model M_t likelihood (Eq. 4) is substituted directly for $[\mathbf{x} | \boldsymbol{\theta}]$ in Eq. 11. A full description of our Bayesian analysis, including uninformative prior and full conditional distributions, is provided in Appendix A.

We examined the performance of analysis methods under several different sampling scenarios. We generated data using different combinations of underlying abundance (N), detection probability (p), and bilateral encounter probabilities (δ^L , δ^R , δ^B) for $T = 5$ occasions (see Appendix B). For each scenario, we performed 100 simulations, to which we applied four estimation procedures: (1) inference based on Eq. 11 using the LRB data type; (2) inference based on Eq. 11 using the LR data type; (3) a “traditional” analysis using one side of the animal only (the side resulting in the highest number of observations was used); and (4) an analysis based on the likelihood formulation of Corkrey et al. (2008). To enable meaningful comparisons among estimation procedures, each was fit using MCMC with the same set of prior distributions (see Appendix A for further details).

We found both of the proposed integrated estimators and the one-sided estimator to be generally unbiased across all simulated scenarios with credible intervals for N having coverage rates that were close to nominal (see Appendix B). As expected among these estimators, the integrated models for the LRB and LR data types were more precise than the one-sided estimator. The inclusion of B detections afforded relatively small gains in precision relative to the LR data type, and all three estimators performed similarly for the larger sample simulations with $N = 500$ and $p = 0.4$. The estimator based on Corkrey et al. (2008) was

typically biased with poor coverage and low coefficient of variation.

EXAMPLE: BOBCATS IN SOUTHERN CALIFORNIA

R. S. Alonso, B. T. McClintock, L. M. Lyren, E. E. Boydston, and K. R. Crooks (*unpublished manuscript*) conducted a population study of bobcats (*Lynx rufus*) in southern California, where $T = 8$ camera trap surveys were conducted concurrent with a GPS telemetry study from July 2006 to January 2007. Bobcat pelt patterns are bilaterally asymmetrical, and abundance was originally estimated from these data using one-sided mark-recapture analysis methods. However, detection probabilities were low, and coefficients of variation for abundance estimates exceeded 0.20 using traditional methods. We therefore investigated the potential gains in precision afforded by using our bilateral photo-identification model for these data.

The use of a single camera at each camera station would normally have made B detections (and matching of left- to right-sided photographs) impossible for this study. However, because all GPS-collared individuals were photographed during capture, a photo catalog of both sides was available for these $m = 27$ collared individuals during the camera trap surveys. The photo-identification data arising from the camera trap surveys therefore consisted of true encounter histories (ω) for the collared individuals (i.e., 0, L, R, and B encounters with no uncertainty about the number of unique individuals encountered), and recorded histories ($\tilde{\omega}$) of the LR data type for the individuals without collars.

We applied our bilateral model for the LR data type to estimate N using the recorded histories for the $N - m$ non-collared individuals and the complete encounter histories for the m collared individuals. Using the same Bayesian analysis methods described in Appendix A, we simply constrained the latent history frequencies $\mathbf{x} = (x_1, \dots, x_{4^T})$ to be greater than or equal to the known frequencies for the collared individuals. For example, because 12 collared individuals were never detected at a camera station and had encounter history $\omega_1 = 00000000$, we constrained $x_1 \geq 12$ within our MCMC algorithm. We investigated four candidate models using different parameterizations for detection and bilateral encounter probabilities. These candidate models included (1) temporal variation in \mathbf{p} with $\delta^L \neq \delta^R$; (2) constant detection probability p with $\delta^L \neq \delta^R$; (3) temporal variation in \mathbf{p} with $\delta^L = \delta^R$; and (4) constant detection probability p with $\delta^L = \delta^R$. We evaluated the posterior model probability (PMP) for these four candidates models using reversible jump model updates (e.g., Green 1995) within our MCMC algorithm. More detailed descriptions of the data and analysis are in Appendix C.

After a burn in of 20 000 iterations, it required about 10 hours on our computer (3.4GHz Intel Core i7 processor, 16Gb RAM) to obtain a chain of 100 000 iterations from which posterior summaries were calculated. We ran three chains with a range of starting values

and achieved similar results. We found the most support for models including a constant detection probability, with model 4 receiving 0.77 PMP and model 2 receiving 0.23 PMP. Model-averaged estimates (with 95% credible interval) for δ were $\hat{\delta}_L = 0.46$ (0.37–0.53), $\hat{\delta}_R = 0.46$ (0.38–0.54), and $\hat{\delta}_B = 0.08$ (0.02–0.19), with $\hat{p} = 0.16$ (0.12–0.21) and an abundance point estimate (with 95% credible interval; coefficient of variation) of $\hat{N} = 49$ (45–61; 0.09). Analogous one-sided analyses for models with and without temporal variation in detection probability yielded model-averaged $\hat{N} = 47$ (41–70; 0.16) and $\hat{N} = 45$ (40–66; 0.15) from the left and right sides, respectively.

Our bilateral model produced abundance estimates that were about twice as precise as those from analogous one-sided models. The minimum number of bobcats known to be alive during the study was 41 individuals, and only the credible interval for N from the integrated bilateral analysis exceeded this minimum. When discarding the additional information provided by the collared individuals (i.e., B detections, left- to right-sided matches, and all-zero encounter histories) and using only the recorded histories of the LR data type for all individuals, our model yielded model-averaged $\hat{\delta}^L = 0.40$ (0.27–0.50), $\hat{\delta}^R = 0.41$ (0.27–0.51), and $\hat{\delta}^B = 0.19$ (0.01–0.44), with $\hat{p} = 0.21$ (0.13–0.30) and $\hat{N} = 36$ (28–51; 0.17). Similarly, when ignoring the collared individuals, the analogous one-sided analyses yielded model-averaged $\hat{N} = 38$ (27–61; 0.23) and $\hat{N} = 36$ (26–58; 0.22) for the left- and right-side analyses, respectively. This situation typifies the data often available when conducting photographic mark-recapture experiments, and we suspect the lower abundance estimates when excluding the additional information provided by collared individuals may be an indication of individual capture heterogeneity (a common culprit in the underestimation of abundance). Using traditional mark-recapture methods, R. S. Alonso, B. T. McClintock, L. M. Lyren, E. E. Boydston, and K. R. Crooks (*unpublished manuscript*) found evidence of individual capture heterogeneity in their right-sided analysis, but not in their left-sided analysis. Extension of our model to accommodate individual variation in parameters is a focus of current research.

DISCUSSION

Technological advances, including automated pattern recognition software (e.g., Arzoumanian et al. 2005) have dramatically reduced error rates in matching photographs for certain species and will likely lead to greater application of photo-identification analyses in the future. Cameras are also becoming more affordable, and felid studies have used two cameras per station to try and capture both flanks (e.g., Karanth and Nichols 1998). When remote camera stations are used, this may reduce the need for integrated approaches such as that proposed here. However, when designing remote camera studies, power analyses could prove very useful in determining whether there may be benefits to deploying one camera at more sites (and using an integrated

analysis) rather than multiple cameras at fewer sites (and using a more traditional analysis).

Our approach does not distinguish between both-sided encounters resulting from a single image with both sides visible or from separate left-sided and right-sided images of the same individual. Under certain conditions, there may be some advantages to distinguishing between the different ways a both-sided encounter can occur. A referee noted implicit assumptions about the conditional probability of recording a B given both sides were photographed (α). In the absence of B detections, we assume $\alpha = 0$ for models using the LR data type. Conversely, we assume $\alpha = 1$ for models using the LRB data type. Our bobcat example was somewhat unusual because $\alpha = 0$ for the non-collared individuals and $\alpha = 1$ for the collared individuals. Assuming $\alpha = 1$ for all individuals in a population may be appropriate under certain sampling scenarios, such as when sampling occasions are instantaneous. For example, aerial or “snapshot” surveys will often produce at most one detection for each individual per sampling occasion and B encounters can only occur when both sides are simultaneously visible. Regardless, we found only a slight advantage to the inclusion of perfectly observed B detections (i.e., when $\alpha = 1$) in our simulation study. Similarly, we found little advantage to including partially observed B detections in an extended model that accounts for $0 < \alpha < 1$ (see Appendix D). This was an interesting result, and may not necessarily hold for other mark–recapture models, such as those including survival (S. J. Bonner and J. Holmberg, unpublished manuscript) or individual variation in parameters.

Similar to the misidentification error parameter of Link et al. (2010), it is perhaps not intuitive that information about δ^B is provided by the LR data type. To illustrate, consider the extreme case where we are fortunate enough to observe all animals, yielding only the recorded histories LLL and RRR (both with frequency N). In this case, both p and δ^B must be close to one for sufficiently large N ; otherwise, some non-detections (0) would have been recorded. Using uninformative priors, our simulations corroborated that δ^B can be informed by the LR data type in this manner.

There are many potential refinements and extensions to our integrated photo-identification approach. Our simulation study and bobcat example focused on a closed population abundance model, but any other mark–recapture model could be modified accordingly and used for $[x | \theta]$ in Eq. 11, including open population models, such as the Cormack–Jolly–Seber (Eq. 6) and more recent multi-state and robust design formulations (e.g., Williams et al. 2002). As suggested by our bobcat example, the incorporation of individual capture heterogeneity would be a very worthwhile extension. Individual covariates might also be useful for limiting the number of possible matches between recorded and latent histories. Spatially explicit mark–recapture meth-

ods (e.g., Royle et al. 2009) could be useful for modeling a particular type of capture heterogeneity that is induced by the differential exposure of individuals to cameras (e.g., due to home range behavior). Spatially explicit methods could also be used to more realistically map recorded histories to latent histories (e.g., see discussion by Wright 2011:192–201). For example, recorded histories $0L_i0L_i$ and R_j0R_j0 (where subscripts indicate location) would be unlikely to arise from latent history $R_jL_iR_jL_i$ if locations i and j are not within the same individual’s home range or are relatively far apart given the movement characteristics of the species.

Similar to Link et al. (2010), the computational demands of using basis vectors to propose \mathbf{x} from the feasible set satisfying $\mathbf{f} = \mathbf{A}'\mathbf{x}$ can be impractical for large T . These computational demands can be somewhat reduced by eliminating basis vectors that will always produce negative latent history frequencies for any given \mathbf{f} , as demonstrated in the provided R code in the Supplement. However, in the absence of gains in computing power, more efficient methods for evaluating Eq. 11 will likely be needed for $T > 10$.

Our integrated approach provides a means for practitioners to revisit previous analyses that were limited to left- or right-side data only. This may provide more precise and consistent inferences, particularly if one-sided sample sizes were small or the independent analyses produced conflicting results. If detections on both sides never occurred (or were not originally recorded as such), a new analysis using our integrated model for the LR data type would not require raw photographic records to be revisited. However, as demonstrated by our simulation studies, the incorporation of B encounters could slightly increase precision. Because photo-identification records are often used to monitor species of concern (e.g., Wilson et al. 1999, Holmberg et al. 2008, Berrow et al. 2012, Nair et al. 2012), re-visiting previous analyses using the proposed methodology may help reduce uncertainty in the management of these species.

Although motivated by bilateral photo-identification records, the proposed methodology can be used to combine and jointly analyze other sources of mark–recapture data. For example, Madon et al. (2011) estimated humpback whale abundance from photo and genetic records using a modified Jolly–Seber model (but see Bonner 2013). To utilize our approach for such a sampling scenario, one would simply redefine the encounter types (e.g., L, R) and encounter probabilities (δ) accordingly.

ACKNOWLEDGMENTS

We thank USGS, E. Boydston, and L. Lyren for pivotal roles in the bobcat study. J. Laake, W. Link, J. Ver Hoef, and an anonymous referee for comments on a previous draft. Several colleagues (S. Bonner, R. Fewster) have communicated that they are independently researching methods for analyzing bilateral photo-identification data; we recommend considering their papers for simultaneous citation. The findings and

conclusions in the paper are those of the authors and do not necessarily represent the views of any government agency.

LITERATURE CITED

- Arzoumanian, Z., J. Holmberg, and B. Norman. 2005. An astronomical pattern-matching algorithm for computer-aided identification of whale sharks *Rhincodon typus*. *Journal of Applied Ecology* 42:999–1011.
- Berrow, S., J. O'Brien, L. Groth, A. Foley, and K. Voigt. 2012. Abundance estimate of bottlenose dolphins (*Tursiops truncatus*) in the Lower River Shannon candidate Special Area of Conservation, Ireland. *Aquatic Mammals* 38:136–144.
- Bonner, S. J. 2013. Response to: A new method for estimating animal abundance with two sources of data in capture–recapture studies. *Methods in Ecology and Evolution*. *In press*.
- Calambokidis, J., and J. Barlow. 2004. Abundance of blue and humpback whales in the eastern North Pacific estimated by capture–recapture and line-transect methods. *Marine Mammal Science* 20:63–85.
- Corkrey, R., S. Brooks, D. Lusseau, K. Parsons, J. W. Durban, P. S. Hammond, and P. M. Thompson. 2008. A Bayesian capture–recapture population model with simultaneous estimation of heterogeneity. *Journal of the American Statistical Association* 103:948–960.
- Dempster, A., N. Laird, and D. Rubin. 1977. Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B* 39:1–38.
- Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732.
- Hiby, L., W. D. Paterson, P. Redman, J. Watkins, S. D. Twiss, and P. Pomeroy. 2013. Analysis of photo-id data allowing for missed matches and individuals identified from opposite sides. *Methods in Ecology and Evolution* 4:252–259.
- Holmberg, J., B. Norman, and Z. Arzoumanian. 2008. Robust, comparable population metrics through collaborative photo-monitoring of whale sharks *Rhincodon typus*. *Ecological Applications* 18:222–233.
- Karanth, K. I. L., and J. D. Nichols. 1998. Estimation of tiger densities in India using photographic captures and recaptures. *Ecology* 79:2852–2862.
- Link, W. A., J. Yoshizaki, L. L. Bailey, and K. H. Pollock. 2010. Uncovering a latent multinomial: analysis of mark–recapture data with misidentification. *Biometrics* 66:178–185.
- Mackey, B. L., J. W. Durban, S. J. Middlemas, and P. M. Thompson. 2008. A Bayesian estimate of harbour seal survival using sparse photo-identification data. *Journal of Zoology* 274:18–27.
- Madon, B., O. Gimenez, B. McArdle, C. S. Baker, and C. Garrigue. 2011. A new method for estimating animal abundance with two sources of data in capture–recapture studies. *Methods in Ecology and Evolution* 2:390–400.
- Nair, T., J. B. Thorbjarnarson, P. Aust, and J. Krishnaswamy. 2012. Rigorous gharial population estimation in the Chambal: implications for conservation and management of a globally threatened crocodilian. *Journal of Applied Ecology* 49:1046–1054.
- Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical-inference from capture data on closed animal populations. *Wildlife Monographs* 62:7–135.
- R Development Core Team. 2007. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.r-project.com
- Royle, J. A., K. U. Karanth, A. M. Gopalaswamy, and N. S. Kumar. 2009. Bayesian inference in camera trapping studies for a class of spatial capture–recapture models. *Ecology* 90:3233–3244.
- Samuel, M. D., D. H. Rusch, and S. Craven. 1990. Influence of neck bands on recovery and survival rates of Canada geese. *Journal of Wildlife Management* 54:45–54.
- Soisalo, M. K., and S. M. C. Cavalcanti. 2006. Estimating the density of a jaguar population in the Brazilian Pantanal using camera-traps and capture–recapture sampling in combination with GPS radio-telemetry. *Biological Conservation* 129:487–496.
- Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego, California, USA.
- Wilson, B., P. S. Hammond, and P. M. Thompson. 1999. Estimating size and assessing trends in a coastal bottlenose dolphin population. *Ecological Applications* 9:288–300.
- Wright, J. A. 2011. Incorporating genotype uncertainty into mark–recapture-type models for estimating abundance using DNA samples. Dissertation. University of Otago, Dunedin, New Zealand.

SUPPLEMENTAL MATERIAL

Appendix A

Details of Bayesian analysis for bilateral photo-identification data ([Ecological Archives E094-132-A1](#)).

Appendix B

Simulation methods and results ([Ecological Archives E094-132-A2](#)).

Appendix C

Description of the bobcat data and integrated analysis ([Ecological Archives E094-132-A3](#)).

Appendix D

Extension for partially observed B encounters ([Ecological Archives E094-132-A4](#)).

Supplement

Example code for constructing A matrix and calculating basis vectors ([Ecological Archives E094-132-S1](#)).