

Multivariate State Hidden Markov Models for Mark-Recapture Data

Devin S. Johnson, Jeff L. Laake, Sharon R. Melin and Robert L. DeLong

Abstract. State-based Cormack–Jolly–Seber (CJS) models have become an often used method for assessing states or conditions of free-ranging animals through time. Although originally envisioned to account for differences in survival and observation processes when animals are moving through various geographical strata, the model has evolved to model vital rates in different life-history or diseased states. We further extend this useful class of models to the case of multivariate state data. Researchers can record values of several different states of interest, for example, geographic location and reproductive state. Traditionally, these would be aggregated into one state with a single probability of state uncertainty. However, by modeling states as a multivariate vector, one can account for partial knowledge of the vector as well as dependence between the state variables in a parsimonious way. A hidden Markov model (HMM) formulation allows straightforward maximum likelihood inference. The proposed HMM models are demonstrated with a case study using data from a California sea lion vital rates study.

Key words and phrases: Capture-recapture, Cormack–Jolly–Seber, hidden Markov model, multivariate, partial observation, state uncertainty.

1. INTRODUCTION

The seminal papers by Cormack (1964), Jolly (1965) and Seber (1965) initiated 50 years of active development of capture-recapture theory and application. The Cormack–Jolly–Seber (CJS) model for survival estimation and the Jolly–Seber model for survival and abundance estimation and their extensions are still the most widely used capture-recapture models. They have been implemented in the computer software MARK (White and Burnham, 1999), POPAN (Arnason and Schwarz, 1999), SURGE (Lebreton et al., 1992), M-SURGE (Choquet et al., 2004), E-SURGE (Choquet, Rouan and Pradel, 2009) and, more recently, marked (Laake, Johnson and Conn, 2013) and multimark (McClintock, 2015).

Devin S. Johnson is Statistician, Jeff L. Laake is Statistician, Sharon R. Melin is Wildlife Research Biologist, Robert L. DeLong is Program Leader, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, Washington, USA (e-mail: devin.johnson@noaa.gov; jeff.laake@noaa.gov; sharon.melin@noaa.gov; bob.delong@noaa.gov).

Expansion of CJS models to account for movement between areas (multistate) was initiated by Darroch (1958) and Arnason (1973) and further developed by Hilborn (1990), Hestbeck, Nichols and Malecki (1991), Schwarz, Schweigert and Arnason (1993) and Brownie et al. (1993). Nichols et al. (1992) used the multistate model to describe survival with state transitions based on mass and this has been followed by numerous other examples using the multistate model (Lebreton and Pradel, 2002; Nichols and Kendall, 1995; White, Kendall and Barker, 2006). An important advance to the multistate CJS generalization was made by Kendall et al. (2004) who considered models with state uncertainty (unknown state) and Pradel (2005) who cast a multistate model with state uncertainty as a hidden Markov model (Zucchini and MacDonald, 2009). The state uncertainty models are often termed multievent models. In a further step, Kendall et al. (2012) combined a robust design with a multistate hidden Markov model to improve precision in the face of state uncertainty. We use the general term *state-based* to refer to all of these types of CJS models.

For traditional CJS analysis, the detection portion of the model is often considered to be a nuisance with

little scientific interest and often determined largely by the resampling methods used by researchers to resight individuals. The achievement made by the development of the original CJS model was that it permitted estimation of survival even when individuals were not observed on every occasion. The state-based extensions allowed survival inference to account for time-varying heterogeneity or deviations from CJS model assumptions induced by states as well as to make inference to the state process itself (Lebreton and Pradel, 2002). With the recently developed extensions, one can often find state-based CJS analyses where survival is also considered a nuisance process and primary scientific interest lies with the transitions between states. See Gourlay-Larour et al. (2014) for a recent example. With this in mind, practicing ecologists are using the state-based CJS framework and the associated HMM formulation to analyze complex state transitions such as the combination of geographic location and reproductive or disease status. This presents the problem that a state may be only partially observed when an individual is resighted. For example, location will be known but reproductive status may not be observed. To further generalize the state-based CJS models, King and McCrea (2014) developed a closed-form likelihood expression to handle partial knowledge of the state of an individual. However, a general framework for parametrizing complex state-based CJS models is not examined. Laake et al. (2014) made initial strides into multivariate state CJS models by proposing a bivariate state model to account for dependence in double-mark loss. Here we seek to augment the work of King and McCrea (2014) and Laake et al. (2014) to provide a general method to construct complex state-based CJS models.

In this paper we propose a general modeling framework for multivariate state-based CJS models in which the state is defined by one or more discrete categorical variables and each variable may be unknown when the animal is resighted. The modeling framework extends the state-based models in MARK by allowing any or all of the state variables to be uncertain rather than just a single uncertain state. We have implemented the model in the `marked` package (Laake, Johnson and Conn, 2013) for the R statistical environment (R Development Core Team, 2015) using a hidden Markov model (HMM) formulation for maximum likelihood estimation of the parameters. General multivariate HMMs have been proposed in the past (Ghahramani and Jordan, 1997, Brand, Oliver and Pentland, 1997), but here we model the state vector in a log-linear framework

(Christensen, 1997) to easily allow all levels of dependence among state variables. The proposed model is illustrated with a case study using 18 years of annual resighting data collected on a single cohort of California sea lion (*Zalophus californianus*) pups that were both branded and double tagged on their flippers in 1996 at San Miguel Island near Santa Barbara, California (Melin et al., 2011). The animal's state was defined based on location (San Miguel or Año Nuevo Islands), which was always known when the animal was resighted and the, possibly unknown, states of the two flipper tags (present or absent due to tag loss).

2. METHODS

To begin the description of multivariate state modeling of mark-recapture data, we must begin with the state and data structure. The general structure is based on the capture histories of individuals, indexed $i = 1, \dots, I$, over capture occasions, indexed $j = 1, \dots, J$. The state of an individual on a particular occasion is determined by the cross-classification of that individual to a cell in a multiway table where the margins represent state variables of interest. Specifically, the state $\mathbf{s} = (s_1, \dots, s_K)$ is a vector where the entries correspond to K classification criteria, for example, location, reproductive status or disease presence. Each state variable has a set of possible values $s_k \in \mathcal{S}_k = \{1, \dots, m_k\}$, $k = 1, \dots, K$. The space of the state vector, \mathbf{s} , is thus $\mathcal{S} = \mathcal{S}_1 \times \dots \times \mathcal{S}_K$. The number of possible states is denoted $M = |\mathcal{S}| = \prod_k m_k$. We also need to augment the state-space, \mathcal{S} , with a "death" state, such that the individual can transition between all $\mathbf{s} \in \mathcal{S}$ as well as death. We represent the death state with $\mathbf{s} = \mathbf{0}$, so the augmented state space is $\mathcal{S}^\dagger = \mathcal{S} \cup \{\mathbf{0}\}$. The observed data for an individual on a particular (re)capture occasion is denoted as $\mathbf{c} = (c_1, \dots, c_K)$. The observed state can differ from the true state in two ways. First, an individual may not be observed (detected) on a particular capture occasion. Second, it is possible, and highly probable in many situations, that even if an individual is physically observed, state specification may not be completely determined by the researcher. Thus, the space of the observation vector is augmented with two additional levels corresponding to "not observed" and "unknown," that is, $c_k \in \{0, 1, \dots, m_k, m_k + 1\}$, where 0 is for undetected individuals and $m_k + 1$ is associated with an unknown level for that state variable. In heuristic descriptions we will use $c_k = u$ to be clear we are referring to the unknown state. In addition, we take the

convention that if $c_k = 0$ for any k , it is assumed that all $c_k = 0$. We use c_{ij} to denote the observed state of individual i on occasion j . Table A.1 of Supplement A (Johnson et al., 2016a) provides a glossary of the notation used in this paper.

2.1 Two-State Variable Model

Specifying a model for a general multivariate discrete state is notationally cumbersome due to the complex dependence structure. Therefore, we begin with a description for a bivariate state model, and then follow up in the next section with a general model formulation for >2 state variables with additional covariates. The general parameter types are the same as in any state-based CJS model: survival (ϕ), detection (p), and state transition (ψ) from one occasion to the next. In addition, we have parameters associated with the inability to fully assess the state of an individual (δ) and the probability that an individual is in a particular state at first capture (π). Typically, each of these parameter types depends on the state of the individual; for example, survival of individual i from occasion j to occasion $j + 1$ depends on the state of the individual on occasion j . Each multivariate state variable may contribute independently to survival, or may interact at various levels to further enhance or degrade survival. Thus, models for these parameters must accommodate this interaction (dependence) when modeling state transition. We take a log-linear approach (Christensen, 1997) for modeling multiway contingency tables which allows for all levels of dependence between state variables.

We begin with modeling state transitions from one occasion to the next. Suppose individual i is alive on occasion $j < J$, then the probability of a transition from state $\mathbf{r} = (r_1, r_2) \in \mathcal{S}$ on occasion j to state $\mathbf{s} = (s_1, s_2) \in \mathcal{S}$ on occasion $j + 1$ is given by the conditional log-linear model,

$$(2.1) \quad \psi(\mathbf{s}; \mathbf{r}) \propto \exp\{\beta^{(1)}(s_1; \mathbf{r}) + \beta^{(2)}(s_2; \mathbf{r}) + \beta^{(1,2)}(s_1, s_2, \mathbf{r})\},$$

where, for each $\mathbf{r} \in \mathcal{S}$, $\beta^{(1)}(s_1; \mathbf{r})$, $\beta^{(2)}(s_2; \mathbf{r})$ and $\beta^{(1,2)}(s_1, s_2; \mathbf{r})$ are parameters which depend on the previous state, \mathbf{r} , and are indexed by the values of s_1 , s_2 and the double index of (s_1, s_2) , respectively. For example, the collection of $\beta^{(1)}$ parameters is $\{\beta^{(1)}(1; \mathbf{r}), \dots, \beta^{(1)}(m_1; \mathbf{r})\}$. If we denote Ψ to be the transition matrix describing the probability of transitioning from any state on occasion j to any other state on occasion $j + 1$, equation (2.1) evaluated for

all $\mathbf{s} \in \mathcal{S}$ would form the rows associated with each possible current state \mathbf{r} . There are $m_1 + m_2 + m_1m_2$ parameters for each \mathbf{r} . Thus, there are $m_1m_2 \times (m_1 + m_2 + m_1m_2)$ β parameters within the $m_1m_2 \times m_1m_2$ transition matrix Ψ . Hence, the model is overparameterized. To alleviate this problem, it is customary to pick a reference cell, say, $\mathbf{s}^* \in \mathcal{S}$, and fix all β parameters to 0 when $s_1 = s_1^*$ or $s_2 = s_2^*$. For example, if $\mathbf{s}^* = (r_1, r_2)$ (the current state), $\beta^{(1)}(r_1; \mathbf{r}) \equiv 0$, $\beta^{(2)}(r_2; \mathbf{r}) \equiv 0$, $\beta^{(1,2)}(s_1, r_2; \mathbf{r}) \equiv 0$ for any s_1 value, and $\beta^{(1,2)}(r_1, s_2; \mathbf{r}) \equiv 0$ for any s_2 value. Of course, it is not necessary to use the current state as the reference state, but it has some advantages, namely, the β parameters are interpreted as controlling movement away from the current state and the current state is usually an acceptable possibility for a realization at the next occasion (i.e., nonzero probability of occurring). This may not be the case if you define a fixed state as the reference. See Table 1 for an example of $\psi(\mathbf{s}; \mathbf{r})$ specification with the reference cell constraint.

Markovian dependence in time is explicitly included by allowing the β parameters to depend on the previous state. In addition, dependence between state variables is accommodated via the $\beta^{(1,2)}$ interaction parameters. If $\beta^{(1,2)}(s_1, s_2; \mathbf{r}) = 0$ for all values of s_1 and s_2 , then this is a necessary and sufficient condition for s_1 and s_2 to be independent given the previous state \mathbf{r} . Sufficiency is easily observed by the fact that under that assumption

$$(2.2) \quad \begin{aligned} \psi(\mathbf{s}; \mathbf{r}) &\propto \exp\{\beta^{(1)}(s_1; \mathbf{r}) + \beta^{(2)}(s_2; \mathbf{r})\} \\ &\propto \psi^{(1)}(s_1; \mathbf{r}) \times \psi^{(2)}(s_2; \mathbf{r}). \end{aligned}$$

Thus, the probability of transition to state \mathbf{s} is the product of independent transitions from $r_1 \rightarrow s_1$ and $r_2 \rightarrow s_2$. Proof of necessity is quite cumbersome and beyond the scope of this paper. See Lauritzen (1996) for more theoretical details of log-linear modeling. One caveat of these independence results is that they assume there are no structural zeros (impossible state transitions). If there are, one should check that independence still holds if it is a desirable property. Otherwise, zero-valued interaction terms can just be used for parsimonious modeling.

The next set of parameters of scientific interest is survival. Given that individual i is in state \mathbf{s} on occasion j , the probability of survival from occasion j to $j + 1$ is described by

$$(2.3) \quad \begin{aligned} \phi(\mathbf{s}) &= \ell^{-1}\{\eta_{\emptyset} + \eta^{(1)}(s_1) + \eta^{(2)}(s_2) \\ &\quad + \eta^{(1,2)}(s_1, s_2)\}, \end{aligned}$$

TABLE 1
Example transition probabilities, $\psi(\mathbf{s}; \mathbf{r})$, for a hypothetical bivariate state \mathbf{s} composed of state variables $s_1 \in \{1, 2, 3\}$ and $s_2 \in \{1, 2\}$. Here we use the reference cell $\mathbf{s}^ = (1, 1)$. The ψ probabilities are presented in their unnormalized form as in equation (2.1). The necessary normalizing constant is the sum over all the table entries*

		s_2	
		1	2
s_1	1	1	$\exp\{\beta^{(2)}(2; \mathbf{r})\}$
	2	$\exp\{\beta^{(1)}(2; \mathbf{r})\}$	$\exp\{\beta^{(1)}(2; \mathbf{r}) + \beta^{(2)}(2; \mathbf{r}) + \beta^{(1,2)}(2, 2; \mathbf{r})\}$
	3	$\exp\{\beta^{(1)}(3; \mathbf{r})\}$	$\exp\{\beta^{(1)}(3; \mathbf{r}) + \beta^{(2)}(2; \mathbf{r}) + \beta^{(1,2)}(3, 2; \mathbf{r})\}$

where ℓ is a link function such as logit or probit, and the η parameters are functions of (indexed by) the current state, \mathbf{s} . The η parameters have the same structure and reference cell constraints as the β parameters in the ψ model. Readers should also observe that there is a η_{\emptyset} parameter. This term is not indexed by the state and represents baseline survival. Even though the notation may appear different for some readers, it should be noted that this ϕ model is identical to a model with s_1 and s_2 being thought of as known categorical covariates with interaction effects included. So, as far as the survival model is concerned, s_1 and s_2 are equivalent to (individual \times occasion) indexed factor covariates.

We now move to the observation and detection portion of the model. Given that individual i is in state \mathbf{s} on occasion j , the probability that the individual is detected (i.e., recaptured or resighted) by the researcher is given by

$$(2.4) \quad p(\mathbf{s}) = \ell^{-1} \{ \zeta_{\emptyset} + \zeta^{(1)}(s_1) + \zeta^{(2)}(s_2) + \zeta^{(1,2)}(s_1, s_2) \}.$$

As with the survival model, the state variables can be considered as factor variables for the purposes of conceptualizing the model. There is one more set of parameters working in concert with p ; these are related to the ability of the researcher to jointly classify a detected individual according to all of the state variables. If an individual i is detected on occasion j , then the joint probability of the state observation, $\mathbf{c} = (c_1, c_2)'$, conditioned on the true state is modeled via

$$(2.5) \quad \delta(\mathbf{c}; \mathbf{s}) \propto \exp\{ \alpha^{(1)}(c_1; \mathbf{s}) + \alpha^{(2)}(c_2; \mathbf{s}) + \alpha^{(1,2)}(c_1, c_2; \mathbf{s}) \}, \quad \mathbf{c} \neq \mathbf{0}.$$

Note that we do not consider misidentification of states; a state variable is either observed correctly

($c_k = s_k$) or it is not observed and the value is unknown ($c_k = m_k + 1$ or u). As with the multivariate transition model, ψ , a reference cell, say \mathbf{c}^* , is necessary for parameter identifiability. One could use $\mathbf{c}^* = (m_1 + 1, m_2 + 1)$, the completely unknown state, to be consistent with previous univariate state uncertainty models (e.g., Laake, 2013, Kendall et al., 2012), but it is not necessary and may be undesirable if the completely unknown state is not observable, for example, double-tag studies where no third permanent mark is available. We suggest using the current state, $\mathbf{c}^* = \mathbf{s}$, as the reference. As with the ψ models, one can incorporate dependence in recording the level of each state variable. That dependence implies, for example, that if the researcher is able to record the level for one state variable, the researcher is more (less) likely to record the level of the other. There may be some state variables which are always recorded with certainty, for example, location. In that situation, one can simply remove any \mathbf{c} that contains an unknown observation for those variables, or equivalently set the corresponding δ value to zero. As with the state transition probabilities, ψ , if these structural zeros exist, then the dependence interpretation needs to be investigated because conditional dependence can be induced by the zeros.

Finally, the last set of parameters is associated with the initial capture of the individual. We are considering only CJS type models here, thus, we are still conditioning the model on the occasion when an individual is first observed. When an individual is first captured (marked), however, the researcher may not be able to observe all state variables. Therefore, we need to model the probability that individual i , who is first marked on occasion j , is in state \mathbf{s} . So, we describe this probability via

$$(2.6) \quad \pi(\mathbf{s}) \propto \exp\{ \lambda^{(1)}(s_1) + \lambda^{(2)}(s_2) + \lambda^{(1,2)}(s_1, s_2) \}.$$

Here, there is no previous cell from which to form the reference cell. Therefore, the researcher will have to decide on the most appropriate state to serve as the reference. The π model parameters, however, may not necessarily have to be estimated. For example, if it is certain that all state variables will be observed at first marking, then it can be fixed to $\pi(\mathbf{s}) = 1$, where \mathbf{s} is the state individual i was in when it was first marked, and $\pi(\mathbf{r}) = 0$ for all $\mathbf{r} \neq \mathbf{s}$ (see Section 3). Or, a researcher might specify, say, $\pi(\mathbf{s}) = 1/M$, where M is the number of cells in \mathcal{S} . The π probabilities can be interpreted as a prior distribution, in the Bayesian sense, on the state of a newly marked individual, thus, $1/M$ would be a uniform prior over the state-space.

2.2 General Multivariate State Model

In this section we generalize the bivariate state models presented in the previous section. Here, researchers can consider state specifications that index a cell in a general K -dimensional hypercube and models can depend on covariates as well. To accomplish this, we introduce the notation $\mathcal{V} = \{1, \dots, K\}$ to be the set which indexes the collection of K state variables and $v \subseteq \mathcal{V}$ is a subset. The partial state $\mathbf{s}^{(v)}$ represents only those elements of \mathbf{s} whose indexes are contained in $v \subseteq \mathcal{V}$. The state subspace $\mathcal{S}^{(v)}$ denotes the set of all possible values of $\mathbf{s}^{(v)}$.

Now, the fully specified model describing the transition of individual i in state $\mathbf{r} = (r_1, \dots, r_K)$ on occasion j to state $\mathbf{s} = (s_1, \dots, s_K)$ on occasion $j + 1$ is given by

$$(2.7) \quad \psi_{ij}(\mathbf{s}; \mathbf{r}) \propto \exp \left\{ \sum_{v \subseteq \mathcal{V}} \mathbf{x}_{ij}^{(v)} \cdot \boldsymbol{\beta}_{ij}^{(v)}(\mathbf{s}; \mathbf{r}) \right\},$$

where $\mathbf{x}_{ij}^{(v)}$ is a vector of covariates, the notation “ $\mathbf{x} \cdot \boldsymbol{\beta}$ ” represents the dot product $\mathbf{x}'\boldsymbol{\beta}$, and $\boldsymbol{\beta}_{ij}^{(v)}(\mathbf{s}; \mathbf{r})$ is a coefficient vector that depends on the state, $\mathbf{s} \in \mathcal{S}$, only through the value of the substate $\mathbf{s}^{(v)} \in \mathcal{S}^{(v)}$ and $\mathbf{r} \in \mathcal{S}$. As with the bivariate model in the previous section, we must satisfy the reference cell constraint to have an identifiable set of parameters. Therefore, all $\boldsymbol{\beta}^{(v)}(\mathbf{s}; \mathbf{r}) \equiv \mathbf{0}$ if any element of the substate $\mathbf{s}^{(v)}$ is equal to the corresponding elements of the reference cell, \mathbf{s}^* . Independence of state variables is more complex than described in the previous section due to a number of possible interactions between different sets of variables. To provide a generalization, let v_1, v_2 and v_3 be subsets of state variables that partition \mathcal{V} . Then $\mathbf{s}^{(v_1)}$ and $\mathbf{s}^{(v_3)}$ are conditionally independent given $\mathbf{s}^{(v_2)}$ (and \mathbf{r}) if and only if $\boldsymbol{\beta}^{(v)}(\mathbf{s}; \mathbf{r}) = \mathbf{0}$ for all \mathbf{s} , where v

contains elements of both v_1 and v_3 (i.e., $v \cap v_1 \neq \emptyset$ and $v \cap v_3 \neq \emptyset$) (Frydenberg, 1990). For example, suppose $\mathcal{V} = \{1, 2, 3\}$, then s_1 is conditionally independent of s_3 given s_2 (and \mathbf{r}) if $\boldsymbol{\beta}^{(1,3)}(\mathbf{s}; \mathbf{r}) = \boldsymbol{\beta}^{(1,2,3)}(\mathbf{s}; \mathbf{r}) = \mathbf{0}$ for all \mathbf{s} . Further, if we assume that $\boldsymbol{\beta}^{(1,2)}(\mathbf{s}; \mathbf{r}) = \mathbf{0}$, then $\mathbf{s}^{(1)} = s_1$ is independent (given \mathbf{r}) of $\mathbf{s}^{(2,3)} = (s_2, s_3)$. In that case, we can follow the example in (2.2) and write

$$(2.8) \quad \psi_{ij}(\mathbf{s}; \mathbf{r}) = \psi_{ij}^{(1)}(\mathbf{s}; \mathbf{r}^{(1)}) \psi_{ij}^{(2,3)}(\mathbf{s}; \mathbf{r}^{(2,3)}).$$

This independence factorization can extend to more than two groups using the Gibbs factorization theorem (see Frydenberg, 1990). Again, independence theorems are based on the absence of structural zeros.

We can progress through the other parameter groups in a similar fashion to produce a general multivariate state model. The survival portion of the model can be generalized to

$$(2.9) \quad \phi_{ij}(\mathbf{s}) = \ell^{-1} \left\{ \sum_{v \subseteq \mathcal{V}} \mathbf{x}_{ij}^{(v)} \cdot \boldsymbol{\eta}_{ij}^{(v)}(\mathbf{s}) \right\},$$

where the covariate vector $\mathbf{x}_{ij}^{(v)}$ need not be the same as that shown in equation (2.7); we use the same notation simply to avoid extra clutter. Cycling through the remaining parameters, we have

$$(2.10) \quad p_{ij}(\mathbf{s}) = \ell^{-1} \left\{ \sum_{v \subseteq \mathcal{V}} \mathbf{x}_{ij}^{(v)} \cdot \boldsymbol{\zeta}_{ij}^{(v)}(\mathbf{s}) \right\}, \quad \mathbf{s} \neq \mathbf{0},$$

$$(2.11) \quad \delta_{ij}(\mathbf{c}; \mathbf{s}) \propto \exp \left\{ \sum_{v \subseteq \mathcal{V}} \mathbf{x}_{ij}^{(v)} \cdot \boldsymbol{\alpha}_{ij}^{(v)}(\mathbf{c}; \mathbf{s}) \right\}, \quad \mathbf{c} \neq \mathbf{0} \text{ and } \mathbf{s} \neq \mathbf{0},$$

where we suggest the reference cell $\mathbf{c}^* = \mathbf{s}$. In general, there will be 2^d possible values of \mathbf{c} associated with each \mathbf{s} , where d is the number of state variables for which a u can be recorded. Finally, the probability distribution of states upon first capture is

$$(2.12) \quad \pi_{ij}(\mathbf{s}) \propto \exp \left\{ \sum_{v \subseteq \mathcal{V}} \mathbf{x}_{ij}^{(v)} \cdot \boldsymbol{\lambda}_{ij}^{(v)}(\mathbf{s}) \right\}.$$

This completes the general model for a K -dimensional state vector. In the next section we focus on methods for efficient parameter estimation and inference. Similar to the ψ models, the δ and π parameters can be factored into independent products as well. This will facilitate easier overall model construction if certain state variables can be assumed to be independent, making it easier to parameterize dependence over a small subset of variables.

2.3 Statistical Inference

Statistical inference for multistage mark-recapture models can be challenging (McCrea and Morgan, 2014). We take the approach described by Laake (2013) who uses a hidden Markov model (HMM) formulation. By framing a state-based CJS model as a specific HMM, the efficient *forward algorithm* can be used to maximize the likelihood (Zucchini and MacDonald, 2009). Laake (2013) provides a detailed description of HMM formulated mark-recapture models and likelihood calculation, but we briefly revisit it here to place the forward algorithm in the context of multivariate state CJS models.

To begin the description of the forward algorithm in the multivariate state CJS model case, we must define three matrices. The transition probability matrix for movement within \mathcal{S}^\dagger is given by

$$(2.13) \quad \mathbf{\Gamma}_{ij} = \begin{bmatrix} \phi_{ij}(\mathbf{r}_1)\psi_{ij}(\mathbf{s}_1; \mathbf{r}_1) & \cdots & \vdots & \vdots \\ \phi_{ij}(\mathbf{r}_M)\psi_{ij}(\mathbf{s}_1; \mathbf{r}_M) & \cdots & 0 & 0 \\ \phi_{ij}(\mathbf{r}_1)\psi_{ij}(\mathbf{s}_M; \mathbf{r}_1) & 1 - \phi_{ij}(\mathbf{r}_1) & \vdots & \vdots \\ \phi_{ij}(\mathbf{r}_M)\psi_{ij}(\mathbf{s}_M; \mathbf{r}_M) & 1 - \phi_{ij}(\mathbf{r}_M) & \vdots & \vdots \\ \dots & \dots & \dots & 1 \end{bmatrix},$$

where M is the size of the state-space (recall $M = \prod_k m_k$), \mathbf{r}_m , $m = 1, \dots, M$, refers to the current state on occasion j , \mathbf{s}_m refers to the state to which the animal is transitioning on occasion $j + 1$, and the last column (row) is associated with transition to (from) the death state. In addition to the state-space augmentation, we define the observation probability matrix which describes the probability of observing state \mathbf{c}_n , $n = 0, \dots, N$, for animal i on occasion j (rows) given it was actually in state $\mathbf{s} \in \mathcal{S}^\dagger$ (columns):

$$(2.14) \quad \mathbf{D}_{ij} = \begin{bmatrix} p_{ij}(\mathbf{s}_1)\delta_{ij}(\mathbf{c}_1; \mathbf{s}_1) & \cdots & \vdots & \vdots \\ p_{ij}(\mathbf{s}_1)\delta_{ij}(\mathbf{c}_N; \mathbf{s}_1) & \cdots & 1 - p_{ij}(\mathbf{s}_1) & \cdots \\ p_{ij}(\mathbf{s}_M)\delta_{ij}(\mathbf{c}_1; \mathbf{s}_M) & 0 & \vdots & \vdots \\ p_{ij}(\mathbf{s}_M)\delta_{ij}(\mathbf{c}_N; \mathbf{s}_M) & 0 & 1 - p_{ij}(\mathbf{s}_M) & 1 \end{bmatrix},$$

where observed state $\mathbf{c}_0 = (0, \dots, 0)$, the undetected state, is represented by the last row. For use in the

forward algorithm, we also define the $K \times K$ matrix, $\mathbf{P}(\mathbf{c}_{ij})$, to be a diagonal matrix with the row of \mathbf{D}_{ij} associated with the observed state for individual i on occasion j as the diagonal entries.

Using the $\mathbf{\Gamma}_{ij}$ and $\mathbf{P}(\mathbf{c}_{ij})$ matrices, we can describe the forward algorithm specific to maximum likelihood estimation in the multivariate state MR model. The algorithm proceeds as follows for individual $i = 1, \dots, I$:

1. Initial conditions:
 - Set $w_{if_i} = \boldsymbol{\pi}'_{if_i} \mathbf{P}(\mathbf{c}_{if_i}) \mathbf{1}$, where $f_i = 1, \dots, J - 1$ is the occasion on which individual i was first captured/marked, $\boldsymbol{\pi}_{if_i} = (\pi_{if_i}(\mathbf{s}_1), \dots, \pi_{if_i}(\mathbf{s}_M))'$, and $\mathbf{1}$ is a vector of all 1's.
 - Set $\boldsymbol{\mu}'_{if_i} = \boldsymbol{\pi}_{if_i} \mathbf{P}(\mathbf{c}_{if_i}) / w_{if_i}$.
 - Set $L_i = \log w_{if_i}$.
2. For $j = f_i + 1, \dots, J$:
 - Set $w_{ij} = \boldsymbol{\mu}'_{i,j-1} \mathbf{\Gamma}_{ij} \mathbf{P}(\mathbf{c}_{ij}) \mathbf{1}$.
 - Set $\boldsymbol{\mu}'_{ij} = \boldsymbol{\mu}_{i,j-1} \mathbf{\Gamma}_{ij} \mathbf{P}(\mathbf{c}_{ij}) / w_{ij}$.
 - Set $L_i = L_i + \log w_{ij}$.

The forward probabilities $\boldsymbol{\mu}_{ij}$ represent the conditional probability distribution of the state of individual i on occasion j , given the observed states up to and including occasion j and the parameters. After the algorithm has been run for every individual, the log-likelihood is given by

$$(2.15) \quad L = \sum_{i=1}^I L_i.$$

The log-likelihood can be maximized to provide parameter estimates and biological inference. In the next section we will provide a real-world example where multivariate states are collected in a mark-resight study of California sea lions.

3. EXAMPLE: MOVEMENT, SURVIVAL AND MARK LOSS IN CALIFORNIA SEA LIONS

3.1 Data and Model Description

As an example of a multivariate state CJS analysis we use 18 years of annual resighting data collected on a single cohort of California sea lion (*Zalophus californianus*) pups. In 1996, a total of 485 4–5 month old pups born on San Miguel Island (SMI), off the west coast near Santa Barbara, California (Melin et al., 2011) were herded into a pen over a total period of five days. The sex and weight of each pup was recorded and a unique permanent number was applied as hot brand to

their left side. In addition, a uniquely numbered yellow roto-tag was applied to each of their foreflippers. Pups were resighted by their brand during a three month period from 15 May to 15 August each year at SMI and Año Nuevo Island (ANI). SMI is a breeding rookery, whereas ANI is a haul out where animals rest and only a few pups are born and raised. When the animal was resighted, the presence (+) or absence (−) of each tag was noted. In many cases, however, the status was not recorded for one or both of the tags. Although the flipper tags were not directly relevant to the resighting effort, their presence allowed a rare opportunity to assess loss and dependence for other pinniped studies that rely solely on flipper tags for marking (e.g., Testa et al., 2013).

From the sea lion data, a capture history was constructed with annual occasions and each entry was composed of 3 characters, with the first for the area, the second for the left tag and the third for the right tag. There were 8 different possible states for a live animal ($a++$, $a-+$, $a+-$, $a--$, $s++$, $s-+$, $s+-$ and $s--$). Here we are using symbolic representation of the states for ease of discussion. In terms of the notation presented in the previous sections, s_1 is the location state ($s_1 = 1$ for “a” and 2 for “s”), and s_2 and s_3 are the left and right tag status variables ($s_k = 1$ for “+” and 2 for “−”; $k = 2, 3$).

To begin the model specification used to analyze these data, we will start with the observation portions of the model (p and δ). The possible capture history values were 0 if not seen, the 8 possible states if seen and each tag status was recorded, and an additional 10 observations with unknown tag status ($a+u$, $au+$, $a-u$, $au-$, auu , $s+u$, $su+$, $s-u$, $su-$, suu). For these data, location is always known with certainty when the animal is detected but each tag status could be unknown. For the δ model we used the fully known tag status as the reference cell, which implies that the parameters can be interpreted as controlling whether tag status was not obtained (i.e., unknown). The following formulation was used:

$$(3.1) \quad \delta_{ij}(\mathbf{c}; \mathbf{s}) = \delta^{(2,3)}(\mathbf{c}^{(2,3)}, \mathbf{s}) \quad \text{for } c_1 = s_1, \\ \propto \exp\{\alpha^{(2)}(\mathbf{c}) + \alpha^{(3)}(\mathbf{c}) + \alpha^{(2,3)}(\mathbf{c})\}.$$

Table A.2 in Supplement A (Johnson et al., 2016a) illustrates the observation model parameters. For detection, we used the occasion \times area model

$$(3.2) \quad \text{logit } p_{ij}(\mathbf{s}) = \zeta_0 + \zeta_j + \zeta^{(1)}(\mathbf{s}) + \zeta_j^{(1)}(\mathbf{s}),$$

where ζ_j is an occasion effect, $\zeta^{(1)}(\mathbf{s})$ is an area effect, and $\zeta_j^{(1)}(\mathbf{s})$ is an occasion \times area interaction. Resighting of animals was based on the permanent brand mark, thus, tag status was not included in the model.

Now, we describe the more scientifically relevant, process portions of the model. First, all animals started on SMI with two tags [i.e., $\mathbf{s}_{i1} = (s++)$] on the first release occasion. Thus, for this example $\pi_{i1}(\mathbf{s}) = 1$ for $\mathbf{s} = (s++)$. The survival model was formulated with sex, age and area-specific effects,

$$(3.3) \quad \text{logit } \phi_{ij}(\mathbf{s}) = \mathbf{x}_{ij} \cdot \boldsymbol{\eta}_{\emptyset} + B_{\text{sex}(i),j} + \eta^{(1)}(\mathbf{s}),$$

where $\mathbf{x}_{ij} = (1, w_{ij})$, where w_{ij} is the mass anomaly (difference from sex-specific mean) if $j = 1$, zero otherwise (i.e., only pup survival is influenced by mass), $B_{\text{sex}(i),j}$ is a sex-specific b -spline smooth (df = 3; Hastie, 1992) over age, and $\eta^{(1)}(\mathbf{s})$ is an area effect. Here, mass anomaly was used as a proxy for body condition. Males tend to be heavier on average and we did not want sex differences due to other factors to be hidden within an absolute mass effect. For the state transitions we used the previous state as the reference state for the following occasion, that is, $\mathbf{s}^* = \mathbf{r}$ for transitions from $\mathbf{r} \rightarrow \mathbf{s}$. Thus, the parameters control movement away from the current area (s_1) and loss of flipper tags (s_2, s_3). The state transitions were modeled by

$$(3.4) \quad \psi_{ij}(\mathbf{r}, \mathbf{s}) \propto \exp\{B_{\text{sex}(i),j}^{(1)}(\mathbf{s}; \mathbf{r}) + \mathbf{x}_j \cdot \boldsymbol{\beta}^{(2)}(\mathbf{s}; \mathbf{r}) \\ + \mathbf{x}_j \cdot \boldsymbol{\beta}^{(3)}(\mathbf{s}; \mathbf{r}) + \beta^{(2,3)}(\mathbf{s}; \mathbf{r})\},$$

where $B_{\text{sex}(i),j}^{(1)}(\mathbf{s}; \mathbf{r})$ is a sex- and area-specific b -spline (df = 3) over age, $\mathbf{x}_j = (1, \text{age}_j)$, and $\beta^{(2,3)}(\mathbf{s}; \mathbf{r})$ is an interaction term between the two tags which models dependence in loss. Although not immediately obvious, the b -spline effect, $B_{\text{sex}(i),j}^{(1)}(\mathbf{s}; \mathbf{r})$, can be written in the form $\mathbf{b}_j \cdot \boldsymbol{\beta}_{\text{sex}(i)}^{(1)}(\mathbf{s}; \mathbf{r})$, where \mathbf{b}_j is a vector of the b -spline basis function values for occasion j . Therefore, it fits into the proposed log-linear framework. Also, because the same type of tag was placed on each flipper, we constrained the marginal loss rates to be identical for each side. This is accomplished by setting the tag loss parameters $\boldsymbol{\beta}^{(2)}(-; +) = \boldsymbol{\beta}^{(3)}(-; +)$. The probabilities for impossible transitions [i.e., going from “−” to “+” for either flipper tag variable (s_2, s_3)] were fixed to zero. The tag loss portion of the state transition generally follows the tag loss model proposed by Laake et al. (2014). Note that the structural zeros only affect the $\mathbf{r}^{(2,3)} \rightarrow \mathbf{s}^{(2,3)}$ transitions and there is no interaction term for $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2,3)}$; therefore, we can write $\psi_{ij}(\mathbf{s}; \mathbf{r}) = \psi_{ij}^{(1)}(\mathbf{s}; \mathbf{r})\psi_{ij}^{(2,3)}(\mathbf{s}; \mathbf{r})$. Table A.3 in Supple-

ment A (Johnson et al., 2016a) illustrates the transition probabilities for each current state, \mathbf{r} , to each possible state, \mathbf{s} .

The full model was fitted with maximum likelihood methods using the forward HMM algorithm in Section 2.3. The ability to fit multivariate state models was added to the R package marked (Laake, Johnson and Conn, 2013). R code used for this example is available in Supplement B (Johnson et al., 2016b). Confidence intervals (CI) for derived quantities were approximated using a multivariate normal parametric bootstrap with mean equal to the maximum likelihood estimate and covariance matrix equal to the negative Hessian of the log-likelihood function (see Supplement B; Johnson et al., 2016b). The delta method could be used. However, with complex parameter functions, forming the gradient can be challenging, thus we opted for the ease of the bootstrap.

3.2 Results

The 1996 cohort of sea lion pups on SMI experienced a very strong El Niño in 1997 which both influenced the movement of young animals to ANI and created a high level of mortality before the sea lions reached age 2. With a single cohort and a high level of early mortality, the sample sizes were small at older age classes and the confidence intervals became very wide and that should be considered in evaluating any patterns across age. This unusual event may have also influenced differences in survival across sex and age. However, as an example, it does illustrate the usefulness and flexibility of building multivariate state models.

We begin the results description with the observation model, δ . This is a somewhat unusual example with regard to observation of each state's status. Observers use the brand for resighting and the tag status was often not observed. Observers were more likely to record status of left tag because the brand is on the sea lion's left side (Figure 1). The odds of missing the second side when the first is missed increases 5-fold [3.5–8.2] (95% confidence intervals are contained within square brackets). This again is due to use of brand for resighting. If the observer didn't record the status of the left side, then they were unlikely to get the status of the right side. The fitted detection model, p , is illustrated in Figure A.1 of Supplement A (Johnson et al., 2016a). It is relatively standard with respect to multistate (and CJS) modeling, so we do not elaborate further here.

Male pups had a slightly better overall pup survival (Figure 2). Although confidence intervals over-

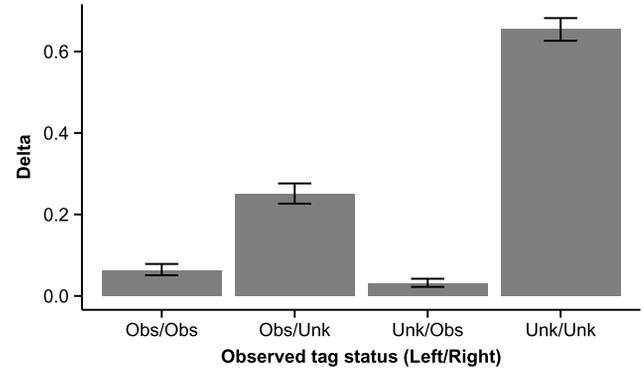


FIG. 1. Estimated δ parameters for the California sea lion model. If the tag status was observed, then it appears as “Obs” in the plot. Thus, for example, the column “Obs/Unk” indicates the probability that the left tag was observed and the right tag was not observed and its status is unknown. The error bars represent 95% CIs.

lapped considerably. For both sexes, pup survival increased with increasing mass, which has also been observed in northern fur seals (Baker and Fowler, 1992; Baker, Fowler and Antonelis, 1994), Steller sea lions (Hastings, Gelatt and King, 2009) and Hawaiian monk seals (Baker, Fowler and Antonelis, 1994; Craig and Ragen, 1999).

For nonpups (age > 0) survival increased to adult age (about 5 years old) and then declined in older ages (Figure 3). The decline was much more dramatic for males whose cost of reproduction is high due to their polygynous breeding system in which breeding males have to defend a territory (Peterson and Bartholomew, 1967; Johnson, 1968). For both sexes, higher survival at ANI was likely due to two factors. First, in El Niño years prey resources are distributed farther north, closer to ANI, thus animals not constrained by repro-

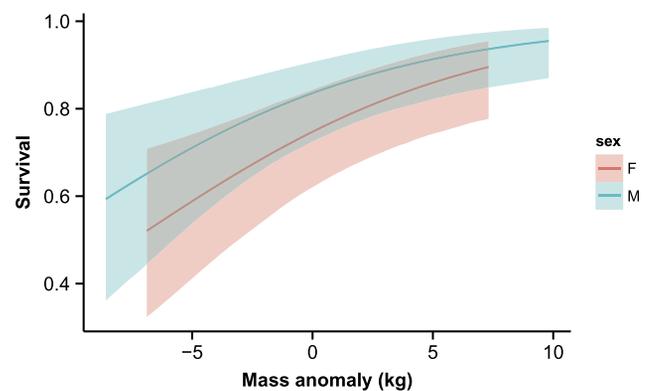


FIG. 2. Mass- and sex-specific survival probability for California sea lion pups on San Miguel Island, California. Mass anomaly is the deviation from the sex-specific mean mass. Colored envelopes represent 95% CIs.

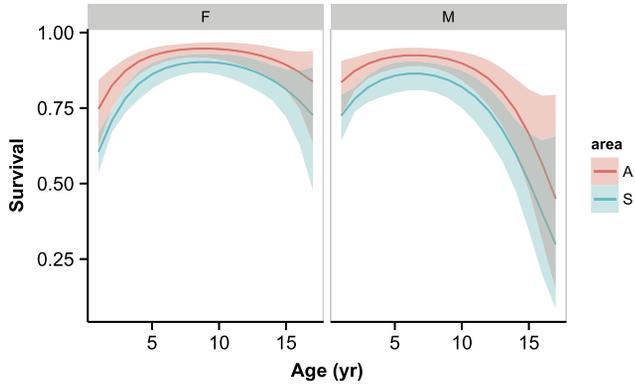


FIG. 3. Sex- and age-specific annual survival probabilities for California sea lion nonpups (age > 0). Separate curves are plotted for each area A = Año Nuevo I. and S = San Miguel I. Although plotted as a solid curve, animals transitioned between the curves depending on the area occupied. Colored envelopes represent 95% CIs.

ductive commitments at SMI had better access to food. Second, animals at ANI during the breeding season were likely nonreproductive (at least for a given year) and, thus, did not have to invest in costly activities such as holding territories (males) or lactation (females), which allowed them to preserve or invest in better body condition.

With respect to movement between SMI and ANI, we could look at the fitted age- and sex-specific annual area transition probabilities (Figure A.2; Supplement A, Johnson et al., 2016a). However, a more biologically interesting, derived quantity is the sex-specific proportion of animals that are located in each area for a given age, π_{ij} . This is because animals at SMI during the breeding season were likely reproductive. This derived quantity can be calculated by

$$(3.5) \quad \hat{\pi}_{ij} = \pi_{i1} \cdot \prod_{j=1}^{J-1} \hat{\Psi}_{ij}^{(1)},$$

where, for the sea lion data, $\pi_{i1} = (1, 0)'$ because all pups began at SMI on occasion 1, and $\hat{\Psi}_{ij}^{(1)}$ is a 2×2 matrix with $\hat{\psi}_{ij}^{(1)}(\mathbf{s}; \mathbf{r})$ entries. Figure 4(A) illustrates the occupancy rate of SMI by both sexes. In general, females began returning to SMI at around age 3 and occupancy stabilized to ≈ 0.7 by age 5. Thus, about 70% of females from the 1996 cohort were reproductively active each year based on location alone. This is similar to the 0.77 average natality estimated by Melin et al. (2011) for several cohorts. A large proportion of males stayed at ANI until the age of 10 or 11 when they began to return and to hold territories. Because a large proportion of males were located at ANI during the breeding

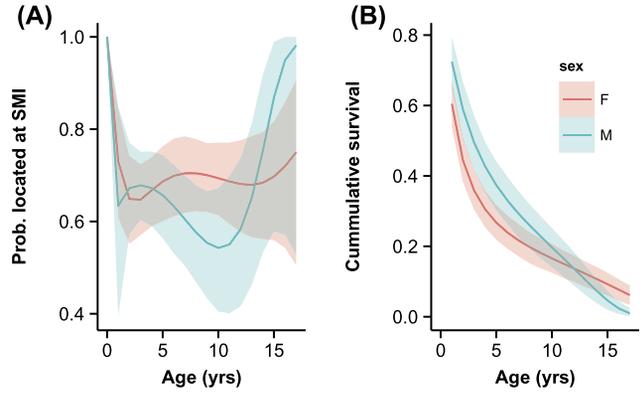


FIG. 4. (A) Age- and sex-specific probability that an individual was located at San Miguel I. during the breeding season. Solid lines and envelopes represent probability that an animal was located at SMI and 95% CIs. (B) The Age- and sex-specific cumulative marginal survival, that is, the expected proportion of animals that were alive at a given age.

season, they experienced better annual survival (for the reasons discussed previously) and, hence, had higher survivorship than females [Figure 4(B)]. By age 12–13, however, the lower survival for breeding males on SMI resulted in equal survivorship.

Finally, although it did not affect the estimation of survival in this case due to the permanent brand, the flipper tag results were quite revealing for pinniped studies where a third mark is not possible. First, tag loss rate increased with age. Marginally, the odds of losing a specific tag (left or right) increased 1.56-fold [1.38–1.75] every year. There was also substantial dependence between tags. The odds of losing a tag increased 406-fold [76–2204] when the other had been lost. Therefore, tags were almost always lost simultaneously. The probability of losing both tags at or before the animal was 5 years old was 0.49 [0.38–0.61]; by age 9, the probability increased to 0.91 [0.86–0.95] (Figure 5; CDF). The most probable age for an animal to become $(-, -)$ tag status was 5–6, when approximately 40% of animals entered double loss status (Figure 5; PMF). This was probably due to the fact that females and males are in periods of strong growth and development at these ages as they prepare for reproductive activities.

4. DISCUSSION

In the ongoing development of state-based CJS modeling we present a generalization of the multievent CJS models to the case where states are multivariate vectors of categorical variables. This allows the user to analyze complex state transitions and handle partial knowledge

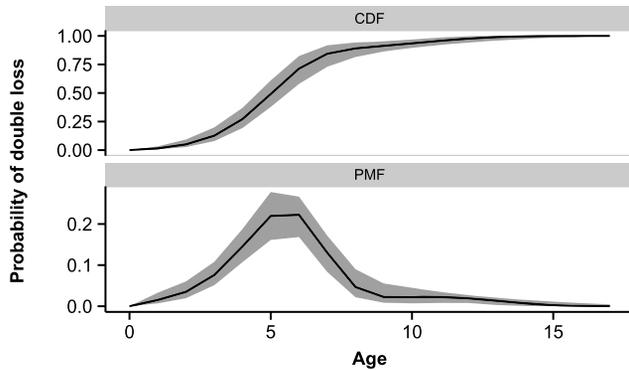


FIG. 5. Age-specific probabilities of double tag loss. The upper plot (CDF) is the cumulative probability distribution of entering the double tag loss state at or before each age. The lower plot (PMF) is the difference of the upper plot and provides the probability mass function that an animal enters the double tag loss state at each age. Gray envelopes represent 95% CIs.

of states upon observation of the individual. Using a log-linear model parameterization for multiway contingency table data allows parsimonious modeling of multivariate state transitions by enforcing desired independence (or dependence) between state variables by exclusion (inclusion) of specified interaction terms. Because of the necessary and sufficient conditions for determining independence from log-linear models, inference on dependence of state variables can also be tested by examining for significant deviation from zero for the appropriate interaction terms. We use the term “testing” in a loose sense, in that one might ascertain those effects through a model selection procedure (Burnham and Anderson, 2002; Hooten and Hobbs, 2015). Because we can efficiently calculate the likelihood for the multivariate state model, all of the usual information criteria and Bayesian methods for model selection are applicable.

The parsimonious construction and hypothesis testing properties were both demonstrated in the example analysis of the sea lion data. We assumed that tag loss was primarily due to growth of the foreflippers from the time the tags were placed on the animals as pups. Therefore, it is reasonable to conclude that loss of tags will be independent of the location of the animal (i.e., tag loss variables s_2 and s_3 are independent of the area s_1). Therefore, we specifically omitted any interaction terms that would imply dependence of tag loss on area. Second, we examined the parameter measuring interaction between the loss of each tag, $\beta^{(2,3)}(\mathbf{s}; \mathbf{r})$, for significant deviation from zero and found that there is a high degree of dependence in the loss of flipper tags

in California sea lions. This conclusion has strong implications for future studies that do not use a permanent third mark such as a brand, as survival estimation may be negatively biased, depending on the assumed tag loss process (Laake et al., 2014). By using the multivariate state framework we were able to directly extend the double tag loss model of Laake et al. (2014) to account for movement between different areas as well. Another practical extension multivariate models provide over the single multievent framework is that the probability of state uncertainty can depend on the true state. For example, although we did not examine it here, it is possible that the probability of an observer missing the status of a flipper tag depends on whether it is present or not. Or, if there are different observer teams in each area, one team may ignore the tag status, while the other dutifully records it. Thus, probability of correct observation of one state variable may depend on another state variable.

Although the proposed multivariate framework is quite general, there are further extensions that are straightforward to develop given the HMM formulation. First, our model was based on the assumption that the state variable is either observed correctly or it is not observed at all and, hence, unknown. There is no reason that this cannot be relaxed. All that needs to be done is to augment the observation space for each variable for which errors can be made. For example, $C_k = \{0, 1, \dots, m, u, e_1, \dots\}$, where e_1, \dots refers to states that are not biological in origin, but observational errors. In the tag loss example, an observer might see a tag is physically present, but fail to read it. In this case it is uncertain whether or not the tag should be considered lost (i.e., unreadable = lost), but there is more information than a simple “unknown” observation. In addition, it is also possible that an observer might declare a tag to be lost when in fact it is present or vice versa. This is accomplished by removing the constraint that $\delta_{ij}(\mathbf{c}; \mathbf{s}) = 0$ when any $c_k \neq s_k$ or u . Finally, continuously valued states could be incorporated using the discrete approximation method of Langrock and King (2013). This creates a large state-space for the continuous variable, thus, careful consideration of the interaction parameterization may be necessary. Each of these three extensions are directly handled by the HMM inference framework. Some of these extensions, however, might produce a model of such complexity that some parameters become unidentifiable or the model fitting is computationally unfeasible. Placement of the multivariate model in a robust design framework similar to Kendall et al. (2012) might improve estimation of parameters, however, careful parameterization may

be the only way to handle large state-spaces. Robust design analysis can be accomplished in the multivariate setting (and in the `marked` package in general) by constraining parameters to be constant (most likely 0 or 1) within secondary sampling periods, but free between primary periods.

In addition to model extensions, there is also a need to develop methods to assess goodness of fit (GOF) for complex CJS models such as those presented here. We do not provide specific recommendations here, as this is a large topic in its own right. There are, however, some avenues worth pursuing. Zucchini and MacDonald (2009) and Titman and Sharples (2008) provide general frameworks for assessing model fit in HMMs. Those general approaches may be adapted for the specifics of the state-based CJS models. In addition, the minimally sufficient statistics GOF testing approach of Pradel, Gimenez and Lebreton (2005) might be investigated using the work of King and McCrea (2014) to derive the necessary sufficient statistics.

We, again, would like to acknowledge the seminal work by the Cormack (1964), Jolly (1965) and Seber (1965) papers which initiated 50 years of development and analysis of capture-recapture data. Through the years, the same basic framework has been generalized to add additional inference capability as well as overcome assumptions of the original CJS models, but the same basic framework exists in each of these models, including the extension proposed herein.

ACKNOWLEDGMENTS

The findings and conclusions in the paper are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service, NOAA. Reference to trade names does not imply endorsement by the National Marine Fisheries Service, NOAA.

SUPPLEMENTARY MATERIAL

Supplement A: Notation summary and additional sea lion analysis details (DOI: [10.1214/15-ST542SUPPA](https://doi.org/10.1214/15-ST542SUPPA); .pdf). Contains additional tables that summarize notation used throughout the paper and provides additional details and results for the analysis of sea lion data in Section 3.

Supplement B: R code used to analyze sea lion data (DOI: [10.1214/15-ST542SUPPB](https://doi.org/10.1214/15-ST542SUPPB); .pdf). Contains the R code used to run the sea lion example analysis in Section 3. The most up-to-date version of the `marked` package can be installed directly using

the R package `devtools` with the command: `devtools::install_github("jlaake/marked/`
`marked")`. However, in order to install this version, users need to ensure that their machines are equipped to compile R packages with source code (namely, Fortran and C++). Using the R console, the command, `install.packages("marked")`, will install a pre-compiled binary version from CRAN.

REFERENCES

- ARNASON, A. N. (1973). The estimation of population size, migration rates and survival in a stratified population. *Researches on Population Ecology* **15** 1–8.
- ARNASON, A. N. and SCHWARZ, C. J. (1999). Using POPAN-5 to analyse banding data. *Bird Study* **46** 157–168.
- BAKER, J. D. and FOWLER, C. W. (1992). Pup weight and survival of northern fur seals *Callorhinus ursinus*. *Journal of Zoology* **227** 231–238.
- BAKER, J. D., FOWLER, C. W. and ANTONELIS, G. A. (1994). Body weight and growth of juvenile male northern fur seals, *Callorhinus ursinus*. *Marine Mammal Science* **10** 151–162.
- BRAND, M., OLIVER, N. and PENTLAND, A. (1997). Coupled hidden Markov models for complex action recognition. In *Computer Vision and Pattern Recognition, 1997 Proceedings, 1997 IEEE Computer Society Conference on* 994–999. IEEE, New York.
- BROWNIE, C., HINES, J. E., NICHOLS, J. D., POLLOCK, K. H. and HESTBECK, J. B. (1993). Capture-recapture studies for multiple strata including non-Markovian transitions. *Biometrics* **49** 1173–1187.
- BURNHAM, K. P. and ANDERSON, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. Springer, New York. [MR1919620](https://doi.org/10.1007/978-1-4939-9726-2)
- CHOQUET, R., ROUAN, L. and PRADEL, R. (2009). Program E-surge: A software application for fitting multievent models. In *Modeling Demographic Processes in Marked Populations* (D. Thomson, E. G. Cooch and M. J. Conroy, eds.). *Environmental and Ecological Statistics* **3** 845–865. Springer, Berlin.
- CHOQUET, R., REBOULET, A. M., PRADEL, R., GIMENEZ, O. and LEBRETON, J. D. (2004). M-SURGE: New software specifically designed for multistate capture-recapture models. *Animal Biodiversity and Conservation* **27** 207–215.
- CHRISTENSEN, R. (1997). *Log-Linear Models and Logistic Regression*, 2nd ed. Springer, New York. [MR1633357](https://doi.org/10.1007/978-1-4939-9726-2)
- CORMACK, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika* **51** 429–438.
- CRAIG, M. P. and RAGEN, T. J. (1999). Body size, survival, and decline of juvenile Hawaiian monk seals, *Monachus schauinslandi*. *Marine Mammal Science* **15** 786–809.
- DARROCH, J. N. (1958). The multiple-recapture census. I. Estimation of a closed population. *Biometrika* **45** 343–359. [MR0119360](https://doi.org/10.1093/biomet/45.3-4.343)
- FRYDENBERG, M. (1990). The chain graph Markov property. *Scand. J. Stat.* **17** 333–353. [MR1096723](https://doi.org/10.2307/2346723)
- GHAHRAMANI, Z. and JORDAN, M. I. (1997). Factorial hidden Markov models. *Mach. Learn.* **29** 245–273.
- GOURLAY-LAROUR, M., PRADEL, R., GUILLEMAIN, M., GUI-TTON, J., L'HOSTIS, M., SANTIN-JANIN, H. and CAIZ-ERGUES, A. (2014). Movement patterns in a partial mi-

- grant: A multi-event capture-recapture approach. *PLoS ONE* **9** e96478.
- HASTIE, T. J. (1992). Generalized additive models. In *Statistical Models in S* (J. M. Chambers and T. J. Hastie, eds.) Chapter 7. Wadsworth & Brooks, Pacific Grove, CA.
- HASTINGS, K. K., GELATT, T. S. and KING, J. C. (2009). Post-branding survival of Steller sea lion pups at Lowrie Island in Southeast Alaska. *The Journal of Wildlife Management* **73** 1040–1051.
- HESTBECK, J. B., NICHOLS, J. D. and MALECKI, R. A. (1991). Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. *Ecology* **72** 523–533.
- HILBORN, R. (1990). Determination of fish movement patterns from tag recoveries using maximum likelihood estimators. *Canadian Journal of Fisheries and Aquatic Sciences* **47** 635–643.
- HOOTEN, M. B. and HOBBS, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs* **85** 3–28.
- JOHNSON, A. M. (1968). Annual mortality of territorial male fur seals and its management significance. *The Journal of Wildlife Management* **32** 94–99.
- JOHNSON, D. S., LAAKE, J. L., MELIN, S. R. and DELONG, R. L. (2016a). Supplement to “Multivariate State Hidden Markov Models for Mark-Recapture Data.” DOI:10.1214/15-STS542SUPPA.
- JOHNSON, D. S., LAAKE, J. L., MELIN, S. R. and DELONG, R. L. (2016b). Supplement to “Multivariate State Hidden Markov Models for Mark-Recapture Data.” DOI:10.1214/15-STS542SUPPB.
- JOLLY, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* **52** 225–247. MR0210227
- KENDALL, W. L., LANGTIMM, C. A., BECK, C. A. and RUNGE, M. C. (2004). Capture-recapture analysis for estimating manatee reproductive rates. *Marine Mammal Science* **20** 424–437.
- KENDALL, W. L., WHITE, G. C., HINES, J. E., LANGTIMM, C. A. and YOSHIZAKI, J. (2012). Estimating parameters of hidden Markov models based on marked individuals: Use of robust design data. *Ecology* **93** 913–920.
- KING, R. and MCCREA, R. S. (2014). A generalised likelihood framework for partially observed capture-recapture-recovery models. *Stat. Methodol.* **17** 30–45. MR3133584
- LAAKE, J. L. (2013). Capture-recapture analysis with hidden Markov models. Technical Report No. AFSC Processed Rep. 2013-04. Seattle, WA.
- LAAKE, J. L., JOHNSON, D. S. and CONN, P. B. (2013). Marked: An R package for maximum likelihood and Markov chain Monte Carlo analysis of capture-recapture data. *Methods in Ecology and Evolution* **4** 885–890.
- LAAKE, J. L., JOHNSON, D. S., DIEFENBACH, D. R. and TERNENT, M. A. (2014). Hidden Markov model for dependent mark loss and survival estimation. *J. Agric. Biol. Environ. Stat.* **19** 524–540. MR3292571
- LANGROCK, R. and KING, R. (2013). Maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. *Ann. Appl. Stat.* **7** 1709–1732. MR3127965
- LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. Clarendon Press, Oxford. MR1419991
- LEBRETON, J. D. and PRADEL, R. (2002). Multistate recapture models: Modelling incomplete individual histories. *J. Appl. Stat.* **29** 353–369. MR1909755
- LEBRETON, J. D., BURNHAM, K. P., CLOBERT, J. and ANDERSON, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs* **62** 67–118.
- MCCCLINTOCK, B. T. (2015). Multimark: An R package for analysis of capture-recapture data consisting of multiple “non-invasive” marks. *Ecology and Evolution* **5** 4920–4931.
- MCCREA, R. S. and MORGAN, B. J. (2014). *Analysis of Capture-Recapture Data*. CRC Press, Boca Raton, FL.
- MELIN, S. R., LAAKE, J. L., DELONG, R. L. and SINIFF, D. B. (2011). Age-specific recruitment and natality of California sea lions at San Miguel Island, California. *Marine Mammal Science* **28** 751–776.
- NICHOLS, J. D. and KENDALL, W. L. (1995). The use of multi-state capture-recapture models to address questions in evolutionary ecology. *J. Appl. Stat.* **22** 835–846.
- NICHOLS, J. D., SAUER, J. R., POLLOCK, K. H. and HESTBECK, J. B. (1992). Estimating transition-probabilities for stage-based population projection matrices using capture recapture data. *Ecology* **73** 306–312.
- PETERSON, R. and BARTHOLOMEW, G. (1967). The natural history and behavior of the California sea lion. Special publication No. 1. The American Society of Mammalogists.
- PRADEL, R. (2005). Multievent: An extension of multistate capture-recapture models to uncertain states. *Biometrics* **61** 442–447. MR2140915
- PRADEL, R., GIMENEZ, O. and LEBRETON, J. D. (2005). Principles and interest of GOF tests for multistate capture-recapture models. *Animal Biodiversity and Conservation* **28** 189–204.
- R DEVELOPMENT CORE TEAM (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- SCHWARZ, C. J., SCHWEIGERT, J. F. and ARNASON, A. N. (1993). Estimating migration rates using tag-recovery data. *Biometrics* **49** 177–193.
- SEBER, G. A. F. (1965). A note on the multiple-recapture census. *Biometrika* **52** 249–259. MR0210228
- TESTA, J. W., THOMASON, J. R., REAM, R. R. and GELATT, T. S. (2013). Demographic studies of northern fur seals on the Pribilof Islands, Alaska, 2007–2012. In *Fur Seal Investigations, 2012* (J. W. Testa, ed.) U.S. Dept. of Commerce, NOAA Tech. Memo. NMFS-AFSC-257, 90 p. Washington, DC.
- TITMAN, A. C. and SHARPLES, L. D. (2008). A general goodness-of-fit test for Markov and hidden Markov models. *Stat. Med.* **27** 2177–2195. MR2439893
- WHITE, G. C. and BURNHAM, K. P. (1999). Program MARK: Survival estimation from populations of marked animals. *Bird Study* **46** 120–139.
- WHITE, G. C., KENDALL, W. L. and BARKER, R. J. (2006). Multistate survival models and their extensions in program MARK. *Journal of Wildlife Management* **70** 1521–1529.
- ZUCCHINI, W. and MACDONALD, I. L. (2009). *Hidden Markov Models for Time Series: An Introduction Using R. Monographs on Statistics and Applied Probability* **110**. CRC Press, Boca Raton, FL. MR2523850