EXTENDING THE LATENT MULTINOMIAL MODEL WITH COMPLEX ERROR PROCESSES AND DYNAMIC MARKOV BASES

By Simon J Bonner[*], Matthew R Schofield[†], Patrik Noren[‡] and Steven J Price[*]

University of Kentucky[*], University of Otago[†], and Institute of Science and Technology Austria[‡]

The latent multinomial model (LMM) of Link et al. (2010) provides a framework for modelling mark-recapture data with potential identification errors. Key is a Markov chain Monte Carlo (MCMC) scheme for sampling configurations of the latent counts of the true capture histories that could have generated the observed data. Assuming a linear map between the observed and latent counts, the MCMC algorithm uses vectors from a basis of the kernel to move between configurations of the latent data. Schofield and Bonner (2015) shows that this is sufficient for some models within the framework but that a larger sets called Markov bases are required with complex types of errors. We address two further challenges: 1) that models with complex error mechanisms may not fit within the LMM framework and 2) that Markov bases can be difficult to compute for studies of even moderate size. We extend the framework to model the capture/demographic and error processes separately and develop a new MCMC algorithm using dynamic Markov bases. Our work is motivated by a study of Queen snakes (*Regina septemvittata*) and we use simulation to compare estimates of survival rates when snakes are marked with PIT tags which have perfect identification versus brands which are prone to error.

1. Introduction. Standard models for data from studies of marked individuals require that researchers are able to identify captured individuals uniquely and without error. However, these assumptions may be violated in many ways. Researchers may misread marks and provide partial identifications based on visual sightings or poor quality photographs (McClintock et al., 2014; Morrison et al., 2011), allelic dropout may lead to incorrect identifications from DNA samples (Barker et al., 2014; Lukacs and Burnham, 2005; Wright et al., 2009; Yoshizaki et al., 2011), man-made tags may be lost or degrade (Cowen and Schwarz, 2006), and natural marks may evolve over time (Yoshizaki et al., 2012). This paper continues our investigation of the application of methods from algebraic statistics to models allowing for possible identification errors in mark-recapture type data.

Our work is based on the latent multinomial model (LMM) first intro-
duced by Link et al. (2010). Key to this approach was a novel Markov chain
Monte Carlo (MCMC) algorithm for sampling configurations of the true
captures consistent with the observed data. In Schofield and Bonner (2015),
we showed that the original algorithm proposed by Link et al. (2010) may
produce Markov chains that are not irreducible and hence fail to sample
from the set of all configurations with positive probability under the poste-
rior distribution. We described an extended MCMC algorithm incorporating
Markov bases to address this issue. As one example, we presented simula-
tion results from a model which we called the band-read error (BRE) model.
Our original manuscript did not provide details on the implementation of
the BRE model and, for simplicity, we ignored population demographics as-
suming instead that all configurations of the latent counts consistent with
the observed data were equally likely. In fact, the BRE model cannot easily
be fit with the original LMM. Here we describe an extension of the LMM
framework that is needed to fit the BRE model and other models with more
complex types of errors than those considered by Link et al. (2010). We also
address a second, practical challenge which is that Markov bases for the BRE
model can only be computed for very small experiments with four capture
occasions or less. Although we focus on the BRE model, the methods we
develop are applicable to all models within the LMM framework including
model $M_{t\alpha}$ and the multiple mark models of Bonner and Holmberg (2013);
McClintock et al. (2013).

2. The Latent Multinomial Model. The LMM of Link et al. (2010)
accounts for possible errors in the data by recasting the mark-recapture
model as a missing data problem. Suppose that individuals are sampled from
the population on T capture occasions. On each occasion, the individuals
are identified, marked if necessary, and returned to the population. Let n_{Tot}
be the number of distinct individuals captured. The raw data consist of n_{Tot}
vectors in $\{0, 1\}^T$, called the capture histories, such that $\omega_{it} = 1$ if individual
i was captured on occasion j. If $T = 5$ then the history $\boldsymbol{\omega}_i = 01010$ indicates
that the ith marked individual was captured on occasions 2 and 4. We use I to
denote the number of unique capture histories that can be observed during
the experiment. If the population is homogeneous then the probabilities
assigned to each capture history are the same for all individuals and the
I-vector of counts, $\boldsymbol{n}$, recording the number of times each unique history
was observed is sufficient. The likelihood is then defined by the distribution
of $\boldsymbol{n}$.

When errors occur the distribution of $\boldsymbol{n}$ may be difficult to compute

⁷⁹ directly. To make the likelihood tractable, the LMM introduces a set of
⁸⁰ $J > I$ latent histories identifying the true captures for each individual and
⁸¹ describing what errors occurred. Let $\boldsymbol{x}$ be the unobserved J-vector of counts
⁸² for the latent histories which is modelled by $f(\boldsymbol{x}|\boldsymbol{\theta})$[1] for some vector of
⁸³ parameters, $\boldsymbol{\theta}$. The likelihood can then be computed by summing $f(\boldsymbol{x}|\boldsymbol{\theta})$
⁸⁴ over all values of $\boldsymbol{x}$ consistent with $\boldsymbol{n}$. In particular, the LMM assumes that
⁸⁵ $\boldsymbol{n}$ is a linear function of $\boldsymbol{x}$ so that $\boldsymbol{n} = \boldsymbol{A}\boldsymbol{x}$ for some known $I \times J$ matrix $\boldsymbol{A}$
⁸⁶ and

$$(1) \qquad f(\boldsymbol{n}|\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathbb{N}^J} \mathbb{1}(\boldsymbol{n} = \boldsymbol{A}\boldsymbol{x})f(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{n}}} f(\boldsymbol{x}|\boldsymbol{\theta})$$

⁸⁷ where $\mathcal{F}_{\boldsymbol{n}} = \{\boldsymbol{x} \in \mathbb{N}^J : \boldsymbol{n} = \boldsymbol{A}\boldsymbol{x}\}$ denotes the inverse image of $\boldsymbol{n}$ (called the
⁸⁸ $\boldsymbol{n}$-fibre in algebraic statistics) and $\mathbb{N} = \{0, 1, 2, \ldots\}$.

⁸⁹ As an example, Link et al. (2010) considered a closed population model
⁹⁰ which they called model $M_{t\alpha}$. This model assumes that that all individuals
⁹¹ have the same probability of capture on occasion t, denoted by p_t, that
⁹² errors occur independently with probability $1 - \alpha$ each time an individual is
⁹³ captured, and that these errors result in new identities that are not observed
⁹⁴ otherwise. The latent histories for this model include the 3^T strings formed
⁹⁵ by the events:

⁹⁶ 0 – indicating that the individual was not captured,

⁹⁷ 1 – indicating that the individual was captured and correctly identified, and

⁹⁸ 2 – indicating that the individual was captured and incorrectly identified.

⁹⁹ For example, the latent history $\boldsymbol{\nu}_i = 01020$ indicates that individual i was
¹⁰⁰ captured and correctly identified on occasion 2 and recaptured and misiden-
¹⁰¹ tified on occasion 4. It is assumed that $\boldsymbol{x}|\boldsymbol{p}, \alpha$ follows a multinomial dis-
¹⁰² tribution conditional on the true population size, N, with cell probability
¹⁰³ corresponding to the latent history $\boldsymbol{\nu}$ computed as

$$f(\boldsymbol{\nu}|\boldsymbol{p}, \alpha) = \prod_{t=1}^{T} \left[p_t^{\mathbb{1}(\nu_t > 0)}(1 - p_t)^{\mathbb{1}(\nu_t = 0)}\alpha^{\mathbb{1}(\nu_t = 1)}(1 - \alpha)^{\mathbb{1}(\nu_t = 2)} \right]$$

¹⁰⁴ where $\mathbb{1}(\cdot)$ is the indicator function.

¹⁰⁵ Although equation (1) makes it easier to compute $f(\boldsymbol{n}|\boldsymbol{\theta})$ in theory, $\mathcal{F}_{\boldsymbol{n}}$
¹⁰⁶ is often so large that exact computation is not practical. Instead, Link et al.
¹⁰⁷ (2010) proposed a Bayesian solution implemented by sampling from the
¹⁰⁸ joint posterior distribution of $\boldsymbol{x}$ and $\boldsymbol{\theta}$. The specific MCMC algorithm uses

---

[1]Throughout we use $f(\boldsymbol{y}|\boldsymbol{\beta})$ to denote the assumed model of $\boldsymbol{y}$ given parameters $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta})$ the prior distribution for $\boldsymbol{\beta}$, and $\pi(\boldsymbol{\beta}|\boldsymbol{y})$ the posterior distribution of $\boldsymbol{\beta}$ given $\boldsymbol{y}$.

109 a block Metropolis-Hastings (MH) approach and the main challenge lies in
110 constructing proposals of $\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{n}$ which fall inside the fibre. The algorithm
111 starts by defining a lattice basis for the kernel of $\boldsymbol{A}$; that is, a linearly
112 independent set $\mathcal{B} = \{b_1, \ldots, b_K\}$ such that $\forall \boldsymbol{b} \in \ker(\boldsymbol{A}) \bigcap \mathbb{Z}^J$ there exists
113 $c_1, \ldots, c_K \in \mathbb{Z}$ so that

$$\boldsymbol{b} = \sum_{k=1}^{K} c_k \boldsymbol{b}_k.$$

114 A proposal, $\boldsymbol{x}^{\mathrm{prop}}$, is then constructed by adding an integer multiple of one
115 of the basis vectors to the current value of $\boldsymbol{x}$. The magic of this approach is
116 that any proposal is guaranteed to satisfy the linear constraint and to have
117 integer entries. Note, that $\boldsymbol{x}^{\mathrm{prop}}$ may still fall outside $\mathcal{F}_{\boldsymbol{n}}$ since there is no
118 guarantee that $x_j^{\mathrm{prop}} \geq 0$ for every j.
119   Link et al. (2010) implied that Markov chains constructed with this algo-
120 rithm would connect all elements in $\mathcal{F}_{\boldsymbol{n}}$ and hence be irreducible. In Schofield
121 and Bonner (2015), we showed that this is true for model $M_{t\alpha}$ provided that
122 the right lattice basis is chosen and extended this result to a broader class
123 of models containing what we called simple corruptions. However, we also
124 provided examples of more complicated models for which the algorithm does
125 not produce irreducible Markov chains, including the BRE model. The cen-
126 tral problem is that some pairs of elements in $\mathcal{F}_{\boldsymbol{n}}$ may be connected by this
127 algorithm only by passing through intermediate configurations containing
128 negative entries. These elements have zero probability under the posterior,
129 and so the chain will never follow these paths.
130   Irreducible chains can always be produced by adding linear combinations
131 of all elements in $\mathcal{B}$ simultaneously, but the resulting proposals are likely
132 to fall outside of $\mathcal{F}_{\boldsymbol{n}}$ and Diaconis and Sturmfels (1998) reported that this
133 method is not efficient. Instead, Diaconis and Sturmfels (1998) suggested
134 using the one-at-a-time algorithm but drawing the elements from a larger
135 subset $\mathcal{M} \subset \ker(\boldsymbol{A})$ chosen to ensure that it is possible to move between any
136 two elements of $\mathcal{F}_{\boldsymbol{n}}$. Diaconis and Sturmfels (1998) called $\mathcal{M}$ a Markov basis
137 and the elements of $\mathcal{M}$ moves, and provided methodology for computing this
138 set based on the theory of toric ideals. For simplicity, we consider the special
139 case of the algorithm presented in Schofield and Bonner (2015) in which
140 one element is selected from $\mathcal{M}$ on each iteration of the MCMC algorithm
141 and either added to or subtracted from the current configuration without a
142 multiplier. Details are provided in Algorithm 1.

143   **3. Data.** As an example, we consider data from a study of queen snakes
144 conducted in Jessamine County, Kentucky. An initial sample of 61 snakes
145 was captured and marked in the fall of 2013 and a second sample of 41 snakes

Define a Markov basis, $\mathcal{M}$.
Initialise $\boldsymbol{\theta}^{(0)}$ and $\boldsymbol{x}^{(0)}$ so that $\boldsymbol{n} = \boldsymbol{A}\boldsymbol{x}^{(0)}$.
Set $k = 1$.

1. Update $\boldsymbol{\theta}$ conditional on $\boldsymbol{x}^{(k-1)}$. Call the result $\boldsymbol{\theta}^{(k)}$.

2. Update $\boldsymbol{x}$ conditional on $\boldsymbol{\theta}^{(k)}$.

    i) Sample $\boldsymbol{b}$ uniformly from $\mathcal{M}$ and c uniformly from $\{-1, 1\}$.

    ii) Set $\boldsymbol{x}^{\text{prop}} = \boldsymbol{x}^{(k-1)} + c\boldsymbol{b}$.

    iii) Calculate the Metropolis acceptance probability:

$$r(\boldsymbol{x}^{(k-1)}, \boldsymbol{x}^{\text{prop}} | \theta^{(k)}) = \min\left\{ 1, \frac{\pi(\boldsymbol{x}^{\text{prop}} | \boldsymbol{\theta}^{(k)})}{\pi(\boldsymbol{x}^{(k-1)} | \boldsymbol{\theta}^{(k)})} \cdot \frac{q(\boldsymbol{x}^{(k-1)} | \boldsymbol{x}^{\text{prop}})}{q(\boldsymbol{x}^{\text{prop}} | \boldsymbol{x}^{(k-1)})} \right\}.$$

    where $q(\boldsymbol{x}'|\boldsymbol{x})$ is the probability of proposing $\boldsymbol{x}'$ given the current state $\boldsymbol{x}$.

    iv) Set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{\text{prop}}$ with probability r. Otherwise, set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)}$.

3. Increment k.

Algorithm 1: MCMC algorithm for sampling from the joint posterior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{x}$ given a fixed Markov basis, $\mathcal{M}$.

was marked in the spring of 2014. All snakes were implanted with PIT tags and a subset of 73 snakes were also branded with unique marks as described in Winne et al. (2006). In the summer of 2014, two technicians visited the site to locate and identify snakes approximately every two weeks. On each visit the technicians conducted searches using a PIT receiver and attempted to physically capture any snakes that were detected so that their brands could be read. The 102 snakes were re-encountered 191 times in total, an average of 1.87 per snake. The study is aimed primarily at modelling the survival and movements of the snakes in this population and in understanding impacts of snake fungal disease, an emerging pathogen about which little is yet known (Allender et al., 2013; Sleeman, 2013). For illustration, we focus on modelling the apparent over-wintering survival, the probability that a snake marked in the fall of 2013 is still in the population in 2014.

Previous studies have found that snakes may expel PIT tags (e.g. Roark and Dorcas, 2000) and some loose tags were found at the study site. However, we believe that the rate of expulsion is small and there is no reason to think that PIT tags are ever misidentified. With these assumptions capture histories formed using the PIT tag encounters can be modelled with standard Cormack-Jolly-Seber (CJS) type models ignoring identification errors and tag loss (see Lebreton et al., 1992; Seber, 2002; Williams, Nichols and Conroy, 2002, and references therein). An introduction to the CJS model is provided in Appendix A. We use the common notation p_t and ϕ_t to denote

168 the capture and survival probabilities.

169   In comparison, brands can be difficult to read and the identification of
170 physically captured snakes is prone to error. A total of 9 branded snakes
171 were recaptured physically during the summer of 2014. By comparing with
172 the PIT tag records we knew that the first technician identified 8 of 9 (89%)
173 correctly while the second technician identified only 6 of 9 (67%) correctly.
174 The small number of physical recaptures did not allow us to compare results
175 based on the PIT tag and brand data directly. Instead, we examine the
176 feasibility of branding snakes by analysing simulated data generated with
177 survival and capture probabilities obtained from the PIT tag data and error
178 rates matching those observed from the two technicians.

179   **4. Model.** The specific model we consider both for generating and
180 analysing the simulated data combines the standard CJS model for the de-
181 mographic and capture processes and the BRE model of the errors. We call
182 this combined model the CJS/BRE model. Suppose that researchers visit a
183 location on T occasions. On each visit they capture a number of unmarked
184 individuals, mark them, and return them to the population. At the same
185 time, the researchers also conduct surveys to identify previously marked
186 individuals. The assumptions of the BRE model are that:

187   1. all individuals are correctly identified when first captured and marked,
188   2. recaptured individuals are correctly identified with probability α on
189   each occasion,
190   3. errors cause one marked individual to be misidentified as another
191   marked individual, and
192   4. each individual can be involved in only one event on each occasion. In
193   particular, it is not possible to mistake individual j for individual i if
194   individual i has been captured on the same occasion.

195 Assumption 3 contrasts directly with the assumptions of model $M_{t\alpha}$ and
196 is justified by the differences between man-made marks and natural marks.
197 Model $M_{t\alpha}$ is intended for use with natural marks including genotypes and
198 pigmentation patterns. The set of possible natural marks is usually unknown
199 and the number of possible marks is so large that it is unlikely for an er-
200 ror to reproduce the identity of another individual exactly. On the other
201 hand, the BRE model is intended for use with man-made marks. The set
202 of possible marks is known when using man-made marks, and this means
203 that erroneous sightings of marks which have never been released can be
204 detected and removed from the data prior to the analysis. The only errors
205 that cannot be detected occur when one marked individual is mistaken for
206 another marked individual. The fourth assumption is not realistic, but sim-

207 plifies the model and we will work to relax this in future work. We present
208 the likelihood for this model in Section 5.1 after introducing the extended
209 modelling framework.

210 5. Methods.

211 5.1. *Extended Framework.* The first challenge is that the CJS/BRE model
212 does not fit easily in the framework of the LMM. Link et al. (2010) focused on
213 models, like $M_{t\alpha}$, for which $\boldsymbol{x}$ follows a multinomial distribution. Although
214 they suggested that the methods could be applied more generally examples
215 were not provided. The CJS/BRE model does not result in a multinomial
216 distribution for $\boldsymbol{x}$, and it is difficult to determine the density of $\boldsymbol{x}$ explicitly.
217 To address this, we extend the LMM to include a second vector of latent
218 counts. This allows the mark-recapture process and the error mechanism
219 to be modelled separately. Suppose, for example, that an experiment has
220 $T = 2$ occasions and individual i is captured on both occasions, correctly
221 identified on the first occasion, and identified as an entirely new individual
222 on the second occasion (this is the error mechanism for model $M_{t\alpha}$). In the
223 terminology of Link et al. (2010), individual i would have latent history
224 $\boldsymbol{\nu}_i = 12$ and would produce the recorded histories $\boldsymbol{\omega}_{i1} = 10$ and $\boldsymbol{\omega}_{i2} = 01$.
225 The original LMM assigns probabilities to the latent histories, $\boldsymbol{\nu}_i$, directly by
226 simultaneously modelling the capture and error processes. Our formulation
227 introduces a second latent history, $\boldsymbol{\xi}_i$, identifying the occasions on which
228 the individual was truly captured but ignoring the errors. The new latent
229 history would be $\boldsymbol{\xi}_i = 11$ since the individual was truly captured on both
230 occasions. We then model the joint distribution of $\boldsymbol{\nu}_i$ and $\boldsymbol{\xi}_i$ by assigning
231 probabilities first to $\boldsymbol{\xi}_i$ and second to $\boldsymbol{\nu}_i$ given $\boldsymbol{\xi}_i$. We distinguish between
232 the two sets of latent histories by calling $\boldsymbol{\nu}_i$ the latent error history and $\boldsymbol{\xi}_i$
233 the latent capture history.

234 Generally, we let $\boldsymbol{n}$ be the I-vector of counts for the observable histories,
235 $\boldsymbol{x}$ the J-vector of counts for the latent error histories, and $\boldsymbol{z}$ the K-vector of
236 counts for the latent capture histories. As in Link et al. (2010), we assume
237 that $\boldsymbol{n} = \boldsymbol{A}\boldsymbol{x}$ for some known matrix $\boldsymbol{A}$. Further, we assume that $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}$ for
238 some known matrix $\boldsymbol{B}$. The complete data likelihood is then constructed in
239 two stages: 1) modelling the process of capturing, marking, and recapturing
240 individuals to define $f(\boldsymbol{z}|\boldsymbol{\theta})$ and 2) modelling the error process conditional
241 on the true captures to define $f(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta})$. We expect the parameters in the
242 two components to be disjoint and label them as $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. The posterior
243 distribution of the complete data and parameters is

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\boldsymbol{n}) \propto \mathbb{1}(\boldsymbol{n} = \boldsymbol{A}\boldsymbol{x})f(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta}_2)f(\boldsymbol{z}|\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2)$$

where $\pi(\boldsymbol{\theta}_1)$ and $\pi(\boldsymbol{\theta}_2)$ represent priors assumed to be independent. For convenience we identify entries in the vectors of counts, $\boldsymbol{n}$, $\boldsymbol{x}$, and $\boldsymbol{z}$, both by index and by the corresponding history. For example, n_i represents the count for the ith element of $\boldsymbol{n}$ using some implicit ordering while $n_{\boldsymbol{\omega}}$ represents the count of history $\boldsymbol{\omega}$. This allows us to define sums in two equivalent ways as either $n_{\text{Tot}} = \sum_{i=1}^{I} n_i$ or $n_{\text{Tot}} = \sum_{\boldsymbol{\omega} \in \{0,1\}^T/\boldsymbol{0}} n_{\boldsymbol{\omega}}$. A table summarising our notation for the extended LMM is provided in Appendix B.

To fit the CJS/BRE into the extended framework we need to 1) identify the sets of observable histories, latent error histories, and latent capture histories, 2) construct the constraint matrices, and 3) define the components of the likelihood function. As with the CJS model, the set of observable histories includes the $I = 2^T - 2$ in $\{0,1\}^T$ excluding the zero history and the history ending with a single capture. The latent capture histories also belong to the same set so that $K = 2^T - 2$ as well. In defining the latent error histories, four events can occur on each occasion after an individual is marked. The ith individual may be not resighted (event 0), resighted and correctly identified (event 1), or resighted and incorrectly identified (event 2). Finally, another marked individual may be captured and incorrectly identified as individual i (event 3). Events 2 and 3 represent false negative and false positive resightings. A total of 4^T possible histories can be constructed from these events but many of these can be ignored in the likelihood. We first remove the zero history and the three histories with a single non-zero event on the final occasion, since these histories do not contribute to the likelihood of the CJS model. Assuming that individuals are correctly identified when first captured we can also ignore any history whose leading non-zero entry is not 1. This leaves $J = (4^T - 4)/3$ that contribute to the likelihood.

Next, we construct the constraint matrices. One factor that makes the CJS/BRE model more complicated than model $M_{t\alpha}$ is that it contains constraints on $\boldsymbol{x}$ beyond those imposed by the observed counts. In particular, the number of false positive and false negative captures on occasion t must be the same for all $t = 2, \ldots, T$. The $\boldsymbol{A}$ matrix is constructed as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{bmatrix}$$

where $\boldsymbol{A}_1$ is a $(2^T - 2) \times J$ matrix modelling the relationship between $\boldsymbol{x}$ and $\boldsymbol{n}$ that is defined similar to the matrix $\boldsymbol{A}'$ in Link et al. (2010), and $\boldsymbol{A}_2$ is a $(T-1) \times J$ matrix constraining the number of false positives and negatives on the final $T-1$ occasions. Mathematically,

$$A_{1ij} = \begin{cases} 1 & \text{if } \omega_{it} = \mathbb{1}(\nu_{jt} = 1) + \mathbb{1}(\nu_{jt} = 3) \text{ for all } t = 1, \ldots, T \\ 0 & \text{otherwise} \end{cases}$$

279 and

$$
A_{2tj} = \begin{cases} -1 & \text{if } \nu_{j,t+1} = 2 \\ 1 & \text{if } \nu_{j,t+1} = 3 \\ 0 & \text{otherwise} \end{cases}
$$

280 The tth row of $\boldsymbol{A}_2$ computes the difference between the number of 2s and
281 3s in the latent error histories, and the vector $\boldsymbol{n}$ must also be extended by
282 concatenating $T-1$ extra 0s corresponding to the added constraints. The
283 matrix $\boldsymbol{B}$ is defined such that $B_{jk} = 1$ if the jth latent capture history has
284 the same pattern of captures as the kth latent error history. That is

$$
B_{jk} = \begin{cases} 1 & \text{if } \xi_{kt} = \mathbb{1}(\nu_{jt} = 1) + \mathbb{1}(\nu_{jt} = 2) \text{ for all } t = 1, \ldots, T \\ 0 & \text{otherwise} \end{cases} .
$$

285   Finally, we define the distributions of $\boldsymbol{z}$ and $\boldsymbol{x}|\boldsymbol{z}$. For the CJS/BRE model
286 $\boldsymbol{\theta}_1 = \{\boldsymbol{\phi}, \boldsymbol{p}\}$ and $\boldsymbol{\theta}_2 = \{\alpha\}$. Let a_t denote the number of individuals first cap-
287 tured and marked on occasion t, M_t the number of individuals marked before
288 occasion t, and m_t the number of these individuals resighted on occasion t.
289 Then $\boldsymbol{z}$ is product multinomial with density

$$
(2) \qquad f(\boldsymbol{z}|\boldsymbol{\phi}, \boldsymbol{p}) = \frac{\prod_{t=1}^{T-1} a_t!}{\prod_{k=1}^{K} \boldsymbol{z}_k!} \prod_{k=1}^{K} f(\boldsymbol{\xi}_k|\boldsymbol{\phi}, \boldsymbol{p})^{z_k}
$$

290 where $f(\boldsymbol{\xi}_k|\boldsymbol{\phi}, \boldsymbol{p})$ denotes the probability assigned to history $\boldsymbol{\xi}_k$ by the stan-
291 dard CJS model. To construct the second component of the likelihood we
292 consider occasions $t = 2, \ldots, T$ separately first modelling the number of er-
293 rors that occur, $e_t(\boldsymbol{x}) = \sum_{j=1}^{J} x_j \mathbb{1}(\nu_{jt} = 2) = \sum_{j=1}^{J} x_j \mathbb{1}(\nu_{jt} = 3)$, and then
294 modelling the exact configuration of false positives and false negatives given
295 $\boldsymbol{e} = (e_1, \ldots, e_T)'$ to obtain a specific configuration of $\boldsymbol{x}$. Under the assump-
296 tions in Section 4, $e_t \leq m_t^* = \min(m_t, M_t - m_t)$ and follows the (possibly)
297 truncated binomial with density

$$
f(e_t|\boldsymbol{z}, \alpha) \propto \binom{m_t}{e_t}(1-\alpha)^{e_t}\alpha^{m_t - e_t}, \quad e_t = 0, \ldots, m_t^*.
$$

298 We further assume that all assignments of false positives and false negatives
299 are equally likely conditional on e_t. For each $t = 2, \ldots, T$ there are $\binom{m_t}{e_t}$ and
300 $\binom{M_t - m_t}{e_t}$ ways to select the false negatives and false positives and so

$$
f(\boldsymbol{x}|\boldsymbol{e}, \boldsymbol{z}) = \mathbb{1}(\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}) \prod_{t=2}^{T} \left[\binom{m_t}{e_t} \binom{M_t - m_t}{e_t} \right]^{-1}.
$$

301 The second component of the likelihood is

$$f(\boldsymbol{x}|\boldsymbol{z}, \alpha) = \mathbb{1}(\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}) \frac{\prod_{k=1}^{K} \boldsymbol{z}_k!}{\prod_{j=1}^{J} \boldsymbol{x}_j!} \prod_{t=2}^{T} \left[\frac{(1-\alpha)^{e_t} \alpha^{m_t - e_t}}{\binom{M_t - m_t}{e_t} \sum_{e_t=0}^{m_t^*} \binom{m_t}{e_t} (1-\alpha)^{e_t} \alpha^{m_t - e_t}} \right]$$

302 where the initial term accounts for the many relabellings of the marked
303 individuals that would produce the same counts in $\boldsymbol{x}$ and $\boldsymbol{z}$.

304   The joint posterior distribution is completed by specifying a prior distri-
305 bution. Link et al. (2010) noted that the observed histories contain almost
306 no information about the error rate and so assigned α a very informative
307 prior. Alternatively, information about α could be obtained from double
308 observers or double tags, as in the queen snake study. For convenience,
309 we simply fix α to the known value in our analysis of the simulated data
310 and remove α from the posterior distribution. Finally, we assume that the
311 prior chosen for $\boldsymbol{\phi}$ and $\boldsymbol{p}$ is positive over the entire unit hypercube so that
312 $\pi(\boldsymbol{\phi}, \boldsymbol{p}) > 0$ if and only if $\boldsymbol{\phi} \in (0,1)^{T-1}$ and $\boldsymbol{p} \in (0,1)^{T-1}$. In particular,
313 we assume independent, uniform priors in our simulation study such that
314 $\boldsymbol{\pi}(\boldsymbol{\phi}, \boldsymbol{p}) \propto \mathbb{1}\left(\boldsymbol{\phi} \in (0,1)^{T-1}\right) \cdot \mathbb{1}\left(\boldsymbol{p} \in (0,1)^{T-1}\right)$.

315   5.2. *Dynamic Markov Bases.*   The second challenge in fitting the CJS/BRE
316 model is that the Markov basis grows very quickly with the number of oc-
317 casions and could only be computed for small values of T with 4ti2, a com-
318 monly used free software package for algebraic statistics (Hemmecke et al.,
319 2013). Addition of the second vector of latent counts does not complicate
320 matters because $\boldsymbol{z}$ is a deterministic function of $\boldsymbol{x}$. Consistent proposals for
321 $\boldsymbol{x}$ and $\boldsymbol{z}$ could, in theory, be constructed by defining a Markov basis, $\mathcal{M}$,
322 sampling a move $\boldsymbol{b} \in \mathcal{M}$, and setting $\boldsymbol{x}^{\text{prop}} = \boldsymbol{x} + \boldsymbol{b}$ and $\boldsymbol{z}^{\text{prop}} = \boldsymbol{B}\boldsymbol{x}^{\text{prop}}$.
323 However, 4ti2 ran out of memory on a computer with 8 GB of RAM before
324 completing the calculations when $T \geq 5$.

325   We avoid this problem by using dynamic Markov bases. Dobra (2012) de-
326 fined a dynamic Markov basis to be a collection of sets of local moves, $\mathcal{M}(\boldsymbol{x})$,
327 which connect each $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{n}}$ to a relatively small number of neighbours. A
328 proposal is generated on the kth iteration of the MCMC algorithm by sam-
329 pling a move from $\mathcal{M}(\boldsymbol{x}^{(k-1)})$. This avoids the need to compute the entire
330 Markov basis *a priori*. The method Dobra (2012) described for constructing
331 $\mathcal{M}(\boldsymbol{x})$ applies generally to all models with constraints of the form $\boldsymbol{n} = \boldsymbol{A}\boldsymbol{x}$
332 and, possibly, further bounds on the counts in individual cells. In short, local
333 moves are generated by permuting the counts in a random selection of "free"
334 cells – entries of $\boldsymbol{x}$ whose values are not fixed by the information in $\boldsymbol{n}$ and
335 the cell bounds. The method we describe is specific to the mark-recapture
336 framework but produces a dynamic Markov basis containing intuitive moves

that are easily sampled. Further comparison with Dobra (2012) is provided in Section 8.

The dynamic Markov basis we propose allows the chain to move through $\mathcal{F}_{\boldsymbol{n}}$ using operations that either add or remove errors from the current configuration. Each of these operations modifies four entries in the vector of counts for the latent error histories, increasing two counts and decreasing two counts. Errors are added by decreasing the counts for a pair of histories with a 0 and 1 on some occasion, t, and increasing the counts of the corresponding histories formed by changing the 0 in the first history to a 2 and the 1 in the second history to a 3. Errors are removed using the opposite operation.

To make this explicit, define

$$\mathcal{X}_{vt}(\boldsymbol{x}) = \{\boldsymbol{\nu} : \nu_s = 1 \text{ for some } s < t, \nu_t = v, \text{ and } x_{\boldsymbol{\nu}} > 0\}$$

to be the set of latent error histories with an initial capture before occasion t, event v on occasion t, and positive entry in $\boldsymbol{x}$. Moves in $\mathcal{M}(\boldsymbol{x})$ modify the counts for one history drawn from each of the sets $\mathcal{X}_{0t}(\boldsymbol{x})$, $\mathcal{X}_{1t}(\boldsymbol{x})$, $\mathcal{X}_{2t}(\boldsymbol{x})$, and $\mathcal{X}_{3t}(\boldsymbol{x})$, for some common t and are divided into two classes: $\mathcal{M}_1(\boldsymbol{x})$ containing the moves that add errors and $\mathcal{M}_2(\boldsymbol{x})$ containing the moves that remove errors. A key advantage of this approach is that moves in $\mathcal{M}_1(\boldsymbol{x})$ and $\mathcal{M}_2(\boldsymbol{x})$ can be sampled without ever having to construct the entire sets. Moves in $\mathcal{M}_1(\boldsymbol{x})$ are denoted by $\boldsymbol{b}^+(\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3)$ and are generated by sequentially sampling

1. $\boldsymbol{\nu}_1 \in \chi_{1\cdot}(\boldsymbol{x}) = \bigcup_{t=2}^{T} \mathcal{X}_{1t}(\boldsymbol{x})$
2. $s \in \{t : \nu_{1t} = 1\}$
3. $\boldsymbol{\nu}_0 \in \mathcal{X}_{0s}(\boldsymbol{x})$,

all uniformly, and setting

$$\nu_{2t} = \left\{ \begin{array}{ll} 2 & \text{if } t = s \\ \nu_{0t} & \text{otherwise} \end{array} \right. \text{ and } \nu_{3t} = \left\{ \begin{array}{ll} 3 & \text{if } t = s \\ \nu_{1t} & \text{otherwise} \end{array} \right. .$$

More compactly, $\boldsymbol{\nu}_2 = \boldsymbol{\nu}_0 + 2\boldsymbol{\delta}_t$ and $\boldsymbol{\nu}_3 = \boldsymbol{\nu}_1 + 2\boldsymbol{\delta}_t$ where $\boldsymbol{\delta}_t$ represents the J-vector with a single 1 in entry t. Indexing by name, as described in Section 5.1, the corresponding move has entries

$$(3) \qquad b_{\boldsymbol{\nu}}^+(\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3) = \left\{ \begin{array}{ll} -1 & \text{if } \boldsymbol{\nu} = \boldsymbol{\nu}_0 \text{ or } \boldsymbol{\nu} = \boldsymbol{\nu}_1 \\ 1 & \text{if } \boldsymbol{\nu} = \boldsymbol{\nu}_2 \text{ or } \boldsymbol{\nu} = \boldsymbol{\nu}_3 \\ 0 & \text{otherwise} \end{array} \right. .$$

Similarly, moves in $\mathcal{M}_2(\boldsymbol{x})$ are denoted by $\boldsymbol{b}_{\boldsymbol{\nu}}^-(\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3)$ and are generated by sequentially sampling:

1. $\boldsymbol{\nu}_2 \in \chi_{2\cdot}(\boldsymbol{x}) = \bigcup_{t=1}^{T} \mathcal{X}_{2t}(\boldsymbol{x})$

2. $s \in \{t : \nu_{2t} = 2\}$

3. $\boldsymbol{\nu}_3 \in \mathcal{X}_{3s}(\boldsymbol{x})$

and setting $\boldsymbol{\nu}_0 = \boldsymbol{\nu}_2 - 2\boldsymbol{\delta}_t$ and $\boldsymbol{\nu}_1 = \boldsymbol{\nu}_3 - 2\boldsymbol{\delta}_t$. The corresponding move has entries

$$(4) \qquad b_{\boldsymbol{\nu}}^-(\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3) = \left\{ \begin{array}{rl} 1 & \text{if } \boldsymbol{\nu} = \boldsymbol{\nu}_0 \text{ or } \boldsymbol{\nu} = \boldsymbol{\nu}_1 \\ -1 & \text{if } \boldsymbol{\nu} = \boldsymbol{\nu}_2 \text{ or } \boldsymbol{\nu} = \boldsymbol{\nu}_3 \\ 0 & \text{otherwise} \end{array} \right. .$$

On the kth iteration of our MCMC algorithm a proposal, $\boldsymbol{x}^{\text{prop}}$, is generated by choosing whether to add or remove an error from the current configuration, $\boldsymbol{x}^{(k-1)}$ with equal probability and sampling a move as described above. The proposal density when adding an error is

$$(5) \qquad q(\boldsymbol{x}^{\text{prop}}|\boldsymbol{x}^{(k-1)}) = \frac{.5}{\#\chi_{1\cdot}(\boldsymbol{x}^{(k-1)}) \cdot \#\{t : \nu_{1t} = 1\} \cdot \#\chi_{0s}(\boldsymbol{x}^{(k-1)})}$$

and when removing an error is

$$(6) \qquad q(\boldsymbol{x}^{\text{prop}}|\boldsymbol{x}^{(k-1)}) = \frac{.5}{\#\chi_{2\cdot}(\boldsymbol{x}^{(k-1)}) \cdot \#\{t : \nu_{2t} = 2\} \cdot \#\chi_{3s}(\boldsymbol{x}^{(k-1)})}$$

where $\#\mathcal{S}$ denotes the cardinality of $\mathcal{S}$. If we propose to add an error and $\boldsymbol{x}^{(k-1)}$ contains no errors, $\chi_{3\cdot}(\boldsymbol{x}^{(k-1)}) = \emptyset$, or if we propose to remove an error and $\boldsymbol{x}$ contains no correct identifications, $\chi_{3\cdot}(\boldsymbol{x}^{(k-1)}) = \emptyset$, then we set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)}$ and continue to the next iteration. Full details are provided in Algorithm 2. Proof that the resulting chains converge to the joint posterior distribution of $\boldsymbol{x}$, $\boldsymbol{\phi}$, and $\boldsymbol{p}$ is provided Supplement A.

6. Results. In our analysis of the queen snake data we fit an initial CJS model to the original PIT tag data (Model 1). We then simulated data mimicking what might be observed from the branding data by generating new data from the estimated demographic parameters and adding errors following the BRE model using the observed identification rates, $\alpha = 8/9$ and $\alpha = 6/9$. We refit the CJS model to each data set to assess the effects of errors that are not modelled (Model 2), and then fit the CJS/BRE model to the each data set using the methods described in Section 5 (Model 3). One hundred simulated data sets were generated for each value of α.

Analysis of the PIT tag data using maximum likelihood methods in Program MARK (White and Burnham, 1999) strongly supported a simplified CJS model which allowed the capture probabilities to vary independently

Initialise $\boldsymbol{\phi}^{(0)}$ and $\boldsymbol{p}^{(0)}$.
Initialise $\boldsymbol{x}^{(0)}$ so that $\boldsymbol{n} = \boldsymbol{A}\boldsymbol{x}^{(0)}$ and set $\boldsymbol{z}^{(0)} = \boldsymbol{B}\boldsymbol{x}^{(0)}$.
Set $k = 1$.

1. Update $\boldsymbol{\phi}$ and $\boldsymbol{p}$ conditional on $\boldsymbol{z}^{(k-1)}$. Call the results $\boldsymbol{\phi}^{(k)}$ and $\boldsymbol{p}^{(k)}$.

2. Update $\boldsymbol{x}$ and $\boldsymbol{z}$ conditional on $\boldsymbol{\phi}^{(k)}$ and $\boldsymbol{p}^{(k)}$ as follows.

   i) With probability .5 sample $\boldsymbol{b}$ from $\mathcal{M}_1(\boldsymbol{x}^{(k-1)})$. If $\mathcal{M}_1(\boldsymbol{x}^{(k-1)}) = \emptyset$ then set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)}$ and continue to step v).
   Otherwise sample $\boldsymbol{b}$ from $\mathcal{M}_2(\boldsymbol{x}^{(k-1)})$. If $\mathcal{M}_2(\boldsymbol{x}^{(k-1)}) = \emptyset$ then set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)}$ and continue to step v).

   ii) Set $\boldsymbol{x}^{\text{prop}} = \boldsymbol{x}^{(k-1)} + \boldsymbol{b}$.

   iii) Calculate the Metropolis acceptance probability:

   $$r(\boldsymbol{x}, \boldsymbol{x}^{\text{prop}} | \boldsymbol{\phi}^{(k)}, \boldsymbol{p}^{(k)}, \alpha) = \min\left\{ 1, \frac{\pi(\boldsymbol{x}^{\text{prop}} | \boldsymbol{n}, \boldsymbol{\phi}^{(k)}, \boldsymbol{p}^{(k)}, \alpha)}{\pi(\boldsymbol{x}^{(k-1)} | \boldsymbol{n}, \boldsymbol{\phi}^{(k)}, \boldsymbol{p}^{(k)}, \alpha)} \cdot \frac{q(\boldsymbol{x}^{(k-1)} | \boldsymbol{x}^{\text{prop}})}{q(\boldsymbol{x}^{\text{prop}} | \boldsymbol{x}^{(k-1)})} \right\}.$$

   iv) Set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{\text{prop}}$ with probability $r(\boldsymbol{x}, \boldsymbol{x}^{\text{prop}} | \boldsymbol{\phi}^{(k)}, \boldsymbol{p}^{(k)}, \alpha)$ and $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)}$ otherwise.

   v) Set $\boldsymbol{z}^{(k)} = \boldsymbol{B}\boldsymbol{x}^{(k)}$

3. Increment k.

Algorithm 2: Proposed algorithm for sampling from the posterior distribution of the CJS/BRE model using the dynamic Markov basis.

across all occasions but constrained survival to be equal on the final 8 occasions. The estimated survival probabilities were $\hat{\phi}_1 = 0.66$, $\hat{\phi}_2 = 1.00$, and $\hat{\phi}_3 = \ldots = \hat{\phi}_9 = 0.93$. Clearly the overwintering survival rate, ϕ_1, has the largest effect on the population and is of most interest.

Figure 1 compares the bias of the posterior means and the width and coverage of the central 95% credible intervals (CIs) for the three models. Fitting the standard CJS model to the data without errors (Model 1) provides an estimate of ϕ_1 that is almost unbiased and 95% CIs with coverage above the nominal value. Fitting the same CJS model to the data with errors (Model 2) produced very poor results. When $\alpha = 8/9$ the estimated bias of $\hat{\phi}_1$ was 0.15 (23%) and the coverage of ϕ_1 was only 47%. The bias increased one and a half times to 0.24 (36%) and the coverage dropped to only 6% when α decreased to 6/9. In comparison, the posterior mean of ϕ_1 from the CJS/BRE model (Model 3) was negligibly biased for both levels of error and coverage of the 95% CI again exceeded the nominal rate. As expected, credible intervals from Model 3 were wider than those from Model 1 to account for the extra uncertainty introduced by the errors. All models produced estimates of ϕ_2 that were biased because the true parameter lies
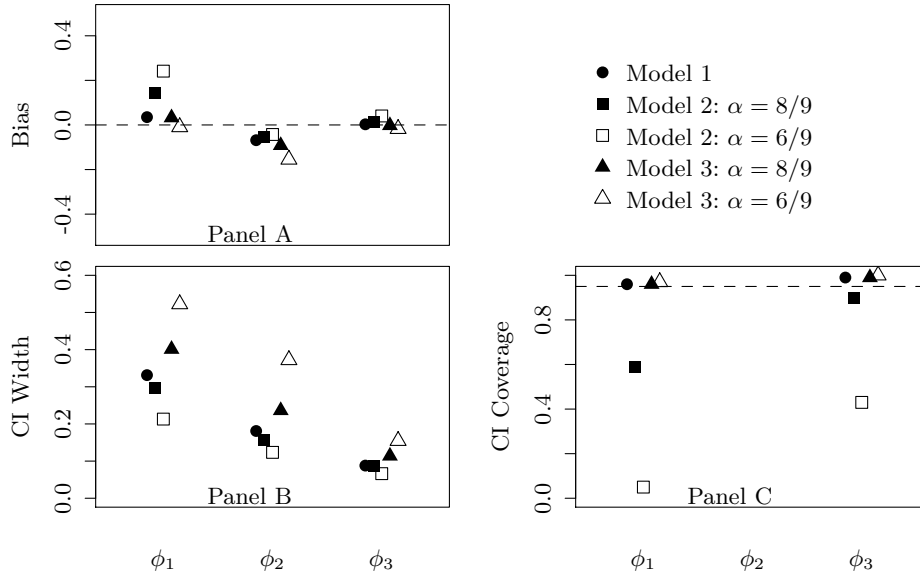
Fig 1. *Results of the analysis of the queen snake data. The three panels present the estimate bias of the posterior means (Panel A) and the estimated width (Panel B) and coverage probability (Panel C) of the 95% credible interval for the survival probabilities for the three models described in Section 6. The different models are indicated by the shape of the plotting symbol. The rates of error are for Models 2 and 3 indicated by the colour of the symbol. Coverage of ϕ_2 is not reported because the true parameter lies on the boundary of the parameter space.*

on the boundary of the parameter space. Coverage of this parameter was zero for all models and is not reported. The posterior mean of ϕ_2 from Model 3 was significantly more biased than that of Model 1, underestimating ϕ_2 by 9% when $\alpha = 8/9$ and 16% when $\alpha = 6/9$. This is due to there being more significant shrinkage toward the prior mean of 0.50 when there is more uncertainty in the data.

7. Computational Efficiency. Not only does the dynamic Markov basis allow us to fit the CJS/BRE model when T is large, but it also leads to more efficient sampling when the full Markov basis can be computed. To illustrate this, we present results from analysing a single simulated data set with $T = 4$ capture occasions (the largest number for which we can compute the Markov basis using 4ti2). Data was generated for a sample of 30 individuals with constant survival probability $\phi_1 = \phi_2 = \phi_3 = .8$, constant capture probability $p_2 = p_3 = p_4 = .5$, and error rate $\alpha = .5$.
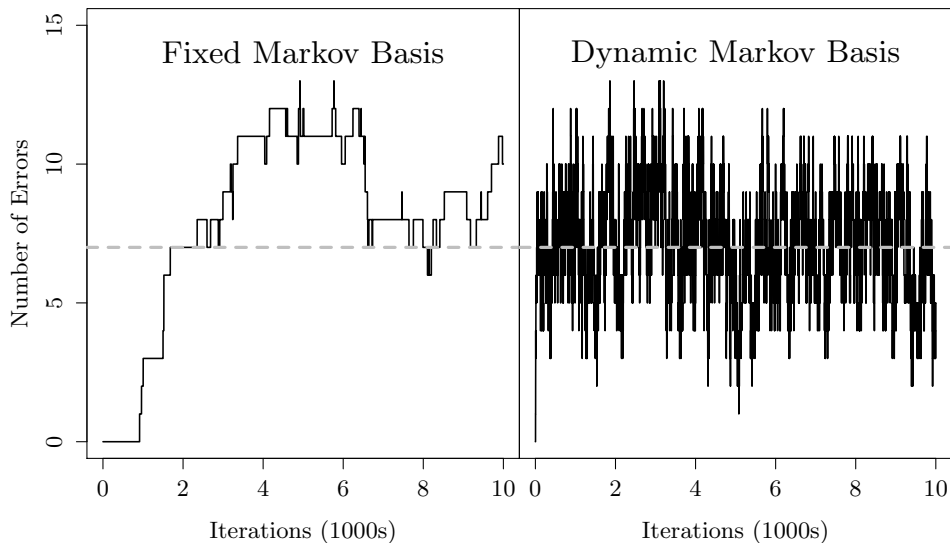
FIG 2. *Comparison of the chains sampling from the posterior distribution of the CJS/BRE model applied to the simulated data. The figures trace the number of errors in $\boldsymbol{x}$ for the algorithms using the fixed basis (left) and dynamic basis (right). The grey dotted lines represent the true number of errors in the data set.*

Samples from the joint posterior distribution of $\boldsymbol{x}$ and $\boldsymbol{z}$ were then drawn using Step 1 Algorithm 1 and Step 2 of of Algorithm 2 while keeping $\boldsymbol{\phi}$ and $\boldsymbol{p}$ fixed at their true values.

We assessed how well the chains mixed by comparing the acceptance rates and the number of unique solutions for $\boldsymbol{x}$ identified. The chain constructed using Algorithm 1 identified a total of 79 unique configurations among the 7,500 values of $\boldsymbol{x}$ sampled after the burn-in phase. Less than 1% of the proposed configurations were accepted. In comparison, the chain constructed with Algorithm 2 identified 2548 unique configurations and 38% of the proposed configurations were accepted. Figure 2 provides traceplots of the chains using the number of errors in the configurations sampled on each accept/reject step as a metric. These summaries all make it clear that the chain constructed from Algorithm 2 is mixing and moving through the fibre much more quickly than the chain constructed from Algorithm 1.

8. Discussion. The results in Section 6 clearly illustrate the problems with misidentification. The overwintering survival probability was overestimated by 23% or 36% depending on the error rate and coverage of the 95%

CIs was always less than 50%. In comparison, estimates from the CJS/BRE model were almost unbiased and the credible intervals had above nominal coverage. The extra uncertainty in the error does increase the posterior variances and the obvious recommendation is to reduce error rates experimentally by using marks that are clearer or tagging individuals twice. Uncertainty could also be reduced by pairing observers or simply by increasing the number of hours spent in the field to raise capture rates. Of course, these measures would increase expense, and we are currently assessing the costs and benefits of these options.

Although we have focused on the CJS/BRE model, we believe that the methods presented in Section 5 should be applicable to broad range of mark-recapture models with possible errors. As a second example, we describe the application of these methods to model $M_{t\alpha}$ in Supplement B. However, the framework described in Section 5.1 can incorporate more complex models of both the capture and error processes than the original LMM and is particularly useful when the distribution of the joint histories described by the combining processes is intractable. The algorithm based on dynamic Markov bases presented in Section 5.2 essentially entails moving through $\mathcal{F}_{\boldsymbol{n}}$ by adding or removing errors one at a time, and we expect that the same procedure can be applied to an even broader set of models, with two important caveats. First, it must be possible to write the model in terms of the two linear constraints described in the extended framework. This will not always be the case and does not happen if we extend the BRE model so that individual i can be captured on occasion t and another individual can be captured and identified as individual i at the same time. We are working to extend these models to allow for such events. The second caveat is that the Markov chains derived from the new algorithm may not be irreducible if the posterior distribution assigns probability zero to some elements in $\mathcal{F}_{\boldsymbol{n}}$. This might occur if certain configurations of the errors can be ruled out *a priori*, and would require the Markov basis to be expanded further.

An important issue that remains is how the connectivity of $\mathcal{F}_{\boldsymbol{n}}$ and the efficiency of the chains are affected by different dynamic Markov bases. In the methods of Dobra (2012), moves are generated on each iteration of the MCMC algorithm by sampling $M \in \{1, 2, \ldots, J\}$ according to some density $g(\cdot)$, randomly sampling a permutation function $\delta(\cdot)$ from the set of all permutations of the indices of $\boldsymbol{x}$, and sequentially resampling the counts in the first M cells in $\delta(\boldsymbol{x})$ to maintain the linear constraint. The moves in our dynamic Markov basis comprise a subset of these moves for which 1) $M = 4$, 2) the set of permutations is restricted so that the first four cells belong to $\xi_{0t}(\boldsymbol{x})$, $\xi_{1t}(\boldsymbol{x})$, $\xi_{2t}(\boldsymbol{x})$, and $\xi_{3t}(\boldsymbol{x})$ respectively for some t, and 3) the counts in

these cells are modified by adding or subtracting the vector $(-1, -1, 1, 1)'$. Both bases connect the fibre and produce irreducible Markov chains. The basis of Dobra (2012) contains many more moves and has the advantage that the chains will sometimes make larger jumps in $\mathcal{F}_{\boldsymbol{n}}$. However, this will probably reduce the acceptance rate. Exploring the balance between these extremes to produce efficient samples is a topic requiring further research.

APPENDIX A: CORMACK-JOLLY-SEBER MODEL

The basic assumptions of the CJS model are that (see e.g. Seber (2002, pg. 196)):

1. Each individual alive on occasion t survives to occasion $t + 1$ with probability ϕ_t.
2. Each individual alive on occasion t is captured with probability p_t.
3. All individuals are correctly identified when captured (i.e., marks are not lost or misread).
4. All events are independent.

Given these assumptions, probabilities are assigned to the capture histories conditional on the first release of each individual. For example, the probability assigned to the history $\omega = 01010$ is

$$f(\omega|\boldsymbol{\phi}, \boldsymbol{p}) = \phi_2(1 - p_3)\phi_3 p_4(\phi_4(1 - p_5) + (1 - \phi_4)).$$

The final term accounts for the possibility that the individual was not observed on occasion 5 either because it did not survive or survived and was not captured. The likelihood can then be written as a product multinomial so that:

$$f(\boldsymbol{n}|\boldsymbol{\phi}, \boldsymbol{p}) \propto \prod_{i=1}^{I} f(\omega_i|\boldsymbol{\phi}, \boldsymbol{p})^{n_i}.$$

APPENDIX B: NOTATION

Algebraic Statistics and Markov Bases:

$\mathcal{F}_{\boldsymbol{n}}$	$\boldsymbol{n}$-fibre, $\mathcal{F}_{\boldsymbol{n}} = \{\boldsymbol{x} \in \mathbb{N}^J : \boldsymbol{n} = \boldsymbol{A}\boldsymbol{x}\}$.
$\mathcal{B}$	Lattice basis for $\ker(\boldsymbol{A})$.
$\mathcal{M}$	Markov basis for $\ker(\boldsymbol{A})$.
$\mathcal{M}(\boldsymbol{x})$	Dynamic Markov basis for $\ker(\boldsymbol{A})$ computed at $\boldsymbol{x}$.

Extended LMM:

n_{Tot}	Number of distinct individuals captured and marked.
$\boldsymbol{\omega}_i$	Observed capture history for the ith marked individual.
$\boldsymbol{\nu}_i$	Latent error history for the ith marked individual.
$\boldsymbol{\xi}_i$	Latent capture history for the ith marked individual.
$\boldsymbol{n}$	Observed vector of counts for the observable histories (indexed by either i and $\boldsymbol{\omega}$).
$\boldsymbol{x}$	Unknown vector of counts for the latent error histories (indexed by either j and $\boldsymbol{\nu}$).
$\boldsymbol{z}$	Unknown vector of counts for the latent capture histories (indexed by either k and $\boldsymbol{\xi}$).
I	Length of $\boldsymbol{n}$. For the CJS/BRE model $I = 2^T - 2$.
J	Length of $\boldsymbol{x}$. For the CJS/BRE model $J = (4^T - 1)/3 - 1$.
K	Length of $\boldsymbol{z}$. For the CJS/BRE model $K = 2^T - 2$.
$\boldsymbol{A}$	$I \times J$ matrix mapping $\boldsymbol{x}$ onto $\boldsymbol{n}$, $\boldsymbol{n} = \boldsymbol{A}\boldsymbol{x}$.
$\boldsymbol{B}$	$K \times J$ matrix mapping $\boldsymbol{x}$ onto $\boldsymbol{z}$, $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}$.
$\boldsymbol{\theta}_1$	Parameters in the model of $\boldsymbol{z}$.
$\boldsymbol{\theta}_2$	Parameters in the conditional model of $\boldsymbol{x}$ given $\boldsymbol{z}$.

Band-Read Error Model:

p_t	Capture probability: the probability that an individual alive on occasion t is captured, $t = 2, \ldots, T$.
ϕ_t	Survival probability: the probability that an individual is alive on occasion $t+1$ given that it was alive on occasion t, $t = 1, \ldots, T-1$.
α	Correct identification rate: the probability that a captured individual is identified correctly.

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIAL

Supplement A: Proof of Convergence
(http://www.e-publications.org/ims/support/dowload/imsart-ims.zip). Proof that the chains generated by Algorithm 2 converge to the correct distribution.

Supplement B: Model $M_{t\alpha}$
(http://www.e-publications.org/ims/support/dowload/imsart-ims.zip). Ap-

plication of the extended framework with dynamic Markov bases to model $M_{t\alpha}$.

REFERENCES

ALLENDER, M. C., DRESLIK, M. J., WYLIE, D. B., WYLIE, S. J., SCOTT, J. W., and PHILLIPS, C. A. (2013). Ongoing health assessment and prevalence of *Chrysosporium* in the Eastern Massasauga (*Sistrurus catenatus catenatus*). *Copeia* **1** 97–102.

BARKER, R. J., SCHOFIELD, M. R., WRIGHT, J. A., FRANTZ, A. C. and STEVENS, C. (2014). Closed-population capturerecapture modeling of samples drawn one at a time. *Biometrics* **70** 775–782.

BONNER, S. J. and HOLMBERG, J. (2013). Mark-recapture with multiple non-invasive marks. *Biometrics* **69** 766–775.

COWEN, L. and SCHWARZ, C. J. (2006). The Jolly-Seber model with tag loss. *Biometrics* **62** 699–705.

DIACONIS, P. and STURMFELS, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Annals of Statistics* **26** 363–397.

DOBRA, A. (2012). Dynamic Markov bases. *Journal of Computational and Graphical Statistics* **21** 496–517.

HEMMECKE, R., HEMMECKE, R., KOEPPE, M., MALKIN, P. and WALTER, M. (2013). User's guide for `4ti2` version 1.6.

LEBRETON, J.-D., BURNHAM, K. P., CLOBERT, J. and ANDERSON, D. R. (1992). Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* **62** 67–118.

LINK, W. A., YOSHIZAKI, J., BAILEY, L. L. and POLLOCK, K. H. (2010). Uncovering a latent multinomial: analysis of mark-recapture data with misidentification. *Biometrics* **66** 178–185.

LUKACS, P. M. and BURNHAM, K. P. (2005). Estimating population size from DNA-based closed capture-recapture data incorporating genotyping error. *Journal of Wildlife Management* **69** 396—-403.

MCCLINTOCK, B. T., CONN, P., ALONSO, R. and CROOKS, K. R. (2013). Integrated modeling of bilateral photo-identification data in mark-recapture analyses. *Ecology* **94** 1464-1471.

MCCLINTOCK, B. T., HILL, J. M., FRITZ, L., CHUMBLEY, K., LUXA, K. and DIEFENBACH, D. R. (2014). Mark-resight abundance estimation under incomplete identification of marked individuals. *Methods in Ecology and Evolution* **5** 1294–1304.

MORRISON, T. A., YOSHIZAKI, J., NICHOLS, J. D. and BOLGER, D. T. (2011). Estimating survival in photographic capture-recapture studies: overcoming misidentification error. *Methods in Ecology and Evolution* **2** 454–463.

ROARK, A. W. and DORCAS, M. E. (2000). Regional body temperature variation in corn snakes measured using temperature-sensitive passive integrated transponders. *Journal of Herpetology* **34** 481-485.

SCHOFIELD, M. and BONNER, S. (2015). Connecting the latent multinomial. *Biometrics*. In press.

SEBER, G. A. F. (2002). *The Estimation of Animal Abundance and Related Parameters*, 2nd ed. The Blackburn Press, New Jersey, USA.

SLEEMAN, J. (2013). Snake fungal disease in the United States Bulletin No. 2013-02, USGS National Wildlife Health Center.

WHITE, G. C. and BURNHAM, K. P. (1999). Program MARK: survival estimation from populations of marked animals. *Bird Study* **46** 120–139.

560   WILLIAMS, B. K., NICHOLS, J. D. and CONROY, M. J. (2002). *Analysis and Management*
561   *of Animal Populations.* Academic Press, San Diego, California.
562   WINNE, C. T., WILLSON, J. D., ANDREWS, K. M. and REED, R. N. (2006). Efficacy of
563   marking snakes with disposable medical cautery units. *Herpetological Review* **37** 52–54.
564   WRIGHT, J. A., BARKER, R. J., SCHOFIELD, M. R., FRANTZ, A. C., BYROM, A. E. and
565   GLEESON, D. M. (2009). Incorporating genotype uncertainty into mark-recapture-type
566   models for estimating abundance using DNA samples. *Biometrics* **65** 833–40.
567   YOSHIZAKI, J., BROWNIE, C., POLLOCK, K. H. and LINK, W. A. (2011). Modeling
568   misidentification errors that result from use of genetic tags in capture-recapture studies.
569   *Environmental and Ecological Statistics* **18** 27–55.
570   YOSHIZAKI, J., POLLOCK, K. H., BROWNIE, C. and WEBSTER, R. A. (2012). Modeling
571   misidentification errors in capture-recapture studies using photographic identification
572   of evolving marks. *Ecology* **90** 3–9.

573   DEPARTMENT OF STATISTICS   DEPARTMENT OF MATHEMATICS AND STATISTICS
      UNIVERSITY OF KENTUCKY   UNIVERSITY OF OTAGO
      LEXINGTON, KENTUCKY 40536-0082 USA   DUNEDIN 90504, NEW ZEALAND
      E-MAIL: simon.bonner@uky.edu   E-MAIL: mschofield@maths.otago.ac.nz

574   INSTITUTE OF SCIENCE AND TECHNOLOGY AUSTRIA DEPARTMENT OF FORESTRY
      AM CAMPUS 1   UNIVERSITY OF KENTUCKY
      3400 KLOSTERNEUBURG, AUSTRIA   LEXINGTON, KENTUCKY 40546-0073 USA
      E-MAIL: patrik.noren@ist.ac.at   E-MAIL: steven.price@uky.edu

SUPPLEMENT TO
"EXTENDING THE LATENT MULTINOMIAL MODEL WITH COMPLEX ERROR PROCESSES AND DYNAMIC MARKOV BASES"

By Simon J Bonner[*], Matthew R Schofield[†], Patrik Noren[‡] and Steven J Price[*]

University of Kentucky[*], University of Otago[†], and Institute of Science and Technology Austria[‡]

Supplement A: Proof of Convergence. To prove that chains generated from Algorithm 2 from the manuscript converge to the correct distribution, we need to satisfy three conditions: 1) that Step 1 produces chains which converge to $\pi(\boldsymbol{\phi}, \boldsymbol{p}|\boldsymbol{z})$ for any $\boldsymbol{z}$ such that $\boldsymbol{z} = \boldsymbol{Bx}$ for some $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{n}}$, 2) that Step 2 produces chains which converge to $\pi(\boldsymbol{x}|\boldsymbol{n}, \boldsymbol{\phi}, \boldsymbol{p}, \alpha)$ for any $\boldsymbol{\phi}$ and $\boldsymbol{p}$ in the parameter space, and 3) that the joint posterior distribution satisfies the positivity condition of (Robert and Casella, 2010, pg. 345). Sampling from $\pi(\boldsymbol{\phi}, \boldsymbol{p}|\boldsymbol{z})$ is equivalent to sampling from the posterior distribution for a simple CJS model. This is now standard (Link et al., 2002, see e.g.), and so we assume that Condition 1 is satisfied. It is also simple to show that the positivity constraint is satisfied given that the prior distributions for $\boldsymbol{\phi}$ and $\boldsymbol{p}$ are positive over all of $(0,1)^T \times (0,1)^{T-1}$, as assumed in Section 5 of the original manuscript. It remains to show that Condition 2 holds.

We assume here that $\mathcal{F}_{\boldsymbol{n}}$ contains at least two elements. The fibre always contains at least one element with no errors which we denote by $\boldsymbol{x}^{\emptyset}$. The entries of this element are

$$x_{\boldsymbol{\nu}}^{\emptyset} = \begin{cases} n_{\boldsymbol{\nu}} & \text{if } \boldsymbol{\nu} \text{ is observable} \\ 0 & \text{otherwise} \end{cases} .$$

Cases in which $\mathcal{F}_{\boldsymbol{n}} = \{x^{\emptyset}\}$ arise when no errors could have occurred, for example, if no individuals were ever recaptured. These situations are easily identified and there is no need to sample from the joint posterior of both $\boldsymbol{x}$ and $\boldsymbol{\theta}$ in such cases since $\boldsymbol{x} = x^{\emptyset}$ with probability one.

Some useful results that are easy to prove are:

1. that any configuration of the latent error histories within the fibre has positive probability under the conditional posterior for all values of the parameters in the parameter space,

LEMMA 1. *If $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{n}}$ then $\pi(\boldsymbol{x}|\boldsymbol{n}, \boldsymbol{\phi}, \boldsymbol{p}, \alpha) > 0$ for all values of $\boldsymbol{\phi}$ and $\boldsymbol{p}$ in the parameter space.*

2. that the local sets within the dynamic Markov basis are symmetric,

LEMMA 2. *Let $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{n}}$. If $\boldsymbol{b}^+ \in \mathcal{M}_1(\boldsymbol{x})$ then $-\boldsymbol{b}^+ \in \mathcal{M}_2(\boldsymbol{x}+\boldsymbol{b}^+)$ and if $\boldsymbol{b}^- \in \mathcal{M}_2(\boldsymbol{x})$ then $-\boldsymbol{b}^- \in \mathcal{M}_1(\boldsymbol{x}+\boldsymbol{b}^-)$.*

3. that all proposals remain inside $\mathcal{F}_{\boldsymbol{n}}$,

LEMMA 3. *Let $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{n}}$. If $\boldsymbol{b} \in \mathcal{M}(\boldsymbol{x}) = \mathcal{M}_1(\boldsymbol{x}) \bigcup \mathcal{M}_2(\boldsymbol{x})$ then $\boldsymbol{x}+\boldsymbol{b} \in F_{\boldsymbol{n}}$.*

4. that there is a unique element $\boldsymbol{x}^\emptyset$ in $\mathcal{F}_{\boldsymbol{n}}$ with no errors.

LEMMA 4. *Suppose that $\boldsymbol{x}^\emptyset \in \mathcal{F}_{\boldsymbol{n}}$. Then $e_t(\boldsymbol{x}^\emptyset) = 0 \ \forall t = 2, \ldots, T$ if and only if*

$$x_{\boldsymbol{\nu}}^\emptyset = \begin{cases} n_{\boldsymbol{\nu}} & \text{if } \boldsymbol{\nu} \text{ is observable} \\ 0 & \text{otherwise} \end{cases}.$$

First we establish irreducibility. Proposition 1 implies that there is a path connecting any two elements in the fibre while Proposition 2 implies that each step, and hence the entire path, has positive probability under the transition kernel. Together, these show that that the chains are irreducible.

PROPOSITION 1. *For any distinct $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{F}_{\boldsymbol{n}}$ there exists a sequence of moves $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_L$ such that:*

1. $\boldsymbol{b}_{L'} \in \mathcal{M}\left(\boldsymbol{x}_1 + \sum_{l=1}^{L'-1} \boldsymbol{b}_l\right)$ for all $L' = 1, \ldots, L$

2. $\boldsymbol{x}_1 + \sum_{l=1}^{L'} \boldsymbol{b}_l \in \mathcal{F}_{\boldsymbol{n}}$ for all $L' = 1, \ldots, L-1$, and

3. $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \sum_{l=1}^{L} \boldsymbol{b}_l$,

where we take $\boldsymbol{x}_1 + \sum_{l=1}^{0} \boldsymbol{b}_l = \boldsymbol{x}_1$.

PROOF. Our proof follows by (reverse) induction on the number of errors. Suppose that $e_t(\boldsymbol{x}_1) > 0$ for some t. Then $\mathcal{X}_{2t}(\boldsymbol{x}_1)$ and $\mathcal{X}_{3t}(\boldsymbol{x}_1)$ are both non-empty and $\exists \boldsymbol{b}_{11}^- \in \mathcal{M}_2(\boldsymbol{x}_1)$. Then $e_t(\boldsymbol{x}_1+\boldsymbol{b}_{11}^-) = e_t(\boldsymbol{x}_1)-1$ and $\boldsymbol{x}_1+\boldsymbol{b}_{11}^- \in \mathcal{F}_{\boldsymbol{n}}$ by Lemma 3. Repeating this procedure $L_1 = \sum_{t=2}^{T} e_t(\boldsymbol{x}_1)$ times, we find $\boldsymbol{b}_{11}^-, \ldots, \boldsymbol{b}_{1L_1}^-$ such that

1. $\boldsymbol{b}_{1L'}^- \in \mathcal{M}_2(\boldsymbol{x}_1 + \sum_{l=1}^{L'-1} \boldsymbol{b}_{1l}^-)$ for $L' = 1, \ldots, L_1$,

2. $\boldsymbol{x}_1 + \sum_{l=1}^{L'} \boldsymbol{b}_{1l}^- \in \mathcal{F}_{\boldsymbol{n}}$ for all $L' = 1, \ldots, L_1$, and

61 3. $e_t(\boldsymbol{x}_1 + \sum_{l=1}^{L_1} \boldsymbol{b}_{1l}^-) = 0$.

62 It follows from Lemma 4 that $\boldsymbol{x}_1 + \sum_{l=1}^{L_1} \boldsymbol{b}_{1l}^- = \boldsymbol{x}^\emptyset$. By the same argument,
63 $\exists \boldsymbol{b}_{21}^-, \ldots, \boldsymbol{b}_{2L_2}^-$ such that

64 1. $\boldsymbol{b}_{2l}^- \in \mathcal{M}_2(\boldsymbol{x}_2 + \sum_{l=1}^{L'-1} \boldsymbol{b}_{2l}^-)$ for $L' = 1, \ldots, L_2$,,
65 2. $\boldsymbol{x}_2 + \sum_{l=1}^{L'} \boldsymbol{b}_{2l}^- \in \mathcal{F}_{\boldsymbol{n}}$ for all $L' = 1, \ldots, L_2$, and
66 3. $\boldsymbol{x}_2 + \sum_{l=1}^{L_2} \boldsymbol{b}_{2l}^- = \boldsymbol{x}^\emptyset$.

67 Moreover, $-\boldsymbol{b}_{2,L_2-l+1}^- \in M_1(\boldsymbol{x}^\emptyset + \sum_{l=0}^{L'-1} -\boldsymbol{b}_{2,L_2-l}^-)$ for all $L' = 1, \ldots, L_2$
68 by Lemma 2. Then the sequence $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_L$ where $L = L_1 + L_2$, $\boldsymbol{b}_l = \boldsymbol{b}_{1l}^-$
69 for $l = 1, \ldots, L_1$, and $\boldsymbol{b}_{L_1+l} = -\boldsymbol{b}_{2,L_2-l+1}^-$ for $l = 1, \ldots, L_2$ satisfies the
70 conditions of the proposition. Note that half of this argument suffices if
71 either $\boldsymbol{x}_1 = \boldsymbol{x}^\emptyset$ or $\boldsymbol{x}_2 = \boldsymbol{x}^\emptyset$. $\qquad\square$

72 PROPOSITION 2. *Let $\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{n}}$. If $\boldsymbol{b} \in \mathcal{M}(\boldsymbol{x})$ then $P(\boldsymbol{x}^{(k+1)} = \boldsymbol{x}+\boldsymbol{b}|\boldsymbol{x}^{(k)} =$*
73 *$\boldsymbol{x}) > 0$.*

74 PROOF. Suppose that $\boldsymbol{b} \in \mathcal{M}_1(\boldsymbol{x})$ and let $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{b}$. Then $-\boldsymbol{b} \in \mathcal{M}_2(\boldsymbol{x}')$
75 by Lemma 2. Direct calculation of equations (5) and (6) shows that both
76 $q(\boldsymbol{x}'|\boldsymbol{x}) > 0$ and $q(\boldsymbol{x}|\boldsymbol{x}') > 0$. Combined with Lemma 1 it follows that
77 $r(\boldsymbol{x}, \boldsymbol{x}'|\boldsymbol{\phi}^{(k)}, \boldsymbol{p}^{(k)}, \alpha)$ (defined in Step 2, Substep iii of Algorithm 2) is positive
78 and hence that $P(\boldsymbol{x}^{(k+1)} = \boldsymbol{x}'|\boldsymbol{x}^{(k)} = \boldsymbol{x}) = q(\boldsymbol{x}'|\boldsymbol{x}) \cdot r(\boldsymbol{x}, \boldsymbol{x}'|\boldsymbol{\phi}^{(k)}, \boldsymbol{p}^{(k)}, \alpha) > 0$.
79 A similar argument shows that $P(\boldsymbol{x}^{(k+1)} = \boldsymbol{x} + \boldsymbol{b}|\boldsymbol{x}^{(k)} = \boldsymbol{x}) > 0$ for all
80 $\boldsymbol{b} \in \mathcal{M}_2(\boldsymbol{x})$. $\qquad\square$

81 We establish aperiodicity by showing that there is positive probability of
82 holding at $\boldsymbol{x}^\emptyset$.

83 PROPOSITION 3. *If $\boldsymbol{x}^{(k)} = \boldsymbol{x}^\emptyset$ then $P(\boldsymbol{x}^{(k+1)} = \boldsymbol{x}_0|\boldsymbol{x}^{(k)} = \boldsymbol{x}_0) \geq .5$.*

84 PROOF. The set $\mathcal{M}_2(\boldsymbol{x}^\emptyset)$ is empty since there are no errors to remove from
85 $\boldsymbol{x}^\emptyset$. However, Algorithm 2 still proposes to draw a move from $\mathcal{M}_2(\boldsymbol{x}^\emptyset)$ with
86 probability .5. When this occurs we set $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)}$ so that $P(\boldsymbol{x}^{(k+1)} =$
87 $\boldsymbol{x}^\emptyset|\boldsymbol{x}^{(k)} = \boldsymbol{x}^\emptyset) \geq .5$. $\qquad\square$

88 This shows that $\boldsymbol{x}^\emptyset$ is an aperiodic state and hence that the entire chain is
89 aperiodic (Cinlar, 1975, pg. 125)
90 Since the fibre is finite, irreducibility and aperiodicty are sufficient to en-
91 sure that the chains have a unique stationary distribution which is also the
92 limiting distribution (see Cinlar, 1975, Corollary 2.11). That this distribu-
93 tion is equal to the target distribution is guaranteed by the detailed balance

94   condition of the MH algorithm which holds under Proposition 4 (Liu, 2004,
95   pg. 111).

96   PROPOSITION 4.   *If $q(\boldsymbol{x}'|\boldsymbol{x}) > 0$ then $q(\boldsymbol{x}|\boldsymbol{x}') > 0$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{F}_{\boldsymbol{n}}$.*

97   PROOF. Suppose that $q(\boldsymbol{x}'|\boldsymbol{x}) > 0$. Then either $\boldsymbol{x}' - \boldsymbol{x} \in \mathcal{M}_1(\boldsymbol{x})$ or $\boldsymbol{x}' -$
98   $\boldsymbol{x} \in \mathcal{M}_2(\boldsymbol{x})$. If $\boldsymbol{x}' - \boldsymbol{x} \in \mathcal{M}_1(\boldsymbol{x})$ then $\boldsymbol{x} - \boldsymbol{x}' \in \mathcal{M}_2(\boldsymbol{x}')$ by Lemma 2 and
99   $q(\boldsymbol{x}|\boldsymbol{x}') > 0$. Similarly, if $\boldsymbol{x}' - \boldsymbol{x} \in \mathcal{M}_2(\boldsymbol{x})$ then $\boldsymbol{x} - \boldsymbol{x}' \in \mathcal{M}_1(\boldsymbol{x}')$ and
100 $q(\boldsymbol{x}|\boldsymbol{x}') > 0$. $\square$

101 This completes our proof that the Markov chains produced by Algorithm
102 2 have unique limiting distribution $\pi(\boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{p}|\boldsymbol{n}, \alpha)$ so that realisations from
103 the tail of a converged chain can be used to approximate properties of the
104 joint posterior distribution of $\boldsymbol{x}$, $\boldsymbol{\phi}$, and $\boldsymbol{p}$.

105   **Supplement B: Model $M_{t\alpha}$.**   Here we show how the model described
106 by Link et al. (2010) can be fit into the extended model using a dynamic
107 Markov basis to sample from the posterior distribution. Model $M_{t\alpha}$ extends
108 the standard closed population model with time dependent capture prob-
109 abilities by allowing for individuals to be misidentified. Specifically, $M_{t\alpha}$
110 assumes that individuals are identified correctly with probability α on each
111 capture and that errors do not duplicate observed marks in that one marked
112 individual cannot be identified as another marked individual and that the
113 same error can never occur twice. This means that each error creates a new
114 recorded history with a single non-zero entry. For example, suppose that
115 $T = 3$ and that individual i is captured on the first occasion, recaptured
116 and misidentified on the second occasion, and not captured on the third oc-
117 casion. The true capture history for this individual, including errors, is 120,
118 where the event 2 denotes that the individual was captured and misidenti-
119 fied. The individual would then contribute two recorded histories, 100 and
120 010, to the observed data set.

121   *Extended Formulation.*   As with the CJS/BRE model, we cast model
122 $M_{t\alpha}$ into the new framework by 1) identifying the sets of observable capture
123 histories, latent error histories, and latent capture histories, 2) constructing
124 the linear constraints matrices for the corresponding count vectors, and 3)
125 identifying the components of the likelihood function. For an experiment
126 with T occasions, the set of observable capture histories contains all 2^{T-1}
127 histories in $\{0, 1\}^T$ excluding the unobservable zero history, the set of latent
128 error histories includes all 3^T histories in $\{0, 1, 2\}^T$, and the set of latent
129 capture histories includes all 2^T histories in $\{0, 1\}^T$. The matrix $\boldsymbol{A}$ is defined

130 exactly as in Link et al. (2010): $A_{ij} = 1$ if the ith observable history is
131 observed from the jth latent error history and $A_{ij} = 0$ otherwise. Similarly,
132 $B_{kj} = 1$ if the kth latent capture history has the same pattern of captures as
133 the jth latent error history. Mathematically, let $\boldsymbol{\omega}_i$, $\boldsymbol{\nu}_j$, and $\boldsymbol{\xi}_k$ represent the
134 i^{th}, j^{th}, and k^{th} observable history, latent error history, and latent capture
135 history for some implicit orderings. Then

$$
A_{ij} = \begin{cases} 1 & \text{if } \omega_{it} = I(\nu_{jt} = 1) \text{ for all } t = 1, \ldots, T \\ & \text{or if } \boldsymbol{\omega}_i = \boldsymbol{\delta}_t \text{ and } \nu_{jt} = 2 \text{ for some } t \in \{1, \ldots, T\} \\ \\ 0 & \text{otherwise} \end{cases}
$$

136 and

$$
B_{kj} = \begin{cases} 1 & \xi_{kt} = I(\nu_{jt} = 1) + I(\nu_{jt} = 2) \text{ for all } t = 1, \ldots, T \\ \\ 0 & \text{otherwise} \end{cases}.
$$

137 Here $\boldsymbol{\delta}_t$ represents t^{th} column of the $T \times T$ identity matrix (i.e., the vector
138 with a single 1 in the tth entry). Finally, the two components of the likelihood
139 function are

$$
\pi(\boldsymbol{z}|\boldsymbol{p}) = \frac{N!}{\prod_{\boldsymbol{\xi} \in \mathcal{Z}} z_{\boldsymbol{\xi}}!} \prod_{\boldsymbol{\xi} \in \mathcal{Z}} \left[\prod_{k=1}^{K} p_k^{I(\xi_k=1)} (1 - p_k)^{I(\xi_k=0)} \right]^{z_{\boldsymbol{\xi}}}
$$

140 and

$$
\pi(\boldsymbol{x}|\boldsymbol{z}, \alpha) = \frac{\prod_{\boldsymbol{\xi} \in \mathcal{Z}} z_{\boldsymbol{\xi}}!}{\prod_{\boldsymbol{\nu} \in \mathcal{X}} x_{\boldsymbol{\nu}}!} \prod_{\boldsymbol{\nu} \in \mathcal{X}} \left[\prod_{k=1}^{K} \alpha^{I(\nu_k=1)} (1 - \alpha)^{I(\nu_k=2)} \right]^{x_{\boldsymbol{\nu}}}
$$

141 Here $N = \sum_{\boldsymbol{\xi} \in \mathcal{Z}} z_{\boldsymbol{\xi}}^T$ represents the true population size. Note that the prod-
142 uct of these two contributions exactly reconstructs the single likelihood func-
143 tion for $M_{t\alpha}$ defined by Link et al. (2010, eqns 6 and 7).

144   *Dynamic Moves.* We can again generate a dynamic Markov basis by
145 selecting from a set of moves which add or remove errors from the current
146 configuration. An extra step is also needed on each iteration of the algorithm
147 to update the count of the unobserved individuals, $\boldsymbol{\nu_0}$.
148   Let $\chi_{\nu t}(\boldsymbol{x}) = \{\boldsymbol{\nu} : \nu_t = \nu \text{ and } x_{\boldsymbol{\nu}} > 0\}$ be the set of latent error histories
149 with event ν on occasion t and positive counts in $\boldsymbol{x}$. As with the CJS/BRE
150 model, we define $\mathcal{M}(\boldsymbol{x})$ as the union of two sets: $\mathcal{M}_1(\boldsymbol{x})$ containing moves

which add errors and $\mathcal{M}_2(\boldsymbol{x})$ containing moves which remove errors. Each move in the dynamic Markov basis modifies the counts of three latent error histories. Moves in $\mathcal{M}_1(\boldsymbol{x})$ are defined by sampling one history $\boldsymbol{\nu}_0 \in \chi_{0t}(\boldsymbol{x})$ for some t and are denoted by $\boldsymbol{b}(\boldsymbol{\nu}_0)$. The other two latent error histories, $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$, are defined by setting $\boldsymbol{\nu}_1 = \delta_t$ and $\boldsymbol{\nu}_2 = \boldsymbol{\nu}_0 + 2\delta_t$. The move $\boldsymbol{b}(\boldsymbol{\nu}_0)$ is then defined by setting

$$b_{\boldsymbol{\nu}}(\boldsymbol{\nu}_0) = \begin{cases} -1 & \text{if } \boldsymbol{\nu} = \boldsymbol{\nu}_0 \text{ or } \boldsymbol{\nu} = \boldsymbol{\nu}_1 \\ 1 & \text{if } \boldsymbol{\nu} = \boldsymbol{\nu}_2 \end{cases}.$$

Similarly, moves in $\mathcal{M}_2(\boldsymbol{x})$, denoted by $\boldsymbol{b}(\boldsymbol{\nu}_2)$, are defined by sampling a history $\boldsymbol{\nu}_2 \in \chi_{2t}(\boldsymbol{x})$ for some t, setting

$$\boldsymbol{\nu}_1 = \delta_t$$

and

$$\nu_{0s} = \begin{cases} 0 & s = t \\ \nu_{2s} & \text{otherwise} \end{cases} \quad t = 1, \ldots, T.$$

The move $\boldsymbol{b}(\boldsymbol{\nu}_2)$ is then defined by setting

$$b_{\boldsymbol{\nu}}(\boldsymbol{\nu}_2) = \begin{cases} -1 & \text{if } \boldsymbol{\nu} = \boldsymbol{\nu}_2 \\ 1 & \text{if } \boldsymbol{\nu} = \boldsymbol{\nu}_0 \text{ or } \boldsymbol{\nu} = \boldsymbol{\nu}_1 \end{cases}.$$

If we assume that the decision to add or remove an error are each chosen with probability .5 and that histories are sampled uniformly from $\chi_{0\cdot} = \bigcup_{t=1}^{T} \chi_{0t}$ and $\chi_{2\cdot} = \bigcup_{t=1}^{T} \chi_{2t}$, when adding or removing an error respectively, then the proposal densities for the moves $\boldsymbol{x}' = \boldsymbol{x}^{(k-1)} + \boldsymbol{b}(\boldsymbol{\nu}_0)$ and $\boldsymbol{x}' = \boldsymbol{x}^{(k-1)} + \boldsymbol{b}(\boldsymbol{\nu}_2)$ are given by

$$q(\boldsymbol{x}'|\boldsymbol{x}^{(k-1)}) = \frac{.5}{\#\chi_{0\cdot}(\boldsymbol{x}^{(k-1)}) \cdot \#\{t : \nu_{0t} = 1\}}$$

and

$$q(\boldsymbol{x}'|\boldsymbol{x}^{(k-1)}) = \frac{.5}{\#\chi_{2\cdot}(\boldsymbol{x}^{(k-1)}) \cdot \#\{t : \nu_{2t} = 2\}}.$$

As in the algorithm for the CJS/BRE, we retain $x^{(k-1)}$ with probability 1 if an empty set is encountered in one of these processes or if the selected move reduces any of the counts in $\boldsymbol{x}$ below zero. Details of the full algorithm for sampling from the posterior distribution of model $M_{t\alpha}$ are given in Algorithm 3.

Initialise $\boldsymbol{p}^{(0)}$ and α.
Initialise $\boldsymbol{x}^{(0)}$ so that $\boldsymbol{n} = \boldsymbol{A}\boldsymbol{x}^{(0)}$ and set $\boldsymbol{z}^{(0)} = \boldsymbol{B}\boldsymbol{x}^{(0)}$.
Set $k = 1$.

1. Update $\boldsymbol{p}$ and α conditional on $\boldsymbol{x}^{(k-1)}$ and $\boldsymbol{z}^{(k-1)}$. Call the results $\boldsymbol{p}^{(k)}$ and $\alpha^{(k)}$.

2. Update $\boldsymbol{x}$ and $\boldsymbol{z}$ conditional on and $\boldsymbol{p}^{(k)}$ and $\alpha^{(k)}$ as follows.

   i) With probability .5 sample $\boldsymbol{b}$ from $\mathcal{M}_1(\boldsymbol{x}^{(k-1)})$. If $\mathcal{M}_1(\boldsymbol{x}^{(k-1)}) = \emptyset$ then set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)}$ and continue to step v).
   Otherwise sample $\boldsymbol{b}$ from $\mathcal{M}_2(\boldsymbol{x}^{(k-1)})$. If $\mathcal{M}_2(\boldsymbol{x}^{(k-1)}) = \emptyset$ then set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)}$ and continue to step v).

   ii) Set $\boldsymbol{x}^{\mathrm{prop}} = \boldsymbol{x}^{(k-1)} + \boldsymbol{b}$. If $x_j' < 0$ for any $j = 1, \ldots, J$ set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)}$ and continue to step v).

   iii) Calculate the Metropolis acceptance probability:

   $$r(\boldsymbol{x}, \boldsymbol{x}^{\mathrm{prop}}|\boldsymbol{p}^{(k)}, \alpha) = \min\left\{1, \frac{\pi(\boldsymbol{x}^{\mathrm{prop}}|\boldsymbol{n}, \boldsymbol{p}^{(k)}, \alpha)}{\pi(\boldsymbol{x}^{(k-1)}|\boldsymbol{n}, \boldsymbol{p}^{(k)}, \alpha)} \cdot \frac{q(\boldsymbol{x}^{(k-1)}|\boldsymbol{x}^{\mathrm{prop}})}{q(\boldsymbol{x}^{\mathrm{prop}}|\boldsymbol{x}^{(k-1)})}\right\}.$$

   iv) Set $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{\mathrm{prop}}$ with probability $r(\boldsymbol{x}, \boldsymbol{x}^{\mathrm{prop}}|\boldsymbol{\phi}^{(k)}, \boldsymbol{p}^{(k)}, \alpha)$ and $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)}$ otherwise.

   v) Set $\boldsymbol{z}^{(k)} = \boldsymbol{B}\boldsymbol{x}^{(k)}$

3. Increment k.

Algorithm 3: Proposed algorithm for sampling from the posterior distribution of Model $M_{t\alpha}$ using the dynamic Markov basis.

References.

CINLAR, E. (1975). *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, NJ.

LINK, W. A., CAM, E., NICHOLS, J. D. and COOCH, E. G. (2002). Of BUGS and Birds: Markov chain Monte Carlo for hierarchical modeling in wildlife research. *Journal of Wildlife Management* **66** 277–291.

LINK, W. A., YOSHIZAKI, J., BAILEY, L. L. and POLLOCK, K. H. (2010). Uncovering a latent multinomial: analysis of mark-recapture data with misidentification. *Biometrics* **66** 178–185.

LIU, J. S. (2004). *Monte Carlo Strategies in Scientific Computing*. Springer, New York, NY.

ROBERT, C. P. and CASELLA, G. (2010). *Monte Carlo Statistical Methods*, 2nd ed. Springer-Verlag, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF KENTUCKY
LEXINGTON, KENTUCKY 40536-0082 USA
E-MAIL: simon.bonner@uky.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF OTAGO
DUNEDIN 90504, NEW ZEALAND
E-MAIL: mschofield@maths.otago.ac.nz

INSTITUTE OF SCIENCE AND TECHNOLOGY AUSTRIA
AM CAMPUS 1
3400 KLOSTERNEUBURG, AUSTRIA
E-MAIL: patrik.noren@ist.ac.at

DEPARTMENT OF FORESTRY
UNIVERSITY OF KENTUCKY
LEXINGTON, KENTUCKY 40546-0073 USA
E-MAIL: steven.price@uky.edu