



2016

# Extending the Latent Multinomial Model with Complex Error Processes and Dynamic Markov Bases

Simon J. Bonner

*University of Western Ontario, Canada*

Matthew R. Schofield

*University of Otago, New Zealand*

Patrik Noren

*North Carolina State University*

Steven J. Price

*University of Kentucky, [steven.price@uky.edu](mailto:steven.price@uky.edu)*

**Right click to open a feedback form in a new tab to let us know how this document benefits you.**

Follow this and additional works at: [https://uknowledge.uky.edu/forestry\\_facpub](https://uknowledge.uky.edu/forestry_facpub)



Part of the [Applied Statistics Commons](#), and the [Forest Sciences Commons](#)

## Repository Citation

Bonner, Simon J.; Schofield, Matthew R.; Noren, Patrik; and Price, Steven J., "Extending the Latent Multinomial Model with Complex Error Processes and Dynamic Markov Bases" (2016). *Forestry and Natural Resources Faculty Publications*. 18.

[https://uknowledge.uky.edu/forestry\\_facpub/18](https://uknowledge.uky.edu/forestry_facpub/18)

This Article is brought to you for free and open access by the Forestry and Natural Resources at UKnowledge. It has been accepted for inclusion in Forestry and Natural Resources Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

---

**Extending the Latent Multinomial Model with Complex Error Processes and Dynamic Markov Bases**

**Notes/Citation Information**

Published in *The Annals of Applied Statistics*, v. 10, no. 1, p. 246-263.

© Institute of Mathematical Statistics, 2016

The copyright holder has granted the permission for posting the article here.

**Digital Object Identifier (DOI)**

<https://doi.org/10.1214/15-AOAS889>

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

BY SIMON J. BONNER<sup>\*</sup>, MATTHEW R. SCHOFIELD<sup>†</sup>,  
PATRIK NOREN<sup>‡</sup> AND STEVEN J. PRICE<sup>§</sup>

*University of Western Ontario<sup>\*</sup>, University of Otago<sup>†</sup>,  
North Carolina State University<sup>‡</sup> and University of Kentucky<sup>§</sup>*

The latent multinomial model (LMM) of Link et al. [*Biometrics* **66** (2010) 178–185] 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 [*Biometrics* **71** (2015) 1070–1080] shows that this is sufficient for some models within the framework but that a larger set called a Markov basis is required when errors are more complex. 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 (2015)], 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 introduced by Link et al. (2010). Key to this approach was a novel Markov chain Monte Carlo

---

Received April 2015; revised September 2015.

*Key words and phrases.* Bayesian inference, Markov basis, Markov chain Monte Carlo, mark-recapture, misidentification, queen snake (*Regina septemvittata*).

(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 posterior distribution. We described an extended MCMC algorithm incorporating Markov bases to address this issue. As one example, we presented simulation 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, assuming 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  $M_{I\alpha}$  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  $n_{\text{Tot}}$  be the number of distinct individuals captured. The raw data consist of  $n_{\text{Tot}}$  vectors in  $\{0, 1\}^T$ , called the capture histories, such that  $\omega_{it} = 1$  if individual  $i$  was captured on occasion  $t$ . If  $T = 5$  then the history  $\omega_i = 01010$  indicates that the  $i$ th 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,  $\mathbf{n}$ , recording the number of times each unique history was observed is sufficient. The likelihood is then defined by the distribution of  $\mathbf{n}$ .

When errors occur the distribution of  $\mathbf{n}$  may be difficult to compute directly. To make the likelihood tractable, the LMM introduces a set of  $J > I$  latent histories identifying the true captures for each individual and describing what errors occurred. Let  $\mathbf{x}$  be the unobserved  $J$ -vector of counts for the latent histories which is modelled by  $f(\mathbf{x}|\boldsymbol{\theta})$ <sup>1</sup> for some vector of parameters,  $\boldsymbol{\theta}$ . The likelihood can then be computed by summing  $f(\mathbf{x}|\boldsymbol{\theta})$  over all values of  $\mathbf{x}$  consistent with  $\mathbf{n}$ . In particular,

---

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

the LMM assumes that  $\mathbf{n}$  is a linear function of  $\mathbf{x}$  so that  $\mathbf{n} = \mathbf{Ax}$  for some known  $I \times J$  matrix  $\mathbf{A}$  and

$$(1) \quad f(\mathbf{n}|\boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathbb{N}^J} \mathbb{1}(\mathbf{n} = \mathbf{Ax}) f(\mathbf{x}|\boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathcal{F}_{\mathbf{n}}} f(\mathbf{x}|\boldsymbol{\theta}),$$

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

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

- (0) indicating that the individual was not captured,
- (1) indicating that the individual was captured and correctly identified, and
- (2) indicating that the individual was captured and incorrectly identified.

For example, the latent history  $\mathbf{v}_i = 01020$  indicates that individual  $i$  was captured and correctly identified on occasion 2 and recaptured and misidentified on occasion 4. It is assumed that  $\mathbf{x}|\mathbf{p}, \alpha$  follows a multinomial distribution conditional on the true population size,  $N$ , with cell probability corresponding to the latent history  $\mathbf{v}$  computed as

$$f(\mathbf{v}|\mathbf{p}, \alpha) = \prod_{t=1}^T [p_t^{\mathbb{1}(v_t>0)} (1 - p_t)^{\mathbb{1}(v_t=0)} \alpha^{\mathbb{1}(v_t=1)} (1 - \alpha)^{\mathbb{1}(v_t=2)}],$$

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

Although equation (1) makes it easier to compute  $f(\mathbf{n}|\boldsymbol{\theta})$  in theory,  $\mathcal{F}_{\mathbf{n}}$  is often so large that exact computation is not practical. Instead, Link et al. (2010) proposed a Bayesian solution implemented by sampling from the joint posterior distribution of  $\mathbf{x}$  and  $\boldsymbol{\theta}$ . The specific MCMC algorithm uses a block Metropolis–Hastings (MH) approach and the main challenge lies in constructing proposals of  $\mathbf{x}|\boldsymbol{\theta}, \mathbf{n}$  which fall inside the fibre. The algorithm starts by defining a lattice basis for the kernel of  $\mathbf{A}$ ; that is, a linearly independent set  $\mathcal{B} = \{b_1, \dots, b_K\}$  such that  $\forall \mathbf{b} \in \ker(\mathbf{A}) \cap \mathbb{Z}^J$  there exists  $c_1, \dots, c_K \in \mathbb{Z}$  so that

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

A proposal,  $\mathbf{x}^{\text{prop}}$ , is then constructed by adding an integer multiple of one of the basis vectors to the current value of  $\mathbf{x}$ . The magic of this approach is that any proposal is guaranteed to satisfy the linear constraint and to have integer entries. Note, that  $\mathbf{x}^{\text{prop}}$  may still fall outside  $\mathcal{F}_{\mathbf{n}}$  since there is no guarantee that  $x_j^{\text{prop}} \geq 0$  for every  $j$ .

Link et al. (2010) implied that Markov chains constructed with this algorithm would connect all elements in  $\mathcal{F}_n$  and hence be irreducible. In Schofield and Bonner (2015), we showed that this is true for model  $M_{I\alpha}$ , provided that the right lattice basis is chosen, and extended this result to a broader class of models containing what we called simple corruptions. However, we also provided examples of more complicated models for which the algorithm does not produce irreducible Markov chains, including the BRE model. The central problem is that some pairs of elements in  $\mathcal{F}_n$  may be connected by this algorithm only by passing through intermediate configurations containing negative entries. These elements have zero probability under the posterior, and so the chain will never follow these paths.

Irreducible chains can always be produced by adding linear combinations of all elements in  $\mathcal{B}$  simultaneously, but the resulting proposals are likely to fall outside of  $\mathcal{F}_n$  and Diaconis and Sturmfels (1998) reported that this method is not efficient. Instead, Diaconis and Sturmfels (1998) suggested using the one-at-a-time algorithm but drawing the elements from a larger subset  $\mathcal{M} \subset \ker(\mathbf{A})$  chosen to ensure that it is possible to move between any two elements of  $\mathcal{F}_n$ . Diaconis and Sturmfels (1998) called  $\mathcal{M}$  a Markov basis and the elements of  $\mathcal{M}$  moves, and provided methodology for computing this set based on the theory of toric ideals. For simplicity, we consider the special case of the algorithm presented in Schofield and Bonner (2015) in which one element is selected from  $\mathcal{M}$  on each iteration of the MCMC algorithm and either added to or subtracted from the current configuration without a multiplier. Details are provided in Algorithm 1.

---

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

---

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

1. Update  $\theta$  conditional on  $\mathbf{x}^{(k-1)}$ . Call the result  $\theta^{(k)}$ .
2. Update  $\mathbf{x}$  conditional on  $\theta^{(k)}$ .
  - (i) Sample  $\mathbf{b}$  uniformly from  $\mathcal{M}$  and  $c$  uniformly from  $\{-1, 1\}$ .
  - (ii) Set  $\mathbf{x}^{\text{prop}} = \mathbf{x}^{(k-1)} + c\mathbf{b}$ .
  - (iii) Calculate the Metropolis acceptance probability:

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

- where  $q(\mathbf{x}'|\mathbf{x})$  is the probability of proposing  $\mathbf{x}'$  given the current state  $\mathbf{x}$ .
- (iv) Set  $\mathbf{x}^{(k)} = \mathbf{x}^{\text{prop}}$  with probability  $r$ . Otherwise, set  $\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)}$ .
  3. Increment  $k$ .
-

**3. Data.** As an example, we consider data from a study of queen snakes conducted in Jessamine County, Kentucky. An initial sample of 61 snakes was captured and marked in the fall of 2013 and a second sample of 41 snakes 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 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  $\phi_t$  to denote the capture and survival probabilities.

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

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

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

Assumption 3 contrasts directly with the assumptions of model  $M_{t\alpha}$  and is justified by the differences between man-made marks and natural marks. Model  $M_{t\alpha}$  is intended for use with natural marks including genotypes and pigmentation patterns. The set of possible natural marks is usually unknown and the number of possible marks is so large that it is unlikely for an error to reproduce the identity of another individual exactly. On the other hand, the BRE model is intended for use with man-made marks. The set of possible marks is known when using man-made marks, and this means that erroneous sightings of marks which have never been released can be detected and removed from the data prior to the analysis. The only errors that cannot be detected occur when one marked individual is mistaken for another marked individual. The fourth assumption is not realistic, but simplifies the model and we will work to relax this in the future. We present the likelihood for this model in Section 5.1 after introducing the extended modelling framework.

## 5. Methods.

5.1. *Extended framework.* The first challenge is that the CJS/BRE model does not fit easily in the framework of the LMM. Link et al. (2010) focused on models, like  $M_{t\alpha}$ , for which  $\mathbf{x}$  follows a multinomial distribution. Although they suggested that the methods could be applied more generally, examples were not provided. The CJS/BRE model does not result in a multinomial distribution for  $\mathbf{x}$ , and it is difficult to determine the density of  $\mathbf{x}$  explicitly.

To address this, we extend the LMM to include a second vector of latent counts. This allows the mark-recapture process and the error mechanism to be modelled separately. Suppose, for example, that an experiment has  $T = 2$  occasions and individual  $i$  is captured on both occasions, correctly identified on the first occasion, and identified as an entirely new individual on the second occasion (this is the error mechanism for model  $M_{t\alpha}$ ). In the terminology of Link et al. (2010), individual  $i$  would have latent history  $\mathbf{v}_i = 12$  and would produce the recorded histories  $\omega_{i1} = 10$  and  $\omega_{i2} = 01$ . The original LMM assigns probabilities to the latent histories,  $\mathbf{v}_i$ , directly by simultaneously modelling the capture and error processes. Our formulation introduces a second latent history,  $\xi_i$ , identifying the occasions on which the individual was truly captured but ignoring the errors. The new latent history would be  $\xi_i = 11$  since the individual was truly captured on both occasions. We then model the joint distribution of  $\mathbf{v}_i$  and  $\xi_i$  by assigning probabilities



first to  $\xi_i$  and second to  $\nu_i$  given  $\xi_i$ . We distinguish between the two sets of latent histories by calling  $\nu_i$  the latent error history and  $\xi_i$  the latent capture history.

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

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

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,  $\mathbf{n}$ ,  $\mathbf{x}$ , and  $\mathbf{z}$ , both by index and by the corresponding history. For example,  $n_i$  represents the count for the  $i$ th element of  $\mathbf{n}$  using some implicit ordering while  $n_\omega$  represents the count of history  $\omega$ . This allows us to define sums in two equivalent ways as either  $n_{\text{Tot}} = \sum_{i=1}^I n_i$  or  $n_{\text{Tot}} = \sum_{\omega \in \{0,1\}^T / \mathbf{0}} n_\omega$ . 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  $i$ th 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  $\mathbf{x}$  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  $t = 2, \dots, T$ . The  $\mathbf{A}$  matrix is constructed as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix},$$

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

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

and

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

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

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

Finally, we define the distributions of  $\mathbf{z}$  and  $\mathbf{x}|\mathbf{z}$ . For the CJS/BRE model  $\theta_1 = \{\phi, \mathbf{p}\}$  and  $\theta_2 = \{\alpha\}$ . Let  $a_t$  denote the number of individuals first captured and marked on occasion  $t$ ,  $M_t$  the number of individuals marked before occasion  $t$ , and  $m_t$  the number of these individuals resighted on occasion  $t$ . Then  $\mathbf{z}$  is a product multinomial random variable with density

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

where  $f(\xi_k|\phi, \mathbf{p})$  denotes the probability assigned to history  $\xi_k$  by the standard CJS model. To construct the second component of the likelihood we consider occasions  $t = 2, \dots, T$  separately first modelling the number of errors that occur,  $e_t(\mathbf{x}) = \sum_{j=1}^J x_j \mathbb{1}(v_{jt} = 2) = \sum_{j=1}^J x_j \mathbb{1}(v_{jt} = 3)$ , and then modelling the exact configuration of false positives and false negatives given  $\mathbf{e} = (e_1, \dots, e_T)'$  to obtain a specific configuration of  $\mathbf{x}$ . Under the assumptions in Section 4,  $e_t \leq m_t^* = \min(m_t, M_t - m_t)$  and follows the (possibly) truncated binomial with density

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

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

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

The second component of the likelihood is

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

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

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

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

We avoid this problem by using dynamic Markov bases. Dobra (2012) defined a dynamic Markov basis to be a collection of sets of local moves,  $\mathcal{M}(\mathbf{x})$ , which connect each  $\mathbf{x} \in \mathcal{F}_{\mathbf{n}}$  to a relatively small number of neighbours. A proposal is generated on the  $k$ th iteration of the MCMC algorithm by sampling a move from  $\mathcal{M}(\mathbf{x}^{(k-1)})$ . This avoids the need to compute the entire Markov basis a priori. The method Dobra (2012) described for constructing  $\mathcal{M}(\mathbf{x})$  applies generally to all models with constraints of the form  $\mathbf{n} = \mathbf{A}\mathbf{x}$  and, possibly, further bounds on the counts in individual cells. In short, local moves are generated by permuting the counts in a random selection of “free” cells—entries of  $\mathbf{x}$  whose values are not

fixed by the information in  $\mathbf{n}$  and the cell bounds. The method we describe is specific to the mark-recapture 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  $\mathcal{F}_{\mathbf{n}}$  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

$$\mathcal{X}_{v_t}(\mathbf{x}) = \{ \mathbf{v} : v_s = 1 \text{ for some } s < t, v_t = v, \text{ and } x_v > 0 \}$$

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

1.  $\mathbf{v}_1 \in \mathcal{X}_{1 \cdot}(\mathbf{x}) = \bigcup_{t=2}^T \mathcal{X}_{1t}(\mathbf{x})$ ,
2.  $s \in \{t : v_{1t} = 1\}$ ,
3.  $\mathbf{v}_0 \in \mathcal{X}_{0s}(\mathbf{x})$ ,

all uniformly, and setting

$$v_{2t} = \begin{cases} 2, & \text{if } t = s, \\ v_{0t}, & \text{otherwise} \end{cases} \quad \text{and} \quad v_{3t} = \begin{cases} 3, & \text{if } t = s, \\ v_{1t}, & \text{otherwise.} \end{cases}$$

More compactly,  $\mathbf{v}_2 = \mathbf{v}_0 + 2\delta_t$  and  $\mathbf{v}_3 = \mathbf{v}_1 + 2\delta_t$  where  $\delta_t$  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

$$b_{\mathbf{v}}^+(\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \begin{cases} -1, & \text{if } \mathbf{v} = \mathbf{v}_0 \text{ or } \mathbf{v} = \mathbf{v}_1, \\ 1, & \text{if } \mathbf{v} = \mathbf{v}_2 \text{ or } \mathbf{v} = \mathbf{v}_3, \\ 0, & \text{otherwise.} \end{cases}$$

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

1.  $\mathbf{v}_2 \in \mathcal{X}_{2 \cdot}(\mathbf{x}) = \bigcup_{t=1}^T \mathcal{X}_{2t}(\mathbf{x})$ ,
2.  $s \in \{t : v_{2t} = 2\}$ ,
3.  $\mathbf{v}_3 \in \mathcal{X}_{3s}(\mathbf{x})$

and setting  $\mathbf{v}_0 = \mathbf{v}_2 - 2\delta_t$  and  $\mathbf{v}_1 = \mathbf{v}_3 - 2\delta_t$ . The corresponding move has entries

$$b_{\mathbf{v}}^-(\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \begin{cases} 1, & \text{if } \mathbf{v} = \mathbf{v}_0 \text{ or } \mathbf{v} = \mathbf{v}_1, \\ -1, & \text{if } \mathbf{v} = \mathbf{v}_2 \text{ or } \mathbf{v} = \mathbf{v}_3, \\ 0, & \text{otherwise.} \end{cases}$$

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

$$(3) \quad q(\mathbf{x}^{\text{prop}}|\mathbf{x}^{(k-1)}) = \frac{0.5}{\#\chi_{1\cdot}(\mathbf{x}^{(k-1)}) \cdot \#\{t : v_{1t} = 1\} \cdot \#\chi_{0s}(\mathbf{x}^{(k-1)})}$$

and when removing an error is

$$(4) \quad q(\mathbf{x}^{\text{prop}}|\mathbf{x}^{(k-1)}) = \frac{0.5}{\#\chi_{2\cdot}(\mathbf{x}^{(k-1)}) \cdot \#\{t : v_{2t} = 2\} \cdot \#\chi_{3s}(\mathbf{x}^{(k-1)})},$$

where  $\#\mathcal{S}$  denotes the cardinality of  $\mathcal{S}$ . If we propose to add an error and  $\mathbf{x}^{(k-1)}$  contains no errors,  $\chi_{3\cdot}(\mathbf{x}^{(k-1)}) = \emptyset$ , or if we propose to remove an error and  $\mathbf{x}$  contains no correct identifications,  $\chi_{3\cdot}(\mathbf{x}^{(k-1)}) = \emptyset$ , then we set  $\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)}$  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  $\mathbf{x}$ ,  $\phi$ , and  $\mathbf{p}$  is provided in the supplementary material (Supplement A [Bonner et al. (2015)]).

**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 each data set using the methods described in Section 5 (Model 3). One hundred simulated data sets were generated for each value of  $\alpha$ .

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 across all occasions but constrained survival to be equal on the final eight occasions. The estimated survival probabilities were  $\hat{\phi}_1 = 0.66$ ,  $\hat{\phi}_2 = 1.00$ , and  $\hat{\phi}_3 = \dots = \hat{\phi}_9 = 0.93$ . Clearly the overwintering survival rate,  $\phi_1$ , 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  $\phi_1$  that is almost unbiased and 95% CIs with coverage above the nominal value. Fitting

---

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

---

Initialise  $\phi^{(0)}$  and  $\mathbf{p}^{(0)}$ .

Initialise  $\mathbf{x}^{(0)}$  so that  $\mathbf{n} = \mathbf{A}\mathbf{x}^{(0)}$  and set  $\mathbf{z}^{(0)} = \mathbf{B}\mathbf{x}^{(0)}$ .

Set  $k = 1$ .

1. Update  $\phi$  and  $\mathbf{p}$  conditional on  $\mathbf{z}^{(k-1)}$ . Call the results  $\phi^{(k)}$  and  $\mathbf{p}^{(k)}$ .
2. Update  $\mathbf{x}$  and  $\mathbf{z}$  conditional on  $\phi^{(k)}$  and  $\mathbf{p}^{(k)}$  as follows.
  - (i) With probability 0.5 sample  $\mathbf{b}$  from  $\mathcal{M}_1(\mathbf{x}^{(k-1)})$ . If  $\mathcal{M}_1(\mathbf{x}^{(k-1)}) = \emptyset$  then set  $\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)}$  and continue to step (v).  
 Otherwise sample  $\mathbf{b}$  from  $\mathcal{M}_2(\mathbf{x}^{(k-1)})$ . If  $\mathcal{M}_2(\mathbf{x}^{(k-1)}) = \emptyset$  then set  $\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)}$  and continue to step (v).
  - (ii) Set  $\mathbf{x}^{\text{prop}} = \mathbf{x}^{(k-1)} + \mathbf{b}$ .
  - (iii) Calculate the Metropolis acceptance probability:

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

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

3. Increment  $k$ .

---

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  $\phi_1$  was only 47%. The bias increased one and a half times to 0.24 (36%) and the coverage dropped to only 6% when  $\alpha$  decreased to 6/9. In comparison, the posterior mean of  $\phi_1$  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  $\phi_2$  that were biased because the true parameter lies on the boundary of the parameter space. Coverage of this parameter was zero for all models and is not reported. The posterior mean of  $\phi_2$  from Model 3 was significantly more biased than that of Model 1, underestimating  $\phi_2$  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

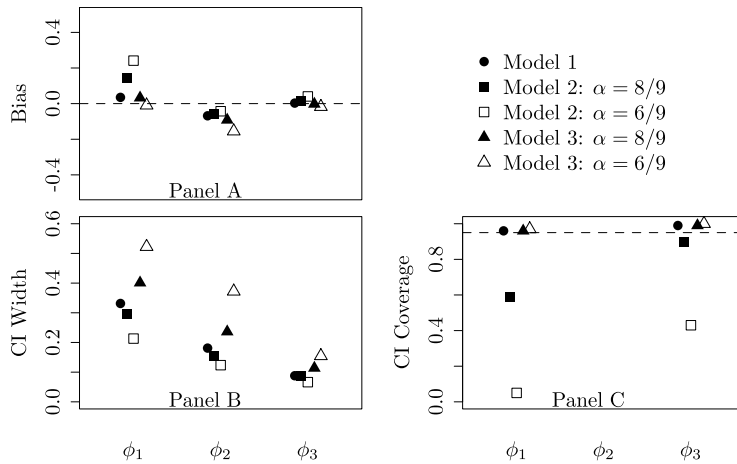


FIG. 1. Results of the analysis of the queen snake data. The three panels present the estimated bias of the posterior means (Panel A) and the estimated width (Panel B) and coverage probability (Panel C) of the 95% credible interval for the survival probabilities for the three models described in Section 6. The different models are indicated by the shape of the plotting symbol. The rates of error for Models 2 and 3 are indicated by the colour of the symbol. Coverage of  $\phi_2$  is not reported because the true parameter lies on the boundary of the parameter space.

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  $\phi_1 = \phi_2 = \phi_3 = 0.8$ , constant capture probability  $p_2 = p_3 = p_4 = 0.5$ , and error rate  $\alpha = 0.5$ . Samples from the joint posterior distribution of  $\mathbf{x}$  and  $\mathbf{z}$  were then drawn using Step 1 of Algorithm 1 and Step 2 of Algorithm 2 while keeping  $\phi$  and  $\mathbf{p}$  fixed at their true values.

We assessed how well the chains mixed by comparing the acceptance rates and the number of unique solutions for  $\mathbf{x}$  identified. The chain constructed using Algorithm 1 identified a total of 79 unique configurations among the 7500 values of  $\mathbf{x}$  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 were always less than 50%. In comparison, estimates from the CJS/BRE model were almost

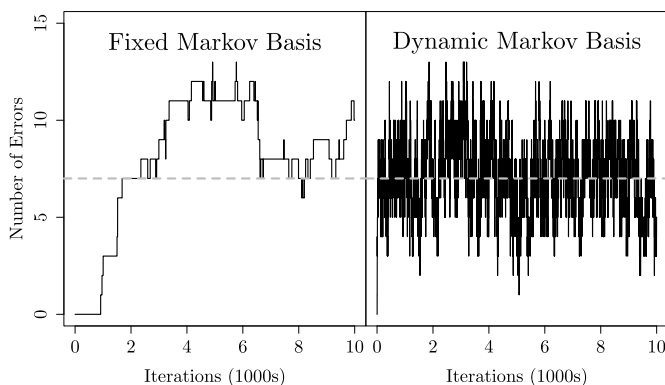


FIG. 2. Comparison of the chains sampling from the posterior distribution of the CJS/BRE model applied to the simulated data. The figures trace the number of errors in  $\mathbf{x}$  for the algorithms using the fixed basis (left) and dynamic basis (right). The grey dotted lines represent the true number of errors in the data set.

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 a 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 the supplementary material (Supplement B [Bonner et al. (2015)]). 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  $\mathcal{F}_{\mathbf{n}}$  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  $\mathcal{F}_{\mathbf{n}}$ . 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  $\mathcal{F}_{\mathbf{n}}$  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  $M \in \{1, 2, \dots, J\}$  according to some density  $g(\cdot)$ , randomly sampling a permutation function  $\delta(\cdot)$  from the set of all permutations of the indices of  $\mathbf{x}$ , and sequentially resampling the counts in the first  $M$  cells in  $\delta(\mathbf{x})$  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  $\xi_{0t}(\mathbf{x})$ ,  $\xi_{1t}(\mathbf{x})$ ,  $\xi_{2t}(\mathbf{x})$ , and  $\xi_{3t}(\mathbf{x})$ , 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  $\mathcal{F}_{\mathbf{n}}$ . 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 \(2002\)](#), page 196]:

1. Each individual alive on occasion  $t$  survives to occasion  $t + 1$  with probability  $\phi_t$ .
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

$$f(\omega|\boldsymbol{\phi}, \mathbf{p}) = \phi_2(1 - p_3)\phi_3p_4(\phi_4(1 - p_5) + (1 - \phi_4)).$$

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

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

#### APPENDIX B: NOTATION

*Algebraic Statistics and Markov Bases:*

$\mathcal{F}_{\mathbf{n}}$       $\mathbf{n}$ -fibre,  $\mathcal{F}_{\mathbf{n}} = \{\mathbf{x} \in \mathbb{N}^J : \mathbf{n} = \mathbf{A}\mathbf{x}\}$ .

$\mathcal{B}$        Lattice basis for  $\ker(\mathbf{A})$ .

$\mathcal{M}$        Markov basis for  $\ker(\mathbf{A})$ .

$\mathcal{M}(\mathbf{x})$    Dynamic Markov basis for  $\ker(\mathbf{A})$  computed at  $\mathbf{x}$ .

*Extended LMM:*

- $n_{\text{Tot}}$  Number of distinct individuals captured and marked.  
 $\omega_i$  Observed capture history for the  $i$ th marked individual.  
 $\mathbf{v}_i$  Latent error history for the  $i$ th marked individual.  
 $\xi_i$  Latent capture history for the  $i$ th marked individual.  
 $\mathbf{n}$  Observed vector of counts for the observable histories (indexed by either  $i$  and  $\omega$ ).  
 $\mathbf{x}$  Unknown vector of counts for the latent error histories (indexed by either  $j$  and  $\mathbf{v}$ ).  
 $\mathbf{z}$  Unknown vector of counts for the latent capture histories (indexed by either  $k$  and  $\xi$ ).  
 $I$  Length of  $\mathbf{n}$ . For the CJS/BRE model  $I = 2^T - 2$ .  
 $J$  Length of  $\mathbf{x}$ . For the CJS/BRE model  $J = (4^T - 1)/3 - 1$ .  
 $K$  Length of  $\mathbf{z}$ . For the CJS/BRE model  $K = 2^T - 2$ .  
 $\mathbf{A}$   $I \times J$  matrix mapping  $\mathbf{x}$  onto  $\mathbf{n}$ ,  $\mathbf{n} = \mathbf{A}\mathbf{x}$ .  
 $\mathbf{B}$   $K \times J$  matrix mapping  $\mathbf{x}$  onto  $\mathbf{z}$ ,  $\mathbf{z} = \mathbf{B}\mathbf{x}$ .  
 $\theta_1$  Parameters in the model of  $\mathbf{z}$ .  
 $\theta_2$  Parameters in the conditional model of  $\mathbf{x}$  given  $\mathbf{z}$ .

*Band-Read Error Model:*

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

**Acknowledgments.** We are grateful for the input of the Associate Editor and two reviewers whose constructive comments greatly improved this manuscript.

## SUPPLEMENTARY MATERIAL

**Supplement to “Extending the latent multinomial model with complex error processes and dynamic Markov bases”** (DOI: [10.1214/15-AOAS889SUPP](https://doi.org/10.1214/15-AOAS889SUPP); .pdf). Supplement A: Proof of convergence. Proof that the chains generated by Algorithm 2 converge to the correct distribution. Supplement B: Model  $M_{t\alpha}$ . Application 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 capture–recapture modeling of samples drawn one at a time. *Biometrics* **70** 775–782. [MR3295738](#)
- BONNER, S. J. and HOLMBERG, J. (2013). Mark-recapture with multiple, non-invasive marks. *Biometrics* **69** 766–775. [MR3106605](#)
- BONNER, S. J., SCHOFIELD, M. R., NOREN, P. and PRICE, S. (2016). Supplement to “Extending the latent multinomial model with complex error processes and dynamic Markov bases.” DOI:10.1214/15-AOAS889SUPP.
- COWEN, L. and SCHWARZ, C. J. (2006). The Jolly–Seber model with tag loss. *Biometrics* **62** 699–705. [MR2247196](#)
- DIACONIS, P. and STURMFELS, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.* **26** 363–397. [MR1608156](#)
- DOBRA, A. (2012). Dynamic Markov bases. *J. Comput. Graph. Statist.* **21** 496–517. [MR2945478](#)
- 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. [MR2756704](#)
- LUKACS, P. M. and BURNHAM, K. P. (2005). Estimating population size from DNA-based closed capture–recapture data incorporating genotyping error. *J. Wildl. Manag.* **69** 396–403.
- MCCLINTOCK, B. T., CONN, P. B., ALONSO, R. S. 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. *J. Herpetol.* **34** 481–485.
- SCHOFIELD, M. and BONNER, S. (2016). Connecting the latent multinomial. *Biometrics*. **17** 1070–1080.
- 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.
- WILLIAMS, B. K., NICHOLS, J. D. and CONROY, M. J. (2002). *Analysis and Management of Animal Populations*. Academic Press, San Diego, CA.
- WINNE, C. T., WILLSON, J. D., ANDREWS, K. M. and REED, R. N. (2006). Efficacy of marking snakes with disposable medical cautery units. *Herpetological Review* **37** 52–54.
- WRIGHT, J. A., BARKER, R. J., SCHOFIELD, M. R., FRANTZ, A. C., BYROM, A. E. and GLEESON, D. M. (2009). Incorporating genotype uncertainty into mark-recapture-type models for estimating abundance using DNA samples. *Biometrics* **65** 833–840. [MR2649856](#)
- YOSHIZAKI, J., BROWNIE, C., POLLOCK, K. H. and LINK, W. A. (2011). Modeling misidentification errors that result from use of genetic tags in capture–recapture studies. *Environ. Ecol. Stat.* **18** 27–55. [MR2783681](#)

YOSHIKAZI, J., POLLOCK, K. H., BROWNIE, C. and WEBSTER, R. A. (2012). Modeling misidentification errors in capture–recapture studies using photographic identification of evolving marks. *Ecology* **90** 3–9.

S. J. BONNER  
DEPARTMENT OF STATISTICAL  
AND ACTUARIAL SCIENCES  
UNIVERSITY OF WESTERN ONTARIO  
LONDON, ONTARIO N6A 5B7  
CANADA  
E-MAIL: [sbonner6@uwo.ca](mailto:sbonner6@uwo.ca)

P. NOREN  
DEPARTMENT OF MATHEMATICS  
NORTH CAROLINA STATE UNIVERSITY  
BOX 8205  
RALEIGH, NORTH CAROLINA 27695  
USA  
E-MAIL: [pgnoren2@ncsu.edu](mailto:pgnoren2@ncsu.edu)

M. R. SCHOFIELD  
DEPARTMENT OF MATHEMATICS  
AND STATISTICS  
UNIVERSITY OF OTAGO  
PO BOX 56  
DUNEDIN 9054  
NEW ZEALAND  
E-MAIL: [mschofield@maths.otago.ac.nz](mailto:mschofield@maths.otago.ac.nz)

S. J. PRICE  
DEPARTMENT OF FORESTRY  
UNIVERSITY OF KENTUCKY  
LEXINGTON, KENTUCKY 40546  
USA  
E-MAIL: [steven.price@uky.edu](mailto:steven.price@uky.edu)