A Test for the Underlying State-Structure of Hidden Markov Models: Partially Observed Capture-Recapture Data

Anita Jeyam, Rachel Mccrea, Roger Pradel

frontiers

A test for the underlying state-structure of Hidden Markov models: partially observed capture-recapture data

Anita Jeyam [1], **Rachel McCrea** [1,*] **and Roger Pradel** [2]

[1] *Statistical Ecology @ Kent, National Centre for Statistical Ecology, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, UK*

[2] *CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France~~Centre d'Ecologie Functionelle et Evolutive, Montpellier, France~~*

Correspondence*:
R.S.McCrea@kent.ac.uk

ABSTRACT

Hidden Markov models (HMMs) are being widely used in the field of ecological modelling, however determining the number of underlying states in an HMM remains a challenge. Here we examine a special case of ~~partially observed~~ capture-recapture models for open populations, where some animals are observed but it is not possible to ascertain their state (partial observations), whilst the other animals' states are assigned without error (complete observations). We propose a mixture test of the underlying state structure generating the partial observations, which assesses whether they are compatible with the set of states ~~directly~~ observed in the complete observations~~capture-recapture experiment~~. We demonstrate the good performance of the test using simulation and through application to a data set of Canada Geese. ~~This paper provides a novel method to offer practical insight to a large class of HMM applications.~~

Keywords: Multievent model, Capture-recapture, Partial observations, Mixture of multinomials

1 INTRODUCTION

Besides its known use for the estimation of the size of a closed population (Bartolucci and Pennoni, 2007; Yang and Chao, 2005; Pledger, 2000) originating in the work of Otis *et al* (1978), capture-recapture is also a

widely used technique to follow the dynamics of open animal populations ~~:~~(Cormack, 1964; Williams *et al*, 2002). The protocol remains the same: animals are uniquely marked, then released and resighted/recaptured at subsequent sampling occasions. In ~~a~~the multi-state framework (Lebreton *et al*, 2009), at each occasion, individual animals' states are recorded upon resighting; if an animal is not seen at a given occasion, this is denoted by a 0. If it is seen, a code, commonly a number, specifies the state (see example data set in supplementary material). Hence, the data resulting from a multi-state capture-recapture experiment consists of individual encounter histories, formed by the series of records made for each animal. Multi-state models allow the estimation of the survival and transition probabilities of animals between the states, whilst accounting for imperfect detection. ~~However,~~ Within this modelling framework, states are assumed to be assigned without error (Kendall, 2004). However, this assumption can be unrealistic in certain situations such as the assessment of sex in a monomorphic species or of health status when biological testing is not possible in the field. Pradel (2005) developed multievent models ~~, which belong to the family of Hidden Markov Models (Zucchini *et al*, 2016)~~ to account for the uncertainty in state assignment. These models belong to the family of Hidden Markov Models (Zucchini *et al*, 2016) and distinguish the events, which are observed, from the states, which are underlying. ~~In this framework, events are observed whilst the states are underlying.~~ The process governing the transitions between states is Markovian (generally assumed of order 1) and the events are generated by the states. Multievent models have a structural absorbing state (death). Transitions are almost systematically time-dependent, which precludes the consideration that the system has reached an equilibrium. Also, because the chance that an individual is missed is state dependent, non-observations cannot be considered as data missing at random. They are informative events like any other outcome of the experiment.

In this paper we focus on a special case of multievent models, where, at a given occasion, the state cannot be ascertained for a proportion of the observed animals, leading to partial observations, whilst the underlying states are directly observable for the other observed animals (complete observations). In analysing this type of data, it is usually assumed that the range of potential states is limited to the set of states observed ~~directly~~in the complete observations (see Figure 1). However, some states may not be directly observable, yet capable of generating partial observations (see Figure 2). We propose a new diagnostic tool to assess whether the partial observations are consistent with being generated only by the directly observable states (H_0) or whether partial observations may be generated by at least one additional unidentified state never directly observed (H_1). For instance, in a study of movements, animals may move between the set of monitored sites, where observations are made, and an additional unmonitored site (see scenarios 2PO and 3PO of the Canada geese example below). ~~This is useful for defining the set of underlying states for the multievent model: are the directly observed states sufficient or do additional~~

~~latent states need to be defined~~? Such a test is currently lacking in the literature and pragmatic approaches need to be taken, see for example Pohle *et al* (2017).

Our test builds on the approach used by Pradel *et al* (2003) to construct a mixture test for the multi-state framework, as well as the sufficient statistics and likelihood components developed by King and McCrea (2014) for the special case of partial observations. Indeed, we show that if partial observations are generated only by the directly observable states, the number of animals ~~previously released,~~ partially observed at a given occasion i and re-observed later in a known state, follows a conditional multinomial distribution, which is a mixture of the conditional multinomial distributions followed by the number of animals released at occasion i in the observable states. Based on this mixture property, we then use usual goodness-of-fit measures to assess the fit of a model where only the directly observable states generate the partial observations.

We use simulation to empirically assess the test and apply it to a Canada Geese, *Branta canadensis*, dataset (Hestbeck *et al*, 1991), in which we artificially create partial observations. This demonstrates that the test can work well under practical settings and sample size.

2 PARTIALLY OBSERVED CAPTURE-RECAPTURE DATA AND MIXTURE PROPERTIES

Consider a capture-recapture experiment with T sampling occasions and R live states. If individuals are assigned to state r upon capture, this is done with certainty and the corresponding event is denoted by r: "observed in state r". When an individual's state cannot be determined, the corresponding event, a partial observation, is denoted by U: "observed with state unknown" and the animal can be in any one of the underlying R states.

The state and time-dependent parameters of the partial observation capture-recapture model (King and McCrea, 2014) are defined by:

- ϕ_t^r is the probability an individual in state r at time t survives until $t+1$, for $t = 1, \ldots, T-1$ and $r = 1, \ldots, R$.
- p_t^r is the probability of recapture at time t for an individual in state r, for $t = 2, \ldots, T$.
- $\psi_t^{r,s}$ is the probability an individual is in state s at time $t+1$ given that it was in state r at time t and is alive at $t+1$, for $t = 1, \ldots, T-1$, $r = 1, \ldots, R$ and $s = 1, \ldots, R$.
- α_t^r is the probability an individual is assigned to state r given it was recaptured at time t and in state r at that time, for $t = 2, \ldots, T$ and $r = 1, \ldots, R$. $\beta_t^r = 1 - \alpha_t^r$ is then defined as the probability an individual is assigned as unknown (U) at time t given the individual is recaptured, and in state r at this

time, for $t = 2, \ldots, T$ and $r = 1, \ldots, R$. An animal is either assigned to the correct state or unassigned but there are no assignment error.

- π_t^r is the initial state probability of individuals in an unknown state when first observed. This corresponds to the probability an individual is in state r at time t, given it was first observed in U at t, for $t = 1, \ldots, T-1$.

The sufficient statistics are based on partitioning the encounter histories (EH) into the following pieces: the EH between observations in two known states; the EH between first observation in unknown state and first re-observation in a known state; the EH following the last observation in a known state; and the EH following the first observation in an unknown state, for animals who are never seen in a known state (Table 1 provides examples). We define the following sufficient statistics:

- $n_{t_1,t_2+1}^{r,z_{(t_1+1):(t_2)},s}$ denotes the number of animals observed at time t_1 in known state r, next observed in known state s at $t_2 + 1$ with partial capture history $z_{(t_1+1):(t_2)}$ between these two time points. Note that when $t_1 = t_2$, $z_{(t_1+1):(t_2)}$ is denoted by $-$.
- $w_{t_1,t_2+1}^{U,z_{(t_1+1):(t_2)},s}$ denotes the number of animals observed for the first time at t_1 in an unknown state, re-observed for the first time in known state s at time $t_2 + 1$ with partial capture history $z_{(t_1+1):(t_2)}$ between these two time points.
- $v_{t_1}^r$ is the number of animals observed in known state r at t_1 and never seen again in a known state (i.e. never seen again or only ever re-observed in an unknown state).
- $b_{t_1}^U$ is the number of animals first observed in an unknown state at t_1 and never seen again in a known state.

Building upon the notation and probabilities introduced in the previous section, we will demonstrate that the number of animals *partially observed* at time i and later seen again in a known state, follows a multinomial distribution which is a mixture of the multinomial distributions of the animals released in a *known state* at time i and seen again in a known state later. The multinomial cells correspond to the time and state of the first re-observation in a known state after time i.

The mixture property is illustrated for a simple example in Table 2 for occasion $i = 2$ of a $T = 4$ occasion capture-recapture study with two live states A and B. The number of animals released in state A at occasion 1 first re-captured in a known state at the different occasions, and those never seen again in a known state, follow a multinomial distribution (row 1). Similarly for those released in state B at occasion 1 (row 2), and those first released in an unknown state at occasion 1 (row 3) and at occasion 2 (row 4).

When the number of sampling occasions increases, capture histories are long and there are a great number of possible intermediate capture histories, formed of combinations of 0s and Us, before the first observation

in a known state appears. In order to lower the chances of a sparse table, we opt to build the multinomials based on the time and state of the first known re-observed state, thus pooling over all possible intermediate capture histories.

In supplementary material Section 2 we show that the number of animals previously released in a known state r, partially observed at occasion i and re-observed later in a known state, follows a conditional multinomial distribution, which is a mixture of the conditional multinomial distributions followed by the animals released at occasion i in the observable states. We also show that the number of animals first released before i or at i in an unknown state, partially observed at occasion i and re-observed later in a known state, follows a conditional multinomial distribution (denoted in blue in Table 1), which is a mixture of the conditional multinomial distributions followed by the animals released at i in the observable states (denoted in red in Table 1).

Using the following property cited from Pradel *et al* (2003): "*if* $B1$ *and* $B2$ *are mutually independent stochastic vectors, which are multinomially distributed, and if* $M1$ *and* $M2$ *are mutually independent stochastic vectors whose distributions are separately mixtures of the distributions of* $B1$ *and* $B2$*, then the distribution of* $M1+M2$ *is itself a mixture of the distributions of* $B1$ *and* $B2$", the conditional multinomials of the animals released in a known state or first released in an unknown state before or at i, and partially observed at i can be pooled as shown in Table 3. Thus, the table used to test the mixture property of partial observations at occasion i is given in Table 3.

3 TESTING THE UNDERLYING STATE STRUCTURE GENERATING THE PARTIAL OBSERVATIONS

Based on the mixture property of partial observations at a given occasion demonstrated in the previous section, we use the Multinomial Maximum Likelihood Mixture approach (MMLM) developed by Yantis *et al* (1991) to assess the goodness-of-fit of a model where the partial observations are generated only by the directly observable states. The MMLM approach is targeted to mixtures of multinomial distributions and is used when independent samples are available from both the mixtures and their associated components. This approach consists of two steps: first estimating the cell probabilities of the mixture components and the mixing weights via maximum-likelihood, then assessing the goodness-of-fit of the hypothesised model structure (mixtures and associated components) using a classical measure of comparison between observed and expected frequencies.

Hence, based on the mixture property of the partial observations demonstrated in the supplementary material and reported in Section 2, there is no need to estimate the numerous capture-recapture parameters

for the purpose of the test, the information needed is summarised in simpler terms: one parameter per component-cell and the mixing weights as illustrated in Table 4.

For the goodness-of-fit assessment, various statistics based on the distance between expected values under the model and observed values may be considered: Pearson's χ^2, the log-likelihood ratio statistic G^2 (Cressie and Read, 1988, p. 10); and more generally, due to the different properties of these statistics depending on the alternatives or sparseness of the table, the power-divergence family of statistics (Cressie and Read, 1988), which encompasses G^2 and χ^2 as special cases. Within this paper, we present the results obtained with Pearson's χ^2 as all the various statistics used gave similar results.

Under the null hypothesis, animals partially observed at i and re-observed later in a known state are consistent with being a mixture of animals observed in the directly observable states at i and re-observed in the same conditions: the partial observations are generated solely by the observable states (Figure 1). Using the usual H_0 notation for the null hypothesis and H_1 for the alternative, $H_1 = \bar{H}_0$. A large array of situations come under the alternative hypothesis: from the partial observations being generated by the directly observable states *and* another state which is never directly observable (Figure 2) to the most extreme case of partial observations all being generated only by one (or more) states which are never directly observable.

Under the null hypothesis, the Pearson goodness-of-fit statistic presented above follows a χ^2 distribution (Cressie and Read, 1984) with $K - p - 1$ degrees of freedom (Moore, 1986, p. 66) where K denotes the number of observed frequencies and p denotes the number of parameters in the model. In order for the asymptotic distributions to hold, expected frequencies in each cell should be at least 2 for a level $\alpha = 0.05$ (Moore, 1986, p. 71).

The tables used at each occasion i condition on known states. Therefore, the test-statistics obtained at each occasion are independent and a global test-statistic can be computed by summing up the tests for each occasion. This global test-statistic follows, under the null hypothesis, a chi-square distribution with the number of degrees of freedom being the sum of the degrees of freedom of the test-statistics per occasion.

4 APPLICATIONS

4.1 Simulation results

In order to minimise the chances of sparse data and verify that the test works as expected in theory, we first used simulation with very large sample size (N=25,000 animals newly released at each occasion), whilst also focusing on an extreme case of the alternative hypothesis (results not presented here). We then simulated the same scenarios under more realistic settings as detailed below. First, we present

simulations for two-state capture-recapture data under the null hypothesis, arising from two directly observable states, with $K = 5$ sampling occasions, under two sample size settings: N=5000 and N=1000 animals newly released per occasion. The capture, survival and transition probabilities, are respectively set as $p^A = p^B = 0.6$, $\phi^A = 0.6$, $\phi^B = 0.9$, $\psi^{AB} = 0.8$, $\psi^{BA} = 0.7$. This scenario is denoted by $2S$. In order to introduce partial observations, we set to unknown at random a varying percentage of the observed states~~observations as missing completely at random~~ (MCAR). More specifically, we ran a binomial on each observed state in scenario 2S to decide whether it should be kept as 'observed in the relevant known state' or changed to 'observed in unknown state'. We also simulated data under the alternative hypothesis, where the partial observations are not generated by either of the two directly observable states, but by a third state C which is never directly observable, this scenario is denoted by $3S$. Using standard multievent notation (see for example Pradel, 2005), the survival matrix is denoted by $\mathbf{\Phi_t}$ with the diagonal terms being the probability that an animal in state r at time t survives until $t + 1$ and the last column being the probability of dying,

$$\mathbf{\Phi_t} = \begin{array}{c} \\ A \\ B \\ C \\ Dead \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & Dead \end{array} \\ \begin{pmatrix} 0.7 & 0 & 0 & 0.3 \\ 0 & 0.8 & 0 & 0.2 \\ 0 & 0 & 0.9 & 0.1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

for $t = 1, \ldots, 4$; the transition matrix with the (r, s)th element being $\psi_t^{r,s}$, the probability that an animal is in state s at time $t + 1$, given it was in state r at t and that it is alive at $t + 1$, is denoted by

$$\mathbf{\Psi_t} = \begin{array}{c} \\ A \\ B \\ C \\ Dead \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & Dead \end{array} \\ \begin{pmatrix} 0.1 & 0.3 & 0.6 & 0 \\ 0.3 & 0.15 & 0.55 & 0 \\ 0.4 & 0.4 & 0.2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

for $t = 1, \ldots, 4$ and finally, the event matrix with the (r, e)th element being the probability of observing event e for an animal in state r at time t is denoted by

$$\mathbf{B_t} = \begin{array}{c} \\ A \\ B \\ C \\ Dead \end{array} \begin{array}{c} \begin{array}{cccc} 0 & A & B & U \end{array} \\ \begin{pmatrix} 0.45 & 0.55 & 0 & 0 \\ 0.45 & 0 & 0.55 & 0 \\ 0.45 & 0 & 0 & 0.55 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

for $t = 1, \ldots, 5$. Here the events (corresponding to the columns) are, *not observed*, *observed in state A*, *observed in state B* and *observed in unknown state denoted by U*.

We examine this scenario for the following numbers of animals newly released at each occasion: N=100, N=250, N=500, and N=1000, ~~N=2500 and N=5000~~. We simulate 600 datasets for each scenario. If any of the expected values are lower than two, the corresponding test is deemed Non Applicable (NA). Since sparse data were extremely likely to arise for the smaller sample sizes, we automatically applied pooling strategies before performing the maximum likelihood test: pooling across columns while the number of columns is greater than the number of components plus one, and across the lines: all the mixtures are pooled together to form just one mixture. The results obtained are given in terms of percentage of significant test results out of the number of applicable tests, at a 5% level, in Table 5.

In order to examine how the test would perform in the more challenging situation where some partial observations are generated by the observable states, we also examined for the sample size N=1000 a variant of the 3S scenario where, in addition to the partial observations corresponding to state C, 30% of the observations generated by the observable states A and B are set to partial at random (unknown state).

The simulation results show that for the datasets simulated under the null hypothesis (scenario 2S), the Type I error rate is close to 5%, whatever the percentage of partial observations. Importantly, the test showed good power for the datasets simulated under an alternative hypothesis (scenario 3S), with close to 50% of tests being significant for a sample size as small as 100 animals newly released per occasion (i.e. 500 animals altogether) and close to 100% of the global test being significant for 250 animals released per occasion. The simulation results show that the test reacts as expected from the derivation made in the previous sections, when the partial observations are not generated by the directly observable states, and that it can work well for realistic sample sizes. When part of the partial observations are generated by the observable states, the test is not as powerful as could be expected but nonetheless rejects H_0.

4.2 Canada Geese

We have shown theoretically and empirically that our test has the ability to assess whether partial observations can be adequately modelled as stemming solely from the directly observable states in a

capture-recapture experiment. In this section, we apply the test to an ecological dataset, chosen so that the underlying state structure is actually known.

We use the Canada geese dataset from Hestbeck *et al* (1991) which consists of 21,435 migrant geese individually marked with neck-bands and re-observed at their wintering locations each year, between 1984 and 1989 (Hestbeck *et al*, 1991; Rouan *et al*, 2009). These wintering sites constituted the states in the capture-recapture experiment: mid-Atlantic (New York, Pennsylvania, New Jersey), Chesapeake (Delaware, Maryland, Virginia), and Carolinas (North and South Carolina). Since the tables needed for the test were quite sparse, we therefore used the following pooling strategy: on the columns, pooled to the maximum until there was one degree of freedom left for the test (the column with the minimal sum is pooled with the column with the second minimal sum and so on) whilst on the rows, all the rows corresponding to mixtures are pooled so that there is just one mixture left to test for.

We examine the Canada geese dataset under both the null and alternative hypotheses by artificially creating these situations within the data. First, in order to create partial observations generated by the observable states (H_0), we set some ~~a varying percentage of the~~ observed geese's states to unknown (MCAR). We considered varying percentages to see how the test reacts to the amount of partial observations: 15%, 25% and 45%~~, so that the partial observations are generated only by the observable states (H_0)~~. These situations are respectively denoted by MCAR15, MCAR25 and MCAR45 in Table 6. Then we examine situations that come under the alternative hypotheses (H_1) by setting all of the observations from a particular state to "unknown" so that this particular state becomes unobservable while the states remaining observable do not generate any partial observations. We considered 2 situations: all observations in state 2 are set to "unknown" (situation 2PO), or all those in state 3 are set to "unknown" (situation 3PO). Eventually, we considered the hybrid situation where, in addition to the partial observations generated by the unobservable state 3 as in scenario 3PO, 25% then 45% of the observations generated by state 2 are also set to partial: scenarios Hyb25 and Hyb45~~. Hence, for situations 2PO and 3PO the partial observations stem, respectively, only from states 2 and 3. Note that the state set to "unknown" is never directly observed in each of these situations. This allows us to test the performance of the test as we know that all unknown events correspond to a different state from the ones "observed".~~

The p-values obtained from applying the mixture test to all these configurations of the geese dataset are given in Table 6. These results are very promising, with the test reacting as it should under the different configurations examined. Under all the null hypothesis configurations, the directly observable states as sole underlying states for the partial observations, there is insufficient evidence to reject the null hypothesis. For the configurations under the alternative, the null hypothesis is strongly rejected, with $p<0.001$ for almost all of the tests examined (by occasion and global). The non-significant test at occasion 2 under

scenario 3PO is due to the small number of individuals captured in state 3 at this occasion, resulting in insufficient power to detect the different properties of that state. Hence, the results from configurations 2PO and 3-PO lead to the conclusion that the directly observable states do not provide an adequate underlying state-structure for the partial observations. When some partial observations are generated by the observable states (Hyb25 and Hyb45), there is a clear loss of statistical power. The global tests are still very close to significance at the 5% level, but more than 5 years of study would have been necessary to detect the presence of the third unmonitored location.

5 DISCUSSION

We have derived a mixture test that assesses whether partial observations in a capture-recapture study are generated solely from the directly observable states. This test is based on distributional properties which we have demonstrated. It has been shown to perform well in theory, through simulation and for real-data applications. Regarding the interpretation of the test, if the null hypothesis is not rejected, the observable states provide an adequate underlying structure for the partial observations. However, similarly to classical goodness-of-fit tests, the interpretation of a significant test result is not as straightforward as the range of alternatives to be considered is quite large. For example, if the set of observable states are inadequate, it is not known how many additional states should be considered for the underlying structure and how the partial observations should be modelled. Both of these questions do not have obvious answers at this stage and constitute an area of future research.

Partial observations might also stem from alternatives less extreme than those considered in our applications: they could be generated by one of the directly observable states and an additional state that is never observable directly. Going further, they may also stem from all the observable states and another state which is never observable directly. In theory, the test will react to this situation too. However, in practice, we surmise that the other state would have to present different enough properties from the directly observable states for the test to be powerful enough to detect it.

Finally, determining a minimum sample size for which the test is powerful enough is more complex than usual in this framework, as it is not only the total sample size which matters but also the proportion of partial observations, which will depend on combinations of the parameter values. From a modelling perspective, we would recommend fitting a model with one additional state when the test is found to be significant.

This new test has sound theoretical basis, we showed it can work well even with small sample sizes, and we believe that it will be useful in a multi-state capture-recapture model, in statistical ecology and also other areas of application. Hidden Markov models are used for a range of purposes in capture-recapture

modelling, (see for example Langrock and King (2013); Zhou *et al* (2019); Worthington *et al* (2019)), and the work of this paper will considerably contribute to the theoretical tools available for a wide range of applications. It will enable practitioners to consider better fitting models and will also give practical insight as to the existence of at least one state where the animals go, that is different from those directly observed.

Clearly it is desirable to consider whether the approach presented in this paper can be extended to other applications of HMMs in ecology, for example in application to movement models (Langrock *et al*, 2012), and beyond, and this is a current area of research.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

RM and RP conceived of the presented idea. AJ developed the theory and performed the computations. RP verified the theory and analytical results. All authors discussed the results and contributed to the final manuscript.

FUNDING

RM was funded by NERC fellowship grant NE/J018473/1 when conducting the research of this paper and by EPSRC grant EP/S020470/1 during the writing of it. AJ was funded by the School of Mathematics, Statistics and Actuarial Science of the University of Kent (UK) and National Centre for Statistical Ecology EPSRC/NERC grant EP/I000917/1.

ACKNOWLEDGMENTS

We would like to thank Olivier Gimenez for letting us use part of his R code from his R2UCare package: https://github.com/oliviergimenez/R2ucare/tree/master/R

DATA AVAILABILITY STATEMENT

The Canada goose data set used within this paper has been provided as a supplementary file.
The R code used to implement the test described in this paper has been provided as supplementary files.

REFERENCES

BARTOLUCCI, F. AND PENNONI, F. (2007) A class of latent Markov models for capture- recapture data allowing for time, heterogeneity and behavior effects. *Biometrics*. **63,** 568-578.

CORMACK, R. M. (1964) Estimates of survival from sighting of marked animals. *Biometrika*. **51,** 429-438.

CRESSIE, N. AND READ, T. R. C. (1984) Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society. Series B (Methodological)*. **46,** 440–464.

CRESSIE, N. AND READ, T. R. C. (1988) *Goodness-of-fit statistics for discrete multivariate data*. Springer, New York, USA.

HESTBECK, J. B. AND NICHOLS, J. D. AND MALECKI, R. A. (1991) Estimates of movement and site fidelity using mark-resight data of wintering Canada geese *Animal Biodiversity and Conservation* **27,** 97-107.

KENDALL, W. L. (2004) Coping with unobservable and mis–classified states in capture–recapture studies. *Ecology* **72,** 523–533.

KING, R. AND MCCREA, R. S. (2014) A generalised likelihood framework for partially observed capture–recapture–recovery models. *Statistical Methodology* **17,** 30–45.

LANGROCK, R., KING, R., MATTHIOPOULOS, J., THOMAS, L., FORTIN, D. AND MORALES, J. M. (2012) Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, **93,** 2336–2342.

LEBRETON, J. D. , NICHOLS, J. D. , BARKER, R. J. , PRADEL, R. AND SPENDELOW, J. A. (2009) Modeling individual animal histories with multistate capture–recapture models. *Advances in Ecological Research* **41,** 87–173.

LANGROCK, R. AND KING, R. (2013) Maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. *Annals of Applied Statistics* **7,** 1709–1732.

MOORE, D.S. (1986) Tests of chi-squared type. In *Goodness-of-fit techniques* (Eds. D'Agostino, R. B. and Stephens, M. A.) CRC Press, New York.

OTIS, D. L. , BURNHAM, K. P. , WHITE, G. C. AND ANDERSON, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62,** 1–135.

PLEDGER, S. (2000) Unified maximum likelihood estimates for closed capture recapture models using mixtures. *Biometrics*. **56,** 434-442.

POHLE, J., LANGROCK, R. VAN BEEST, F. M. AND SCHMIDT, N. M. (2017). Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological, and Environmental Statistics* **22,** 270-293.

PRADEL, R. (2005). Multievent: An Extension of Multistate Capture-Recapture Models to Uncertain States. *Biometrics* **61,** 442–447.

PRADEL, R. AND WINTREBERT, C. M. A. AND GIMENEZ, O. (2003). A Proposal for a Goodness-of-Fit Test to the Arnason-Schwarz Multisite Capture-Recapture Model. *Biometrics* **59,** 43–53.

ROUAN, L. AND CHOQUET, R. AND PRADEL, R. (2009). A general framework for modeling memory in capture-recapture data. *Journal of Agricultural, Biological, and Environmental Statistics* **14,** 338–355.

WILLIAMS, B. K., NICHOLS, J. D. AND CONROY, M. J. (2002) *Analysis and management of animal populations.* Academic Press, San Diego, California, USA.

WORTHINGTON, H., MCCREA, R. S. , KING, R. AND GRIFFITHS, R. A. (2019) Estimating abundance from multiple sampling capture-recapture data using hidden Markov models in the presence of discrete-state information. *Annals of Applied Statistics.* In press.

YANG H. C. AND CHAO, A. (2005) Modeling animals behavioral response by Markov chain models for capture-recapture experiments. *Biometrics*. **61,** 1010-1017.

YANTIS, S. AND MEYER, D. E. AND SMITH, J. K. (1991). Analyses of multinomial mixture distributions: new tests for stochastic models of cognition and action. *Psychological Bulletin* **110,** 350.

ZHOU, M., MCCREA, R. S. , MATECHOU, E., COLE. D. J. AND GRIFFITHS, R. A. (2019) Removal models accounting for temporary emigration. *Biometrics* **75,** 24-35.

ZUCCHINI, W. AND MACDONALD, I. L. AND LANGROCK, R. (2016). *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC Press.

FIGURES AND TABLES

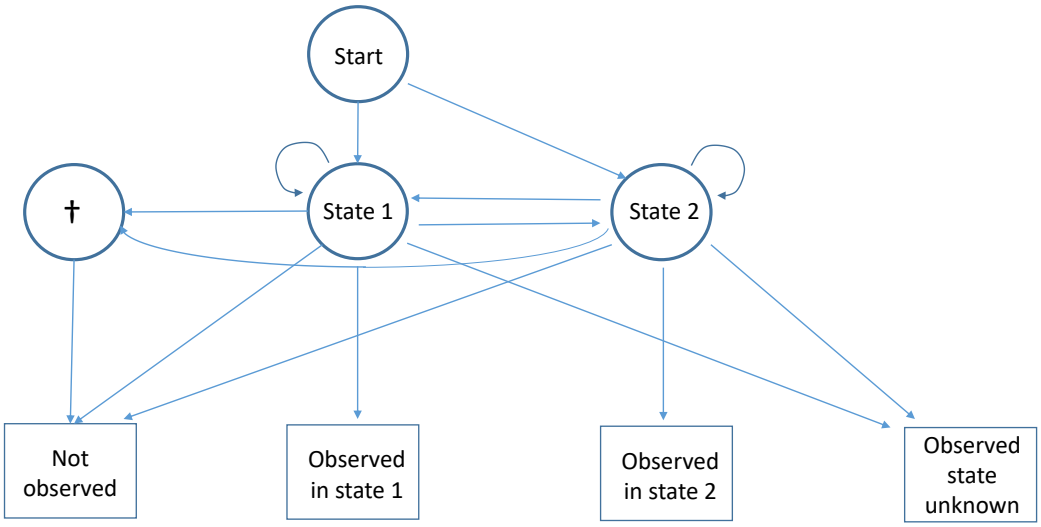

Figure 1. Diagram of the capture recapture multievent model for partial observations with two observable live states under the null hypothesis. The state 'dead' is represented by †. Four events are generated by the three states: 'Not observed', which is obligatory for the state 'dead'; two complete observations, 'Observed in state 1' and 'Observed in state 2'; and the partial observation 'Observed state unknown', which may be generated by either live state.

347

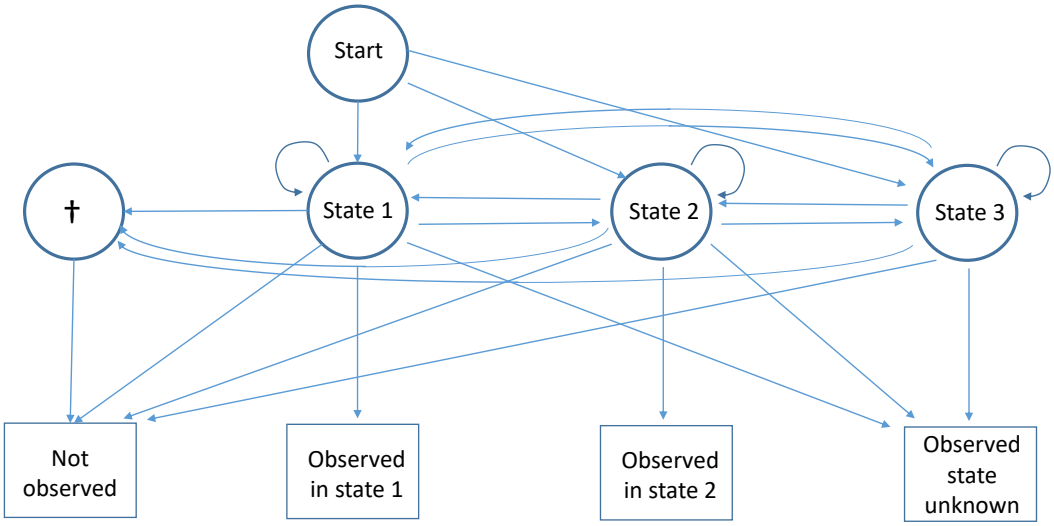

Figure 2. Diagram of the capture recapture multievent model for partial observations with two observable live states under the alternative hypothesis where there is one additional non-observable live state (state 3). This last state is never recognized upon observation. See Figure 1 for more details

Table 1. Illustrating how example individual capture histories contribute to the sufficient statistic terms, for a capture-recapture experiment with two observable states A, B and five sampling occasions. Partial observations are denoted by U. The elements of capture history determining the indices within the statistics are denoted in bold.

Capture History					sufficient statistic
U	**A**	U	U	**B**	$w_{1,2}^{U,-,A}$, $n_{2,5}^{A,UU,B}$
A	U	U	U	**A**	$n_{1,5}^{A,UUU,A}$
A	U	0	U	0	v_1^A
U	U	U	U	**B**	$w_{1,5}^{U,UUU,B}$
0	0	**U**	0	**1**	$w_{3,5}^{U,0,1}$
0	**A**	**B**	U	U	$n_{2,3}^{A,-,B}$, v_3^B
0	**U**	0	U	U	b_2^U

Table 2. The sufficient statistics for multinomial distributions corresponding to individuals released before or at $i = 2$ in an capture-recapture experiment with 4 occasions where individuals can be in any of 2 live states~~: sufficient-statistic terms~~. At each time for each individual, one of 4 events occurs: the individual is not encountered (code 0), the individual is encountered but its state is not recognized (event U), the individual is encountered and recognized to be in state A (code A), the individual is encountered and recognized to be in state B (code B). In the electronic version of the paper the terms constitutive of mixtures are denoted in blue whilst those constituting components are denoted in red. The terms in black will be conditioned upon. b_i^U-terms are the counts of animals with a first partial observation at i (initial event U) that are never completely observed. $w_{i,j}^{U,h,S}$-terms are the counts of animals with a first partial observation at i and a first complete observation at j in state S with intervening capture history h (- stands for the empty capture history). $n_{i,j}^{R,h,S}$-terms are the counts of animals with two successive complete observations respectively at times i and j in states R and S with intervening capture history h. v_i^S-terms are the counts of animals observed completely for the last time at i in state S.

$n_{1,2}^{A,-,A}$	$n_{1,2}^{A,-,B}$	$n_{1,3}^{A,0,A}$	...	$n_{1,4}^{A,0U,B}$	$n_{1,3}^{A,U,A}$	$n_{1,3}^{A,U,B}$	$n_{1,4}^{A,U0,A}$	$n_{1,4}^{A,U0,B}$	$n_{1,4}^{A,UU,A}$	$n_{1,4}^{A,UU,B}$	v_1^A
$n_{1,2}^{B,-,A}$	$n_{1,2}^{B,-,B}$	$n_{1,3}^{B,0,A}$	...	$n_{1,4}^{B,0U,B}$	$n_{1,3}^{B,U,A}$	$n_{13}^{B,U,B}$	$n_{14}^{B,U0,A}$	$n_{14}^{B,U0,B}$	$n_{14}^{B,UU,A}$	$n_{14}^{B,UU,B}$	v_1^B
$w_{12}^{U,-,A}$	$w_{12}^{U,-,B}$	$w_{13}^{U,0,A}$	...	$w_{14}^{U,0U,B}$	$w_{13}^{U,U,A}$	$w_{13}^{U,U,B}$	$w_{14}^{U,U0,A}$	$w_{14}^{U,U0,B}$	$w_{14}^{U,UU,A}$	$w_{14}^{U,UU,B}$	b_1^U
					$w_{23}^{U,-,A}$	$w_{23}^{U,-,B}$	$w_{24}^{U,0,A}$	$w_{24}^{U,0,B}$	$w_{24}^{U,U,A}$	$w_{24}^{U,U,B}$	b_2^U
-	-	-	-	-	$n_{23}^{A,-,A}$	$n_{23}^{A,-,B}$	n_{24}^{A0A}	n_{24}^{A0B}	n_{24}^{AUA}	n_{24}^{AUB}	v_2^A
-	-	-	-	-	$n_{23}^{B,-,A}$	$n_{23}^{B,-,B}$	$n_{24}^{B,0,A}$	$n_{24}^{B,0,B}$	$n_{24}^{B,U,A}$	$n_{24}^{B,U,B}$	v_2^B

Table 3. Table used for testing the mixture property of partial observations at occasion i in a capture-recapture experiment with T occasions where individuals can be in any of R live states. Notations are as in Table 2. The columns correspond to the circumstances (time and state) of the first reobservation in a known state after i. They are pooled over the different intervening partial histories (. notation), $h(i) = U$ denotes that the animals are seen in U at i. For individuals seen in U at i, the rows are pooled by last recognized state ~~at last release~~ (first R rows) and when there are no ~~certain~~complete observations prior to i+1 (row R+1). For instance, the first row is for animals seen in U at i and with A as their last recognized state; the summation is over the timing of this last previous complete observation.

$j =$	$i+1$			$\dots$	T		
$s =$	A	$\dots$	R	$\dots$	A	$\dots$	R
	$\sum_{f=1}^{i-1} n_{f,i+1}^{A,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,i+1}^{A,.,R,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,T}^{A,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,T}^{A,.,R,h(i)=U}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\sum_{f=1}^{i-1} n_{f,i+1}^{R,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,i+1}^{R,.,R,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,T}^{R,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i-1} n_{f,T}^{R,.,R,h(i)=U}$
	$\sum_{f=1}^{i} w_{f,i+1}^{U,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i} w_{f,i+1}^{U,.,R,h(i)=U}$	$\dots$	$\sum_{f=1}^{i} w_{f,T}^{U,.,A,h(i)=U}$	$\dots$	$\sum_{f=1}^{i} w_{f,T}^{U,.,R,h(i)=U}$
	$\mathrm{n}_{i,i+1}^{A,A}$	$\dots$	$\mathrm{n}_{i,i+1}^{A,R}$	$\dots$	$\mathrm{n}_{i,T}^{A,.,A}$	$\dots$	$\mathrm{n}_{i,T}^{A,.,R}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\mathrm{n}_{i,i+1}^{R,A}$	$\dots$	$\mathrm{n}_{i,i+1}^{R,R}$	$\dots$	$\mathrm{n}_{i,T}^{R,.,A}$	$\dots$	$\mathrm{n}_{i,T}^{R,.,R}$

Table 4. Simple model structure of mixtures and associated components used to test the mixture property. In the electronic version of the paper the mixing weights are denoted in blue and the component cell-probabilities in red. B_r is the basis corresponding to animals released at i in state r, $r = 1, \ldots, R$. M_r is the mixture corresponding to animals partially observed at i and most lately completely observed in state r, $r = 1, \ldots, R$. Only animals completely reobserved at some point after i are used in the bases and mixtures. The cells of the multinomials correspond to the time and state of the first complete observation after i. They are ordered by states within times for a total of $R \times (T - i)$ cells. $p_i^{B_r}$ is the probability associated to cell i of basis B_r, $i = 1, \ldots, R \times (T - i)$, $r = 1, \ldots, R$. γ_r, $r = 1, \ldots, R$, are the mixing weights for M_1. π_r, $r = 1, \ldots, R$, are the mixing weights for M_R.

$j =$	$i+1$			$\ldots$	T		
$s =$	A	$\ldots$	R	$\ldots$	A	$\ldots$	R
M_1	$\gamma_1 p_1^{B_1} + \ldots + \gamma_R p_1^{B_R}$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\gamma_1 p_{R\times(T-i)}^{B_1} + \ldots + \gamma_R p_{R\times(T-i)}^{B_R}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$							
M_R	$\pi_1 p_1^{B_1} + \ldots + \pi_R p_1^{B_R}$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\pi_1 p_{R\times(T-i)}^{B1} + \ldots + \pi_R p_{R\times(T-i)}^{B_R}$
B_1	$p_1^{B_1}$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$p_{R\times(T-i)}^{B_1}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$							
B_R	$p_1^{B_R}$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$p_{R\times(T-i)}^{B_R}$

Table 5. Testing the mixture property of partial observations: simulation results. For H_0, we generated 2-state capture histories (scenario 2S) examining 2 sample sizes (1000 and 5000 animals newly released per occasion) and 2 percentages of observations rendered partial by setting the state to unknown (%MCAR). Different values of the binomial parameter were considered. For H_1, we generated 3-state capture histories (scenario 3S) examining 4 sample sizes: 2 states were fully observable while the third, never observed, gave rise to all the partial observations. Values of the detection, survival, and transition parameters for scenarios 2S and 3S are given in section 4.1. Under a variant of scenario 3S with the largest sample size, 30% of the observations generated by the 2 observable states are also made partial at random. In all cases, 600 replicates were simulated. Results are given as percentage of significant test results out of the number of applicable tests (all expected values ≥ 2). G denotes the global test, i the sampling occasion and $\%MCAR$ the percentage of observations set to "Unknown" and N denotes the number of applicable tests. ~~The sample size examined is indicated next to the relevant scenario -~~ When 50% or more of the test-results were significant, this is indicated in bold.

Scenario	sample size	% MCAR	i	$\chi^2(\%)$	N ~~(non NA)~~
			2	4.83	600
		45	3	3.33	600
			G	5.00	600
			2	3.50	600
		25	3	4.17	600
			G	4.50	600
	5000				
			2	4.77	600
		45	3	3.63	600
			G	4.93	600
		-	2	26.44	435
	100	-	3	45.84	373
		-	G	42.73	550
		-	2	**52.75**	582
	250	-	3	**92.60**	581
		-	G	**94.33**	600
		-	2	**83.47**	599
3S H_1	500	-	3	**99.83**	600
		-	G	**100.00**	600
		-	2	**96.50**	600
		-	3	**100.00**	600
		-	G	**100.00**	600
	1000				
			2	**76.30**	600
		30	3	**97.50**	600
			G	**99.80**	600

Table 6. Using different configurations of the Canada geese dataset to assess the performance of the new mixture test for assessing the underlying state structure of partial observations, under real-life conditions. Starting from an original data set where individually identified Canada geese have been observed at 3 locations during 6 consecutive wintering seasons, we artificially generated 3 scenarios under H_0 by setting 15%, 25%, and 45% of the observed geese's locations to unknown : scenarios MCAR15, MCAR25, MCAR45 respectively, and 2 scenarios under H_1 by setting all the observations at location 2 (resp. 3) to unknown: scenarios 2PO (resp. 3PO). The p-value obtained at each occasion i is presented and the associated global tests are denoted by G.

	Configuration	i	p-value	df ~~dof~~
H0	MCAR15	2	0.14	1
		3	0.14	1
		4	0.60	1
		G	0.21	3
	MCAR25	2	0.57	1
		3	0.09	1
		4	0.85	1
		G	0.35	3
	MCAR45	2	0.84	1
		3	0.82	1
		4	0.85	1
		G	0.99	3
H_1	2PO	2	<**0.001**	1
		3	<**0.001**	1
		4	<**0.001**	1
		G	<**0.001**	3
	3PO	2	0.13	1
		3	<**0.001**	1
		4	<**0.001**	1
		G	<**0.001**	3
	Hyb25	2	0.25	1
		3	0.28	1
		4	0.04	1
		G	0.07	3
	Hyb45	2	0.37	1
		3	0.08	1
		4	0.07	1
		G	0.06	3