

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

Anita Jeyam, Rachel Mccrea, Roger Pradel

► **To cite this version:**

Anita Jeyam, Rachel Mccrea, Roger Pradel. A Test for the Underlying State-Structure of Hidden Markov Models: Partially Observed Capture-Recapture Data. *Frontiers in Ecology and Evolution*, Frontiers Media S.A, 2021, 9 (598325), 10.3389/fevo.2021.598325 . hal-03149918

**HAL Id: hal-03149918**

**<https://hal.archives-ouvertes.fr/hal-03149918>**

Submitted on 23 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

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

Anita Jeyam<sup>1</sup>, Rachel McCrea<sup>1,\*</sup> and Roger Pradel<sup>2</sup>

<sup>1</sup>*Statistical Ecology @ Kent, National Centre for Statistical Ecology, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, UK*

<sup>2</sup>*CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France* *Centre d'Ecologie Fonctionnelle et Evolutive, Montpellier, France*

Correspondence\*:

R.S.McCrea@kent.ac.uk

## 2 ABSTRACT

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

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

## 1 INTRODUCTION

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

16 widely used technique to follow the dynamics of open animal populations (Cormack, 1964; Williams *et al.*,  
17 2002). The protocol remains the same: animals are uniquely marked, then released and resighted/recaptured  
18 at subsequent sampling occasions. In the multi-state framework (Lebreton *et al.*, 2009), at each occasion,  
19 individual animals' states are recorded upon resighting; if an animal is not seen at a given occasion, this  
20 is denoted by a 0. If it is seen, a code, commonly a number, specifies the state (see example data set  
21 in supplementary material). Hence, the data resulting from a multi-state capture-recapture experiment  
22 consists of individual encounter histories, formed by the series of records made for each animal. Multi-state  
23 models allow the estimation of the survival and transition probabilities of animals between the states, whilst  
24 accounting for imperfect detection. However, Within this modelling framework, states are assumed to be  
25 assigned without error (Kendall, 2004). However, this assumption can be unrealistic in certain situations  
26 such as the assessment of sex in a monomorphic species or of health status when biological testing is not  
27 possible in the field. Pradel (2005) developed multievent models, which belong to the family of Hidden  
28 Markov Models (Zucchini *et al.*, 2016) to account for the uncertainty in state assignment. These models  
29 belong to the family of Hidden Markov Models (Zucchini *et al.*, 2016) and distinguish the events, which are  
30 observed, from the states, which are underlying. In this framework, events are observed whilst the states  
31 are underlying. The process governing the transitions between states is Markovian (generally assumed of  
32 order 1) and the events are generated by the states. Multievent models have a structural absorbing state  
33 (death). Transitions are almost systematically time-dependent, which precludes the consideration that the  
34 system has reached an equilibrium. Also, because the chance that an individual is missed is state dependent,  
35 non-observations cannot be considered as data missing at random. They are informative events like any  
36 other outcome of the experiment.

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

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

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

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

## 2 PARTIALLY OBSERVED CAPTURE-RECAPTURE DATA AND MIXTURE PROPERTIES

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

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

- 70 •  $\phi_t^r$  is the probability an individual in state  $r$  at time  $t$  survives until  $t + 1$ , for  $t = 1, \dots, T - 1$  **and**  
 71  $r = 1, \dots, R$ .
- 72 •  $p_t^r$  is the probability of recapture at time  $t$  for an individual in state  $r$ , for  $t = 2, \dots, T$ .
- 73 •  $\psi_t^{r,s}$  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  
 74 alive at  $t + 1$ , for  $t = 1, \dots, T - 1$ ,  $r = 1, \dots, R$  **and**  $s = 1, \dots, R$ .
- 75 •  $\alpha_t^r$  is the probability an individual is assigned to state  $r$  given it was recaptured at time  $t$  **and in state**  
 76  **$r$  at that time**, for  $t = 2, \dots, T$  **and**  $r = 1, \dots, R$ .  $\beta_t^r = 1 - \alpha_t^r$  is then defined as the probability an  
 77 individual is assigned as unknown (U) at time  $t$  given the individual is recaptured, and in state  $r$  at this

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

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

83 The sufficient statistics are based on partitioning the encounter histories (EH) into the following pieces:  
84 the EH between observations in two known states; the EH between first observation in unknown state and  
85 first re-observation in a known state; the EH following the last observation in a known state; and the EH  
86 following the first observation in an unknown state, for animals who are never seen in a known state (Table  
87 1 provides examples). We define the following sufficient statistics:

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

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

103 The mixture property is illustrated for a simple example in Table 2 for occasion  $i = 2$  of a  $T = 4$  occasion  
104 capture-recapture study with two live states A and B. The number of animals released in state A at occasion  
105 1 first re-captured in a known state at the different occasions, and those never seen again in a known state,  
106 follow a multinomial distribution (row 1). Similarly for those released in state B at occasion 1 (row 2), and  
107 those first released in an unknown state at occasion 1 (row 3) and at occasion 2 (row 4).

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

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

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

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

### 3 TESTING THE UNDERLYING STATE STRUCTURE GENERATING THE PARTIAL OBSERVATIONS

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

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

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

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

147

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

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

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

## 4 APPLICATIONS

### 165 4.1 Simulation results

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

170 simulations for two-state capture-recapture data under the null hypothesis, arising from two directly  
 171 observable states, with  $K = 5$  sampling occasions, under two sample size settings:  $N=5000$  and  $N=1000$   
 172 animals newly released per occasion. The capture, survival and transition probabilities, are respectively  
 173 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  
 174 order to introduce partial observations, we set to unknown at random a varying percentage of the observed  
 175 states observations as missing completely at random (MCAR). More specifically, we ran a binomial on  
 176 each observed state in scenario  $2S$  to decide whether it should be kept as 'observed in the relevant known  
 177 state' or changed to 'observed in unknown state'. We also simulated data under the alternative hypothesis,  
 178 where the partial observations are not generated by either of the two directly observable states, but by a  
 179 third state  $C$  which is never directly observable, this scenario is denoted by  $3S$ . Using standard multievent  
 180 notation (see for example Pradel, 2005), the survival matrix is denoted by  $\Phi_t$  with the diagonal terms  
 181 being the probability that an animal in state  $r$  at time  $t$  survives until  $t + 1$  and the last column being the  
 182 probability of dying,

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

for  $t = 1, \dots, 4$ ; 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

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

183

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



$$\mathbf{B}_t = \begin{matrix} & 0 & A & B & U \\ \begin{matrix} A \\ B \\ C \\ Dead \end{matrix} & \begin{pmatrix} 0.45 & 0.55 & 0 & 0 \\ 0.45 & 0 & 0.55 & 0 \\ 0.45 & 0 & 0 & 0.55 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

186

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

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

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

201 The simulation results show that for the datasets simulated under the null hypothesis (scenario 2S), the  
 202 Type I error rate is close to 5%, whatever the percentage of partial observations. Importantly, the test  
 203 showed good power for the datasets simulated under an alternative hypothesis (scenario 3S), with close to  
 204 50% of tests being significant for a sample size as small as 100 animals newly released per occasion (i.e.  
 205 500 animals altogether) and close to 100% of the global test being significant for 250 animals released  
 206 per occasion. The simulation results show that the test reacts as expected from the derivation made in the  
 207 previous sections, when the partial observations are not generated by the directly observable states, and  
 208 that it can work well for realistic sample sizes. When part of the partial observations are generated by the  
 209 observable states, the test is not as powerful as could be expected but nonetheless rejects  $H_0$ .

## 210 4.2 Canada Geese

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

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

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

224 We examine the Canada geese dataset under both the null and alternative hypotheses by artificially creating  
225 these situations within the data. ~~First, in order to create partial observations generated by the observable~~  
226 ~~states ( $H_0$ ), we set some a varying percentage of the observed geese's states to unknown (MCAR). We~~  
227 ~~considered varying percentages to see how the test reacts to the amount of partial observations: 15%,~~  
228 ~~25% and 45%, so that the partial observations are generated only by the observable states ( $H_0$ ).~~ These  
229 situations are respectively denoted by MCAR15, MCAR25 and MCAR45 in Table 6. Then we examine  
230 situations that come under the alternative hypotheses ( $H_1$ ) by setting all of the observations from a  
231 particular state to “unknown” ~~so that this particular state becomes unobservable while the states remaining~~  
232 ~~observable do not generate any partial observations. We considered 2 situations: all observations in state~~  
233 ~~2 are set to “unknown” (situation 2PO), or all those in state 3 are set to “unknown” (situation 3PO).~~  
234 ~~Eventually, we considered the hybrid situation where, in addition to the partial observations generated by~~  
235 ~~the unobservable state 3 as in scenario 3PO, 25% then 45% of the observations generated by state 2 are also~~  
236 ~~set to partial: scenarios Hyb25 and Hyb45. Hence, for situations 2PO and 3PO the partial observations~~  
237 ~~stem, respectively, only from states 2 and 3. Note that the state set to “unknown” is never directly observed~~  
238 ~~in each of these situations. This allows us to test the performance of the test as we know that all unknown~~  
239 ~~events correspond to a different state from the ones “observed”.~~

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

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

## 5 DISCUSSION

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

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

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

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

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

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

## CONFLICT OF INTEREST STATEMENT

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

## AUTHOR CONTRIBUTIONS

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

## FUNDING

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

## ACKNOWLEDGMENTS

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

## DATA AVAILABILITY STATEMENT

295 The Canada goose data set used within this paper has been provided as a supplementary file.  
296 [The R code used to implement the test described in this paper has been provided as supplementary files.](#)

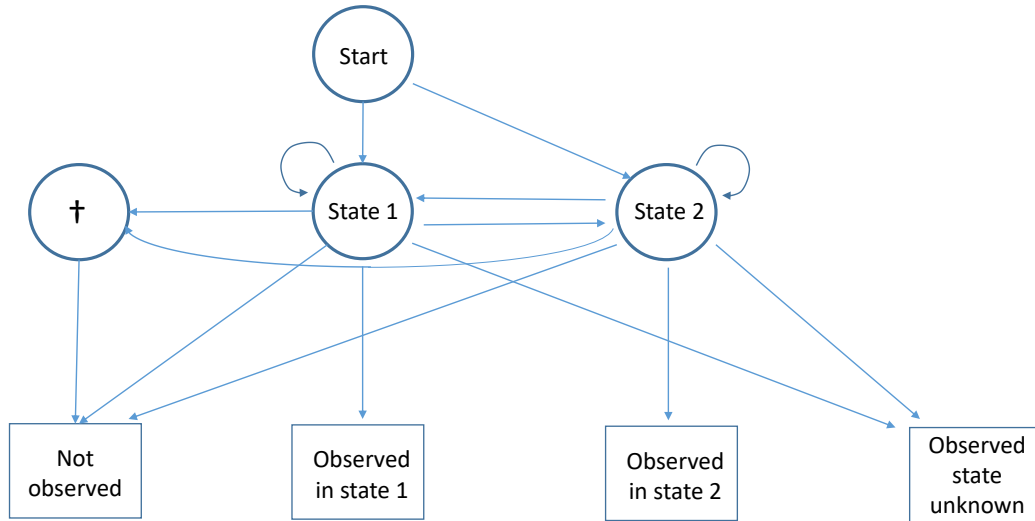
## REFERENCES

- 297 BARTOLUCCI, F. AND PENNONI, F. (2007) A class of latent Markov models for capture- recapture data  
298 allowing for time, heterogeneity and behavior effects. *Biometrics*. **63**, 568-578.  
299 CORMACK, R. M. (1964) Estimates of survival from sighting of marked animals. *Biometrika*. **51**, 429-438.

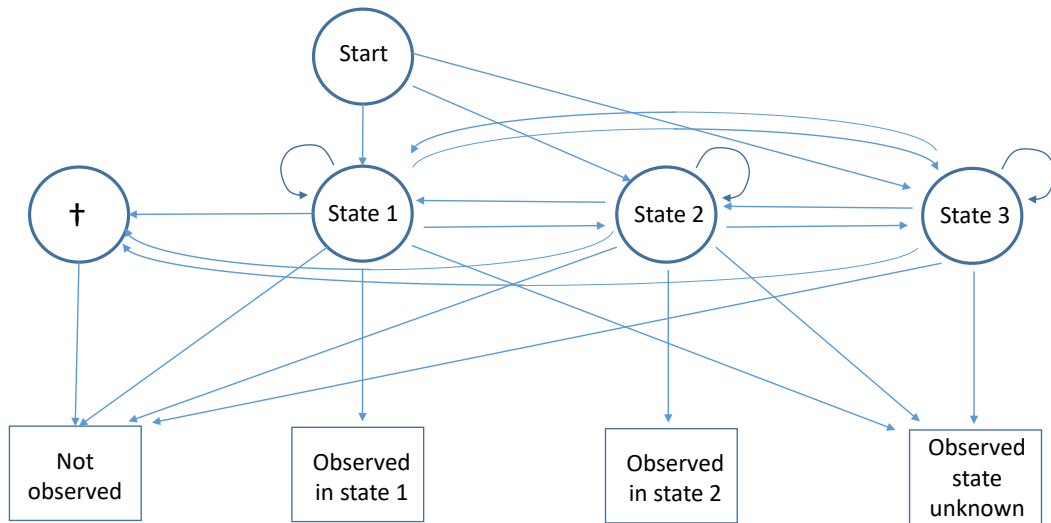
- 300 CRESSIE, N. AND READ, T. R. C. (1984) Multinomial goodness-of-fit tests. *Journal of the Royal*  
301 *Statistical Society. Series B (Methodological)*. **46**, 440–464.
- 302 CRESSIE, N. AND READ, T. R. C. (1988) *Goodness-of-fit statistics for discrete multivariate data*. Springer,  
303 New York, USA.
- 304 HESTBECK, J. B. AND NICHOLS, J. D. AND MALECKI, R. A. (1991) Estimates of movement and site  
305 fidelity using mark-resight data of wintering Canada geese *Animal Biodiversity and Conservation* **27**,  
306 97-107.
- 307 KENDALL, W. L. (2004) Coping with unobservable and mis-classified states in capture–recapture studies.  
308 *Ecology* **72**, 523–533.
- 309 KING, R. AND MCCREA, R. S. (2014) A generalised likelihood framework for partially observed  
310 capture–recapture–recovery models. *Statistical Methodology* **17**, 30–45.
- 311 LANGROCK, R., KING, R., MATTHIOPOULOS, J., THOMAS, L., FORTIN, D. AND MORALES, J. M.  
312 (2012) Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions.  
313 *Ecology*, **93**, 2336–2342.
- 314 LEBRETON, J. D. , NICHOLS, J. D. , BARKER, R. J. , PRADEL, R. AND SPENDELOW, J. A. (2009)  
315 Modeling individual animal histories with multistate capture–recapture models. *Advances in Ecological*  
316 *Research* **41**, 87–173.
- 317 LANGROCK, R. AND KING, R. (2013) Maximum likelihood estimation of mark-recapture-recovery models  
318 in the presence of continuous covariates. *Annals of Applied Statistics* **7**, 1709–1732.
- 319 MOORE, D.S. (1986) Tests of chi-squared type. In *Goodness-of-fit techniques* (Eds. D’Agostino, R. B.  
320 and Stephens, M. A. ) CRC Press, New York.
- 321 OTIS, D. L. , BURNHAM, K. P. , WHITE, G. C. AND ANDERSON, D. R. (1978). Statistical inference  
322 from capture data on closed animal populations. *Wildlife Monographs* **62**, 1–135.
- 323 PLEDGER, S. (2000) Unified maximum likelihood estimates for closed capture recapture models using  
324 mixtures. *Biometrics*. **56**, 434-442.
- 325 POHLE, J., LANGROCK, R. VAN BEEST, F. M. AND SCHMIDT, N. M. (2017). Selecting the number of  
326 states in hidden Markov models: pragmatic solutions illustrated using animal movement. *Journal of*  
327 *Agricultural, Biological, and Environmental Statistics* **22**, 270-293.
- 328 PRADEL, R. (2005). Multievent: An Extension of Multistate Capture-Recapture Models to Uncertain  
329 States. *Biometrics* **61**, 442–447.
- 330 PRADEL, R. AND WINTREBERT, C. M. A. AND GIMENEZ, O. (2003). A Proposal for a Goodness-of-Fit  
331 Test to the Arnason-Schwarz Multisite Capture-Recapture Model. *Biometrics* **59**, 43–53.
- 332 ROUAN, L. AND CHOQUET, R. AND PRADEL, R. (2009). A general framework for modeling memory in  
333 capture-recapture data. *Journal of Agricultural, Biological, and Environmental Statistics* **14**, 338–355.

- 334 WILLIAMS, B. K., NICHOLS, J. D. AND CONROY, M. J. (2002) *Analysis and management of animal*  
335 *populations*. Academic Press, San Diego, California, USA.
- 336 WORTHINGTON, H., MCCREA, R. S. , KING, R. AND GRIFFITHS, R. A. (2019) Estimating abundance  
337 from multiple sampling capture-recapture data using hidden Markov models in the presence of discrete-  
338 state information. *Annals of Applied Statistics*. In press.
- 339 YANG H. C. AND CHAO, A. (2005) Modeling animals behavioral response by Markov chain models for  
340 capture-recapture experiments. *Biometrics*. **61**, 1010-1017.
- 341 YANTIS, S. AND MEYER, D. E. AND SMITH, J. K. (1991). Analyses of multinomial mixture distributions:  
342 new tests for stochastic models of cognition and action. *Psychological Bulletin* **110**, 350.
- 343 ZHOU, M., MCCREA, R. S. , MATECHOU, E., COLE. D. J. AND GRIFFITHS, R. A. (2019) Removal  
344 models accounting for temporary emigration. *Biometrics* **75**, 24-35.
- 345 ZUCCHINI, W. AND MACDONALD, I. L. AND LANGROCK, R. (2016). *Hidden Markov models for time*  
346 *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.



**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  $S$ .

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

**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 ( $\cdot$  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.

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

**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  $r$ ,  $r = 1, \dots, R$ .  $M_r$  is the mixture corresponding to animals partially observed at  $i$  and most lately completely observed in state  $r$ ,  $r = 1, \dots, R$ . 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  $B_r$ ,  $i = 1, \dots, R \times (T - i)$ ,  $r = 1, \dots, R$ .  $\gamma_r$ ,  $r = 1, \dots, R$ , are the mixing weights for  $M_1$ .  $\pi_r$ ,  $r = 1, \dots, R$ , are the mixing weights for  $M_R$ .

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

**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  $\geq 2$ ).  $G$  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)	
2S $H_0$	5000	45	2	4.83	600	
			3	3.33	600	
			G	5.00	600	
		25	2	3.50	600	
			3	4.17	600	
			G	4.50	600	
	45	2	4.77	600		
		3	3.63	600		
		G	4.93	600		
	3S $H_1$	100	-	2	26.44	435
			-	3	45.84	373
			-	G	42.73	550
250		-	2	<b>52.75</b>	582	
		-	3	<b>92.60</b>	581	
		-	G	<b>94.33</b>	600	
500		-	2	<b>83.47</b>	599	
		-	3	<b>99.83</b>	600	
		-	G	<b>100.00</b>	600	
1000		-	2	<b>96.50</b>	600	
			3	<b>100.00</b>	600	
			G	<b>100.00</b>	600	
	30	2	<b>76.30</b>	600		
		3	<b>97.50</b>	600		
		G	<b>99.80</b>	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	$df$
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	