

## ARTICLE

# Using multistate mark-recapture mixture models to account for latent subpopulation structure in survival and state transition probabilities: North Atlantic right whales<sup>†</sup>

Paul B. Conn<sup>1\*</sup>, David L. Miller<sup>2</sup>, Devin S. Johnson<sup>1</sup>, Richard M. Pace III<sup>2</sup>, Paul R. Wade<sup>1</sup>, Kenady Wilson<sup>1</sup>, and Peter J. Corkeron<sup>2</sup>

<sup>1</sup>Marine Mammal Laboratory, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, 7600 Sand Point Way NE, Seattle, WA 98115 USA; <sup>2</sup> Northeast Fisheries Science Center, National Marine Fisheries Service, NOAA NOAA, 166 Water Street, Woods Hole, MA 02543 USA

---

## Summary

1. Animal ecologists often use multistate mark-recapture models to analyze survival and state transition probabilities in individually identifiable animal populations. These models often assume that individuals belong to a single population with homogeneous vital rates and state transition probabilities, except where static individual covariates (e.g. sex) can be used to partition the population.
2. Some populations are comprised of several population subgroups subscribing to different migratory strategies. In such cases, traditional multistate mark-recapture models will only provide estimates of mean population dynamics relative to the marked sample. Further, heterogeneity in detection can be expected, and may bias inference.
3. We develop a multistate mark-recapture modeling framework based on finite mixtures that can be used to relax the homogeneity assumption of traditional multistate models. This framework, together with commonly available model selection and averaging tools, can be used to identify a reasonable number of population subgroups and investigate their relative dynamics (i.e., survival and state transition probabilities). We use simulation to verify that parameters can be identified within this framework.
4. We apply our approach to North Atlantic right whales (*Eubalaena glacialis*) on the east coast of North America. This population is of considerable conservation interest, but difficult to study because individuals have different migratory strategies that make detection probabilities highly heterogeneous.

Word count: xxx

**Key-words:** clustering, *Eubalaena glacialis*, mixture model, multistate mark-recapture, population substructure, survival, North Atlantic right whale

\*Correspondence author. E-mail: paul.conn@noaa.gov

## 1 Introduction

2 Ecologists often use multistate mark-recapture models (hereafter, MSMR models [Hestbeck et al. 1991](#); [Brownie et al. 1993](#); [Schwarz et al. 1993](#)) to study survival and state transition probabilities in wild animal populations. State transition probabilities can include  
3 movement among geographical strata, or transitions among other ecologically relevant states (e.g., disease status or breeding state).  
4 Importantly, MSMR models can be used to parameterize population models needed for conservation and management ([Nichols et al. 1992](#); [Caswell 2001](#)).

5 One assumption of canonical MSMR models is that individuals within a given state (geographical strata, disease state,  
6 breeding/nonbreeding) have homogeneous survival and state transition probabilities. This assumption can be relaxed if there are  
7 static covariates (e.g. sex) that can be used to partition the marked population into groups. For instance, different state transition  
8 matrices could be estimated for males and females if sex was observable.

9 For many populations, latent subpopulation structure may affect survival and state transition probabilities. For instance, many  
10 species of waterfowl, passerines, cervids, and salmonids are composed of both migratory and resident individuals, and it may be  
11 difficult to discriminate between the two groups based on appearance alone. One possibility in such cases is to study populations in  
12 locations or times where overlap of subpopulations is minimized (e.g. studying residents in the winter when migrants are not present  
13 [Hestbeck et al. 1991](#)). However, this is not always possible.

14 Various model-based solutions for dealing with subpopulation overlap have been proposed, depending on the unique sampling  
15 situation. For instance, researchers have developed mixture models to account for the presence of transients and/or migrants while  
16 estimating survival ([Pradel et al. 1997](#); [Fieberg & Conn 2014](#)) and abundance ([Conn et al. 2011](#)) with capture-recapture data. These  
17 models allow uncertainty about class membership (e.g. resident vs. transient) when conspecifics of both types can be sampled in the  
18 same location. However, to our knowledge, the general question of how to fit multistate mark-recapture models with multiple latent  
19 population subgroups has not been investigated. One possibility is to formulate a hidden Markov modeling framework for such data,  
20 where group membership of individuals with different patterns of encounters are modeled probabilistically. This idea is similar to the  
21 finite mixture modeling framework sometimes used to account for heterogeneity in detection when estimating single site abundance  
22 or survival (cf. [Pledger 2000](#); [Pledger et al. 2003, 2010](#)).

23 The endangered North Atlantic right whale (*Eubalaena glacialis*; hereafter, NARW) are one population which would benefit  
24 from such a generalized multistate analysis. NARW survival has historically been studied using Cormack-Jolly-Seber (CJS) mark-  
25 recapture models ([Caswell et al. 1999](#)) and traditional MSMR models ([Fujiwara & Caswell 2002a,b](#)) applied to NARW photo-  
26 identification records. Results of such analyses were important for establishing a decline in adult survival from the early 1980s to late  
27 1990s, and for raising conservation concerns about population viability ([Caswell et al. 1999](#); [Fujiwara & Caswell 2001](#)).

28 Although CJS and MSMR estimates have proved useful for conservation and management, there has been some speculation that  
29 the analyses conducted with these models may not adequately capture heterogeneity in resighting and availability patterns ([Clapham et al. 2002](#)). For instance, original estimates of survival from CJS models ([Caswell et al. 1999](#)) evidenced strong support for models  
30 in which capture probability was a logit-linear function of an "offshore index" - the proportion of years for which a particular whale  
31 had only been observed in offshore sites. Such sites received less survey effort than sites closer to shore, particularly in the mid-1990s.  
32 In an unpublished manuscript, [Wade & Clapham \(Unpublished\)](#) suggested that Caswell et al.'s (1999) approach may have not been  
33 sufficient to adequately address heterogeneity in the seasonal patterns of attendance exhibited by whales with different migratory  
34 strategies. Instead, Wade and Clapham (Unpublished) employed cluster analysis to group whales with similar migratory behavior  
35 prior to running a CJS-based survival analysis, finding support for models with four clusters of whales with distinct movement and  
36 observation patterns. These groups also exhibited biologically meaningful differences in survival.

37 Despite the importance of detection heterogeneity on survival, which can result in negatively biased survival estimates ([Abadi et al. 2013](#)), there has been little published work in the primary literature investigating the consequences of detection heterogeneity  
38 on NARW survival since the first estimates of [Caswell et al. \(1999\)](#). Part of the reason for this might have to do with violation of  
39 statistical mechanics. Specifically, it is generally not appropriate to construct a predictive covariate for use in a statistical model based  
40 on knowledge of the response variable. For instance, the clustering approach used by [Wade & Clapham \(Unpublished\)](#) could possibly  
41 confound detection patterns with mortality if whales that died and had fewer detections were more often assigned to one particular  
42 group.

cluster over another. The “offshore index” covariate used by Caswell *et al.* (1999) also suffers from this issue. For instance, animals who are seen once by definition have an index covariate of 0.0 or 1.0, and will be assigned either maximum or minimum survival.

Recent developments in mark-recapture modeling suggest a viable alternative. For instance, capture-recapture data are increasingly analyzed using hidden Markov models (HMMs Zucchini & MacDonald 2009) that allow the detection process to be linked to survival and movement processes in a more general way. First applied to capture-recapture data by Pradel (2005), these models decouple detection types from underlying “states” of the animal and formally permit uncertainty about an animal’s underlying state. Recently, Johnson *et al.* (2016) extended this modeling type to multivariate states, which may or may not interact (e.g., location and tagging state of sea lions). We propose to use this framework for estimation of survival and state transition probabilities when the membership among different latent population groups is unknown.

This paper is structured as follows. First, we introduce a general framework for conducting multistate mark-recapture analysis with multiple population subgroups. Second, we use simulation to verify that our estimation approach can recover the survival and state transition matrices used to generate data. Next, we apply our modeling framework to a photograph-based NARW mark-recapture dataset. Finally, we discuss relevance of our research - both as a general modeling framework as well as specific implications for NARW conservation and management.

## Materials and methods

### MODEL DEVELOPMENT

#### Sampling

As with canonical MSMR models, we suppose that the investigator individually identifies animals through artificial or natural markings at a sequence of discrete sampling occasions (call these  $t \in \{1, 2, \dots, T\}$ ). Animals may be encountered in one of  $s \in \{1, 2, \dots, S\}$  states, which may be geographical locations or dynamical states such as disease or breeding status. For instance, if sampling is conducted in three geographical states (A, B, and C) for 5 sampling occasions, the encounter history

$$H_i = 0AB0C$$

would indicate an animal that was first encountered in state A at time 2, was observed in state B at time 3, was not observed at time 4, and was observed in state C at time 5. Our objective will be to estimate survival, state transition probabilities, and detection probabilities from these data when there is latent subpopulation structure.

#### Notation and basic model structure

We describe multistate mixture models in similar notation to that typically used for MSMR data (Table 1). As with other extensions of the CJS model, we condition on the time of first capture when modeling the probability of each encounter history. We assume that mixture (i.e. subpopulation) membership does not change over time; thus, transition probabilities among states are governed by mixture-specific survival and state transition probabilities. However, we must account for initial state probabilities since there is uncertainty about mixture membership. For instance, we write the probability of the history in Eq. 1 as

$$\Pr(H_i = 0AB0C) = \sum_m \pi_{A,2,m} \phi_{2,m}^{AB} p_{B,2,m} \left[ \sum_s \psi_{3,m}^{B,s} (1 - p_{B,3,m}) \psi_{4,m}^{s,C} \right] p_{C,5,m}.$$

Statistical inference about the parameters in Table 1 can be performed using a product multinomial likelihood function,

$$L = \prod_i \Pr(H_i).$$

Efficient evaluation of this likelihood can be accomplished using the so-called forward algorithm used to fit HMMs (Zucchini & MacDonald 2009; Laake 2013).

## Model assumptions

No switching among population subgroups; “instantaneous” sampling

## Derived parameters

## Computing

Several software packages are available to fit hidden Markov models to capture-recapture data, either in a univariate (?; e.g. E-SURGE) or a multivariate state framework (marked Laake et al. 2013; Johnson et al. 2016). Analyses reported in this paper use the marked package and can be implemented in the R programming environment (?).

## SIMULATION STUDY

## EXAMPLE: NORTH ATLANTIC RIGHT WHALES

## Results

## Discussion

## Conclusion

## Acknowledgements

Views expressed are those of the authors and do not necessarily represent findings or policy of any government agency. Use of trade or brand names does not indicate endorsement by the U.S. government.

## Data accessibility

R scripts and data necessary to recreate analyses have been collated into an R package, which is currently available at <https://github.com/pconn/MSmixture>. We plan to publish the package to an online archive/repository upon acceptance.

## References

- Abadi, F., Botha, A. & Altwegg, R. (2013) Revisiting the effect of capture heterogeneity on survival estimates in capture-mark-recapture studies: does it matter? *PloS one*, **8**(4), e62636.
- Brownie, C., Hines, J.E., Nichols, J.D., Pollock, K.H. & Hestbeck, J.B. (1993) Capture-recapture studies for multiple strata including non-markovian transitions. *Biometrics*, **49**, 1173–1187.
- Caswell, H. (2001) *Matrix Population Models, 2nd Edition*. Sinauer, Sunderland, MA.
- Caswell, H., Fujiwara, M. & Brault, S. (1999) Declining survival probability threatens the north atlantic right whale. *Proceedings of the National Academy of Sciences*, **96**(6), 3308–3313.
- Clapham, P., Brault, S., Caswell, H., Fujiwara, M., Kraus, S., Pace, R. & Wade, P. (2002) Report of the working group on survival estimation for north atlantic right whales. *Available from the Northeast Fisheries Science Center*, **166**.
- Conn, P.B., Gorgone, A.M., Jugovich, A.R., Byrd, B.L. & Hansen, L.J. (2011) Accounting for transients when estimating abundance of bottlenose dolphins in choctawhatchee bay, florida. *Journal of Wildlife Management*, **75**, 569–579.
- Fieberg, J.R. & Conn, P.B. (2014) A hidden markov model to identify and adjust for selection bias: an example involving mixed migration strategies. *Ecology and evolution*, **4**(10), 1903–1912.
- Fujiwara, M. & Caswell, H. (2001) Demography of the endangered north atlantic right whale. *Nature*, **414**(6863), 537–541.
- Fujiwara, M. & Caswell, H. (2002a) Estimating population projection matrices from multi-stage mark–recapture data. *Ecology*, **83**(12), 3257–3265.
- Fujiwara, M. & Caswell, H. (2002b) A general approach to temporary emigration in mark–recapture analysis. *Ecology*, **83**(12), 3266–3275.
- Hestbeck, J.B., Nichols, J.D. & Malecki, R.A. (1991) Estimates of movement and site fidelity using mark-resight data of wintering canada geese. *Ecology*, **72**, 523–533.
- Johnson, D.S., Laake, J.L., Melin, S.R., DeLong, R.L. et al. (2016) Multivariate state hidden markov models for mark-recapture data. *Statistical Science*, **31**(2), 233–244.
- Laake, J.L. (2013) Capture-recapture analysis with hidden Markov models. AFSC Processed Rep 2013-04, Alaska Fisheries Science Center, NOAA, National Marine Fisheries Service, Sand Point Way NE, Seattle WA 98115 USA.
- Laake, J.L., Johnson, D.S. & Conn, P.B. (2013) marked: an r package for maximum likelihood and markov chain monte carlo analysis of capture–recapture data. *Methods in Ecology and Evolution*, **4**(9), 885–890.
- Nichols, J.D., Sauer, J.R., Pollock, K.H. & Hestbeck, J.B. (1992) Estimating transition probabilities for stage-based population projection matrices using capture-recapture data. *Ecology*, **73**, 306–312.
- Pledger, S. (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics*, **56**, 434–442.
- Pledger, S., Pollock, K.H. & Norris, J.L. (2003) Open capture-recapture models with heterogeneity: I. cormack-jolly-seber model. *Biometrics*, **59**(4), 786–794.
- Pledger, S., Pollock, K.H. & Norris, J.L. (2010) Open capture–recapture models with heterogeneity: II. jolly–seber model. *Biometrics*, **66**(3), 883–890.
- Pradel, R. (2005) Multievent: An extension of multistate capture-recapture models to uncertain states. *Biometrics*, **61**, 442–447.

- 127 Pradel, R., Hines, J.E., Lebreton, J.D. & Nichols, J.D. (1997) Capture-recapture survival models taking account of transients. *Biometrics*, 60–72.
- 128 Schwarz, C.J., Schweigert, J.F. & Arnason, A.N. (1993) Estimating migration rates using tag-recovery data. *Biometrics*, **49**, 177–193.
- 129 Wade, P.R. & Clapham, P.J. (Unpublished) The influence of spatial distribution on survival estimates: use of cluster analysis to address heterogeneity in north atlantic right
- 130 whales. *Unpublished manuscript*.
- 131 Zucchini, W. & MacDonald, I.L. (2009) *Hidden Markov models for time series: an introduction using R*, vol. 150. CRC press.

**Table 1.** Definitions of parameters and statistics used in the multistate mixture model.

Quantity	Definition
<b>A. Statistics</b>	
$H_i$	Encounter history for individual $i$
$S$	Number of states (with state $S$ being reserved for the absorbing state, “dead”)
$T$	Number of sampling occasions
$M$	Number of mixtures (population subgroups)
<b>B. Parameters</b>	
$\pi_{s,t,m}$	Probability that an individual first captured in state $s$ at time $t$ is a member of population subgroup $m$
$\theta_{s,t,m}$	Probability that an animal alive and in state $s$ at time $t$ and from population subgroup $m$ is alive at time $t + 1$ .
$\psi_{t,m}^{s_1,s_2}$	Probability that an animal of population subgroup $m$ that is alive and in state $s_1$ at time $t$ is in state $s_2$ at time $(t + 1)$ , given that it survives from $t \rightarrow (t + 1)$
$\phi_{t,m}^{s_1,s_2}$	It will sometimes be convenient to use the product $\phi_{t,m}^{s_1,s_2} = S_{s,t,m} \psi_{t,m}^{s_1,s_2}$
$p_{s,t,m}$	Probability that a member of population subgroup $m$ is detected at time $t$ given that they are in state $s$ .