

Incorporating individual variability into mark–recapture models

Jessica H. Ford^{1,2*}, Mark V. Bravington³ and Jooke Robbins⁴

¹CSIRO Wealth from Oceans National Research Flagship, Marine Laboratory, Castray Esplanade, Hobart, 7001, TAS, Australia; ²Institute for Marine and Antarctic Studies, University of Tasmania, Sandy Bay, 7005, TAS, Australia; ³CSIRO Mathematics, Informatics, and Statistics, Marine Laboratory, Castray Esplanade, Hobart, 7001, TAS, Australia; and

⁴Provincetown Center for Coastal Studies, 5 Holway Avenue, Provincetown, MA, 02657, USA

Summary

1. Understanding individual variation is a key challenge in ecology. Inherent individual differences in movement and behaviour pose fundamental problems in the analysis of mark–recapture data as unmodelled individual differences can bias estimates of population size and survival rates. Multi-state mark–recapture models have been the focus of much recent research but have yet to explicitly incorporate individual variability.

2. We use a multi-state mark–recapture model with individual-level random effects, built in ADMB-RE, a software tool that automatically provides an accurate analytical approximation of the likelihood which is otherwise intractable. We tested the model using simulation studies and applied the model to data from North Atlantic humpback whales in the Stellwagen Bank National Marine Sanctuary where heterogeneity is apparent in both sighting probability and site preference.

3. Simulation studies demonstrated accurate estimation of true parameter values with random effects models but bias sometimes resulted from fitting simpler models.

4. In application to data from the North Atlantic humpback whales, we were able to estimate both annual variation in the local population and three measures of individual-level variation. Results indicate considerable heterogeneity within this population in both sighting probability and site preference. Ignoring random effects led to bias in estimates of proportion of time within a marine reserve.

Key-words: hidden Markov model, Individual heterogeneity, marine reserve, mark–recapture, multi-state, North Atlantic humpback whales

Introduction

Variability between individuals is an important statistical and ecological issue for mark–recapture studies. Individual heterogeneity within species can include factors such as prey preference, foraging techniques (Bolnick *et al.* 2003) and inherent individual differences in behaviour (Hammond 1990). If measurable individual-level covariates (e.g. sex, age) adequately explain individual variability, standard covariate-based extensions of mark–recapture can be used. However, it is well known that abundance estimates are biased if there is unmodelled heterogeneity in capture probability between individuals (Seber 1982). This paper describes a statistical and computational approach to handling such unmodelled heterogeneity, tests it using simulations and explores it through application to real data.

Heterogeneity and its representation in mark–recapture analysis through the incorporation of random effects have received much attention (Burnham & Overton 1978; Huggins & Yip 2001; Barry *et al.* 2003; Maunders *et al.* 2008;

Royle & Dorazio 2008; Lebreton *et al.* 2009), yet few models have included individual-level random effects (Pledger, Pollock & Norris 2003; Royle 2008; Gimenez & Choquet 2010). In many mark–recapture studies, including ours, some capture histories are too short, and the number of individuals too large to allow latent heterogeneity to be modelled as individual-level fixed effects. We use the term latent heterogeneity to refer to individual heterogeneity that is not described by any covariate in the model. One approach to modelling individual heterogeneity is to use a discrete-valued prior supported on a pre-specified number of points (Pledger & Phillpot 2008). However, it may be more appropriate to assume a continuous distribution for individual-level random effects, for example with a Normal prior whose variance reflects between-individual variability. In this case, the variance needs to be estimated, which entails integration across all possible values of the individual random effects. This cannot be made analytically, so an approximate technique is required. Markov chain Monte Carlo is one possibility, but can be difficult to compute. Instead, our focus here is to show that automatic (albeit approximate) maximum-likelihood estimates of individual heterogeneity can readily be obtained with the software Automatic Differentiation Model

*Correspondence author. E-mail: jessica.ford@csiro.au

Builder (ADMB) (Skaug & Fournier 2006; Fournier *et al.* 2012), which performs the necessary integration using Laplace approximation.

Multi-state mark-recapture models, first developed by Arnason (1972, 1973), extend the traditional Cormack–Jolly–Seber models by allowing animals to be in different ‘states’. Individual animals are allowed to transition from one state to another through time, and different states are often associated with different detection probability. Multi-state mark-recapture models have been the focus of much research in recent years, and the publication by Lebreton *et al.* (2009) provides a thorough synthesis. Multi-state models are expressed in the form of a hidden Markov model. Hidden Markov models are a class of state-space models, with discrete rather than continuous underlying hidden states, which can be defined as any model with an observation and state process, where the true dynamics of the system are not directly observed (Zucchini, Raubenheimer & MacDonald 2008; Conn & Cooch 2009).

Our motivating example is a study of a subpopulation of North Atlantic humpback whales (North Atlantic HW) sighted in the Stellwagen Bank National Marine Sanctuary (SBNMS) off the northeast coast of the United States. Individual humpback whales have been intensively studied in this region since the late 1970s. The SBNMS (Fig. 1) is one of the several feeding sites of North Atlantic HW that summer in the Gulf of Maine. The SBNMS covers only a small part of the population’s summer range, and although some individuals are seen regularly there during the summer, none are thought to remain permanently within its boundaries. This presents a challenge when studying the vital rates of whales using the area and the effectiveness of management initiatives.

Two primary sources of heterogeneity exist for the North Atlantic HW in this study: sighting probability and site fidelity. Heterogeneity in sighting probability is a well-known phenomenon for North Atlantic HW (Hammond 1986, 1990). It is also plausible that individual whales vary in their propensity to use the SBNMS as opposed to other parts of the feeding range. In the Methods section, we present a hidden Markov model that allows for individual heterogeneity in sightability and/or site fidelity.

When contemplating fitting a variety of models, there are always costs or trade-offs to consider: fitting a model that is too simple may result in increased bias or variance, whereas fitting a model that is too complex may result in high prediction error or computational difficulties. One aspect of complexity is how many model parameters are allowed to have latent heterogeneity, that is, individual-level random effects. In this paper, we use simulations of a simplified version of our SBNMS situation, to investigate the costs and trade-offs associated with fitting too complex or too simple a model by testing the effect of model mis-specification (i.e. mistakenly assuming no latent heterogeneity, or assuming latent heterogeneity when none is in fact present). We also examine the effect of assuming a Normal prior on individual variability when the true variability is discrete and bimodal.

Materials and methods

This section actually describes our model for simulated data, although it is closely aligned to the real SBNMS situation. Analysing the real data entails attention to a few extra details, omitted here for clarity but described in a subsequent section.

As usual with hidden Markov or multi-state models, the overall model is split into a process part and an observation part. For the process model, we assume that at time t an animal i can be in any of three states S_{it} : Here (in SBNMS), Away (outside it) and Dead, or H/A/D for short. As usual in mark-recapture, ‘Dead’ includes ‘permanent emigration’: in this case, the possibility that an animal has switched allegiance to a completely different summer feeding area, such as off Iceland. Changes in the state over time are governed by a Markov process with transition matrix γ , so (omitting dependence on i for now) for any two states s and s^* , we have

$$\mathbb{P}[S_{t+1} = s^*] = \gamma^{ss^*} \mathbb{P}[S_t = s]$$

The nine elements of γ can be written in terms of just three parameters γ^{HH} , γ^{AA} and γ^D (respectively the probabilities of staying *Here*, staying *Away* and *Dying*) as follows:

$$\gamma = \begin{pmatrix} \gamma^{HH}(1-\gamma^D) & (1-\gamma^{HH})(1-\gamma^D) & \gamma^D \\ (1-\gamma^{AA})(1-\gamma^D) & \gamma^{AA}(1-\gamma^D) & \gamma^D \\ 0 & 0 & 1 \end{pmatrix}$$

where we assume that probability of death (which is very low relative to the other transition rates) does not depend on whether the animal is *Here* or *Away*. Note that γ can depend on covariates, including individual-level and time.

For the observation model, there are ‘capture attempts’ (photo-ID expeditions) at each t in which an animal *may* be seen if and only if it is *Here*. Our data for animal i are thus a time series $Y_{it}; T$ of 0s (not seen) and 1s (seen) where t_{1i} denotes the first observation of the animal and T the most recent expedition. If $Y_{it} = 1$ then we know $S_{it} = H$, but if $Y_{it} = 0$, the state cannot be determined for certain. Formally, the probability of observation given state is expressed in terms of a parameter π by

$$\mathbb{P}[Y_{it} = 1 | S_{it} = s] = \begin{cases} \pi_{it} & s = H \\ 0 & s \in \{A, D\} \end{cases}$$

$$\mathbb{P}[Y_{it} = 0 | s] = 1 - \mathbb{P}[Y_{it} = 1 | s]$$

Although there is potentially some information in the ‘pre-history’ of an animal before it is first seen, in practice there is little gain and much pain from including it; an animal may not be seen simply because it has not yet recruited to the population, and our focus here is not on the recruitment process. We therefore start each animal’s series at its first sighting and condition on $S_{t_{1i}} = 1$.

The likelihood of the observed history for each animal is constructed as for a multi-state mark-recapture model. If the entire state history of the animal was known, that is, the true state at each time, then it would be straightforward to compute the probabilities of each observation given each state, which could then be multiplied to get the probability of the observation history itself. The probability of the state history could then be computed from the transition matrices, because the probability distribution of each state depends only on the previous state, which is known. The likelihood of the joint data (state history, observation history) could be computed as $\mathbb{P}[\text{observation history} | \text{state history}] \times \mathbb{P}[\text{state history}]$.

However, because the state history here is unknown, it is conceptually necessary to sum over the entire collection of possible state histories. This is completely infeasible computationally, but because of the

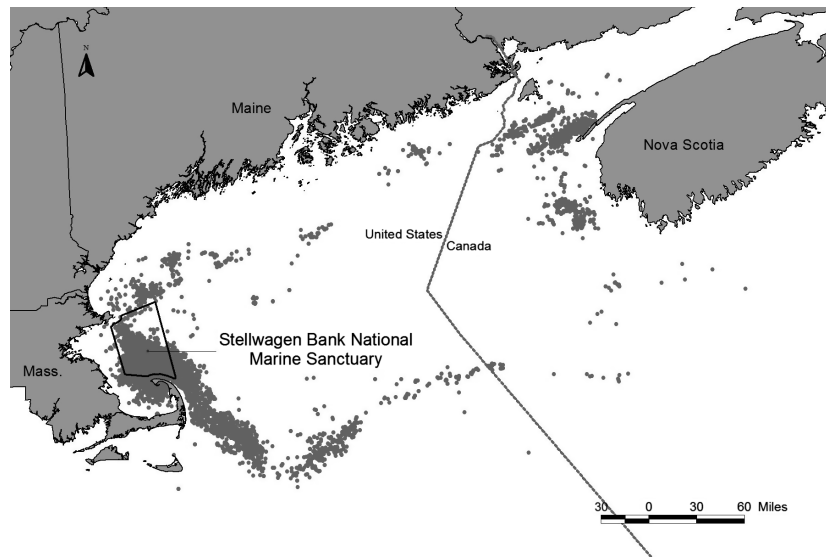


Fig. 1. Stellwagen Bank National Marine Sanctuary (study site) in relation to the greater Gulf of Maine humpback whale feeding ground. Grey circles represent sightings of individually identified humpback whales made by the Provincetown Center for Coastal Studies, 1979–2005. Sightings are not corrected for effort.

assumed Markov nature of the state-switching process, it turns out that there is an efficient way to compute the overall probability sequentially, by iterating a sequence of pairs of computations: compute the probability distribution of the state S_{it} given all the observation history up to the previous time step, $y_{i1:t-1}$, and then compute the probability of the observation y_{it} for each possible value of S_{it} .

Formally, the log-likelihood L_i for one animal i can therefore be written in terms of its entire data y_i and its transition and observation probabilities $\pi_i \triangleq \{\pi_{i1}, \dots, \pi_{iT}\}$ and γ_i (defined analogously) as follows:

$$L_i(\pi_i, \gamma_i; y) = \sum_{t=t_{i1}}^T \log \sum_{j \in \{HAD\}} \mathbb{P}[y_{it} | S_{it} = j; \pi_i] \mathbb{P}[S_{it} = j | y_{i1:t-1}; \pi_i, \gamma_i]$$

where $\mathbb{P}[S_{i1} = H] = 1$, and the term $\mathbb{P}[S_{it} = j | y_{i1:t-1}; \pi_i, \gamma_i]$ is computed successively over t , using standard and straightforward equations for hidden Markov models; see Zucchini, Raubenheimer & MacDonald (2008) and Patterson *et al.* (2009) for details. Animals are assumed to have independent histories conditional on parameters, so the log-likelihood for all animals is simply $L = \sum_i L_i$.

To actually fit these models, we need to express the π and γ parameters in terms of underlying parameters. This can be carried out similarly to a GLM or GLMM, using logit-links:

$$\begin{aligned} \text{logit} \pi_{it} &= \alpha_i^\pi + X_{it}^\pi b^\pi \\ \text{logit} \gamma_{it}^{ss} &= \alpha_i^{ss} + X_{it}^{ss} b^s \end{aligned}$$

where $s \in \{H, A\}$, α_i describes latent heterogeneity via an individual-level effect (which might be set at zero, a fixed effect, or a random effect), X is a design matrix and b are population-level fixed effects. For parameters that are constant across time, the corresponding X is just a column of 1s.

If no random effects are used, then it is straightforward to re-express L in terms of $\{\alpha, b, \gamma^D\}$ (i.e. including all combinations of α and γ) and then to estimate those parameters by maximum likelihood. However, random effects require a different approach. First, extra parameters are required for the variance of the random effects, such that:

$$\alpha_i^R \sim N(0, \sigma_R^2)$$

where $R \in \{\pi, H, A\}$. The joint log-posterior of the data and random effects becomes

$$L_i^*(\alpha_i, b, \gamma_D, \sigma^2; y_i) = L_i(\alpha, b, \gamma_D; y_i) + \ell(\alpha_i | \sigma^2) \quad \text{eqn 1}$$

where $\ell(\cdot)$ is the Normal log-density, which will have 1–3 terms depending on which of π , γ^{HH} and γ^{AA} have associated random effects. Again, data from animals are conditionally independent given the parameters, so $L^* = \sum_i L_i^*$. Unfortunately, this joint log-posterior L^* is not directly suitable for likelihood-based inference, essentially because the usual asymptotic approximations do not apply; for example, the maximum is always at $\hat{\sigma}^2 = 0$. One could choose to tackle the whole problem by MCMC, but if one prefers to follow the likelihood-maximization paradigm, then it is necessary to first ‘marginalize’ eqn (1) by integrating out the random effects α , to leave $L^{**}(b, \gamma^D, \sigma^2)$, which turns out to have behave well under maximization. Once the random effects variance(s) have been estimated, respectable inferences for α (and b, γ^D) can be made conditional on the point estimate $\hat{\sigma}^2$ (but see Discussion). Unfortunately, the integral is analytically intractable, so some form of approximation is required to estimate b, γ^D and σ^2 by maximum likelihood.

Automatic Differentiation Model Builder with Random Effects (ADMB-RE) uses an approximation of the integral to fit nonlinear random effects models with multiple parameters. The program combines a flexible mathematical modelling language (C++) with a powerful function minimizer based on automatic differentiation (Fournier *et al.* 2012). Briefly, the Laplace approximation uses a Taylor expansion of random effects distributions around their posterior mode within the likelihood to analytically approximate the integral. For further information, we direct the readers to the paper by Fournier *et al.* (2012), who provide a thorough synthesis of theoretical and computational aspects of ADMB.

SIMULATION OF SYNTHETIC DATA SETS

We used simulated data to test how well random effects could be estimated, and what the implications might be of choosing to model them or not. We used varying length of observation history (50, 100, 250) for 50 individuals. Although for some applications a series of 250 might

seem unreasonably long, the informativeness of any given data series is influenced by the number of transitions.

We constructed six models with increasing complexity. Model 0, our base model, has fixed effects only. Models 1–5 have random effects on the parameters as indicated by an ‘RE’ in Table 1.

We refer to the simulated data corresponding to each model as data 0–5, and we tested all combinations of models 0–5 against data 0–5. In other words, we tested models that were ‘too simple’, ‘too complex’ or ‘just right’ for the data. Testing all combinations of data and models allows us to identify any costs associated with fitting the wrong model to the data.

To test the effects of sampling bias in the simulated data, we simulated 100 data sets for each combination of model, data and length of observation history. We used the following fixed parameter values:

Table 1. The following table indicates incorporation of individual-level random effects onto a combination of the three parameters π , γ^{HH} and γ^{AA} . RE denotes the presence of an individual-level random effects on the relevant parameter. This allows for individual-level latent heterogeneity on sighting probability, probability of remaining *Here* and probability of remaining *Away*.

	Seen (π)	Stay Here (γ^{HH})	Stay Away (γ^{AA})
Model 1	RE		
Model 2	RE	RE	
Model 3		RE	
Model 4			RE
Model 5	RE	RE	RE

$b^\pi = 0.8$, $\gamma^{HH} = 0.85$, $\gamma^{AA} = 0.95$ and γ^D varying with length of observation history to generate approximately 10 deaths at each length. Individual random effects were simulated from a $N(0, 0.4^2)$ distribution.

Further, to explore the ability of ADMB-RE to pick up non-Normal variability (e.g. bimodality in the posterior estimates of random effects), we simulated individual random effects to be one of the two discrete values: -1 and 1 .

APPLICATION TO REAL DATA

Individual HW can be identified from their natural markings, especially the ventral pigmentation of the flukes and the shape and size of the dorsal fin (Katona & Whitehead 1981). We used a photo-identification data set of 1147 individuals with near-daily sighting opportunities in this area from April through October, 1979–2005. Various forms of heterogeneity are known to exist within this subpopulation and are evident from a plot of observation histories for 47 whales first seen in 1979 (Fig. 2). We note that in this study, we have used data on the Gulf of Maine HW which are a consistent subset of the population of North Atlantic HW.

We analysed observation histories on a weekly basis, so that each whale in each week is either seen at least once or not seen at all. We note that given another data set and study question, this transition step could be, for example, hourly, daily or even annual.

Only data from the first sighting of each whale onwards are used. As sighting effort is focused in the middle of the year, we included all sightings from the 18th week of the year through to the 43rd week. We assumed that the probability of survival, P_{surv} , over the remaining

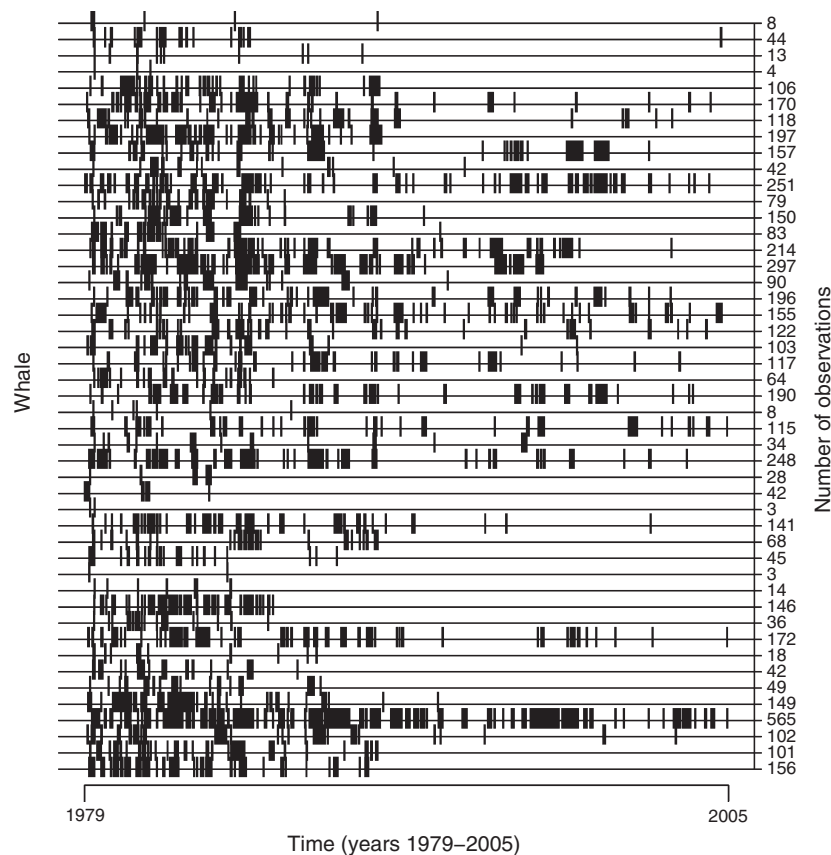


Fig. 2. Observation histories for 47 whales first seen in 1979. | Represent an observation.

26-week period was $P_{\text{surv}} = (1 - \gamma^D)^{26}$. We also assumed that state at the beginning of the season was independent of state at the end of the previous season (except for the Dead state), because almost all whales leave the SBNMS over winter. An extra parameter q was introduced for the chance of being present in the SBNMS at the start of the season. We calculated the probability of each state in the first week of the new year to be:

$$\mathbb{P}(S_t) = \begin{pmatrix} q * P_{\text{surv}} & (1-q) * P_{\text{surv}} & 1 - P_{\text{surv}} \\ q * P_{\text{surv}} & (1-q) * P_{\text{surv}} & 1 - P_{\text{surv}} \\ 0 & 0 & 1 \end{pmatrix} * \mathbb{P}(S_{t-1})$$

where $\mathbb{P}(S_{t-1})$ is the vector of state probabilities in the last week of the previous year.

We included individual-level random effects on each of π , γ^{HH} and γ^{AA} . We assume individual-level parameters to be consistent over time but have allowed for annual variation, at the population level, in γ^{HH} .

One important summary of an hidden Markov model allows us to characterize the proportion of time spent in the SBNMS. The stationary distribution (Zucchini, Raubenheimer & MacDonald 2008) of the underlying Markov chain for γ^{HH} and γ^{AA} indicates, conditional on not being dead, the long-term proportion of time spent in each state. Given our model structure, this allows us to characterize both population-level inter-annual and individual-level variability in usage of the SBNMS.

Results

SIMULATION RESULTS

All figures for simulation results are presented in Appendix S1.

Our simulation studies were designed to explore the costs associated with using too complex, or too simple, a model for the data at hand. We intentionally cross-compared the data and models to identify costs, such as increased bias or variance if we (for example) mistakenly modelled for no heterogeneity when it was in fact present.

From the simulation study on data 0 (only fixed effects), there did not appear to be any cost associated with fitting more complex models (models 1 through 5) to this data. Fixed parameter values were accurate at observation histories of length 50. Random effects were estimated with σ_π and σ_γ close to zero and converging to zero with increasing length of observation history. We note that for short series, there is a moderate bias in estimated survival rate when no random effects are fitted, even when there is no latent heterogeneity in the simulation. This bias disappears with longer series, so it appears to be a small-effect bias from maximum-likelihood estimation. In itself this is unsurprising given the imperfect observations and very small number of actual deaths in our simulations, but it is interesting that the bias appears to be removed by fitting random effects even when they should be superfluous. It does not seem advisable to attempt to generalize this conclusion beyond this study.

In fitting the most complex data (data 5 – with individual-level random effects on π , γ^{HH} and γ^{AA}), there appeared to be some cost for all but the true model. Fitting simpler models to this complex data resulted in bias in γ^D for the base model, consistent bias in γ^{HH} for model 1, and consistent bias in γ^{AA} for

models 1–3 as these models did not account for individual variation in γ^{AA} .

Across all simulations, at shorter time series, the models were occasionally unable to accurately assign random effects, resulting in $\hat{\sigma}_\pi$, $\hat{\sigma}_H$ or $\hat{\sigma}_A$ close to zero. This was less of a problem with increasing length of data series, as values converged close to true values. More complex models fit to simpler data did not lead to problems in estimation of fixed parameters and $\hat{\sigma}$ were well estimated in all cases with values converging to zero when they were not present in the data. In comparison, consistent bias was evident when simpler models were fit with the more complex data.

Despite the assumptions of a normal distribution to model variability, results for the exploratory simulation study using discrete random effects suggest that ADMB-RE is able to estimate bimodality from longer observation histories (Appendix S1). Fixed parameter estimates were unbiased at short-observation histories of 50, and density plots of the posterior estimates for the random effects indicate that the model was able to capture bimodality in the data with time series of 250. However, in our simulations, all animals have equally long histories: in real data sets, many may have short data series, and the resulting posterior modes would cluster around zero, which would tend to mask any bimodality.

During model exploration, we found state flipping rates were a major determinant of accuracy of parameter estimation. Longer observation histories were required for convergence with high-state flipping rates (e.g. $\gamma^{HH} \leq 0.6$ and $\gamma^{AA} \leq 0.6$), which was unsurprising given the lack of information in observation histories with such frequent transitions; conversely, we found convergence at shorter observation histories with infrequent transitions (e.g. $\gamma^{HH} \geq 0.9$ and $\gamma^{AA} \geq 0.9$). We did not experience convergence problems in these simulations; however, we note from exploratory work that increased variance in random effects resulted in an increased rate of non-convergence for the most complex models. This was most apparent at shorter capture histories when random effects were included on both the observation and process components of the model and is likely due to the fundamental difficulty in distinguishing between a whale that is rarely seen versus rarely there.

APPLICATION TO REAL DATA

In application to 1147 North Atlantic HW, we used a three-state model with individual-level random effects on π , γ^{HH} , and γ^{AA} and an annual population-level term on γ^{HH} . We found \hat{b}^π (population-level probability of observation) was estimated at 0.54, with the posterior modes of the observation component of individual-level variability $\hat{\sigma}_\pi = 0.72$. Figure 3 indicates considerable individual-level heterogeneity in probability of observation and ranged from 0.166 to 0.792 (1st Quantile: 0.423, mean 0.474, 3rd Quantile: 0.527).

For this population, the annual term on the first state ranged from a low of 0.119 in year 1995 to a high of 0.947 in year 1983 (Fig. 4). The posterior estimate for the individual-level variability on the first state was $\hat{\sigma}_H = 0.467$. Posterior estimates

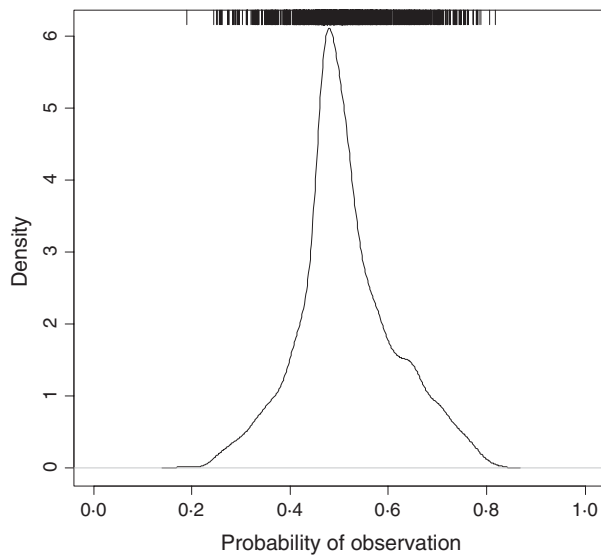


Fig. 3. Distribution of posterior modes of animal-level random effects for probability of observation for all 1147 North Atlantic humpback whales for the years 1979–2005. The rug plot indicates the 1147 individual values for probability of observation, with strong evidence of heterogeneity in probability of observation within this population.

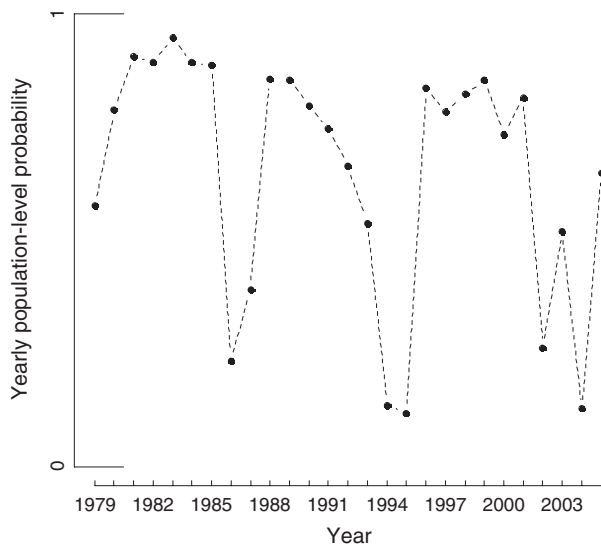


Fig. 4. Yearly population-level probability of γ^{HH} . The year 1983 recorded the highest population-level probability of γ^{HH} and 1995 the lowest.

for the second state, $\hat{\gamma}^{AA} = 0.976$, $\hat{\sigma}_A = 0.802$, indicate that once an animal leaves the SBNMS, they are likely to remain away for an extended period.

Taking the stationary distribution of the underlying Markov chain for γ^{HH} and γ^{AA} helps to describe the long-term behaviour of the animals. In our application to North Atlantic HW, the stationary distribution conditional on not being Dead, indicates the long-term proportion of time spent within the SBNMS. Figure 5 indicates both annual population-level heterogeneity and individual-level heterogeneity in the proportion of time present in SBNMS: 1983 has considerably higher

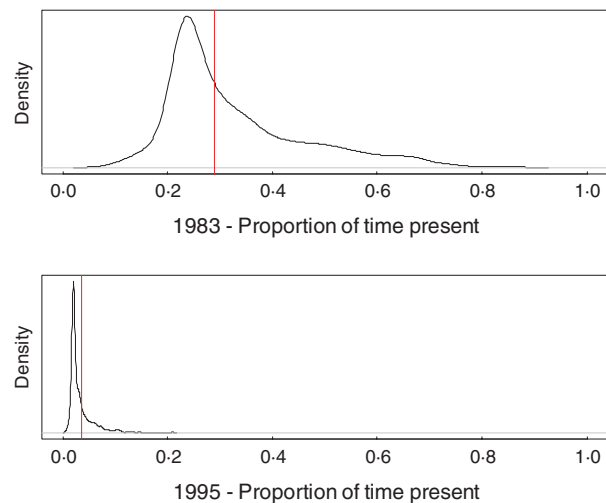


Fig. 5. The top figure indicates the individual-level variation in proportion of time spent in the Stellwagen Bank National Marine Sanctuary (SBNMS) in 1983 for all 1147 North Atlantic humpback whales, and the bottom figure the proportion of time spent in the SBNMS in 1995. These results demonstrate a considerable individual-level and annual-level propensity to use the marine sanctuary. The vertical lines on each plot indicate results when no individual-level heterogeneity was modelled.

annual probability on γ^{HH} , which corresponded with greater heterogeneity in proportion of time spent in the SBNMS. These results are based on the posterior modes and do not incorporate uncertainty about the value of each posterior mode.

To understand the cost associated with a model that does not account for individual heterogeneity, we ran the same three-state model with yearly variation on γ^{HH} , but no individual-level heterogeneity. The following changes were evident in the model: a 7.45% decrease in γ^D ; a 3.6% increase in π ; and a slight decrease in γ^{AA} . The vertical lines in Fig. 5 indicate that even small changes in the estimates for γ^{HH} and γ^{AA} correspond to an appreciably higher estimated long-term proportion of time present in the SBNMS.

Discussion

As the focus of our paper is on ways to handle latent variability, we have avoided analysing and discussing biologically complex models. Instead, we have used a simple description of the movement of North Atlantic HW in the SBNMS to highlight the potential applications of the model. Individual heterogeneity is known to exist in multiple forms within this whale population (Hammond 1990), and our results indicate considerable individual-level latent heterogeneity in both sighting probability and site fidelity. Substantial annual variation in propensity to use the SBNMS in any given year was evident through the inclusion of the annual term in the model. The abundance of HW is known to vary annually at sites within the Gulf of Maine, and this has been linked to local prey abundance (Payne *et al.* 1990; Weinrich *et al.* 1997).

To illustrate one potential application of the model, we calculated the stationary distribution for the first two states in the

model (γ^{HH} and γ^{AA}) which gives an estimate for the long-term proportion of time spent within the SBNMS. Our results indicate the loss of information when individual variability was not modelled. With no latent heterogeneity, the proportion of time spent in the SBNMS was the same for all whales. The importance of accounting for individual-level heterogeneity is well known in the mark–recapture field, and the focus of much research, but as far as we know, has not been applied in this way to issues of reserve design for marine mammals. Our study demonstrates the importance of accounting for individual-level latent heterogeneity in application to marine reserve use and highlights implications associated with the loss of information if ignored.

In comparison with MCMC approaches, ADMB-RE is automatic and quick, but approximate. Nevertheless, in our simulations, we found little bias. MCMC is theoretically exact, but can be notoriously hard to get working and requires multiple manual checks. This can make MCMC methods hard to use in exploratory analysis and difficult to test through simulations similar to those employed in this paper.

Automatic Differentiation Model Builder with Random Effects uses, at least by default, an Empirical Bayes approach to random effects, in that inferences are conditioned on the point estimate of the random effects variance, in other words assuming that the estimated variance is exact. A fully Bayesian approach would instead place a prior on that variance and would then ‘integrate’ its inferences on random and the fixed effects across the posterior distribution of that variance. The Empirical Bayes approximation has often been effective in statistics (e.g. the generally low bias and reasonable precision of fixed effect estimates in our simulations) and avoids the problem of how to choose a prior for a variance. However, for some data sets and some types of inference, the Empirical Bayes approach may not be adequate, for example if the focus is on particular random effects, or the proportion of individuals with individual effects above some threshold. A particular conundrum arises if the point estimate of random effects variance is zero; should one then assert that there is no individual variability at all? At least for the shortest data sets in our simulations, variances close to zero were common, although it did not happen with the (much larger) real data set. If faced with a zero-random-effects-variance estimate in practice, and if the desired inference specifically involves the random effects, then further efforts would be needed; for example, one could make several sets of inferences for values of σ^2 close to zero, then weighting each set by the marginal likelihood of that σ^2 (which corresponds to placing a post hoc uniform prior on the random effects variance). Further case-specific simulations of the impact on the inference of most interest would also be advisable.

Although we investigated bimodality only for exploratory purposes, simulations indicate that at longer observation histories, ADMB-RE, despite a Normal prior, was able to estimate bimodality in the posterior distribution of individual random-effect values. Although the Laplace approximation assumes a Normal prior, the posterior distribution of individual modes need not be normal. Given enough data on each individual

whale, it would not be necessary to use random effects. The benefit of random effects lies in the ability for individuals with little data to draw strength, and information, from those with enough data to form a clear understanding of the underlying process. In our case, the appeal and justification for the use of random effects lies in the large number of whales with uninformative observation histories.

Heterogeneity between individuals in parameters such as observability and site fidelity is sometimes of interest in its own right and is sometimes also a biasing factor that needs to be accounted for when estimating other parameters such as abundance. Whether it is necessary to allow for heterogeneity in mark–recapture studies depends on the extent of the heterogeneity and the objective of inference. In this paper, we have focused on quantities such as survival and mean residence time for which the impact of heterogeneity (in simulations) is appreciable but not immense; however, if similar data were being used for mark–recapture estimation of abundance, then the impact of the degree of heterogeneity in probability of observation seen in Fig. 3 would certainly be very large.

If heterogeneity is likely to be important enough to be worth modelling in principle, then the question arises of whether and how to do so. If all animals have very short-observation histories, there may be little gain in modelling heterogeneity at all. This is because there will likely not be enough data to estimate it, and it may only be practical to acknowledge qualitatively the possibility that the unaccounted heterogeneity has influenced the other results. On the other hand, if all animals have sufficiently long observation histories, then heterogeneity can easily be handled through individual-level fixed effects. However, when some individuals have long capture histories but others have very short histories, then heterogeneity is in principle estimable, but fixed-individual-effect models will run into problems with the short-history animals, because of small sample sizes. Assuming the model can be made to fit in the first place, the point estimates of individual effects will be very noisy (and possibly biased because of small-sample effects) for the short-history animals and that noise makes it difficult to disentangle the underlying distribution across the population. In this situation, individual-level random-effect models should be particularly beneficial, despite the extra computational workload, because the short-history animals ‘borrow strength’ from those with longer histories. The inferred population-level distribution of the random effect (i.e. in this paper the particular inverse-logit-Normal distribution implied by the mean and variance of the random effects) does not suffer from corruption by noisy point estimates. Further, the distribution of the posterior modes themselves, which is a useful check on the adequacy of the assumed form of the random effects, does not seem to be much affected by noise, presumably because of the shrinkage applied to individuals with short histories.

The focus of this paper is primarily to demonstrate a computationally feasible approach to dealing with individual-level heterogeneity, and therefore, we have used equal-length histories in the simulations; but for that very reason, our simulations may tend to understate the benefits of individual-level

random-effect models for heterogeneity in long-term field studies.

We have demonstrated that random effects can be easily used to incorporate latent individual-level effects into a model that allows capture probabilities and site fidelity to vary according to individuals. Our simulation results explore the cost associated with not accounting for individual variability in models. There was little cost associated with using a model that was more complex than the data at hand, as fixed and random effects were accurately estimated, whether present or not. In contrast, using too simple a model often resulted in bias and, at times, increased variance in estimation for both the fixed and random effects. ADMB-RE has proven a useful, flexible and accurate platform for applying these statistical methods.

Acknowledgements

We thank Toby Patterson and Chris Wilcox for much intellectual input and discussion, and three anonymous reviewers.

References

- Arnason, A.N. (1972) Parameter estimates from mark-recapture experiments on two populations subject to migration and death. *Researches on Population Ecology*, **13**, 97–113.
- Arnason, A.N. (1973) The estimation of population size, migration rates, and survival in a stratified population. *Researches on Population Ecology*, **15**, 1–8.
- Barry, S.C., Brooks, S.P., Catchpole, E. and Morgan, B. (2003) The analysis of ring-recovery data using random effects. *Biometrics*, **59**, 54–65.
- Bolnick, D., Svanbäck, R., Fordyce, J., Yang, L., Davis, J., Hulsey, C. and Forister, M. (2003) The ecology of individuals: incidence and implications of individual specialization. *The American Naturalist*, **161**, 1–28.
- Burnham, K.P. and Overton, W.S. (1978) Estimation of size of a closed population when capture probabilities vary among animals. *Biometrika*, **65**, 625–633.
- Conn, P.B. and Cooch, E.G. (2009) Multistate capture-recapture analysis under imperfect state observation: an application to disease models. *Journal of Applied Ecology*, **46**, 486–492.
- Fournier, D.A., Skaug, H., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M., Nielsen, A. and Sibert, J. (2012) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, **27**, 233–249.
- Gimenez, O. and Choquet, R. (2010) Individual heterogeneity in studies on marked animals using numerical integration: capture-recapture mixed models. *Ecology*, **91**, 951–957.
- Hammond, P.S. (1986) Estimating the size of naturally marked whale populations using capture-recapture techniques. *Reports of the International Whaling Commission, Special Issue*, **8**, 253–282.
- Hammond, P.S. (1990) Heterogeneity in the Gulf of Maine? Estimating humpback whale population size when capture probabilities are not equal. *Reports of the International Whaling Commission, Special Issue*, **12**, 135–139.
- Huggins, R.M. and Yip, P.S.F. (2001) Note on nonparametric inference for capture-recapture experiments with heterogeneous capture probabilities. *Statistica Sinica*, **11**, 843–853.
- Katona, S.K. and Whitehead, H.P. (1981) Identifying humpback whales using their natural markings. *Polar Record*, **20**, 439–441.
- Lebreton, J.D., Nichols, J.D., Barker, R., Pradel, R. and Spendelov, J. (2009) Modeling individual animal histories with multistate capture-recapture models (ed. H. Caswell). *Advances in Ecological Research*, **41**, 87–173.
- Maunder, M., Skaug, H., Fournier, D. and Hoyle, S. (2008) Comparison of estimators for mark-recapture models: random effects, hierarchical Bayes, and AD Model Builder. *Modeling Demographic Processes in Marked Populations* (eds D.L. Thomson, E.G. Cooch & M.J. Conroy). *Environmental and Ecological Statistics*, **3**, 917–948.
- Patterson, T.A., Basson, M., Bravington, M. and Gunn, J. (2009) Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology*, **78**, 1113–1123.
- Payne, P., Wiley, D., Young, S., Pittman, S., Clapham, P. and Jossi, J. (1990) Recent fluctuations in the abundance of baleen whales in the southern Gulf of Maine in relation to changes in prey abundance. *Fish. Bull.*, **84**, 41–47.
- Pledger, S. and Phillpot, P. (2008) Using mixtures to model heterogeneity in ecological capture-recapture studies. *Biometrical Journal*, **50**, 1022–1034.
- Pledger, S., Pollock, K.H. and Norris, J. (2003) Open capture-recapture models with heterogeneity: I. Cormack–Jolly–Seber model. *Biometrics*, **59**, 786–794.
- Royle, J.A. (2008) Modeling individual effects in the Cormack–Jolly–Seber model: a state-space formulation. *Biometrics*, **64**, 364–370.
- Royle, J.A. and Dorazio, R.M. (2008) *Hierarchical Modeling and Inference in Ecology: The Analysis of Data From Populations, Metapopulations and Communities*. Academic Press Elsevier, London.
- Seber, G. (1982) *The Estimation of Animal Abundance and Related Parameters*, 2nd edn, pp. 654. MacMillan Publishing Co, New York.
- Skaug, H.J. and Fournier, D.A. (2006) Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. *Computational Statistics and Data Analysis*, **51**, 699–709.
- Weinrich, M., Martin, M., Griffiths, R., Bove, J., and Schilling, M. (1997) A shift in distribution of humpback whales, *Megaptera novaeangliae*, in response to prey in the southern Gulf of Maine. *Fishery Bulletin*, **95**, 826–836.
- Zucchini, W., Raubenheimer, D. and MacDonald, I. (2008) Modeling time series of animal behavior by means of a latent-state model with feedback. *Biometrics*, **64**, 807–815.

Received 18 October 2011; accepted 15 July 2012

Handling Editor: Jane Molofsky

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Simulation results.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.