# ShinyWhale

Is a Bayesian multi-state mark-recapture-recovery model

### Overview

ShinyWhale was developed with the goal of providing a simplified tool in which estimates for the abundance and demographic information (survival, recapture, site fidelity and dead recovery) of baleen whales can be generated. This application was initially designed for the North Atlantic right whale, however the application can be applied to similar baleen whale species with long-term sighting histories. ShinyWhale was developed with the aim of providing managers, NGO’s, government organisations and conservationists a tool to assess the changes in population demographic trends, without the need to understand the complete workings of Bayesian mark-recapture techniques, or strong programming skills. More information on this model is available in the Github repository.

### Background

ShinyWhale relies on the following packages:

* Data prep: unix, dplyr, reshape2, MASS, stringr
* Shiny: shiny, shinybusy
* Model: R2jags
* Outputs: MCMCvis, ggplot2

#### **Capture histories**

Observations of individuals (i) in a given year (t) are compressed into a single value:

1. seen alive
2. recovered dead, or
3. neither seen nor recovered

The culmination of observations for each individual during the sampling periods for the study forms the capture history. Each year of the capture history represents a year for the study population, which I set as the first month of the reproductive period, which can be set by the user in the Capture Histories tab of the app. The capture history is then combined with an additional period at the beginning of the study period where all individuals are assigned a 3 (neither seen nor recovered) which results in the observation matrix.

#### **Data augmentation**

The data set can also be adjusted to include data augmentation. This is the inclusion of additional individuals into the observation matrix, who are in state 3 (neither seen nor recovered). This is to account for individuals who could exist in the population over the study period, though have never been sighted. Individuals added in data augmentation are not all going to enter the population, but can provide more realistic estimates for abundance if it is believed that not every individual has been sighted. More information on data augmentation can be found here: (Royle 2008; Royle, Dorazio, and Link 2007; Royle and Dorazio 2010).

#### **Model**

The model used in ShinyWhale is a Bayesian multi-event Jolly-Seber framework (Thomson, Cooch, and Conroy 2009; Schaub and Royle 2013) fit with a mark-recapture-recovery model (Barker 1999; Liljestrand, Wilberg, and Schueller 2019). In the model, we considered five true biological states:

1. not yet entered the population (NE)
2. alive within the study area (AI)
3. alive outside the study area (AO)
4. recovered dead (RD) and
5. dead (D)

The focal temporal parameters that are estimated in this model are:

* true survival (*s*)
* recapture (*p*)
* site fidelity (*F*)
* dead recovery (*r*) and
* abundance (*N*)

The state of all individuals in the first occasion was set so that they were considered to not have entered the population (*NE*):

The state of individual from the second occasion until they enter the population is:

While the subsequent states are dependent on that state of the previous time period, thus the state model is denoted as:

where denotes the state of individual at time , and denotes state membership over time, where is the number of true states.

The observation model is denoted as:

where denotes the observation of individual at time , and  denotes the observational process, linking the true states to the observed states , where is the number of observed states.

The number of individuals entering the population at time can be calculated as:

while the total population size at time is:

Therefore, the full conditional probability of the model can be expressed as:

### Getting Started

**Before you start you must download** [**JAGS (Just Another Gibbs Sampler)**](https://sourceforge.net/projects/mcmc-jags/files/JAGS/4.x/)**.**

The ShinyWhale app is divided up into tabs in order to make it easier to digest. The tabs are broken down into sections which are detailed below.

### Data

#### Import Data

ShinyWhale requires you to have access to a long-term sightings catalog or record that is in *.csv* format. The raw data should be in long format, with column headings describing what each column contains and associated data in the corresponding rows going down. If the data is in the correct format, it can be uploaded to the app.

ShinyWhale requires you to select the appropriate columns from your sightings data set for each of the following sections:

| Column | Description |
| --- | --- |
| Year | This is a column within your sighting histories that contains the information on which year each sighting was made. |
| Month | This is a column within your sighting histories that contains the information on which month of the year each sighting was made. |
| Unique Identification Number | This is a column within your sightings history that contains the unique identification number for each individual. |
| Mortality | This is a column in your sightings history that contains information on if you have recorded a dead whale - must be recorded as *DEAD* in the sighting history. This may be recorded with other information such as behaviors or births. |

#### Capture Histories

Now that you have the data loaded and have selected the columns that will be needed for the model, it is time to define some constraints for the data. You will need to enter or select from a drop down menu the response to the following questions in the app.

* Select the first year you want to model from - This is the year you want to model from. Your data set may cover a large time period, and perhaps you are only interested in the last decade. Perhaps your data set contains early records with low effort that you do not wish to include in your analysis.
* Select the last year you want to model - This is the final year you wish to model. You may wish to not include the final year of data from your data set as it includes incomplete sightings information, so you can just select the previous year here to exclude it.
* Select the month to be used as the start of the year - This is the month you want to be the beginning of a sighting period. This could be the start of the breeding or feeding period when you see most individuals arriving in your study area, or it could be the start of the calender year. What ever makes the most sense for the species you are trying to model.
* Enter the number of additional individuals you want to include for data augmentation - Data augmentation is a way that the model can account for the presence of individuals within the population, who for what ever reason have never been sighted. This number can be as large or small as makes sense for the population or species, however if too small, the population size may be under estimated. The number of individuals you enter as data augmentation will not all be added to the total abundance.

Now that all the constraints have been set, we can now create the capture-history by pressing the **Create capture histories** button. This will create the capture history that you can view in the tab of the same name to ensure that it looks correct.

Now we can move on to setting up the model.

### Model

#### Model Settings - Markov chain Monte Carlo (MCMC)

ShinyWhale uses MCMC to sample from the data and estimate the parameters of interest. In order to run the MCMC we need to provide some basic settings for the model which will depend on the quality and quantity of the data being used. See below for a description of the different settings, and what values can be adjusted depending on the data that is being supplied.

1. **Number of iterations:** The number of times you wanted to model to be run with the unique initial values. The model needs to be run for enough iterations for it to converge (reach a stable estimate for the parameters). We suggest a minimum of 2000 iterations with good sightings records, and this can be increased if sighting records are patchy.
2. **Amount of thinning:** This is how many iterations you wish to keep from your sample, i.e. 1 - keeps every iterations, 10 - keeps every ten iterations. By storing less iterations you can save on computer memory needed to run the model, however you can loose precision.
3. **Number of burn-in’s:** These are the initial iterations of the model that are when the parameter estimation is stabilizing. These iterations will be discarded so they wont influence the uncertainty surrounding the parameter estimation. We suggest this be a third to half the amount of iterations you have selected. If you have 3000 iterations, set the burn-in between 1000-1500.
4. **Number of chains:** The number of unique starting points for the model. Each chain begins with a unique initial value for the parameters of interest. We suggest a minimum of ***3*** chains be used.

Now that the data is prepared and we have selected the constraints for the model, we can now run the model by pressing the **“Run Model”** button at the bottom of the page. Once the model is running, a status bar will flash across the top of the page indicating the model is running.

Once the model has finished running, a message will appear at the bottom of the page that will read either:

* Model Complete - in which case the model has successfully run and you can move on to the next step of retrieving the output and plotting the data.
* Rerun model - in which case the model did not successfully converge and will need to be rerun. If you selected a minimum of 3 chains and no thinning (i.e. set to 1), then this problem can likely be fixed by increasing the number of iterations and increasing the burn-in period accordingly (i.e. if you add another 1000 iterations, add another 300-500 to the burn-in).

### Output

Now that the model has successfully run, you will be able to access the Output tabs. These tabs are all have the same structure, with a figure of the output and a table of the summary of the output for the different parameters:

* Abundance
* Survival
* Recapture
* Site Fidelity
* Dead Recovery

For each of the different parameters, a *.CSV* can be downloaded of the data by clicking the “**Download the data**” button, and the plot can be downloaded by clicking the “**save plot**” button.

Barker, Richard J. 1999. “Joint Analysis of Markrecapture, Resighting and Ring-Recovery Data with Age-Dependence and Marking-Effect.” *Bird Study* 46 (sup1): S82–91. <https://doi.org/10.1080/00063659909477235>.

Liljestrand, Emily M., Michael J. Wilberg, and Amy M. Schueller. 2019. “Multi-State Dead Recovery Mark-Recovery Model Performance for Estimating Movement and Mortality Rates.” *Fisheries Research* 210 (February): 214–23. <https://doi.org/10.1016/j.fishres.2018.10.014>.

Royle, J. Andrew. 2008. “Analysis of Capture-Recapture Models with Individual Covariates Using Data Augmentation.” *Biometrics* 65 (1): 267–74. <https://doi.org/10.1111/j.1541-0420.2008.01038.x>.

Royle, J. Andrew, and Robert M. Dorazio. 2010. “Parameter-Expanded Data Augmentation for Bayesian Analysis of Capturerecapture Models.” *Journal of Ornithology* 152 (S2): 521–37. <https://doi.org/10.1007/s10336-010-0619-4>.

Royle, J. Andrew, Robert M Dorazio, and William A Link. 2007. “Analysis of Multinomial Models With Unknown Index Using Data Augmentation.” *Journal of Computational and Graphical Statistics* 16 (1): 67–85. <https://doi.org/10.1198/106186007x181425>.

Schaub, Michael, and J. Andrew Royle. 2013. “Estimating True Instead of Apparent Survival Using Spatial Cormack-Jolly-Seber Models.” Edited by Richard Barker. *Methods in Ecology and Evolution* 5 (12): 1316–26. <https://doi.org/10.1111/2041-210x.12134>.

Thomson, David L, Evan G. Cooch, and Michael J. Conroy, eds. 2009. *Modeling Demographic Processes in Marked Populations*. Springer US. <https://doi.org/10.1007/978-0-387-78151-8>.