

# 1 The movement model

We present a spatially explicit movement model for sablefish based on the model of McGarvey and Feenstra (2002). The main difference between the model of McGarvey and Feenstra (2002) and the model presented here is that ours is time-averaged over the length of the mark-recapture experiment, rather than using a discrete time step with all movement assumed to take place on the same day each year. Should we check how their model works out?. This class of model makes the assumption that natural mortality and the reporting rate of tags is spatially constant, however the latter assumption is reasonable given that the fisheries studied have 100% at-sea observer programmes. Unlike the Peterson class of mark-recapture models, this class estimates movement rates as probabilities and not population abundance.

This is very much a work in progress. I'm just trying to get the model in here, we can tidy up order later. I'm trying to move as much definition of terms to the table at the end, a la Schnute's model presentation technique.

## 1.1 Data Structure

There are two main sources of data which are used in the model:

1.  $N = (n_{i,j})_{i,j}$ , a matrix of tag recoveries indexed by release inlet  $i$  and recovery area  $j$ ;
2.  $\bar{E}_j''$ , a measure of average relative fishing effort for each recovery area  $j$ .

### 1.1.1 Tag recoveries

The PacSableTag database, maintained at the Pacific Biological Station in Nanaimo, British Columbia, contains tag release and recovery data from 1995 through to 2014

Confirm these dates, do we need a reference?. These data were extracted for use in this model. I'm going to leave this for now, below I've listed things they restricted, and we should talk about whether or not to include them

- excluded same year recoveries
- multiple recoveries
- restricted to only trap gear - to remove selectivity issues Can we account for this with a selectivity function?
- did not include survey recoveries (argued that these are negligible)
- Only included survey tagged fish in inlets I think we talked about expanding this

We should include a table of tag recoveries, which I can make in R and export to latex.

### 1.1.2 Fishing Effort

Fishing effort  $E_{j,t}$  is defined as number of traps deployed Does soak time play a factor here? Should we include it in the effort metric? in the offshore commercial trap fishery and inlet tagging surveys in each area and year. This is considered as a raw metric of fishing effort, and is time-averaged and normalised to produce the

time-averaged relative fishing effort  $\bar{E}_j''$  for each of the nine tag recovery areas. We use the following three steps to time average and normalise the raw effort  $E_{j,t}$  over time and area fished.

1.  $E_{j,t}$  is the sum of the number of traps deployed in area  $j$  in year  $t$ .
2.  $A_{j,t}$  is the area fished in  $\text{km}^2$ , computed by applying a gear-fish attraction width to each fishing set. We assume that sablefish are attracted to the baited traps from 0.5km in any direction, giving a diameter of 1.0km. Movement probabilities estimated by our model are sensitive to the values for fishing effort so insert sensitivity analyses here (Wyeth et al., 2006). To normalise the fishing effort density take

\*

$$E'_{j,t} = \frac{E_{j,t}}{A_{j,t}}.$$

3. Finally we time average the  $E'_{j,t}$  over all the years to get

$$\bar{E}_j' = \frac{1}{T} \sum_{t=1}^T E'_{j,t},$$

which is then averaged over all areas to get the relative average fishing effort

$$\bar{E}_j'' = \frac{\bar{E}_j'}{\sum_{j=1}^J \bar{E}_j'}$$

## 1.2 Model Structure

We estimate the parameters  $P_{i,j}$ , which are defined as the probabilities of a fish tagged and released in inlet  $i$  being recovered in, or moving to, area  $j$ . These probabilities are estimated by comparing the observed number of recaptures  $n_{i,j}$  (Table 3) with the predicted probabilities of movement through a Bayesian posterior distribution function, as outlined in Section ??.

We define  $\hat{n}_{i,j}$  as the predicted number of fish recaptured in area  $j$  that were tagged and released in area  $i$  in Table 3 as a function of the parameters  $P_{i,j}$ . When substituted into the function  $\mathcal{R}(j \mid i)$ , given in Table 3, the releases term  $R_i$  cancels and the function simplifies to

$$\mathcal{R}(j \mid i) = \frac{P_{i,j} (1 - \exp(-Z_j)) \frac{F_j}{Z_j}}{\sum_{j=1}^J P_{i,j} (1 - \exp(-Z_j)) \frac{F_j}{Z_j}}. \quad (1)$$

Equation (1) is a time-averaged version of equation (5) in McGarvey and Feenstra (2002), and all terms are defined in Table 3. The fishing mortality in each area is given as

$$F_j = \bar{F} \cdot \bar{E}_j'', \quad (2)$$

with the time-averaged relative effort  $\bar{E}_j''$  defined in Section 1.1.2.

## 1.3 Parameter Estimation

The parameter matrix  $\Theta$  for our model satisfies the same assumptions as a Markovian transition matrix:

- The entries are all probabilities,  $P_{i,j} \in [0, 1]$ ;
- The sum of the entries in each row of  $\Theta$  is equal to 1,  $\sum_{j=1}^J P_{i,j} = 1$ .

We estimate the distributions of the entries in  $\Theta$  by using a Gibbs sampler to numerically integrate the Bayes posterior distribution

$$\mathbb{P}(\Theta \mid N) \propto \mathcal{P}(\Theta) \cdot \mathcal{L}(N \mid \Theta).$$

I think we should mention somewhere that the posterior for each row has no bearing on the others, so we can compute them independently - well, I think I just said it but I haven't worded it correctly yet. \*

We use the Dirichlet distribution as a prior on each row of  $\Theta$ , as this incorporates the Markovian conditions on the entries. For a fixed  $i$ , we have the prior probability distribution function

$$\mathcal{P}_i(\Theta_{i,\vec{j}}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_J)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_J)} \Theta_{i,1}^{\alpha_1-1} \dots \Theta_{i,J}^{\alpha_J-1}. \quad (3)$$

The parameters of the dirichlet distribution are chosen as  $\alpha_j = 1$  for all  $j$ , so that it collapses to a uniform distribution on the discrete space of recovery probabilities and is therefore fairly uninformative.

The likelihood of observing the data  $N$  given our hypothesis  $\Theta$  is also computed for each tagging area (row). For this we use a multinomial distribution, which is appropriate when each recapture event is independent of the others. This is a strong assumption, but I'm not sure what else to do about it. \*

fixed tagging area  $i$  is the probability mass function of the multinomial distribtuion

$$\mathcal{L}_i(N_{i,\vec{J}} \mid \Theta_{i,\vec{J}}) = \frac{\sum_{j=1}^J n_{i,j}}{n_{i,1}! \cdots n_{i,J}!} \prod_{j=1}^J \mathcal{R}(j \mid i)^{n_{i,j}}. \quad (4)$$

The posterior for each tagging area is then given by

$$\mathbb{P}_i(\Theta \mid N) = \mathcal{P}_i(\Theta) \cdot \mathcal{L}_i(N \mid \Theta), \quad (5)$$

which is combined with the other, independent ? tagging areas to give the posterior function \*

$$\mathbb{P}(\Theta \mid N) = \prod_{i=1}^I \mathcal{P}_i(\Theta) \cdot \mathcal{L}_i(N \mid \Theta). \quad (6)$$

The posterior distributions for the model parameters are then found by numerically integrating the posterior  $\mathbb{P}(\Theta \mid N)$ , or equivalently, the log-posterior

$$\log \mathbb{P}(\Theta \mid N) = \sum_{i=1}^I (\log \mathcal{P}_i(\Theta) + \log \mathcal{L}_i(N \mid \Theta)). \quad (7)$$

## 2 A latex'ed version of the model used in the Cleary et al paper.

Let's build the Schnute style table of the model:

I'm playing with table formats below, this one didn't work out :P \*

Parameter	Description
$P_{i,j}$	Probability that a fish tagged in inlet $i$ moves to area $j$ , leading
$\Theta$	Parameter matrix of probabilities $(P_{i,j})_{i,j}$
$\bar{F}$	Average fishing mortality over all areas (0.2) <i>source?</i>
$M$	Natural mortality of sablefish (0.08, Kronlund et al, 2003) <i>update?</i>
$\alpha_j$	Parameters of the Dirichlet prior distribution for the parameters $P_{i,j}$ ( $\alpha_j = 1$ )
$d$	Sample size of multinomial jumping distribution for MCMC integration of posterior dist

Table 1: The parameters of the model.

Index	Description
$i$	Inlet where fish were tagged and released
$j$	Area where tagged fish recovered
$I$	Total number of inlets (release sites), $I = 4$
$J$	Total number of recovery sites (inlets and offshore fishing areas), $J = 9$
$\vec{I}$	Vector of release site indices $\vec{I} = (1, 2, \dots, I)$
$\vec{J}$	Vector of recovery site indices $\vec{J} = (1, 2, \dots, J)$
$k$	Iteration of MCMC chain used in numerical integration of posterior distributions of paramet
$t$	Year
$T$	Total number of years ( $T = 20$ )

Table 2: Indices of variables in the movement model

## References

- McGarvey, R. and Feenstra, J. E. (2002). Estimating rates of fish movement from tag recoveries: conditioning by recapture. *Canadian Journal of Fisheries and Aquatic Sciences*, 59(6):1054–1064.
- Wyeth, M., Kronlund, A. R., Elfert, M., and Station, P. B. (2006). *Summary of the 2004 British Columbia sablefish (*Anoplopoma fimbria*) research and assessment survey*. Fisheries and Oceans Canada, Science Branch, Pacific Region, Pacific Biological Station.

Variable	Description
$n_{i,j}$	Number of fish tagged in inlet $i$ and recovered in area $j$ over $T$ years
$N = (n_{i,j})_{i,j}$	Matrix of count data
$R_i$	Number of fish tagged in inlet $i$ (cancels, so not needed)
$E_{j,t}$	Fishing effort (traps deployed) in area $j$ during year $t$
$A_{j,t}$	Geographic coverage of area fished in $km^2$
$E'_{j,t} = \frac{E_{j,t}}{A_{j,t}}$	Fishing effort scaled to area fished
$\bar{E}'_j = \frac{1}{T} \sum_{t=1}^T E'_{j,t}$	Average scaled fishing effort,
$\bar{E}''_j = \frac{\bar{E}'_j}{\sum_j \bar{E}'_j}$	Relative fishing effort for area $j$ , averaged over area fished and $T$ years <i>Choose better notation for these.</i>
$F_j = \bar{F} \cdot \bar{E}''_j$	Fishing mortality in area $j$
$Z_j = F_j + M$	Total mortality in area $j$
$\hat{n}_{i,j} = P_{i,j} (1 - \exp(-Z_j)) \frac{F_j}{Z_j} R_i$	Predicted number fish tagged in $i$ and recovered in $j$
$\mathcal{R}(j   i) = \frac{\hat{n}_{i,j}}{\sum_j \hat{n}_{i,j}}$	Predicted probability of recovering in area $j$ a fish tagged in area $i$
$Y_{i,\vec{J},k}$	Dummy candidate vector of release inlet $i$ of length $J$ for MCMC iteration $k$
$P_{i,vJ,k}^*$	Candidate vector of movement probabilities for release inlet $i$ of length $J$ for MCMC iteration $k$

Table 3: Variables used in the specification of the movement model

Function	Description
$\mathcal{L}(N   \Theta)$	Model likelihood pdf of data $N$ given the hypothesis $\Theta$
$\mathcal{P}(\Theta)$	Prior distribution pdf on the hypothesis $\Theta$
$\mathbb{P}(\Theta   N)$	Posterior distribution pdf of the hypothesis $\Theta$ given the data $N$
$\Gamma$	Gamma function used in the Dirichlet distribution
$\mathcal{M}$	Multinomial jumping distribution for MCMC integration

Table 4: Functions used in the movement model