Dissecting SAM 1: Processes

Anders Nielsen, Ethan Lawler, & Sean Anderson

an@aqua.dtu.dk

The parts of SAM

Processes:

- The three main processes are: Recruitment ($N_{1,y}$), survival ($N_{>1,y}$), fishing ($F_{a,y}$).
- These are treated as unobserved random effects in the model
- The processes describe the development of the system we are monitoring
- Observations related to the system are used to predict these processes

Observations:

- Anything we can observe, which can help to inform about the processes
- Common options are catch-at-age $C_{a,y}$, survey index-at-age $I_{a,y}$, total catches, biomass index, tagging, lengths ...
- From the process (and a few estimated model parameters) we should be able to predict the observations

Parameters:

- Fixed effects model parameters to be estimates
- E.g. catchabilities, variance parameters, and stock-recruitment parameters.

Here we will look at the process part

Recruitment

- In a state-space assessment model we want to setup a recruitment process
- The simplest possible option could be a random walk, where:

$$\log R_y = \log R_{y-1} + \varepsilon_y \ , \quad \text{where } \varepsilon_y \sim \mathcal{N}(0, \sigma_R^2)$$

- Another option could be to use the spawning stock biomass (SSB) to predict the rectuitment (if recruitment is at age 1 we need to use the SSB from the year before)
- Popular options are the functions:
  - Ricker: $R = \alpha \mathsf{SSB} e^{-\beta \mathsf{SSB}}$
  - Beverton-Holt: $R = \frac{\alpha \mathsf{SSB}}{1+\beta \mathsf{SSB}}$
- To use these we can setup the process like:

$$\log R_y = \log \mathsf{SR}(\mathsf{SSB}_{y-1}) + \varepsilon_y \ , \quad \text{where } \varepsilon_y \sim \mathcal{N}(0, \sigma_R^2)$$

  where the SR() function is the stock-recruitment function assumed.
- Variance parameter σ_R^2 is objectively estimated via maximum likelihood
- Prediction is straight forward

Recruitment exercise

- The data set `Robs.RData` contains three vectors `year`, `ssb`, and `Robs`.
- Implement the state-space model corresponding to the 'true' unobserved recruitment following a random walk.
- In this exercise we consider `Robs` to be observations of recruitment subject to measurement noise.
- Is the model describing data well (plot)?
- Consider how we could implement the Ricker and Berverton-Holt state-space versions (we will get back to that later).
- Some stocks have extreme recruitment events — could we somehow adapt the model to that?

Survival

- Models use the stock equation:

$$N_{a,y} = N_{a-1,y-1} e^{-F_{a-1,y-1} - M_{a-1,y-1}}$$

- If the oldest age group contains fish age A and older (a so-called plus-group), then

$$N_{A,y} = N_{A-1,y-1} e^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1} - M_{A,y-1}}$$

- Even with perfect knowledge of F_{ay} and M_{ay} we should still expect some uncertainty

In state-space models:

- F_{ay} and M_{ay} are considered rates in a process, e.g. as:

$$\log N_{a,y} = \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \xi_{ay} \ , \quad \text{where } \xi_{ay} \sim \mathcal{N}(0, \sigma_S^2)$$

- Can also be formulated such that $N_{a+1,y+1} < N_{ay}$ always. Here considered a ‘feature’
- Very small σ_S^2 not necessarily a problem
- Large σ_S^2 or one-sided deviations from stock equation can be used to diagnose problems

Survival exercise

- The data set `Nobs.RData` contains 'observations' of N_{ay} (`Nobs`) with observation noise.
- The data set also contain a few helper variables (`F`, `M`)
- Here we will implement a multivariate state-space model for $\log N$, which uses (`Nobs`) as observations.
- For the youngest age group use a random walk recruitment model.
- For the following age groups use the stock-equation and adjust for plus group for oldest age.
- The observation equation is simple, because we simply assume

$$\log N_{ay}^{(obs)} = \log N_{a,y} + \varepsilon_{ay} \quad , \quad \text{where } \varepsilon_{ay} \sim \mathcal{N}(0, \sigma^2)$$

- Estimate the three variance parameters in the model (for recruitment, survival, and observation)
- Plot the fitted versus the observed $\log N$

Fishing mortality

- Different approaches in use to define F_{ay}, e.g:
- Deterministic (assuming C_{ay} know without error) possibly with ad-hoc smoothing
- Multiplicative $F_{ay} = S_a f_y$
- Block wise multiplicative $F_{a,y} = S_a f_y$ with separate S_a in time blocks
- Splines with fixed degree of smoothness or penalized deviances

In state-space models:

- Define $F_y = (F_{1y}, \ldots, F_{Ay})$
- Formulate a process model for F_y, e.g. as:

$$\log F_{y+1} = \log F_y + \psi_y \ , \quad \text{where } \psi_y \sim \mathcal{N}(0, \Sigma)$$

- Time-varying selectivity is a side-effect (in this formulation)
- Notice that we can set up correlated F_{ay} processes

Suggested options for correlated process

- Remember:

$$\log F_{y+1} = \log F_y + \psi_y \ , \quad \text{where } \psi_y \sim \mathcal{N}(0, \Sigma)$$

- For all combination of ages ($a \neq \tilde{a}$):

  A) Parallel: $\Sigma_{a,\tilde{a}} = \sqrt{\Sigma_{a,a}\Sigma_{\tilde{a},\tilde{a}}}$

  B) Independent: $\Sigma_{a,\tilde{a}} = 0$

  C) Compound symmetry: $\Sigma_{a,\tilde{a}} = \rho\sqrt{\Sigma_{a,a}\Sigma_{\tilde{a},\tilde{a}}}$

  D) AR(1): $\Sigma_{a,\tilde{a}} = \rho^{|a-\tilde{a}|}\sqrt{\Sigma_{a,a}\Sigma_{\tilde{a},\tilde{a}}}$

- Now let's see which one is better.

Fishing mortality exercise

- The data set `Fobs.RData` contains 'observations' of F_{ay} (`Fobs`) with observation noise
- Implement the process models for $\log F$ with Σ defined as in B), C), and D).
- The observation equation is simple, because we assume

$$\log F_{a,y}^{(obs)} = \log F_{a,y} + \varepsilon_{a,y} \quad , \quad \text{where } \varepsilon_{a,y} \sim \mathcal{N}(0, \sigma^2)$$

- Notice that you can implement each model in a separate file, or you can add a `corMode` flag to your model and implement it so can simply change that option in the input.
- Can you approximate the likelihood of option A) via the code already written (and clever use of map and initial values)?
- Which model is the best description of the data?

Adding stock-recruitment models (extra exercise)

- The data set `Nobs2.RData` contains the same observations as the previous exercise about N, which we will be extending, but in addition data on stock mean weight, fraction maturity, fraction F applied before spawning, and fraction M applied before spawning (`SW`, `MO`, `PF`, and `PM`). All of this is needed to calculate spawning stock biomass SSB.
- For a given year y SSB is defined as:

$$\mathsf{SSB}_y = \sum_{a=0}^{A} \mathtt{MO}_{ay}\mathtt{SW}_{ay}N_{ay}e^{-\mathtt{PF}_{ay}F_{ay}-\mathtt{PM}_{ay}M_{ay}}$$

- We need to be able to calculate SSB, because the Ricker and Beverton-Holt stock-recruitment functions depend on it. Write a function to calculate SSB.
- Implement options to switch to Ricker or Berverton-Holt stock-recruitment (from the random walk we first did).
- Which model is the best description of the data?