Stock Assessment Tutorial

Eura Nama

# Stock Assessment Model (SAM) Tutorial Project

# Tutorial Objectives and Overview

1. There are 3 main components to this tutorial that you should ask me lots of questions about until you really understand what is happening!
   1. *Simulation*
      * This will provide us with *fake* data that was generated with set parameters and results in our *known* data to compare against
      * The biomass of the adults and the recruits generated in this simulation is then distributed across the survey domain.
      * You do not need to do anything for this as we have all the parameters set and this will run without you having to change anything
   2. *Survey Design*
      * Here we sample from the simulated data above, you can determine the number of stations to sample and the design of the survey
      * The survey data is then used to create indices of adult Biomass and Recruits which are inputs into the Stock Assessment Model (SAM)
      * By changing the survey design you can see how sample size and survey type influence the estimation of biomass in the Stock Assessment Model (SAM)
   3. *Stock Assessment Model (SAM)*
      * The indices from the survey are inputted into the Stock Assessment Model to get an estimate of the model biomass
      * Here you have control of how the model will estimate 2 of the input parameters
        + *Catchability (q)*  and *Natural mortality (m)*.
      * By varying these input parameters you can explore how being *wrong* about these parameters would impact your believe about the status of the stock
2. Much of the heavy lifting is done behind the scenes and your goal will be to understand how changing the survey design and/or the 2 model parameter impacts the biomass estimates from our Stock Assessment Model (SAM)

# What are we doing?

1. Today you are going to run a simple Stock Assessment Model (SAM) for the Dusky Scallop Shark (*Dustious maximus*) stock in the Gulf of Maine using a Delay Difference model
   1. Instead of getting one year of data, this tutorial will automatically simulate the biomass dynamics of this stock between 1990 and 2021
      * The simulation is used to make ‘fake’ real world data which we then use to understand how survey design and model parameter uncertainty can influence our understanding of population status
   2. Using this simulated data you will design a survey to sample the biomass in the area (similar to Tutorial #2) in each year
   3. Then this survey data will be used in the Delay Difference Stock Assessment Model (SAM) to get an estimate of biomass for each year
   4. You can vary several parameters to see how they influence the model results.
2. The stock and the survey is the same as used in Tutorial/Lecture #2, we just have multiple years of data as part of this survey
3. For the survey design we will again have 2 variables as part of the survey design that you can control
   1. You can vary the number of survey tows
      * You have the same constraints as in Lecture 2, I’d suggest you try a low value and a high value as based on what you found in Tutorial #2
   2. You can decide if you want to have a *Random* or stratified survey
      * The only stratification option this time is using the *NAFO* sub-areas
4. The survey indices will then be used as the inputs into the Delay Difference Stock Assessment Model (SAM). There will be two model inputs you can adjust
   1. The survey *catchability prior*.
      * Survey catchability can be difficult to estimate in the real world, but it is a critical parameter to scale up from a survey index to a biomass index
      * The Stock Assessment Model (SAM) requires ‘prior’ knowledge of what we believe survey catchability is.
      * In our simulation we have set *catchability* to be a constant at 0.3
        + Explore how changing the *catchability prior* effects the estimate of biomass (and other paratmeters) in the SAM
   2. The *natural mortality prior*
      * In our simulation we have set natural mortality to = 0.2 on average, but it varies over time
      * We know natural mortality varies from year to year, but we don’t have any data from our survey that informs *what* m\*.
        + So we do not know what the mean *m* is
        + Nor do we know how *m* varies over time
      * As a result we need to give the model our *best guess* of what *m* is using a *natural mortality prior*
        + Explore how changing the *natural mortality prior* effects the estimate of biomass (and other parameters) in the SAM.

# Questions to Consider

1. How well does the survey index compare to the ‘actual biomass’?
   * Does changing the number of survey stations or the stratification (NAFO vs Random) influence this?
   * Given these results and your experience from Lecture 2, would you recommend the NAFO or Random survey?
   * How many stations would you recommend?
2. How well does the Biomass form the SAM compare to the ‘actual biomass’?
3. How does varying the prior on the catchability impact the Stock Assessment results?
   * What are the consequences of being wrong about catchability?
4. How does varying the prior on natural mortality impact the Stock Assessment results?
   * What are the consequences of being wrong about natural mortality?
5. How do you think having a fixed natural mortality term is impacting the Stock Assessment results?

# The Tutorial Tutorial

1. Open the file “Stock\_Assessment\_tutorial.Rmd” in R-Studio
2. On line XX of this file you can change the number of tows used in the survey. Default is 20 tows
3. On line XX-YY of this file you can change the survey stratification scheme, options are “Random” and “NAFO”
4. On line XX of this file you can change the number of times you simulate the Biomass time series and subsequent survey sampling
5. On line XX of this file you can change the *Catchability Prior* to see how this influences the SAM results
6. On line XX of this file you can chnage the *Natural mortality Prior* to see how this influences the SAM results
7. Press the “Knit” button in the tool bar
   * Make sure that you closed (and saved if you want to keep the results) the word document.

Table 1: A Table of your input values for the current run of your simulation

|  |  |
| --- | --- |
| Parameter | Value |
| Number of Tows | 5 |
| Simulated Catchability | 0.3 |
| Mean Simulated Natural Mortality | 0.2 |
| Area swept by a tow | 10000 m² |
| Survey Design | Random |
| Median of Catchability Prior | 0.3 |
| Median of Natural Mortality Prior | 0.11 |

How long did it take to run the model?

## Time difference of 18.69715 secs

So the first thing we want to look at is how well the survey indices for Biomass and Recruitment track the actual Biomass and Recruitment. To do this we need to account for the *catchability* of the stock in the survey. Since we *know* this value in this case we can create this plot to see how well our survey is capturing the underlying dynamics. Each blue dot represents the realization for the specific year.

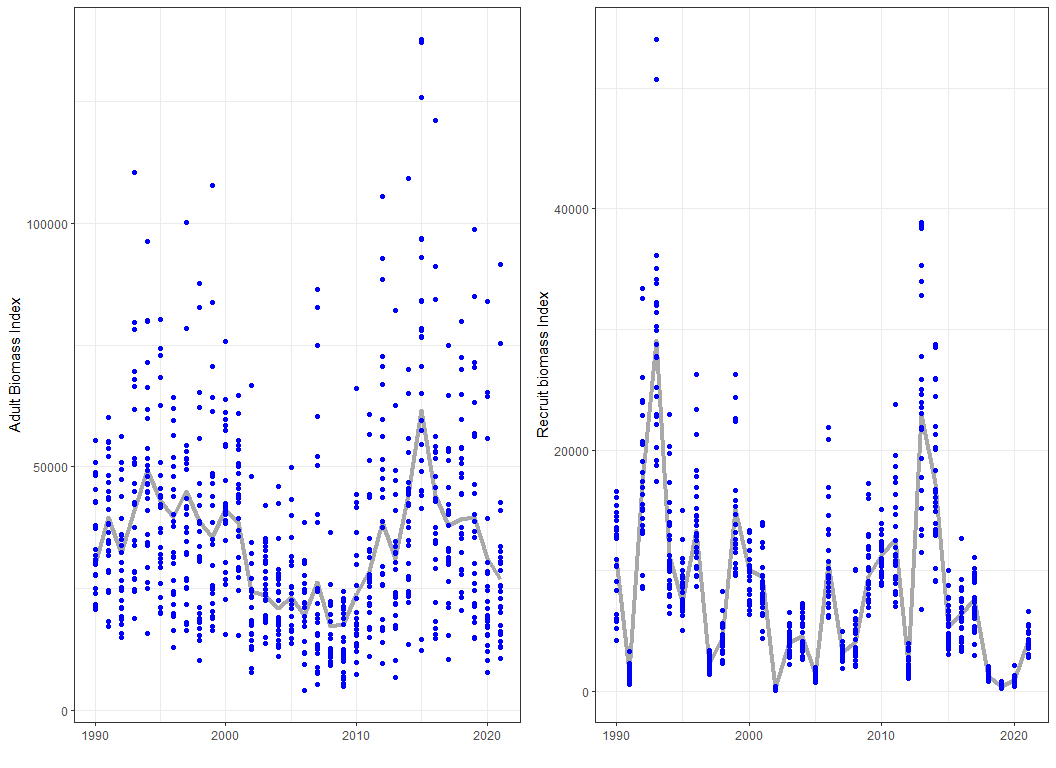


Figure 1: Time series of the survey biomass index from each realization (blue points) and the actual q-corrected biomass (thick grey line) used for the survey simulation.

Table 2: A Table showing how well the survey biomass index estimates the acutal biomass index using various metrics

|  |  |  |
| --- | --- | --- |
| Comparsion | Tonnes | Percentage |
| Mear Error Adult Biomass Index (Survey - Actual) | 980 | 0.22 |
| Mean Error Adult Biomass Index |Survey - Actual| | 13000 | 3.30 |
| Absolute Mean Error Recruit Biomass Index (Survey - Actual) | 180 | 0.17 |
| Absolute Mean Error Recruit Biomass Index |Survey - Actual| | 2100 | 2.20 |

Priors are essentially a guide for the model to help inform likely values for the parameters, the model can decide the value for the parameter is most likely very different from the prior. Sometimes they are well known sometimes they are not much more than an ‘educated’ guess, either way, if the parameter estimate is similar to the prior it is known as an *informative prior*. In our example you should tweak the priors for catchability (*q*) and natural mortality (*m*) to see how these influence the estimate of both these parameters and the other variables/parameters in the model. You should be thinking *are their consequences to getting these wrong and if so what are they*.

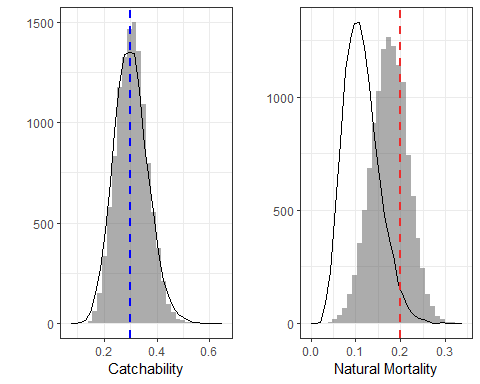


Figure 2: The Prior and Posterior distributions for catchability and mortality. The Prior distribution is approximated by the line, while the model realizations (aka Posteriors) are given by the histogram (bars). In the left figure, the blue dashed line is the catchability (q) you defined for the simulations, this is a fixed unchanging value. In the figure on the right, the red dashed line is the average natural mortality (m) you defined, note that natural mortality varies each year in our simulations, but our Stock Assessment Model (SAM) is constrained to estimate one m for the entire time series.

Here are some useful biomass figures. The first is the q-corrected Model Estimated biomass compared to the *survey* biomass index. The second is the Stock Assessment estimated biomass compared to the *Actual* simulated biomass. In the bottom panel we have the same data but calculated the percent difference between the “Actual Biomass”. **Recall, the Actual simulated biomass would be ‘unknown’ in a real world Stock Assessment, here because we are just doing simulations we can use this as a check of how well our models have done**.

So how well is the Stock Assessment Model (SAM) doing at recreating the survey adult and recruit biomass indicies?

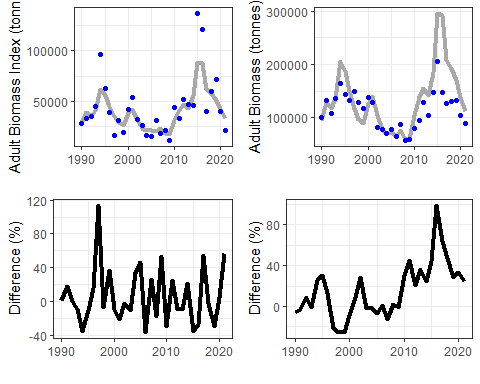


Figure 3: The SAM adult biomass results compared to (TOP LEFT) the survey index adult biomass, here we have q-corrected the SAM adult biomass (the grey line) and compared it to the survey adult biomass index. (TOP RIGHT) we have the SAM adult biomass (grey line) compared to the Actual Adult Biomass. (BOTTOM LEFT) is the percent difference between the SAM adult biomass and the survey index adult biomass. (BOTTOM RIGHT) is the percent difference between the SAM adult biomass and the Actual adult biomass.

Table 3: A Table showing how well SAM adult biomass estiamtes align with the input survey data and the acutal adult biomass

|  |  |  |
| --- | --- | --- |
| Comparsion | Tonnes | Percentage |
| Mean Adult Biomass (SAM - Survey) | -2900 | 6.3 |
| Mean Adult Biomass (SAM - Actual) | 20000 | 16.0 |
| Mean Adult Biomass |SAM - Survey| | 10000 | 25.0 |
| Mean Adult Biomass |SAM - Actual| | 29000 | 23.0 |

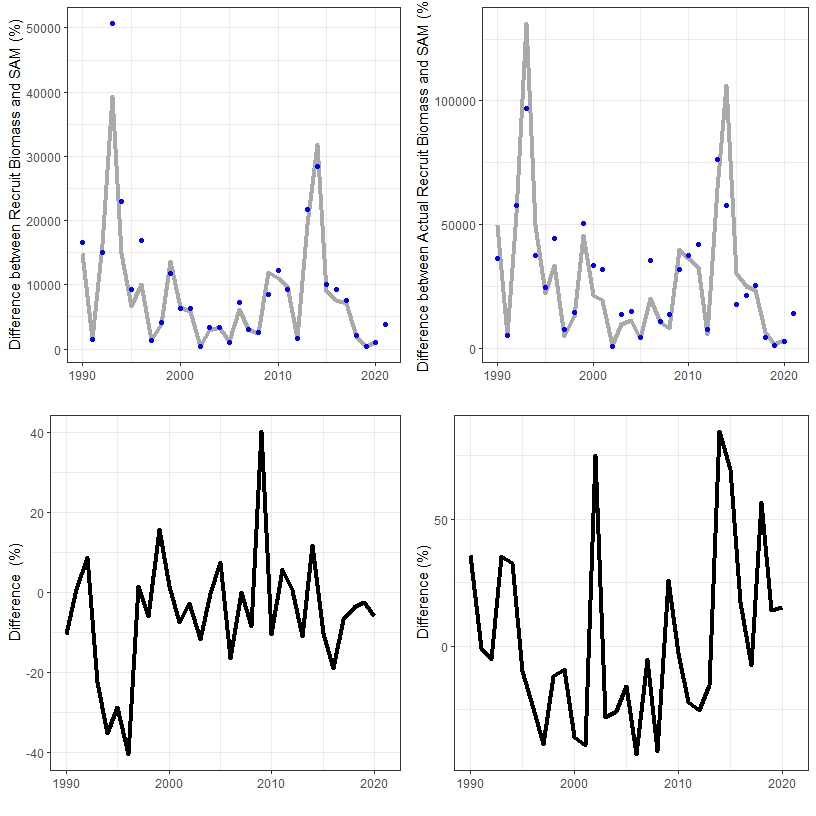


Figure 4: The SAM recruit biomass results compared to (TOP LEFT) the survey index recruit biomass, here we have q-corrected the SAM recruit biomass (the grey line) and compared it to the survey recruit biomass index. In (TOP RIGHT) we have the SAM recruit biomass (grey line) compared to the actual recruit Biomass. (BOTTOM LEFT) is the percent difference between the SAM Recruit biomass and the survey index Recruit biomass. (BOTTOM RIGHT) is the percent difference between the SAM Recruit biomass and the Actual Recruit biomass.

Table 4: A Table showing how well SAM recruit biomass estiamtes align with the input survey data and the acutal recruit biomass

|  |  |  |
| --- | --- | --- |
| Comparsion | Tonnes | Percentage |
| Mean Recruit Biomass (SAM - Survey) | -940 | -5.2 |
| Mean Recruit Biomass (SAM - Actual) | 960 | 1.7 |
| Mean Recruit Biomass Error |SAM - Survey| | 1600 | 11.0 |
| Mean Recruit Biomass Error |SAM - Actual| | 7800 | 28.0 |

Next we can look at the actual natural mortality and compare it to the mean of our prior distribution. *Hint* Think about how the variability in the actual natural mortality that isn’t being captured in our model would impact the model results.

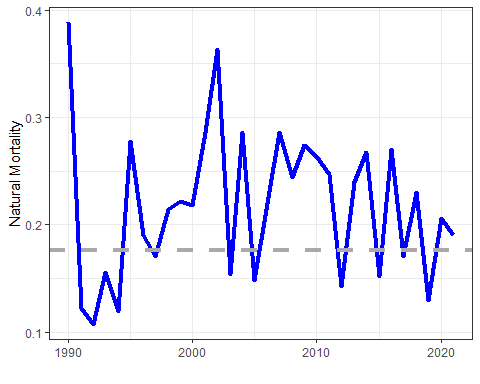


Figure 5: The actual natural mortality (blue line) time series, the grey dashed line is the median of the natural mortality parameter used in the SAM.

We can also explore how different the exploitation rate from the SAM model is to the actual exploitation rate.

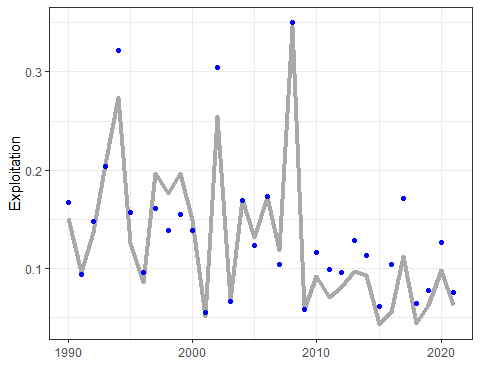


Figure 6: The SAM exploitation rate time series (grey line) compared to the actual exploitation rate (blue points).

# Terms and Jargon

*Important*: Pretty much all of these ‘definitions’ are simplifications that serve our purposes, but if you go deeper into Fisheries Science you will see that it is often much more complex than this.

* *Adults:* Mature individuals that are targeted by the fishery
* *Recruits:* Individuals that will mature and be targeted by the fishery next year
* *Biomass:* The Abundance of individual in a certain size/age class multiplied by the weight of those indivduals. Usually the weight is taken as some sort of average of the population
* *Growth:* The increase in size (mass) of a class of fish in a given year
* *Survey:* The process in which a stock is sampled to get an estimate of population status (e.g. abundance index) and life history parameters (e.g. growth)
* *Survey Design:* The method used to carry our a survey. For our purposes this includes the number of sampling stations and the type of stratification (if any) to use.
* *Stratification:* A method used to divide up a survey into areas with ‘similar’ characteristics. If done properly it will reduce the uncertainty of your survey indices compared to a survey in which the stations are simply randomly allocated.
* *Stock Assessment Model:* A model which uses data from a survey and/or fishery data to get metric(s) of population status (usually an estimate of stock biomass)
* *Parameter:* An input to or output from a model, generally parameters are ‘fixed’ values (e.g catchability is a parameter).
* *Variable:* An input to or output from a model, generally variables can vary (e.g. biomass is a variable).
* *Prior:* Used to inform your model what the most likely range of values are for a particular parameter. This is a “Bayesian statistics” concept.
* *Catchability:* The proportion of the individuals in the area sampled that are captured by a survey tow
* *Natural Mortality:* The proportion of the population that dies from ‘natural causes’ in a given year.
* *Exploitation:* The proportion of the population that is captured by the fishery in a given year.
* *Simulation:* Process of using certain mathematical/statistical equations and tools to develop data and test the impact of varying inputs and assumptions on the end results. Stochastic simulations incorporate statistical uncertainty in model parameters to generate multiple outcomes by accounting for various types of uncertainty.
* *Realizations:* When running a stochastic simulation a realization is one set of results, there can be many realizations.