Simple Statistical Catch at Age model

Anders Nielsen & Olav Nikolai Breivik

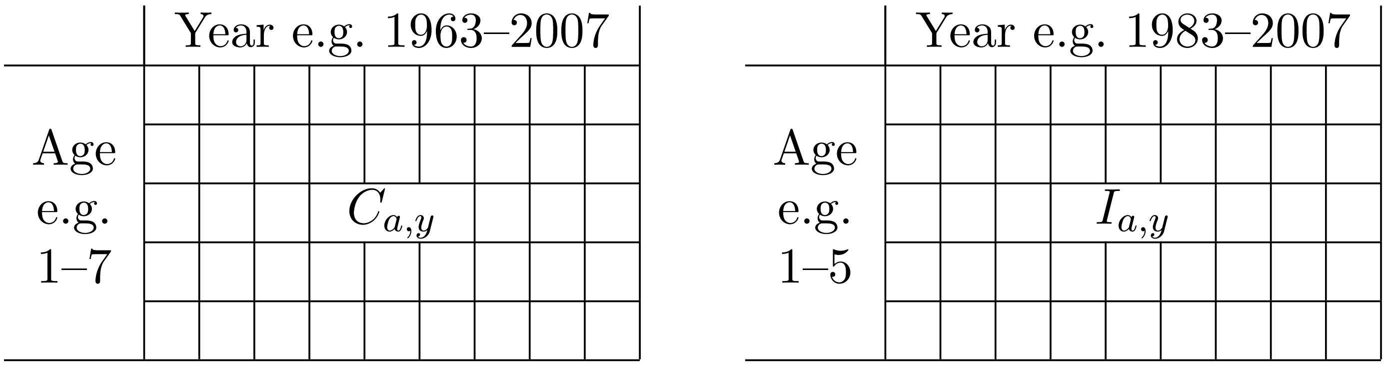
Fish Stock Assessment

Problem: How many fish (relative or absolute) are left in the ocean?

Data:

$C_{a,y}$**:** Yearly catches (divided into age–classes)

$I_{a,y}$**:** Scientific surveys

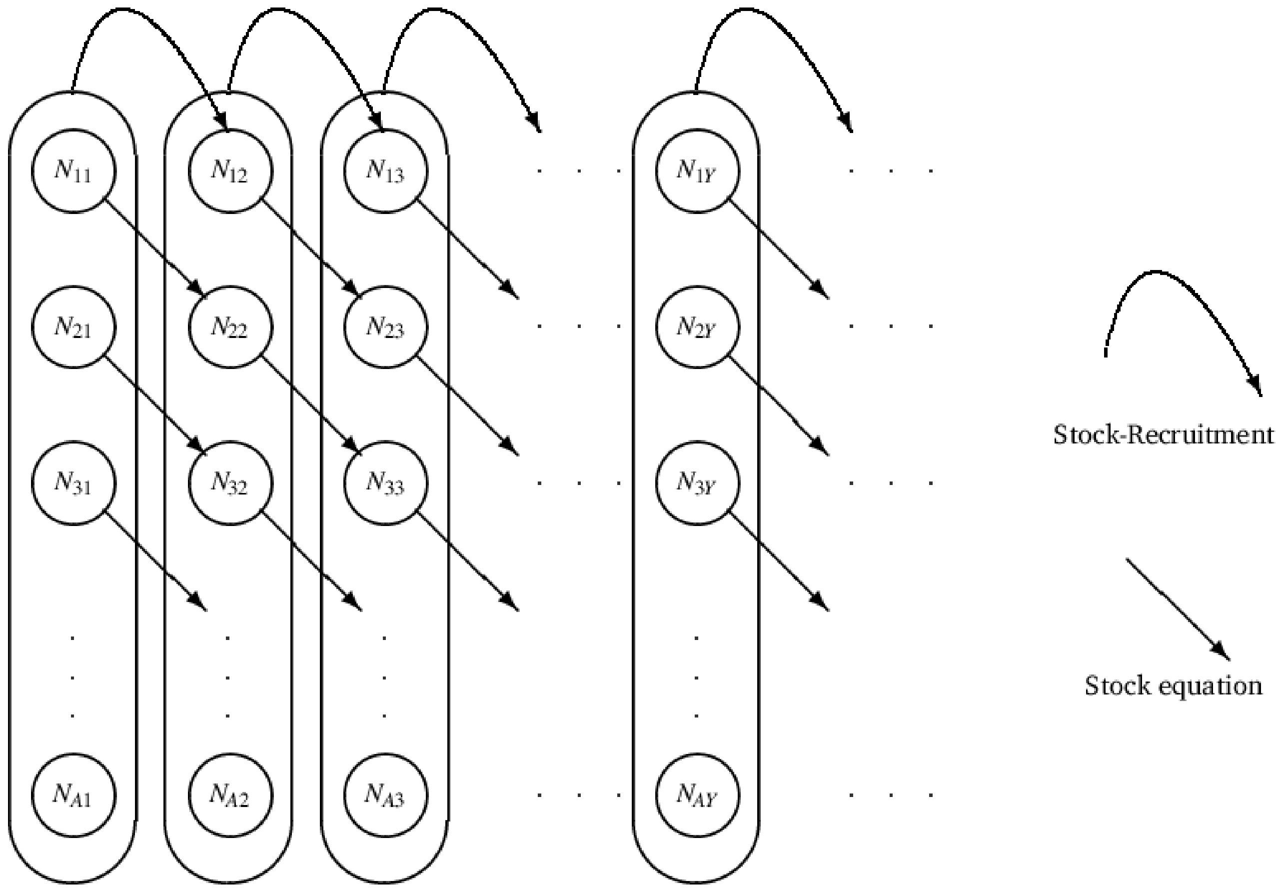


Often we have catches ($C_{f,a,y}$) from more than one fleet and indices ($I_{s,a,y}$) from more than one survey, but here we keep it simple.

Important output: SSB, $\overline{F}$, "Reference points" ...

Assessment Models

- Based on a standard set of equations describing the structure of the system
- Can include different assumptions about selectivity
- Can include different assumptions about 'natural' mortality
- Can include different effort numbers
- Can include different assumptions about recruitment



- But after all that has been taken into account it is a simple function connecting catch to number of fish

Basic equations

Stock equation: The number of fish in a cohort is expected to follow:

$$\frac{dN_t}{dt} = -(F_t + M_t)N_t$$

If F and M are assumed constant within each year we get:

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

Catch equation: The number of fish in a cohort after one year can be separated into:

$$N_{a,y} = \underbrace{N_{a+1,y+1}}_{\text{survived}} + \underbrace{C_{a,y} + D_{a,y}}_{\text{died}}$$

The part of the cohort that was caught during the year is:

$$C_{a,y} = \frac{F_{a,y}}{F_{a,y} + M_{a,y}}\left(1 - e^{-(F_{a,y}+M_{a,y})}\right)N_{a,y}$$

Stock–recruitment equation: Obviously connected — Different opinions about how

Deterministic models

- A deterministic model is a model where observation noise is ignored
- Typically catches are assumed known without error
- Most commonly applied fish stock assessment models are (semi-)deterministic
- These algorithms work (very simply put) by:

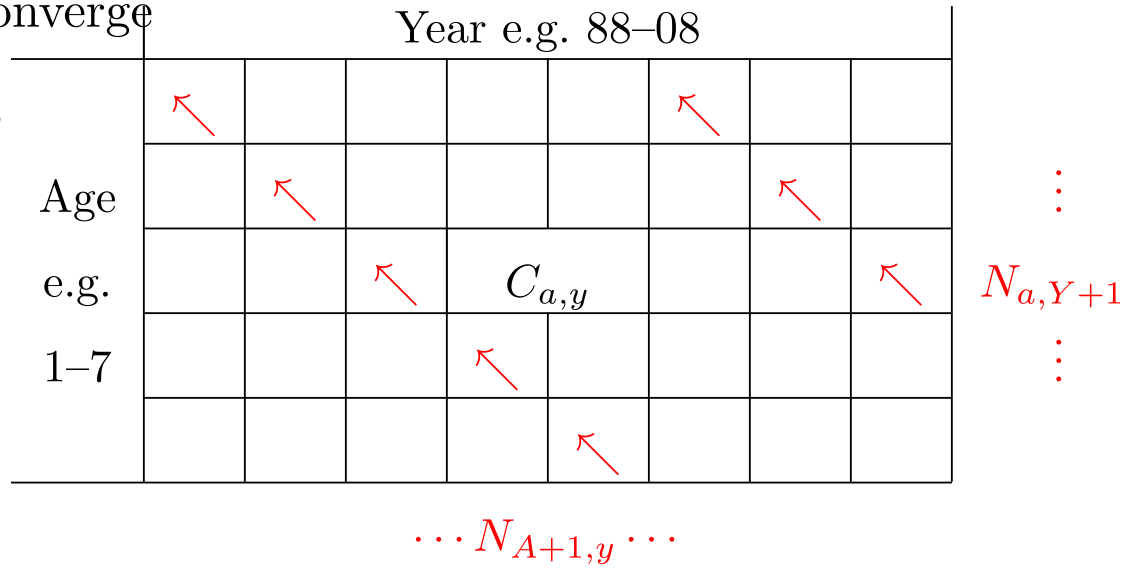
  0: Guess the number of survivors $N_{A+1,y}$ and $N_{a,Y+1}$

  1: Back calculate ($\nwarrow$) all $N_{a,y}$ by subtracting catch and natural mortality

  2: Use surveys to adjust all $N_{a,y}$ and update survivors accordingly

  3: Repeat 1-3 until survivors converge

- Doing 0-1 just ones is known as Virtual Population Analysis



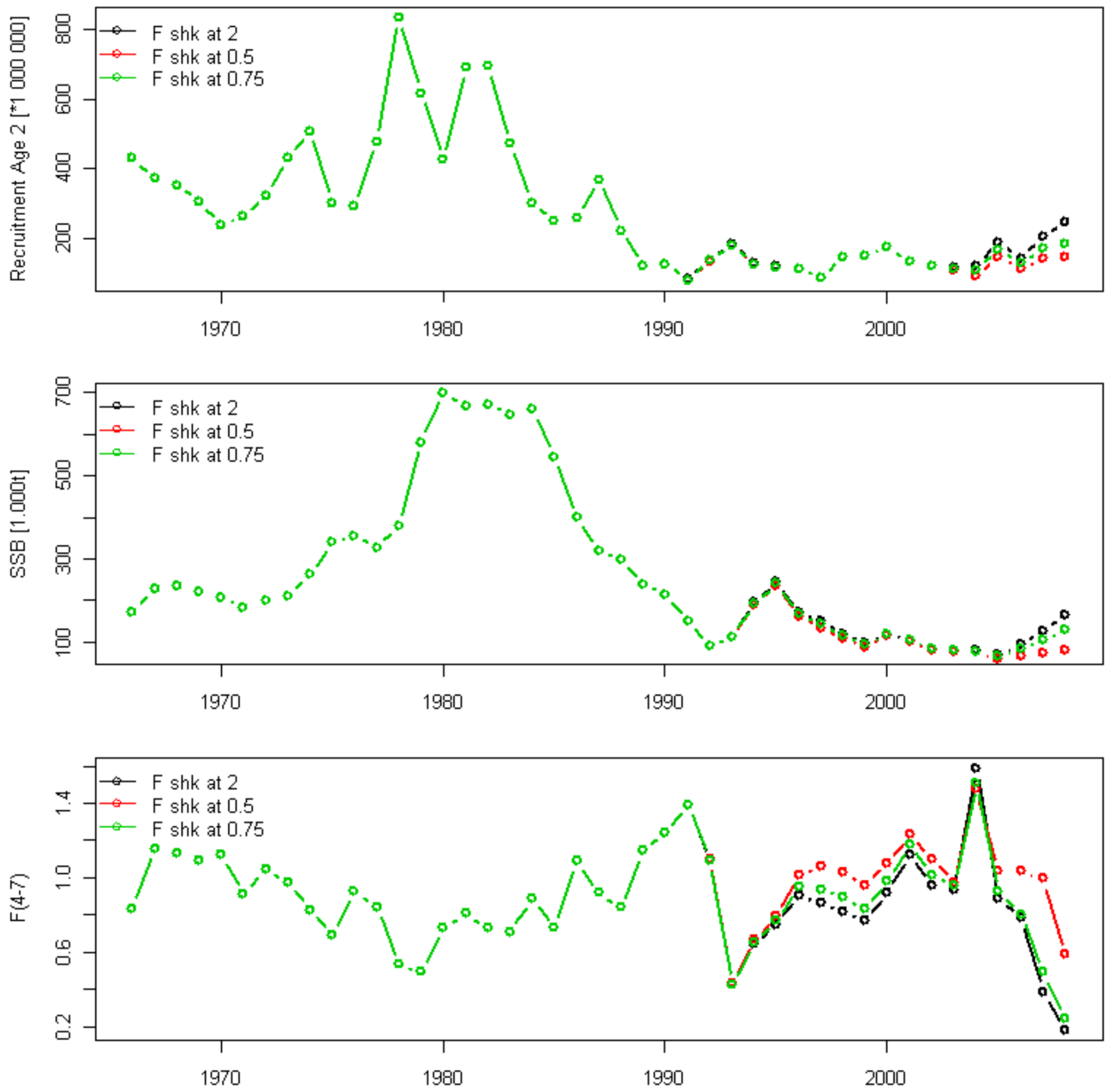
Features of deterministic models

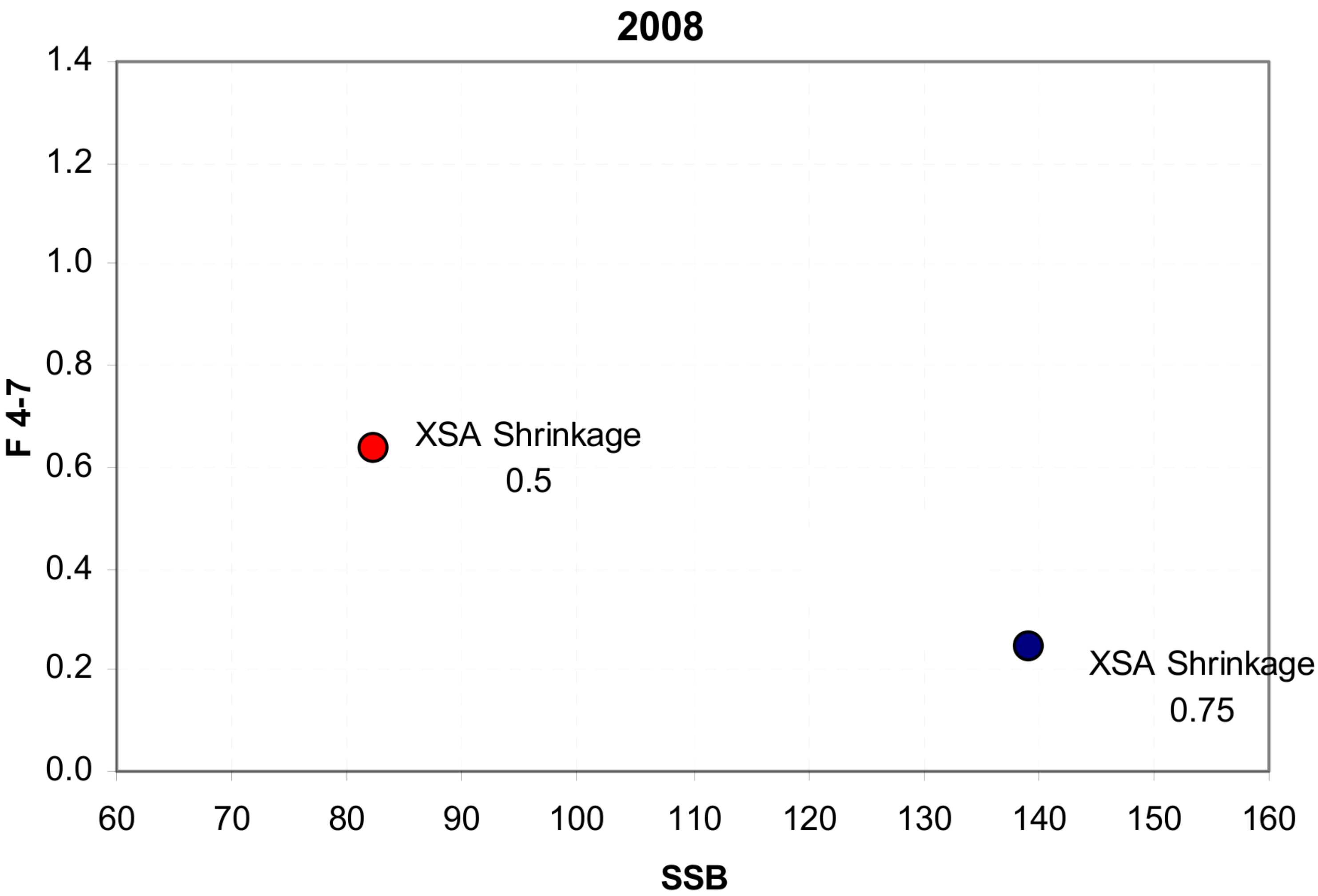
+ Super fast to compute
+ Fairly simple to explain the path from data to stock numbers (especially VPA)
– Difficult to explain why it works (converges), and what a solution mean
– These algorithms contain many ad-hoc settings (shrinkage, tapered time weights, ...) that makes them less objective
– No quantification of uncertainties within model

? What exactly is the model
  - The assumptions are difficult to identify and verify
  - With no clearly defined model more ad-hoc methods are needed to make predi- ctions

- No framework for comparing models (different settings)

Example: F-shrinkage for Eastern Baltic Cod





- These differences are not small and theoretical
- There are no objective way to choose between these two deterministic approaches
- Things would be simpler if we had a statistical model

A full parametric statistical model

- The log catches are assumed to follow:

$$\log(C_{a,y}) \sim \mathcal{N}\left(\log\left(\frac{F_{a,y}}{Z_{a,y}}(1-e^{-Z_{a,y}})N_{a,y}\right), \sigma_c^2\right) \quad , \text{ where}$$

$$F_{a,y} = F_y F_a, \text{ with } F_{a=5} = F_{a=6} = F_{a=7} = 1, \text{ and } Z_{a,y} = F_{a,y} + M_{a,y}$$

- The log catches from the survey are assumed to follow:

$$\log(I_{a,y}) \sim \mathcal{N}\left(\log\left(Q_a e^{-Z_{a,y}T} N_{a,y}\right), \sigma_s^2\right) \quad , \text{ where}$$

  T is the fraction into the year where the survey is taken, and Q_a is catchability parameter.

- The stock sizes are assumed to follow:

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

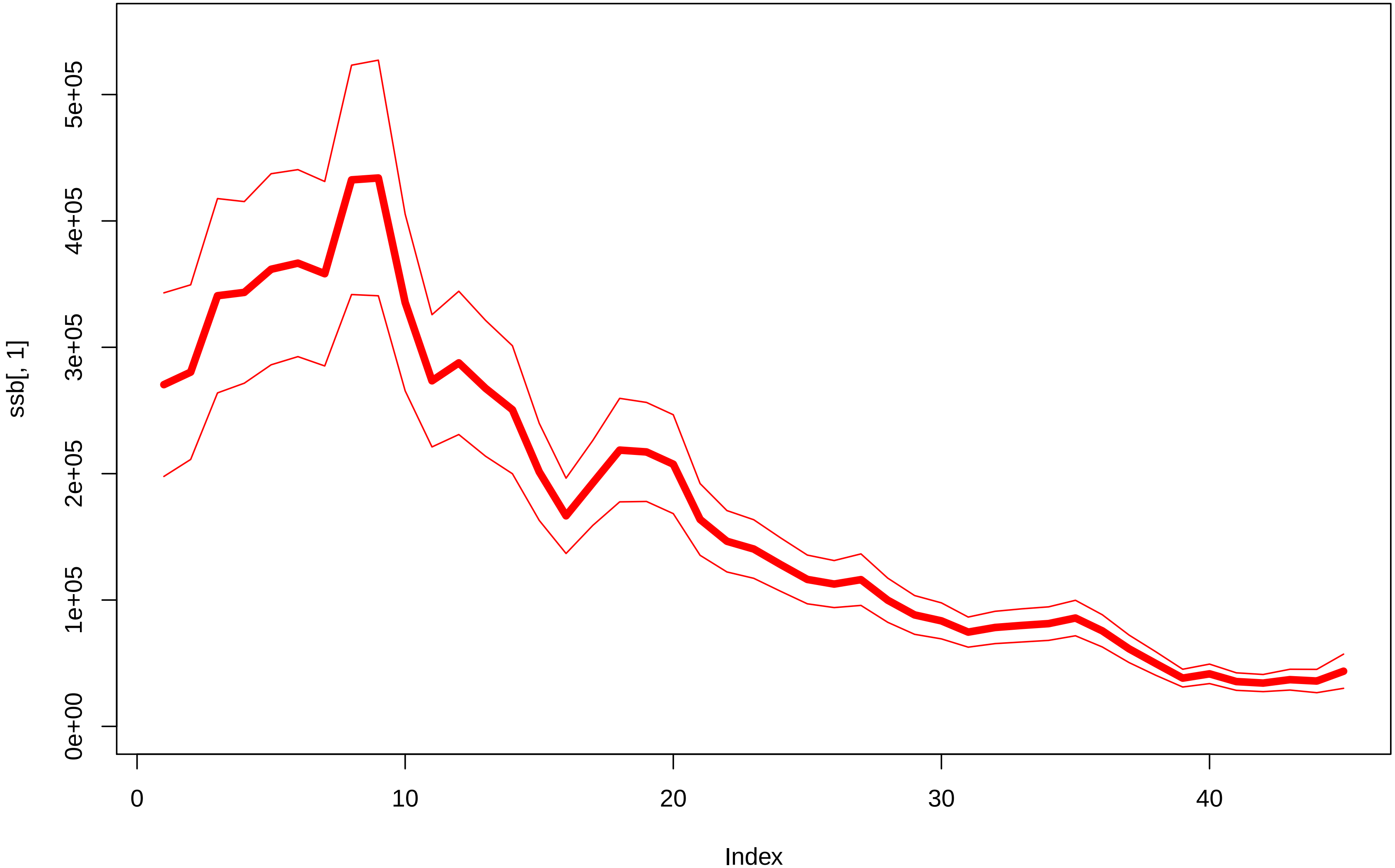
  Notice that it does not define N in the first year and for the youngest age.

- So the model parameters are the undefined N's, F_y, F_a, Q_a, σ_c, and σ_s

Implementation

- For this example this model has 107 model parameters (try to count them).
- Optimizing the likelihood for such a model is in principle the same as all the models we already have seen.
- With TMB `http://www.tmb-project.org` it is not much more difficult
- The data file for this example `fsa.RData` is in the shared folder
- The model implementation `fsa.R` and `fsa.cpp` is in the shared folder
- Now let's look into the code.

Result



Fully parametrized statistical assessment models

- A statistical[a] model acknowledges observation noise
- The error structure is part of the model description
- To find the quantities of interest (e.g. $N_{a,y}$, $F_{a,y}$, and observation uncertainties) the likelihood of the actual observations is optimized w.r.t. the model parameters.
- Parametrized statistical assessment models have a number of benefits:
  - + All model assumptions are transparent
  - + Different model assumptions can be tested against each other (e.g. is $F_5 = F_6$?)
  - + Different data sources can be included and correctly and objectively weighted
  - + Estimation of uncertainties are an integrated part of the model
- But also a few difficulties:
  - – Trade-off between the number model parameters and flexibility of the model (e.g. $F_{a,y}$ vs. $F_{a,y} = S_a f_y$)
  - – More advanced software needed

[a] a.k.a. stochastic

Exercise

- Get the data and the model file from the shared folder. Compile and run it on your own computer.
- The current simple implementation assumes that the catches from age group seven consists of only 7 year old fish. In reality catches from age group seven consists of fish of ages 7 and above. We say group 7 is a plus group. This means that for the last age A the stock equation should be updated to:

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

  make the necessary changes.
- Catches of one year old fish are often to a large extent determined by discard estimates. Extend the model to use a separate variance for the log-catches of one year old fish.
- To test if the survey catchabilities have changed over time, try to extend the model to use one set of catchabilities before year 2000, and a different set after and including.