

Simple Statistical Catch at Age model

Anders Nielsen & Arni Magnusson

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

	Year e.g. 1963–2007								
Age e.g. 1–7									
				$C_{a,y}$					

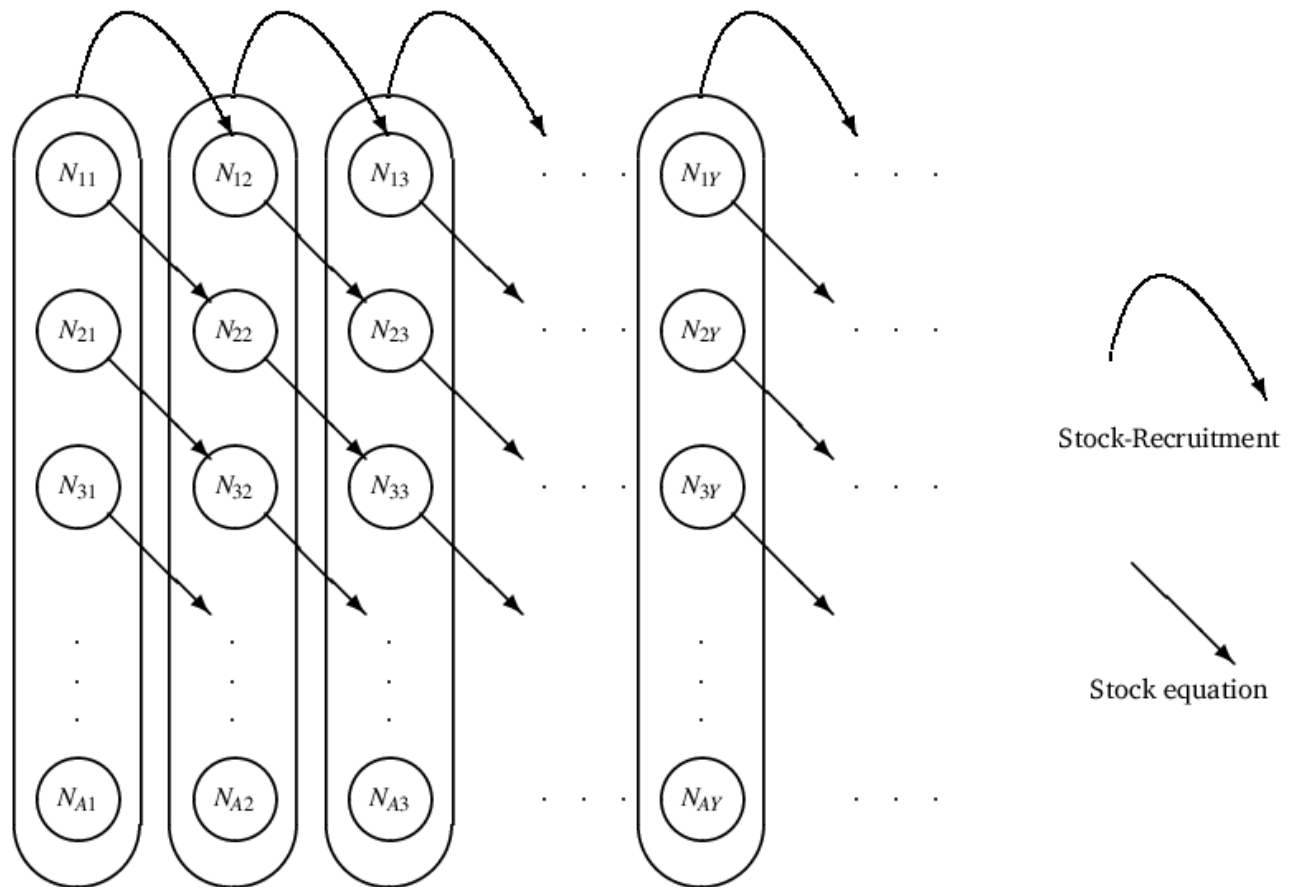
	Year e.g. 1983–2007								
Age e.g. 1–5									
				$I_{a,y}$					

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: “Reference points” SSB, \bar{F} , ...

Assessment Models

- Based on a standard set of equations describing the structure of the system
- Can include selectivity
- Can include different assumptions about ‘natural’ mortality
- Can include different effort numbers
- Can include different assumptions about recruitment
- When all of these things have been taken into account:



The number of fish is assumed to be proportional to the catch

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

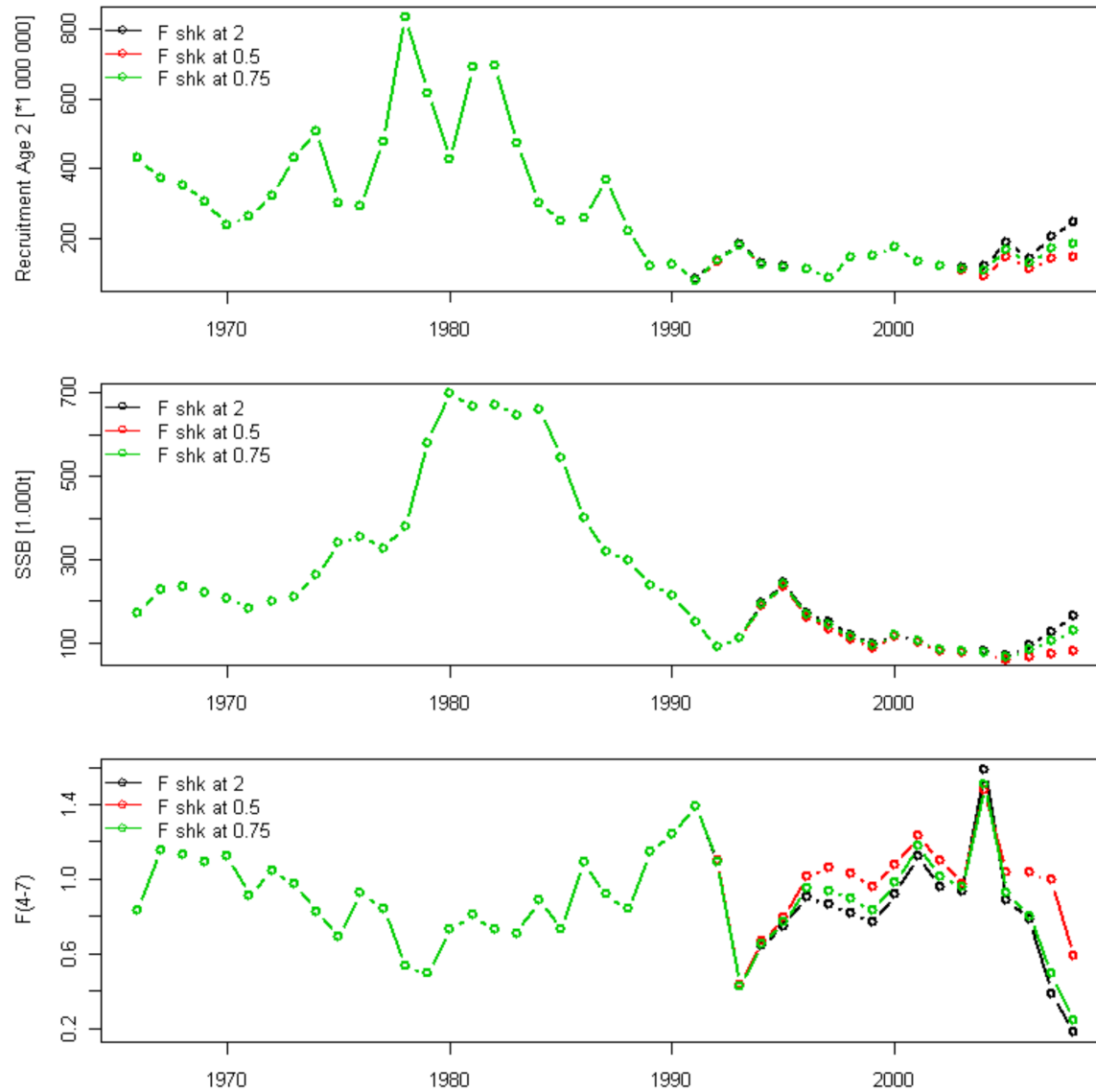
- Doing 0-1 just ones is known as Virtual Population Analysis

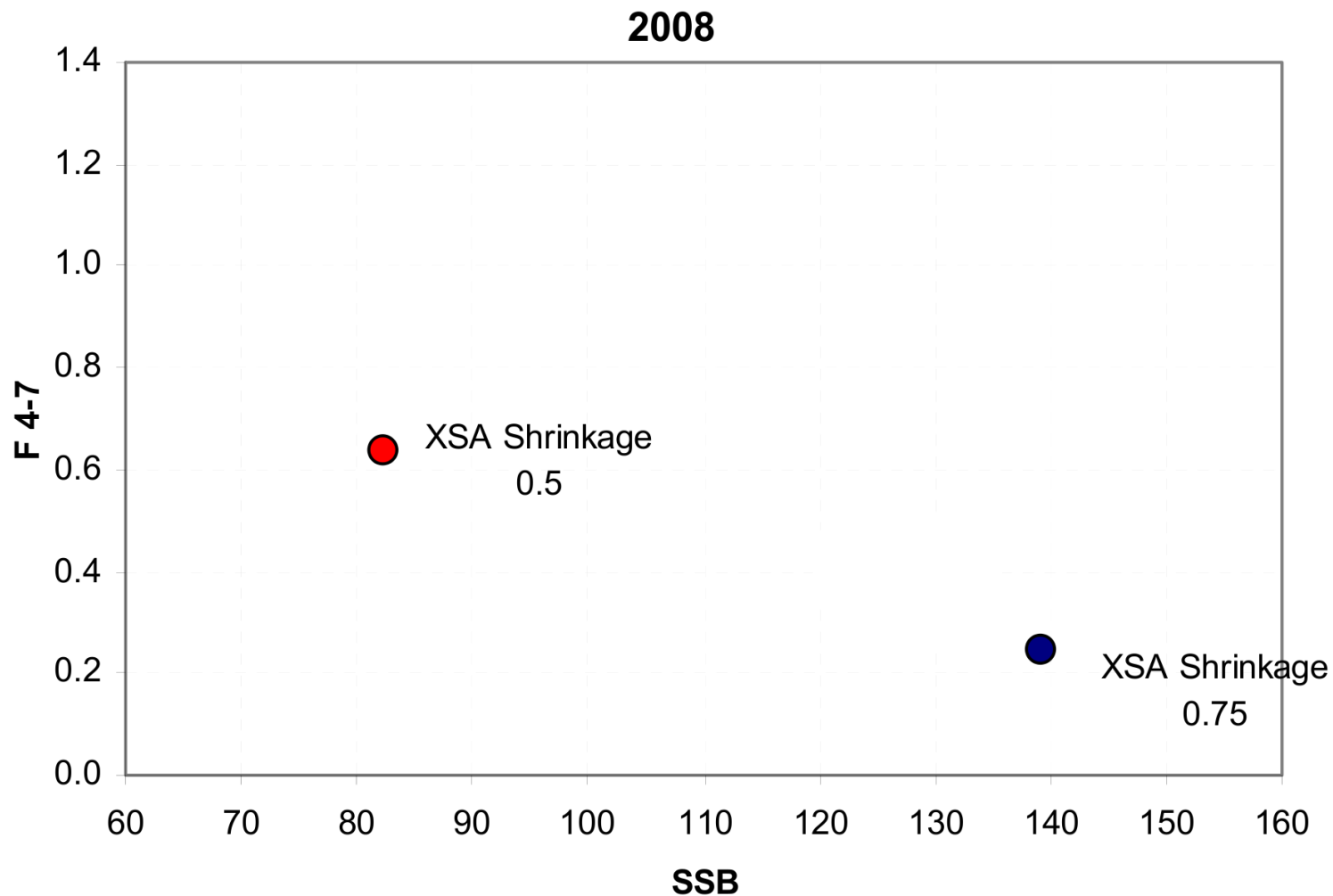
		Year e.g. 88–08								$N_{a,Y+1}$ \vdots
Age e.g. 1–7		\nwarrow					\nwarrow			
			\nwarrow					\nwarrow		
				\nwarrow	$C_{a,y}$				\nwarrow	
					\nwarrow					
						\nwarrow				
$\cdots N_{A+1,y} \cdots$										

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 predictions
 - 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:

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

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

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

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

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:

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

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 requires specialized software, but in principle it is the same as for all the models we already have seen.
- Such a tool is AD Model Builder <http://www.admb-project.org>
- The data file for this example is here:
<http://www.nielsensweb.org/TCADSA/sca/fsa.dat>
- The model implementation is here:
<http://www.nielsensweb.org/TCADSA/sca/fsa.tpl>

DATA_SECTION

DATA_SECTION

```
init_int minAge
init_int maxAge
init_int minYear
init_int maxYear
init_matrix catchNo(minAge,maxAge,minYear,maxYear)
init_matrix stockMeanWeight(minAge,maxAge,minYear,maxYear)
init_matrix propMature(minAge,maxAge,minYear,maxYear)
init_matrix M(minAge,maxAge,minYear,maxYear)

init_int minAgeS
init_int maxAgeS
init_int minYearS
init_int maxYearS
init_number surveyTime
init_matrix survey(minAgeS,maxAgeS,minYearS,maxYearS)

matrix logC(minAge,maxAge,minYear,maxYear)
!! logC=log(catchNo);
matrix logSurvey(minAgeS,maxAgeS,minYearS,maxYearS)
!! logSurvey=log(survey);
```

PARAMETER_SECTION

PARAMETER_SECTION

```
init_vector logN1Y(minAge,maxAge)
init_vector logN1A(minYear+1,maxYear)
init_vector logFY(minYear,maxYear)
init_vector logFA(minAge,maxAge-3)
init_number logVarLogCatch
init_vector logQ(minAgeS,maxAgeS)
init_number logVarLogSurvey

objective_function_value nll

matrix F(minAge,maxAge,minYear,maxYear)
matrix logN(minAge,maxAge,minYear,maxYear)
matrix predLogC(minAge,maxAge,minYear,maxYear)
matrix predLogSurvey(minAgeS,maxAgeS,minYearS,maxYearS)
number ss
vector logFAextend(minAge,maxAge)
sdreport_vector ssb(minYear,maxYear)
sdreport_vector fbar24(minYear,maxYear)
```

PROCEDURE_SECTION

```
logFAextend.initialize();
logFAextend(minAge,maxAge-3)=logFA;
F=outer_prod(exp(logFAextend),exp(logFY));

logN.colfill1(minYear,logN1Y);
for(int y=minYear+1; y<=maxYear; ++y){
    logN(minAge,y)=logN1A(y);
    for(int a=minAge+1; a<=maxAge; ++a){
        logN(a,y)=logN(a-1,y-1)-F(a-1,y-1)-M(a-1,y-1);
    }
}

predLogC=log(F)-log(F+M)+log(1.0-exp(-F-M))+logN;

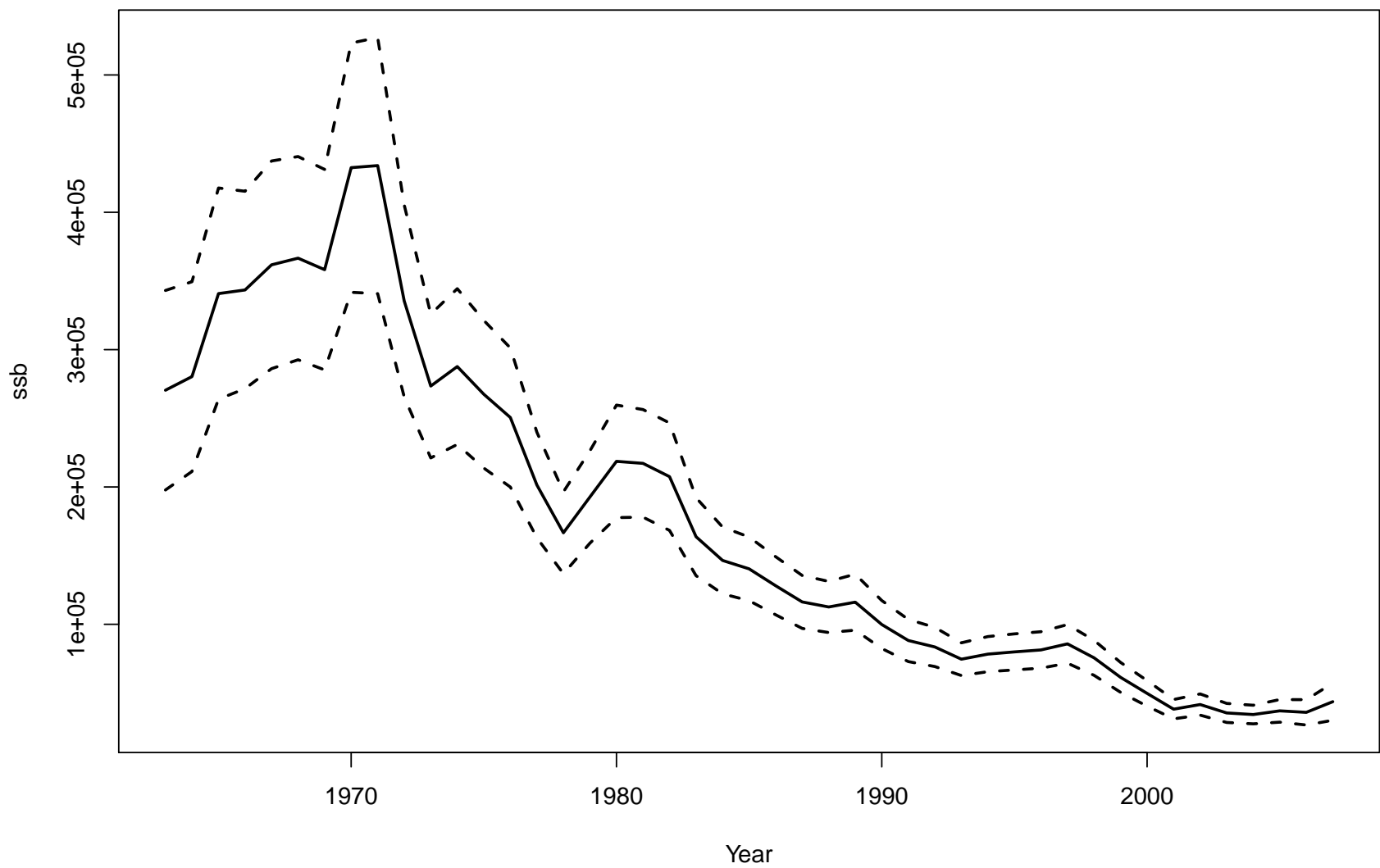
ss=exp(logVarLogCatch);
int N=(maxYear-minYear+1)*(maxAge-minAge+1);
nll=0.5*(N*log(2.0*M_PI*ss)+sum(square(logC-predLogC))/ss);

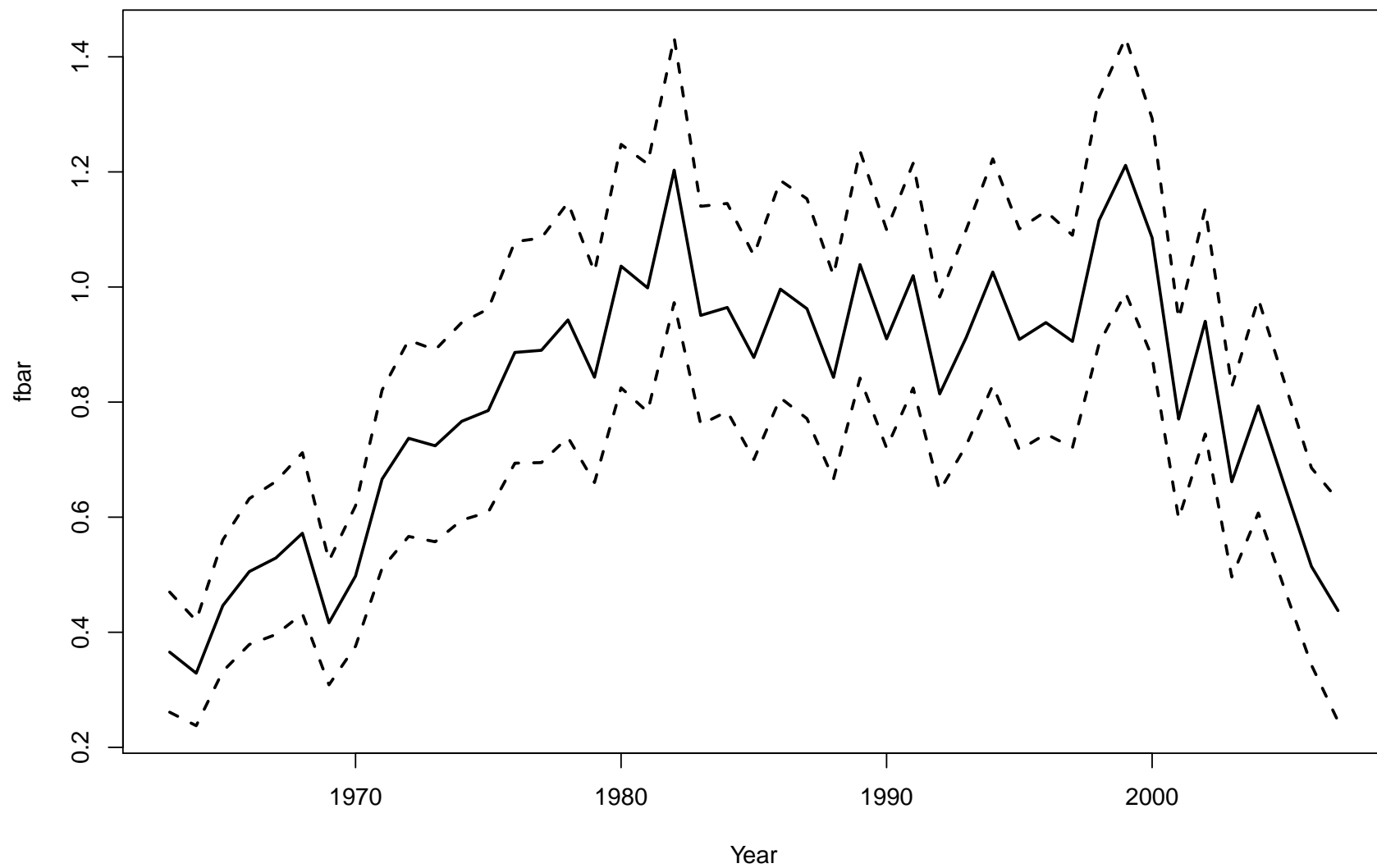
ss=exp(logVarLogSurvey);
for(int y=minYearS; y<=maxYearS; ++y){
    for(int a=minAgeS; a<=maxAgeS; ++a){
        predLogSurvey(a,y)=logQ(a)-(F(a,y)+M(a,y))*surveyTime+logN(a,y);
        nll+=0.5*(log(2.0*M_PI*ss)+square(logSurvey(a,y)-predLogSurvey(a,y))/ss);
    }
}

ssb=colsum(elem_prod(elem_prod(exp(logN),propMature),stockMeanWeight));
fbar24=colsum(F.sub(2,4))/3.0;
```

Results

- When this is run (takes less than a second) we get few result files
- The main one is:
<http://www.nielsensweb.org/TCADSA/sca/fsa.std>
- Is it OK to assume catches are known without error?
- Is the survey helpful?





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

^aa.k.a. stochastic

Exercise

- Get the data and the model file from:
 - <http://www.nielsensweb.org/TCADSA/sca/fsa.dat>
 - <http://www.nielsensweb.org/TCADSA/sca/fsa.tpl>

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.