UALG

Statistical catch at age models: Step by step guide in Excel

Einar Hjörleifsson

A short comment first

- For line code people Excel may be cumbersome, inefficient and (for some) confusing.
- However wide availability, almost daily usage, integrated instant graphical display and the existence of Solver make it an ideal teaching and testing medium.
- The model coding we do here is probably the only time most of us will ever go in this field.
- The intent with doing it at least once is to disentangle the mathematical formulations that are the basis for statistical catch at age models.

The model in math

$$s_a = \begin{cases} e^{\frac{-(a-a_{full})^2}{\sigma_L}} & \text{for } a \le a_{full} \\ e^{\frac{-(a-a_{full})^2}{\sigma R}} & \text{for } a > a_{full} \end{cases}$$

$$F_{ay} = s_a F_y$$

$$N_{a,y} = \begin{cases} R_{a,y} & a = 1 \quad \text{or } y = 1 \\ N_{a-1,y-1} e^{-(F_{a-1,y-1} + M_{a-1,y-1})} & 1 < a < a_{plus} \\ N_{a-1,y-1} e^{-(F_{a-1,y-1} + M_{a-1,y-1})} + N_{a,y-1} e^{-(F_{a,y-1} + M_{a,y-1})} & a = a_{plus} = 11 \end{cases}$$

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

$$\hat{U}_{ay} = q_a N_{ay}$$

$$\min SSE = \lambda_C \sum_y \sum_a \omega_a \left(\ln C_{ay} - \ln \hat{C}_{ay}\right)^2 + \lambda_U \sum_y \sum_a \rho_a \left(\ln U_{ay} - \ln \hat{U}_{ay}\right)^2$$

Can you disentangle this?

The approach

- I hear - I forget
- I see - I remember
- I do - I understand

The model in words

- Make a separable model having:
  - A fixed (constant through time) selection pattern (s_a) for each age, assume selection pattern follows double-half Gaussian
    - Fixed selectivity with time is commonly referred to as a separable model.
  - Fishing mortality (for some reference age) for each year (F_y)
  - Numbers of fish that enter the stock each year (year class size, recruitment, $N_{1,y}$) and in the first year ($N_{a,1}$)
  - A plus group: Catches of the oldest age groups are summed - Needs to be taken into account in the model
- Calculate:
  - The number of fish caught each year and age by the fishermen (C_{ay}-hat). This is the modeled Cay number.
  - The number of fish caught each year and age by the scientist (U_{ay}-hat). This is the modeled Uay number.
    - Assume the relationship between stock size and survey indices as: $Uay = qN^{\beta}$
- Set up an objective function:
  - Constrain the model such that we minimize the squared difference between observed values (Cay and Uay) and predicted values (Cay-hat and Uay-hat)

The model as a map

einar

The separable part

- Selectivity describes the relative fishing mortality within each age group.
- In this simplest model setup we assume that the selectivity is the same in all years. Fishing mortality by age and year can thus be described by:

$$F_{ay} = s_a F_y$$

  - F_{ay}: Fishing mortality of age a year y
  - s_a: Selectivity of age a
  - F_y: Fishing mortality (of some reference age) in year y
    - Note: The separability assumption reduces the number fishing mortality parameters from:
      n = (#age groups x #years) to
      n = (#age groups + #years)

Assume double half-Gaussian

einar

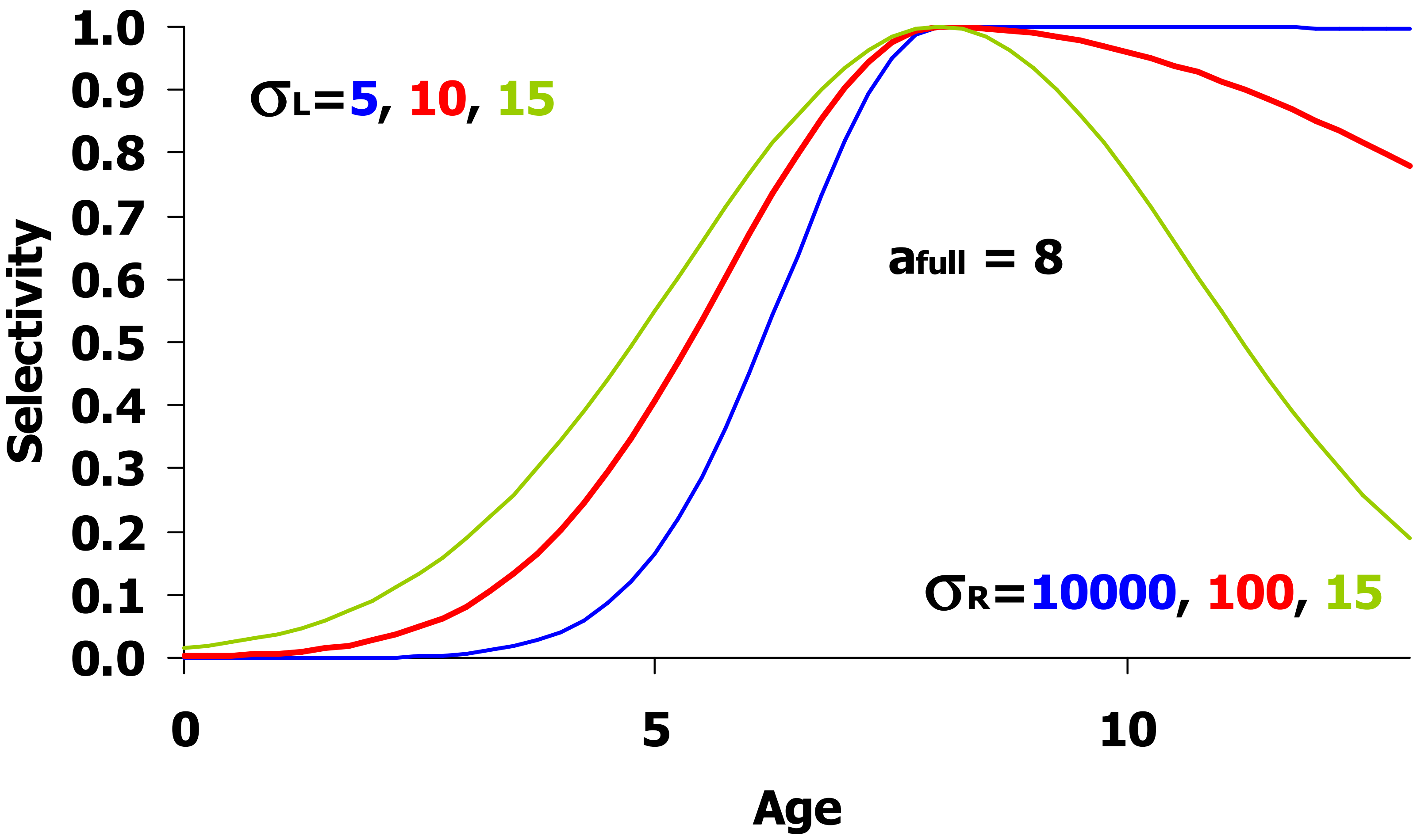
- Lets make a further assumption here by letting selectivity follow:

$$s_a = \begin{cases} e^{\frac{-(a-a_{full})^2}{\sigma_L}} & \text{for } a \le a_{full} \\ e^{\frac{-(a-a_{full})^2}{\sigma R}} & \text{for } a > a_{full} \end{cases}$$

  - a_{full}: age at full selectivity
  - σ_R: Shape factor (standard deviation) for right hand curve
  - σ_L: Shape factor (standard deviation) for left side of curve
  - Note: by using this selection function we reduce the number of parameters from whatever number of age groups we have, to only 3 parameters.
  - But could just as well just estimate each Sa without resorting to a particular function.
- The σ_R, σ_L and Afull are parameters that we estimate

Selectivity - double half-Gaussian

Note asymmetry



The map

9. Solver

Store Afull, σL and σR here

1. Parameters

Year | Age -----> | V
2. Selection pattern
Calculate sa here

Year | Age -----> | V
3. Fishing mortality

Year | Age -----> | V
Natural mortality

Year | Age -----> | V
4a. Population numbers

Year | Age -----> | V
4b. Nay at survey 1 time

Year | Age -----> | V
4c. Nay at survey 2 time

Year | Age -----> | V
5. Predicted catch

Year | Age -----> | V
6a. Predicted survey 1 indices

Year | Age -----> | V
6b. Predicted survey 2 indices

Year | Age -----> | V
Observed catch at age

Year | Age -----> | V
Observed survey 1 indices

Year | Age -----> | V
Observed survey 2 indices

Year | Age -----> | V
7. Catch residuals

Year | Age -----> | V
8a. Survey 1 residuals

Year | Age -----> | V
8b. Survey 2 residuals

Year | Age -----> | V
7. Catch residuals squared

Year | Age -----> | V
8a. Survey 1 residuals squared

Year | Age -----> | V
8b. Survey 2 residuals squared

Color code
Population model
Observation mo
Measurements
Objectives
Parameters

The map: S_{ay} details

einar

Estimated parameters

Param.
σ_L
σ_R
a_{full}
Fy
..
..
F_{y+8}
N_{ay}
N_{ay+1}
N_{ay+2}
N_{ay+3}
N_{ay+4}
...

y/a	a	a+1	a+2	a+3	a+4	a+5
y	S_a	S_{a+1}	S_{a+2}	S_{a+3}	S_{a+4}	S_{a+5}
y+1	S_a	S_{a+1}	S_{a+2}	S_{a+3}	S_{a+4}	S_{a+5}
y+2	S_a	S_{a+1}	S_{a+2}	S_{a+3}	S_{a+4}	S_{a+5}
y+3	S_a	S_{a+1}	S_{a+2}	S_{a+3}	S_{a+4}	S_{a+5}
y+4	S_a	S_{a+1}	S_{a+2}	S_{a+3}	S_{a+4}	S_{a+5}
y+5	S_a	S_{a+1}	S_{a+2}	S_{a+3}	S_{a+4}	S_{a+5}
y+6	S_a	S_{a+1}	S_{a+2}	S_{a+3}	S_{a+4}	S_{a+5}
y+7	S_a	S_{a+1}	S_{a+2}	S_{a+3}	S_{a+4}	S_{a+5}
y+8	S_a	S_{a+1}	S_{a+2}	S_{a+3}	S_{a+4}	S_{a+5}

$$s_a = \begin{cases} e^{\frac{-(a-a_{full})^2}{\sigma_L}} & \text{for } a \le a_{full} \\ e^{\frac{-(a-a_{full})^2}{\sigma R}} & \text{for } a > a_{full} \end{cases}$$

Excel speak: =exp(-((a-afull)^2/if(a=<afull;sL;sR)))

Side step

A word on nomenclature

- Often make the following distinction:
  - Selectivity: The probability of catching an individual of a given age scaled to the maximum probability over all ages, given that all animals are available to be caught by a certain gear in a certain plaice.
    - This is what gear technologist study at lengths! when they are studying the properties of various gears.
  - Availability: The relative probability, as a function of age, of being in the area in which catching occurs.
  - Vulnerability: The combination of selectivity and availability.
    - Thus should really refer to vulnerability but lets stick with the more ambiguous word selectivity, the reason being its wide usage.

Setting up F_y and calculating F_{ay}

- The fishing mortality each year (F_y) are parameters of the model that we want to estimate.
- Since we already calculated s_a we can calculate fishing mortality by age and year from:
  - $F_{ay} = s_a F_y$

The map

einar

9. Solver

1. Para

Store F_y here

Year Age -----> 2. Selection pattern	Year Age -----> 3. Fishing mortality **Calculate F_{ay} here**	Year Age -----> Natural mortality
Year Age -----> 4a. Population numbers	Year Age -----> 4b. Nay at survey 1 time	Year Age -----> 4c. Nay at survey 2 time
Year Age -----> 5. Predicted catch	Year Age -----> 6a. Predicted survey 1 indices	Year Age -----> 6b. Predicted survey 2 indices
Year Age -----> Observed catch at age	Year Age -----> Observed survey 1 indices	Year Age -----> Observed survey 2 indices
Year Age -----> 7. Catch residuals	Year Age -----> 8a. Survey 1 residuals	Year Age -----> 8b. Survey 2 residuals
Year Age -----> 7. Catch residuals squared	Year Age -----> 8a. Survey 1 residuals squared	Year Age -----> 8b. Survey 2 residuals squared

Color code
Population model
Observation mo
Measurements
Objectives
Parameters

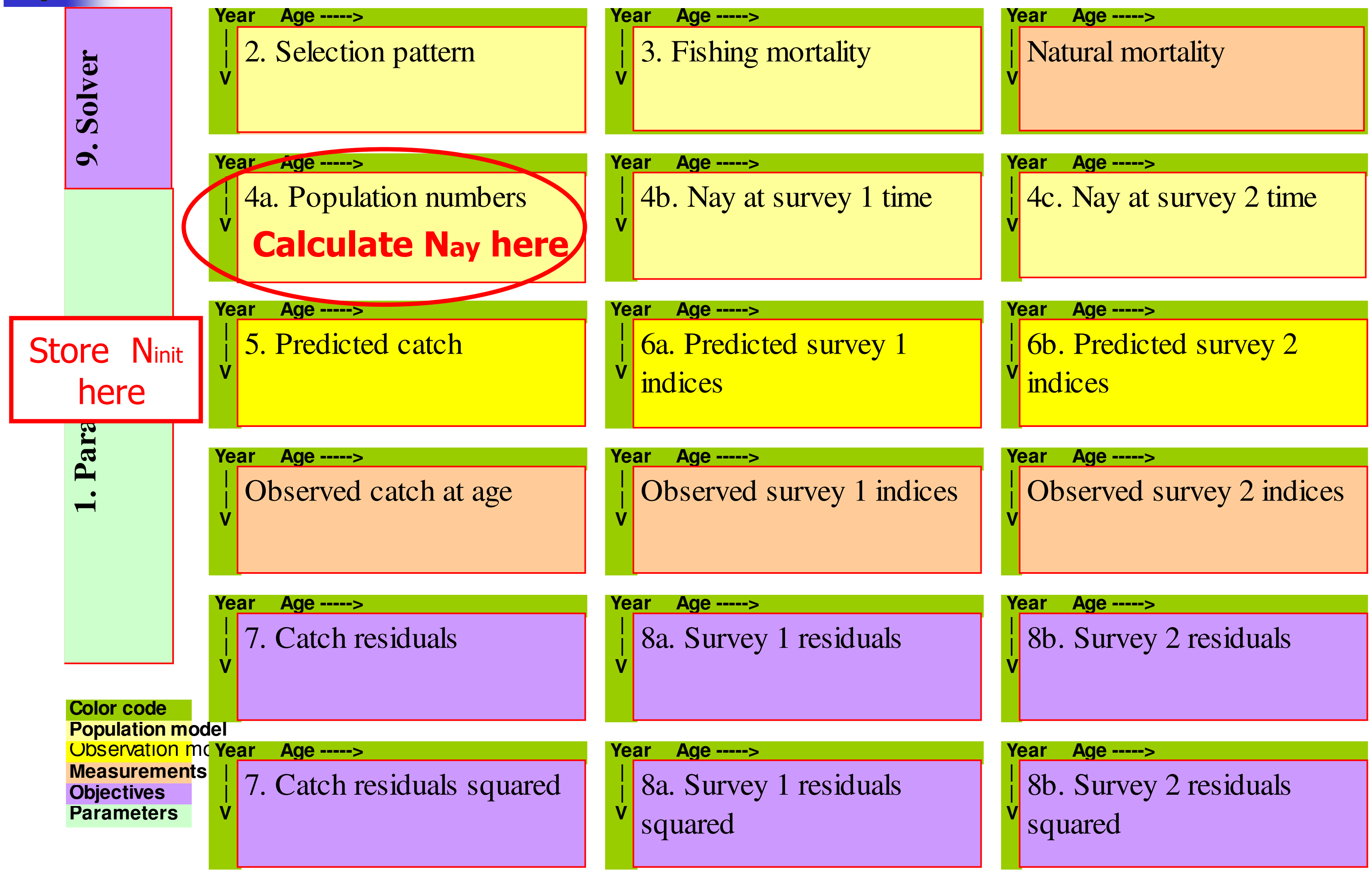
Setting up N_{init} and calculating N_{ay}

- The number of fish entering the system in first year and in the first age (N_{init}) are parameters of the model that we want to estimate. Need:
  - The number of fish in each age group in the first year ($N_{a,1}$)
  - The number of recruits entering each year ($N_{1,y}$)
- Given the above we can then fill in the abundance matrix by the conventional stock equation

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

The map

einar



The map: Nay details

Green area: Estimated parameters

Param.
σ_L
σ_R
a_{full}
Fy
..
..
F_{y+8}
N_{ay}
N_{ay+1}
N_{ay+2}
N_{ay+3}
N_{ay+4}
...

y/a	a	a+1	a+2	a+3	a+4	a+5
y	$N_{a,y}$	$N_{a+1,y}$	$N_{a+2,y}$	$N_{a+3,y}$	$N_{a+4,y}$	$N_{a+5,y}$
y+1	$N_{a,y+1}$	$N_{a+1,y+1}$	$N_{a+2,y+1}$	$N_{a+3,y+1}$	$N_{a+4,y+1}$	$N_{a+5,y+1}$
y+2	$N_{a,y+2}$	$N_{a+1,y+2}$	$N_{a+2,y+2}$	$N_{a+3,y+2}$	$N_{a+4,y+2}$	$N_{a+5,y+2}$
y+3	$N_{a,y+3}$	$N_{a+1,y+3}$	$N_{a+2,y+3}$	$N_{a+3,y+3}$	$N_{a+4,y+3}$	$N_{a+5,y+3}$
y+4	$N_{a,y+4}$	$N_{a+1,y+4}$	$N_{a+2,y+4}$	$N_{a+3,y+4}$	$N_{a+4,y+4}$	$N_{a+5,y+4}$
y+5	$N_{a,y+5}$	$N_{a+1,y+5}$	$N_{a+2,y+5}$	$N_{a+3,y+5}$	$N_{a+4,y+5}$	$N_{a+5,y+5}$
y+6	$N_{a,y+6}$	$N_{a+1,y+6}$	$N_{a+2,y+6}$	$N_{a+3,y+6}$	$N_{a+4,y+6}$	$N_{a+5,y+6}$
y+7	$N_{a,y+7}$	$N_{a+1,y+7}$	$N_{a+2,y+7}$	$N_{a+3,y+7}$	$N_{a+4,y+7}$	$N_{a+5,y+7}$
y+8	$N_{a,y+8}$	$N_{a+1,y+8}$	$N_{a+2,y+8}$	$N_{a+3,y+8}$	$N_{a+4,y+8}$	$N_{a+5,y+8}$

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

The map: N-plus group details

Param.
σL
σR
a_{50}
F_y
F_{y+1}
..
F_{y+8}
N_{ay}
N_{ay+1}
N_{ay+2}
N_{ay+3}
N_{ay+4}
...

y/a	a	a+1	a+2	..	a+10	a-plus
y	$N_{a,y}$	$N_{a+1,y}$	$N_{a+2,y}$		$N_{a+10,y}$	$N_{a+,y}$
y+1	$N_{a,y+1}$	$N_{a+1,y+1}$	$N_{a+2,y+1}$		$N_{a+10,y+1}$	$N_{a+,y+1}$
y+2	$N_{a,y+2}$	$N_{a+1,y+2}$	$N_{a+2,y+2}$		$N_{a+10,y+2}$	$N_{a+,y+2}$
y+3	$N_{a,y+3}$	$N_{a+1,y+3}$	$N_{a+2,y+3}$		$N_{a+10,y+3}$	$N_{a+,y+3}$
y+4	$N_{a,y+4}$	$N_{a+1,y+4}$	$N_{a+2,y+4}$		$N_{a+10,y+4}$	$N_{a+,y+4}$
y+5	$N_{a,y+5}$	$N_{a+1,y+5}$	$N_{a+2,y+5}$		$N_{a+10,y+5}$	$N_{a+,y+5}$
y+6	$N_{a,y+6}$	$N_{a+1,y+6}$	$N_{a+2,y+6}$		$N_{a+10,y+6}$	$N_{a+,y+6}$
y+7	$N_{a,y+7}$	$N_{a+1,y+7}$	$N_{a+2,y+7}$		$N_{a+10,y+7}$	$N_{a+,y+7}$
y+8	$N_{a,y+8}$	$N_{a+1,y+8}$	$N_{a+2,y+8}$		$N_{a+10,y+8}$	$N_{a+,y+8}$

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

Side step

Note on parameters

What is Na+5,y+8?

$$
\begin{aligned}
N_{a+5,y+8} &= N_{a+4,y+7}e^{-(s_{a+4}F_{y+7}+M)} \\
&= N_{a+3,y+6}e^{-(s_{a+3}F_{y+6}+M)}e^{-(s_{a+4}F_{y+7}+M)} \\
&= N_{a+2,y+5}e^{-(s_{a+2}F_{y+5}+M)}e^{-(s_{a+3}F_{y+6}+M)}e^{-(s_{a+4}F_{y+7}+M)} \\
&= N_{a+1,y+4}e^{-(s_{a+1}F_{y+4}+M)}e^{-(s_{a+2}F_{y+5}+M)}e^{-(s_{a+3}F_{y+6}+M)}e^{-(s_{a+4}F_{y+7}+M)} \\
&= N_{a,y+3}e^{-(s_{a}F_{y+3}+M)}e^{-(s_{a+1}F_{y+4}+M)}e^{-(s_{a+2}F_{y+5}+M)}e^{-(s_{a+3}F_{y+6}+M)}e^{-(s_{a+4}F_{y+7}+M)}
\end{aligned}
$$

Estimated parameters

I.e.: Stock size (Nay) for each year and age is only a function of recruitment and cumulative mortality.

Thus Na+5,y+8 is NOT a parameters. One could actually say that the model never "sees" this value, only the catches: Ca+5,y+8

Estimating catch C_{ay}-hat

- Once the population matrix is calculated it is simple to calculate the predicted catch (C_{ay}-hat) according to the catch equation:

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

  - The C-hats are values that we will later "confront" with the measurements that we have.

The map

einar

9. Solver

1. Parameters

Each box is labelled: Year (down the side, | | V) and Age -----> (across the top).

2. Selection pattern	3. Fishing mortality	Natural mortality
4a. Population numbers	4b. Nay at survey 1 time	4c. Nay at survey 2 time
5. Predicted catch **Calculate Cay-hat**	6a. Predicted survey 1 indices	6b. Predicted survey 2 indices
Observed catch at age	Observed survey 1 indices	Observed survey 2 indices
7. Catch residuals	8a. Survey 1 residuals	8b. Survey 2 residuals
7. Catch residuals squared	8a. Survey 1 residuals squared	8b. Survey 2 residuals squared

Color code
- Population model
- Observation mc
- Measurements
- Objectives
- Parameters

Confronting the model with data

- Until now we have only set up equations that follow the progression of each year class and calculated catch.
  - This is more or less what a kind of a the population simulator.
    - If we let recruitment be a function of **SSB** and we add some noise to recruitment we have a closed system and thus almost a medium/long term simulator
  - This is also more or less the same thing as we do when we do a short term projection.
  - Or for that matter in a yield per recruit analysis, except that there we focus only on one cohort (here one diagonal line).
- At present we are only interested in fitting the model to observations (measurements). Need thus some kind of objective function.

The objective function in words

- Find the value of the parameters:
  - fishing pattern by age (controlled by σ_L, σ_R and A_{full})
  - yearly fishing mortality (F_y)
  - population number in the first year ($N_{a,1}$)
  - recruitment ($N_{1,y}$) in each year
- that minimize the squared deviation of estimated catch (C_{ay}-hat) and measured catch (C_{ay}).

$$\min SSE_C = \sum_y \sum_a \left(\ln C_{ay} - \ln \hat{C}_{ay} \right)^2$$

  - Note that here we assume a log-normal error distribution. Could easily be replace with other type of error structure.

The map

einar

Different weights by age

- The catch of different age groups are often measured with different accuracy. Thus often set different weights to the residuals, so that the information from age groups that are measured with the most accuracy weigh more in the objective function:

$$\min SSE_C = \sum_y \sum_a \omega_a \left(\ln C_{ay} - \ln \hat{C}_{ay} \right)^2$$

More on ω later: it is inversely related to the variance

If we only have C_{ay}

- If there are no other available data for a stock than catch at age one could attempt to fit the model to catches alone.
- May need a extra “stabilizer”: The brave one may assume that fishing mortality does not change much between consecutive years:

$$
\begin{aligned}
{}_{\min} SSE &= SSE_C + SSE_F \\
&= SSE_C + \sum_y \left(\ln F_y - \ln F_{y-1}\right)^2
\end{aligned}
$$

Use with extreme caution!

Tuning with survey indices

- If additional information are available it is relatively easy to add them to the model. If age-based survey indices are available one may use:

$$\hat{U}_{ay} = q_a N_{ay}$$

- where q_a is a parameter (catchability). The minimization is by (again assuming log-normal errors):

$$\min SSE_U = \sum_y \sum_a \rho_a \left(\ln U_{ay} - \ln \hat{U}_{ay}\right)^2$$

Population numbers at survey time

- If survey time is not in the beginning of the year we need to take that into account by:

$$N'_{ay} = N_{ay} e^{-p(Fay+May)}$$

$$\hat{U}_{ay} = q_a N'_{ay}$$

- Where
  - N' is the population size at survey time
  - p is the fraction of the year when the survey takes place.

The map

einar

Objective function I

- Simple to combine the two objective functions:

$$\min SSE = SSE_C + SSE_U$$
$$= \sum_y \sum_a \omega_a \left(\ln C_{ay} - \ln \hat{C}_{ay} \right)^2$$
$$+ \sum_y \sum_a \rho_a \left(\ln U_{ay} - \ln \hat{U}_{ay} \right)^2$$

Lets not worry about weighing for now

Objective function II

- Often put different weights to the data:

$$\min SSE = SSE_C + SSE_U$$

$$= \lambda_C \sum_y \sum_a \omega_a \left(\ln C_{ay} - \ln \hat{C}_{ay} \right)^2$$

$$+ \lambda_U \sum_y \sum_a \rho_a \left(\ln U_{ay} - \ln \hat{U}_{ay} \right)^2$$

- λ weights are most often set externally
- At minimum should check the sensitivity to λ

Getting it all together

einar

THE MINIMIZATION STUFF

Sum of squares		Lambda
C@A	48.419	1
U@A - Survey 1	53.900	1
U@A - Survey 2	49.084	1
SSE total	151.4023	

PARAMETERS

Name	Ln(parameter)	Switches	Parameter
Ln Afull	2.3900		10.91
Ln σL	1.4481		4.26
Ln σR	5.0000		148.41

The heart of the setup lies in the left side of the spread-sheet. There we have the the objective functions, Penalties and weighing factors and parameters in one place. The only thing left is to setup the solver such that it minimizes the total SSE by changing the para-meters.

A trick to get things working

PARAMETERS

Name	Ln(parameter)	Switches	Parameter
Ln Afull	2.3900		10.91
Ln σL	1.4481		4.26
Ln σR	5.0000		148.41
LnF 0	-1.6871		0.19
LnF 1	-2.3877		0.09
……	-1.9916		0.14
LnN 1,0	6.9367		1029.39
LnN 2,0	6.2462		516.05
….			
LnN 0,1	7.4796		1771.57
LnN 0,1	7.4774		1767.62
..	5.7345		
Ln q 0	-4.0738		0.0170
Ln q 1	-2.6097	→	=EXP(D17)
…..			

What Solver "sees"

Values that are used in the model coding

Estimate logarithms for most parameters because:
1) Precludes searching nonsensical negative space
2) In logarithmic form parameters become scaled.

Leave space for switches, used later on when modify the model

Setup of xModel

- The following steps may be followed in order to setup the model in Excel:
  - 1. set up some sensible initial values for the parameters.
  - 2. calculate selectivity (s_a)
  - 3. calculate fishing mortality (F_{ay})
  - 4. calculate abundance (N_{ay})
  - 5. calculate estimated catch (C_{ay}-hat)
  - 6. calculate estimated survey indices (U_{ay}-hat)
  - 7. calculate residuals in catch $(\ln C_{ay}-\ln C_{ay}\text{-hat})^2$
  - 8. calculate residuals in survey $(\ln U_{ay}-\ln U_{ay}\text{-hat})^2$
  - 9. use solver to minimize 7. and 8.
- Population model: 1-4
- Observation model: 5-6
- Objective function: 7-9

The approach

- I hear - I forget
- I see - I remember
- I do - I understand

Lets start the work .....

Excel translation of Shakespeare!

=OR(B2,NOT(B2))