

Advanced configurations

Olav N.R. Breivik

Thanks to Anders Nielsen for borrowing course material



Norsk Regnesentral
ANVENDT DATAFORSKNING

Advanced options

- All of the basic options are things we need to consider
- Some of the advanced options address issues that we mostly don't have to consider
- But the dividing line is not absolutely clear

Basic configurations

We will now elaborate configurations for:

- Density dependent catchability
- Catch scaling
- Biomass and related indices
- Log normal or additive logistic normal
- Brekas in constant recruitment
- Using robust distribution for observations
- Using robust distribution for process increments
- Using model for biological parameters
- Fixing model parameters to a value

Density dependent catchability

We assume the stock equations:

$$\log N_{1,y} = \log R(\mathbf{N}_{y-1}) + \eta_{1,y}$$

$$\log N_{a,y} = \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y}$$

$$\log N_{A,y} = \log(N_{A-1,y-1} e^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1} - M_{A,y-1}}) + \eta_{A,y}$$

were

$$\log \mathbf{F}_y = \log \mathbf{F}_{y-1} + \boldsymbol{\xi}_y.$$

Observe:

$$\log C_{a,y} \sim N(\mu_{C_{a,y}}, \sigma_{C,a}^2)$$

$$\log I_{a,y}^{(s)} \sim N(\log(q_a^{(s)} \tilde{N}_{a,y}^{p_a^{(s)}}), \sigma_{s,a}^2)$$

```
1 $keyQpow
2 # Density dependent catchability power parameters (if any).
3 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
4 0 0 0 0 0 0 -1 -1 -1 -1 -1
5 1 1 1 1 1 2 2 -1 -1 -1 -1
```

Density dependent catchability

- Normally survey observations are predicted by:

$$E(\log I_{a,y}^{(s)}) = \log(q_a^{(s)} \tilde{N}_{a,y})$$

where $\tilde{N}_{a,y}$ is $N_{a,y}$ calculated at time of survey

- In situations where we think the survey density dependent we can estimate a power on N , so the relationship becomes:

$$E(\log I_{a,y}^{(s)}) = \log(q_a^{(s)} \tilde{N}_{a,y}^p)$$

- The option can be configured by:

```
1 $keyQpow
2 # Density dependent catchability power parameters (if any).
3 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
4 0 0 0 0 0 0 -1 -1 -1 -1
5 1 1 1 1 1 2 2 -1 -1 -1
```

Catch scaling

- If catches in some selected years are not representative we can estimate the mismatch by:

```
1 $noScaledYears
2 # Number of years where catch scaling is applied.
3 5
4
5 $keyScaledYears
6 # A vector of the years where catch scaling is applied.
7 2001 2002 2003 2004 2005
8
9 $keyParScaledYA
10 # A matrix specifying the couplings of scale parameters (nrow = no scaled years, ncols =
    no ages) .
11 0 0 0 0 0 0
12 1 1 1 1 1 1
13 2 2 2 2 2 2
14 3 3 3 3 3 3
15 4 4 4 4 4 4
```

- Then the corresponding catches $C_{a,y}$ are predicted by the normal catch equation prediction $\hat{C}_{a,y}$ divided by the estimated catch scaling $S_{a,y}$, so:

$$E(\log(C_{a,y})) = \log(\hat{C}_{a,y}/S_{a,y})$$

- To be identifiable it requires a period with surveys and unbiased catches
- Option considered based on knowledge of fishery

Treatment of biomass and related indices

- If a survey fleet is read in with one column, where the age is set to -1 , then it is a code to SAM that it should be treated as an not related to a specific age (e.g. a biomass index)
- Then this option is used to specify the type e.g. as below where the third fleet is an index of SSB:

```
1 $keyBiomassTreat
2 # To be defined only if a biomass survey is used (0 SSB index, 1 catch index, 2 FSB
   index, 3 total catch,
3 # 4 total landings, 5 TSB index, 6 TSN index, and 10 Fbar idx).
4 -1 -1 0 -1 -1
```

- The different options are described above
- If an option is called an index, then the corresponding `$keyLogFpar` and `$keyVarObs` first column needs to be configured also (here element (3,1))
- If the option is not called an index (option 3 total catch and 4 total landings), then those fields should be undefined -1 (because numbers are assumed absolute)
- As example: an SSB index is modelled as:
$$\log(\widehat{\text{SSB}_y}) \sim \mathcal{N}(\log(q^{(s)} \widehat{\text{SSB}_y}), (\sigma^{(s)})^2)$$

Log normal or additive logistic normal

- Observation vectors are normally assumed to follow log-normal distributions
- It is possible to switch to additive logistic for compositions and log-normal for total

```
1 $obsLikelihoodFlag
2 # Option for observational likelihood | Possible values are: "LN" "ALN"
3 "LN" "LN" "LN"
```

- The ALN option is not used often in SAM (see next slide)

Table 1: Overview of the observational models used in the case studies and some properties: if zero observations are allowed; whether the Baranov catch equation determines the mean, median or location; the number of estimated observational parameters per age (a) and fleet (f); and whether a correlation parameter is estimated. The models are divided in to model classes: Univariate numbers-at-age (UN@A), multivariate numbers-at-age (MN@A), proportions-at-age with log-normal total numbers (P@AwN), and proportions-at-age with log-normal total weight (P@AwW).

Model	Distribution	Class	Allows 0	Baranov	Est. par.s	Est. cor.
M_1	log-Normal	UN@A	No	Median	1 a f^1	No
M_2	Gamma	UN@A	Some	Mean	1 a f	No
M_3	Generalized Gamma	UN@A	Some	Location	2 a f	No
M_4	Normal	UN@A	Yes	Mean	1 a f	No
M_5	Left Truncated Normal	UN@A	Yes	Location	1 a f	No
M_6	log-Student's t	UN@A	No	Location	2 a f	No
M_7	Multivariate log-Normal	MN@A	No	Median	1 a $f+1$ f^2	Yes
M_8	Additive Logistic Normal	P@AwN	No	Location	1 a $f+1$ f	Yes
M_9	Multiplicative Logistic Normal	P@AwN	No	Location	1 a $f+1$ f	Yes
M_{10}	Dirichlet	P@AwN	No	Mean	1 f	No
M_{11}	Additive Logisitic Normal	P@AwW	No	Location	1 a $f+1$ f	Yes
M_{12}	Multiplicative Logistic Normal	P@AwW	No	Location	1 a $f+1$ f	Yes
M_{13}	Dirichlet	P@AwW	No	Mean	1 f	No

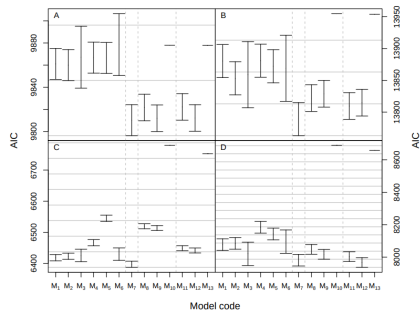


Figure 3: AIC intervals for models M_1 to M_{13} (Table 1) in the case studies: Blue Whiting (A), North-East Arctic Haddock (B), North Sea Cod (C), and Northern Shelf Haddock (D). The horizontal grey lines indicate AIC differences of 50 starting at the lowest lower bound of the models. Vertical dashed grey lines separates the models in model classes (Table 1).

● From paper:

Choosing the observational likelihood in state-space stock assessment models CM Albertsen, A Nielsen, UH Thygesen - *Canadian Journal of Fisheries and Aquatic Sciences*, 2016

Breaks in constant recruitment or spline nodes

- If `$stockRecruitmentModelCode` is set to 3 (a constant level of recruitment), several domains can be defined

```
1 $constRecBreaks
2 # For stock-recruitment code 3: Vector of break years between which recruitment is at
   constant level. The break year is included in the left interval.
3 # For spline stock-recruitment: Vector of log-ssb knots. (This option is only used in
   combination with stock-recruitment code 3, 90-92, and 290)
4 1983 1997
```

- This configuration would estimate 3 mean levels of recruitment one before and including 1983, one from 1984-1997, and one from 1998 and forward.
- This option can be considered if different recruitment regimes are considered plausible.
- Results from other recruitment models (e.g. random walk) or recruitment residuals may also indicate this type of model
- (The spline options are pretty experimental, but the node-years define their flexibility)

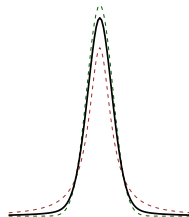
Using robust distribution for observations

- The normal distribution is fairly sensitive to outliers
- Standard model uses pdf: $\phi\left(\frac{x-\mu}{\sigma}\right) \frac{1}{\sigma}$
- Robust model will use pdf:

$$\left((1-p)\phi\left(\frac{x-\mu}{\sigma}\right) + p\psi\left(\frac{x-\mu}{\sigma}\right) \right) \frac{1}{\sigma}$$

, where the p is an input (not estimated)

- Where ϕ is pdf of $N(0,1)$ and ψ is pdf of heavy-tailed distribution (t_3).



```
1 $fracMixObs
2 # A vector with same length as number of fleets, where each element is the
3 # fraction of t(3) distribution used in the distribution of that fleet
4 0.05 0 0
```

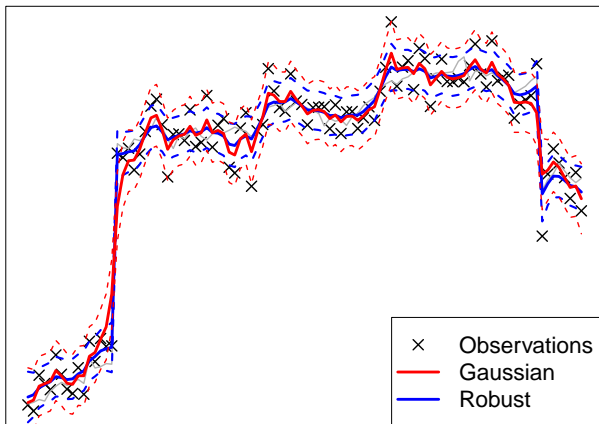
- In the configuration above a mixture distribution is used for the catches
- Can be considered if a few unexplained outliers are in the dataset
- **Joint class exercise:** Try introducing an outlier in an assessment and compare with and without robustifying

Using robust distributions for process increments

- Robust distributions can also be used for increments in N , and F processes

```
1 $fracMixN
2 # The fraction of t(3) distribution used in logF increment distribution
3 0
4
5 $fracMixF
6 # The fraction of t(3) distribution used in logN increment distribution (for each age
   group)
7 0
```

- Works as for observations, but for increments (allows big jumps)
- Can be problematic w.r.t. convergence
- To be considered experimental



Time

- There is no option for time-defined breakpoints in observation standard-deviations
- but the following can be (mis-)used for that

```
1 $keyXtraSd
2 # An integer matrix with 4 columns (fleet year age coupling), which allows additional
   uncertainty to be estimated for the specified observations
3 1  2021  1  0
4 1  2021  2  0
5 1  2021  3  0
6 1  2021  4  0
7 1  2021  5  0
```

- Extra variance estimated for catches ages 1-5 in 2021
- Very flexible, but on the other hand configuration requires a lot of lines...
- To be considered if we have reason to think some observations are less reliable than others based on observation time
- Residual plot can also inform us
- (when setting this up the function `expand.grid` in R can be helpful)

Using model for biological parameters

- See presentation `bioparupdate.pdf`

```
1 $stockWeightModel
2 # Integer code describing the treatment of stock weights in the model (0 use as known,
3 # 1 use as observations to inform stock weight process (GMRF with cohort and within year
   correlations), 2 to add extra correlation to plusgroup)
4 1
5 $keyStockWeightMean
6 # Coupling of stock-weight process mean parameters (not used if stockWeightModel==0)
7 0 1 2 3 4 5 6 7 8 9 10
8 $keyStockWeightObsVar
9 # Coupling of stock-weight observation variance parameters (not used if stockWeightModel
   ==0)
10 0 0 0 0 0 0 0 0 0 0 0
11
12 $matureModel
13 # Integer code describing the treatment of proportion mature in the model (0 use as
   known, 1 use as observations to inform proportion mature process (GMRF with cohort
   and within year correlations on logit(proportion mature)), 2 to add extra
   correlation to plusgroup)
14 1
15 $keyMatureMean
16 # Coupling of mature process mean parameters (not used if matureModel==0)
17 0 1 2 3 4 5 6 7 8 9 10
```

- Used to fill in missing biological parameter observations
- Used to be able to predict biological parameters
- Used to filter out observation noise and propagate uncertainties

Fixing model parameters to a value

- No built in option for fixing parameters (should there be?)
- There is however a very primitive way:

```
1 library(stockassessment)
2 ## ... all the usual stuff
3 par$logSdLogN[2] <- -3                # set value you want to fix at
4 fixmap <- list(logSdLogN=factor(c(1,NA))) # map must be as long as parameter
5 # NA where fixed unique values elsewhere
6 fit <- sam.fit(dat,conf,par, map=fixmap)
```

- Primitive, but works (demonstrate)